Dissecting Cell Lineage Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell Transcriptomics

STEVANT, Isabelle, et al.

Abstract

Sex determination is a unique process that allows the study of multipotent progenitors and their acquisition of sex-specific fates during differentiation of the gonad into a testis or an ovary. Using time series single-cell RNA sequencing (scRNA-seq) on ovarian Nr5a1-GFP+ somatic cells during sex determination, we identified a single population of early progenitors giving rise to both pre-granulosa cells and potential steroidogenic precursor cells. By comparing time series single-cell RNA sequencing of XX and XY somatic cells, we provide evidence that gonadal supporting cells are specified from these early progenitors by a non-sex-specific transcriptomic program before pre-granulosa and Sertoli cells acquire their sex-specific identity. In XX and XY steroidogenic precursors, similar transcriptomic profiles underlie the acquisition of cell fate but with XX cells exhibiting a relative delay. Our data provide an important resource, at single-cell resolution, for further interrogation of the molecular and cellular basis of mammalian sex determination.

Reference


DOI : 10.1016/j.celrep.2019.02.069
PMID : 30893600

Available at: http://archive-ouverte.unige.ch/unige:115788

Disclaimer: layout of this document may differ from the published version.
Dissecting Cell Lineage Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell Transcriptomics

**Highlights**

- XX *Nr5a1* progenitors give rise to pre-granulosa and steroidogenic precursor cells
- Supporting lineages commit similarly with the exception of *Sry* in XY cells
- Sertoli and granulosa cell differentiation is temporally asymmetric
- XY and XX progenitors progressively acquire a steroidogenic precursor fate

**In Brief**

Using single-cell RNA sequencing of *Nr5a1*-expressing gonadal somatic cells during female and male sex determination, Stévant et al. deconvoluted the cell lineage specification process and sex-specific differentiation of both the supporting and the steroidogenic cell lineages at a transcriptomic level.
Dissecting Cell Lineage Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell Transcriptomics

Isabelle Stévant,1,2,3 Françoise Kühne,1 Andy Greenfield,4 Marie-Christine Chaboissier,5 Emmanouil T. Dermitzakis,1,2,3 and Serge Nef1,2,6,*

1Department of Genetic Medicine and Development, University of Geneva, 1211 Geneva, Switzerland
2GE3, Institute of Genetics and Genomics of Geneva, University of Geneva, 1211 Geneva, Switzerland
3SIB, Swiss Institute of Bioinformatics, University of Geneva, 1211 Geneva, Switzerland
4Mammalian Genetics Unit, Medical Research Council, Harwell Institute, Oxfordshire OX11 0RD, UK
5Université Côte d’Azur, CNRS, Inserm, iBV, Nice, France
6Lead Contact
*Correspondence: serge.nef@unige.ch
https://doi.org/10.1016/j.celrep.2019.02.069

SUMMARY

Sex determination is a unique process that allows the study of multipotent progenitors and their acquisition of sex-specific fates during differentiation of the gonad into a testis or an ovary. Using time series single-cell RNA sequencing (scRNA-seq) on ovarian Nr5a1-GFP+ somatic cells during sex determination, we identified a single population of early progenitors giving rise to both pre-granulosa cells and potential steroidogenic precursor cells. By comparing time series single-cell RNA sequencing of XX and XY somatic cells, we provide evidence that gonadal supporting cells are specified from these early progenitors by a non-sex-specific transcriptomic program before pre-granulosa and Sertoli cells acquire their sex-specific identity. In XX and XY steroidogenic precursors, similar transcriptomic profiles underlie the acquisition of cell fate but with XX cells exhibiting a relative delay. Our data provide an important resource, at single-cell resolution, for further interrogation of the molecular and cellular basis of mammalian sex determination.

INTRODUCTION

Testes and ovaries have the same developmental origin: the gonadal primordia. These start developing around embryonic day (E) 9.5 in mice by thickening and proliferating coelomic epithelia on the ventromedial surface of the mesonephroi (By-skov, 1986). Before sex determination, the gonadal primordia, also called bipotential gonads, are composed of multipotent somatic progenitor cells that are competent to adopt one or the other sex-specific cell fate and of migrating primordial germ cells. Sex determination is initiated around E11.5, leading to the supporting cell lineage differentiating as Sertoli cells in XY gonads following the expression of Sry (Albrecht and Eicher, 2001; Sekido et al., 2004) or as pre-granulosa cells in XX with the stabilization of WNT/β-catenin signaling (Chassot et al., 2008, 2012). Following supporting cell differentiation, this sex fate decision propagates to the germ cells and the other somatic lineages, including the steroidogenic cells (Leydig cells in XY and theca cells in XX) that later drive acquisition of primary and secondary sexual characteristics through hormonal control.

Over the past two decades, intensive efforts have been made to investigate the genetic network at play during sex determination and gonadal cell differentiation, using large-scale transcriptomic methods, such as microarrays, on whole gonads or sorted cells (Beverdam and Koopman, 2006; Bouma et al., 2007, 2010; Jameson et al., 2012; Munger et al., 2013; Nef et al., 2005; Small et al., 2005). More recently, RNA sequencing (RNA-seq) and single-cell RNA sequencing (scRNA-seq) have been used (Inoue et al., 2016; McClelland et al., 2015; Stévant et al., 2018; Zhao et al., 2018). These studies revealed the bipotential state of the XX and XY gonads before sex determination and highlighted the gene expression dynamics and mutually antagonistic genetic programs controlling the establishment of sexual identity in the supporting cell lineage. However, because of the lack of specific markers, little is known about the establishment of gonadal cell lineages before sex determination.

Transcriptomic investigation of sex determination faces two main challenges. First, the gonadal cells differentiate in a non-synchronous manner, as demonstrated by the wave of expression of Sry from the center to the pole of the gonad between E10.5 and E12.5 (Bullejos and Koopman, 2001; Jeske et al., 1995; Koopman et al., 1990; Warr et al., 2012). Second, the gonad is composed of multiple cell types, including potentially as-yet-uncharacterized subpopulations, particularly in the bipotential gonads and in the testicular interstitial and ovarian stromal compartments. By using bulk transcriptomic assays on whole gonads or purified cell populations, gene expression measurements are derived from a mix of heterogeneous cell types of varying differentiation status, resulting in an averaged picture of the genetic program that does not accurately reflect the complexity of events. By applying time series single-cell RNA sequencing on XX Nr5a1-GFP gonadal cells, in combination with our previously published XY Nr5a1-GFP single-cell RNA sequencing data (Stévant et al., 2018), we captured
transcriptomic snapshots of the most abundant gonadal somatic cell populations as they differentiate in both sexes. By focusing on the differentiation status and not on the embryonic stages, we reconstructed the chronology of transcriptomic events underlying cell lineage specification and the appearance of sexual dimorphisms in the differentiating gonads.

We show that commitment to supporting cell fate from early gonadal progenitor cells is a dynamic process involving the up-regulation of around 200 genes in a non-sex-specific manner, with the exception of genes from the sex chromosomes, before their differentiation as either Sertoli or granulosa cells. We also show that Nr5a1-GFP progenitor cells of the ovarian stroma exhibit the same characteristics as testicular and steroidalogenic progenitor cells and gradually express steroidalogenic precursor marker genes. Our analysis provides a precise view of cell lineage specification and the establishment of sexual dimorphism at a cellular level in the most abundant gonadal somatic cell populations, and it represents an important resource that complements previously published transcriptomic analyses.

RESULTS

**XX Somatic Cells Are Classified in Four Transcriptionally Distinct Cell Populations**

To characterize and reconstruct the somatic cell lineages of the developing XX gonad as ovary development proceeds, we purified the somatic cells at six developmental stages of gonadal differentiation (E10.5, E11.5, E12.5, E13.5, E16.5, and post-natal day 6 [P6]) using the Tg(Nr5a1-GFP) transgenic mouse (Figure 1A) (Stallings et al., 2002). Nr5a1 is expressed specifically in gonadal somatic cells giving rise to the supporting and the steroidalogenic lineages, in both male (XY) and female (XX) gonads, and in a time window large enough to cover the bipotential state before sex determination and the whole of gonadal development (Nef et al., 2005; Stévant et al., 2018). Thus, the Tg(Nr5a1-GFP) transgenic mouse constitutes a powerful tool for isolating gonadal somatic cells and studying their differentiation (Figures 1B and 1C; Figure S1). At each relevant embryonic stage, gonads from Tg(Nr5a1-GFP) animals were dissociated and the Nr5a1-GFP+ cells were sorted by fluorescence-activated cell sorting (FACS) (Figure S1). GFP+ cells were isolated and processed with the Fluidigm C1 Autoprep system at a sample size allowing the detection of cell populations representing a minimum of 10% of the total cells and were sequenced at saturation (full-length RNA sequencing, 100 bp paired-end reads, 10 million reads per cell) (Figure 1D). A total of 563 cells remained after filtering based on various quality control metrics (STAR Methods).

We classified the somatic cell populations present in the developing ovary using the same method we developed for the analysis of testis development; i.e., we selected the highly variable genes and performed a principal-component analysis (PCA) and hierarchical clustering on the significant principal components (Stévant et al., 2018) (STAR Methods). We obtained four cell clusters combining different embryonic stages (clusters C1 to C4) (Figures 1E and 1F). Expression enrichment of known markers and differentially expressed genes (Figures 1G and 1H; Figure S2; Data S1) allowed us to assign the identity of the cell clusters. Clusters C1 and C2 represent Nr2f2-expressing progenitor cells. Cluster C1 forms the early progenitor cell population, with cells from E10.5 to E13.5, and expresses markers found in our previous study, such as Nr2f1 (Stévant et al., 2019); cluster C2 contains cells from E13.5 onward, which are stromal progenitor cells that express the stromal marker gene Mal (DeFalco et al., 2011) (Figure 1G) and, surprisingly, some markers of fetal Leydig cell progenitors, Tcf21 and Pdgfra (Brennan et al., 2003; Cui et al., 2004; Stévant et al., 2018) (Figure 1H).

In contrast, cells from clusters C3 and C4 coexpress the supporting cell marker Amhr2 (Baarends et al., 1995), Foxl2 (Mork et al., 2012; Schmidt et al., 2004; Uda et al., 2004), and Fst (Bourma et al., 2007) and represent granulosa cells at different stages of their differentiation (Figures 1G and 1H). Cluster C3 contains fetal cells from E11.5 to E16.5 and some from post-natal day 6 and expresses genes related to pre-granulosa cells, such as Lgr5 (Figure 1H; Figure S2). The C4 cluster contains mostly post-natal day 6 cells that express estrogen receptor Erś2 and the steroid-related genes Hsd3b1 and Hsd17b1. Therefore, we named the C3 cluster “pre-granulosa” and the C4 cluster “granulosa” to distinguish these two populations.

Over-represented Gene Ontology (GO) terms associated with the differentially expressed genes in each cell cluster (Figure 1; Data S2) indicate that the early progenitor cells are proliferating cells, whereas the stromal progenitors express genes related to morphological organization (“extracellular matrix organization” and “positive regulation of cell migration”). The pre-granulosa and granulosa cells express genes related to lipid metabolic processes and hormone secretion, indicating that these cells have the capacity to produce hormones as early as fetal life (Dutta et al., 2014).

To summarize, the time series single-cell RNA sequencing of the Nr5a1-GFP+ cells of the developing XX gonad from E10.5 to post-natal day 6 identified four cell populations, including early and stromal progenitor cells, pre-granulosa cells, and post-natal granulosa cells. The absence of a cell cluster with a theca cell signature at post-natal day 6 indicates that either theca cells are not yet differentiated at post-natal day 6 or they constitute a rare cell population (<10% of the Nr5a1-GFP+ cells) and were not captured by our sample size as a consequence.

**Cell Lineage Reconstruction Identifies the Dynamics of Gene Expression during Ovarian Fate Commitment**

The reconstruction of the Nr5a1-GFP+ cell lineages in the developing XX gonad allowed us to identify transition states leading to the differentiation of the granulosa cells and the stromal cells from a common progenitor cell population (Figures 2A and 2B) (STAR Methods). We observe that the early progenitor cells give rise subsequently to the granulosa cell lineage and the stromal progenitor cell lineage around E11.5–E12.5.

By ordering the cells along a pseudotime (Street et al., 2018), we identified genes that are dynamically expressed during the specification of the two cell lineages (differentially expressed genes along the pseudotime, q < 0.05) (Figure 2C) (STAR Methods). A total of 3,916 genes reveal a dynamic expression profile through time, with 1,733 genes restricted to the granulosa cell lineage, 1,059 genes restricted to the stromal progenitor cell lineage, and 1,124 genes dynamically expressed in the two cell lineages.
lineages (Data S3). We represented the smoothed gene expression level (loess regression) of the two cell lineages with a double heatmap, in which the center represents the starting point of the lineage (pseudotime 0) and the extremities represent the lineage endpoints (pseudotime 100) of the stromal progenitors (left) and the granulosa cell lineages (right), respectively (Figure 2C). We classified these genes by expression patterns (G1–G17), and by cell-specific expression categories (a–g) to eventually identify enrichment of biological processes through a GO term over-representation test (Figure 2D; Data S4). The heatmap revealed that granulosa cells diverge from the early progenitor cells with a robust differentiation program mediated by thousands of genes.
compared to the stromal progenitor cells, which exhibit many fewer dynamic genes.

The genes that are overexpressed in the common progenitor cells of the undifferentiated gonad (patterns G1 to G4, or category a) (Figure 2C) are related to mitotic cell division, mesonephros development, positive regulation of stem cell development, and epithelium morphogenesis, consistent with their coelomic epithelial origin (Figure 2D; Data S4). Their expression level decreases during commitment to both pre-granulosa and stromal progenitor fate, suggesting a cell identity conversion.

Figure 2. Cell Lineage Reconstruction and Identification of the Genetic Program Driving Granulosa and Stromal Cells
(A and B) Diffusion map of the most variable genes and reconstruction of the cell lineages. Dots represent cells, and black lines represent estimated cell lineages. (A) is colored by embryonic stages and (B) is colored by cell clusters.
(C) Heatmap representing the dynamics of gene expression in the progenitor (left) and granulosa (right) cell lineage through the pseudotime. Early progenitor cells common to both lineages are located in the center (0) of the map, with divergence of granulosa cells to the right and evolution of progenitor cells to the left. Gene expression was normalized to the mean, and classified with k-means (k = 17). They were categorized according to the cell types in which they are overexpressed (a–g). The dotted lines mark the limit of the cell clusters.
(D) Top results from the Gene Ontology (GO) enrichment test showing the terms associated with the gene categories defined in (C).
(E and F) Expression profiles of genes that become restricted to granulosa cells (E) or stromal progenitors (F). The solid line represents the loess regression, and the fade band is the 95% confidence interval of the model.
We noted that this cell conversion occurs around E11.5–E12.5 in the granulosa lineage, while it occurs from E13.5 in the stromal progenitor lineage (Figures 2A–2C). This suggests that the progenitor cells remain competent to be recruited as pre-granulosa cells as late as E12.5.

The differentiation of the granulosa cell lineage is characterized by highly dynamic transcriptomic profiles involving transient and permanent activation of genes. Genes overexpressed in pre-granulosa cells only (category c, 608 genes) are classified in three expression profiles (G6 to G8) (Figure 2C; Data S3). The G6 profile is composed of 62 genes that are transiently overexpressed at E11.5 and E12.5, at the onset of pre-granulosa cell differentiation. Among these genes, we found the male testis-maintenance gene Dmrt1 (Lei et al., 2007); Cyp11a1, as previously observed in the pre-Sertoli cells (Stévant et al., 2018; Val et al., 2007); and Lgr4, known to act as a receptor of RSPO1 and a promoter of Wnt/β-catenin signaling (Koizumi et al., 2015) (Data S3). The profiles G9 to G11, or category d, contain 368 genes that are expressed as soon as the pre-granulosa cells differentiate and are maintained after birth (Figure 2C; Data S4). These genes are mostly related to lipid metabolic processes (Figure 2D; Data S4). In this category, we also found genes known to be involved in ovarian development, such as Kiti (Hutt et al., 2006; Jones and Pepling, 2013), Fst (Kashimada et al., 2011), and the steroidogenic acute regulatory protein Star (Caron et al., 1997). Finally, the profiles G12 and G13, or category e, contain 242 genes that are overexpressed in the post-natal granulosa cells at post-natal day 6 (Figure 2C; Data S3). This includes genes that have been previously described as expressed in fetal Sertoli cells, such as Amh, which is also known to control primordial follicle recruitment (Durlinger et al., 1999, 2002); Aard (Bouma et al., 2010); and Mro (Smith et al., 2003, 2006). We also found the expression of Igtf1r (Baugarten et al., 2017), which is required for steroidogenesis, and Inha and Inhbb (Findlay, 1993; Mather et al., 1997; Weng et al., 2006) (Data S3). Figure 2E provides examples of genes upregulated in developing granulosa cells, including Runx1, Cdkn1b, Fox2l, and Lgr5. Although FOXL2 is known to be strongly expressed in granulosa cells, the expression profile reported here is consistent with reports indicating additional Fox2 expression in some postnatal cells (Matson et al., 2011). There is also published evidence that FOXL2-positive gonadal precursor cells give rise to both granulosa cells and theca cells, based on classical lineage-tracing experiments (Jilnenhaut et al., 2009).

In contrast to the highly dynamic program mediating granulosa cell differentiation, the progenitor cell lineage displays less variation in gene expression during ovarian development. Gene category f regroups 173 genes that are expressed in the stromal progenitor lineage, with the pattern G14 containing genes expressed as early as E10.5 and overexpressed at post-natal day 6 and the pattern G15 containing genes overexpressed at post-natal day 6. Among them, we found genes known as markers of steroidogenic cell precursors, such as Wnt5a (Stévant et al., 2018), Pdgfra (Brennan et al., 2003), Tcf21 (Bhandari et al., 2012; Cui et al., 2004), Gli2 (Barsoum and Yao, 2011), Arx (Miyabayashi et al., 2013), and the secreted negative regulator of the Wnt signaling pathway, Sfrp1 (Warr et al., 2009) (Figure 2F). These results suggest that the progenitor cell lineage undergoes transcriptional changes that restrict its competence toward a steroidogenic fate required for the differentiation of theca cells.

Overall, the reconstruction of the Nr5a1-GFP+ cell lineage reveals that before female sex determination, there exists a single, highly proliferative progenitor cell population. A subset of this population differentiates first as pre-granulosa cells and ultimately as granulosa cells, driven by a dynamic transcriptional program composed of the transient expression of hundreds of genes, including genes implicated in the Wnt signaling pathway. Conversely, the remaining progenitor cell population exhibits transcriptomic changes that restrict its competence toward a steroidogenic fate during fetal ovarian development, ultimately expressing genes known as markers of steroidogenic cell precursors in the testis.

**Integration of Single-Cell RNA Sequencing Data from Both XX and XY Nr5a1-GFP+ Cells Reveals the Establishment of Sexual Dimorphism**

One of the biggest advantages of single-cell RNA sequencing compared to cell-sorted bulk transcriptomic assays for studying sex determination is that it avoids the confounding effects of asynchronous cell differentiation, allowing identification of transcriptomic events associated with lineage specification and sex-specific cell differentiation. By combining single-cell RNA sequencing data from 400 XY Nr5a1-GFP+ cells (Stévant et al., 2018) with the present data from 563 XX Nr5a1-GFP+ cells, we investigated the differences and similarities of cell lineage specification, as well as the establishment of sexual dimorphism in both sexes.

With clustering, we merged the two datasets and applied the same clustering method used previously to classify cells by transcriptomic similarity without a priori knowledge concerning the genetic sex, and we obtained five major cell clusters (clusters D1 to D5) (Figures 3A–3C) (STAR Methods). The early progenitor populations from XX and XY gonads cluster together (cluster D1), as well as the XX stromal and the XY interstitial progenitors (cluster D2), even though we observe a tendency to segregate by sex in the t-distributed stochastic neighbor embedding (t-SNE) representation (Figure 3C). This suggests that the progenitor cell lineages of both XX and XY gonads do not display sufficient sexual dimorphism to permit segregation and be considered different cell types, even late in development. Cluster D3 contains not only pre-granulosa cells from E11.5 to post-natal day 6 but also E11.5 pre-Sertoli cells. Clusters D4 and D5 contain granulosa cells at post-natal day 6 and Sertoli cells from E12.5 to E16.5, respectively. This suggests that the supporting cells emerge from the progenitors with a similar transcriptomic program and that Sertoli cells differentiate with more pronounced and dynamic transcriptomic changes when compared to pre-granulosa cells, which complete their differentiation after birth.

The diffusion map and the lineage reconstruction (Figures 3D–3F) confirm what was observed using clustering. The progenitor cell lineage combines both sexes from E10.5 to late stages (Figures 3D and 3E), and the supporting cell lineage diverges from progenitors independently of genetic sex (branchpoint 1 [BP1]) and subsequently gives rise to Sertoli and granulosa cells with distinct timings (branchpoint 2).
Commitment to the Supporting Cell Lineage Involves a Common Intermediate Differentiation Step before the Emergence of Sexual Dimorphism

We analyzed how the supporting cell lineage emerges and acquires its respective sex-specific cell types. We first evaluated when the supporting cell lineage transcriptomes become sexually dimorphic. To do so, we used cell lineage reconstruction to divide the supporting cell lineage into three developmental windows (a, b, and c) defined by the branchpoints of the lineage trajectories (Figures 3F and 4A), and we determined whether genes display sexual dimorphisms within these windows (differentially expressed genes, q < 0.05) (Figures 4A and 4B; Data S5). Before the supporting cell commitment (branchpoint 1), the early progenitor cells display relatively few sexually dimorphic genes (12 overexpressed in XX and 65 overexpressed in XY). Except for four genes located on the Y chromosome, these genes are expressed in both sexes but at different levels; later, they are not sexually dimorphic, suggesting that parts of these genes are potential false positives (Figure 4B). Although cells commit to the supporting cell lineage, between branchpoint 1 and branchpoint 2, they display few sexually dimorphic genes (11 overexpressed genes in XX and 40 overexpressed genes in XY). Among them, we found the earliest testis-determining factors Sry and Sox9 in XY and Fst in XX. 27 of the 40 XY overexpressed genes remain sexually dimorphic after branchpoint 2, against 5 of 11 genes for XX. The number of sexually dimorphic genes dramatically increases as cells specify as Sertoli or granulosa after branchpoint 2, with 656 genes overexpressed in XX and 808 genes overexpressed in XY. These results demonstrate that the establishment of sexual dimorphism in the supporting cell lineage mainly occurs after branchpoint 2, which corresponds to the transition between pre-Sertoli and Sertoli cells from E12.5 in XY and around E12.5–E13.5 in pre-granulosa cells in XX (Figures 4A and 4B).

We then compared the expression dynamics of Sertoli and granulosa cell differentiation by classifying dynamically expressed genes by expression profiles in both sexes (k-means, k = 25) (Data S5), and we represented on a double heatmap...
the averaged $Z$ scores of each of the expression profile clusters (F1–F25) (Figure 4C). We see that the genes expressed in early progenitor cells are downregulated during the commitment of the supporting cells of both sexes (gene profiles F1 to F10) (Figure 4C). Then, we observed transient upregulation of 168 genes at the onset of both pre-Sertoli and pre-granulosa cell differentiation (gene profile F11) (Figures 4C and 4D). These genes represent a narrow window of expression in E11.5 pre-Sertoli cells, while they are overexpressed from E11.5 to E16.5 in the pre-granulosa cells. In this expression profile, we found the transient receptor potential cation channel $Trp7$, $Gadd45g$, which is required for normal Sry expression via mitogen-activated protein kinase (MAPK) signals (Gierl et al., 2012; Warr et al., 2012), and the pancreatic lipase $Pnlip$ (Figure 4D). We also found in this profile genes that are upregulated at the onset of supporting cell commitment but that exhibit a delay in downregulation in XX compared to XY, including $Runx1$ (Munger et al., 2013; Nef et al., 2005), $Ifitm1$, and $Podxl$ (Herrera et al., 2005) but also the early gonadal ridge promoter genes $Emx2$ and $Lhx9$ (Birk et al., 2000; Miyamoto et al., 1997) (Figure 4E; Data S5). These data suggest that pre-granulosa cells maintain a progenitor-like state longer than pre-Sertoli cells. With this analysis, we also observed a considerable number of genes that share the same expression behavior during differentiation of Sertoli and granulosa cells (gene profiles F12 to F16, 931 genes) (Figure 4C).

The differentiation of supporting cells during gonad development consists first of the commitment of the bipotential supporting cells from the early progenitor cell population, involving the upregulation of around 200 genes in both sexes. Apart for the up-regulation of Sry and the activation of the Sox9 regulatory network in XY gonads and the upregulation of $Fst$ in XX gonads, the commitment of the supporting cell lineage is not a sexually dimorphic process. This step might be necessary to establish the bipotential state of the supporting cells to prepare them for a sex fate decision in the presence or the absence of the Y chromosome, consistent with the observation that pre-granulosa cells are also able to activate the Sry promoter in the same time window as pre-Sertoli cells (Albrecht and Eicher, 2001; Harikae et al., 2013). We also noticed that pre-granulosa cells exhibit
this bipotential state for a day longer (E11.5 and E12.5 cells) (Figure 4C) compared to pre-Sertoli cells (E11.5 cells mainly) and that a significant proportion of pre-granulosa cells maintains a bipotential supporting cell signature throughout fetal life until this is lost after birth. In contrast, the bipotential signature of the Sertoli cell lineage is lost from around E12.5, right after the peak of Sry expression, reflecting the rapid effect of Sry on the activation of the Sertoli cell differentiation program.

Stromal and Interstitial Cells Display Progressive Sexual Dimorphism and Differ in Timing of Expression

We repeated our analysis of the progenitor cell lineage to examine the emergence of sexual dimorphism as the XY interstitial and the XX stromal cells progress during gonadal development. We observe that the progenitor cells display sexual dimorphism after branchpoint 1, with 748 genes overexpressed in XY (Figure 5D). We also identified a few genes that become sexually dimorphic, such as Foxl2 and Cfh, which are expressed in postnatal day 6 XX stromal cells, and Eit4 and Inhba, which are overexpressed in XY interstitial cells only from around E12.5–E13.5 (Figures 5E and 5F; Data S6).

These results suggest that the progenitor cell lineage commits to the steroidogenic fate earlier in the XY gonad to allow these cells to differentiate as fetal Leydig cells, whereas XX stromal cells commit to steroidogenic progenitor cells from E16.5 onward to ultimately differentiate as theca cells after birth, when granulosa cells also complete their differentiation.

DISCUSSION

With this study, we aimed to analyze the transcriptomic programs at play during sex determination in Nr5a1+ somatic cells...
of XX and XY mouse gonads. Subsequent to our previously published analysis of the early testis development (Stévant et al., 2018), we sequenced and analyzed the transcriptome of hundreds of individual XX Nr5a1-GFP+ gonadal somatic cells, from the bipotential gonads at E10.5 to post-natal ovaries at post-natal day 6. By combining these transcriptomic data from XX and XY cells, we were able to reconstruct the sequence of transcriptomic events underlying the cell lineage specification and the sex-specific cell differentiation of both the supporting cell lineage and the steroidogenic precursor cell lineage from a common multipotent progenitor cell population.

In contrast to the testis, fetal ovarian development is characterized by modest and late morphological changes; as a consequence, early ovarian development was not well understood. The cell type contribution of the expression of critical factors such as Fst, Wnt4, or Rsop1 remained unclear in the absence of cell lineage tracing. Our unsupervised analyses of XX Nr5a1-GFP+ somatic cells from E10.5 to post-natal day 6 allow the identification of the most abundant somatic cell populations present during early ovarian development and the study of their lineage specification, their transcriptomic signatures, and the expression dynamics as they differentiate. We found a similar lineage specification pattern to that observed during early testis development. In XY gonads, this bipotential state corresponds to E11.5 pre-Sertoli cells, while in XX, it corresponds to an extended period, from E11.5 to E16.5, in pre-granulosa cells. The supporting cell precursors acquire their respective sex-specific identities around E13.5, which specify their identity as potential steroidogenic cell precursors. This study constitutes an important high-resolution, single-cell transcriptomic resource of somatic cell lineage commitment and differentiation during early ovarian development.

By combining single-cell RNA sequencing data from XX and XY Nr5a1-GFP+ gonadal somatic cells from E10.5 to late developmental stages (E16.5 in XY and post-natal day 6 in XX), we provide a more comprehensive picture of cell lineage specification during sex determination. Our experimental design, coupling a single-cell transcriptomic approach with a focus on the Nr5a1+ somatic cells of XX and XY mouse gonads, offers several advantages compared to classical bulk RNA sequencing or lineage-tracing experiments. First, characterizing the transcriptome of individual cells detaches the analysis from embryonic time and circumvents problems caused by the asynchrony in cell differentiation at these stages. Thus, this approach allows reconstruction of the chronology of transcriptomic events underlying cell lineage specification and the establishment of sexual dimorphism. Second, comparison of XX and XY cells allows precise dissection of the transcriptomic changes that are sex-specific or common to both sexes during the process of cell lineage specification. We show that the supporting and interstitial or stromal cell lineages both derive from a common Nr5a1+ early progenitor cell population, present at E10.5. We also provide clear evidence that XY and XX multipotent progenitors that adopt a supporting cell fate share a similar transcriptomic identity, before initiating robust sex-dependent genetic programs leading to their differentiation into Sertoli and granulosa cells, respectively. This suggests that supporting cell commitment is initially sex independent and disconnected from sex-specific differentiation. Although the transcriptomes of XX and XY supporting cells are similar before sex-specific Sertoli cell and granulosa cell differentiation, some genes are already sexually dimorphic, including Sry and Sox9, consistent with the literature.

In addition, the transient bipotential state of supporting cell precursors reveals a temporal asymmetry between XY and XX development. In XY gonads, this bipotential state corresponds to E11.5 pre-Sertoli cells, while in XX, it corresponds to an extended period, from E11.5 to E16.5, in pre-granulosa cells. The supporting cell precursors acquire their respective sex-specific genes from E12.5 onward in Sertoli cells and from E13.5 onward in pre-granulosa cells. These results are consistent with the competence windows of supporting cells of both sexes to express Sry (Harikae et al., 2013). However, while Sertoli cells downregulate the bipotential supporting cell precursor genes, we found that pre-granulosa cells maintain expression of stem cell-related genes until as late as E16.5, indicating that pre-granulosa cells remain in an early stage of their differentiation for several days and continue their differentiation from folliculogenesis onward.

Interstitial or stromal progenitor cells commit from early progenitors from around E12.5 in XY and around E13.5 in XX. This temporal asymmetry is consistent with observations made in the supporting cell lineage. Surprisingly, interstitial or stromal progenitors share a common cell identity, regardless of the progressive appearance of sexual dimorphism, even late in embryonic development. Both XY and XX progenitor cells acquire a steroidogenic precursor fate by progressively expressing Pdgfra, Arx, or Ptc1. The supporting cells of both sexes might similarly control the specification of steroidogenic precursor cells, because they control the differentiation of theca and Leydig cells via the Hedgehog signaling pathway (Liu et al., 2015; Yao et al., 2002).

Our data provide an important resource, at single-cell resolution, for further studies on the molecular and cellular programs of testis and ovary development. They also raise multiple questions, including the identity of the signals or factors that control the specification toward either the supporting or the steroidogenic fate. Several parameters may influence this choice. First, it is possible that the fate of individual Nr5a1+ multipotent progenitors is affected by the local environment and interaction with neighboring cells. Second, it remains plausible that the multipotent progenitor population described here as a homogeneous population at the transcriptomic level is already heterogeneous at the chromatin level and composed of slightly different subpopulations primed to differentiate as supporting or steroidogenic cells. The methylation status and chromatin accessibility of these progenitor cells have not yet been investigated. Similar to embryonic stem cells (Atliasi and Stunnenberg, 2017), it has been hypothesized that the chromatin landscape

3280 Cell Reports 26, 3272–3283, March 19, 2019
in XX and XY progenitor cells of the gonad has an open configuration that confers on these multipotent cells a unique plasticity that enables differentiation into different lineages (Garcia-Moreno et al., 2018). Following cell fate commitment and sex-specific differentiation, the chromatin landscape becomes more restricted, canalizing the developmental program toward either the male or the female fate and repressing the alternative pathway. We expect that the advent of new single-cell chromatin accessibility sequencing methods for open chromatin study, combined with single-cell transcriptomic data, will be instrumental in advancing our understanding of gene regulatory network at play in each somatic cell lineage during mammalian sex determination.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Transgenic Mice
- METHOD DETAILS
  - Isolation of purified Nr5a1-GFP positive cells
  - Tissue processing and immunological analyses
  - Single-cell capture and cDNA synthesis
  - Library preparation and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.celrep.2019.02.069.

ACKNOWLEDGMENTS

This work was supported by grants from the Swiss National Science Foundation (31003A_173070 and 31PH0-141994) and by the Département de l’Instruction Publique of the State of Geneva (to S.N.). We thank Luciana Romano and Deborah Penet for the sequencing. We thank also the members of the Nef and Dermitzakis laboratories for helpful discussion and critical reading of the manuscript, Cécile Gameiro from the flow cytometry facility of the iGE3 Genomics Platform for the single-cell capture.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 12, 2018
Revised: January 7, 2019
Accepted: February 19, 2019
Published: March 19, 2019

REFERENCES


## STAR METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFP Polyclonal Antibody</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# A10262, RRID:AB_2534023</td>
</tr>
<tr>
<td>Human COUP-TF II/NR2F2 Antibody</td>
<td>Perseus Proteomics</td>
<td>Cat# PP-H7147-00, RRID:AB_2155627</td>
</tr>
<tr>
<td>Rabbit anti-FOXL2</td>
<td>Gift from Dagmar Wilhelm</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trypsin-EDTA (0.05%), phenol red</td>
<td>Thermo Fisher Scientific</td>
<td>25300054</td>
</tr>
<tr>
<td>DPBS, no calcium, no magnesium</td>
<td>Thermo Fisher Scientific</td>
<td>14190144</td>
</tr>
<tr>
<td>Fetal bovine serum</td>
<td>Thermo Fisher Scientific</td>
<td>26140087</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1 IFC for mRNA-seq (5-10 μM)</td>
<td>Fluidigm</td>
<td>1005759</td>
</tr>
<tr>
<td>C1 Single-Cell Auto Prep Kit for mRNA Seq</td>
<td>Fluidigm</td>
<td>100-6209</td>
</tr>
<tr>
<td>C1 Single-Cell Auto Prep Reagent Kit for mRNA Seq</td>
<td>Fluidigm</td>
<td>100-6201</td>
</tr>
<tr>
<td>SMARTer Ultra Low RNA Kit for the Fluidigm C1 System</td>
<td>Fluidigm</td>
<td>634833</td>
</tr>
<tr>
<td>Nextera XT DNA Sample Preparation Kit</td>
<td>Illumina</td>
<td>FC-131-1096</td>
</tr>
<tr>
<td>Nextera XT DNA Sample Preparation Index Kits A-D</td>
<td>Illumina</td>
<td>FC-131-2001-4</td>
</tr>
<tr>
<td>Agencourt AMPure XP beads</td>
<td>Beckman Coulter</td>
<td>A63880</td>
</tr>
<tr>
<td>Qubit dsDNA High Sensitivity</td>
<td>Life Technologies</td>
<td>Q32854</td>
</tr>
<tr>
<td>Agilent High Sensitivity DNA Kit Reagents</td>
<td>Agilent</td>
<td>5067-4626</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw data, normalized counts matrix for XY data</td>
<td>GEO</td>
<td>GSE97519</td>
</tr>
<tr>
<td>Raw data, normalized counts matrix for XX data</td>
<td>GEO</td>
<td>GSE119766</td>
</tr>
<tr>
<td><strong>Scripts for analysis</strong></td>
<td>GitHub</td>
<td><a href="https://github.com/IStevant/mouse-gonad-scRNA-seq">https://github.com/IStevant/mouse-gonad-scRNA-seq</a></td>
</tr>
<tr>
<td><strong>Experimental Models: Organisms/Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mus musculus</em>: CD-1</td>
<td>Charles River</td>
<td>Strain code 022</td>
</tr>
<tr>
<td><em>Mus musculus</em>: CD-1-Tg(NR5a1-GFP)</td>
<td>Stallings et al., 2002</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primers for sex genotyping</td>
<td>McFarlane et al., 2013</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Software and Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gem mapper (version 1.7.1)</td>
<td>Marco-Sola et al., 2012</td>
<td>N/A</td>
</tr>
<tr>
<td>SamTools (version 0.1.19)</td>
<td>Li et al., 2009</td>
<td><a href="http://www.htslib.org/">http://www.htslib.org/</a></td>
</tr>
<tr>
<td>R (version 3.4.3)</td>
<td>N/A</td>
<td><a href="https://www.R-project.org/">https://www.R-project.org/</a></td>
</tr>
<tr>
<td>FactoMineR</td>
<td>Lê et al., 2008</td>
<td><a href="http://factominer.free.fr/">http://factominer.free.fr/</a></td>
</tr>
<tr>
<td>Destiny</td>
<td>Angerer et al., 2016</td>
<td><a href="https://github.com/theislab/destiny">https://github.com/theislab/destiny</a></td>
</tr>
<tr>
<td>Slingshot</td>
<td>Street et al., 2018</td>
<td><a href="https://github.com/kstreet13/slingshot">https://github.com/kstreet13/slingshot</a></td>
</tr>
<tr>
<td>Monocle2</td>
<td>Qiu et al., 2017</td>
<td><a href="https://github.com/cole-trapnell-lab/monocle-release">https://github.com/cole-trapnell-lab/monocle-release</a></td>
</tr>
<tr>
<td>Ggplot2</td>
<td>Wickham, 2009</td>
<td><a href="https://ggplot2.tidyverse.org/">https://ggplot2.tidyverse.org/</a></td>
</tr>
<tr>
<td>Pheatmap</td>
<td>N/A</td>
<td><a href="https://github.com/raivokolde/pheatmap">https://github.com/raivokolde/pheatmap</a></td>
</tr>
</tbody>
</table>

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Serge Nef (serge.nef@unige.ch).
**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

**Transgenic Mice**
Wild-type CD1 and Tg(Nr5a1-GFP) in CD-1 background were used in this study. Animals were housed and cared for according to the ethical guidelines of the Direction Générale de la Santé of the Canton de Genève (experimentation number:GE-122-15).

**METHOD DETAILS**

**Isolation of purified Nr5a1-GFP positive cells**
CD-1 female mice were bred with heterozygous Tg(Nr5a1-GFP) transgenic male mice (Stallings et al., 2002). Adult females were timed and checked for the presence of vaginal plugs the next morning (E0.5). On the relevant days of gestation (E10.5, E11.5, E12.5, E13.5, E16.5, and post-natal day 6), the presence of the Nr5a1-GFP transgene in the embryos was assessed under a fluorescent binocular microscope by the presence of GFP fluorescence in the urogenital ridges. Age of the embryos was assessed by counting the tail somites (ts); embryos at 8 ts (±2 ts) were considered as E10.5, 19 ts (±2 ts) as E11.5 and 30 ts (±3 ts) as E12.5. Sexing of the embryos at E10.5 and E11.5 was performed by PCR according to McFarlane et al. (McFarlane et al., 2013) in parallel of the dissection. For each embryonic stage, XX and XY genital ridges (mesonephroi and gonad, except E16.5 and post-natal day 6, gonad only) were collected in PBS and pooled separately. PBS was removed and tissues were digested 10 minutes at 37°C with trypsin/EDTA 0.05% (GIBCO), mechanically dissociated with gentle up-and-down pipetting, and filtered through a 40μm cell strainer to obtain single cell suspension as previously described (Nef et al., 2005; Pitetti et al., 2013; Stévant et al., 2018). Nr5a1-GFP+ cells were then sorted by fluorescent-activated cell sorting (BD FACSCARIA II), excluding cell doublets, and dead cells with Draq7™ dye staining. Cells were collected in PBS and centrifuged at low speed to be concentrated at an average of 400 million cells per mL prior cell capture. The genetic sex of Nr5a1-GFP+ cells was confirmed in silico by looking at the expression of Xist, the long non-coding RNA responsible for the X chromosome inactivation in XX. As expected, the XY cells did not display any Xist expression, while all the XX cells expressed Xist at high levels, confirming the absence of contamination due to inefficiency of sex genotyping.

**Tissue processing and immunological analyses**
Embryos at relevant stages of development were collected, fixed in 4% paraformaldehyde overnight at 4°C. Tissue processing and immunological analyses embryos at E10.5 and E11.5 was performed by PCR according to McFarlane et al. (McFarlane et al., 2013) in parallel of the dissection. For each embryonic stage, XX and XY genital ridges (mesonephroi and gonad, except E16.5 and post-natal day 6, gonad only) were collected in PBS and pooled separately. PBS was removed and tissues were digested 10 minutes at 37°C with trypsin/EDTA 0.05% (GIBCO), mechanically dissociated with gentle up-and-down pipetting, and filtered through a 40μm cell strainer to obtain single cell suspension as previously described (Nef et al., 2005; Pitetti et al., 2013; Stévant et al., 2018). Nr5a1-GFP+ cells were then sorted by fluorescent-activated cell sorting (BD FACSCARIA II), excluding cell doublets, and dead cells with Draq7™ dye staining. Cells were collected in PBS and centrifuged at low speed to be concentrated at an average of 400 million cells per mL prior cell capture. The genetic sex of Nr5a1-GFP+ cells was confirmed in silico by looking at the expression of Xist, the long non-coding RNA responsible for the X chromosome inactivation in XX. As expected, the XY cells did not display any Xist expression, while all the XX cells expressed Xist at high levels, confirming the absence of contamination due to inefficiency of sex genotyping.

**Single-cell capture and cDNA synthesis**
The average diameter of the Nr5a1-GFP cells was assessed after FACS using Tali Image-Based Cytometer (Life Technology). Single cells were captured with small size IFC microfluidic chips using the C1 Single-Cell Auto Prep System (Fluidigm) following the manufacturer recommended protocol. Cell capture sites were visually checked on an inverted microscope and captured cell positions were recorded. Reverse-transcription and pre-amplification of the single-cell cDNAs were achieved within the IFC chip using the SMARTer Ultra Low RNA kit for Illumina (Clontech) according to the C1 protocol. We performed two independent captures for each embryonic stage to reach a reasonable number of cells except for E10.5 where we capture enough cells in one experiment (capture rate around 60 cells out of 96 chambers per experiment, Table S1).

**Library preparation and sequencing**
The concentration of the harvested cDNAs for each cell was normalized to 0.15 ng/μL and cells from the different experiments were randomized prior to library preparation to avoid library preparation batch effects. RNA sequencing libraries of the single-cell cDNA were prepared using the Illumina Nextera XT DNA Sample Preparation kit using the modified protocol described in the C1 documentation. Libraries were sequenced in two sequencing runs. Libraries were multiplexed within each sequencing run, distributing the embryonic stages on all the lanes and sequenced as 100bp paired-end reads using the Illumina HiSEQ2000 platform at a depth of 10M reads per cell. Due to a technical issue, the libraries of the second run were sequenced twice to obtain the expected number of read per library. The FastQ files from were subsequently merged after careful screening for batch effects.

**QUANTIFICATION AND STATISTICAL ANALYSIS**
The computations were performed at the Vital-IT (http://www.vital-it.ch) Centre for high-performance computing of the SIB Swiss Institute of Bioinformatics. FastQ files were mapped with GemTools (version 1.7.1) to the mouse reference genome (GRCm38.p3).
and the genome annotation (version M4, modified to integrate the GFP transgene and to correct Foxl2 annotation) downloaded from GENCODE (Mudge and Harrow, 2015). Bam files were generated using SamTools (version 0.1.19). Demultiplexing, mapping, read count per gene, per exon, and RPKM (Reads Per Kilobase of exon per Million reads mapped) were computed with an in-house pipeline. Quality check of the RNA sequencing was assessed by excluding samples with too low total mapped reads (< 4M mapped reads) and with the too low percentage of exonic reads (< 40%) and aberrant proportion of mitochondrial gene RPKM (> 50%) (Stévant et al., 2018). Genes from mitochondria were removed from the data prior to analysis. Data were analyzed with R version 3.4.3 (2017-11-30) - “Kite-Eating Tree.”

To classify the cells, we first selected the highly variable genes using glmgam.fit (generalized linear model by Fisher scoring with levenberg damping) (Brennecke et al., 2013). We performed a PCA with the obtained gene set, and assess the most significant PCs with Jackstraw R package (Chung and Storey, 2015). We ran HCPC clustering from FactoMineR R package (Le et al., 2008) on the significant PCs (p value < 0.05) with default parameters and performed a t-SNE with Rtsne R package to visualize the clustering result (default perplexity, using the previously calculated PCA as input).

Differential expression analysis was performed using negative binomial tests with Monocle2 (Qiu et al., 2017; Trapnell et al., 2014) with the read counts. Genes with less than 5 reads and expressed in less than 10 cells were removed. Genes with a corrected p value (or q-value) less than 0.05 were considered as significantly differentially expressed. GO term enrichment analysis was performed with clusterProfiler R package (Yu et al., 2012) and with PANTHER overrepresentation test (GO Ontology database Released 2017-08-14) (Mi et al., 2013).

Cell lineage reconstruction was performed using diffusion map Destiny R package (Angerer et al., 2016) on the set of highly variable genes. Trajectories and pseudotime were calculated with Slingshot R package (Street et al., 2018) using the diffusion map as input.

Genes differentially expressed along the pseudotime were calculated with Monocle2 using the smoothing parameter with three degrees of freedom (q-value < 0.05). For the visualization of the heatmaps and the individual gene profile graphs, the RPKMs were log transformed and was smoothed using Loess regression with the degree of smoothing (span) set to 0.5. Heatmaps and gene expression profile clustering were obtained using Pheatmap R package. Graphics were generated with Ggplot2 and figures were prepared with Inkscape.

DATA AND SOFTWARE AVAILABILITY

The accession number for the XX single-cell RNA sequencing data reported in this paper is GEO: GSE119766. The accession number for the XY single cell RNA sequencing data reported in this paper is GEO: GSE97519. R scripts generated for the analysis are available on GitHub (https://github.com/ISTevant/mouse-gonad-scRNA-seq). Both XX and XY gene expression data are included in ReproGenomics Viewer (Darde et al., 2015, 2019): http://rgv.genouest.org