Divergent regulatory element programs steer sex-specific supporting cell differentiation along mouse gonadal development

Isabelle Stévant[1#], Elisheva Abberbock[1#], Meshi Ridnik[1#], Roni Weiss[1], Linoy Swisa[1], Christopher R Futtner[2], Danielle Maatouk[3], Robin Lovell-Badge[4], Valeriya Malysheva[5,6,7], Nitzan Gonen[1*]

[1] The Mina and Everard Goodman Faculty of Life Sciences and the Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat Gan 5290002, Israel.

[2] Department of Endocrinology, Northwestern University, Chicago IL 60611, USA

[3] The author is deceased

[4] The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

[5] VIB Center for Molecular Neurology, VIB, Antwerp, Belgium

[6] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

[7] VIB Center for AI and Computational Biology, VIB, Leuven, Belgium

# These authors contributed equally

Contact information:

*Corresponding Author: nitzan.gonen@biu.ac.il

Short title: The gene regulatory landscape of sex determination

Keywords:

Testis, Ovary, Sertoli cells, pre-granulosa cells, gene regulatory networks, cis-regulatory elements, Transcription Factor Binding Sites enrichment

Abstract

Gonadal sex determination relies on tipping a delicate balance involving the activation and repression of several transcription factors and signalling pathways. This is likely mediated by numerous non-coding regulatory elements that shape sex-specific transcriptomic programs. To explore the dynamics of these in detail, we performed paired time-series of transcriptomic and chromatin accessibility assays on pre-granulosa and Sertoli cells throughout their development in the embryo, making use of new and existing mouse reporter lines. Regulatory elements were associated with their putative target genes by linkage analysis, and this was complemented and verified experimentally using promoter capture Hi-C. We identified the transcription factor motifs enriched in these regulatory elements along with their occupancy, pinpointing LHX9/EMX2 as potentially critical regulators of ovarian development. Variations in the DNA sequence of these regulatory elements are likely to be responsible for many of the unexplained cases of individuals with Differences of Sex Development.

Teaser

Multiomics analysis revealed the regulatory elements and transcription factors responsible for gonadal sex determination.

Introduction

In eutherian mammals, sex determination occurs during embryo development with the bipotential gonad committing to either testicular or ovarian cell fates. This choice relies on an initially extremely delicate balance between the expression and repression of several key transcription factors (TFs) and signalling pathways. There is then a period of reinforcement involving redundant mechanisms, perhaps to ensure the whole gonad follows one fate, before specific factors can become predominant postnatally (*1*, *2*). The gonadal sex determination process is, therefore, an ideal one in which to study cell fate decisions and the roles played by gene regulatory networks over time (*3*).

In mice, the bipotential gonad begins to develop at embryonic day 10 (E10.0) as a thickening layer positioned on the ventromedial surface of the mesonephros (*4–7*). The early bipotential gonad in the mouse comprises two distinct populations of progenitor cells: those derived from the coelomic epithelium, which can be considered multipotent because it is the origin of most of the somatic cell types, and the primordial germ cells, which differentiate according to the type of gonad that forms into spermatogonia in the testis, or oogonia in the ovary. The somatic cell types include the bipotential supporting cell precursors that differentiate into Sertoli cells (male) or pre-granulosa cells (female), and the steroidogenic cell precursors, which give rise to Leydig (male) or theca (female) cells (*5–10*).

In XY gonads, the sex-determining gene *Sry*, located on the Y chromosome, together with its direct downstream gene *Sox9*, initiates a genetic cascade at E11.5 leading to the differentiation of the supporting cell precursors into Sertoli cells (*11*, *12*). Sertoli cell identity is further maintained by several other TFs and signalling pathways including *Nr5a1*, *Gata4*, *Wt1*, *Dmrt1*, *Sox8* and *Fgf9*, which together reinforce Sertoli cell fate and repress the pre-granulosa cell pathway (*5–7*). Once established, Sertoli cells instruct the steroidogenic cell precursors to differentiate into Leydig cells (*13*), and also encapsulate the germ cells and promote their differentiation into prospermatogonia (*14*).

In XX gonads, without SRY to upregulate *Sox9* expression by E11.5, the activity of ovarian promoting factors leads to pre-granulosa cell differentiation and ovary development. These include the ovarian determining factor $WT1^{-KTS}$ isoform, the WNT4/RSPO1/β-Catenin signalling pathway and several TFs, notably FOXL2 and RUNX1 (*15–20*). The expression of these factors within the supporting cell precursors also act to repress the Sertoli cell

differentiation program, for example high levels of WT1^{-KTS} interfere with *Sry* expression (*12*), while WNT signalling interferes with *Sox9* upregulation (*21*, *22*).

Despite identifying many key pro-testicular and pro-ovarian factors, the gene regulatory networks they govern and the genomic elements they utilize to coordinate gene expression during sex determination remain poorly understood. *Cis*-regulatory elements are defined as regions of non-coding of DNA (usually 500–1,500 bp) that have the capacity to control gene expression. Such elements are highly enriched in transcription factor binding sites and can be located upstream or downstream and at varying distances of the transcriptional start site(s) of the genes they regulate (*23–25*). *Cis*-regulatory elements are usually classified as promoters, enhancers, silencers, and insulators. Within the field of sex determination, few *cis*-regulatory regions have been characterized in detail to date. We have previously identified Enh13, a distal enhancer of the *Sox9* gene, as a critical enhancer for male sex determination. We and others have shown that deletions, duplications and even micro-deletions in two transcription factor binding sites (TFBS) of this enhancer can significantly alter *Sox*9 expression levels, leading to complete sex reversal in both mice and humans (*26–29*). Thus, we believe Enh13 is one of many regulatory elements in which variant DNA sequences/mutations may underlie numerous unexplained cases of Differences (or Disorders) of Sex Development (DSDs), and which offer insights into the complex gene regulatory networks driving sex determination and gonadal differentiation.

Pioneer attempts to characterize the gene regulatory landscape of purified Sertoli and pre-granulosa cells have used DNaseI-seq (*30*) and ATAC-seq (*31*) on TESCO-CFP (*32*) and TESMS-CFP (*27*) transgenic lines. While these approaches identified key elements like Enh13 (*27*), they provided limited temporal resolution. Indeed, DNaseI-seq was restricted to Sertoli cells at E13.5 and E15.5, and ATAC-seq examined supporting cell precursors at E10.5 and Sertoli and pre-granulosa cells at E13.5. However, single-cell transcriptomics analyses indicate that critical gonadal gene expression changes occur between E11.5–E12.5 (*9*, *10*, *33*), likely accompanied by dynamic patterns of chromatin accessibility. The lack of such data between E10.5 and E13.5 may obscure transient regulatory elements active during this critical window. Thus, a detailed timeline of chromatin accessibility, supported by high-quality, cell type-specific, and time-resolved bulk RNA datasets, is essential to fully capture and begin to understand the complexity of *cis*-regulation during sex determination.

Here, we generated a novel reporter mouse strain, *Enh8-mCherry,* which allowed us to efficiently purify pre-granulosa cells. Using both the *Enh8*-mCherry and *Sox9*$^{IRES\text{-}GFP}$ mouse strains, we purified pre-granulosa and Sertoli cells at four developmental time points in which sex determination occurs and performed bulk RNA-seq and ATAC-seq. This data constitutes the first bulk RNA-seq data of these critical cell types and provides a wealth of information of their dynamic transcriptomic profiles. The chromatin accessibility data enables the identification of numerous putative regulatory elements that may function during sex determination, among which many exhibit sex- and stage-specific patterns. We employed computational approaches to provide linkage correlation between accessible regions and gene expression profiles allowing us to identify putative enhancers and silencers of key sex determination genes. This was validated *in vivo* by performing Promoter Capture Hi-C (PCHi-C) on purified Sertoli and pre-granulosa cells at E13.5. We find enrichment of our ATAC-seq candidate regulatory elements within the PCHi-C data, suggesting that many of the elements we identified are physically bound to the promoters of the neighbouring genes that they regulate within these cells. Furthermore, we performed transcription factor motif and footprinting analyses on the open chromatin regions in each sex, allowing us to identify the regulators that are bound to them and therefore likely to control the process of sex determination. In Sertoli cells we see enriched binding of SOX/SRY/DMRT1/GATA/WT1 and NR5A1 TFs, all of which are known to be critical for testis development (*34*). Strikingly, in pre-granulosa cells we see all of the known pro-female factors as WT1/FOXL2/RUNX1/TCF and GATA, but the most enriched motif is the motif bound by EMX2 and LXH9, suggesting that they may play an additional, yet unidentified role during ovary differentiation and not only during early gonad development (*6*, *35*, *36*). Altogether, this study constitutes an extensive analysis of the gene regulatory networks that operate during mouse sex determination, and it is highly likely that variants within the human homologues of these regulatory elements are responsible for many of the undiagnosed cases of DSD.

Results

Generation and use of transgenic mouse lines that allow efficient purification of pre-granulosa and Sertoli cells

To explore the gene regulatory network and *cis*-regulatory elements controlling gonadal sex determination, we first needed to purify gonadal supporting cells from both sexes. Therefore, we needed transgenic mouse lines that allow for the efficient purification of pre-granulosa and Sertoli cells at key developmental stages. Existing pre-granulosa cell reporter used in seminal studies, such as the TESMS-CFP (*31*) or the Sry-GFP (*37*, *38*) presents limitations in term of percentage of positive cells and sorting efficiency due to low fluorescent signal. To overcome these limitations, we generated a novel transgenic mouse line in which the *Sox9* enhancer 8 (referred to as Enh8) was cloned upstream of the *hsp68* minimal promoter and the *mCherry* gene (**Fig. 1A**, termed *Tg(Enh8-mCherry)*, or in short *Enh8-mCherry*). Enh8 is a 672 bp-long enhancer, located 838 kb upstream to the *Sox9* transcription start site (TSS) (*27*). This enhancer has been shown to be active and capable of driving LacZ reporter expression in the mouse embryonic ovary. Indeed, although *Sox9* is highly expressed in Sertoli cells and faintly in pre-granulosa cells, several studies performing ChIP-seq in both embryonic and adult ovaries have found binding of FOXL2 and RUNX1, two pre-granulosa cell-specific transcription factors, to the *Sox9* Enh8 (**Fig. 1A**) (*3*, *39–42*). The binding of these factors explains why this enhancer is active in pre-granulosa cells. To characterize the expression profile of the *Enh8-mCherry* mouse line, we dissected XX and XY embryonic gonads at E11.5, 12.5, 13.5 and 15.5. As expected, mCherry expression was evident in ovaries from E11.5 onward (**Fig. 1B**). Surprisingly, and conversely to what we observed with the LacZ reporter (*27*), mCherry expression was also evident in the E11.5-E15.5 testes (**Fig. S1A**).

To explore whether the mCherry-positive cells are indeed pre-granulosa cells in the ovary, we performed co-immunostaining with antibodies against mCherry, FOXL2 (pre-granulosa cells) and TRA98 (germ cells). As demonstrated in **Fig. 1C** and **Fig**. **S1B**, the cytoplasmic mCherry staining overlaps nicely with the nuclear FOXL2 staining suggesting that the mCherry labels pre-granulosa cells. No overlap is seen with the TRA98-positive germ cells. To further confirm that the XX mCherry-positive cells are pre-granulosa cells, we performed bulk RNA-seq on E11.5, E12.5 and E13.5 mCherry-positive sorted cells. We compared the expression of different cell-type-specific genes (*9*, *10*) in our RNA-seq data to that of embryonic whole gonad RNA-seq data (*43*). As demonstrated in **Figure 1D**, expression profiles of the

mCherry-positive cells resemble pre-granulosa cells with enriched expression of markers of the pre-supporting cells, but also many markers known to be expressed in pre-granulosa cells. No expression and negative enrichments were found with markers of both germ cells and stromal cells, indicating the absence of these cells in the mCherry-positive sorted cells (**Fig. 1D**). All of the above strongly suggest that this new *Enh8-mCherry* reporter mouse line is an efficient genetic tool to allow the sorting of pure population of pre-granulosa cells from embryonic XX gonads.

To analyse which cells are mCherry-positive in embryonic testes, we performed immunostaining with antibodies against mCherry, SOX9 (Sertoli cells) and TRA98 (germ cells) (**Fig. S1C**). While mCherry overlapped with SOX9, suggesting it is labelling Sertoli cells, there was also expression outside the tubules, within the interstitium. Co-staining with mCherry and 3βHSD (fetal Leydig cells) indicated that the mCherry is also labelling fetal Leydig cells (**Fig. S1C**). Hence, this mouse line cannot be used to purify Sertoli cells from embryonic gonads. Instead, we used the well-established *Sox9*$^{IRES\text{-}GFP}$ reporter strain (*44*) where GFP labels the *Sox9*-expressing Sertoli cells (**Fig. S1D**).

We and others have previously used the TESCO-CFP and TESMS-CFP reporter mouse lines to purify embryonic Sertoli and pre-granulosa cells, respectively (*27*, *30*, *31*). While these lines allow the purification of Sertoli and pre-granulosa cells, the percentage of CFP-positive cells out of the entire embryonic gonad was significantly lower than we find with the *Enh8-mCherry* or *Sox9*$^{IRES\text{-}GFP}$ lines (~33-40% mCherry-positive cells in *Enh8-mCherry* ovaries, ~14% of GFP- positive cells in *Sox9*$^{IRES\text{-}GFP}$ testes, ~5-7% CFP positive cells with the TESCO-CFP/TESMS-CFP reporters) (**Fig. S2**). ScRNA-seq studies show that the actual proportion of supporting cells in the embryonic gonads is similar to those obtained after sorting with the *Enh8-mCherry* or the *Sox9*$^{IRES\text{-}GFP}$ lines (**Fig. S2D**), suggesting that these allow more representative capture of Sertoli and pre-granulosa cell populations at these stages.

Exploring the transcriptomics and chromatin accessibility of purified pre-granulosa and Sertoli cells from embryonic gonads

To investigate the establishment of the *cis*-regulatory element landscape that drives pre-supporting cell differentiation as pre-granulosa and Sertoli cells in embryonic gonads, we performed paired time-series transcriptomic (RNA-seq) and chromatin accessibility assays

(ATAC-seq) at four time points (E11.5, E12.5, E13.5 and E15.5) covering cell fate commitment and differentiation in both sexes using the *Enh8-mCherry* for XX and *Sox9*$^{IRES\text{-}GFP}$ for XY gonads (**Fig. 1E**, **Fig. S3**, methods).

The bulk transcriptomes on sorted Sertoli and pre-granulosa cells provided the expression level of 12,058 protein-coding genes, and the ATAC-seq detected a total of 87,988 high confidence and non-overlapping open chromatin regions (stages and sexes combined), which represent 2.8% of the mouse genome. Principal component analysis (PCA) reveals that transcriptomic and chromatin accessibility changes along supporting cell differentiation are following fairly similar trajectories and are associated first with the sex (PC1), and second with the embryonic developmental stages (PC2) (**Fig. 1F and G**). This illustrates that the rearrangement of the chromatin accessibility is occurring in concert with the establishment of sex-specific gene expression programs during gonadal supporting cell differentiation.

Next, we wanted to compare our bulk RNA-seq on purified pre-granulosa and Sertoli cells (E11.5-E15.5) to a bulk RNA-seq dataset that was done on whole embryonic gonads (E11.5-E13.5) (*43*). It is evident that the expression levels of pre-granulosa cell markers (*Runx1* and *Fst)* and Sertoli cell markers (*Sox9* and *Amh)* differ significantly between the two datasets, showing higher expression in our data (solid lines) compared to the whole gonad data (dashed lines) (**Fig. 1H)** without any enrichment in other cell populations **(Fig. S4**). This is likely due to dilution of supporting cell gene expression when many other cell types are present in the whole gonad. Hence, our data constitute an important resource and the first time-series bulk transcriptomic data of the supporting cell population during the developmental window of sex determination. To facilitate further research, we offer a web application to enable the community to easily plot the expression level of their genes of interest (Link).

These paired time-series RNA-seq and ATAC-seq on sorted gonadal supporting cells from both sexes allow the in-depth characterization of the relationship between gene expression temporal dynamics and chromatin accessibility in regard to the differentiation of the pre-supporting cells into pre-granulosa and Sertoli cells.

The sexual dimorphism of pre-granulosa and Sertoli cell transcriptomes increases with time

We undertook to characterize the transcriptomes of the differentiating supporting cells in both sexes to discern the genes that are expressed in a sex- and time-specific manner with a focus on transcription factors (TFs) as they directly contribute to gene regulatory networks.

We first conducted differential expression analysis to identify the genes exhibiting sexual dimorphism at each embryonic stage (**Fig. 2A, Fig. S5, data file S1**). Supporting cells progressively acquire a growing number of genes expressed in a sex-biased manner as they differentiate (**Fig. 2A**), from 1,776 and 1,424 at E11.5, to 3,713 and 3,649 at E15.5 pre-granulosa and Sertoli cells, respectively. The enriched GO (Gene Ontology) terms associated with sexually dimorphic genes at each embryonic stage align with our knowledge of the supporting cell differentiation processes. Genes expressed at higher levels in pre-granulosa cells are involved in epithelial morphogenesis, cell differentiation and WNT signalling pathways (*17*, *18*, *45*), while those with higher expression in Sertoli cells are predominantly linked to mitotic cell cycle and epithelial morphogenesis (*46*, *47*) (**Fig. S5, data file S1**). The number of TFs exhibiting sex-biased expression also increases progressively as cells differentiate (**data file S1**). Across all stages, pre-granulosa cells exhibit higher levels of transcripts for 552 TFs compared to Sertoli cells, of these, 84 have been associated with gonadal or infertility phenotypes in the Mouse Genome Informatics (MGI) phenotype database (*48*) (**data file S1 and S2**). These include well-known critical gonadal factors such as *Tcf21* (also known as *Pod1*), *Foxl2*, *Nr0b1* (also known as *Dax1*), but also less described factors like *Osr1* that causes genital ridge hypoplasia when mutated in mice (*49*), and *Pbx3* which leads to the absence of ovaries in adults when deleted, as reported by the International Mouse Phenotyping Consortium (IMPC) (*50*). Similarly, Sertoli cells express higher levels of 471 TFs compared to pre-granulosa cells, with 76 of these associated with a gonadal phenotype upon mutation. These include well-known testicular factors *Sry*, *Sox9 and Dmrt1*, as well as lesser-known factors mainly associated with infertility, such as *Patz1* (*51*), *Pick1*(*52*) and *Hmga1* (*53*) (**data file S1**).

We next explored the dynamics of the supporting cell transcriptomes in each sex separately to identify the genes whose expression changes throughout the differentiation process. Pre-granulosa cell differentiation involves 6,345 genes with expression level changes, while Sertoli cell differentiation involves modulation of 5,475 genes. These genes were

subsequently classified based on their expression profiles (groups *a* to *h* in pre-granulosa cells, and *a* to *g* in Sertoli cells) and GO term enrichment analysis was performed on each of the gene profiles (**Fig. 2B and C, data file S3**). Among these dynamically expressed genes, we identified 574 TFs in pre-granulosa cells, with 89 of them reported causing a gonadal phenotype when mutated. In Sertoli cells, 530 dynamically expressed TFs were found, 86 of which are associated with gonadal phenotypes upon mutation (**data file S3**), some of them are highlighted in **Figure 2B and C**.

When overlapping the genes exhibiting expression changes during cell differentiation of both sexes, we observed that approximately half of the dynamically expressed genes in pre-granulosa cells were also found to change in Sertoli cells, and *vice versa* (**Fig. 2D**, **data file S3**). As examples, the dosage sensitive sex-determining factor *Nr0b1* (also known as *Dax1*) (*54*), and *Cyp11a1*, a cholesterol cleavage enzyme (*55*) expressed in pre-supporting cells around E11.5 (*10*) are similarly expressed in both cell types along their differentiation (**Fig. 2D**). In contrast, *Gata4* (*56*, *57*) and *Dnmt3a* (*58*) expression is relatively stable in Sertoli cells but changes in pre-granulosa cells (**Fig. 2D**). Conversely, the genes *Ctnnd1* (also known as δ-catenin*)* and *Inha* are stably expressed in pre-granulosa cells but change in Sertoli cells (**Fig. 2D**).

This characterization of the transcriptome during supporting cell differentiation identified a total of 802 TFs with differential expression either by sex or at specific embryonic stages as cells differentiate into pre-granulosa or Sertoli cells. Mutations of 118 of these are known to lead to a gonadal phenotype or infertility when mutated in mice, but many others have not yet been studied in the context of sex determination and gonadal development, though they may play crucial roles.

Profiling sexually dimorphic chromatin region accessibility along supporting cell differentiation

To identify the putative *cis*-regulatory elements involved in supporting cell differentiation such as potential promoters, enhancers, silencers, and insulators, we analysed the time-series ATAC-seq data using a quantitative approach. We quantified the ATAC-seq signal of the 87,988 gonadal open chromatin regions found in all sexes and stages and assessed their level of accessibility across sexes and embryonic stages. We then performed differential accessible

region (DAR) analysis to identify sexually dimorphic open chromatin regions during differentiation.

As shown in **Figure 3A**, pre-granulosa cells present fewer sex-biased accessible regions than Sertoli cells (21,608 and 31,081 in total, respectively), which is consistent with previous observations (*31*). Like the transcriptomes, the open chromatin landscape exhibits increasing sexual dimorphism as cells differentiate, from 4,206 to 16,619 regions more accessible in pre-granulosa cells and 12,679 to 24,703 regions more accessible in Sertoli cells between E11.5 and E15.5 (**Fig. 3A, data file S4**). A major proportion of the pre-granulosa-biased accessible regions (7,476) appears from E12.5 onward (**Fig. 3B**), while 9,039 Sertoli-biased regions are already established from E11.5 (**Fig. 3C**), suggesting the sex-specific accessible chromatin landscape mediating pre-granulosa cell differentiation is delayed compared to Sertoli cells, as suggested by previous transcriptomic studies (*10*, *15*). The sexually dimorphic open chromatin regions are highly enriched in intronic and intergenic regions when compared to all the open chromatin regions (**Fig. 3D, Fig. S3E**). Therefore, the sex differences in chromatin accessibility between the differentiating supporting cells are explained by the increase in accessibility of sex-specific enhancers, silencers, or insulators rather than gene promoters.

Figures 3E-G exhibit examples of interesting accessible regions presenting a sex-specific pattern. The vicinity of the female-expressed gene *Rspo1* presents six genomic regions that are only accessible in pre-granulosa cells, but not in Sertoli cells (highlighted in yellow). These regions are likely to be redundantly involved in the establishment of *Rspo1* expression in pre-granulosa cells (**Fig 3E**). Likewise, exploring the *Sox8* genomic locus, we identified three male-specific regions (highlighted in light blue) which concord with the Sertoli-specific *Sox8* expression, although the *Sox8* gene promoter remains accessible in both sexes. Interestingly, our data can also identify sex-specific regions within genes expressed in both sexes. The *Zfpm2* gene (also known as *Fog2*), which is a critical co-factor of the GATA4 protein (*59*, *60*) is expressed in both Sertoli and pre-granulosa cells. While we can identify several open chromatin regions that behave similarly in both sexes (highlighted in grey), we can also identify a male-specific region located in the second intron (highlighted in light blue) (**Fig. 3G**). This may suggest that a gene expressed in both Sertoli and pre-granulosa cells can be regulated by different, sex-specific, set of regulatory regions.

We next looked at the enrichment of TF-binding motifs in the sexually dimorphic accessible chromatin regions. We focused our analysis on TFs that were found expressed in the supporting cells based on our RNA-seq data (**Fig. 3H and I, data file S5**). Both pre-granulosa and Sertoli-biased open chromatin regions are enriched for the motifs recognized by the non-sex-specific gonadal factors NR5A1, GATAs, WT1, and NR2F2. Among the pre-granulosa biased open chromatin enriched motifs, we also find known ovary-specific factors such as FOXL2, RUNX1, TCFs and HES1 (*61*). Similarly, we find enrichment of the testis-specific factors DMRTs and SOX-SRY in the Sertoli-biased open chromatin regions. We notice, however, that the GATA factors are much more enriched in pre-granulosa biased regions compared to Sertoli, and that the motif recognized by many factors including EMX2 and LHX9, which are critical factors for the genital ridge development, is among the topmost enriched motifs in pre-granulosa cells.

Taken together, the results demonstrate that supporting cells operate major sex-specific chromatin rearrangements along their differentiation. These sex-specific rearrangements are concomitant with the increase in accessibility of TF-binding motifs related to their respective sex-specific factors, but also factors with yet no identified role in the context of gonadal development. It remains to decipher whether the sex-specific chromatin accessibility changes are a cause or the consequence of the sex-specific TF-binding.

Chromatin accessibility landscapes transition along supporting cell differentiation

We then focused on the open chromatin regions that change in accessibility along supporting cell development in each sex. We found 8,298 regions that change in accessibility in pre-granulosa cells and 20,607 in Sertoli cells along their differentiation, demonstrating that Sertoli cells operate a massive rearrangement of their chromatin landscape (**Fig. 4A and B, data file S6**). Among them, only 2,266 are found in common in both sexes, suggesting that the change in accessibility is mostly driven by sex-specific regions (**Fig. 4C**). We classified the chromatin regions according to the dynamics of their accessibility (groups *a* to *d* for both sexes). The chromatin accessibility events in pre-granulosa cells can be divided into two main modules. The first consists of chromatin regions that decrease in accessibility from either E12.5 (group *a*) or from E13.5 (group *b*) as cells differentiate. The second module represents chromatin regions that increase in accessibility, from E13.5 (group *c*) and from E15.5 (group *d*) (**Fig. 4A**). This suggests that the pre-granulosa chromatin landscape is transitioning from

committed to differentiated cells between E12.5 and E13.5. Sertoli cells present more gradual chromatin accessibility changes, with regions from groups *a* and *b* that decrease in accessibility, and groups *c* and *d* that become more accessible (**Fig. 4B**). However, unlike pre-granulosa cells, the increase in accessibility is initiating from E12.5 onward, supporting the idea that Sertoli specific chromatin landscape established earlier than pre-granulosa cells, similar to the transcriptomic profiles (*10*, *15*).

These genomic regions are mainly located in intronic and intergenic regions as observed with the sexually dimorphic accessible regions (**Fig. 4D**). **Figure 4E-F** presents several interesting examples of changes in chromatin accessibility over time. For example, we observe that the *Wnt4* gene, which is expressed in pre-granulosa cells, presents an open chromatin region in its first intron that increases in accessibility as cells differentiate (highlighted in yellow), while the promoter of the *Wnt4* gene remains stably accessible over time (**Fig. 4E**). Similarly, we identified two intronic open chromatin regions on the *Fshr* male specific gene that increase in accessibility as cells differentiate and are likely to be putative enhancers (**Fig. 4F**). Among the dynamically accessible regions present around gonadal critical gonadal genes, we could observe more complex patterns. *Dmrt1*, which is expressed in both pre-granulosa and Sertoli at E11.5 and become exclusively expressed in Sertoli cells from E12.5 onward, presents one open chromatin region upstream of the promoter that gradually becomes more accessible specifically in Sertoli cells (highlighted in light blue), while another region in the second intron shows a decrease in accessibility in pre-granulosa cells while remain stably accessible in Sertoli cells (highlighted in yellow) (**Fig. 4G**). Altogether this suggests a complicated gene regulatory network both temporally and also between sexes.

The changes in the open chromatin landscape also reflect a change of accessibility of specific TF-binding motifs. Differential TF-binding motif enrichment analysis showed that pre-granulosa cell regions that are more accessible at early stages (groups *a* and *b*) are enriched for motifs recognized by NR2F2 and NR5A1, but also DMRT1. The group *d* is strongly enriched for motifs recognized by different factors including FOS, which induce impaired ovarian folliculogenesis with atretic follicles in adult mice when mutated (*62*) (**Fig. 4H, data file 7**). In Sertoli cells (**Fig. 4I, data file 7**), we see that the group *a* □ *i.e.* the regions that are more accessible at E11.5 □ is enriched in motifs for LHX9 and RUNX1, TFs whose expression is down-regulated in Sertoli cells after E11.5. The group *c* □ regions that are gradually more accessible from E12.5 □ is enriched in motifs for GATAs, NR2F2 and

NR5A1, DMRT1, and the SOXs; and the group *d* also show enrichment for DMRT1, a TF known to be crucial for sex maintenance in the testis.

Taken together, the ATAC-seq data analysis of the differentiating pre-granulosa and Sertoli cells allowed the identification of genomic loci presenting a difference in accessibility between sexes and embryonic stages. These regions are enriched in motifs for known critical sex-specific TFs but also many others that have not yet been described in the context of gonadal development. This data also provides a wide atlas for the identification of stage- and sex-specific regulatory elements that may be involved in the tight and precise regulation of gene expression during sex determination, mutations in which may lead to DSD.

Predicting the target genes of the putative supporting cell *cis*-regulatory regions

After having characterized the transcriptome and the open chromatin landscape of the supporting cells, we set out to predict the target genes of the putative *cis*-regulatory elements. One of the most commonly used proxies to achieve this is to look for the correlation between gene expression and accessibility of the open chromatin regions in the +/-500 kb region from their TSS. A positive correlation, or link, corresponds to a chromatin region accessibility that mirrors the nearby gene expression and therefore suggests a putative enhancer function. Conversely, a negative correlation means that the chromatin accessibility is the opposite of the nearby gene expression suggesting it may function as a silencer (**Fig. 5A**). Using this method, we could identify 44,116 putative regulatory regions (50.1% of all the open chromatin regions) whose accessibility correlates positively or negatively with 10,685 genes, which represents 88.6% of all the expressed protein coding genes (**data file S8**). Each gene was linked to an average of seven putative *cis*-regulatory elements, four positively and three negatively (**Fig. 5A, Fig. S6**) and each linked open chromatin region was connected to a median of two genes (**Fig. S6**). The linked open chromatin regions are mainly located within intragenic regions, including within the gene that they are linked with, and intergenic regions (**Fig. S6**).

In **Figures 5 B**, **C** and **D**, we show three examples of sex-specific genes involved in gonadal development and some of their regulatory elements that present either a positive or negative correlation. We found only one open chromatin region that was linked to the testis determining factor *Sry*, located 84 kb upstream of its TSS (**Fig. 5B**). This region corresponds to the locus containing transcriptionally active sequences derived from transposable elements

we previously identified (*63*). *Lef1*, a downstream gene of the WNT signalling, presents 15 open chromatin regions that correlate with its expression in a window of +/-500 kb from its TSS (**data file S8**). These include 10 putative enhancers, *i.e.* open in pre-granulosa but not in Sertoli cells, and five putative silencers, *i.e.* close in pre-granulosa but open in Sertoli cells. **Figure 5C** shows two upstream and four intronic open chromatin regions that are likely to be *Lef1* enhancers, and one downstream region that which accessibility anti-correlates with *Lef1* expression and could act as a silencer (**Fig. 5C**). Similarly, we found nine putative *Fgf9* enhancers, and one potential silencer (**data file S8**). **Figure 5D** shows four distal downstream putative *Fgf9* enhancer regions and the potential downstream silencer (**Fig. 5D**). Interestingly, deletion of 306 kb region that includes the most downstream enhancer resulted in XY male-to-female sex reversal in mice (*64*).

To complement the linkage prediction analysis, we performed promoter capture Hi-C (PCHi-C) on purified E13.5 Sertoli and pre-granulosa cells (**Fig. 5E**). PCHi-C enables to profile chromatin interactions between promoters of protein-coding genes and their regulatory elements (*65*, *66*). Significant contacts were detected with CHiCAGO (*67*) at a 5 kb resolution using two approaches (original bait and 5 kb extended bait) according to the best practice guidelines, with and without inclusion of promoters in the binning process to increase the detection sensitivity for proximal and distal interactions (*65*) (see Methods, **Fig. S7A**). In total, we detected 82,532 and 108,206 contacts (interaction score > 3) in pre-granulosa and Sertoli cells, respectively, between promoters and promoter-interacting regions (PIRs) at 5 kb resolution, 60,909 of them being commonly found in both sexes (interaction score > 3 in both sexes) (**data file S9**). Pre-granulosa and Sertoli interactions were enriched for markers of accessible and/or active enhancers (ATAC, H3K27ac) and active transcription (H3K4me3), previously described in pre-granulosa and Sertoli cells (*31*) ("active PIRs", **Fig. S7B**). Moreover, we observed a positive correlation between mean gene expression and the number of promoter-interacting regions, which suggest that highly expressed genes tend to be controlled by more *cis*-regulatory regions (**Fig. S7C**).

Overall, we found 3,603 PCHi-C interactions that overlap with 3,484 links predicted with our linkage analysis. **Figure 5F** and **G** show examples of physical interactions that we also predicted around two sexually dimorphic genes. We identified a physical interaction between a distal pre-granulosa-specific open chromatin region downstream of *Foxl2* and its promoter (**Fig. 5F**), strongly suggesting this region functions as an enhancer for *Foxl2*. Similarly, we

observed five Sertoli-specific open chromatin regions that physically interact with the *Serpine2* gene, highly expressed specifically in Sertoli cells (**Fig. 5G**).

This data provides the first promoter-targeted chromatin interactome in Sertoli and pre-granulosa cells, shedding light on the regulatory signalling involved in sex determination.

Different sets of TFs are physically bound to regulatory elements in pre-granulosa and Sertoli cells

For deciphering comprehensive gene regulatory networks, it is crucial to identify regulatory elements, physically associate them to their target genes and identify the TFs that bind these enhancers. While we identified the accessible chromatin regions in pre-granulosa and Sertoli cells and were able to predictively and physically associate them to their target genes, we still aim to identify which TFs bind to these regulatory regions to control precise gene expression patterns. While few TF ChIP-seq experiments were performed on gonadal cells (*3*, *40*, *41*, *68*), it remains a major challenge due to the highly limited number of gonadal cells at embryonic stages. To overcome this challenge and identify potential TFs that are physically bound to accessible chromatin regions, we conducted ATAC footprinting analysis on the open chromatin regions identified in both cell types at all developmental stages.

To that aim, we first looked at the enrichment of TF-binding motifs in the sexually dimorphic accessible chromatin regions. We performed differential TF-binding motifs enrichment on the sexually dimorphic accessible chromatin regions to detect motifs that are more accessible in one sex compared to the other (**Fig. 6A, Fig. S8, data file S5**). Next, we used the ATAC-seq data to detect TF-binding motif occupancy or footprinting. This represents physical binding of protein onto the open chromatin DNA in a way that confers small blockage to Tn5 digestion (*69*). We measured the difference of occupancy of the motifs of the expressed TFs between pre-granulosa and Sertoli cells in the sex-biased open chromatin regions and confirmed that most of the sexually dimorphic enriched motifs are also differentially occupied by a transcription factor (**Fig. 6A, data file S10**). Surprisingly, although the RFX1/5 and ZBTB14 motifs are among the top three enriched transcription factor motifs of pre-granulosa-biased open chromatin regions compared to Sertoli, they are not the most differentially bound motifs. The motif recognized by EMX2, LHX9 and MSX1, among

others, is the most differentially bound compared to Sertoli (**Fig. 6A and B**), and its score increases with time, suggesting that these factors are potentially the most important for the establishment of the pre-granulosa specific gene regulatory network. Conversely, the motifs recognized by the DMRT1 and SOX TFs are the most differentially bound in Sertoli cells (**Fig. 6A and C**), confirming their key role in controlling Sertoli cell differentiation and identity maintenance (*2*, *70*).

We focused our interest on the factors that are the most differentially bound in the pre-granulosa cell-biased open chromatin regions (ARID3B, EMX2, LHX9, ISX and MSX1). We checked the expression level of their genes in our RNA-seq data (**Figure S9A**) and in the single-cell RNA-seq atlas of the developing gonad (*9*) (**Figure S9B**) to identify which factors are the most highly and specifically expressed in the pre-granulosa cells. We found that *Arid3b* and *Isx* are lowly expressed in pre-granulosa cells in both dataset (**Figure S9**). *Msx1* is detected at a higher level in our bulk RNA-seq (**Figure S9A**) than in the single-cell data (**Figure S9B**), and is also expressed in the female germ cells. Finally, *Emx2* and *Lhx9*, while they are expressed in both sexes in the early progenitors (prior to supporting cell commitment) and the pre-supporting cells (E11.5 supporting cells), their expression decreases in Sertoli cells and become almost restricted to pre-granulosa cells (**Figure S9**). These observations suggest that EMX2 and LHX9 could play a crucial role in the pre-granulosa cell commitment.

Finally, we looked at TF footprint at a locus resolution around known enhancers like Enh8 and Enh13 of the *Sox9* gene (**Fig. 6D and E)** but also within candidate enhancers of interesting sex determination genes to identify potential gonadal factor binding sites (highlighted in **Fig. 6A**) (**Figure 6D to J**, **data file S10**). We can see footprinting of SOX genes to Enh13 (**Fig. 6D**) as well as footprinting of several pro-female factors to Enh8 as RUNX1, GATAs, NR5A1 and LHX9/EMX2 (**Fig. 6E**). Interestingly, we can identify binding of SOX/DMRT and NR5A1 to the downstream enhancer of *Fgf9* gene (**Fig. 6I**) but also footprinting of GATAs, NR5A1 and RUNX1 to the upstream enhancer of *Sry* (**Fig 6J**). This analysis can pinpoint the key factors bound to each regulatory element of each gene.

Altogether, we find that the sex-biased open chromatin regions are enriched in TF motifs for the known sex determining factors of the respective sex, but also for many less known TFs that may have a role in gonad development. Thus, these regions constitute putative genomic targets of the sex-determining factors. Furthermore, it is possible that factors with known

function at early gonad development such as EMX2 and LHX9 also play a critical role at later stages during pre-granulosa cell development.

Discussion

Decades of research on mice sex determination and human DSD individuals have unravelled many of the TFs and signalling pathways involved in the process. Yet, it is striking that ~70% of the 46,XX DSD and ~50% of 46,XY DSD patients fail to receive genetic diagnosis following whole exome sequencing (*71*, *72*). This suggests that variants in the 98% of the non-coding genome are highly likely to explain many of the unresolved DSD cases. This emphasises the need to identify the *cis*-regulatory elements that control the delicate, antagonistic gene networks at play during mammalian sex determination.

In this study, we aimed to decipher the *cis*-regulatory elements that participate in the process of gonadal supporting cell differentiation, the genes they regulate, and the transcriptomic outcome. To achieve this, we purified fetal pre-granulosa and Sertoli cells using the newly generated Enh8-mCherry and the established *Sox9*$^{IRES\text{-}GFP}$ mouse lines at four embryonic stages covering the process of sex fate decision and the subsequent supporting cell differentiation. One limitation of the use of the *Sox9*$^{IRES\text{-}GFP}$ to purify Sertoli cells is that we preferentially enriched our multiomics data in already committed Sertoli cells at E11.5, as shown by the low expression of *Sry* in our dataset compared to whole gonads (**Fig. S4C**). However, the use Enh8-mCherry and the *Sox9*$^{IRES\text{-}GFP}$ mouse lines allowed us to increase our cell sorting yield by five-fold compared to the previously used *TESCO*-CFP and *TESMS*-CFP transgenic mouse lines (*30*, *31*) (**Fig. S2D**). Moreover, the proportion of cells we obtained using these two mouse lines is close to what was observed in the whole gonads, according to scRNA-seq study (*9*). This suggests that our study is based on representative fetal gonad supporting cell populations.

In recent years, several studies have generated time-course transcriptomic profiles of mouse gonadal development using both whole gonad bulk RNA-seq (*43*) and scRNA-seq (*9*, *10*, *33*). While these studies have been transformative to the field and enabled the identification of novel and rare cell populations, they also hold some limitations. Bulk RNA-seq of whole gonads lacks the resolution needed to characterize gene expression within specific gonadal cells, a challenge that can be overcome by scRNA-seq technologies. However, scRNA-seq techniques are limited in sensitivity, making it difficult to detect low -abundant transcripts (*73*). Our bulk RNA-seq data represent the first transcriptomic analysis of purified Sertoli and pre-granulosa cells along embryonic fetal gonad development. Although averaging cells in

which differentiation is asynchrony, the sensitivity of bulk RNA-seq makes this method as a good complement to scRNA-seq data for lowly expressed genes.

Previous ATAC-seq analysis was performed on E10.5 supporting cell precursors and E13.5 Sertoli and pre-granulosa cells (*31*), leaving a big gap of the most critical stages in which the sex is being determined. Here, we performed RNA-seq, as well as ATAC-seq on four critical developmental stages covering the entire window of sex determination. Overall, we found that 2.8% of the 98% non-coding DNA are accessible during male and female supporting cell differentiation. Our analysis indicates that while many *cis*-regulatory elements exhibit sexual dimorphism in accessibility staring at E11.5 in Sertoli cells, most pre-granulosa-biased open chromatin regions are established from E12.5, reinforcing the idea that pre-granulosa cell commitment is delayed compared to Sertoli cells (*10*, *38*).

TF-binding motif enrichment analysis and footprinting indicate that most of the Sertoli-cell biased regions are bound by DMRT and the SOX factors. SOX9 has been described as being a pioneer TF that is able to bind close chromatin, inducing *cis*-regulatory element opening that leads to cell reprogramming in embryonic epidermal stem cells and umbilical vein endothelial cells (*74*, *75*). All these data confirm the importance of SOX9 as the master regulator for the onset of Sertoli cell fate decision by establishing the chromatin landscape necessary for the activation of the male-specific gene program. It was also shown that ectopic DMRT1 in the ovary acts as a pioneer factor to induce *Sox9* expression but also repress *Foxl2* (*76*, *77*). Interestingly, depletion of *Dmrt1* in males does not affect testicular development (*78*), which argues in favour of a role in the maintenance of Sertoli cell identity rather than a determining gene. However, its expression in E11.5 pre-granulosa cells and the enrichment of its binding motif in early pre-granulosa cells (when *Foxl2* is not yet expressed), can fit with its role to negatively regulate *Foxl2*, and, more globally, female-specific gene expression. This data may also potentially explain the delay in pre-granulosa cell differentiation compared to Sertoli cells.

Regarding pre-granulosa cells, we observed that RUNX1 and FOX motifs were not the most enriched in the pre-granulosa-biased open chromatin regions at the onset of their differentiation. This is consistent with the fact that deletion of these TF in fetal gonads leads to a delayed female-to-male sex reversal, either in postnatal ovaries when only *Foxl2* is deleted (*19*), or in late fetal stage, around E15.5, when both factors are deleted (*41*). As such, FOXL2 and RUNX1 are required for granulosa cell fate maintenance rather than being

female determining factors. However, our study sheds light on two factors that may act on pre-granulosa cell commitment: EMX2 and LHX9. While they have been shown to be crucial for the bipotential gonad formation (*35*, *36*), their role at later stages has never been characterized. The identification of the *Wt1*$^{-KTS}$ isoform as the female determining factor is a clear illustration that early factors can also have critical roles at later stages (*15*). *Emx2* and *Lhx9* expression become restricted to the pre-granulosa cells soon after the supporting cell commitment, from E11.5. Concomitantly, the pre-granulosa cell genome present increasing enrichment and binding footprints of the motif recognized by either of these two factors. Although our analysis cannot decipher which of these two factors is actually binding on the pre-granulosa putative *cis*-regulatory elements, EMX2 and LHX9 represent the best candidate TFs that could regulate the pre-granulosa cell differentiation genetic program. Moreover, EMX2 has been shown to play important roles in the mouse neuroblast proliferation, migration and differentiation (*79*), while LHX9 has not been identified in cell fate decision role in other systems.

Integration of both RNA and ATAC-seq data allowed for the linking of putative *cis*-regulatory elements to their target genes, allowing the identification of thousands of pre-granulosa- and Sertoli-specific putative *cis*-regulatory elements. Yet, it is likely that this is an understatement of the real state, as we limited the linkage analysis to +/- 500 kb, and critical enhancers like Enh13 (*Sox9*) and ZRS (*Shh*) are often located over 500 kb away from their target genes (*27*, *80*). Similar findings in other systems suggest a complex regulatory network of enhancers and silencers, many acting redundantly to control gene expression. While redundancy is common, single enhancers like Enh13 and ZRS can fully regulate target genes at specific stages, with deletions causing dramatic phenotypes.

Our PCHi-C analysis corroborated some of the *cis*-regulatory element-gene association predicted by the linkage analysis. Yet, while able to detect specific enhancer-promoter interactions, we could not detect the interaction of Enh13 to the *Sox9* promoter in Sertoli cells. This suggests that the PCHi-C, while revealing interesting interactions, is not comprehensive enough and needs to be further refined. Indeed, PCHi-C is normally being performed using millions of cells, and here we did it using 100K sorted gonadal cells. It is possible that using a larger number of cells could improve the resolution.

Altogether, this study serves as an important layer to unravel the gene regulatory networks that are at play during mammalian sex determination. It uncovers the *cis*-regulatory elements

that function during sex determination, along with the binding motifs enriched and occupied in them, the physical binding of regulatory elements and target genes, and how all this eventually translates into unique gene expression profiles. To complete our understanding of the gene regulatory networks, it will be important to have additional layers of information such as histone post-transcriptional modifications as well as binding of TFs and histone modifiers.

Materials and Methods

Mice and ethics

All animals were maintained with appropriate husbandry according to Bar-Ilan University ethics protocols 11-02-2020, 57-08-19, 2305-111-1 and 2306-117-1. All mouse strains were maintained on an F1 (C57BL6/J x CBA-Ca) genetic background. *Tg(Enh8-hsp68-mCherry* mice were generated at the Francis Crick Institute transgenic facility by zygote microinjection and imported to Israel.

The X-GFP (*Tg*(*CAG-EGFP*)*D4Nagy*) (*81*) and *Sox9*$^{IRES\text{-}GFP/+}$ (*44*) mice were used. Embryos and animals used in this study were either *Tg(Enh8-hsp68-mCherry)* heterozygotes for Cherry or *Sox9*$^{IRES\text{-}GFP}$ heterozygotes for GFP. Primers used for genotyping these mice strains are listed in **Table S1**.

Cloning of Enh8-hsp68-mCherry vector

Enh8 (mm9: 111,805,552-111,806,224; 672 bp long, located 838 kb upstream of the *Sox9* gene) was amplified by PCR and cloned into the AseI restriction sites of the pmCherry-N1 vector (Kan resistance) reporter vector (Takara Cat No. 632523) using In- Fusion HD (Clontech). The *hsp68* sequence was amplified from the pSfi-Hsp68-LacZ reporter vector (Addgene #33351) via In- Fusion HD and cloned into the AseI and AgeI restriction sites of the pmCherry-N1 vector. The primers used for the cloning are presented in **Table S1**. To release the entire Enh8-hsp68-mCherry construct, the AseI and NotI restriction enzymes were used. All plasmids were verified using Sanger sequencing.

Generation of the *Tg(Enh8-mCherry)* mice

To prepare DNA for zygote injection, 50 μg of the Enh8-hsp68-mCherry plasmid was digested with AseI and NotI and the insert was gel purified by electroelution. The DNA was phenol-chloroform extracted, ethanol precipitated and resuspended in TE buffer (10 mM Tris-HCl, 1 mM disodium EDTA, pH 8.0). The DNA was further purified on a DNA-cleanup column (Qiagen PCR Purification Kit). 5 ng/μl of the purified plasmid was injected into pronuclei of F1 (C57BL/6JxCBA) zygotes. These were transferred on the same day into the oviducts of pseudopregnant CD1 females. Injections were performed by the Crick Genetic Manipulation Service to generate stable transgenic lines. Cherry positive founders were bred

to wild type (C57BL/6JxCBA) F1 females to first examine expression profiles in the gonads of E13.5 embryos and then perform germline transmission. *Tg(Enh8-hsp68-mCherry)* line No. 3 was found to show robust expression in the gonads and was bred for a few generations to "clean" the line from extra transgenes until a stable line was established. This line is termed Enh8-mCherry (*Tg(Enh8-hsp68-mCherry))*.

gDNA isolation and transgenic mice genotyping

Genomic DNA (gDNA) was extracted from tail tissue of embryos or ear punch tissue of adult animals. gDNA isolation from adult earpiece included 15 min incubation at 95ºC with lysis buffer composed of 10 mM NaOH, 0.1 mM EDTA pH 8 followed by the addition of 40 mM Tris-HCl pH 5. For embryo samples, the PCRBIO rapid extract lysis kit was used (PCRBIO, PB15.11-S).

All PCR reactions were performed using 2X of Dream-Taq PCR mix (Thermo, K1082) according to the manufacturer's instructions. All mice and embryos were also genotyped for the chromosomal sex (Genotyping primers are listed in **Table S1**).

Time mating, gonad harvesting, tissue preparation and imaging

Embryos and animals carrying the desired transgene were produced by crossing a *Sox9*$^{IRES\text{-}GFP}$ male or a *Tg*(*CAG-EGFP*)*D4Nagy; Tg(Enh8-hsp68-mCherry)* male mice on F1 (C57BL/6J x CBA) genetic background female mice. Pregnant females were sacrificed by CO_2 inhalation and embryos were harvested at different stages of the pregnancy (E11.5, E12.5, E13.5 and E15.5). Day 0.5 was determined by the presence of a vaginal plug (VP). For E11.5 or E12.5 gonads, we considered the embryos that had between 18 and 21 tail somites (TS) as E11.5 and embryos that had 27-30 TS as E12.5. The sex of gonads at E13.5 and E15.5 was determined via visual inspection. The sex of the E11.5 and E12.5 *Tg*(*CAG-EGFP*)*D4Nagy; Tg(Enh8-hsp68-mCherry)* embryos (where the Cherry appears in both XX and XY) was determined by having the X-GFP transgene where GFP-positive embryos are XX and GFP-negative embryos are XY (*81*).

All bright field and fluorescent images of gonads were taken using the Nikon Eclipse Ts2R microscope. Exposure times: 10 ms for BF, 5-7s for mCherry, 2-3 seconds for GFP. Images were analysed using the NIS-Elements D software.

For immunostaining, gonads from embryos were harvested and fixed overnight in 4% paraformaldehyde (PFA) (Sigma Aldrich, P6148) in phosphate buffered saline (PBS) at 4ºC, washed three times with PBST (PBS with 0.1% Triton (Sigma Aldrich, 9002-93-1)) at room temperature, incubated with 20% sucrose (Fisher BioReagents, BP220-1) overnight at 4ºC and then embedded in OCT (Leica, 14020108926) and stored at -80ºC until further use.

Immunofluorescence staining

Immunofluorescence staining was performed on 10 µm-thick sagittal cryostat sections (Leica, CM3050-S). Antigen retrieval was performed with DAKO (Target retrieval solution, Agilent, S1699) at 65ºC for 30 min. Samples were then blocked in PBST containing 10% donkey serum (Sigma Aldrich, D9663) for 1hr and incubated with primary antibodies (diluted in PBST containing 1% donkey serum) overnight at 4ºC (All primary and secondary antibodies as well as dyes used are listed in **Table S2**). Following three PBST washes, secondary antibodies were added and incubated for 1hr at room temperature (RT). Slides were then washed, dried and mounted (Polysciences 18606-20). All immunofluorescence slides were also stained with 4 ,6-diamidino-2- phenylindole (DAPI; Invitrogen, D1306) to visualize nuclear DNA. Images were obtained with a Leica Microsystems SP8 confocal microscope.

FACS of Sertoli and pre-granulosa cells

Following dissection, gonads were separated from the mesonephros using a G25 needle and dissociated into single cells by culture in 0.045% trypsin- EDTA (Thermo, cat. 25300062) along with 0.25% collagenase (Wortington, cat. LS004176) in a total volume of 500 µl for 8 min at 37°C in a 24-well plate. 200 µl of DPBS (without Calcium and Magnesium) containing 3% Bovine serum albumin (BSA) (Sigma-Aldrich, cat. A3311) were added to quench the trypsin action and immediately aspirated to remove the trypsin. Gonads were then resuspended in 300µl of DPBS containing 3% BSA and mechanically dissociated by gentle pipetting. Cells were filtered through a 30µm strainer (Miltenyi Biotec, cat. 130-041-407) and collected into a Polypropylene FACS tube (BD Falcon, cat. 352063). The well and filter are washed with additional 200µl of DPBS containing 3% BSA. Cells were sorted using the BD FACS Aria III cell sorter with an 85-micron nozzle. Texas Red filter (excitation 561 nm, emission 610 nm) was used to detect mCherry and FITC (excitation 495 nm, emission 519 nm) filters to detect GFP. For ATAC sequencing, positive cells were sorted into Eppendorf tubes containing 150µl of DPBS + 3% BSA. For RNA sequencing positive cells

were collected into 150µl QIAzol lysis reagent (Qiagen, 79306). FACS raw data were analysed using FlowJo v10 for visualization (**Fig. S2A- C**).

RNA-seq and library prep

Total RNA was isolated from ~40,000 sorted Sertoli or pre-granulosa cells of E11.5, E12.5, E13.5 and E15.5 gonads via FACS. RNA purification was performed using the RNeasy plus micro kit (Qiagen, 74034) according to the manufacturer instructions. cDNA libraries were prepared using the NEB Next Single Cell/Low Input RNA Library Prep Kit for Illumina (NEB, E6420) according to the manufacturer's protocol. cDNAs were synthesized by reverse transcriptase enzyme, followed by cDNA amplification and cDNA clean-up using AMPure beads (Beckman coulter, Bc-a63881). cDNA was then fragmented, and adapters were ligated with unique indexes for each sample using the NEBNext® Multiplex Oligos for Illumina (NEB, E7335). cDNA and library concentrations were analysed using Qubit ds HS Assay Kit (Invitrogen, 2326054) and sample size distribution using Tapestation with high sensitivity D1000 tape (Agilent, 5067-5585). Samples were pooled together to create a 4nM library and sequenced with 83 bp single end (SE) reads on the Illumina Next-Seq 500 platform at the Bar-Ilan Sequencing Unit with roughly 25-30 million reads per sample.

RNA-seq mapping and quality controls

FastQ files were processed using nf-core/rnaseq pipeline v3.12.0 (*82*). Briefly, read quality controls were performed with FastQC. Sequencing adapters were removed with TrimGalore!. Reads were mapped on the mm10/GRCm38 reference genome from Gencode (M25) with STAR. Gene quantification as read counts and TPM (Transcript Per Million) was obtained using RSEM. Read count matrix was filtered to exclude lowly expressed genes (genes with less than 15 reads and/or TPM value less than 5). Non-protein-coding genes were filtered out in the subsequent analysis. Sample correlation (Spearman) and PCA were performed with R (corr and prcomp functions) using the filtered read counts normalized by library size (sizeFactor) with DESeq2 (*83*). After inspection of the correlation and the PCA, we excluded the XX E11.5 replicate 2 because it was more similar to XX E12.5 samples than the E11.5 samples. This is probably due to embryos that were older than expected the day of the collection (**Fig. S3A- B**).

Differential expression analysis

Sex differential expression analysis was performed on the filtered read count matrix using DESeq2 with ~Sex_stage as model and using the “Wald” test. The number of sexually dimorphic genes per stage was retrieved using the contrasts. Genes with log2(FoldChange) < -0.5 and log2(FoldChange) > 0.5, as well as an adjusted *p*-value < 0.01 were considered as differentially expressed.

Stage differential expression analysis was performed on each sex separately using DESeq2 with ~Stage as model and using “LRT” (Likelihood Ratio Test). As previously, genes with log2(FoldChange) < -0.5 and log2(FoldChange) > 0.5, as well as an adjusted *p*-value < 0.01 were considered as differentially accessible. The overlap of the sex differentially expressed genes across stages was computed and represented as an upset plot using eulerr and UpsetR packages (*84*, *85*).

Gene expression values (filtered read count matrix normalized by library size) from the stage differentially expressed genes were transformed as z-scores, clustered into groups according to their expression profiles using hclust and "ward.D2" method, and represented as heatmap using ComplexeHeatmap (*34*). The optimal number of clusters were assessed using the best.cutree function from the Jlutils R package (https://github.com/larmarange/JLutils).

The overlap between the sex and the stage differentially expressed genes was computed and represented as a venn diagram using the eulerr package. The gene expression visualisation website was developed using Shiny and Plotly R packages.

GO term enrichment analysis

Gene Ontology enrichment analyses were performed using ClusterProfiler (*86*). Biological process GO terms with an enrichment *p*-value < 0.01 and *q*-value < 0.05 were reduced by similarity using the “simplify” function with a similarity cutoff of 0.7.

TF and phenotype annotation

Differentially expressed genes were annotated whether they are transcription factors and whether a mutated allele in mice induced a gonadal or fertility-related phenotype.

The list of mouse transcription factors was taken from https://resources.aertslab.org/cistarget/tf_lists/allTFs_mm.txt.

Gonadal phenotype annotation was constituted using the Mouse Phenotype OBO database v1.2 (Open Biological and Biomedical Ontology) from MGI (Mouse Genome Informatics) (https://www.informatics.jax.org/downloads/reports/Mpheno_OBO.ontology). Phenotype names containing the patterns "gonad*", "testi*", "ovar*", "fertility", and "sex*" were selected. Genes associated with the selected phenotype IDs were retrieved from the gene-phenotype database from MGI (https://www.informatics.jax.org/downloads/reports/MGI_GenePheno.rpt).

The list of the 1,860 mouse transcription factors and the 22,215 genes associated with a gonadal phenotype is provided in the **data file S3**.

ATAC-seq, library prep and sequencing

Sorted Sertoli and pre-granulosa cells (~60,000) were isolated from E11.5, E12.5, E13.5 and E15.5 gonads via FACS. ATAC-seq library preparation was performed using the Active Motif kit (Active Motif, cat. 53150) according to the manufacturer instructions. Sample concentration was measured using Qubit ds HS Assay Kit (Invitrogen, 2326054) and sample distribution size using Tapestation with high sensitivity D1000 tape (Agilent, 5067-5585). Samples were then pooled together accordingly into a 4nM library and sequenced with 37 bp paired end (PE) reads on the Illumina Next-Seq 500 platform at the Bar-Ilan Sequencing Unit with 25-40 million reads per sample.

ATAC-seq mapping, peak calling and quality controls

FastQ files were processed using nf-core/atacseq pipeline v2.1.2 (*87*). Briefly, read quality controls were performed with FastQC. Sequencing adaptors were removed with Cutadapt. Reads were mapped on the mm10/GRCm38 reference genome from Gencode (M25) with BWA. Mapped reads were filtered to remove the unpaired reads, the mitochondrial reads, the duplicated reads, the multimapped reads, the fragments with insert size > 2 kb, and the reads mapping to the blacklisted regions (https://github.com/Boyle-Lab/Blacklist/blob/master/lists/mm10-blacklist.v2.bed.gz). Blacklist region file has been modified to allow peak calling on the Y chromosome to inspect the open chromatin regions around the *Sry* gene.

Peaks were called with MACS2 using the "narrow_peaks" parameter. For each condition, peaks with FDR<0.01 and found in at least two replicates were merged as consensus open

chromatin regions. The obtained consensus regions were then combined and merged to constitute the set of non-overlapping open chromatin regions present in any conditions and was used for the downstream analysis. Read quantification was performed on the obtained open chromatin regions using featureCounts. Regions from which the maximum coverage is fewer than 50 reads across all samples were discarded from the read quantification matrix for the subsequent analysis.

Bigwig files normalized by million mapped reads from the nf-core/atacseq pipeline were corrected to be normalized by the size factors calculated from reads in peaks from DESeq2 in order to make the peak height reflecting the downstream analysis.

Consensus peak annotation for each condition was assessed using ChIPseeker (*88*). Sample correlation (Spearman) and PCA was performed using consensus region read counts normalized with DESeq2 with VST (Variance Stabilizing Transformation).

Differential chromatin accessibility analysis

Differential chromatin accessibility analysis was performed similarly to the differential expression analysis. Sex differential accessibility analysis was performed on the filtered read quantification matrix using DESeq2 with ~Sex_stage as model and using the “Wald” test, while stage-specific differential accessibility analysis was performed using the “LTR” test with ~Stage as model. In both analyses, regions with log2(FoldChange) < -1 and log2(FoldChange) > 1, as well as an adjusted *p*-value < 0.01 were considered as differentially accessible.

Accessibility values (filtered read count matrix normalized by library size) from the stage differentially accessible regions were transformed as z-scores, clustered into groups according to their accessibility profiles using hclust and "ward.D2" method, and represented as heatmap using ComplexeHeatmap. The optimal number of clusters were assessed by visual inspection.

Open chromatin region and gene expression correlations

Link between open chromatin region and gene expression was calculated using the same method used in the cisDynet R package (*89*). Briefly, for each protein-coding gene, we selected the open chromatin regions within 500 kb upstream and downstream of the canonical TSS of the gene as potential regulatory regions. Then we calculated the Pearson

correlation coefficients between the gene expression and the chromatin accessibility of each region (all stages and sexes together). To control for false positives, we selected 10,000 random open chromatin regions and calculated the correlation coefficients between the regions and expression. We then tested the significance of each region-to-gene correlation using the *Z*-test method and obtained a *p*-value that is corrected for false discovery rate (FDR). Finally, we considered a significant region-to-gene link with a Pearson correlation coefficient >0.7 (positive link) and <0.7 (negative link) and a FDR <0.01. We excluded open chromatin regions overlapping TSS of alternative transcripts from our analysis to avoid links between genes and their alternative promoters. We visualized the region-to-gene links using Gviz and GenomicInteraction packages (*90*, *91*).

Promoter Capture Hi-C

Capture Hi-C libraries (3 libraries of E13.5 *Sox9*$^{IRES\text{-}GFP}$ Sertoli cells and 2 libraries of E13.5 Enh8-mCherry pre-granulosa cells) were generated from 100,000 sorted cells, crosslinked as previously described (*92*, *93*). Following lysis, the nuclei were permeabilized and digested with DpnII (NEB) overnight. The restriction overhangs were filled in using a biotinylated dATP (Jena Bioscience) and ligation was performed for 4 hours at 16°C (T4 DNA ligase; Life Technologies). The crosslinks were reversed using proteinase K and overnight incubation at 65°C, followed by purification with SPRI beads (AMPure XP; Beckman Coulter). Short fragments up to 1000 bp were produced via tagmentation and the biotinylated restriction junctions were then pulled down using MyOne C1 streptavidin beads (Life Technologies). PCR amplification (5 cycles) was performed on the libraries directly bound to the C-1 beads and the libraries were purified using SPRI beads as before. Promoter Capture was performed using the mouse custom-designed Agilent SureSelect system, following the manufacturer's protocol, followed by 7 PCR cycles. The libraries were sequenced using 150 bp paired-end sequencing on an Illumina NovaSeq (Novogene UK) with a sequencing depth of 1bln paired reads for each biological replicate.

PCHi-C data processing and detection of significant contacts

PCHi-C reads from biological replicates were merged using *cat* before processing further. Read processing, alignment and filtering was performed using a modified version of the Hi-C User Pipeline (HiCUP) v0.7.4 (*94*), HiCUP Combinations (https://github.com/StevenWingett/HiCUP/tree/combinations), which first creates all possible

combinations of ditags from the paired reads prior to mapping, then performs standard filtering for common Hi-C artefacts. At this point, alignment files for technical replicates were merged per sample. To generate the consensus dataset, CHiCAGO input (.chinput) files were generated per replicate, using *bam2chicago.sh* from chicagoTools v1.13 (*67*). CHiCAGO was run at 5 kb bins resolution, on two modalities. One where we only considered interactions with the original bait (termed "original bait") and another where the bait bin was expanded to a 5 kb bin resolution (termed "5 kb extended bait"), as we have previously described (*67*, *93*). We performed tests for enrichment at promoter interacting regions in both modes, using the *peakEnrichment4Features* function in ChiCAGO. ChIP-seq data for H3K4me3, H3K27ac and H3H27me3 from E13.5 pre-granulosa and Sertoli cells was obtained from GSE118755 and GSE130749) and reanalysed using the nf-core/chipseq v2.0.0 (*95*) as previously described (*63*).

TFBS motif analysis

Differential transcription factor motif enrichment analysis was performed with monaLisa (*96*) using vertebrate TFBS matrices of transcription factors from JAPSAR2024 (*97*) extracted using TFBSTools (*98*).

TFBS motifs differentially enriched in the sexually dimorphic accessible regions were analysed by merging the sexually dimorphic regions from all stages and by binning them by sex-specific accessibility. TFBS motifs differentially enriched in the dynamically accessible regions were analysed by binning the region by their cluster of dynamics profile.

Enrichment test was run against the bins as well as against a random genomic background (corrected for the GC content of the tested regions) to avoid GC content biases.

TFBS motifs from transcription factors not expressed in the Sertoli and pre-granulosa cell samples (TPM<10) were filtered out.

TFBS motifs were clustered by enrichment score (log2(enrichment)) using hclust (“ward.D” method). For each enrichment cluster, TFBS motifs were grouped by similarity using motifStack (*99*) with cutoffPval = 0.001. The result is visualized as heatmaps of the log2(enrichment) showing the logos and the names of the merged TFBSs using ComplexeHeatmap (*100*).

ATAC-seq footprint analysis

ATAC-seq footprint analysis was performed using TOBIAS v0.17.0 (*69*). We merged the bam files from the different replicates prior to the analysis. We ran ATACorrect and ScoreBigwig commands with default parameters. We then ran BINDetect on the combined sexually dimorphic regions from all stages and using TFBS motifs merged by similarity using motifStack as described above. Differential binding scores between pre-granulosa and Sertoli at each developmental stage cells were plot using ggplot2.

Single-cell RNA-seq expression profiles

Single-cell RNA-seq data were obtained from GSE184708 (*9*). Data were loaded in Seurat v5.1.0 (*101*) using the gene, barcode, expression matrix and metadata files provided on GEO. Data was log-normalized and gene expression plotted using the VlnPlot function on selected cell types.

Acknowledgements

We are grateful to the BIU Life Sciences Kanbar Core Facility unit of NGS, Flow Cytometry and Microscopy units. We are also grateful to the Animal Research Facility at BIU. We are thankful to the Genetic Modification Service of the Francis Crick Institute, and to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing the computing resources necessary for the current study. We thank members of our lab for advice, support, and helpful comments.

Funding

This study was funded by the European Union (ERC, EnhanceSex, 101039928), and the Israel Science Foundation (ISF No.710_2020). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. VM was funded by VIB internal core funding. RLB is funded by the Francis Crick Institute which receives its core funding from Cancer Research UK (CC2116), the UK Medical Research Council (CC2116), and the Wellcome Trust (CC2116). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contribution

IS, EA, MR and NG conceived the experiments; MR, EA and LS performed experiments and IS analysed the data. RW developed the RNA-seq visualization platform. CRF and DM cloned the Enh8-mCherry vector. The Enh8-mCherry line was generated at the Crick in the lab of RLB. VM performed and analysed the PCHi-C. The manuscript was written by IS, EA, MR and NG. All authors have read and accepted the data being presented in the manuscript.

Competing declaration

The authors declare that they have no competing interests.

Data and materials availability

All data is available in the manuscript or the supplementary materials. RNA-seq, ATAC-seq and PCHi-C data have been deposited in the Gene Expression Omnibus under accession number GSE277647 and GSE277646 and GSEXXXXXX, respectively. The data is only accessible to reviewers at the moment. The code produced to analyze the data is available on GitHub: https://github.com/Istevant/SupportingCRE.

References

1. N. H. Uhlenhaut, S. Jakob, K. Anlag, T. Eisenberger, R. Sekido, J. Kress, A. C. Treier, C. Klugmann, C. Klasen, N. I. Holter, D. Riethmacher, G. Schütz, A. J. Cooney, R. Lovell-Badge, M. Treier, Somatic Sex Reprogramming of Adult Ovaries to Testes by FOXL2 Ablation. *Cell* **139**, 1130–1142 (2009).

2. C. K. Matson, M. W. Murphy, A. L. Sarver, M. D. Griswold, V. J. Bardwell, D. Zarkower, DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* **476**, 101–4 (2011).

3. R. Migale, M. Neumann, R. Mitter, M.-R. Rafiee, S. Wood, J. Olsen, R. Lovell-Badge, FOXL2 interaction with different binding partners regulates the dynamics of ovarian development. *Sci. Adv.* **10**, eadl0788 (2024).

4. J. Karl, B. Capel, Sertoli Cells of the Mouse Testis Originate from the Coelomic Epithelium. *Dev. Biol.* **203**, 323–333 (1998).

5. D. M. Maatouk, B. Capel, Sexual development of the soma in the mouse. *Curr. Top. Dev. Biol.* **83**, 151–183 (2008).

6. S. Nef, I. Stévant, A. Greenfield, "Characterizing the bipotential mammalian gonad" in *Current Topics in Developmental Biology* (Academic Press, 2019; https://pubmed.ncbi.nlm.nih.gov/30999975/)vol. 134, pp. 167–194.

7. D. Wilhelm, S. Palmer, P. Koopman, Sex determination and gonadal development in mammals. *Physiol. Rev.* **87**, 1–28 (2007).

8. H. Ademi, C. Djari, C. Mayère, Y. Neirijnck, P. Sararols, C. M. Rands, I. Stévant, B. Conne, S. Nef, Deciphering the origins and fates of steroidogenic lineages in the mouse testis. *Cell Rep.* **39** (2022).

9. C. Mayère, V. Regard, A. Perea-Gomez, C. Bunce, Y. Neirijnck, C. Djari, P. Sararols, R. Reeves, S. Greenaway, M. Simon, others, Origin, specification and differentiation of a rare supporting-like lineage in the developing mouse gonad. (2021).

10. I. Stévant, F. Kühne, A. Greenfield, M.-C. C. Chaboissier, E. T. Dermitzakis, S. Nef, Dissecting Cell Lineage Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell Transcriptomics. *Cell Rep.* **26**, 3272-3283.e3 (2019).

11. P. Koopman, A. Münsterberg, B. Capel, N. Vivian, R. Lovell-Badge, Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* **348**, 450–452 (1990).

12. A. H. Sinclair, P. Berta, M. S. Palmer, J. R. Hawkins, B. L. Griffiths, M. J. Smith, J. W. Foster, A.-M. Frischauf, R. Lovell-Badge, P. N. Goodfellow, A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240–244 (1990).

13. I. B. Barsoum, H. H.-C. H.-C. H.-C. H.-C. H.-C. Yao, Fetal Leydig Cells: Progenitor Cell Maintenance and Differentiation. *J. Androl.* **31**, 11–15 (2010).

14. T. Svingen, P. Koopman, Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. *Genes Dev.* **27**, 2409–26 (2013).

15. E. P. Gregoire, M.-C. De Cian, R. Migale, A. Perea-Gomez, S. Schaub, N. Bellido-Carreras, I. Stévant, C. Mayère, Y. Neirijnck, A. Loubat, P. Rivaud, M. L. Sopena, S. Lachambre, M. M. Linssen, P. Hohenstein, R. Lovell-Badge, S. Nef, F. Chalmel, A. Schedl, M.-C. Chaboissier, The −KTS splice variant of WT1 is essential for ovarian determination in mice. *Science* **382**, 600–606 (2023).

16. B. Nicol, M. A. Estermann, H. H.-C. Yao, N. Mellouk, Becoming female: Ovarian differentiation from an evolutionary perspective. *Front. Cell Dev. Biol.* **10**, 944776 (2022).

17. A.-A. Chassot, S. T. Bradford, A. Auguste, E. P. Gregoire, E. Pailhoux, D. G. de Rooij, A. Schedl, M.-C. Chaboissier, WNT4 and RSPO1 together are required for cell proliferation in the early mouse gonad. *Dev. Camb. Engl.* **139**, 4461–72 (2012).

18. A.-A. Chassot, F. Ranc, E. P. Gregoire, H. L. Roepers-Gajadien, M. M. Taketo, G. Camerino, D. G. de Rooij, A. Schedl, M.-C. Chaboissier, Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary. *Hum. Mol. Genet.* **17**, 1264–77 (2008).

19. M. Uda, C. Ottolenghi, L. Crisponi, J. E. Garcia, M. Deiana, W. Kimber, A. Forabosco, A. Cao, D. Schlessinger, G. Pilia, Foxl2 disruption causes mouse ovarian failure by pervasive blockage of follicle development. *Hum. Mol. Genet.* **13**, 1171–1181 (2004).

20. C. Ottolenghi, S. Omari, J. E. Garcia-Ortiz, M. Uda, L. Crisponi, A. Forabosco, G. Pilia, D. Schlessinger, Foxl2 is required for commitment to ovary differentiation. *Hum. Mol. Genet.* **14**, 2053–2062 (2005).

21. R. Hiramatsu, S. Matoba, M. Kanai-Azuma, N. Tsunekawa, Y. Katoh-Fukui, M. Kurohmaru, K.-I. Morohashi, D. Wilhelm, P. Koopman, Y. Kanai, A critical time window of Sry action in gonadal sex determination in mice. *Dev. Camb. Engl.* **136**, 129–38 (2009).

22. D. M. Maatouk, L. Dinapoli, A. Alvers, K. L. Parker, M. M. Taketo, B. Capel, Stabilization of β-catenin in XY gonads causes male-to-female sex-reversal. *Hum. Mol. Genet.* **17**, 2949–2955 (2008).

23. C. C. Bolt, D. Duboule, The regulatory landscapes of developmental genes. *Dev. Camb. Engl.* **147**, dev171736 (2020).

24. S. Chatterjee, N. Ahituv, Gene Regulatory Elements, Major Drivers of Human Disease. *Annu. Rev. Genomics Hum. Genet.* **18**, 45–63 (2017).

25. H. K. Long, S. L. Prescott, J. Wysocka, Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).

26. B. Croft, T. Ohnesorg, J. Hewitt, J. Bowles, A. Quinn, J. Tan, V. Corbin, E. Pelosi, J. van den Bergen, R. Sreenivasan, I. Knarston, G. Robevska, D. C. Vu, J. Hutson, V. Harley, K. Ayers, P. Koopman, A. Sinclair, Human sex reversal is caused by duplication or deletion of core enhancers upstream of SOX9. *Nat. Commun.* **9**, 5319 (2018).

27. N. Gonen, C. R. Futtner, S. Wood, S. A. Garcia-Moreno, I. M. Salamone, S. C. Samson, R. Sekido, F. Poulat, D. M. Maatouk, R. Lovell-Badge, Sex reversal following deletion of a single distal enhancer of Sox9. *Science* **360**, 1469–1473 (2018).

28. Y. Ogawa, M. Terao, S. Hara, M. Tamano, H. Okayasu, T. Kato, S. Takada, Mapping of a responsible region for sex reversal upstream of Sox9 by production of mice with serial deletion in a genomic locus. *Sci. Rep.* **8**, 17514 (2018).

29. M. Ridnik, E. Abberbock, V. Alipov, S. Z. Lhermann, S. Kaufman, M. Lubman, F. Poulat, N. Gonen, Two redundant transcription factor binding sites in a single enhancer are essential for mammalian sex determination. *Nucleic Acids Res.* **52**, 5514–5528 (2024).

30. D. M. Maatouk, A. Natarajan, Y. Shibata, L. Song, G. E. Crawford, U. Ohler, B. Capel, Genome-wide identification of regulatory elements in Sertoli cells. *Development* **144**, 720–730 (2017).

31. S. A. Garcia-Moreno, C. R. Futtner, I. M. Salamone, N. Gonen, R. Lovell-Badge, D. M. Maatouk, Gonadal supporting cells acquire sex-specific chromatin landscapes during mammalian sex determination. *Dev. Biol.* **446**, 168–179 (2019).

32. R. Sekido, R. Lovell-Badge, Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**, 930–934 (2008).

33. I. Stévant, Y. Neirijnck, C. Borel, J. Escoffier, L. B. Smith, S. E. Antonarakis, E. T. Dermitzakis, S. Nef, Y. Neirjinck, C. Borel, J. Escoffier, L. B. Smith, S. E. Antonarakis, E. T. Dermitzakis, S. Nef, Y. Neirijnck, C. Borel, J. Escoffier, L. B. Smith, S. E. Antonarakis, E. T. Dermitzakis, S. Nef, Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing. *Cell Rep.* **22**, 1589–1599 (2018).

34. I. Stévant, S. Nef, Genetic Control of Gonadal Sex Determination and Development. *Trends Genet.* **35**, 346–358 (2019).

35. N. Miyamoto, M. Yoshida, S. Kuratani, I. Matsuo, S. Aizawa, Defects of urogenital development in mice lacking Emx2. *Dev. Camb. Engl.* **124**, 1653–64 (1997).

36. O. S. Birk, D. E. Casiano, C. A. Wassif, T. Cogliati, L. Zhao, Y. Zhao, A. Grinberg, S. Huang, J. A. Kreidberg, K. L. Parker, F. D. Porter, H. Westphal, The LIM homeobox gene Lhx9 is essential for mouse gonad formation. *Nature* **403**, 909–13 (2000).

37. K. H. Albrecht, E. M. Eicher, Evidence That Sry Is Expressed in Pre-Sertoli Cells and Sertoli and Granulosa Cells Have a Common Precursor. *Dev. Biol.* **240**, 92–107 (2001).

38. S. A. Jameson, A. Natarajan, J. Cool, T. DeFalco, D. M. Maatouk, L. Mork, S. C. Munger, B. Capel, Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS Genet.* **8**, e1002575 (2012).

39. A. Georges, D. L'Hôte, A. L. Todeschini, A. Auguste, B. Legois, A. Zider, R. A. Veitia, The transcription factor FOXL2 mobilizes estrogen signaling to maintain the identity of ovarian granulosa cells. *eLife* **3**, e04207 (2014).

40. B. Nicol, S. A. Grimm, A. Gruzdev, G. J. Scott, M. K. Ray, H. H. C. Yao, Genome-wide identification of FOXL2 binding and characterization of FOXL2 feminizing action in the fetal gonads. *Hum. Mol. Genet.* **27**, 4273–4287 (2018).

41. B. Nicol, S. A. Grimm, F. Chalmel, E. Lecluze, M. Pannetier, E. Pailhoux, E. Dupin-De-Beyssat, Y. Guiguen, B. Capel, H. H. C. Yao, RUNX1 maintains the identity of the fetal ovary through an interplay with FOXL2. *Nat. Commun. 2019 101* **10**, 1–14 (2019).

42. M. Rossitto, S. Déjardin, C. M. Rands, S. Le Gras, R. Migale, M. R. Rafiee, Y. Neirijnck, A. Pruvost, A. L. Nguyen, G. Bossis, F. Cammas, L. Le Gallic, D. Wilhelm, R. Lovell-Badge, B. Boizet-Bonhoure, S. Nef, F. Poulat, TRIM28-dependent SUMOylation protects the adult ovary from activation of the testicular pathway. *Nat. Commun. 2022 131* **13**, 1–19 (2022).

43. L. Zhao, C. Wang, M. L. Lehman, M. He, J. An, T. Svingen, C. M. Spiller, E. T. Ng, C. C. Nelson, P. Koopman, Transcriptomic analysis of mRNA expression and alternative splicing during mouse sex determination. *Mol. Cell. Endocrinol.* **478**, 84–96 (2018).

44. L. Nel-Themaat, T. J. Vadakkan, Y. Wang, M. E. Dickinson, H. Akiyama, R. R. Behringer, Morphometric analysis of testis cord formation in sox9-eGFP Mice. *Dev. Dyn.* **238**, 1100–1110 (2009).

45. O. Habara, C. Y. Logan, M. Kanai-Azuma, R. Nusse, H. M. Takase, WNT signaling in pre-granulosa cells is required for ovarian folliculogenesis and female fertility. *Dev. Camb. Engl.* **148**, dev198846 (2021).

46. J. Schmahl, Fgf9 induces proliferation and nuclear localization of FGFR2 in Sertoli precursors during male sex determination. *Development* **131**, 3627–3636 (2004).

47. J. Schmahl, E. M. Eicher, L. L. Washburn, B. Capel, Sry induces cell proliferation in the mouse gonad. *Dev. Camb. Engl.* **127**, 65–73 (2000).

48. R. M. Baldarelli, C. L. Smith, M. Ringwald, J. E. Richardson, C. J. Bult, Mouse Genome Informatics Group, Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* **227**, iyae031 (2024).

49. Q. Wang, Y. Lan, E.-S. Cho, K. M. Maltby, R. Jiang, Odd-skipped related 1 (Odd1) is an essential regulator of heart and urogenital development. *Dev. Biol.* **288**, 10.1016/j.ydbio.2005.09.024 (2005).

50. T. Groza, F. L. Gomez, H. H. Mashhadi, V. Muñoz-Fuentes, O. Gunes, R. Wilson, P. Cacheiro, A. Frost, P. Keskivali-Bond, B. Vardal, A. McCoy, T. K. Cheng, L. Santos, S. Wells, D. Smedley, A.-M. Mallon, H. Parkinson, The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **51**, D1038–D1045 (2023).

51. M. Fedele, R. Franco, G. Salvatore, M. P. Paronetto, F. Barbagallo, R. Pero, L. Chiariotti, C. Sette, D. Tramontano, G. Chieffi, A. Fusco, P. Chieffi, PATZ1 gene has a critical role in the spermatogenesis and testicular tumours. *J. Pathol.* **215**, 39–47 (2008).

52. J. Jin, K. Li, Y. Du, F. Gao, Z. Wang, W. Li, Multi-omics study identifies that PICK1 deficiency causes male infertility by inhibiting vesicle trafficking in Sertoli cells. *Reprod. Biol. Endocrinol.* **21**, 114 (2023).

53. J. Liu, J. F. Schiltz, H. R. Ashar, K. K. Chada, Hmga1 is required for normal sperm development. *Mol. Reprod. Dev.* **66**, 81–89 (2003).

54. L. M. Ludbrook, V. R. Harley, Sex determination: a “window” of DAX1 activity. *Trends Endocrinol. Metab. TEM* **15**, 116–121 (2004).

55. A. H. Payne, G. L. Youngblood, L. Sha, M. Burgos-Trinidad, S. H. Hammond, Hormonal regulation of steroidogenic enzyme gene expression in Leydig cells. *J. Steroid Biochem. Mol. Biol.* **43**, 895–906 (1992).

56. Y.-C. Hu, L. M. Okumura, D. C. Page, Gata4 is required for formation of the genital ridge in mice. *PLoS Genet.* **9**, e1003629 (2013).

57. N. L. Manuylov, B. Zhou, Q. Ma, S. C. Fox, W. T. Pu, S. G. Tevosian, Conditional ablation of Gata4 and Fog2 genes in mice reveals their distinct roles in mammalian sexual differentiation. *Dev. Biol.* **353**, 229–241 (2011).

58. Z. Chen, Y. Zhang, Role of Mammalian DNA Methyltransferases in Development. *Annu. Rev. Biochem.* **89**, 135–158 (2020).

59. M. Anttonen, I. Ketola, H. Parviainen, A.-K. Pusa, M. Heikinheimo, FOG-2 and GATA-4 Are coexpressed in the mouse ovary and can modulate mullerian-inhibiting substance expression. *Biol. Reprod.* **68**, 1333–1340 (2003).

60. S. G. Tevosian, K. H. Albrecht, J. D. Crispino, Y. Fujiwara, E. M. Eicher, S. H. Orkin, Gonadal differentiation, sex determination and normal Sry expression in mice require direct interaction between transcription partners GATA4 and FOG2. *Dev. Camb. Engl.* **129**, 4627–34 (2002).

61. I. Manosalva, A. González, R. Kageyama, Hes1 in the somatic cells of the murine ovary is necessary for oocyte survival and maturation. *Dev. Biol.* **375**, 140–151 (2013).

62. R. S. Johnson, B. M. Spiegelman, V. Papaioannou, Pleiotropic effects of a null mutation in the c-fos proto-oncogene. *Cell* **71**, 577–586 (1992).

63. I. Stévant, N. Gonen, F. Poulat, Transposable elements acquire time- and sex-specific transcriptional and epigenetic signatures along mouse fetal gonad development. *Front. Cell Dev. Biol.* **11** (2024).

64. I. Mota-Gómez, J. A. Rodríguez, S. Dupont, O. Lao, J. Jedamzick, R. Kuhn, S. Lacadie, S. A. García-Moreno, A. Hurtado, R. D. Acemel, B. Capel, M. A. Marti-Renom, D. G. Lupiáñez, Sex-determining 3D regulatory hubs revealed by genome spatial auto-correlation analysis. *bioRxiv*, 2022.11.18.516861 (2022).

65. P. Freire-Pritchett, H. Ray-Jones, M. Della Rosa, C. Q. Eijsbouts, W. R. Orchard, S. W. Wingett, C. Wallace, J. Cairns, M. Spivakov, V. Malysheva, Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools. *Nat. Protoc.* **16**, 4144–4176 (2021).

66. S. Schoenfelder, M. Furlan-Magaril, B. Mifsud, F. Tavares-Cadete, R. Sugar, B.-M. Javierre, T. Nagano, Y. Katsman, M. Sakthidevi, S. W. Wingett, E. Dimitrova, A. Dimond, L. B. Edelman, S. Elderkin, K. Tabbada, E. Darbo, S. Andrews, B. Herman, A. Higgs, E. LeProust, C. S. Osborne, J. A. Mitchell, N. M. Luscombe, P. Fraser, The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).

67. J. Cairns, P. Freire-Pritchett, S. W. Wingett, C. Várnai, A. Dimond, V. Plagnol, D. Zerbino, S. Schoenfelder, B.-M. Javierre, C. Osborne, P. Fraser, M. Spivakov, CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).

68. M. Rahmoun, R. Lavery, S. Laurent-Chaballier, N. Bellora, G. K. Philip, M. Rossitto, A. Symon, E. Pailhoux, F. Cammas, J. Chung, S. Bagheri-Fam, M. Murphy, V. Bardwell, D. Zarkower, B. Boizet-Bonhoure, P. Clair, V. R. Harley, F. Poulat, In mammalian foetal testes, SOX9 regulates expression of its target genes by binding to genomic regions with conserved signatures. *Nucleic Acids Res.* **45**, 7191–7211 (2017).

69. M. Bentsen, P. Goymann, H. Schultheis, K. Klee, A. Petrova, R. Wiegandt, A. Fust, J. Preussner, C. Kuenne, T. Braun, J. Kim, M. Looso, ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).

70. F. J. Barrionuevo, A. Hurtado, G.-J. Kim, F. M. Real, M. Bakkali, J. L. Kopp, M. Sander, G. Scherer, M. Burgos, R. Jiménez, Sox9 and Sox8 protect the adult testis from male-to-female genetic reprogramming and complete degeneration. *eLife* **5** (2016).

71. E. C. Délot, E. Vilain, Towards improved genetic diagnosis of human differences of sex development. *Nat. Rev. Genet.* **22**, 588–602 (2021).

72. S. Eggers, T. Ohnesorg, A. Sinclair, Genetic regulation of mammalian gonad development. *Nat. Rev. Endocrinol.* **10**, 673–683 (2014).

73. T. M. Yamawaki, D. R. Lu, D. C. Ellwanger, D. Bhatt, P. Manzanillo, V. Arias, H. Zhou, O. K. Yoon, O. Homann, S. Wang, C.-M. Li, Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling. *BMC Genomics* **22**, 66 (2021).

74. B. M. Fuglerud, S. Drissler, J. Lotto, T. L. Stephan, A. Thakur, R. Cullum, P. A. Hoodless, SOX9 reprograms endothelial cells by altering the chromatin landscape. *Nucleic Acids Res.* **50**, 8547–8565 (2022).

75. Y. Yang, N. Gomez, N. Infarinato, R. C. Adam, M. Sribour, I. Baek, M. Laurin, E. Fuchs, The pioneer factor SOX9 competes for epigenetic factors to switch stem cell fates. *Nat. Cell Biol.* **25**, 1185–1195 (2023).

76. R. E. Lindeman, M. D. Gearhart, A. Minkina, A. D. Krentz, V. J. Bardwell, D. Zarkower, Sexual cell-fate reprogramming in the ovary by DMRT1. *Curr. Biol.* **25**, 764–771 (2015).

77. R. E. Lindeman, M. W. Murphy, K. S. Agrimson, R. L. Gewiss, V. J. Bardwell, M. D. Gearhart, D. Zarkower, The conserved sex regulator DMRT1 recruits SOX9 in sexual cell fate reprogramming. *Nucleic Acids Res.* **49**, 6144–6164 (2021).

78. C. S. Raymond, M. W. Murphy, M. G. O'Sullivan, V. J. Bardwell, D. Zarkower, Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14**, 2587–2595 (2000).

79. C. Cecchi, Emx2: a gene responsible for cortical development, regionalization and area specification. *Gene* **291**, 1–9 (2002).

80. F. Lim, J. J. Solvason, G. E. Ryan, S. H. Le, G. A. Jindal, P. Steffen, S. K. Jandu, E. K. Farley, Affinity-optimizing enhancer variants disrupt development. *Nature* **626**, 151–159 (2024).

81. A. K. Hadjantonakis, M. Gertsenstein, M. Ikawa, M. Okabe, A. Nagy, Non-invasive sexing of preimplantation stage mammalian embryos. *Nat. Genet.* **19**, 220–222 (1998).

82. H. Patel, P. Ewels, A. Peltzer, O. Botvinnik, G. Sturm, D. Moreno, P. Vemuri, M. U. Garcia, silviamorins, L. Pantano, M. Binzer-Panchal, nf-core bot, R. Syme, M. Zepper, G. Kelly, F. Hanssen, J. A. F. Yates, C. Cheshire, rfenouil, J. Espinosa-Carrasco, marchoeppner, E. Miller, A. Talbot, P. Zhou, S. Guinchard, M. Hörtenhuber, G. Gabernet, C. Mertes, D. Straub, P. D. Tommaso, nf-core/rnaseq: nf-core/rnaseq v3.12.0 - Osmium Octopus, version 3.12.0, Zenodo (2023); https://doi.org/10.5281/zenodo.7998767.

83. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

84. J. Conway, N. Gehlenborg, UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets, version 1.4.0 (2019); https://cran.r-project.org/web/packages/UpSetR/index.html.

85. J. Larsson, A. J. R. Godfrey, P. Gustafsson, D. H. E. (geometric algorithms), E. H. (root solver code), F. Privé, eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses, version 7.0.2 (2024); https://cran.r-project.org/web/packages/eulerr/index.html.

86. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–7 (2012).

87. H. Patel, J. Espinosa-Carrasco, B. Langer, P. Ewels, nf-core bot, M. U. Garcia, R. Syme, A. Peltzer, A. Talbot, D. Behrens, G. Gabernet, M. Jin, M. Hörtenhuber, J. G. Rodriguez, K. Menden, Ö. An, nf-core/atacseq: [2.1.2] - 2022-08-07, version 2.1.2, Zenodo (2023); https://doi.org/10.5281/zenodo.8222875.

88. Q. Wang, M. Li, T. Wu, L. Zhan, L. Li, M. Chen, W. Xie, Z. Xie, E. Hu, S. Xu, G. Yu, Exploring Epigenomic Datasets by ChIPseeker. *Curr. Protoc.* **2**, e585 (2022).

89. T. Zhu, X. Zhou, Y. You, L. Wang, Z. He, D. Chen, cisDynet: An integrated platform for modeling gene-regulatory dynamics and networks. *iMeta* **2**, e152 (2023).

90. F. Hahne, R. Ivanek, "Visualizing Genomic Data Using Gviz and Bioconductor" in *Statistical Genomics: Methods and Protocols*, E. Mathé, S. Davis, Eds. (Springer, New York, NY, 2016; https://doi.org/10.1007/978-1-4939-3578-9_16)*Methods in Molecular Biology*, pp. 335–351.

91. N. Harmston, E. Ing-Simmons, M. Perry, A. Barešić, B. Lenhard, GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data. *BMC Genomics* **16**, 963 (2015).

92. J. S. Y. Ho, B. W.-Y. Mok, L. Campisi, T. Jordan, S. Yildiz, S. Parameswaran, J. A. Wayman, N. N. Gaudreault, D. A. Meekins, S. V. Indran, I. Morozov, J. D. Trujillo, Y. S. Fstkchyan, R. Rathnasinghe, Z. Zhu, S. Zheng, N. Zhao, K. White, H. Ray-Jones, V. Malysheva, M. J. Thiecke, S.-Y. Lau, H. Liu, A. J. Zhang, A. C.-Y. Lee, W.-C. Liu, S. Jangra, A. Escalera, T. Aydillo, B. S. Melo, E. Guccione, R. Sebra, E. Shum, J. Bakker, D. A. Kaufman, A. L. Moreira, M. Carossino, U. B. R. Balasuriya, M. Byun, R. A. Albrecht, M. Schotsaert, A. Garcia-Sastre, S. K. Chanda, E. R. Miraldi, A. D. Jeyasekharan, B. R. TenOever, M. Spivakov, M. T. Weirauch, S. Heinz, H. Chen, C. Benner, J. A. Richt, I. Marazzi, TOP1 inhibition therapy protects against SARS-CoV-2-induced lethal inflammation. *Cell* **184**, 2618-2632.e17 (2021).

93. V. Malysheva, H. Ray-Jones, T. A. Cazares, O. Clay, D. Ohayon, P. Artemov, J. A. Wayman, M. D. Rosa, C. Petitjean, C. Booth, J. I. J. Ellaway, W. R. Orchard, X. Chen, S. Parameswaran, T. Nagano, P. Fraser, S. Schoenfelder, M. T. Weirauch, L. C. Kottyan, D. F. Smith, N. Powell, J. Weimer, C. Wallace, E. R. Miraldi, S. Waggoner, M. Spivakov, High-resolution promoter interaction analysis in Type 3 Innate Lymphoid Cells implicates Batten Disease gene CLN3 in Crohn's Disease aetiology. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.10.19.512842.

94. S. Wingett, P. Ewels, M. Furlan-Magaril, T. Nagano, S. Schoenfelder, P. Fraser, S. Andrews, HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* **4**, 1310 (2015).

95. P. Ewels, A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M. Ulysse Garcia, P. Di Tommaso, S. Nahnsen, The nf-core framework for community-curated bioinformatics pipelines., Zenodo (2022); https://doi.org/10.5281/zenodo.7139814.

96. D. Machlab, L. Burger, C. Soneson, F. M. Rijli, D. Schübeler, M. B. Stadler, monaLisa: an R/Bioconductor package for identifying regulatory motifs. *Bioinformatics* **38**, 2624–2625 (2022).

97. I. Rauluseviciute, R. Riudavets-Puig, R. Blanc-Mathieu, J. A. Castro-Mondragon, K. Ferenc, V. Kumar, R. B. Lemma, J. Lucas, J. Chèneby, D. Baranasic, A. Khan, O. Fornes, S. Gundersen, M. Johansen, E. Hovig, B. Lenhard, A. Sandelin, W. W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **52**, D174–D182 (2024).

98. G. Tan, B. Lenhard, TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).

99. J. Ou, S. A. Wolfe, M. H. Brodsky, L. J. Zhu, motifStack for the analysis of transcription factor binding site evolution. *Nat. Methods* **15**, 8–9 (2018).

100. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

101. A. Butler, S. Choudhary, D. Collins, C. Darby, J. Farrell, I. Grabski, C. Hafemeister, Y. Hao, A. Hartman, P. Hoffman, J. Jain, L. Jiang, M. Kowalski, S. Li, G. Molla, E. Papalexi, P. Roelli, R. Satija, K. Shekhar, A. Srivastava, T. Stuart, K. Torkenczy, S. Zheng, S. L. and Collaborators, Seurat: Tools for Single Cell Genomics, version 5.1.0 (2024); https://cran.r-project.org/web/packages/Seurat/index.html.

Figures

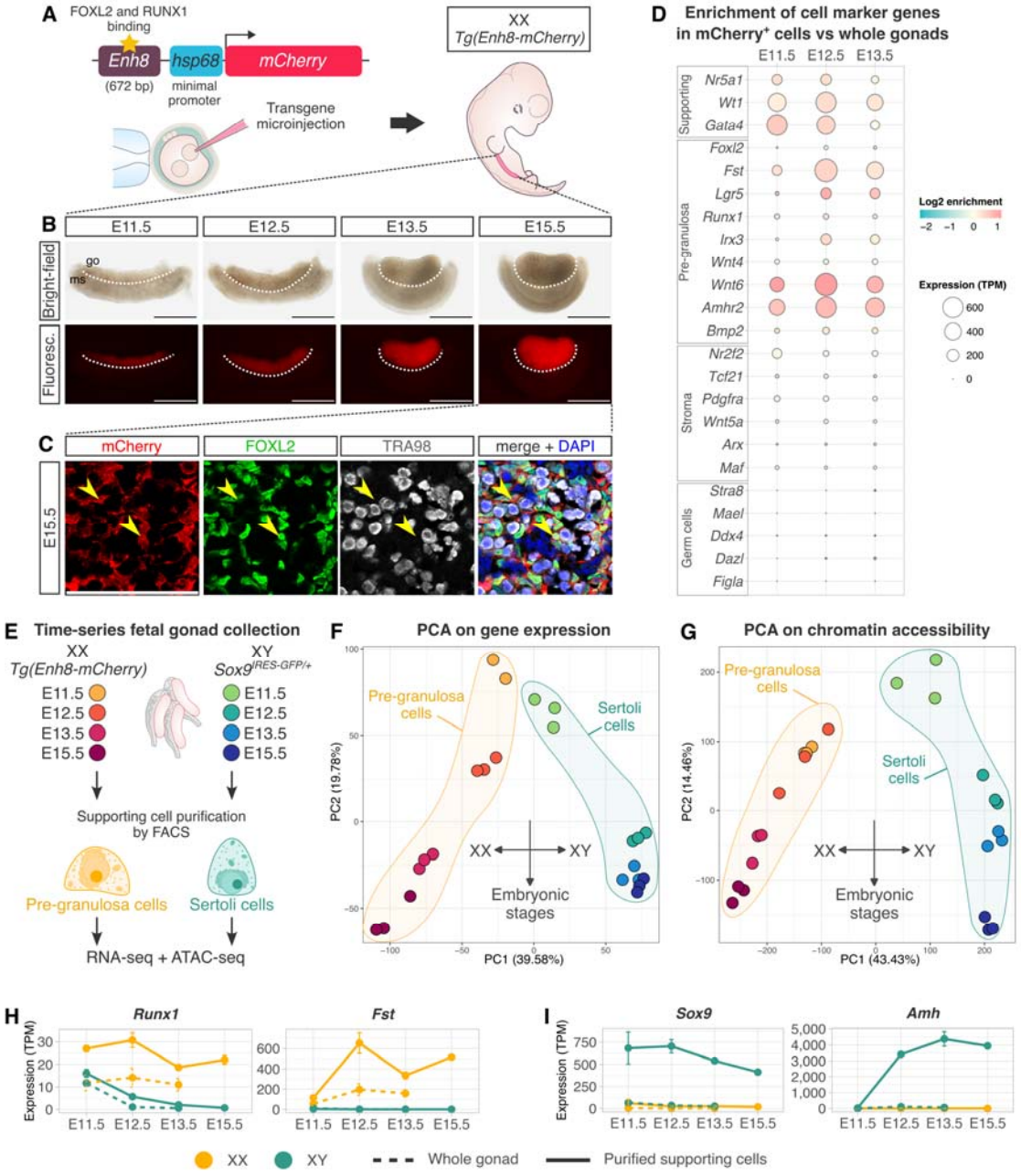


Figure 1. Generation of a new mouse line allowing pre-granulosa cell purification for multi-omics analysis

(**A**) Enh8 was cloned upstream of the mCherry reporter gene controlled by the *hsp68* minimal promoter. The obtained vector was injected into zygote mouse embryos by microinjection to obtain the *Enh8-mCherry (Tg(Enh8-mCherry))* mouse line. (**B**) Binocular pictures of dissected XX *Enh8-mCherry* gonads at E11.5, E12.5, E13.5 and E15.5 in bright-field and

fluorescence. go=gonad, ms=mesonephros. Scale bar: 500 µm. (**C**) Immunofluorescent staining of E15.5 *Enh8-mCherry* ovary sections. mCherry (transgene) is labelled in red, FOXL2, a marker of pre-granulosa cells is labelled in green, and TRA98, a marker of germ cells is labelled in green. Some mCherry expressing pre-granulosa cells are highlighted with yellow arrowheads. Scale bar: 100 µm. (**D**) Gene expression enrichment of ovarian marker genes in *Enh8-mCherry* sorted cells at E11.5, E12.5 and E13.5 RNA-seq compared to whole gonad RNA-seq (from Zhao et al. 2018). Enrichment (log2 of the ratio between *Enh8-mCherry* sorted cells and whole gonad gene expression) is represented with a blue-to-red gradient, and the expression levels of the *Enh8-mCherry* sorted cells are represented as TPM (transcript per million) with the size of the dots. (**E**) Schematic experimental design. XX gonads from *Enh8-mCherry* and XY gonads from *SOX*9$^{\text{IRES-GFP/+}}$ mouse lines were collected at E11.5, E12.5, E13.5 and E15.5. Pre-granulosa and Sertoli cells were purified by FACS at each embryonic stage, and the sorted cells were subjected to RNA-seq and ATAC-seq to constitute a time-series paired gene expression and chromatin accessibility data collection. (**F**) and (**G**) PCA (principal component analysis) of the obtained RNA-seq and ATAC-seq data. Samples are coloured by sex and embryonic stage. Pre-granulosa samples were circled in yellow, and Sertoli cell samples in green. (**H**) and (**I**) Expression profiles of *Runx1*, *Fst* (pre-granulosa specific genes), and *Sox9* and *Amh* (Sertoli specific genes) in purified cells (continuous line) and in whole gonads (from Zhao et al. 2018, dashed line) along embryonic stages in both sexes (XX in yellow, XY in green).

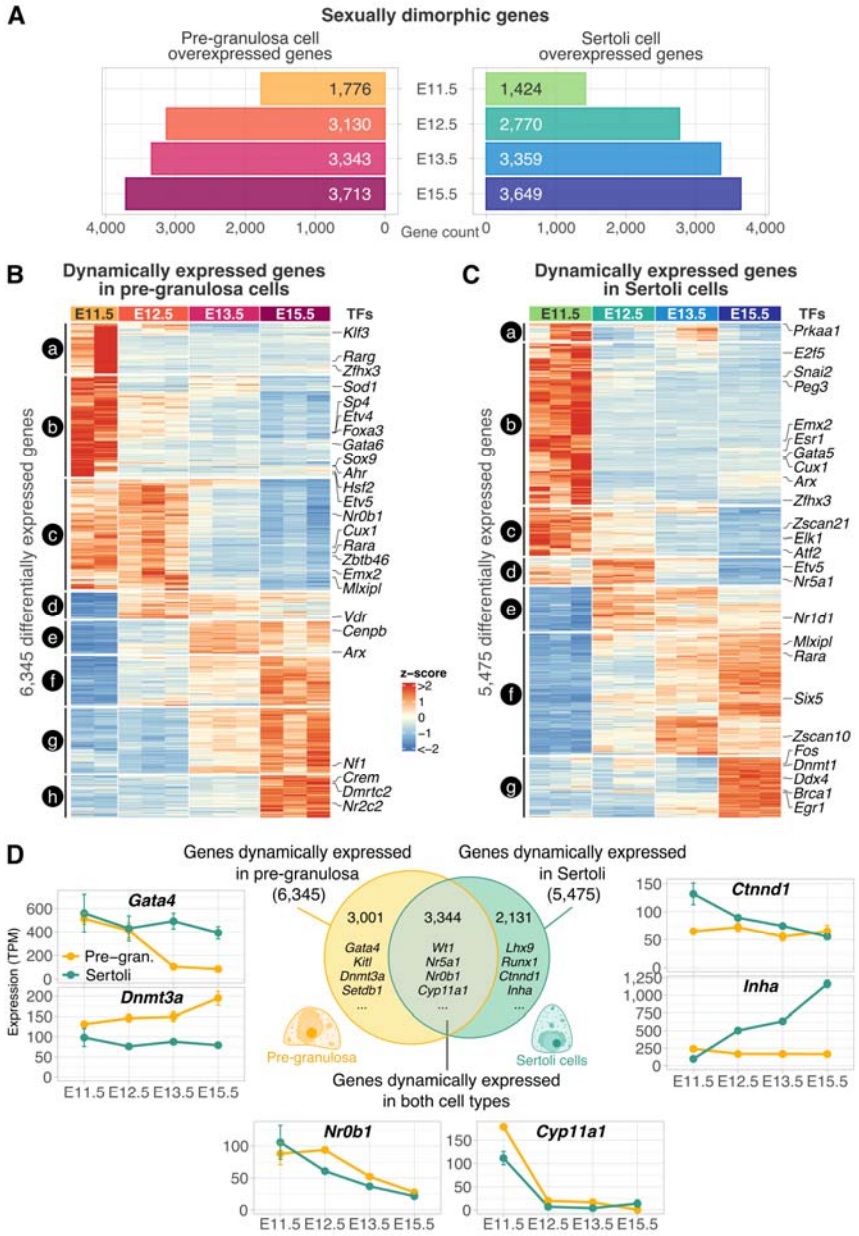


Figure 2. Characterization of the supporting cell transcriptome during their differentiation

(**A**) Number of overexpressed genes in pre-granulosa and Sertoli cells at each embryonic
1280 stage compared to the opposite sex. (**B**) and (**C**) Heatmaps representing the expression changes (z-score) of the genes dynamically expressed during pre-granulosa and Sertoli cells, respectively. Genes were clustered by expression profiles. The clusters were labelled with letters on the left side of the heatmaps. 25 transcription factors known to cause a gonadal or

infertility phenotype were labelled on the right side of the heatmaps. (**D**) Venn diagram
1285 showing the overlap of the dynamically expressed genes in both sexes with examples of gene names present in each intersection. Expression (TPM) profiles of genes from the different sets are shown as examples.

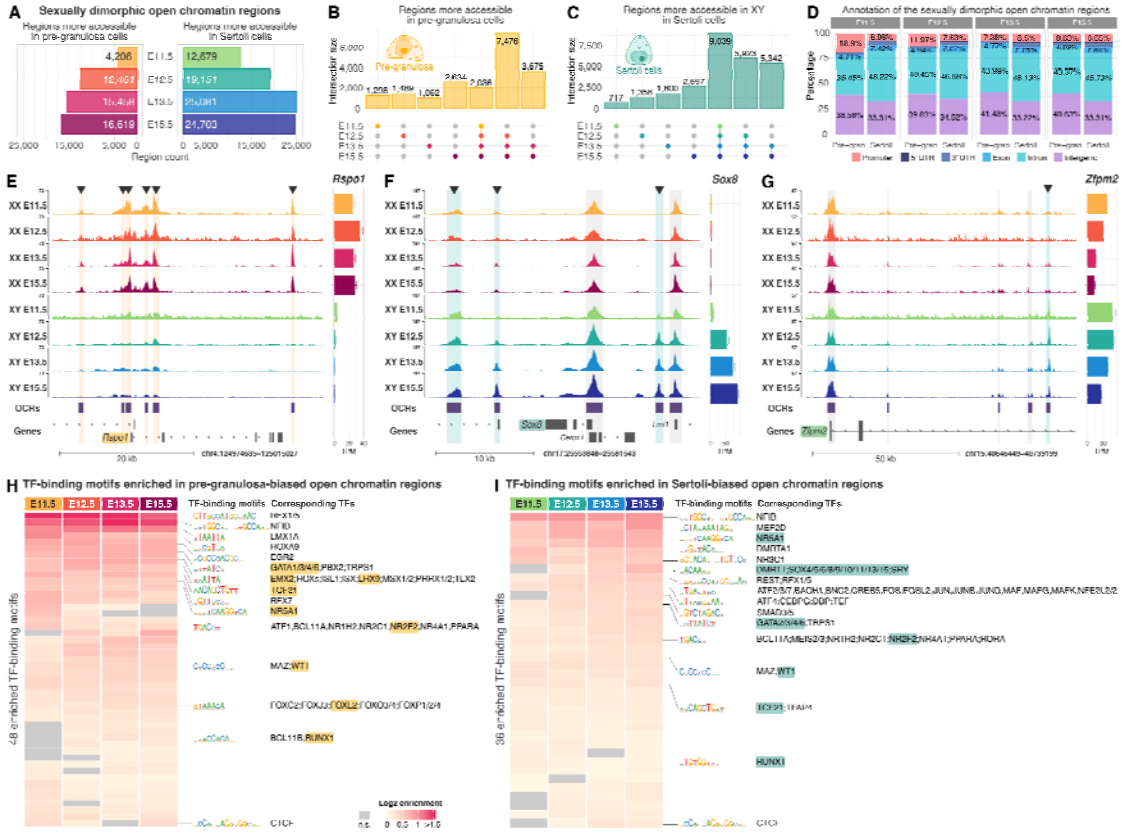


Figure 3. Establishment of sexually dimorphic open chromatin regions in supporting cells during their differentiation

(**A**) Number of differentially accessible regions in pre-granulosa and Sertoli cells at each embryonic stage compared to the opposite sex. (**B**) and (**C**) Upset plot representing the distribution of the sex-biased open chromatin regions across embryonic stages in pre-granulosa and Sertoli cells, respectively. Each column represents a specific intersection between sets. Dots connected by lines (bottom) indicate which sets are involved in the intersection. The height of the bars on the y-axis (top) quantifies the number of elements in each intersection. (**D**) Genomic features overlapped by the sex-biased open chromatin regions across embryonic stages in pre-granulosa and Sertoli cells. (**E**) to (**G**) Genomic tracks showing the normalized ATAC-seq signal of sexually dimorphic open chromatin regions (OCRs) around the pre-granulosa specific factor *Rspo1*, the Sertoli specific factor *Sox8* and the critical gonadal factor *Zfmp2* (also known as *Fog2*). Non-sex specific OCRs are highlighted in grey, sex-biased OCRs are marked by an arrowhead on top, pre-granulosa-biased OCRs are highlighted in yellow, and Sertoli-biased in blue. The bar plot on the right-

hand side shows the expression level in TPM of the gene of interest. The error bars represent the standard deviation between the replicates. (**H**) and (**I**) Transcription factor binding motif enrichment analysis in pre-granulosa and Sertoli-biased open chromatin regions respectively. Only transcription factors found expressed in the supporting cells are shown. Motifs were merged by sequence similarity and the consensus logo is shown. Known gonadal transcription factors are highlighted in yellow and blue. Non-significant enrichments are coloured in grey. n.s.: non-significant.

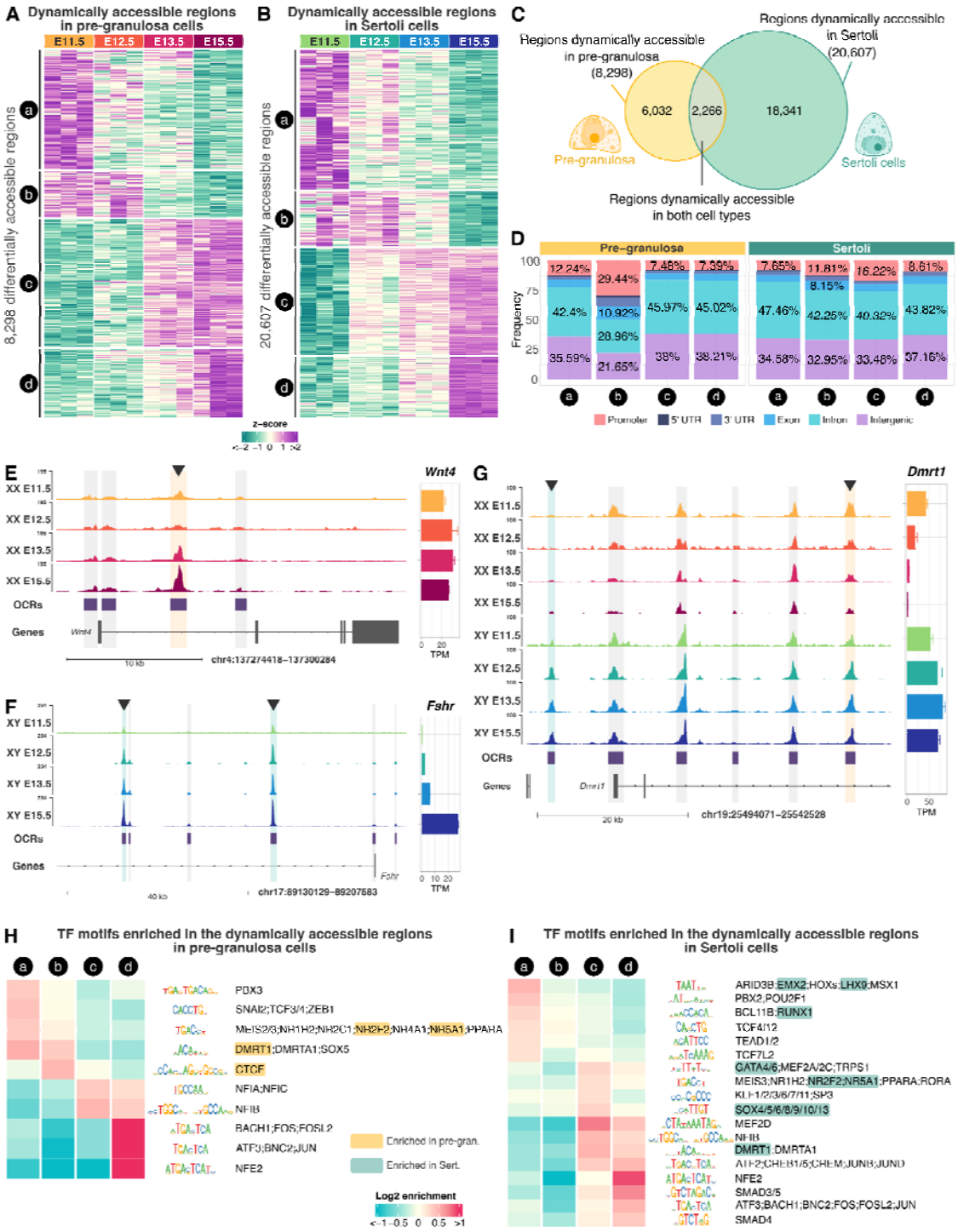

A
Dynamically accessible regions in pre-granulosa cells
E11.5
E12.5
E13.5
E15.5
8,298 differentially accessible regions
B
Dynamically accessible regions in Sertoli cells
20,607 differentially accessible regions
z-score
C
Regions dynamically accessible in pre-granulosa (8,298)
Regions dynamically accessible in Sertoli (20,607)
6,032
2,266
18,341
Pre-granulosa
Sertoli cells
Regions dynamically accessible in both cell types
D
Pre-granulosa
Sertoli
Frequency
Promoter
5' UTR
3' UTR
Exon
Intron
Intergenic
E
Wnt4
XX E11.5
XX E12.5
XX E13.5
XX E15.5
OCRs
Genes
chr4:137274418–137300284
TPM
F
Fshr
XY E11.5
XY E12.5
XY E13.5
XY E15.5
chr17:89130129–89207583
G
Dmrt1
XY E11.5
XY E12.5
XY E13.5
XY E15.5
chr19:25494071–25542528
H
TF motifs enriched in the dynamically accessible regions in pre-granulosa cells
PDX3
SNAI2;TCF3/4;ZEB1
MEIS2/3;NR1H2;NR2C1;NR2F2;NR4A1;NR5A1;PPARA
DMRT1;DMRTA1;SOX5
CTCF
NFIA;NFIC
NFIB
BACH1;FOS;FOSL2
ATF3;BNC2;JUN
NFE2
Enriched in pre-gran.
Enriched in Sert.
Log2 enrichment
I
TF motifs enriched in the dynamically accessible regions in Sertoli cells
ARID3B;EMX2;HOXs;LHX9;MSX1
PBX2;POU2F1
BCL11B;RUNX1
TCF4/12
TEAD1/2
TCF7L2
GATA4/6;MEF2A/2C;TRPS1
MEIS3;NR1H2;NR2F2;NR5A1;PPARA;RORA
KLF1/2/3/6/7/11;SP3
SOX4/5/6/8/9/10/13
MEF2D
NFIB
DMRT1;DMRTA1
ATF2;CHEB1/5;CHEM;JUNB;JUND
NFE2
SMAD3/5
ATF3;BACH1;BNC2;FOS;FOSL2;JUN
SMAD4

Figure 4. Dynamics of open chromatin regions in supporting cells during their differentiation

(A) and (**B**) Heatmaps representing the accessibility changes (z-score) of the open chromatin regions during pre-granulosa and Sertoli cells, respectively. Regions were clustered by accessibility profiles. The clusters were labelled with letters on the left side of the heatmaps. (**C**)Venn diagram showing the overlap between pre-granulosa and Sertoli cell dynamic open chromatin regions. (**D**) Genomic features where the dynamic open chromatin regions are found in goth pre-granulosa and Sertoli cells at each stage. (**E**) to (**G**) Genomic tracks showing the normalized ATAC-seq signal of genomic regions loci containing dynamic open chromatin regions (OCRs) in pre-granulosa and/or in Sertoli cells in the vicinity of known gonadal genes. Non-dynamic OCRs are highlighted in grey, significantly dynamic OCRs are marked by an arrowhead on top, pre-granulosa-dynamic OCRs are highlighted in yellow, and Sertoli-dynamic in blue. The bar plot on the right-hand side shows the expression level in TPM of the gene of interest. The error bars represent the standard deviation between the replicates. (**H**) and (**I**) Heatmap of the transcription factor motifs differentially enrichment between the different open chromatin region clusters from pre-granulosa and Sertoli cells, respectively. Only transcription factors found expressed in the supporting cells are shown. Motifs were merged by sequence similarity and the consensus logo is shown. Known gonadal transcription factors are highlighted.

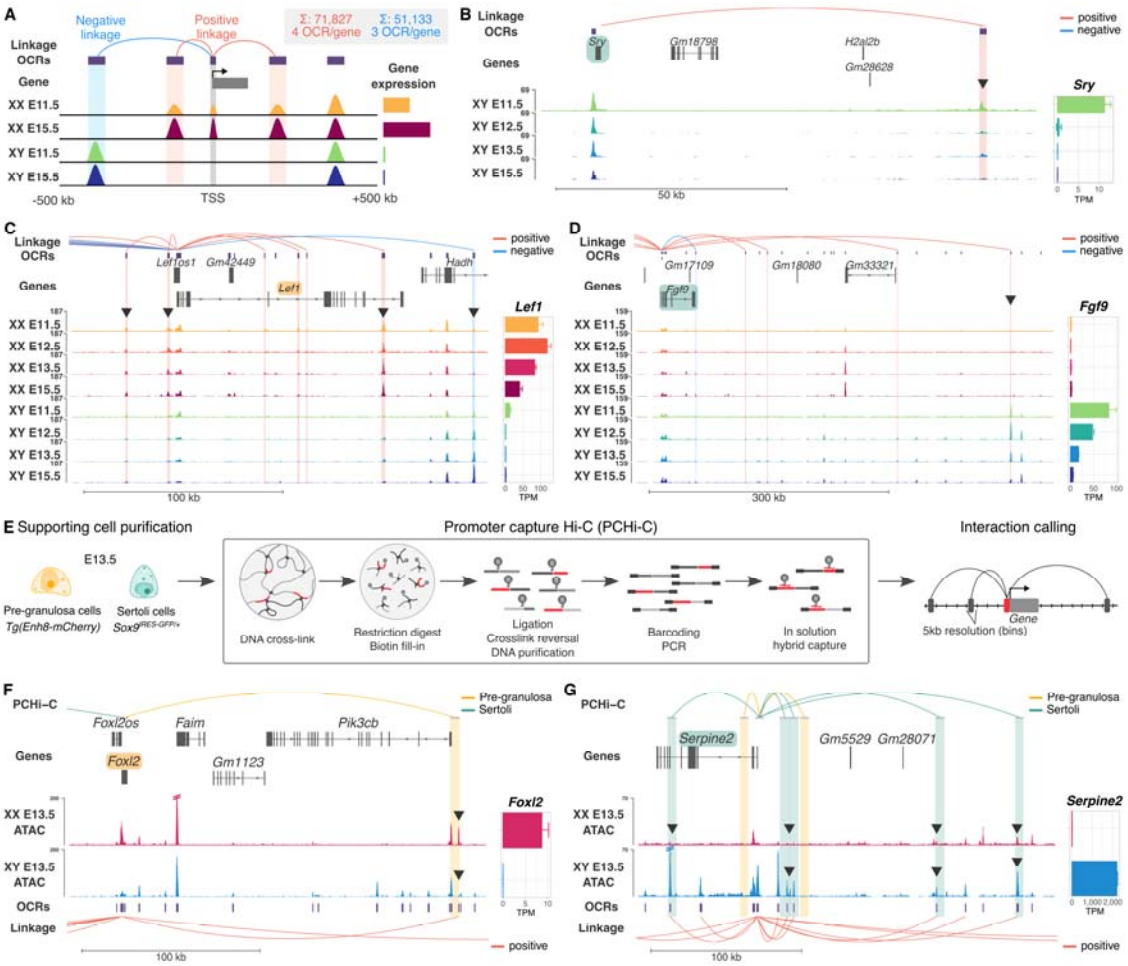


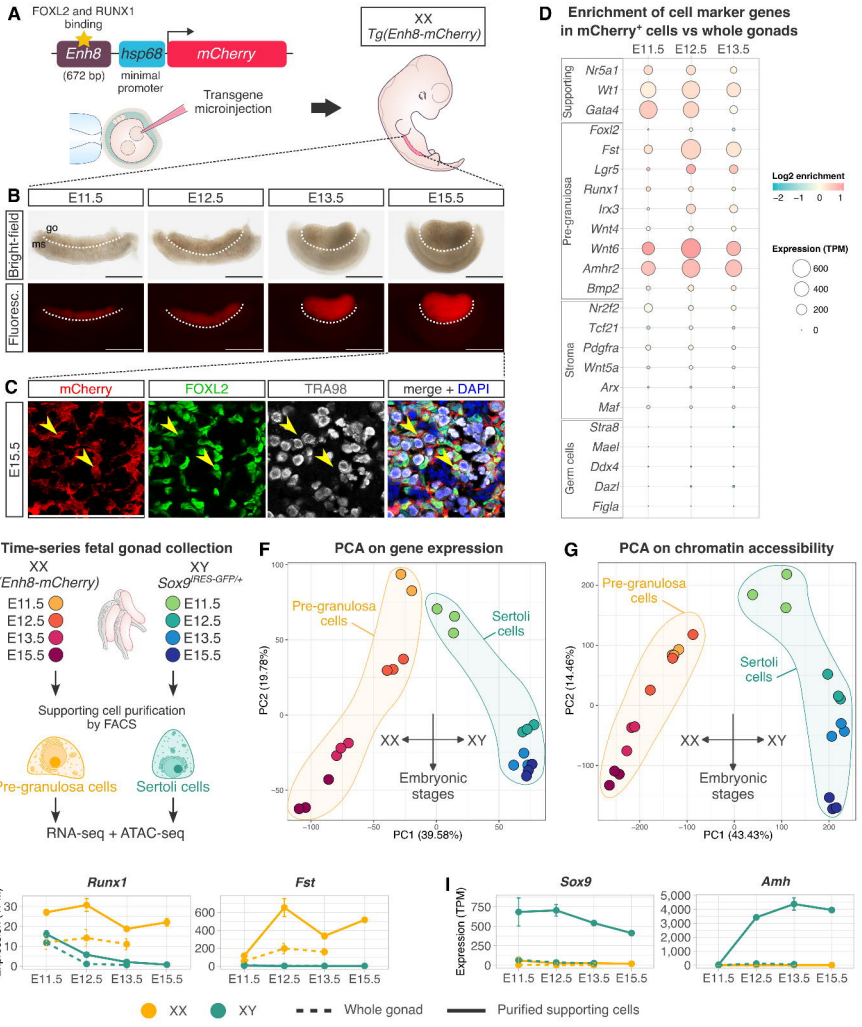
Figure 5. Prediction of gene-cis-regulatory region associations

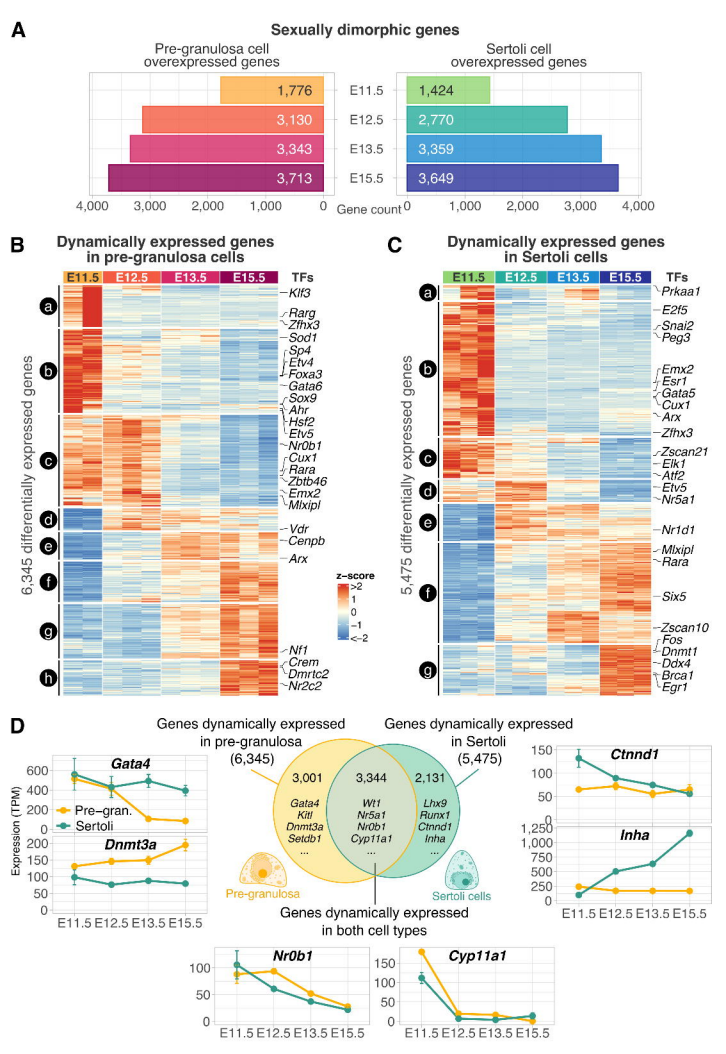
(**A**) Schematic representation describing the strategy to link open chromatin regions with their target genes. Open chromatin regions located within a ±500 kb window of a gene's transcription start site (TSS), whose accessibility correlates with gene expression, are identified as putative enhancers. Conversely, open chromatin regions that anti-correlate with gene expression are linked as putative silencers. The number of positive and negative link, as well as the average number of linked open chromatin regions (OCR) per gene are indicated. (**B-D**) Genomic tracks showing the predicted links between open chromatin regions and gene expression. The positive links are represented as red line arcs, and negative links as blue line arcs. The genomic tracks represent normalized ATAC-seq signal of genomic regions loci in pre-granulosa and/or in Sertoli cells. Noticeable genomic loci are indicated with an arrowhead. The bar plot on the right-hand side shows the expression level in TPM of the gene of interest. The error bars represent the standard deviation between the replicates. (**E**)

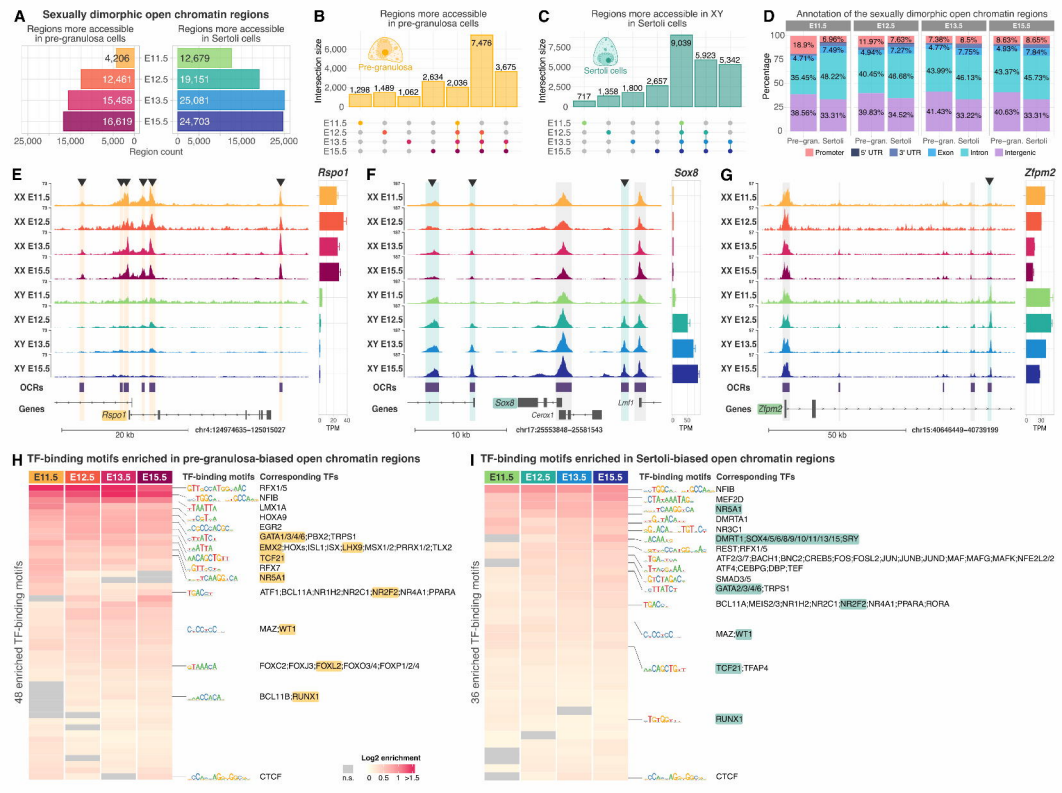
Experimental design of the promoter capture Hi-C (PCHi-C) experiment. E13.5 pre-granulosa and Sertoli cells have been purified by FACS, fixed, and have been subjected to PCHi-C. Interactions between gene promoters and genomic regions were called using a 5 kb bin resolution. (**F-G**) Genomic tracks showing the PCHi-C interactions found in either pre-
1350 granulosa or Sertoli cells around two sexually dimorphic genes at E13.5. The interactions contain open chromatin regions and overlap with the RNA-ATAC linkage analysis. The dash lines on top of the ATAC-seq signal signify that the scale has been cropped.

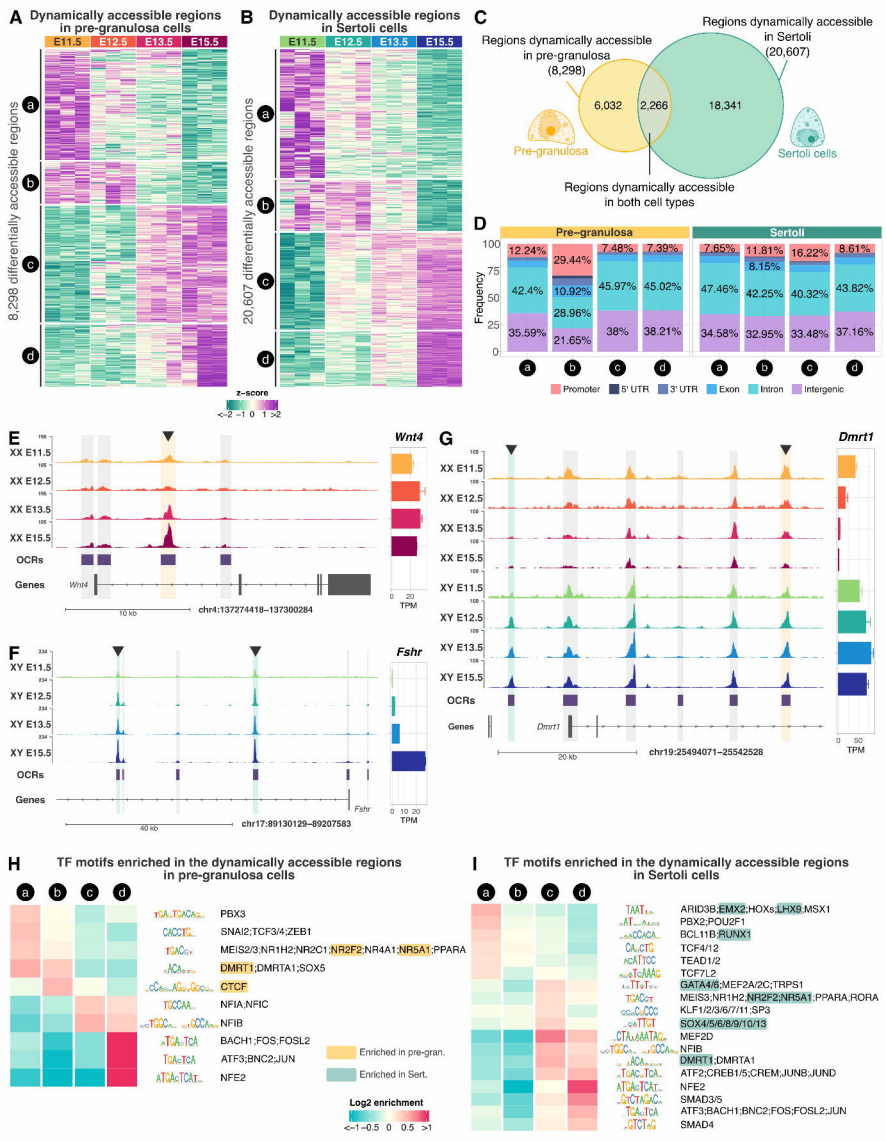
Figure 6. ATAC TF footprints on sexually dimorphic open chromatin regions

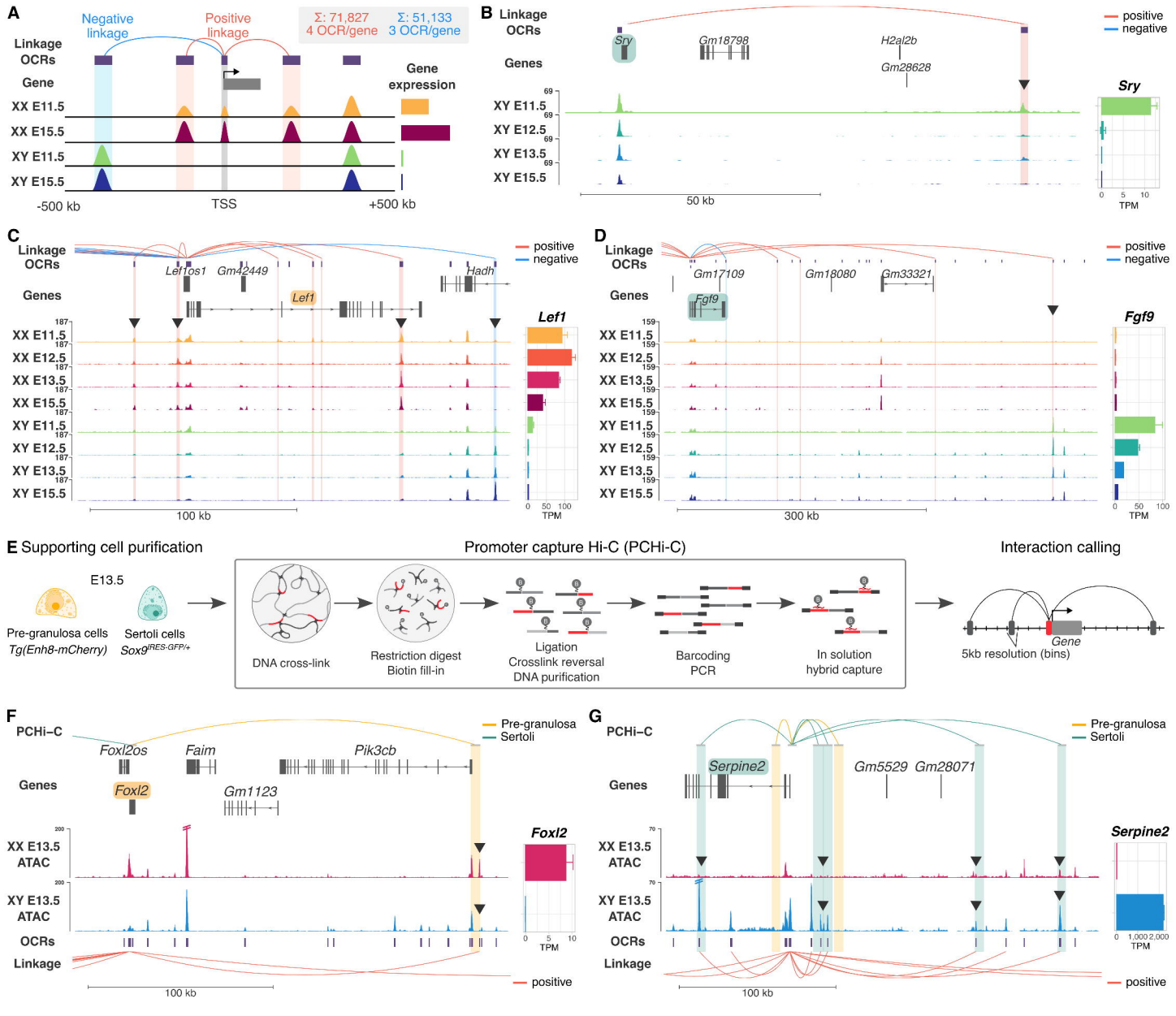
(**A**) Differential transcription factor binding motif enrichment and occupancy (ATAC-seq footprints) in the combined (all developmental stages) sex-biased open chromatin regions when comparing one sex to the other. Only transcription factors found expressed in the supporting cells are shown. Motifs were merged by sequence similarity and the consensus logo is shown. Known gonadal transcription factors are highlighted. (**B**) and (**C**) Comparison of the aggregated ATAC-seq footprint signals at E13.5 in both sexes for the top pre-granulosa and Sertoli cell differentially bound TF-binding motifs recognized by EMX2 and LHX9, and DMRT1 and SOX and SRY factors, respectively. The number of bound motifs is indicated. The dashed lines represent the motif location. (**D**) to (**J**) Genomic tracks showing TF footprints of the gonadal factors highlighted in figure (**A**) on different sex-biased accessible loci. Footprints found in any of the studied embryonic stages were aggregated. For concision, we only show the name of the known gonadal TFs and not the full list of TFs able to bind the occupied motifs.


A
FOXL2 and RUNX1 binding
Enh8
(672 bp)
hsp68
minimal promoter
mCherry
XX
Tg(Enh8-mCherry)
Transgene microinjection
B
E11.5
E12.5
E13.5
E15.5
Bright-field
Fluoresc.
go
ms
C
mCherry
FOXL2
TRA98
merge + DAPI
E15.5
D
Enrichment of cell marker genes in mCherry+ cells vs whole gonads
E11.5
E12.5
E13.5
Supporting
Nr5a1
Wt1
Gata4
Pre-granulosa
Foxl2
Fst
Lgr5
Runx1
Irx3
Wnt4
Wnt6
Amhr2
Bmp2
Stroma
Nr2f2
Tcf21
Pdgfra
Wnt5a
Arx
Maf
Germ cells
Stra8
Mael
Ddx4
Dazl
Figla
Log2 enrichment
−2
−1
0
1
Expression (TPM)
600
400
200
0
E
Time-series fetal gonad collection
XX
Tg(Enh8-mCherry)
XY
Sox9IRES-GFP/+
E11.5
E12.5
E13.5
E15.5
Supporting cell purification by FACS
Pre-granulosa cells
Sertoli cells
RNA-seq + ATAC-seq
F
PCA on gene expression
Pre-granulosa cells
Sertoli cells
XX
XY
Embryonic stages
PC2 (19.78%)
PC1 (39.58%)
G
PCA on chromatin accessibility
Pre-granulosa cells
Sertoli cells
XX
XY
Embryonic stages
PC2 (14.46%)
PC1 (43.43%)
H
Runx1
Fst
Expression (TPM)
E11.5
E12.5
E13.5
E15.5
I
Sox9
Amh
Expression (TPM)
E11.5
E12.5
E13.5
E15.5
XX
XY
Whole gonad
Purified supporting cells


A
Sexually dimorphic genes
Pre-granulosa cell overexpressed genes
Sertoli cell overexpressed genes
E11.5
E12.5
E13.5
E15.5
1,776
3,130
3,343
3,713
1,424
2,770
3,359
3,649
4,000 3,000 2,000 1,000 0
Gene count
0 1,000 2,000 3,000 4,000
B
Dynamically expressed genes in pre-granulosa cells
E11.5 E12.5 E13.5 E15.5
TFs
6,345 differentially expressed genes
a
b
c
d
e
f
g
h
Klf3
Rarg
Zfhx3
Sod1
Sp4
Etv4
Foxa3
Gata6
Sox9
Ahr
Hsf2
Etv5
Nr0b1
Cux1
Rara
Zbtb46
Emx2
Mlxipl
Vdr
Cenpb
Arx
Nf1
Crem
Dmrtc2
Nr2c2
z-score
>2
1
0
−1
<−2
C
Dynamically expressed genes in Sertoli cells
E11.5 E12.5 E13.5 E15.5
TFs
5,475 differentially expressed genes
a
b
c
d
e
f
g
Prkaa1
E2f5
Snal2
Peg3
Emx2
Esr1
Gata5
Cux1
Arx
Zfhx3
Zscan21
Elk1
Atf2
Etv5
Nr5a1
Nr1d1
Mlxipl
Rara
Six5
Zscan10
Fos
Dnmt1
Ddx4
Brca1
Egr1
D
Genes dynamically expressed in pre-granulosa (6,345)
Genes dynamically expressed in Sertoli (5,475)
3,001
Gata4
Kitl
Dnmt3a
Setdb1
...
3,344
Wt1
Nr5a1
Nr0b1
Cyp11a1
...
2,131
Lhx9
Runx1
Ctnnd1
Inha
...
Pre-granulosa
Sertoli cells
Genes dynamically expressed in both cell types
Gata4
Dnmt3a
Expression (TPM)
Pre–gran.
Sertoli
E11.5 E12.5 E13.5 E15.5
Ctnnd1
Inha
Nr0b1
Cyp11a1


A
Sexually dimorphic open chromatin regions
Regions more accessible in pre-granulosa cells
Regions more accessible in Sertoli cells
4,206
12,461
15,458
16,619
12,679
19,151
25,081
24,703
E11.5
E12.5
E13.5
E15.5
25,000 15,000 5,000 0
0 5,000 15,000 25,000
Region count
B
Regions more accessible in pre-granulosa cells
Intersection size
Pre-granulosa
1,298 1,489 1,062 2,634 2,036 7,478 3,675
C
Regions more accessible in XY in Sertoli cells
Sertoli cells
717 1,358 1,800 2,657 9,039 5,923 5,342
D
Annotation of the sexually dimorphic open chromatin regions
Percentage
Pre−gran. Sertoli
18.9% 8.96% 11.97% 7.63% 8.63% 8.65%
4.71% 7.49% 4.94% 7.27% 4.83% 7.84%
35.45% 48.22% 40.45% 48.68% 43.37% 45.73%
38.56% 33.31% 39.83% 34.52% 40.63% 33.31%
4.77% 8.5% 7.98% 7.75% 46.13% 43.99% 33.22% 41.43%
Promoter
5' UTR
3' UTR
Exon
Intron
Intergenic
E
Rspo1
XX E11.5
XX E12.5
XX E13.5
XX E15.5
XY E11.5
XY E12.5
XY E13.5
XY E15.5
OCRs
Genes
20 kb
chr4:124974635−125015027
TPM
F
Sox8
Cerox1
Lmf1
10 kb
chr17:25553848−25581543
G
Zfpm2
50 kb
chr15:40646449−40739199
H
TF-binding motifs enriched in pre-granulosa-biased open chromatin regions
48 enriched TF-binding motifs
TF-binding motifs
Corresponding TFs
RFX1/5
NFIB
LMX1A
HOXA9
EGR2
GATA1/3/4/6;PBX2;TRPS1
EMX2;HOXs;ISL1;SX;LHX9;MSX1/2;PRRX1/2;TLX2
TCF21
RFX7
NR5A1
ATF1;BCL11A;NR1H2;NR2C1;NR2F2;NR4A1;PPARA
MAZ;WT1
FOXC2;FOX;FOXL2;FOXO3/4;FOXP1/2/4
BCL11B;RUNX1
CTCF
Log2 enrichment
n.s. 0 0.5 1 >1.5
I
TF-binding motifs enriched in Sertoli-biased open chromatin regions
36 enriched TF-binding motifs
NFIB
MEF2D
NR5A1
DMRTA1
NR3C1
DMRT1;SOX4/5/6/8/9/10/11/13/15;SRY
REST;RFX1/5
ATF2/3/7;BACH1;BNC2;CREB5;FOS;FOSL2;JUN;JUNB;JUND;MAF;MAFG;MAFK;NFE2L2/2
ATF4;CEBPG;DBP;TEF
SMAD3/5
GATA2/3/4/6;TRPS1
BCL11A;MEIS2/3;NR1H2;NR2C1;NR2F2;NR4A1;PPARA;RORA
MAZ;WT1
TCF21;TFAP4
RUNX1
CTCF


A
Dynamically accessible regions in pre-granulosa cells
E11.5
E12.5
E13.5
E15.5
8,298 differentially accessible regions
a
b
c
d
B
Dynamically accessible regions in Sertoli cells
20,607 differentially accessible regions
z-score
<-2 -1 0 1 >2
C
Regions dynamically accessible in pre-granulosa (8,298)
Regions dynamically accessible in Sertoli (20,607)
6,032
2,266
18,341
Pre-granulosa
Sertoli cells
Regions dynamically accessible in both cell types
D
Pre-granulosa
Sertoli
Frequency
12.24%
29.44%
7.48%
7.39%
42.4%
10.92%
28.96%
45.97%
45.02%
35.59%
21.65%
38%
38.21%
7.65%
11.81%
16.22%
8.61%
8.15%
47.46%
42.25%
40.32%
43.82%
34.58%
32.95%
33.48%
37.16%
Promoter
5' UTR
3' UTR
Exon
Intron
Intergenic
E
Wnt4
XX E11.5
XX E12.5
XX E13.5
XX E15.5
OCRs
Genes
10 kb
chr4:137274418–137300284
TPM
F
Fshr
XY E11.5
XY E12.5
XY E13.5
XY E15.5
40 kb
chr17:89130129–89207583
G
Dmrt1
20 kb
chr19:25494071–25542528
H
TF motifs enriched in the dynamically accessible regions in pre-granulosa cells
PBX3
SNAI2;TCF3/4;ZEB1
MEIS2/3;NR1H2;NR2C1;NR2F2;NR4A1;NR5A1;PPARA
DMRT1;DMRTA1;SOX5
CTCF
NFIA;NFIC
NFIB
BACH1;FOS;FOSL2
ATF3;BNC2;JUN
NFE2
Enriched in pre-gran.
Enriched in Sert.
Log2 enrichment
<-1 -0.5 0 0.5 >1
I
TF motifs enriched in the dynamically accessible regions in Sertoli cells
ARID3B;EMX2;HOXs;LHX9;MSX1
PBX2;POU2F1
BCL11B;RUNX1
TCF4/12
TEAD1/2
TCF7L2
GATA4/6;MEF2A/2C;TRPS1
MEIS3;NR1H2;NR2F2;NR5A1;PPARA;RORA
KLF1/2/3/6/7/11;SP3
SOX4/5/6/8/9/10/13
MEF2D
NFIB
DMRT1;DMRTA1
ATF2;CREB1/5;CREM;JUNB;JUND
NFE2
SMAD3/5
ATF3;BACH1;BNC2;FOS;FOSL2;JUN
SMAD4


A
Negative linkage
Positive linkage
Σ: 71,827 4 OCR/gene
Σ: 51,133 3 OCR/gene
Linkage
OCRs
Gene
Gene expression
XX E11.5
XX E15.5
XY E11.5
XY E15.5
-500 kb
TSS
+500 kb
B
Linkage
OCRs
Genes
positive
negative
Sry
Gm18798
H2al2b
Gm28628
XY E11.5
XY E12.5
XY E13.5
XY E15.5
50 kb
TPM
C
Linkage
OCRs
Genes
Lef1os1
Gm42449
Lef1
Hadh
XX E11.5
XX E12.5
XX E13.5
XX E15.5
XY E11.5
XY E12.5
XY E13.5
XY E15.5
100 kb
D
Gm17109
Gm18080
Gm33321
Fgf9
300 kb
E Supporting cell purification
E13.5
Pre-granulosa cells
Tg(Enh8-mCherry)
Sertoli cells
Sox9IRES-GFP/+
Promoter capture Hi-C (PCHi-C)
DNA cross-link
Restriction digest
Biotin fill-in
Ligation
Crosslink reversal
DNA purification
Barcoding
PCR
In solution
hybrid capture
Interaction calling
Gene
5kb resolution (bins)
F
PCHi-C
Pre-granulosa
Sertoli
Foxl2os
Faim
Pik3cb
Foxl2
Gm1123
XX E13.5 ATAC
XY E13.5 ATAC
OCRs
Linkage
100 kb
G
Serpine2
Gm5529
Gm28071

A
TF-binding motifs
Pre–gr.
Sert.
More bound in XX
More bound in XY
Corresponding TFs
RFX1/5
ARID3B;EMX2;HOXs;ISX;LHX9;MSX1
NFIB;SRF
ZBTB14
GATA1/3/4/6;HOXs;PBX2;TRPS1
ETS1
TCF12/21;TFAP4
FOXs
BCL11B;NFIA;RUNX1
KLF1/2/6/7
IRF1;STAT2
OSR2
PPARA
PLAGL1;ZFX
MEF2A/2C;YY1
NR5A1
MEIS3
PBX3;PKNOX1
OSR1
SMAD2/4
ZBTB12
ZKSCAN3
SMAD3/5;ZSCAN29
DMRT1;SOX4/5/6/8/9/10/11/13/15
DMRTA1
Log2 enrichment
Differential binding score
XX E11.5
XX E12.5
XX E13.5
XX E15.5
XY E11.5
XY E12.5
XY E13.5
XY E15.5
Enriched in pre-gran.
Enriched in Sert.
B
ARID3B;EMX2;HOXs;ISX;LHX9;MSX1
n=13,867
Normalized Tn5 cut signal
bp from center
XX E13.5
XY E13.5
C
DMRT1;SOX4/5/6/8/9/10/11/13/15
n=8,628
D
chr11:111943502–111945409
Sox9 Enh8
Sox9 >>
OCRs
XX footprints
XY footprints
GATAs
NR5A1
EMX2–LHX9
RUNX1
1 kb
E
chr11:112216859–112218018
Sox9 Enh13
DMRT1–SOXs
900 bp
F
positive
negative
chr4:124975800–125014500
Linkage
Genes
Rspo1
30 kb
G
chr2:133549206–133573437
Bmp2
Mir3090
FOXs
DMRT1–SOXs
20 kb
H
chr17:25565109–25613145
Sox8
Cerox1
40 kb
I
chr14:58069277–58114058
Fgf9
Gm17109
RUNX1
20 kb
chr14:58507365–58508623
DMRT1–SOXs
NR5A1
500 bp
J
chrY:2661356–2663744
Sry
2 kb
chrY:2745428–2747605