

RESEARCH

Open Access



The single-cell chromatin landscape in gonadal cell lineage specification

Hoi Ching Suen^{1†}, Fanghong Ou¹, Kai-kei Miu¹, Zhangting Wang¹, Wai-yee Chan¹ and Jinyue Liao^{2*†}

Abstract

Gonad development includes sex determination and divergent maturation of the testes and ovaries. Recent advances in measuring gene expression in single cells are providing new insights into this complex process. However, the underlying epigenetic regulatory mechanisms remain unclear. Here, we profiled chromatin accessibility in mouse gonadal cells of both sexes from embryonic day 11.5 to 14.5 using single-cell assay for transposase accessible chromatin by sequencing (scATAC-seq). Our results showed that individual cell types can be inferred by the chromatin landscape, and that cells can be temporally ordered along developmental trajectories. Integrative analysis of transcriptomic and chromatin-accessibility maps identified multiple putative regulatory elements proximal to key gonadal genes *Nr5a1*, *Sox9* and *Wt1*. We also uncover cell type-specific regulatory factors underlying cell type specification. Overall, our results provide a better understanding of the epigenetic landscape associated with the progressive restriction of cell fates in the gonad.

Keywords Gonad development, Sex determination, scATAC-seq, Chromatin accessibility, Regulatory genomics, Single-cell biology, Single-cell sequencing

Introduction

Gonadal differentiation has a decisive influence on sexual development. Defects in gonadal development cause Disorders of sex development (DSD), conditions with ambiguous genitalia, impaired steroid hormone production, reduced or null fertility, partial to complete sex reversal, and severe psychological pressure. The dysfunction of the gonadal system is also implicated in fertility problems, cancer, and polycystic ovary syndrome [8, 27].

Bipotential gonads, also known as genital ridges, are paired structures raised from proliferating coelomic epithelium (CE) on the ventromedial surface of the mesonephros at about four weeks post-coitum in human or embryonic day (E) 9.5 in mice [40, 52, 62]. Lineage progression in the gonads of both XX and XY embryos contains two major somatic cell populations—the supporting cells and the steroidogenic cells—both of which are raised from WT1- and NR5A1-positive progenitors [28, 33]. Supporting-like cells (SLCs), a newly discovered cell type, are the first somatic cell lineage to be specified in the bipotential gonad at E10.5, which have been shown to contribute to the formation of the rete testis and rete ovarii, as well as the pool of Sertoli and pregranulosa cells [42]. In XY mouse gonads, Sertoli cell lineage is specified by the expression of *Sry* at around E10 to E10.5, which in turn upregulates *Sox9* and leads to the formation of Sertoli cell precursors at E11.5 [3, 31]. Some somatic progenitors maintain their cellular states

[†]Hoi Ching Suen and Jinyue Liao have contributed equally to this work and share first authorship.

*Correspondence:

Jinyue Liao

liaojinyue@link.cuhk.edu.hk; liaojinyue@gmail.com

¹Developmental and Regenerative Biology Program, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

²Department of Chemical Pathology, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China



at this stage and differentiate into interstitial precursor cells later, which keep differentiating from E12.5 to E15.5 to produce sufficient steroidogenic fetal Leydig cells for male reproductive tract development [1, 2, 5, 9, 38, 64]. In XX mouse gonads, the -KTS alternatively spliced isoform of *Wt1* has been identified as a key determinant of female sex determination [26]. The somatic progenitor cells commit to supporting pre-granulosa cells or stromal progenitor cells. Embryonic granulosa cell recruitment happens between E11.5 and E14.5, contributing to the medullary follicles that are activated at birth [45]. The stromal progenitor cells remain undifferentiated in the fetal ovary and most of them transform into the steroidogenic theca cells when receiving Hedgehog signaling from granulosa cells near the time of birth [52]. Sex determination of the somatic cells further instruct germ cell differentiation as gonocytes in XY or oocytes in XX gonads [53].

With the advance of single cell sequencing technology, the gene expression profiles of sex-specific gonadal cell types in human, mouse, and chicken embryos have been well characterized at the single-cell level [18, 60, 61, 63]. Despite the increasing insights into gene functions in gonadal development, the underlying mechanism of most gonadal disorders still cannot be explained clearly. Understanding of gonadal development from an epigenetic perspective is important. One well-documented example is distal enhancers of haploinsufficient gene *SOX9* related to sex reversal [12]. Different from scRNA-seq, single-cell transposase accessible chromatin sequencing (scATAC-seq) uses Tn5 transposase to cleave and tag open DNA regions with sequencing adaptors in a single-cell droplet, thus providing the chromatin accessibility landscape of the whole genome. scATAC-seq is an effective tool for uncovering cell type-specific active transcription factors and associating genetic variants with disease causal cell types [11, 14].

Here, we characterize the cell type-specific epigenetic regulatory network and cellular differentiation trajectory of both sexes during mouse gonadal development using scATAC-seq. Using these knowledge and resources, we are able to gain a deeper understanding of epigenetic regulation underlying gonad development and lay the foundation for diagnosing unexplained gonadal disorders and designing effective therapies and in vitro gametogenesis in the future.

Results

Single-cell ATAC-seq revealed cellular heterogeneity in the XX and XY gonads

To resolve heterogeneous cell populations and delineate the chromatin dynamics in gonads during gonad development, we profiled the chromatin accessibility landscapes of female and male mouse gonads at four time

points (E11.5, E12.5, E13.5, and E14.5). We dissected the gonad/mesonephros complex and gently removed the mesonephros (except for E11.5). The morphology of the gonad/mesonephros complex was cross-referenced with previous morphological studies of gonad development to ensure the collection of gonads at the appropriate timing [37]. We then applied scATAC-seq using the BioRad SureCell platform (Fig. 1a). Altogether, we profiled the chromatin accessibility in 9494 individual cells after stringent quality control and heterotypic doublet removal (Supplementary Figure S1, Supplementary Table S1). Initial clustering by SnapATAC revealed 14 clusters, with various proportions from E11.5 to E14.5 (Fig. 1b and Supplementary Figure S2a). The sex of the cells was determined in silico by quantifying the reads mapped to the Y chromosome and normalizing to the sequencing depth to give the chrY score (Supplementary Figure S2b-c). Notably, Cluster 3 had a high chrY score, whereas Clusters 4 and 6 had a low chrY score. The remaining clusters exhibited a modest chrY score due to a mix of male and female cells. The clusters will be referred to by cluster number (0 to 14) and, where possible, are also named by their deduced cell types as described below.

We inferred the gene activity scores (which predict gene expression) compiled from chromatin accessibility signals within the gene body and promoter [25]. We then conducted label transfer from a recently published scRNA-seq reference dataset on mouse gonadal cells based on the coordination of gene score from scATAC-seq and gene expression levels from scRNA-seq [42]. Cellular identities were further confirmed by leveraging previous knowledge of the expression status of genes relevant to gonadal development, which were determined by transcriptomic and lineage tracing studies [18, 30, 43, 56, 61]. These include *Six2* in mesonephric cells primarily from E11.5 (Cluster 1), *Myh11* and *Acta2* in interstitial and stromal cells (Clusters 2 and 7), *Sox9* and *Dmrt1* in pre-Sertoli cells (Cluster 3), *Foxl2* in pre-granulosa cells (Cluster 4), *Upk3b* in coelomic epithelium (Cluster 5), *Dazl* and *Oct4 (Pou5f1)* in germ cells (Cluster 8), *Insl3* and *Star* in fetal Leydig cells (Cluster 11), *Cdh5* in endothelial cells (Cluster 12), *Car2* in erythrocyte progenitors (Cluster 13), as well as *Icam1* in immune cells (Cluster 14) (Fig. 1c).

Taken together, these results show that scATAC-seq is capable of distinguishing all the major cell types in gonads.

A high-resolution cis-regulatory atlas of the developing mouse gonad

To further explore our data, we examined the genomic regions with cell-type differential chromatin accessibilities. We first identified confident ATAC-seq peaks in each cluster using MACS2, and then we combined these

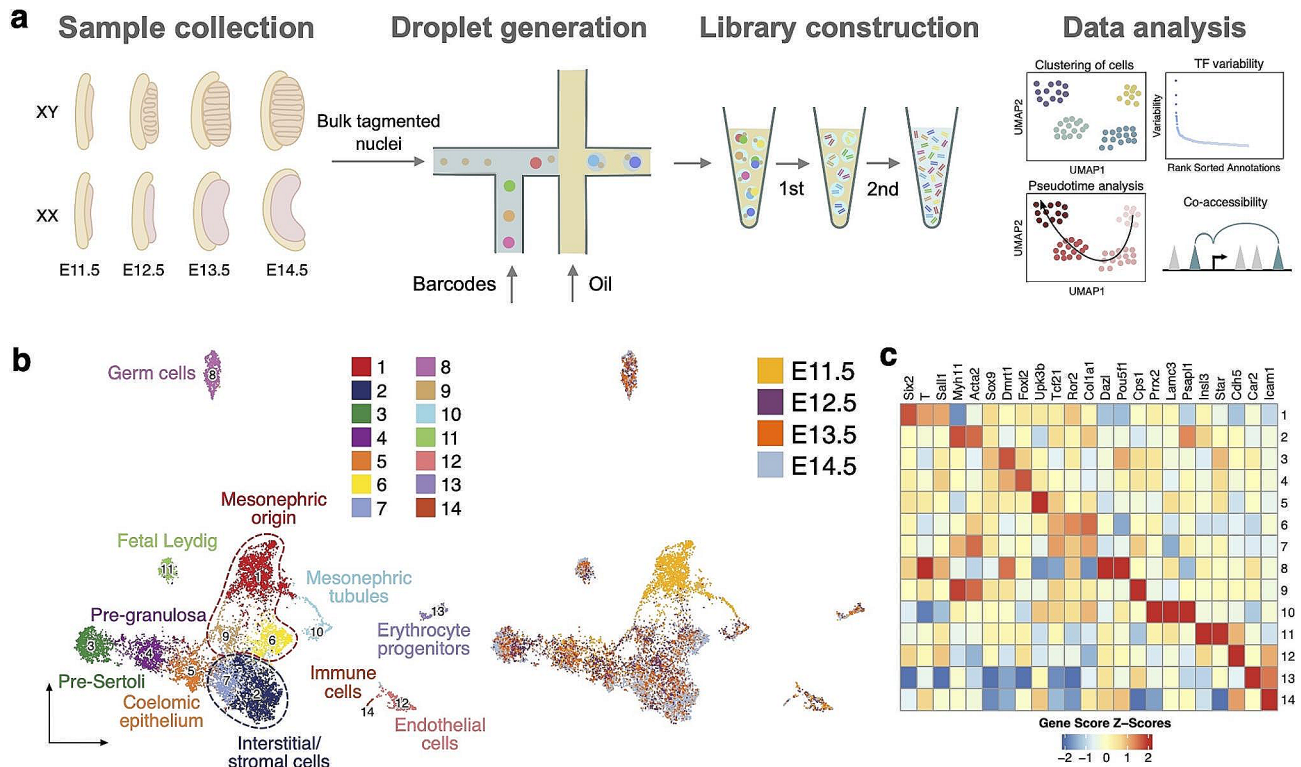


Fig. 1 Single-cell ATAC-seq revealed cellular heterogeneity in the XX and XY gonads. **(a)** The workflow of gonad sample collection and scATAC-seq to measure single nuclei accessibility on BioRad SureCell ATAC-seq platform. **(b)** UMAP representation of all cells captured from all four time points. Cells are colored by annotated cell types (left) and time points (right). **(c)** Heatmap showing the gene score of selected cell type marker genes

peaks to generate a peak set that depicted the entire epigenetic diversity ($n=186,471$ peaks). We discovered 101,160 differential accessibility regions (DARs) using Wilcoxon testing, taking into account the bias of transcription start site (TSS) enrichment and the number of unique fragments per cell ($FDR \leq 0.1$ and $\log_2FC \geq 0.5$) (Fig. 2a, Supplementary Table S2). Next, we associated these DARs with their neighboring genes and performed GO analysis (Fig. 2b). Notably, the Cluster 5 CE progenitor cells showed specific accessible loci related to genes of WNT and BMP signaling pathway. The Cluster 12 endothelial cells showed enrichment of GO terms related to small GTPase mediated signal transduction. The Cluster 4 pre-granulosa cells showed increased accessibility in regions associated with hormone secretion and transport.

Our scATAC-seq data allowed us to probe the specific transcriptional regulatory mechanisms of each cell type. We employed two complementary methods to assess TF motif activity: TF motifs enriched in the cell type-specific DARs (Fig. 2c) and chromVAR deviations (Fig. 2d). chromVAR predicts enrichment of TF activity on a per-cell basis by calculating accessibility deviation across multiple conditions for each TF motif. Using hierarchical clustering of calculated deviations of the top 100 most variable TFs, we identified TFs specific to

particular cell subpopulations (Fig. 2d). Overlapping TF bindings identified by both methods signify confident TFs, including known TFs in gonad development such as TCF21 in interstitial progenitors and NR5A1 in fetal Leydig cells (Fig. 2c and d). Additionally, CE cells exhibited enrichment for binding motifs associated with ARX and LHX9, known regulators of Wnt signaling and gonad development. Top enriched TF motifs in pre-Sertoli cells included members of the ESRR, RAR, and SOX families. Notably, SOX4 has a well-described role in male development [65]. We also observed the enrichment of ZEB1 and SNAI2 in germ cells, as well as ETV2 and ELF1 in endothelial cells (Fig. 2c and d). The DNA regions occupied by TF binding are protected from transposition by Tn5, which can be visualized by plotting the ‘footprint’ pattern of each TF as the local chromatin accessibility surrounding the motif midpoint. Footprint occupancy of selected cell type-dependent TFs validated the robustness of our approach (Fig. 2e).

Chromatin accessibility landscape during somatic cell lineage specification

Although somatic cell commitment has been characterized in depth based on scRNA-seq profiling [18, 61, 63], the underlying chromatin accessibility dynamics remain unclear. Therefore, we tracked the temporal events

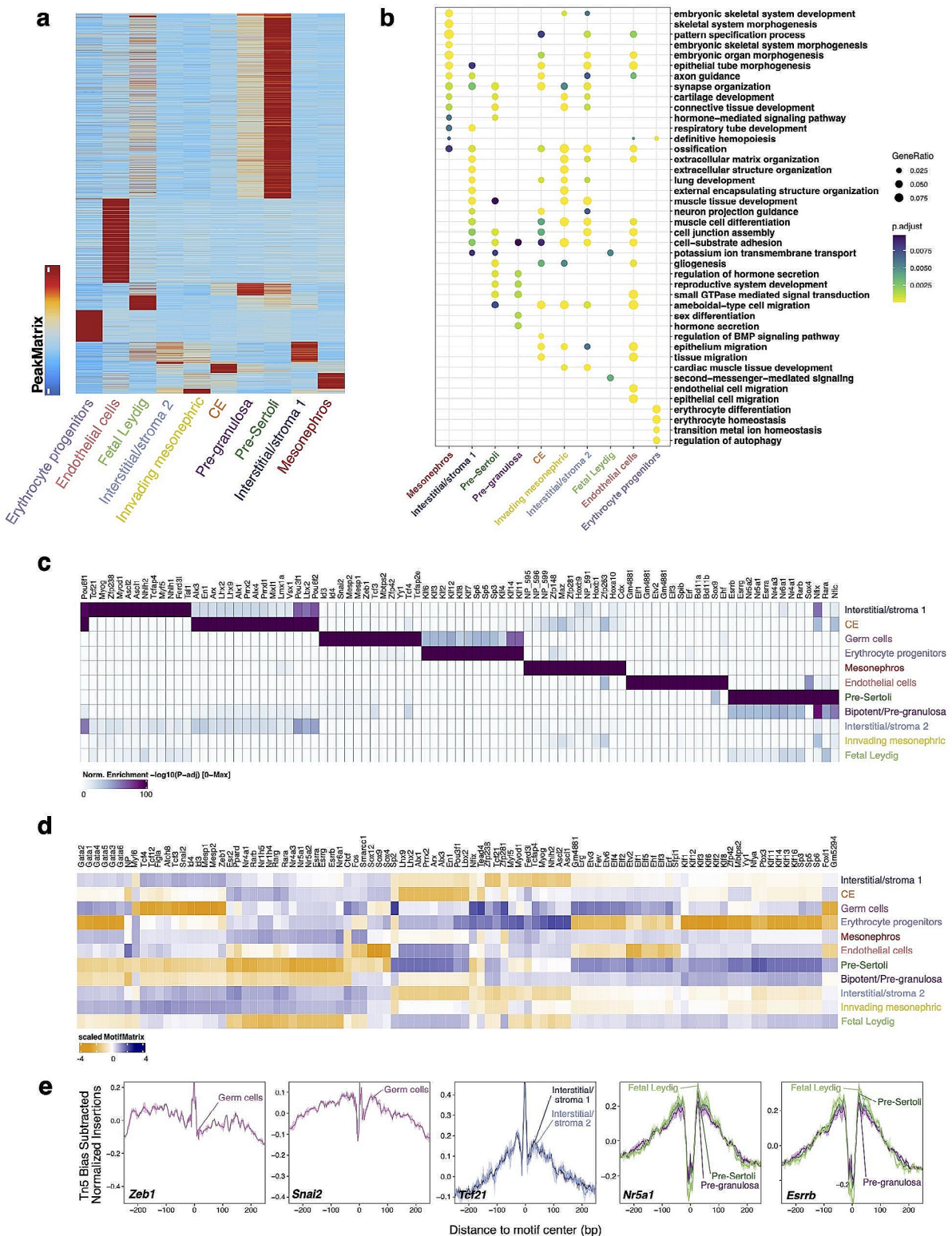


Fig. 2 A high-resolution cis-regulatory atlas of the developing mouse gonad. **(a)** Heatmap showing 101,160 marker peaks across cell types (FDR <= 0.1 & Log2FC >= 0.5). **(b)** Top results from the Gene Ontology (GO) enrichment test showing the terms associated with cell type-specific differential accessibility regions (DARs) by associating them to the neighboring gene. **(c)** Heatmap showing the transcription factor (TF) motifs enriched in cell type-specific DARs as shown in Fig. 2a. **(d)** Heatmap showing the chromVAR deviations of top 100 variable TF motifs in each cell type. **(e)** TF footprints (average ATAC-seq signal around predicted binding sites) for selected TFs

associated with cell fate progression using our scATAC-seq data. We excluded mesonephric cells, germ cells, endothelial cells and cells from the immune system. We also excluded Clusters 6 and 9 due to their possible mesonephric origin. Re-clustering of these cells generated six clusters, namely somatic clusters (SC) 1 to 6 (Fig. 3a).

At E11.5, somatic cells were already heterogeneous at the chromatin level as three distinct developmental populations emerged (SC2, SC3 and SC5) (Fig. 3b). SC5 cells (shown as Cluster 5 in Fig. 1b) were CE-derived progenitor cells (including gonadal surface epithelium) that showed a high gene score of *Wt1* (Figs. 1c and 3c). They were predisposed to develop into supportive cell lineages in SC2 and SC3, and the XY interstitial or XX stromal

cells in SC1 and SC4 at later time points (Fig. 3d). We regarded SC2 cells at E11.5 as progenitors of the “supporting cell lineage”, which was established prior to sex determination (Fig. 3b and d). The absence of any discernible pattern in the scattering of XX and XY cells in UMAP indicated that the chromatin accessibility landscapes in male and female supporting cell progenitors were highly comparable. This similarity may facilitate equitable access to promoters and regulatory elements of sex-determining genes in both sexes, aligning with findings from previous ATAC-seq studies (Garcia-Moreno et al., 2019). SC3 (pre-Sertoli) cells were already apparent by E11.5, indicating that the XY gonad was not fully undifferentiated at this time point, at least at the chromatin

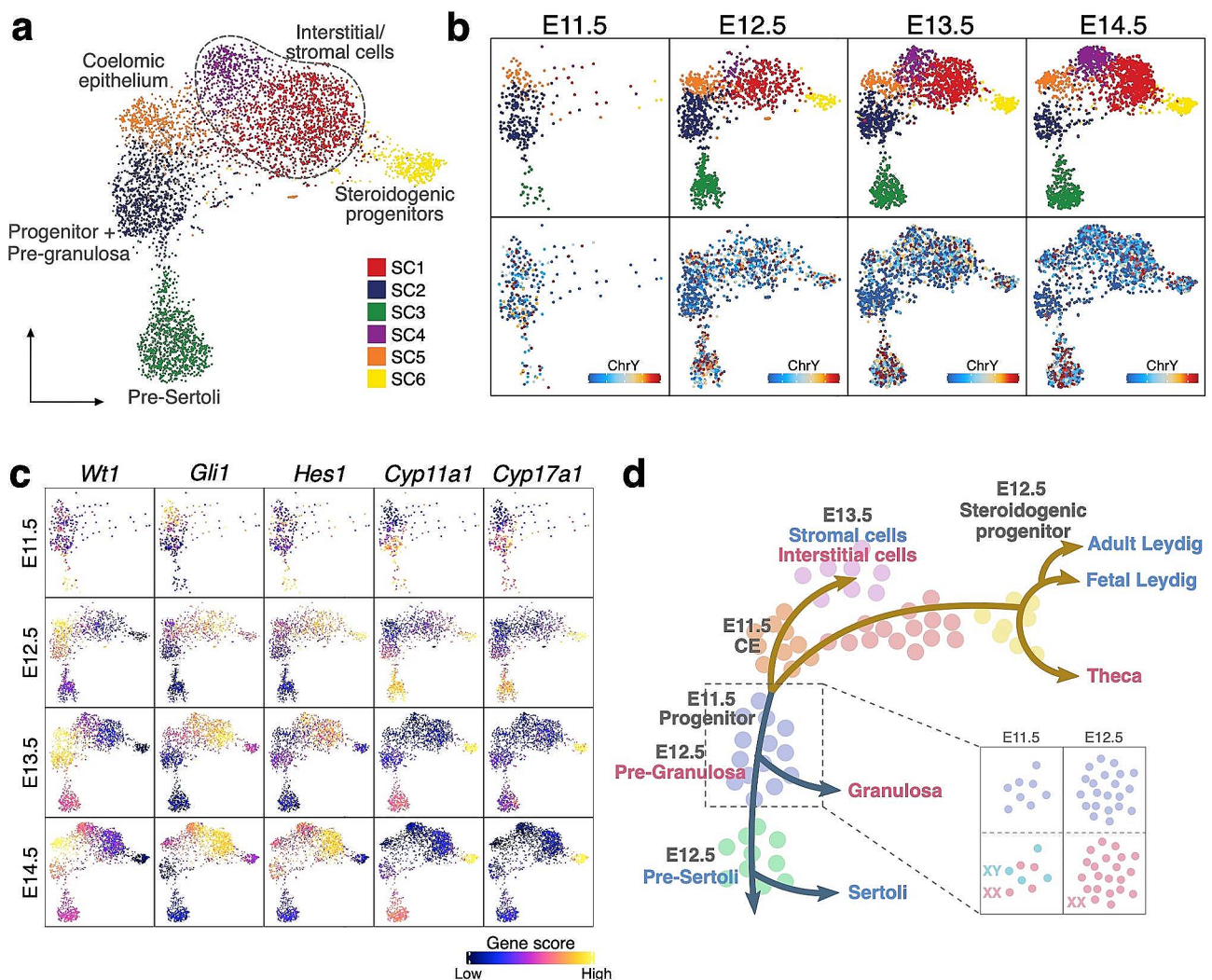


Fig. 3 Chromatin accessibility landscape during somatic cell lineage specification. **(a)** UMAP representation of gonadal somatic cells captured from all four time points. Cells are colored by cell clusters. **(b)** UMAP representation of gonadal somatic cells of the four time points. Cells are colored by cell clusters (top panel) and ChrY score (bottom panel). **(c)** UMAP representations with somatic cells colored by the gene score of specific genes at four time points. **(d)** Model of somatic cell lineage specification during sex determination. All somatic gonadal cells are derived from progenitors at coelomic epithelium (CE) at E11.5. For the supporting cell lineage, pre-Sertoli and pre-granulosa cells, deriving from a group of progenitors, start to emerge at E12.5. For the steroidogenic lineage, steroidogenic progenitors emerge at E12.5 and later differentiate to Leydig or theca cell lineages. Interstitial and stromal cells start to be abundant at around E13.5

level. This also reflected the rapid chromatin accessibility remodeling upon activation of the Sertoli cell fate at this developmental stage. On the contrary, pre-granulosa cells were initially located near the supporting cell progenitor cluster until a slight separation occurred after E13.5, as depicted in the UMAP representation (Fig. 3b). This observation suggests that pre-granulosa cells maintain the progenitor chromatin accessibility landscape for a longer time than pre-Sertoli cells (Fig. 3d). During E11.5 to E12.5, these cell populations expanded, concurrent with the emergence of SC1 and SC6, which were both a mixture of male and female cells based on the ChrY score (Fig. 3b). SC1 represented the interstitial (XY) or stromal (XX) progenitors as they showed increased accessibility at *Gli1* and *Hes1* and loss of accessibility at *Wt1* (Fig. 3c). Some of the cells began to specify their identity as potential steroidogenic cell precursors (SC6), which showed higher gene scores for *Cyp17a1* and *Cyp11a1* (Fig. 3c). Between E12.5 and E13.5, another cell population developed, corresponding to XY interstitial and XX stromal cells (SC4). After E13.5, all cells continued to mature with progressive chromatin accessibility changes.

Integration of scRNA-seq identified cis-regulatory elements of cell type-specific genes and positive TF regulators

To identify cis-regulatory elements (CREs) of transcription that are linked to somatic lineage-specific gene expression, we reanalyzed, annotated and integrated published scRNA-seq data of mouse gonads from E11.5 to E13.5 [58]. We performed label transfer from this reference scRNA-seq dataset to our scATAC-seq data (Supplementary Figure S4a-c). We then linked the distal peaks to genes in cis, based on the coordination of chromatin accessibility information from scATAC-seq and gene expression levels from scRNA-seq (peak-to-gene links). To demonstrate the utility of peak-to-gene linkage analysis in discovering potential enhancers, we examined the regions upstream of the *Sox9* gene. We found that the Enh13 enhancer region, previously identified in mouse and overlapping within the human XYSR enhancer, showed stronger accessibility in pre-Sertoli cells when compared with other clusters [12, 23]. More importantly, Enh13 was also identified as a significant region linked to the *Sox9* gene with a significant correlation (correlation=0.62, FDR=2.09727e-52) (Supplementary Figure S4d). This result highlights the potential of our resource to identify candidate enhancers for genes of interest in the field of reproductive biology.

We then systematically identified 30,561 peak-to-gene (P2G) links (correlation>0.5, FDR<0.01) by correlating ATAC peak accessibility within 250 kb of the gene promoter with the mRNA expression of the gene (Fig. 4a). We performed clustering analysis to identify the cell

type-specific P2G links: Clusters (C) 1 and 2, interstitial/stromal progenitors; C3, CE; C4, pre-Sertoli cells; C5, pre-Sertoli/pre-granulosa cells; and C6, steroidogenic progenitors (Fig. 4b). We validated that the enrichments of gene ontologies (GO) for each P2G cluster matched the functions of the respective cell types (Supplementary Figure S5). It is reassuring that this approach discovered previously reported functionally relevant regulatory regions, including enhancers for *Nr5a1* and *Wt1* (discussed in more detail below) [39, 57]. We also identified novel putative regulatory elements for important cell type-specific regulators, including *Cyp11a1* in steroidogenic progenitors, *Tcf21* in interstitial progenitor cells, and *Wnt6* in pre-Sertoli cells (Supplementary Figure S6). The full list of P2G links in each cluster can be found in (Supplementary Table S3).

Genes often harbor several enhancers in their close proximity, and one or another enhancer or enhancer combination can be used to ensure a proper spatio-temporal expression pattern depending on the cell type or developmental stage [16, 20, 32, 48]. The *Wt1* locus shows at least eight P2G links, and both the intensity and apparent combinatorial usage of these elements vary depending on the cell type (Fig. 4c). Genome browser tracks depicted a combination of predicted enhancers in C3 that were preferentially accessible in CE and supporting cell progenitors/granulosa cells but not detected in pre-Sertoli cells, while C5 putative enhancers were more accessible in pre-Sertoli and pre-granulosa cells. We also observed distal elements that were differentially co-accessible within the *Nr5a1* locus between pre-Sertoli/pre-granulosa cells and steroidogenic progenitors. Careful examination of *Nr5a1* locus revealed steroidogenic progenitor specific peaks (C5 P2G) linked to *Nr5a1*, which overlapped with the previously reported CR8-CR10 fetal Leydig cell enhancer (FLE) (Fig. 4c) [57]. C6 P2G links of *Nr5a1* were more accessible in pre-Sertoli/pre-granulosa cells, suggesting possible cell type-specific enhancer usage.

It is challenging to identify the specific TF binding to the genome and driving gene expression based only on motif enrichment analysis, as TFs from the same family often share a similar motif. Taking the advantage of the integrated data, we identified putative positive TF regulators determined from the correlation between gene expression and the chromVAR motif activity score across single cells, reasoning that the active TF (high motif activity score) should also be actively expressed to maintain its abundance (high mRNA level). Clustering analysis of positive regulators confirmed that the motif binding activity of the identified TFs across cell types closely mirrored their respective gene expression. Reassuringly, this approach also accurately identified known cell type-specific regulators, such as *Sox9* and *Dmrt1* in

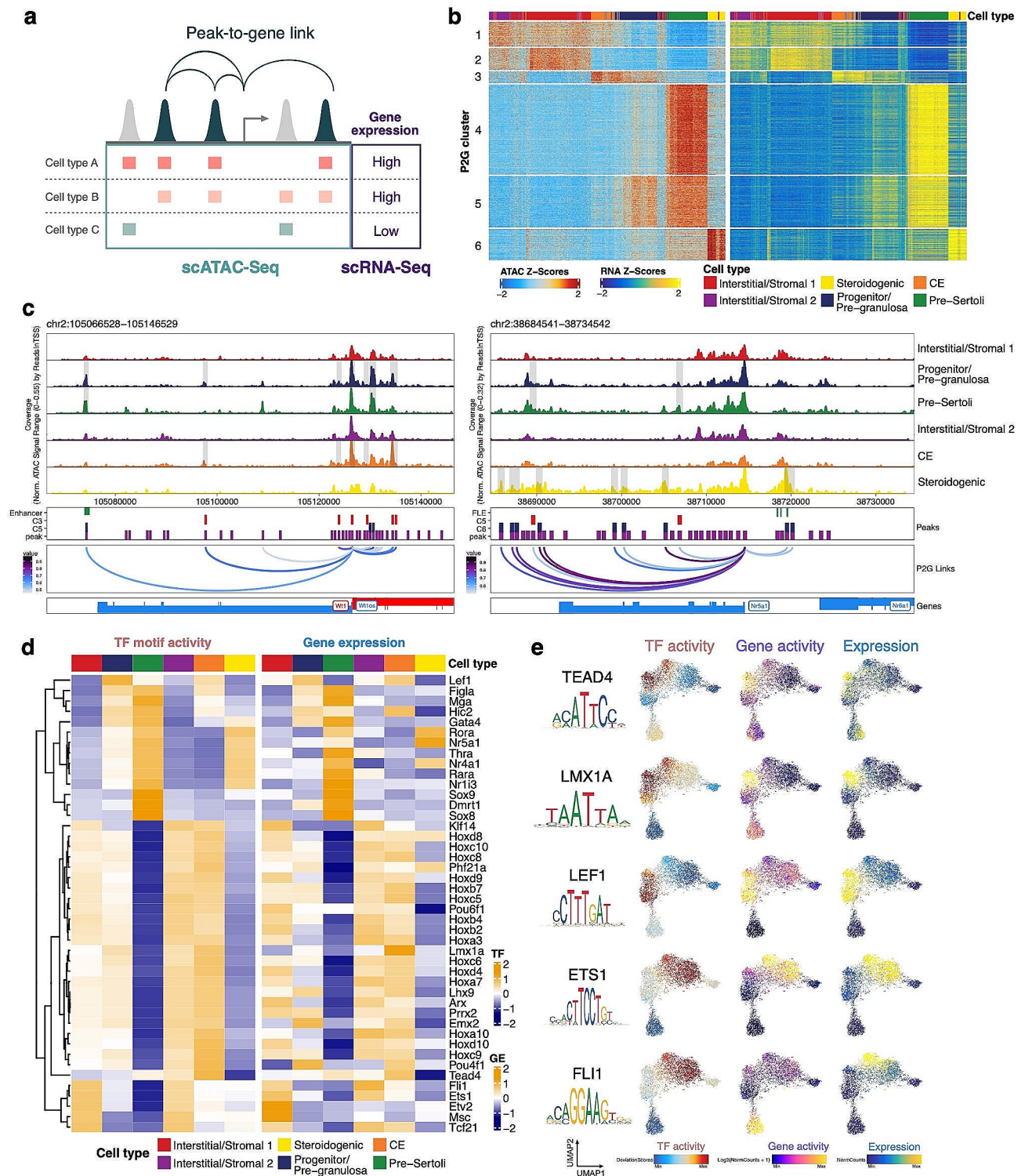


Fig. 4 Integration of scRNA-seq identified cis-regulatory elements of somatic cell type-specific genes and positive transcription factor regulators. **(a)** Schematic for identifying significant peak-to-gene linkages by correlating accessible peaks (scATAC-seq data) to gene expression (integrating scATAC seq data and scRNA-seq data). **(b)** Heatmaps showing transcription factor (TF) motif activity (left) and gene expression (right) of 30,561 peak-to-gene linkages across somatic cell types. **(c)** Aggregated scATAC-seq profiles showing peak-to-gene links to the *Wt1* locus overlapped with known enhancer regions (left) and peak-to-gene links to the *Nr5a1* locus overlapped with known fetal Leydig enhancer (FLE) regions (right). **(d)** Heatmaps showing differential TF motif activity (left) and gene activity (right) of positive TF regulators across somatic cell clusters (correlation > 0.3, adjusted p-value < 0.05). **(e)** UMAP embedding of TF chromVAR deviations (left), gene activity scores (middle) and gene expression (right) for selected TFs

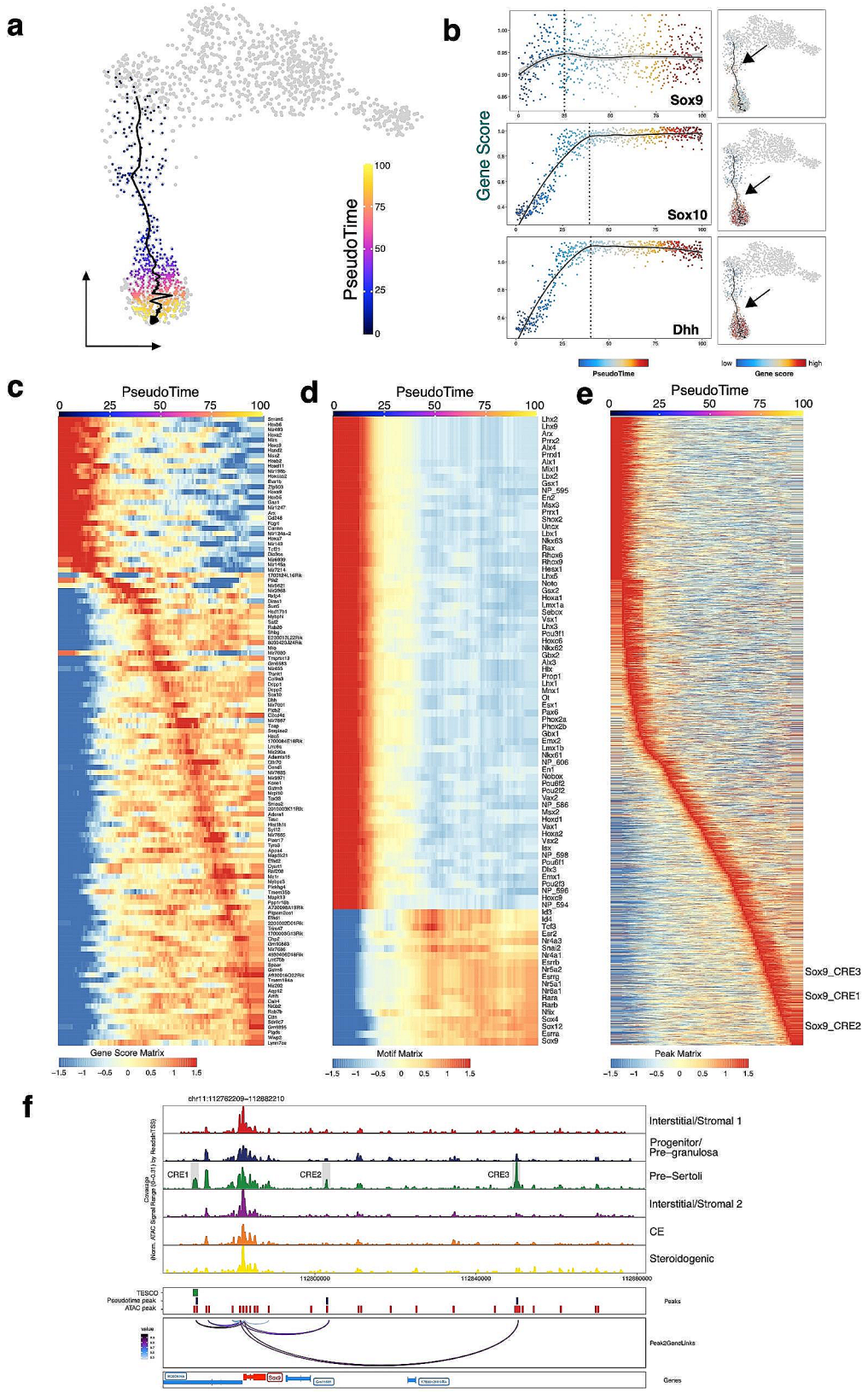


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Chromatin accessibility landscape associated with male supporting progenitor lineage. **(a)** scATAC-seq profiles are ordered by pseudotime, corresponding to the Sertoli cell differentiation trajectory. **(b)** Gene scores of selected genes ordered by pseudotime. The dashed lines and arrows point to the location along pseudotime where the highest gene scores are observed. **(c)** Smoothened heatmap showing dynamic gene score of indicated genes along pseudotime. **(d)** Smoothened heatmap showing dynamic motif accessibility of indicated transcription factors along pseudotime. The heatmap was generated using `plotTrajectoryHeatmap()` function with a `varCutOff` threshold of 0.9. **(e)** Smoothened heatmap showing peaks with differential accessibility along pseudotime. The peaks labeled with `Sox9_CRE1` to 3 correspond to the cis-regulatory element (CRE) regions in Fig. 5e. **(f)** Aggregated scATAC-seq profiles showing peak-to-gene links to the *Sox9* locus overlapped with TESCO and the three potential CREs (CRE1 to 3)

the pre-Sertoli cluster, *Tcf21* in interstitial/stroma clusters, *Gata4* in pre-granulosa and pre-Sertoli cells, and *Nr5a1* in steroidogenic cells (Fig. 4d). Furthermore, it identified novel TF candidates, such as *Tead4* and *Lmx1a* in CE, *Lef1* in CE/granulosa cells, and *Ets1/Fli1* in interstitial/stromal progenitors (Fig. 4e).

Taken together, single-cell multimodal analysis shed new light on the underlying cis-regulatory DNA interactions and the employment of positive TF regulators during gonadal somatic cell type development.

Chromatin accessibility landscape associated with the male supporting progenitor lineage

We next examined the early development of Sertoli cells in more detail, because of the more profound chromatin changes in the male supporting cell lineage. We reconstructed the developmental trajectories by ordering the XY cells according to their identity and time points and then generated a pseudotime trajectory (Fig. 5a). The developmental trajectory started from the E11.5 *Wt1*+early progenitor cells (SC5), passed through E11.5 XY cells (SC2), and transitioned to presumptive pre-Sertoli cells (SC3, E11.5–E14.5). We observed that several Sertoli cell-associated genes changed chromatin accessibility (gene score) along the pseudotime, such as *Sox10* and *Dhh* (Fig. 5b and c). *Sox10* began to be expressed in pre-Sertoli cells shortly after *Sox9* upregulation and *Sox10* gain-of-function causes XX sex reversal [51]. *Dhh* initiated expression shortly after the activation of *Sry* in pre-Sertoli cells and is one of the earliest indications of male sexual differentiation [7]. Consistently, both genes exhibited peak accessibility in the middle of the trajectory, noticeably occurring later than *Sox9*, further highlighting the robustness of pseudotime analysis (Fig. 5b). Interestingly, the scATAC-seq data revealed that numerous miRNA loci are associated with dynamic chromatin accessibility along the trajectory. Since differences in open chromatin correlate with changes in miRNA expression levels [66], these results hinted at miRNA candidates with potential roles in Sertoli cell differentiation, which complements standard single-cell RNA-seq techniques that do not capture miRNAs due to the lack of a polyA tail (Fig. 5c).

We further identified TFs with changes in motif accessibility along the trajectory (Fig. 5d). The motif accessibility landscape can be briefly divided into three stages. Before Sertoli cell specification, the CE-enriched

transcription factors, including *Arx* and *Lhx9*, showed higher motif accessibility. The motif of homeobox protein *Emx2*, which is essential for the formation of genital ridge, is also more enriched [34]. In addition, other homeobox genes, such as *Hoxa1*, *Hoxc6* and *Hoxd1*, are more active before Sertoli cell specification. During the transitional stage, there is an enrichment of *Id3* and *Id4* motifs. The *Drosophila* homolog of the ID proteins, extramacrochaetae (*emc*), is involved in sex determination [59]. *Id3* and *Id4* may have similar regulatory roles in the development of mouse gonads. Moreover, *Tcf3*, *Esr2*, and *Snai2* also showed an increase of motif activity in the middle of the trajectory, suggesting that this may contribute to Sertoli cell fate specification. In accordance with this observation, *Snai2* mutant mice displayed gonadal defects [50]. At late stage, the retinoic acid (RA) receptors *Rara* and *Rarb* showed increased motif accessibility, consistent with RA's role to induce Sertoli cell proliferation [47]. *Sox9* also showed increased motif activity at late stage, concordant with the role of SOX9 in the maintenance of the Sertoli cell fate [4]. We further explored the regulation of the differentiation process by correlating ChromVAR TF motif activity with integrated TF mRNA expression along this pseudotime trajectory. This analysis allowed us to prioritize positive regulators of the differentiation process beyond TF motif activity alone (Supplementary Figure S7).

We then tracked the individual peaks that showed accessibility changes along the pseudotime of Sertoli cell differentiation (correlation>0.95) (Fig. 5e, Supplementary Table S5). During male sex determination, SRY activates male-specific transcription of *Sox9* and the *Sox9* mRNA expression becomes detectable in XY pre-Sertoli cells at E11.5 [44, 55]. We readily recovered a region 13 kb upstream of the *Sox9* TSS change along the pseudotime, which overlapped with the testis-specific enhancer of *Sox9* (TES), including the core element of TES, TESCO (Fig. 5e). Previous research revealed that the TES accounted for only approximately half of the normal levels of *Sox9* expression in the mouse testis, and that one or more other *Sox9* enhancers remain to be discovered [22]. We identified two additional regions located 21 kb 3' and 68 kb 3' of *Sox9*. Importantly, these two regions were identified as peaks linking *Sox9* mRNA expression in the P2G analysis (Fig. 5f), which represent potential regulatory elements involved in both *Sox9* gene regulation and Sertoli cell differentiation. Furthermore,

we focused on the 682 distal intergenic regions to better understand the cis-regulatory mechanisms that underlie gene expression dynamics (Supplementary Table S4). ToppGene suite analysis of the associated genes revealed that they were enriched in the Notch pathway (e.g., *Yy1* and *Notch1*) and the Hippo pathway (e.g., *Smad3* and *Smad7*) (Supplementary Figure S8).

Taken together, we have demonstrated that the development of Sertoli cells involves drastic and distinct chromatin remodeling together with unique TF regulation.

TF candidates in early lineage specification of steroidogenic populations

The steroidogenic progenitor cluster (SC6) emerged at E12.5, together with the interstitial somatic cell cluster (SC1) (Fig. 3b). The mix of both XX and XY cells in these clusters indicates that the chromatin accessibility landscapes of XY interstitial and XX stromal progenitors are similar and the resemblance is maintained after both XY and XX progenitor cells acquire a steroidogenic precursor fate (Fig. 3b). Steroidogenic progenitors can originate from the cells in the interstitial compartment derived from the coelomic epithelium, as well as from the mesonephros [1, 15, 17]. To explore the lineage specification of steroidogenic cells, we conducted pseudotime analysis focusing on SC5, SC1, and SC6 at E12.5, corresponding to the progression from the CE to interstitial cells and finally to steroidogenic progenitors (Fig. 6a). We observed elevated *Tcf21* TF motif activity when cells transitioned to SC6, suggesting *Tcf21*'s role in interstitial/stromal progenitor specification (Fig. 6b and c). Various studies have shown TCF21's role in male sex determination and testis somatic cell differentiation [6, 13]. We identified higher activity of *Nr5a1* in the late stage of the trajectory, consistent with its role in steroidogenic cell differentiation [35]. Novel candidates include *Tcfap4* and the retinoic acid receptors (*Rara*, *Rarb* and *Rarg*), while the nuclear estrogen receptor *Esr2* and Peroxisome Proliferator-activated Receptor-D (*Ppard*) also merit further investigation (Fig. 6b and c). Through our analysis of positive regulators along pseudotime, we were able to further prioritize potential candidates, including *Meis3* and *Nr3c1*, as potential drivers of the differentiation process (Supplementary Figure S9).

Distinct TF dynamics and chromatin regulation during male and female germ cell development

We then investigated the development of the female and male germ cells. The germ cells from E11.5 to E14.5 were re-clustered into five cell clusters, namely GC1 to GC5 (Fig. 7a, Supplementary Figure S10a). The cells were branched from GC5, indicating that sex specification at the chromatin level of germ cells begins at around E11.5 and E12.5. By inferring the chrY scores, GC2 and GC3

were the male germ cells while GC1 and GC4 were the female germ cells (Supplementary Figure S10b). We then reconstructed the pseudotime developmental trajectories by ordering the female and male clusters following developmental time points (Fig. 7b). We also integrated our scATAC-seq data with previous scRNA-seq data [41]. To determine the TFs driving germ cell development in males and females, we first identified a list of genes with dynamic expression and a list of TF motifs with differential binding activity throughout the two trajectories. We then correlated the gene expression (inferred by scRNA-seq) and TF binding activity (chromaVAR score inferred by scATAC-seq) to identify a list of high-confidence TFs with dynamic regulation along the trajectories (Fig. 7c and d).

Comparing the heatmap side-by-side reveals a distinct pattern of TF reconfiguration in female and male germ cell differentiation. In females, there is a two-step process involving the down-regulation of pluripotency markers (*Oct4* and *Lef1*) among other TFs (e.g., *Rhox6* and *Rhox9*), followed by the upregulation of another set of TFs, including *Yy1* and *Stat3*. *Stat3* has been shown to be differentially distributed in the cytoplasm of mature oocytes [46]. In contrast, male germ cell differentiation appears to involve a more complex, multi-step process. Male germ cell differentiation also involved down-regulation of pluripotency markers (*Oct4* and *Lef1*) but followed by a transient activation of TFs including *Id4* and *Six4*, before later stage upregulation of TFs such as *Yy1* and *Foxo4*. *Foxo4* is essential for spermatogonial stem cell self-renewal [21].

Utilizing the aforementioned analytical framework, we identified the P2G links of germ cells (Supplementary Table S5). A total of 23,004 germ cell-specific P2G links were clustered into seven clusters and GO analysis was performed (Fig. 7e, Supplementary Figure S10c). As expected, the GO terms related to meiosis were enriched in the genes of the P2G links of Clusters 1 and 2, which were mainly upregulated in E14.5 female germ cells. For the male germ cells, the GO terms associated with the P2G links of Clusters 3 to 6 included RNA processing, cell cycle and ribosome biogenesis, which is consistent with the fact that male germ cells are undergoing active cell division during this period. Our analysis further revealed putative regulatory elements specific to male (*Msx1* and *Id1*) and female (*Lhx1* and *Wnt3*) germ cells (Fig. 7f).

Discussion

In addition to the transcriptome, the epigenome is another element crucial to understanding the mechanisms underlying many developmental processes, including gonad development [29]. To date, our knowledge about the epigenetic regulation of gonadal development

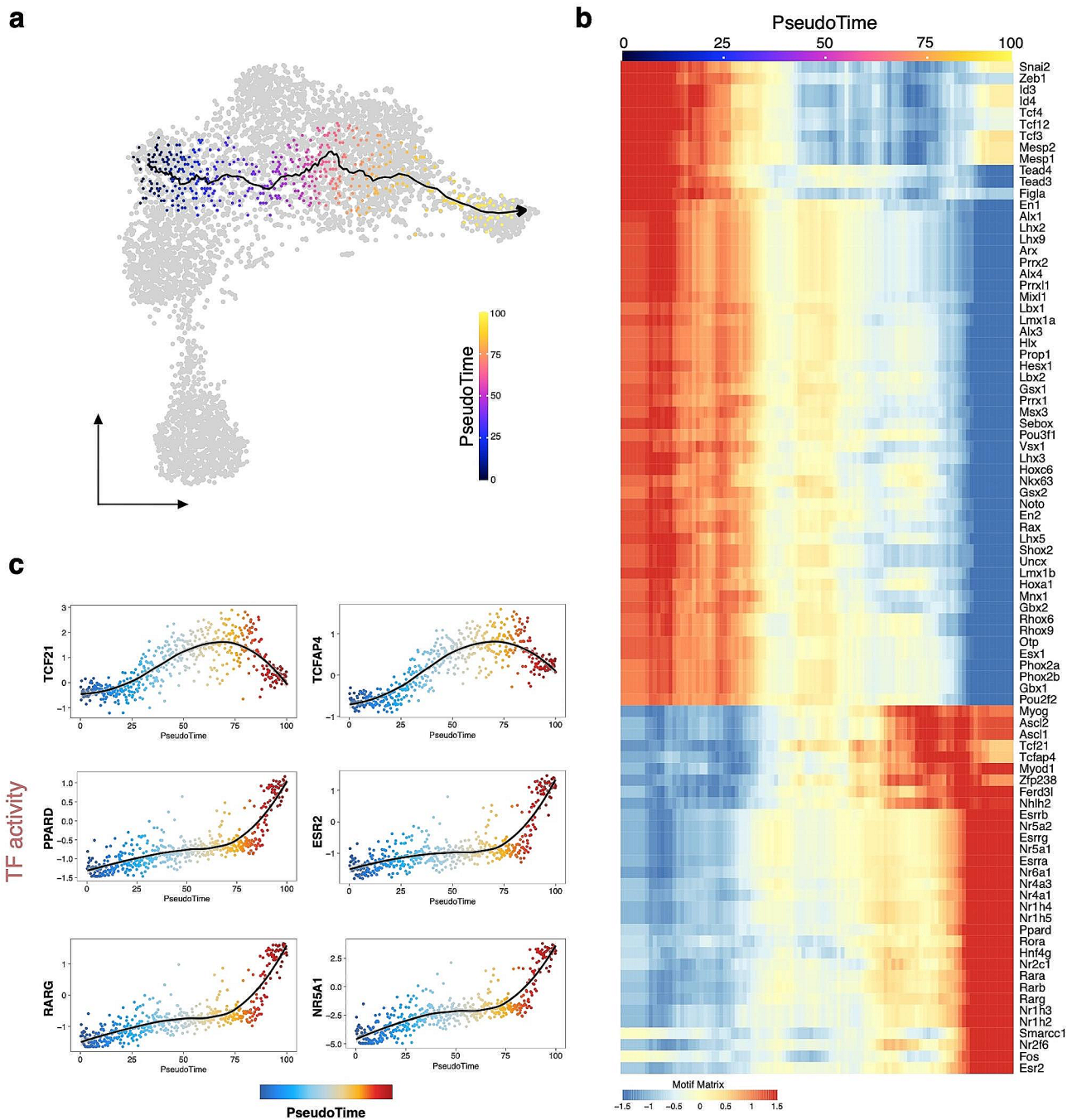


Fig. 6 Transcription factor candidates in early lineage specification of steroidogenic populations. **(a)** scATAC-seq profiles are ordered by pseudotime, corresponding to the steroidogenic lineage specification trajectory. **(b)** Smoothed heatmap showing dynamic motif accessibility of indicated transcription factors (TFs) along pseudotime. **(c)** Motif accessibility of selected TFs ordered by pseudotime

at the single-cell level is still limited, especially the intersection between transcriptome and chromatin accessibility. In this study, we conducted an in-depth analysis of the temporal dynamics of the chromatin accessibility landscape at single cell resolution for mouse embryonic gonads.

scATAC-seq analysis offers several insights into the epigenomic states underlying cellular transition during

gonadogenesis. First, chromatin accessibility landscapes of XX and XY supporting cell progenitors prior sex differentiation at E11.5 are similar. Hence, the specification of supporting cell lineages is independent of sex. Secondly, the chromatin reconfiguration during Sertoli and granulosa cell lineage specification shows a temporal asymmetry since pre-Sertoli cells have already appeared in E11.5 bipotent gonads. Third, the shared chromatin

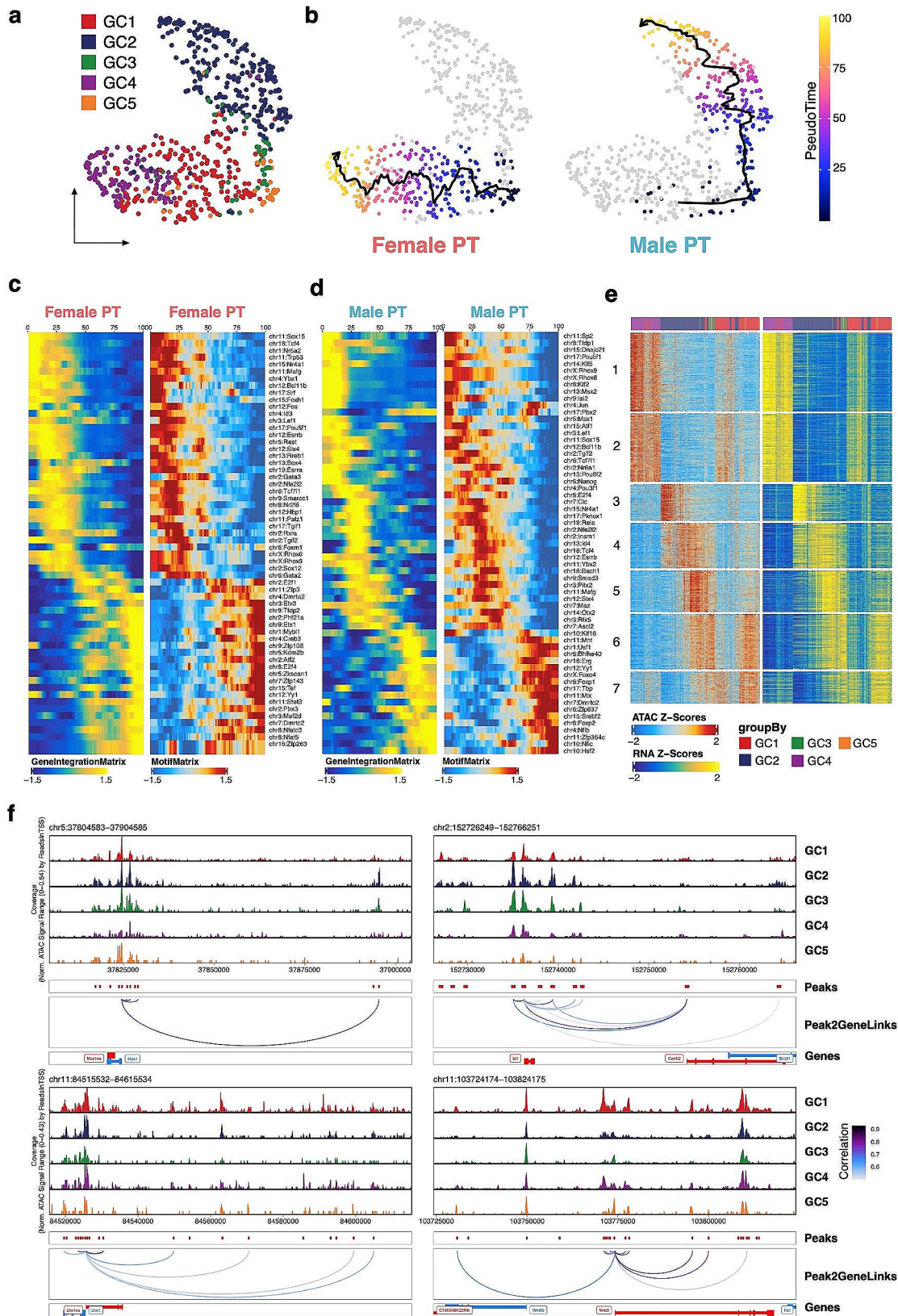


Fig. 7 (See legend on next page.)

(See figure on previous page.)

Fig. 7 Transient patterns of transcription factor expression and regulatory elements during germ cell sex specification. **(a)** UMAP representation of all female and male germ cells. Cells are colored by clustering based on integration with scRNA-seq data. **(b)** scATAC-seq profiles are ordered by pseudotime, corresponding to the female (left) and male (right) germ cell development trajectory. **(c)** Heatmap showing dynamic gene expression (left) and transcription factor (TF) motif activity (right) of indicated TFs along pseudotime for gene-motif pairs of the female germ cell trajectory. **(d)** Smoothed heatmap showing dynamic gene expression (left) and TF motif accessibility (right) of indicated TFs along pseudotime for gene-motif pairs of the male germ cell trajectory. **(e)** Heatmaps showing TF motif activity (left) and gene expression (right) of 21,710 peak-to-gene links across germ cell clusters (corCutOff = 0.6). **(f)** Aggregated scATAC-seq profiles showing peak-to-gene links to the *Msx1* (upper left), *Id1* (upper right), *Lhx1* (lower left) and *Wnt3* (lower right) loci

accessibility landscape between the pre-granulosa cell state (E12.5 to E14.5) and the progenitor state (E11.5), as opposed to Sertoli cells, suggesting the supporting cell progenitors exhibit a bias toward a female chromatin state before subsequently being diverted in XX or XY gonads. Fourth, both sexes have similar patterns of chromatin accessibility during interstitial or stromal lineage commitment and differentiation.

Low-abundance essential genes, including many TFs, are particularly affected by the limitations of scRNA-seq such as dropouts (false zero expression estimates) and high variability [36]. This means that the biological information conveyed by lowly expressed TFs is lost. scATAC-seq allows the detection of TF activity in single cells, which significantly increases the ability to search for transcriptional regulators. Through integrated analysis of both chromatin accessibility and gene expression, our study provided a holistic view of temporal TF regulation during Sertoli and steroidogenic cell differentiation, which serves as a fundamental base to reveal the mechanisms involved in gonadogenesis. Examples include *Tcf3*, *Esr2* and *Snai2* for Sertoli differentiation, and *Tcf21* and *Tcfap4* for steroidogenic differentiation. Hopefully, these TF candidates will be useful for improving the in vitro induced system for Sertoli and steroidogenic cells, which may provide potential treatments for male infertility in the future.

Integrating previous work profiling the transcriptome of gonad development, our multiomics analysis not only allowed us to prioritize TFs that are both expressed and active but also to identify the probable CREs of critical cell type-specific genes based on the P2G linkages. More importantly, our study highlights the benefit of scATAC-seq analysis in identifying cell type-specific enhancers, exemplified by the re-discovery of the Enh13 enhancer of *Sox9*. Complex spatially and temporally defined expression patterns of gonadal genes have been observed in development. For instance, *Nr5a1* expression starts at the progenitor cell stage and reaches a plateau in matured steroidogenic cells and supporting cells. *Wt1* is also expressed in different gonadal cell types including the CE, supporting cell progenitors, Sertoli cells and granulosa cells [49, 54]. The multiple cell type-specific CREs of *Nr5a1* and *Wt1* that have been identified may be responsible for this complex expression pattern, each acting independently and controlling transcription in different cell types at different developmental stages. Researchers

interested in how gene regulation is mediated in the developing gonad can benefit immediately from the identification and categorization of these candidate CREs across cell types.

Taken together, our findings establish a groundwork for future clarification of mechanisms in mammalian gonad development, as well as a better knowledge of the genetic basis of human sex development disorders and infertility.

Limitations of the study

The relatively small number of cells included in our study limited our ability to explore only major cell types in the gonad. To capture the full complexity of gonad development, larger-scale studies or pre-enrichment methods followed by scATAC-seq are necessary. The mix of male and female gonads during sampling, along with homologous genes on X and Y chromosomes and potential alignment errors, introduces challenges in accurately assigning sex to cells based on sex chromosome reads bioinformatically. As a result, while we have explored certain aspects of sex-specific processes, a complete understanding of the intricacies of sex determination requires further investigation. Our study focuses on chromatin accessibility, one of the many attributes of chromatin status. We expect that the advent of new single-cell sequencing methods for the study of DNA and histone modification, combined with single-cell transcriptome/chromatin accessibility data, will be instrumental in advancing our understanding of the epigenetic regulation at play in each somatic cell lineage during mammalian gonad development. Performing scRNA-seq and scATAC-seq in the same cells will further enhance the accuracy of linking the chromatin landscape with gene expression. Furthermore, analysis of human gonad development using the same framework is critical to validate our results on mice. Finally, we recognize that further validation and functional studies will be necessary to confirm the regulatory roles of the identified TFs and cis-regulatory elements. We believe that these limitations provide opportunities for future research to build upon the findings of our study and to further enhance our understanding of the complex regulatory networks that govern gonad development.

Methods

scATAC-seq analysis

Sample collection

Mouse strain C57BL/6J was purchased from the Jackson Laboratory. Pregnant females were euthanized by cervical dislocation at embryonic days: E11.5, E12.5, E13.5 and E14.5, which the day with the presence of vaginal plugs was defined as E0.5. For the embryos at E11.5, whole genital ridges (including gonad and mesonephros) were dissected. For the embryos at E12.5, E13.5 and E14.5, gonads were dissected free of the mesonephros. The sexing step was omitted to reduce the sample processing time, and the gonads from both sexes were pooled in equal numbers for subsequent steps. The XX and XY gonads at the same time point were pooled together. Each time point contained at least 6 embryos.

Cell lysis and tagmentation

Cell tagmentation was performed according to SureCell ATAC-Seq Library Prep Kit (17,004,620, Bio-Rad) User Guide (10,000,106,678, Bio-Rad) and the protocol based on Omni-ATAC was followed [10]. In brief, washed and pelleted cells were lysed with the Omni-ATAC lysis buffer containing 0.1% NP-40, 0.1% Tween-20, 0.01% digitonin, 10 mM NaCl, 3 mM MgCl₂ and 10 mM Tris-HCl pH 7.4 for 3 min on ice. The lysis buffer was diluted with ATAC-Tween buffer that contains 0.1% Tween-20. Nuclei were counted and examined under microscope to ensure successful isolation. Nuclei were resuspended in tagmentation mix, buffered with 1×PBS supplemented with 0.1% BSA and agitated on a ThermoMixer for 30 min at 37 °C. Equal number of nuclei and equal ratio of cells/Tn5 transposase were used to minimize potential batch effect. Tagmented nuclei were kept on ice before encapsulation.

scATAC-seq library preparation and sequencing

Tagmented nuclei were loaded onto a ddSEQ Single-Cell Isolator (Bio-Rad). scATAC-seq libraries were prepared using the SureCell ATAC-Seq Library Prep Kit (17,004,620, Bio-Rad) and SureCell ATAC-Seq Index Kit (12,009,360, Bio-Rad). Bead barcoding and sample indexing were performed with PCR amplification as follows: 37 °C for 30 min, 85 °C for 10 min, 72 °C for 5 min, 98 °C for 30 s, eight cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 60 s, and a single 72 °C extension for 5 min to finish. Emulsions were disrupted and products were cleaned up using Ampure XP beads. Barcoded amplicons were further amplified for 8 cycles. PCR products were purified using Ampure XP beads and quantified on an Agilent Bioanalyzer (G2939BA, Agilent) using the High-Sensitivity DNA kit (5067–4626, Agilent). Libraries were sequenced on HiSeq 2000 with 150 bp paired-end reads.

Sequencing reads preprocessing

The Bio-Rad ATAC-seq Analysis Toolkit was used to process the sequencing data. This streamlined computational pipeline includes tools for FASTQ debarcoding, read trimming, alignment, bead filtration, bead deconvolution, cell filtration, and peak calling. The reference index was built upon the mouse genome mm10. First, barcodes were parsed out of the R1 reads of the FASTQ files and appended to the read name for all reads with valid barcodes. Reads failed debarcoding were discarded. The alignment step aligned debarcoded reads to the reference genome using BWA (<http://bio-bwa.sourceforge.net/>). Reads with low mapping quality, those not properly paired, and those mapped to the mitochondrial genome were removed. The BioRad SureCell scATAC-seq kit used in this study allows multiple beads per droplet, necessitating barcode merging. Bead-based ATAC-seq processing (BAP) was employed (<https://github.com/caleblareau/bap>) to identify systematic biases (i.e., reads aligning to an inordinately large number of barcodes), deduplicate reads and perform merging of multiple bead barcode instances associated with the same cell (“bead deconvolution”). For generation of the fragments file, which contain the start and end genomic coordinates of all aligned sequenced fragments, sorted bam files were further processed with “bap-frag” module of BAP. Downstream analysis was performed with SnapATAC [19] and ArchR [24]. Sorted bam file of each sample was structured into hdf5 snapfiles (single-nucleus accessibility profiles) using snaptools version 1.4.1. Cell-by-bin matrices were created with a bin size of 5,000 bp. SnapATAC package was used to merge snapfiles based on the same bin coordinates. Peak coordinates obtained from Bio-Rad ATAC-seq Analysis Toolkit were loaded on snapfiles to generate a cell-by-peak matrix. Fragment files were used to create the Arrow file in the ArchR package.

Clustering analysis

We filtered out low-quality nuclei with stringent selection criteria, including read depth per cell (>2,000) and TSS enrichment score (>4). Potential doubles were further removed based on the ArchR method. Bin regions were cleaned by eliminating bins overlapping with ENCODE Blacklist regions, mitochondrial DNA as well as the top 5% of invariant features (house-keeping gene promoters). Dimensionality reduction was performed using a nonlinear diffusion map algorithm available in the SnapATAC 1.0.0 package to produce 50 eigen-vectors of which the first 10 were selected in order to generate K Nearest Neighbor graph with K=15. The clustering was performed using Leiden Algorithm available in the package leidenalg version 0.7.0 with a resolution of 0.8 and UMAP embedding were generated using umap-learn.

Gene score and transcription factor activity estimation

ArchR was used to estimate gene expression for genes and TF motif activity from single cell chromatin accessibility data. Gene scores were calculated using the `addGeneScoreMatrix` function with gene score models implemented in ArchR. `addDeviationsMatrix` function was used to compute enrichment of TF activity on a per-cell basis across all motif annotations based on chromVAR (chromVAR_0.3). The chrY scores were calculated based on the number of reads mapped to the Y chromosome, normalized to sequencing depth per 10 K reads. Cells were designated as male if ChrY reads were present (chrY score > 0).

Marker gene and DAR analysis

To identify marker genes based on gene scores or marker peaks, we used the `getMarkerFeatures` function in ArchR with `useMatrix = "GeneScoreMatrix"` or `"PeakMatrix"`, respectively. ArchR performed Wilcoxon testing for comparison after accounting for the biases (i.e., "TSSE enrichment" score and `log10(nFragments)`). These p-values are then adjusted for multiple hypothesis using the FDR method. Cell type-specific DARs were visualized as a heatmap using the `markerHeatmap()` function. Motif enrichment on marker peaks were identified using `getMarkerFeatures()`. Motif footprinting was generated using the `getFootprints()` function. DARs were annotated using ChIPseeker and the GO term enrichment of DARs associated genes were performed using clusterProfiler.

Integration with scRNA-seq data

The integration of scATAC-seq with published time-series scRNA-seq data of all gonadal cells (GEO: GSE184708), somatic cell subsets in the gonads (DNA Data Bank of Japan: DRA011172) and germ cells (GEO: GSE136220) were carried out using ArchR [41, 58]. A new `GeneIntegrationMatrix` corresponding to the predicted linked gene expression from scRNA-seq was added using `addGeneIntegrationMatrix()`. Putative cis-regulatory elements were identified based on correlated peak accessibility and linked gene expression (peak-to-gene linkages) across different cell types using `addPeak2GeneLinks()`.

Trajectory analysis

Trajectory analysis was performed in ArchR. `addTrajectory()` function in ArchR was used to construct trajectory on the UMAP embedding. Pseudo-time versus the gene score, peak accessibility and Motif activity were visualized using `plotTrajectory()` or `plotTrajectoryHeatmap()` function. To identify positive TF regulators by integration of gene scores with motif accessibility across pseudo-time, we used the `correlateTrajectories()` function which

takes two SummarizedExperiment objects retrieved from the `getTrajectories()` function.

Statistical analysis

Assessment of statistical significance was performed using two-tailed unpaired t-tests, one-way ANOVA with Tukey multiple comparisons tests or Chi-squared tests. Statistical analysis was performed using GraphPad Prism v8. Associated P values are indicated as follows: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; not significant (ns) $P > 0.05$.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10376-1>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Author contributions

HCS and JL conceived and designed the study. HCS, FO and JL performed experiments and wrote the manuscript. HCS and JL performed single-cell sequencing analysis and interpreted data. HCS, FO, KM, ZW, WYC and JL reviewed and edited the manuscript.

Funding

JL was supported by the Chinese University of Hong Kong, Department of Chemical Pathology's Faculty Startup Fund. This work was supported by the Chinese University of Hong Kong (CUHK) – Shandong University (SDU) Joint Laboratory on Reproductive Genetics, the CUHK Vice Chancellor Discretionary Fund to support the Hong Kong Branch of CAS Center for Excellence in Animal Evolution and Genetics (8601011), State Ministries Special Budget to support MOE Key Laboratory for Regenerative Medicine (CUHK-Jinan University) Project Code 2622009 provided to WYC and Department of Chemical Pathology (CUHK)'s Faculty Startup Fund provided to JL.

Data availability

Raw data, ATAC-seq fragment files, and associated metadata from this study have been deposited in the NCBI gene expression omnibus (GEO) under accession number GSE218995. Code for producing the majority of analysis from this paper is available on GitHub at <https://github.com/liaojinyue/mouse-gonad-scATAC-seq>.

Declarations**Ethics approval and consent to participate**

This study was carried out in compliance with the ARRIVE guidelines 2.0 (<https://arriveguidelines.org/>). All the animal experiments were performed according to the protocols approved by the Animal Experiment Ethics Committee (AEEC) of The Chinese University of Hong Kong (CUHK) and followed the Animals (Control of Experiments) Ordinance (Cap. 340) licensed from the Department of Health, the Government of Hong Kong Special Administrative Region. C57BL/6J mice were maintained in CUHK Laboratory Animal Services Center. All the mice were housed under a cycle of 12-hour light/dark and kept in ad libitum feeding and controlled temperature of 22–24 °C.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 31 October 2023 / Accepted: 3 May 2024

Published online: 13 May 2024

References

- Ademi H, Djari C, Mayère C, Neirijnck Y, Sararols P, Rands CM, Stévant I, Conne B, Nef S. Deciphering the origins and fates of steroidogenic lineages in the mouse testis. *Cell Rep.* 2022;39:110935. <https://doi.org/10.1016/j.celrep.2022.110935>.
- Anamthathmakula P, Miryala CSJ, Moreci RS, Kyathanahalli C, Hassan SS, Condon JC, Jeyasuria P. Steroidogenic factor 1 (nr5a1) is required for sertoli cell survival post sex determination. *Sci Rep.* 2019;9:4452. <https://doi.org/10.1038/s41598-019-41051-1>.
- Barrionuevo F, Bagheri-Fam S, Klattig J, Kist R, Taketo MM, Englert C, Scherer G. Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. *Biol Reprod.* 2006;74:195–201. <https://doi.org/10.1095/biolreprod.105.045930>.
- Barrionuevo FJ, Hurtado A, Kim G-J, Real FM, Bakkali M, Kopp JL, Sander M, Scherer G, Burgos M, Jiménez R. Sox9 and Sox8 protect the adult testis from male-to-female genetic reprogramming and complete degeneration. *eLife.* 2016;5. <https://doi.org/10.7554/eLife.15635>.
- Barsoum IB, Yao HH-C. Fetal Leydig cells: progenitor cell maintenance and differentiation. *J Androl.* 2010;31:11–5. <https://doi.org/10.2164/jandrol.109.008318>.
- Bhandari RK, Sadler-Riggleman I, Clement TM, Skinner MK. Basic helix-loop-helix transcription factor TCF21 is a downstream target of the male sex determining gene SRY. *PLoS ONE.* 2011;6:e19935. <https://doi.org/10.1371/journal.pone.0019935>.
- Bitgood MJ, Shen L, McMahon AP. Sertoli cell signaling by Desert Hedgehog regulates the male germline. *Curr Biol.* 1996;6:298–304. [https://doi.org/10.1016/s0960-9822\(02\)00480-3](https://doi.org/10.1016/s0960-9822(02)00480-3).
- Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod.* 2016;31:2841–55. <https://doi.org/10.1093/humrep/dew218>.
- Brennan J, Tilmann C, Capel B. Pdgfr-alpha mediates testis cord organization and fetal Leydig cell development in the XY gonad. *Genes Dev.* 2003;17:800–10. <https://doi.org/10.1101/gad.1052503>.
- Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods.* 2017;14:959–62. <https://doi.org/10.1038/nmeth.4396>.
- Corces MR, Shcherbina A, Kundu S, Gloudemans MJ, Fréard L, Granja JM, Louie BH, Eulalio T, Shams S, Bagdatli ST, et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet.* 2020;52:1158–68. <https://doi.org/10.1038/s41588-020-00721-x>.
- Croft B, Ohnesorg T, Hewitt J, Bowles J, Quinn A, Tan J, Corbin V, Pelosi E, van den Bergen J, Sreenivasan R, et al. Human sex reversal is caused by duplication or deletion of core enhancers upstream of SOX9. *Nat Commun.* 2018;9:5319. <https://doi.org/10.1038/s41467-018-07784-9>.
- Cui S, Ross A, Stallings N, Parker KL, Capel B, Quaggin SE. Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice. *Development.* 2004;131:4095–105. <https://doi.org/10.1242/dev.01266>.
- Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filipova GN, Huang X, Christiansen L, DeWitt WS, et al. A single-cell atlas of *Vivo* Mammalian Chromatin Accessibility. *Cell.* 2018;174:1309–e132418. <https://doi.org/10.1016/j.cell.2018.06.052>.
- DeFalco T, Takahashi S, Capel B. Two distinct origins for Leydig cell progenitors in the fetal testis. *Dev Biol.* 2011;352:14–26. <https://doi.org/10.1016/j.ydbio.2011.01.011>.
- Dickel DE, Ypsilanti AR, Pla R, Zhu Y, Barozzi I, Mannion BJ, Khin YS, Fukuda-Yuzawa Y, Plajzer-Frick I, Pickle CS, et al. Ultraconserved enhancers are required for normal development. *Cell.* 2018;172:491–e49915. <https://doi.org/10.1016/j.cell.2017.12.017>.
- Estermann MA, Major AT, Smith CA. Gonadal sex differentiation: supporting versus steroidogenic cell lineage specification in mammals and birds. *Front Cell Dev Biol.* 2020a;8:616387. <https://doi.org/10.3389/fcell.2020.616387>.
- Estermann MA, Williams S, Hirst CE, Roly ZY, Serralbo O, Adhikari D, Powell D, Major AT, Smith CA. Insights into gonadal sex differentiation provided by single-cell transcriptomics in the Chicken embryo. *Cell Rep.* 2020b;31:107491. <https://doi.org/10.1016/j.celrep.2020.03.055>.
- Fang R, Preissl S, Hou X, Lucero J, Wang X, Motamedi A, Shiau AK, Mukamel EA, Zhang Y, Behrens MM, et al. Fast and accurate clustering of single cell Epigenomes reveals Cis-Regulatory Elements in Rare Cell types. *BioRxiv.* 2019. <https://doi.org/10.1101/615179>.
- Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature.* 2010;466:490–3. <https://doi.org/10.1038/nature09158>.
- Goertz MJ, Wu Z, Gallardo TD, Hamra FK, Castrillon DH. Foxo1 is required in mouse spermatogonial stem cells for their maintenance and the initiation of spermatogenesis. *J Clin Invest.* 2011;121:3456–66. <https://doi.org/10.1172/JCI57984>.
- Gonen N, Quinn A, O'Neill HC, Koopman P, Lovell-Badge R. Normal levels of sox9 expression in the developing mouse testis depend on the TES/TESCO enhancer, but this does not act alone. *PLoS Genet.* 2017;13:e1006520. <https://doi.org/10.1371/journal.pgen.1006520>.
- Gonen N, Futtner CR, Wood S, Garcia-Moreno SA, Salamone IM, Samson SC, Sekido R, Poulat F, Maatouk DM, Lovell-Badge R. Sex reversal following deletion of a single distal enhancer of Sox9. *Science.* 2018;360:1469–73. <https://doi.org/10.1126/science.aas9408>.
- Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang H, Greenleaf W. ArchR: an integrative and scalable software package for single-cell chromatin accessibility analysis. *BioRxiv.* 2020. <https://doi.org/10.1101/2020.04.28.066498>.
- Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, Greenleaf WJ. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet.* 2021;53:403–11. <https://doi.org/10.1038/s41588-021-00790-6>.
- Gregoire EP, De Cian M-C, Migale R, Perea-Gomez A, Schaub S, Bellido-Carreras N, Stévant I, Mayère C, Neirijnck Y, Loubat A, et al. The -KTS splice variant of WT1 is essential for ovarian determination in mice. *Science.* 2023;382:600–6. <https://doi.org/10.1126/science.add8831>.
- Hughes IA, Houk C, Ahmed SF, Lee PA, Consensus Group LWPES, ESPE Consensus Group. Consensus statement on management of intersex disorders. *Arch Dis Child.* 2006;91:554–63. <https://doi.org/10.1136/adc.2006.098319>.
- Ikeda Y, Shen WH, Ingraham HA, Parker KL. Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases. *Mol Endocrinol.* 1994;8:654–62. <https://doi.org/10.1210/mend.8.5.8058073>.
- Jiang S, Huang Z, Li Y, Yu C, Yu H, Ke Y, Jiang L, Liu J. Single-cell chromatin accessibility and transcriptome atlas of mouse embryos. *Cell Rep.* 2023;42:112210. <https://doi.org/10.1016/j.celrep.2023.112210>.
- Kanamori-Katayama M, Kaiho A, Ishizu Y, Okamura-Oho Y, Hino O, Abe M, Kishimoto T, Sekihara H, Nakamura Y, Suzuki H, et al. LRRN4 and UPK3B are markers of primary mesothelial cells. *PLoS ONE.* 2011;6:e25391. <https://doi.org/10.1371/journal.pone.0025391>.
- Kashimada K, Koopman P. Sry: the master switch in mammalian sex determination. *Development.* 2010;137:3921–30. <https://doi.org/10.1242/dev.048983>.
- Kim S, Shendure J. Mechanisms of interplay between transcription factors and the 3D genome. *Mol Cell.* 2019;76:306–19. <https://doi.org/10.1016/j.molcel.2019.08.010>.
- Kreidberg JA, Sariola H, Loring JM, Maeda M, Pelletier J, Housman D, Jaenisch R. WT-1 is required for early kidney development. *Cell.* 1993;74:679–91. [https://doi.org/10.1016/0092-8674\(93\)90515-r](https://doi.org/10.1016/0092-8674(93)90515-r).
- Kusaka M, Katoh-Fukui Y, Ogawa H, Miyabayashi K, Baba T, Shima Y, Sugiyama N, Sugimoto Y, Okuno Y, Kodama R, et al. Abnormal epithelial cell polarity and ectopic epidermal growth factor receptor (EGFR) expression induced in Emx2 KO embryonic gonads. *Endocrinology.* 2010;151:5893–904. <https://doi.org/10.1210/en.2010-0915>.
- Lala DS, Ikeda Y, Luo X, Baity LA, Meade JC, Parker KL. A cell-specific nuclear receptor regulates the steroid hydroxylases. *Steroids.* 1995;60:10–4. [https://doi.org/10.1016/0039-128x\(94\)00002-t](https://doi.org/10.1016/0039-128x(94)00002-t).
- Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The human transcription factors. *Cell.* 2018;175:598–9. <https://doi.org/10.1016/j.cell.2018.09.045>.

37. Li Y, Taketo T, Lau Y-FC. Isolation of fetal gonads from embryos of timed-pregnant mice for morphological and molecular studies. *Methods Mol Biol.* 2012;825:3–16. https://doi.org/10.1007/978-1-61779-436-0_1.
38. Lin Y-T, Barske L, DeFalco T, Capel B. Numb regulates somatic cell lineage commitment during early gonadogenesis in mice. *Development.* 2017;144:1607–18. <https://doi.org/10.1242/dev.149203>.
39. Maatouk DM, Natarajan A, Shibata Y, Song L, Crawford GE, Ohler U, Capel B. Genome-wide identification of regulatory elements in sertoli cells. *Development.* 2017;144:720–30. <https://doi.org/10.1242/dev.142554>.
40. Makiyan Z. Studies of gonadal sex differentiation. *Organogenesis.* 2016;12:42–51. <https://doi.org/10.1080/15476278.2016.1145318>.
41. Mayère C, Neirijnck Y, Sararols P, Rands CM, Stévant I, Kühne F, Chassot A-A, Chaboissier M-C, Dermitzakis ET, Nef S. Single-cell transcriptomics reveal temporal dynamics of critical regulators of germ cell fate during mouse sex determination. *FASEB J.* 2021;35:e21452. <https://doi.org/10.1096/fj.202002420R>.
42. Mayère C, Regard V, Perea-Gomez A, Bunce C, Neirijnck Y, Djari C, Bellido-Carreras N, Sararols P, Reeves R, Greenaway S, et al. Origin, specification and differentiation of a rare supporting-like lineage in the developing mouse gonad. *Sci Adv.* 2022;8:eabm0972. <https://doi.org/10.1126/sciadv.abm0972>.
43. Mazaud S, Oréal E, Guigon CJ, Carré-Eusèbe D, Magre S. Lhx9 expression during gonadal morphogenesis as related to the state of cell differentiation. *Gene Expr Patterns.* 2002;2:373–7. [https://doi.org/10.1016/S1567-133X\(02\)00050-9](https://doi.org/10.1016/S1567-133X(02)00050-9).
44. Morais da Silva S, Hacker A, Harley V, Goodfellow P, Swain A, Lovell-Badge R. Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nat Genet.* 1996;14:62–8. <https://doi.org/10.1038/ng0996-62>.
45. Mork L, Maatouk DM, McMahon JA, Guo JJ, Zhang P, McMahon AP, Capel B. Temporal differences in granulosa cell specification in the ovary reflect distinct follicle fates in mice. *Biol Reprod.* 2012;86:37. <https://doi.org/10.1095/biolreprod.111.095208>.
46. Murphy K, Carvajal L, Medico L, Pepling M. Expression of Stat3 in germ cells of developing and adult mouse ovaries and testes. *Gene Expr Patterns.* 2005;5:475–82. <https://doi.org/10.1016/j.modgep.2004.12.007>.
47. Nicholls PK, Harrison CA, Rainczuk KE, Vogl W, A., and, Stanton PG. Retinoic acid promotes sertoli cell differentiation and antagonises activin-induced proliferation. *Mol Cell Endocrinol.* 2013;377:33–43. <https://doi.org/10.1016/j.mce.2013.06.034>.
48. Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature.* 2018;554:239–43. <https://doi.org/10.1038/nature25461>.
49. Pelletier J, Schalling M, Buckler AJ, Rogers A, Haber DA, Housman D. Expression of the Wilms' tumor gene WT1 in the murine urogenital system. *Genes Dev.* 1991;5:1345–56. <https://doi.org/10.1101/gad.5.8.1345>.
50. Pérez-Losada J, Sánchez-Martin M, Rodríguez-García A, Sánchez ML, Orfao A, Flores T, Sánchez-García I. Zinc-finger transcription factor slug contributes to the function of the stem cell factor c-kit signaling pathway. *Blood.* 2002;100:1274–86. https://doi.org/10.1182/blood.V100.4.1274.h81602001274_1274_1286.
51. Polanco JC, Wilhelm D, Davidson T-L, Knight D, Koopman P. Sox10 gain-of-function causes XX sex reversal in mice: implications for human 22q-linked disorders of sex development. *Hum Mol Genet.* 2010;19:506–16. <https://doi.org/10.1093/hmg/ddp520>.
52. Rotgers E, Jørgensen A, Yao HH-C. At the crossroads of fate-somatic cell lineage specification in the fetal gonad. *Endocr Rev.* 2018;39:739–59. <https://doi.org/10.1210/er.2018-00010>.
53. Saga Y. How germ cells determine their own sexual fate in mice. *Sex Dev.* 2022;16:329–41. <https://doi.org/10.1159/000520976>.
54. Sasaki K, Oguchi A, Cheng K, Murakawa Y, Okamoto I, Ohta H, Yabuta Y, Iwatani C, Tsuchiya H, Yamamoto T, et al. The embryonic ontogeny of the gonadal somatic cells in mice and monkeys. *Cell Rep.* 2021;35:109075. <https://doi.org/10.1016/j.celrep.2021.109075>.
55. Sekido R, Lovell-Badge R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature.* 2008;453:930–4. <https://doi.org/10.1038/nature06944>.
56. Shen Y-C, Shami AN, Moritz L, Larose H, Manske GL, Ma Q, Zheng X, Sukhwani M, Czerwinski M, Sultan C, et al. TCF21 + mesenchymal cells contribute to testis somatic cell development, homeostasis, and regeneration in mice. *Nat Commun.* 2021;12:3876. <https://doi.org/10.1038/s41467-021-24130-8>.
57. Shima Y, Miyabayashi K, Baba T, Otake H, Katsura Y, Oka S, Zubair M, Morohashi K. Identification of an enhancer in the Ad4BP/SF-1 gene specific for fetal leydig cells. *Endocrinology.* 2012;153:417–25. <https://doi.org/10.1210/en.2011-1407>.
58. Shimada R, Koike H, Hirano T, Kato Y, Saga Y. NANOS2 suppresses the cell cycle by repressing mTORC1 activators in embryonic male germ cells. *iScience.* 2021;24:102890. <https://doi.org/10.1016/j.isci.2021.102890>.
59. Sikder HA, Devlin MK, Dunlap S, Ryu B, Alani RM. Id proteins in cell growth and tumorigenesis. *Cancer Cell.* 2003;3:525–30. [https://doi.org/10.1016/S1535-6108\(03\)00141-7](https://doi.org/10.1016/S1535-6108(03)00141-7).
60. Stévant I, Neirijnck Y, Borel C, Escoffier J, Smith LB, Antonarakis SE, Dermitzakis ET, Nef S. Deciphering cell lineage specification during male sex determination with single-cell RNA sequencing. *Cell Rep.* 2018;22:1589–99. <https://doi.org/10.1016/j.celrep.2018.01.043>.
61. Stévant I, Kühne F, Greenfield A, Chaboissier M-C, Dermitzakis ET, Nef S. Dissecting cell lineage specification and sex fate determination in gonadal somatic cells using single-cell transcriptomics. *Cell Rep.* 2019;26:3272–e32833. <https://doi.org/10.1016/j.celrep.2019.02.069>.
62. Tanaka SS, Nishinakamura R. Regulation of male sex determination: genital ridge formation and sry activation in mice. *Cell Mol Life Sci.* 2014;71:4781–802. <https://doi.org/10.1007/s00018-014-1703-3>.
63. Wang R, Liu X, Li L, Yang M, Yong J, Zhai F, Wen L, Yan L, Qiao J, Tang F. Dissecting human gonadal cell lineage specification and sex determination using a single-cell RNA-seq Approach. *Genomics Proteom Bioinf.* 2022;20:223–45. <https://doi.org/10.1016/j.gpb.2022.04.002>.
64. Wen Q, Cheng CY, Liu Y-X. Development, function and fate of fetal leydig cells. *Semin Cell Dev Biol.* 2016;59:89–98. <https://doi.org/10.1016/j.semcdb.2016.03.003>.
65. Zhao L, Arsenault M, Ng ET, Longmuss E, Chau TC-Y, Hartwig S, Koopman P. SOX4 regulates gonad morphogenesis and promotes male germ cell differentiation in mice. *Dev Biol.* 2017;423:46–56. <https://doi.org/10.1016/j.ydbio.2017.01.013>.
66. Zhou H, Xiang Y, Hu M, Xu Y, Hou Y, Qi X, Fu L, Luan Y, Wang Z, Li X, et al. Chromatin accessibility is associated with the changed expression of miRNAs that target members of the Hippo pathway during myoblast differentiation. *Cell Death Dis.* 2020;11:148. <https://doi.org/10.1038/s41419-020-2341-3>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.