1 Transcriptome Landscape of Human Oocytes and Granulosa Cells Throughout 2 Folliculogenesis

3 Yaoyao Zhang^{1,2,3,8}, Zhiqiang Yan^{1,5,8}, Qingyuan Qin^{1,3,4,8}, Vicki Nisenblat^{1,4}, Yang
4 Yu^{1,2,3}, Tianren Wang^{1,2,3}, Cuiling Lu^{1,2,3}, Ming Yang^{1,2,3}, Shuo Yang^{1,2}, Ying Yao⁵,
5 Xiaohui Zhu^{1,3,4}, Xi Xia^{1,2,3}, Yujiao Dang^{1,3,4}, Yixin Ren^{1,3,4}, Peng Yuan^{1,3,4}, Rong
6 Li^{1,2,3}, Ping Liu^{1,2,3}, Hongyan Guo⁵, Jinsong Han⁵, Haojie He⁵, Yu Wu⁵, Meng Li⁵,
7 Kun Zhang⁵, Yiting Wang⁵, Jie Qiao^{1,2,3,5,6,9*}, Jie Yan^{1,2,3*}, Liying Yan^{1,2,3*}

8 ¹ Center for Reproductive Medicine, Department of Obstetrics and Gynecology,
9 Peking University Third Hospital, No.49 North HuaYuan Road, HaiDian District,
10 Beijing 100191, China;

11 ² National Clinical Center for Obstetrics and Gynecology, Beijing 100191, China;

12 ³ Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing 100191,
13 China;

14 ⁴ Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproduction,
15 Beijing 100191, China;

16 ⁵ Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871,
17 China

18 ⁶ Department of Obstetrics and Gynecology, Peking University Third Hospital,
19 No.49 North HuaYuan Road, HaiDian District, Beijing 100191, China;

20 ⁷ Beijing Advanced Innovation Center for Genomics (ICG), Peking University,
21 Beijing 100871, China;

22 ⁸ Co-first author

23 ⁹ Lead Contact

24 * Corresponding: jie.qiao@263.net (J.Q.), yanjiebjmu@bjmu.edu.cn (J.Y.),
25 yanliyingkind@aliyun.com (L.Y.).

26

SUMMARY

Folliculogenesis is a highly regulated process that involves bidirectional interactions of the oocytes and surrounding granulosa cells (GCs). Little is unknown, however, about the transcriptomic profiles of human oocytes and GCs throughout folliculogenesis. Here we performed a high resolution RNA-Seq of human oocytes and GCs at each follicular stage, which revealed unique transcriptional profiles, stage-specific signature genes, oocyte- and GC-derived genes that reflect ovarian reserve. We identified reciprocal cell-to-cell interactions between oocytes and GCs, including NOTCH, TGF- β signaling and gap junctions and determined the expression patterns of maternal-effect genes involved in folliculogenesis and early embryogenesis. Finally, we demonstrated robust differences between human and mice oocyte transcriptomes. This is the first comprehensive overview of the transcriptomic signatures governing the stepwise human folliculogenesis in-vivo that provides a valuable resource for basic and translational research in human reproductive biology.

INTRODUCTION

Human folliculogenesis is a remarkably complex, well-orchestrated process that relies on a synchronization between oocyte maturation and proliferation of the neighboring granulosa cells (GCs)^{1,2}. Follicle growth and oocyte maturation are associated with dynamic transcriptional events in both oocyte and GC compartments of the follicle, featured by high transcriptional activity of growing follicles post recruitment and transcriptional silencing of mature oocytes³. Despite the impressive body of data produced in recent years on oocyte biology, many questions regarding the key differentiation events occurring during folliculogenesis in humans remain unanswered. Oocyte-GCs bidirectional communications via signal transduction or direct cell-to-cell contact provide molecular and structural basis for effective oocyte-GC crosstalk required for adequate follicular development. It is poorly understood, however, what initiates⁴ the communication between human follicle compartments and how this interaction is regulated and maintained.

Genes that accumulate in oocytes during maturation and influence early embryonic development are referred to as maternal-effect genes⁵. The events associated with maternal-effect genes and the process of maternal-to-zygotic transition have been studied in animal models⁶, albeit the expression pattern and the role of these genes during oocyte development in humans remain unclear. In women, reproductive potential and reproductive lifespan depend on the follicle number available for fertilization and the rate of follicular loss during reproductive years, both of which

determine ovarian reserve ⁷. Even though the levels of tests such as anti-Müllerian hormone and follicle stimulating hormone (FSH), and the antral follicle count measured by transvaginal ultrasound has been widely used to predict ovarian reserve, the universally accepted markers that accurately predict fertility potential and ovarian reserve are still lacking ⁸⁻¹⁰.

Till now, most data on follicular transcriptome are derived from the studies on animal models due to restricted availability of human follicles for research. While animal studies have advanced our understanding of ovarian biology, expression patterns of oocyte-specific genes are known to display considerable interspecies variations ¹¹. In humans, transcriptomic studies of ovarian follicle rely on evaluations of the oocytes at limited stages of development or on analysis of unseparated follicle compartments ¹²⁻¹⁴, which undermines the applicability of the findings to understanding of a coherent whole of folliculogenesis. Application of single-cell RNA sequencing (scRNA-Seq) and other high resolution sequencing techniques allows an unbiased evaluation of cells of any size and efficient amplification of low-abundant transcripts, which makes these methods a powerful tool in studying the biological process of interest¹⁵. Therefore, characterizing human oocyte and GC transcriptome during different stages of development at high resolution is key to understanding the events that coordinate human oocyte maturation and follicle growth, which in turn is expected to provide remarkable opportunities for developing novel diagnostic and therapeutic approaches to improve fertility.

This work aimed at analyzing the gene expression patterns throughout folliculogenesis by exploring the transcriptome of human oocytes and GCs at five key stages of follicular development in-vivo, from primordial to preovulatory stage. Using scRNA-Seq approach, we identified dynamic expression patterns with distinct transcriptome signatures in oocytes and GCs, recognized the oocyte-GC communications during folliculogenesis and revealed putative biomarkers to predict ovarian reserve. In addition, we investigated the expression patterns of maternal-effect genes in human oocytes and propose their role in follicular development. Finally, we compared our data with the published oocyte transcriptome profiles from rodents and demonstrated robust interspecies differences.

RESULTS

Global Transcriptome Profiling of Human Oocytes and GCs

A total of 83 oocytes and 92 GC samples (each GC sample comprised randomly selected 10 cells per follicle owing to low abundancy of RNA in these cells) were obtained (Figure 1A). The data for MII oocytes were downloaded from our previous research¹³ and merged with the newly generated data for the subsequent analyses. A down sampling analysis revealed adequate sequencing depth for gene expression detection (Figure S1D). After quality control and filtration of the RNA-Seq data, 80 oocytes and 71 GC samples were retained for the analysis (Figures S1). The unsupervised principal component analysis (PCA) showed there was a clear separation between 80 oocytes and 71 GCs with distinction between the developmental stages (Figure 1B). Next we analyzed the expression patterns of the known cell-type markers in the oocyte and GC clusters (Figure 1C and Figure 1D), which confirms the validity of PCA classification. Moreover, we selected the most variable genes that contributed to PC1 and could distinguish between the oocytes and GCs. In oocytes, the most variable genes included *NOL4*, *CTCTL*, *ITIH2*, *SNAP91* and *NAALAD2*. In GCs, the most variable genes were: *ZEB2*, *CD44*, *HSPG2*, *KDSR* and *SRRM3*. The top oocyte or GCs-expressed variable genes are presented in Figures 1E and 1F, respectively. The above observations suggest that the most variable genes between oocytes and GCs could be candidate cell type-specific markers.

Gene Expression Dynamics and Transcriptional Characteristics of Oocytes during Folliculogenesis

To explore the gene expression patterns in oocytes during folliculogenesis, we performed PCA analysis within the cohort of oocytes and observed five distinct subpopulations which corresponded morphological classification of follicular development (Figure 2A). The stage of follicular development was much more powerful discriminator than the sampling entity (Figure S1E), indicating that transcriptome profiles of oocytes reflect the physiological status of maturation, rather than genetic background. We then characterized the differentially expressed genes (DEGs) between the stages of follicle development in oocytes ($FDR < 0.05$; FC of $\log_2$ transformed FPKM > 1.5) (Figure 2B). A large number of these genes were under-expressed in primordial follicles with progressively increasing number of activated transcripts with follicular growth and maturation. The expression profiles of DEGs *RBM24*, *GPD1*, *NTF4* and *LCP2* were validated by using immunohistochemistry (IHC) on ovarian tissue from the scRNA-Seq sample set. Results showed that *RBM24* was expressed only in primordial and primary stage, *GPD1* was restricted secondary and antral stage, while *NTF4* and *LCP2* were specifically expressed in antral stage. These results are in agreement with the scRNA-Seq results (Figure 2E). To investigate possible biological functions of DEGs involved in folliculogenesis, we performed Gene Ontology (GO) analysis of DEGs in oocytes and revealed enriched biological processes in association with each stage of follicular development ($p < 0.05$) (Figure 2C).

We also investigated a number of germ cell markers and revealed some exhibited stable expression levels in oocytes during follicle development (e.g. *DDX4* and *ZP3*, Figure S2A), while others showed progressive upregulation with oocyte maturation (e.g. *ZP1*, *GDF9* and *HIFOO*). The expression of *ZP2*, *ZP3* and *ZP4* remained stable across all stages. Our findings suggest that *ZP1* might play a role in acquisition of oocyte competence and *ZP2*, *ZP3* and *ZP4* may act across all developmental stages in human oocytes. In contrast with our previous findings in fetal ovary ^{16,17}, this study showed that the expression levels of pluripotency markers *NANOG* and *POU5F1* were constantly low during oocyte development, suggesting their different roles in fetal and adult ovary.

Next, we focused on investigating the expression patterns of meiosis-related genes in oocytes. The sets of 195 genes involved in meiosis I and 13 genes involved in meiosis II were gathered from the GO Consortium. Of these, 52 genes were differentially expressed across different stages of oocyte maturation. Most of the meiosis-specific genes demonstrated shift towards upregulation as maturation of oocytes proceeded, and exhibited strong over-expression in antral and preovulatory follicles (Figure 2D).

There is an ongoing lack of clarity regarding the levels of DNA methylation that determine change in DNA activity during human folliculogenesis. Here we demonstrated high expression levels of *DNMT1*, *DNMT3A* and *DNMT3B* in oocytes at all stages of folliculogenesis with mounting abundance (Figure S2A), which implies that DNA methylation level might be continually elevated with oocyte maturation. In

contrast to our previous study in PGCs¹⁶, our results showed that ten-eleven translocation (TET) family genes were under-expressed in oocytes at all stages of development (Figure S2A). Together, overexpression of DNMT genes with loss of TET activity indicates an active methylation transition in human oocyte during folliculogenesis.

Dynamic Expression and Transcriptional profiles of GCs during Folliculogenesis

In GCs, PCA of all samples showed clustering into five groups according to stage of follicular development (Figure 3A). An observed overlap between the clusters of primary and secondary follicles indicated less pronounced difference at the transcriptome level between these two stages. We then examined DEGs (FDR < 0.05; FC of log₂ transformed FPKM > 1.5) in GCs (Figure 3B). Interestingly, some of the DEGs have been previously reported in association with oocyte competence, including quality, maturity or fertilization rate, embryo quality and pregnancy outcomes^{18,19} as detailed in Table S2. GO analysis of DEGs (Figure 3C) revealed the crucial functional roles of these genes during folliculogenesis. The expression profiles of the DEGs *CDCA3*, *BNIP1* and *TST* were validated by using immunohistochemistry (IHC) on ovarian tissue from the scRNA-Seq sample set (Figure 3E), which are in concordance with the scRNA-Seq results.

Analysis of the cell-cycle genes involved in the G1/M and G2/S phases²⁰ in GCs Revealed that at primordial stage, GCs were found to be relatively quiescent and maintained low proliferative activity (Figure 3D). Both G1/S and G2/M-specific genes were up-regulated in the GCs from primary, secondary and antral follicles. The

cell-cycle genes were predominantly abundantly expressed in the GCs from antral follicles, which designates their high proliferative activity at this stage. In contrast, the cell-cycle genes were down-regulated in the GCs from preovulatory follicles, which implies decreased degree of proliferation and differentiation in GCs at this stage.

Next, we focused on genes that have been previously associated with steroidogenesis (Figure S2B). *CYP11A1*, *CYP19A1*, *HSD3B1* and *HSD17B2*, encoding for steroidogenic enzymes, showed up-regulation in the antral follicles and reached peak expression levels in the preovulatory follicles. Their expression patterns indicate progressive increase in steroidogenic activity in maturing follicle, which culminates before ovulation. *NR5A1*, a key regulator of steroidogenic enzyme-encoding genes that has been previously described in sheep GCs ²¹, exhibited up-regulation with the progression of follicle growth. Both *ESR1* and *AR*, encoding for estradiol (E2) and androgen hormone receptors, respectively, were expressed in GCs at the primary follicle stage and reached highest expression level in antral follicles. This suggests E2 and androgens might control GC growth from the primordial follicle stage through these receptors.

Stage-Specific Signature Genes Identified in Oocytes and GCs

We identified a subset of stage-specific oocyte genes that were exclusively pertained to a single stage of folliculogenesis, which we referred to as “signature genes”. Conjointly, 382 signature genes were identified comprising the genes expressed in the oocytes from each stage (Figure S3A, Table S3). Distinct stage-specific expression patterns of *NTF4* and *LCP2* in oocytes were further confirmed by IHC staining (Figure 2E), which

showed that the expressions of these genes were restricted in antral stage. Of note, *NTF4*, proposed to facilitate follicle development in mouse by inducing FSH receptor (*FSHR*)²², was exclusively expressed in oocytes from antral follicles (Figure 2E). We found that *FSHR* gene expression was pertained to GCs in antral follicles and was concordant to that of *NTF4* (Figure S3D). Our data suggested that *NTF4* might upregulates *FSHR* expression in human GCs and contributes to oocyte-to-GC crosstalk, thus plays an important role in oocyte development. For GCs, we also identified a subset of stage-specific signature genes during folliculogenesis (Figures S3C and S3D, Table S3). The expressions of signature genes *CDCA3* and *BNIP1* were validated by IHC (Figure 3E). Interestingly, *FSHR*, previously observed in somatic cells from early preantral to mature follicles in sheep²³, was exclusively expressed in the GCs from antral follicles in our study (Figure S3D), suggesting considerable inter-species differences. The full list of the signature genes in oocytes and GCs is presented in Table S3. These identified signature genes (Figure S3 and Table S3) could serve as candidate cell-specific markers for each follicle stage, allowing to establish the objective criteria for selecting competent oocytes in vitro.

Secretory Protein Coding Signature Genes

Certain specific oocyte- or neighboring GCs-derived genes could reflect the quantity and the quality of the follicles. Consequently, the peripherally secreted products of these genes may reveal prognostic information concerning ovarian reserve. Therefore, we searched for the secretory protein-coding signature genes that were exclusively pertained to either oocytes or GCs and were expressed either at early or at more

advanced stages of follicle development (as described in Methods). In total, we identified 51 protein-coding genes in oocytes and 17 in GCs that clustered into five groups (Figure S4). Cluster 1 comprised 16 oocyte genes with specific expression in the oocytes of primordial and primary follicles, suggesting they may reflect the primordial follicle pool. Cluster 2 included 19 oocyte genes with specific expression in the oocytes from secondary, antral and preovulatory follicles. Cluster 3 consisted of 27 oocyte genes that were represented across folliculogenesis. Cluster 4 (n = 4) and cluster 5 (n = 13) included GC-derived genes with specific expression at preovulatory and secondary-antral stages, respectively. The expression patterns of the genes from each cluster are exemplified in Figure S4B. These findings represent a potentially valuable source of candidate molecular biomarkers of ovarian reserve. Further evaluation of their expressions in biological fluids could contribute to development of ovarian reserve test.

Transcription Factors Regulatory Networks in Oocytes and GCs

To find the master regulators and construct the transcriptional regulatory network along the steps of human folliculogenesis, we utilized the ARACNe method to analyze all 1,469 known TFs from Animal Transcription Factor Database (TFDB v2.0)^{24,25}. In the oocytes, the expressions of *GTF2I*, *CSDE1*, *SOHLH2*, *SMARCE1*, *TUB*, *HBPI*, *SOX30* and *HIF1A* were up-regulated in primary follicles, indicating that these TFs may play a critical role in the transition from primordial to primary stage (Figure 4). *KLF2*, *YBX2*, *FOXO6*, *SOX13*, *ETV5*, *TEAD2* and *OTX2* were over-expressed in the oocytes from secondary follicles compared to those from primary follicles, implying

they are likely the regulators of primary-to-secondary stage transition. *PINX1*, *PBX1*, *MTF1*, *SOX15*, *UBTF*, *SOX13*, and *POU2F1* had higher expressions in the oocytes from antral compared to secondary follicles, indicating their crucial regulatory roles in cytoplasmic and nuclear maturation of oocyte during antral stage. *ATF2* and *EOMES* were abundant in MII oocytes of preovulatory follicles, which indicates their potential role as TFs that possibly initiate the unique transcription networks involved in meiosis progression. Of note, we found that *SOX13* and *SOX15*, members of the SOX family of TFs, were up-regulated in the oocytes from secondary to antral follicles. While *SOX30* was found abundantly expressed in the oocytes during transition from primordial to primary follicle stages. This indicates their potential role in regulating transcription networks during primordial follicle activation (*SOX30*) and antral formation (*SOX13*, *SOX15*), respectively.

In the GCs, we found that *CREB1*, *NFKB1*, *MEF2A*, *PIAS1*, *FOSL2*, *KLF13* and *PRDM4* potentially regulated activation from the primordial to primary follicle stage, whereas *CRX*, *HES2*, *ZNF554*, *ZKSCAN3*, *LMX1B*, *FOXK1* and *SIX4* were the top candidate TFs possibly involved in transition from the primary to secondary follicle stage. *MBD1*, *FIZ1*, *GABPA*, *TGIF2*, *PIAS3*, *E4F1* and *IRF3* were likely driving initiation of the transcription network in antral follicles. *MEIS3*, *PRDM15* and *VTN* expressed in GCs of preovulatory follicles, could be the key regulators of cumulus cells progression in preovulatory follicles (Figure S5). Further evaluation of these TFs, their upstream signaling pathways, ligands, receptors and downstream targets in oocytes and GCs will provide insight into the transcriptional control of human folliculogenesis.

Key Signaling Pathways in Oocytes and GCs during Folliculogenesis

The Gene Set Enrichment Analysis (GSEA) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were applied to perform the pairwise comparisons of the follicular stages associated with each transition. There were 43 pathways overrepresented in the oocytes and 79 pathways enriched in the GCs from the primary follicles versus the primordial follicles ($p < 0.05$) (Table S4). Of these, the functional pathways that were significantly overrepresented both in oocytes and GCs ($p < 0.05$) and described in a context of folliculogenesis, included mTOR, GnRH, Neurotrophin and Insulin signaling (Figures S6A and S6B). These pathways probably mediate primordial-primary follicle transition and indicate the concerted action of the two follicular compartments. We identified 21 pathways enriched in the oocytes and 60 pathways in GCs that were overrepresented in antral, but not in secondary follicles ($p < 0.05$) (Table S4). In GCs, the steroid hormone biosynthesis was among the enriched pathways at the secondary-to-antral follicle transition, which indicates GCs play a key role in steroidogenesis at antral stage²⁶. In addition, mTOR, Insulin and Neurotrophin pathways were over-represented in both primordial-to-primary and secondary-to-antral stages, suggesting a possibility for their involvement in these two transitions (Figures S6B, S6C and S6D). Interestingly, the enrichment of Neurotrophin pathway was concordant with the expression of its ligand *NTF4*, which was identified in this study as an antral stage-specific signature gene in oocyte (Figure 2E).

Bidirectional Interactions between the Oocyte and GC Compartments

To investigate interactions between oocyte and GCs, first we analyzed the expression of

the components of cell signaling pathways, including ligands, receptors and target genes. NOTCH pathway was one of the significantly enriched in antral follicle GCs (Table S4). Expression of the components of NOTCH signaling revealed that ligands *DLL3* and *JAG1* were predominantly expressed in early oocytes before preovulatory stage. Their receptors *NOTCH2*, *NOTCH3* and the downstream target gene *HES1* were highly expressed in GCs (Figure 5A). This finding highlights the role of NOTCH signaling in oocyte-controlled proliferation and differentiation of GCs.

Then we analyzed the expression of the key components of TGF- β signaling pathway in oocytes and GCs. Our data revealed that the expression levels of *GDF9* in oocytes were high at all stages of follicular development, whereas *BMP15* was highly expressed only in the oocytes from antral and preovulatory follicles (Figure 5B). This observation points out that *GDF9* is likely to activate the TGF- β all through folliculogenesis, while *BMP15* mostly exerts its role at more advanced stages. In contrast, animal studies demonstrated that activation of the primordial follicles is mediated by *BMP15*²⁷, suggesting it might play different roles in folliculogenesis across the species. Interestingly, BMP type II receptor gene *BMPR2* and its target *ID3* were expressed in both oocytes and GCs. The presence of gene and its target in two different cell types, herein, allows to assume an existence of both autocrine and paracrine mechanisms in BMP signaling.

In addition, we identified the components of GC-derived signaling involved in folliculogenesis. For example, *KITLG* and its receptor *KIT*, previously implicated in paracrine signaling in folliculogenesis, were expressed in GCs and oocytes,

respectively, which is consistent with previous findings in animal models²⁸. *KITLG* was expressed in primordial and upregulated in primary follicles with a subsequent down-regulation, suggesting a possible role in primordial follicle activation (Figure S2). Together, our findings provide evidence that human folliculogenesis is coordinated by both autocrine and paracrine signaling pathway mechanisms that could be initiated by either oocytes or GCs.

To understand the distribution of gap junctions in human follicles, we screened both oocytes and GCs for the connexin (gap junction components) encoding genes, previously reported in mammalian ovary, namely *GJB2* (Cx26), *GJB4* (Cx30.3), *GJB1* (Cx32), *GJA4* (Cx37), *GJA5* (Cx40), *GJA1* (Cx43), *GJC1* (Cx45) and *GJA10* (Cx57)²⁹. Three out of these genes exhibited compartment-specific pattern across the follicle stages, while others were largely under-expressed. We identified that *GJC1* (Cx45) was expressed only in the oocytes and *GJA1* (Cx43) was exclusively expressed in the GCs at all follicle stages, while *GJA5* (Cx40) was pertained to the GCs of antral and preovulatory follicles (Figure 5C). We speculate that the heterotypic (Cx43/Cx45 or Cx40/Cx45) connexins might generate gap junction channels between oocyte and GCs, while homotypic (Cx43 or Cx40) connexins may generate the channels between GCs.

Global Expression Patterns of Maternal Effect Genes

To explore the global expression profiles of maternal-effect genes, we integrated the oocyte transcriptome data derived from this study with the transcriptome of human preimplantation embryos that had been identified in our previous work³⁰. We searched for the genes that were expressed in MII oocytes and in zygotes, but not expressed in

8-cell embryos, as these genes are carried over from oocyte to early embryo and then get degraded following zygotic genome activation (ZGA). Overall, we identified 1,785 putative maternal genes. The unsupervised cluster analysis identified 4 clusters of genes according to expression patterns at specific stages of follicular growth (Figure S7A, Table S5). The genes in cluster 1 (n = 47) and cluster 2 (n = 588) were not expressed in primordial stage and upregulated during advanced stages of follicular maturation with peak expression in MII oocytes. However, the expression patterns in cluster 3 (n = 987) and cluster 4 (n = 163) showed that these transcripts could accumulate before primordial follicle assembly and potentially could be involved in early folliculogenesis and embryo development. The GO analysis revealed biological functions that have been reported in association with meiosis were among the overrepresented terms in cluster 2. While in cluster 3 and 4, genes were enriched in ubiquitous functions (Figure S7B).

Comparison of Human and Mouse Oocyte Transcriptomic Profiles during Folliculogenesis

Appraisal of the inter-species differences in oocyte transcriptome dynamics during all the stages of folliculogenesis is important to better understand the translational limitations of animal oocyte models, but such comparative data are largely lacking. Recent meta-analysis demonstrated substantial variation in oocyte transcriptome across different species³¹. To better understand the interspecies complexity of oocyte transcriptome in human and mice, and to provide a comparative view of gene expression in human and mice oocytes across folliculogenesis, we compared our

findings in human oocytes with reported RNA-Seq datasets produced from mice oocytes collected from primordial to antral stage. Mice RNA-Seq data were downloaded from Gahurova's study³². The standardized analytical approach to the merged raw data revealed a total of 23,091 genes expressed in human (our data) and 11,585 genes expressed in mice oocytes (Figure 6A).

We focused further analyses on 16,175 one-to-one homologous genes shared by human and mice, as per Mouse Genome Informatics (MGI) database³³, of which 14,431 genes were expressed in human oocytes and 9,964 genes were expressed in mice oocytes. There were 9,698 homologous genes that overlapped between human and mice oocyte samples. Of these, there were 6,115 house-keeping genes, 2,127 consistently expressed genes and 1,456 DEGs identified in this study as differentially expressed between different stages of folliculogenesis (Figures 6A). GO analysis of the co-expressed DEGs revealed that these genes exhibited ubiquitous functions, and none were specific for oocyte development (Figure 6B), which suggested that distinct molecular mechanisms may be involved during folliculogenesis in human and mouse. Moreover, several homologous DEGs implicated in oocyte development, such as *POU5F1*, *GJC1*, *TEAD2* and *BMP15*, showed different expression patterns in human and mice oocytes (Figure 6C).

Next, we assumed that oocyte genes with concordant expression pattern and high degree of correlation ($r > 0.8$) are involved in conserved molecular mechanisms in human and mice oocytes (Figure 6D). Notably, consistent homologous genes (*ZP3*, *RAC1*, *PATL2*), annotated in association with "oocyte maturation defect" in Online

Mendelian Inheritance in Man (OMIM) database, showed stable expression across folliculogenesis in human and mice oocytes (Figure 6E), which proposes conserved functional importance of these genes for oocyte development in both species. These shared expression patterns suggested existence of evolutionary conserved mechanisms involved in folliculogenesis and highlights the potentially significant translational value of these conserved genes to the future research.

Discussion

This work describes the transcriptional landscape of the human oocytes and their surrounding GCs across all major stages of folliculogenesis. To the best of our knowledge, this is the first comprehensive investigation of human transcriptome from both the germ cell and somatic follicle compartments in the adult ovary at single-cell resolution. Previously, we have characterized the transcriptome profiles of human pre-implantation embryos and PGCs with their gonadal niche at several stages of development^{16,17,30}. It is well established that oogenesis, folliculogenesis and embryogenesis are sequential interrelated events, thus the transcriptome features of both fetal oogenesis and adult folliculogenesis confer the underlying molecular mechanisms of oocyte development. This body of work complements the findings from our previous studies and contributes to a comprehensive overview of the transcriptional events and their dynamics throughout the course of female gametogenesis with a subsequent link to early pre-implantation embryo.

Previously, characterization of ovarian follicle in humans relied on evaluations of the

oocytes at certain stages of development or of the unseparated follicle compartments¹²⁻¹⁴. Although previous studies provided important insights into molecular events involved in folliculogenesis, generalizing the conclusions is made challenging by the multiple technical aspects related to different testing platforms and analytical variables. Tissue heterogeneity is one of the main predicaments of biological research when it is crucial to distinguish cell-specific profiles that reflect the in-vivo status. Laser Capture Microdissection (LCM) is increasingly used to harvest the cells of interest and to improve experimental precision, although faces challenges of compromised RNA quality and cell integrity³⁴. To date, there have been no other studies that evaluated the transcriptomic profile across all the key stages of folliculogenesis and that performed parallel evaluations in both oocytes and GCs.

For the first time we demonstrated that each stage of folliculogenesis, defined by the morphological characteristics of both follicle compartments, had distinct transcriptome profile with different gene expression dynamics in oocytes and GCs. This finding supports the proposed, but not entirely proven concept that morphological classification of ovarian follicles reflects complex intrinsic transcriptional changes during folliculogenesis.

We also identified the candidate compartment-specific and stage-specific genes in both oocytes and GCs that provide a valuable clue for future more directed functional research. These findings may have important implication for development of essential genetic tools for cell-type-specific or stage-specific labeling and manipulations, which could be utilized in human follicle functional studies³⁵. Further, an unmet need for the

reliable predictors of oocyte developmental competency drives discovery of new biomarkers that would allow to establish more objective criteria for selecting competent oocytes³⁶. Previously reported biomarkers associated with successful embryo development and pregnancy outcomes were mainly derived from GCs and showed little consistency across the studies, which could be explained by the differences in experimental methodology^{18,19}. Some of the already proposed biomarkers exhibited distinct expression patterns in our study, which makes them particularly attractive candidates for further investigations. Of note, the functional characteristics for majority of the gene markers identified in this study are unrecognized and await further investigation.

In addition, we identified a subset of regulators of oocyte and GC transcriptome that may be involved in follicle development and acquisition of oocyte competency. Further evaluation of these TFs, their upstream signaling pathways, ligands, receptors and downstream targets in oocytes and GCs will provide more understanding of the transcriptional control of human folliculogenesis. These TFs may evolve as potential targets for future therapeutic interventions to modulate follicle development.

It has been widely acknowledged that oocyte-granulosa bidirectional communications are one of the core mechanisms involved in oocyte acquisition of developmental competence⁴. Communication via signal transduction (e.g. signaling pathways) and via direct cell-to-cell contact (e.g. gap junctions) provide molecular and structural basis for effective oocyte-GC crosstalk required for follicular development. However, precisely how oocyte-GC interact and how bi-directional communications

between the follicle compartments are initiated and maintained, is still unclear. In this study we uncovered the signaling pathways that are coordinately and reciprocally regulated in human oocytes and their surrounding GCs, by identifying the ligands and their receptors that were derived from the reciprocal compartments, but shared the expression pattern. This indicates that signaling pathways are activated by the ligands derived from the germ cells that act on the neighboring gonadal somatic cells, and vice versa. For example, in this study NOTCH signaling pathway, known important regulator of cell-cell communication³⁷, was shown to be activated in GCs via oocyte-driven mechanisms, which is in line with previously observed by our group in fetal ovary¹⁷. In contrast, KITLG-KIT pathway was GC-driven, while TGF- β pathway appeared to be activated through both autocrine and paracrine regulations. Single-cell resolution made possible evaluation of the distinct gap junctional channels in human follicles. We identified three different connexins, known to contribute to gap junctions in animals, and we infer their involvement in gap junctions in humans, which so far has been only demonstrated in animal models. Together, our findings provide further insight into molecular basis of bidirectional interactions in both cell signaling pathway and gap junction communications in human follicle. We support the notion that human folliculogenesis is coordinated by both autocrine and paracrine signaling pathway mechanisms that could be initiated by either oocytes or GCs.

In this study, we propose a set of candidate biomarkers that could be a valuable source of information on reproductive potential and ovarian reserve. Ideally, an ovarian reserve test should be easy to perform, reproducible and, importantly, should clearly

show evidence of improvements in patient reproductive outcomes following test-directed interventions. However, none of the currently available ovarian reserve tests, namely anti-Mullerian hormone (AMH)^{38,39}, follicle stimulating hormone (FSH) and the antral follicle count (AFC) measured by transvaginal ultrasound^{40,41} seem to meet the above criteria. It is now understood that while ovarian reserve markers can moderately predict response to stimulation in assisted reproductive technology (ART) cycles, none of these tests determine the pregnancy potential and their role in general population or in infertile women not undergoing ART treatment is not confirmed^{10,42}. For instance, AMH, originating from the GCs of secondary and antral follicles, has been initially accepted as the endocrine marker of FOR in humans⁴³. However, it has been recently proposed that AMH is not associated with the reproductive potential in women of late reproductive age⁸ and is a moderate predictor of menopause in general population⁹. The main limitation of the current approach is inability to directly measure ovarian reserve, since the biomarkers to determine the number or quality of primordial follicles, a non-renewable pool that determines reproductive life span⁴⁴, have yet to be identified. In addition, majority of the available biomarkers are secreted by GCs and currently there is little information on oocyte-derived markers that provide information on a status of follicle pool. Furthermore, it is possible to assume that ‘all-purpose biomarker’ does not exist, and different markers would be required to address the specific outcome of interest. For example, reproductive potential in general population is likely to be represented by the markers expressed at all stages of follicle development, while age of menopause could be more accurately predicted by the

markers that originate in primordial pool. It is also possible that response to fertility treatments and pregnancy outcomes may be better predicted by the markers confined to the antral follicles, the main targets of ovarian stimulation gonadotropin treatments, while the predictors of natural pregnancy outcomes may be secreted by the more advanced follicles. Until new data emerge, these questions will remain open. Further research to elucidate the biological roles of the proposed candidates, in conjunction with comprehensive evaluation of their expression levels in biological fluids could contribute to advancements in development of clinically informative ovarian reserve test.

In this study we offer further insights on species-specific gene expression during folliculogenesis. Mouse models are a mainstay of translational biomedical research and are widely implicated in the investigations of ovarian transcriptome. Mice are poly-ovulating species, breed readily, inexpensive and easy to handle, which makes them ideal candidates for laboratory experiments. Besides, mice and humans are at least 95% genetically identical, which allows to assume conservation of fundamental biological processes across the species. However, it has been previously reported that the oocyte transcriptome is highly variable across mammals and the human oocyte is likely to have a greater complexity than other mammals^{31,45}. Appraisal of the inter-species differences in oocyte transcriptome dynamics during all the stages of folliculogenesis is important to better understand the translational limitations of animal oocyte models, but such comparative data are largely lacking. We evidenced strong transcriptional activity in both human and mouse oocytes, which dominated in humans

and demonstrated considerable variability in oocyte transcriptome between human and mouse. We revealed different gene expression dynamics during folliculogenesis in both species and thus, propose a cautious approach when mice oocyte data are applied to the human domain. We also showed some degree of similarity in gene expression between human and mice oocytes, which highlights the potentially significant translational value of these conserved genes to the future research. We do not suggest that the mouse is an invalid experimental system for studying human oocytes. Rather, we provide a database for further investigations of the molecular mechanisms associated with oocyte development through mouse models, and improve the understanding on how well the oocyte transcriptome data translate from mouse to human.

It is important to mention that our study is not devoid of limitations. Firstly, despite rigorous experimental approach and robust analyses, our findings are not sufficient to confirm the functional characteristics of the identified transcripts. Rather, our data offer a comprehensive resource for future more directed functional studies in oocyte research. Next, our methods relied on the mechanical or enzymatic dissociation of follicular compartments, thus the information on temporal and spatial regulation of gene expression in GCs was not reflected. The recently proposed Geo-Seq protocol allows the profiling of transcriptome information from only a small number cells and retains their native spatial information³⁴, which will help in future investigations of gene expression with positional information. Finally, low number of MII oocytes included in this study could undermine the reliability of the results for this group, although satisfactory gene number and considerable homogeneity of the expression

data from these cells suggest adequate quality of the data.

In summary, understanding molecular events involved in folliculogenesis remains a key challenge for reproductive biology research with limited availability of human follicles for research being the main constrain. By using a high-resolution transcriptome analysis for single oocyte and its surrounding GCs, we recapitulated a cascade of the molecular events involved in folliculogenesis. This work contributes to the collaborative effort to strengthen our understanding of reproductive function, which may assist with the development of more targeted future interventions to improve oocyte competence in-vitro and in-vivo.

METHODS

Collection of Human Ovarian Tissue Samples

With oral and written informed consent, fresh ovarian tissues were taken from 7 female donors who underwent ovariectomy for the following indications: sex reassignment surgery (n = 1), fertility preservation for cervical cancer (n = 1), endometrial cancer (n = 2), benign ovarian mass (n = 2) and lymphoma (n = 1). All the donors were of reproductive age, ranging 24 to 32 years (median age of 28 years). All the participants had regular menstrual cycles, had no history of autoimmune or genetic conditions. None of the tissue donors were on hormonal treatment at least 6 months before surgery, had no previous ovarian surgery and were not exposed to any cytotoxic agents or radiotherapy. All the samples had normal histopathology (described below). The ovarian tissue samples were collected in operation theatre during the procedure, immediately transferred to the laboratory and treated as previously described ⁴⁶. Briefly, the human ovarian tissues were collected in the operation theatre and transported to the laboratory in Leibovitz's L-15 medium (Sigma-Aldrich) on ice with the supplementation of 1% human serum albumin (Life Technologies, Carlsbad, CA), 100 IU/mL streptomycin (Sigma, St. Louis, MO) and 100 µg/mL penicillin (Sigma, St. Louis, MO). Scalpel and surgical scissors were used to enucleate the medulla tissues. Then scalpel was used to cut the ovarian cortical tissues into small ovarian pieces with a size of 5 mm×5 mm×1 mm.

Ovarian Histology Assessment

Histological assessment was performed on all the ovarian tissue samples by using

hematoxylin and eosin (HE) staining as described elsewhere⁴⁷. After fixation, the samples were paraffin embedded and cut into serial sections of 5- μ m-thick. All the prepared tissue sections were reviewed by two independent pathologists and confirmed normal ovarian tissues.

Immunohistochemistry

Immunohistochemistry staining of the ovarian tissue samples that were utilized in this study for RNA-Seq experiments was performed by using the ABC Staining System (Zhongshan Golden Bridge Biotechnology, Inc., Beijing, China), as previously described⁴⁸. Brown staining of the cytoplasm or nucleus of the cells was considered as positive.

Isolation of Human Oocytes and GCs

Human follicles were isolated from fresh ovarian tissues as described previously^{49,50}. Briefly, after removal of medulla tissues, the ovarian cortical pieces were placed in a tissue sectioner (McIlwain Tissue Chopper, The Mickle Laboratory, Guildford, UK) and cut into 0.5 $\times$ 0.5 $\times$ 1 mm pieces. Then the tissue pieces were enzymatically digested by a mixed digestion medium, which included α MEM (Sigma-Aldrich) media, 0.04 mg/ml Liberase DH (Dispase High; Roche Diagnostics GmbH, Mannheim, Germany), 10 IU/ml DNase I (Sigma-Aldrich), 100 IU/ml penicillin and 100 μ g/ml streptomycin (Invitrogen) and incubated for 75 min on a shaker at 37°C (Thermo Fisher, Marietta, OH, USA). The incubation was terminated by double wash with DPBS (Sigma-Aldrich) supplemented with 10% HSA (LifeGlobal). Then the follicles were isolated mechanically using 29G needles and were transferred to culture

medium. Subsequently, the single oocytes and GCs were enzymatically separated with
 accutase (Sigma-Aldrich) and were mechanically isolated by using 29G needles. GC
 samples comprised randomly selected 10 cells per sample because of low abundancy of
 RNA in these cells. GCs from the antral and preovulatory follicles included cumulus
 cells isolated from the cumulus-oocyte complex (COC) as following: 6 samples of
 cumulus cells from the antral follicles and 16 samples from the preovulatory follicles.
 Follicle stages were classified according to the criteria described by Gougeon ⁵¹. The
 diameters of follicles and oocytes were measured under light microscope (Nikon).

Single-Cell cDNA Libraries Construction from Oocytes and GCs

The oocytes and GCs that were isolated from follicles were analyzed by scRNA-Seq as
 previously described ^{30,52}. Briefly, the single oocytes or 10 randomly selected GCs were
 transferred into the lysis buffer quickly using a mouth pipette. Then we performed
 reverse transcription on the cell lysate and terminal deoxynucleotidyl transferase was
 adopted to add a poly A tail to the 3' end of the first-strand cDNAs, next we performed
 20 cycles of PCR to amplify the single-cell cDNA library. qPCR analysis was
 conducted to check the quality of the cDNA libraries using housekeeping genes
 (GAPDH and RPS24). The RNA-Seq libraries were constructed by a Kappa Hyper
 Prep Kit (Kappa Biosystems).

RNA-Seq Data Processing

The analysis of single-cell RNA-Seq data was carried out as previously described
^{16,17,53}. Briefly, RNA-Seq raw reads with 10% low quality bases, adapters and artificial
 sequences (including UP1, UP2, polyA sequences) introduced during the experimental

processes were trimmed by in house scripts. Next, the trimmed clean reads were aligned to the UCSC human hg19 reference using Tophat2 (v2.1.0) with default settings⁵⁴. Cufflinks (v2.2.1) was further used to quantify transcription levels of annotated genes⁵⁵.

Previous published data, including those from human MII oocytes³⁰, human pre-implantation embryos³⁰, and mouse oocytes^{56,57}, were downloaded from the GEO data sets, and the raw fastq reads were obtained and incorporated into our analysis. For all sequenced cells, we counted the number of genes detected in each cell. Cells with fewer than 2,400 genes or 500,000 mapped reads were filtered out. In total, 80 oocytes and 71 GCs at five developmental stages passed the filter standards. To ensure the accuracy of estimated gene expression levels, only genes with FPKM > 1 in at least one cell were analyzed⁵⁸. Expression levels of each gene were plus one then log2 transformed in the following analysis. Expression levels of each oocyte and GC were in Table S1.

Principal Component Analysis (PCA)

The Seurat method was applied to analyze the single-cell data (77 unmaturred oocytes, three matured MII oocytes and 71 GCs) to observe the whole clustering profile⁵⁹. Only highly variable genes (coefficient of variation > 0.5) were used as inputs for PCA. And the marker genes in PCA plot were plotted by the FeaturePlot function in Seurat package. To complement the PCA clustering more accurately, we also clustered the oocytes and GCs separately using the FactoMineR package in R⁶⁰.

Identification of DEGs and Gene Ontology Analysis

Multiple t-test was used to obtain the statistical significance of differentially expressed genes in each stage. Only the genes with significant p-values and false discovery rate (FDR) less than 0.05 with a fold change of log2 transformed FPKM larger than 1.5 were considered to be differentially expressed. Gene ontology analysis of differentially expressed genes was performed using DAVID ⁶¹.

Identifying Expression Patterns of Maternal Effect Gene

Maternal effect genes play a critical role before zygote genome activation (ZGA). It's reported that ZGA mainly happened at 4 cells to 8 cells stages ^{30,62}. To explore the expression patterns of maternal effect genes during folliculogenesis, we searched for genes carried from oocytes to early embryo and degraded before ZGA. Genes expressed (FPKM > 1) at MII and zygote stage but relatively low expressed (FPKM < 1) at 8 cells stage were taken as candidates. The candidate genes were clustered by expression correlation and cut into four clusters with cutree function in R.

Identifying Secretory Protein Coding Signature Genes

To identify candidate biomarkers for predicting ovarian reserve, we focused on the oocyte- or GC-derived secretory proteins. The secretory protein encoding genes were downloaded from the Human Protein Atlas database (<https://www.proteinatlas.org/>). The GTEx v6 database (<https://www.gtexportal.org>) was utilized to obtain information on gene expression level in tissues of interest. We aimed to evaluate only protein secreting genes expressed in ovary and therefore we focused on genes that were previously identified only in gonads and could be present in brain, but not in any

other tissues⁶³. We assumed that brain-blood (B-B) barrier prevents secretion of the proteins into systemic circulation and thus reduces confounding effect on peripheral levels of these proteins. This assumption was based on the previously observed high levels of AMH in brain, which supports the idea of restrictive function of B-B barrier⁶⁴.

Transcription Factor Network Construction

Transcription factors play a key role in regulating development. To find the driver factors and construct their regulatory network in each two consecutive stages, we used ARACNe to perform the regulatory network analysis as described previously¹⁷. First, the 1,469 human transcription factors in AnimalTFDB²⁵ and gene expression matrix were taken as inputs for ARACNe-AP software²⁵. Then, viper package in R was used to visualize the transcription factors and their target genes in each consecutive stage⁶⁵. Regulators with p-values less than 0.01 were inferred as driver factors in each two consecutive stages.

GSEA Analysis

To identify the significant enriched pathways in each two consecutive development stages, we used Gene Set Enrichment Analysis (GSEA, <http://software.broadinstitute.org/gsea/index.jsp>) to perform enrichment analysis, and the KEGG pathway was used⁶⁶. The gene sets that showed nominal p-value less than 0.05 were chosen as enriched.

Analysis of Conservation Between Human and Mouse

Human and mouse homologous genes were downloaded from Vertebrate Homology

673 Database (<http://www.informatics.jax.org/homology.shtml>). Housekeeping genes
674 were obtained from Human Protein Atlas. Human oocytes expressed genes were
675 overlapped with mouse oocytes expressed homologous genes to find genes expressed
676 in both human and mouse oocytes, and the housekeeping genes were filtered out from
677 the overlapped gene set. The retained genes were divided as homologous DEGs
678 (genes in DEGs of human oocytes) and homologous non-DEGs. The genes involved
679 in oocyte maturation defect were downloaded from Online Mendelian Inheritance in
680 Man (OMIM) database (<https://www.omim.org/>) and incorporated in to analysis.
681

References:

- 1 Eppig, J. J., Wigglesworth, K. & Pendola, F. L. The mammalian oocyte orchestrates the rate of ovarian follicular development. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2890-2894, doi:10.1073/pnas.052658699 (2002).
- 2 Moley, K. H. & Schreiber, J. R. Ovarian follicular growth, ovulation and atresia. Endocrine, paracrine and autocrine regulation. *Advances in experimental medicine and biology* **377**, 103-119 (1995).
- 3 Sanchez, F. & Smits, J. Molecular control of oogenesis. *Biochimica et biophysica acta* **1822**, 1896-1912, doi:10.1016/j.bbadis.2012.05.013 (2012).
- 4 Li, R. & Albertini, D. F. The road to maturation: somatic cell interaction and self-organization of the mammalian oocyte. *Nature reviews. Molecular cell biology* **14**, 141-152, doi:10.1038/nrm3531 (2013).
- 5 Zhang, K. & Smith, G. W. Maternal control of early embryogenesis in mammals. *Reproduction, fertility, and development* **27**, 880-896, doi:10.1071/RD14441 (2015).
- 6 Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. *Developmental cell* **42**, 316-332, doi:10.1016/j.devcel.2017.07.026 (2017).
- 7 Gleicher, N., Weghofer, A. & Barad, D. H. Defining ovarian reserve to better understand ovarian aging. *Reproductive biology and endocrinology : RB&E* **9**, 23, doi:10.1186/1477-7827-9-23 (2011).
- 8 Steiner, A. Z. *et al.* Association Between Biomarkers of Ovarian Reserve and Infertility Among Older Women of Reproductive Age. *Jama* **318**, 1367-1376, doi:10.1001/jama.2017.14588 (2017).
- 9 Depmann, M. *et al.* Does anti-Mullerian hormone predict menopause in the general population? Results of a prospective ongoing cohort study. *Human reproduction* **31**, 1579-1587, doi:10.1093/humrep/dew112 (2016).
- 10 Jirge, P. R. Ovarian reserve tests. *Journal of human reproductive sciences* **4**, 108-113, doi:10.4103/0974-1208.92283 (2011).
- 11 Knight, P. G. & Glister, C. TGF-beta superfamily members and ovarian follicle development. *Reproduction* **132**, 191-206, doi:10.1530/rep.1.01074 (2006).
- 12 Grondahl, M. L. *et al.* The dormant and the fully competent oocyte: comparing the transcriptome of human oocytes from primordial follicles and in metaphase II. *Molecular human reproduction* **19**, 600-617, doi:10.1093/molehr/gat027 (2013).

714 13 Kristensen, S. G., Ebbesen, P. & Andersen, C. Y. Transcriptional profiling of five isolated
715 size-matched stages of human preantral follicles. *Molecular and cellular endocrinology* **401**,
716 189-201, doi:10.1016/j.mce.2014.12.012 (2015).

717 14 Markholt, S. *et al.* Global gene analysis of oocytes from early stages in human
718 folliculogenesis shows high expression of novel genes in reproduction. *Molecular human*
719 *reproduction* **18**, 96-110, doi:10.1093/molehr/gar083 (2012).

720 15 Picelli, S. Single-cell RNA-sequencing: The future of genome biology is now. *RNA biology* **14**,
721 637-650, doi:10.1080/15476286.2016.1201618 (2017).

722 16 Guo, F. *et al.* The Transcriptome and DNA Methylome Landscapes of Human Primordial
723 Germ Cells. *Cell* **161**, 1437-1452, doi:10.1016/j.cell.2015.05.015 (2015).

724 17 Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and
725 Gonadal Niche Interactions. *Cell stem cell* **20**, 858-873 e854, doi:10.1016/j.stem.2017.03.007
726 (2017).

727 18 Huang, Z. & Wells, D. The human oocyte and cumulus cells relationship: new insights from
728 the cumulus cell transcriptome. *Molecular human reproduction* **16**, 715-725,
729 doi:10.1093/molehr/gaq031 (2010).

730 19 Parks, J. C. *et al.* Corona cell RNA sequencing from individual oocytes revealed transcripts
731 and pathways linked to euploid oocyte competence and live birth. *Reproductive biomedicine*
732 *online* **32**, 518-526, doi:10.1016/j.rbmo.2016.02.002 (2016).

733 20 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell
734 RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

735 21 Bonnet, A. *et al.* An overview of gene expression dynamics during early ovarian
736 folliculogenesis: specificity of follicular compartments and bi-directional dialog. *BMC*
737 *genomics* **14**, 904, doi:10.1186/1471-2164-14-904 (2013).

738 22 Kerr, B., Garcia-Rudaz, C., Dorfman, M., Paredes, A. & Ojeda, S. R. NTRK1 and NTRK2
739 receptors facilitate follicle assembly and early follicular development in the mouse ovary.
740 *Reproduction* **138**, 131-140, doi:10.1530/REP-08-0474 (2009).

741 23 Tisdall, D. J., Watanabe, K., Hudson, N. L., Smith, P. & McNatty, K. P. FSH receptor gene
742 expression during ovarian follicle development in sheep. *Journal of molecular endocrinology*
743 **15**, 273-281 (1995).

744 24 Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory
745 networks in a mammalian cellular context. *BMC bioinformatics* **7 Suppl 1**, S7,
746 doi:10.1186/1471-2105-7-S1-S7 (2006).

- 747 25 Zhang, H. M. *et al.* AnimalTFDB: a comprehensive animal transcription factor database.
748 *Nucleic acids research* **40**, D144-149, doi:10.1093/nar/gkr965 (2012).
- 749 26 Mingoti, G. Z., Garcia, J. M. & Rosa-e-Silva, A. A. Steroidogenesis in cumulus cells of
750 bovine cumulus-oocyte-complexes matured in vitro with BSA and different concentrations of
751 steroids. *Animal reproduction science* **69**, 175-186 (2002).
- 752 27 Kim, J. Y. Control of ovarian primordial follicle activation. *Clinical and experimental*
753 *reproductive medicine* **39**, 10-14, doi:10.5653/term.2012.39.1.10 (2012).
- 754 28 Thomas, F. H. & Vanderhyden, B. C. Oocyte-granulosa cell interactions during mouse
755 follicular development: regulation of kit ligand expression and its role in oocyte growth.
756 *Reproductive biology and endocrinology : RB&E* **4**, 19, doi:10.1186/1477-7827-4-19 (2006).
- 757 29 Gershon, E., Plaks, V. & Dekel, N. Gap junctions in the ovary: expression, localization and
758 function. *Molecular and cellular endocrinology* **282**, 18-25, doi:10.1016/j.mce.2007.11.001
759 (2008).
- 760 30 Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and
761 embryonic stem cells. *Nature structural & molecular biology* **20**, 1131-1139,
762 doi:10.1038/nsmb.2660 (2013).
- 763 31 Biase, F. H. Oocyte Developmental Competence: Insights from Cross-Species Differential
764 Gene Expression and Human Oocyte-Specific Functional Gene Networks. *Omics : a journal*
765 *of integrative biology* **21**, 156-168, doi:10.1089/omi.2016.0177 (2017).
- 766 32 Gahurova, L. *et al.* Transcription and chromatin determinants of de novo DNA methylation
767 timing in oocytes. *Epigenetics Chromatin* **10**, 25, doi:10.1186/s13072-017-0133-5 (2017).
- 768 33 Blake, J. A. *et al.* Mouse Genome Database (MGD)-2017: community knowledge resource for
769 the laboratory mouse. *Nucleic acids research* **45**, D723-D729, doi:10.1093/nar/gkw1040
770 (2017).
- 771 34 Chen, J. *et al.* Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq.
772 *Nature protocols* **12**, 566-580, doi:10.1038/nprot.2017.003 (2017).
- 773 35 Zuccotti, M., Merico, V., Cecconi, S., Redi, C. A. & Garagna, S. What does it take to make a
774 developmentally competent mammalian egg? *Human reproduction update* **17**, 525-540,
775 doi:10.1093/humupd/dmr009 (2011).
- 776 36 Patrizio, P., Fragouli, E., Bianchi, V., Borini, A. & Wells, D. Molecular methods for selection
777 of the ideal oocyte. *Reproductive biomedicine online* **15**, 346-353 (2007).
- 778 37 Vanorny, D. A. & Mayo, K. E. The role of Notch signaling in the mammalian ovary.
779 *Reproduction* **153**, R187-R204, doi:10.1530/REP-16-0689 (2017).

- 780 38 Freeman, E. W., Sammel, M. D., Lin, H. & Gracia, C. R. Anti-mullerian hormone as a
781 predictor of time to menopause in late reproductive age women. *The Journal of clinical*
782 *endocrinology and metabolism* **97**, 1673-1680, doi:10.1210/jc.2011-3032 (2012).
- 783 39 Dolleman, M. *et al.* Anti-Mullerian hormone is a more accurate predictor of individual time to
784 menopause than mother's age at menopause. *Human reproduction* **29**, 584-591,
785 doi:10.1093/humrep/det446 (2014).
- 786 40 Hendriks, D. J. *et al.* Repeated clomiphene citrate challenge testing in the prediction of
787 outcome in IVF: a comparison with basal markers for ovarian reserve. *Human reproduction*
788 **20**, 163-169, doi:10.1093/humrep/deh553 (2005).
- 789 41 Wellons, M. F. *et al.* Antral follicle count predicts natural menopause in a population-based
790 sample: the Coronary Artery Risk Development in Young Adults Women's Study. *Menopause*
791 **20**, 825-830, doi:10.1097/GME.0b013e31827f06c2 (2013).
- 792 42 Gleicher, N., Weghofer, A. & Barad, D. H. The role of androgens in follicle maturation and
793 ovulation induction: friend or foe of infertility treatment? *Reproductive biology and*
794 *endocrinology : RB&E* **9**, 116, doi:10.1186/1477-7827-9-116 (2011).
- 795 43 Peluso, C. *et al.* AMH: An ovarian reserve biomarker in assisted reproduction. *Clinica*
796 *chimica acta; international journal of clinical chemistry* **437**, 175-182,
797 doi:10.1016/j.cca.2014.07.029 (2014).
- 798 44 Wallace, W. H. & Kelsey, T. W. Human ovarian reserve from conception to the menopause.
799 *PloS one* **5**, e8772, doi:10.1371/journal.pone.0008772 (2010).
- 800 45 Sylvestre, E. L. *et al.* Evolutionary conservation of the oocyte transcriptome among
801 vertebrates and its implications for understanding human reproductive function. *Molecular*
802 *human reproduction* **19**, 369-379, doi:10.1093/molehr/gat006 (2013).
- 803 46 Zhang, Y. *et al.* Mesenchymal stem cell-derived angiogenin promotes primordial follicle
804 survival and angiogenesis in transplanted human ovarian tissue. *Reproductive biology and*
805 *endocrinology : RB&E* **15**, 18, doi:10.1186/s12958-017-0235-8 (2017).
- 806 47 Xia, X. *et al.* Mesenchymal Stem Cells Enhance Angiogenesis and Follicle Survival in Human
807 Cryopreserved Ovarian Cortex Transplantation. *Cell transplantation* **24**, 1999-2010,
808 doi:10.3727/096368914X685267 (2015).
- 809 48 Gao, J. M. *et al.* Improvement in the quality of heterotopic allotransplanted mouse ovarian
810 tissues with basic fibroblast growth factor and fibrin hydrogel. *Human reproduction* **28**,
811 2784-2793, doi:10.1093/humrep/det296 (2013).
- 812 49 Wang, T. R. *et al.* Basic fibroblast growth factor promotes the development of human ovarian

813 early follicles during growth in vitro. *Human reproduction* **29**, 568-576,
814 doi:10.1093/humrep/det465 (2014).

815 50 Vanacker, J. *et al.* Enzymatic isolation of human primordial and primary ovarian follicles with
816 Liberase DH: protocol for application in a clinical setting. *Fertility and sterility* **96**, 379-383
817 e373, doi:10.1016/j.fertnstert.2011.05.075 (2011).

818 51 Gougeon, A. Dynamics of follicular growth in the human: a model from preliminary results.
819 *Human reproduction* **1**, 81-87 (1986).

820 52 Wang, T. R. *et al.* Human single follicle growth in vitro from cryopreserved ovarian tissue
821 after slow freezing or vitrification. *Human reproduction* **31**, 763-773,
822 doi:10.1093/humrep/dew005 (2016).

823 53 Zhou, F. *et al.* Tracing haematopoietic stem cell formation at single-cell resolution. *Nature*
824 **533**, 487-492, doi:10.1038/nature17997 (2016).

825 54 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq.
826 *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

827 55 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated
828 transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**,
829 511-515, doi:10.1038/nbt.1621 (2010).

830 56 Veselovska, L. *et al.* Deep sequencing and de novo assembly of the mouse oocyte
831 transcriptome define the contribution of transcription to the DNA methylation landscape.
832 *Genome biology* **16**, 209, doi:10.1186/s13059-015-0769-z (2015).

833 57 Hikabe, O. *et al.* Reconstitution in vitro of the entire cycle of the mouse female germ line.
834 *Nature* **539**, 299-303, doi:10.1038/nature20104 (2016).

835 58 Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using
836 single-cell RNA-seq. *Nature* **509**, 371-375, doi:10.1038/nature13173 (2014).

837 59 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells
838 Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

839 60 Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using
840 Principal Component Analysis and heatmap. *Nucleic acids research* **43**, W566-570,
841 doi:10.1093/nar/gkv468 (2015).

842 61 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large
843 gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57,
844 doi:10.1038/nprot.2008.211 (2009).

845 62 Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs between the four- and

846 eight-cell stages of preimplantation development. *Nature* **332**, 459-461,
847 doi:10.1038/332459a0 (1988).

848 63 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419,
849 doi:10.1126/science.1260419 (2015).

850 64 Haddad-Tovoli, R., Dragano, N. R. V., Ramalho, A. F. S. & Velloso, L. A. Development and
851 Function of the Blood-Brain Barrier in the Context of Metabolic Control. *Frontiers in*
852 *neuroscience* **11**, 224, doi:10.3389/fnins.2017.00224 (2017).

853 65 Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using
854 network-based inference of protein activity. *Nature genetics* **48**, 838-847, doi:10.1038/ng.3593
855 (2016).

856 66 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
857 interpreting genome-wide expression profiles. *Proceedings of the National Academy of*
858 *Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102
859 (2005).

860

861

Figure legends

Figure 1. The Global Transcriptome Patterns of Human Oocytes and granulosa cells (GCs).

A) Schematic illustration of the study workflow.

Primordial - primordial follicles; Primary - primary follicles; Secondary - secondary follicles; Antral - antral follicles; Preovulatory - preovulatory follicles; GCs - granulosa cells.

B) Principal component analysis (PCA) of the transcriptome of scRNA-Seq data from all oocytes and GCs included in this study. The PC1 separates two follicular compartments (oocytes vs. GCs). The PC2 separates the samples according to their follicular stages. Red color indicates oocytes, blue color indicates GCs. The sizes of the points represent follicles of different sizes.

C) Expression patterns of oocyte marker genes exhibited on PCA plots; a gradient of gray to red indicates the low to high gene expression level.

D) Expression patterns of GC marker genes exhibited on PCA plots.

E) Expression of the candidate cell-type specific markers of oocytes on PCA plots.

F) Expression of the candidate cell-type specific markers of GCs on PCA plots.

Figure 2. Gene Expression Dynamics and Transcriptional Characteristics of Oocytes.

A) Principal component analysis (PCA) of the transcriptome of scRNA-Seq data from 80 oocytes collected from follicles at different stages of development. Oocytes are clustered into five subpopulations corresponding to morphological stages.

Different colors and sizes of the points represent follicles of different stages and sizes, respectively. See also Figure S1E for the results of PCA that portrays sampling entities within the oocytes.

B) Heatmap of all the differentially expressed genes (DEGs) in oocytes at five stages of folliculogenesis. The numbers of identified DEGs are indicated on the y-axis, and the stages of follicle development are presented along the x-axis. The color key from blue to red indicates the relative gene expression level from low to high, respectively.

C) Significantly enriched GO terms (biological processes) of DEGs in oocytes at five stages of folliculogenesis.

D) Heatmap of genes involved in meiosis that are differentially expressed in oocytes at five stages of folliculogenesis.

E) Immunohistochemistry staining of selected DEGs. The boxplot demonstrating gene expression level is presented on the left of each corresponding immunohistochemistry panel. White triangles indicate the follicles. Scale bar, 100 μ m.

Figure 3. Dynamic Gene Expression Patterns and Transcriptional Features in GCs.

A) Principal component analysis (PCA) of the transcriptome of scRNA-Seq data from 71 GCs collected from follicles at different stages of development. GCs are clustered into five subpopulations. Different colors and sizes of the points represent follicles of different stages and sizes, respectively.

B) Heatmap of all the differentially expressed genes (DEGs) in GCs at five stages of

folliculogenesis. The numbers of identified DEGs are indicated on the y-axis, and the stages of follicle development are presented along the x-axis. The color key from blue to red indicates the relative gene expression level from low to high, respectively.

C) Significantly enriched GO terms (biological processes) of DEGs in GCs at five stages of folliculogenesis.

D) Heatmap of cell-cycle related genes that are expressed in GCs at five stages of folliculogenesis.

E) Immunohistochemistry staining of selected DEGs. The boxplot demonstrating gene expression level is presented on the left of each corresponding immunohistochemistry panel. White triangles indicate the follicles. Scale bar, 100 μ m.

Figure 4. Inferred Key Transcriptional Factors in Oocytes at Each Stage-to-stage Transition of Folliculogenesis.

A) MARINA plots of targets for each candidate master regulator. Red vertical bar represents the activated targets, blue represents the repressed targets. On the x axis, genes are rank-sorted according to the significance of differential expression between the two developmental stages. The candidate master regulators are displayed on the right. The corresponding p-values of these master regulators indicate the significance of enrichment are displayed on the left. See also Figure S4 for the analogous information in GCs.

B) Violin plots show the relative expression levels ($\log_2$ [FPKM+1]) of each master regulator in two consecutive stages. See also Figure S4 for the analogous information in

GCs.

Figure 5. Signaling Pathways and Gap Junction Involved in Oocyte-GC Crosstalk.

A) NOTCH signaling pathway involved in oocyte-GC crosstalk in folliculogenesis. The relative expression levels ($\log_2$ [FPKM+1]) of the specific ligands, receptors, and target genes are shown. The diagrams at the left show the relationship among these genes.

B) TGF- β signaling pathway involved in oocyte-GC crosstalk.

C) Gap junction involved in oocyte-GC crosstalk. The relative expression levels ($\log_2$ [FPKM+1]) of the connexin genes are shown. The diagrams at the left show the putative relationship among these genes.

Figure 6. Comparison of Human and Mouse Oocyte Transcriptomic Profiles during Folliculogenesis.

A) Venn diagram of total expressed genes in human oocytes (peach) and mouse oocytes (light green) demonstrates overlap between the gene populations in two species and also overlap with the human housekeeping genes (deep orange). Pie chart represents the non-housekeeping human-mouse oocyte-derived homologous genes (3,583), including homologous DEGs (red) and homologous ubiquitous genes (green).

B) Heatmap of homologous DEGs in oocytes at five stages of folliculogenesis. The numbers of identified DEGs are indicated on the y-axis, and the stages of follicle

development are presented along the x-axis. The color key from blue to red indicates the relative gene expression level from low to high, respectively. The enriched GO terms of human oocytes are shown on the right.

C) Bar plots demonstrate a comparative analysis of the selected oocyte-derived homologous DEGs that show the distinct expression patterns in human (left) and mouse (right) oocytes during folliculogenesis. Different colors represent different stages of folliculogenesis.

D) Bar plots demonstrate a comparative analysis of four oocyte-derived homologous DEGs that show the similar expression pattern in human (left) and mouse (right) oocytes during folliculogenesis.

E) Bar plots demonstrate a comparative analysis of three homologous non-DEGs annotated as oocyte maturation defect genes (OMIM) in human (left) and mouse (right) oocytes during folliculogenesis.

Supplementary Figure Legends

Figure S1. Sample Collection and Quality Control. Related to Figure 1 and 2.

A) Table demonstrating number of samples per two follicle cell-type specific compartments at five follicle stages.

B) Representation of the follicles in primordial, primary, secondary and antral stages.

Scale bar, 100 μ m.

C) The number of detected genes (FPKM > 1) in each samples included in this study.

D) Saturation analysis for scRNA-Seq data. X-axis is mapped reads for each down-sampling data, y-axis is number of genes detected.

E) Principal component analysis (PCA) of the transcriptome of human oocytes that portrays sampling entities. Different colors and sizes of the points represent oocytes of different stages and sizes, respectively. Different shapes of the points represent different sampling entities (individuals).

Figure S2. The Expression Patterns of Selected Marker Genes in Human Oocytes and GCs. Related to Figure 2 and 3.

A) Boxplots of selected marker genes in oocytes at each follicle stage.

B) Boxplots of selected hormone receptors and genes involved in steroidogenesis in GCs at each follicle stage.

Figure S3. The Expression Patterns of Signature Genes in Oocytes and GCs at Each Stage. Related to Figure 2 and 3.

A) Heatmap of signature genes in oocytes at five stages of folliculogenesis. The numbers of identified signature genes are indicated on the y-axis, and the stages of

follicle development are presented along the x-axis. The color key from blue to red indicates the relative gene expression level from low to high, respectively.

B) Boxplots of the relative expression levels ($\log_2$ [FPKM+1]) of selected signature genes in oocytes to exemplify the expression patterns in each stage.

C) Heatmap of signature genes in GCs at different stages of folliculogenesis.

D) Boxplots of the relative expression levels ($\log_2$ [FPKM+1]) of selected signature genes in GCs to exemplify the expression patterns in each stage.

Figure S4. Secretory Protein Coding Signature Genes for Ovary Reserve Prediction.

A) Heatmap of the secretory protein encoding genes expressed in oocytes (left) and GCs (right). The list of these genes is indicated on the y-axis according to the five clusters by similarity of expression patterns, and the stages of follicle development are presented along the x-axis. The color key from blue to red indicates the relative gene expression level from low to high, respectively.

B) Boxplots of the relative expression levels ($\log_2$ [FPKM+1]) of selected secretory proteins encoding genes to exemplify the expression pattern in each cluster.

Figure S5. Inferred Key Transcriptional Factors in GCs at Each Stage-to-stage Transition of Folliculogenesis. Related to Figure 4.

A) MARINA plots of targets for each candidate master regulator. Red vertical bar represents the activated targets, blue represents the repressed targets. On the x axis, genes were rank-sorted according to the significance of differential expression between the two developmental stages. The candidate master regulators were

displayed on the right. The corresponding p-values of these master regulators indicate the significance of enrichment were displayed on the left.

B) Violin plots show the relative expression levels ($\log_2$ [FPKM+1]) of each master regulator in two consecutive stages.

Figure S6 Signaling Pathways Enriched in Follicle Recruitment and Antral Formation by GSEA/KEGG Analysis.

A) Heatmaps of DEGs between primordial and primary stage in oocytes and GCs. The numbers of identified DEGs are indicated on the y-axis, and the stages of follicle development are presented along the x-axis. The color key from blue to red indicates the relative gene expression level from low to high, respectively.

B) GSEA enrichment plots of KEGG signaling pathways in oocytes and GCs between primordial and primary stage.

C) Heatmaps of DEGs between secondary and antral stage in oocytes and GCs.

D) GSEA enrichment plots of KEGG signaling pathways in oocytes and GCs between secondary and antral stage.

Figure S7. The Expression of Putative Maternal-Effect Genes in Folliculogenesis and Early Embryonic Development.

A) Heatmap of the putative maternal-effect genes expressed in folliculogenesis and early embryonic development. The identified genes were clustered into four clusters by similarity of expression patterns. The stages of follicle development and early embryonic development are presented along the x-axis. Right panel show relative

1027 expression level of each cluster. The color key from blue to red indicates the relative

1028 gene expression level from low to high, respectively.

1029 B) Bar plots show the enriched GO terms of each cluster genes.

1030

Figure 2

Figure 5

A NOTCH signal pathway

B TGF-beta signal pathway

C Gap junction

Figure 6

A

B

C

D

E

A

Follicle stages	Primordial	Primary	Secondary	Antral	Preovulatory	Total
Oocytes	17	25	12	23	3	80
GCs	8	15	6	24	18	71

B

C

D

E

A

B

A

B

A

B