

ARTICLE

Open Access

# Dissecting the fate of *Foxl2*-expressing cells in fetal ovary using lineage tracing and single-cell transcriptomics

Jingjing Zhou<sup>1,2,3,4</sup>, Xiangxiang Jiang<sup>1,2,5</sup>, Haowei Wu<sup>1,2,3</sup>, Lianjun Zhang<sup>1,6</sup>, Min Chen<sup>1</sup>, Min Chen<sup>1,2,7</sup>, Zhiming Shen<sup>1,2,3</sup>, Xudong Guo<sup>8</sup>, Hongmei Wang<sup>1,2,3</sup> and Fei Gao<sup>1,2,3</sup>

## Abstract

Gonad somatic cells acquire sex-specific fates during sex determination. In XX gonad, a subset of somatic cells expresses *Foxl2* after sex determination which is considered the progenitor of granulosa cells. However, whether these cells also contribute to other cell types at later developmental stages is unknown. In the present study, the cell fate of *Foxl2*-expressing cells in fetal ovaries was analyzed by lineage tracing and single-cell transcriptomics. We found that *Foxl2*-expressing cells gave rise to three cell types at later developmental stages, including granulosa cells, theca-interstitial cells, and stromal cells. Series single-cell RNA sequencing revealed FOXL2-positive cells were divided into two clusters at P0. One group further differentiated into granulosa cells and Theca-G (Theca-interstitial cells derived from granulosa) at P14. Another group was classified as stromal cell lineage, then a small portion of them further differentiated into  $3\beta$ -HSD-positive Theca-S (Theca-interstitial cells derived from stroma). *Cyp17a1* was expressed in Theca-S, but not in Theca-G. This study demonstrated that *Foxl2*-expressing cells in XX gonad after sex determination are multipotent and theca-interstitial cells are derived from different progenitors. Our data provided an important resource, at single-cell resolution, for a better understanding of somatic cell differentiation in ovary development.

## Introduction

In mammals, both testes and ovaries are derived from the genital ridge, which forms as a thickening of the epithelial layer on the ventromedial surface of the mesonephros<sup>1,2</sup>. Mammalian sex is determined in the fetal gonad by the presence or absence of the Y-linked high mobility group (HMG) domain transcription factor *Sry* gene<sup>3</sup>. *Sry* directs the Sertoli cell differentiation in XY gonads by inducing *Sox9* expression<sup>3–5</sup>. In XX gonads, which lack *Sry* expression, the somatic cells express FOXL2 under the

regulation of the RSPO1/WNT4- $\beta$ -catenin (CTNNB1) signaling pathway. These cells are considered the progenitors of granulosa cells and differentiate into granulosa cells in developing ovarian follicles at later developmental stage<sup>6–8</sup>. Forkhead box L2 (*FoxL2*) encodes the forkhead transcription factor 2, which is a member of the winged helix/forkhead transcription factors<sup>9</sup> and plays an essential role in female reproduction. Mutations of the *Foxl2* gene in humans are associated with Blepharophimosis Ptosis Epicanthus Inversus Syndrome (BPES). One of the phenotypes associated with type 1 BPES is a premature ovarian failure in females<sup>10</sup>. Deletion of *Foxl2* leads to ovary dysgenesis and infertility in female mice. *Foxl2* mutant granulosa cells are unable to undergo the squamous-to-cuboidal transition, which in turn causes progressive follicular depletion and infertility<sup>11,12</sup>.

Leydig cells and theca-interstitial cells are steroidogenic cells in male and female gonads, respectively. The steroid

Correspondence: Xudong Guo (bixdguo@imu.edu.cn) or Hongmei Wang (wanghm@ioz.ac.cn) or Fei Gao (gaof@ioz.ac.cn)  
<sup>1</sup>State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China  
<sup>2</sup>Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, China  
Full list of author information is available at the end of the article  
These authors contributed equally: Jingjing Zhou, Xiangxiang Jiang, Haowei Wu

© The Author(s) 2022

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

hormones produced by steroidogenic cells play essential roles in the development of germ cells and secondary sexual characteristics. Leydig cells first appear in testes at E12.5 (12.5 days post coitus), whereas theca-interstitial cells are observed postnatally in the ovaries along with the formation of primary follicles. The origins of Leydig cells<sup>13,14</sup> and theca-interstitial cells<sup>15</sup> have been investigated previously. It has been reported that theca-interstitial cells in the ovary are derived from two sources: WT1-positive cells indigenous to the ovary and GLI1-positive mesenchymal cells that migrate from the mesonephros<sup>15</sup>. Our previous studies also demonstrate that steroidogenic cells in female gonads are derived from WT1-positive cells<sup>16</sup>. Our recent study finds that the inactivation of WT1 in FOXL2-positive cells also causes these cells to transform into  $3\beta$ -HSD-positive steroidogenic cells<sup>17</sup>. These results suggest that FOXL2-positive cells in fetal gonads have the potential to develop into steroidogenic cells.

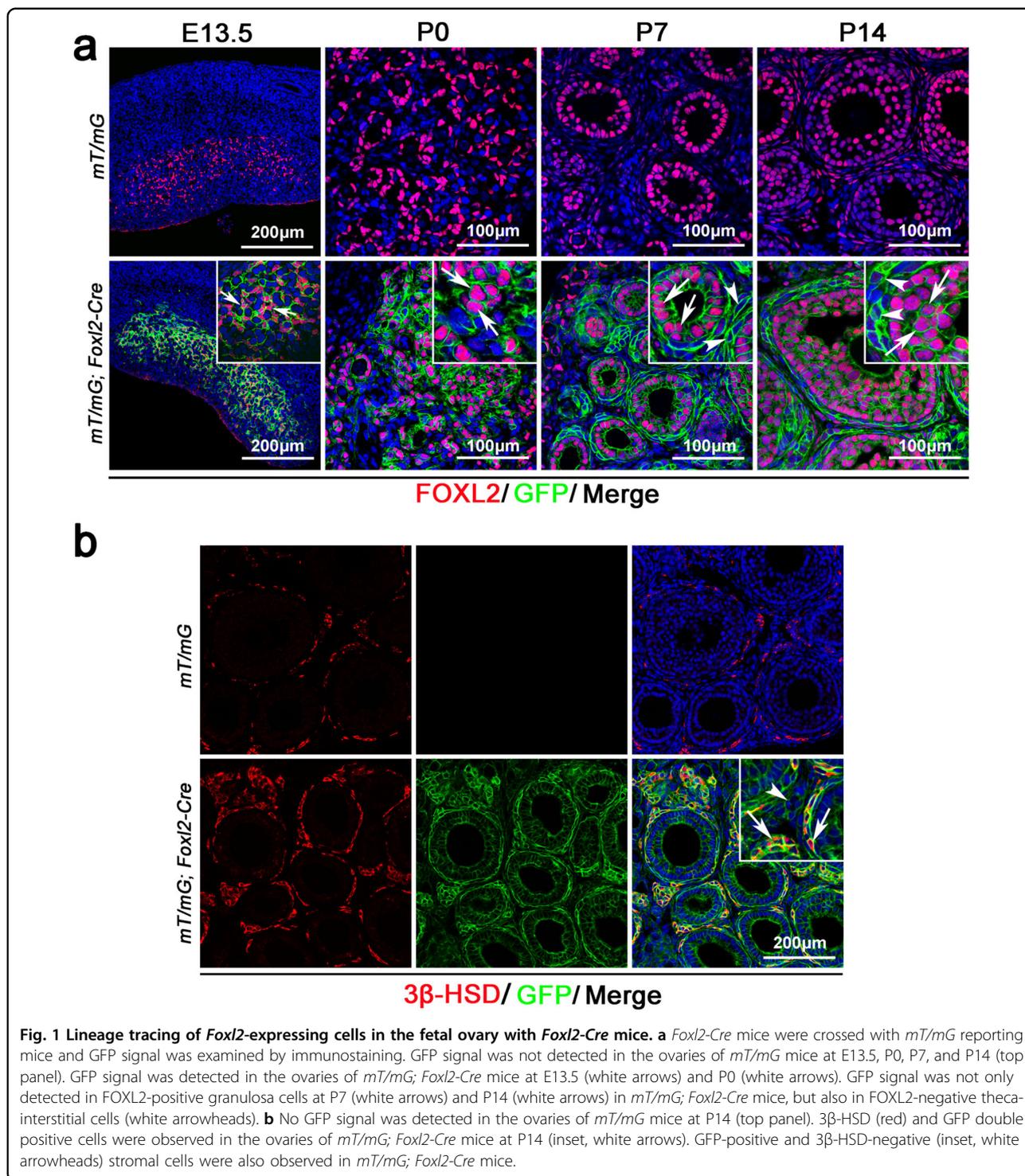
In the present study, the cell fate of FOXL2-positive cells in the fetal ovary was examined by lineage tracing and single-cell transcriptomics. We found that other than granulosa cells, these cells also gave rise to theca-interstitial cells and ovarian stromal cells. Series single-cell RNA sequencing revealed that a subset of theca-interstitial cells (Theca-G) were directly differentiated from granulosa cell lineage, whereas other  $3\beta$ -HSD-positive theca-interstitial cells (Theca-S) were derived from ovarian stromal cells which were separated from granulosa lineage at P0 (postnatal day 0). We also found that Theca-S were CYP17A1-positive, whereas Theca-G did not express CYP17A1.

## Results

### GFP signal was detected in both granulosa cells and theca-interstitial cells of *mT/mG; Foxl2-Cre* mice

To trace the fate of *Foxl2*-expressing cells in fetal ovaries, *mT/mG* reporting mice were crossed with *Foxl2-Cre* mice and the expression of GFP was examined at different developmental stages. *mT/mG* reporting mouse is a double-fluorescent Cre reporter mouse that expresses membrane-targeted tandem dimer Tomato (mT) prior to Cre-mediated excision and membrane-targeted green fluorescent protein (mG) after activation of cyclization recombination enzyme (CRE).<sup>18</sup> To test *mT/mG; Foxl2-Cre* mouse model, an immunofluorescence experiment was performed in the ovaries of *mT/mG; Foxl2-Cre* mice at different developmental stages. In *mT/mG* ovaries, all the cells were Tomato positive (red), including FOXL2-positive (green) granulosa cells (Supplementary Fig. S1a, top panel). In *mT/mG; Foxl2-Cre* ovaries, no Tomato signal (red) was detected in FOXL2-positive (green) granulosa cells (Supplementary Fig. S1a, bottom panel), indicating that Tomato signal was turned off in Cre-

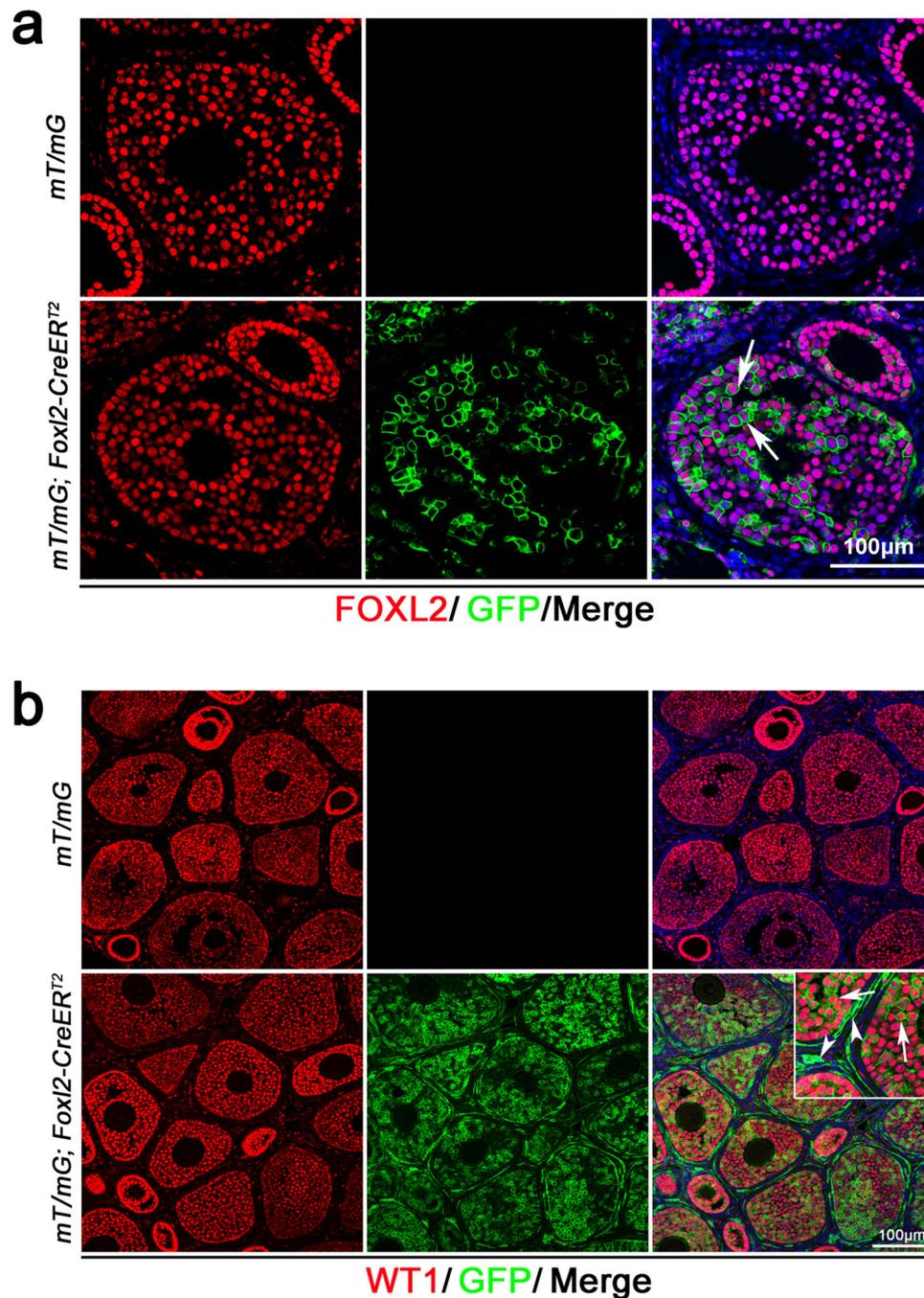
expressing cells. As shown in Supplementary Fig. S1b, no GFP signal was detected in the ovaries of *mT/mG; Foxl2-Cre* mice at E11.5, indicating that *Foxl2-Cre* was not activated at this stage. FOXL2 signal (red) was detected in *mT/mG; Foxl2-Cre* mice at E12.5, while no GFP signal was detected, indicating that *Foxl2-Cre* was not activated at E12.5. A large number of FOXL2 and GFP double positive cells were detected in *mT/mG; Foxl2-Cre* mice at E13.5 (Supplementary Fig. S1b, inset, white arrows), indicating that *Foxl2-Cre* was activated at this stage. We also found some cells surrounded by the GFP signal were negative for the FOXL2 signal (Supplementary Fig. S1b, inset, white arrowheads). Based on the nuclear morphology, these cells are most likely germ cells, not somatic cells. Since the germ cells and somatic cells were mixed together at this stage, the membrane-targeted GFP signal was most likely from adjacent somatic cells. To further confirm this result, MVH and GFP co-immunostaining was performed at E13.5 (Supplementary Fig. S1c). The germ cells were labeled with MVH (red), and no GFP signal was detected in control ovaries (Supplementary Fig. S1c, top panel). In *mT/mG; Foxl2-Cre* mice, some MVH-positive germ cells surrounded by GFP signal were observed (Supplementary Fig. S1c, inset, white arrows). Given the fact that *Foxl2-Cre* is only activated in somatic cells, we speculated that these GFP signal was located at the membrane of adjacent somatic cells. In the ovaries of control *mT/mG* mice from E13.5 to P14 (postnatal 14 days), no GFP signal was detected (Fig. 1a). In *mT/mG; Foxl2-Cre* mice, a GFP signal was detected in the gonad somatic cells at E13.5 and the GFP signal was co-localized with FOXL2 protein (Fig. 1a), indicating that GFP signal was specifically activated in *Foxl2*-expressing cells of fetal gonads. GFP signal was continually expressed in FOXL2-positive cells at P0. Interestingly, not only the FOXL2-positive granulosa cells were GFP-positive at P7 and P14, GFP signal was also observed in many FOXL2-negative cells. To further confirm this result, GFP was co-stained with WT1, which is specifically expressed in granulosa cells. Consistent with FOXL2 staining, a GFP signal was detected in both WT1-positive granulosa cells and WT1-negative somatic cells (Supplementary Fig. S2a). Based on the location and morphology, these WT1-negative cells were most likely theca-interstitial and stromal cells. To verify the cell identity, we stained GFP with steroidogenic enzyme  $3\beta$ -HSD. As shown in Fig. 1b, GFP and  $3\beta$ -HSD double positive cells were observed in the ovaries of *mT/mG; Foxl2-Cre* mice at P14, indicating that these cells were theca-interstitial cells which were responsible for steroid hormone production in female gonads. However, a small portion of GFP<sup>+</sup> $3\beta$ -HSD<sup>-</sup> cells were also observed in the interstitium of *mT/mG; Foxl2-Cre* ovaries at P14. These results indicated that FOXL2-positive cells in fetal ovaries might give rise to WT1-positive granulosa cells,



3β-HSD-positive theca-interstitial cells, and 3β-HSD-negative stromal cells.

It has been reported that a low level of FOXL2 is also expressed in theca-interstitial cells<sup>11</sup>. To exclude the possibility that the GFP signal detected in theca-interstitial is due to the ectopic *Foxl2* expression, we

crossed *mT/mG* mice with *Foxl2-CreER<sup>T2</sup>* mice and Cre activity was induced by Tamoxifen injection at 3 weeks of age. As shown in Fig. 2a, the GFP signal was exclusively detected in FOXL2-positive granulosa cells, not in theca-interstitial and stromal cells. By contrast, when Cre activity was activated at E13.5, a GFP signal was detected



**Fig. 2** Lineage tracing of *Foxl2*-expressing cells in the fetal ovary with *Foxl2-CreER<sup>T2</sup>* mice. **a** No GFP were observed in theca-interstitial cells when *Foxl2-CreER<sup>T2</sup>* mice was induced with tamoxifen at 3 weeks of age. *mT/mG; Foxl2-CreER<sup>T2</sup>* mice were induced with tamoxifen at 3 weeks of age and the GFP signal was examined by immunostaining. No GFP signal was detected in the ovaries of *mT/mG* mice after tamoxifen induction. GFP signal was only detected in FOXL2-positive granulosa cells (bottom panel, white arrows) of *mT/mG; Foxl2-CreER<sup>T2</sup>* mice and no GFP signal was detected in theca-interstitial cells. **b** GFP-positive cells were observed as both granulosa cells and theca-interstitial cells of *mT/mG; Foxl2-CreER<sup>T2</sup>* mice when Cre was induced with tamoxifen at E13.5. *mT/mG; Foxl2-CreER<sup>T2</sup>* mice were induced with tamoxifen at E13.5 and GFP signal was examined by immunostaining at 3 weeks of age. No GFP signal was detected in the ovaries of *mT/mG* mice after tamoxifen induction (top panel). GFP signal was detected in both WT1-positive granulosa cells (inset, white arrows) and WT1-negative theca-interstitial cells (inset, white arrowheads).

in both granulosa cells and theca-interstitial cells at 3 weeks of age (Fig. 2b). Furthermore, FOXL2 and GFP co-immunostaining was performed with P0 ovary of *mT/mG; Foxl2-creER<sup>T2</sup>* mice. As shown in Supplementary Fig. S2b, FOXL2<sup>+</sup>GFP<sup>+</sup> (Supplementary Fig. S2b, inset, white arrows) and FOXL2<sup>+</sup>GFP<sup>+</sup> (Supplementary Fig. S2b, inset, white arrowheads) were detected, indicating that FOXL2-positive cells have been separated into two groups at P0. These results further confirmed that FOXL2-positive cells in fetal gonads not only contributed to granulosa cells, but also other cell populations at later developmental stages.

### The progeny of *Foxl2*-expressing cells was classified into ten transcriptionally distinct cell populations along with ovary development

To characterize and reconstruct the differentiation of *Foxl2*-expressing cells as ovary development proceeds, we crossed *mT/mG* reporting mice with *Foxl2-CreER<sup>T2</sup>* mice and *Foxl2*-expressing cells were labeled by injecting Tamoxifen at E13.5. GFP-positive cells were isolated at E14.5, P0, P7, and P14 by Flow Cytometry. Single-cell RNA-sequencing experiment was performed using the 10× Chromium system. 23,534 GFP-positive cells passed stringent filtering (methods), with a median of ~2300 genes detected per cell (Supplementary Fig. S3). GFP-positive cells present in the developing ovaries were classified based on the highly variable genes, the time points of sample collection and the principle component analysis (Fig. 3a–c). We obtained ten cell clusters (C1 to C10) combining different developmental stages (Fig. 3b, c).

Partition-based graph abstraction and pseudotime reconstruction showed the relationship and the developmental process of these 10 cell clusters (Fig. 3a–c). Expression enrichment of known markers and differentially expressed genes (DEGs) allowed us to assign the identity of the cell clusters and enriched items of gene ontology (GO) biological process for corresponding genes, indicating the biological process of C1 to C10 (Fig. 3d).

### *Foxl2*-expressing progenitor cells were separated into two lineages during ovary development

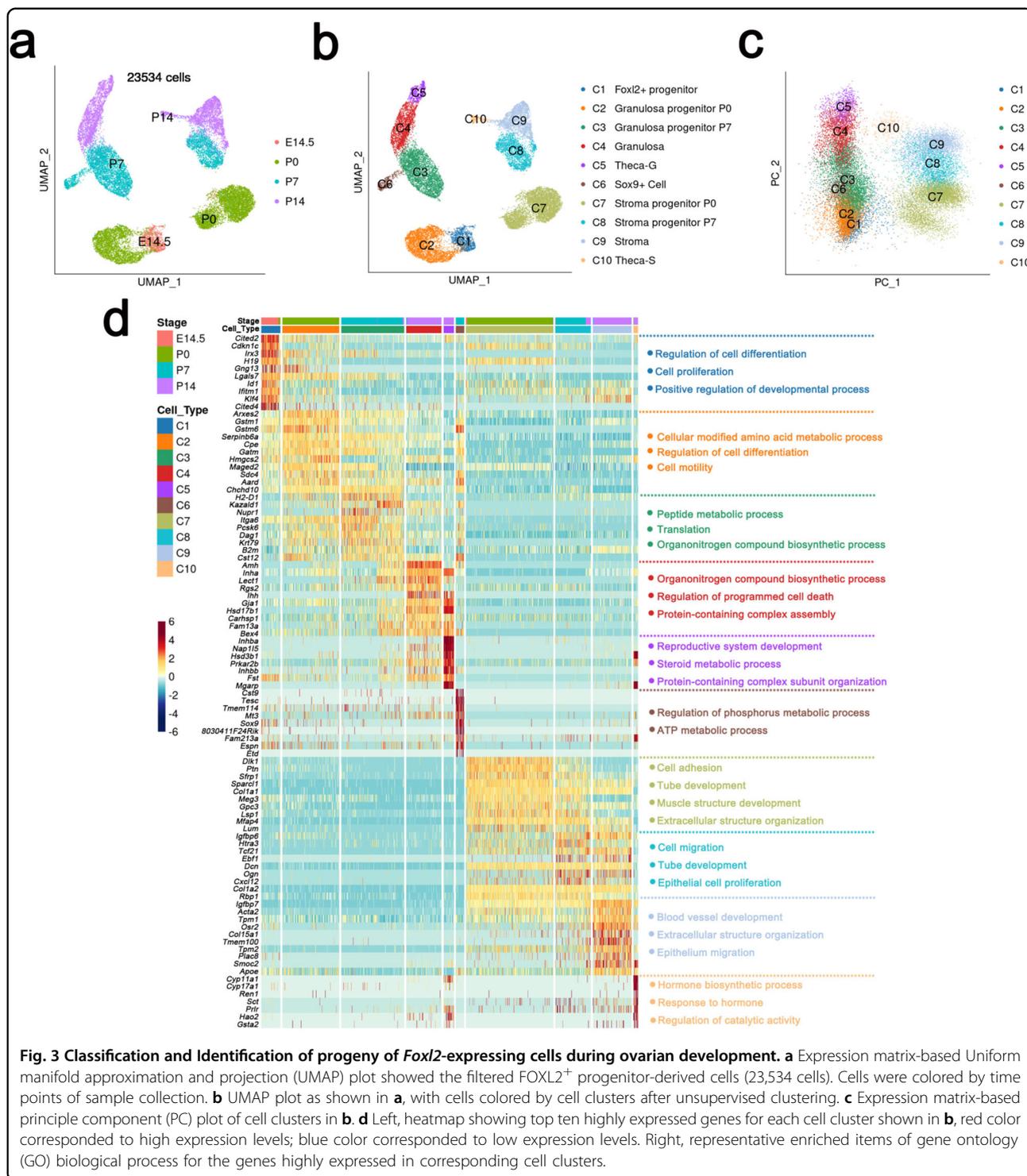
C1 represented the early progenitor cell population with cells mainly from E14.5 (Fig. 3a, b). The GFP-positive cells from the P0 ovary were classified into two groups, C2 and C7 (Fig. 3a, b). C3, C6, and C8 were from P7 ovaries (Fig. 3a, b). The GFP-positive cells from P14 ovaries were classified into four groups, C4, C5, C9, and C10 (Fig. 3a, b). Principle component analysis (PCA) result indicated that C1–C5 and C7–C10 belonged to two different branches (Fig. 3c). Partition-based graph abstraction (PAGA) plot showed the relationship and trajectories of those 10 cell clusters. C1, C2, C3, C4, and C5 were closely related, whereas C7,

C8, C9, and C10 were closely related (Fig. 4a). The pseudotime trajectories reconstructed by monocle3 further indicated the developmental process of cell clusters from C1 to C5 and C7 to C10.

Granulosa cell-specific genes such as *Amh*<sup>19</sup>, *Fst*<sup>20</sup>, *Inha*, and *Ihh*<sup>21</sup> were expressed in C4 (Figs. 3d, 5d and Supplementary Fig. S6a), indicating this cluster was granulosa cells. *Amh*, *Inha*, and *Fst* were also expressed in C2 and C3, but the expression levels were lower than those in C4 (Fig. 3d and Supplementary Fig. S6a). The results of pseudotime prediction and lineage reconstruction also showed that the cells in C4 were derived from C2 and C3. Based on these results, we speculated that C2, C3, and C4 belonged to the granulosa cell lineage, and C2 and C3 were the intermediate stages of pre-granulosa cells. *Arxes2*, *Gstm1*, *Gstm6*, *Serpina6a*, etc, were abundantly expressed in C2 and the top expressed genes in C3 were *H2-D1*, *Chchd10*, *Kazald1*, *Nupr1*, *Itga6*, etc (Fig. 3d and Supplementary Fig. S6a). However, the functions of these genes in granulosa cell development have not been reported previously.

GO analysis based on highly expressed genes in each cell cluster indicated that the early progenitor cells were mainly proliferating cells (C1, cell proliferation, and regulation of cell differentiation). The biological processes of “organonitrogen compound biosynthetic process”, “peptide metabolic process” and “regulation of cell differentiation” were mostly enriched in pre-granulosa cells (C2 and C3) and granulosa cells (C4) (Fig. 3d). C5 was closely related to C4 and both of them came from P14 ovaries (Fig. 3a–c). High level of steroidogenic enzyme genes, *Hsd3b1* and *Hsd17b1*, were expressed in this cluster (Fig. 3d and Supplementary Fig. S6a), indicating that C5 were theca-interstitial cells which were most likely derived from granulosa cell lineage (Theca-G). C6 was closely related to C3, but the cell number was much less. Surprisingly, Sertoli cell-specific gene *Sox9* was detected in this cluster (Fig. 3d). The results of immunohistochemistry also showed that a small number of cells in ovaries at P7 were Sox9-positive (Supplementary Fig. S4, inset). The exact functions of these cells were unclear and need further investigation.

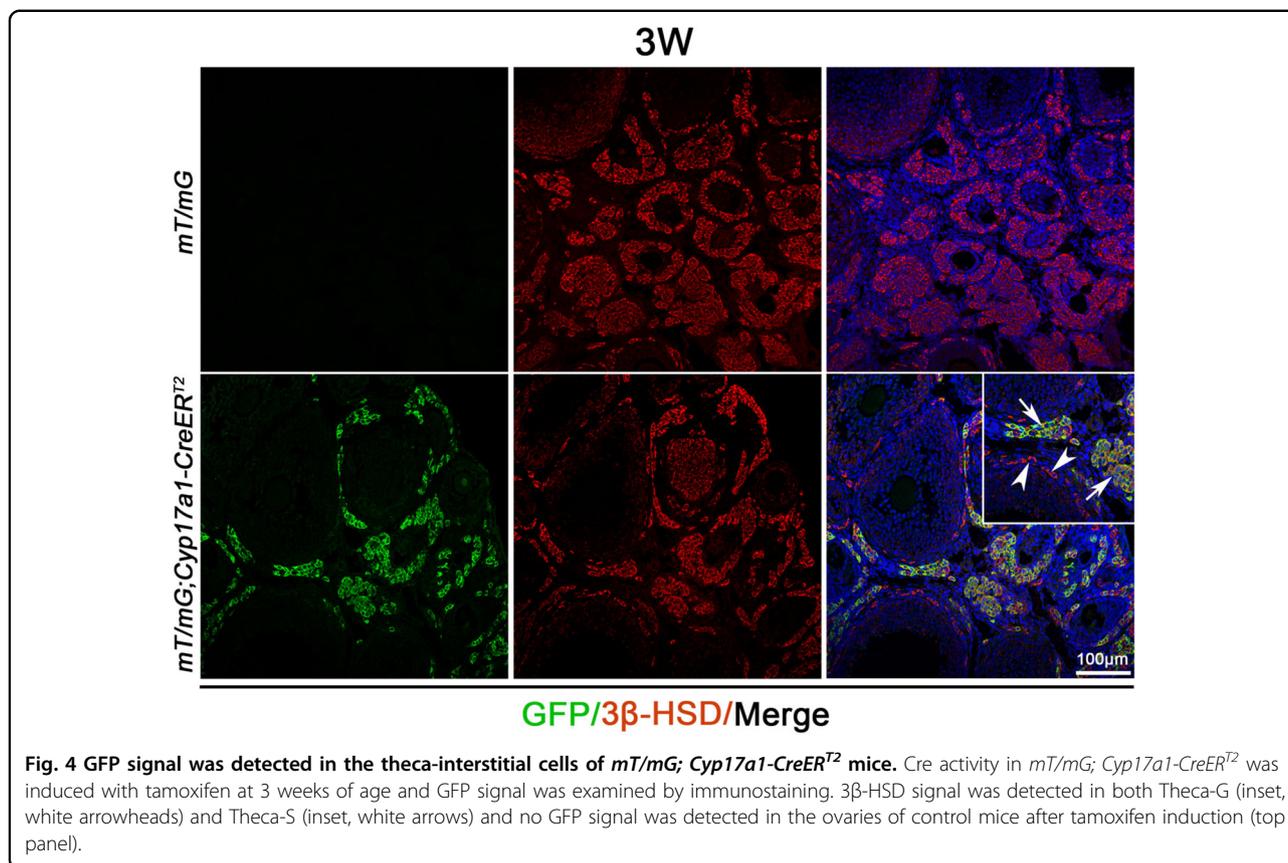
Stromal cell-associated genes *Ptn*<sup>22</sup>, *Sfrp1*<sup>23</sup>, *Col1a1*<sup>24</sup>, and *Mfap4*<sup>25</sup> were abundantly expressed in C7 (Fig. 6d and Supplementary Fig. S6a). The cells in C8 and C9 also expressed stromal cell-associated genes, such as *Col1a1* and *Mfap4* (Figs. 3d, 6d and Supplementary Fig. S6a). Thus, C7–C9 were annotated as stroma progenitors and stromal cells. The stroma progenitor-expressing genes related to the morphological organization (C7 and C8, “extracellular matrix organization”, “cell migration”, “cell adhesion”, and “tube development”) were enriched in these cell clusters. Cluster 10 was developmentally related



to Cluster 9 and also came from P14 ovaries. Interestingly, high levels of steroidogenic enzyme genes, *Hsd3b1* and *Cyp11a1*, were expressed in C10. Based on gene expression, cluster 10 was also annotated as theca-interstitial cells, which were most likely derived from stroma cell lineage (Theca-S).

**Theca-G and Theca-S were derived from different progenitor cells and gene expressing was different**

Based on gene expression, both C5 and C10 were theca-interstitial cells. However, C5 was derived from granulosa cells lineage (Theca-G), and still expressed granulosa-specific genes, such as *Fst* and *Ihh* (Supplementary Fig.



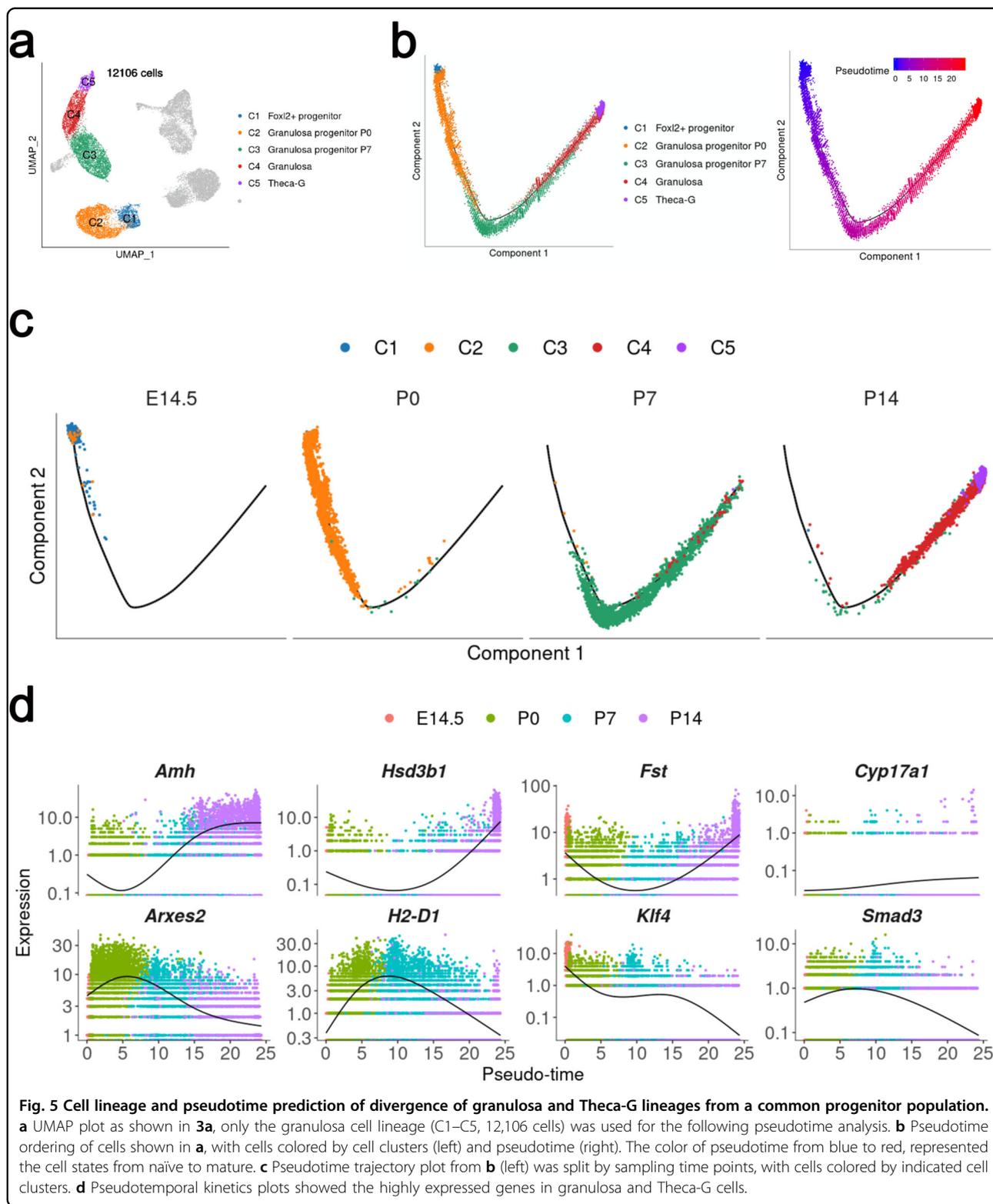
S5d). C10 was differentiated from stroma cell lineage (Theca-S), and still expressed stromal cell-associated genes, such as *Igfbp7* and *Smoc2* (Supplementary Fig. S5e). Meanwhile, *Cyp17a1* was exclusively expressed in C10, and *Inhba* was specifically expressed in C5 (Supplementary Fig. S5d, e). To further identify these two groups of steroidogenic cells, *mT/mG* reporting mice were crossed with *Cyp17a1-CreER<sup>T2</sup>* mice and Cre activity was induced by Tamoxifen induction at 3 weeks of age. As shown in Fig. 4, the GFP signal was detected in the 3β-HSD-positive theca-interstitial cells (Fig. 4, inset, white arrows), which were located in the interstitium, but not in the 3β-HSD-positive theca-interstitial cells (Fig. 4, inset, white arrowheads) which surrounded the ovarian follicles. These results suggested that 3β-HSD-positive theca-interstitial cells with different locations in the ovary were derived from different progenitor cells and the gene-expressing profiles were also different.

#### The development of granulosa cell and stromal cell lineage

To further confirm our results, pseudotime prediction was conducted. For granulosa lineage, C1–C5 cells were selected to construct pseudotime trajectory (Fig. 5a). Pseudotime trajectories showed the differentiation of granulosa and Theca-G cells (Fig. 5b, c). There was no

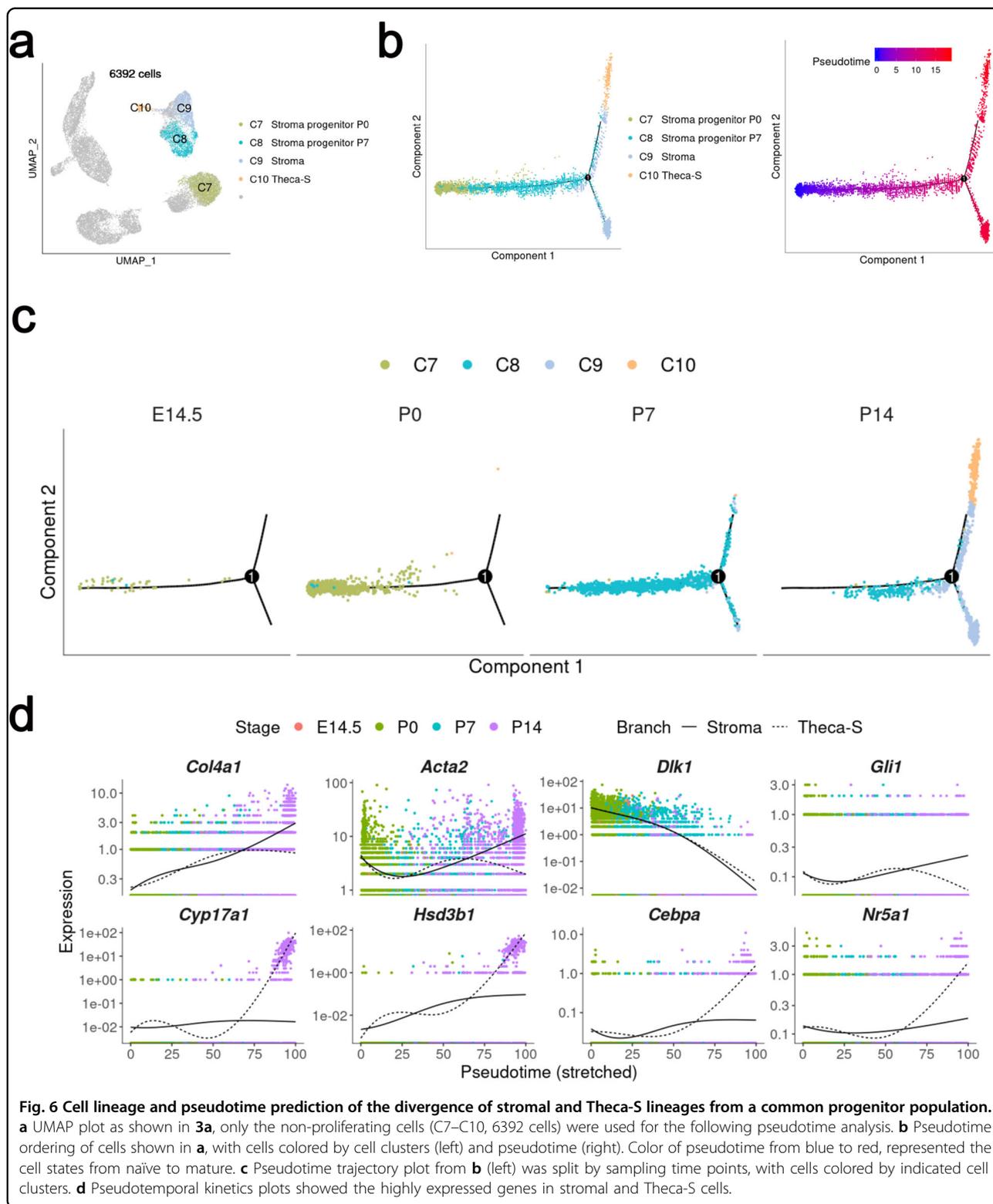
divergence in the pseudotime trajectories of the granulosa lineage, and C5 (Theca-G) appeared later than C4 (granulosa), indicating that Theca-G was differentiated from granulosa. Pseudotime profiles were scrutinized based on genes are known to be expressed in granulosa (*Amh*, *Fst*) and theca cells (*Hsd3b1*). The gene expression profile showed that *Amh* expression was increased from E14.5 to P14 in the granulosa cell lineage (Fig. 5d). Interestingly, we found that the expression of *Klf4* was significantly decreased from E14.5 to P14. The expression of *Arxcs2* and *H2-D1* was increased first, then decreased in this lineage (Fig. 5d), suggesting that these genes are probably also involved in the regulation of granulosa cell differentiation. Expression of genes specific to granulosa (C4) and Theca-G (C5) was skewed preferentially toward the beginning and end of the trajectories (Fig. 5d). In this process, *Smad3*, *Nfe2l1*, *Mef2d*, *Rreb1*, *Foxp4*, et al., were highly activated in C3 at P7, which indicated that these transcription factors could regulate the differentiation of granulosa and Theca-G (Supplementary Figs. S5f–i). *Tcf7* was essential to the cell fate maintenance of granulosa cell, and *Hif1a*, *Creb3l2*, *Zfp711*, and *Srebf1* play roles in the functions of Theca-G (Supplementary Figs. S5f–i).

Non-proliferating cells in C7–C10 were selected to construct pseudotime trajectory (Fig. 6a). Cells were



ordered from the beginning (blue) to the end (red) as they appeared by sampling time in the corresponding UMAP plots (Fig. 6b, c). The result indicated that stroma lineage

cells were divided into two routes during ovary development. One portion of stroma lineage cells developed into stromal cells and the others differentiated into Theca-S



(Fig. 6b, c). We selected cell cluster-enriched genes and analyzed their dynamic expression along the predicted pseudotime (Fig. 6d). The theca-specific genes (*Cyp17a1*,

*Nr5a1*, *Hsd3b1*) and the stromal cell-specific gene (*Col4a1*) were gradually increased from E14.5 to P14. In this process, transcription factors *Prrx2*, *Rarb*, *Gata2*, and

*Lhx8* were highly activated in cluster 9 and 10, suggesting that these genes might be involved in regulating the differentiation of stromal cells. *Cebpa*, *Nr5a1*, *Arntl*, *Six4*, *Crem*, and *Zfp931* were abundantly expressed in cluster 10, indicating that these genes might play important roles in Theca-S differentiation or lineage maintenance (Supplementary Fig. S5f–i).

To further explore the relationship among those transcription factors in differentiation or cell fate maintenance of stromal and granulosa lineages, we conducted the regulation network of key transcriptional regulators in Theca-G (Supplementary Fig. S6b) and Theca-S (Supplementary Fig. S6c). *Cebpa* and *Nr5a1* were the key regulators in both Theca-G and Theca-S. In Theca-G, *Cebpa* controlled the expression of *Txndc11*, while the key target of *Nr5a1* was *Nap115*. *Fasn* was the only common target of *Nr5a1* and *Cebpa* in Theca-G. In contrast, more common targets of *Nr5a1* and *Cebpa* (*Hsd3b1*, *Cyp17a1*, *Wnt6*, et al.) were identified in Theca-S.

## Discussion

Ovarian morphogenesis is a highly orchestrated process in which ovarian follicles are formed through intricate communication between germ cells and somatic cells. Defects in this process lead to reproductive disease and infertility in females<sup>26,27</sup>. *Foxl2* has been considered a granulosa-specific marker gene and the FOXL2-positive somatic cells in fetal ovaries will differentiate into granulosa cells in developing ovarian follicles. In this study, we found that FOXL2-positive cells in fetal gonads give rise to three different cell types, including granulosa cells, theca-interstitial cells, and ovarian stromal cells. Theca-interstitial cells are steroidogenic cells. They express enzymes for the biosynthesis of steroid hormones. However, the location and morphology of these cells are different. Theca cells are typical squamous epithelial cells and are separated from granulosa cells by a basal membrane<sup>27</sup>. In contrast, interstitial cells are stroma-like cells located in the mesenchyme between ovarian follicles.

Although the morphology and location of theca and interstitial cells are different, whether the origin and functions of these cells are different have not been investigated previously. In this study, we demonstrated that theca and interstitial cells are derived from different cell lineages. Both Theca and interstitial cells are derived from FOXL2-positive cells in fetal ovaries. However, Theca are closely related to granulosa cells lineage and directly differentiates from granulosa cells, so we called these cells Theca-G. By contrast, interstitial cells come from stromal cells, which are separated from granulosa cells lineage at an early developmental stage, then we called them Theca-S. Although our study demonstrated that steroidogenic cells in ovaries were derived from *Foxl2*-expressing cells, we could not exclude the

possibility that the cells from mesonephros and other sources also contribute to theca-interstitial cell lineage.

CYP17A1 is a cytochrome P450 enzyme with 17- $\alpha$ -hydroxylase and C17, 20-lyase activities, which converts pregnenolone and progesterone to dehydroepiandrosterone (DHEA), cortisol, testosterone, and estradiol<sup>28,29</sup>. The results of single-cell transcriptomics showed that the 3 $\beta$ -HSD-positive cells derived from granulosa cell lineage were negative for CYP17A1, whereas the steroidogenic cells derived from stromal cell lineage expressed CYP17A1. These results were further confirmed using *Cyp17a1-CreER<sup>T2</sup>* and *mT/mG* reporting mice. Our results showed that CYP17A1 was expressed in Theca-S, not in Theca-G, which was consistent with the previous study<sup>30</sup>. These results indicate that the gene expression profile is different between Theca-S and Theca-G. Whether the functions between these two cell types is also different is unclear, which needs further investigation.

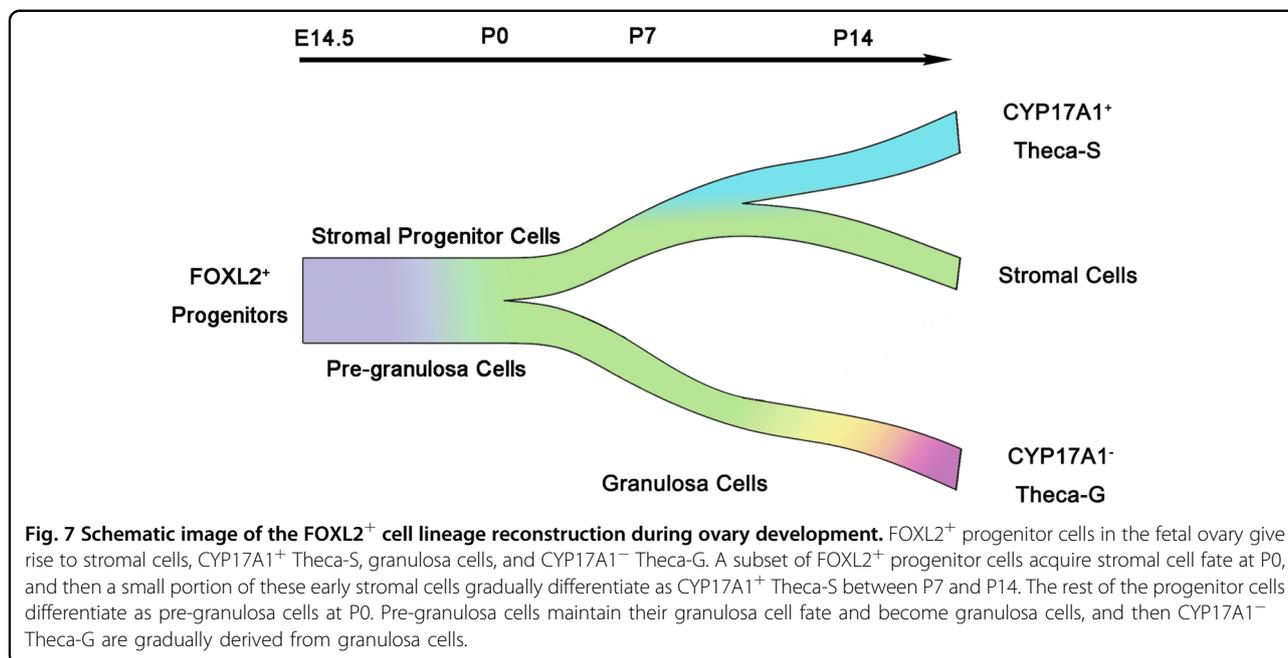
The reconstruction of the *Foxl2*-expressing cell lineages in the developing XX gonad allowed us to identify transition states leading to the differentiation of the granulosa cells, theca-interstitial cells, and the stromal cells from a common progenitor cell population. Several intermediate stages of granulosa cells were identified by single-cell transcriptomics analysis. *AMH* is a granulosa cell-specific gene and first expresses in the granulosa cells of primary follicles<sup>19</sup>. As expected, its expression was significantly increased in granulosa lineage from E14.5 to P14. Interestingly, we also found *Arxcs* and *Klf4* were gradually reduced from E14.5 to P14. The expression of *H2-D1* was increased first and then decreased along with granulosa cell differentiation, suggesting that these genes are also involved in the lineage specification of granulosa cells.

In this study, we demonstrated that *Foxl2*-expressing somatic cells in fetal ovaries gave rise to granulosa cells, theca-interstitial cells, and stromal cells. Theca-S and Theca-G are derived from different progenitor cells. CYP17A1-negative Theca-G cells were directly differentiated from granulosa cells, whereas CYP17A1-positive Theca-S were derived from ovarian stromal cells, which were separated from granulosa lineage at P0 (Fig. 7). We also found the gene expression between Theca-S and Theca-G was different. Our study provides an important resource for better understanding somatic cell differentiation in ovary development.

## Methods

### Animals

The mice were maintained in the specific pathogen-free animal facility of the Experimental Animal Center of the Institute of Zoology, Chinese Academy of Sciences, with a humidity of 35%  $\pm$  4%, a stable temperature of 24  $\pm$  1  $^{\circ}$ C, and 12/12 h light/dark cycle. All the animal experiments



were performed under the guidelines of the Animal Research Committee of the Institute of Zoology, Chinese Academy of Sciences. *mT/mG*; *Foxl2-Cre* female mice were obtained by crossing *mT/mG* mice<sup>18</sup> with *Foxl2-Cre* mice<sup>17</sup>. *mT/mG* reporting mouse is a double-fluorescent Cre reporter mouse, which expresses membrane-targeted tandem dimer Tomato (mT) prior to Cre-mediated excision and membrane-targeted green fluorescent protein (mG) after activation of cyclization recombination enzyme (Cre).<sup>18</sup> *mT/mG*; *Foxl2-CreER<sup>T2</sup>* mice were obtained by crossing *mT/mG* mice with *Foxl2-CreERT2* mice<sup>31</sup>. *mT/mG*; *Cyp17a1-CreER<sup>T2</sup>* mice were obtained by crossing *mT/mG* mice with *Cyp17a1-CreERT2* mice. The genotype was examined by PCR using DNA isolated from tail tips. The primer sequence used were as follows:

*Foxl2-Cre-Forward*, 5'-GAGAAGAGAGTGAGAGCCGC-3',

*Foxl2-Cre-Reverse*, 5'-GCCAGCAGGGTCCCCGCCGTGTCT-3',

*Foxl2-CreER<sup>T2</sup>-Forward*, 5'-TCCAATTTACTGACCG-TACACCAA-3',

*Foxl2-CreER<sup>T2</sup>-Reverse*, 5'-CCTGATCCTGGCAATTCGGCTA-3',

*Cyp17a1-CreER<sup>T2</sup>-Forward*, 5'-TGGCCCGGCAGGAGCTCTTTATCTTCAT-3'

*Cyp17a1-CreER<sup>T2</sup>-Reverse*, 5'-ACAGGGAGGGCAGGCAGGTTTTGGTG-3'

*mT/mG-Forward*, 5'-CTCTGCTGCCTCCTGGCTTCT-3',

*mT/mG-Reverse1*, 5'-CGAGGCGGATCACAAGCAATA-3',

*mT/mG-Reverse2*, 5'-TCAATGGGCGGGGGTCGTT-3'.

## Immunofluorescence

The ovaries were fixed in 4% paraformaldehyde, dehydrated in series ethanol, and embedded in paraffin. About 5- $\mu$ m sections were prepared by microtome, dewaxed in dimethylbenzene, and re-hydrated in series ethanol. Sections were blocked for 1 h at room temperature in 5% BSA with 0.3% TritonX-100. After blocking, sections were incubated with primary antibodies diluted in 1% BSA for 1 h at room temperature. Then, FITC or TRITC conjugated secondary antibodies were employed for 1 h at room temperature and DAPI were used to stain nuclei for 15 min. Primary antibodies used are listed in the following table. Immunolabelled sections were imaged with a confocal laser scanning microscope (Zeiss 880).

Antibody	Catalog	Dilution	Company
anti-3 $\beta$ -HSD	sc-30820	1:400	Santa Cruz
anti-SOX9	AB5535	1:200	Chemicon/Millipore
anti-WT1	ab89901	1:400	Abcam
anti-GFP	sc-9996	1:100	Santa Cruz
anti-FOXL2	ab5096	1:100	Abcam
anti-RFP	600-401-379 S	1:1000	Rockland

## Isolation of GFP-positive cells and single-cell RNA sequencing

The gonads from *mT/mG*; *Foxl2-CreER<sup>T2</sup>* female mice were collected at E14.5, P0, P7, and P14, and digested with trypsin and collagenase for 5–8 min at 37 °C to obtain single-cell suspension. The cell suspension was filtered

through a 40- $\mu$ m cell strainer to remove undigested cell clusters. GFP-positive cells were sorted by FACS. Single-cell libraries were constructed using the 10  $\times$  single-cell 3' Library & Gel Bead Kit v2 according to the manufacturer's instruction<sup>32</sup>. In short, cell counts were assessed with a hemocytometer (Luna-fl, Logos Biosystems) and the cell concentration was adjusted to about 600 cells/ $\mu$ L. All cells from E14.5 gonads were added to the channel of the 10 $\times$  chip. For other stages, 12,000 cells were added to each channel, and about 7000 cells were captured. Captured cells were lysed, and the released RNA was barcoded through reverse transcription. Then, the cDNA was amplified for the library construction, and the qualities of cDNA and cDNA libraries were assessed by Agilent 2100. Finally, libraries were sequenced on an Illumina HiSeq X Ten platform (CapitalBio, Beijing).

### 10 $\times$ Genomics data pre-processing

Reads processing was performed using the 10 $\times$  Genomics Cell Ranger (3.1.0) workflow<sup>32</sup> with default mapping arguments. Briefly, the cellranger count command line was used to assign barcodes, map reads, and quantify UMI for each sample. Reads were mapped to the mm10 genome (including eGFP sequence) and counted with GRCm38.84 annotation (including eGFP information). The cellranger aggr command line was executed to uniform the sequencing depth for all 10 $\times$  samples in this study, and the mean reads per cell were more than 30,000 post-normalization.

### Filtering of the scRNA-seq Data

For all sequenced single cells, we first filtered out the cells with a percentage of mitochondria genes of more than 10%. Then Scrublet python package was exploited to compute multiple cells for each sample<sup>33</sup>. All samples were used to create the Seurat object<sup>34</sup>, and only genes expressed in more than three cells were retained. After further excluding the doublet cells, low-level GFP expressing cells (including germ cells and macrophages), and cells with low quality (cells with fewer than 500 genes or the percentage of mitochondria genes less than 1%), 23,534 cells remained to do the following analysis.

### Clustering, DEG, and GO analysis

Single-cell data were analyzed with the Seurat v3 R package<sup>34</sup>. 23,534 cells were used to create the Seurat object, and only genes expressed in more than three cells were retained. Variable genes were computed with the "mean.var.plot" method. "Elbow Plot" was used to determine an appropriate number of dimensions to perform the nonlinear dimensional reduction (UMAP). Unsupervised clustering analysis was performed by the "Find Clusters" function to identify cell clusters. The Seurat "Find all Markers" function was used to compute DEGs between cell

clusters. The biological process aspect of GO analysis was performed using the clusterProfiler R package<sup>35</sup>.

### PAGA analysis

The Seurat object was converted to a loom file by the as.loom function of loomR package<sup>34</sup>, then the loom file was loaded to the python environment by the sc.read\_loom function of Scanpy python package<sup>36</sup>. After setting the neighbor argument, the relationship between different cell clusters was evaluated by the partition-based graph abstraction (PAGA) analysis with the Scanpy python package<sup>36</sup>.

### Pseudotime analysis

The Monocle2 R package<sup>37</sup> was used to compute the developmental pseudotime of granulosa and stromal lineage cells, respectively. After selecting the cells of the individual lineage, the UMI count data and metadata were exported from the Seurat3 object. The New Seurat2 object was built based on the UMI count data and metadata, and only the genes expressed in more than three cells were retained. The Seurat2 object was then converted to the Monocle2 object by the importCDS function of Monocle2<sup>37</sup>. Variable genes computed by the "mean.var.plot" method of seurat package were used to do the unsupervised ordering of the cells with default settings of Monocle2 R package<sup>37</sup>.

### SCENIC analysis

SCENIC analysis was carried out following the SCENIC command line protocol<sup>38</sup>. SCENIC command line version was used to do gene regulatory network inference, regulon prediction, and cellular enrichment (Area Under the Curve, AUC) processes with the filtered 23,534 GFP<sup>+</sup> cells. SCENIC UMAP was computed based on the AUC matrix. Regulon specificity scores (RSS) were computed based on the cell clusters identified by Seurat. We chose the top regulons for each cell cluster. Finally, the SCENIC AUC heatmap was plotted with reported regulons and top regulons of each cell cluster by the pheatmap R package.

### Acknowledgements

We thank Mr. Xue Liu and Mr. Hongyang Ma for the technique help. This work was supported by the National Key R&D Program of China (2018YFA0107702), the National Natural Science Foundation of China (31970785 and 31671496), and the Jining Medical College High-level Talents Start-up Fund.

### Author details

<sup>1</sup>State Key Laboratory of Stem cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. <sup>2</sup>Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, China. <sup>3</sup>University of Chinese Academy of Sciences, Beijing, China. <sup>4</sup>The Collaborative Innovation Center, Jining Medical University, Jining, Shandong, China. <sup>5</sup>NHC Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Anhui Medical University, Hefei, China. <sup>6</sup>Department of Hematological Malignancies Translational Science, Hematologic Malignancies and Stem Cell Transplantation Institute, Beckman Research Institute, City of Hope National Medical Center, Duarte, CA, USA. <sup>7</sup>Guangdong and Shenzhen Key Laboratory

of Male Reproductive Medicine and Genetics, Institute of Urology, Peking University Shenzhen Hospital, Shenzhen Peking University-The Hong Kong University of Science and Technology Medical Center, Shenzhen, China. <sup>6</sup>State Key Laboratory of Reproductive Regulation & Breeding of Grassland Livestock, Inner Mongolia University, Hohhot, China

#### Author contributions

F.G., H. Wang, and X.G. conceived and supervised this study; J.Z. and H. Wu performed tissue collection, immunofluorescence, and RNA extraction with the help of Min Chen<sup>1</sup>, Min Chen<sup>1,2,6</sup>, Z.S., and L.Z.; X.J. analyzed the bioinformatic data; J.Z., X.J., and F.G. wrote the original draft; J.Z., X.J., H. Wu, L.Z., Min Chen<sup>1</sup>, Min Chen<sup>1,2,6</sup>, Z.S., X.G., H. Wang, and F.G. reviewed and edited the manuscript; Funding acquisition, F.G.; Resources, F.G.

#### Data availability

All data generated or analyzed during this study are included in this article and its supplementary information files. The data reported in this paper were available at the NCBI Gene Expression Omnibus (GEO) with the accession GSE163879. The softwares used in this paper were Seurat, Monocle2, clusterProfiler, pheatmap R packages, and Scrublet, Scanpy, and SCENIC python packages.

#### Conflict of interest

The authors declare no competing interests.

#### Publisher's note

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

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41421-022-00492-1>.

Received: 20 December 2021 Accepted: 25 October 2022

Published online: 27 December 2022

#### References

- Byskov, A. G. Differentiation of mammalian embryonic gonad. *Physiol. Rev.* **66**, 71–117 (1986).
- Capel, B., Albrecht, K. H., Washburn, L. L. & Eicher, E. M. Migration of mesonephric cells into the mammalian gonad depends on *Sry*. *Mech. Dev.* **84**, 127–131 (1999).
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351**, 117–121 (1991).
- Gubbay, J. et al. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245–250 (1990).
- Koopman, P., Bullejos, M. & Bowles, J. Regulation of male sexual development by *Sry* and *Sox9*. *J. Exp. Zool.* **290**, 463–474 (2001).
- Bernard, P. & Harley, V. R. *Wnt4* action in gonadal development and sex determination. *Int. J. Biochem. Cell Biol.* **39**, 31–43 (2007).
- Chassot, A. A., Gillot, I. & Chaboissier, M. C. *R-spondin1*, *WNT4*, and the *CTNNB1* signaling pathway: strict control over ovarian differentiation. *Reproduction* **148**, R97–R110 (2014).
- Tomizuka, K. et al. *R-spondin1* plays an essential role in ovarian development through positively regulating *Wnt-4* signaling. *Hum. Mol. Genet.* **17**, 1278–1291 (2008).
- Hannenhalli, S. & Kaestner, K. H. The evolution of *Fox* genes and their role in development and disease. *Nat. Rev. Genet.* **10**, 233–240 (2009).
- De Baere, E. et al. Spectrum of *FOXL2* gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype-phenotype correlation. *Hum. Mol. Genet.* **10**, 1591–1600 (2001).
- Schmidt, D. et al. The murine winged-helix transcription factor *Foxl2* is required for granulosa cell differentiation and ovary maintenance. *Development* **131**, 933–942 (2004).
- Uda, M. et al. *Foxl2* disruption causes mouse ovarian failure by pervasive blockage of follicle development. *Hum. Mol. Genet.* **13**, 1171–1181 (2004).
- Barsoum, I. B. & Yao, H. H. Fetal Leydig cells: progenitor cell maintenance and differentiation. *J. Androl.* **31**, 11–15 (2010).
- DeFalco, T., Takahashi, S. & Capel, B. Two distinct origins for Leydig cell progenitors in the fetal testis. *Dev. Biol.* **352**, 14–26 (2011).
- Liu, C., Peng, J., Matzuk, M. M. & Yao, H. H. Lineage specification of ovarian theca cells requires multicellular interactions via oocyte and granulosa cells. *Nat. Commun.* **6**, 6934 (2015).
- Chen, M. et al. *Wt1* directs the lineage specification of sertoli and granulosa cells by repressing *Sf1* expression. *Development* **144**, 44–53 (2017).
- Cen, C. et al. Inactivation of *Wt1* causes pre-granulosa cell to steroidogenic cell transformation and defect of ovary development. *Biol. Reprod.* **103**, 60–69 (2020).
- Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
- Mullen, R. D., Ontiveros, A. E., Moses, M. M. & Behringer, R. R. *AMH* and *AMHR2* mutations: a spectrum of reproductive phenotypes across vertebrate species. *Dev. Biol.* **455**, 1–9 (2019).
- Liu, C. F., Liu, C. & Yao, H. H. Building pathways for ovary organogenesis in the mouse embryo. *Curr. Top. Dev. Biol.* **90**, 263–290 (2010).
- Liu, C., Rodriguez, K. F., Brown, P. R. & Yao, H. H. Reproductive, physiological, and molecular outcomes in female mice deficient in *Dhh* and *lh*. *Endocrinology* **159**, 2563–2575 (2018).
- Yu, H. F. et al. *Ptn* functions downstream of *C/EBPβ* to mediate the effects of cAMP on uterine stromal cell differentiation through targeting *Hand2* in response to progesterone. *J. Cell Physiol.* **233**, 1612–1626 (2018).
- Jabara, S. et al. Stromal cells of the human postmenopausal ovary display a distinctive biochemical and molecular phenotype. *J. Clin. Endocrinol. Metab.* **88**, 484–492 (2003).
- Hatzirodos, N. et al. Transcript abundance of stromal and thecal cell related genes during bovine ovarian development. *PLoS ONE* **14**, e0213575 (2019).
- Knight, J. M. et al. Comprehensive site-specific whole genome profiling of stromal and epithelial colonic gene signatures in human sigmoid colon and rectal tissue. *Physiol. Genomics* **48**, 651–659 (2016).
- Matzuk, M. M. & Lamb, D. J. The biology of infertility: research advances and clinical challenges. *Nat. Med.* **14**, 1197–1213 (2008).
- Edson, M. A., Nagaraja, A. K. & Matzuk, M. M. The mammalian ovary from genesis to revelation. *Endocr. Rev.* **30**, 624–712 (2009).
- Nakajin, S., Hall, P. F. & Onoda, M. Testicular microsomal cytochrome P-450 for C21 steroid side chain cleavage. Spectral and binding studies. *J. Biol. Chem.* **256**, 6134–6139 (1981).
- Nakajin, S., Shively, J. E., Yuan, P. M. & Hall, P. F. Microsomal cytochrome P-450 from neonatal pig testis: two enzymatic activities (17 alpha-hydroxylase and c17,20-lyase) associated with one protein. *Biochemistry* **20**, 4037–4042 (1981).
- Nicol, B. et al. *RUNX1* maintains the identity of the fetal ovary through an interplay with *FOXL2*. *Nat. Commun.* **10**, 5116 (2019).
- Zheng, W. et al. Two classes of ovarian primordial follicles exhibit distinct developmental dynamics and physiological functions. *Hum. Mol. Genet.* **23**, 920–928 (2014).
- Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
- Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 e289 (2019).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 e1821 (2019).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
- Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
- Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
- Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).