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

Adriana K. Alexander[1], Karina F. Rodriguez[1], Yu-Ying Chen[1], Ciro M. Amato[1], Martin A. Estermann[1], Barbara Nicol[1], Xin Xu[2], and Humphrey Hung-Chang Yao[1*]

[1]Reproductive Developmental Biology Group, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

[2]Epigenetics & Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

[*]Corresponding author:

Humphrey H-C Yao

Reproductive Developmental Biology Group

National Institute of Environmental Health Sciences (NIEHS/NIH)

111 T.W. Alexander Dr, Mail Drop C4-10

Research Triangle Park, NC 27709

U.S.A.

Tel: 984-287-4004

Email: humphrey.yao@nih.gov

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.

46 **Introduction**

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

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

69 *Sox2*[8]. Consequently, there are multiple layers of gene regulation required for sex
70 determination of PGCs.
71 Beyond the patterns of gene expression in PGCs, relatively little is known about how
72 signals from the gonadal environment activate the expression of sexually dimorphic TFs
73 and genes in PGCs. First, the gene regulatory networks, *i.e.* TFs and their predicted target
74 genes, specific to XX and XY PGCs are not well defined. Second, it remains unclear how the
75 chromatin environment is temporally regulated during PGC sex determination. Finally,
76 additional data are needed on the patterns of ligand-receptor expression in gonadal
77 supporting cells and PGCs. Previous reports have used bulk gene regulation and
78 expression genomics assays to investigate the transcriptional programs underlying PGC
79 development[15-18]. However, these assays may not have the resolution or sensitivity
80 required to detect the transient changes in gene regulation among PGC subpopulations that
81 are essential to sex determination.
82 In the present study, we employed the integrative genomics method, combined
83 single-nucleus transcriptome and chromatin accessibility sequencing from the same cell, to
84 decipher various layers of gene regulation during PGC sex determination in mice. We
85 comprehensively profiled 3,054 XX and XY PGCs at embryonic days (E) E11.5, E12.5, and
86 E13.5, which covers the developmental time frame from bipotential to sexually
87 differentiated PGCs. Single-nucleus sequencing enabled the detection of sex-enriched
88 regulatory loci, TFs, and gonadal cues that may be responsible for initiating gene
89 expression in individual PGCs. We first systematically examined the molecular signatures
90 of PGC subpopulations to identify the genes and accessible chromatin regions underpinning
91 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 70–74 °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

121 away from the mesonephros, which was done using needles. Isolated fetal mouse gonads
122 were isolated in PBS, and pairs of gonads or groups pooled between fetuses of the same sex
123 and age were immediately frozen in liquid nitrogen for storage. Nuclei isolation was
124 performed using the 10X Genomics Chromium Nuclei Isolation Kit (catalog #1000494)
125 following the manufacturer's protocol. The total numbers of gonads used per sex and age
126 are listed in **Supplementary Table 1**. The 10X Genomics Chromium Next GEM Single Cell
127 Multiome ATAC + Gene Expression Library Preparation Kit (catalog #1000284) was used
128 to generate a 10X barcoded library of mRNA and transposed DNA from individual nuclei.
129 The 10X barcoded single-nuclei RNA and ATAC-seq libraries were obtained from 8,687
130 average (1,555-20,000) nuclei per sample and sequenced with Illumina NOVAseq to a
131 minimum sequencing depth of 129,120,442 raw reads and 22,406 read pairs/nucleus
132 (**Supplementary Table 1**). Two independent technical replicates of multiome sequencing
133 were performed for each sex by embryonic stage combination. The two independent
134 technical replicates comprised different pools of paired gonads.

135 **Supplementary Table 1. Sequencing and QC statistics for single-nucleus multiome**
136 **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

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).

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

157 subset of the snATAC-seq data was generated using the following cutoffs: nCount_ATAC <
158 100,000 & nCount_ATAC > 1000 & nucleosome_signal < 2 & TSS.enrichment > 1
159 (**Supplementary Figure 1b-d**). Analysis of the snATAC-seq peak feature distribution
160 demonstrated that most peaks occurred in distal intergenic or promoter regions of the
161 genome (**Supplementary Figure 1e**).
162 **Sex filtering.** To ensure that the correct sex was analyzed in all downstream analyses, we
163 applied computational filters to confirm the sex of the cells in each dataset. Cells were
164 filtered based on Y-linked gene expression and fragments mapped to the Y chromosome
165 (chrY). The *UCell* package[28] (v. 2.0.1) was used to create a module score of the Y-linked
166 genes *Kdm5d*, *Eif2s3y*, *Uty*, and *Ddx3y*. Module scores of E11.5 and E12.5 cells above one
167 standard deviation of the mean module score of all XY E13.5 cells were determined to be
168 chromosomally male. Module scores of E11.5 and E12.5 cells below one standard deviation
169 of the mean module score of all XX E13.5 cells were determined to be chromosomally
170 female. Next, we examined the peaks called on chrY outside of the pseudoautosomal region.
171 In total, 21 peaks were identified between chrY 1-90,000,000 bp. Cells with 0 fragment
172 counts within our defined chrY peak region or chrY gene expression module score < 1
173 standard deviation were removed from the male datasets. Cells with > 1 fragment counts
174 within our defined chrY peak region or chrY gene expression module score > 1 standard
175 deviation were removed from the female datasets.
176 **Cell type identification**
177 To identify the cellular origin of the snRNA-seq clusters, principal component analysis
178 (PCA) was run on the scaled data and visualization was performed using UMAP with dims =
179 18 and resolution = 0.3. Primordial germ cell (PGC) cluster identification was performed

180 using the established germ cell markers *Ddx4* and *Pou5f1*[29]. To remove potential PGC and
181 gonadal somatic cell doublets, only cells with RNA count < 1 for the following genes were
182 retained as PGCs: the granulosa cell marker genes *Foxl2* and *Runx1*[30]; Sertoli cell marker
183 *Sox9*[30]; Leydig cell marker *Insl3*[30]; stromal cell marker *Wt1*[31]; endothelial cell marker
184 *Plvap*[32]; immune-related cell marker *Mafb*[33]; interstitial cell markers *Pdgfra* and *Nr2f2*[30];
185 supporting cell marker *Tspan8*[34]; and epithelial cell marker *Krt19*[35]. Somatic supporting
186 cell cluster identification was performed using the established female and male support cell
187 markers *Runx1* and *Sox9*, respectively. To remove potential gonadal supporting cell
188 doublets formed with germ cells or interstitial cells, only cells with RNA count > 1 for *Wt1*
189 and RNA count < 1 for the following genes were retained as support cells: the germ cell
190 marker genes *Ddx4* and *Pou5f1*; endothelial cell markers *Plvap* and *Pecam1*[36]; and
191 interstitial cell markers *Pdgfra* and *Nr2f2*.

192 **Clustering and pseudotime trajectory of primordial germ cells**

193 snRNA-seq reclustering of PGCs was done with PCA on scaled PGC RNA data and
194 visualization was performed using UMAP with dimensions (dims) = 18 and resolution =
195 0.3. snATAC-seq clustering of PGCs was done first by calling peaks using *MACS2*[37] (v.
196 2.2.7.1). Peaks on nonstandard chromosomes were removed using the
197 "keepStandardChromosomes" function in *Signac* with the option 'pruning.mode = "coarse"'
198 and peaks in mm10 genomic blacklist regions were removed. Counts in each peak were
199 determined using *Signac* functionalities. The data was normalized by latent semantic
200 indexing (LSI) using the "RunTFIDF", "FindTopFeatures", and "RunSVD" functions in *Signac*.
201 Visualization was performed using UMAP with dims = 2:30, reduction = 'lsi', and algorithm
202 = 3. The joint neighbor graph or joint UMAP was generated using the

203 “FindMultiModalNeighbors” function in Signac and reduction.name = “wnn.umap”. A Seurat
204 object from the clustered PGCs was used as input for pseudotime analysis with Monocle3 to
205 determine the differentiation trajectory of PGCs using both female and male E11.5 PGCs as
206 time ‘0’ in pseudotime.

207 **Identification of differentially expressed genes and differentially accessible peaks**

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

214 **Correlation of peak accessibility with gene expression**

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

219 **Motif enrichment analysis of transcription factor binding sites**

220 DNA sequence motif analysis of peaks linked to DE genes was performed using the
221 “FindMotifs” function in *Signac*. Motifs were obtained from the JASPAR 2020 database
222 (collection = ‘Core’ and taxonomic group = ‘vertebrates’). Background peaks were selected
223 to match the GC content in the peak set by using the “AccessiblePeaks,” “GetAssayData”,
224 and “MatchRegionStats” functions in *Signac*. *MotifScan*[38] (v. 1.3.0) was used to determine
225 the genomic position of linked peaks to DE genes to identify predicted target genes of

226 transcription factors. Per-cell motif activity was computed using the "RunChromVAR"
227 function in *Signac*.

228 **Module score calculation for gene expression**

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

231 **Peak feature distribution**

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

236 **GO annotation and enrichment**

237 The GO enrichment analyses were carried out using the R package *clusterProfiler*[41] (v.
238 4.0.5) with the "enrichGO" function. The Benjamini-Hochberg method (a = 0.05) was used
239 to control for False Discovery Rate. GO annotation enrichment for biological processes
240 were plotted using the "dotplot" or "cnetplot" function of the *Enrichplot*[42] (v1.12.2) R
241 package.

242 **CistromeDB regulatory potential score**

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

247 **Enrichment of signaling pathways between gonadal support cells and primordial**
248 **germ cells**

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

255 **Immunofluorescence**

256 Gonads were collected from E13.5 XX and XY embryos and fixed in 4% PFA for 1h at room
257 temperature. Gonads were then washed in PBS and stored at 4°C until embedding in
258 paraffin and sectioned at 5 μm thickness. Following deparaffinization, sections were
259 rehydrated in decreasing alcohol concentration gradients. Antigen retrieval was performed
260 with 0.1 mM citric acid (Vector Labs) for 20 min in the microwave at 10% power and then
261 cooled to room temperature. Samples were blocked in PBS-TRITON X-100, 0.1% solution
262 with 5% normal donkey serum for 1h. Samples were incubated with primary antibodies
263 diluted in blocking buffer (PBS-TRITON X-100, 0.1% solution with 5% normal donkey
264 serum) at 4°C overnight. Samples were then washed with PBS-TRITON X-100, 0.1%
265 solution and incubated with secondary antibodies diluted in blocking buffer at room
266 temperature for 1h. The slides were washed with PBS-TRITON X-100, 0.1% solution 3x and
267 incubated with the Vector TrueView Autofluorescence Quenching Kit (cat. # SP-8400) for 2
268 min. Following incubation with the quenching kit, the slides were washed with PBS-
269 TRITON X-100, 0.1% solution 2x, PBS 1x, and incubated with DAPI (Invitrogen cat. #
270 D1306). The slides were then mounted with slide mounting media (Vectashield Vibrance
271 Antifade Mounting Medium TrueView Kit SP-8400). Primary antibodies used included:

272 Anti-TRA98 (1:500; rat, BioAcademia cat. # 73-003), Anti-NR2F2 (1:300, mouse, R&D
273 Systems cat. # PP-H7147-00), Anti-PORCN (1:200, rabbit, ThermoFisher cat. # PA5-43423),
274 and Anti-AP-2γ (TFAP2c) (1:200, mouse, Santa Cruz cat. # sc-12762). Secondary antibodies
275 were used at 1:200 dilution (Invitrogen/Life Technology): Donkey anti-rat Alexa 568
276 (A78946), Donkey anti-mouse Alexa 647 (A31571), and Donkey anti-rabbit Alexa 488
277 (A21206). Imaging for sections was performed on a Zeiss LSM 900 confocal microscope
278 using Zen software. Brightness and contrast of images were adjusted using ImageJ
279 (National Institutes of Health, USA).

280 **Statistical analyses**

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

284 **Results**

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

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

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

329 2 and 3, and displayed enriched expression of ribosome biogenesis genes and genes with
330 reported roles in reproduction, such as the mitochondrial leucyl-tRNA synthetase *Lars2*[50]
331 and the DNA repair protein *Mgmt*[51] (**Fig. 1d; Supplementary Figure 2**). Together, these
332 data indicate that the transcriptomic profiles of PGCs are largely homogenous within each
333 sex at E11.5. Subsequently, XX PGCs maintain their transcriptional heterogeneity at E12.5-
334 E13.5, whereas XY PGCs transcriptionally converge on a single identity beginning at E12.5.
335 The snRNA-seq clustering revealed transcriptomic characteristics of PGCs. Next, we
336 performed unbiased clustering of PGCs according to their aggregate profiles of snATAC-seq
337 peaks (210,666 total peaks) to detect the patterns of chromatin accessibility associated
338 with PGC development (**Fig. 1e and 1f**). The snATAC-seq-based clustering of PGCs showed
339 that the chromatin status of E11.5 XX and XY PGCs was highly similar (**Fig. 1e**). By contrast,
340 E12.5-E13.5 XX and XY PGCs formed distinct groupings at opposite poles of the UMAP (**Fig.
341 1e**). Unbiased graph-based clustering of PGCs uncovered 9 clusters of PGCs with unique
342 stage- and sex-specific snATAC-seq peak profiles (**Fig. 1f**). First, E11.5-E12.5 XX and XY
343 PGCs together predominantly consisted of accessibility-based clusters 0, 1, and 2,
344 indicating that these germ cells share a highly similar chromatin profile prior to
345 differentiation (**Fig. 1f and g**). Second, E13.5 XX PGCs were separated into clusters 3, 4, and
346 5 (**Fig. 1f and g**), which may be due to differences in meiotic entry or progression within
347 these populations of germ cells. Third, E12.5-E13.5 XY PGCs comprised snATAC-seq
348 clusters 6 and 7, suggesting that XY PGCs may adopt a sex-specific chromatin status
349 beginning at E12.5 (**Fig. 1f and g**). Finally, cluster 8, which contained PGCs from both sexes
350 at each stage, was separate from all other accessibility-based clusters (**Fig. 1f and g**). Taken
351 together, our findings indicate that the chromatin accessibility profiles of E11.5 PGCs and

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

375 developmental trajectories of PGCs that are clustered with multi-modal data, but also
376 empirically trace their sex-specific fates based on changes in gene regulation.
377 Clustering of PGCs using combined snRNA- and snATAC-seq data identified eleven
378 distinct groups of PGCs and yielded enhanced cluster resolution of PGC subpopulations
379 (**Fig. 1i**). This analysis resulted in clusters of PGCs that were predominantly sex-specific
380 and enabled the detection of subpopulations of each PGC type (**Fig. 1i**). To better
381 understand the population structure of individual clusters, we plotted the percentage of
382 each PGC type (*x*-axis) for each cluster identified in **Fig. 1i** (*y*-axis) (**Fig. 1j**). First, E11.5 XX
383 PGCs clustered into one main population, cluster 1, and later diverged into six
384 subpopulations, clusters 0 through 5 at E12.5 (**Fig. 1i-j**). Clusters 0 - 2 were more similar to
385 E11.5 PGCs, whereas clusters 3 – 5 were composed of E12.5 XX PGCs that were
386 developmentally closer to E13.5 XX PGCs (**Fig. 1i-j**). Finally, we found that E13.5 XX PGCs
387 converged onto either one of two developmental fates in clusters 6 and 7, with cluster 7
388 exhibiting a large spread along the UMAP2 axis (**Fig. 1i-j**). These findings indicate that
389 E12.5-E13.5 XX PGCs may require extensive chromatin remodeling and transcriptional
390 reprogramming to adopt the oogenic fate. By comparison, XY PGCs grouped into clusters 0
391 and 1 at E11.5 and appeared to converge on a single developmental fate at later stages,
392 with E12.5 XY PGCs comprising clusters 8 and 9 and E13.5 XY PGCs belonging to cluster 10
393 (**Fig. 1i-j**). Given that neither expression- nor accessibility-based clustering alone resolved
394 the granularity of PGC populations to the level described here, these data underscore the
395 empirical power of simultaneously measuring multiple modalities from the same cell to
396 define the developmental fate of PGCs.
397

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

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

413 To identify potential mechanisms regulating the expression of XY PGC functional
414 genes, we searched for candidate transcription factors (TFs) with XY-enriched gene
415 expression. We first linked snATAC-seq peaks to XY PGC functional genes (listed
416 previously), hereafter termed peak-to-gene linkages, in XY PGCs. Peak-to-gene linkages
417 were determined using Signac functionalities and were derived from the correlation
418 between peak accessibility and the intensity of gene expression. Next, we performed a TF
419 motif enrichment analysis on significant peak-to-gene linkages to identify potential TFs.
420 Finally, we compared the expression levels of potential TFs between XX and XY PGCs to

421 identify TFs with XY-enriched gene expression. This strategy uncovered the Krüppel-like
422 zinc finger family member ZKSCAN5 (Zinc finger with KRAB and SCAN domains 5) as a
423 potential regulator of XY PGC functional genes (**Fig. 2a**). *Zkscan5* showed XY biased
424 expression (e.g., there is a 1.5-fold increase in the expression of *Zkscan5* between XY and XX
425 PGCs at E13.5), and the level of *Zkscan5* expression increased from E11.5 to E13.5 in XY
426 PGCs (**Fig. 2a**). In addition, the ZKSCAN5 motif was not enriched in XX PGCs. ZKSCAN5 is
427 not only highly expressed in the adult testis[56] but is also proposed to regulate cellular
428 proliferation and growth[57]. Thus, although little is known about the function of ZKSCAN5,
429 we hypothesize ZKSCAN5 may possess a regulatory role during male germ cell
430 development. These data highlight the strength of single-nucleus multiomics in discovering
431 candidate factors that may regulate the expression of a cohort of developmental genes and,
432 consequently, direct cell fate decisions.

433 Among the functionally important PGC genes, *Bnc2* expression was most strongly
434 enriched in XY PGCs. We therefore investigated the regulation of *Bnc2* at the chromatin
435 level as an example to demonstrate the utility of combined multiome datasets. The
436 chromatin accessibility of the *Bnc2* promoter was highest in XY PGCs and increased
437 temporally from E11.5 to E13.5 (**Fig. 2b**). Furthermore, the level of chromatin accessibility
438 at the *Bnc2* locus was positively correlated with *Bnc2* expression in XY PGCs (**Fig. 2b-c**). We
439 then applied the Cistrome Data Browser (Cistrome DB)[44] to identify putative regulators of
440 *Bnc2* expression based on transcription factor binding motifs detected within 10 kb of the
441 *Bnc2* promoter (**Fig. 2d**). The top candidate regulators of the *Bnc2* locus were determined
442 according to a regulatory potential score derived from a combination of TF binding motif
443 information and previously published genomics data (*i.e.* ChIP-seq, DNase-seq, and ATAC-

444 seq) (Fig. 2d). Of the top candidate regulators from the Cistrome DB analysis, expression of
445 *Ctcf*, *Nrf1*, *Polr2a*, and *Brd4* was enriched in XY PGCs at E13.5 (Fig. 2e). In particular, BRD4
446 is known to activate transcription elongation by recruiting kinases to the largest RNA
447 polymerase II (Pol II) subunit, POLR2A, thereby releasing Pol II from its promoter-proximal
448 bound state[58]. Taken together, these findings provide insight into potential XY-specific
449 regulatory mechanisms at the *Bnc2* locus in PGCs.
450
451 **Characterization of genes known for their roles in XX PGC development**
452 We next performed the same analyses of XX PGC functional genes, *i.e.* those that
453 have been shown to be functionally important for XX PGC development (Fig. 3a). Overall,
454 as expected, we found that E13.5 XX PGCs expressed genes involved in XX germ cell
455 development (*i.e., Rec8, Sycp2, Sycp3, Ctnnb1, Meioc, Rnf2, Stra8, Ythdc2,* and *Zglp1*)[8]
456 substantially higher than E13.5 XY PGCs (Fig. 3a). Moreover, *Rnf2* and *Zglp1* were strongly
457 expressed at E12.5 in XX PGCs, and their expression occurred immediately prior to the
458 activation of all other genes in the female cohort (Fig. 3a). Although *Zglp1* was expressed in
459 fewer XX PGCs than *Rnf2*, *Zglp1* expression increased from E12.5 to E13.5 whereas *Rnf2*
460 expression decreases (Fig. 3a). These data support previous studies that *Rnf2* and *Zglp1*
461 control the timing of sexual differentiation of XX PGCs[10,59]. By comparison, the expression
462 of *Ctnnb1* diminished in both XX and XY PGCs at E12.5, as expected given that negative
463 regulation of WNT/β-catenin is required for XX PGCs to enter meiosis[60] (Fig. 3a). In sum,
464 these results provide insight into the timing, expression levels, and pervasiveness of female
465 gene activation in XX PGC populations.

466 We next identified potential regulators of XX PGC functional genes using the same
467 approach for XY PGCs (**Fig. 2**). Our analysis identified significant enrichment of the binding
468 motif for DMRT1 in the regulatory elements linked to XX PGC functional genes (**Fig. 3a**).
469 DMRT1 has been shown to control *Stra8* expression in a sex-specific manner, such that
470 *Stra8* is activated in the fetal ovary and repressed in the fetal testis[61]. Our snRNA-seq data
471 also demonstrated *Dmrt1* expression in both XX and XY PGCs (**Fig. 3a**), which is consistent
472 with previous findings[61].
473 Since *Stra8* plays an essential role in controlling meiotic entry[8], we next sought to
474 further understand the gene regulatory mechanisms underlying *Stra8* activation. First, we
475 not only found that accessibility of the *Stra8* promoter was highest in XX PGCs, but also that
476 *Stra8* chromatin accessibility was positively correlated with *Stra8* gene expression (**Fig.**
477 **3b-c**). Second, the population of E13.5 XX PGCs displaying the strongest *Stra8* expression
478 levels corresponded to the same population of XX PGCs with the highest module score of
479 early meiotic prophase I genes (**Fig. 3c; Supplementary Fig. 3a-b**). Next, to determine
480 putative regulators of *Stra8* expression, we identified enriched TF binding motifs within 10
481 kb of the *Stra8* promoter using Cistrome DB[44] (**Fig. 3d**). Of the factors with the highest
482 Cistrome DB Regulatory Potential Score, we observed XX-enriched expression of *Rxra,*
483 *H2afy, Tcf12, Rad21, Smc3,* and *Smad3* (**Fig. 3e**). Several of these factors, such as RAD21 and
484 SMC3, are members of the cohesin complex, which serves both a structural and gene
485 regulatory role during meiotic prophase I[62,63]. Moreover, RXRA has been shown to
486 positively influence *Stra8* expression levels during oocyte development in the mouse[64].
487 Taken together, these data provide a proof-of-principle example for the application of
488 single-nucleus multiomics in identifying putative regulators of germ cell fate.

489

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

491 Since recent reports have demonstrated that meiosis occurs normally in the fetal
492 ovary of mice lacking key members of the retinoic acid (RA) signaling pathway[65], we next
493 asked whether we could detect any evidence for gene regulation by RA receptors in PGCs.
494 Previous studies have demonstrated that RA from the mesonephros induces meiosis in the
495 fetal ovary whereas in the fetal testis, RA is degraded by the enzyme CYP26B1, therefore
496 preventing PGC entry into meiosis[66,67]. However, this model was challenged by the findings
497 that meiosis, and consequently oogenesis, still occurs in XX gonads of mice lacking the RA-
498 synthesizing enzyme, ALDH1, or all RA receptors[65,68]. Nevertheless, *Stra8* expression was
499 significantly reduced in E13.5 fetal ovaries without *Aldh1a1-3* or all RA receptors[65,68], thus
500 indicating that *Stra8* expression is likely controlled by the RA signaling pathway, in
501 addition to others. We therefore investigated the expression and *in silico* chromatin binding
502 patterns of members of the RA signaling pathway (*i.e. Rara, Rarb, Rarg, Rxra, Rxrb, Rxrg,*
503 *Crabp1*, and *Crabp2*)[69] (**Supplementary Figure 4a-b**). We found that expression of the
504 retinoic acid receptor beta, *Rarb*, and cellular retinoic acid binding protein 1, *Crabp1*, were
505 enriched in XX PGCs beginning at E12.5 and continuing to E13.5 (**Supplementary Figure**
506 **4a-b**). Similarly, E13.5 XX PGCs showed enriched expression of the retinoid X receptor
507 alpha, *Rxra*, and cellular retinoic acid binding protein 2, *Crabp2* (**Supplementary Figure**
508 **4a-b**). By contrast, the expression of *Rara, Rarg,* and *Rxrb* was enriched in XY PGCs, and
509 *Rxrg* was lowly expressed in both XX and XY PGCs (**Supplementary Figure 4a-b**). We also
510 observed significant enrichment of the 'motif activity score', an *in silico* measurement of the
511 contribution of a specific motif to gene regulation[70], of all RA receptors in E13.5 XX PGCs

512 (**Supplementary Figure 4c-d**). These data indicate that the RA receptors display strong
513 evidence of *in silico* chromatin binding in E13.5 XX PGCs but not E13.5 XY PGCs.
514 Furthermore, the binding motifs for the RA receptors were predominantly located in E13.5
515 XX PGC chromatin peaks that were distal intergenic, intronic, or within promoters
516 (**Supplementary Figure 4e;** motifs for RARB and RXRA provided as examples). These data
517 indicate that the binding motifs for the RA receptors are predominantly enriched in both
518 distal and proximal regulatory elements in E13.5 XX PGCs. The RA receptor motifs were
519 also statistically linked to genes involved in 'response to retinoic acid', 'negative regulation
520 of cell cycle', and 'meiotic cell cycle' (**Supplementary Figure 4f;** motifs for RARB and RXRA
521 provided as examples). Finally, when we searched for the presence of RA receptor motifs in
522 peaks linked to genes related to meiosis and female sex determination, we found that *Stra8,*
523 *Rec8, Rnf2, Sycp1, Sycp2, Ccnb3,* and *Zglp1* contain the RA receptor motifs in their
524 regulatory sequences (**Supplementary Figure 4g**). Together, these *in silico* experiments
525 indicate that the RA signaling pathway directly regulates the expression of several meiosis-
526 related genes.
527

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

530 We further investigated our multiome dataset to discover genes and *cis*-regulatory
531 elements enriched in PGCs in a sex-specific manner. First, we analyzed which annotated
532 protein-coding genes showed transcriptional changes between XX and XY PGCs at each
533 embryonic stage. Differential expression analysis ($\log_2$fold-change > 0.25) of snRNA-seq
534 data identified 87, 301, and 585 XX PGC-enriched genes and 75, 344, and 544 XY PGC

535 enriched genes at E11.5, E12.5, and E13.5, respectively (False Discovery Rate-adjusted *p*-
536 value < 0.05; **Fig. 4a**). The number of differentially expressed genes (DEGs) increased by
537 ~3.5-fold during the transition from E11.5 to E12.5 PGCs and by 1.75-fold during the
538 transition from E12.5 to E13.5 PGCs (**Fig. 4a**). These findings demonstrate that the number
539 of DEGs steadily increased over developmental time in both XX and XY PGCs.

540 An in-depth analysis of the DEGs between XX and XY germ cells identified sex-
541 enriched genes for PGCs at each embryonic stage (**Supplementary Figure 5a**). For
542 example, we found elevated expression of the WNT protein regulator *Porcn*[71], the beta-
543 catenin binding protein *Grip1*[72], and the growth regulator *Phlda2*[73] in XX PGCs
544 (**Supplementary Figure 5a**). These observations are consistent with the fact that WNT/β-
545 catenin signaling is involved in specifying the female fate of PGCs[8]. We also identified XY
546 enriched DEGs, which included the histone demethylase *Kdm5d*[74], the regulator of
547 pluripotency *Dppa5*[75], and the forkhead transcription factor *Foxp1*[76] (**Supplementary**
548 **Figure 5a**). Gene ontology (GO) analyses of DEGs revealed several biological processes that
549 show sex-enriched expression patterns (**Supplementary Figure 5b**). For example, the
550 DEGs identified in XX PGCs were assigned to the 'meiotic chromosome organization', 'sister
551 chromatid cohesion', and 'WNT signaling pathway' gene ontologies (**Supplementary**
552 **Figure 5b**). The DEGs identified in XY PGCs were related to 'mitotic cell cycle', 'TGFβ
553 signaling pathway', and 'regulation of translation' gene ontologies (**Supplementary Figure**
554 **5b**). In sum, the DEGs between XX and XY PGCs were characterized by their sex-specific
555 roles in meiotic/mitotic cell cycle regulation, post-transcriptional processing of mRNA, and
556 cell-cell signaling properties (**Supplementary Figure 5a-b**). These analyses provide a

557 clear description of the transcriptional differences between XX and XY PGCs underlying the
558 transition from bipotential germ cells to sex-specific oogonium or gonocytes.

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

570 We next sought to associate sex-specific changes in chromatin accessibility with
571 changes in gene expression for PGCs. To determine peak-gene associations, we correlated
572 snATAC-seq peak accessibility with the expression of nearby genes, with the goal to
573 identify sex- and stage-specific chromatin peaks positively correlated with mRNA
574 expression (p-value < 0.05; z-score > 0). This analysis detected putative gene regulatory
575 interactions based on correlations between chromatin peak accessibility and gene
576 expression levels for all peaks and genes within 500 kb of each other[77]. Given the notable
577 increase in DA peaks between XX and XY PGCs at E13.5 (**Fig. 4b**), we first identified the
578 predicted target genes of all sex-specific DA peaks at this stage (**Supplementary Figure**
579 **6a-b**). For E13.5 XX PGCs, we observed positive DA peak-gene correlations for genes

580 enriched for biological functions such as 'WNT signaling pathway', 'chromosome
581 segregation', 'chromatin organization', 'meiotic cell cycle', and 'RNA stabilization'
582 (**Supplementary Figure 6a**). A similar analysis of E13.5 XY PGCs revealed positive DA
583 peak-gene correlations for genes involved in 'ncRNA metabolic process', 'mitotic cell cycle',
584 'DNA repair', 'regulation of translation', and 'RNA localization' (**Supplementary Figure**
585 **6b**). These findings demonstrate that we can use chromatin accessibility information to
586 identify the regulatory elements linked to sex-specific genes in PGCs.
587 We next investigated the associations of differentially accessible (DA) chromatin
588 peaks with differentially expressed (DE) gene expression for each PGC type. At E11.5, we
589 found that DA peak-to-gene linkages were associated with ~20% of DE genes in XX PGCs
590 and ~1% of DE genes in XY PGCs (**Supplementary Figure 6c**). The overlap of DA peak-to-
591 gene linkages with DE genes increased steadily at E12.5 and E13.5, with ~35% and ~60%
592 of DA peaks in E13.5 XX and XY PGCs, respectively, positively correlated with DE gene
593 expression (**Supplementary Figure 6c**). For example, at E13.5, we identified the DA peaks
594 positively correlated with the DE genes *Porcn* (involved in the processing of WNT
595 proteins)[71] and *Rec8* (involved in meiotic recombination)[8] in XX PGCs and *Rimbp1*
596 (regulates cytosolic calcium ion concentration)[78] in XY PGCs (**Fig. 4c-d; Supplementary**
597 **Figure 6d**). Immunofluorescence staining also confirmed strong expression of *Porcn* in
598 E13.5 fetal XX germ cells and diminishing expression of *Porcn* in E13.5 fetal XY germ cells
599 (**Fig. 4e**), demonstrating that single-nucleus multiomics can identify DE genes and the DA
600 chromatin peaks that are associated with their expression.
601

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

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

618 In XX PGCs, we uncovered several candidate TF regulators of E11.5-E13.5 XX-
619 enriched DEGs (**Supplementary Figure 7a-c**). We plotted the mRNA expression levels of
620 TFs and the *in silico* chromatin binding score, termed 'motif activity', of all significantly
621 enriched binding motifs (**Supplementary Figure 7a-c**). For example, in E11.5 XX PGCs,
622 RREB1 is a differentially expressed TF showing high gene expression, but negative levels of
623 *in silico* chromatin binding (**Supplementary Figure 7a**). This finding indicates that
624 although RREB1 is a highly expressed TF in E11.5 XX PGCs, RREB1 does not likely promote

625 increased chromatin accessibility either because it does not bind to chromatin or a critical
626 co-factor is unavailable (**Supplementary Figure 7a**). Following this analytical framework,
627 we focused our analysis on TFs that showed high levels of gene expression, were
628 differentially expressed in XX PGCs, and displayed a positive *in silico* chromatin binding
629 score (**Supplementary Figure 7a-c**). The TFs meeting these criteria were KLF12, NR6A1,
630 and ZFX in E11.5 XX PGCs; TFAP2c in E12.5 XX PGCs; and BACH2, CREM, GATA2, GLI3,
631 MGA, NFKB1, PLAGL1, TBX4, TCFL5, TFAP2c, ZBTB7c, and ZFX in E13.5 XX PGCs
632 (**Supplementary Figure 7a-c**). Of these sixteen TF candidates with PGC-enriched
633 transcript expression, only seven (*Gata2, Mga, Nr6a1, Tbx4, Tcfl5, Tfap2c,* and *Zfx*) showed
634 enriched expression in germ cells when compared to the somatic compartment of the
635 gonad, and thus met our requirements for further *in silico* investigation (**Supplementary**
636 **Figure 8a**). These candidate TFs have reported roles in cell fate determination, growth and
637 fertility regulation, and meiotic progression[79,80], all of which are in line with the cellular
638 events that occur as PGCs commit to the female pathway. Nonetheless, the chromatin
639 binding and gene expression of GATA2, MGA, TBX4, TCFL5, and TFAP2c showed the
640 clearest patterns of sexual dimorphism in E13.5 PGCs and represented the most compelling
641 TF candidates (**Fig. 5b-c; Supplementary Figure 8b-f**).
642 The binding motif for TFAP2c (also known as *AP2γ*) showed prominent TF-
643 associated accessibility, especially when compared to all other significantly enriched TF
644 binding motifs (**Supplementary Figure 7b-c**). Immunofluorescence staining of TFAP2c
645 also showed enrichment of TFAP2c in E13.5 XX PGCs when compared to the somatic
646 compartment and E13.5 XY PGCs (**Supplementary Figure 9a**). Compared to XY PGCs, XX
647 PGCs had enriched *Tfap2c* expression (p-value < 0.05; log2FC > 0.25) and enriched

648 accessibility of the TFAP2c motif at E12.5-E13.5 (**Fig. 5b**). The TFAP2c motif was close to
649 genes related to glycogen metabolism, carbohydrate metabolism, and chromatin
650 modification (**Fig. 5d**). The predicted target genes of TFAP2c accounted for 75% (372/494)
651 of all DEGs with peak-to-gene linkages in XX PGCs (**Fig. 5e**), suggesting that TFAP2c is
652 likely a high-level regulator of female PGC sex determination.

653 In addition to TFAP2c, we also investigated the regulatory potential of TCFL5
654 (Transcription Factor Like 5) in greater detail. TCFL5 was the most differentially expressed
655 in E13.5 XX PGCs (p-value < 0.05; log2FC > 0.25) (**Fig. 5c**). Accessibility of the TCFL5 motif
656 was also enriched in E13.5 XX PGCs, indicating that chromatin binding of TCFL5 is likely
657 strongest at this stage (**Fig. 5c**). The predicted target genes of TCFL5 totaled 74%
658 (367/494) of all DEGs with peak-to-gene linkages in XX PGCs (**Fig. 5e**). A large majority of
659 the predicted target genes of TCFL5 were also predicted to be the target genes of the
660 enriched TFs presented in Fig. 5e, e.g., the predicted target genes of these TFs overlapped
661 with 4%-100% of the predicted target genes of TCFL5. The TCFL5-associated genes were
662 related to chromosome segregation, meiotic cell cycle, sister chromatid cohesion, histone
663 modification, regulation of TF activity, and cell-cell signaling by WNT. Our findings are
664 supported by a recent report on the role of TCFL5 in XX PGCs, which demonstrated the
665 requirement of TCFL5 in meiotic progression and female fertility[81].

666 Given that the motifs for TFAP2c and TCFL5 were not detected at every XX-enriched
667 DEG, we hypothesize that a network of TFs, such as those identified in **Supplementary**
668 **Figure 8a**, could work in cooperativity to regulate XX PGC gene expression. For example,
669 we found that the predicted target genes of ZFX, MGA, NR6A1, TBX4, and GATA2 accounted
670 for 73% (359/454), 31% (152/494), 20% (100/494), and 6% (32/494), respectively, of all

671 DEGs with peak-to-gene linkage in XX PGCs (Fig. 5e). We also found that TFAP2c, TCFL5,
672 ZFX, MGA, and NR6A1 may regulate the expression of each other (Fig. 5f). For instance, the
673 motif for ZFX was located in the enhancers or promoters for TFAP2c, TCFL5, MGA, and
674 NR6A1 (Fig. 5f). Likewise, NR6A1 is predicted to control the expression of TFAP2C, TCFL5,
675 and TBX4 (Fig. 5f). Furthermore, the motifs for the candidate TFs were statistically linked
676 to genes involved in 'meiotic cell cycle' and the 'WNT signaling pathway' (Fig. 5g-h).
677 Meiosis related genes such as *Tex15, Sycp1, Dmc1,* and *Hormad1* contained the motifs for at
678 least three of the predicted TFs in their regulatory sequences (Fig. 5g). A similar pattern
679 was also found for genes related to WNT signaling, e.g. *Pkd1, Notum, Porcn, Fzd3,* and *Eda*
680 were predicted to be regulated by the binding of at least two of the putative TFs in XX PGCs
681 (Fig. 5h). These data indicate that there may be unknown and undefined TF regulatory
682 networks underlying XX PGC identity and function.
683

684 **Identification of candidate TF regulators of XY PGC differentiation**

685 To identify candidate TF regulators involved in XY PGC sex determination, we
686 performed the same analysis as described for XX PGCs (Fig. 5a). Our analysis focused
687 strictly on differentially expressed TFs that show evidence of chromatin binding based on
688 the empirical *in silico* chromatin binding score (Supplementary Fig. 10a-c). In contrast to
689 XX PGCs (Supplementary Fig. 7), we did not identify any significantly enriched TF binding
690 motifs (p-value > 0.05) meeting our criteria in E11.5 XY PGCs (Supplementary Fig. 10a).
691 Only three XY-enriched TFs, FOXP1, POU6F2, and TCF4, were differentially expressed
692 between XY and XX PGCs at E12.5 (Supplementary Fig. 10b), and only TCF4 showed a
693 positive *in silico* chromatin binding score in E12.5 XY PGCs (Supplementary Fig. 10b). For

694 E13.5 XY PGCs, we detected five significantly enriched motifs that bound differentially
695 expressed TFs (p-value < 0.05): FOXK2, FOXO1, FOXP1, POU6F1, and TCF4
696 (**Supplementary Fig. 8c**). Of these six XY-enriched TF candidates from E11.5 to E13.5,
697 FOXO1, FOXK2, and POU6F1/2 emerged as the top candidates based on germline enriched
698 expression when compared to somatic cells, differential expression in XY PGCs, and
699 positive *in silico* chromatin binding scores (**Supplementary Fig. 11a-b; Fig. 5i-j**).
700 Together, these data indicate that XY PGCs rely less on differentially expressed TF-
701 mediated gene regulation when compared to XX PGCs.

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

717 increase in the number of DA peaks in E13.5 XY PGCs (**Fig. 4b**), we hypothesize that the
718 pioneering properties of the FOX family promotes chromatin remodeling as XY PGCs
719 transition to gonocytes.

720 We also detected significant enrichment of the POU domain, class 6 (POU6) TF
721 family in XY PGCs (**Supplementary Fig. 10b-c; Supplementary Fig. 11a & c; Fig. 5j**). The
722 POU family plays important roles in regulating cellular differentiation and the timing of
723 cellular events[83]. We found XY-enriched expression of POU6F1 and POU6F2 in PGCs,
724 however, only POU6F2 displayed germline enrichment when compared to the somatic cell
725 populations of the gonad (**Supplementary Fig. 11a & c; Fig. 5j**). Furthermore, we
726 observed the strongest POU6F1 and POU6F2 *in silico* chromatin binding scores in XY PGCs
727 at E13.5 (**Supplementary Fig. 11c; Fig. 5j**). The POU6F1/2 motif was associated with
728 19% (40/208) of all DEGs with peak-to-gene linkages in XY PGCs (**Fig. 5l**). POU6F1/2-
729 associated genes were related to voltage-gated chloride channel activity (*Ano6*)[84],
730 regulation of cell proliferation (*Appl1*)[85], regulation of actin filaments (*Fmnl2*)[86], Notch
731 signaling (*Maml3*)[87], lipid metabolism (*Osbpl9*)[88], and ubiquitin ligase activity (*Usp7*)[89] (**Fig.
732 5m**). Thus, the POU6 TF family is a top candidate to coordinate gene expression activity in
733 XY PGCs.

734 We next compared the number and identity of FOXO1-, FOXK2-, and POU6F1/2-
735 associated genes in XY PGCs. First, we did not detect the FOXO1, FOXK2, or POU6F1/2
736 motifs in the promoters/enhancers for each of these TFs. These findings indicate that these
737 TFs do not regulate the expression of each other, and other mechanisms or regulatory
738 pathways are required to activate the expression of candidate TFs in XY PGCs.
739 Furthermore, the predicted target genes of the FOX and POU6 TF families shared little

740 overlap with each other (11/87) (Fig. 5m). The genes that were shared between FOXO1,
741 FOXK2, and POU6F1/2 had reported functions in maintaining survival after DNA damage
742 (*Alkbh8*)[90], TGFβ signaling (*Bmp7*)[91], and initiation of mitosis (*Nek7*)[92] (Fig. 5m). By
743 comparison, the predicted target genes of FOXO1 and FOXK2 showed considerable overlap
744 (42/47) and included genes related to RNA splicing (*Aqr*)[93], cell cycle progression (*Cdk6*)[94],
745 cellular metabolism (*Eno1* and *Odc1*)[95,96], cell death (*Pml*)[97], and negative regulation of the
746 IGF1 signaling pathway (*Socs2*)[98] (Fig. 5m). The motifs for FOXO1, FOXK2, and POU6F1/2
747 were statistically associated with genes involved in 'mitotic cell cycle' and 'signal
748 transduction' (Fig. 5n-o). Mitosis-related genes such as *Cradd, Cdk6, Exoc6b,* and *Pml* were
749 potentially regulated by FOXO1 and FOXK2 (Fig. 5n). In addition, we noted that the
750 POU6F1/2 motif was associated with *Cradd* and *Psmg2* (Fig. 5n). We observed a similar
751 pattern for genes related to signal transduction (Fig. 5o). For example, *Bmp7, Map3k5,*
752 *Akt3,* and *Sulf1* were associated with FOXO1 and FOXK2, whereas *Bmp7, Map3k5,* and
753 *Sfprp1* were associated with POU6F1/2 motif activity (Fig. 5o). Taken together, these data
754 suggest that in XY PGCs, although FOXO1, FOXK2, and POU6F1/2 regulate genes with
755 similar properties, the FOX and POU6 TF families are unlikely to collaboratively regulate
756 gene expression.
757

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

760 Sexual dimorphism of germ cell differentiation is thought to be controlled by a
761 meiosis-inducing substance produced by the somatic cells of the fetal ovary and/or by a
762 meiosis-preventing substance produced by the fetal testis[99]. In experimentally generated

763 XX/XY chimeras, both XX and XY PGCs can initiate meiosis when residing in the fetal ovary
764 or enter mitotic arrest when colonizing the fetal testis[100]. In the fetal ovary, PGC meiosis is
765 induced by retinoic acid (RA) produced by ovarian and mesonephric somatic cells[8].
766 Consistent with this model, RA signaling induces expression of *Stra8* and *Meiosin*, both of
767 which are critical for promoting entry into meiosis[8]. Signaling by WNT and BMP also
768 influences the timing of meiotic onset in XX PGCs[8], and inhibin/activin signaling regulates
769 XX germ cell growth and proliferation[101]. In the fetal testis, RA degradation by CYP26B1
770 and FGF9 signaling from XY supporting cells result in reduced *Stra8* expression,
771 consequently blocking meiotic entry in XY PGCs[8]. Thus, PGC sex determination is
772 dependent on external signals from the gonadal environment rather than chromosomal
773 sex.

774 To uncover potential instructive signals from the supporting somatic cells (Sertoli
775 cells in the fetal testis or granulosa cells in the fetal ovary) to PGCs, we searched for
776 supporting cell-derived ligands and the corresponding receptors in PGCs (**Fig. 6a**), using
777 CellphoneDBv5.0 for cell communication analysis[45]. We first identified the supporting cell
778 clusters from the snRNA-seq dataset of whole fetal gonads using the established female and
779 male supporting cell markers *Runx1* and *Sox9*, respectively[30,102]. We next quantified the
780 number of significant interactions between supporting cells and PGCs at each embryonic
781 stage (p-value < 0.05; **Fig. 6b**). Such analyses revealed that the number of interactions
782 gradually increased from E11.5 to E13.5 for both XX and XY gonads (**Fig. 6b**). We also
783 observed that the number of ligand-receptor pairs was 15-38% higher in XX gonads,
784 compared to XY gonads at the same embryonic stage (**Fig. 6b**). To validate our
785 computational approach, we next searched for known pathways associated with PGC

786 differentiation, such as signaling by WNT, inhibin/activin, and BMP (**Fig. 6c**). The number
787 of significant interactions per pathway increased in both XX and XY gonads from E11.5 to
788 E13.5 (**Fig. 6c; Supplementary Fig. 12a-b**). We also observed a moderate increase in the
789 number of interactions for the inhibin/activin and BMP pathways in XX gonads when
790 compared to XY gonads at the same embryonic stage (**Fig. 6c**). Finally, we found that the
791 number of interactions related to WNT signaling were nearly 2-3 times greater in XX than
792 XY gonads (**Fig. 6c; Supplementary Fig. 12a-b**). Together, these data demonstrate that
793 sexually dimorphic patterns of cell signaling events were present between supporting cells
794 and PGCs.

795 We next searched for new sex-specific enrichment of ligand-receptor pairs between
796 supporting cells and PGCs (**Fig. 6d-i**). For each embryonic stage, we first identified which
797 signaling pathways were detected in XX alone, XY alone, or both XX and XY gonads that met
798 our statistical threshold (p-value < 0.05; **Fig. 6d, f, & h**). At E11.5, XX and XY gonads shared
799 the greatest overlap of supporting cell-derived ligands and PGC-expressing receptors (**Fig.
800 6d**). The shared pathways between XX and XY gonads were largely related to adhesion by
801 collagen/integrin, fibronectin, and cadherin. E11.5 XX gonads had enrichment of 19 ligand-
802 receptor pairs related to, for example, BMP, ephrin, integrin, FGF, and WNT signaling (**Fig.
803 6d**). By contrast, E11.5 XY gonads had significant enrichment of 10 ligand-receptor pairs
804 predominantly related to cell adhesion pathways like ephrin, nectin, and integrin (**Fig. 6d**).
805 To highlight the general trends of expression between sexes, we compared the mean
806 expression of all representative ligand-receptor pairs from each group in Fig. 6d, f, & h (**Fig.
807 6e, g, & i**). In E11.5 gonads, we observed dimorphic expression of XX-enriched interactors,
808 such as BMP2/BMR1A/AVR2B and WNT4/FZD10/LRP5, but not the XY-enriched

809 interactors CADM1/NECTIN3 and EFNB1/EPHB3 (Fig. 6e). Furthermore, we did not
810 observe a sex bias for the mean expression of interacting pairs for the shared pathway
811 DLK1:NOTCH3 between E11.5 XX and XY gonads (Fig. 6e). These data indicate that at
812 E11.5, sexual dimorphism of signaling pathways, in terms of number and mean expression,
813 is most prominent between XX supporting cells and XX PGCs.

814 In E12.5 and E13.5 gonads, we observed a stronger XX bias in ligand-receptor
815 interactions between supporting cells and PGCs (Fig. 6f-i). At E12.5, we observed 49 and
816 24 significant interactions in XX and XY gonads, respectively (Fig. 6f). In E12.5 XX gonads,
817 BMP, Insulin-like growth factor (IGF), R-spondin, and WNT signaling were the most
818 frequently observed pathways and displayed strong XX-enriched mean expression (Fig. 6f-
819 g). In XY gonads, detection of new signaling pathways emerged from E11.5 to E12.5,
820 including signaling by BMP6, chemokines, Matrix metalloproteinases (MMP), and
821 Neuregulin (Fig. 6f). Compared to E12.5, E13.5 XX gonads continued to have enrichment
822 for the BMP, IGF, and WNT signaling pathways with additional pathways related to midkine
823 (MDK), Calsyntenin (Clstn), and SLIT and NTRK-like protein signaling detected (Fig. 6f and
824 h). The BMP5:BMR1A:AVR2B and MDK:PTPRZ1 ligand-receptor interactors also showed
825 XX-biased mean expression (Fig. 6i). Similarly in E12.5 and E13.5 XY gonads, we observed
826 maintenance of the ligand-receptor pairs, BMP6:ACVR:1A2B and BMP6:BMR1A:AVR2B
827 (Fig. 6f and h). However, enrichment of signaling by Desert hedgehog, Notch, and WNT
828 inhibition became significant in E13.5 XY gonads but not E12.5 (Fig. 6f and h). High mean
829 expression of DHH:PTCH:GAS1 and DLK1:NOTCH3 interactors was detected in E13.5 XY
830 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.

835 **Discussion**

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

851

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

854 We first identified the sub-groupings of E11.5-E13.5 XX and XY PGCs and resolved
855 the patterns of gene expression and chromatin accessibility underlying each PGC
856 population. We observed that the combined analysis of integrated snRNA- and snATAC-seq
857 data produced better clustering and separation of PGCs than either assay alone. For

858 example, we found that at E11.5, XX and XY PGCs share highly similar transcriptomic and
859 open chromatin profiles, indicating that PGCs at this stage are bipotential at the level of
860 gene regulation and chromatin accessibility. This observation agrees with the study that
861 showed the global overlap of transcriptomes between E11.5 PGCs from both sexes, in
862 which they detected only 101 DEGs between XX and XY PGCs with functional annotations
863 related to chaperonin complexes and proteasome formation[104]. At E12.5, subpopulations of
864 PGCs clustered separately within each sex, supporting the model that PGCs do not commit
865 to a sex-specific fate in a synchronous manner[54]. At E13.5, two sub-clusters of XX PGCs
866 appear, consistent with the asynchronous entry of XX PGCs into meiosis[54]. Others also
867 reported substantial heterogeneity of human and mouse PGC populations, often arising
868 from intrinsic variation in the epigenetic reprogramming required for sex
869 differentiation[105,106]. These data underscore the discovery potential of simultaneously
870 measuring multiple layers of gene regulation to define the developmental fate of PGCs.
871

872 **Molecular signatures of differentiating PGCs revealed disease-related genes**
873 **associated with PGC clusters.**

874 When examining genes that could be responsible for the expression-based
875 clustering of PGCs, we found TFs, cell cycle regulators, chromatin modifiers, and ribosome
876 biogenesis genes underlying the diverse transcriptional profiles of differentiating PGCs.
877 Several of the expressed genes enriched in individual PGC clusters were involved in
878 infertility or disease processes when nonfunctional. For example, mutations in *Lars2*, the
879 mitochondrial leucyl-tRNA synthetase, is associated with premature ovarian failure in
880 women[50] and aberrant methylation at the promoter of *Mgmt,* a gene that is required for

881 DNA repair and linked to the formation of testicular germ cell tumors in humans[51]. The
882 dynamic and varied regulatory profiles of PGCs that we observed highlight the necessity of
883 pairing chromatin accessibility information with gene expression data for elucidating the
884 mechanisms underlying PGC sex determination.
885
886 **Sexual dimorphism in chromatin accessibility increases temporally during PGC**
887 **development.**
888 The chromatin landscape of PGCs is the most sexually dimorphic at E13.5 when
889 compared to E11.5 and E12.5. The substantial increase in DAPs at E13.5 leads to several
890 hypotheses that could account for the changes to PGC chromatin architecture during sex
891 determination. First, it is possible that the chromatin of bipotential E11.5 PGCs is primed
892 for activating either sex-specific differentiation pathway. Chromatin priming occurs when
893 chromatin is maintained in an accessible state prior to gene expression, and is evidenced in
894 a number of developmental contexts, such as during specification of hematopoietic cells
895 and lineage commitment of embryonic stem cells[107,108]. During development, chromatin
896 priming fine-tunes gene activation by enabling a rapid response to external cues that
897 induce cell differentiation[107]. Second, we posit that E11.5 PGCs have fewer DA peaks
898 because they are actively dividing. Mitotic chromosome condensation is known to induce a
899 state of heterochromatin and the eviction of TFs from chromatin in actively dividing
900 cells[109]. Third, the increased number of DA peaks at E13.5 may be the result of changes to
901 chromatin structure as XX PGCs enter meiotic prophase I, an event that requires dramatic
902 changes to meiotic chromosomes as they undergo DNA double-strand breaks, synapsis, and
903 crossing over[47]. Indeed, increased chromatin accessibility was found in mouse

904 spermatocytes during prophase I of meiosis[110]. Finally, the upregulated expression of
905 forkhead TFs in male PGCs at E12.5 and E13.5 may induce the *de novo* establishment of DA
906 peaks through their pioneering activity. Nonetheless, our results are consistent with
907 previous reports that demonstrate the global decrease in DNA methylation levels, or
908 increased chromatin accessibility, in mouse PGCs beginning at E9.5 and continuing through
909 E13.5[111].
910
911 **Single-nucleus multiomics identifies evidence for chromatin priming during PGC**
912 **lineage commitment.**
913 Moreover, not all sex- and stage-specific DA chromatin peaks account for changes in
914 differential gene expression for PGCs. Chromatin accessibility and gene expression are
915 often asynchronously correlated during cellular differentiation, such as during chromatin
916 priming[77]. For example, accessibility of regulatory loci involved in cell-fate decisions may
917 precede gene expression to prime cells towards one developmental lineage versus
918 another[77]. In the context of germ cell differentiation, PGCs must also undergo extensive
919 chromatin remodeling events, including genome-wide DNA demethylation and a reduction
920 in both repressive and activating histone modifications, prior to sex determination[112].
921 Indeed, chromatin priming has been observed in PGCs from other species, such as
922 macaques, pigs, and goats[105]. Thus, it is likely that the DA chromatin peaks observed for
923 E11.5 and E12.5 PGCs in mice may either reflect global chromatin remodeling events
924 associated with sex determination or foreshadow future cell fate decisions by remaining
925 poised for TF binding. Nevertheless, a comprehensive profile of 3D chromatin structure

926 and enhancer-promoter contacts in differentiating PGCs is needed to fully understand how
927 changes to chromatin facilitate PGC sex determination.
928
929 **The gene regulatory networks of XX PGCs are enriched for the transcription factors,**
930 **TFAP2c, TCFL5, GATA2, MGA, NR6A1, TBX4, and ZFX**
931 Our multiomics approach uncovered compelling TF candidates that may be involved
932 in PGC development and differentiation. In XX PGCs, seven TF prospects were identified:
933 GATA2, MGA, NR6A1, TBX4, ZFX, TCFL5, and TFAP2c. Of these TFs, TCFL5 and TFAP2C
934 showed the strongest enriched expression in XX PGCs. TCFL5 binds to the promoters of
935 meiotic genes such as *Sycp1* (synapsis) and *Stag3* (cohesion) during spermatogenesis in the
936 mouse[79]. *Tcfl5* is indispensable for the completion of meiosis and spermatid maturation[79]
937 in male germ cells, and meiosis of female germ cells[81]. Our multiomics data also
938 demonstrate strong *Tfap2c* expression and motif enrichment in XX PGCs beginning at E12.5
939 and continuing through E13.5. These findings are corroborated by other single-cell RNA-
940 seq analyses that showed *Tfap2c* expression in human and mouse PGCs and gonocytes[80,113-
941 116]. Additionally, TFAP2c has several reported roles in cell-type specification in both mouse
942 and human, including germline specification and maintenance of germ cell identity[80],
943 remodeling of naïve enhancers[117], and hormone responsiveness[118]. Loss of TFAP2c in mice
944 also leads to a reduction in germ cell number by either PGC apoptosis or PGC
945 differentiation into somatic cells[119]. Considering that changes in gene regulation leading to
946 the female and male pathways begin as early as E12.5[3], understanding how TCFL5 and
947 TFAP2c activity is induced and investigating the role of TFAP2c in PGC sex determination
948 are important future directions.

949

950 **XY PGCs showed sex-specific enrichment of transcription factors associated with**

951 **reproductive dysfunction.**

952 In XY PGCs, we detected significant enrichment of the forkhead box and POU6

953 families of TFs. In particular, FOXO1, FOXK2, and POU6F1/2 exhibited XY biased

954 expression and *in silico* chromatin binding. Although the function of *Foxo1* has been studied

955 in gonocytes[120], little is known about the roles of *Foxk2* and *Pou6f1/2* in PGCs. Loss of *Foxo1*

956 in embryonic gonocytes leads to defective spermatogenesis in adulthood[120], and

957 inactivation of *Foxo1* in spermatogonia results in germ cell loss and male infertility[121].

958 FOXO1 activity is also widely involved in regulating metabolic homeostasis, redox balance,

959 cell death, and DNA damage repair[122]. Likewise, FOXK2 functions in similar roles as FOXO1,

960 such as in cellular metabolism, cell survival, and apoptosis[123]. Furthermore, the POU6

961 family consists of POU6F1 and POU6F2, which share highly conserved DNA motif

962 recognition sequences and function in similar cellular pathways[124]. Both POU6F1 and

963 POU6F2 have been reported to suppress tumor proliferation[125], and POU6F2 mutations

964 have been implicated in the progression of Wilms Tumor, hypogonadism, and pubertal

965 failure in human patients[126]. Investigating the cooperativity of these TFs will greatly

966 improve our understanding of how mitotic arrest and the commitment to spermatogenesis

967 are regulated during germ cell development.

968

969 **Single-nucleus RNA-seq maps the temporal expression patterns of WNT, BMP, and**

970 **RA signaling during PGC sex determination.**

971 During sex determination, PGCs respond to signals from the supporting cells of the
972 gonads to determine their sex-specific fate. In the fetal ovary, *Runx1+* supporting cells
973 produce WNT4 to maintain germ cell pluripotency and prevent precocious meiotic
974 entry[8,60]. We observed signaling by WNT4, WNT5A/B, and WNT6 in both XX and XY gonads
975 as early as E11.5. The number of interactions related to WNT signaling also increased from
976 E11.5 to E13.5 in both XX and XY gonads even though diminished WNT signaling is
977 required for meiotic entry[60]. However, XY PGCs showed enrichment of the WNT agonist
978 SFRP2[127] at E13.5, indicating that WNT signaling may be inhibited in XY PGCs earlier than
979 XX PGCs. BMP secretion by ovarian supporting cells also plays a critical role in regulating
980 meiotic entry[8]. Even though BMP6 signaling was observed in XY PGCs, BMP2 and BMP5
981 signaling were specific to XX PGCs beginning at E11.5. Our findings are consistent with
982 functional studies showing that BMP2 signaling induces *Zglp1* expression and promotes the
983 oogenic fate[10]. Additionally, RA produced by both ovarian and mesonephric somatic cells is
984 required to initiate expression of *Stra8* and *Meiosin*, two TFs that are essential for
985 completion of meiosis[8,9]. Significant enrichment of the RA signaling pathway was not
986 detected between supporting cells and PGCs in XX or XY gonads at any embryonic stage
987 with CellphoneDB. Given that RA is primarily produced by the mesonephros[66], detection of
988 RA signaling between gonadal supporting cells and PGCs is unlikely. However, *in silico*
989 chromatin binding of RA receptors was enriched in XX PGCs and the RA receptor motifs
990 were found in the regulatory loci for *Rec8, Stra8, Sycp2, Sycp3,* and *Zglp1*, which is
991 consistent with previous studies[128]. Taken together, our findings provide insight into the
992 temporal patterns of WNT, BMP, and RA signaling during PGC differentiation.
993

994 **Discovery analysis identifies potentially new signaling pathways between gonadal**
995 **supporting cells and PGCs.**

996 In addition to these pathways known for their roles in PGC sex determination, we
997 identified several uncharacterized ligand-receptor pairs with sex-specific enrichment. For
998 example, midkine (MDK) signaling was enriched in E13.5 XX gonads. Although the role of
999 MDK signaling in PGCs is not understood, MDK is known to be important for maintaining
1000 cell survival and growth and promoting cell migration[129]. MDK is a heparin-binding
1001 cytokine and, together with pleiotrophin, regulates female reproductive activities such as
1002 oocyte maturation and the estrous cycle[130]. Furthermore, in E13.5 XY gonads, we found
1003 enrichment of the Desert Hedgehog (DHH) signaling pathway between *Sox9+* cells and XY
1004 PGCs. In addition to regulating fetal Leydig cell differentiation[131], DHH signaling has also
1005 been shown to regulate germ cell maturation and spermatogenesis[132]. Ablation of *Dhh* in
1006 adult XY mice causes infertility due to a loss of mature sperm[132], however, whether *Dhh*
1007 regulates germ cell development in fetal XY mice remains to be determined. Overall, we
1008 posit that these and other signaling pathways are important for promoting PGC survival
1009 and differentiation.

1010 Our study demonstrated the benefit of using multiomics approaches to probe the
1011 cellular and developmental pathways involved in cell fate decisions, gametogenesis, and
1012 PGC differentiation. Unlike traditional assays, joint RNA and ATAC analyses from individual
1013 PGCs enabled the detection of paired chromatin accessibility and gene expression
1014 information from the same cell to discover new gene regulatory interactions. We show that
1015 our multiomics data accurately identified enrichment of known pathways involved in PGC
1016 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.

1026 **Acknowledgments**

1027 We thank the late Dr. Keith Parker (UT Southwestern Medical Center, USA) for the Nr5a1-
1028 cre mice. Valuable discussions with F. DeMayo, C. Williams, M. Morgan, C. Guardia, and all
1029 members of the Yao lab during the course of this project contributed greatly to our
1030 manuscript. We are especially grateful for P. Brown for her advice and assistance with
1031 experiments and editorial comments in the writing process. We are grateful to the NIEHS
1032 Comparative Medicine Branch for mouse colony maintenance and to the Epigenomics and
1033 DNA Sequencing Core and the Integrative Bioinformatics Support Group for their
1034 assistance with sequencing and data analysis. This work was supported by the Intramural
1035 Research Program (ES102965 to H.H.-C.Y.) of the NIH, National Institute of Environmental
1036 Health Sciences and the Postdoctoral Research Associate Training (PRAT) Program (Fi2)
1037 Fellowship of the NIH, National Institute of General Medical Sciences awarded to A.K.A.

1038

1039 **Conflict of interest Statement (if requested by the journal)**

1040 The authors declare no competing financial or non-financial interests.

1041 **Figure legends**

1042 **Figure 1: Single-nucleus multiomics sequencing of E11.5-E13.5 XX and XY PGCs.** (a)
1043 Workflow of single-nucleus multiomics analyses. (b-c) Dimensional reduction (UMAP) and
1044 cell-type specific clustering of snRNA-seq data collected from E11.5-E13.5 XX and XY PGCs.
1045 Each dot represents an individual cell color coded by either embryonic stage and sex (a) or
1046 cluster number (c). (d) Cell type composition of snRNA-seq clusters. The embryonic stage
1047 and sex of the PGC population is indicated on the *x*-axis and the unbiased clustering
1048 number is represented on the *y*-axis. (e-f) UMAP of snATAC-seq data collected from E11.5-
1049 E13.5 XX and XY PGCs color coded by either embryonic stage and sex (e) or cluster number
1050 (f). (g) Cell type composition of snATAC-seq clusters. (h-i) UMAP visualization of E11.5-
1051 E13.5 XX and XY PGCs using a joint, or weighted, measurement of snRNA- and snATAC-seq
1052 modalities. (h) Trajectory of transcriptomic changes (also known as pseudo-time analysis
1053 in Monocle) among PGCs overlaid on the joint UMAP. Arrows indicate the path of the
1054 developmental trajectory. Numbers on the pseudotime trajectory indicate the following: '0'
1055 represents the starting cell population; '1' indicates the branching point in the trajectory
1056 path; and '2' represents the terminal cell populations of the trajectory paths. Each PGC is
1057 color coded by embryonic stage and sex. (i) Unbiased clustering of PGCs using combined
1058 snRNA- and snATAC-seq data. Each PGC is color coded by cluster number. (j) Cell type
1059 composition of joint UMAP clusters. The number of PGCs per sex and embryonic stage are:
1060 375 E11.5 XX PGCs; 1,106 E12.5 XX PGCs; 750 E13.5 XX PGCs; 110 E11.5 XY PGCs; 465
1061 E12.5 XY PGCs; and 348 E13.5 XY PGCs.

1062

1063 **Figure 2: Molecular characterization of genes known for their roles in male PGC**
1064 **differentiation.** (a) Dotplot of the average expression and percentage of cells expressing
1065 *Rb1*, *Rbl2*, *Cdkn1b*, *Cdkn2b*, *Bnc2*, *Cnot1*, *Dnd1*, *Nanos2*, *Nanos3*, and *Zkscan5*. The DNA
1066 binding motif for ZKSCAN5 is indicated above. The color scale represents the average
1067 expression level, and the size of the dot represents the percentage of cells expressing the
1068 gene. (b) *Left:* Coverage plot of the normalized snATAC-seq signal at the *Bnc2* locus. *Right:*
1069 Violin plot of *Bnc2* expression in E11.5-E13.5 XX and XY PGCs. (c) Joint UMAP showing *Bnc2*
1070 expression levels for E11.5-E13.5 XX and XY PGCs. Individual cells are color coded by the
1071 level of *Bnc2* expression. (d) Candidate factors (*x*-axis) regulating *Bnc2* expression based on
1072 their Cistrome DB regulatory potential score (*y*-axis). (e) Dotplot of the average expression
1073 and percentage of cells expressing the candidate factors that potentially regulate *Bnc2*
1074 expression. The color scale represents the average expression level, and the size of the dot
1075 represents the percentage of cells.
1076
1077 **Figure 3: Molecular characterization of genes known for their roles in female PGC sex**
1078 **determination.** (a) Dotplot of the average expression and percentage of cells expressing
1079 *Rec8*, *Sycp2*, *Sycp3*, *Ctnnb1*, *Meioc*, *Rnf2*, *Stra8*, *Ythdc2*, *Zglp1*, and *Dmrt1*. The DNA binding
1080 motif for DMRT1 is indicated. The color scale represents the average expression level, and
1081 the size of the dot represents the percentage of cells expressing the gene. (b) *Left:* Coverage
1082 plot of the normalized snATAC-seq signal at the *Stra8* locus. *Right:* Violin plot of *Stra8*
1083 expression in E11.5-E13.5 XX and XY PGCs. (c) Joint UMAP showing *Stra8* expression levels
1084 for E11.5-E13.5 XX and XY PGCs. Individual cells are color coded by the level of *Stra8*
1085 expression. (d) Candidate factors (*x*-axis) regulating *Stra8* expression based on their

1086 Cistrome DB regulatory potential score (*y*-axis). (e) Dotplot of the average expression and
1087 percentage of cells expressing the candidate factors that potentially regulate *Stra8*
1088 expression. The color scale represents the average expression level, and the size of the dot
1089 represents the percentage of cells.
1090
1091 **Figure 4: Identification of differentially expressed genes and differentially accessible**
1092 **peaks in E11.5-E13.5 XX and XY PGCs.** (a) Bar graph showing the number of differentially
1093 expressed genes (DEGs) between XX and XY PGCs at E11.5-E13.5. (b) Bar graph showing
1094 the number of differentially accessible chromatin peaks (DAPs) between XX and XY PGCs at
1095 E11.5-E13.5. (c) *Left:* Coverage plot of the normalized snATAC-seq signal at the DEG *Porcn*
1096 locus. Peak-to-gene linkages are indicated by the 'links' line. *Right:* Violin plot of *Porcn*
1097 expression in E11.5-E13.5 XX and XY PGCs. (d) *Left:* Coverage plot of the normalized
1098 snATAC-seq signal at the DEG *Rimbp1* locus. Peak-to-gene linkages are indicated by the
1099 'links' line. *Right:* Violin plot of *Rimbp1* expression in E11.5-E13.5 XX and XY PGCs. (e) DAPI
1100 staining (grey) and immunofluorescence staining of TRA98 (pink), PORCN (yellow), and
1101 NR2F2 (blue) in E13.5 XX and XY gonads from C57Bl/6 and CD1 mice.
1102
1103 **Figure 5: Analysis of gene networks in E11.5-E13.5 XX and XY PGCs.** (a) Illustration of
1104 the analytical flow to infer gene networks (*i.e.* transcription factors (TFs) and their
1105 predicted target genes) involved in PGC sex determination. (b-c; i-j) Line plots showing the
1106 expression levels and *in silico* chromatin binding, termed 'motif activity', of *Tfap2c* (b), *Tcfl5*
1107 (c), *Foxk2* (i), and *Pou6f2* (j) in E11.5-E13.5 XX (pink) and XY (teal) PGCs. The *p*-value of
1108 motif enrichment and the TF binding motif are indicated. (d and k) Enrichment analysis of

1109 biological process gene ontology (GO) of the predicted target genes for TFAP2c (d) and
1110 FOXK2 (k). The color scale represents the *p*-value of GO term enrichment, and the size of
1111 the dot indicates the gene ratio for each GO term. (e and l) The number of predicted target
1112 genes for enriched TFs in XX PGCs (e) and XY PGCs (l). (f) Circos plot showing whether the
1113 motifs for TFAP2C, TCFL5, ZFX, MGA, or NR6A1 (indicated in black) are detected in the
1114 regulatory loci linked to *Tbx4, Nr6a1, Mga, Tcfl5,* or *Tfap2c* (indicated in grey). (m) Venn
1115 diagram showing the overlap of predicted target genes for POU6F1/2, FOXK2, and FOXO1.
1116 (g-h) Icon array showing the presence (pink/purple) or absence (light grey) of the motifs
1117 for TFAP2c, TCFL5, ZFX, MGA, NR6A1, TBX4, and GATA2 in peaks linked to genes related to
1118 meiosis (g) or WNT signaling (h). (n-o) Icon array showing the presence (teal) or absence
1119 (light grey) of the motifs for FOXO1, FOXK2, and POU6F1/2 in peaks linked to genes related
1120 to mitosis (n) or signal transduction (o).
1121
1122 **Figure 6: Sexually dimorphic interactions between PGCs and gonadal supporting**
1123 **cells.** (a) Illustration of cell communication between PGCs and gonadal supporting cells. (b)
1124 Number of significant interactions or ligand-receptor pairs (*p*-value < 0.05) between
1125 supporting cells and PGCs in E11.5-E13.5 XX and XY gonads. (c) Number of significantly
1126 enriched ligand-receptor pairs, or interactions, per signaling pathways induced by either
1127 WNTs, inhibins/activins, or BMPs in E13.5 XX and XY gonads. (d, f, & h) Venn diagrams
1128 showing the overlap of significant interactions between supporting cells and PGCs in XX
1129 and XY gonads at E11.5 (d), E12.5 (f), and E13.5 (h). (e, g, & i) Dotplots showing the mean
1130 expression for all interacting partners in representative ligand-receptor pairs in XX and XY
1131 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.

References

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

2 Hancock, G. V., Wamaitha, S. E., Peretz, L. & Clark, A. T. Mammalian primordial germ cell specification. *Development* **148**, dev189217 (2021).

3 Spiller, C. M. & Bowles, J. Sex determination in mammalian germ cells. *Asian Journal of Andrology* **17**, 427 (2015).

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

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

6 Czukiewska, S. M. & de Sousa Lopes, S. M. C. in *Seminars in Cell & Developmental Biology*. 58-65 (Elsevier).

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

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

9 Ishiguro, K.-i. *et al*. MEIOSIN directs the switch from mitosis to meiosis in mammalian germ cells. *Developmental cell* **52**, 429-445. e410 (2020).

10 Nagaoka, S. I. *et al.* ZGLP1 is a determinant for the oogenic fate in mice. *Science* **367**, eaaw4115 (2020).

11 Vanhoutteghem, A. *et al.* 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 (2014).

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

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

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

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

16 Jameson, S. A. *et al.* Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS genetics* **8**, e1002575 (2012).

17 Lesch, B. J., Dokshin, G. A., Young, R. A., McCarrey, J. R. & Page, D. C. 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 (2013).

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

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

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

21 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 14049 (2017).

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

23 Griffiths, J. A., Richard, A. C., Bach, K., Lun, A. T. & Marioni, J. C. Detection and removal of barcode swapping in single-cell RNA-seq data. *Nature communications* **9**, 2667 (2018).

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

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

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

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

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

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

30 Stévant, I. *et al.* Dissecting cell lineage specification and sex fate determination in gonadal somatic cells using single-cell transcriptomics. *Cell reports* **26**, 3272-3283. e3273 (2019).

31 Zhao, C. *et al.* 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 (2009).

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

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

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

35 Kuony, A. & Michon, F. Epithelial markers aSMA, Krt14, and Krt19 unveil elements of murine lacrimal gland morphogenesis and maturation. *Frontiers in physiology* **8**, 739 (2017).

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

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

38 Sun, H. *et al.* 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**, 38 (2018).

39 Wang, Q. *et al.* Exploring epigenomic datasets by ChIPseeker. *Current Protocols* **2**, e585 (2022).

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

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

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

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

44 Mei, S. *et al.* Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic acids research*, gkw983 (2016).

45 Troulé, K. *et al.* CellPhoneDB v5: inferring cell-cell communication from single-cell multiomics data. *arXiv preprint arXiv:2311.04567* (2023).

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

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

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

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

50 Pierce, S. B. *et al.* 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 (2013).

51 da Silva Martinelli, C. M. *et al.* MGMT and CALCA promoter methylation are associated with poor prognosis in testicular germ cell tumor patients. *Oncotarget* **8**, 50608 (2017).

52 Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902. e1821 (2019).

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

54 Soygur, B. *et al.* Intercellular bridges coordinate the transition from pluripotency to meiosis in mouse fetal oocytes. *Science advances* **7**, eabc6747 (2021).

55 Suzuki, A., Niimi, Y. & Saga, Y. 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 (2014).

56 Dreyer, S. D., Zheng, Q., Zabel, B., Winterpacht, A. & Lee, B. 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 (1999).

57 Li, J. *et al.* ZKSCAN5 activates VEGFC expression by recruiting SETD7 to promote the Lymphangiogenesis, tumour growth, and metastasis of breast cancer. *Frontiers in Oncology* **12**, 875033 (2022).

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

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

60 Le Rolle, M. *et al.* 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**, e2023376118 (2021).

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

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

63 Garcia-Cruz, R. *et al.* 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 (2010).

64 Endo, T., Mikedis, M. M., Nicholls, P. K., Page, D. C. & de Rooij, D. G. Retinoic acid and germ cell development in the ovary and testis. *Biomolecules* **9**, 775 (2019).

65 Vernet, N. *et al.* Meiosis occurs normally in the fetal ovary of mice lacking all retinoic acid receptors. *Science advances* **6**, eaaz1139 (2020).

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

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

68 Chassot, A.-A. *et al.* Retinoic acid synthesis by ALDH1A proteins is dispensable for meiosis initiation in the mouse fetal ovary. *Science advances* **6**, eaaz1261 (2020).

69 Napoli, J. L. & Yoo, H. S. in *Methods in enzymology* Vol. 637 55-75 (Elsevier, 2020).

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

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

72 Li, H., Kim, J. H., Koh, S. S. & Stallcup, M. R. 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 (2004).

73 Ma, Z., Lou, S. & Jiang, Z. PHLDA2 regulates EMT and autophagy in colorectal cancer via the PI3K/AKT signaling pathway. *Aging (Albany NY)* **12**, 7985 (2020).

74 Navarro-Costa, P. *et al.* Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. *Nature communications* **7**, 12331 (2016).

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

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

77 Ma, S. *et al.* Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103-1116. e1120 (2020).

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

79 Galán-Martínez, J. *et al.* TCFL5 deficiency impairs the pachytene to diplotene transition during spermatogenesis in the mouse. *Scientific Reports* **12**, 10956 (2022).

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

81 Zhang, F.-L. *et al.* Single cell epigenomic and transcriptomic analysis uncovers potential transcription factors regulating mitotic/meiotic switch. *Cell Death & Disease* **14**, 134 (2023).

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

83 Xiao, W. *et al.* POU6F1 cooperates with RORA to suppress the proliferation of lung adenocarcinoma by downregulating HIF1A signaling pathway. *Cell Death & Disease* **13**, 427 (2022).

84 Yu, K. *et al.* Identification of a lipid scrambling domain in ANO6/TMEM16F. *elife* **4**, e06901 (2015).

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

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

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

88 Lehto, M. & Olkkonen, V. M. 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 (2003).

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

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

91 Jia, S. & Meng, A. TGFβ family signaling and development. *Development* **148**, dev188490 (2021).

92 Bachus, S. *et al.* In Mitosis You Are Not: The NIMA Family of Kinases in Aspergillus, Yeast, and Mammals. *International Journal of Molecular Sciences* **23**, 4041 (2022).

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

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

95 Song, Q. *et al.* Knockout of ENO1 leads to metabolism reprogramming and tumor retardation in pancreatic cancer. *Frontiers in Oncology* **13**, 1119886 (2023).

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

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

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

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

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

101 Da Silva, S. M. *et al.* 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 (2004).

102 Nicol, B. *et al.* RUNX1 maintains the identity of the fetal ovary through an interplay with FOXL2. *Nature communications* **10**, 5116 (2019).

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

104 Mayère, C. *et al.* Single-cell transcriptomics reveal temporal dynamics of critical regulators of germ cell fate during mouse sex determination. *The FASEB Journal* **35**, e21452 (2021). https://doi.org:https://doi.org/10.1096/fj.202002420R

105 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).

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

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

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

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

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

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

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

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

114 Chen, D. *et al.* Human primordial germ cells are specified from lineage-primed progenitors. *Cell reports* **29**, 4568-4582. e4565 (2019).

115 Schemmer, J. *et al.* Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice. *PloS one* **8**, e71113 (2013).

116 Schäfer, S. *et al.* 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 (2011).

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

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

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

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

121 Shen, Z. *et al.* The function of Foxo1 in spermatogonia development is independent of PI3K/PTEN signaling. *The FASEB Journal* **36**, e22522 (2022).

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

123 Kang, Y., Zhang, K., Sun, L. & Zhang, Y. Regulation and roles of FOXK2 in cancer. *Frontiers in Oncology* **12**, 967625 (2022).

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

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

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

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

128 Koubova, J. *et al.* Retinoic acid activates two pathways required for meiosis in mice. *PLoS genetics* **10**, e1004541 (2014).

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

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

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

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

Figure 1

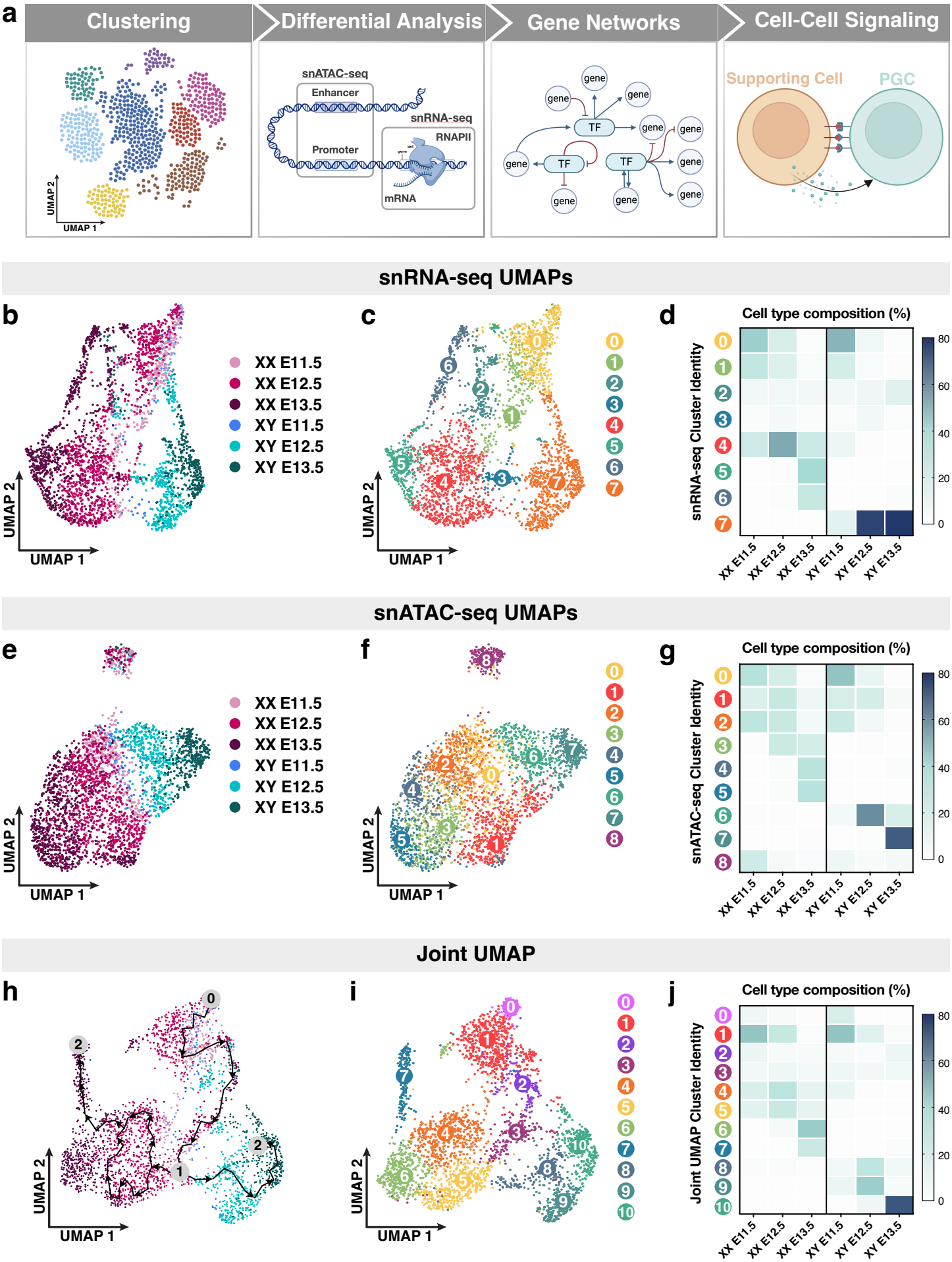

a
Clustering
Differential Analysis
Gene Networks
Cell-Cell Signaling
UMAP 2
UMAP 1
snATAC-seq
Enhancer
Promoter
snRNA-seq
RNAPII
mRNA
gene
TF
Supporting Cell
PGC
snRNA-seq UMAPs
b
XX E11.5
XX E12.5
XX E13.5
XY E11.5
XY E12.5
XY E13.5
c
d
Cell type composition (%)
snRNA-seq Cluster Identity
snATAC-seq UMAPs
e
f
g
snATAC-seq Cluster Identity
Joint UMAP
h
i
j
Joint UMAP Cluster Identity

Figure 2

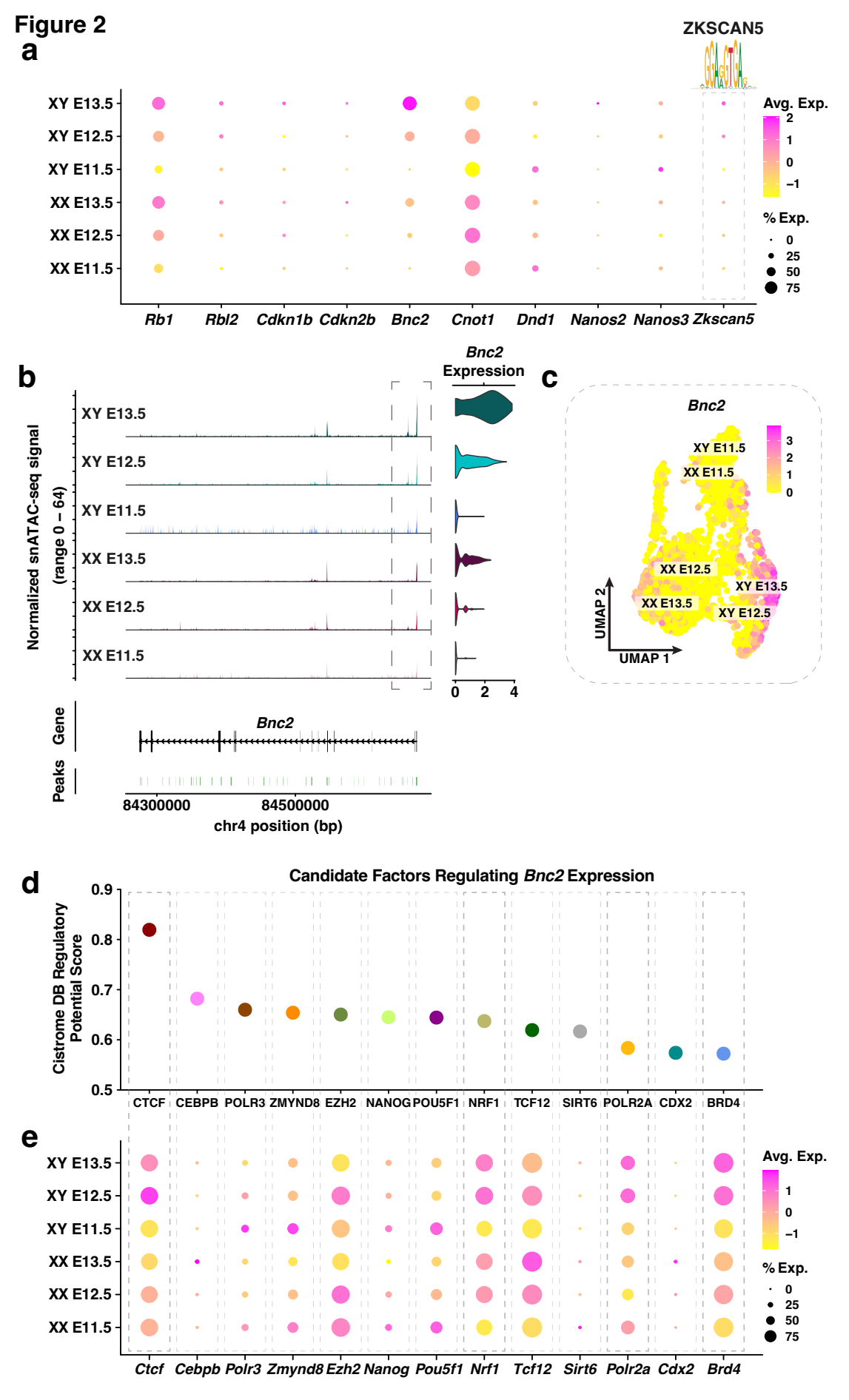

a
ZKSCAN5
XY E13.5
XY E12.5
XY E11.5
XX E13.5
XX E12.5
XX E11.5
Rb1
Rbl2
Cdkn1b
Cdkn2b
Bnc2
Cnot1
Dnd1
Nanos2
Nanos3
Zkscan5
Avg. Exp.
% Exp.
b
Bnc2 Expression
Normalized snATAC-seq signal (range 0 – 64)
Gene
Peaks
84300000
84500000
chr4 position (bp)
c
Bnc2
UMAP 1
UMAP 2
d
Candidate Factors Regulating Bnc2 Expression
Cistrome DB Regulatory Potential Score
CTCF
CEBPB
POLR3
ZMYND8
EZH2
NANOG
POU5F1
NRF1
TCF12
SIRT6
POLR2A
CDX2
BRD4
e
Ctcf
Cebpb
Polr3
Zmynd8
Ezh2
Nanog
Pou5f1
Nrf1
Tcf12
Sirt6
Polr2a
Cdx2
Brd4

Figure 3

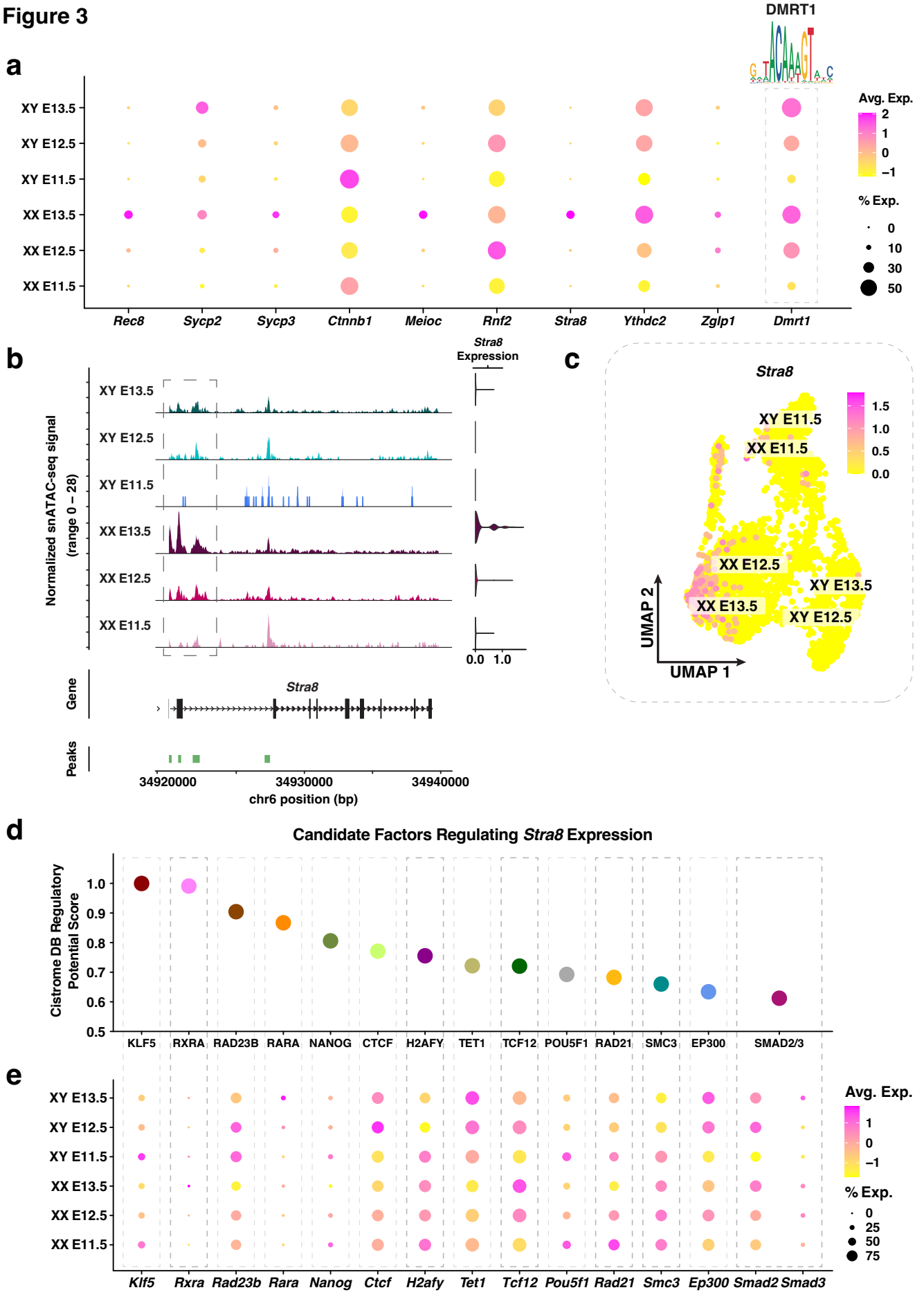
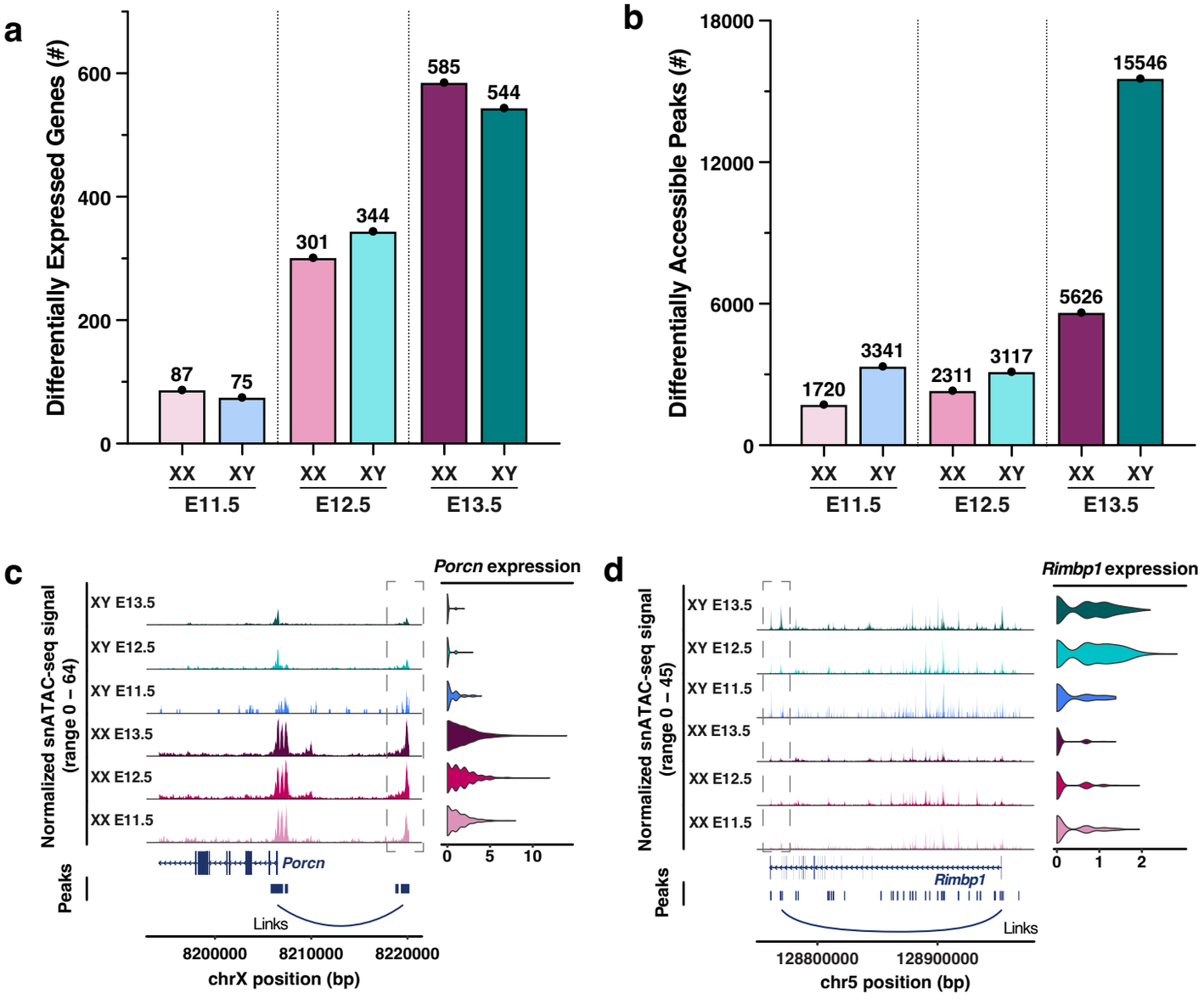

a
DMRT1
XY E13.5
XY E12.5
XY E11.5
XX E13.5
XX E12.5
XX E11.5
Rec8
Sycp2
Sycp3
Ctnnb1
Meioc
Rnf2
Stra8
Ythdc2
Zglp1
Dmrt1
Avg. Exp.
2
1
0
−1
% Exp.
0
10
30
50
b
Stra8 Expression
Normalized snATAC-seq signal (range 0 – 28)
0.0
1.0
Gene
Stra8
Peaks
34920000
34930000
34940000
chr6 position (bp)
c
Stra8
1.5
1.0
0.5
0.0
UMAP 2
UMAP 1
d
Candidate Factors Regulating Stra8 Expression
Cistrome DB Regulatory Potential Score
1.0
0.9
0.8
0.7
0.6
0.5
KLF5
RXRA
RAD23B
RARA
NANOG
CTCF
H2AFY
TET1
TCF12
POU5F1
RAD21
SMC3
EP300
SMAD2/3
e
Klf5
Rxra
Rad23b
Rara
Nanog
Ctcf
H2afy
Tet1
Tcf12
Pou5f1
Rad21
Smc3
Ep300
Smad2
Smad3
25
50
75

Figure 4


a
Differentially Expressed Genes (#)
600
400
200
0
87
75
301
344
585
544
XX
XY
XX
XY
XX
XY
E11.5
E12.5
E13.5
b
Differentially Accessible Peaks (#)
18000
12000
6000
0
1720
3341
2311
3117
5626
15546
XX
XY
XX
XY
XX
XY
E11.5
E12.5
E13.5
c
Porcn expression
Normalized snATAC-seq signal (range 0 – 64)
XY E13.5
XY E12.5
XY E11.5
XX E13.5
XX E12.5
XX E11.5
0
5
10
Porcn
Peaks
Links
8200000
8210000
8220000
chrX position (bp)
d
Rimbp1 expression
Normalized snATAC-seq signal (range 0 – 45)
XY E13.5
XY E12.5
XY E11.5
XX E13.5
XX E12.5
XX E11.5
0
1
2
Rimbp1
Peaks
Links
128800000
128900000
chr5 position (bp)

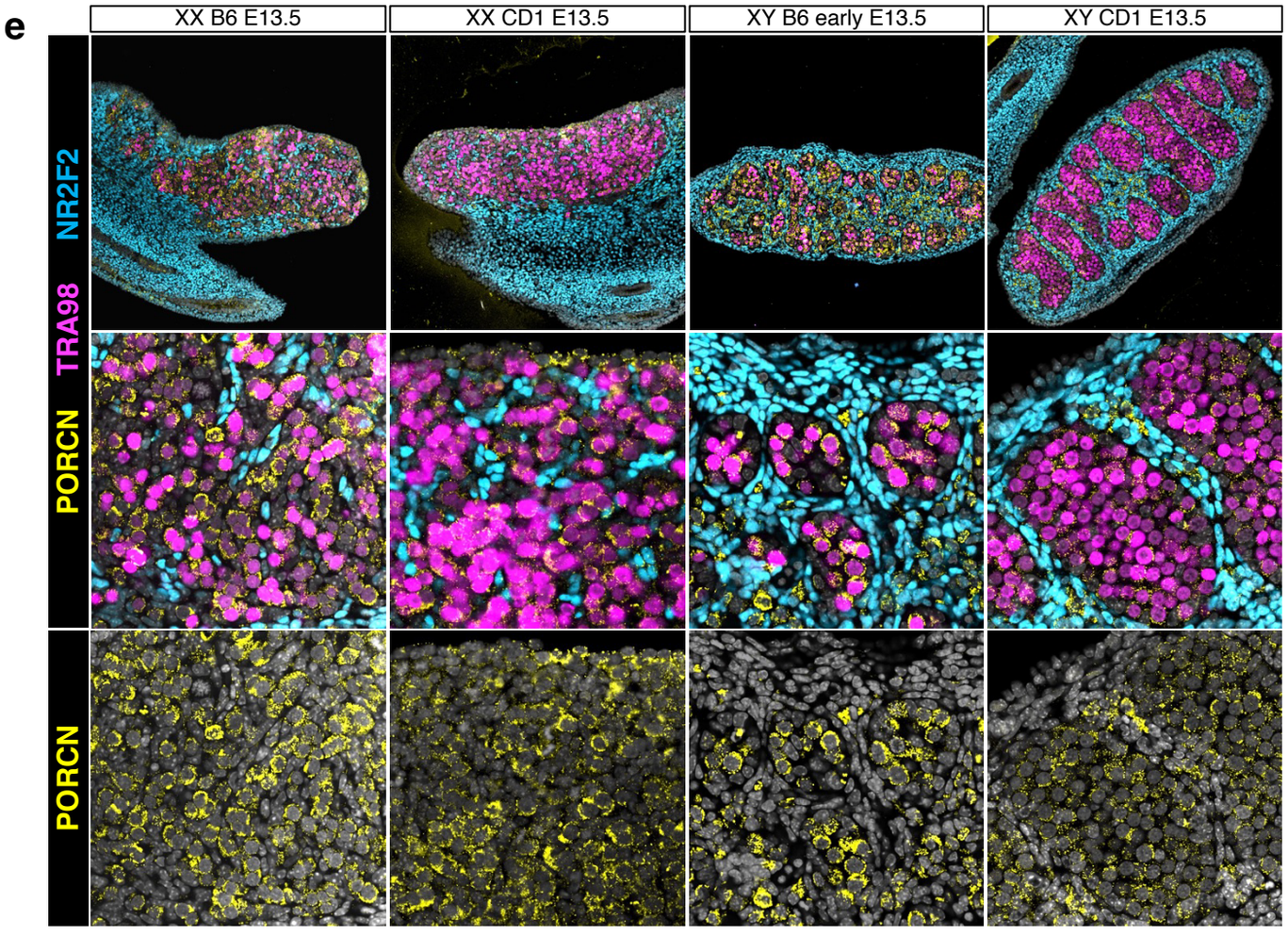

e
XX B6 E13.5
XX CD1 E13.5
XY B6 early E13.5
XY CD1 E13.5
NR2F2
TRA98
PORCN
PORCN

Figure 5

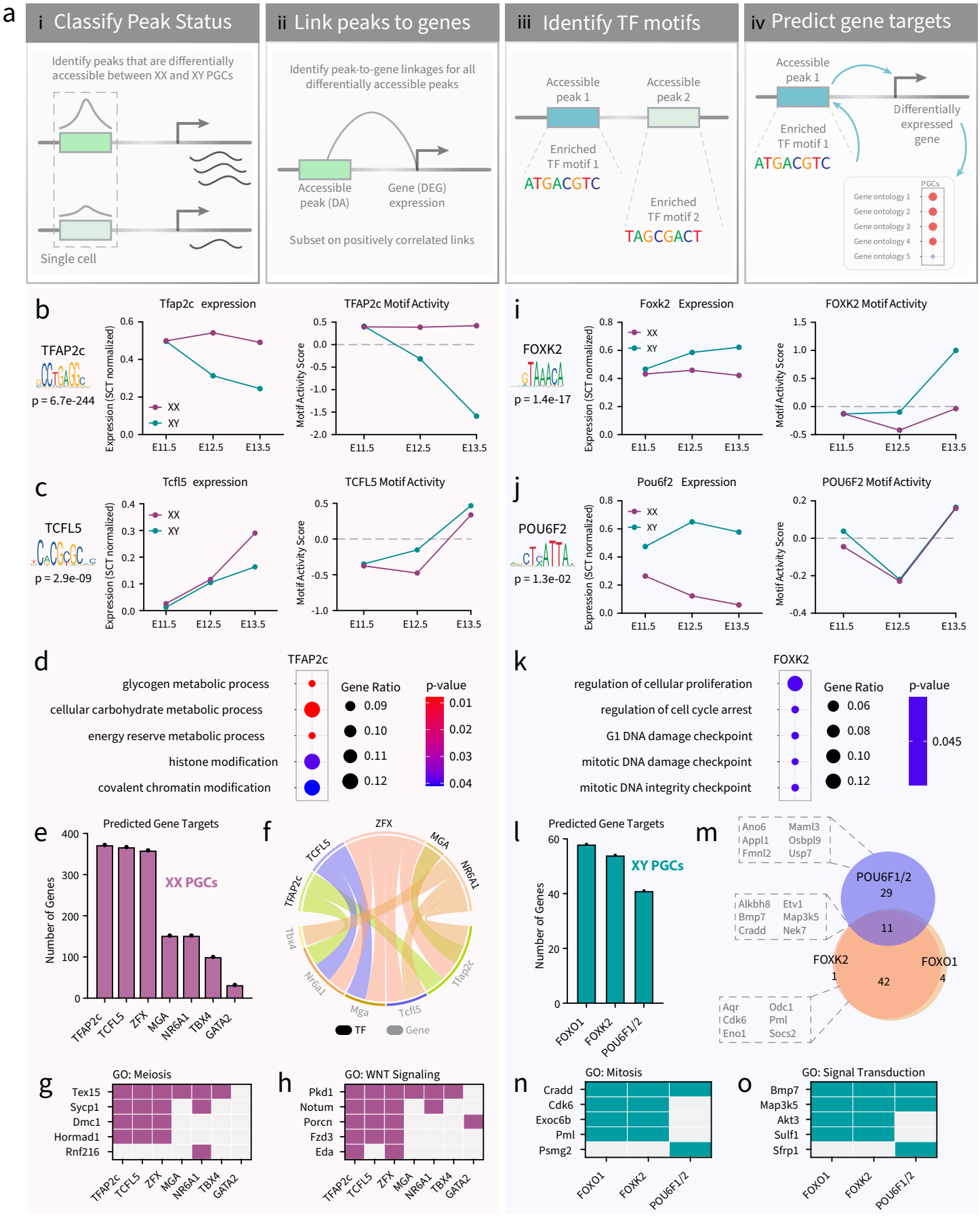

a
i Classify Peak Status
Identify peaks that are differentially accessible between XX and XY PGCs
Single cell
ii Link peaks to genes
Identify peak-to-gene linkages for all differentially accessible peaks
Accessible peak (DA)
Gene (DEG) expression
Subset on positively correlated links
iii Identify TF motifs
Accessible peak 1
Accessible peak 2
Enriched TF motif 1
ATGACGTC
Enriched TF motif 2
TAGCGACT
iv Predict gene targets
Accessible peak 1
Differentially expressed gene
Enriched TF motif 1
ATGACGTC
PGCs
Gene ontology 1
Gene ontology 2
Gene ontology 3
Gene ontology 4
Gene ontology 5
b
TFAP2c
p = 6.7e-244
Tfap2c expression
Expression (SCT normalized)
XX
XY
E11.5
E12.5
E13.5
TFAP2c Motif Activity
Motif Activity Score
c
TCFL5
p = 2.9e-09
Tcfl5 expression
TCFL5 Motif Activity
d
TFAP2c
glycogen metabolic process
cellular carbohydrate metabolic process
energy reserve metabolic process
histone modification
covalent chromatin modification
Gene Ratio
0.09
0.10
0.11
0.12
p-value
0.01
0.02
0.03
0.04
e
Predicted Gene Targets
Number of Genes
XX PGCs
TFAP2c
TCFL5
ZFX
MGA
NR6A1
TBX4
GATA2
f
ZFX
MGA
NR6A1
TCFL5
TFAP2c
Tbx4
Nr6a1
Mga
Tcfl5
Tfap2c
TF
Gene
g
GO: Meiosis
Tex15
Sycp1
Dmc1
Hormad1
Rnf216
h
GO: WNT Signaling
Pkd1
Notum
Porcn
Fzd3
Eda
i
FOXK2
p = 1.4e-17
Foxk2 Expression
FOXK2 Motif Activity
j
POU6F2
p = 1.3e-02
Pou6f2 Expression
POU6F2 Motif Activity
k
FOXK2
regulation of cellular proliferation
regulation of cell cycle arrest
G1 DNA damage checkpoint
mitotic DNA damage checkpoint
mitotic DNA integrity checkpoint
Gene Ratio
0.06
0.08
0.10
0.12
p-value
0.045
l
Predicted Gene Targets
XY PGCs
FOXO1
FOXK2
POU6F1/2
m
Ano6
Appl1
Fmnl2
Maml3
Osbpl9
Usp7
Alkbh8
Bmp7
Cradd
Etv1
Map3k5
Nek7
Aqr
Cdk6
Eno1
Odc1
Pml
Socs2
POU6F1/2
29
11
FOXK2
1
42
FOXO1
4
n
GO: Mitosis
Cradd
Cdk6
Exoc6b
Pml
Psmg2
o
GO: Signal Transduction
Bmp7
Map3k5
Akt3
Sulf1
Sfrp1

Figure 6

a
Supporting Cell
PGC
b
Significant Interactions (p < 0.05)
Supporting Cell to PGC
Number of Interactions
150
100
50
0
E11.5 XX
E11.5 XY
E12.5 XX
E12.5 XY
E13.5 XX
E13.5 XY
c
Number of Interactions per Pathway
Signaling by WNT
Signaling by Inhibin/Activin
Signaling by BMP
E13.5 XX
E13.5 XY
0
10
20
30
40
# of Sig. Interactions (p < 0.05)
Ligand-Receptor Pairs at E11.5
d
E11.5 XX
19
47
E11.5 XY
10
BMP2_BMR1A_AVR2B
EFNA2_EPHA4
FN1_integrin_aVb1_complex
NCAM1_FGFR1
PPIA_BSG
WNT4_FZD10_LRP5
CADM1_NECTIN3
EFNB1_EPHB3
NECTIN1_NECTIN3
SPP1_integrin_aVb3_complex
e
Mean Expression Values for All Interacting Partners
E11.5 XX
Runx1+ to PGC
E11.5 XY
Sox9+ to PGC
BMP2 BMR1A AVR2B
WNT4 FZD10 LRP5
DLK1 NOTCH1
CADM1 NECTIN3
EFNB1 EPHB3
Mean Value
0.50
1.00
1.50
1.75
Ligand-Receptor Pairs at E12.5
f
E12.5 XX
49
42
E12.5 XY
24
BMP2_BMR1A_AVR2B
BMP5_ACVR1_BMPR2
IGF2_IGF2R
RSPO1_Kremen1_LRP6
WNT5A_ROR2
BMP6_ACVR_1A2B_receptor
CX3CL1_CX3CR1
MMP2_integrin_aVb3_complex
NRG1_ERBB3
NTN4_NTRK2
g
Mean Expression Values for All Interacting Partners
E12.5 XX
Runx1+ to PGC
E12.5 XY
Sox9+ to PGC
BMP5 ACVR1 BMPR2
WNT5A ROR2
COL4A1 ADGRG6
BMP6 ACVR 1A2BRr
MMP2 integrin _aVb3 complex
Mean Value
0.50
1.00
1.50
1.75
Ligand-Receptor Pairs at E13.5
h
E13.5 XX
57
43
E13.5 XY
30
BMP5_BMR1A_AVR2B
MDK_PTPRZ1
CLSTN1_NRXN1
SLIT2_ROBO1
WNT4_FZD3_LRP5
BMP6_BMR1A_AVR2B
DHH_PTCH_GAS1
DLK1_NOTCH3
WNT5B_SFRP2
WNT6_SFRP2
i
Mean Expression Values for All Interacting Partners
E13.5 XX
Runx1+ to PGC
E13.5 XY
Sox9+ to PGC
BMP5 BMR1A AVR2B
MDK PTPRZ1
NECTIN1 NECTIN3
DHH PTCH GAS1
DLK1 NOTCH3
Mean Value
0.50
1.00
1.50
1.75