Single-cell roadmap of human gonadal development

Roser Vento-Tormo (✉ rv4@sanger.ac.uk)

Wellcome Sanger Institute   https://orcid.org/0000-0002-9870-8474

Luz Garcia-Alonso

Wellcome Sanger Institute, Cambridge, UK

Valentina Lorenzi

Wellcome Genome Campus, Hinxton, Cambridge

Cecilia Mazzeo

Wellcome Trust Sanger Institute

Carmen Sancho-Serra

Wellcome Sanger Institute, Cambridge, UK

Kenny Roberts

Wellcome Sanger Institute   https://orcid.org/0000-0001-6155-0821

Justin Engelbert

Newcastle University

João Alves-Lopes

Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK.

Magda Marečková

Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, United Kingdom

Rachel Botting

Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK   https://orcid.org/0000-0001-9595-4605

Tong Li

Wellcome Sanger Institute

Berta Crespo

Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH, UK

Stijn van Dongen

Wellcome Sanger Institute

Vladimir Kiselev

Sanger Institute

Elena Prigmore

Wellcome Trust Sanger Institute

Ashley Moffett

University of Cambridge   https://orcid.org/0000-0002-8388-9073

Mary Herbert
  Institute of Genetic Medicine, Newcastle University

Omer Ali Bayraktar
  Wellcome Sanger Institute   https://orcid.org/0000-0001-6055-277X

Azim Surani
  Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK.

Muzlifah Haniffa
  Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom.

Biological Sciences - Article

Single-cell roadmap of human gonadal development

Luz Garcia-Alonso 1†, Valentina Lorenzi 1†, Cecilia Icoresi Mazzeo 1, Carmen Sancho-Serra 1, Kenny Roberts 1, Justin Engelbert 2, João Pedro Alves-Lopes 3, Magda Marečková 1,4, Rachel A Botting 2, Tong Li 1, Berta Crespo 5, Stijn van Dongen 1, Vladimir Kiselev 1, Elena Prigmore 1, Ashley Moffett 6, Mary Herbert 2, Omer Bayraktar 1, Azim Surani 3, Muzlifah Haniffa 1,2, Roser Vento-Tormo 1*

Affiliations:

1 Wellcome Sanger Institute, Cambridge, UK
2 Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom.
3 Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK.
4 Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, United Kingdom
5 Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH, UK.
6 University of Cambridge Centre for Trophoblast Research, Cambridge, UK; Department of Pathology, University of Cambridge, Cambridge, UK.

* corresponding author. Email: rv4@sanger.ac.uk
† equal contribution

Summary: Gonadal development is a complex process that involves sex determination followed by divergent maturation into ovaries or testes. Historically, limited tissue accessibility and lack of reliable *in vitro* models have impeded our understanding of human gonadogenesis, despite its importance in gonadal pathologies and infertility. Here, we generated a comprehensive map of first- and second-trimester gonadal development using a combination of single-cell and spatial transcriptomics, chromatin accessibility assays and imaging. Using this approach, we identified novel transcription factors and cell states in human germ and supporting cell lineages. We compared them with other mammalian species and found primate-specific regulatory programmes. Our data identified cell context–specific interactions shaping sex specification and development of human germ cells. We defined a novel bipotent progenitor cell (*LGR5+, TSPAN8+*) in late embryos that can differentiate into early Sertoli in males or pre-granulosa cells in females. In fetal ovaries, we defined two subsets of pre-granulosa cells supporting germ-cell differentiation and distributed across the cortico-medullary axis. We also found a subset of developing granulosa cells appearing during the second trimester of pregnancy that is involved in follicular assembly. In fetal testes, we defined a novel supporting population (sPAX8 cells) located at the poles of the developing testis cords. We also found two tissue-resident myeloid populations that we named microglia-like and *SIGLEC15+* fetal testicular macrophages. This study provides an unprecedented spatiotemporal map of human gonadal differentiation that can be utilised as a blueprint for *in vitro* gametogenesis.

Gonadal development is a complex process requiring tight coordination of multiple cell lineages in time and space, some of which originate outside the gonads. Reproductive health of adults depends on these early developmental stages. Alterations in gonadal development cause Differences in Sex Development (DSD), conditions with discrepancies between chromosomal, gonadal and phenotypic sex that are linked to infertility and, to a lesser extent, cancer [12]. In addition, gonadal dysfunction is associated with infertility, cancer and polycystic ovary syndrome, the most prevalent endocrinopathy among women of reproductive age [3,4] . Limited access to fetal tissue, lack of *in vitro* models, and critical differences between humans and other mammals have hampered our knowledge of human gonadogenesis [5] . Understanding this process will establish the foundation necessary for *in vitro* gametogenesis [6–9], improve diagnosis of gonadal disorders [10], and lead to design of effective therapies.

Primordial germ cells (PGCs) are the precursors of female and male gametes that transmit genetic information to the next generation. Specification of PGCs occurs during gastrulation at ~2.5 weeks after fertilisation (post-conceptional weeks, PCW); thereafter PGCs migrate to the genital ridge 3-5 weeks after fertilisation [5,11]. Sex determination is initiated in the somatic lineage by the action of *SRY*, the sex-determining gene located on the Y chromosome [12]. *SRY* expression leads to differentiation of Sertoli cells (supporting lineage), testosterone-producing fetal Leydig cells (steroidogenic lineage) and other non-steroidogenic interstitial cells in XY gonads [13]. In XX gonads, somatic cells differentiate into pre-granulosa (supporting lineage) *via* Wnt4/Rspo1/β-catenin signaling and into ovarian stromal interstitial cells[14,15]. Sex determination in the somatic compartment instructs germ cell differentiation as pre-spermatogonia in XY or oogonia and oocytes in XX gonads. Male germ cells enter mitotic arrest, with germ cells and pre-Sertoli cells organising into cord-like structures surrounded by Leydig and other interstitial cells. At puberty, the testicular cords differentiate into seminiferous tubules where the male gametes reside. In contrast, female germ cells enter an asynchronous transition from mitosis to meiosis after ~10 PCW. Initially, oogonia develop in clusters named germ-cell nests that begin to break apart at ~17 PCW, when granulosa cells surround primary oocytes to form primordial follicles that further differentiate into mature follicles at puberty.

Single-cell transcriptomics analyses on gonadal development have started to shed light on this process in humans [16–19]. These studies describe the main transcriptional changes in germ and testicular somatic cell differentiation. However, holistic studies integrating multiple regulatory layers in space and time at the single-cell level are still missing. In addition, although much is known about development at puberty [20,21], a detailed map of the anatomical organisation of the testis and ovary at a cellular level during fetal development is lacking. Understanding the close communication between germ and somatic cells that shapes sex specification and differentiation is important, as well as identifying additional somatic cell lineages, such as tissue-resident macrophages, known to play a role in gonadogenesis in mice [22].

Coupled analysis of single-cell and spatial transcriptomic atlases of tissues have laid the groundwork to understanding human development [23,24], mapping disease [25,26] and engineering organoids [27,28]. In addition, emerging studies of chromatin accessibility are revealing the first regulatory networks that have the potential to fine-map active transcription factors (TFs) during cell specification and associate genetic variants with cell type specificity [29–31]. Here, we profile 43 samples of human fetal gonadal and extragonadal tissue from the first and second trimester using a multi-omics approach that combines single-cell and spatial transcriptomics, single-cell open chromatin profiling, and imaging. We compare our regulatory profiles to other mammals, gaining insight into the unique cell states and regulatory programmes activated in humans. We

identify novel somatic cells: a bipotent supporting progenitor in the early gonad; two subsets of pre-granulosa and a subset of developing granulosa cells in the developing ovary; a supporting cell population plumbing the poles of the forming cords in the developing testis; and two embryonic-derived testicular macrophages. We propose functions for the novel subsets according to their cellular neighbourhood and identify their potential implications in distinct gonadal conditions.

Results

A single-cell multiomics and spatial map of developing gonads and extragonadal tissue

We acquired gonad and extragonadal tissue from 43 embryos/fetuses in the first and second trimester of gestation (Carnegie Stage, CS16 -around 6 PCW- to 21 PCW) **(Fig. 1a-b)**. Samples were classified as female (n=23) or male (n=20) based on karyotype and sex-specific gene expression **(Supplementary Table 1)**. We used multiple genomics methods: i) single-cell RNA sequencing (scRNA-seq), ii) spatial transcriptomics (10X Genomics Visium), iii) single-cell accessible chromatin sequencing (scATAC-seq) and iv) combined single nuclei RNA and ATAC sequencing (snRNA-seq/scATAC-seq). Gonadal and extragonadal tissues were processed in parallel and loaded in the same channel of the 10x Genomics chip.

After quality control, we recovered 319,081 single-cell gene expression profiles that we integrated using Harmony [32] **(Fig. 1c-d, Supplementary Fig. 1a-b, Supplementary Table 2)**. We identify twelve main populations based on their expression of known markers **(Supplementary Fig. 1b, Supplementary Table 3)**. Some subsets are exclusively present in the gonads: (i) germ, (ii) developing pre-granulosa and other supporting (enriched in females), (iii) Sertoli (males-specific) and (iv) mesothelial (including coelomic epithelium). Other subsets are only existent in extragonadal tissue: (i) epithelial, and (ii) metanephric cells. Finally, other subsets are common to both gonads and mesonephros: (i) endothelial, (ii) perivascular (PV), (iii) erythroid, (iv) immune and (v) mesenchymal cells. Neural and metanephric cells are mostly in samples at CS16-18 (=<7 PCW) **(Figure 1e, Supplementary Fig. 1b)**. Further sub-clustering of the major populations reveals 65 cell states that we annotated using known markers and the most distinctly expressed gene **(Supplementary Fig. 1c; Supplementary Fig. 2a-b)**. Our comprehensive single-cell atlas of first- and second-trimester gonadal development is accessible at www.reproductivecellatlas.org (user: rca, password: $Uj6mPXA).

Although we find typical endothelial, perivascular and erythroid cell populations in both gonadal and extragonadal tissue, mesenchymal cells differ based on location. We leveraged Visium to spatially resolve transcripts in mesenchymal cells that reside inside and outside the gonadal tissue. We find that the expression of *GATA4, LHX9* and *ARX* is exclusive to the gonads while expression of *GATA2* and *NR2F1* is restricted to extragonadal tissue **(Fig. 1f-g, Supplementary Fig. 3a)**. Based on spatial information, and the expression of known and novel markers, we further annotated the mesenchymal clusters **(Supplementary Fig. 2b)**. We refined our annotations by mapping the cells back into the tissue using cell2location [33] **(Supplementary Fig. 3b)**. Within the gonadal mesenchymal cells we find testis-specific

3

interstitial cells in males, including fetal Leydig cells, ovary-specific stromal cells in females and a population located at the dorsal mesentery shared by both sexes.

145

To gain insight into the chromatin landscape that shapes developing gonadal cells, we profiled 67,202 and 30,276 cells for female and male gonads, respectively, using scATAC-seq **(Supplementary Table 1-2)**. Batch effects were corrected using Harmony [32] and cell populations were annotated leveraging scRNA-seq data from matched individuals [34] **(Fig. 1h-**

150 **i, Supplementary Fig. 4a-d)**. Cell identity assignment was validated by snRNA-seq data from the combined snRNA-seq/snATAC-seq profiling **(Supplementary Fig. 4e)**. We annotated a total of 17 populations in females and 15 in males and used chromatin accessibility peaks to: i) identify TF motifs that are differentially accessible in the distinct cell lineages [35] and ii) define genomic regulatory cis-co-accessibility networks (CCANs) specific to each cell type

155 using Cicero [36] **(Supplementary Fig. 4f-g, Supplementary Table 4)**. Our scATAC-seq data confirms that accessibility of *GATA4, LHX9, ARX, GATA2* and *NR2F1* clearly differs between gonadal and extragonadal mesenchymal populations in both females and males **(Supplementary Fig. 4h-i)**.

160 Furthermore, we made use of our multiomics map to assign cell type specificity to the genes that are frequently mutated in Differences in Sex Development (DSD) [1] **(Supplementary Fig. 3c-d)**. We find expression and motif activity of DSD-associated genes enriched in pre-granulosa (*EMX2, FOXL2*), Sertoli (*SOX9/10*) and germ cells (*DMRT1*). Additional cell type-specific genes are described in further sections of our manuscript as we delve into each cell

165 lineage.

Multiomics integration uncovers unique TF signatures modulating germ cell differentiation in humans

170 Our reanalysis of single-cell transcriptomics and chromatin accessibility of germ cells uncovers multiple cell states: i) primordial germ cells (PGCs); ii) fetal germ cells (FGC; *NANOG-*, *DDX4+*); iii) pre-spermatogonia (*PIWIL4+*); iv) oogonia, including pre-meiotic (*STRA8+*) and those initiating meiosis (*SYCP1+*); and v) primary oocytes (*GDF9+*) **(Fig. 2a-d)**. PGCs are found at all stages of development, but their proportion decreases with fetal age, representing

175 100% of total germ cells at 6-7 PCW and dropping to 12% and 40% at 17-21 PCW in females and males, respectively. In females, the *STRA8* surge, which marks the initiation of meiosis [37], appears at 11 PCW. Oocytes are apparent by 14 PCW and increase in frequency as they develop, accounting for 30% of female germ cells by 17-21 PCW. In males, pre-spermatogonia appear at 11 PCW and increase to nearly 60% by 17-21 PCW **(Fig. 2b)**.

180

We used spatial transcriptomics and cell2location [33] to map cell states to the tissue architecture. The fetal ovarian cortex remains PGC rich at least until 17 PCW (latest stage analysed by Visium) **(Fig. 2e)**. The ovarian medulla evolves to embrace a heterogeneous composition of states including mature germ cells and retinoic acid signaling-responsive oogonia at 11 PCW,

185 meiotic oogonia at 14 PCW and oocytes at 17 PCW **(Fig. 2e, Supplementary Fig. 5a)**. Female germ cells differentiation across the forming cortico-medullary axis was validated by single-molecule fluorescence *in situ* hybridisation (smFISH) with RNAscope probes **(Fig. 2f)**. In contrast, PGCs in males are located in the forming testicular cords **(Fig. 2f, Supplementary Fig. 5b)**.

190

We then measured the regulatory dynamics of the TFs mediating germ-cell differentiation by taking into account their expression and activity, the latter measured by (i) the expression of consensus targets; and/or (ii) the chromatin accessibility of their binding sites in the germ cell scATAC-seq data **(Fig. 2g, Supplementary Fig. 5c-e, Supplementary Table 5)**. As expected, pluripotency factors (eg. *POU5F1, KLF4*) are expressed and active in PGCs, while genes regulating oogenesis (eg. *FIGLA, SOHLH1*) and spermatogenesis (eg. *ID4, TCF3*) are only active in female or male germ cells, respectively **(Fig. 2g)**. Overall, there is a drastic change in the TFs and chromatin landscape when transitioning from PGCs to pre-spermatogonia. Conversely, the process of oogenesis is accompanied by more gradual changes in TF and chromatin accessibility dynamics **(Supplementary Fig. 5f)**.

We find novel TFs mediating germ cell differentiation in humans and compare their expression in other mammalian species by re-analysing publicly available scRNA-seq datasets of developing testes (E10 to E16) and ovaries (E10 to E16 and E11.5 to P5) from mouse, as well as developing ovaries from macaque (E84 and E116) [38–40] **(Supplementary Fig. 5g-j, Supplementary Table 6)**. Murine scRNA-seq data included pre- and postnatal samples to accommodate the different dynamics of germ cell differentiation and follicle development [41]. One of the TFs, *SOX4,* is expressed in PGCs and spermatogonia but is downregulated during oogenesis **(Fig. 2h)**. This expression pattern is conserved in mouse and macaque. The transition from PGCs to pre-spermatogonia involves the activation of the novel regulators *EGR4, KLF7, TWIST1* and *NFIL3*, displaying the same expression pattern in mice.

Fetal oocyte differentiation is more complex than its male counterpart, as it involves committing to the oogenic fate and transitioning from mitosis to meiosis. At the time of the *STRA8* surge, we find activation of *ZGLP1*, the oogenic TF recently described in mice [42]. The homeobox factors *HOXA* are upregulated in humans, a result that has not been previously described to our knowledge **(Fig. 2g)**. The activation of these factors is not conserved in macaque and mice **(Fig. 2h)**. At the meiotic stage, the TFs *DMRTC2* and *ZNF711*, previously described in mice, are active. We also identify another conserved DMRT member, *DMRTB1*, which is upregulated in macaque and, to a lesser extent, in mouse oogonia **(Fig. 2g-h)**. In oocytes, we find activation of *TP63, TSC22D1,* conserved in macaque and mouse, and *ZHX3*, conserved in macaque but not mouse **(Fig. 2g-h)**. *TP63* is a gene associated with ovarian insufficiency [43], showing the relevance of our TF roadmap to human germ cell development.

Human bipotent supporting cells are defined by co-expression of LGR5 and TSPAN8

Gonadal supporting cells guide mammalian germ cell differentiation and maturation. In mice, supporting cells arise in, at least, two waves [39,44,45]. During the first wave, bipotent progenitor cells derived from the coelomic epithelium differentiate into pre-granulosa cells (females) or alternatively, if the Y-linked gene *SRY* is expressed, into Sertoli cells (males). This first wave of supporting cells determines the sex specification of the gonad, which will in turn trigger the subsequent developmental waves. In humans, the bipotent cells giving rise to the first wave of male and female supporting cells have not yet been defined. Amongst the cells of the supporting lineage, we find a new population, sLGR5, which expresses known stem cells markers (eg. *LGR5, TSPAN8*) and other gonadal TFs (eg. *GATA4*) **(Fig. 3a-c, Supplementary Table 7)**. sLGR5 is enriched in the late stages of embryo development CS16-CS22 (~6–8 PCW) and over-expresses *SRY* in males **(Fig. 3b-c; Supplementary Fig. 6a, Supplementary Table 7)**. Other sexually dimorphic genes present in sLGR5 include *CAV1* in females, an inhibitor of

TGFβ signalling [46] and a checkpoint in epithelial-mesenchymal transition (EMT) [47] **(Fig. 3d, Supplementary Table 7)**. Additionally, sLGR5 expresses *DMRT1* and *NR2F2*, both associated with DSD **(Supplementary Fig. 6b)**.

The expression profile of sLGR5 points at these cells as the bipotent progenitor that differentiate into Sertoli cells or the first wave of pre-granulosa cells. We then set out to validate its spatial location in developing testes and ovaries at early stages (CS19-CS20) using multiplexed RNA Scope **(Fig. 3e; Supplementary Fig. 6c)**. sLGR5 is mainly located in the medulla, co-existing with the first wave of pre-granulosa cells (*OSR1+*) in females, and in close contact to developing Sertoli cells (*SOX9+*, *LGR5-*) in males. Therefore, the spatiotemporal dynamics of sLGR5 is in keeping with this population being the bipotent progenitor.

By 8 PCW, there are increased numbers of sex-specific supporting cells in males and females **(Fig. 3b)**. In ovaries, three main subsets within the developing granulosa lineage (*FOXL2+, IRX3+*) are present (see below). In testes, an early Sertoli population (Sertoli WFDC2) enriched between 8-10 PCW is connected to more advanced Sertoli cells **(Fig. 3a-b)**. Sertoli cells have a unique extracellular matrix (ECM) composition facilitating interactions with germ cells to form the testicular cords **(Fig. 3f, Supplementary Table 8)**. They express integrin alpha-6/beta-1 and *NCAM1* that interact with junction adhesion molecules (JAM) *JAM2/JAM3*, as well as other ECM proteins such as *LAMC1*, collagen IX (*COL9A1/COL9A3*) and *AGRN*. In the assembling testicular cords, specific signals triggered by the expression of *SOX9* in the supporting cells orchestrate specification to the male lineage **(Fig. 3f)**. Sertoli-to-germ cell signaling in the testicular cords is restricted to Inhibin β (*INHA/INHBB*) and *TGFB1*, known regulators of spermatogenesis. Indeed, cognate receptors for Inhibin β (*BMPR2* and *ACVR2A*) and *TGFB1* (*TGBR1-R2* complex) are upregulated in spermatogonia. *WNT5B*, which can activate the non-canonical WNT pathway in Sertoli cells, was one spermatogonia-specific ligand we identified. Sertoli cells also express oxytocin receptor, *OXTR*, which binds to its cognate ligand, *OXT*, expressed by all germ cells. In females, WNT signaling dominates in the distinct pre-granulosa subsets. In contrast, WNT ligands in the coelomic epithelium (*WNT2B, SOSTDC1, WNT5A)* are repressed in male supporting cells, with the exception of the non-canonical WNT ligand *WNT6* **(Supplementary Fig. 6d)**. In line with this, Sertoli cells do not express *AXIN2* , a well-defined target of the WNT canonical pathway [48] **(Supplementary Fig. 6d)**.

A novel supporting PAX8+ population controlling the gonadal-mesonephric interface

We find a supporting population in the gonads showing expression of *PAX8* and other specific somatic gonadal genes *(eg. GATA4)* that we name supporting *PAX8* (sPAX8) **(Fig. 3a)**. We define two main states for this population, sPAX8 early (sPAX8e) and sPAX8 male (sPAX8m), that have distinct temporal dynamics. sPAX8e is enriched in CS16-CS22 (~6–8 PCW) in both sexes, decreases in late 8 PCW and is barely detected after 13 PCW **(Fig. 3b, Supplementary Fig. 6a)**. sPAX8m is only present in males after 9 PCW **(Fig. 3b, Supplementary Fig. 6a)**. Although *PAX8* expression is characteristic of Mullerian and Wolffian ductal epithelium [49], both sPAX8 subsets express lower levels of *EPCAM* and do not cluster with ductal epithelial cells when analysed in the same manifold **(Supplementary Fig. 7a-d)**. Therefore, sPAX8 is a new supporting population, distinct from the ductal epithelial PAX8+ cells.

We used smFISH to characterise sPAX8 in male and female gonads. In the early samples (CS19 and CS20), *PAX8* expression is localised at the gonadal-mesonephric interface, between the mesonephric tubules and the gonadal medulla **(Fig.3g)**. Gonadal *PAX8+* cells are *EPCAM*low, similar to the other gonadal supporting and germ cells. This is different from the *EPCAM*high expression pattern characteristic of the epithelial cells in the mesonephric tubules. The proximity of sPAX8 (*EPCAM*low, *PAX8+*) and mesonephric tubules (*EPCAM*high, *PAX8+*) in early stages, forming a continuum, suggests that epithelial-mesenchymal transition (EMT) could be taking place as cells migrate into the gonad **(Fig.3g)**. In line with this, sPAX8 expresses the TF *SNAI2* involved in EMT [50], as well as the detachment-induced protein *PERP* [51], the suppressor of anoikis (apoptosis induced by detachment) *NTRK2* [52], and the protease *PLAU*, all involved in controlling cell adhesion and apoptosis **(Fig. 3c)**. We find differences between the location of sPAX8 in males and females in later stages (late 8 PCW onwards) **(Fig. 3h, Supplementary Fig. 7e-f)**. In developing testes, sPAX8 cells are in the medulla, at the poles of the developing testicular cords, a location consistent with the *rete testis* **(Fig. 3h-i, Supplementary Fig. 7e)**. In developing human ovaries, few sPAX8 cells are observed, maintaining their location at the ovarian-metanephric interface, near the hilum **(Supplementary Fig. 7f)**. Indeed, the presence of rudimentary *rete ovarii* has been described at this interface in early development, but degenerates in later stages [53].

The gonadal-mesonephric interface, where the sPAX8 is located, is a site of extensive tissue remodelling. At this interface, the movement of cells and their products within the gonads and formation of the *rete testis* are controlled **(Fig. 3i)**. Accordingly, sPAX8 cells display a unique transcriptional pattern of morphogens, suggesting a structural and supporting role for this population during development **(Fig. 3h)**. Early sPAX8e expresses *CXCL12/14* and its receptor *CXCR4*; the latter is also expressed by endothelial, epithelial and sLGR5, suggesting a chemotactic role for these populations **(Fig. 3j-k)**. Male sPAX8m expresses *NRP2*, which binds VEGF proteins expressed by endothelial cells and *SEMAB/C* on the epithelial cells **(Fig. 3j-k)**. Male sPAX8m also expresses additional molecules such as the somatostatin *SST* and *IGFBP3*, whose receptors are broadly expressed in Sertoli, pre-spermatogonia (*TMEM219*) and epithelial cells from the mesonephric duct (*SSTR2*) **(Fig. 3j-k)**.

We then aligned our single-cell signatures with publicly available mouse female gonad scRNA-seq data using a support vector machine (SVM) model trained on human supporting cells **(Supplementary Fig. 7g-h, Supplementary Table 9)**. We find a probable murine sPAX8 cell population that shares expression of *TBX1*, *TBX2* and *LYPD1* with its human counterpart and also distinctly expresses other markers, such as *Pdgfgl* and *Aldh1a3*, which are absent in humans **(Supplementary Fig. 7i)**. In mice, sPAX8 are also located inside the gonads close to the interface with extragonadal tissue **(Supplementary Fig. 7j)**. These results suggest main features of sPAX8 cells are conserved in mammals.

TFs defining the two waves of medullary and cortical pre-granulosa cells in humans

Unlike testis determination, where action of *SRY* induces formation of testicular cords by Sertoli cells, morphological changes in the ovaries only begin in mid-second trimester (~17 PCW) when granulosa cells surround the oocyte to form the primordial follicle. We investigated the transcriptional profile of the developing granulosa lineage in human ovaries and found 5 substates organised in 3 major states: pre-granulosa (preGC) I (preGC-I); preGC-

II (preGC-IIa, preGC-IIb, preGC-IIc); and developing granulosa **(Fig. 3a, Fig. 4a)**. preGC-I is a steroidogenic $FOXL2^{high}$ population that expresses *OSR1* and emerges at CS19 (~7.5 PCW) **(Fig. 4a, Supplementary Fig. 6a)**. This subset also expresses *WNT4* and *RSPO1*, both upstream regulators of granulosa lineage determination *via* the WNT/β-catenin pathway [15]. Indeed, the WNT-target, *AXIN2*, peaks in this subset **(Supplementary Fig. 6d)**. Right after this stage, CS23 (~8.5 PCW), we find preGC-IIa/b, which are $FOXL2^{low}$ and express *CYP26B1*, an enzyme that inactivates retinoic acid -a key molecule for induction of meiosis **(Fig. 4a, Supplementary Fig. 6a)**. PreGC-IIb also expresses WNT-related specific transcription factors, such as *CSRNP1*, as well as Nr4a nuclear hormone receptors (*NR4A1-3*) and early growth response (Egr) factors (*EGR1-4*). PreGC-IIc appear later (11 PCW) and are $FOXL2^{high}$ expressing *BMP2+* , a factor involved in oogenic induction **(Fig. 4a, Supplementary Fig. 6a)**. Developing granulosa cells are the last subset to appear, they are the most abundant subset in 17-21 PCW, and express *BMP2,* folliculogenesis markers (*NOTCH3, HEYL*) and *RDH10* **(Fig. 4a, Supplementary Fig. 6a)**.

To anatomically locate these three pre-granulosa subsets in fetal ovaries, we integrated single-cell and spatial transcriptomics data using cell2location [33] **(Fig. 4b, Supplementary Fig. 8a)**. Early preGC-I are located in the medulla, while preGC-IIa/b mapped to the outer cortex, subjacent to the ovarian surface epithelium (*LRRN4+, KLK11+*). preGC-IIc and developing granulosa are located in the transition zone between the outer cortex and the medulla. This region in the inner cortex will eventually become the location of the primordial follicles after ~17 PCW. The spatial relationship among female supporting populations is reflected in their global chromatin accessibility patterns **(Fig. 4c-d, Supplementary Fig. 8b-d, Supplementary Table 7)**. All pre-granulosa subsets share a similar co-accessibility profile as the ovarian surface epithelium, with the exception of early medullary pre-granulosa (preGC-I). Thus, as previously described in mice, there appear to be two waves of pre-granulosa cells (medullary and cortical) during human ovarian development [39].

We then aligned the female supporting populations with publicly available female gonadal scRNAseq data from macaque and mouse using a support vector machine (SVM) model **(Supplementary Fig. 8e-h, Supplementary Table 9)**. Overall, humans and macaques share similar pre-granulosa cell populations, but, intriguingly, we could not establish direct correspondence with mouse pre-granulosa cells. We also could not establish a match of our bipotent supporting cell sLGR5 in mice. *LGR5* expression is restricted to bipotent progenitors in humans, while in mice *LGR5* is expressed by the second wave of pre-granulosa cells, which are derived from the ovarian surface **(Supplementary Fig. 8i)**. This suggests that, despite the presence of similar developmental waves of pre-granulosa cells in mice and humans, the regulatory networks giving rise to them are distinct.

We therefore explored the unique TF modules regulating human pre-granulosa development integrating transcriptomics and chromatin accessibility, as we did with the germ cells **(Fig. 4e, Supplementary Table 10)**. Some of the TFs reflect distinct lineage relationships of preGCs. The first wave of medullary preGC-I activates the master regulator of the granulosa lineage, *FOXL2* [54] , and the novel TFs *STAT1*, *OSR1*, *ELF3* and *MAFB*, which are shared with the sLGR5 progenitor in females. The second wave of pre-GCs (preGC-II) expresses the previously described *ARX, LHX2/9* [55] and the novel *EMX2* and *ASCL1*, shared with the ovarian surface epithelium. Other TFs reflect environmental and functional differences. For example, preGC-I and preGC-IIc express WNT-induced TFs (*FOXO1, FOXP1, HIF1* or *FOXJ3),* suggesting there is a higher WNT environment deeper in the ovary. In support of this, the WNT modulators *SFRP1/2* are highly expressed in the ovarian surface and absent in any other cell

380   state in the medulla **(Supplementary Fig. 6d)**. The developing granulosa cells express TFs previously associated with germline nest breakdown and primordial follicle formation in mice, including NOTCH-induced TFs (*HEY2* and *HES4*) and *PEG3,* as well as novel TFs such as *TSHZ1*, *SOX5* or *TEF* **(Fig. 4a and 4e)**.

385   In humans, the numbers and quality of primordial follicles formed during fetal life determines the reproductive lifespan and fertility of women in adulthood [56]. Polycystic ovary syndrome (PCOS) is the commonest cause of anovulation in women of reproductive age and is associated with disordered growth and development of the follicles [57]. In order to explore the link between genetics and fetal origin of adult PCOS, we aligned cell type–specific co-accessible regions
390   from intragonadal female cells with polymorphisms associated with PCOS risk from the GWAS Catalog [58]. Out of 43 PCOS-associated variants exceeding genome-wide significance, 30 map to our co-accessible regions, with some regions containing multiple variants. The majority of these regulatory regions are enriched in preGC-II, developing granulosa and ovarian surface epithelium, but to a lesser extent in preGC-I, indicating that the cellular target
395   of PCOS could be supporting cells in the cortical and transition zones **(Fig. 4f)**. In addition, fine mapping of one variant, rs8043701-A, reveals that it might affect the co-accessibility pattern of *TOX3*, which is upregulated in preGC-II and developing granulosa cells **(Fig. 4g-h)**.

Crosstalk between germ and supporting cells shapes ovarian architecture

400   Our spatial analysis reveals a gradient of pre-granulosa and germ cells in the forming ovary. To systematically map gonadal cell–cell communication in the ovarian microenvironments we updated CellPhoneDB by doubling the number of manually curated interactions. We also included interactions mediated by small molecules such as steroid hormones. To do this, we considered expression of the last enzyme of the metabolic pathway synthesizing the small
405   molecule **(Supplementary Table 11)**. In addition, we developed CellSign, a novel module of CellPhoneDB v4, that links receptors with their downstream TFs, as assessed by manual curation. TFs are used as sensors of pathway activities, allowing us to identify sets of cell–cell interactions that affect downstream signaling predicted in the previous sections with a high degree of certainty **(Fig. 5a-b)**.
410
       In the outer cortex, primordial germ cells co-exist with cortical preGC (preGC-IIa and b), coelomic epithelium and the ovarian surface. CellPhoneDB reveals that cortical supporting cells are a major source of paracrine factors for the PGCs, such as *BMP4* (coelEpi) and *KITLG* (preGC-II), both required for the chemotaxis and maintenance of PGCs **(Fig. 5c)**. *STAT1* and
415   *STAT3*, downstream of *KIT*, are accordingly active in PGCs **(Fig. 5d)**. We also detect novel interactions between supporting cortical cells and PGCs mediated by ephrins, *EFNB2/EFNB3* (coelomic epithelium), with their downstream effector *ELK1* activated in PGCs **(Fig. 5c-d)**. Moreover, we observe a distinctive pattern of interacting extracellular matrix proteins between the coelomic epithelium and PGCs, suggesting a possible mechanism for PGCs accumulation
420   **(Fig. 5e)**. Among these adhesion molecules, we find several collagens (such *COL5A2*, *COL8A1* or *COL9A3*) and integrin β4 (*ITGB4*) **(Fig. 5e).**

       PreGC-I in the medulla as well as preGC-IIc and developing granulosa cells in the transition zone express molecules that mediate oocyte differentiation **(Fig. 5f)**. preGC-I upregulate the
425   enzymes involved in the production of estrogen, *HSD17B1* and *CYP19A1*, which could bind the estrogen receptor *ESR1* expressed by oogonia STRA8+ **(Fig. 5c)**. preGC-IIc expresses the enzyme *ALDH1A2* involved in the synthesis of retinoic acid, whose product will bind to the

retinoic acid receptor *RARA* and transporter *CRABP2* expressed by oogonia STRA8+ **(Fig. 5c)**. PreGC-IIc and developing granulosa cells also overexpress *BMP2*, recently shown in mice to activate the oogenesis factor *ZGLP1* [42], which we predict as specifically active in the oogonia STRA8+ **(Fig. 5c-d)**. Activation of BMP-2 signaling is associated with parathyroid hormone ligand, *PTHLH*, and its receptor, *PTH1R*. PreGC-IIc and oogonia overexpress *PTHLH* and *PTH1R,* respectively [59,60].

Around 17 PCW, granulosa cells in the transition zone surround germ cells and begin to form follicles, a process that involves bidirectional communication between the two cell types. For example, developing granulosa cells express *NOTCH2/3* receptors, which are inhibited *via* *DLL3* by PGCs, regulated *via* *DLK1* by oogonia, and finally activated *via* *JAG1* by oocytes to induce folliculogenesis in females **(Fig. 5c)**. Other factors expressed by oocytes are *GDF9*, key for folliculogenesis, and enzymes involved in histamine and steroid hormone biosynthesis. Their cognate receptors are exclusively expressed by developing granulosa cells. Finally, we find a unique expression pattern of extracellular matrix proteins in the oocytes and developing granulosa that will likely participate in the formation of the follicle **(Fig. 5c)**. For example, oocytes in the medulla upregulate *ITGA4, ITGA11* and E-cadherin (*CDH1*) to attach to the developing granulosa cells *via* collagen IV (*COL4A1-4*) and integrin α-2/β-1 (*ITGA2/ITGB1*) **(Fig. 5e)**.

Finally, we investigated if the main roles of the pre-granulosa lineage subsets are preserved in mice, despite the transcriptomic differences we have shown between rodents and primates. We plotted the main identified ligand/receptor interactions in the medullary and cortical pre-granulosa subsets from the mouse dataset. Despite major differences between murine and human pre-granulosa cells, the first-wave of murine medullary pre-granulosa cells expresses enzymes involved in estrogen synthesis (eg. *HSD17B1*) while the second wave of murine cortical pre-granulosa cells expresses the oogenic factor *BMP2* **(Supplementary Fig. 8j)**. This suggests that the location and main functional modules that define the first and second wave of pre-granulosa cells in primates and rodents are conserved.

Two subsets of tissue-resident macrophages in the developing testis

Tissue-resident immune cells are thought to be important in gonadal development and function [22,61]. To further characterize the immune compartment of the fetal gonads, we sorted cells from 11 samples using the pan-leukocyte marker CD45 **(Supplementary Fig. 9a, Supplementary Table 1-2)**. Our integrated immune dataset defines 19 clusters, including haematopoietic stem cells/multipotent progenitors (HSC/MPP), megakaryocytes and mast cells, myeloid cells, B cells and other innate lymphocyte lineages **(Fig. 6a; Supplementary Fig. 9b-d, Supplementary Table 12)**. Myeloid cells are the most abundant during the first and second trimester, accounting for over 50% of CD45+ cells in the gonads **(Supplementary Fig. 9e)**.

We find a unique composition of tissue-resident gonadal macrophages. In both sexes, a population is present with a tissue-repair phenotype (*LYVE1, F13A1, STAB1, FOLR2, CD163*) previously defined in other organs [62–64] **(Fig. 6b-c; Supplementary Table 13)**. One subset of tissue-repair macrophages expresses genes induced by IFNγ (*CXCL9, CXCL10, CXCL11, GBP1*) and TFs activated by IFNγ (*STAT1, STAT2, IRF1, RFX5*) **(Fig. 6b-c; Supplementary Fig. 9f-g)**. Unique to the developing testes, we find a new population of macrophages expressing genes characteristic of microglia (*CX3CR1, P2RY12, C3, OLFML3, HTRA1,*

TREM2) [63,65,66] that we named fetal microglia-like cells **(Fig. 6b-d; Supplementary Fig. 9f-g)**. Projecting fetal microglial cells [63] onto our dataset further confirms that the cluster of microglia-like testicular macrophages has this unusual signature **(Fig. 6e)**. Interestingly, we also define *SIGLEC15+* testicular macrophages in the developing testis **(Fig. 6b-c; Supplementary Fig. 9f-g)**. In addition to expressing *SIGLEC15*, this unique testicular macrophage population shares additional markers with osteoclasts, including *MMP9, ACP5, CTSK and ATP6V0D2* [67–70] **(Fig. 6c)**. The expression patterns of our novel tissue-resident testicular macrophages do not match with surface markers defined for peritubular and interstitial populations in postnatal mice [22,71–73] **(Fig. 6f)**.

To study the unique profile of gonadal macrophages, we integrated our dataset with publicly available scRNA-seq datasets for myeloid cells in the fetal liver, skin, kidney, yolk sac, gut, thymus and placenta [62,74–76] **(Fig. 6g; Supplementary Fig. 9h-i, Supplementary Table 13)**. As expected, most fetal macrophages share the characteristic tissue-repair signature **(Fig. 6h)**. In the absence of scRNA-seq data from fetal bone, *SIGLEC15+* macrophages are exclusively present in developing testes. Microglia-like testicular macrophages cluster together with a similar subset in skin, suggesting these unusual immune cells might have a broader role beyond their function in the brain.

Using smFISH, we confirmed the presence of tissue-repair macrophages in developing gonads of both sexes and further investigated the anatomical location of microglia-like and *SIGLEC15+* macrophages in the fetal testes. Tissue-repair macrophages in the developing testes and ovaries are interstitial, co-existing with other interstitial and mesenchymal lineages **(Supplementary Fig. 10a)**. *SIGLEC15+* macrophages are found outside the testicular cords, in close proximity to endothelial cells **(Fig. 6i; Supplementary Fig. 10b-c)**. Microglia-like macrophages are usually found inside the testicular cords, with the potential to contact Sertoli and germ cells **(Fig. 6j; Supplementary Fig. 10d-e)**.

CellPhoneDB v4 identifies specific interactions between *SIGLEC15+* macrophages and intragonadal mesenchymal and endothelial cells through the expression of extracellular matrix proteins **(Supplementary Table 14)**. For example, *SIGLEC15+* macrophages uniquely express *COL1A2*, which interacts with the integrin complexes (α1/β1, α2/β1, α10/β1, and α11/β1) expressed by endothelial and mesenchymal cells **(Fig. 6k)**. This suggests a role for *SIGLEC15+* macrophages in promoting mesonephric endothelial cell migration [77]. During early development, endothelial cells enter the gonad to generate the coelomic vessel that is required for testicular cord formation [77] **(Fig. 6l)**. Accordingly, *SIGLEC15+* macrophages seem to disappear after 14 PCW, as observed by smFISH **(Supplementary Fig. 10b-c)**. Microglia-like macrophages are predicted to have distinctive cell-cell communication with Sertoli and germ cells *via* the interaction between *TREM2* and apolipoproteins (*CLU, APOA1, APOE*) **(Fig. 6k)**. Together with the expression of immunomodulatory molecules (*IL10, TGFB1, ENTPD1*), a potential function for microglia-like macrophages is in supporting an immunoregulatory environment inside the developing cords to avoid inflammation or oxidative stress that could damage maturing germ cells **(Fig. 6l)**.

Discussion

Generating a cell atlas of gonadal development is essential to define the cell states and pathways involved in sex specification and divergent differentiation of ovaries *versus* testes. Our multiomics map, profiling more than one third of a million cells, will improve understanding of gonadal conditions including DSD, infertility and cancer. Studying gonadal development is challenging, because it is a highly dynamic process that involves tight coordination of multiple cell lineages, including PGCs and the formation of oogonia and spermatogonia in females and males, respectively. Here, we have used single-cell transcriptomics, chromatin accessibility and spatial transcriptomics methods to generate a comprehensive dataset of gonadal and extragonadal cells throughout embryonic and fetal development browsable online at reproductivecellatlas.org (user: rca, password: $Uj6mPXA). We discover novel supporting and immune cells, compare them to other mammalian species and thereby reveal primate-specific regulatory programmes. We identify the key interactions between somatic and germ cells that orchestrate the formation of male and female gonadal structures during fetal development, and use this information to describe the potential function of new cell types.

Building on previous work profiling the transcriptome of human germ-cells [5,78,79], our joint analysis of transcriptome and chromatin accessibility reveals novel TFs involved in germ cell differentiation. Our multiomics analysis allowed us to prioritise TFs that are both expressed and active. Our multi-species comparison highlights a conserved TF network in primates with multiple differences from their murine counterparts, in agreement with previous murine-human comparisons of primordial germ cells [5] and spermatogonia [80]. Our work also reinforces the relevance of development in postnatal disease. For example, our analysis highlights the expression of *TP63* in oocytes, a gene associated with ovarian insufficiency [43].

When developing germ cells enter the gonadal ridge (~5 PCW), sex-specific programmes are activated in somatic cells of the developing testes and ovaries (~6 PCW). We define for the first time in humans a bipotent supporting progenitor expressing unique markers (*TSPAN8+, LGR5hi, GATA4+*) between 6-8 PCW. *SRY* expression peaks in this population in males. This bipotent population is likely to be derived from the coelomic epithelium [81] and gives rise to either the first wave of pre-granulosa cells or, if *SRY* is expressed, to Sertoli cells. A bipotent supporting progenitor has been previously described in mice [12], although the murine markers (*WT1+, NR5A1+*) are broadly expressed by all supporting populations in humans.

The formation of testicular cords is a major event, indicating the start of male sex specification. We have identified essential interactions between Sertoli and germ cells that orchestrate differentiation and sex specification. Formation of the cords critically depends on testicular vascular and mesenchymal cell migration from the adjacent mesonephros to surround Sertoli-germ cell aggregates [77,82,83]. This process has to occur in a well-organised manner. We identify a supporting sPAX8 population (*WNT6+, GATA4+*) that is located in the gonadal-mesonephric border, peaks at 8 PCW, and is conserved in mice. After 13 PCW, this population is virtually absent in females and remains close to the developing cords in males. sPAX8 has a unique expression of chemokines and paracrine factors indicating this population has a role in structural organisation by creating gradients of morphogens. Specifically, sPAX8 is located in one of the poles of the forming cords, suggesting it is involved in the formation of the *rete testis,* a complex network of tubules, into which the testis cords anastomose [84,85]. Indeed, PAX8+ cells are present in the adult *rete testis* [49], the site of initial seminiferous tubule's fluid reabsorption. Altogether, this population may have a structural role organising the endothelial, mesenchymal and epithelial cells that will form and connect the seminiferous tubules, *rete testis*, efferent ductules and the head of the epididymis.

In addition, we find that the developing testes harbor two unusual tissue-resident macrophages that are likely to participate in the testicular cord formation, and that we named microglia-like and *SIGLEC15+* macrophages. *SIGLEC15+* macrophages, found in the peritubular space surrounding the testicular cords, interact with mesenchymal and endothelial cells *via* extracellular matrix proteins and are transcriptomically similar to osteoclasts, macrophages involved in bone reabsorption. Our results echo previous research showing that osteoclasts contribute to bone angiogenesis through a mechanism involving metalloproteinase-9 [86], an extracellular matrix protein that is specifically expressed by *SIGLEC15+* testicular macrophages. The spatial location and predicted interactions of these *SIGLEC15+* macrophages suggest a possible role in mesonephric endothelial cell migration [77]. Indeed, *SIGLEC15+* macrophages peak around 11-12 PCW, coinciding with the period of testicular endothelial remodelling.

The other macrophage population, microglia-like cells, are located primarily inside the testicular cords in close contact with Sertoli and germ cells and express immunomodulatory molecules. The transcriptome of these microglia-like cells is similar to those of fetal microglia and a subset of macrophages in the fetal skin. This transcriptional programme shared across diverse tissues suggests that they have a protective function in dampening damaging immune responses. The testes need to maintain an immunomodulatory environment to prevent immune cells from attacking neoantigens appearing with the onset of spermatogenesis, which only starts during puberty [87,88]. Similarly, the anagen hair follicle in adults is considered a relatively immunosuppressive environment where antigen-presentation responses are strongly downregulated [89]. This is the first description of immune cells derived from the embryo inside the human fetal cords. Future work will determine if these macrophages play a role protecting germ cells in postpubertal samples.

In females, the developing ovaries undergo dynamic cellular and morphological changes to form a highly organised tissue structure that contains the primordial follicle reserve required for sustained female fertility. To systematically quantify the dynamic cell-cell interactions in the distinct spatial microenvironments, and to improve the accuracy of our predictions, we have developed a novel resource named CellSign, which links receptors to their downstream transcription factors. This module is now incorporated into CellPhoneDB.org v4.0.

After their arrival at the gonadal ridge, PGCs remain in the outer ovarian cortex, which is rich in extracellular matrix proteins and chemoattractants. Their pattern of adhesion molecules then alters, allowing them to migrate towards the outer medulla, where they will further differentiate. We show that pre-granulosa cells form a gradient in the ovarian tissue that will regulate the distinct stages of germ cell development. As previously described in mice [39,44,45], we define two waves of pre-granulosa cells. The first wave (preGC-I) derives from bipotent sLGR5 and will likely form the primary sex cords in females [90]. preGC-I cells produce estrogen, which is recognized by oogonia, indicated by the expression of estrogen receptors in oogonia STRA8+ (pre-meiotic). The second wave of *LHX2+* pre-granulosa cells, derived from the ovarian cortex, is distributed in the cortex (preGC-IIa-b) and transition zone, close to the medulla (preGC-IIc and developing granulosa). These will develop into the secondary sex cords. Cortical preGC-II (preGC-IIa-b) express *CYP26B1*, which inactivates retinoic acid, and ligands whose cognate receptors are expressed in PGCs. In contrast, transition zone preGC-IIc express ligands involved in initiation of meiosis (eg. retinoic acid by *ALDH1A2*) and oogenesis (eg. *BMP2, PTHLH*). Markers of folliculogenesis (eg. *NOTCH2/3*) are only active in developing granulosa cells at 17 PCW or older. Human pre-granulosa subsets show a different

transcriptomic profile compared to their murine counterparts, reinforcing the importance of continued study of their regulatory networks. Interestingly, we see functional conservation in
625 estrogen and *BMP2* production between murine and human pre-granulosa subsets, indicating that the spatiotemporal functional compartmentalisation in the ovary is preserved across primates and rodents, even if the cell states and molecular regulators are different.

In conclusion, we have defined a comprehensive map of cells, molecular and cellular
630 programmes of gonadal differentiation as a unique resource to understand gonadal function and pathology. DSD are rare conditions associated with infertility and cancer. Many of these disorders involve pathology in other organs, (eg. kidney, heart, peripheral and central nervous systems) [91]. Our resource will help with prompt diagnosis of DSD by providing an accurate interpretation of sequencing datasets, finding novel therapeutic candidates, and improving
635 clinical outcomes and quality of life. Our study will also help understand the pathogenesis of reproductive cancers; for example, ovarian cancer recapitulates fetal programmes and germ-cell tumours originate during fetal development [92]. Finally, our atlas can be used as a blueprint for the design of protocols for *in vitro* gametogenesis from reprogrammed somatic cells, with relevance for infertility treatments.

640

Figure legends

Fig. 1 Single-cell profiling of gonadal and extra-gonadal tissue. a, Schematic illustration of
645 gonadal development showing the main structures of the XX and XY gonads. **b,** Diagram summarising stages and samples collected in our study along with major events occuring during gonadogenesis. **c,** UMAP (uniform manifold approximation and projection) projections of scRNA-seq data from 43 individuals (n = 319,081). Clusters for supporting, Sertoli and mesenchymal cells were defined in independent per-lineage re-analyses for these populations
650 (see Supplementary Figure 1-2). **d,** UMAP coloured per sex. Pink = XX cells, blue = XY cells. **e,** Barplot showing the percentage of cells per PCW classified by cell types identified in the main UMAP. **f,** Dot plot showing variance-scaled, log-transformed expression of transcription factors (TFs) differentially expressed in gonadal and extragonadal mesenchymal cells. **g,** Spatial plot showing log-transformed expression of TFs in each Visium spot shown over the
655 H&E image of a female 11 PCW (above) and a male 12 PCW (below) sample including gonadal and extragonadal tissue. **h,** UMAP projections of scATAC-seq data from females (left) and males (right). **i,** Heatmap showing label transfer scores from scRNA-seq to scATAC-seq data of male samples. PV = Perivascular cell; coelEpi = Coelomic epithelium; sLGR5 = supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pre-
660 granulosa cells; M_ = mesenchymal from the mesonephros.

Fig. 2 Transcriptional signatures in human germ cells. a, UMAP projections of scRNA-seq data from female and male germ cells (n = 8,869). **b,** Barplot showing the proportions of germ cells and their annotation classified by sex and developmental stage. **c,** UMAP projections of
665 scATAC-seq data from female and male germ cells (n = 2,017). **d,** Dot plot showing the log-transformed expression of genes characteristic of the germ cell states. **e,** cell2location estimated amount of mRNA (colour intensity) contributed by each female germ cell population to each Visium spot (colour) shown over the H&E image of 14 and 17 PCW ovaries. **f,** High-resolution, large-area imaging of representative gonadal sections of two human fetal ovaries (11 and 14

670 PCW) and two testes (8 and 12 PCW), with intensity proportional to smFISH signal for *FIGLA* (cyan), *DAZL* (yellow), *POU5F1* (red). Dashed lines on magnified fields describe developing gonadal structures. **g,** Heatmaps showing TFs differentially expressed in germ cell states. Colour proportional to log-transformed expression. "o" = TFs whose binding motifs are differentially accessible (*i.e.* TF can bind their potential targets); "a" = TFs whose targets are

675 also differentially expressed (i.e. differentially activated TFs); and asterisk (*) = TFs that meet both "o" and "a" conditions. **h,** Dot plot showing the log-transformed expression of novel TFs relevant for human germ cell states in macaque and mouse scRNAseq datasets. PGC = Primordial germ cells; FGC = Fetal germ cells; DEG = differentially expressed gene; TC = Testicular cords.

680

Fig. 3 Novel subsets in human supporting lineage. a, UMAP projections of scRNA-seq data from female and male supporting and coelomic epithelial cells (n = 83,233). **b,** UMAP coloured per PCW and sex. **c,** Dot plot showing the log-transformed expression of genes characteristic of the supporting cell states. **d,** Volcano plot showing genes differentially expressed between

685 males and females within the sLGR5 population, highlighting SRY gene. **e,** High-resolution large-area imaging of representative gonadal sections of a human fetal ovary (7PCW, CS19) and testis (8PCW, CS20), with intensity proportional to smFISH signal for *KLK11* (cyan), *LGR5* (yellow), *OSR1* or *SOX9* (red); red blood cells appear as bright autofluorescent cells. White dashed rectangles highlight enlarged gonadal regions. Dashed lines on magnified fields

690 in the XY sample describe developing testicular cords. **f,** Dot plots showing scaled log-transformed expression of genes coding for interacting ligand-receptor proteins in supporting and germ states. List of interactions available in Supplementary Table 8. **g,** High-resolution large-area imaging of representative gonadal sections of a human fetal ovary (7PCWs, CS19) and testis (8 PCWs, CS20), with intensity proportional to smFISH signal for *EPCAM* (red),

695 *NR5A1* (cyan) and *PAX8* (yellow); red blood cells appear as bright autofluorescent cells. **h,** High-resolution large-area imaging of representative gonadal sections of two human fetal testes -8 PCWs (> CS23) and 12PCW-, with intensity proportional to smFISH signal for *EPCAM* (red), *NR5A1* (cyan) and *PAX8* (yellow). White dashed rectangles highlight enlarged gonadal regions with *PAX8*^{high}/*EPCAM*^{low} expression. White dashed lines on magnified fields describe

700 developing testicular cords. **i,** Schematic representation of sPAX8 cells in the human fetal testis at two developmental stages. **j,** Dot plot showing the log-transformed expression of genes coding for sPAX8 ligands and receptor proteins in the supporting cells. **k,** Dot plot showing the log-transformed expression of genes coding for sPAX8 receptor proteins in the germ and endothelial cells. coelEpi = Coelomic epithelium; ovarianSurf = Ovarian surface; sLGR5 =

705 supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pre-granulosa cells; ECM = extracellular matrix; MD = Mullerian Duct; WD = Wolffian Duct; WT = Wolffian Tubules; DMN = Degenerating Mesonephric Nephron; UC = Urogenital Connection; RT = Rete Testis; ED = Efferent Ductule; DE = Ductus Epididimydis; DMD = Degenerating Mullerian Duct; TC = Testicular Cord.

710

Fig. 4 Transcriptional signatures and location of human pre-granulosa cells. a, Dot plot showing the log-transformed expression of genes characteristic of the supporting cell states. **b,** Estimated amount of mRNA (colour intensity) contributed by ovarian surface epithelium, preGC-I, most frequent preGC-II and/or developing granulosa states to each Visium spot

715 (colour) shown over the H&E image of a 11, 14 and 17 PCW ovaries. **c,** UMAPs projections of scATAC-seq data from female and male supporting cells (n = 34,015). **d,** Hierarchical clustering of enrichment z-scores for peaks contained within each cis-co-accessibility network (CCAN) with respect to the cell states identified in scATAC-seq data of human supporting cells. **e,** Heatmaps showing TFs differentially expressed in female supporting cell states. Colour

proportional to scaled log-transformed expression. "o" = TFs whose binding motifs are differentially accessible (i.e. TF can bind their potential targets), "a" = TFs whose targets are also differentially expressed (i.e. differentially activated TFs), and asterisk (*) = TFs that meet both "o" and "a" conditions. **f,** Enrichment z-scores for peaks contained within CCANs including PCOS-associated single nucleotide polymorphisms (SNPs) with respect to the cell states identified in scATAC-seq data of human supporting cells. CCANs that contain multiple SNPs are marked with a number of asterisks (*) that corresponds to the number of SNPs. **g,** Co-accessibility and coverage plot of the genomic region of the CCAN including the SNP rs8043701-A and gene *TOX3*. Viewpoint in the co-accessibility plot is set to the genomic coordinates of the SNP rs8043701-A. **h,** Dot plot showing the variance-scaled, log-transformed expression of *TOX3* characteristic of pre-granulosa cell states. coelEpi = Coelomic epithelium sLGR5 = supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pre-granulosa cells; PV = Perivascular cells; DEG = differentially expressed gene.

Fig. 5 Cell-cell communication networks. a, (left) Diagram showing the information aggregated within the updated version of CellPhoneDB, which includes: (i) 534 novel (1,852 total) ligand-receptor interactions; (ii) 194 novel interaction mediated by small molecules; (iii) 186 novel curated links between ligand-receptor and transcription factors. (right) Diagram showing the new statistical framework to infer active cell-cell interaction partners. It includes an additional step to indicate active ligand-receptor partners in our data based on the activation of downstream signals on the receiver cell. Downstream signals are calculated based on TF expression and TF activity from scRNA-seq and scATAC-seq data. **b,** (Top) Schematic representation of pre-granulosa and germ cell states in the human fetal ovary. (Bottom) H&E imaging of a representative XX human fetal ovary section (21PCW). White dashed rectangles highlight follicles at different stages of maturation. Note the different morphology of Developing Granulosa cells surrounding the oocytes **c,** Dot plots showing scaled log-transformed expression of genes coding for interacting ligand-receptor proteins in supporting and germ cells states, in the cortex, medulla and formation of follicles. **d,** Heatmap showing the activity (i.e. expression level of bona fide targets) of TFs downstream receptors in germ and supporting cells. Colour proportional to scaled normalized enrichment score. **e,** Dot plots showing scaled log-transformed expression of genes coding for extracellular matrix (ECM) and adhesion molecules in germ and supporting cells states. **f,** Schematic illustration of main TFs, receptors, ligands and extracellular molecules regulating germ cell differentiation. PGC = Primordial germ cells; coelEpi = Coelomic epithelium; sLGR5 = supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pre-granulosa cells; ECM = extracellular matrix. Scale Bar = 250μm.

Fig. 6 Tissue-resident macrophages in the developing testes. a, UMAP projections of immune cells (n=19,538) in the gonads and mesonephros labeled by cell state. 11 samples were enriched for immune (CD45+) cells. **b,** Bar plot showcasing the proportion of male and female cells in each macrophage subset. **c,** Dot plot showing variance-scaled, log-transformed expression of marker genes for the identified macrophage subsets. **d,** Dot plot showing the variance-scaled, log-transformed expression of microglia-like markers in both sexes reveals that the few female cells that belong to this cluster do not express the key markers. **e,** Projection of microglia cells from Bian *et al.*, 2020 onto our gonadal immune manifold using scmap (left) and a SVM classifier (right). **f,** Dot plot showing the variance-scaled, log-transformed expression of marker genes of interstitial and peritubular mouse macrophages in our human gonadal macrophages. **g,** UMAP projections of myeloid cells from embryonic/fetal gonads integrated using single-cell Variational Inference (scVI) with myeloid cells from embryonic/fetal gut, kidney, liver, lung, placenta, skin, thymus and yolk sac (n=52,363). **h,**

770   Embryonic/fetal myeloid cell type abundance (% cells) in different organs. *SIGLEC15+* macrophages and microglia-like macrophages are two novel subsets present in XY gonads. *SIGLEC15+* macrophages are exclusive to XY gonads while microglia-like macrophages are found primarily in XY gonads and skin. **i,** High-resolution imaging of a representative XY gonadal section (12 PCW), with intensity proportional to smFISH signal to *PDGFRA* (green,

775   mesenchymal), *CDH5* (cyan, endothelial cells), *CD68* (red, macrophages), *SIGLEC15* (yellow, *SIGLEC15+* macrophages). *SIGLEC15+* macrophages distribute outside of the testicular cords in proximity to endothelial cells (as shown by white arrows and close up marked by white dashed rectangle). **j,** High-resolution large-area imaging of representative gonadal sections of one fetal testis (8 PCW), with intensity proportional to smFISH signal to *SOX9* (purple, Sertoli

780   cells), *POU5F1* (purple, primordial germ cells), *CD68* (red, macrophages), *P2RY12* (yellow, microglia-like macrophages), *PDGFRA* (cyan, mesenchymal). Microglia-like macrophages (white arrows) are observed adjacent to the germ and Sertoli cells compartments. White dashed rectangles highlight gonadal regions magnified. **k,** Dotplots showing variance-scaled, log-transformed expression of ligands and receptors involved in the interactions between

785   microglia-like and *SIGLEC15+* macrophages and gonadal cells in their proximity as observed with high-resolution imaging. **l,** Schematics illustrating the spatial location of the distinct testicular macrophage populations. Mac = Macrophages; cDC = conventional Dendritic cells; pDC = plasmacytoid Dendritic cell; Mono = monocytes; NK = Natural Killer cells.

790

Supplementary Figures

Supplementary Fig. 1 Quality control of the scRNA-seq dataset and round 1 per-lineage reanalysis.
a, Schematic representation of the computational workflow used to analyse scRNA-seq data.
b, UMAP projections of the scRNAseq data showing unbiased Louvain clusters, scrublet scores, sex, PCW, donor and given annotation for each cell. Dotplot showing the variance-scaled, log-transformed expression of marker genes in the main population. **c,** UMAP projections of the per-lineage scRNAseq reanalysis (round 1) showing unbiased Louvain clusters, sex, donor, PCW, cell cycle phase and scrublet scores for each cell in the germ, supporting, mesenchymal, epithelial and endothelial lineages.

Supplementary Fig. 2 Quality control of the scRNA-seq round 2 per-lineage reanalysis.
a, UMAP projections of the per-lineage scRNAseq reanalysis (round 2) showing unbiased Louvain clusters, sex, donor and PCW for each cell in the germ, supporting, mesenchymal, epithelial and endothelaial lineages. **b,** Dotplot showing the log-transformed expression of marker genes in the identified sub-populations for the mesenchymal, epithelial and endothelial lineages. Dotplots for germ and supporting cells are in Fig. 2 and Fig. 3-4, respectively.

Supplementary Fig. 3 Quality control of visium datasets and DSD genes.
a, H&E images matching the six tissue sections used for Visium spatial transcriptomics including two 11, 14 and 17 PCW ovaries and two 12 PCW testis. **b,** cell2location estimated amount of mRNA (colour intensity) contributed by each mesenchymal subpopulation to each Visium spot (colour) shown over the H&E images. **c**, Dotplot showing the variance-scaled, log-transformed expression of main DSD-associated genes in the identified gonadal populations. **d**, Heatmap showing the enrichment z-scores of the accessibility of transcription factor motifs associated with DSD with respect to the identified gonadal populations in scATAC-seq data from female (top) and male (bottom) samples.

Supplementary Fig. 4 Chromatin accessibility landscape of the developing ovaries and testes.
a, Schematic representation of the computational workflow used to analyse scATAC-seq data.
b, UMAP projections of scATAC-seq data from female samples labeled by unbiased Leiden clustering, donor and PCW. **c**, Heatmap reporting label transfer scores from scRNA-seq to scATAC-seq data of matched individuals (females). **d**, UMAP projections of scATAC-seq data from male samples labeled by unbiased Leiden clustering, donor and PCW. **e**, UMAP projections of scATAC-seq data labeled by cell type identified in snRNA-seq data from the combined snRNA-seq/snATAC-seq profiling (males). Only gonads were included in the combined snRNA-seq/snATAC-seq assay. **f**, Heatmap showing enrichment z-scores for the top transcription factor (TF) motif per identified cell type in scATAC-seq data from female (left) and male samples (right). **g**, Hierarchical clustering of enrichment z-scores for peaks contained within each cis-co-accessibility network (CCAN) with respect to the cell types identified in scATAC-seq data from female (left) and male samples (right). **h**, Coverage plots of the genomic regions including the TFs *GATA4, LHX9, ARX, GATA2, NR2F1* with respect to the mesenchymal cell types identified in scATAC-seq data from female samples. **i,** Coverage plots of the genomic regions including the TFs *GATA4, LHX9, ARX, GATA2, NR2F1* with respect to the mesenchymal cell types identified in scATAC-seq data from male samples.

Supplementary Fig. 5 Spatial location, chromatin accessibility and cross species comparison of germ cells.
a, cell2location estimated amount of mRNA (colour intensity) contributed by each germ cell state to each Visium spot (colour) shown over the H&E image of 11, 14 and 17 PCW ovaries.
b, cell2location estimated amount of mRNA (colour intensity) contributed by each germ cell

state to each Visium spot (colour) shown over the H&E image of a 12 PCW testis. **c,** UMAP projections of scATAC-seq germ cells data labeled by unbiased Leiden clustering, sex, donor and PCW. **d,** Heatmap reporting label transfer scores from scRNA-seq to scATAC-seq germ cell data of matched individuals. **e,** UMAP projections of scATAC-seq data labeled by cell state identified in snRNA-seq germ cell data from the combined snRNA-seq/snATAC-seq profiling. **f,** Hierarchical clustering of enrichment z-scores for peaks contained within each cis-co-accessibility network (CCAN) with respect to the cell states identified in scATAC-seq germ cell data. **g,** UMAP projections of monkey ovarian germ cells re-analysed from Zhao *et al.*, 2020 labeled by cell state and developmental stage (E84 and E116). **h,** UMAP projections of mouse ovarian germ cells re-analysed from Niu and Spradling, 2020 labeled by cell state and developmental stage (E11.5 to P5). **i,** UMAP projections of mouse ovarian and testicular germ cells re-analysed from Mayere *et al.*, 2021 labeled by cell state and developmental stage (E10 to E16). **j,** Dot plot showing the log-transformed expression of known genes mediating oogenesis and spermatogenesis in mammalian germ cells from our human dataset, Zhao *et al.*, 2020, Niu and Spradling, 2020, Mayere *et al.*, 2021.

Supplementary Fig. 6 Supporting progenitor population.
a, Barplot showing the number of cells per identified supporting subpopulation classified by sex and PCW. **b,** Dot plot showing the log-transformed expression of DSD-associated genes characteristic of the sLGR5 subpopulation. **c,** High-resolution large-area imaging of representative gonadal sections of a human fetal testis (8PCW, CS20) and three fetal ovaries (7PCW, CS19), with intensity proportional to smFISH signal for *KLK11* (cyan), *LGR5* (yellow), *OSR1* or *SOX9* (red). **d** Dot plot showing the log-transformed expression of *SFRP1*, *SFRP2* and *AXIN2* in each supporting subpopulation.

Supplementary Fig. 7 sPAX8.
a, UMAP projections of scRNA-seq data from epithelials and sPAX8 cells (n = 10,794). **b,** UMAP coloured per donor, PCW and unbiased Louvain clustering. **c,** Barplot showing the number of cells per subpopulation classified by sex and PCW. **d,** Dot plot showing the log-transformed expression of genes characteristic of the epithelial and sPAX8 cell states. **e,** High-resolution imaging of representative gonadal sections of three human fetal testes (8, 11 and 12 PCW), with intensity proportional to smFISH signal for *EPCAM* (red), *PAX8* (yellow), *NR5A1* (cyan) and *KLK11* (green). White dashed rectangles highlight enlarged gonadal regions with PAX8high/EPCAMlow expression; Mesonephric tubules (EPCAMhigh, PAX8+) marked with white asterisk. Scale bars = 100um. **f,** High-resolution imaging of representative gonadal sections of three fetal ovaries (9, 11 and 14 PCW), with intensity proportional to smFISH signal for *EPCAM* (red), *NR5A1* (cyan), *KLK11* (green) and *PAX8* (yellow). White dashed rectangles highlight enlarged gonadal regions with *PAX8high*/*EPCAMlow* expression. Scale bars = 100um unless specified. **g,** UMAP projections of mouse ovarian supporting cells re-analysed from Niu and Spradling, 2020 labeled by cell state and developmental stage (E11.5 to P5) **h,** UMAP projections of mouse ovarian supporting cells re-analysed from Niu and Spradling, 2020 showing the label transfer probabilities for sPAX8e using a SVM model trained on human ovarian supporting cells. **i,** UMAP projections of mouse ovarian supporting cells re-analysed from Niu and Spradling, 2020 showing the log-transformed expression of Pax8, Lypd1, Txb1, Tbx2, Aldh1a3 and Pdgfrl. **j,** High-resolution large-area imaging of one mouse fetal ovary (E13.5), with intensity proportional to smFISH signal for Lgr5 (yellow), Pax8 (red), Hmgcs2 (green, pre-granulosa medullary marker) and Gng13 (purple, pre-granulosa cortical marker). White dashed rectangles highlight enlarged gonadal regions with PAX8high/EPCAMlow expression. Scale bars = 100um. MD = Mullerian Duct.

Supplementary Fig. 8 Spatial location, chromatin accessibility and cross species comparison of supporting cells.
a, cell2location estimated amount of mRNA (colour intensity) contributed by each preGC-II (preGCa/b/c) and developing granulosa population to each Visium spot (colour) shown over

the H&E image of a 11, 14 and 17 PCW ovaries. **b**, UMAP projections of scATAC-seq supporting cells data labeled unbiased Leiden clustering, sex, donor and PCW. **c,** Heatmap reporting label transfer scores from scRNA-seq to scATAC-seq supporting cells data of matched individuals. **d**, UMAP projections of scATAC-seq data labeled by cell state identified in snRNA-seq supporting cells data from the combined snRNA-seq/snATAC-seq profiling. **e**, UMAP projections of monkey ovarian supporting cells re-analysed from Zhao *et al.*, 2020 labeled by cell state and developmental stage (E84 and E116). **f**, UMAP projections of monkey ovarian supporting cells re-analysed from Zhao *et al*, 2020 showing the label transfer probabilities using a SVM model trained on human ovarian supporting cells. **g**, UMAP projections of mouse ovarian supporting cells re-analysed from Niu and Spradling, 2020 labeled by cell state and developmental stage (E11.5 to P5). **h**, UMAP projections of mouse ovarian supporting cells re-analysed from Niu and Spradling, 2020 showing the label transfer probabilities using a SVM model trained on human ovarian supporting cells. **i**, UMAP projections of mouse ovarian supporting cells re-analysed from Niu and Spradling, 2020 showing the log-transformed expression of *LGR5* (mouse marker of cortical pre-granulosa) and *FOXL2* (mouse marker of medullary pre-granulosa). **j**, UMAP projections of mouse ovarian supporting cells re-analysed from Niu and Spradling, 2020 showing the log-transformed expression of functional markers of the two waves of pre-granulosa cells which are conserved across species.

Supplementary Fig. 9 Tissue-resident macrophages in the developing testes. a, Schematics illustrating the CD45+ enrichment strategy for gonadal and extragonadal samples. **b,** UMAP projections of immune cells labeled by unbiased Leiden clustering, sex, donor and PCW. **c,** Heatmap showing label transfer scores from the fetal liver hematopoiesis dataset (Popescu *et al.*, 2019) to our gonadal immune dataset using a Support Vector Machine (SVM) classifier. Low probabilities assigned to neutrophils, which were not defined in the liver dataset, and to macrophages **d,** Dot plot showing variance-scaled, log-transformed expression of marker genes expressed in the identified immune subsets. **e,** Bar plot showcasing the proportion of cells belonging to the main immune lineages during the first and second trimester of pregnancy. **f,** Transcription factor activities computed with Dorothea for the identified macrophage subsets. **g,** Representative Gene Ontology (Biological Process) enriched terms for the identified macrophages subsets. **h,** UMAP projections of the multi-organ integrated fetal myeloid dataset labeled by unbiased Leiden clustering, donor and tissue. **i,** Dot plot showing variance-scaled, log-transformed expression of marker genes expressed in the identified cell populations from the multi-organ integrated fetal myeloid dataset.

Supplementary Fig. 10 Macrophages smFISH panels
a, High-resolution imaging of representative gonadal sections of a fetal ovary (14 PCW) and a fetal testis (12PCW), with intensity proportional to smFISH signal to *EPCAM* (cyan, high = epithelial cells; low = sertoli and germ cells), *CD68* (red, macrophages), *F13A1* (yellow, tissue-repair macrophages). White dashed rectangles highlight gonadal regions magnified. **b,** High-resolution imaging of representative gonadal sections of three fetal testes (11, 12 and 19 PCW), with intensity proportional to smFISH signal to *PDGFRA* (green, mesenchymal), *EPCAM* (cyan, high = epithelial cells; low = sertoli and germ cells), CD68 (red, macrophages), *SIGLEC15* (yellow, *SIGLEC15*+ macrophages). White dashed rectangles highlight gonadal regions magnified. **c,** High-resolution imaging of representative gonadal sections of three fetal testes (8, 14 and 15 PCW), with intensity proportional to smFISH signal to *PDGFRA* (green, mesenchymal), *CDH5* (cyan, endothelial cells), *CD68* (red, macrophages), *SIGLEC15* (yellow, *SIGLEC15*+ macrophages). White dashed rectangles highlight gonadal regions magnified. **d,** High-resolution imaging of representative gonadal sections of two fetal testes (12 PCW), with intensity proportional to smFISH signal to *SOX9* (purple, Sertoli cells), *CD68* (red, macrophages), *P2RY12* (yellow, microglia-like macrophages). White dashed rectangles highlight gonadal regions magnified. **E,** High-resolution imaging of a representative gonadal sections of fetal testis (12 PCW), with intensity proportional to smFISH signal to *PDGFRA*

(cyan, mesenchymal), *POU5F1* (purple, primordial germ cells), *CD68* (red, macrophages), *P2RY12* (yellow, microglia-like macrophages). White dashed rectangles highlight gonadal regions magnified.
Scale bars =100um unless indicated

985

Materials and methods

Patient samples

All tissue samples used for this study were obtained with written informed consent from all participants in accordance with the guidelines in The Declaration of Helsinki 2000.
990   Human embryo and fetal samples were obtained from the MRC and Wellcome-funded Human Developmental Biology Resource (HDBR43, http:// www.hdbr.org), with appropriate maternal written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee (08/H0906/21+5). The HDBR is regulated by the UK Human Tissue Authority (HTA; www.hta.gov.uk) and operates in accordance with the relevant HTA Codes
995   of Practice.

Assignment of fetal developmental stage

Embryos up to eight post-conception weeks were staged using the Carnegie staging method [93]. At fetal stages beyond eight post-conception weeks, age was estimated from measurements of foot length and heel-to-knee length and compared with the standard growth chart [94]. A piece of
1000 skin, or where this was not possible, chorionic villi tissue was collected from every sample for Quantitative Fluorescence-Polymerase Chain Reaction analysis using markers for the sex chromosomes and the following autosomes 13, 15, 16, 18, 21, 22, which are the most commonly seen chromosomal abnormalities. All samples were karyotypically normal.

Tissue processing

1005 All tissues for sequencing and spatial work were collected in saline (HypoThermosol biopreservation media) and stored 4°C until processing. Tissue dissociation was conducted within 24 hours of tissue retrieval with the exception of tissues that were cryopreserved and stored in -80°C **(see Supplementary Table 1)**.

We used the previous protocol optimised for gonadal dissociation [5] and available at protocols.io
1010 [95]. In short, tissues were cut into <1 mm^3 segments before being transferred to a 15ml tube. Tissues were digested with Trypsin/EDTA 0.25% between 5-15 minutes at 37°C with intermittent shaking. Digested tissue was passed through a 100 μm filter, and cells collected by centrifugation (500 x g for 5 minutes at 4°C). Cells were washed with flow buffer (PBS containing 5 % (v/v) FBS and 2 mM EDTA) prior to cell counting.

1015 ## Cell sorting

Dissociated cells were incubated at 4 °C with 2.5 μl of antibodies in 1% FBS in DPBS without calcium and magnesium (Thermo Fisher Scientific, 14190136). DAPI was used for live versus dead discrimination. Cells were sorted using a Becton Dickinson (BD) FACS Aria Fusion with

5 excitation lasers (355 nm, 405 nm, 488 nm, 561 nm and 635 nm red), and 18 fluorescent detectors, plus forward and side scatter. The sorter was controlled using BD FACS DIVA software (version 7).

Single nuclei suspension

Single-nuclei suspension was produced from dissociated cells for doing scATAC-seq, following the manufacturers' instructions, and from frozen tissue sections for doing the multiomic snRNA-seq/scATAC-seq. For the latter, thick (300 μm) sections were cryosectioned, and kept in a tube on dry ice until subsequent processing. Nuclei were released *via* Dounce homogenisation as described in detail in the protocols.io [96].

Tissue cryopreservation

Fresh tissue was cut into <1 mm^3 segments before being resuspended with 1 ml of ice cold Cryostor solution (CS10) (C2874-Sigma). The tissue was frozen at -80 °C decreasing the temperature approximately 1 °C per minute. Detailed protocol available at https://www.protocols.io/view/tissue-freezing-in-cryostor-solution-processing-bgsnjwde

Tissue freezing

Fresh tissue samples of human fetal gonads were embedded in cold OCT medium and flash frozen using a dry ice-isopentane slurry. Protocol available at protocols.io [97].

Ovarian tissue collection from mouse embryos

Developing ovaries were collected from E13.5 mouse embryos carrying the Oct4ΔPE-GFP transgene. Tissues were fixed in 4% (w/v) formaldehyde solution for 2 hours and at 4 °C. Samples were washed with PBS and afterwards sequentially incubated with 10% and 20% (w/v) sucrose at 4°C. After, samples were embedded in an optimal cutting temperature compound (OCT) and subsequently flash frozen using a dry ice-isopentane slurry. All experimental procedures were carried out in agreement with the project licence PE596D1FE issued by the UK Home Office and carried out in a Home Office designated facility.

Haematoxylin and Eosin (H&E) staining and imaging

Slides holding fresh frozen sections were removed from -80°C storage and let to air dry before being immersed in 10% Neutral Buffered Formalin for 5 minutes. After rinsing with RO water, slides were dipped in Mayer's Haematoxylin solution for 1.30 minutes. A series of submersions in RO water - with 4-5 changes of water and until water runs clear- was used to both rinse slides as well as blueing haematoxylin stain. 1% Aqueous Eosin was manually applied on top of the sections with a pipette and rinsed with RO water after 1-3 seconds. Slides were next immersed in a series of ethanol solutions (2 x 70% EtOH and 2 x 100% EtOH) to dehydrate tissue sections, before proceeding to 2 x Xylene baths. Slides were finally coverslipped and let to air dry before being imaged on Hamamatsu Nanozoomer 2.0HT digital slide scanner.

Multiplexed smFISH and high-resolution imaging

Large tissue section staining and fluorescent imaging was conducted largely as described previously [98]. Sections were cut from fresh frozen or fixed frozen samples embedded in OCT

at a thickness of 10 μm using a cryostat, placed onto SuperFrost Plus slides (VWR) and stored at -80°C until stained. For FFPE samples, sections were cut at a thickness of 5 μm using a microtome, placed onto SuperFrost Plus slides (VWR), and left at 37°C overnight to dry and 1060 ensure adhesion. Tissue sections were then processed using a Leica BOND RX to automate staining with the RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (Advanced Cell Diagnostics, Bio-Techne), according to the manufacturers' instructions. Probes may be found in Table S15. Prior to staining, human fresh frozen sections were post-fixed in 4% paraformaldehyde in PBS for 15 minutes at 4°C, then dehydrated through a series of 50%, 1065 70%, 100%, and 100% ethanol, for 5 minutes each. Following manual pre-treatment, automated processing included epitope retrieval by protease digestion with Protease IV for 30 minutes prior to probe hybridisation. Mouse fixed frozen sections were subjected to the same manual pre-treatment described above. Subsequently, the automated processing for these sections included heat-induced epitope retrieval at 95°C for 5 minutes in buffer ER2 and digestion with 1070 Protease III for 15 minutes prior to probe hybridisation. Upon this treatment no endogenous fluorescence from the Oct4ΔPE-GFP transgene is observed. For FFPE sections, automated processing included baking at 60°C for 30 minutes and dewaxing, as well as heat-induced epitope retrieval at 95°C for 15 minutes in buffer ER2 and digestion with Protease III for 15 minutes prior to probe hybridisation. Tyramide signal amplification with Opal 520, Opal 570, 1075 and Opal 650 (Akoya Biosciences) and TSA-biotin (TSA Plus Biotin Kit, Perkin Elmer) and streptavidin-conjugated Atto 425 (Sigma Aldrich) was used to develop RNAscope probe channels.

Stained sections were imaged with a Perkin Elmer Opera Phenix High-Content Screening System, in confocal mode with 1 μm z-step size, using a 20X (NA 0.16, 0.299 μm/pixel); 40X 1080 (NA 1.1, 0.149 μm/pixel); or 63X (NA 1.15, 0.091 μm/pixel) water-immersion objective. Channels: DAPI (excitation 375 nm, emission 435-480 nm), Atto 425 (ex. 425 nm, em. 463-501 nm), Opal 520 (ex. 488 nm, em. 500-550 nm), Opal 570 (ex. 561 nm, em. 570-630 nm), Opal 650 (ex. 640 nm, em. 650-760 nm).

10x Genomics Chromium GEX library preparation and sequencing

1085 For the scRNA-seq experiments, cells were loaded according to the manufacturer's protocol for the Chromium Single Cell 5′ Kit v.1.0, v1.1 and v.2 (10X Genomics) to attain between 2,000 and 10,000 cells per reaction. Library preparation was carried out according to the manufacturer's protocol. Libraries were sequenced, aiming at a minimum coverage of 20,000 raw reads per cell, on the Illumina HiSeq 4000 or Novaseq 6000 systems; using the sequencing 1090 formats; read 1: 26 cycles; i7 index: 8 cycles, i5 index: 0 cycles; read 2: 98 cycles (3' Kit v.2) or read 1: 28 cycles; i7 index: 8 cycles, i5 index: 0 cycles; read 2: 91 cycles (3' Kit v.3).

For the scATAC-seq experiments, cells were loaded according to the manufacturer's protocol for the Chromium Single Cell ATAC v1.0 and v to attain between 2,000 and 10,000 cells per 1095 well. Library preparation was carried out according to the manufacturer's protocol. Libraries were sequenced on Illumina NovaSeq 6000, aiming at a minimum coverage of 10,000 fragments per cell, using the sequencing formats; read 1: 50 cycles; i7 index: 8 cycles, i5 index: 16 cycles; read 2: 50 cycles.

10x Genomics Visium library preparation and sequencing

1100 Ten micron cryosections were cut and placed in duplicate on Visium slides (beta product version). These were processed according to the manufacturer's instructions. Briefly, sections were fixed with cold methanol, stained with haematoxylin and eosin and imaged on a

Hamamatsu NanoZoomer S60 before permeabilisation, reverse transcription and cDNA synthesis using a template-switching protocol. Second-strand cDNA was liberated from the slide and single-indexed libraries prepared using a 10x Genomics PCR-based protocol. Libraries were sequenced (1 per lane on a HiSeq4000), aiming for 300M raw reads per sample with read lengths 28cy R1, 8cy i7 index, 0cy i5 index, 91cy read 2.

Alignment and quantification of scRNA-seq data

For each sequenced library, we performed read alignment to the 10x Genomics' GRCh38 3.1.0 reference genome, quantification and initial quality control (QC) using the Cell Ranger Software (version 3.1; 10X Genomics) using default parameters. Cell Ranger filtered count matrices were used for downstream analysis.

Downstream scRNA-seq analysis

Quality filters, doublet detection, alignment of data across different batches, and clustering

We used Scrublet for cell doublet calling on a per-library basis. We used a two-step diffusion doublet identification followed by Bonferroni-FDR correction and a significance threshold of 0.01, as described in [62]. Predicted doublets were not excluded from the initial analysis, but used afterwards to flag clusters with high doublet score.

We integrated the filtered count matrices from Cell Ranger and analysed them with Scanpy v.1.7.0, with the pipeline following their recommended standard practices. Briefly, we excluded genes expressed by less than three cells and excluded cells with fewer than 300 genes or with more than 20% mitochondrial reads. After converting the expression space to $\log(\mathrm{CPM}/100 + 1)$, the object was transposed to gene space to identify cell cycling genes in a data-driven manner, as described in [62,74]. After performing PCA, neighbour identification and Louvain clustering, the members of the gene cluster including known cycling genes (*CDK1*, *MKI67*, *CCNB2* and *PCNA*) were flagged as the data-derived cell cycling genes. Next, we identified the identifying highly variable genes, which were used for Principal Components Analysis (PCA). We corrected for the donor's effect using Harmony [32] on PCA space with theta equal to 0, keeping the remaining parameters default. Finally, we used the Harmony-corrected PCA matrix for neighbour identification, Louvain clustering and Uniform Manifold Approximation and Projection (UMAP) visualisation.

Annotation of scRNA-seq datasets

Identification, labeling and naming of the major cell types was carried by manual inspection of marker genes and interpretation of these based on the literature. We used the TF-IDF approach from the *SoupX* R package v.1.5.0 [99] to identify genes specifically expressed in a cluster. Clusters with overall high doublet score or low counts number and no distinctive differentially expressed gene were flagged and discarded in further analysis.

Lineage-specific analysis
Germ cells, supporting and mesothelial, mesenchymal, endothelial and epithelial cells were reanalysed independently using Seurat 3.2.2. For each lineage, we performed two analysis rounds. In both rounds, cells from a specific lineage (defined in the main analysis) were subsetted. Genes expressed by less than 3 cells were discarded. Counts were normalized, PCA was carried out on highly variable genes using Seurat defaults, and the donor effect was corrected by the Harmony tool (theta = 0). Corrected PCA dimension was used for

neighbourhood identification, Louvain clustering and UMAP visualisation. We used TF-IDF [99] to identify cluster-specific genes. The first round of per-lineage analysis aimed to detect doublets and low QC cells not detected in the main analysis, and was characterized by high resolution clustering and identification of low quality cells. Briefly, we further annotated: 1) doublets, by identifying clusters with high scrublet score and expressing genes specific of multiple lineages; 2) low QC clusters, by identifying clusters with low counts number and without any distinctive differentially expressed genes; 3) low QC clusters with low counts number and high expression of genes specific of the erythroid lineage; and 4) cycling clusters composed mostly of cells in G2/M and S phase, as estimated with *CellCycleScoring* function in Seurat. The remaining cells (*aka* the clean per-lineage dataset) were re-analysed as described previously and annotated using known marker genes (when available) or the most distinctively expressed gene according to the TF-IDF approach.

Immune cells analysis

Cell Ranger filtered count matrices of CD45+ enriched samples were processed using the Scanpy v.1.7.0 workflow described above for the main scRNA-seq analysis (Doublet detection, alignment of data across different batches, and clustering). These cells were then merged with the cluster of immune cells from the non-enriched samples. We corrected for donor effect using Harmony (theta = 0). We used the corrected PCA space to compute the neighborhood graph, Leiden clustering and UMAP visualisation. The resulting clustered manifold was preliminary annotated by transferring labels from a publicly available dataset of human fetal liver hematopoiesis[62]. Fetal liver raw sequencing data were downloaded from ArrayExpress (E-MTAB-7407), processed with Scanpy v.1.7.0 workflow described above for the main scRNA-seq analysis and filtered based on the expression of CD45 (*PTPRC*) to exclude non-immune cells. We then trained a Support Vector Machine (SVM) model (sklearn.svm.SVC) on the filtered liver dataset and used it to predict cell types on our gonadal immune dataset. Predicted cell type annotations were validated or disproved by looking at the expression of known marker genes.

To further characterize the novel gonadal macrophage populations, we performed functional enrichment analysis of Gene Ontology (Biological Process) terms with the Python interface of the g:Profiler [100] toolkit (gprofiler-official v.1.0.0) using the top 100 marker genes identified with TF-IDF as query and all the genes as background. Projection of fetal microglia cells [63] onto our immune dataset was done using a SVM model as well as scmap [101] following the tutorial (http://bioconductor.org/packages/release/bioc/vignettes/scmap/inst/doc/scmap.html).

Fetal multi-organ myeloid cells integration

To study the unique profile of our identified subsets of gonadal macrophages, we integrated myeloid cells from multiple fetal tissues: liver, skin, kidney, yolk sac, gut, thymus and placenta [62,74–76]. Raw sequencing data were downloaded from ArrayExpress (E-MTAB-7407, E-MTAB-8901, E-MTAB-8581, E-MTAB-0701) and processed with the Scanpy v.1.7.0 workflow described above for the main scRNA-seq analysis. Myeloid cells from all the datasets were selected based on the expression of established myeloid markers (*CD14, CD68, CSF1R*). Data integration was performed using single-cell Variational Inference (n_hidden = 128, n_latent = 55, n_layers = 3) [102] with a combined batch of donor and library.

Cross-species comparison

We compared our single-cell signatures of germ and supporting cells to those of other mammalian species by re-analysing two published datasets of mouse and one of macaque fetal

gonads. The macaque dataset included fetal ovaries at stages E84 and E116 [38]. One mouse dataset included fetal ovaries from E11.5 to P5 [39], while the other contained germ cells from both fetal ovaries and testes from E10 to E16 [40]. The distinct datasets were downloaded from the Gene Expression Omnibus database (GSE136220, GSE136441, GSE149629). We re-analysed the datasets using the Scanpy v.1.7.0 workflow, including batch correction with Harmony, as described above for the main scRNA-seq analysis. Macaque and mouse genes were converted to human genes using ENSEMBL Biomart multi species comparison filter. Main cell populations were annotated based on the expression of marker genes described by the authors. For the comparison of germ cells, we subsetted the datasets to keep germ cells only, and plotted our fetal human signatures onto them. For the comparison of ovarian supporting cells, we trained a SVM model (sklearn.svm.SVC) on human ovarian supporting cells and projected the data onto the macaque and mouse datasets. Additionally, we plotted human genes in the mouse and macaque dataset to validate our models, and found additional functional relationships between the development of gonadal supporting cells in mammals.

Alignment, quantification, and quality control of ATAC data

We processed scATAC-seq libraries (read filtering, alignment, barcode counting, and cell calling) with 10x Genomics Cell Ranger ATAC pipeline (version 1.2.0) using the pre-built 10x's GRCh38 genome (version 3.1.0) as reference. We called the peaks using an in house implementation of the approach described in Cusanovich et al.[103] (available at https://github.com/cellgeni/cellatac, revision 21-099). In short, the genome was broken into 5 kb windows and then each cell barcode was scored for insertions in each window, generating a binary matrix of windows by cells. Matrices from all samples were concatenated into a unified matrix, which was filtered to retain only the top 200K most commonly used windows per sample. Using Signac (https://satijalab.org/signac/ version 0.2.5), the binary matrix was normalised with TF-IDF followed by a dimensionality reduction step using Singular Value Decomposition (SVD). Latent Semantic Indexing (LSI) were clipped at +/-1.5. The first LSI component was ignored as it usually correlates with sequencing depth (technical variation) rather than a biological variation [103]. The 2-30 top remaining components were used to perform graph-based Louvain clustering. Next, peaks were called separately on each cluster using macs2 [104]. Finally, peaks from all clusters were merged into a master peak set (i.e. peaks overlapping in at least one base pair were aggregated) and used to generate a binary peak by cell-matrix, indicating any reads occurring in each peak for each cell.

Downstream scATAC-seq analysis

Quality filters, alignment of data across different batches, and clustering
To obtain a set of high quality peaks for downstream analysis, we filtered out peaks that (i) are included in the ENCODE blacklist, (ii) have a width outside the 210-1500bp range and (iii) are accessible in less than 4% of cells from a *cellatac* cluster. Low quality cells were also removed by setting to 5.5 the minimum threshold for log1p transformed total counts per cell.

We adopted the cisTopic approach [105,106] for the core of our downstream analysis. cisTopic employs Latent Dirichlet Allocation (LDA) to estimate the probability of a region belonging to a regulatory topic (region-topic distribution) and the contribution of a topic within each cell (topic-cell distribution). The topic-cell matrix was used for constructing the neighborhood graph, computing UMAP projections and clustering with the Leiden algorithm. Donor effects

were corrected using Harmony (theta = 0), with the exception of the germ cells dataset which already exhibited good batch mixing in the uncorrected manifold. Cell doublets were identified and removed using scrublet [107].

Gene activity scores

Next, we generated a denoised accessibility matrix (predictive distribution) by multiplying the topic-cell and region-topic distribution and used it to calculate gene activity scores. To integrate them with scRNA-seq data, gene activity scores were rounded and multiplied by a factor of 10^7, as previously described [106].

Cell type annotation

To annotate cell types in scATAC-seq data, we first performed label transfer from scRNA-seq data of matched individuals. We used Canonical Correlation Analysis as dimensionality reduction method, *vst* as selection method, 3000 variable features and 25 dimensions for finding anchors between the two datasets and transferring the annotations [34]. The predicted cell type annotations by label transfer were validated by importing annotations of the combined snRNA-seq/snATAC-seq profiling data.

Cell type-specific cis-regulatory networks

Co-accessible peaks in the genome and cis-co-accessibility networks (CCANs) were estimated using the R package Cicero [36] v. 1.3.4.11 with default parameters. We then filtered the denoised accessibility matrix from cisTopic to keep only the peaks included in CCANs. The resulting matrix was further processed to average cells by cell type and peaks by CCAN. Finally, we z-scored the matrix across CCANs and visualized the separation of CCANs by cell type by hierarchical clustering and plotting the heatmap.

GWAS enrichment with single-cell annotations

We downloaded the list of polycystic ovary syndrome associated SNPs from the GWAS Catalog (Experimental Factor Ontology id = EFO_0000660) and filtered it to keep only the variants exceeding genome-wide significance (p-value = 5 x 10^8). Bed files were created from both the SNPs and the CCANs, and we used bedtools[108] to intersect the two sets of bed files. Z-scores of CCANs including a PCOS-associated SNP were visualized with a heatmap to assess which cell types are most likely to be affected by these SNPs. Moreover, we specifically focused on the CCAN containing the variant rs8043701-A and investigated how this variant could disrupt the co-accessibility pattern of the CCAN (using the plot_connections function in Cicero and setting as viewpoint the genomic coordinates of variant rs8043701-A).

Alignment, quantification, and quality control of Visium data

For each 10X Genomics Visium sequencing data, we used Space Ranger Software Suite (v.1.2.1) to align to the GRCh38 human reference genome (official Cell Ranger reference, version 2020-A) and quantify gene counts. Spots were automatically aligned to the paired H&E images by Space Ranger software. All spots under tissue detected by Space Ranger were included in downstream analysis.

Downstream analysis of 10x Genomics Visium data

Location of cell types in Visium data

To spatially locate the cell states on the Visium transcriptomics slides, we used the cell2location tool v.0.3 [33]. As reference, we used scRNAseq data from individuals of the same sex and gestational stage. We used general cell annotations from the main analysis, with the exception of the major gonadal lineages (germ, supporting and mesenchymal) for which we considered the identified sub-populations. We used default parameters with the exception of *cells_per_spot* which was set to 20. Each Visium section was analysed separately. Results were visualized following the cell2location tutorial.

CellPhoneDB and CellSign

We updated the CellphoneDB database to include: 1) more manually curated protein cell-cell interactions (n = 1,852 interactions) and 2) cell-cell interactions involving non-protein ligands such as steroid hormones and other small molecules (n = 194). For the latter, we used the last *bona fide* enzyme in the synthesis pathway. Additionally, we have added a new module called CellSign which links receptors in CellphoneDB to their known downstream transcription factor, which were manually curated and the relevant pubmed reference number recorded. The code is available at https://github.com/ventolab/Cellphonedbv3, and a Python package will be provided in the revised version.

To retrieve interactions between supporting and other cell populations identified in our gonadal samples, we used an updated version of our CellPhoneDB[76,109] approach described in [27]. In short, we retrieved the interacting pairs of ligands and receptors meeting the following requirements: 1) all the protein members were expressed in at least 10% of the cell type under consideration; and 2) at least one of the protein members in the ligand or the receptor is a differentially expressed gene, with an adjusted p-value below 0.01 and a log2 fold change above 0.2. To account for the distinct spatial location of cells, we further classified the cells according to their location in the fetal ovaries (cortex, medulla) as observed by Visium and smFISH. We filtered cell-cell interactions to exclude cell pairs that do not share the same location.

Transcription Factor (TF) analysis

To prioritize the TFs relevant for a cell state in a lineage, we integrate three measurements: (i) expression levels of the TF and (ii) the activity status of the TF measured from (iia) the expression levels of their targets (described below in "Transcription factor activities derived from scRNAseq") and/or (iib) the chromatin accessibility of their binding motifs (described below in "Transcription factor motif activity analysis from scATACseq"). Plots include TFs those meeting the following criteria: 1) TF is differentially expressed (i), with log2 fold change greater than 0.5 and adjusted p-value < 0.05 and 2) TF is differentially active (ii), with log2 fold change greater than 1 and adjusted p-value < 0.05 in at least one of the TF activity measurements (iia/iib).

Transcription factor differential expression (from scRNAseq)

We compute differential expression using the Logistic Regression approach implemented in the *FindAllMarkers* function in Seurat, in a one-vs-all fashion.

1330

<u>Transcription factor activities derived from scRNAseq</u>
We estimated protein-level activity for human Transcription factors (TF) as *a proxy* of the combined expression levels of their targets. Target genes were retrieved from *Dorothea* [110], an orthogonal collection of TF targets compiled from a range of different sources. Next, we
1335   estimated TF activities for each cell using *Viper* [111], a GSEA-like approach, as implemented in the *Dorothea* R package and tutorial [112]. Finally, to identify TFs whose activity is upregulated in a specific cell type, we applied Logistic Regression from *Seurat* onto the z-transformed *cell x TF* activity matrix in a one-vs-all fashion. TFs with adjusted p-value below 0.05 and average log2 fold change above 1 in a cell type were considered as differentially active .

1340

<u>Transcription factor motif activity analysis from scATACseq</u>
Transcription factor motif activities were computed using chromVar [35] v. 1.12.2 with positional weight matrices from JASPAR2018 [113], HOCOMOCOv10 [114], SwissRegulon [115], HOMER [116]. chromVar returns a matrix with binding activity estimates of each TF in each cell, which we
1345   used to test for differential TF binding activity between cell types in a one-vs-all fashion with Logistic Regression (*FindAllMarkers* function in Seurat).

Image analysis

<u>Image stitching</u>

1350   Confocal image stacks were stitched as two-dimensional maximum intensity projections using the BIOP Perkin Elmer Acapella Stitcher (EPFL, Lausanne; https://www.perkinelmer.com/PDFs/downloads/TCH-Workflows-In-Depth-High-Content-Analysis-Operetta.pdf).

1355

References

1.   Bashamboo, A. & McElreavey, K. Human sex-determination and disorders of sex-development (DSD). *Semin. Cell Dev. Biol.* **45**, 77–83 (2015).
2.   Hughes, I. A. *et al.* Consensus statement on management of intersex disorders. *Arch. Dis.*
1360   *Child.* **91**, 554–563 (2006).
3.   Bozdag, G., Mumusoglu, S., Zengin, D., Karabulut, E. & Yildiz, B. O. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum. Reprod.* **31**, 2841–2855 (2016).
4.   Legro, R. S. *et al.* Diagnosis and Treatment of Polycystic Ovary Syndrome: An Endocrine
1365   Society Clinical Practice Guideline. *J. Clin. Endocrinol. Metab.* **98**, 4565–4592 (2013).
5.   Tang, W. W. C. *et al.* A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* **161**, 1453–1467 (2015).
6.   Sybirna, A., Wong, F. C. K. & Surani, M. A. Genetic basis for primordial germ cells specification in mouse and human: Conserved and divergent roles of PRDM and SOX
1370   transcription factors. *Curr. Top. Dev. Biol.* **135**, 35–89 (2019).
7.   Kobayashi, T. *et al.* Blastocyst complementation using Prdm14-deficient rats enables

efficient germline transmission and generation of functional mouse spermatids in rats. *Nat. Commun.* **12**, 1328 (2021).

8. Hackett, J. A. *et al.* Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat. Commun.* **9**, 4292 (2018).

9. Hamazaki, N. *et al.* Reconstitution of the oocyte transcriptional network with transcription factors. *Nature* **589**, 264–269 (2021).

10. Hiort, O. *et al.* Addressing gaps in care of people with conditions affecting sex development and maturation. *Nat. Rev. Endocrinol.* **15**, 615–622 (2019).

11. Witschi, E. Migration of the germ cells of human embryos from the yolk sac to the primitive gonadal folds. *Contrib. Embryol.* **32**, 67–80 (1948).

12. Albrecht, K. H. & Eicher, E. M. Evidence that Sry is expressed in pre-Sertoli cells and Sertoli and granulosa cells have a common precursor. *Dev. Biol.* **240**, 92–107 (2001).

13. Liu, C., Rodriguez, K. & Yao, H. H.-C. Mapping lineage progression of somatic progenitor cells in the mouse fetal testis. *Development* **143**, 3700–3710 (2016).

14. Chassot, A.-A. *et al.* Activation of -catenin signaling by Rspo1 controls differentiation of the mammalian ovary. *Human Molecular Genetics* vol. 17 1264–1277 (2008).

15. Vainio, S., Heikkilä, M., Kispert, A., Chin, N. & McMahon, A. P. Female development in mammals is regulated by Wnt-4 signalling. *Nature* **397**, 405–409 (1999).

16. Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 858–873.e4 (2017).

17. Guo, J. *et al.* Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment. *Cell Stem Cell* **28**, 764–778.e4 (2021).

18. Chitiashvili, T. *et al.* Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nat. Cell Biol.* **22**, 1436–1446 (2020).

19. Vértesy, Á. *et al.* Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells. *Nat. Commun.* **9**, 1873 (2018).

20. Wagner, M. *et al.* Single-cell analysis of human ovarian cortex identifies distinct cell populations but no oogonial stem cells. *Nat. Commun.* **11**, 1147 (2020).

21. Fan, X. *et al.* Single-cell reconstruction of follicular remodeling in the human adult ovary. *Nat. Commun.* **10**, 3164 (2019).

22. Mossadegh-Keller, N. & Sieweke, M. H. Testicular macrophages: Guardians of fertility. *Cell. Immunol.* **330**, 120–125 (2018).

23. Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826.e23 (2021).

24. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647–1660.e19 (2019).

25. Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-019-0392-8.

26. Maniatis, S. *et al.* Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89–93 (2019).

27. Garcia-Alonso, L. *et al.* Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Cold Spring Harbor Laboratory* 2021.01.02.425073 (2021) doi:10.1101/2021.01.02.425073.

28. He, Z. *et al.* Lineage recording reveals dynamics of cerebral organoid regionalization. *bioRxiv* 2020.06.19.162032 (2020) doi:10.1101/2020.06.19.162032.

29. Zhang, K. *et al.* A cell atlas of chromatin accessibility across 25 adult human tissues.

bioRxiv 2021.02.17.431699 (2021) doi:10.1101/2021.02.17.431699.

30. Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**, (2020).

31. King, H. W. *et al.* Integrated single-cell transcriptomics and epigenomics reveals strong germinal center-associated etiology of autoimmune risk loci. *bioRxiv* 2021.03.16.435578 (2021) doi:10.1101/2021.03.16.435578.

32. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

33. Kleshchevnikov, V. *et al.* Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. *Cold Spring Harbor Laboratory* 2020.11.15.378125 (2020) doi:10.1101/2020.11.15.378125.

34. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).

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

36. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858–871.e8 (2018).

37. Anderson, E. L. *et al.* Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14976–14980 (2008).

38. Zhao, Z.-H. *et al.* Single-cell RNA sequencing reveals regulation of fetal ovary development in the monkey (Macaca fascicularis). *Cell Discovery* **6**, 1–16 (2020).

39. Niu, W. & Spradling, A. C. Two distinct pathways of pregranulosa cell differentiation support follicle formation in the mouse ovary. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 20015–20026 (2020).

40. Mayère, C. *et al.* Single-cell transcriptomics reveal temporal dynamics of critical regulators of germ cell fate during mouse sex determination. *FASEB J.* **35**, e21452 (2021).

41. Sarraj, M. A. & Drummond, A. E. Mammalian foetal ovarian development: consequences for health and disease. *Reproduction* **143**, 151–163 (2012).

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

43. Tucker, E. J. *et al.* TP63-truncating variants cause isolated premature ovarian insufficiency. *Hum. Mutat.* **40**, 886–892 (2019).

44. Mork, L. *et al.* Temporal differences in granulosa cell specification in the ovary reflect distinct follicle fates in mice. *Biol. Reprod.* **86**, 37 (2012).

45. Rastetter, R. H. *et al.* Marker genes identify three somatic cell types in the fetal mouse ovary. *Dev. Biol.* **394**, 242–252 (2014).

46. Lönn, P., Morén, A., Raja, E., Dahl, M. & Moustakas, A. Regulating the stability of TGFbeta receptors and Smads. *Cell Res.* **19**, 21–35 (2009).

47. Strippoli, R. *et al.* Caveolin-1 deficiency induces a MEK-ERK1/2-Snail-1-dependent epithelial-mesenchymal transition and fibrosis during peritoneal dialysis. *EMBO Mol. Med.* **7**, 357 (2015).

48. Lustig, B. *et al.* Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol. Cell. Biol.* **22**, 1184–1193 (2002).

49. Ozcan, A. *et al.* PAX 8 expression in non-neoplastic tissues, primary tumors, and metastatic tumors: a comprehensive immunohistochemical study. *Mod. Pathol.* **24**, 751–764 (2011).

50. Yokoyama, K. *et al.* Increased invasion and matrix metalloproteinase-2 expression by Snail-induced mesenchymal transition in squamous cell carcinomas. *Int. J. Oncol.* **22**, 891–898 (2003).

51. Khan, I. A. *et al.* ErbB2-dependent downregulation of a pro-apoptotic protein Perp is

required for oncogenic transformation of breast epithelial cells. *Oncogene* **35**, 5759–5769 (2016).

52. Douma, S. *et al.* Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. *Nature* **430**, 1034–1039 (2004).

53. Pansky, B. Review of MEDICAL EMBRYOLOGY. (1982).

54. Ottolenghi, C. *et al.* Foxl2 is required for commitment to ovary differentiation. *Hum. Mol. Genet.* **14**, 2053–2062 (2005).

55. Bowen, N. J. *et al.* Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC Med. Genomics* **2**, 71 (2009).

56. Findlay, J. K., Hutt, K. J., Hickey, M. & Anderson, R. A. How Is the Number of Primordial Follicles in the Ovarian Reserve Established?1. *Biol. Reprod.* **93**, (2015).

57. Wolf, W., Wattick, R., Kinkade, O. & Olfert, M. Geographical Prevalence of Polycystic Ovary Syndrome as Determined by Region and Race/Ethnicity. *International Journal of Environmental Research and Public Health* vol. 15 2589 (2018).

58. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

59. Liu, F. *et al.* Involvement of parathyroid hormone-related protein in vascular calcification of chronic haemodialysis patients. *Nephrology* **17**, 552–560 (2012).

60. Minina, E. *et al.* BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation. *Development* **128**, 4523–4534 (2001).

61. Shechter, R., London, A. & Schwartz, M. Orchestrated leukocyte recruitment to immune-privileged sites: absolute barriers versus educational gates. *Nat. Rev. Immunol.* **13**, 206–218 (2013).

62. Popescu, D.-M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).

63. Bian, Z. *et al.* Deciphering human macrophage development at single-cell resolution. *Nature* **582**, 571–576 (2020).

64. Blériot, C., Chakarov, S. & Ginhoux, F. Determinants of Resident Tissue Macrophage Identity and Function. *Immunity* **52**, 957–970 (2020).

65. Kracht, L. *et al.* Human fetal microglia acquire homeostatic immune-sensing properties early in development. *Science* **369**, 530–537 (2020).

66. Gosselin, D. *et al.* An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, (2017).

67. Hayman, A. R. *et al.* Mice lacking tartrate-resistant acid phosphatase (Acp 5) have disrupted endochondral ossification and mild osteopetrosis. *Development* vol. 122 3151–3162 (1996).

68. Vu, T. H. *et al.* MMP-9/gelatinase B is a key regulator of growth plate angiogenesis and apoptosis of hypertrophic chondrocytes. *Cell* **93**, 411–422 (1998).

69. Gelb, B. D., Shi, G. P., Chapman, H. A. & Desnick, R. J. Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. *Science* **273**, 1236–1238 (1996).

70. Frattini, A. *et al.* Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis. *Nat. Genet.* **25**, 343–346 (2000).

71. Lokka, E. *et al.* Generation, localization and functions of macrophages during the development of testis. *Nat. Commun.* **11**, 1–16 (2020).

72. DeFalco, T., Bhattacharya, I., Williams, A. V., Sams, D. M. & Capel, B. Yolk-sac–derived macrophages regulate fetal testis vascularization and morphogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2384–E2393 (2014).

73. Mossadegh-Keller, N. *et al.* Developmental origin and maintenance of distinct testicular

macrophage populations. *J. Exp. Med.* **214**, 2829–2841 (2017).

74. Park, J.-E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, (2020).

75. Elmentaite, R. *et al.* Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease. *Dev. Cell* **55**, 771–783.e5 (2020).

76. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).

77. Combes, A. N. *et al.* Endothelial cell migration directs testis cord formation. *Dev. Biol.* **326**, 112–120 (2009).

78. Gkountela, S. *et al.* DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell* **161**, 1425–1436 (2015).

79. Guo, F. *et al.* The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* **161**, 1437–1452 (2015).

80. Shami, A. N. *et al.* Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Dev. Cell* **54**, 529–547.e12 (2020).

81. Karl, J. & Capel, B. Sertoli cells of the mouse testis originate from the coelomic epithelium. *Dev. Biol.* **203**, 323–333 (1998).

82. Bott, R. C., McFee, R. M., Clopton, D. T., Toombs, C. & Cupp, A. S. Vascular Endothelial Growth Factor and Kinase Domain Region Receptor Are Involved in Both Seminiferous Cord Formation and Vascular Development During Testis Morphogenesis in the Rat1. *Biol. Reprod.* **75**, 56–67 (2006).

83. Cool, J., DeFalco, T. J. & Capel, B. Vascular-mesenchymal cross-talk through Vegf and Pdgf drives organ patterning. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 167–172 (2011).

84. Male Reproductive System. in *Comparative Anatomy and Histology* 335–363 (Academic Press, 2018).

85. Rete Testis: Structure, Cell Biology and Site for Stem Cell Transplantation. 263–269 (2018).

86. Cackowski, F. C. *et al.* Osteoclasts are important for bone angiogenesis. *Blood* **115**, 140–149 (2010).

87. Chen, Q., Deng, T. & Han, D. Testicular immunoregulation and spermatogenesis. *Semin. Cell Dev. Biol.* **59**, 157–165 (2016).

88. Meinhardt, A. & Hedger, M. P. Immunological, paracrine and endocrine aspects of testicular immune privilege. *Mol. Cell. Endocrinol.* **335**, 60–68 (2011).

89. Paus, R., Ito, N., Takigawa, M. & Ito, T. The hair follicle and immune privilege. *J. Investig. Dermatol. Symp. Proc.* **8**, 188–194 (2003).

90. Moore, K. L. & Persaud, T. V. N. *Review of Medical Embryology.* (Elsevier Health Sciences, 2003).

91. Cox, K. *et al.* Novel associations in disorders of sex development: findings from the I-DSD Registry. *J. Clin. Endocrinol. Metab.* **99**, E348–55 (2014).

92. Oosterhuis, J. W. & Looijenga, L. H. J. Human germ cell tumours from a developmental perspective. *Nat. Rev. Cancer* **19**, 522–537 (2019).

93. Harper, J. Human Embryology and Teratology. Second Edition. By Ronan O'Rahilly and Fabiola Muller. *Ann. Hum. Genet.* **60**, 533–533 (1996).

94. Hern, W. M. Correlation of fetal age and measurements between 10 and 26 weeks of gestation. *Obstet. Gynecol.* **63**, 26–32 (1984).

95. Hoo, R., Vento-Tormo, R. & Sancho, C. Human embryonic gonad dissociation with Trypsin-EDTA. *protocols.io* doi:10.17504/protocols.io.66fhhbn.

96. Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).

97. Roberts K, L. T. Embedding and freezing fresh human tissue in OCT using isopentane V.3. *protocols.io* (2019) doi:10.17504/protocols.io.95mh846.

98. Bayraktar, O. A. *et al.* Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. *Nat. Neurosci.* (2020) doi:10.1038/s41593-020-0602-1.

99. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, (2020).

100. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* vol. 47 W191–W198 (2019).

101. Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15**, 359–362 (2018).

102. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

103. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309–1324.e18 (2018).

104. Gaspar, J. M. Improved peak-calling with MACS2. doi:10.1101/496521.

105. González-Blas, C. B. *et al.* cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).

106. Bravo González-Blas, C. *et al.* Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. *Mol. Syst. Biol.* **16**, e9438 (2020).

107. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).

108. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

109. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).

110. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).

111. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).

112. Holland, C. H. *et al.* Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36 (2020).

113. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).

114. Kulakovskiy, I. V. *et al.* HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.* **44**, D116–25 (2016).

115. Pachkov, M., Erb, I., Molina, N. & van Nimwegen, E. SwissRegulon: a database of genome-wide annotations of regulatory sites. *Nucleic Acids Research* vol. 35 D127–D131 (2007).

116. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
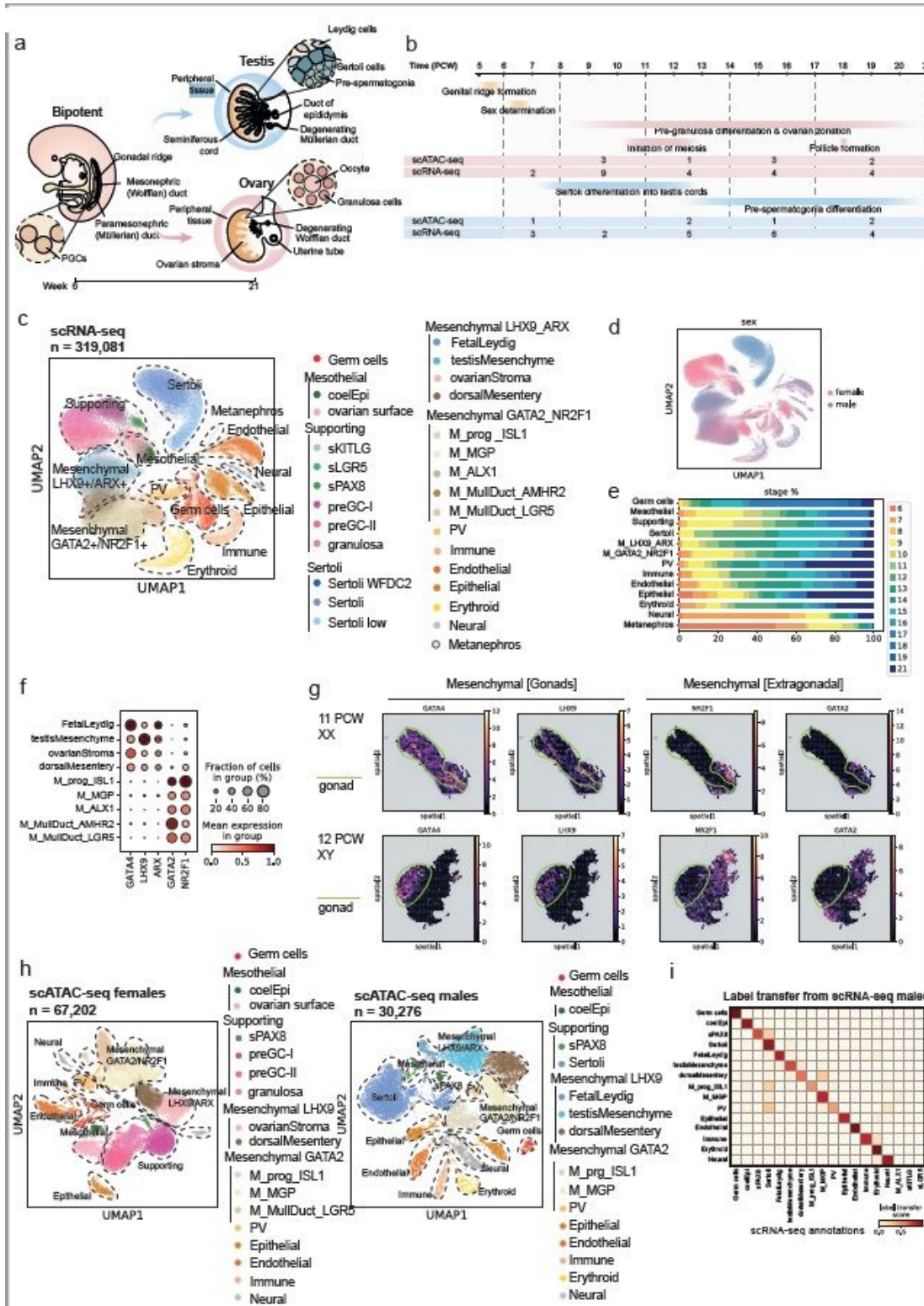
Figures

Figure 1

Single-cell profiling of gonadal and extra-gonadal tissue. a, Schematic illustration of gonadal development showing the main structures of the XX and XY gonads. b, Diagram summarising stages and samples collected in our study along with major events occuring during gonadogenesis. c, UMAP

(uniform manifold approximation and projection) projections of scRNA-seq data from 43 individuals (n = 319,081). Clusters for supporting, Sertoli and mesenchymal cells were defined in independent per-lineage re-analyses for these populations (see Supplementary Figure 1-2). d, UMAP coloured per sex. Pink = XX cells, blue = XY cells. e, Barplot showing the percentage of cells per PCW classified by cell types identified in the main UMAP. f, Dot plot showing variance-scaled, log-transformed expression of transcription factors (TFs) differentially expressed in gonadal and extragonadal mesenchymal cells. g, Spatial plot showing log-transformed expression of TFs in each Visium spot shown over the H&E image of a female 11 PCW (above) and a male 12 PCW (below) sample including gonadal and extragonadal tissue. h, UMAP projections of scATAC-seq data from females (left) and males (right). i, Heatmap showing label transfer scores from scRNA-seq to scATAC-seq data of male samples. PV = Perivascular cell; coelEpi = Coelomic epithelium; sLGR5 = supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pre granulosa cells; M_ = mesenchymal from the mesonephros.
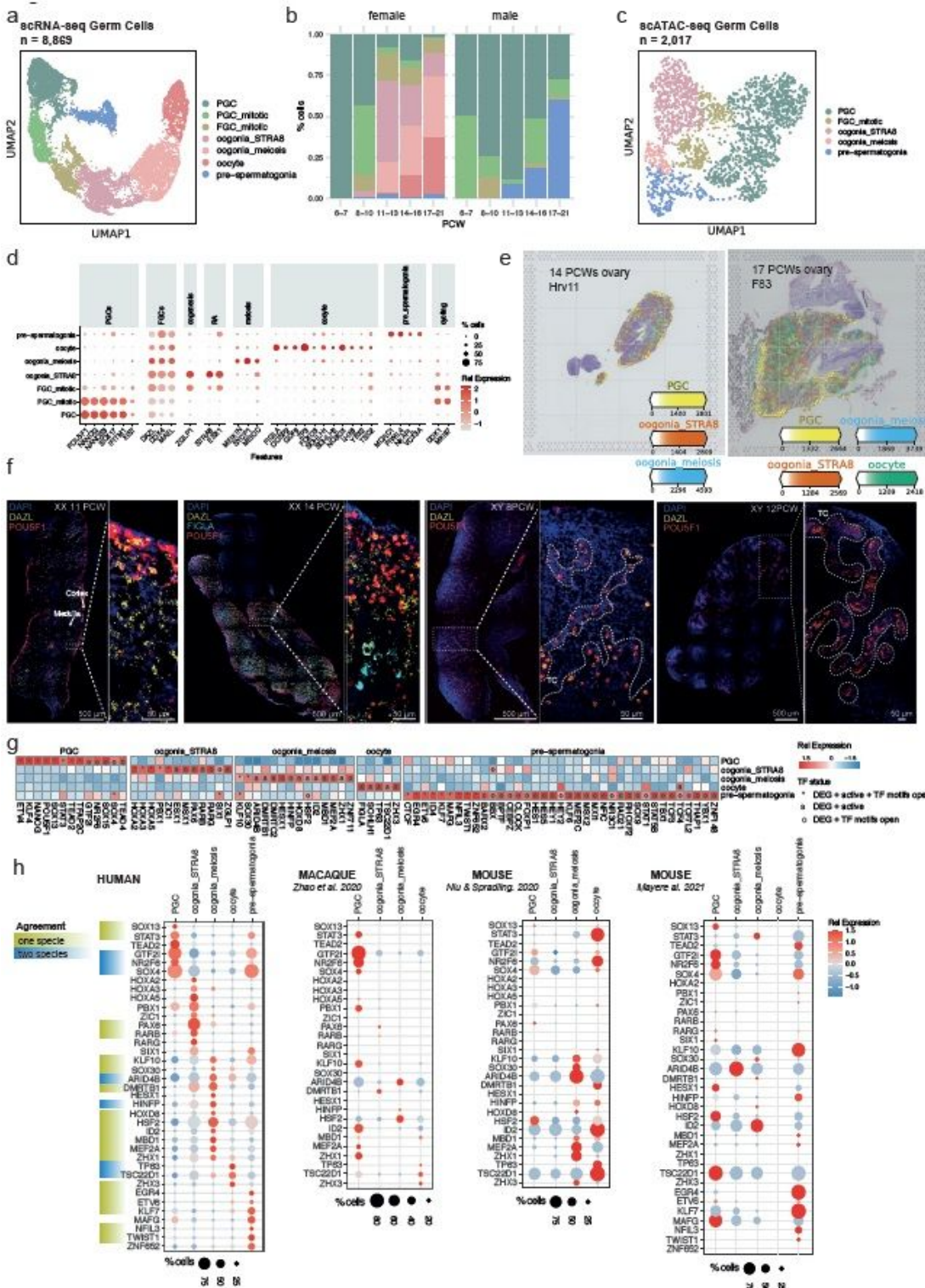
Figure 2

Transcriptional signatures in human germ cells. a, UMAP projections of scRNA-seq data from female and male germ cells (n = 8,869). b, Barplot showing the proportions of germ cells and their annotation classified by sex and developmental stage. c, UMAP projections of scATAC-seq data from female and male germ cells (n = 2,017). d, Dot plot showing the logtransformed expression of genes characteristic of the germ cell states. e, cell2location estimated amount of mRNA (colour intensity) contributed by each

female germ cell population to each Visium spot (colour) shown over the H&E image of 14 and 17 PCW ovaries. f, High-resolution, large-area imaging of representative gonadal sections of two human fetal ovaries (11 and 14 PCW) and two testes (8 and 12 PCW), with intensity proportional to smFISH signal for FIGLA (cyan), DAZL (yellow), POU5F1 (red). Dashed lines on magnified fields describe developing gonadal structures. g, Heatmaps showing TFs differentially expressed in germ cell states. Colour proportional to log-transformed expression. "o" = TFs whose binding motifs are differentially accessible (i.e. TF can bind their potential targets); "a" = TFs whose targets are also differentially expressed (i.e. differentially activated TFs); and asterisk (*) = TFs that meet both "o" and "a" conditions. h, Dot plot showing the log-transformed expression of novel TFs relevant for human germ cell states in macaque and mouse scRNAseq datasets. PGC = Primordial germ cells; FGC = Fetal germ cells; DEG = differentially expressed gene; TC = Testicular cords.
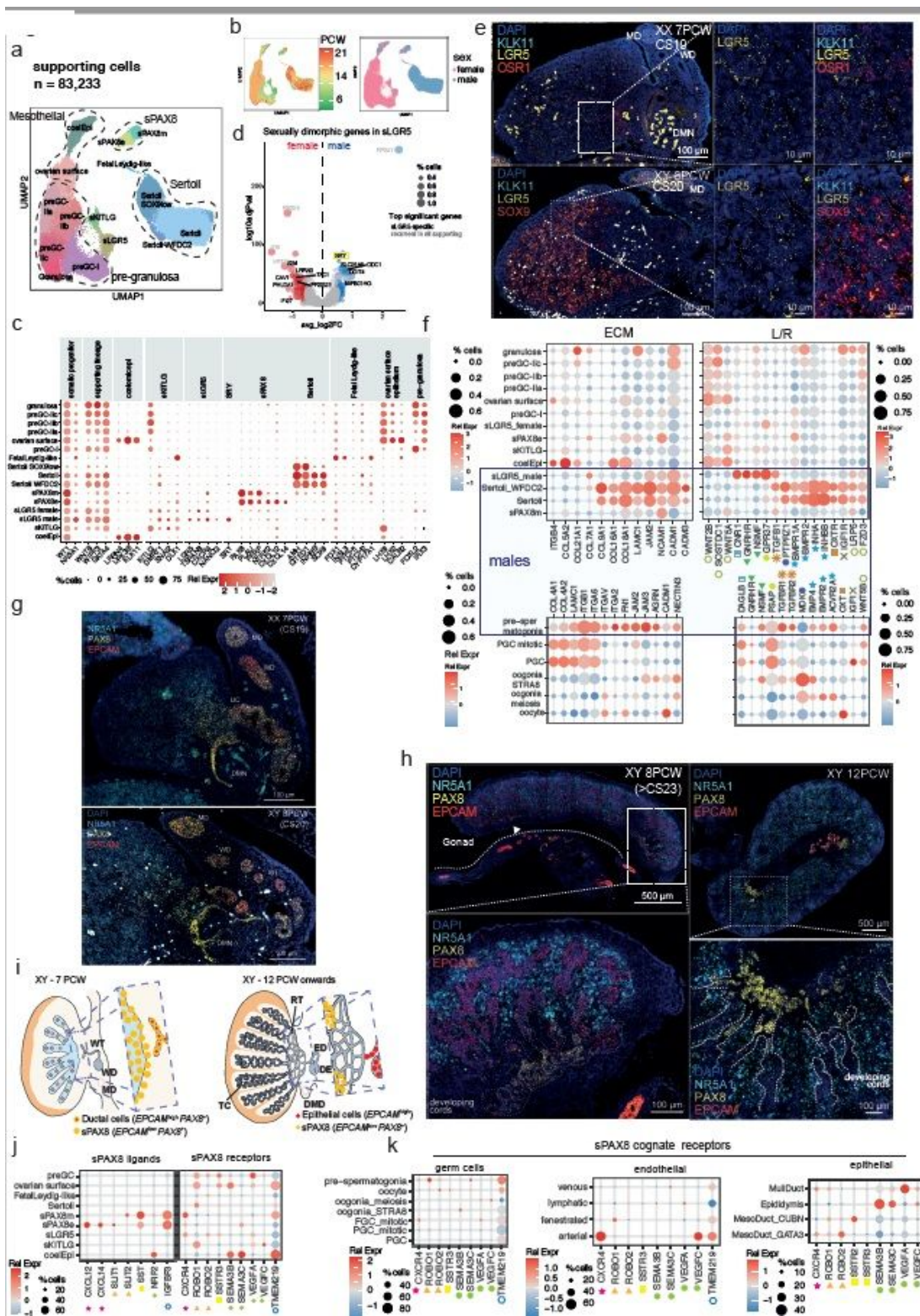
Figure 3

Novel subsets in human supporting lineage. a, UMAP projections of scRNA-seq data from female and male supporting and coelomic epithelial cells (n = 83,233). b, UMAP coloured per PCW and sex. c, Dot plot showing the log-transformed expression of genes characteristic of the supporting cell states. d, Volcano plot showing genes differentially expressed between males and females within the sLGR5 population, highlighting SRY gene. e, High-resolution large-area imaging of representative gonadal sections of a

human fetal ovary (7PCW, CS19) and testis (8PCW, CS20), with intensity proportional to smFISH signal for KLK11 (cyan), LGR5 (yellow), OSR1 or SOX9 (red); red blood cells appear as bright autofluorescent cells. White dashed rectangles highlight enlarged gonadal regions. Dashed lines on magnified fields in the XY sample describe developing testicular cords. f, Dot plots showing scaled logtransformed expression of genes coding for interacting ligand-receptor proteins in supporting and germ states. List of interactions available in Supplementary Table 8. g, High-resolution large-area imaging of representative gonadal sections of a human fetal ovary (7PCWs, CS19) and testis (8 PCWs, CS20), with intensity proportional to smFISH signal for EPCAM (red), NR5A1 (cyan) and PAX8 (yellow); red blood cells appear as bright autofluorescent cells. h, High-resolution large-area imaging of representative gonadal sections of two human fetal testes -8 PCWs (> CS23) and 12PCW-, with intensity proportional to smFISH signal for EPCAM (red), NR5A1 (cyan) and PAX8 (yellow). White dashed rectangles highlight enlarged gonadal regions with PAX8high/EPCAMlow expression. White dashed lines on magnified fields describe developing testicular cords. i, Schematic representation of sPAX8 cells in the human fetal testis at two developmental stages. j, Dot plot showing the log-transformed expression of genes coding for sPAX8 ligands and receptor proteins in the supporting cells. k, Dot plot showing the log-transformed expression of genes coding for sPAX8 receptor proteins in the germ and endothelial cells. coelEpi = Coelomic epithelium; ovarianSurf = Ovarian surface; sLGR5 = supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pregranulosa cells; ECM = extracellular matrix; MD = Mullerian Duct; WD = Wolffian Duct; WT = Wolffian Tubules; DMN = Degenerating Mesonephric Nephron; UC = Urogenital Connection; RT = Rete Testis; ED = Efferent Ductule; DE = Ductus Epididimydis; DMD = Degenerating Mullerian Duct; TC = Testicular Cord.
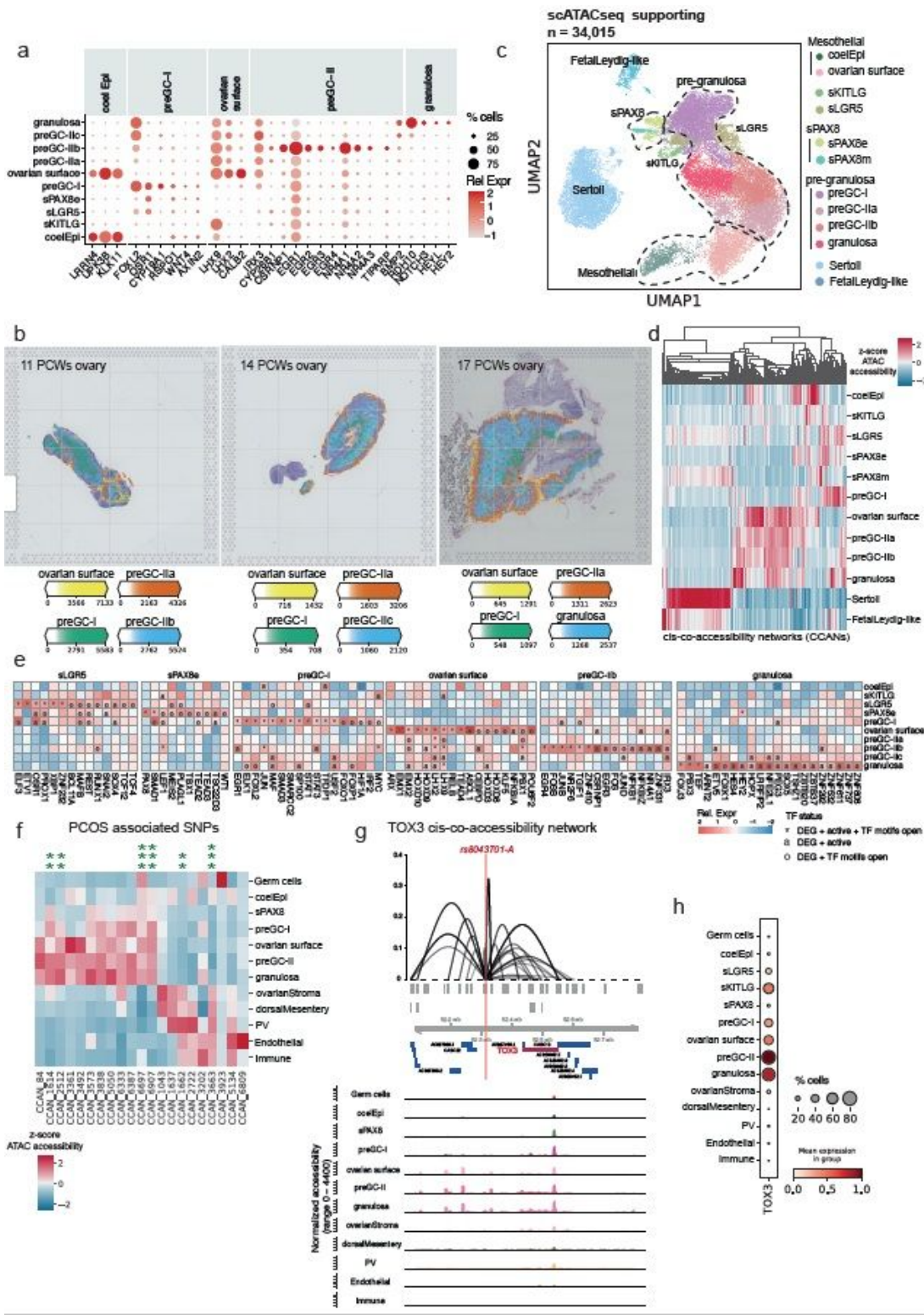
Figure 4

Transcriptional signatures and location of human pre-granulosa cells. a, Dot plot showing the log-transformed expression of genes characteristic of the supporting cell states. b, Estimated amount of mRNA (colour intensity) contributed by ovarian surface epithelium, preGC-I, most frequent preGC-II and/or developing granulosa states to each Visium spot (colour) shown over the H&E image of a 11, 14 and 17 PCW ovaries. c, UMAPs projections of scATAC-seq data from female and male supporting cells (n =

34,015). d, Hierarchical clustering of enrichment z-scores for peaks contained within each cis-co-accessibility network (CCAN) with respect to the cell states identified in scATAC-seq data of human supporting cells. e, Heatmaps showing TFs differentially expressed in proportional to scaled log-transformed expression. "o" = TFs whose binding motifs are differentially accessible (i.e. TF can bind their potential targets), "a" = TFs whose targets are also differentially expressed (i.e. differentially activated TFs), and asterisk (*) = TFs that meet both "o" and "a" conditions. f, Enrichment z-scores for peaks contained within CCANs including PCOS-associated single nucleotide polymorphisms (SNPs) with respect to the cell states identified in scATAC-seq data of human supporting cells. CCANs that contain multiple SNPs are marked with a number of asterisks (*) that corresponds to the number of SNPs. g, Co-accessibility and coverage plot of the genomic region of the CCAN including the SNP rs8043701-A and gene TOX3. Viewpoint in the co-accessibility plot is set to the genomic coordinates of the SNP rs8043701-A. h, Dot plot showing the variance-scaled, log-transformed expression of TOX3 characteristic of pre-granulosa cell states. coelEpi = Coelomic epithelium sLGR5 = supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pre-granulosa cells; PV = Perivascular cells; DEG = differentially expressed gene.
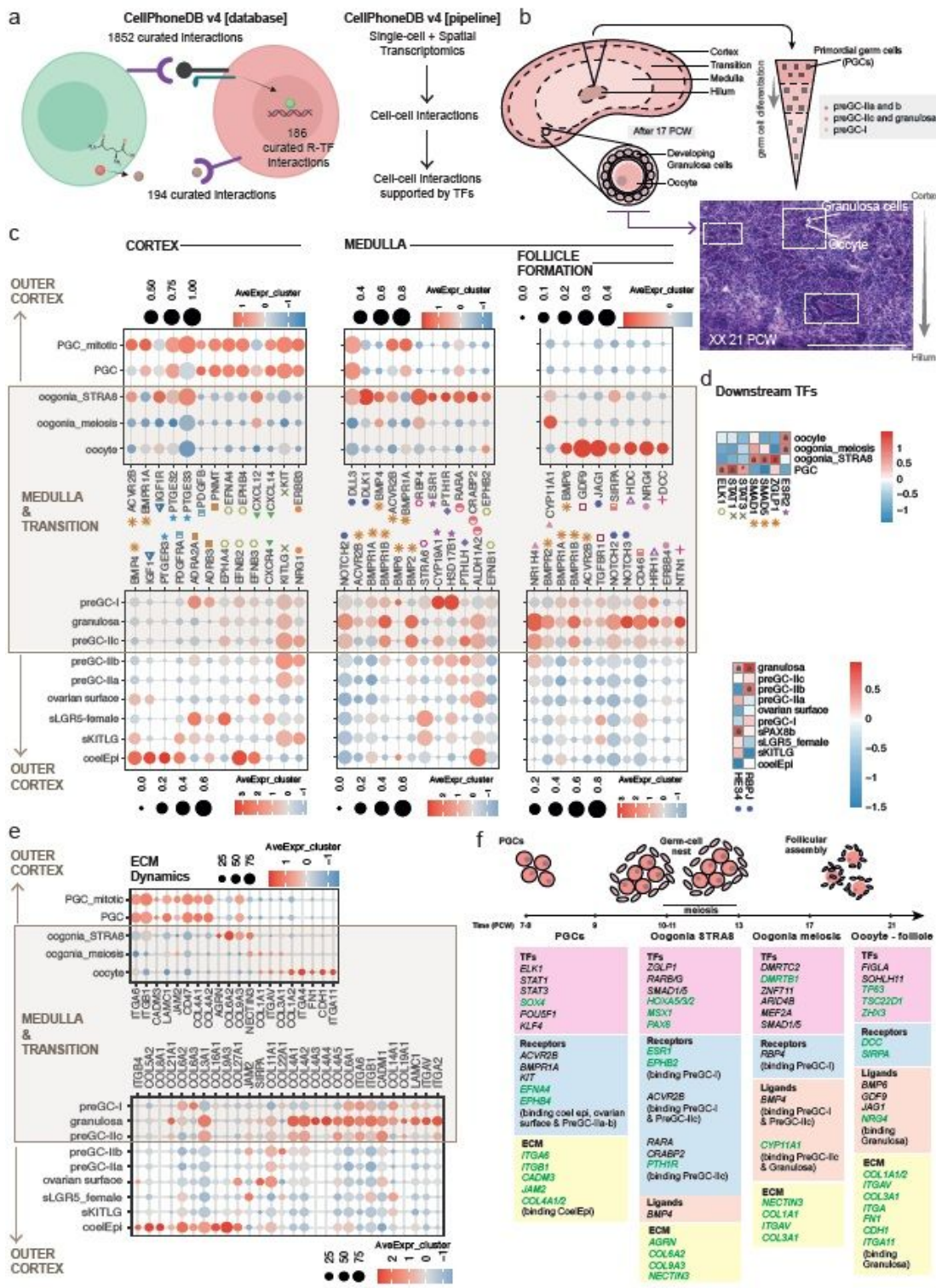
Figure 5

Cell-cell communication networks. a, (left) Diagram showing the information aggregated within the updated version of CellPhoneDB, which includes: (i) 534 novel (1,852 total) ligand-receptor interactions; (ii) 194 novel interaction mediated by small molecules; (iii) 186 novel curated links between ligand-receptor and transcription factors. (right) Diagram showing the new statistical framework to infer active cell-cell interaction partners. It includes an additional step to indicate active ligand-receptor partners in

our data based on the activation of downstream signals on the receiver cell. Downstream signals are calculated based on TF expression and TF activity from scRNA-seq and scATAC-seq data. b, (Top) Schematic representation of pre-granulosa and germ cell states in the human fetal ovary. (Bottom) H&E imaging of a representative XX human fetal ovary section (21PCW). White dashed rectangles highlight follicles at different stages of maturation. Note the different morphology of Developing Granulosa cells surrounding the oocytes c, Dot plots showing scaled logtransformed expression of genes coding for interacting ligand-receptor proteins in supporting and germ cells states, in the cortex, medulla and formation of follicles. d, Heatmap showing the activity (i.e. expression level of bona fide targets) of TFs downstream receptors in germ and supporting cells. Colour proportional to scaled normalized enrichment score. e, Dot plots showing scaled log-transformed expression of genes coding for extracellular matrix (ECM) and adhesion molecules in germ and supporting cells states. f, Schematic illustration of main TFs, receptors, ligands and extracellular molecules regulating germ cell differentiation. PGC = Primordial germ cells; coelEpi = Coelomic epithelium; sLGR5 = supporting LGR5; sPAX8 = supporting PAX8; sKITLG = supporting KITLG; preGC = pre-granulosa cells; ECM =extracellular matrix. Scale Bar = 250μm.
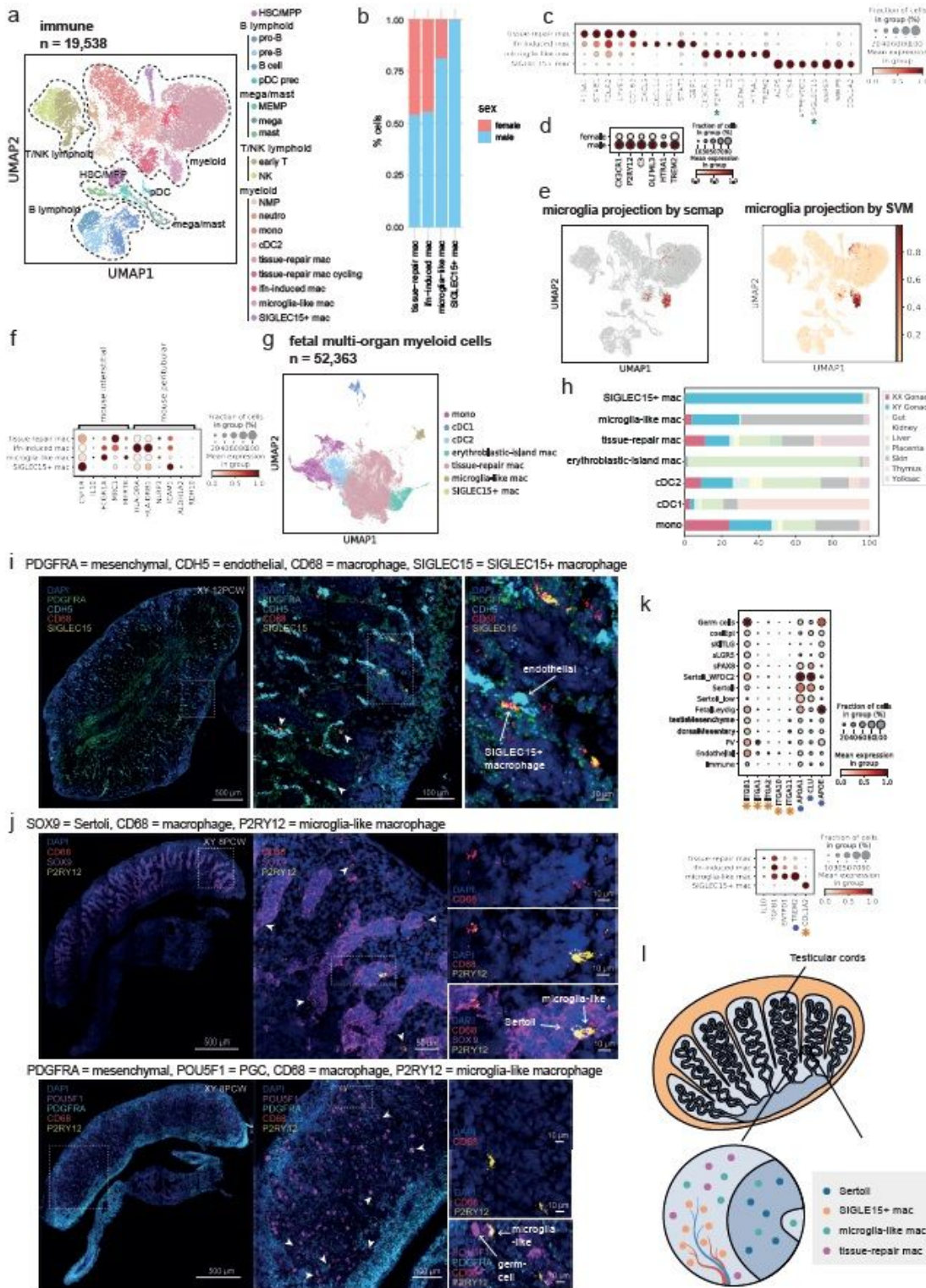
Figure 6

Tissue-resident macrophages in the developing testes. a, UMAP projections of immune cells (n=19,538) in the gonads and mesonephros labeled by cell state. 11 samples were enriched for immune (CD45+) cells. b, Bar plot showcasing the proportion of male and female cells in each macrophage subset. c, Dot plot showing variance-scaled, log-transformed expression of marker genes for the identified macrophage subsets. d, Dot plot showing the variance-scaled, log-transformed expression of microglia-like markers in

both sexes reveals that the few female cells that belong to this cluster do not express the key markers. e, Projection of microglia cells from Bian et al., 2020 onto our gonadal immune manifold using scmap (left) and a SVM classifier (right). f, Dot plot showing the variance-scaled, log-transformed expression of marker genes of interstitial and peritubular mouse macrophages in our human gonadal macrophages. g, UMAP projections of myeloid cells from embryonic/fetal gonads integrated using single-cell Variational Inference (scVI) with myeloid cells from embryonic/fetal gut, kidney, liver, lung, placenta, skin, thymus and yolk sac (n=52,363). h, Embryonic/fetal myeloid cell type abundance (% cells) in different organs. SIGLEC15+ macrophages and microglia-like macrophages are two novel subsets present in XY gonads. SIGLEC15+ macrophages are exclusive to XY gonads while microglia-like macrophages are found primarily in XY gonads and skin. i, High-resolution imaging of a representative XY gonadal section (12 PCW), with intensity proportional to smFISH signal to PDGFRA (green, mesenchymal), CDH5 (cyan, endothelial cells), CD68 (red, macrophages), SIGLEC15 (yellow, SIGLEC15+ macrophages). SIGLEC15+ macrophages distribute outside of the testicular cords in proximity to endothelial cells (as shown by white arrows and close up marked by white dashed rectangle). j, High-resolution large-area imaging of representative gonadal sections of one fetal testis (8 PCW), with intensity proportional to smFISH signal to SOX9 (purple, Sertoli cells), POU5F1 (purple, primordial germ cells), CD68 (red, macrophages), P2RY12 (yellow, microglia-like macrophages), PDGFRA (cyan, mesenchymal). Microglia-like macrophages (white arrows) are observed adjacent to the germ and Sertoli cells compartments. White dashed rectangles highlight gonadal regions magnified. k, Dotplots showing variance-scaled, logtransformed expression of ligands and receptors involved in the interactions between microglia-like and SIGLEC15+ macrophages and gonadal cells in their proximity as observed with high-resolution imaging. l, Schematics illustrating the spatial location of the distinct testicular macrophage populations. Mac = Macrophages; cDC = conventional Dendritic cells; pDC = plasmacytoid Dendritic cell; Mono = monocytes; NK = Natural Killer cells.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInfoupload.pdf
- SupplementaryTable2upload.xlsx
- SupplementaryTable3upload.xlsx
- SupplementaryTable4upload.xlsx
- SupplementaryTable5upload.xlsx
- SupplementaryTable6upload.xlsx
- SupplementaryTable7upload.xlsx
- SupplementaryTable8upload.xlsx
- SupplementaryTable9upload.xlsx
- SupplementaryTable10upload.xlsx

- SupplementaryTable11upload.xlsx
- SupplementaryTable12upload.xlsx
- SupplementaryTable13upload.xlsx
- SupplementaryTable14upload.xlsx
- SupplementaryTable16upload.xlsx