Controlled X-chromosome dynamics defines meiotic potential of female mouse in vitro germ cells

Jacqueline Severino¹,‡, Moritz Bauer⁴,†, Tom Mattimoe¹,†, Niccolò Arecco¹,†, Luca Cozzuto¹,†, Patricia Lorden²,†, Norio Hamada³, Yoshiaki Nosaka⁴,⁵,†, So I Nagaoka⁴,⁵,†, Pauline Audergon¹,†, Antonio Tarruell¹,†, Holger Heyn²,⁷,†, Katsuhiko Hayashi⁸,†, Mitinori Saitou⁴,⁵,⁶,† & Bernhard Payer¹,⁷,*

Abstract

The mammalian germline is characterized by extensive epigenetic reprogramming during its development into functional eggs and sperm. Specifically, the epigenome requires resetting before parental marks can be established and transmitted to the next generation. In the female germline, X-chromosome inactivation and reactivation are among the most prominent epigenetic reprogramming events, yet very little is known about their kinetics and biological function. Here, we investigate X-inactivation and reactivation dynamics using a tailor-made in vitro system of primordial germ cell-like cell (PGCLC) differentiation from mouse embryonic stem cells. We find that X-inactivation in PGCLCs in vitro and in germ cell-competent epiblast cells in vivo is moderate compared to somatic cells, and frequently characterized by escaping genes. X-inactivation is followed by step-wise X-reactivation, which is mostly completed during meiotic prophase I. Furthermore, we find that PGCLCs which fail to undergo X-inactivation or reactivate too rapidly display impaired meiotic potential. Thus, our data reveal fine-tuned X-chromosome remodelling as a critical feature of female germ cell development towards meiosis and oogenesis.

Keywords  epigenetic reprogramming; in vitro model; meiosis; primordial germ cells; X-chromosome inactivation

Subject Categories  Chromatin, Transcription & Genomics; Development; Stem Cells & Regenerative Medicine

Introduction

The germ cell lineage is unique in its critical function to transmit genetic and epigenetic information from one generation to the next. In mice, primordial germ cells (PGCs), the precursors of eggs and sperm, are specified during early postimplantation development from somatic precursors in the proximal epiblast by inductive signals (Lawson et al., 1999; Ohinata et al., 2005, 2009). Thereafter, PGCs migrate and enter the future gonads where they receive sex-specific somatic signals, which determine the germ cell sex and promote differentiation towards a spermatogenic or oogenic fate (Miyauchi et al., 2017; Spiller et al., 2017). While in males, germ cells enter mitotic arrest and differentiate into prospermatogonia, in females, germ cells instead progress into meiosis and oogenesis.

A hallmark feature of early germ cell development is the extensive epigenetic reprogramming (Kurimoto & Saitou, 2019), characterized by global changes in histone marks (Seki et al., 2005; Hajkova et al., 2008), DNA demethylation and erasure of genomic imprints (Hajkova et al., 2002; Seisenberger et al., 2012; Shirane et al., 2016). This establishes an epigenetic naive state (Ohata et al., 2017), which is required in order for PGCs to progress towards gonadal germ cell fate (Hill et al., 2018) and to control their timing to enter female meiosis (Yokobayashi et al., 2013). Ultimately, this erasure of parental information allows the reestablishment of new maternal and paternal marks during spermatogenesis and oogenesis, respectively, which are critical for the competence of egg and sperm to facilitate embryonic development in the next generation (Reik & Surani, 2015; Ohta et al., 2017).
In addition to these global changes, another important epigenetic reprogramming event takes place in the female germline; the reversal of silencing of the inactive X chromosome by X-chromosome reactivation. While X-chromosome inactivation (Lyon, 1961; Payer & Lee, 2008; Galupa & Heard, 2018) is the process by which female mammals (XX) achieve X-linked gene dosage parity with males (XY), X-reactivation takes place specifically in pluripotent epiblast cells of the mouse blastocyst (Mak et al., 2004; Borensztein et al., 2017) and in PGCs during their migration and upon their entry into the gonads (Sugimoto & Abe, 2007; Chuva de Sousa Lopes et al., 2008). Therefore, while X-inactivation is associated with pluripotency exit and the differentiated state (Schulz et al., 2014), X-reactivation is a key feature of naïve pluripotency and germ cell development (Pasque et al., 2014; Payer, 2016; Janiszewski et al., 2019; Panda et al., 2020; Bauer et al., 2021; Talon et al., 2021).

X-reactivation in mouse PGCs is a multistep process, which initiates during PGC migration with downregulation of Xist, the long non-coding master regulator RNA of X-inactivation and concomitant loss of the associated histone H3K27me3 mark from the inactive X (Sugimoto & Abe, 2007; Chuva de Sousa Lopes et al., 2008). This process is regulated by repression of the Xist gene by the germ cell transcription factor PRDM14 (Payer et al., 2013; Mallol et al., 2019) and potentially by other members of the pluripotency network such as NANOG or OCT4 (Navarro et al., 2008), which are all expressed during PGC development. Subsequently, X-linked genes become progressively reactivated during migration, with the process being completed after PGCs have reached the gonads, and following the initiation of oogenesis and meiosis (Sugimoto & Abe, 2007; Sangrithi et al., 2017). X-linked gene reactivation is thereby thought to be enhanced by a female-specific signal from gonadal somatic cells (Chuva de Sousa Lopes et al., 2008). Although the molecular nature of the X-reactivation-promoting signal is currently unknown, the timing of X-linked gene reactivation around meiotic entry and the dependency of both processes on a female somatic signal suggest a potential mechanistic link. Until now it has not been formally tested, if, and to which degree, the X-inactivation status might impact the meiotic and oogenic potential of germ cells. Furthermore, previous studies on the X-inactivation and -reactivation dynamics during mouse germ cell development have been limited to few individual genes (Sugimoto & Abe, 2007) or have not been allelically resolved and therefore been unable to discriminate between transcripts expressed from either one or two X chromosomes (Sangrithi et al., 2017). Therefore, a comprehensive analysis of X-inactivation and -reactivation kinetics and its functional relation to germ cell developmental progression is necessary to gain mechanistic insight.

Based on in vitro germ cell differentiation from mouse embryonic stem cells (ESCs) (Hayashi et al., 2011, 2012; Nakaki et al., 2013), we developed an X-chromosome reporter system (XRep) to study the kinetics of X-inactivation and -reactivation during germ cell development. We thereby provide a high-resolution allelic analysis of X-chromosome dynamics and discovered that germ cells with high meiotic and oogenic competence are characterized by a moderate degree of X-inactivation and gradual X-reactivation kinetics. In contrast, germ cells that failed to undergo X-inactivation or which reactivated the X chromosome too rapidly displayed abnormal gene expression and differentiation characteristics. Thus, we found first evidence that a controlled sequence of X-inactivation followed by X-reactivation to be a characteristic hallmark of normal female germ cells. This suggests that both dosage control and epigenetic reprogramming of the X chromosome may be critical indicators for female germ cells’ developmental potential to progress towards meiosis and oogenesis.

**Results**

XRep, a tailor-made system for tracing X-chromosome dynamics during in vitro germ cell development

In order to achieve a better understanding of the X-chromosome dynamics during mouse germ cell development, we created a tailor-made in vitro model system called XRep (Fig 1A). XRep combines the following features. First, it is based on a hybrid female embryonic stem cell (ESC) line containing one *Mus musculus* (*Xmus*) and one *Mus castaneus* (*Xcas*) X chromosome (Lee & Lu, 1999; Ogawa et al., 2008), allowing allele-specific determination of gene expression. Moreover, this line was shown to be karyotypically highly stable (Lee & Lu, 1999; Bauer et al., 2021), therefore preventing X-loss, a crucial prerequisite for X-inactivation and -reactivation studies. Additionally, the cell line contains a Tsix truncation (TST) on *Xmus*, forcing non-random X-inactivation of the *Xmus* upon cell differentiation (Luikenhuis et al., 2001; Ogawa et al., 2008). This enabled us to study the X-inactivation and -reactivation dynamics specifically of the *Xmus* chromosome, while the *Xcas* would remain constitutively active. Second, primordial germ cell-like cells (PGCLCs) can be traced by dual X-linked reporter genes placed in the *Hprt* locus (Wu et al., 2014), a GFP reporter on *Xmus* (XGFP) and a tdTomato reporter on *Xcas* (*XTomato*). This allows us to isolate distinct populations of cells, harbouring either two active X chromosomes (XGFP+/XTomato+) or one inactive and one active X (XGFP−/XTomato+), using fluorescence-activated cell sorting (FACS). This gives us a unique advantage over in vivo studies, as it enables us to test the importance of X-inactivation and -reactivation for germ cell development by isolating and further culturing cells of different X-inactivation states. Taken together, this tailor-made system allows us to assess X-chromosome dynamics and its importance for female mouse germ cell development in vitro.

We first set out to assess competence for PGCLC differentiation of our XRep cell line. We slightly adapted published protocols (Hayashi & Saitou, 2013; Nakaki et al., 2013), by differentiating ESCs into epiblast-like cells (EpiLCS) for 4 days, as differentiation for 2 days, as demonstrated in said previous studies, did not yield PGCLCs with our XRep cells likely due to their specific genetic background (Fig EV1A). Furthermore, we extended the induction time of PGCLC generation from 4 to 5 days to ensure sufficient time to undergo X-inactivation (Fig 1B). We quantified PGCLC induction efficiency by FACS analysis, using SSEA1 and CD61 double-positive staining to mark successfully induced PGCLCs (Fig 1C). At PGCLC day 5, we found ~ 60% of the cell population to be double-positive for SSEA1/CD61, indicating a very high PGCLC induction efficiency when compared to the cytokine-based protocol (Hayashi & Saitou,
and in line with previous observations on transcription factor-based PGCLC induction (Nakaki et al., 2013). To further assess the quality of our PGCLCs, we stained cryosections of PGCLC bodies at day 5 of induction for SOX2 and TFAP2C, both germ-line expressed transcription factors. We observed that > 50% of cells were double-positive for SOX2 and TFAP2C (Fig 1D), confirming PGCLC cell identity. We next wanted to assess X-inactivation kinetics using our XGFP and XTomato reporters. As expected, XTomato stayed active
throughout the differentiation (Fig 1E and F). In contrast, we observed downregulation of the XGFP reporter at day 2 of PGCLC differentiation, with the XGFP− population gradually increasing until day 5 (Fig 1E and F). Nevertheless, even at day 5, up to 40% of PGCLCs remained XGFP+ in our system (Fig EV1B). Despite this, the large majority of EpilCs showed H3K27me3 foci (Fig EV1D–F), which suggests that both XGFP− and XGFP+ PGCLCs originated from EpilCs that had initiated X-inactivation. We do, however, note that EpilCs still retained XGFP protein staining (Figs 1E and F, and EV1D and F), while XGFP transcripts were being downregulated (Fig EV1C), indicating that protein stability of GFP gives a delayed read-out of X-inactivation kinetics. Nevertheless, at the PGCLC stage, the XGFP signal faithfully reflected the X-inactivation state, as XGFP was only detected in cells without the X-inactivation-specific H3K27me3 spot (Fig EV1D–F).

In summary, using our tailor-made XRep cell line, we could show that X-inactivation initiates early during PGCLC differentiation. Additionally, our system enables the isolation of distinct PGCLC populations, either having undergone X-inactivation or harbouring two active X chromosomes, suggesting that PGCLC specification can occur in the absence of X-inactivation as well.

**XGFP+ and XGFP− PGCLCs define distinct subpopulations**

Having identified two distinct PGCLC populations, we set out to characterize the transcriptional changes taking place during differentiation. We induced EpilCs from ESCs for 4 days and subsequently induced PGCLCs for 5 days, at which stage we isolated XGFP+ and XGFP− PGCLCs by FACS (Fig EV2A). With these samples, we performed allele-specific RNA-sequencing on two biological replicates (different clones) with two technical replicates each. Principal component analysis (PCA) of the expression profiles showed a high coherence between replicates, with ESCs, EpilCs and PGCLCs occupying distinct clusters (Fig 2A). Moreover, we observed that XGFP+ and XGFP− PGCLCs clustered separately, indicating distinct expression profiles of the two populations. To exclude the possibility that the distinct clustering of PGCLC populations was influenced by the different X-status of the two, we repeated the PCA while eliminating X-chromosome-linked genes from the analysis. We observed a highly similar clustering of samples with minimal changes in component variances (Fig EV2B). In order to assess whether transcriptional differences in XGFP+ and XGFP− PGCLCs could be explained by differences in developmental timing, we took advantage of published datasets of female in vivo PGCs from E9.5, E10.5, E11.5 and E12.5 embryos (Nagaoka et al, 2020) and compared expression profiles to our in vitro derived PGCLCs (Fig 2B). PCA revealed a trajectory where PC1 defined the developmental timing of in vivo and in vitro samples, whereas PC2 did not greatly contribute to the separation of our in vitro PGCLCs (PC loadings in Dataset EV4). We found that both PGCLC populations clustered around E10.5, with XGFP+ cells corresponding to a slightly advanced developmental stage. Therefore, as XGFP+ and XGFP− PGCLCs seemed to correspond to a similar developmental time point, we wanted to characterize their transcriptional differences in more detail. We performed differential gene expression analysis and could identify 2,684 upregulated and 2,437 downregulated genes in XGFP− PGCLCs, when compared to XGFP+ PGCLCs (Figs 2C–E and EV2C and D). Among the genes significantly upregulated in XGFP− PGCLCs, we found early germ cell genes including Blimp1 (Prdm1), Prdm14 and Tdap2c (Figs 2D and EV2C). In contrast, in XGFP+ PGCLCs, we observed higher expression of pluripotency genes such as Esrtrb and Zfp42 and a subset of late germ cell genes like Dazl.

Moreover, when we performed functional annotation by gene ontology (GO) term analysis, we observed enrichment for genes involved in urogenital system development, MAPK regulation and WNT signalling in XGFP− PGCLCs, while genes upregulated in XGFP+ PGCLCs were enriched for DNA methylation involved in gamete generation, meiotic cell cycle and response to LIF signalling (Figs 2F and EV2D). MAPK signalling is known to be inhibited by double X-dosage (Schulz et al, 2014; Song et al, 2019; Genolet et al, 2021), which might explain enrichment of this pathway in our XGFP− PGCLCs. LIF signalling on the other hand, which is enriched in our XGFP+ PGCLCs, is known to enable expression of the naive pluripotency network, which repressiones Xist, thereby promoting the active X state (Payer & Lee, 2014; Panda et al, 2020). Furthermore, enrichment for meiotic cell cycle genes in XGFP+ PGCLCs such as Aurkc, Dazl and Puri1 (Fig EV2D), suggests a premature activation of a subset of meiotic genes in XGFP+ PGCLCs.

**Figure 2. Gene expression analysis reveals two PGCLC subpopulations.**

A. PCA of gene expression dynamics during PGCLC differentiation. Four biological replicates are shown. n = top 500 most variable genes. PGCLCs were sorted for SSEA1 and CD61 expression and further divided into XGFP+ and XGFP−. Axes indicate the variance. Arrows indicate hypothetical trajectory. Shapes indicate the biological clone (clone A11 = square, clone E9 = rhombus).

B. PCA of gene expression dynamics compared to in uivo samples from Nagaoka et al (2020). n = top 500 most variable genes, calculated including in uivo samples. Grey arrow indicates putative developmental trajectory. Shapes indicate the replicates (clone A11 = square, clone E9 = rhombus, n in uivo samples = circle).

C. MA plot of differential gene expression changes between XGFP− and XGFP+ PGCLCs as determined by RNA-seq. Log2-mean expression (log2-normalized counts from DESeq2) on the X-axis and the log2-fold change on the Y-axis are shown. Significantly upregulated and downregulated genes are highlighted in red and green respectively. False discovery rate (FDR) < 0.001. Non-significant genes with log2-mean expression between 0 and 0.2 were removed for easier plot visualization.

D. Expression levels (normalized DESeq2 counts) of selected differentially expressed genes between XGFP− and XGFP+ PGCLCs during the differentiation time course. Genes with FDR < 0.001 were considered significantly differentially expressed. Points indicate expression of individual biological replicates.

E. Heatmap of RNA-seq normalized counts showing the Z-score across PGCLC induction time points of 31 manually selected and manually ordered marker genes belonging to the categories reported on the side.

F. Selected GO terms enriched in XGFP− PGCLCs and XGFP+ PGCLCs.

G. FACS analysis of cell cycle using DAPI. Numbers indicate the percentage of cells in G1, S and G2/M respectively.

H. Alkaline phosphatase staining for ESCs, XGFP+ PGCLCs and XGFP− PGCLCs grown for 7 days in 2/LIF medium on immortalized mouse embryonic fibroblasts.

I. Barplot indicating the absolute numbers of alkaline phosphatase (AP)-positive colonies in each cell type after 7 days of culture in 2/LIF medium on immortalized mouse embryonic fibroblasts. Y-axis is in square root scale (sqrt) for better plot visualization. Each white dot represents one technical replicate.

Source data are available online for this figure.
One characteristic feature of PGCLCs is changes in cell cycle progression and proliferation upon differentiation (Ohta et al., 2017), both of which are known to be affected by MAPK, as well as LIF signalling pathways (Meloche & Pouysségur, 2007; Onishi & Zandstra, 2015). We therefore performed cell cycle analysis using DAPI and found that ESCs, EpiLCs and XGFP+ PGCLCs shared similar cell cycle distributions.

**Figure 2.**

- **A** shows a PCA plot of cell cycle distribution for ESCs, EpiLCs, and XGFP+ PGCLCs. The plot is divided into in vitro and in vivo conditions.
- **B** displays a heatmap of log2 fold change in gene expression between XGFP- PGCLCs and XGFP+ PGCLCs, with 2684 genes upregulated in XGFP- PGCLCs and 2437 upregulated in XGFP+ PGCLCs.
- **C** presents a volcano plot of log2 fold change in gene expression for PGCLC differential gene expression.
- **D** illustrates the expression levels of selected genes, including Prdm14, Tfap2c, Blimp1, Zfp42, Esrrb, and Dazl, showing differences in expression between ESCs, EpiLCs, and XGFP+ PGCLCs.
- **E** displays a heatmap of gene expression levels for ESCs, EpiLCs, XGFP+ PGCLCs, and XGFP- PGCLCs, with a focus on pluripotency and germ cell markers.
- **F** shows selected GO terms enriched in XGFP- PGCLCs, including urogenital system development, WNT signaling, MAPK regulation, and response to LIF.
- **G** presents a heatmap of gene expression levels for ESCs, EpiLCs, XGFP+ PGCLCs, and XGFP- PGCLCs, with a focus on gene expression levels for ESCs, EpiLCs, XGFP+ PGCLCs, and XGFP- PGCLCs.
- **H** shows a bar graph of AP+ colony counting for ESCs, XGFP+ PGCLCs, and XGFP- PGCLCs.
highly similar profiles, with the majority of cells (>60%) residing in S phase. In contrast, XGFP– PGCLCs showed a decreased number of cells in S phase, concomitant with an increase in cells in G1, suggesting a slower proliferation of this population (Fig 2G).

As our transcriptomics and cell cycle analysis suggested that XGFP+ PGCLCs could correspond to an aberrant PGCLC state with properties related to ESCs, we set out to address if this would also lead to an advantage in growth and survival under physiological conditions favouring ground-state pluripotent stem cells. We therefore isolated XGFP+ and XGFP– PGCLCs at day 5 and seeded them (1,000 cells per six-well) on irradiated mouse embryonic fibroblasts in 2i/LIF medium (Fig 2H and I), which previously has been shown to allow the establishment of pluripotent embryonic germ cell (EGC) lines from in vivo mouse PGCs (Leitch et al, 2010). When we then compared EGC colony formation capacity, we found that while almost no colonies (2 from 1,000 seeded cells) originated from XGFP– PGCLCs, we observed a substantially higher number of colonies (n = 84) from XGFP+ PGCLCs, albeit still fewer than when re-plating ESCs (633 colonies). Importantly, both ESCs and XGFP+ PGCLCs retained two active X chromosomes, while only a subset of XGFP– PGCLCs had undergone X-reactivation during EGC colony formation (Fig EV2E and F).

In summary, RNA expression analysis of XGFP+ and XGFP– PGCLCs showed a PGC-like transcriptome of both populations, further suggesting that X-inactivation and PGCLC induction can be uncoupled in our system. However, we observed that XGFP+ PGCLCs displayed higher expression of several naive pluripotency genes as well as premature expression of a subset of mitotic genes and a rapid cell cycle. Moreover, considering their higher ability to form EGC colonies under ground-state pluripotency conditions, this suggests that XGFP+ PGCLCs may correspond to an aberrant PGCLC state with pluripotent stem cell-related features. This indicates that X-inactivation could be necessary for correct PGCLC maturation, or alternatively, be a sensitive biomarker of high-quality PGCLCs.

**Heterogeneous and moderate X-inactivation is a feature of germ cell fate in vitro and in vivo**

To this point, due to the lack of an allele-specific transcriptomic analysis, the X-inactivation and -reactivation dynamics during mouse PGC development in vivo and in vitro have not been assessed on a chromosome-wide level. Therefore, to determine X-chromosome-wide gene inactivation kinetics during PGCLC differentiation, we assessed the allelic expression ratio between the inactive X\(^{inu}\) and the active X\(^{ac}\). We performed PCA of the allelic ratio of our samples in addition to neural progenitor cells (NPCs) from the same parental clone (Bauer et al, 2021) to include a cell type shown to have undergone complete X-inactivation (Fig 3A). We observed that the PCA of the PCA defined the degree of X-inactivation, separating samples with two active Xs on the left (ESCs and XGFP+ PGCLCs), and with one inactive X on the right (XGFP– PGCLCs and NPCs). Moreover, we noticed that EpiLCs were positioned at the centre, suggesting an intermediate degree of X-inactivation. We next determined X-inactivation kinetics, while focussing on genes biallelically expressed in ESCs (Fig EV3A) (allelic expression ratio > 0.3 and < 0.7) and established an X-inactivation cut-off of an allelic ratio of 0.135, according to the distribution of genes in NPCs and the local minimum (Fig EV3B), and similar to cut-offs used in previous studies (Peeters et al, 2014; Borensztein et al, 2017; Xu et al, 2017; Janiszewski et al, 2019; Bauer et al, 2021). As a control, we assessed the allelic expression ratio of the fully hybrid chromosome 13, which maintained biallelic expression throughout the time course (Fig EV3C). In contrast, we observed initiation of X-linked gene silencing in EpiLCs, progressing further in XGFP– PGCLCs, while XGFP+ PGCLCs showed biallelic expression, similar to ESCs (Fig 3B). To assess X-inactivation dynamics in more detail, we grouped X-linked genes according to their silencing kinetics (Fig 3C). We found 62 genes to have undergone X-inactivation (XCI) in EpiLCs (termed early XCI), and 138 genes to have undergone inactivation in PGCLCs (late XCI). Moreover, to our surprise, we observed a large number of genes (93) to still be active in XGFP– PGCLCs (escapees). In comparison, we observed 46 genes

![Figure 3. Characterization of X-inactivation dynamics during PGCLC induction.](image-url)
escaping X-inactivation in NPCs (Bauer et al., 2021), out of which 37 were also found to be escapees in PGCLCs (Fig 3D). While a certain degree of escape from X-inactivation is expected, the percentage of escapees we observed for PGCLCs here is at 32%, which is considerably higher than reported for other cell types (Fig EV3F) (Peeters et al., 2014; Marks et al., 2015; Balaton et al., 2021).

Figure 3.
Given these results, we wondered how this large degree of escape from X-inactivation might be explained. We assessed Xist expression levels and could observe high levels in EpiLCs, reaching levels comparable to those in NPCs (Fig 3E). However, expression levels in XGFP– PGCLCs were considerably decreased, which might be explained by the high expression of Prdm14 in PGCLCs, a known repressor of Xist (Payer et al., 2013). This is also in line with in vivo data (Sugimoto & Abe, 2007), where Xist has been shown to be completely downregulated in E10.5 PGCs of equivalent stage (Fig 2B). Furthermore, we wanted to know which features might distinguish escapees from inactivated genes in our system. We measured gene expression levels from the Xmut allele in ESCs and found escapees to be significantly higher expressed, while early inactivating genes, in contrast, showed the lowest expression levels (Fig EV3D). Similarly, expression of escapees from the Xmut allele was also elevated in XGFP– PGCLCs (Fig EV3E). When we compared our gene categories of early, late and escaping X-inactivation from XGFP– PGCLCs with published data of X-inactivation kinetics during embryoid body (EB) differentiation (Marks et al., 2015), the overlap of genes was relatively modest (Fig EV3F). This suggests that the timing of inactivation of individual X-linked genes varies between different cell types and differentiation systems.

Having observed that both their precursors, EpiLCs, as well as PGCLCs themselves show a distinct moderate X-inactivation with a high number of escapees, we wondered if our in vitro findings were reflecting the situation in vivo during mouse development. We therefore re-analysed publicly available allelically resolved single-cell RNA-seq data of E6.5 embryos (Cheng et al., 2019) which corresponds to the developmental time point that includes the in vivo equivalents of EpiLCs (Fig 3F and G; Han et al., 2010; Hayashi et al., 2011). We focussed on female cells from a CAST(F) X C57(M) cross, based clustering (Fig3G). This yielded six clusters, which according to applied modularity optimization-based clustering (Fig 3G). This yielded six clusters, which according to marker gene expression (Fig EV3G) corresponded to the extraembryonic cell types of anterior visceral endoderm (Amm+, Cer1+, Gata4+, Otx2+; cluster 1), posterior visceral endoderm (Amm+, Gata4+, Wnt5+; cluster 2) and parietal endoderm (Fst+; cluster 3) and the embryonic cell types of proximal epiblast (Dppa4+, Ifitm3+/ Fragilis+, Pou5f1+; cluster 4), transition epiblast (Otx2+, Pou5f1+; cluster 5) and posterior epiblast (Gata4+, Hand1+, Pou5f1+, T+, Wnt5+; cluster 6). To assess the degree of X-inactivation, we calculated the average allelic ratio per cell (Fig 3H), which, as expected, highlighted imprinted X-inactivation in the extraembryonic cell types; where the paternal C57 allele was always silenced, and showed random X-inactivation in the epiblast clusters. However, within the epiblast we noticed striking differences, with a considerably larger fraction of cells from the proximal epiblast—the Ifitm3/Fragilis+ region with competence for PGC specification (Lawson & Hage, 1994; Saitou et al., 2002)—not having undergone X-inactivation (Fig 3I), when compared with the other non-PGC-competent epiblast clusters. Specifically, 25% of proximal epiblast cells displayed an average allelic ratio between 0.4 and 0.6 indicating a lack of X-inactivation compared to < 10% for transition and posterior epiblast cells. Moreover, when we then compared the degree of gene silencing for cells with an average allelic ratio < 0.4 (Fig 3J), to allow comparison with imprinted XCI cells, we found that 56% of proximal epiblast genes escaped / displayed incomplete X-inactivation at that stage. This strongly contrasts to < 8% for all other cell types, including extraembryonic and other epiblast clusters (Fig 3K). This could partially be explained by the higher pluripotency factor expression in the proximal epiblast when compared with the posterior epiblast (Fig EV3H), which may contribute to the reduced degree of X-inactivation specifically in PGC-competent precursor cells.

Taken together, we find that EpiLCs and PGCLCs undergo a moderate degree of X-inactivation, characterized by a large percentage of escapees / incompletely silenced genes. Moreover, our analysis suggests that low expression of Xist in PGCLCs might lead to a failure of gene silencing of highly expressed genes, leading to a large percentage of escapees. Furthermore, our analysis of in vivo E6.5 data shows heterogeneity and a moderate degree of X-inactivation specifically in the PGC-competent cells of the proximal epiblast, highlighting how this distinct X-inactivation state is a conserved hallmark of germ line competence both in vitro and in vivo.

Single-cell RNA-seq analysis of meiotic entry of in vitro-derived germ cells reveals clusters of distinct developmental progression

After having established the degree of X-inactivation during PGCLC specification, we wanted to address the further developmental progression of PGCLCs depending on their X-chromosome status. Having identified and isolated distinct PGCLC types with either two active X-chromosomes (XGFP+ PGCLCs) or one active and one inactivated X-chromosome (XGFP– PGCLCs) (Fig 2), we were able to assess whether the X-inactivation status of PGCLCs had an impact on germ cell maturation. Furthermore, we sought to investigate to which degree X-reactivation and meiotic entry were intrinsically coupled processes.

To this end, we differentiated XGFP+ and XGFP– PGCLCs using an in vitro reconstituted ovary (rOvary) protocol (Hayashi & Saitou, 2013) and performed single-cell RNA-sequencing (scRNA-seq) using the SMART-Seq v5 Ultra Low Input RNA (SMARTer) kit for Sequencing (Takara Bio) (Karimi et al., 2021). Briefly, we aggregated in vitro derived PGCLCs, originating from either XGFP+ or XGFP– populations, for 6 days with somatic cells isolated from E13.5 female embryonic gonads plus mesonephros in order to mimic the female urogenital environment and provide in vitro-derived germ cells with the appropriate signalling niche (Chuva de Sousa Lopes et al., 2008; Hayashi et al., 2012) to facilitate their meiotic entry and X-reactivation (Fig 4A and B). We then sorted single cells of the following populations on which we performed scRNA-seq. Derived from XGFP– PGCLC rOvaries, we collected three populations: XGFP high reactivated (144 cells, XTOMato+/XGFP+), XGFP intermediate (XTOMato+/XGFPInt., 144 cells) and XGFP low (XTOMato+/XGFPlow, 136 cells). From the constitutively active XGFP+ PGCLC rOvaries, we collected one population: XGFP+ constitutive (XTOMato+/XGFP+, 188 cells) (Fig 4A and B). In total, we obtained 391 million reads, with an average of 740,000 reads per cell. Next, to ensure that our analysis focussed on germ cells of appropriate quality, we only included cells with the germ cell
marker Dazl expression > 1 (log2 counts per 10,000) and with sufficient allelic information (see methods). This left us with 379 cells in total and 15,583 informative genes.

To characterize cellular heterogeneity, we performed dimensionality reduction using UMAP on genome-wide single-cell expression data, and projected the four FACS populations (Fig 4B and C) on
our UMAP plot. We observed a clear separation into two major groups, which coincided well with the levels of XGFP fluorescence. One group predominantly included the XGFPlow and XGFPintermediate germ cells (originating both from the XGFP+ PGCLCs) on the left, and another group was constituted from the XGFPhigh reactivated and XGFP+ constitutive germ cells on the right. We then set out to answer whether our XGFP+ and XGFP− PGCLCs, which were the starting material for our rOvaries (Fig 4A and B), showed a differential developmental profile, and in particular, if the meiotic germ cells originated preferentially from XGFP+ or XGFP− PGCLCs. We applied SNN modularity optimization-based clustering which returned five clusters (Fig 4D) that showed distinct patterns according to the expression of mitotic and meiotic germ cell marker genes (Fig 4E and F). We identified two mitotic clusters termed “Mitotic 1” and “Mitotic 2,” showing expression of the PGC marker Stella (also known as Dppa3) as well as mitotic PGC markers Morc1 and Nanog. Moreover, we identified two pre-meiotic clusters termed “Pre-meiotic 1” and “Pre-meiotic 2,” defined by the initial expression of both Stella and Ddx4, and lastly one meiotic cluster termed “Meiotic” characterized by expression of the meiotic genes Prdm9 and Sycp3. Next, we wanted to assess whether a directionality within the clusters and eventually among the two groups could be observed. Pseudo-time analysis using RNA velocity (La Manno et al., 2018) placed the meiotic cluster at the apex of a path which revealed a differentiation trajectory directed towards meiosis, initiating from the pre-meiotic clusters (Fig 4F). Moreover, comparison to in vivo data (Zhao et al., 2020) showed that our mitotic clusters corresponded to E12.5 germ cells, whereas pre-meiotic and meiotic clusters corresponded to later time points; E14.5 and E16.5 (Fig EV4A–C), confirming that our in vitro clusters followed an in vivo-like developmental trajectory. Intriguingly, both pre-meiotic and meiotic germ cells almost exclusively originated from XGFP− PGCLCs, whereas mitotic germ cells consisted of XGFPhigh reactivated and XGFP+ constitutive germ cells.

Considering this, we wanted to assess whether our XGFP− PGCLCs could therefore mature further and initiate oogenesis. We took advantage of a published in vitro differentiation protocol and aggregated XGFP− PGCLCs with embryonic-derived somatic gonadal cells, forming an rOvary, followed by the culture of the rOvary onto a transwell to allow in vitro differentiation (IVDi) of PGCLCs (Fig 4G) (Hayashi et al., 2017). However, to perform the experiment in a more physiological niche, without external cues, no retinoic acid was added to the IVDi culture and the IVDi tissue was cultured for 11 days until primary follicles had formed. We then stained the entire whole-mount tissue for DAZL and SYCP3 to identify mature (DAZL+) and meiotic (SYCP3+) germ cells and could observe on average 200 oocytes in cysts per aggregate and moreover, around 50 primary follicles (Fig 4H and I), similar to what has been observed previously (Hamada et al., 2020), showing that our XGFP− PGCLCs could indeed mature further.

Taken together, germ cells seem to adopt highly similar transcriptomes when two active X chromosomes are present, irrespective of their parental condition of origin and hence regardless of whether cells underwent X-inactivation followed by X-reactivation (XGFPhigh), or were constitutively X-active (XGFP+). Moreover, germ cells can undergo X-reactivation in the absence of the meiotic gene expression programme, suggesting that X-reactivation is not dependent on meiotic entry. However, our data suggest that X-inactivation may be either functionally important, or, alternatively, a predictive indicator for subsequent germ cell maturation and entry into meiosis, as germ cells originating from constitutively active XGFP+ PGCLCs failed to acquire a meiotic transcriptional profile.

**Figure 4. Single-cell RNA-seq of maturing germ cells using the rOvary system.**

A Schematic illustration of the single-cell RNA-seq experiment and the isolated populations during germ cell maturation in rOvaries. The first 24h of culture are indicated as d0. rOvary, reconstituted Ovary; d, day of rOvary culture; XGFPlow, XGFPintermediate.

B (i) Imaging of XGFP and XTomato reporters in rOvaries ds5 aggregated with E13.5 gonadal and mesonephric cells. Scale bars = 50 μm. BF, bright field. (ii) FACS gating strategy for single cells sorted XTomato+ cells against XGFP intensities. Numbers indicate the percentage of gated live cells over the total population. Numbers in brackets indicate the percentage of gated cells over the XTomato+ population.

C UMAP projection labelled with FACS sorted populations. XGFPlow, XGFPintermediate.

D UMAP embedding based on shared nearest-neighbour (SNN) modularity clustering identified five clusters, termed Mitotic 1 (n = 77), Mitotic 2 (n = 97), Pre-meiotic 1 (n = 62), Pre-meiotic 2 (n = 90) and Meiotic (n = 53) labelled with different colours. Arrows indicating cell trajectories, inferred by RNA velocity analysis.

E Marker gene expression projected onto the UMAP plot.

F Heatmap of gene expression dynamics throughout germ cell maturation clusters. Selected genes belong to the category “early germ cell” and “meiotic.” Z-score is shown.

G Schematic representation of the IVDi (in vitro differentiation) maturation system. The stages of oogenesis in culture for 11 days are indicated. The condition of culture is indicated above. rOvary, reconstituted Ovary. Agg, aggregation day; AA, Ascorbic Acid.

H Immunofluorescence images of SYCP3 (red), DAZL (yellow) and DAPI at agg11 of IVDi tissue maturation from XGFP− PGCLCs. IVDi, in vitro differentiation. White squares indicate the positions of the magnified section shown below. Top panel scale bar = 100 μm. Middle panel scale bar = 10 μm. Bottom panel scale bar = 50 μm.

I Quantification of SYCP3+ cells (oocytes in cyst and primary follicles) in IVDi tissues at agg11. Each dot represents one IVDi tissue performed in three biological replicates.
despite originating from mostly XGFPlow and XGFPintermediate populations, showed close to biallelic expression at an average allelic ratio of ~ 0.4, as the sensitivity of the XGFP reporter was insufficient to mark cells as reactivated if they had low levels of X-inactivation (Fig 5B). We therefore assessed the X-status on a gene-by-gene level and compared it to the data of ESCs, EpiLCs and

**Figure 5.**

- **A** 220 X-linked genes
- **B** R = 0.74 p < 2.2 e-16
- **C** X-inactive \( \rightarrow \) X-active
- **D** Escapees Early XCR Late XCR
- **E** 1.3e-10 p < 2.22e-16

*Jacqueline Severino et al.* *The EMBO Journal* 41 e109457 | 2022 11 of 23
XGFP– PGCLCs (Fig 5C and D). In addition to 78 escapees, being active throughout the differentiation, we observed early X-chromosome reactivation (early XCR) of 58 genes in pre-meiotic cells. Therefore, the vast majority of genes (85%) had escaped X-inactivation in the first place, or had undergone reactivation, before the onset of meiosis. Moreover, 17 genes reactivated as cells underwent meiosis (late XCR), with only 8 genes still being inactive in meiotic cells (no XCR). Furthermore, we observed that early reactivating genes displayed higher allelic ratios in XGFP reporter quantified by BD FACSDiva Software, plotted against the X chromosome average allelic ratio per cell. R and P-values calculated by Pearson’s correlation are shown. Black line represents linear regression fitting.

Figure 5: Transcriptional reactivation of X-linked genes.

A Average allelic ratio of single cells projected onto the UMAP plot. n = 220 X-linked genes per single cell. X-inactivation (average ratio < 0.135) in red and X-reactivation (average ratio from 0.135 to 0.8) in green. Labels indicate the five different previously identified clusters. Dashed lines indicate the position of cluster borders (clusters obtained from Fig 4G–I).

B Distribution of single cells based on fluorescence intensity of XGFP reporter quantified by BD FACSDiva Software, plotted against the X chromosome average allelic ratio per cell. R and P-values calculated by Pearson’s correlation are shown. Black line represents linear regression fitting.

C Heatmaps of allele-specific ratios of X-linked genes in ESC, EpiLC, XGFP– PGCLC, pre-meiotic, meiotic and mitotic clusters. X-inactive genes are shown in red (ratio ≤ 0.135), X-active genes in green (ratio > 0.135) and mono-allelic X expression in blue (ratio between 0.5 and 1). Colour gradients used in between and above these two values as shown in the legend. Genes are ordered by genomic position and grouped according to the category to which they belong, indicated on the left side of the heatmap. n = 161 genes.

D Average allelic ratios of X-linked genes within each category (escapees, early XCR and late XCR) in ESC, EpiLC, XGFP– PGCLC, pre-meiotic and meiotic clusters. Shading denotes lower and upper Gaussian confidence limits based on the t-distribution.

E Each dot indicates the average allelic ratio of a single X-linked gene belonging to the indicated category in XGFP– PGCLC. The numbers above the bars indicate P-values (two-sample unpaired Wilcoxon–Mann–Whitney test with R defaults). Box plots depict the first and third quartiles as the lower and upper bounds of the box, with a band inside the box showing the median value and whiskers representing 1.5x the interquartile range.

X-inactivation in PGCLCs is associated with an increased meiotic competence

Our single-cell RNA-seq analysis showed an exclusive ability for XGFP– over XGFP+ PGCLCs to differentiate into mature germ cells with a meiotic transcriptional profile in the rOvary system. We therefore wanted to functionally compare their ability to enter meiosis and their capability to differentiate to more mature stages. To be able to assess in more detail the extent of prophase I progression, we cultured XGFP– and XGFP+ PGCLCs for an additional 9 days on immortalized m220 stromal feeder cells in the presence of BMP2 and retinoic acid (Fig 6A), which was previously shown to facilitate entry into meiosis (Miyauchi et al, 2017). During this expansion culture, we observed a progressive accumulation of SYCP3, which shows a distinctive pattern according to the different prophase stages (Fig 6E). The EMBO Journal

Discussion

While X-chromosome inactivation has been a long-studied phenomenon (Lyon, 1961) and has been shown to play an important biological role for embryonic development (Marahrens et al, 1997) and pluripotency exit (Schulz et al, 2014), its reversal by X-reactivation and its biological function during germ cell development have remained elusive to date. Previous studies on X-chromosome dynamics during female mouse germ cell development have been hampered by a lack of allelic resolution, a low number of genes assessed, as well as an inability to directly trace the X-chromosome status of single cells (Sugimoto & Abe, 2007; Chuva de Sousa Lopes et al, 2008). To overcome these limitations, we generated an in vitro system, which allowed us to reveal the X-chromosome inactivation and reactivation cycle and its functional relation to germ cell development and meiotic progression. Although our XGFP marker has its limitations by showing a slight delay in response during kinetic changes, such as downregulation during X-inactivation and upregulation during X-reactivation, it allowed us to isolate PGCLC populations with distinct developmental capacity. We thereby uncovered that X-inactivation is an important hallmark of proper PGCLC differentiation in order to progress at later stages towards meiotic entry (Fig 7). X-reactivation, on the other hand,
coincides temporally with meiotic maturation. This is in line with the timing of X-reactivation in mouse germ cells in vivo (Sugimoto & Abe, 2007; Chuva de Sousa Lopes et al, 2008; Sangrithi et al, 2017), where it takes place gradually, initiating during germ cell migration and peaking after colonization of the gonads around the time of meiotic entry. Additionally, our in vitro system enabled the isolation of PGCLCs harbouring two active X, a unique advantage over in vivo systems, as it allowed us to compare the differentiation potential of...
PGCLCs with and without X-inactivation. While our results suggest that PGCLC specification can occur in the absence of X-inactivation, we found that germ cells, which had never undergone X-inactivation in the first place, or in which X-reactivation occurred pre-emptively, displayed a mitotic germ cell character, did not enter a normal meiotic trajectory on a transcriptomic level and showed a significantly reduced meiotic differentiation capacity. This further highlights how timely X-inactivation and -reactivation might be necessary for proper germ cell maturation (Fig 7). Moreover, while we acknowledge that our findings are based on data generated in vitro, we note that allele-specific single-cell RNA-seq of E5.5–E6.5 epiblast cells, the precursors of PGCs, revealed a considerable heterogeneity in X-inactivation progression at this developmental time window (Mohammed et al., 2017; Cheng et al., 2019; preprint: Naik et al., 2021; Lentini et al., 2022), which could potentially allow cells to give rise to XaXa PGCs, similar to our XGFP PGCLCs. Indeed, when we analysed data from E6.5 epiblast in vivo (Cheng et al., 2019), we detected specifically in the proximal epiblast, which contains the competent precursor cells for PGC development (Lawson & Hage, 1994; Saitou et al., 2002), a high fraction of cells that have not gone through X-inactivation (Fig 3H and I). Furthermore, proximal epiblast cells which did go through X-inactivation displayed a lower degree of gene silencing than other cells of the epiblast or extraembryonic tissues at that stage (Fig 3J and K). Thus, our data support the idea of a potential functional link between appropriate X-chromosome dosage compensation kinetics and developmental progression during mammalian germ cell maturation. Nevertheless, we acknowledge that it would be important to validate our results during later stages of germ cell development in vivo. Particularly, whether PGCs that have failed to go through X-inactivation exist, and to what fate these cells would commit.

It remains an open question, what could be the potential role of X-inactivation for proper PGCLC development and if it is a driver or, alternatively, a diagnostic mark for meiotic competence of germ cells. We observed that XGFP+ PGCLCs, which failed to undergo X-inactivation, differed from XGFP− PGCLCs on multiple accounts. Albeit sharing an overall similar transcriptome signature with their XGFP− germ cell counterparts, XGFP+ PGCLCs displayed ESC-like features including a higher expression of naive pluripotency genes, shortened cell cycle and propensity to form pluripotent EGC colonies when cultured under 2i/LIF conditions. An explanation for this pluripotency-related phenotype could be the two-fold expression of critical X-linked dosage-sensitive genes, which need to be silenced...
by X-inactivation to allow normal pluripotency exit during ESC differentiation (Schulz et al., 2014). For example, Dusp9, an X-linked regulator of MAPK signalling, has been shown to be responsible for the lower DNA-methylation levels of XX pluripotent stem cells, when compared with XY and XO cells (Choi et al., 2017; Song et al., 2019; Genolet et al., 2021). In germ cell development, DNA methylation safeguards repression of late germ cell / meiotic genes during early germ cell stages and demethylation of their promoters is required for their upregulation during germ cell maturation and meiotic entry (Yamaguchi et al., 2012; Hill et al., 2018). Along those lines, we observed that XGFP⁺ PGCLCs also displayed precocious expression of a subset of late germ cell markers, which remained repressed in XGFP⁻ PGCLCs. Importantly, demethylation of late germ cell genes alone has been shown to only lead to partial activation of some germ cell genes, while not being sufficient for their full expression in the absence of meiosis-inducing signals (Miyauchi et al., 2017; Ohta et al., 2017). This would explain our observation of a relatively mild upregulation of late germ cell genes in our XGFP⁺ PGCLCs and why this was not sufficient to aid entrance of XGFP⁺ cells into a full meiotic trajectory after their aggregation with gonadal somatic cells.

**Materials and Methods**

**Cell culture**

**Embryonic stem cell culture: Serum/LIF**

Embryonic Stem Cells (ESCs) were maintained and expanded on 0.2% gelatin-coated dishes in DMEM (Thermo Fisher Scientific, 31966021) supplemented with 10% Fetal Bovine Serum (FBS) (ES qualif, Thermo Fisher Scientific, 16141079), 1,000 U/ml LIF (ORF Genetics, 01-A1140-0100), 1 mM Sodium Pyruvate (Thermo Fisher Scientific, 11360070), 1x MEM Non-Essential Amino Acids Solution (Thermo Fisher Scientific, 11140050), 50 U/ml penicillin/streptomycin (bien Tech, P06-07100) and 0.1 mM 2-mercaptoethanol (Thermo Fisher Scientific, 31350010). Cells were cultured at 37°C with 5% CO2. Medium was changed every day and cells were passaged using 0.05% Trypsin-EDTA (Thermo Fisher Scientific, 25300054) and quenched 1:5 in DMEM supplemented with 10% FBS (Life Technologies, 10270106). Cells were monthly tested for mycoplasma contamination by PCR.

**Embryonic stem cell culture: 2i/LIF**

ESCs were cultured for 24 h prior to the start of the primordial germ cell-like cell induction in 2i/LIF medium. Briefly, a home-made version of the N2B27 medium was prepared based on previous reports (Ying et al., 2008) with additional modifications reported in Hayashi and Saitou (2013) containing two chemical inhibitors 0.4 μM PD032591 (Selleck Chemicals, S1036) and 3 μM CHIR99021 (SML1046, SML1046) together with 1,000 U/ml LIF (ORF Genetics, 01-A1140-0100). ESCs were seeded on a dish coated with 0.01% poly-L-ornithine (Sigma-Aldrich, P3655) and 500 ng/ml laminin (Corning, 354232).

**XRep cell line generation**

We used the female F2 ESC line EL16.7 TST (obtained from Jeannie Lee, Massachusetts General Hospital (Boston, USA)), derived from a cross of Mus musculus musculus with Mus musculus castaneus (Ogawa et al., 2008). As a result, cells contain one X chromosome
from *M. m. musculus* (X\textsubscript{mus}) and one from *M. m. castaneus* (X\textsubscript{cas}). Moreover, EL16.7 TST contains a truncation of Tsix on X\textsubscript{mus} (Tsx\textsubscript{XST}/Tsx\textsubscript{XREP}), which abrogates Tsix expression and leads to the non-random inactivation of X\textsubscript{mus} upon differentiation. XGFP and XtdTomato vectors were integrated first, followed by integration of rtTA and last of germ cell transcription factor vectors.

**XGFP and XtdTomato dual-colour reporter**

A GFP reporter construct (Wu et al., 2014) was targeted in the second exon of Hprt on X\textsubscript{mus} as described in Bauer et al. (2021). The same strategy was used to simultaneously target a tdTomato reporter construct in the second exon of Hprt on X\textsubscript{cas} and a GFP reporter on X\textsubscript{mus}. Briefly, 5 × 10\textsuperscript{5} EL16.7 TST ESCs were nucleofected with the AMAXA Mouse Embryonic Stem Cell Nucleofector kit (LONZA, VPH-1001) using program A-30 with 1.6 µg each of GFP and tdTomato circularized targeting vectors and 5 µg single gRNA vector PX459 (5'-TATACCTAATCATTATTTGCCG-3') (Addgene, 48139, a gift from Feng Zhang). Homology arms flanking the target site were amplified from genomic DNA and cloned into pBluescript II SK(+) (Addgene, 212205) by restriction enzyme-based cloning and the hS4-CAG-nlstdTomato-hS4 and hS4-CAG-nlsGFP-hS4 constructs, kindly provided by J. Nathans (Wu et al., 2014), were cloned between the two homology arms. 7.5 µM of RS-1 (Merck, 553510) was added to enhance homology-directed repair. To select for the homologous disruption of Hprt, cells were grown in the presence of 10 µM 6-thiouguanine (Sigma-Aldrich, A4882-250MG) for 6 days, and GFP+/tdTomato+ cells were isolated by FACS using a BD Influx (BD Biosciences). Single clones were screened by Southern blot hybridization as described in (Bauer et al., 2021).

**Rosa26 rtTA**

One microgram of R26P-M2rtTA-targeting vector (Addgene, 47381) and 5 µg of PX459 gRNA vector (5'-GACTCCAGTCTTCTTAGAAG-3') were nucleofected with the AMAXA Mouse Embryonic Stem Cell Nucleofector kit (LONZA, VPH-100) using program A-30 in the XRep. Cells were selected with 3 µg/ml puromycin (Ibian tech., ant-pr-1) for 5 days, with medium being changed daily. Single clones were screened for rtTA expression by quantitative RT–PCR and by Southern blot hybridization, with genomic DNA being digested by EcoRV.

**Germ cell transcription factors overexpression**

PB-TET vectors containing key germ cell factors Blimp1, Tfap2c and Prdm14 (Nakaki et al., 2013) were kindly given by F. Nakaki. Cells were transfected with 3 µg each of PB-TET vectors, pBPAC-hpg and a PiggyBac transposase vector using the AMAXA Mouse Embryonic Stem Cell Nucleofector kit (LONZA, VPH-1001). Transfected cells were selected with 200 µg/ml hygromycin B (Gold (Ibian Tech., ant-hg-1)) for 10 days and genotyped by PCR for transgenes. The primer sequences are shown in Table 1.

Copy number integration was estimated by Southern blot hybridization. Briefly, 15 µg of genomic DNA were digested with BamHI. DNA fragments were electrophoresed in 0.8% agarose gel and transferred to an Amersham Hybond X membrane (GE Healthcare, RN303S). The b-geo probe was designed downstream of the BamHI site, obtained by digesting the PB-TET-Avi-Blimp1 plasmid with Cpol/Smal, labelled with dCTP [x-32P] (Perkin Elmer, NEG513H250UC) using High Prime (Roche, 11585592001), purified with an Illustra ProbeQuant G-50 Micro Column (GE Healthcare, 28903408) and hybridization performed in Church buffer. Radioisotope images were captured with a Phosphorimager Typhoon Trio.

**Epiblast-like cell and primordial germ cell-like cell induction**

XRep ESCs were induced into PGCLCs as described previously (Hayashi & Saitou, 2013) with the following modifications as this condition was most efficient in generating PGCLCs. ESCs were

<table>
<thead>
<tr>
<th>Target Transcript</th>
<th>Forward (5’-3’)</th>
<th>Reverse (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbp</td>
<td>CAACCTGACCCAGAACAAAGGAG</td>
<td>AATTAACGACGCTGACCTGTTC</td>
</tr>
<tr>
<td>Xist mus/cas</td>
<td>ATCATCTAAAGCCACACAAGAATTC</td>
<td>ATTTGATTTGCAAGGCTGAT</td>
</tr>
<tr>
<td>Amot mus/cas</td>
<td>TTTGCTCCACCTGGTCACAAGAG</td>
<td>GACACGTTTGAGAGGAGAC</td>
</tr>
<tr>
<td>Prdx4 mus/cas</td>
<td>TGAGCTCTTCAAGGTATACCTG</td>
<td>TGAAGTGTCAAGATCTCCTGTT</td>
</tr>
<tr>
<td>Prxk mus/cas</td>
<td>TCGAAATGGAACACAAAGGC</td>
<td>CCACACTACGAGTATGAG</td>
</tr>
<tr>
<td>Rnf4</td>
<td>TGCTGCTGGTGACTGTGG</td>
<td>GCTCCCCGTTGAGTACCTT</td>
</tr>
<tr>
<td>Dnd1</td>
<td>GCTGGCTACAGTTACCCAC</td>
<td>GAAGTGTGCTTGGACTG</td>
</tr>
<tr>
<td>Zfp42 (Rest)</td>
<td>CCCTGCAGAAGCTGATCTCAAA</td>
<td>TGGGGCTATCTCACTATGGT</td>
</tr>
<tr>
<td>Dnmt3b</td>
<td>CTCCGGAAGTTCGGGCTTCTTAC</td>
<td>CTTGGCATCTTCAGTCCC</td>
</tr>
<tr>
<td>Dnmt3l</td>
<td>CGCGGAGAATTCTCTCCAGGTC</td>
<td>TGAAGTGCAGACAGGAC</td>
</tr>
<tr>
<td>Blimp1</td>
<td>ACCATGACCTCAGGACACC</td>
<td>CTCACACTCTCATGTAAGGAC</td>
</tr>
<tr>
<td>Prdm14</td>
<td>ACAACCAACGAAAATCTGCAC</td>
<td>TTACCTGCGTATTCCAGG</td>
</tr>
<tr>
<td>Tfap2c</td>
<td>GGGCTTTCTCTCTGTGGT</td>
<td>TCCACACTCCACACAGAA</td>
</tr>
<tr>
<td>Aur-Blimp1</td>
<td>TGCTGCTCCATGAGTCAAC</td>
<td>GCCCAGAAATTTGATCAC</td>
</tr>
<tr>
<td>SexFLAG-Prdm14</td>
<td>TCTGGCAGAAGGCTTTC</td>
<td>ACTAAGTACGGCCCATCAG</td>
</tr>
<tr>
<td>V5-Tfap2c</td>
<td>ATTCAGGCAAGACTGAGAG</td>
<td>GCCGAAATTTGATCAC</td>
</tr>
<tr>
<td>rtTA</td>
<td>CTACCACCGATTATGCCC</td>
<td>CCCTTTGAGAGTATCCTG</td>
</tr>
</tbody>
</table>

Copy number integration was estimated by Southern blot hybridization. Briefly, 15 µg of genomic DNA were digested with BamHI. DNA fragments were electrophoresed in 0.8% agarose gel and transferred to an Amersham Hybond X membrane (GE Healthcare, RN303S). The b-geo probe was designed downstream of the BamHI site, obtained by digesting the PB-TET-Avi-Blimp1 plasmid with Cpol/Smal, labelled with dCTP [x-32P] (Perkin Elmer, NEG513H250UC) using High Prime (Roche, 11585592001), purified with an Illustra ProbeQuant G-50 Micro Column (GE Healthcare, 28903408) and hybridization performed in Church buffer. Radioisotope images were captured with a Phosphorimager Typhoon Trio.
thawed on 0.2% gelatin in serum/LIF and after 24 h seeded at a density of 0.6 \times 10^{3} \text{cells/cm}^2 in 2i/LIF medium on a dish coated with 0.01% poly-L-ornithine (Sigma-Aldrich, P3655) and 500 ng/ml laminin (Corning, 354232). Twenty-four hours later, ESCs were dissociated with TrypLE express for 5 min at 37°C and induced into EpiLCs by addition of human recombinant basic fibroblast growth factor (bFGF) (Invitrogen, 13256-029) and activin A (Peprotech, 120-14P) and seeding on 16.7 μg/ml human plasma fibronectin-coated plates (Merck Millipore, FC010). After 48 h, EpiLCs were split using TrypLE Express (Life Technologies 12604013) and re-seeded at 0.2 \times 10^{5} \text{cells/cm}^2 on 16.7 μg/ml human plasma fibronectin-coated plates. After an additional 48 h, EpiLCs were aggregated in U-bottom 96-well Lipidure-Coat plate (Thermo Fisher Scientific, 81100525) at 2,000 cells per aggregate in GK15 medium (GMEM (Life Technologies, 11710035), 15% KnockOut Serum Replacement (KSR) (Thermo Fisher, 10828028), 0.1 mM nonessential amino acids (NEAA) (Thermo Fisher Scientific, 11140050), 1 mM sodium pyruvate (Thermo Fisher Scientific, 11360), 2 mM Glutamax (Life Technologies, 35050061), 0.1 mM 2-mercaptoethanol (Thermo Fisher Scientific, 21985-023) and 100 U/ml penicillin and 0.1 mg/ml streptomycin (Thermo Fisher Scientific, 15140) with 1.5 μg/ml doxycycline (Tocris, 4090/50) for 5 days.

**Embryonic Germ Cell (EGC) colony-forming assay**

Primordial germ cell-like cells were sorted by FACS at day 5 of differentiation for their XGFP status and seeded onto immortalized mouse embryonic fibroblasts at 1,000 cells/well of a six-well plate. Cells were cultured for 7 days in 2i/LIF medium thereby facilitating their transition into pluripotent EGC colonies (Leitch et al., 2010), changing the medium every 24 h. EGC colony-forming capacity of XGFP+ and XGFP– PGCLCs was compared to ESCs replated at equal numbers and scored by alkaline phosphatase (AP) staining. Cells were fixed in 4% paraformaldehyde before adding the AP solution (1 mg/ml Fast Red (Sigma, F8764) and 0.01% w/v Naphthol AS-MX phosphate (Sigma, 8551). The reaction was stopped after 10 min by removing the AP solution and washing the cells with MillQ water. Each well was imaged with a digital camera and EGC colonies were counted manually.

**PGCLCs mitotic expansion**

PGCLCs mitotic expansion culture was performed as previously described (Ohta et al., 2017) with few modifications. Briefly, 5 days after PGCLC induction, SSEA1+/CD61+ PGCLCs were sorted by flow cytometry onto m220 feeder cells, which constitutively express a membrane-bound form of mouse stem cell factor (Dolci Majumdar et al., 2018). Cells were dissociated and quenched as described above. Cells were then fixed for 10 min at room temperature with freshly prepared 1% formaldehyde in PBS (Sigma-Aldrich, F8775-4X25ML) and the reaction was quenched by addition of 0.2 M glycine (NZYTech, MB01401) for 15 min on ice. 1 \times 10^{6} cells/ml were permeabilized using 0.1% saponin (Sigma-Aldrich, 47036-50-G) containing 10 μg/ml DAPI (Thermo Fisher Scientific, D1306) and 100 μg/ml treatment of 300 ng/ml BMP2 (R&D Systems, 355-BM) and 100 nM RA. Medium was replaced completely every 2 days until the end of the culture period.

**rOvary reconstitution**

A total of 10,000 sorted SSEA1+/CD61+ PGCLCs were mixed with 75,000 freshly thawed E13.5 female somatic gonadal and mesonephric cells (SSEA1–/CD31–) or E12.5 female somatic gonadal cells from CD1/ICR strain mice and cultured in Lipidure-Coat plates at 37°C in a 5% CO2 incubator for 6 days for the scRNAseq protocol or for 2 days for the IVDi as described in Hayashi et al (2017). Mouse care and procedures were conducted according to the protocols approved by the Ethics Committee on Animal Research of the Parc de Recerca Biomèdica de Barcelona (PRBB) and by the Departament de Territori i Sostenibilitat of the Generalitat de Catalunya (Ref. No. 10469).

**Oocyte in vitro differentiation (IVDi) culture**

IVDi culture was performed as previously described (Hayashi et al., 2017). Briefly, one single rOvary was placed in the middle of a 24-well Transwell-COL membrane (Corning, CLS3470-48EA) and cultured in alpha-MEM (Life Technologies, 12571063) with 0.15 mM ascorbic acid (Sigma-Aldrich, A7506), 2% FBS, 2 mM Glutamax (Life Technologies, 35050061), 0.1 mM 2-mercaptoethanol and 50 U/ml penicillin/streptomycin (Thermo Fisher Scientific, 15140) with 1.5 μg/ml doxycycline (Tocris, 4090/50) for 5 days.

**Fluorescence-activated cell sorting (FACS)**

After 5 days of culture, PGCLC aggregates were dissociated using TrypLE Express (Thermo Fisher Scientific, 12604021) for 8 min at 37°C, with periodical tap mixing. The reaction was quenched 1:5 with wash buffer DMEM/F12 (Thermo Fisher Scientific, 11320-082) containing 0.1% bovine serum albumin (BSA) fraction V (Thermo Fisher Scientific, 15260-037) and 30 mM HEPES (Gibco, 15630-056) containing 0.1 mg/ml of DNase I (Sigma-Aldrich, DN25-10MG). The cell suspension was centrifuged at 300 g for 5 min, resuspended in FACS buffer (0.1% BSA in PBS) and passed through a 70 μm cell strainer (Corning, 352350). Cells were stained with 1:100 SSEA1-eFluor 660 (Thermo Fisher Scientific, 50-8813-42) and 1:10 CD61-stain (Corning, 352350). Cells were washed thrice with FACS Buffer, stained with 1:1,000 DAPI (Thermo Fisher Scientific, D1306) and then FACS sorted using a BD FACSAria II or a BD Influx. Double-positive population of PGCLCs was collected in GK15 medium. Data were analysed with Flowjo (Tree Star) software.

**Cell cycle analysis**

Identification of G1, S and G2/M cell cycle phases was based on DNA content and performed as described previously (Bonev et al., 2017) with minor modifications. Briefly, ESCs, EpiLC and PGCLCs were dissociated and quenched as described above. Cells were then fixed for 10 min at room temperature with freshly prepared 1% formaldehyde in PBS (Sigma-Aldrich, F8775-4X25ML) and the reaction then quenched by addition of 0.2 M glycine (NZYTech, MB01401) for 15 min on ice. 1 \times 10^{6} cells/ml were permeabilized using 0.1% saponin (Sigma-Aldrich, 47036-50-G) containing 10 μg/ml DAPI (Thermo Fisher Scientific, D1306) and 100 μg/ml
RNase A (Thermo Fisher Scientific, EN0531) for 30 min at room temperature, protected from light with slight agitation. After washing once with cold PBS, samples were resuspended in cold 0.5% BSA in PBS at a concentration of 1 × 10^6 cells/ml and immediately analysed using a BD LSRFortessa.

**Immunofluorescence of PGCLC bodies and rOvaries**

Immunofluorescence analysis of PGCLC bodies or rOvaries was performed on cryosections prepared as follows: Aggregates were fixed with 4% paraformaldehyde (PFA) (Electron Microscopy Science, 15713) in PBS at room temperature for 30 min, followed by three washes in PBS and submerged in serial concentrations of 10 and 30% of sucrose (Sigma-Aldrich, S0389) in PBS, 15 mins and overnight at 4°C respectively. The samples were embedded in OCT compound (Sakura Finetek, 4583), snap-frozen in liquid nitrogen and cryo-sectioned at a thickness of 10 μm at −20°C on a cryostat (Leica, CM1850). The sections were placed on a coated glass slide (MAS-GP type A; Matsunami, S9901) and dried completely. For immunostaining, the slides were blocked with PBS containing 10% normal goat serum (NGS) (Abcam, ab7481), 3% BSA (Sigma-Aldrich, A3311) and 0.2% Triton X-100 (Sigma-Aldrich, T9284) for 1 h at room temperature, followed by incubation with the primary antibodies diluted in a 1:1 solution of blocking buffer to PBS with 0.2% Tween (PBST) (Sigma-Aldrich, P7949) overnight at room temperature. The slides were washed three times with PBST, then incubated with the secondary antibodies diluted as the primary, with DAPI at 1 μg/ml for 1 h at room temperature. Following three washes in PBST, the samples were mounted in VECTASHIELD with DAPI (Vector Laboratories, H1200) and observed under a Leica SP8 confocal microscope. All images were analysed using Fiji/ImageJ software (Schindelin et al., 2012). All antibodies used in this study are listed in Table 2.

**Immunofluorescence of cultured PGCLC-derived cells**

Immunofluorescence analysis of cultured PGCLC-derived cells was performed as described in (Nagaoka et al., 2020). Briefly, PGCLCs were cultured on m220 feeder cells seeded on a 0.1% gelatin-coated plate used specifically for imaging (Nunc, 165305). PGCLC-derived cells were fixed at c5, c7 or c9 with 4% PFA (Electron Microscopy Sciences, 15713) with 0.2% Triton X-100 (pH 9.2) overnight. After three washes in PBST, cells were incubated with the secondary antibodies and DAPI at room temperature overnight. For immunostaining, the slides were blocked with PBS containing 10% NGS, 3% BSA and 0.2% Triton X-100 for 1 h, then incubated with the primary antibodies diluted in a 1:1 solution of blocking buffer to PBS with 0.2% Tween (PBST) (Sigma-Aldrich, P7949) overnight at room temperature. After three washes in PBST, cells were incubated with the secondary antibodies and DAPI at room temperature for 2 h and washed three times in PBST. Finally, the well was filled with VECTASHIELD without DAPI (Vector laboratories, H1000). Immunostained samples were observed with a Leica SP8 confocal microscope.

**Meiotic cell spreads**

Cultured PGCLC-derived cells were harvested by TrypLE Express at 37°C for 5 min, quenched with 1:1 TrypLE wash buffer (DMEM/F12 containing 0.1% BSA fraction V, 30 mM HEPES), filtered through a 70 μm strainer and centrifuged at 300 g for 5 min. Cell pellets were dislodged by tapping and washed once in PBS. Cells were then treated with a hypotonic solution (30 mM Tris–HCl, 50 mM sucrose (Sigma, S0389), 17 mM trisodium citrate, 5 mM ethylenediaminetraacetic acid (EDTA), 2.5 mM dithiothreitol (DTT) (Sigma, D0632) and 0.5 mM phenylmethylsulfonylfluoride (PMSF) (Sigma, P7626)), pH 8.2-8.4 at room temperature for 20 min. Cells were spun down 3 min at 300 g, resuspended in 100 mM sucrose and the cell suspension distributed onto slides (Matsunami, S9901) covered with 1% PFA in H2O (Electron Microscopy Science, 15713) with 0.2% Triton X-100 (pH 9.2–9.4). The slides were incubated at room temperature overnight in a humidified chamber. Finally, the slides were air-dried and washed with 0.5% Kodak Photo-Flo 200 (Kodak, B00K335F65) for 2 min at room temperature. The spread slides were blocked in PBS containing 10% NGS, 1% BSA for 1 h and then incubated with the primary antibodies diluted in a 1:1 solution of blocking buffer to PBS with 0.2% Tween (PBST) at room temperature overnight. After three washes in PBST, cells were incubated with the secondary antibodies and DAPI at room temperature for 2 h, washed three times in PBST and mounted in VECTASHIELD mounting medium with DAPI (Vector Laboratories, h1200). Immunostained cells were observed under a Leica SP8 confocal microscope.

**Table 2. Antibodies used in this study.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Dilution</th>
<th>Company</th>
<th>Catalogue#</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary antibody</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Sox2</td>
<td>Rabbit polyclonal</td>
<td>100x</td>
<td>Abcam</td>
<td>ab97959</td>
</tr>
<tr>
<td>Anti-Tfap2 (6E/4/4)</td>
<td>Mouse monoclonal</td>
<td>300x</td>
<td>Santa Cruz</td>
<td>SC12762</td>
</tr>
<tr>
<td>Anti-H3K27me3</td>
<td>Mouse monoclonal</td>
<td>500x</td>
<td>Active Motif</td>
<td>61017</td>
</tr>
<tr>
<td>Anti-Sycp3</td>
<td>Mouse monoclonal</td>
<td>100x</td>
<td>Abcam</td>
<td>ab97672</td>
</tr>
<tr>
<td>Anti-γH2AX S139</td>
<td>Rabbit polyclonal</td>
<td>100x</td>
<td>Abcam</td>
<td>ab11174</td>
</tr>
<tr>
<td>Anti-Dazl</td>
<td>Rabbit polyclonal</td>
<td>200x</td>
<td>Abcam</td>
<td>ab34139</td>
</tr>
<tr>
<td>Anti-GFP</td>
<td>Chicken polyclonal</td>
<td>500x</td>
<td>Abcam</td>
<td>ab13970</td>
</tr>
<tr>
<td><strong>Surface markers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSEA1-eFluor 660</td>
<td>Mouse monoclonal</td>
<td>50x</td>
<td>Thermos</td>
<td>S0-8813-42</td>
</tr>
<tr>
<td>CD61-PE-Vio770</td>
<td>Hamster monoclonal</td>
<td>10x</td>
<td>Miltenyi Biotec</td>
<td>130-102-627</td>
</tr>
<tr>
<td><strong>Secondary antibody</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-chicken IgY</td>
<td>Goat polyclonal / Alexa488</td>
<td>500x</td>
<td>Life Technologies</td>
<td>A11039</td>
</tr>
<tr>
<td>Anti-rabbit IgG</td>
<td>Goat polyclonal / Alexa488</td>
<td>500x</td>
<td>Life Technologies</td>
<td>A11034</td>
</tr>
<tr>
<td>Anti-mouse IgG</td>
<td>Goat polyclonal / Alexa555</td>
<td>500x</td>
<td>Life Technologies</td>
<td>A21424</td>
</tr>
<tr>
<td>Anti-rabbit IgG</td>
<td>Donkey polyclonal / Alexa647</td>
<td>500x</td>
<td>Life Technologies</td>
<td>A31573</td>
</tr>
</tbody>
</table>
**Immunofluorescence of IVDi tissues**
Day 11 IVDi tissues were treated while still attached to the transwell member as follow: culture medium was carefully removed from the transwell and the whole membrane was fixed in 4% PFA (Electron Microscopy Science, 15713) in PBS for 30 min at room temperature, washed twice with PBS and blocked overnight at room temperature in 10% NGS, 1% BSA and 0.2% Triton X-100. Primary antibodies were diluted in a 1:1 solution of blocking buffer to PBS with 0.2% Tween (PBST) and incubated overnight. After three washes with PBST, secondary antibodies and DAPI diluted as the primary were incubated an additional overnight, washed thrice and the whole membrane mounted on VECTASHIELD with DAPI (Vector Laboratories, H1200). Immunostained tissues were observed under a Leica SP8 confocal microscope.

**Tilescan analysis**
All images were analysed using Fiji/Image J software (Schindelin et al., 2012). XtdTomato fluorescence was used to determine aggregate size. First, Gaussian Blur with Sigma (Radius) 5.00 was applied. Then, a threshold with 10–255 and settings “dark background” and “B&W” was set. Finally “Analyze Particles” was used to measure aggregate size.

**RNA fluorescent in situ hybridization and immunofluorescence**
Cells were fixed with 3% paraformaldehyde PFA (Electron Microscopy Science, 15713) for 10 min with 2 mM Ribonucleoside-Vanadyl Complex (New England Biolabs, SI4025) at room temperature and then permeabilized for 5 min on ice in 0.5% Triton-X with 2 mM RVC. Cells were then blocked in 3% BSA/PBS with 2 mM RVC for 1 h at room temperature, incubated with primary antibodies diluted in blocking solution with 2 mM RVC overnight at 4°C. The secondary antibodies were diluted in blocking buffer and incubated 1 h at room temperature. Cells were then again fixed in 3% PFA for 10 min at room temperature. Strand-specific RNA FISH was performed with fluorescently labelled oligonucleotides (IDT) as described previously (Del Rosario et al., 2017). Briefly, probe mix was prepared by mixing 10 ng/ml equimolar amounts of Cy5 labelled Xist probes BD384-Xist-Cy5-3’ (5’-ATG ACT CTG GAA GTC AGT ATG GAG /3Cy5Sp/ -3’), BD417-5’Cy5-Xist-Cy5-3’ (5’- /5Cy5/ label) and Xist probes BD384-Xist-Cy5-3’ (5’-ATG ACT CTG GAA GTC AGT ATG GAG /3Cy5Sp/ -3’), BD417-5’Cy5-Xist-Cy5-3’ (5’- /5Cy5/ label). This probe mix was labeled with Cy5 dye and the labeled probes were then hybridized to the samples at 90°C and then slow cooled to room temperature. Following hybridization, the slides were incubated in 37°C for 1 h to allow for proper hybridization. After hybridization, the slides were washed in 2xSSC pH 7 at room temperature overnight. Slides were then washed in 25% formamide 2xSSC pH 7 at room temperature, incubated an additional overnight, washed thrice and the whole membrane mounted with VECTASHIELD with DAPI (Vector Laboratories, H1200). Images were acquired using a Zeiss Cell Observer.

**RNA extraction, cDNA synthesis and qPCR analysis**
Total RNA was isolated from ESCs, EpilCs and PGCLCs (two biological replicates each, corresponding to two different transwell cultures) and was used for qRT–PCR analysis in triplicate reactions with Power SYBR Green PCR Master Mix (Thermo Fisher Scientific, 4367659). The gene expression levels are presented as ΔΔCt normalized with the mean Ct values of one housekeeping gene, Arbp, in a normalization sample (ESCs). The primer sequences used in this study are listed in Table 1.

**Bulk RNA-seq analysis**
RNA libraries were prepared using the TruSeq Stranded Total RNA Library Preparation kit (Illumina, 20020596) followed by 125 bp paired-end sequencing on an Illumina HiSeq 2500.

**Allele-specific analysis**
FastQ files that passed quality control were aligned to the mm10 reference genome containing CAST/EU and 129S1/SvImJ SNPs positions masked. The positions of all 36 mouse strains SNPs were downloaded from ftp: https://ftp-mouse.sanger.ac.uk/REL-1505-SNPs_indels/mbp.v5_merged.snp_all.dbSNP142.vcf.gz.tbi. From here, we generated a VCF file containing only the SNPs for the strains of interest, CAST/EU and 293S1/SvImJ. Reads with ≥ 1 SNPs were retained and aligned using STAR (Dobin et al., 2013) implementing the WASP method (van de Geijn et al., 2015) for filtering of allele-specific alignments.

The generated bam files were used for counting reads using the HTseq tool (v0.6.1) (Anders et al., 2015). All of the steps above were performed using a customized Nextflow pipeline (Di Tommaso et al., 2017). We obtained between 50x10⁶ and 75x10⁶ reads per replicate. Coherence among samples, time points and replicates was verified by principal component analysis (PCA). Batch effects in principal component analysis (PCA) for comparison to in vivo samples were corrected using the R package limma (Ritchie et al., 2015).

Differential expression analysis was performed using the R package DESeq2 (v1.16) (Love et al., 2014). Briefly, differentially expressed genes were called by comparing XGFP+ PGCLCs and XGFP- PGCLCs or XGFP+ PGCLCs to ESCs. The DESeqDataSet (dds) was generated considering the dataset in its entirety while the DEseq analysis was conducted on dataset filtered as follows: Read counts were normalized by library size using “estimateSizeFactors;” were filtered for having a mean across the samples > 10 (a more stringent cut-off than the sum across the samples > 10) and poorly annotated genes on chromosomal patches were removed. The resulting 16,289 genes were kept for downstream analysis. Log2-fold change was shrunk using the “normal” parameter.

Gene ontology enrichment analysis performed on top and bottom differentially expressed genes defined as FDR < 0.001 e log2-fold change > |1| using the Gorilla. Over-represented categories were simplified using Revigo (http://revigo.irb.hr/) using a similarity of 0.4 as threshold. As background, all identified genes were used.

**Single-cell RNA-seq analysis**
Full-length single-cell RNA-seq libraries were prepared using the SMART-Seq v5 Ultra Low Input RNA (SMARTer) kit for Sequencing (Takara Bio). All reactions were downscaled to one quarter of the original protocol and performed following thermal cycling manufacturer’s conditions. Cells were sorted into 96-well plates containing 2.5 μl of the reaction buffer (1× Lysis Buffer, RNase Inhibitor 1 U/μl). Reverse transcription was performed using 2.5 μl of the RT MasterMix (SMART-Seq v5 Ultra Low Input RNA kit for Sequencing, Takara Bio). cDNA was amplified using 6 μl of the PCR MasterMix (SMART-Seq v5 Ultra Low Input RNA kit for Sequencing, Takara Bio) following the manufacturer’s protocol. cDNA was amplified using 6 μl of the PCR MasterMix (SMART-Seq v5 Ultra Low Input RNA kit for Sequencing, Takara Bio) following the manufacturer’s protocol.
Bio) with 25 cycles of amplification. Following purification with Agencourt Ampure XP beads (Beckmann Coulter), product size distribution and quantity were assessed on a Bioanalyzer using a High Sensitivity DNA kit (Agilent Technologies). A total of 140 pg of the amplified cDNA was fragmented using Nextera XT (Illumina) and amplified with double indexed Nextera PCR primers (IDT). Products of each well of the 96-well plate were pooled and purified twice with Agencourt Ampure XP beads (Beckmann Coulter). Final libraries were quantified and checked for fragment size distribution using a Bioanalyzer High Sensitivity DNA kit (Agilent Technologies). Pooled sequencing of Nextera libraries was carried out using a HiSeq4000 (Illumina) obtaining between 0.5 × 10^6 and 1.5 × 10^6 reads per cell. Sequencing was carried out as paired-end (PE75) reads with library indexes corresponding to cell barcodes.

Allele-specific alignment was done as described for bulk RNA-seq analysis using STAR and WASP. Data processing and visualization was performed using the R package Seurat (v4.0) (Stuart et al., 2019).

**In vitro single-cell analysis**

The non-allelic gene expression matrix was filtered for protein-coding and non-coding transcripts using annotations from mmuseculus_gene_ensembl version 67. Low-quality cells with less than 4,000 identified genes, less than 10,000 RNA molecules or more than 5% mitochondrial reads were removed. Data were log normalized and the top 2,000 highly variable features were selected for downstream analysis. The expression matrix was then scaled and linear dimensional reduction was performed. To ensure that our analysis would not be confounded by in vitro differentiation artefacts, we focussed our analysis on germ cells by subsetting for cells with a normalized and scaled Dazl expression greater than 1 (60 of 460 sorted germ cells did not pass this criterium). Moreover, only cells that passed our allele expression QC (explained below) were retained. Clusters were subsequently identified using “FindClusters” at a resolution of 0.8 on the first 20 principal components and visualized as UMAP projections using “RunUMAP.” Clusters were annotated based on marker gene expression. Processing of allelic data was performed for all cells that passed the Seurat QC. Cells that passed the following criteria were considered for downstream analysis: More than 3,500 total allelic reads (sum of mus and cas), a minimum of 25 allelically expressed genes as well as a minimum of 3% of total allelic reads from either genotype. Moreover, a gene was considered informative if the sum of its allelic reads was higher than 10 and if it was expressed in at least 25% of cells. This resulted in 379 cells that passed all our quality control steps.

**In vivo single-cell analysis**

Allelic single-cell data of E6.5 embryos were obtained from GEO GSE169201 (Cheng et al., 2019; Data ref: Deng & Cheng, 2019), non-allelic analysis was performed as described above for in vitro. Processing of allelic data was performed for all cells that passed the Seurat QC. Cells that passed the following criteria were considered for downstream analysis: 400 total allelic reads (sum of cas and C57) and a minimum of 25 allelically expressed genes. Moreover, a gene was considered informative if the sum of its allelic reads was higher than 10 and if it was expressed in at least 20% of cells. This gave 239 cells in total and 16,003 informative genes.

Single-cell data of in vivo female germ cells were obtained from GEO GSE130212 (Zhao et al., 2020a; Data ref: Zhao et al., 2020b).

Non-allelic analysis was performed as described above for in vitro with the following exceptions: low-quality cells with less than 2,000 identified genes, less than 2,000 RNA molecules were removed. To ensure that the analysis would not be confounded by somatic cells, we focussed it on germ cells by subsetting for cells with a normalized and scaled Dazl expression > 0.5 and removed somatic cells by subsetting for cells with expression < 0.5 for Nr5a1, Axl, Fosb, Emx2 and Gata4. This gave 18,417 cells in total and 14,133 informative genes.

**RNA velocity analysis**

Non-allelic-specific RNA velocity analysis was performed as follows: Briefly, loom files only of Dazl-positive cells were generated from the non-allelic-specific BAM files from STAR using velocyto run-smartseq2 version v0.17.17 using the default parameters, mouse genome assembly mm10 and the UCSC repeat genome masked regions using custom-made scripts.

Subsequently, the loom files were imported into Python version 3.7 and processed using scVelo v0.2.3 (Bergen et al., 2020). The metadata, the clusters and the UMAP dimensionality reduction coordinates from Seurat were imported, then the single-cell data were filtered and normalized with a minimum of 20 counts and 2,000 top genes. The moments for velocity estimations were computed with 20 principal components and 30 neighbours. The genes’ full splicing kinetics were recovered before estimating the velocities using the dynamical model. The RNA velocity was visualized using velocity_embedding_stream colour coding cells by their Seurat cluster.

**Integration with in vivo datasets**

Single-cell data of in vivo female germ cells (Zhao et al., 2020) were analysed as described above. Normalized and scaled in vivo and in vitro data from this study were merged by canonical correlation analysis (CCA) using the Seurat function RunCCA. UMAP was then performed using CCA.

**Statistical analysis**

Statistical analysis of replicate data was performed using appropriate strategies in R. Number of independent experiments (n), type of statistical test, definition of significance and measurements are defined in figure legends. No statistical method was used to predetermine sample size and no data were excluded from the analyses. Samples were not randomized and investigators were not blinded to group allocation during data collection and analysis.

**Data availability**


**Expanded View for this article is available online.**

**Acknowledgements**

We are grateful to present and previous B.P. lab members for discussions and input. We thank L. Velten for advice on single-cell RNA-seq analysis; F. Nakaki for input on integration of transcription factor vectors and H. Ohta for help in...
establishing the m220 expansion system. We furthermore acknowledge critical technical support by the CRG core facilities including the CRG Genomics Unit, the CRG/UPF FACS Unit; the Bioinformatics Unit; the Advanced Light Microscopy Unit and the PRBB Animal Facility. This work was supported by the Spanish Ministry of Science, Innovation and Universities (BFU2014–55275–P, BFU2017–88407–P to B.P.), the Agencia Estatal de Investigación (AEI) (EURO2019–103817 to B.P.), the AXA Research Fund (to B.P.) and the Agencia de Gestio d’Ajuts Universitaris i de Recerca (AGAUR, 2017 SGR 346 to B.P.). We would like to thank the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership and to the “Centro de Excelencia Severo Ochoa.” We also acknowledge support of the CERCA Programme of the Generalitat de Catalunya. N.A. is supported by an EMBO postdoctoral fellowship (LTF 695–2019), J.S. and M.B. were supported by La Caixa International PhD Fellowships and J.S. by a travel grant from the Company of Biologists (Development Journal). P.A. has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 751651.

Author contributions
Jacqueline Severino: Conceptualization; Data curation; Software; Formal analysis; Investigation; Visualization; Methodology; Writing—original draft; Writing—review & editing. Moritz Bauer: Conceptualization; Data curation; Software; Formal analysis; Investigation; Visualization; Methodology; Writing—original draft; Writing—review & editing. Tom Mattimoe: Data curation; Formal analysis; Investigation; Methodology; Writing—review & editing.

Niccolo Arrecco: Data curation; Software; Formal analysis; Investigation; Methodology; Writing—review & editing. Luca Cazzuto: Software; Methodology. Patricia Lorden: Investigation. Norio Hamada: Resources; Methodology. Yoshiaki Nosaka: Resources; Methodology. So I Nagaoka: Resources; Methodology. Pauline Audergon: Methodology. Antonio Tarruell: Methodology. Holger Heyn: Resources; Supervision; Methodology. Katsumi Hayashi: Resources; Supervision; Methodology. Mimitori Saitou: Resources; Supervision; Methodology. Bernhard Payer: Conceptualization; Resources; Supervision; Funding acquisition; Writing—original draft; Writing—review & editing.

In addition to the CReditAuthor contributions listed above, the contributions in detail are:
BP, JS and MB conceived the study and wrote the manuscript with input from TM, JS and TM performed experiments. MB and JS established XRep cell line. JS, MB and NA performed bioinformatic analyses. LC wrote the bioinformatic methods. Bernhard Payer: Conceptualization; Resources; Supervision; Funding acquisition; Writing—original draft; Review & editing.

Disclosure and competing interests statement
HH is co-founder of Omniscope and SAB member of MiRXES.

References

© 2022 The Authors
The EMBO Journal 41: e109457 | 2022 21 of 23


Payer B, Lee JT (2014) Coupling of X chromosome reactivation with the pluripotent stem cell state. RNA Biol 11: 798 – 807

License: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.