eLife

Reviewed Preprint
v2 • January 29, 2025
Revised by authors

Reviewed Preprint
v1 • May 21, 2024

Developmental Biology

Single-nucleus multiomics reveals the gene-regulatory networks underlying sex determination of murine primordial germ cells

Adriana K Alexander, Karina F Rodriguez, Yu-Ying Chen, Ciro M Amato, Martin A Estermann, Barbara Nicol, Xin Xu, Humphrey Hung-Chang Yao

Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States • Epigenetics & Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Durham, United States

eLife Assessment

This **important** study reports single-nucleus multiomics-based profiling of transcriptome and chromatin accessibility of mouse XX and XY primordial germ cells (PGCs). The main conclusions of this study, which will be of interest to developmental and reproductive biologists, as well as andrologists, are supported by **convincing** data.

https://doi.org/10.7554/eLife.96591.2.sa4

Abstract

Accurate specification of female and male germ cells during embryonic development is critical for sexual reproduction. Primordial germ cells (PGCs) are the bipotential precursors of mature gametes that commit to an oogenic or spermatogenic fate in response to sex-determining cues from the fetal gonad. The critical processes required for PGCs to integrate and respond to signals from the somatic environment in gonads are not understood. In this study, we developed the first single-nucleus multiomics map of chromatin accessibility and gene expression during murine PGC development in both XX and XY embryos. Profiling of cell-type specific transcriptomes and regions of open chromatin from the same cell captured the molecular signatures and gene networks underlying PGC sex determination. Joint RNA and ATAC data for single PGCs resolved previously unreported PGC subpopulations and cataloged a multimodal reference atlas of differentiating PGC clusters. We discovered that regulatory element accessibility precedes gene expression during PGC development, suggesting that changes in chromatin accessibility may prime PGC lineage commitment prior to differentiation. Similarly, we found that sexual dimorphism in chromatin accessibility and gene expression increased temporally in PGCs. Combining single-nucleus sequencing data, we computationally mapped the cohort of transcription factors that regulate the expression of sexually dimorphic genes in PGCs. For example, the gene regulatory networks of XX PGCs are enriched for the transcription factors, TFAP2c, TCFL5, GATA2, MGA, NR6A1, TBX4, and ZFX.

Sex-specific enrichment of the forkhead-box and POU6 families of transcription factors was also observed in XY PGCs. Finally, we determined the temporal expression patterns of WNT, BMP, and RA signaling during PGC sex determination, and our discovery analyses identified potentially new cell communication pathways between supporting cells and PGCs. Our results illustrate the diversity of factors involved in programming PGCs towards a sex-specific fate.

Introduction

Proper formation of germ cells during embryonic development is crucial for ensuring the production of functional gametes[1]. The embryonic precursors to mature gametes, primordial germ cells (PGCs), are the bipotential stem cells of the germline that give rise to eggs and sperm[1,2]. During embryonic development in mice, PGCs commit to the oogenic or spermatogenic lineage in response to sex-determining cues from the gonadal environment in a process termed PGC sex determination[3]. During sex determination, XX PGCs in the fetal ovary commit to oogenesis by entering meiosis immediately after pre-granulosa cell specification[4]. By contrast, XY PGCs in the fetal testis initiate the spermatogenic program and arrest mitotically in response to signals from Sertoli cells[3,5]. Defects in germ cell differentiation, particularly during fetal life, often lead to reproductive diseases, such as infertility[6] and the formation of germ cell tumors[7]. Thus, it is necessary to enhance our understanding of germ cell development so we can better identify the etiologies of reproductive dysfunction in humans.

PGC sex determination is induced by the sex-specific activation of transcription factors (TFs) and downstream gene networks[8]. In XX PGCs, the retinoic acid (RA)-responsive TFs STRA8 and MEIOSIN and the bone morphogenetic protein (BMP)-responsive TF ZGLP1 are required for entry into meiosis and oogenesis[8–10]. These TFs also initiate the expression of genes related to meiotic processes, including *Rec8* and *Sycp1-3* [8]. In contrast, XY PGCs require the expression of cell cycle inhibitors, such as *Bnc2* and *Cdkn2b,* and the male-specific genes, *Nanos2* and *Dnd1*, to enter mitotic arrest[11–14]. During the transition from PGC to oogonium or gonocyte, both XX and XY PGCs must also lose their bipotential state by downregulating the pluripotency-related genes, *Pou5f1*, *Nanog*, and *Sox2*[8]. Consequently, there are multiple layers of gene regulation required for sex determination of PGCs.

Beyond the patterns of gene expression in PGCs, relatively little is known about how signals from the gonadal environment activate the expression of sexually dimorphic TFs and genes in PGCs. First, the gene regulatory networks, *i.e.* TFs and their predicted target genes, specific to XX and XY PGCs are not well defined. Second, it remains unclear how the chromatin environment is temporally regulated during PGC sex determination. Finally, additional data are needed on the patterns of ligand-receptor expression in gonadal supporting cells and PGCs. Previous reports have used bulk gene regulation and expression genomics assays to investigate the transcriptional programs underlying PGC development[15–18]. However, these assays may not have the resolution or sensitivity required to detect the transient changes in gene regulation among PGC subpopulations that are essential to sex determination.

In the present study, we employed the integrative genomics method, combined single-nucleus transcriptome and chromatin accessibility sequencing from the same cell, to decipher various layers of gene regulation during PGC sex determination in mice. We comprehensively profiled 3,054 XX and XY PGCs at embryonic days (E) E11.5, E12.5, and E13.5, which covers the developmental time frame from bipotential to sexually differentiated PGCs. Single-nucleus sequencing enabled the detection of sex-enriched regulatory loci, TFs, and gonadal cues that may be responsible for initiating gene expression in individual PGCs. We first systematically examined the molecular signatures of PGC subpopulations to identify the genes and accessible chromatin regions underpinning the sex-specific fates of PGCs. By combining our epigenomic and

transcriptomic data, we predicted *cis*-regulatory elements and sex-enriched TFs to construct the gene regulatory networks unique to XX and XY PGCs. Lastly, we probed the cell-cell communication pathways between supporting cells and PGCs to nominate potentially new ligand-receptor pairs involved in PGC development. Our results provide insights into the cell fate decisions underlying gametogenesis and sex determination of PGCs.

Materials and Methods

Animals

The single-nucleus multiomics experiments described herein used homozygous Rosa-tdTomato9 females (JAX 007909) (B6.Cg-Gt(ROSA)26Sor<tm9(CAG-tdTomato)Hze>/J) crossed to homozygous Nr5a1-cre mice (B6D2-Tg(Nr5a1-cre)2Klp), provided by the late Dr. Keith Parker[19]. The immunofluorescence experiments described herein used mice on the C57Bl/6 and CD1 backgrounds. Female mice were timed-mated and the day of detection of vaginal plug was considered embryonic day (E) E0.5. Mice were housed on a 12 h light:dark cycle, temperature range 70–7411°F and relative humidity range from 40 to 50%. All animal studies were conducted in accordance with the NIH Guide for the Care and Use of Laboratory Animals and approved by the National Institute of Environmental Health Sciences (NIEHS) Animal Care and Use Committee.

Single-nucleus RNA-sequencing and ATAC-sequencing

Whole ovaries and testes were collected at E11.5, E12.5, and E13.5. Plug date was used to determine the stage of embryos collected for single-nucleus RNA-seq and ATAC-seq. The stage of E11.5 embryos was confirmed by counting somites. The stage of embryos collected at E12.5 was confirmed by the morphological presence of the vessel and cords of the testes collected from XY embryos. Similarly, we confirmed the stage of embryos collected at E13.5 by the size of the gonads, the presence of more distinct cords in the testes of XY embryos, and the elongation of the ovaries of XX embryos. The sexes of the embryos were determined using amnion staining, as previously published[20], to identify the presence (XX) or absence (XY) of the Barr body and confirmed with PCR using primers that recognize the Y chromosome. Tomato fluorescence was used to facilitate the dissection of the gonad away from the mesonephros, which was done using needles. Isolated fetal mouse gonads were isolated in PBS, and pairs of gonads or groups pooled between fetuses of the same sex and age were immediately frozen in liquid nitrogen for storage. Nuclei isolation was performed using the 10X Genomics Chromium Nuclei Isolation Kit (catalog #1000494) following the manufacturer's protocol. The total numbers of gonads used per sex and age are listed in **Supplementary Table 1**. The 10X Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Library Preparation Kit (catalog #1000284) was used to generate a 10X barcoded library of mRNA and transposed DNA from individual nuclei. The 10X barcoded single-nuclei RNA and ATAC-seq libraries were obtained from 8,687 average (1,555-20,000) nuclei per sample and sequenced with Illumina NOVAseq to a minimum sequencing depth of 129,120,442 raw reads and 22,406 read pairs/nucleus (**Supplementary Table 1**). Two independent technical replicates of multiome sequencing were performed for each sex by embryonic stage combination. The two independent technical replicates comprised different pools of paired gonads.

Data preprocessing for single-nucleus RNA-seq and ATAC-seq libraries

CellRanger [21,22] (v3.0) was used for count, alignment, filtering, and cell barcode and UMI counting (**Supplementary Table 1**). Barcode swapping correction was performed for all libraries[23]. FASTQ files of the corrected cell count matrices were generated and the data were analyzed using *Seurat* [24] (v. 4.3.0) and *Signac* [25] (v. 1.6.0) on R (v. 4.1.0).

Supplementary Table 1.

Sequencing and QC statistics for single-nucleus multiome libraries of E11.5-E13.5 XX and XY gonads.

Sample	Pooled Paired Gonads (Number)	Raw Reads	Barcode	Mapped	Saturation	Cells	Reads/cell	Median UMI/cell	Median genes/cell
E11.5 F_R1	19	448,115,792	93%	91%	38%	20,000	22,406	3,665	2,030
E11.5 F_R2	11	812,813,546	94%	92%	68%	10,008	81,216	8,184	2,798
E11.5 M_R1	18	193,672,880	93%	94%	35%	8,080	23,969	2,790	1,708
E11.5 M_R2	12	719,219,983	94%	91%	64%	4,960	145,326	5,934	3,512
E12.5 F_R1	12	498,266,895	93%	94%	43%	18,696	26,651	4,620	2,422
E12.5 F_R2	11	654,449,771	94%	91%	66%	4,949	131,946	8,351	3,504
E12.5 M_R1	8	425,422,729	93%	91%	44%	14,897	28,558	5,268	2,682
E12.5 M_R2	10	951,070,062	94%	92%	74%	5,864	162,188	9,194	3,446
E13.5 F_R1	3	135,990,358	95%	92%	86%	1,555	83,036	3,608	2,121
E13.5 F_R2	5	398,709,188	95%	92%	73%	6,369	74,137	6,013	2,831
E13.5 M_R1	3	129,120,442	94%	94%	82%	2,040	66,662	5,132	2,642
E13.5 M_R2	7	472,180,945	94%	91%	71%	6,597	60,438	6,546	2,969

snRNA-seq data preprocessing

Doublets within each dataset were removed with the *scDblFinder* [26] (v. 1.8.0) package using default methods. Ambient RNA was removed using the *celda/decontX* [27] (v. 1.12.0) package, with the maximum iterations of the EM algorithm set at 100. Datasets of gonadal cell populations from each embryonic stage and sex were combined into one *Seurat* object using the "merge" function. As batch effect was not observed in our dataset, we did not perform further integration or batch correction. Cells with nCount_RNA > 1000 & nCount_RNA < 25,000& percent.mt <25 were retained for downstream analyses (**Supplementary Figure 1a**). Mitochondrial genes and transcripts were removed from the snRNA-seq datasets to eliminate any potentially contaminating signal. The data was normalized using the "SCTransform" function in *Seurat.*

snATAC-seq data preprocessing

snATAC-seq peak calls from *CellRanger* were used to generate one combined peak set and cells with peak size > 20 bp & peak size < 10,000 bp were retained for downstream analyses. Fragment objects were created for each library, combined with the respective *Seurat* object, and analyzed using the *Signac* package. A subset of the snATAC-seq data was generated using the following cutoffs: nCount_ATAC < 100,000 & nCount_ATAC > 1000 & nucleosome_signal < 2 & TSS.enrichment > 1 (**Supplementary Figure 1b-d**). Analysis of the snATAC-seq peak feature distribution demonstrated that most peaks occurred in distal intergenic or promoter regions of the genome (**Supplementary Figure 1e**).

Sex filtering

To ensure that the correct sex was analyzed in all downstream analyses, we applied computational filters to confirm the sex of the cells in each dataset. Cells were filtered based on Y-linked gene expression and fragments mapped to the Y chromosome (chrY). The *UCell* package[28] (v. 2.0.1) was used to create a module score of the Y-linked genes *Kdm5d, Eif2s3y, Uty,* and *Ddx3y*. Module scores of E11.5 and E12.5 cells above one standard deviation of the mean module score of all XY E13.5 cells were determined to be chromosomally male. Module scores of E11.5 and E12.5 cells below one standard deviation of the mean module score of all XX E13.5 cells were determined to be chromosomally female. Next, we examined the peaks called on chrY outside of the pseudoautosomal region. In total, 21 peaks were identified between chrY 1-90,000,000 bp. Cells with 0 fragment counts within our defined chrY peak region or chrY gene expression module score < 1 standard deviation were removed from the male datasets. Cells with > 1 fragment counts within our defined chrY peak region or chrY gene expression module score > 1 standard deviation were removed from the female datasets.

Cell type identification

To identify the cellular origin of the snRNA-seq clusters, principal component analysis (PCA) was run on the scaled data and visualization was performed using UMAP with dims = 18 and resolution = 0.3. Primordial germ cell (PGC) cluster identification was performed using the established germ cell markers *Ddx4* and *Pou5f1*[29]. To remove potential PGC and gonadal somatic cell doublets, only cells with RNA count < 1 for the following genes were retained as PGCs: the granulosa cell marker genes *Foxl2* and *Runx1* [30]; Sertoli cell marker *Sox9*[30]; Leydig cell marker *Insl3* [30]; stromal cell marker *Wt1*[31]; endothelial cell marker *Plvap* [32]; immune-related cell marker *Mafb*[33]; interstitial cell markers *Pdgfra* and *Nr2f2* [30]; supporting cell marker *Tspan8* [34]; and epithelial cell marker *Krt19* [35]. Somatic supporting cell cluster identification was performed using the established female and male support cell markers *Runx1* and *Sox9*, respectively. To remove potential gonadal supporting cell doublets formed with germ

cells or interstitial cells, only cells with RNA count > 1 for *Wt1* and RNA count < 1 for the following genes were retained as support cells: the germ cell marker genes *Ddx4* and *Pou5f1*; endothelial cell markers *Plvap* and *Pecam1*[36]; and interstitial cell markers *Pdgfra* and *Nr2f2*.

Clustering and pseudotime trajectory of primordial germ cells

snRNA-seq reclustering of PGCs was done with PCA on scaled PGC RNA data and visualization was performed using UMAP with dimensions (dims) = 18 and resolution = 0.3. snATAC-seq clustering of PGCs was done first by calling peaks using *MACS2*[37] (v. 2.2.7.1). Peaks on nonstandard chromosomes were removed using the "keepStandardChromosomes" function in *Signac* with the option 'pruning.mode = "coarse"' and peaks in mm10 genomic blacklist regions were removed. Counts in each peak were determined using *Signac* functionalities. The data was normalized by latent semantic indexing (LSI) using the "RunTFIDF", "FindTopFeatures", and "RunSVD" functions in *Signac*. Visualization was performed using UMAP with dims = 2:30, reduction = 'lsi', and algorithm = 3. The joint neighbor graph or joint UMAP was generated using the "FindMultiModalNeighbors" function in Signac and reduction.name = "wnn.umap". A Seurat object from the clustered PGCs was used as input for pseudotime analysis with Monocle3 to determine the differentiation trajectory of PGCs using both female and male E11.5 PGCs as time '0' in pseudotime.

Identification of differentially expressed genes and differentially accessible peaks

Differentially expressed genes (DEGs) in each cluster were determined using the "FindAllMarkers" function in *Seurat* based on the Wilcoxon Rank Sum Test. DEGs between female and male PGCs at each embryonic stage were identified with the "FindMarkers" function in *Seurat* with min.pct = 0.25 and logfc.threshold = 0.25. Differentially accessible peaks (DAPs) between female and male PGCs were identified with the "FindMarkers" function in *Signac* with min.pct = 0.001 and logfc.threshold = 0.1.

Correlation of peak accessibility with gene expression

snATAC-seq peaks were linked to gene expression, as imputed from the snRNA-seq data, using the "LinkPeaks" function in *Signac*. Peaks linked to DE genes for each sex and embryonic stage were then subsetted out for further downstream analyses. Plots showing peak-to-gene linkages were generated using the "CoveragePlot" function in *Signac*.

Motif enrichment analysis of transcription factor binding sites

DNA sequence motif analysis of peaks linked to DE genes was performed using the "FindMotifs" function in *Signac*. Motifs were obtained from the JASPAR 2020 database (collection = 'Core' and taxonomic group = 'vertebrates'). Background peaks were selected to match the GC content in the peak set by using the "AccessiblePeaks," "GetAssayData", and "MatchRegionStats" functions in *Signac*. *MotifScan*[38] (v. 1.3.0) was used to determine the genomic position of linked peaks to DE genes to identify predicted target genes of transcription factors. Per-cell motif activity was computed using the "RunChromVAR" function in *Signac*.

Module score calculation for gene expression

Module scores of gene expression were calculated using the "AddModuleScore" function in *Seurat*. Module scores were plotted on UMAPs using the "FeaturePlot" function in *Seurat*.

Peak feature distribution

Pie charts of peak feature distribution were generated with the R package *ChIPseeker* [39,40] (v. 1.38.0). Peaks were annotated with the "annotatePeak" function with options 'tssRegion= c(−3000,3000)', 'TxDb=TxDb.Mmusculus.UCSC.mm10.knownGene', and 'annoDb="org.Mm.eg.db". Pie charts were plotted with the function "plotAnnoPie".

GO annotation and enrichment

The GO enrichment analyses were carried out using the R package *clusterProfiler* [41] (v. 4.0.5) with the "enrichGO" function. The Benjamini-Hochberg method (a = 0.05) was used to control for False Discovery Rate. GO annotation enrichment for biological processes were plotted using the "dotplot" or "cnetplot" function of the *Enrichplot* [42] (v1.12.2) R package.

CistromeDB regulatory potential score

The CistromeDB Regulatory Potential Score[43,44] was used to identify potential regulators of *Bnc2* and *Stra8*. The toolkit (*http://dbtoolkit.cistrome.org*) was used for these analyses by selecting Mouse mm10 as 'Species', transcription factor/chromatin regulator as 'Data Type in Cistrome', and 10 kb as 'The half-decay distance to transcription start site.'

Enrichment of signaling pathways between gonadal support cells and primordial germ cells

CellphoneDB [45] (v. 5.0.0) was used to predict paired ligand-receptor pairs for cell-cell communication between gonadal supporting cells and PGCs. Normalized snRNA-seq counts were used as input data. Significant ligand-receptor pairs were determined independently between gonadal supporting cells and PGCs for each embryonic stage and sex. The statistical analysis (method 2) of *CellphoneDB* and the 'statistical_analysis_method' function in *CellphoneDB* were used to identify significantly enriched signaling pathways.

Immunofluorescence

Gonads were collected from E13.5 XX and XY embryos and fixed in 4% PFA for 1h at room temperature. Gonads were then washed in PBS and stored at 411°C until embedding in paraffin and sectioned at 5 μm thickness. Following deparaffinization, sections were rehydrated in decreasing alcohol concentration gradients. Antigen retrieval was performed with 0.1 mM citric acid (Vector Labs) for 20 min in the microwave at 10% power and then cooled to room temperature. Samples were blocked in PBS-TRITON X-100, 0.1% solution with 5% normal donkey serum for 1h. Samples were incubated with primary antibodies diluted in blocking buffer (PBS-TRITON X-100, 0.1% solution with 5% normal donkey serum) at 411°C overnight. Samples were then washed with PBS-TRITON X-100, 0.1% solution and incubated with secondary antibodies diluted in blocking buffer at room temperature for 1h. The slides were washed with PBS-TRITON X-100, 0.1% solution 3x and incubated with the Vector TrueView Autofluorescence Quenching Kit (cat. # SP-8400) for 2 min. Following incubation with the quenching kit, the slides were washed with PBS-TRITON X-100, 0.1% solution 2x, PBS 1x, and incubated with DAPI (Invitrogen cat. # D1306). The slides were then mounted with slide mounting media (Vectashield Vibrance Antifade Mounting Medium TrueView Kit SP-8400). Primary antibodies used included: Anti-TRA98 (1:500; rat, BioAcademia cat. # 73-003), Anti-NR2F2 (1:300, mouse, R&D Systems cat. # PP-H7147-00), Anti-PORCN (1:200, rabbit, ThermoFisher cat. # PA5-43423), and Anti-AP-2γ (TFAP2c) (1:200, mouse, Santa Cruz cat. # sc-12762). Secondary antibodies were used at 1:200 dilution (Invitrogen/Life Technology): Donkey anti-rat Alexa 568 (A78946), Donkey anti-mouse Alexa 647 (A31571), and Donkey anti-rabbit Alexa 488 (A21206). Imaging for sections was performed on a Zeiss LSM 900 confocal microscope using Zen software. Brightness and contrast of images were adjusted using ImageJ (National Institutes of Health, USA).

Statistical analyses

Libraries were prepared from 3-19 pairs of gonads and two technical replicates per embryonic stage and sex (**Supplementary Table 1**). Statistical analyses were considered significant if $p < 0.05$.

Results

Combined multiome analysis of single-nucleus gene expression and chromatin accessibility of mouse PGCs

To understand how intrinsic transcription networks inside PGCs and external cues from the gonadal environment control PGC sex determination, we performed combined snRNA-seq and snATAC-seq (10x Genomics Multiome) for all cell populations in mouse fetal gonads. This approach allowed us to obtain paired gene expression and chromatin status from the same cell for all cell populations from female (XX) and male (XY) gonads. Whole gonads from Nr5a1-cre x Rosa-tdTomato9 embryos were collected at embryonic days (E) E11.5, E12.5, and E13.5, which encompass the developmental time frame of sex determination-initiated (E11.5), sexually differentiating (E12.5), and morphologically dimorphic (E13.5) gonads. Sequencing libraries of gonads for each sex and embryonic day were generated from two independent technical replicates for an average 8,687 nuclei/sample with a sequencing depth of ~75,544 reads/nucleus and ~5,755 UMIs/nucleus. Based on the well-established germ cell markers *Ddx4*[29] and *Pou5f1*[29], we identified 3,054 germ cells. Somatic supporting cell populations were identified using *Runx1* expression as a marker of pregranulosa cells in our female datasets and *Sox9* expression as a marker of Sertoli cells in our male datasets[30]. A combined 22,382 XX and XY supporting cells were identified. Together, these data allowed us to determine the clustering patterns of differentiating PGCs, identify regulatory loci and transcription factors (TFs) enriched in PGCs, and find previously unreported interactions between PGCs and gonadal supporting cells (**Fig. 1a**).

We first asked whether single-nucleus gene expression and chromatin accessibility information can quantify the similarity or differences within PGC populations. We clustered single PGC nuclei in low-dimensional space based on their snRNA-seq (**Fig. 1b-d**) or snATAC-seq (**Fig. 1e-g**) data using Uniform Manifold Approximation and Projection (UMAP) for dimensional reduction. This graph-based clustering algorithm unbiasedly separated and clustered individual cells with similar transcriptomic profiles[46]. Such analyses revealed that the transcriptomes of E11.5 XX and XY PGCs overlap extensively, whereas E12.5 and E13.5 PGCs principally separated according to their sex (**Fig. 1b**). In addition, we uncovered eight broader clusters of PGCs at a resolution that captured PGC subpopulations (**Fig. 1c**). We determined the cell type composition of the snRNA-seq transcriptome-based clusters by plotting the percentage of PGCs at each embryonic stage and sex (*x*-axis) in each of the 8 transcriptome-based clusters (*y*-axis) (**Fig. 1d**). First, E11.5 and E12.5 XX PGCs primarily consisted of clusters 0, 1, and 4 (**Fig. 1c-d**). At E13.5, XX PGCs separated into three main clusters, 4, 5, and 6, and had significantly higher expression of the synaptonemal complex component and meiotic marker *Sycp1*[47] and the chromatin modifier *Hdac9* [48] (**Fig. 1c-d**; **Supplementary Figure 2**). In XY gonads, PGCs at E11.5 grouped together into three clusters, 0, 1, and 7 (**Fig. 1d**). At E12.5-E13.5, XY PGCs converged onto a single distinct population (cluster 7), indicating less transcriptional diversity among E12.5-E13.5 XY PGCs when compared to E12.5-E13.5 XX PGCs (**Fig. 1d**). Furthermore, we found that E12.5-E13.5 XY PGCs clustered separately from other PGC populations due to elevated expression of *Bnc2*, which is required for mitotic arrest in PGCs[11], and the forkhead-box TF *Foxp1*[49] (**Supplementary Figure 2**). Finally, a small proportion of all PGC populations (E11.5-E13.5 XX and XY PGCs) were detected in clusters 2 and 3, and displayed enriched expression of ribosome biogenesis genes and genes with reported roles in reproduction, such as the mitochondrial leucyl-tRNA synthetase *Lars2* [50] and the DNA

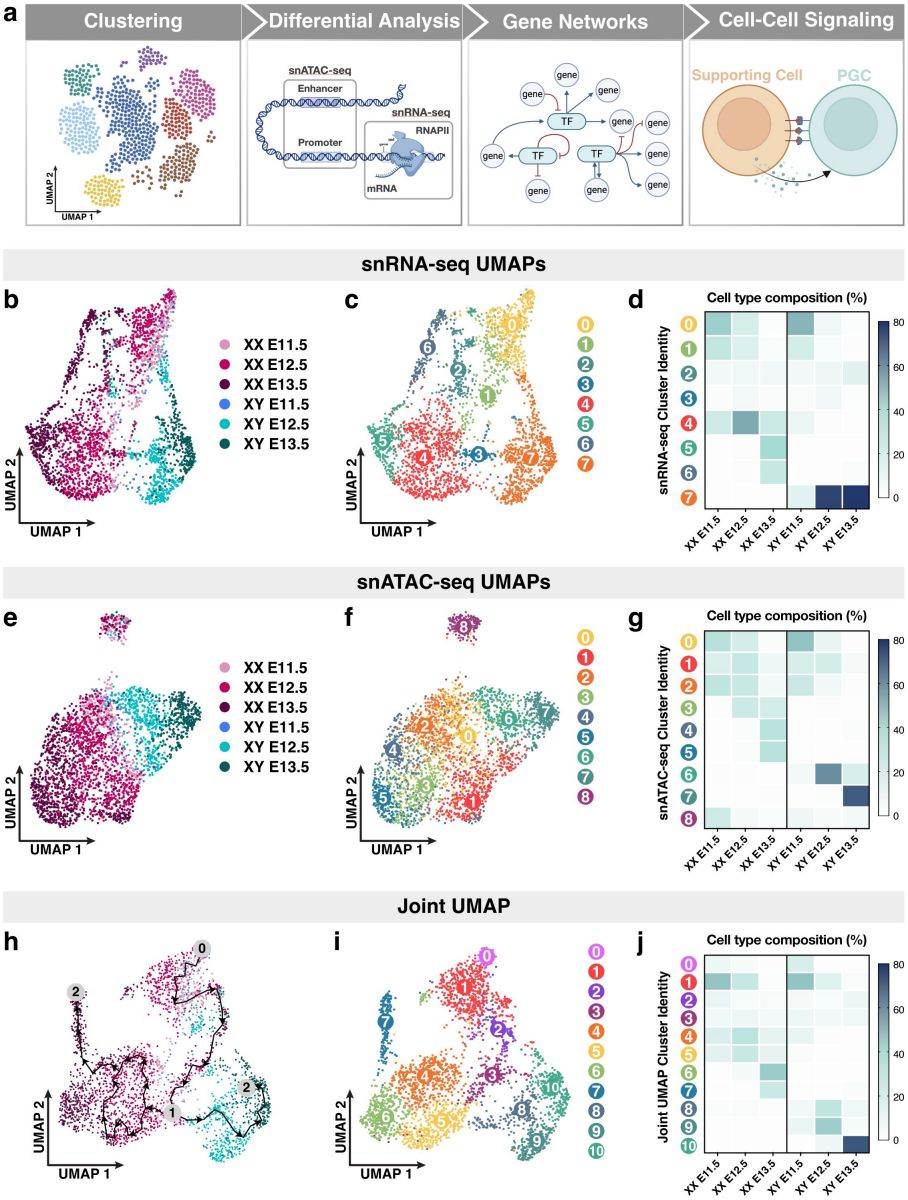

Figure 1

Single-nucleus multiomics sequencing of E11.5-E13.5 XX and XY PGCs.

(a) Workflow of single-nucleus multiomics analyses. (b-c) Dimensional reduction (UMAP) and cell-type specific clustering of snRNA-seq data collected from E11.5-E13.5 XX and XY PGCs. Each dot represents an individual cell color coded by either embryonic stage and sex (a) or cluster number (c). (d) Cell type composition of snRNA-seq clusters. The embryonic stage and sex of the PGC population is indicated on the *x*-axis and the unbiased clustering number is represented on the *y*-axis. (e-f) UMAP of snATAC-seq data collected from E11.5-E13.5 XX and XY PGCs color coded by either embryonic stage and sex (e) or cluster number (f). (g) Cell type composition of snATAC-seq clusters. (h-i) UMAP visualization of E11.5-E13.5 XX and XY PGCs using a joint, or weighted, measurement of snRNA- and snATAC-seq modalities. (h) Trajectory of transcriptomic changes (also known as pseudo-time analysis in Monocle) among PGCs overlaid on the joint UMAP. Arrows indicate the path of the developmental trajectory. Numbers on the pseudotime trajectory indicate the following: '0' represents the starting cell population; '1' indicates the branching point in the trajectory path; and '2' represents the terminal cell populations of the trajectory paths. Each PGC is color coded by embryonic stage and sex. (i) Unbiased clustering of PGCs using combined snRNA- and snATAC-seq data. Each PGC is color coded by cluster number. (j) Cell type composition of joint UMAP clusters. The number of PGCs per sex and embryonic stage are: 375 E11.5 XX PGCs; 1,106 E12.5 XX PGCs; 750 E13.5 XX PGCs; 110 E11.5 XY PGCs; 465 E12.5 XY PGCs; and 348 E13.5 XY PGCs.

repair protein *Mgmt*[51] (**Fig. 1d**; **Supplementary Figure 2**). Together, these data indicate that the transcriptomic profiles of PGCs are largely homogenous within each sex at E11.5. Subsequently, XX PGCs maintain their transcriptional heterogeneity at E12.5-E13.5, whereas XY PGCs transcriptionally converge on a single identity beginning at E12.5.

The snRNA-seq clustering revealed transcriptomic characteristics of PGCs. Next, we performed unbiased clustering of PGCs according to their aggregate profiles of snATAC-seq peaks (210,666 total peaks) to detect the patterns of chromatin accessibility associated with PGC development (**Fig. 1e and 1f**). The snATAC-seq-based clustering of PGCs showed that the chromatin status of E11.5 XX and XY PGCs was highly similar (**Fig. 1e**). By contrast, E12.5-E13.5 XX and XY PGCs formed distinct groupings at opposite poles of the UMAP (**Fig. 1e**). Unbiased graph-based clustering of PGCs uncovered 9 clusters of PGCs with unique stage- and sex-specific snATAC-seq peak profiles (**Fig. 1f**). First, E11.5-E12.5 XX and XY PGCs together predominantly consisted of accessibility-based clusters 0, 1, and 2, indicating that these germ cells share a highly similar chromatin profile prior to differentiation (**Fig. 1f and g**). Second, E13.5 XX PGCs were separated into clusters 3, 4, and 5 (**Fig. 1f and g**), which may be due to differences in meiotic entry or progression within these populations of germ cells. Third, E12.5-E13.5 XY PGCs comprised snATAC-seq clusters 6 and 7, suggesting that XY PGCs may adopt a sex-specific chromatin status beginning at E12.5 (**Fig. 1f and g**). Finally, cluster 8, which contained PGCs from both sexes at each stage, was separate from all other accessibility-based clusters (**Fig. 1f and g**). Taken together, our findings indicate that the chromatin accessibility profiles of E11.5 PGCs and several E12.5 PGC subpopulations do not show a sex bias, whereas E13.5 XX and XY PGCs clustered by sex.

The multiome datasets provided us the opportunity to define PGC identities based on a more biologically relevant status at the combined transcriptomic and chromatin accessibility levels. Specifically, we integrated gene expression and chromatin accessibility data within a single PGC to obtain a joint, or multi-modal, definition of PGC identity (**Fig. 1h and i**). Computational integration methods allowed us to co-visualize the variation in both the snRNA-seq and snATAC-seq profiles of PGCs on a single UMAP[52]. Simultaneously profiling cell-type specific transcriptomes and regions of open chromatin for single PGCs resolved the sex-specific trajectories of PGCs with much greater resolution (**Fig. 1h and i**). In fact, we observed a strong demarcation between XX and XY PGCs beginning at E12.5 with very little to no overlap between XX and XY PGCs at E13.5 (**Fig. 1h-j**). Given that we detected the clearest separation of PGCs when clustering with multi-modal data, we analyzed the trajectory of transcriptomic changes (also known as pseudotime analysis in Monocle[53]) among PGCs overlaid on the joint UMAP (**Fig. 1h**). With E11.5 XX and XY PGCs chosen as the starting point (0) of the trajectory analysis, we observed a distinct branching or 'decision' point (1) between XX and XY PGCs at E12.5 (**Fig. 1h**). Additionally, our unbiased analysis identified two terminal states (2) of PGCs at the E13.5 XX PGC cluster and E13.5 XY PGC cluster, providing additional confirmation that our data accurately reflect the status of PGC sex determination (**Fig. 1h**). We also observed a closed, or circular, branch of the trajectory path among E12.5 XX PGCs (**Fig. 1h**), which may arise from the heterogeneous transcriptomes of female PGCs entering meiosis in an asynchronous manner[54]. These data demonstrate that our approaches enable us to not only visualize the developmental trajectories of PGCs that are clustered with multi-modal data, but also empirically trace their sex-specific fates based on changes in gene regulation.

Clustering of PGCs using combined snRNA- and snATAC-seq data identified eleven distinct groups of PGCs and yielded enhanced cluster resolution of PGC subpopulations (**Fig. 1i**). This analysis resulted in clusters of PGCs that were predominantly sex-specific and enabled the detection of subpopulations of each PGC type (**Fig. 1i**). To better understand the population structure of individual clusters, we plotted the percentage of each PGC type (*x*-axis) for each cluster identified in **Fig. 1i** (*y*-axis) (**Fig. 1j**). First, E11.5 XX PGCs clustered into one main population, cluster 1, and later diverged into six subpopulations, clusters 0 through 5 at E12.5 (**Fig. 1i-j**). Clusters 0 - 2 were more similar to E11.5 PGCs, whereas clusters 3 – 5 were composed of E12.5 XX PGCs that

were developmentally closer to E13.5 XX PGCs (**Fig. 1i-j**). Finally, we found that E13.5 XX PGCs converged onto either one of two developmental fates in clusters 6 and 7, with cluster 7 exhibiting a large spread along the UMAP2 axis (**Fig. 1i-j**). These findings indicate that E12.5-E13.5 XX PGCs may require extensive chromatin remodeling and transcriptional reprogramming to adopt the oogenic fate. By comparison, XY PGCs grouped into clusters 0 and 1 at E11.5 and appeared to converge on a single developmental fate at later stages, with E12.5 XY PGCs comprising clusters 8 and 9 and E13.5 XY PGCs belonging to cluster 10 (**Fig. 1i-j**). Given that neither expression- nor accessibility-based clustering alone resolved the granularity of PGC populations to the level described here, these data underscore the empirical power of simultaneously measuring multiple modalities from the same cell to define the developmental fate of PGCs.

Molecular characterization of genes known for their roles in XY PGC differentiation

We next asked whether our multiomics approaches could gain insight into the regulatory status of genes that are known to be essential for PGC sex determination. We first investigated the patterns of gene expression and chromatin accessibility of male functional genes, *i.e.* those that have been shown to be functionally important for XY PGC development (**Fig. 2**). We found that genes involved in XY PGC development (*i.e., Rb1*, *Rbl2*, *Cdkn1b*, *Cdkn2b*, *Bnc2*, *Cnot1*, *Dnd1*, and *Nanos2*)[11–14,55] showed increasing levels of expression in XY PGCs from E11.5 to E13.5, whereas the expression levels of *Cnot1* and *Nanos3* peaked at E12.5 and E11.5, respectively (**Fig. 2a**). Notable XY-enrichment of *Bnc2* expression and moderately elevated expression of *Rb1*, *Rbl1*, *Cdkn1b*, *Cdkn2b*, *Nanos2*, and *Nanos3* were observed in XY PGCs compared to XX PGCs (**Fig. 2a**). However, expression of *Cnot1* and *Dnd1* showed an XX biased expression pattern (**Fig. 2a**). Overall, these data indicate that the majority of male functional genes show a modest increase in expression levels in XY PGCs over XX PGCs. Among these genes, *Bnc2* expression may serve as a definitive marker of XY PGCs at E13.5.

To identify potential mechanisms regulating the expression of XY PGC functional genes, we searched for candidate transcription factors (TFs) with XY-enriched gene expression. We first linked snATAC-seq peaks to XY PGC functional genes (listed previously), hereafter termed peak-to-gene linkages, in XY PGCs. Peak-to-gene linkages were determined using Signac functionalities and were derived from the correlation between peak accessibility and the intensity of gene expression. Next, we performed a TF motif enrichment analysis on significant peak-to-gene linkages to identify potential TFs. Finally, we compared the expression levels of potential TFs between XX and XY PGCs to identify TFs with XY-enriched gene expression. This strategy uncovered the Krüppel-like zinc finger family member ZKSCAN5 (Zinc finger with KRAB and SCAN domains 5) as a potential regulator of XY PGC functional genes (**Fig. 2a**). *Zkscan5* showed XY biased expression (e.g., there is a 1.5-fold increase in the expression of *Zkscan5* between XY and XX PGCs at E13.5), and the level of *Zkscan5* expression increased from E11.5 to E13.5 in XY PGCs (**Fig. 2a**). In addition, the ZKSCAN5 motif was not enriched in XX PGCs. ZKSCAN5 is not only highly expressed in the adult testis[56] but is also proposed to regulate cellular proliferation and growth[57]. Thus, although little is known about the function of ZKSCAN5, we hypothesize ZKSCAN5 may possess a regulatory role during male germ cell development. These data highlight the strength of single-nucleus multiomics in discovering candidate factors that may regulate the expression of a cohort of developmental genes and, consequently, direct cell fate decisions.

Among the functionally important PGC genes, *Bnc2* expression was most strongly enriched in XY PGCs. We therefore investigated the regulation of *Bnc2* at the chromatin level as an example to demonstrate the utility of combined multiome datasets. The chromatin accessibility of the *Bnc2* promoter was highest in XY PGCs and increased temporally from E11.5 to E13.5 (**Fig. 2b**). Furthermore, the level of chromatin accessibility at the *Bnc2* locus was positively correlated with *Bnc2* expression in XY PGCs (**Fig. 2b-c**). We then applied the Cistrome Data Browser (Cistrome DB)[44] to identify putative regulators of *Bnc2* expression based on transcription factor binding

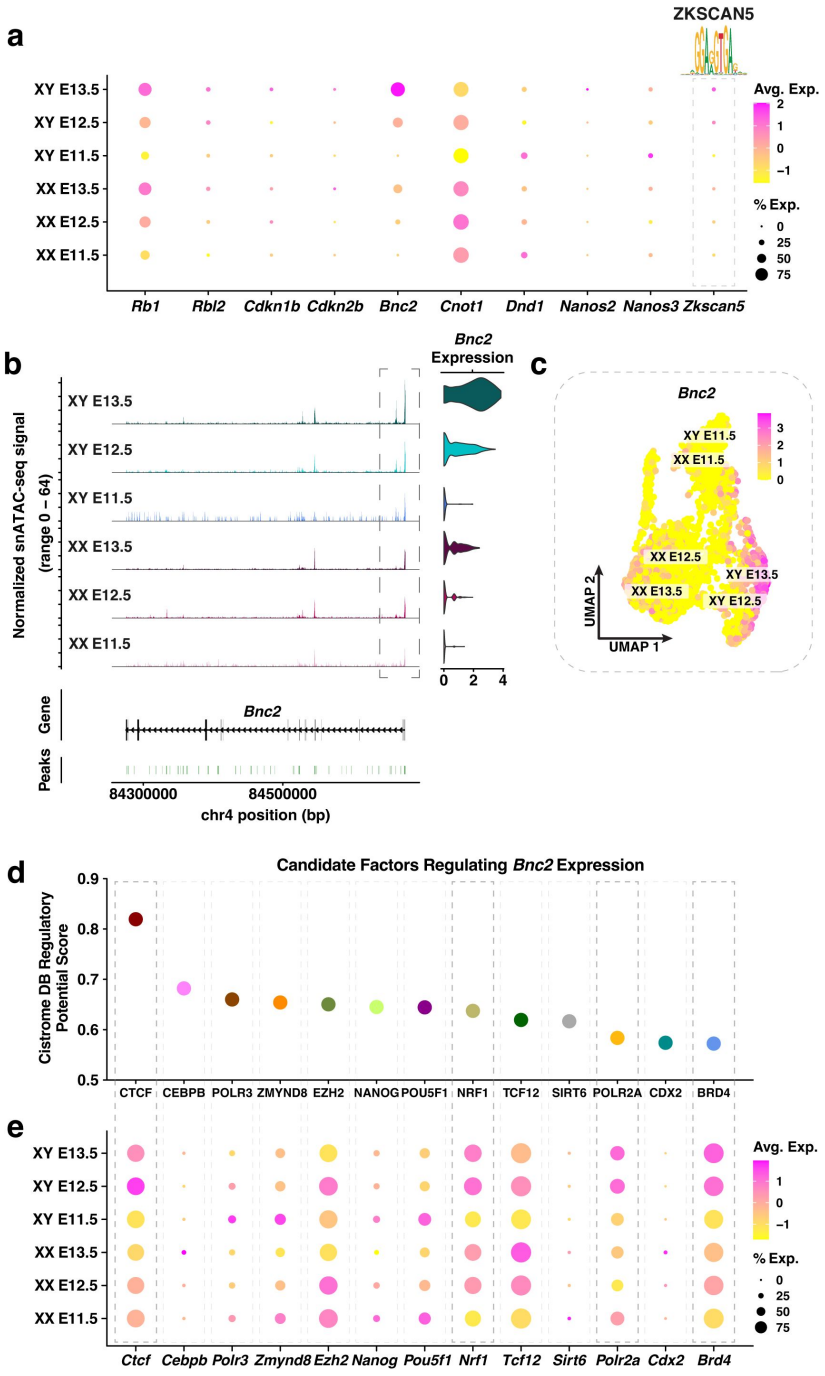

Figure 2

Molecular characterization of genes known for their roles in male PGC differentiation.

(a) Dotplot of the average expression and percentage of cells expressing *Rb1, Rbl2, Cdkn1b, Cdkn2b, Bnc2, Cnot1, Dnd1, Nanos2, Nanos3,* and *Zkscan5.* The DNA binding motif for ZKSCAN5 is indicated above. The color scale represents the average expression level, and the size of the dot represents the percentage of cells expressing the gene. (b) *Left:* Coverage plot of the normalized snATAC-seq signal at the *Bnc2* locus. *Right:* Violin plot of *Bnc2* expression in E11.5-E13.5 XX and XY PGCs. (c) Joint UMAP showing *Bnc2* expression levels for E11.5-E13.5 XX and XY PGCs. Individual cells are color coded by the level of *Bnc2* expression. (d) Candidate factors (*x*-axis) regulating *Bnc2* expression based on their Cistrome DB regulatory potential score (*y*-axis). (e) Dotplot of the average expression and percentage of cells expressing the candidate factors that potentially regulate *Bnc2* expression. The color scale represents the average expression level, and the size of the dot represents the percentage of cells.

motifs detected within 10 kb of the *Bnc2* promoter (**Fig. 2d**). The top candidate regulators of the *Bnc2* locus were determined according to a regulatory potential score derived from a combination of TF binding motif information and previously published genomics data (*i.e.* ChIP-seq, DNase-seq, and ATAC-seq) (**Fig. 2d**). Of the top candidate regulators from the Cistrome DB analysis, expression of *Ctcf, Nrf1, Polr2a,* and *Brd4* was enriched in XY PGCs at E13.5 (**Fig. 2e**). In particular, BRD4 is known to activate transcription elongation by recruiting kinases to the largest RNA polymerase II (Pol II) subunit, POLR2A, thereby releasing Pol II from its promoter-proximal bound state[58]. Taken together, these findings provide insight into potential XY-specific regulatory mechanisms at the *Bnc2* locus in PGCs.

Characterization of genes known for their roles in XX PGC development

We next performed the same analyses of XX PGC functional genes, *i.e.* those that have been shown to be functionally important for XX PGC development (**Fig. 3a**). Overall, as expected, we found that E13.5 XX PGCs expressed genes involved in XX germ cell development (*i.e., Rec8, Sycp2, Sycp3, Ctnnb1, Meioc, Rnf2, Stra8, Ythdc2,* and *Zglp1)*[8] substantially higher than E13.5 XY PGCs (**Fig. 3a**). Moreover, *Rnf2* and *Zglp1* were strongly expressed at E12.5 in XX PGCs, and their expression occurred immediately prior to the activation of all other genes in the female cohort (**Fig. 3a**). Although *Zglp1* was expressed in fewer XX PGCs than *Rnf2*, *Zglp1* expression increased from E12.5 to E13.5 whereas *Rnf2* expression decreases (**Fig. 3a**). These data support previous studies that *Rnf2* and *Zglp1* control the timing of sexual differentiation of XX PGCs[10,59]. By comparison, the expression of *Ctnnb1* diminished in both XX and XY PGCs at E12.5, as expected given that negative regulation of WNT/β-catenin is required for XX PGCs to enter meiosis[60] (**Fig. 3a**). In sum, these results provide insight into the timing, expression levels, and pervasiveness of female gene activation in XX PGC populations.

We next identified potential regulators of XX PGC functional genes using the same approach for XY PGCs (**Fig. 2**). Our analysis identified significant enrichment of the binding motif for DMRT1 in the regulatory elements linked to XX PGC functional genes (**Fig. 3a**). DMRT1 has been shown to control *Stra8* expression in a sex-specific manner, such that *Stra8* is activated in the fetal ovary and repressed in the fetal testis[61]. Our snRNA-seq data also demonstrated *Dmrt1* expression in both XX and XY PGCs (**Fig. 3a**), which is consistent with previous findings[61].

Since *Stra8* plays an essential role in controlling meiotic entry[8], we next sought to further understand the gene regulatory mechanisms underlying *Stra8* activation. First, we not only found that accessibility of the *Stra8* promoter was highest in XX PGCs, but also that *Stra8* chromatin accessibility was positively correlated with *Stra8* gene expression (**Fig. 3b-c**). Second, the population of E13.5 XX PGCs displaying the strongest *Stra8* expression levels corresponded to the same population of XX PGCs with the highest module score of early meiotic prophase I genes (**Fig. 3c**; **Supplementary Fig. 3a-b**). Next, to determine putative regulators of *Stra8* expression, we identified enriched TF binding motifs within 10 kb of the *Stra8* promoter using Cistrome DB[44] (**Fig. 3d**). Of the factors with the highest Cistrome DB Regulatory Potential Score, we observed XX-enriched expression of *Rxra, H2afy, Tcf12, Rad21, Smc3,* and *Smad3* (**Fig. 3e**). Several of these factors, such as RAD21 and SMC3, are members of the cohesin complex, which serves both a structural and gene regulatory role during meiotic prophase I[62,63]. Moreover, RXRA has been shown to positively influence *Stra8* expression levels during oocyte development in the mouse[64]. Taken together, these data provide a proof-of-principle example for the application of single-nucleus multiomics in identifying putative regulators of germ cell fate.

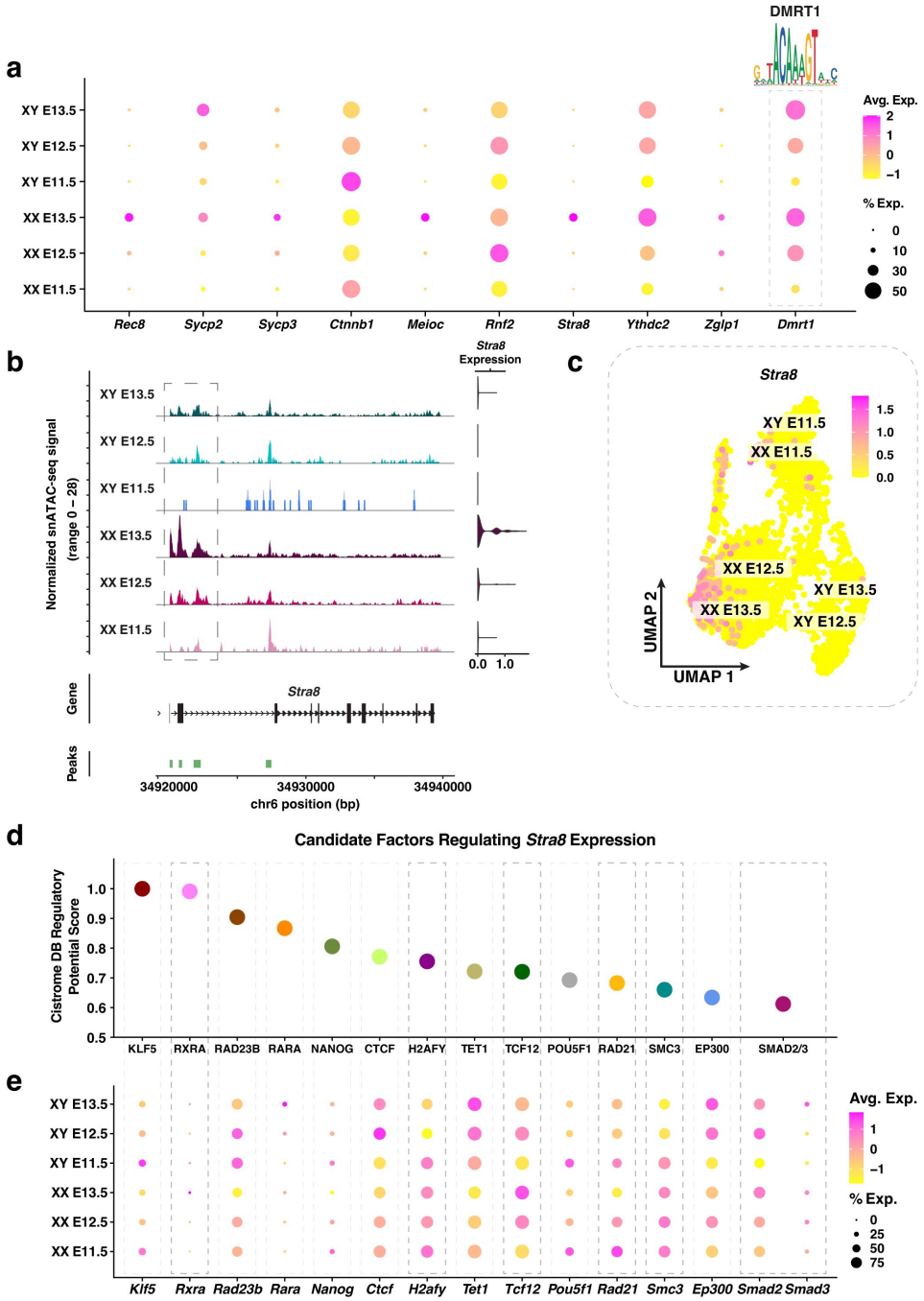

Figure 3

Molecular characterization of genes known for their roles in female PGC sex determination.

(a) Dotplot of the average expression and percentage of cells expressing *Rec8, Sycp2, Sycp3, Ctnnb1, Meioc, Rnf2, Stra8, Ythdc2, Zglp1,* and *Dmrt1*. The DNA binding motif for DMRT1 is indicated. The color scale represents the average expression level, and the size of the dot represents the percentage of cells expressing the gene. (b) *Left:* Coverage plot of the normalized snATAC-seq signal at the *Stra8* locus. *Right:* Violin plot of *Stra8* expression in E11.5-E13.5 XX and XY PGCs. (c) Joint UMAP showing *Stra8* expression levels for E11.5-E13.5 XX and XY PGCs. Individual cells are color coded by the level of *Stra8* expression. (d) Candidate factors (*x*-axis) regulating *Stra8* expression based on their Cistrome DB regulatory potential score (*y*-axis). (e) Dotplot of the average expression and percentage of cells expressing the candidate factors that potentially regulate *Stra8* expression. The color scale represents the average expression level, and the size of the dot represents the percentage of cells.

Expression and *in silico* chromatin binding of retinoic acid receptors in PGCs

Since recent reports have demonstrated that meiosis occurs normally in the fetal ovary of mice lacking key members of the retinoic acid (RA) signaling pathway[65], we next asked whether we could detect any evidence for gene regulation by RA receptors in PGCs. Previous studies have demonstrated that RA from the mesonephros induces meiosis in the fetal ovary whereas in the fetal testis, RA is degraded by the enzyme CYP26B1, therefore preventing PGC entry into meiosis[66,67]. However, this model was challenged by the findings that meiosis, and consequently oogenesis, still occurs in XX gonads of mice lacking the RA-synthesizing enzyme, ALDH1, or all RA receptors[65,68]. Nevertheless, *Stra8* expression was significantly reduced in E13.5 fetal ovaries without *Aldh1a1-3* or all RA receptors[65,68], thus indicating that *Stra8* expression is likely controlled by the RA signaling pathway, in addition to others. We therefore investigated the expression and *in silico* chromatin binding patterns of members of the RA signaling pathway (*i.e. Rara, Rarb, Rarg, Rxra, Rxrb, Rxrg, Crabp1*, and *Crabp2)*[69] (**Supplementary Figure 4a-b**). We found that expression of the retinoic acid receptor beta, *Rarb*, and cellular retinoic acid binding protein 1, *Crabp1*, were enriched in XX PGCs beginning at E12.5 and continuing to E13.5 (**Supplementary Figure 4a-b**). Similarly, E13.5 XX PGCs showed enriched expression of the retinoid X receptor alpha, *Rxra,* and cellular retinoic acid binding protein 2, *Crabp2* (**Supplementary Figure 4a-b**). By contrast, the expression of *Rara, Rarg,* and *Rxrb* was enriched in XY PGCs, and *Rxrg* was lowly expressed in both XX and XY PGCs (**Supplementary Figure 4a-b**). We also observed significant enrichment of the 'motif activity score', an *in silico* measurement of the contribution of a specific motif to gene regulation[70], of all RA receptors in E13.5 XX PGCs (**Supplementary Figure 4c-d**). These data indicate that the RA receptors display strong evidence of *in silico* chromatin binding in E13.5 XX PGCs but not E13.5 XY PGCs. Furthermore, the binding motifs for the RA receptors were predominantly located in E13.5 XX PGC chromatin peaks that were distal intergenic, intronic, or within promoters (**Supplementary Figure 4e**; motifs for RARB and RXRA provided as examples). These data indicate that the binding motifs for the RA receptors are predominantly enriched in both distal and proximal regulatory elements in E13.5 XX PGCs. The RA receptor motifs were also statistically linked to genes involved in 'response to retinoic acid', 'negative regulation of cell cycle', and 'meiotic cell cycle' (**Supplementary Figure 4f**; motifs for RARB and RXRA provided as examples). Finally, when we searched for the presence of RA receptor motifs in peaks linked to genes related to meiosis and female sex determination, we found that *Stra8, Rec8, Rnf2, Sycp1, Sycp2, Ccnb3,* and *Zglp1* contain the RA receptor motifs in their regulatory sequences (**Supplementary Figure 4g**). Together, these *in silico* experiments indicate that the RA signaling pathway directly regulates the expression of several meiosis-related genes.

Discovery of differentially expressed genes and differentially accessible regulatory loci underlying sex determination of PGCs

We further investigated our multiome dataset to discover genes and *cis*-regulatory elements enriched in PGCs in a sex-specific manner. First, we analyzed which annotated protein-coding genes showed transcriptional changes between XX and XY PGCs at each embryonic stage. Differential expression analysis ($\log_2$fold-change > 0.25) of snRNA-seq data identified 87, 301, and 585 XX PGC-enriched genes and 75, 344, and 544 XY PGC enriched genes at E11.5, E12.5, and E13.5, respectively (False Discovery Rate-adjusted p-value < 0.05; **Fig. 4a**). The number of differentially expressed genes (DEGs) increased by ~3.5-fold during the transition from E11.5 to E12.5 PGCs and by 1.75-fold during the transition from E12.5 to E13.5 PGCs (**Fig. 4a**). These findings demonstrate that the number of DEGs steadily increased over developmental time in both XX and XY PGCs.

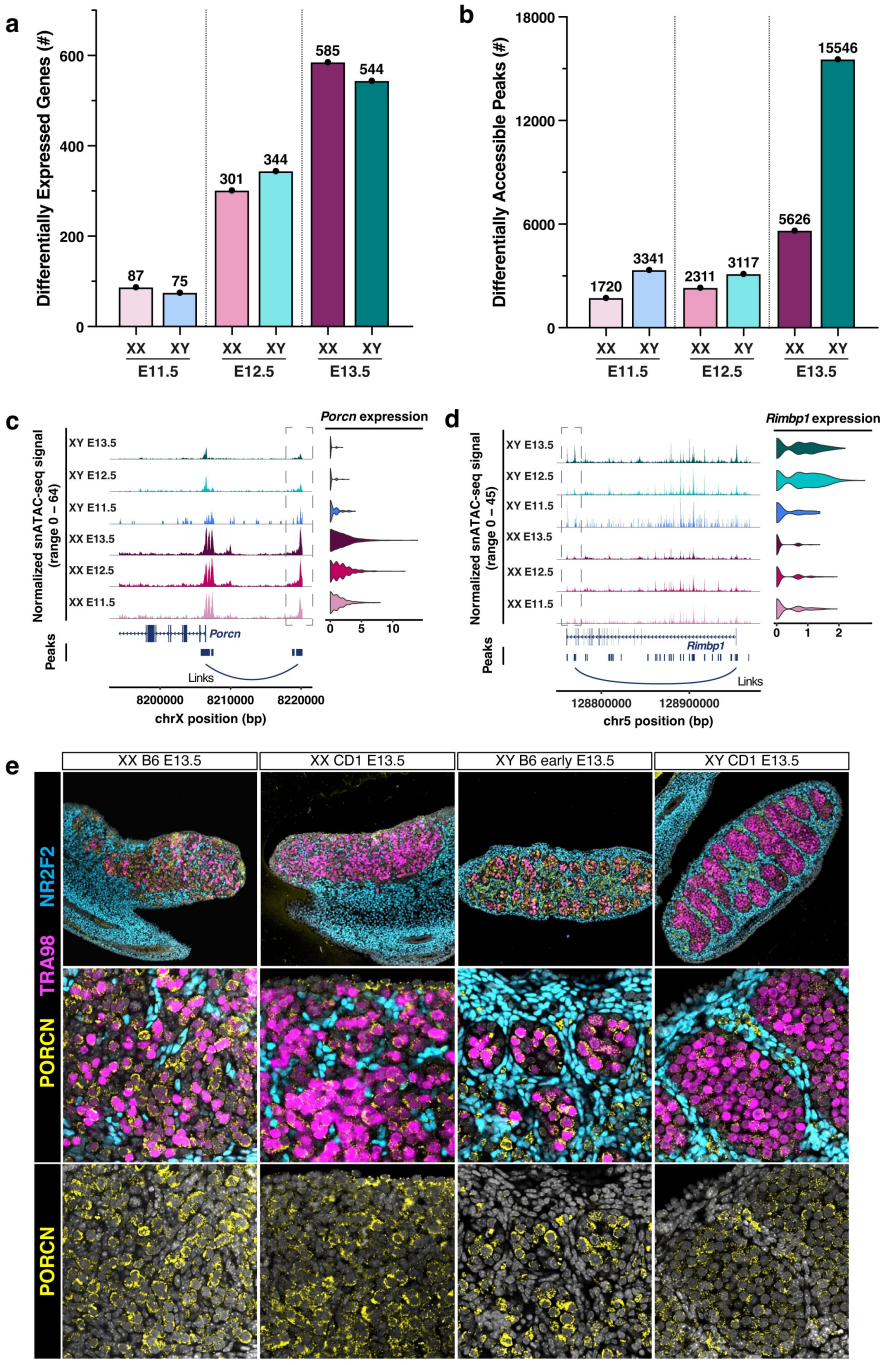

Figure 4

Identification of differentially expressed genes and differentially accessible peaks in E11.5-E13.5 XX and XY PGCs.

(a) Bar graph showing the number of differentially expressed genes (DEGs) between XX and XY PGCs at E11.5-E13.5. (b) Bar graph showing the number of differentially accessible chromatin peaks (DAPs) between XX and XY PGCs at E11.5-E13.5. (c) *Left:* Coverage plot of the normalized snATAC-seq signal at the DEG *Porcn* locus. Peak-to-gene linkages are indicated by the 'links' line. *Right:* Violin plot of *Porcn* expression in E11.5-E13.5 XX and XY PGCs. (d) *Left:* Coverage plot of the normalized snATAC-seq signal at the DEG *Rimbp1* locus. Peak-to-gene linkages are indicated by the 'links' line. *Right:* Violin plot of *Rimbp1* expression in E11.5-E13.5 XX and XY PGCs. (e) DAPI staining (grey) and immunofluorescence staining of TRA98 (pink), PORCN (yellow), and NR2F2 (blue) in E13.5 XX and XY gonads from C57Bl/6 and CD1 mice.

An in-depth analysis of the DEGs between XX and XY germ cells identified sex-enriched genes for PGCs at each embryonic stage (**Supplementary Figure 5a**). For example, we found elevated expression of the WNT protein regulator *Porcn* [71], the beta-catenin binding protein *Grip1*[72], and the growth regulator *Phlda2* [73] in XX PGCs (**Supplementary Figure 5a**). These observations are consistent with the fact that WNT/β-catenin signaling is involved in specifying the female fate of PGCs[8]. We also identified XY enriched DEGs, which included the histone demethylase *Kdm5d* [74], the regulator of pluripotency *Dppa5*[75], and the forkhead transcription factor *Foxp1* [76] (**Supplementary Figure 5a**). Gene ontology (GO) analyses of DEGs revealed several biological processes that show sex-enriched expression patterns (**Supplementary Figure 5b**). For example, the DEGs identified in XX PGCs were assigned to the 'meiotic chromosome organization', 'sister chromatid cohesion', and 'WNT signaling pathway' gene ontologies (**Supplementary Figure 5b**). The DEGs identified in XY PGCs were related to 'mitotic cell cycle', 'TGFβ signaling pathway', and 'regulation of translation' gene ontologies (**Supplementary Figure 5b**). In sum, the DEGs between XX and XY PGCs were characterized by their sex-specific roles in meiotic/mitotic cell cycle regulation, post-transcriptional processing of mRNA, and cell-cell signaling properties (**Supplementary Figure 5a-b**). These analyses provide a clear description of the transcriptional differences between XX and XY PGCs underlying the transition from bipotential germ cells to sex-specific oogonium or gonocytes.

To quantitatively assess chromatin dynamics during PGC sex determination, we next evaluated the number of differentially accessible chromatin peaks (DAPs) between XX and XY germ cells (**Fig. 4b**). The DAP analyses ($(\log_2$fold-change > 0.25) revealed 1,720, 2,311, and 5,626 XX enriched DAPs and 3,341, 3,117, and 15,546 XY enriched DAPs at E11.5, E12.5, and E13.5, respectively (**Fig. 4b**). The largest increase in the number of DA peaks was found between XY PGCs (15,546) and XX PGCs (5,626) at E13.5. The DAPs for XX and XY PGCs at E11.5-E13.5 were predominantly located in promoter, distal intergenic, or intronic regions, indicating that these DAPs likely correspond to gene regulatory sites (**Supplementary Figure 5c**). Together, these data suggest that the chromatin architecture of PGCs, and especially the chromatin of XY PGCs, undergoes significant remodeling during sex determination.

We next sought to associate sex-specific changes in chromatin accessibility with changes in gene expression for PGCs. To determine peak-gene associations, we correlated snATAC-seq peak accessibility with the expression of nearby genes, with the goal to identify sex- and stage-specific chromatin peaks positively correlated with mRNA expression (p-value < 0.05; z-score > 0). This analysis detected putative gene regulatory interactions based on correlations between chromatin peak accessibility and gene expression levels for all peaks and genes within 500 kb of each other[77]. Given the notable increase in DA peaks between XX and XY PGCs at E13.5 (**Fig. 4b**), we first identified the predicted target genes of all sex-specific DA peaks at this stage (**Supplementary Figure 6a-b**). For E13.5 XX PGCs, we observed positive DA peak-gene correlations for genes enriched for biological functions such as 'WNT signaling pathway', 'chromosome segregation', 'chromatin organization', 'meiotic cell cycle', and 'RNA stabilization' (**Supplementary Figure 6a**). A similar analysis of E13.5 XY PGCs revealed positive DA peak-gene correlations for genes involved in 'ncRNA metabolic process', 'mitotic cell cycle', 'DNA repair', 'regulation of translation', and 'RNA localization' (**Supplementary Figure 6b**). These findings demonstrate that we can use chromatin accessibility information to identify the regulatory elements linked to sex-specific genes in PGCs.

We next investigated the associations of differentially accessible (DA) chromatin peaks with differentially expressed (DE) gene expression for each PGC type. At E11.5, we found that DA peak-to-gene linkages were associated with ~20% of DE genes in XX PGCs and ~1% of DE genes in XY PGCs (**Supplementary Figure 6c**). The overlap of DA peak-to-gene linkages with DE genes increased steadily at E12.5 and E13.5, with ~35% and ~60% of DA peaks in E13.5 XX and XY PGCs, respectively, positively correlated with DE gene expression (**Supplementary Figure 6c**). For example, at E13.5, we identified the DA peaks positively correlated with the DE genes *Porcn*

(involved in the processing of WNT proteins)[71] and *Rec8* (involved in meiotic recombination)[8] in XX PGCs and *Rimbp1* (regulates cytosolic calcium ion concentration)[78] in XY PGCs (**Fig. 4c-d**; **Supplementary Figure 6d**). Immunofluorescence staining also confirmed strong expression of *Porcn* in E13.5 fetal XX germ cells and diminishing expression of *Porcn* in E13.5 fetal XY germ cells (**Fig. 4e**), demonstrating that single-nucleus multiomics can identify DE genes and the DA chromatin peaks that are associated with their expression.

A computational approach to identify candidate TFs regulating XX PGC gene expression

To better utilize the multiomics data, we developed an analytical flow to identify potential TFs that are involved in regulating the expression of DEGs in PGCs. First, we classified peak status by identifying peaks that are differentially accessible between XX and XY PGCs following the analytical framework in **Fig. 5a (i)**. Second, we linked peaks to genes to identify positive correlations between DA chromatin peaks and the expression of DEGs at each embryonic stage (*p*-value < 0.05; z-score > 0; **Fig. 5a (ii)**). Third, we identified all enriched TF binding motifs located within DA chromatin peaks and found unique motifs with a significant FDR threshold (*p*-value < 0.05; **Fig. 5a (iii)**). We investigated only the TFs that were differentially expressed between sexes and that showed germline enrichment. Finally, we predicted the target genes of candidate TFs by determining the genomic position of their enriched motifs and the genes linked to their resident peaks (**Fig. 5a (iv)**). This analytical flow allowed us to rank enriched TF binding motifs based on variation in chromatin accessibility, and unbiasedly infer TFs involved in PGC sex determination with greater confidence.

In XX PGCs, we uncovered several candidate TF regulators of E11.5-E13.5 XX-enriched DEGs (**Supplementary Figure 7a-c**). We plotted the mRNA expression levels of TFs and the *in silico* chromatin binding score, termed 'motif activity', of all significantly enriched binding motifs (**Supplementary Figure 7a-c**). For example, in E11.5 XX PGCs, RREB1 is a differentially expressed TF showing high gene expression, but negative levels of *in silico* chromatin binding (**Supplementary Figure 7a**). This finding indicates that although RREB1 is a highly expressed TF in E11.5 XX PGCs, RREB1 does not likely promote increased chromatin accessibility either because it does not bind to chromatin or a critical co-factor is unavailable (**Supplementary Figure 7a**). Following this analytical framework, we focused our analysis on TFs that showed high levels of gene expression, were differentially expressed in XX PGCs, and displayed a positive *in silico* chromatin binding score (**Supplementary Figure 7a-c**). The TFs meeting these criteria were KLF12, NR6A1, and ZFX in E11.5 XX PGCs; TFAP2c in E12.5 XX PGCs; and BACH2, CREM, GATA2, GLI3, MGA, NFKB1, PLAGL1, TBX4, TCFL5, TFAP2c, ZBTB7c, and ZFX in E13.5 XX PGCs (**Supplementary Figure 7a-c**). Of these sixteen TF candidates with PGC-enriched transcript expression, only seven (*Gata2*, *Mga, Nr6a1*, *Tbx4, Tcfl5, Tfap2c*, and *Zfx*) showed enriched expression in germ cells when compared to the somatic compartment of the gonad, and thus met our requirements for further *in silico* investigation (**Supplementary Figure 8a**). These candidate TFs have reported roles in cell fate determination, growth and fertility regulation, and meiotic progression[79,80], all of which are in line with the cellular events that occur as PGCs commit to the female pathway. Nonetheless, the chromatin binding and gene expression of GATA2, MGA, TBX4, TCFL5, and TFAP2c showed the clearest patterns of sexual dimorphism in E13.5 PGCs and represented the most compelling TF candidates (**Fig. 5b-c**; **Supplementary Figure 8b-f**).

The binding motif for TFAP2c (also known as *AP2γ*) showed prominent TF-associated accessibility, especially when compared to all other significantly enriched TF binding motifs (**Supplementary Figure 7b-c**). Immunofluorescence staining of TFAP2c also showed enrichment of TFAP2c in E13.5 XX PGCs when compared to the somatic compartment and E13.5 XY PGCs (**Supplementary Figure 9a**). Compared to XY PGCs, XX PGCs had enriched *Tfap2c* expression (*p*-value < 0.05; log2FC > 0.25) and enriched accessibility of the TFAP2c motif at E12.5-E13.5 (**Fig. 5b**). The TFAP2c motif was close to genes related to glycogen metabolism, carbohydrate metabolism, and chromatin

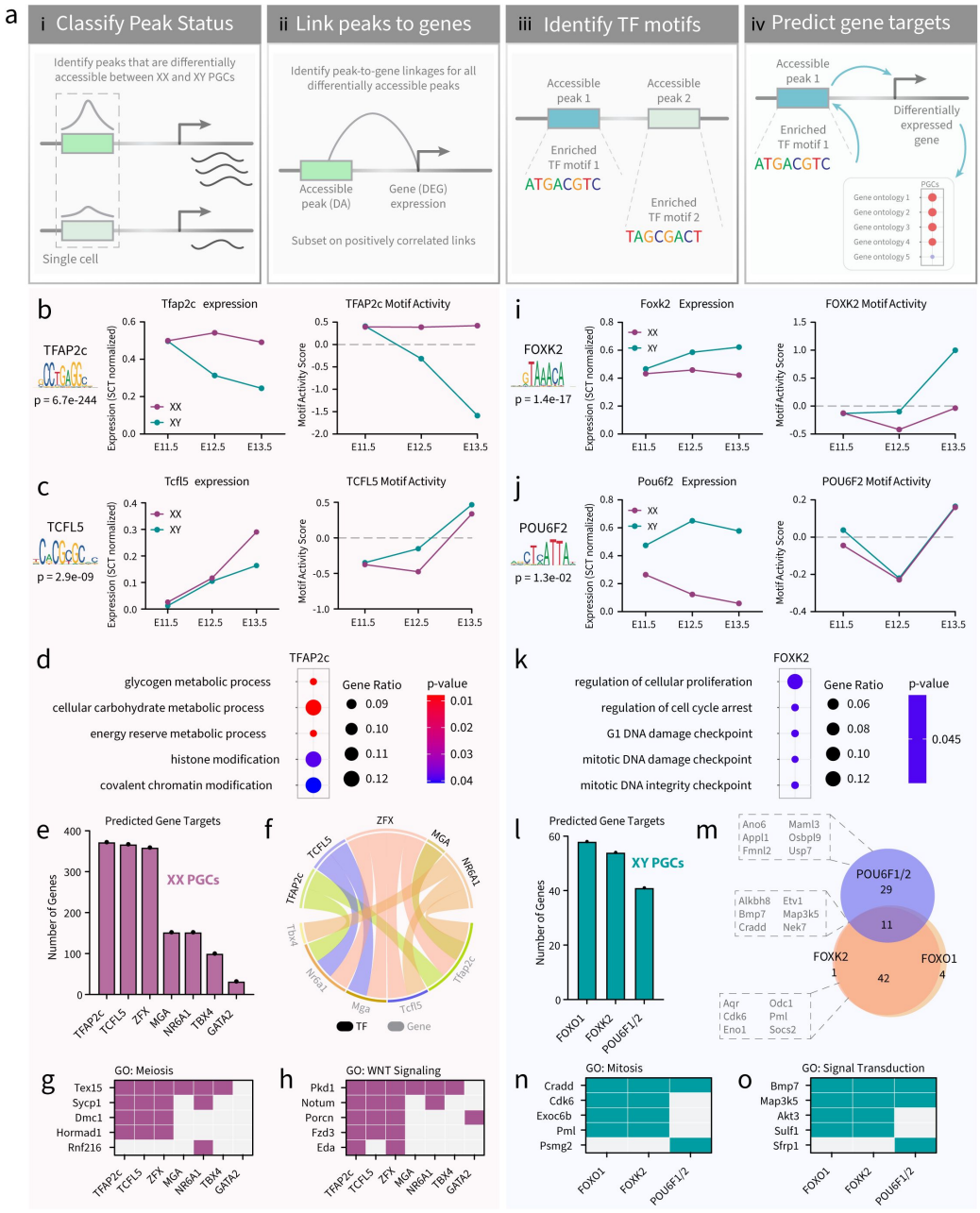

Figure 5

Analysis of gene networks in E11.5-E13.5 XX and XY PGCs.

(a) Illustration of the analytical flow to infer gene networks (*i.e.* transcription factors (TFs) and their predicted target genes) involved in PGC sex determination. (b-c; i-j) Line plots showing the expression levels and *in silico* chromatin binding, termed 'motif activity', of *Tfap2c* (b), *Tcfl5* (c), *Foxk2* (i), and *Pou6f2* (j) in E11.5-E13.5 XX (pink) and XY (teal) PGCs. The *p*-value of motif enrichment and the TF binding motif are indicated. (d and k) Enrichment analysis of biological process gene ontology (GO) of the predicted target genes for TFAP2c (d) and FOXK2 (k). The color scale represents the *p*-value of GO term enrichment, and the size of the dot indicates the gene ratio for each GO term. (e and l) The number of predicted target genes for enriched TFs in XX PGCs (e) and XY PGCs (l). (f) Circos plot showing whether the motifs for TFAP2C, TCFL5, ZFX, MGA, or NR6A1 (indicated in black) are detected in the regulatory loci linked to *Tbx4, Nr6a1, Mga, Tcfl5,* or *Tfap2c* (indicated in grey). (m) Venn diagram showing the overlap of predicted target genes for POU6F1/2, FOXK2, and FOXO1. (g-h) Icon array showing the presence (pink/purple) or absence (light grey) of the motifs for TFAP2c, TCFL5, ZFX, MGA, NR6A1, TBX4, and GATA2 in peaks linked to genes related to meiosis (g) or WNT signaling (h). (n-o) Icon array showing the presence (teal) or absence (light grey) of the motifs for FOXO1, FOXK2, and POU6F1/2 in peaks linked to genes related to mitosis (n) or signal transduction (o).

modification (**Fig. 5d**). The predicted target genes of TFAP2c accounted for 75% (372/494) of all DEGs with peak-to-gene linkages in XX PGCs (**Fig. 5e**), suggesting that TFAP2c is likely a high-level regulator of female PGC sex determination.

In addition to TFAP2c, we also investigated the regulatory potential of TCFL5 (Transcription Factor Like 5) in greater detail. TCFL5 was the most differentially expressed in E13.5 XX PGCs (p-value < 0.05; log2FC > 0.25) (**Fig. 5c**). Accessibility of the TCFL5 motif was also enriched in E13.5 XX PGCs, indicating that chromatin binding of TCFL5 is likely strongest at this stage (**Fig. 5c**). The predicted target genes of TCFL5 totaled 74% (367/494) of all DEGs with peak-to-gene linkages in XX PGCs (**Fig. 5e**). A large majority of the predicted target genes of TCFL5 were also predicted to be the target genes of the enriched TFs presented in **Fig. 5e**, e.g., the predicted target genes of these TFs overlapped with 4%-100% of the predicted target genes of TCFL5. The TCFL5-associated genes were related to chromosome segregation, meiotic cell cycle, sister chromatid cohesion, histone modification, regulation of TF activity, and cell-cell signaling by WNT. Our findings are supported by a recent report on the role of TCFL5 in XX PGCs, which demonstrated the requirement of TCFL5 in meiotic progression and female fertility[81].

Given that the motifs for TFAP2c and TCFL5 were not detected at every XX-enriched DEG, we hypothesize that a network of TFs, such as those identified in **Supplementary Figure 8a**, could work in cooperativity to regulate XX PGC gene expression. For example, we found that the predicted target genes of ZFX, MGA, NR6A1, TBX4, and GATA2 accounted for 73% (359/454), 31% (152/494), 20% (100/494), and 6% (32/494), respectively, of all DEGs with peak-to-gene linkage in XX PGCs (**Fig. 5e**). We also found that TFAP2c, TCFL5, ZFX, MGA, and NR6A1 may regulate the expression of each other (**Fig. 5f**). For instance, the motif for ZFX was located in the enhancers or promoters for TFAP2c, TCFL5, MGA, and NR6A1 (**Fig. 5f**). Likewise, NR6A1 is predicted to control the expression of TFAP2C, TCFL5, and TBX4 (**Fig. 5f**). Furthermore, the motifs for the candidate TFs were statistically linked to genes involved in 'meiotic cell cycle' and the 'WNT signaling pathway' (**Fig. 5g-h**). Meiosis related genes such as *Tex15*, *Sycp1*, *Dmc1*, and *Hormad1* contained the motifs for at least three of the predicted TFs in their regulatory sequences (**Fig. 5g**). A similar pattern was also found for genes related to WNT signaling, e.g. *Pkd1*, *Notum*, *Porcn*, *Fzd3*, and *Eda* were predicted to be regulated by the binding of at least two of the putative TFs in XX PGCs (**Fig. 5h**). These data indicate that there may be unknown and undefined TF regulatory networks underlying XX PGC identity and function.

Identification of candidate TF regulators of XY PGC differentiation

To identify candidate TF regulators involved in XY PGC sex determination, we performed the same analysis as described for XX PGCs (**Fig. 5a**). Our analysis focused strictly on differentially expressed TFs that show evidence of chromatin binding based on the empirical *in silico* chromatin binding score (**Supplementary Fig. 10a-c**). In contrast to XX PGCs (**Supplementary Fig.** 7), we did not identify any significantly enriched TF binding motifs (p-value > 0.05) meeting our criteria in E11.5 XY PGCs (**Supplementary Fig. 10a**). Only three XY-enriched TFs, FOXP1, POU6F2, and TCF4, were differentially expressed between XY and XX PGCs at E12.5 (**Supplementary Fig. 10b**), and only TCF4 showed a positive *in silico* chromatin binding score in E12.5 XY PGCs (**Supplementary Fig. 10b**). For E13.5 XY PGCs, we detected five significantly enriched motifs that bound differentially expressed TFs (p-value < 0.05): FOXK2, FOXO1, FOXP1, POU6F1, and TCF4 (**Supplementary Fig. 8c**). Of these six XY-enriched TF candidates from E11.5 to E13.5, FOXO1, FOXK2, and POU6F1/2 emerged as the top candidates based on germline enriched expression when compared to somatic cells, differential expression in XY PGCs, and positive *in silico* chromatin binding scores (**Supplementary Fig. 11a-b; Fig. 5i-j**). Together, these data indicate that XY PGCs rely less on differentially expressed TF-mediated gene regulation when compared to XX PGCs.

At E13.5, when the most TFs and their motifs were found in XY PGCs, we observed significant enrichment of the FOX (forkhead-box) protein family of TFs (**Supplementary Fig. 10c**). FOX proteins contain pioneering transcription properties that allow for binding to condensed chromatin and enabling competency of gene activity through histone remodeling[82]. Given their role in opening condensed chromatin, FOX TFs play important roles in regulating the expression of genes involved in cellular differentiation, proliferation, plasticity, and growth[49]. In particular, we found that the expression of *Foxo1* and *Foxk2* were specific to XY PGCs when compared to XX germ cells and the XY somatic compartment (**Supplementary Fig. 10c; Supplementary Fig. 11a-b; Fig. 5i**). From our dataset, we found the FOXK2 motif positively associated with genes related to cellular proliferation, cell cycle arrest, mitosis, and DNA damage checkpoint (**Fig. 5k**). In addition, the predicted target genes of FOXK2 totaled 26% (54/208) of all DEGs with peak-to-gene linkages in XY PGCs (**Fig. 5l**). This finding was consistent with the number of predicted target genes of FOXO1, which included 28% (58/208) of all DEGs with peak-to-gene linkages in XY PGCs and may be the result of motif redundancy among FOX proteins. Given the substantial increase in the number of DA peaks in E13.5 XY PGCs (**Fig. 4b**), we hypothesize that the pioneering properties of the FOX family promotes chromatin remodeling as XY PGCs transition to gonocytes.

We also detected significant enrichment of the POU domain, class 6 (POU6) TF family in XY PGCs (**Supplementary Fig. 10b-c; Supplementary Fig. 11a & c; Fig. 5j**). The POU family plays important roles in regulating cellular differentiation and the timing of cellular events[83]. We found XY-enriched expression of POU6F1 and POU6F2 in PGCs, however, only POU6F2 displayed germline enrichment when compared to the somatic cell populations of the gonad (**Supplementary Fig. 11a & c; Fig. 5j**). Furthermore, we observed the strongest POU6F1 and POU6F2 *in silico* chromatin binding scores in XY PGCs at E13.5 (**Supplementary Fig. 11c; Fig. 5j**). The POU6F1/2 motif was associated with 19% (40/208) of all DEGs with peak-to-gene linkages in XY PGCs (**Fig. 5l**). POU6F1/2-associated genes were related to voltage-gated chloride channel activity (*Ano6*)[84], regulation of cell proliferation (*Appl1*)[85], regulation of actin filaments (*Fmnl2*)[86], Notch signaling (*Maml3*)[87], lipid metabolism (*Osbpl9*)[88], and ubiquitin ligase activity (*Usp7*)[89] (**Fig. 5m**). Thus, the POU6 TF family is a top candidate to coordinate gene expression activity in XY PGCs.

We next compared the number and identity of FOXO1-, FOXK2-, and POU6F1/2-associated genes in XY PGCs. First, we did not detect the FOXO1, FOXK2, or POU6F1/2 motifs in the promoters/enhancers for each of these TFs. These findings indicate that these TFs do not regulate the expression of each other, and other mechanisms or regulatory pathways are required to activate the expression of candidate TFs in XY PGCs. Furthermore, the predicted target genes of the FOX and POU6 TF families shared little overlap with each other (11/87) (**Fig. 5m**). The genes that were shared between FOXO1, FOXK2, and POU6F1/2 had reported functions in maintaining survival after DNA damage (*Alkbh8*)[90], TGFβ signaling (*Bmp7*)[91], and initiation of mitosis (*Nek7*)[92] (**Fig. 5m**). By comparison, the predicted target genes of FOXO1 and FOXK2 showed considerable overlap (42/47) and included genes related to RNA splicing (*Aqr*)[93], cell cycle progression (*Cdk6*)[94], cellular metabolism (*Eno1* and *Odc1*)[95,96], cell death (*Pml*)[97], and negative regulation of the IGF1 signaling pathway (*Socs2*)[98] (**Fig. 5m**). The motifs for FOXO1, FOXK2, and POU6F1/2 were statistically associated with genes involved in 'mitotic cell cycle' and 'signal transduction' (**Fig. 5n-o**). Mitosis-related genes such as *Cradd, Cdk6, Exoc6b,* and *Pml* were potentially regulated by FOXO1 and FOXK2 (**Fig. 5n**). In addition, we noted that the POU6F1/2 motif was associated with *Cradd* and *Psmg2* (**Fig. 5n**). We observed a similar pattern for genes related to signal transduction (**Fig. 5o**). For example, *Bmp7, Map3k5, Akt3,* and *Sulf1* were associated with FOXO1 and FOXK2, whereas *Bmp7, Map3k5,* and *Sfprp1* were associated with POU6F1/2 motif activity (**Fig. 5o**). Taken together, these data suggest that in XY PGCs, although FOXO1, FOXK2, and POU6F1/2 regulate genes with similar properties, the FOX and POU6 TF families are unlikely to collaboratively regulate gene expression.

Sex-specific enrichment of ligand-receptor signals between gonadal support cells and PGCs

Sexual dimorphism of germ cell differentiation is thought to be controlled by a meiosis-inducing substance produced by the somatic cells of the fetal ovary and/or by a meiosis-preventing substance produced by the fetal testis[99]. In experimentally generated XX/XY chimeras, both XX and XY PGCs can initiate meiosis when residing in the fetal ovary or enter mitotic arrest when colonizing the fetal testis[100]. In the fetal ovary, PGC meiosis is induced by retinoic acid (RA) produced by ovarian and mesonephric somatic cells[8]. Consistent with this model, RA signaling induces expression of *Stra8* and *Meiosin*, both of which are critical for promoting entry into meiosis[8]. Signaling by WNT and BMP also influences the timing of meiotic onset in XX PGCs[8], and inhibin/activin signaling regulates XX germ cell growth and proliferation[101]. In the fetal testis, RA degradation by CYP26B1 and FGF9 signaling from XY supporting cells result in reduced *Stra8* expression, consequently blocking meiotic entry in XY PGCs[8]. Thus, PGC sex determination is dependent on external signals from the gonadal environment rather than chromosomal sex.

To uncover potential instructive signals from the supporting somatic cells (Sertoli cells in the fetal testis or granulosa cells in the fetal ovary) to PGCs, we searched for supporting cell-derived ligands and the corresponding receptors in PGCs (**Fig. 6a**), using CellphoneDBv5.0 for cell communication analysis[45]. We first identified the supporting cell clusters from the snRNA-seq dataset of whole fetal gonads using the established female and male supporting cell markers *Runx1* and *Sox9*, respectively[30,102]. We next quantified the number of significant interactions between supporting cells and PGCs at each embryonic stage (p-value < 0.05; **Fig. 6b**). Such analyses revealed that the number of interactions gradually increased from E11.5 to E13.5 for both XX and XY gonads (**Fig. 6b**). We also observed that the number of ligand-receptor pairs was 15-38% higher in XX gonads, compared to XY gonads at the same embryonic stage (**Fig. 6b**). To validate our computational approach, we next searched for known pathways associated with PGC differentiation, such as signaling by WNT, inhibin/activin, and BMP (**Fig. 6c**). The number of significant interactions per pathway increased in both XX and XY gonads from E11.5 to E13.5 (**Fig. 6c**; **Supplementary Fig. 12a-b**). We also observed a moderate increase in the number of interactions for the inhibin/activin and BMP pathways in XX gonads when compared to XY gonads at the same embryonic stage (**Fig. 6c**). Finally, we found that the number of interactions related to WNT signaling were nearly 2-3 times greater in XX than XY gonads (**Fig. 6c**; **Supplementary Fig. 12a-b**). Together, these data demonstrate that sexually dimorphic patterns of cell signaling events were present between supporting cells and PGCs.

We next searched for new sex-specific enrichment of ligand-receptor pairs between supporting cells and PGCs (**Fig. 6d-i**). For each embryonic stage, we first identified which signaling pathways were detected in XX alone, XY alone, or both XX and XY gonads that met our statistical threshold (p-value < 0.05; **Fig. 6d, f & h**). At E11.5, XX and XY gonads shared the greatest overlap of supporting cell-derived ligands and PGC-expressing receptors (**Fig. 6d**). The shared pathways between XX and XY gonads were largely related to adhesion by collagen/integrin, fibronectin, and cadherin. E11.5 XX gonads had enrichment of 19 ligand-receptor pairs related to, for example, BMP, ephrin, integrin, FGF, and WNT signaling (**Fig. 6d**). By contrast, E11.5 XY gonads had significant enrichment of 10 ligand-receptor pairs predominantly related to cell adhesion pathways like ephrin, nectin, and integrin (**Fig. 6d**). To highlight the general trends of expression between sexes, we compared the mean expression of all representative ligand-receptor pairs from each group in **Fig. 6d, f, & h** (**Fig. 6e, g, & i**). In E11.5 gonads, we observed dimorphic expression of XX-enriched interactors, such as BMP2/BMR1A/AVR2B and WNT4/FZD10/LRP5, but not the XY-enriched interactors CADM1/NECTIN3 and EFNB1/EPHB3 (**Fig. 6e**). Furthermore, we did not observe a sex bias for the mean expression of interacting pairs for

Figure 6

Sexually dimorphic interactions between PGCs and gonadal supporting cells.

(a) Illustration of cell communication between PGCs and gonadal supporting cells. (b) Number of significant interactions or ligand-receptor pairs (*p*-value < 0.05) between supporting cells and PGCs in E11.5-E13.5 XX and XY gonads. (c) Number of significantly enriched ligand-receptor pairs, or interactions, per signaling pathways induced by either WNTs, inhibins/activins, or BMPs in E13.5 XX and XY gonads. (d, f, & h) Venn diagrams showing the overlap of significant interactions between supporting cells and PGCs in XX and XY gonads at E11.5 (d), E12.5 (f), and E13.5 (h). (e, g, & i) Dotplots showing the mean expression for all interacting partners in representative ligand-receptor pairs in XX and XY gonads at E11.5 (e), E12.5 (g), and E13.5 (i). Ligand-receptor pairs in pink represent XX-specific interactions, ligand-receptor pairs in blue represent XY-specific interactions, and ligand-receptor pairs in grey represent interactions found in both XX and XY gonads. The size of the dot represents the mean expression value.

the shared pathway DLK1:NOTCH3 between E11.5 XX and XY gonads (**Fig. 6e**). These data indicate that at E11.5, sexual dimorphism of signaling pathways, in terms of number and mean expression, is most prominent between XX supporting cells and XX PGCs.

In E12.5 and E13.5 gonads, we observed a stronger XX bias in ligand-receptor interactions between supporting cells and PGCs (**Fig. 6f-i**). At E12.5, we observed 49 and 24 significant interactions in XX and XY gonads, respectively (**Fig. 6f**). In E12.5 XX gonads, BMP, Insulin-like growth factor (IGF), R-spondin, and WNT signaling were the most frequently observed pathways and displayed strong XX-enriched mean expression (**Fig. 6f-g**). In XY gonads, detection of new signaling pathways emerged from E11.5 to E12.5, including signaling by BMP6, chemokines, Matrix metalloproteinases (MMP), and Neuregulin (**Fig. 6f**). Compared to E12.5, E13.5 XX gonads continued to have enrichment for the BMP, IGF, and WNT signaling pathways with additional pathways related to midkine (MDK), Calsyntenin (Clstn), and SLIT and NTRK-like protein signaling detected (**Fig. 6f and h**). The BMP5:BMR1A:AVR2B and MDK:PTPRZ1 ligand-receptor interactors also showed XX-biased mean expression (**Fig. 6i**). Similarly in E12.5 and E13.5 XY gonads, we observed maintenance of the ligand-receptor pairs, BMP6:ACVR:1A2B and BMP6:BMR1A:AVR2B (**Fig. 6f and h**). However, enrichment of signaling by Desert hedgehog, Notch, and WNT inhibition became significant in E13.5 XY gonads but not E12.5 (**Fig. 6f and h**). High mean expression of DHH:PTCH:GAS1 and DLK1:NOTCH3 interactors was detected in E13.5 XY gonads but became indetectable in E13.5 XX gonads (**Fig. 6i**). Taken together, these data suggest that sex-specific enrichment of ligand-receptor interactions become gradually more apparent from E11.5 to E13.5.

Discussion

Male and female germ cells carry the hereditary information required for the generation of a new individual and the perpetuation of sexually reproducing species. Errors in germ cell development and differentiation can result in a variety of reproductive diseases, such as infertility[6]. Disruption of gametogenesis during early embryonic development can even lead to the onset of disease later in life, a concept which is supported by the Developmental Origins of Health and Disease (DOHaD) theory[103]. Thus, it is of critical importance to understand the molecular and epigenetic signatures of developing germ cells, and particularly PGCs. Nonetheless, the heterogeneous nature and low cell numbers of PGC populations often hinders the identification of novel factors that influence PGC fate. To overcome this limitation, we generated the first single-nucleus atlas of gene expression and chromatin accessibility for murine PGCs during sex determination. We demonstrate that single-nucleus multiomics successfully captures PGC sex determination at high resolution, identifies putative TFs involved in PGC sex determination, and provides insight into the cellular communication pathways between PGCs and the supporting cells of the gonads.

Integrated RNA and ATAC data for individual PGCs identified significant heterogeneity in XX PGC populations compared to XY PGCs

We first identified the sub-groupings of E11.5-E13.5 XX and XY PGCs and resolved the patterns of gene expression and chromatin accessibility underlying each PGC population. We observed that the combined analysis of integrated snRNA- and snATAC-seq data produced better clustering and separation of PGCs than either assay alone. For example, we found that at E11.5, XX and XY PGCs share highly similar transcriptomic and open chromatin profiles, indicating that PGCs at this stage are bipotential at the level of gene regulation and chromatin accessibility. This observation agrees with the study that showed the global overlap of transcriptomes between E11.5 PGCs from both sexes, in which they detected only 101 DEGs between XX and XY PGCs with functional annotations

related to chaperonin complexes and proteasome formation[104]. At E12.5, subpopulations of PGCs clustered separately within each sex, supporting the model that PGCs do not commit to a sex-specific fate in a synchronous manner[54]. At E13.5, two sub-clusters of XX PGCs appear, consistent with the asynchronous entry of XX PGCs into meiosis[54]. Others also reported substantial heterogeneity of human and mouse PGC populations, often arising from intrinsic variation in the epigenetic reprogramming required for sex differentiation[105,106]. These data underscore the discovery potential of simultaneously measuring multiple layers of gene regulation to define the developmental fate of PGCs.

Molecular signatures of differentiating PGCs revealed disease-related genes associated with PGC clusters

When examining genes that could be responsible for the expression-based clustering of PGCs, we found TFs, cell cycle regulators, chromatin modifiers, and ribosome biogenesis genes underlying the diverse transcriptional profiles of differentiating PGCs. Several of the expressed genes enriched in individual PGC clusters were involved in infertility or disease processes when nonfunctional. For example, mutations in *Lars2,* the mitochondrial leucyl-tRNA synthetase, is associated with premature ovarian failure in women[50]. and aberrant methylation at the promoter of *Mgmt,* a gene that is required for DNA repair and linked to the formation of testicular germ cell tumors in humans[51]. The dynamic and varied regulatory profiles of PGCs that we observed highlight the necessity of pairing chromatin accessibility information with gene expression data for elucidating the mechanisms underlying PGC sex determination.

Sexual dimorphism in chromatin accessibility increases temporally during PGC development

The chromatin landscape of PGCs is the most sexually dimorphic at E13.5 when compared to E11.5 and E12.5. The substantial increase in DAPs at E13.5 leads to several hypotheses that could account for the changes to PGC chromatin architecture during sex determination. First, it is possible that the chromatin of bipotential E11.5 PGCs is primed for activating either sex-specific differentiation pathway. Chromatin priming occurs when chromatin is maintained in an accessible state prior to gene expression, and is evidenced in a number of developmental contexts, such as during specification of hematopoietic cells and lineage commitment of embryonic stem cells[107,108]. During development, chromatin priming fine-tunes gene activation by enabling a rapid response to external cues that induce cell differentiation[107]. Second, we posit that E11.5 PGCs have fewer DA peaks because they are actively dividing. Mitotic chromosome condensation is known to induce a state of heterochromatin and the eviction of TFs from chromatin in actively dividing cells[109]. Third, the increased number of DA peaks at E13.5 may be the result of changes to chromatin structure as XX PGCs enter meiotic prophase I, an event that requires dramatic changes to meiotic chromosomes as they undergo DNA double-strand breaks, synapsis, and crossing over[47]. Indeed, increased chromatin accessibility was found in mouse spermatocytes during prophase I of meiosis[110]. Finally, the upregulated expression of forkhead TFs in male PGCs at E12.5 and E13.5 may induce the *de novo* establishment of DA peaks through their pioneering activity. Nonetheless, our results are consistent with previous reports that demonstrate the global decrease in DNA methylation levels, or increased chromatin accessibility, in mouse PGCs beginning at E9.5 and continuing through E13.5[111].

Single-nucleus multiomics identifies evidence for chromatin priming during PGC lineage commitment

Moreover, not all sex- and stage-specific DA chromatin peaks account for changes in differential gene expression for PGCs. Chromatin accessibility and gene expression are often asynchronously correlated during cellular differentiation, such as during chromatin priming[77]. For example, accessibility of regulatory loci involved in cell-fate decisions may precede gene expression to

prime cells towards one developmental lineage versus another[77]. In the context of germ cell differentiation, PGCs must also undergo extensive chromatin remodeling events, including genome-wide DNA demethylation and a reduction in both repressive and activating histone modifications, prior to sex determination[112]. Indeed, chromatin priming has been observed in PGCs from other species, such as macaques, pigs, and goats[105]. Thus, it is likely that the DA chromatin peaks observed for E11.5 and E12.5 PGCs in mice may either reflect global chromatin remodeling events associated with sex determination or foreshadow future cell fate decisions by remaining poised for TF binding. Nevertheless, a comprehensive profile of 3D chromatin structure and enhancer-promoter contacts in differentiating PGCs is needed to fully understand how changes to chromatin facilitate PGC sex determination.

The gene regulatory networks of XX PGCs are enriched for the transcription factors, TFAP2c, TCFL5, GATA2, MGA, NR6A1, TBX4, and ZFX

Our multiomics approach uncovered compelling TF candidates that may be involved in PGC development and differentiation. In XX PGCs, seven TF prospects were identified: GATA2, MGA, NR6A1, TBX4, ZFX, TCFL5, and TFAP2c. Of these TFs, TCFL5 and TFAP2C showed the strongest enriched expression in XX PGCs. TCFL5 binds to the promoters of meiotic genes such as *Sycp1* (synapsis) and *Stag3* (cohesion) during spermatogenesis in the mouse[79]. *Tcfl5* is indispensable for the completion of meiosis and spermatid maturation[79] in male germ cells, and meiosis of female germ cells[81]. Our multiomics data also demonstrate strong *Tfap2c* expression and motif enrichment in XX PGCs beginning at E12.5 and continuing through E13.5. These findings are corroborated by other single-cell RNA-seq analyses that showed *Tfap2c* expression in human and mouse PGCs and gonocytes[80,113–116]. Additionally, TFAP2c has several reported roles in cell-type specification in both mouse and human, including germline specification and maintenance of germ cell identity[80], remodeling of naïve enhancers[117], and hormone responsiveness[118]. Loss of TFAP2c in mice also leads to a reduction in germ cell number by either PGC apoptosis or PGC differentiation into somatic cells[119]. Considering that changes in gene regulation leading to the female and male pathways begin as early as E12.5[3], understanding how TCFL5 and TFAP2c activity is induced and investigating the role of TFAP2c in PGC sex determination are important future directions.

XY PGCs showed sex-specific enrichment of transcription factors associated with reproductive dysfunction

In XY PGCs, we detected significant enrichment of the forkhead box and POU6 families of TFs. In particular, FOXO1, FOXK2, and POU6F1/2 exhibited XY biased expression and *in silico* chromatin binding. Although the function of *Foxo1* has been studied in gonocytes[120], little is known about the roles of *Foxk2* and *Pou6f1/2* in PGCs. Loss of *Foxo1* in embryonic gonocytes leads to defective spermatogenesis in adulthood[120], and inactivation of *Foxo1* in spermatogonia results in germ cell loss and male infertility[121]. FOXO1 activity is also widely involved in regulating metabolic homeostasis, redox balance, cell death, and DNA damage repair[122]. Likewise, FOXK2 functions in similar roles as FOXO1, such as in cellular metabolism, cell survival, and apoptosis[123]. Furthermore, the POU6 family consists of POU6F1 and POU6F2, which share highly conserved DNA motif recognition sequences and function in similar cellular pathways[124]. Both POU6F1 and POU6F2 have been reported to suppress tumor proliferation[125], and POU6F2 mutations have been implicated in the progression of Wilms Tumor, hypogonadism, and pubertal failure in human patients[126]. Investigating the cooperativity of these TFs will greatly improve our understanding of how mitotic arrest and the commitment to spermatogenesis are regulated during germ cell development.

Single-nucleus RNA-seq maps the temporal expression patterns of WNT, BMP, and RA signaling during PGC sex determination

During sex determination, PGCs respond to signals from the supporting cells of the gonads to determine their sex-specific fate. In the fetal ovary, *Runx1+* supporting cells produce WNT4 to maintain germ cell pluripotency and prevent precocious meiotic entry[8,60]. We observed signaling by WNT4, WNT5A/B, and WNT6 in both XX and XY gonads as early as E11.5. The number of interactions related to WNT signaling also increased from E11.5 to E13.5 in both XX and XY gonads even though diminished WNT signaling is required for meiotic entry[60]. However, XY PGCs showed enrichment of the WNT agonist SFRP2[127] at E13.5, indicating that WNT signaling may be inhibited in XY PGCs earlier than XX PGCs. BMP secretion by ovarian supporting cells also plays a critical role in regulating meiotic entry[8]. Even though BMP6 signaling was observed in XY PGCs, BMP2 and BMP5 signaling were specific to XX PGCs beginning at E11.5. Our findings are consistent with functional studies showing that BMP2 signaling induces *Zglp1* expression and promotes the oogenic fate[10]. Additionally, RA produced by both ovarian and mesonephric somatic cells is required to initiate expression of *Stra8* and *Meiosin*, two TFs that are essential for completion of meiosis[8,9]. Significant enrichment of the RA signaling pathway was not detected between supporting cells and PGCs in XX or XY gonads at any embryonic stage with CellphoneDB. Given that RA is primarily produced by the mesonephros[66], detection of RA signaling between gonadal supporting cells and PGCs is unlikely. However, *in silico* chromatin binding of RA receptors was enriched in XX PGCs and the RA receptor motifs were found in the regulatory loci for *Rec8, Stra8, Sycp2, Sycp3,* and *Zglp1*, which is consistent with previous studies[128]. Taken together, our findings provide insight into the temporal patterns of WNT, BMP, and RA signaling during PGC differentiation.

Discovery analysis identifies potentially new signaling pathways between gonadal supporting cells and PGCs

In addition to these pathways known for their roles in PGC sex determination, we identified several uncharacterized ligand-receptor pairs with sex-specific enrichment. For example, midkine (MDK) signaling was enriched in E13.5 XX gonads. Although the role of MDK signaling in PGCs is not understood, MDK is known to be important for maintaining cell survival and growth and promoting cell migration[129]. MDK is a heparin-binding cytokine and, together with pleiotrophin, regulates female reproductive activities such as oocyte maturation and the estrous cycle[130]. Furthermore, in E13.5 XY gonads, we found enrichment of the Desert Hedgehog (DHH) signaling pathway between *Sox9+* cells and XY PGCs. In addition to regulating fetal Leydig cell differentiation[131], DHH signaling has also been shown to regulate germ cell maturation and spermatogenesis[132]. Ablation of *Dhh* in adult XY mice causes infertility due to a loss of mature sperm[132], however, whether *Dhh* regulates germ cell development in fetal XY mice remains to be determined. Overall, we posit that these and other signaling pathways are important for promoting PGC survival and differentiation.

Our study demonstrated the benefit of using multiomics approaches to probe the cellular and developmental pathways involved in cell fate decisions, gametogenesis, and PGC differentiation. Unlike traditional assays, joint RNA and ATAC analyses from individual PGCs enabled the detection of paired chromatin accessibility and gene expression information from the same cell to discover new gene regulatory interactions. We show that our multiomics data accurately identified enrichment of known pathways involved in PGC sex determination, suggesting that our computational approaches are valid and capable of discovering potentially new regulators of PGC fate. Indeed, our findings revealed the diversity of factors required for PGC programming. For example, we cataloged the heterogeneity of PGC populations based on chromatin status and transcription, identified chromatin priming as a major regulatory mechanism in PGC lineage

commitment, and revealed a core set of TFs and ligand-receptor pairs unique to XX and XY PGCs. Taken together, our data demonstrate the power of explorative single-nucleus ATAC-seq and RNA-seq to understand the mechanisms inherent to germline differentiation.

Acknowledgements

We thank the late Dr. Keith Parker (UT Southwestern Medical Center, USA) for the Nr5a1-cre mice. Valuable discussions with F. DeMayo, C. Williams, M. Morgan, C. Guardia, and all members of the Yao lab during the course of this project contributed greatly to our manuscript. We are especially grateful for P. Brown for her advice and assistance with experiments and editorial comments in the writing process. We are grateful to the NIEHS Comparative Medicine Branch for mouse colony maintenance and to the Epigenomics and DNA Sequencing Core and the Integrative Bioinformatics Support Group for their assistance with sequencing and data analysis. This work was supported by the Intramural Research Program (ES102965 to H.H.-C.Y.) of the NIH, National Institute of Environmental Health Sciences and the Postdoctoral Research Associate Training (PRAT) Program (Fi2) Fellowship of the NIH, National Institute of General Medical Sciences awarded to A.K.A.

Additional files

Supplementary figures

References

1 McLaren A. (2003) **Primordial germ cells in the mouse** *Developmental biology* **262**:1–15

2 Hancock G. V., Wamaitha S. E., Peretz L., Clark A. T. (2021) **Mammalian primordial germ cell specification** *Development* **148**

3 Spiller C. M., Bowles J. (2015) **Sex determination in mammalian germ cells** *Asian Journal of Andrology* **17**

4 Borum K. (1961) **Oogenesis in the mouse: a study of the meiotic prophase** *Experimental cell research* **24**:495–507

5 Bowles J., Koopman P. (2010) **Sex determination in mammalian germ cells: extrinsic versus intrinsic factors** *Reproduction* **139**:943–958

6 Czukiewska S. M., de Sousa Lopes S. M. C. (2022) **Fetal germ cell development in humans, a link with infertility** *Seminars in Cell & Developmental Biology* Elsevier :58–65

7 Oosterhuis J. W., Looijenga L. H. (2019) **Human germ cell tumours from a developmental perspective** *Nature Reviews Cancer* **19**:522–537

8 Spiller C., Bowles J. (2022) **Instructing mouse germ cells to adopt a female fate** *Sexual Development* **16**:336–348

9 Ishiguro, K.-i.,, et al. (2020) **MEIOSIN directs the switch from mitosis to meiosis in mammalian germ cells** *Developmental cell* **52**:429–445

10 Nagaoka S. I., et al. (2020) **ZGLP1 is a determinant for the oogenic fate in mice** *Science* **367**

11 Vanhoutteghem A., et al. (2014) **The zinc-finger protein basonuclin 2 is required for proper mitotic arrest, prevention of premature meiotic initiation and meiotic progression in mouse male germ cells** *Development* **141**:4298–4310

12 Spiller C. M., Wilhelm D., Koopman P. (2010) **Retinoblastoma 1 protein modulates XY germ cell entry into G1/G0 arrest during fetal development in mice** *Biology of reproduction* **82**:433–443

13 Saba R., Kato Y., Saga Y. (2014) **NANOS2 promotes male germ cell development independent of meiosis suppression** *Developmental biology* **385**:32–40

14 Cook M. S., Munger S. C., Nadeau J. H., Capel B. (2011) **Regulation of male germ cell cycle arrest and differentiation by DND1 is modulated by genetic background** *Development* **138**:23–32

15 Houmard B., et al. (2009) **Global gene expression in the human fetal testis and ovary** *Biology of reproduction* **81**:438–443

16 Jameson S. A., et al. (2012) **Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad** *PLoS genetics* **8**

17 Lesch B. J., Dokshin G. A., Young R. A., McCarrey J. R., Page D. C. (2013) **A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis** *Proceedings of the National Academy of Sciences* **110**:16061–16066

18 Rolland A. D., Lehmann K. P., Johnson K. J., Gaido K. W., Koopman P. (2011) **Uncovering gene regulatory networks during mouse fetal germ cell development** *Biology of reproduction* **84**:790–800

19 Bingham N. C., Verma Kurvari S., Parada L. F., Parker K. L. (2006) **Development of a steroidogenic factor 1/Cre transgenic mouse line** *genesis* **44**:419–424

20 Capel B., Batchvarov J. (2008) **Sex chromatin staining in amnion cells** *CSH protocols* **2008**:pdb.prot5079–pdb.prot5079

21 Zheng G. X., et al. (2017) **Massively parallel digital transcriptional profiling of single cells** *Nature communications* **8**

22 Satpathy A. T., et al. (2019) **Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion** *Nature biotechnology* **37**:925–936

23 Griffiths J. A., Richard A. C., Bach K., Lun A. T., Marioni J. C. (2018) **Detection and removal of barcode swapping in single-cell RNA-seq data** *Nature communications* **9**

24 Satija R., Farrell J. A., Gennert D., Schier A. F., Regev A. (2015) **Spatial reconstruction of single-cell gene expression data** *Nature biotechnology* **33**:495–502

25 Stuart T., Srivastava A., Madad S., Lareau C. A., Satija R. (2021) **Single-cell chromatin state analysis with Signac** *Nature methods* **18**:1333–1341

26 Germain P.-L., Lun A., Meixide C. G., Macnair W., Robinson M. D. (2021) **Doublet identification in single-cell sequencing data using scDblFinder** *F1000Research* **10**

27 Yang S., et al. (2020) **Decontamination of ambient RNA in single-cell RNA-seq with DecontX** *Genome biology* **21**:1–15

28 Andreatta M., Carmona S. J. (2021) **UCell: Robust and scalable single-cell gene signature scoring** *Computational and structural biotechnology journal* **19**:3796–3798

29 van den Bergen J. A., Miles D. C., Sinclair A. H., Western P. S. (2009) **Normalizing gene expression levels in mouse fetal germ cells** *Biology of reproduction* **81**:362–370

30 Stévant I., et al. (2019) **Dissecting cell lineage specification and sex fate determination in gonadal somatic cells using single-cell transcriptomics** *Cell reports* **26**:3272–3283

31 Zhao C., et al. (2009) **Identification of the most sensitive and robust immunohistochemical markers in different categories of ovarian sex cord-stromal tumors** *The American journal of surgical pathology* **33**:354–366

32 Denzer L., Muranyi W., Schroten H., Schwerk C. (2023) **The role of PLVAP in endothelial cells** *Cell and Tissue Research* :1–20

33 Kelly L. M., Englmeier U., Lafon I., Sieweke M. H., Graf T. (2000) **MafB is an inducer of monocytic differentiation** *The EMBO journal* **19**:1987–1997

34 Garcia-Alonso L., et al. (2022) **Single-cell roadmap of human gonadal development** *Nature* **607**:540–547

35 Kuony A., Michon F. (2017) **Epithelial markers aSMA, Krt14, and Krt19 unveil elements of murine lacrimal gland morphogenesis and maturation** *Frontiers in physiology* **8**

36 Watt S. M., Gschmeissner S. E., Bates P. A. (1995) **PECAM-1: its expression and function as a cell adhesion molecule on hemopoietic and endothelial cells** *Leukemia & lymphoma* **17**:229–244

37 Zhang Y., et al. (2008) **Model-based analysis of ChIP-Seq (MACS)** *Genome biology* **9**:1–9

38 Sun H., et al. (2018) **Quantitative integration of epigenomic variation and transcription factor binding using MAmotif toolkit identifies an important role of IRF2 as transcription activator at gene promoters** *Cell discovery* **4**

39 Wang Q., et al. (2022) **Exploring epigenomic datasets by ChIPseeker** *Current Protocols* **2**

40 Yu G., Wang L.-G., He Q.-Y. (2015) **ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization** *Bioinformatics* **31**:2382–2383

41 Wu T., et al. (2021) **clusterProfiler 4.0: A universal enrichment tool for interpreting omics data** *The innovation* **2**

42 Yu G. (2021) **Enrichplot: visualization of functional enrichment result** *R package version* **1**

43 Zheng R., et al. (2019) **Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis** *Nucleic acids research* **47**:D729–D735

44 Mei S., et al. (2016) **Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse** *Nucleic acids research*

45 Troulé K., et al. (2023) **CellPhoneDB v5: inferring cell-cell communication from single-cell multiomics data** *arXiv*

46 Becht E., et al. (2019) **Dimensionality reduction for visualizing single-cell data using UMAP** *Nature biotechnology* **37**:38–44

47 Gray S., Cohen P. E. (2016) **Control of meiotic crossovers: from double-strand break formation to designation** *Annual review of genetics* **50**:175–210

48 Yang C., Croteau S., Hardy P. (2021) **Histone deacetylase (HDAC) 9: versatile biological functions and emerging roles in human cancer** *Cellular Oncology* **44**:997–1017

49 Herman L., Todeschini A.-L., Veitia R. A. (2021) **Forkhead transcription factors in health and disease** *Trends in Genetics* **37**:460–475

50 Pierce S. B., et al. (2013) **Mutations in LARS2, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in Perrault syndrome** *The American Journal of Human Genetics* **92**:614–620

51 da Silva Martinelli C. M., et al. (2017) **MGMT and CALCA promoter methylation are associated with poor prognosis in testicular germ cell tumor patients** *Oncotarget* **8**

52 Stuart T., et al. (2019) **Comprehensive integration of single-cell data** *Cell* **177**:1888–1902

53 Cao J., et al. (2019) **The single-cell transcriptional landscape of mammalian organogenesis** *Nature* **566**:496–502

54 Soygur B., et al. (2021) **Intercellular bridges coordinate the transition from pluripotency to meiosis in mouse fetal oocytes** *Science advances* **7**

55 Suzuki A., Niimi Y., Saga Y. (2014) **Interaction of NANOS2 and NANOS3 with different components of the CNOT complex may contribute to the functional differences in mouse male germ cells** *Biology open* **3**:1207–1216

56 Dreyer S. D., Zheng Q., Zabel B., Winterpacht A., Lee B. (1999) **Isolation, characterization, and mapping of a zinc finger gene, ZFP95, containing both a SCAN box and an alternatively spliced KRAB A domain** *Genomics* **62**:119–122

57 Li J., et al. (2022) **ZKSCAN5 activates VEGFC expression by recruiting SETD7 to promote the Lymphangiogenesis, tumour growth, and metastasis of breast cancer** *Frontiers in Oncology* **12**

58 Jang M. K., et al. (2005) **The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription** *Molecular cell* **19**:523–534

59 Yokobayashi S., et al. (2013) **PRC1 coordinates timing of sexual differentiation of female primordial germ cells** *Nature* **495**:236–240

60 Le Rolle M., et al. (2021) **Arrest of WNT/β-catenin signaling enables the transition from pluripotent to differentiated germ cells in mouse ovaries** *Proceedings of the National Academy of Sciences* **118**

61 Krentz A. D., et al. (2011) **DMRT1 promotes oogenesis by transcriptional activation of Stra8 in the mammalian fetal ovary** *Developmental biology* **356**:63–70

62 Mehta G. D., Kumar R., Srivastava S., Ghosh S. K. (2013) **Cohesin: functions beyond sister chromatid cohesion** *FEBS letters* **587**:2299–2312

63 Garcia-Cruz R., et al. (2010) **Dynamics of cohesin proteins REC8, STAG3, SMC1 β and SMC3 are consistent with a role in sister chromatid cohesion during meiosis in human oocytes** *Human reproduction* **25**:2316–2327

64 Endo T., Mikedis M. M., Nicholls P. K., Page D. C., de Rooij D. G. (2019) **Retinoic acid and germ cell development in the ovary and testis** *Biomolecules* **9**

65 Vernet N., et al. (2020) **Meiosis occurs normally in the fetal ovary of mice lacking all retinoic acid receptors** *Science advances* **6**

66 Bowles J., et al. (2018) **Retinoic acid antagonizes testis development in mice** *Cell reports* **24**:1330–1341

67 Bowles J., et al. (2006) **Retinoid signaling determines germ cell fate in mice** *Science* **312**:596–600

68 Chassot A.-A., et al. (2020) **Retinoic acid synthesis by ALDH1A proteins is dispensable for meiosis initiation in the mouse fetal ovary** *Science advances* **6**

69 Napoli J. L., Yoo H. S. (2020) **Methods in enzymology** Elsevier :55–75

70 Schep A. N., Wu B., Buenrostro J. D., Greenleaf W. J. (2017) **chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data** *Nature methods* **14**:975–978

71 Madan B., et al. (2016) **Wnt addiction of genetically defined cancers reversed by PORCN inhibition** *Oncogene* **35**:2197–2207

72 Li H., Kim J. H., Koh S. S., Stallcup M. R. (2004) **Synergistic effects of coactivators GRIP1 and β-catenin on gene activation: cross-talk between androgen receptor and Wnt signaling pathways** *Journal of Biological Chemistry* **279**:4212–4220

73 Ma Z., Lou S., Jiang Z. (2020) **PHLDA2 regulates EMT and autophagy in colorectal cancer via the PI3K/AKT signaling pathway** *Aging (Albany NY)* **12**

74 Navarro-Costa P., et al. (2016) **Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling** *Nature communications* **7**

75 Klein R. H., Knoepfler P. S. (2021) **DPPA2, DPPA4, and other DPPA factor epigenomic functions in cell fate and cancer** *Stem cell reports* **16**:2844–2851

76 Lam E. W.-F., Brosens J. J., Gomes A. R., Koo C.-Y. (2013) **Forkhead box proteins: tuning forks for transcriptional harmony** *Nature Reviews Cancer* **13**:482–495

77 Ma S., et al. (2020) **Chromatin potential identified by shared single-cell profiling of RNA and chromatin** *Cell* **183**:1103–1116

78 Mencacci N. E., et al. (2021) **Biallelic variants in TSPOAP1, encoding the active-zone protein RIMBP1, cause autosomal recessive dystonia** *The Journal of Clinical Investigation* **131**

79 Galán-Martínez J., et al. (2022) **TCFL5 deficiency impairs the pachytene to diplotene transition during spermatogenesis in the mouse** *Scientific Reports* **12**

80 Kojima Y., et al. (2021) **GATA transcription factors, SOX17 and TFAP2C, drive the human germ-cell specification program** *Life Science Alliance* **4**

81 Zhang F.-L., et al. (2023) **Single cell epigenomic and transcriptomic analysis uncovers potential transcription factors regulating mitotic/meiotic switch** *Cell Death & Disease* **14**

82 Zaret K. S. (2020) **Pioneer transcription factors initiating gene network changes** *Annual review of genetics* **54**:367–385

83 Xiao W., et al. (2022) **POU6F1 cooperates with RORA to suppress the proliferation of lung adenocarcinoma by downregulating HIF1A signaling pathway** *Cell Death & Disease* **13**

84 Yu K., et al. (2015) **Identification of a lipid scrambling domain in ANO6/TMEM16F** *elife* **4**

85 Zhao Z., et al. (2010) **Rab5a overexpression promoting ovarian cancer cell proliferation may be associated with APPL1 related epidermal growth factor signaling pathway** *Cancer science* **101**:1454–1462

86 Pan M.-H., et al. (2020) **FMNL3 regulates FASCIN for actin-mediated spindle migration and cytokinesis in mouse oocytes** *Biology of reproduction* **102**:1203–1212

87 Oyama T., et al. (2011) **Mastermind-like 1 (MamL1) and mastermind-like 3 (MamL3) are essential for Notch signaling in vivo** *Development* **138**:5235–5246

88 Lehto M., Olkkonen V. M. (2003) **The OSBP-related proteins: a novel protein family involved in vesicle transport, cellular lipid metabolism, and cell signalling** *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1631**:1–11

89 Turnbull A. P., et al. (2017) **Molecular basis of USP7 inhibition by selective small-molecule inhibitors** *Nature* **550**:481–486

90 Zheng G., Fu Y., He C. (2014) **Nucleic acid oxidation in DNA damage repair and epigenetics** *Chemical reviews* **114**:4602–4620

91 Jia S., Meng A. (2021) **TGFβ family signaling and development** *Development* **148**

92 Bachus S., et al. (2022) **In Mitosis You Are Not: The NIMA Family of Kinases in Aspergillus, Yeast, and Mammals** *International Journal of Molecular Sciences* **23**

93 Li M. (2020) **AQR Plays a Key Role in Splicing Fidelity in Higher Organisms** *Research Square*

94 Tigan A., Bellutti F., Kollmann K., Tebb G., Sexl V. (2016) **CDK6—a review of the past and a glimpse into the future: from cell-cycle control to transcriptional regulation** *Oncogene* **35**:3083–3091

95 Song Q., et al. (2023) **Knockout of ENO1 leads to metabolism reprogramming and tumor retardation in pancreatic cancer** *Frontiers in Oncology* **13**

96 Choi Y., et al. (2016) **Targeting ODC1 inhibits tumor growth through reduction of lipid metabolism in human hepatocellular carcinoma** *Biochemical and biophysical research communications* **478**:1674–1681

97 Quignon F., et al. (1998) **PML induces a novel caspase-independent death process** *Nature genetics* **20**:259–265

98 Farquharson C., Ahmed S. F. (2013) **Inflammation and linear bone growth: the inhibitory role of SOCS2 on GH/IGF-1 signaling** *Pediatric nephrology* **28**:547–556

99 McLaren A., Southee D. (1997) **Entry of mouse embryonic germ cells into meiosis** *Developmental biology* **187**:107–113

100 Evans E., Ford C., Lyon M. (1977) **Direct evidence of the capacity of the XY germ cell in the mouse to become an oocyte** *Nature* **267**:430–431

101 Da Silva S. M., et al. (2004) **Expression of activin subunits and receptors in the developing human ovary: activin A promotes germ cell survival and proliferation before primordial follicle formation** *Developmental biology* **266**:334–345

102 Nicol B., et al. (2019) **RUNX1 maintains the identity of the fetal ovary through an interplay with FOXL2** *Nature communications* **10**

103 Jazwiec P. A., Sloboda D. M. (2019) **Nutritional adversity, sex and reproduction: 30 years of DOHaD and what have we learned?** *Journal of Endocrinology* **242**:T51–T68

104 Mayère C., et al. (2021) **Single-cell transcriptomics reveal temporal dynamics of critical regulators of germ cell fate during mouse sex determination** *The FASEB Journal* **35** https://doi.org/10.1096/fj.202002420R

105 Chen M., et al. (2022) **Integration of single-cell transcriptome and chromatin accessibility of early gonads development among goats, pigs, macaques, and humans** *Cell Reports* **41**

106 Nguyen D. H., Jaszczak R. G., Laird D. J. (2019) **Heterogeneity of primordial germ cells** *Current topics in developmental biology* **135**:155–201

107 Bonifer C., Cockerill P. N. (2017) **Chromatin priming of genes in development: Concepts, mechanisms and consequences** *Experimental hematology* **49**:1–8

108 Tee W.-W., Reinberg D. (2014) **Chromatin features and the epigenetic regulation of pluripotency states in ESCs** *Development* **141**:2376–2390

109 Raccaud M., Suter D. M. (2018) **Transcription factor retention on mitotic chromosomes: regulatory mechanisms and impact on cell fate decisions** *FEBS letters* **592**:878–887

110 Wiltshire T., Park C., Handel M. A. (1998) **Chromatin configuration during meiosis I prophase of spermatogenesis** *Molecular Reproduction and Development: Incorporating Gamete Research* **49**:70–80

111 Seisenberger S., et al. (2012) **The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells** *Molecular cell* **48**:849–862

112 Huang T.-C., et al. (2021) **Sex-specific chromatin remodelling safeguards transcription in germ cells** *Nature* **600**:737–742

113 Laronda M. M. (2023) **Factors within the Developing Embryo and Ovarian Microenvironment That Influence Primordial Germ Cell Fate** *Sexual Development* :1–11

114 Chen D., et al. (2019) **Human primordial germ cells are specified from lineage-primed progenitors** *Cell reports* **29**:4568–4582

115 Schemmer J., et al. (2013) **Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice** *PloS one* **8**

116 Schäfer S., et al. (2011) **The role of BLIMP1 and its putative downstream target TFAP2C in germ cell development and germ cell tumours** *International journal of andrology* **34**:e152–e159

117 Pastor W. A., et al. (2018) **TFAP2C regulates transcription in human naive pluripotency by opening enhancers** *Nature cell biology* **20**:553–564

118 Woodfield G. W., Horan A. D., Chen Y., Weigel R. J. (2007) **TFAP2C controls hormone response in breast cancer cells through multiple pathways of estrogen signaling** *Cancer research* **67**:8439–8443

119 Weber S., et al. (2010) **Critical function of AP-2gamma/TCFAP2C in mouse embryonic germ cell maintenance** *Biology of reproduction* **82**:214–223

120 Goertz M. J., Wu Z., Gallardo T. D., Hamra F. K., Castrillon D. H. (2011) **Foxo1 is required in mouse spermatogonial stem cells for their maintenance and the initiation of spermatogenesis** *The Journal of clinical investigation* **121**:3456–3466

121 Shen Z., et al. (2022) **The function of Foxo1 in spermatogonia development is independent of PI3K/PTEN signaling** *The FASEB Journal* **36**

122 Xing, Y.-q.,, et al. (2018) **The regulation of FOXO1 and its role in disease progression** *Life sciences* **193**:124–131

123 Kang Y., Zhang K., Sun L., Zhang Y. (2022) **Regulation and roles of FOXK2 in cancer** *Frontiers in Oncology* **12**

124 Malik V., Zimmer D., Jauch R. (2018) **Diversity among POU transcription factors in chromatin recognition and cell fate reprogramming** *Cellular and Molecular Life Sciences* **75**:1587–1612

125 Yoshioka N., Suzuki N., Uekawa A., Kiguchi K., Ishizuka B. (2009) **POU6F1 is the transcription factor that might be involved in cell proliferation of clear cell adenocarcinoma of the ovary** *Human cell* **22**:94–100

126 Cho H.-J., et al. (2023) **POU6F2 mutation in humans with pubertal failure alters GnRH transcript expression** *Frontiers in Endocrinology* **14**

127 van Loon K., Huijbers E. J., Griffioen A. W. (2021) **Secreted frizzled-related protein 2: a key player in noncanonical Wnt signaling and tumor angiogenesis** *Cancer and Metastasis Reviews* **40**:191–203

128 Koubova J., et al. (2014) **Retinoic acid activates two pathways required for meiosis in mice** *PLoS genetics* **10**

129 Filippou P. S., Karagiannis G. S., Constantinidou A. (2020) **Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target** *Oncogene* **39**:2040–2054

130 Muramatsu H., et al. (2006) **Female infertility in mice deficient in midkine and pleiotrophin, which form a distinct family of growth factors** *Genes to Cells* **11**:1405–1417

131 Yao H. H.-C., Whoriskey W., Capel B. (2002) **Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis** *Genes & development* **16**:1433–1440

132 Bitgood M. J., Shen L., McMahon A. P. (1996) **Sertoli cell signaling by Desert hedgehog regulates the male germline** *Current biology* **6**:298–304

Author information

Adriana K Alexander

Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States

Karina F Rodriguez
Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States

Yu-Ying Chen
Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States

Ciro M Amato
Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States

Martin A Estermann
Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States
ORCID iD: 0000-0002-8623-2720

Barbara Nicol
Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States

Xin Xu
Epigenetics & Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Durham, United States

Humphrey Hung-Chang Yao
Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Durham, United States
ORCID iD: 0000-0003-2944-8469

For correspondence: humphrey.yao@nih.gov

Editors

Reviewing Editor
Wei Yan
Washington State University, Pullman, United States of America

Senior Editor
Detlef Weigel
Max Planck Institute for Biology Tübingen, Tübingen, Germany

Reviewer #1 (Public review):

Summary:

This study uses single nucleus multi-omics to profile the transcriptome and chromatin accessibility of mouse XX and XY primordial germ cells (PGCs) at three time points spanning PGC sexual differentiation and entry of XX PGCs into meiosis (embryonic days 11.5-13.5). They find that PGCs can be clustered into sub-populations at each time point, with higher heterogeneity among XX PGCs and more switch-like developmental transitions evident in XY PGCs. In addition, they identify several transcription factors that appear to regulate sex-

specific pathways as well as cell-cell communication pathways that may be involved in regulating XX vs XY PGC fate transitions. The findings are important and overall rigorous. The study could be further improved by better connection to the biological system, including putting the transcriptional heterogeneity of XX PGCs in the context of findings that meiotic entry is spatially asynchronous in the fetal ovary and further addressing the role of retinoic acid signaling. Overall, this study represents and advance in germ cell regulatory biology and will be a highly used resource in the field of germ cell development.

Strengths:

(1) The multi-omics data is mostly rigorously collected and carefully interpreted.

(2) The dataset is extremely valuable and helps to answer many long-standing questions in the field.

(3) In general, the conclusions are well anchored in the biology of the germ line in mammals.

Comments on revised version:

Most of my concerns have been addressed in the revised manuscript. I have one remaining concern but I believe this is important in order for the paper to be fully appreciated:

In Figures 2a, 2e, 3a, and 3e, the visualization scheme is very difficult to follow, and has not been updated or improved in the revised manuscript. It's very hard to see the colors corresponding to average expression for many genes because the circles are so small. The yellow color is hard to see and makes it hard to estimate the size of the circle. This issue is particularly egregious in Figure 2a for the data relating to ZKSCAN5, which is specifically highlighted in the text in lines 421-426. This data must be shown in a more convincing way in order to make the claims. An update to the visualization, including color scheme, is very strongly recommended; it is not difficult and would substantially improve the ability of these panels to communicate their message.

https://doi.org/10.7554/eLife.96591.2.sa3

Reviewer #2 (Public review):

Summary:

This manuscript by Alexander et al describes a careful and rigorous application of multiomics to mouse primordial germ cells (PGCs) and their surrounding gonadal cells during the period of sex differentiation.

Strengths:

In thoughtfully designed figures, the authors identify both known and new candidate gene regulatory networks in differentiating XX and XY PGCs and sex-specific interactions of PGCs with supporting cells. In XY germ cells, novel findings include the predicted set of TFs regulating Bnc2, which is known to promote mitotic arrest, as well as the TFs POU6F1/2 and FOXK2 and their predicted targets that function in mitosis and signal transduction. In XX germ cells, the authors deconstruct the regulation of the premeiotic replication factor Stra8, which reveals TFs involved in meiosis, retinoic acid signaling, pluripotency and epigenetics among predictions; this finding, along with evidence supporting regulatory potential of retinoic acid receptors in meiotic gene expression is an important addition to the debate over the necessity of retinoic acid in XX meiotic initiation. In addition, a self-regulatory network of other TFs is hypothesized in XX differentiating PGCs, including TFAP2c, TCF5, ZFX, MGA and NR6A1, which is predicted to turn on meiotic and Wnt signaling targets. Finally, analysis of PGC-support cell interactions during sex differentiation reveals substantially more

interactions in XX, via WNTs and BMPs, as well as some new signaling pathways that predominate in XY PGCs including ephrins, CADM1, Desert Hedgehog and matrix metalloproteases. This dataset will be an excellent resource for the community, motivating functional studies and serving as a discovery platform.

Weaknesses:

While the authors performed all of their comparisons between XX versus XY datasets at each timepoint, a more systematic analysis of expression and accessibility changes across time for each sex would be valuable. It remains possible that common mechanisms of differentiation to XX and XY could be missing from this analysis that focused on sex-specific differences.

Specific Questions:

(1) Line 461: "the population of E13.5 XX PGCs displaying the strongest Stra8 expression levels corresponded to the same population of XX PGCs with the highest module score of early meiotic prophase I genes (Fig. 3c; Supplementary Fig. 3a-b)" however the Stra8+ XX PGCs that do not robustly express meiotic genes should be examined to understand more about their differentiation potential. The authors are well-poised to identify the likely trajectories available to cell subsets in their dataset, and not doing so is a missed opportunity.

(2) The authors state that "we found that Stra8, Rec8, Rnf2, Sycp1, Sycp2, Ccnb3, and Zglp1 contain the RA receptor motifs in their regulatory sequences (Supplementary Figure 4g)." What is the strength of the RA->meiosis pathway compared to other mechanisms regulating meiosis? Perhaps the authors could take this analysis further with the following questions: (1) ask whether meiotic genes more enriched in RA motifs compared to other expressed genes or other motifs (2) compare the strength of peak-gene correlations for all peaks containing RA receptor motifs vs. those with peaks for Zglp1, Rnf2, etc binding. The strengths of these correlations could provide clues to how much gene expression varies in response to RA exposure vs. modulation of these other factors and thus tell us something about how much RA is playing a role.

(3) In figure 4, the shift from promoters in E11.5 XX PGCs to distal intergenic regions is fascinating. What can we learn about epigenetic reprogramming/methylation changes across gene bodies?

(4) The overlap between gene targets of TCFL5 with other highly expressed TFs differentially upregulated in E13.5 XX PGCs over XY suggests ambiguity regarding its role as a central or high-level regulator of differentiation; as in vivo validation has not been performed, I suggest softening this conclusion.

https://doi.org/10.7554/eLife.96591.2.sa2

Reviewer #3 (Public review):

Summary:

Alexander et al. reported the gene-regulatory networks underpinning sex determination of murine primordial germ cells (PGCs) through single-nucleus multiomics, offering a detailed chromatin accessibility and gene expression map across three embryonic stages in both male (XY) and female (XX) mice. It highlights how regulatory element accessibility may precede gene expression, pointing to chromatin accessibility as a primer for lineage commitment before differentiation. Sexual dimorphism in these elements and gene expression increases over time, and the study maps transcription factors regulating sexually dimorphic genes in PGCs, identifying sex-specific enrichment in various transcription factors.

Strengths:

The study includes step-wise multiomic analysis with some computational approach to identify candidate TFs regulating XX and XY PGC gene expression, providing a detailed timeline of chromatin accessibility and gene expression during PGC development, which identifies previously unknown PGC subpopulations and offers a multimodal reference atlas of differentiating PGC clusters. Furthermore, the study maps a complex network of transcription factors associated with sex determination in PGCs, adding depth to our understanding of these processes.

Weaknesses:

While the multiomics approach is powerful, it primarily offers correlational insights between chromatin accessibility, gene expression, and transcription factor activity, without direct functional validation of identified regulatory networks.

Comments on revised version:

The authors have answered my questions and concerns in the revised manuscript and correspondence.

https://doi.org/10.7554/eLife.96591.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*
>
> *This study uses single nucleus multiomics to profile the transcriptome and chromatin accessibility of mouse XX and XY primordial germ cells (PGCs) at three time-points spanning PGC sexual differentiation and entry of XX PGCs into meiosis (embryonic days 11.5-13.5). They find that PGCs can be clustered into sub-populations at each time point, with higher heterogeneity among XX PGCs and more switch-like developmental transitions evident in XY PGCs. In addition, they identify several transcription factors that appear to regulate sex-specific pathways as well as cell-cell communication pathways that may be involved in regulating XX vs XY PGC fate transitions. The findings are important and overall rigorous. The study could be further improved by a better connection to the biological system, including the addition of experiments to validate the 'omics-based findings in vivo and putting the transcriptional heterogeneity of XX PGCs in the context of findings that meiotic entry is spatially asynchronous in the fetal ovary. Overall, this study represents an advance in germ cell regulatory biology and will be a highly used resource in the field of germ cell development.*
>
> *Strengths:*
>
> *(1) The multiomics data is mostly rigorously collected and carefully interpreted.*
>
> *(2) The dataset is extremely valuable and helps to answer many long-standing questions in the field.*

> *(3) In general, the conclusions are well anchored in the biology of the germ line in mammals.*
>
> *Weaknesses:*
>
> *(1) The nature of replicates in the data and how they are used in the analysis are not clearly presented in the main text or methods. To interpret the results, it is important to know how replicates were designed and how they were used. Two "technical" replicates are cited but it is not clear what this means.*

The two independent technical replicates comprised different pools of paired gonads. This sentence was added to the methods section of the revised manuscript.

> *(2) Transcriptional heterogeneity among XX PGCs is mentioned several times (e.g., lines 321-323) and is a major conclusion of the paper. It has been known for a long time that XX PGCs initiate meiosis in an anterior-to-posterior wave in the fetal ovary starting around E13.5. Some heterogeneity in the XX PGC populations could be explained by spatial position in the ovary without having to invoke novel subpopulations.*

We thank the reviewer for pointing out this important biological phenomenon. We also recognize that transcriptional heterogeneity among XX PGCs is likely due to the anterior-to-posterior wave of meiotic initiation in E13.5 ovaries and highlight this possibility in our manuscript. However, since our study utilizes single-nucleus RNA-sequencing and not spatial transcriptomics, we are not able to capture the spatial location of the XX PGCs analyzed in our dataset. As such, our analysis applied clustering tools to classify the populations of XX PGCs captured in our dataset.

> *(3) There is essentially no validation of any of the conclusions. Heterogeneity in the expression of a given marker could be assessed by immunofluorescence or RNAscope.*

In our revised manuscript, we included immunofluorescence staining of potential candidate factors involved in PGC sex determination, such as PORCN and TFAP2C. Testing and optimizing antibodies for the targets identified in this study are ongoing efforts in our lab and we look forward to sharing our results with the research community.

> *(4) The paper sometimes suffers from a problem common to large resource papers, which is that the discussion of specific genes or pathways seems incomplete. An example here is from the analysis of the regulation of the Bnc2 locus, which seems superficial. Relatedly, although many genes and pathways are nominated for important PGC functions, there is no strong major conclusion from the paper overall.*

In this manuscript, we set out to identify candidate factors, some already known and many others unknown, involved in the developmental pathways of PGC sex determination using computational tools. Our goal, as a research group and with future collaborators, is to screen these interesting candidates and discover their function in the primordial germ cell. Our research, presented in this study, represents a launching pad for which to identify future projects that will investigate these factors in further detail.

> ***Reviewer #2 (Public Review):***
>
> *Summary:*
>
> *This manuscript by Alexander et al describes a careful and rigorous application of multiomics to mouse primordial germ cells (PGCs) and their surrounding gonadal cells during the period of sex differentiation.*

Strengths:

In thoughtfully designed figures, the authors identify both known and new candidate gene regulatory networks in differentiating XX and XY PGCs and sex-specific interactions of PGCs with supporting cells. In XY germ cells, novel findings include the predicted set of TFs regulating Bnc2, which is known to promote mitotic arrest, as well as the TFs POU6F1/2 and FOXK2 and their predicted targets that function in mitosis and signal transduction. In XX germ cells, the authors deconstruct the regulation of the premeiotic replication regulator Stra8, which reveals TFs involved in meiosis, retinoic acid signaling, pluripotency, and epigenetics among predictions; this finding, along with evidence supporting the regulatory potential of retinoic acid receptors in meiotic gene expression is an important addition to the debate over the necessity of retinoic acid in XX meiotic initiation. In addition, a self-regulatory network of other TFs is hypothesized in XX differentiating PGCs, including TFAP2c, TCF5, ZFX, MGA, and NR6A1, which is predicted to turn on meiotic and Wnt signaling targets. Finally, analysis of PGC-support cell interactions during sex differentiation reveals more interactions in XX, via WNTs and BMPs, as well as some new signaling pathways that predominate in XY PGCs including ephrins, CADM1, Desert Hedgehog, and matrix metalloproteases. This dataset will be an excellent resource for the community, motivating functional studies and serving as a discovery platform.

Weaknesses:

My one major concern is that the conclusion that PGC sex differentiation (as read out by transcription) involves chromatin priming is overstated. The evidence presented in the figures includes a select handful of genes including Porcn, Rimbp1, Stra8, and Bnc2 for which chromatin accessibility precedes expression. Given that the authors performed all of their comparisons between XX versus XY datasets at each timepoint, have they missed an important comparison that would be a more direct test of chromatin priming: between timepoints for each sex? Furthermore, it remains possible that common mechanisms of differentiation to XX and XY could be missing from this analysis that focused on sexspecific differences.

We thank the reviewer for their thoughtful assessment and suggestions, as stated here. We note that chromatin priming in PGCs prior to sex determination is a well-documented research finding (see references below), that is further supported by our single-nucleus multiomics data. To support these findings previously stated in the scientific literature, we included data demonstrating the asynchronous correlation between chromatin accessibility and gene expression during PGC sex determination. Specifically, we investigated the associations of differentially accessible chromatin peaks with differentially expressed gene expression for each PGC type (between sexes and across embryonic stages) using computational tools and methods that are well-established and applied by the research community. In our manuscript, we note that the patterns we identified support the potential role of chromatin priming in PGC sex determination. Nevertheless, we further highlight that a comprehensive profile of 3D chromatin structure and enhancer-promoter contacts in differentiating PGCs is needed to fully understand how changes to chromatin facilitate PGC sex determination.

References:

(1) Chen, M., et al. Integration of single-cell transcriptome and chromatin accessibility of early gonads development among goats, pigs, macaques, and humans. Cell Reports 41 (2022).

(2) Huang, T.-C. et al. Sex-specific chromatin remodelling safeguards transcription in germ cells. Nature 600, 737–742 (2021).

> ***Reviewer #3 (Public Review):***
>
> *Summary:*
>
> *Alexander et al. reported the gene-regulatory networks underpinning sex determination of murine primordial germ cells (PGCs) through single-nucleus multiomics, offering a detailed chromatin accessibility and gene expression map across three embryonic stages in both male (XY) and female (XX) mice. It highlights how regulatory element accessibility may precede gene expression, pointing to chromatin accessibility as a primer for lineage commitment before differentiation. Sexual dimorphism in these elements and gene expression increases over time, and the study maps transcription factors regulating sexually dimorphic genes in PGCs, identifying sex-specific enrichment in various transcription factors. Strengths:*
>
> *The study includes step-wise multiomic analysis with some computational approach to identify candidate TFs regulating XX and XY PGC gene expression, providing a detailed timeline of chromatin accessibility and gene expression during PGC development, which identifies previously unknown PGC subpopulations and offers a multimodal reference atlas of differentiating PGC clusters. Furthermore, the study maps a complex network of transcription factors associated with sex determination in PGCs, adding depth to our understanding of these processes.*
>
> *Weaknesses:*
>
> *While the multiomics approach is powerful, it primarily offers correlational insights between chromatin accessibility, gene expression, and transcription factor activity, without direct functional validation of identified regulatory networks.*

As stated in our response above to a similar concern, we note that our research study represents a launching pad for which to identify future projects that will investigate candidates that may be involved in PGC sex determination, in further detail. With this rich dataset in hand, our goal in future research projects is to screen these candidates and discover their function in PGCs.

> ***Response to Recommendations***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *(1) Clarify at first introduction how combined ATAC-seq/RNA-seq mulitomics libraries were prepared, including if ATAC and RNA-seq data are from the same cell.*

This information was added to the introduction of the revised manuscript.

> *(2) Clarify what the two technical replicates represent. Are they two libraries from the same gonad or the same pool of gonads? Are they from 2 different gonads?*

The two independent technical replicates comprised different pools of paired gonads. This sentence was added to the methods section of the revised manuscript.

> *(3) In Supplemental Figure 1, there is substantial variation in the number of unique snATAC-seq fragments between some conditions. Could this create a systematic bias that affects clustering?*

We recognize the concern that substantial variation in the number of unique snATAC-seq fragments between conditions could potentially create a systematic bias that affects

clustering. However, we analyzed our snATAC-seq dataset with Signac, which performs term frequency-inverse document frequency (TF-IDF) normalization. This is a process that normalizes across cells to correct for differences in cellular sequencing depth. Given that sequencing depth was taken into account in our normalization and clustering procedures, and that the unbiased clustering of PGCs also reflects the sex and embryonic stage of PGCs, we are confident that the clustering of the snATAC-seq datasets closely reflects the biological variability present in the PGCs collected.

References:

Signac Website: https://stuartlab.org/signac/articles/pbmc_vignette

Stuart, T., Srivastava, A., Madad, S., Lareau, C. A., & Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nature methods, 18(11), 1333-1341.

> *(4) In Figures 2a, 2e, 3a, and 3e, the visualization scheme is very difficult to follow. It's very hard to see the colors corresponding to average expression for many genes because the circles are so small. In addition, the yellow color is hard to see and makes it hard to estimate the size of the circle since the boundaries can be indistinct. I recommend using a different visualization scheme and/or set of size scales be used.*

In Figures 2a, 2e, 3a, and 3e, we chose this color palette to be inclusive of viewers who are colorblind. The chosen colors are visible on both a computer screen and on printed paper. We also included a legend of the color scale and dot size representing the average expression and percent of cells expressing the gene, respectively. If the color cannot be seen, it is because the cell population is not expressing the gene.

> *(5) Perform in vivo validation (immunofluorescence or RNAscope) of at least some targets implicated in PGC development by this study.*

Such validations (immunofluorescence staining of PORCN and TFAP2C) are now included in Figure 4 and the supplement.

> *(6) In line 351, the authors state that "we observed a strong demarcation between XX and XY PGCs at E12.5-E13.5." But in Figure 1j it looks like a reasonably high fraction of both XX and XY E12.5 cells are in cluster 1, which should mean that there is some overlap.*

While it is true that Figure 1j shows overlap of both XX and XY E12.5 cells in cluster 1, we were commenting on the separation of E12.5 XX (clusters 4 and 5) and E12.5 XY (clusters 8 and 9) PGCs. We have modified the sentence beginning at line 351 to state that the separation between XX and XY PGCs occurs at E13.5.

> *(7) In lines 404-405: "We first linked snATAC-seq peaks to XY PGC functional genes". It is important to know how the peaks were linked to genes.*

We added the following sentence to address this comment: “Peak-to-gene linkages were determined using Signac functionalities and were derived from the correlation between peak accessibility and the intensity of gene expression.”

> *(8) In Supplemental Figure 5c, the XX E11.5 condition has a substantially higher fraction of ATAC peaks at promoter regions compared to the others. Does this have statistical and biological significance?*

This is an interesting observation beyond the scope of our manuscript. Many interesting questions arise from this study and it is our plan to investigate further in the future.

> *(9) Line 885: "The increased number of DA peaks at E13.5 may be the result of changes to chromatin structure as XX PGCs enter meiotic prophase I"; but in Figure 4b, there's only a modest increase in DAP number from E12.5 to E13.5 in XX PGCs, compared to a massive gain in XY PGCs.*

In our manuscript, we comment on both phenomena: the doubling of differentially accessible peaks in XX PGCs from E12.5 to E13.5 and the massive increase in differentially accessible peaks in XY PGCs from E12.5 to E13.5. In our description of these results, we propose several hypotheses leading to these increases in differentially accessible peaks. As such, it cannot be ruled out that the changes to chromatin structure that occur during meiotic prophase I contribute to the gain in differentially accessible peaks in XX PGCs at E13.5, and we included this statement in the manuscript accordingly.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *(1) The methods state at line 141 that nuclei with mitochondrial reads of more than 25% were removed, however our understanding from the Bioconductor manual and companion manuscript (Amezquita, R.A., Lun, A.T.L., Becht, E. et al. Orchestrating single-cell analysis with Bioconductor. Nat Methods 17, 137-145 (2020). https://doi.org/10.1038/s41592-019-0654-x) is that snRNA-seq approaches remove mitochondrial transcripts entirely and datasets containing mitochondrial transcripts are thought to feature incompletely stripped nuclei. It is thought that mitochondrial transcripts participating in nuclear import may remain hanging on to the nuclear envelope and get encapsulated into GEMs. If the mitochondrial read cutoff of 25% was used intentionally to keep this potentially contaminating signal, please justify why this was done for this dataset.*

We agree with the reviewer that the presence of mitochondrial transcripts may be potentially contaminating signal. In our preprocessing steps, we removed the mitochondrial genes and transcripts from our datasets so that they would not influence or affect our analyses. The following sentence was added to the methods section on snRNA-seq data processing: “Mitochondrial genes and transcripts were removed from the snRNA-seq datasets to eliminate any potentially contaminating signal.”

> *(2) Methods line 227: please include log2fold change and p-adjusted value cutoffs for GO enrichment.*

We used clusterprofiler for our GO enrichment analysis. Our GO enrichment analysis did not include a log2fold change analysis and the p-adjusted value cutoff is stated in the methods.

> *(3) Results line 310: the claim that "At E12.5-E13.5, XY PGCs converged onto a single distinct population (cluster 7), indicating less transcriptional diversity among E12.5-E13.5 XY PGCs when compared to E12.5E13.5 XX PGCs (Fig1d)" would be strengthened if the authors quantified transcriptional distance with distance metrics such as euclidean or cosine distance.*

We used a clustering approach to gain insights into the transcriptional diversity of PGC populations. Using an additional metric, such as Euclidean or cosine distance, would not provide meaningful information not already achieved by clustering or change the conclusions presented in the manuscript.

> *(4) Results line 317: the authors allude to Lars2 defining clusters 2 & 3 as a marker gene, but it is not clear why this is highlighted until the reader reaches the discussion, which alludes to the published role of Lars2 in reproduction. Please consider moving this*

> *sentence to the results section for clarity and perhaps expanding the discussion on the meaning.*

To provide clarity, we added the statement “genes with reported roles in reproduction” to the results section.

> *(5) In Figure 2a, why do the authors choose to focus on Zkscan5 in XY PGCs when it is expressed by such a small portion of cells (<25%)? Do they assume that this is due to dropouts?*

We chose to focus on Zkscan5 as an example because of its enriched and differential expression in male PGCs, the motif for Zkscan5 is not enriched in female PGCs, and the reported roles of Zkscan5 in regulating cellular proliferation and growth. Zkscan5 is an example of how candidate genes can be identified for further investigation.

> *(6) Line 461: "the population of E13.5 XX PGCs displaying the strongest Stra8 expression levels corresponded to the same population of XX PGCs with the highest module score of early meiotic prophase I genes (Figure 3c; Supplementary Fig. 3a-b)". However did the authors also consider examining the Stra8+ XX PGCs that do not robustly express meiotic genes to understand more about their differentiation potential?*

We are thankful to the reviewer for this suggestion. However, this research question is beyond the scope of the manuscript. We plan to investigate further in future research studies.

> *(7) Line 505: "when we searched for the presence of RA receptor motifs in peaks linked to genes related to meiosis and female sex determination, we found that Stra8, Rec8, Rnf2, Sycp1, Sycp2, Ccnb3, and Zglp1 contain the RA receptor motifs in their regulatory sequences (Supplementary Figure 4g)." My read of the text is that the authors are not taking a side on the RA and meiosis controversy, but rather trying to reveal what the data can tell us, and the answer is that there is a strong signature linking RA to meiotic genes, which supports this as a valid biological pathway. But what is the strength of the RA>meiosis pathway compared to other mechanisms (which must be functioning in the triple receptor KO)? Perhaps the authors could take this analysis further with the following questions: (1) ask whether meiotic genes are more enriched in RA motifs compared to other expressed genes or other motifs (2) compare the strength of peak-gene correlations for all peaks containing RA receptor motifs vs. those with peaks for Zglp1, Rnf2, etc binding. The strengths of these correlations could provide clues to how much gene expression varies in response to RA exposure vs. modulation of these other factors and thus tell us something about how much RA is playing a role.*

We agree with the reviewer that this is a very interesting and important question. We also thank the reviewer for their thoughtful suggestions on the types of bioinformatics analyses that could answer this question. However, the section on RA signaling during PGC sex determination is only a small part of the manuscript and would be better analyzed in greater detail in a future research study or publication.

> *(8) The shift from promoters in E11.5 XX PGCs to distal intergenic regions is fascinating. What can we learn about epigenetic reprogramming/methylation changes across gene bodies?*

We agree with the reviewer that this is an interesting question about gene regulation in E11.5 XX PGCs. However, we prefer to analyze the epigenetic reprogramming changes across gene bodies in this cell population in additional research studies. Our purpose and goal for this

section was to link differentially accessible chromatin peaks with differentially expressed genes to identify putative gene regulatory networks.

> *(9) Line 581: why did the authors choose to highlight and validate PORCN1 in PGCs? Please elaborate.*

As stated in the manuscript, we chose to highlight and validate PORCN1 in PGCs because of its role in WNT signaling and because of the visibly strong correlation between chromatin accessibility at the XXenriched DAP in Fig. 4c (dashed box) and and gene expression of PORCN1.

> *(10) Figure 5f would be easier to interpret if presented as two columns rather than a circle; show one line of the proteins and the other line with the transcripts so that each is on the same line and there are connections between them.*

This comment is related to stylistic preferences. The purpose of Fig. 5f is to demonstrate that the candidate transcription factors may regulate the expression of other enriched transcription factors. Figure 5f figure accomplishes this goal.

> *(11) Line 640: "The predicted target genes of TCFL5 totaled 74% (367/494) of all DEGs with peak-to-gene linkages in XX PGCs". This seems like a high number and a lot of work for just TCFL5; given the overlap between other TFs and target genes, how many of these 367 target genes overlap with other TFs?*

We agree with the reviewer that this is an important declaration to make. We added the following sentence to the results section on TCFL5: “A large majority of the predicted target genes of TCFL5 were also predicted to be the target genes of the enriched TFs presented in Fig. 5e, e.g., the predicted target genes of these TFs overlapped with 4%-100% of the predicted target genes of TCFL5.”

> *(12) The presentation of TCFL5 in the results section would make more sense with the additional mention of reproductive phenotypes already known (currently in the discussion Lines 914-917). I would furthermore suggest that the discussion goes into more depth on the difference between the regulatory network of TCFL5 in XX meiosis vs XY.*

We thank the reviewer for this comment, however, we already state in the results section that TCFL5 is known to influence XX PGC sex determination.

> *(13) In the Methods, please state more clearly for those not familiar that the genetic background of mice is mixed.*

We described the mice with their official names, which provides the context of their genetic backgrounds.

> *(14) Please specify which morphologic criteria were used to verify the stage of embryos in the methods.*

We added the following text to the methods section of the revised manuscript: “Plug date was used to determine the stage of embryos collected for single-nucleus RNA-seq and ATAC-seq. The stage of E11.5 embryos was confirmed by counting somites. The stage of embryos collected at E12.5 was confirmed by the morphological presence of the vessel and cords of the testes collected from XY embryos. Similarly, we confirmed the stage of embryos collected at

E13.5 by the size of the gonads, the presence of more distinct cords in the testes of XY embryos, and the elongation of the ovaries of XX embryos."

> *(15) The total number of cells and PGCs that passed QC and are included in UMAPS should be stated.*

The requested information was added to the legend for Fig. 1 of the revised manuscript: "The number of PGCs per sex and embryonic stage are: 375 E11.5 XX PGCs; 1,106 E12.5 XX PGCs; 750 E13.5 XX PGCs; 110 E11.5 XY PGCs; 465 E12.5 XY PGCs; and 348 E13.5 XY PGCs."

> *(16) The order of timepoints changes between figures, and this is not for any obvious reason. Please make it consistent. Figures 1 and 6 list XX 11.5, 12.5, 13.5, and the same for XY, but Figures 2, 3, and 4 use the reverse order: XY E13.5, E12.5, E11.5, and then XX.*

We thank the reviewer for this comment. However, we chose this order for each of the figures to match the coordinates of the graphs and where we would expect the reader to begin reading the graph first. For example, in Figure 3a, XX E11.5 is closest to the x-axis and would be expected to be read first.

> *(17) In Figure S2 the colors of clusters are hard to distinguish, and it is suggested that the cluster numbers should be listed above each colored bar to avoid frustration.*

We made the suggested correction to Figure S2.

> *(18) In Figures 2e and 3e: what do the dashed boxes indicate?*

The dashed boxes are to guide the reader's eyes to the fact that the order of transcription factors/genes under the Cistrome DB regulatory potential score and gene expression plots are the same.

> *(19) In Figure 5a: break panels into i-iv so that the in-text call-outs are not all the same.*

We made the suggested correction to Figure 5a and modified the in-text call-outs.

> *(20) Please indicate XX in Figure 5e and XY in Figure 5l.*

We made the suggested correction to Figure 5e and 5l.

> *(21) In Figure S5c: Please reorganize DA chromatin peak charts so that columns are XX and XY with rows at the same timepoint.*

We made the suggested correction to Figure S5c.

> *(22) In Figure S7a: please make images larger so that the overlapping expression of PORCN and TRA98 is more visible, and consider adding a more magnified panel.*

This image is now included in the main text, with expanded panels.

> *(23) Line 742-754: this seems like a long introduction for the results section; please consider tightening it up.*

We believe this text is important and necessary to provide context to the bioinformatics analyses of cell signaling pathways in PGCs. Not all readers will be familiar with the ligand-

receptor signals between gonadal support cells and PGCs, and this text provides details on which signaling pathways are known to direct sex determination of PGCs.

> *(24) For UMAP plots in Figures 2c, 3c, S3b, and S4b, the text overlaid with the timepoints and sexes onto the UMAP plots is misleading, as it allows the reader to presume that the entire group of cells for a given sex/timepoint is located in the location of the text overlay. However, from the UMAP plots in Figure 1i-j, it is clear that the cells from a given sex/timepoint are actually spread across multiple identified clusters. Thus, the overlaid text obscures the important heterogeneity detected. To better represent the actual locations on the UMAP plot of cells from each sex/timepoint, it would be better to show inset density plots alongside these UMAP plots so the reader can locate the cells for themselves.*

We thank the reviewer for this comment. However, we chose this formatting to offer simplicity and ease of understanding to our UMAPs in addition to highlighting the general biological patterns of gene expression. If the reader is interested in discerning more of the heterogeneity of the UMAPs, they may refer back to Figure 1.

> ***Reviewer #3 (recommendations for the authors):***
>
> *There are some errors or places that need clarification or corrections:*
>
> *(1) Figure 1f, according to the graph, it should be 8 clusters, not 9.*

There are 9 clusters because the numbering for the clusters start at '0'.

> *(2) Why did cluster 8 have so many different states of cells from both sexes?*

The identification of cluster 8 is likely an artifact of sequencing, and would require several different analyses to figure out why cluster 8 has many different states of cells from both sexes. While this will address a technical issue associated with the dataset, this will not change any major conclusions of the study.

> *(3) Figure 1i, shouldn't that be ten instead of eleven?*

There are 11 clusters because the numbering for the clusters start at '0'.

> *(4) Figure 2a, zkscan expression level comparison was not so obvious as the bubble size was small. How many folds of differences from xx pgc?*

There is a 1.5 fold increase in the expression of Zkscan5 between XY and XX PGCs at E13.5. We included this information in the revised manuscript.

https://doi.org/10.7554/eLife.96591.2.sa0