Design of a Microphysiologic Ovarian Follicle Culture System Using Multiomics Technology

by

Andrea Suzanne Kuliasha Jones

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Ariella Shikanov, Chair
Professor Jun Z. Li
Professor Vasantha Padmanabhan
Professor David Sept
Professor Lonnie D. Shea

Andrea S.K. Jones

askj@umich.edu

ORCID iD: 0000-0003-0535-7941

Dedication

To my parents, who have unwaveringly and proudly supported my every endeavor, from the earliest piano recitals to flying across the country to accompany me while I present at a conference. As a parent embarking on the lifetime journey of loving and supporting a child, I admire you both deeply for raising me with confidence, grit, and a commitment to doing everything with kindness. The double-dose of persistence you have imparted to me has fueled much of this work and for that I am grateful. I cannot put into words how very blessed I am to be your daughter.

To my husband Christopher, who courageously took on the burden of this work with me a mere month after we married. These five years of marriage, graduate school, and having a baby have forged between us a partnership that is the greatest accomplishment of my life. Your kindness, patience, and love have truly carried me through the most difficult seasons of this work. Thank you for joining me on this journey and encouraging me to settle for nothing less than the work that makes my heart sing.

To my daughter Hazel, Baby J arriving in May 2024, and any other children with whom we may be blessed in the future—it is my most heartfelt prayer that your life will be full of doing the things that you love most, and that you will be surrounded by people who support you on your journey, much like I have been. I will always fiercely and steadfastly support you and your dreams, wherever they may take you. You are the joy of my life and I am so immensely grateful to be your mom.

Acknowledgements

I believe I once explained graduate school as an individual sport to someone… I could not have been more wrong. This work and my entire graduate experience have been born of many collaborations and friendships that contributed to my success. First, I would like to acknowledge my advisor, Dr. Ariella Shikanov. Your guidance and mentorship have shaped me these past five years and inspired me to follow in your footsteps and build a career around training the next generation of scientists. The time I have spent as your trainee has been immensely rewarding and full of laughter, hard work, and learning. Thank you for pushing me beyond my perceived limits and for opening door after door for me to expand my skillset. I cherish the time I have spent under your leadership and hope to share decades ahead as colleagues and friends.

I would also like to thank my committee members Dr. Jun Li, Dr. Vasantha Padmanabhan, Dr. David Sept, and Dr. Lonnie Shea. Your advice and perspective have enriched this work and inspired me on countless occasions. It has been an honor to learn from and work with you.

Thank you to my most formative peer mentors, Dr. Megan Cox and Dr. Claire Nason-Tomaszewski. I am indebted to both of you for sparking in me a love for research and nurturing me through hundreds of hours of teaching over my undergraduate and graduate years. I have been extremely fortunate to learn from two determined, meticulous, and patient scientists in such close succession. I have followed your example these five years and much of my success is a direct result of your investment in me.

Thank you to my labmates. You have been the hands that worked alongside me, hugged me after failed experiments, and spend the night in the lab with me on countless occasions waiting for donor tissue. Thank you for sharing your strength with me in difficult seasons, for imparting wisdom, for going to get coffee day after day, and for always saving your pickles for me. Thank you for supporting me as I became a mom and loving my daughter so well. Special thanks to Margaret, for being my constant companion from the earliest days of graduate school and for being my chosen family—your friendship means more than words can say. I will deeply miss sitting side-by-side day after day, however I am confident that our friendship, fueled by coffee and a truly endless variety of things to share and laugh about, will withstand whatever life throws at us next.

Lastly, thank you to my family and friends who have supported me on this journey. Your words of encouragement, prayers, and love have sustained me these past five years. Thank you to those who in some way or another, provided advice or homework help or computer coding support along the way, and thank you to those who have no idea what I have been up to these last five years, but unwaveringly cheered me on anyway.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Life-saving anti-cancer treatments cause irreversible damage to the ovarian follicular reserve, resulting in infertility for many patients. Development of a translational follicle culture system is stunted by a lack of knowledge on the complex paracrine signaling driving early follicle development. Ovarian follicles are the functional units of the ovary, each containing the oocyte surrounded by specialized hormone-producing cells. During the reproductive lifespan, quiescent primordial follicles are activated in cohorts to join the growing pool. Early-stage follicles are found in the outer 1mm of the ovary (the cortex), while growing follicles expand toward the inner part of the ovary (the medulla) each taking on a multilayered architecture containing the oocyte surrounded by cumulus granulosa, and outer layers of mural granulosa and theca cells separated from the cumulus-oocyte complex by a fluid-filled antrum. The spatial organization of the ovary robustly correlates with its reproductive and endocrine functions but the cellular diversity of the ovary and its complex structures have been difficult to study, largely due to the scarcity of tissue from healthy donors.

Here we performed spatial transcriptomics (ST) and single-cell RNA sequencing (scRNAseq) of ovaries from five reproductive-age women. Using ST, we discovered transcriptional heterogeneity of genes involved in extracellular matrix remodeling (*VIM*, *COL1/8/12/14/16/18*, *TIMP1/2*) and hormone signaling (*NR4A1*, *CEBPD*, *STAR*) in the ovarian cortex. We created a panel of 76 oocyte-specific genes using antibody-guided RNA isolation and developed 96- and 46-gene panels for granulosa and theca cells from antral follicles, providing the

first spatially-driven confirmation of these markers. We also confirmed the gradient of cumulus cell phenotype, where cells near the oocyte showed high expression of protein kinase A signaling genes (*AKAP12, MIF, RDX)* while those near the antrum showed high expression of Wnt signaling genes (*DDIT3, DKK1, LEF1).* scRNAseq revealed four major cell types: stromal, immune, and endothelial cells and pericytes, and four immune cell subtypes: macrophages, and mast, T, and natural killer cells, expanding upon previous studies reporting the cellular makeup of the organ as a whole. While scRNAseq experiments failed to identify rare cell types such as granulosa and theca, we developed an unbiased single cell atlas of the ovary and our ST data provided an unprecedented profile of the rare cell types not identified in scRNAseq.

Utilizing our group follicle culture system we performed time course analysis of murine folliculogenesis *in vitro* and uncovered dynamic expression of genes related to meiosis and steroidogenesis that reflect the carefully orchestrated progression of events observed *in vivo*. We also identified genes related to angiogenesis (*Vegfa, Angpt1/2, Ang*) and Wnt signaling (*Rspo1, Sfrp1/2*) as upregulated in conditions leading to mature oocytes, suggesting these factors are crucial for maturation. Using ligand-receptor analysis of follicular somatic cells and oocytes, we discovered dynamic relationships that support previous studies on bidirectional crosstalk within the follicle. Shotgun proteomics validated our transcriptional findings, leading to a panel of pro-angiogenic factors and Wnt agonists that may serve as targets for exogenous supplementation in a translational follicle culture system.

In conclusion, these studies have uncovered crucial insights into the regulatory mechanisms governing follicle development. Our research identified key pathways, cellular constituents, and novel markers that significantly advance our understanding of ovarian biology.

These findings hold promise for improving follicle culture systems for assisted reproduction and ultimately enhancing the prospects of fertility preservation for a broad range of patients.

Chapter 1 Introduction

Due to improvements in detection and effective anti-cancer treatment, the number of cancer survivors has been increasing in recent decades[1]. These life-saving treatments, however, have lasting effects on patient health and quality of life. Anti-cancer treatments have cytotoxic effects on the ovaries, leading to premature ovarian insufficiency (POI) and infertility in many patients, as the ovarian follicle reserve is non-renewable[2]. Fertility preservation options for patients are limited, with the standard of care being hormone stimulation followed by egg retrieval and cryopreservation of eggs or embryos for later use[3]. But this procedure is contraindicated for patients who cannot delay treatment or whose cancer is hormone responsive, such as breast cancer. Even more significantly, ovarian stimulation cannot be offered to prepubertal girls, because prepubescent ovaries cannot respond to hormonal stimulation. Autotransplantation of cryopreserved ovarian tissue has been successful in adult patients and led to live births, but the risk of re-introducing malignant cells to the body, especially in cases of blood-borne cancers, such as leukemia, significantly limits the options available for pediatric patients[4]. As such, *in vitro* culture of ovarian follicles to obtain mature gametes is a safer method to provide fertility options to cancer survivors without treatment delays, hormone stimulation, or risk of cancer recurrence.

Ovarian folliculogenesis encompasses a wide breadth of cellular processes beginning in the dormant, primordial follicle and culminating in a fully mature and developmentally competent oocyte that is ovulated and becomes available for fertilization. This highly ordered process is influenced by a milieu of factors from various physiological domains, particularly the hypothalamic-pituitary-gonadal (HPG) axis. The HPG axis governs the secretion of hormones that

cause cyclical changes in the reproductive organs. The secreted hormones, beginning with gonadotropin-releasing hormone (GnRH) from the hypothalamus, as well as downstream hormones follicle-stimulating hormone (FSH), luteinizing hormone (LH), estrogen, progesterone, and others, become particularly important during the preovulatory stage of folliculogenesis, after the follicle's antrum has formed and the oocyte has matured. However earlier stages of folliculogenesis can proceed independent of gonadotropins and are instead dependent on intrafollicular signaling between the oocyte and the somatic cells present. Many studies of murine follicles also highlight the importance of extracellular signaling factors within the follicle microenvironment and elucidating the mechanisms by which intra- and inter-follicular cytokines initiate and sustain complex paracrine and autocrine signaling relationships that govern the various processes of folliculogenesis until gonadotropins gain predominate influence[5]. The complexity of intrafollicular crosstalk can be attributed in part to the follicle's 3D architecture and the juxtaposition of the cells within the follicle. As the follicle expands throughout folliculogenesis, the somatic cells acquire distinct phenotypic characteristics due to the gradient that develops via diffusion of systemic factors into the follicle and oocyte-derived factors outward from the oocyte. The importance of these gradients in developing cumulus and mural granulosa cell lineage have been previously studied and there are perhaps many other processes influenced by these intrafollicular gradients that have yet to be uncovered[6].

Most of the follicles residing in cryopreserved ovarian cortical tissue are at the primordial and primary stages, and thus comprise the main target for *in vitro* culture. However, culture of small early-stage follicles remains challenging due to the inability of current culture systems to supply the essential factors to individual primordial and primary follicles that would otherwise be provided through paracrine interactions in the native ovarian stroma and neighboring follicles *in*

vivo[7]. Detailed understanding of the soluble factors critical for development of early stage follicles and signaling pathways necessary for proper development is essential for the development of well-defined methods for follicle culture.

To develop a translational human follicle culture system, the scientific community needs a similarly detailed understanding of the factors critical for development of early stage human follicles. Progress in this area is hampered by the shortage of human tissues for available for research and the differences in the ovarian landscape between age groups—reproductive-aged tissues are more rare than those of postmenopausal women, whose follicle reserves have been depleted. To maximize knowledge gained from precious tissue samples, the research community has employed next-generation sequencing technology such as single-cell RNA sequencing (scRNAseq). scRNAseq is a powerful tool for identifying rare cell populations, understanding the relationships between genes, and tracking cell lineage during differentiation or phenotypic change. One study profiled the trajectories of granulosa and theca cells isolated from antral and preantral follicles from the ovarian medulla, but the tissue samples used in this study excluded dormant primordial follicles, which constitute the ovarian reserve and are most important for fertility preservation[8]. Another recent study used scRNAseq to identify somatic cell populations in the ovary, including distinct immune cell populations and endothelial and smooth muscle cells[9]. Importantly, the majority of tissues samples obtained for this study came from patients undergoing gender reassignment procedures who were on long-term androgen therapy, which has unknown effects on the follicle reserve and the ovarian stroma[9]. The recent emergence of ST also presents a previously unattainable opportunity to profile the cell types of complex spatially-defined structures in the ovary without sacrificing their spatial context. No published studies have yet detailed scRNAseq and ST approaches for analyzing the cell populations of both the cortical and medullary

compartments of the ovary alongside follicles of varied developmental stages, including their oocytes. Progress in this area will identify cell types, signaling pathways, and molecular drivers of follicle development towards development of a standardized follicle culture system for the clinic.

In this dissertation, we present a cohort of biological processes and extra-follicular signals that support ovarian follicle development, in mice and in humans. By applying omics technologies to a biomimetic follicle culture system, we discovered gene- and protein-level drivers of follicle development *in vitro*, then applied scRNAseq and ST to rare human ovarian samples towards developing a similarly comprehensive network of factors and processes essential for follicle maturation. In Chapter 2, we used microarray technology to identify genes and pathways in follicular somatic cells associated with successful follicle development *in vitro*. In Chapter 3, we expanded on this work to evaluate follicles cultured *in vitro* using RNA sequencing and shotgun proteomics, identifying somatic cell- and oocyte-specific pathways essential for follicle maturation, then validated these findings with protein-level comparisons of follicles from successful and suboptimal culture conditions. In Chapter 4, we turned to the human ovary and its cellular constituents to identify cell types and molecular mechanisms that support human follicle development, towards designing a translational human follicle culture system for the clinic. We summarized the implications of this work and its future directions in Chapter 5. This work is significant in that it integrates tissue engineering approaches for follicle culture with computational analysis of the ovary towards developing an indiscriminate fertility preservation option that maximizes the use of cryopreserved ovarian tissue.

Chapter 2 Capitalizing on Transcriptome Profiling to Optimize and Identify Targets for Promoting Early Murine Folliculogenesis *in vitro*

Previously published as: *Sci Rep*, 2021, **11**, 12517 (DOI: 10.1038/s41598-021-92036-y)

2.1 Preface

The identification of factors driving successful follicle development from the primary stage remains a formidable challenge in the development of a translational follicle culture system. In this chapter, we examine the follicular somatic cell transcriptome using microarray analysis and establish pathways and soluble factors that support follicular growth and maturation *in vitro*. Through meticulous examination of gene expression, we outline biological processes driving the progression of follicle development from the primary to antral stages, and specify pathways associated with successful maturation. We close with a comparison to follicles isolated from the *in vivo* microenvironment and cellular processes that differ in the *in vitro* vs. *in vivo* settings.

2.2 Introduction

An inadequate understanding of the paracrine factors that support early folliculogenesis hampers development of a well-defined, translational follicle culture system. Without protocols for isolating, culturing, and studying ovarian follicle development from the primary stage *in vitro*, it is difficult to identify these soluble signals and evaluate their biological significance. One option to address this challenge is to co-culture follicles with "feeder cells" such as murine embryonic fibroblasts (MEFs), stromal cells, ovarian mesenchymal cells, and granulosa cells that secrete factors to promote follicle growth and survival[10–13]. Feeder cell culture, however, has limited

translational potential and fails to isolate the most salient signaling factors needed for folliculogenesis. Alternatively, co-culture of primary murine follicles in groups has been shown to significantly improve growth and survival when compared to culture of individual primary follicles, with groups of 10 (10X) cultured together being more successful than groups of 5X[14]. This system also has limited translational potential, but serves as a model system for studying follicular crosstalk and paracrine signaling without cues from non-follicular feeder cells.

A recent study by our research group probed the secretome and transcriptome signatures of follicles that were co-cultured as 10X and 5X with a goal of identifying factors present in 10X that drive successful folliculogenesis[15]. Follicles cultured in groups of 10 had upregulated activity of key transcription factors as well as a unique secretory signature, revealing some of the biological processes and pathways essential for early follicle development. Specifically, using a targeted approach, this study determined that the mechano- and oxygen-responsive transcription factors, NF-κB and HIF1, were upregulated and correlated with the antrum formation in 10X but not 5X follicle cultures. This work, paired with recent unsupervised transcriptomic analysis of later stages of follicular development[6,16,17], suggests an unparalleled opportunity to profile global transcription activity of early stage follicles cultured in groups of 5 and 10 in a controlled *in vitro* environment. While follicles in the 5X condition survive and grow, they yield oocytes at a significant lower quality and quantity compared to those from the 10X condition. We hypothesized that unbiased transcriptome comparison of follicles grown in inadequate (5X) and adequate (10X) microenvironments will identify key mediators of successful early folliculogenesis. Here we performed microarray analysis of somatic cell samples from primary murine follicles co-cultured in 5X and 10X, each over a 12-day time course, to identify upstream regulators of follicle growth and maturation that contribute to successful follicle culture *in vitro* and to further characterize the

unique synergism of successful culture of follicles in larger groups. The analytical approaches undertaken sought to answer three key biological questions: 1) what temporal transcriptional signatures drive folliculogenesis *in vitro*, 2) what differentially regulated genes and pathways drive better survival, growth, and maturation in the 10X condition, and 3) how do the *in vitro* and *in vivo* follicular transcriptomes compare on a global scale.

2.3 **Experimental Methods**

2.3.1 Follicle Isolation, Encapsulation, and Culture

Whole mouse ovaries were enzymatically digested to harvest large numbers of primary follicles. Enzymatic follicle isolation was performed as previously described[15]. Briefly, ovaries were taken from female C57B6 X CBA/J mice ages 10-12 days after being anesthetized with isofluorane and euthanized. All animal procedures were performed in compliance with the Guidelines for the Care and Use of Animals at the University of Michigan and approved by the Animal Care and Use Committee at the University of Michigan (PRO00008465). After removing the surrounding tissue, ovaries were washed in warm Leibovitz's L-15 medium (L-15) (Thermo Fisher) and transferred to a dish with pre-equilibrated alpha modification of minimum essential medium (αMEM) (Thermo Fisher) with 0.5% (v/v) PenStrep (Thermo Fisher). 10% (v/v) Liberase DH at 13 Wünsch units/mL (Sigma-Aldrich) was added to the dish of ovaries and gently mixed. Ovaries were incubated undisturbed at 37°C for 35-45 minutes followed by 5 minutes of pipetting to break up the enzymatically-digested tissue into individual follicles. The digest was then arrested using 10% (v/v) fetal bovine serum (FBS) (Thermo Fisher).

Follicle encapsulation was performed as previously described[15]. Briefly, primary follicles with diameters ranging 90-110 µm were chosen from the digest and encapsulated in 10 µL alginate beads in either 5 (5X) or 10 (10X) follicles per bead. Beads containing follicles were dropped into

a solution of 50 mM $CaCl_2$ and 140 mM NaCl and allowed to crosslink for 2 to 3 minutes. Alginate beads were cultured in 96 well plates with 150 µL growth media (GM) (α-MEM supplemented with 3 mg/mL bovine serum albumin (BSA, MPBiomedicals, Irvine, CA, USA), 1 mg/mL bovine fetuin, 5 µg/mL insulin, 5 µg/mL transferrin, 5 ng/mL selenium (ITS, Sigma, St. Louis, MO, USA), and 10 mIU/mL highly purified, human-derived, follicle-stimulating hormone (FSH) (Urofollitropin, Ferring Pharmaceuticals, Saint-Prex, Switzerland)) per well. Because of the manual follicle encapsulation, on average 10-20% of hydrogels will not fully encapsulate exactly 5 or 10 follicles. In this case, those samples were removed from culture and could not be used for microarray analysis. By doing so, we developed samples that reflected only a true 5X or 10X transcriptome signature, where each follicle was experiencing the same *in vitro* culture conditions as the other follicles that it was pooled with. Every 48 hours after encapsulation, half of the growth medium (75 µL) was exchanged for fresh medium. Follicle development was tracked by imaging follicles every 48 hours before exchanging growth medium. Follicle diameter from each 48 hour time point was measured using ImageJ by taking the average of two perpendicular measurements across the follicle. 5X or 10X hydrogels with any follicles that fell out of the alginate hydrogel onto the bottom of the well, or hydrogels with follicles dead at the start of culture, were excluded from analysis and microarray sample collection. Follicles were termed “dead” if the oocyte was extruded more than 50% from the follicular structure with no surrounding granulosa cells, or if the follicular structure was dark and the oocyte was not visible during imaging. Only follicles that survived to Day 12 of culture were included for growth analysis as shown in Figure 2.1D. The results presented were pooled from 9 separate culture experiments, with cells from 430 follicles cultured in 10X and 380 follicles cultured in 5X collected to extract RNA. The variations in follicle

numbers and total RNA content across samples do not impact the downstream analyses outlined below because of the amplification employed prior to microarray.

2.3.2 Somatic Cell Isolation, RNA Extraction, and Microarray Analysis

Follicles (n = 380 total for 5X and n = 420 total for 10X) were taken from culture for somatic cell isolation at the following time points: Days 2, 4, 6, 7, 8, 10, and 12. The number of follicles isolated for each microarray sample are outlined in Appendix Table A.1. To isolate somatic cells, alginate beads were digested by incubating with 10 IU/mL alginate lyase for 30 minutes. Follicles were then transferred to warm L-15 and oocytes were mechanically removed from the follicles using insulin syringes (BD 305620). The cells were centrifuged at 300g for 5 minutes, supernatant was removed, and the remaining cell pellet was snap frozen in liquid nitrogen. The process of follicle encapsulation, culture, and cell isolation is visualized in Figure 2.1A. RNA was extracted using the Qiagen RNEasy Micro Kit® following the manufacturer's instructions. Extracted RNA was stored at -80°C before being submitted to the University of Michigan's Advanced Genomics Core. The RNA samples were analyzed on an Affymetrix Mouse Gene ST 2.1 plate using the Affymetrix GeneChip WT-Pico Kit. The chip was run on a GeneTitan Multi-Channel instrument at the Advanced Genomics Core. The process of follicle encapsulation, culture, and cell isolation is outlined in Figure 2.1A.

2.3.3 Oocyte Maturation in vitro

Oocyte maturation in vitro was performed as previously described[18]. In four separate 5X and 10X culture experiments performed as outlined above, follicles were cultured for 10 to 12 days, at which point follicles with a diameter of at least 280 μm were removed from the alginate beads by incubating in 10 IU/mL alginate lyase in media (n = 45 follicles for 5X and n = 320

follicles for 10X). Most 5X follicles in culture, which did not reach this terminal size, were not subjected to *in vitro* maturation (IVM). This contributes to the large difference in the sample size between 5X and 10X. Individual follicles were aspirated from the culture plate, rinsed in warm media, and placed in maturation media (α-MEM supplemented with 10% fetal bovine serum, 5 ng/mL epidermal growth factor, 1.5 IU/mL human chorionic gonadotropin, and 10 mIU/mL FSH) for 20 hours. After incubation, the expanded cumulus cells were removed by adding 0.03% hyaluronidase to each plate. Oocytes were aspirated and any remaining adherent granulosa cells were removed mechanically by pipetting. Oocytes were imaged at 400X and were classified as germinal vesicle/degenerate (GV/D), metaphase I (MI), or metaphase II (MII) based on their morphology. From the total follicles subjected to IVM (n = 45 follicles for 5X and n = 320 follicles for 10X), 39 (5X) and 204 (10X) oocytes were successfully isolated from their follicles to evaluate maturation state. The discrepancy between total follicles subjected to IVM (45 and 320 respectively) and number of oocytes imaged for IVM (39 and 204 respectively) is due to either a) the oocyte being damaged or broken due to pipetting during transfer or b) inability to locate the oocyte or dislodge it from the follicle for imaging.

2.3.4 Computational Methods for Transcriptome Data Normalization and Filtering

Raw CEL files were loaded into R (Version 3.6.1) using the oligo package[19,20]. Raw expression data was normalized using the RMA algorithm (Appendix Figure A.1), then gene expression data was filtered based on normalized intensity (> 3) and variability (standard deviation > 0.3) of expression across samples, reducing overall gene count from 41,345 to 13,313[21–25]. All gene expression data and subsets of the dataset were mean-centered for analysis and visualization unless otherwise noted. Heat maps were produced using the package pheatmap[26]. All gene annotation was performed using the package AnnotationDbi[23].

2.3.5 Principal Component Analysis of Transcriptome Data

All fourteen microarray samples (n = 13,313 genes per sample) were subjected to principal component analysis and plotted in the PC space as shown in Figure 2.2A. PC1 accounts for 24% of the variance among samples while PC2 accounts for 12%. A line was plotted in the PC space to connect adjacent time points and highlight the trajectory of 5X and 10X samples over the course of follicle maturation.

2.3.6 Temporal Analysis of Transcriptome Data

The combined 13,313 gene dataset (5X and 10X samples together) was subjected to k-means clustering (k = 6) to group genes based on their temporal expression profile by combining the 5X and 10X samples (Figure 2.2B). Genes in these six temporal clusters were individually subjected to differential expression (DE) analysis using limma[21], with the High and Low groups defined separately for each cluster according to the observed patterns (described in Appendix Table A.2). Differentially expressed genes (DEGs) with their associated p-value from limma analysis were submitted to a directional gene ontology (GO) term analysis using *LRPath*[27–29], focusing on GO terms with 50 to 500 genes and using a p-value cutoff of 0.05. DEGs were also submitted to *GOrilla*[30,31] for biological process GO term analysis, using DEGs as an unranked list of target genes (not accounting for p-value) and all 13,313 genes as a background list, with a p-value threshold of 0.01. Only genes with ontological annotation were retained from these analyses, excluding predicted genes.

2.3.7 Gene Filtering for Trajectory Smoothness and Differential Transcriptome Patterns between 5X and 10X

To analyze genes with strong dynamic changes, one could simply select those with large variance over the 14 samples. However, a gene may show a large variance simply by having stochastic "flickers" over the 7-point time series without cohesion among adjacent points or having a systematic difference between the 5X and the 10X groups. To rank genes by their relative cohesion, or smoothness, along the temporal trajectories we implemented two metrics. The first is the Percent Variance Explained (PVE) by fitting the 7-point gene expression values (y) with a cubic function, $y = ax^3 + bx^2 + cx + d$, where x is coded as (-3,-2,-1,0,1,2,3) and a-d are coefficients. PVE is defined as $PVE = V_C/V_T$, where V_C is the variance explained the cubic function fit and V_T is the total variance for the gene. Higher order functions (such as by adding a x^4 term) would increase the potential ruggedness of the fitted curves and allow less-smooth genes to be selected. For each gene, PVE was calculated for the 5X and 10X series separately and then averaged. The second metric is a measure of how pairs of adjacent time points are more similar than pairs of more distant time points. For each gene, we calculated a 7-by-7 "distance" matrix for all pairs of absolute differences. We then defined Cohesion as $C = SSD_A/SSD_{N-A}$, where SSD_A is the sum of squares of the six adjacent Δ values for the gene and $SSD_{N\text{-}A}$ is the sum of squares of the fifteen non-adjacent Δ values. PVE and C are highly correlated (not shown), as they capture similar characteristics, i.e., relatively smooth changes rather than random spikes not "supported" by adjacent time points.

To rank genes by differential patterns between the 5X and 10X series we ran Two-factor Analysis of Variance (ANOVA), where the 14 samples are arranged by a Group Factor (two levels: 5X, 10X) and a Time Factor (7 levels). We extracted the F scores for both the main effects and the interaction effect and focused on selecting those with a large F score for Interaction, F_{int}. We used the three aforementioned metrics to retain genes satisfying PVE > 0.6, C < 0.35, and F > 2,

narrowing the total set of 41,345 initial, unfiltered genes to 6,119 genes. This gene list of interest was clustered using a 3x3 self-organizing map[32] producing 9 gene subsets with differing temporal expression patterns. These gene subsets were subjected to GO analysis using *GOrilla*[30,31], filtering for gene ontology terms with less than 500 genes.

2.3.8 Comparisons to Previously Published Transcriptome Data

To compare the *in vitro* follicle culture data to previously published microarray data from freshly isolated follicles of different stages, raw expression data (GSE 97902) from the Gene Expression Omnibus was downloaded using the package GEOquery[31,33]. This dataset reports ovarian follicle gene expression at various stages of development, using RNA collected from freshly isolated follicles from CD-1 mice[33]. From this dataset, the three replicates each of primary, multilayer secondary, and antral follicle transcriptome data were extracted for comparisons to the *in vitro* co-culture data. 11,424 genes were expressed in both datasets (freshly isolated follicles from GSE 97902 or "*in vivo*" and somatic cells from *in vitro* co-culture in 5X and 10X). The datasets were normalized together using quantile normalization from the package preprocessCore[34] (Appendix Figure A.1). To compare gene expression between the *in vivo* and co-culture datasets while still acknowledging the inherent differences between the datasets due to different microarray chip platforms and thus differences in raw intensity reads, the two datasets were separately subjected to DEG analysis using limma[21], with co-culture data being grouped into "Early", "Middle", and "Late" culture as outlined in Figure 2.4A. The mean expression for each gene between 5X and 10X was used as the gene expression value for comparisons to the *in vivo* dataset. From the DEG analysis using limma[21], each t-score was collected and used to represent an upregulation (t-score > 0) or downregulation (t-score < 0) in the gene from one stage of development to the next (Early to Middle, Middle to Late, and Early to Late). Agreement and

disagreement in t-scores was visualized in Figure 2.4B. Genes with t-score agreement between the *in vivo* and co-culture dataset were collected and subjected to *GOrilla*[30] analysis as described above.

2.4 **Results**

2.4.1 Maturation Characteristics of Longitudinal Follicle Co-Culture Experiments

Follicles encapsulated in alginate hydrogels in 5X and 10X conditions were maintained for 12 days (Figure 2.1A). They grew at different rates, reaching a mean diameter of 212.36 ± 10.04 and 300.02 ± 8.00 μm (mean ± SEM) respectively (Figure 2.1B,D), corroborating previous findings[14,15]. There were no statistical differences in follicle survival rate between 5X and 10X conditions (Figure 2.1C). However, maturation of the follicles cultured in groups of 10 that reached a diameter ≥ 280 μm resulted in 50% of the oocytes resuming meiosis and reaching metaphase II stage, which was significantly greater ($p < 0.0001$) than the maturation rates of ~5% for oocytes from follicles cultured in 5X (Figure 2.1E,F).

2.4.2 Trajectories of the 5X and 10X Transcriptomes during Culture

We analyzed 14 RNA samples representing both experimental groups (5X and 10X) during a 7-point time course each, sampled at a regular interval of two days. Each sample contained 40-70 follicles (Appendix Table A.1). While all other samples were collected on even days of culture (2, 4, 6, 8, 10, 12), we included the Day 7 samples because of the dramatic phenotypic changes observed between Days 6 and 8 *in vitro*, including antrum formation and cell proliferation. By adding this additional time point we hoped to capture transcriptional changes at this crucial point in development. After data normalization and filtering for the most variable 13,313 genes, the transcriptional data for the 14 samples were subjected to an unsupervised principal component

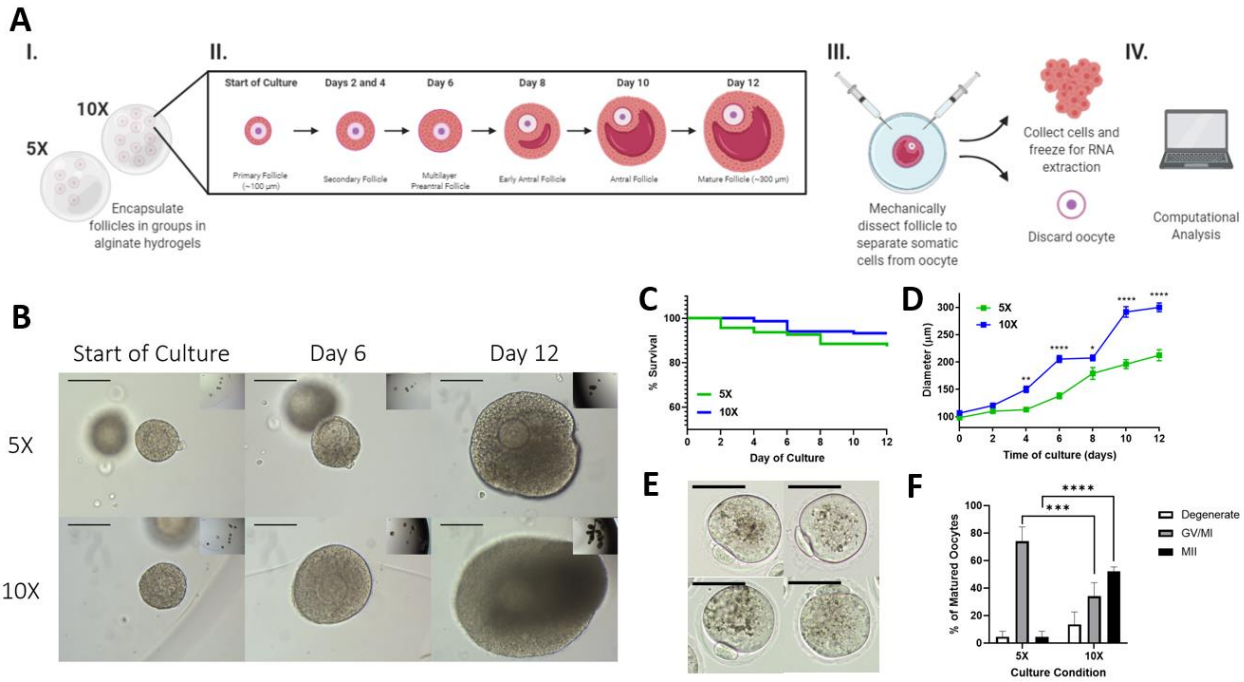


Figure 2.1: Ovarian follicles cultured in 10X have superior outcomes when compared to 5X culture. A) Schematic describing the process of somatic cell isolation, beginning with primary follicles cultured in alginate hydrogels over 12 days, during which follicles are dissected at two-day intervals for cell collection, B) bright field images of individual primary follicles encapsulated in alginate hydrogels in either groups of 5 or 10 (scale bar = 100 µm), C) survival rates of follicles over duration of culture in 5X and 10X, D) growth rates of follicles over duration of culture in 5X and 10X, and E) metaphase II oocytes from follicles cultured in groups (scale bar = 50 µm), and F) in vitro maturation rates for follicles cultured in 5X and 10X. N = 95 follicles for 5X and N = 140 for 10X. N = 39 follicles for 5X and N = 204 for 10X for Fig. 2F. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Figure 2.1A was generated using biorender.com.

analysis (PCA) (Figure 2.2A). In the principal component (PC) space, both 5X and 10X samples travel along PC-1 initially, and then along PC-2 towards the end of culture, reflecting transcriptional similarities between the two culture conditions. The paths of 5X and 10X diverge most dramatically at Day 4 (Figure 2.2A). The trajectories seen in the PC space guided our approach for evaluating differences between the 5X and 10X conditions, where 5X appears to be slower to progress through the middle stage when compared to 10X (Figure 2.2A). With this in mind, our comparisons of the 5X and 10X culture systems (next section) sought to identify genes and pathways that lag behind in 5X when compared to 10X.

2.4.3 Genes and Pathways with Distinct Temporal Patterns

The expression levels for the 13,313 genes were standardized for each gene and shown for the 14 samples in Appendix Figure A.2. To find groups of genes with shared dynamic patterns during the 7-point time series, without accounting for the 5X vs. 10X condition but instead evaluating the *in vitro* microenvironment as a whole, we applied k-means clustering (k = 6) to define gene clusters according to their expression levels in the 14 independent samples (for 5X and 10X, 7 from each condition). In Figure 2.2B, the 5X and 10X series are averaged, resulting in a consensus 7-point profile, ordered by time and separated by the six gene clusters. For genes within each cluster we performed DE analysis to identify those with the most salient changes, with the comparison groups defined for each cluster based on its own patterns (Appendix Table A.2). For example, for the 4th gene cluster (indicated in blue in Figure 2.2B), DE analysis was made by comparing the middle five time points as the High group with the first and last points combined as the Low group. Among the six clusters, four had strongly concerted patterns of up- and down-regulation; while the other two, the 2nd and 3rd cluster, had undefined oscillatory changes throughout the culture. We therefore focused our DE results interpretation on the four clusters with distinct dynamic changes, at both the gene level and pathway level. To identify DE genes, we analyzed each gene using limma[21], and ranked genes within each cluster by log-fold change (logFC). For pathway analysis, DE-ranked genes in a cluster were subjected to GO analysis using *LRPath*[27,28] and *GOrilla*[30], with a focus on Biological Process GO terms. The first cluster (n = 1826 genes, labeled with magenta) contains "early activators": genes with high expression at the beginning of culture, and are downregulated as time in culture progresses (Figure 2.2B). Genes of interest in the top 100 genes according to logFC in this cluster include *Amh* (coding for anti-Müllerian hormone) and *Igfbp5* (coding for insulin-like growth factor-binding peptide 5), both of which have known roles in granulosa cell survival and proliferation[35,36] (Appendix Figure A.3A).

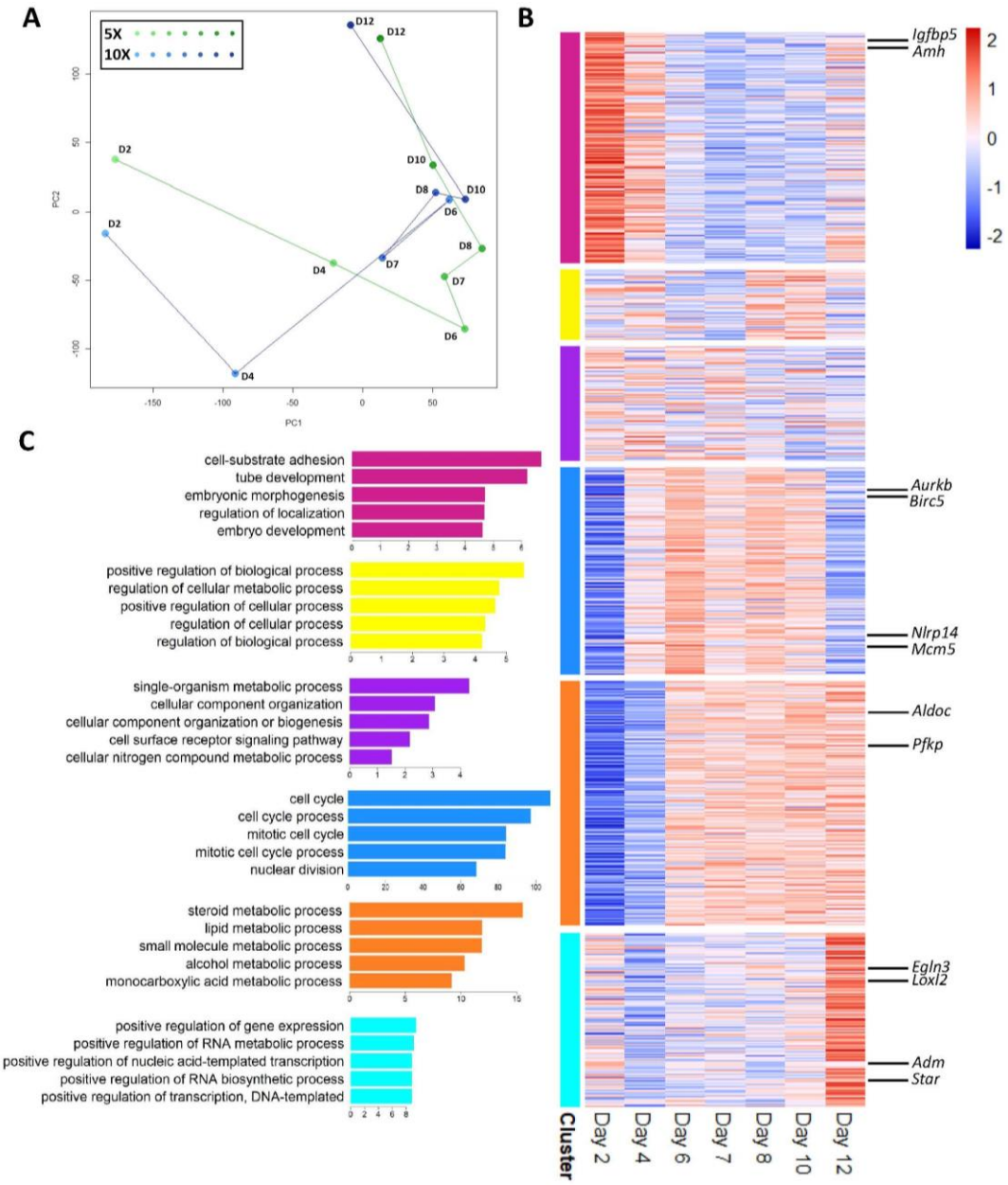


Figure 2.2: Somatic cells from ovarian follicles cultured in groups have dynamic temporal gene expression. A) Principal component plot of the 14 microarray samples, B) heatmap of standardized expression level for genes in the six clusters obtained from k-means analysis (k=6), and C) top five GO terms enriched for differentially expressed genes in each cluster from LRPath analysis (x-axis = -log(P Value)). Figure 2B was generated using pheatmap (version 1.0.12) in R (version 3.6.9, https://www.r-project.org/). Figures 2A and 2C were generated in R (version 3.6.9, https://www.r-project.org/).

The top GO terms for this cluster include "cell-substrate adhesion" (GO:0031589, 53 genes), "embryonic morphogenesis" (GO:0048598, 71 genes), and "embryo development" (GO:009790, 116 genes) (Figure 2.2C). Genes of interest within the cell-substrate adhesion GO term include

genes coding for angiopoietin 1 and 2, collagen types I and VIII, and genes coding for laminin and matrix metalloproteases. GOrilla analysis of this cluster also showed enrichment for terms associated with cell adhesion and migration, cell proliferation, and signal transduction and metabolic processes. The fourth cluster (n = 1630 genes, shown in blue in Figure 2.2B) mainly contains genes with delayed upregulation on Day 4 and a decline in regulation between Days 10 and 12. Genes of interest with high logFC in this cluster include *Aurkb*, *Birc5*, and *Mcm5*, which code for aurora kinase B, baculoviral inhibitor of apoptosis repeat-containing 5 (BIRC5), and the protein MCM5 respectively, all of which have known roles in the cell cycle and apoptosis, and *Nlrp14*, coding for the protein NOD-like receptor family pyrin domain containing 4 (NLRP14), which has been studied in relation to primordial follicle survival and may have other roles in later stages of folliculogenesis[37,38] (Appendix Figure A.3C). All five top GO terms for this cluster were associated with cell cycle activity: "cell cycle" (GO:0007049, 267 genes), "cell cycle process" (GO:0022402, 217 genes), "mitotic cell cycle" (GO:0000278, 173 genes), "mitotic cell cycle process" (GO:1903047, 165 genes), and "nuclear division" (GO:0000280, 147 genes). This period of upregulation from Days 4 to 10 reflects the proliferation of granulosa cells during follicular expansion, which ceases once the follicle has reached a terminal size (Figure 2.2C). *GOrilla* analysis corroborated these findings, also reporting enrichment for terms related to DNA replication and the cell cycle. The fifth cluster (n = 1928 genes, labeled in orange), was comprised of genes with delayed activation, as they were upregulated starting on Day 6 (Figure 2.2B). Genes of interest within this cluster included glycolysis-related genes like *Aldoc* (coding for aldolase C) and *Pfkp* (coding for phosphofructokinase), noteworthy because granulosa cells are known to have highly glycolytic metabolic processes often regulated by oocyte-secreted factors[39] (Appendix Figure A.3D). The top GO terms for this cluster from *LRPath* were all related to metabolism:

"steroid metabolic process" (GO:0008202, 54 genes), "lipid metabolic process" (GO:0006629, 146 genes), "small molecule metabolic process" (GO:0044281, 201 genes), "alcohol metabolic process" (GO:0006066, 50 genes), and "monocarboxylic acid metabolic process" (GO:0032787, 75 genes) (Figure 2.2C). The last cluster (n = 1378 genes, labeled in cyan in Figure 2.2B) was comprised of genes with late activation in the final days of culture. Genes of interest in this cluster included *Adm* (coding for adrenomedullin) and *Star* (coding for steroidogenic acute regulatory protein), both of which drive steroidogenesis in the follicle, and *Egln3* (coding for Egl-9 family hypoxia inducible factor 3) and *Loxl2* (coding for lysyl oxidase like 2), which are both hypoxia-related[40,41] (Appendix Figure A.3E). Top GO terms from *LRPath* and *GOrilla* for this cluster included "positive regulation of gene expression" (GO:0010628, 86 genes) and "positive regulation of transcription, DNA-templated" (GO:0045893, 66 genes) (Figure 2.2C).

2.4.4 Comparison between 5X and 10X Transcriptome Trajectories

While the results above focused on 5X-10X shared patterns, we also contrasted the 5X and 10X transcriptomes to identify genes and pathways with differing patterns of expression. The two datasets were filtered to select genes with dynamic temporal expression and high cohesion across the time series, thus eliminating genes with stochastic "spikes" in expression. The resulting set of 6,119 genes were clustered into nine subsets using a 3x3 self-organizing map[32], then subjected to gene ontology analysis of each subset (Figure 2.3). The gene subsets showed three general patterns of dynamic change: 1) genes with opposite temporal expression patterns between 5X vs. 10X at the beginning, the end, or both (Figure 2.3A), 2) genes with largely similar temporal expression patterns in 5X and 10X (Figure 2.3B), and 3) genes with dynamic expression in the 10X condition in contrast to a more static expression in the 5X condition (Figure 2.3C). These gene subsets were

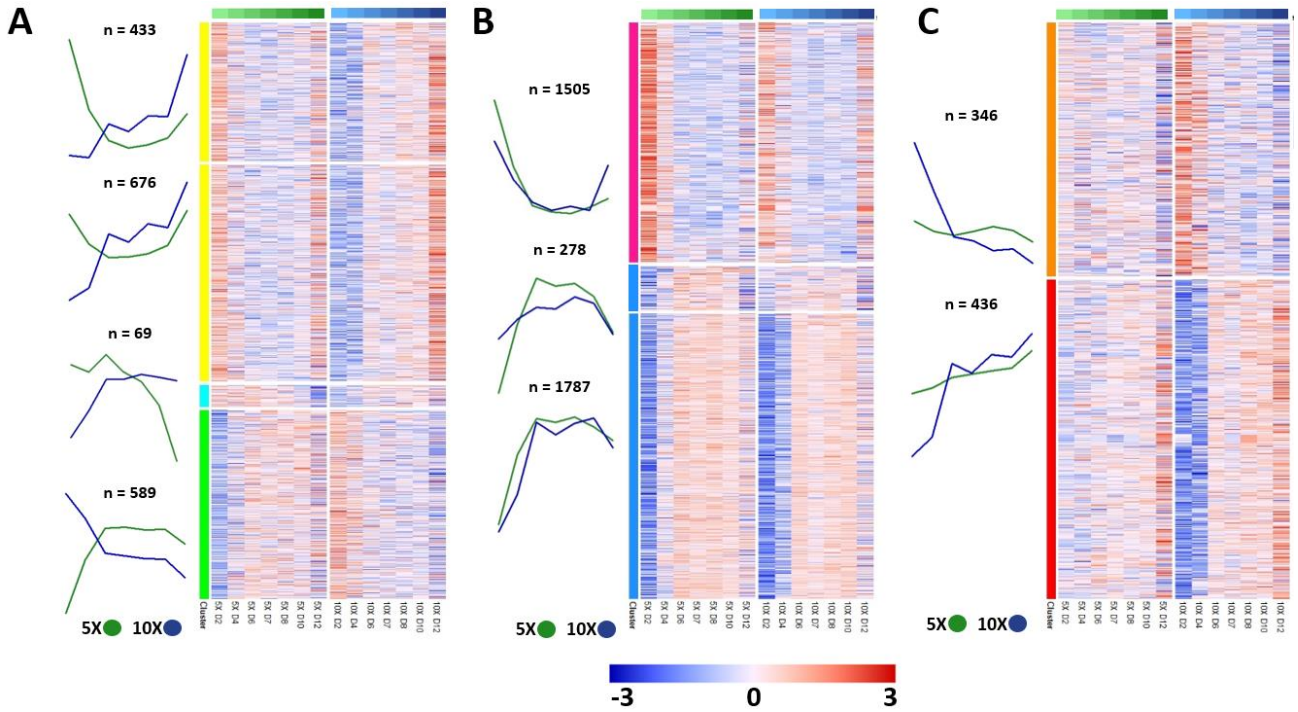


Figure 2.3: Somatic cells from follicles cultured in 5X vs. 10X have distinct gene expression profiles in nine gene subsets. Mean expression plots with gene number and per-gene heatmaps for A) gene subsets with different expression in 5X vs. 10X, B) gene subsets with similar expression in 5X vs. 10X, and C) gene subsets with more dynamic expression in 10X vs. 5X. All plots were generated using pheatmap (version 1.0.12) in R (version 3.6.9, https://www.r-project.org/).

individually subjected to gene ontology analysis, with the exception of the top two subsets in Figure 2.3A, which were merged into a single group due to their similar mean expression patterns.

In the top two subsets in Figure 2.3A (marked in yellow) the gene transcription of follicles cultured in 5X followed a "U-shape" form, starting higher at Day 2 of the culture, dropping to the minimum after 4 days and climbing back up towards the end of the culture. In contrast, the gene transcription of follicles cultured in 10X continuously increased from Day 2 to Day 12. These genes include those related to MAPK signaling, prostaglandins, *Cyp11b1*, which is involved in the biosynthesis of cortisol, and *Star*, which is involved in progesterone synthesis in the follicle. The yellow subset showed enrichment for genes related to calcineurin-NFAT signaling (GO::0051534 and GO::0106057). In the bottom subset (marked in green) the gene transcription of follicles cultured in 5X followed an inverse "U-shape" form, starting low at Day 2 of the culture, climbing to the maximum after 4 days and slightly decreasing towards the end of the culture. In contrast,

the gene transcription of follicles cultured in 10X monotonically decreased from Day 2 to Day 12, following a similar trajectory to 5X in the last few days of the culture. The green subset included genes related to angiogenesis (angiogenins and angiopoietins), as well as Wnt signaling and was enriched for gene ontology terms on G protein-coupled receptor signaling (GO::0007186) and sensory perception of chemical stimulus (GO::0007606). In the middle subset (cyan) the gene transcription of follicles cultured in 5X dramatically decreased since Day 4 of the culture and reached its lowest at the end of the culture. In contrast, the gene transcription of follicles cultured in 10X slowly increased from Day 2 to Day 12, presenting an opposite trajectory to 5X. The cyan subset included prostaglandins and was enriched for terms related to lactation (GO::1903489 and GO::1903487) and JAK/STAT signaling (GO::2000366) (Supplemental Dataset S3).

Among the genes with similar temporal expression in 5X and 10X (Figure 2.3B), the magenta subset was enriched for gene ontology terms related to transmembrane transport and locomotion (GO::0034765, GO::0034762, and GO::0040012). The two blue subsets in Figure 2.3B were enriched for genes related to the cell cycle (GO::0007049, GO::0022402, GO::1903047, GO::0051301, and GO::0051726). The third group of genes with more dynamic expression in 10X compared to 5X (Figure 2.3C) was enriched for terms related to G-protein coupled receptor activity (GO::0007186) and sensory perception of chemical stimulus (GO:: 0007606) in the orange subset, and terms related to protein modification by small protein conjugation (GO::0032446) and protein ubiquitination (GO::0016567) in the red subset (Supplemental Dataset S3).

2.4.5 Comparisons of Transcriptome Data between Freshly Isolated and in vitro Cultured Follicles

To compare the transcriptome profile of somatic cells from follicles co-cultured in vitro to those from freshly isolated follicles in vivo at comparable stages of development, we drew on

previously published data available online on the Gene Expression Omnibus (GSE 97902)[33,42,43]. This published dataset was used to develop a metabolic model for murine ovarian follicles by updating the previously published dataset Mouse Recon 1 to include human homologues and key ovarian follicle development pathways that were not previously represented[33,44]. The dataset used during model development includes transcriptome data from the follicular structure as a whole (including the oocyte and somatic cells) at various stages of follicle development[33]. While the experimental approach used in collecting these samples inevitably differs from the methods we used, parallels can be drawn in follicular development as shown in Figure 2.4A. To compare transcriptional profiles between the two datasets, the "Primary" data from GSE 97902 was considered equivalent to the in vitro Day 0 and Day 2 samples, and subsequently the "Multilayer Secondary" and "Antral" datasets from GSE 97902 were considered equivalent to Days 4 through 8 and Days 10 and 12 respectively (Figure 2.4A). After normalization and filtering, 11,424 genes were shared between the two datasets (Appendix Figure A.4). To compare temporal gene expression patterns between the *in vitro* and *in vivo* datasets, differential gene expression analysis was performed for each developmental transition (Early to Middle, Middle to Late, Early to Late) for the two datasets separately, where the gene expression values for the in vitro dataset were taken as the average gene expression between 5X and 10X. We employed this combined *in vitro* dataset instead of comparing the 5X and 10X datasets to the *in vivo* dataset individually to maximize the statistical power of the analysis and to draw conclusions about the *in vivo* vs. *in vitro* microenvironments, keeping in mind that the *in vivo* microenvironment is highly complex and would be difficult to fully replicate *in vitro*. Because of the differences in sample collection between the two datasets, and the difference in microarray chip formats, we did not directly compare normalized expression values between the two datasets. Instead, the t-score for each gene

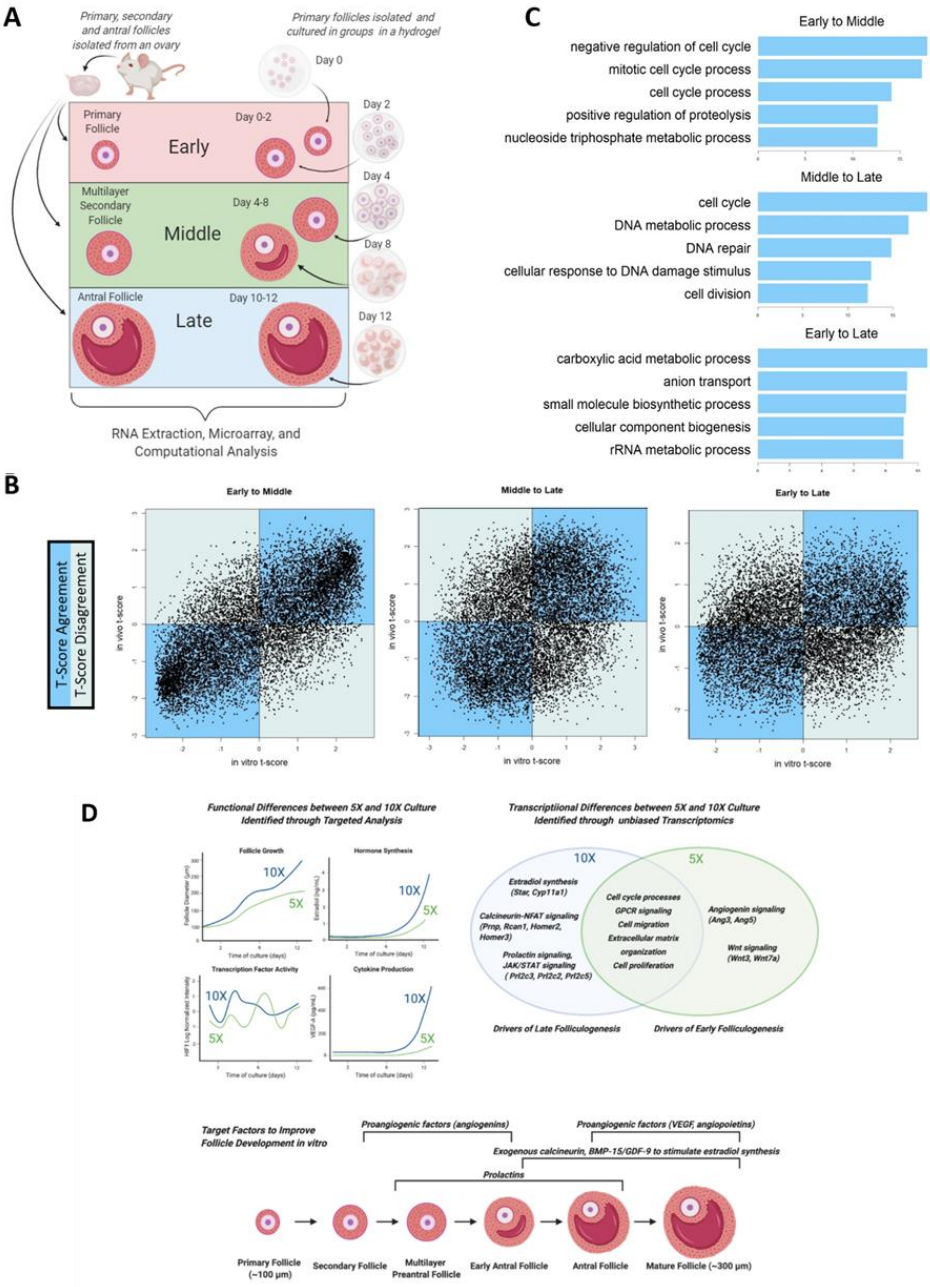


Figure 2.4: Ovarian follicles developing in vitro, compared to those developing in vivo, show significant transcriptional similarities and differences. A) Schematic explaining methods for comparing *in vivo* dataset to 5X and 10X dataset, B) t-score agreement plots (r = 0.704, 0.345, and 0.212 respectively), C) top GO terms from GOrilla analysis, highlighting GO terms from the gene lists with shared t-scores signs between in vitro and in vivo, and D) summary of study findings: previously described targeted analyses from Zhou et al. 2018, combined with transcriptome analysis, leads to identification of targets for improving in vitro follicle culture. Figure 4A was generated using Biorender (biorender.com). Figures 4B and 4C were generated using R (version 3.6.9, https://www.r-project.org/).

from DE analysis was used to compare the changes in gene expression over analogous transitions (i.e. a gene with a positive t-score in both *in vitro* and *in vivo* datasets is one that went up in expression from Time A to Time B in both data subsets). The t-scores from the two datasets were

compared and plotted in Figure 2.4B. The most agreement between the two datasets was observed from the Early to Middle stages of follicle development ($r = 0.704$, $p < 0.0001$ for testing $r = 0$ using Fisher's transformation). The Middle to Late and Early to Late comparisons showed less agreement ($r = 0.345$ and $r = 0.212$ respectively). The t-score agreement and disagreement gene lists were subjected to *GOrilla* analysis to identify GO term enrichment for each dataset and thus cellular processes and pathways that were similar and dissimilar when comparing the *in vitro* and *in vivo* datasets. In both Early~Middle and Middle~Late comparisons, GO terms associated with the cell cycle were enriched, and in the Early~Late comparison various metabolic processes were enriched within the matched gene lists, indicating that these processes behave similarly in the *in vivo* and *in vitro* settings (Figure 2.4C).

2.5 **Discussion**

The group follicle culture system presented here and previously studied by Hornick et al. and Zhou et al. serves as a highly controlled, tunable platform for studying folliculogenesis with the goal of obtaining mature oocytes from cultured follicles[14,15]. Unlike follicles *in vivo*, where the oocyte's maturation potential cannot be determined, the group follicle culture system has been characterized such that suboptimal and optimal culture conditions can be directly compared and samples can be taken from a controlled environment. Previous analyses of group follicle culture have identified 10X culture as superior to 5X with clear secretory and transcriptional distinctions between the two group sizes[14,15]. *In vitro* maturation of oocytes from follicles cultured in groups has also shown the advantage of 10X culture over 5X, which produced significantly fewer mature metaphase II oocytes (Figure 2.1F). Using time-series transcriptomics starting at the primary stage *in vitro*, we have provided a comprehensive view of the dynamic processes involved in folliculogenesis. Understanding temporal expression of genes with known roles in

folliculogenesis, coupled with identification of novel genes and pathways, can enhance optimization of current culture systems by providing a biological basis for supplementing culture media with additional factors identified through analysis of this dataset. Principal component analysis (PCA) of the 5X and 10X microarray data showed that the two systems followed similar trajectories through the reduced dimensional space, reflecting the inherent transcriptional similarities between the two group sizes (Figure 2.2A). Later days of culture (6 through 10) were closely clustered, indicating transcriptional similarities between conditions and across these time points that are reflected phenotypically by follicular expansion and antrum formation seen in both group sizes.

Among the six clusters of genes shown in Figure 2.2B, four had distinct temporal profiles. The cluster termed "early activators" (magenta) (Figure 2.2B,C) showed enrichment for GO terms associated with cell-substrate adhesion, as well as survival and proliferation. This may reflect the follicular cells' adjustment to the *in vitro* culture environment after removal from the ovary and encapsulation in a hydrogel, as well as the initial granulosa cell proliferation during the follicle's transition from the primary to secondary stages. The genes *Amh* and *Amhr2*, which code respectively for anti-Müllerian hormone (AMH) and one of its receptors, belong to this early activator group in magenta (Figure 2.2B,C). AMH has been thoroughly studied as an early regulator of folliculogenesis and is known to inhibit activation of primordial follicles. The early peak of expression and subsequent downregulation observed in our dataset agrees with the expression patterns reported in other studies[45–47]. Interestingly, the fifth cluster, termed "delayed activators" (orange), was enriched for terms related to steroidogenesis—this may reflect the follicle's increased signaling activity as the antrum forms and the follicle becomes hormone-responsive (Figure 2.2B,C). Both *Cyp17a1* and *Cyp19a1* belong to this cluster—the protein

CYP17 is essential for androgen production within the ovary, and in turn, the gene *Cyp19a1* codes for the protein aromatase, which converts androgen into 17β-estradiol[48,49]. The delayed activation of both these genes *in vitro* reflects the timing of hormone production in the follicular structure, which does not occur until the preantral stages and beyond. The transcriptional activity of the transient fourth cluster (blue) peaked in the middle of culture (Figure 2.2B) and was enriched for cell cycle-related genes which agrees with the rapid cell proliferation throughout follicle development in vivo (Figure 2.2C). *Bmp15* was also assigned to the fourth cluster and encodes for bone morphogenetic protein 15 (BMP-15) which has been widely studied, especially in the context of its cooperative role with growth differentiation factor 9 (GDF-9), which is produced by the oocyte[50]. When working as a heterodimer, these factors regulate follicular cell growth and differentiation[50]. Our dataset suggests that transcription of BMP-15 is highest during the middle of culture, when the follicle is expanding, driven by the proliferation and differentiation of its somatic cells.

Current culture systems provide follicles *in vitro* with a "static" nutritional environment, made up of the same supplements at the same concentration throughout culture. But the growing follicular structure may need a more dynamic cocktail of factors – arriving asynchronously, to activate a properly ordered cascade of events - for optimal growth and maturation *in vitro*[51,52]. For example, the cyan cluster (late activators) was enriched for genes with roles in cytokine signaling and angiogenesis, including genes coding vascular endothelial growth factor (VEGF) and angiotensinogen. These factors are not only key for regulating vascular dynamics, but have also been shown to impact steroidogenesis in the antral ovarian follicle[53,54]. VEGF was also identified previously as a top factor distinguishing successful 10X follicle development from suboptimal 5X development[15]. Appropriately-timed supplementation with these pro-angiogenic factors may

enhance follicle development *in vitro* by recapitulating the *in vivo* changes of microenvironment and promoting steroidogenesis late in culture. Supplementation with metabolites, hormone precursors, or factors promoting hormone production may also enhance follicle development or correct dysregulation of folliculogenesis in vitro with careful timing and appropriate dosage.

While follicles cultured in 5X survive to the end of culture and show some growth, their growth and developmental potential is inferior when compared to the 10X condition with regards to terminal diameter and oocyte maturation (Figure 2.1). Comparisons of the somatic cell transcriptomes (Figure 2.3) uncovered important distinctions between the two co-culture conditions that may be leveraged to improve methods for follicle culture in the future. The gene subsets shown in Figure 2.3A highlight pathways and factors with differing temporal expression in 5X and 10X. For example, the yellow subset was enriched for genes related to calcineurin-NFAT signaling. Calcineurin signaling has been characterized for its role in various developmental processes across many organ systems, and in the ovary specifically it has been implicated in modulating FSH-induced upregulation of aromatase and downstream estradiol synthesis[55]. Thus, dysregulated calcineurin signaling in the *in vitro* follicle microenvironment may contribute to delayed follicular expansion and insufficient estradiol synthesis, leading to suboptimal oocyte quality in 5X at the end of culture. Deficient estradiol synthesis was observed in 5X follicles by Zhou et al. in their analysis of the 5X and 10X culture systems[15], and may be mitigated by exogenous stimulation of estradiol production with various factors, such as using exogenous calcineurin or oocyte-secreted factors such as BMP-15 and GDF-9 known to stimulate estradiol production[55,56]. The green subset was enriched for genes related to Wnt signaling and angiogenesis, specifically genes coding for angiogenin proteins. Properly timed activation of angiogenesis-related signaling cascades is not only essential in the follicle's native microenvironment for

vascularization of the theca layer, but may also impact proper steroidogenesis and thus development of the antral follicle. The VEGFs and angiopoietins have been characterized in the context of ovarian follicle development, but other cytokines such as angiogenins have yet to be studied in the context of folliculogenesis. Therefore, follicle development *in vitro* may be enhanced by carefully timed supplementation with pro-angiogenic factors like VEGF or angiopoietin, or by experimental supplementation with angiogenins earlier in culture as evidenced by the decrease in expression of these genes over time in 10X. The cyan cluster was enriched for signaling pathways associated with lactation due to the presence of genes encoding for prolactins. The role of prolactins in ovarian follicle development has not been fully elucidated, although various studies cited the presence of these proteins in the follicular fluid in numerous species[57,58]. Thus, exogenous prolactin supplementation may improve ovarian follicle development *in vitro* despite the lack of knowledge on its mechanisms of action.

The genes represented in Figure 2.3B showed similar temporal transcriptional trajectories in 5X and 10X and highly enriched for terms related to cell cycle, which is reflected in the rapid cell proliferation observed during *in vitro* follicle development. The genes shown in Figure 2.3C showed enrichment for broad gene ontological processes related to G protein-coupled receptor signaling, response to chemical stimulus, and others that may be dysregulated in 5X compared to 10X but are not as dramatically different as those in Figure 2,3A.

The targeted analyses of hormone synthesis, cytokine production, and transcription factor activity from our earlier study[15], combined with the unsupervised transcriptome analysis presented here, highlight biological processes and pathways that may be targeted in a temporal manner to optimize healthy follicular growth in *in vitro* systems. Temporal changes in the transcriptional profile observed in this study are supportive of the superior hormone synthesis, including estradiol

production, we have found in the 10X system[15] linking transcriptional changes with functional changes. Likewise, the cytokine profiles from our earlier study are in agreement with the pathways highlighted in this study, especially with regards to the importance of angiogenic factors in follicle development. With this study, we have expanded on the transcription factor profiling from Zhou et al. and have identified additional target pathways and factors for modulating follicular cell activity. The results of these two studies, as shown in Figure 2.4D, together provide a data-driven basis for modulating these pathways in future studies and using the aforementioned factors to improve ovarian follicle development *in vitro*.

While analysis of culture systems continues to yield important information towards improvements of culture methods, comparing *in vitro*-cultured follicles to those growing in the native ovarian microenvironment may also yield important information on the strengths and shortcomings of current culture methods. By comparing our dataset from follicles cultured in 5X and 10X to follicles freshly isolated from the mouse ovary, we identified similarities and differences that may lead to improvements of *in vitro* culture methods. Transcriptional profiles during early folliculogenesis were similar *in vivo* and *in vitro* and commenced to diverge at a later stages of ovarian follicle development (Figure 2.4B). This may be due to the immature follicle's reliance mainly on paracrine factors provided by the other surrounding follicles, which are present *in vitro* to some extent. However, it is plausible that the *in vitro* microenvironment does not fully recapitulate the endocrine regulation *in vivo* that becomes essential in later stages of follicle development. While there were similar biological processes activated both *in vivo* and *in vitro*, such as cell cycle and metabolic processes, differences in signaling and responses to chemical stimulus were also observed (Figure 2.4C). The lesser agreement in the activated and deactivated biological processes at later stages of follicle development allows us to develop data-driven

hypotheses that can be tested in future research, such as the under activation of steroidogenesis *in vitro* compared with *in vivo*. Nevertheless, it is evident that follicles co-cultured *in vitro* can produce mature oocytes while *in vitro* culture of individual primary follicles failed to do so (Figure 2.1F) and our results reproduced and thus further validated prior observations.

To the best of our knowledge, the results presented here constitute the first transcriptomic time-series dataset comparing insufficient (5X) and optimized (10X) follicle co-culture conditions of primary murine follicles for identifying targets to improve future in vitro culture systems. The timing of culture, tracking follicle development from the primary to fully mature stages, highlights the transcriptional changes in vitro and correlates these changes to phenotypic markers of maturation such as cell proliferation and antrum formation. The longitudinal study, which includes a 7-point time series for both 5X and 10X conditions, provides a detailed picture of transcriptional trajectories in culture. The untargeted genome-wide transcriptional microarray allows for the identification of numerous pathways and biological processes working in concert to drive folliculogenesis, as opposed to targeted approaches such as qPCR. The thorough analytical approach and statistics applied may also serve as a future model for analyzing microarray data with multi-point time series. Conversely, it is important to acknowledge the limitations of transcriptional microarray techniques compared to other transcriptional approaches such as quantitative PCR or next generation sequencing, such as RNAseq or scRNAseq. Single cell sequencing approaches applied to this culture system could uncover more targets and could clarify the identity and relative abundance of major somatic cell types (e.g., theca cells, cumulus and mural granulosa cells). The target factors and pathways presented in this study could be characterized further using other omics techniques that measure downstream consequences of the active signaling and metabolic pathways, such as proteomics and metabolomics profiles. Future

studies evaluating the transcriptional changes of the growing oocyte, in addition to validation of these targets by modulating these pathways for individual follicle culture, will lead to an optimized individual follicle culture system for clinical translation. Moving forward, similar transcriptional approaches should be applied to comparing murine and human folliculogenesis and identifying factors driving human folliculogenesis specifically, to work towards clinical implementation of in vitro follicle culture.

2.6 **Conclusions**

In vitro ovarian follicle culture is an active area of research aimed towards providing survivors of childhood cancer with the opportunity to have their own biological children. Patients' cryopreserved ovarian tissue is largely made up of primordial and primary follicles, the maturation of which less understood than late stage follicles, thus they present a significant challenge to culture *in vitro*. Improved culture methods require a deeper understanding of the mechanisms of early folliculogenesis *in vivo* and how *in vitro* culture may differ. Primary follicle co-culture is a step towards understanding these mechanisms and provides a means to achieve meiotically-competent oocytes, but this method has limited clinical translation. A universal rescue cocktail created based on transcriptional signatures identified in multiple follicle co-culture would provide the individual follicle with the proper nutrients in relevant concentrations and at the appropriate stages of development. Previous targeted secretome analysis identified a few factors that may improve individual follicle growth *in vitro*[15]. However the transcriptome analysis presented here represents a more thorough attempt to identify novel target pathways and cell processes that may be manipulated exogenously to improve *in vitro* culture outcomes. Additionally, comparisons between this dataset and previously published data show transcriptomic similarities and differences between follicles cultured *in vitro* and those isolated from the *in vivo* environment.

Further refinement of media conditions with regards to both relevant signaling pathways in folliculogenesis and necessary metabolites may bring successful early-stage individual follicle culture closer to clinical application in assisted reproductive technologies.

Chapter 3 RNA Sequencing and Proteomics Profiling of Cultured Ovarian Follicles to Decipher Regulatory Cues in Follicle Development

3.1 Preface

In the previous chapter, we used microarray to identify genes and pathways associated with successful follicle development *in vitro*. Here, we build on that work using RNA sequencing of follicular somatic cells and oocytes from follicles cultured *in vitro*. Analysis of these samples supported the most salient findings from our microarray, and provided additional genes and pathways that may serve as candidates for exogenous manipulation in future follicle culture systems. The transcriptome profile of oocytes from early-stage follicles cultured *in vitro* constitutes the first dataset of its kind, and provides a wealth of knowledge on processes leading up to the oocyte's resumption of meiosis at the end of follicle development. We profiled the same culture system using shotgun proteomics, confirming a number of our transcriptome findings and uncovering additional target proteins for modulating folliculogenesis *in vitro*. These studies together have greatly advanced our understanding of the molecular mechanisms driving folliculogenesis in vitro, opening up exciting possibilities for future research and follicle culture system design.

3.2 Introduction

The paracrine milieu that contributes to follicle activation and the earliest stages of development is largely uncharacterized, leading to challenges in fertility preservation and assisted reproduction technologies[59–62]. Recently, our group and others have made efforts to understand

these cues through co-culture of primary murine follicles in groups. This culture system has been shown to significantly improve growth and survival when compared to culture of individual primary follicles, with groups of 10 (10X) cultured together being more successful than groups of 5X[14,15]. This tunable, controlled system recapitulates the follicle's native microenvironment (where early-stage follicles reside in proximity within the ovary) without the complicating factor of non-follicular cell types and allows researchers to explore what factors drive successful follicle development and their temporal dynamics throughout culture.

Previous studies by our research group have used the group follicle culture system to identify transcription factors, cytokines, genes, and pathways associated with the successful follicle development observed in 10X group culture when compared to the suboptimal outcomes of 5X culture[15,63]. 10X follicles had accelerated activity of key transcription factors NF-κB and HIF1, as well as a unique secretory signature, revealing some of the biological processes and pathways essential for early follicle development[15]. Subsequently, our group used microarray analysis of somatic cells from 5X and 10X follicles to profile transcriptional signatures of follicles throughout culture[63]. This analysis revealed a distinct transcriptomic profile in 10X follicles with dynamic expression of genes related to prolactin signaling and angiogenesis, constituting a data-driven list of pathways that may be targeted in a temporal manner to optimize follicle culture systems in the future[63]. Here, we expanded on understanding of the transcriptomes of 5X and 10X follicles using a more robust RNA sequencing approach, separately investigating the profiles of *in vitro* cultured follicular somatic cells and corresponding oocytes for the first time. Analysis of these datasets revealed numerous points of agreement with our previous studies, and uncovered new genes of interest related to fatty acid oxidation and organelle organization in the oocyte. We used ligand-receptor pair analysis to relate the somatic cell and oocyte transcriptomes, identifying

several signaling pathways with no previous link to folliculogenesis, including endogenous opioid peptide signaling and tachykinin signaling. Finally, we used shotgun proteomics to validate and augment our transcriptome analysis, where we observed numerous points of agreement with the RNAseq datasets and established additional biological processes and proteins associated with successful follicle development *in vitro*, including MAPK signaling-related proteins and cell survival proteins. Collectively, this multi-faceted investigation enhances our understanding of the many biological processes driving folliculogenesis *in vitro*. By reporting new genes, proteins, and pathways, our study not only furthers the fundamental understanding of ovarian biology but also presents potential avenues for refining and optimizing follicle culture systems, with far-reaching implications for assisted reproductive technologies and fertility preservation strategies.

3.3 **Experimental Methods**

3.3.1 Follicle Isolation, Encapsulation, and Culture

Primary ovarian follicles were isolated, encapsulated in alginate matrices, and cultured as previously described[15] (Chapter 2). Briefly, whole ovaries were removed from C57B6 X CBA/J mice ages 10-12 days. All animal procedures were performed in compliance with the Guidelines for the Care and Use of animals at the University of Michigan and approved by the Animal Care and Use Committee at the University of Michigan (PRO00009635). Ovaries were washed in warm Leibovitz's L-15 medium (L-15) (Thermo Fisher) and transferred to a dish with pre-equilibrated alpha modification of minimum essential medium (αMEM) (Thermo Fisher) with 0.5% (v/v) PenStrep (Thermo Fisher). To collect large numbers of primary follicles from the ovaries, 10% (v/v) Liberase DH at 13 Wünsch units/mL (Sigma) was added to the dish of ovaries and gently mixed, then incubated undisturbed at 37°C for 35-45 minutes. After incubation, the dishes were

pipetted for 5 minutes to break up the enzymatically-digested tissue, releasing individual follicles, then the digest was arrested using 10% (v/v) fetal bovine serum (FBS) (Thermo Fisher).

Follicles were encapsulated as previously described[15] (Chapter 2). Briefly, primary follicles with diameters ranging 90-100 µm were selected from the ovary digest and encapsulated in 10 µL alginate beads in groups of 5 (5X) or 10 (10X) follicles per bead. To crosslink alginate, beads were dropped into a solution of 50 mM $CaCl_2$ and 140 mM NaCl and allowed to crosslink for 3 minutes. Encapsulated follicles were cultured in 96 well plates with 150 µL growth media (GM) per well, containing αMEM supplemented with 3 mg/mL bovine serum albumin (BSA) (MPBiomedicals), 1 mg/mL bovine fetuin, 5 µg/mL insulin, 5 µg/mL transferrin, 5 ng/mL selenium (ITS, Sigma), and 10 mIU/mL highly purified, human-derived follicle-stimulating hormone (FSH) (Urofollitropin, Ferring Pharmaceuticals). Every 48 hours after encapsulation, half of the GM (75 µL) was replaced with fresh GM. Follicle growth was tracked by imaging follicles every 48 hours before exchanging GM. Follicle diameter from each 48 hour time point was measured using ImageJ by taking the average of two perpendicular measurements across the follicle. Hydrogels without exactly 5 or 10 fully encapsulated follicles were removed from culture. Hydrogels containing follicles that fell out of the alginate hydrogel onto the bottom of the well, or hydrogels with follicles dead at the start of the culture (on day of encapsulation) were excluded from analysis and sample collection for sequencing. Follicles were termed "dead" if the oocyte as extruded more than 50% from the follicular structure with no surrounding granulosa cells, or if the follicular structure was dark and the oocyte was not visible. Only follicles surviving to Day 12 of culture were included for growth analysis, while all follicles were included in survival analysis. The results presented were pooled from 9 separate culture experiments, with a total of 2,920 follicles contributing to RNA samples for sequencing.

3.3.2 Follicle Dissection, RNA Extraction, and Sequencing

Follicles (n = 910 for 5X, n = 1,110 for 10X) were taken from culture for dissection into somatic cell and gamete counterparts at the following time points: Days 2, 4, 6, 8, 10, and 12. Fresh, un-encapsulated follicles were also collected for sequencing as controls for each of the 9 culture experiments. The number of follicles isolated for each RNA sample are outlined in Appendix Table A.3, with an average of 43 follicles/sample for 5X, 53 follicles/sample for 10X, and 100 follicles/sample for controls. To dissect follicles, alginate beads were digested by incubating with 10 IU/mL alginate lyase for 20 minutes. Follicles were then transferred to warm L-15 and oocytes were mechanically removed from the follicles using insulin syringes (BD 305620) then rinsed 3 times in L-15 to remove any remaining somatic cells before collection in a microcentrifuge tube. The remaining somatic cells were collected into a microcentrifuge tube. Oocyte and somatic cell samples were centrifuged at 300g for 5 minutes, supernatant was removed, and cell pellets were snap frozen in liquid nitrogen then stored at -80°C. RNA extraction on each sample was performed using the Qiagen RNEasy Micro Kit following the manufacturer's instructions. Extracted RNA was stored at -80°C before being submitted to the University of Michigan's Advanced Genomics Core for sequencing. Samples were sequenced on the Illumina NovaSeq platform at a sequencing depth of 25 million reads per sample.

3.3.3 RNAseq Data Analysis

The sequence data was analyzed in R (v 4.2.1)[19]. The initial count matrix contained 55,426 genes for 90 samples. We then reduced the number of genes using multiple filtering criteria. Firstly, retaining only genes that were present (non-zero counts) in at least half the samples in oocytes (n = 45) or in cells (n = 45). We retained 25,348 genes for downstream analysis. Next

counts were normalized to counts per million for each sample and then log2 was taken with values floored at 0.02 (chosen because it was half the lowest CPM value of ~0.04).

The data was viewed jointly with both cell types (cells and oocytes) by calculating the principal components (PCs) on the log-transformed CPM matrix. The data was then viewed separately for each cell type. Batch effects were noticed in PC plots when viewing experimental batches in each day set (2-4, 6-8, 10-12). To adjust for this, we took the centroids for each batch and centered them on their day set centroid and applied this linear transformation to individual samples from each batch. PCs and models were then calculated on the batch corrected data.

For the models, the linear model was calculated for each gene by day as a categorical variable, condition (5x vs 10x) and the interaction of day and condition. ANOVA was t run on the results of the linear model to determine differentially expressed genes (DEGs) (Benjamini-Hochberg adjusted p-value < 0.05) for day, condition and the interaction.

Kohonen's self-organising maps (SOM) were then used to group the differentially expressed genes. SOMs were calculated on the scaled average expression for each day and condition combination on each set of DEGs. Maps of dimensions 3 by 3 were used for day DEGs (cells & oocytes) and 2 by 2 for all other DEG sets.

3.3.4 Pathways Analysis

Gene lists identified through SOM were submitted to the PANTHER database for gene ontology enrichment analysis[64]. Genes were filtered through the "biological process" GO terms using gene name, and tested using Fisher's Exact t-test. GO terms with adjusted p-value < 0.5 were considered significant.

3.3.5 Proteomics Sample Preparation and Profiling

Follicles were cultured in 5X and 10X as previously described for liquid chromatography tandem mass spectrometry (LC-MS/MS)-based proteomics. On Days 4, 8, and 12 of culture, follicles were isolated from alginate hydrogels as previously described, rinsed in DPBS-/-, and centrifuged at 300g for 5 minutes to form a cell pellet. Supernatant was removed and pellets were snap frozen and stored at -80°C. Fresh primary follicles (mice ages 10-12 days) were isolated and snap frozen as a control sample (n = 5,000 follicles). The number of follicles isolated for each protein sample, which were compiled over multiple culture experiments, are as follows: D0 control n = , 5X D4 n = 1500, 10X D4 n = 1500, 5X D8 n = 230, 10X D8 n = 225, 5X D12 n = 200, 10X D12 n = 200. Each sample contained approximately 5 million cells based on predicted number of cells in follicles of various diameter (Appendix Table A.4).

Cell pellets were delivered to the University of Michigan Proteomics and Peptide Synthesis Core for analysis. Protein extraction was performed in modi modified RIPA buffer (2% SDS, 150mM NaCl, 50mM Tris pH 8, 1X Roche cOmplete protease inhibitor) using mechanical disruption with NextAdvance buller blender with 1mm stainless steel beads. The protein concentration of each sample was determined by Qubit fluorometry (Invitrogen). 20µg for each sample was processed by 5cm SDS-PAGE using a 10% Bis-Tris Novex mini-gel (Invitrogen) and the MOPS buffer system. The mobility region was excised into 20 equally sized bands and each was processed by in-gel digestion using a robot (DigestPro, CEM) with the following protocol: washing with 25mM ammonium bicarbonate followed by acetonitrile; reduction with 10mM dithiothreitol at 60°C followed by alkylation with 50mM iodoacetamide at RT; digestion with trypsin (Promega) at 37°C for 4 hours; quencing with formic acid. The supernatant was analyzed directly without further processing.

Half of each digested sample was analyzed by nano LC-MS/MS with a Vanquish Neo HPLC system interfaced to a ThermoFisher Fusion Lumos mass spectrometer. Peptides were loaded on a trapping column and eluted over a 75μm analytical column at 350nL/min; both columns were packed with PepMap Neo C18 resin (ThermoFisher). The mass spectrometer was operated in data-dependent mode, with the Orbitrap operating at 60,000 FWHM and 15,000 FWHM for MS and MS/MS respectively. The instrument was run with a 3s cycle for MS and MS/MS. 10 hours of instrument time was used the analysis of each sample.

3.3.6 Proteomics Data Processing

Data were searched using a local copy of Mascot (Matrix Science) with the following parameters: Enzyme: Trypsin/P; Database: SwissProt Mouse (concatenated forward and reverse plus common contaminants); Fixed modification: Carbamidomethyl (C); Variable modifications: Oxidation (M), Acetyl (N-term), Pyro-Glu (N-term Q), Deamidation (N/Q); Mass values: Monoisotopic; Peptide Mass Tolerance: 10 ppm; Fragment Mass Tolerance: 0.02 Da; Max Missed Cleavages: 2. Mascot DAT files were parsed into Scaffold (Proteome Software) for validation, filtering and to create a non-redundant list per sample. Data were filtered using 1% protein and peptide FDR and requiring at least two unique peptides per protein.

3.4 **Results**

3.4.1 Growth and survival kinetics of longitudinal follicle co-culture experiments

To profile the transcriptome of oocytes and somatic cells from follicles cultured in groups, primary follicles were encapsulated in alginate hydrogels in groups of 5 (5X) and 10 (10X) and cultured for up to 12 days (Figure 3.1A-B). We did not observe any statistical difference in survival rate between 5X and 10X conditions (Figure 3.1C). Follicles cultured in 5X and 10X started at

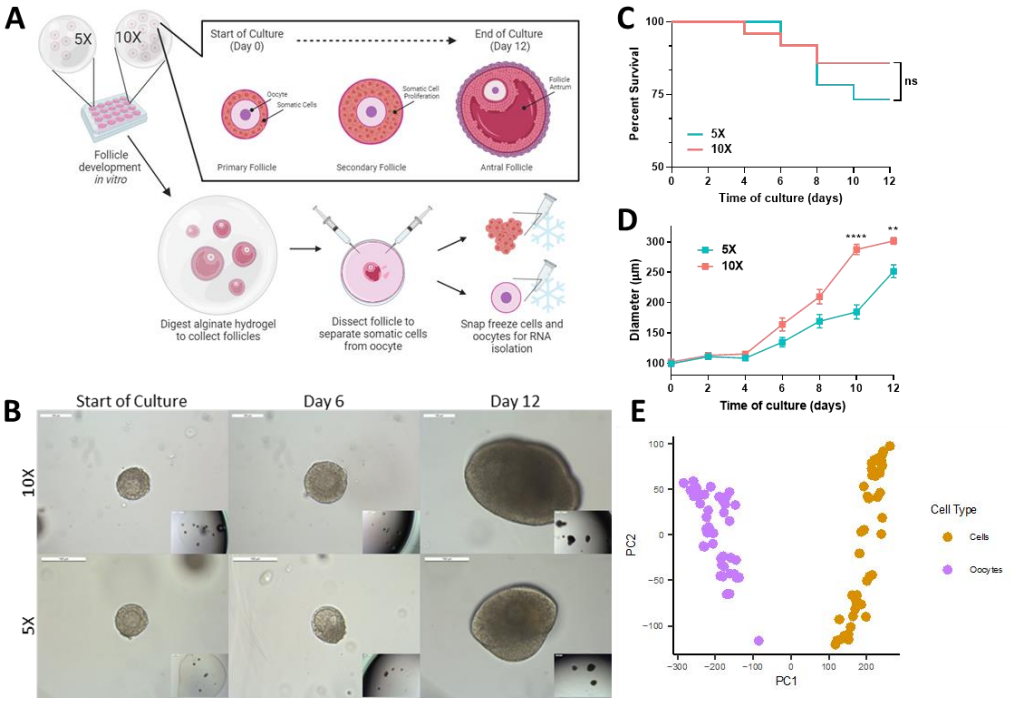


Figure 3.1: Ovarian follicles cultured in 10X outperform 5X culture. A) Schematic describing the process of cell isolation, beginning with primary follicles cultured in alginate hydrogels over 12 days, during which follicles are dissected at two-day intervals for dissection and cell collection, B) bright field images of individual primary follicles encapsulated in alginate hydrogels in either groups of 5 or 10 (scale bar = 100 µm), C) survival rates of follicles over duration of culture in 5X and 10X, D) growth rates of follicles over duration of culture in 5X and 10X, and E) PC1-2 projection of oocyte and somatic cell RNA samples. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Figure 3.1A was generated using biorender.com.

102±0.397 µm in diameter and follicles in 10X grew significantly larger, reaching 301.6±5.492 µm (mean±SEM) in diameter, compared to 5X follicles reaching only 251.6±10.613 µm by Day 12 (Fig. 1B,D; Appendix Figure A.5), which was consistent with previous studies[14,15,63]. Every two days (starting on Day 0) we collected RNA from oocytes and somatic cells cultured in 5X and 10X (Supplementary Fig. 1C). Each RNA sample contained RNA from the oocytes or somatic cells of 30-100 follicles (Appendix Figure A.3). The number of genes for analysis were filtered down to 25,348 and counts were normalized and log-transformed (Experimental Methods). Projection of all 90 samples showed marked separation of the oocyte and somatic cell transcriptomes (Fig. 1E), which were subsequently analyzed individually.

3.4.2 Profiling the 5X and 10X Somatic Cell Transcriptomes during Culture

After batch correction (Experimental Methods), projection of somatic cell samples showed the expected temporal changes in both 5X and 10X, traveling first along PC-1, then along PC-2 towards the end of culture, similar to what was observed previously in our microarray analysis[63] (Figure 3.2A). To identify genes and pathways associated with the progression of follicle development *in vitro*, we filtered out genes with no significant temporal changes and sorted the remaining 15,183 genes using self-organizing maps without accounting for the follicle culture condition (5X or 10X). We sorted genes into four clusters representing different temporal expression patterns (Figure 3.2B) and performed differential expression (DE) analysis using linear models of each cluster to highlight genes with the most striking temporal expression patterns. DE gene lists were submitted to *PANTHER* to identify biological processes enriched in each cluster[64].

The first cluster (yellow, n=649 genes) contained genes that dipped in expression around Day 6, then climbed again towards the end of culture (Figure 3.2B). This group was enriched for genes related to vesicle fusion (*Stx4a/7/16*) and autophagy (*Becn1, Npc1, Sirt2*). Interestingly, numerous studies on extracellular vesicle (EV)-mediated signaling in granulosa cells suggest that EVs shuttle essential, stage-specific signals during follicle development, especially via the follicular fluid late in folliculogenesis. Likewise, autophagy has recently been implicated in the differentiation of granulosa cells where disruption of autophagy-related genes like *Becn1* leads to decrease expression of genes such as *Cyp19a1* and *Fshr*[65]. The second cluster (pink, n=5066 genes) contained genes with high initial expression that decreased as culture progressed (Figure 3.2B). This cluster contained genes related to angiogenic signaling (*Angpt1, Ptpn1, Nedd4*), Wnt signaling (*Rspo1, Sfrp1/2*), and cell proliferation (*Myc, Ctnnb1/ip1, Pdgfa*). These findings agree with expression patterns for angiogenic, Wnt-associated, and proliferation-related genes identified in our previous study[63]. The roles of pro-angiogenic cytokines and canonical Wnt signaling

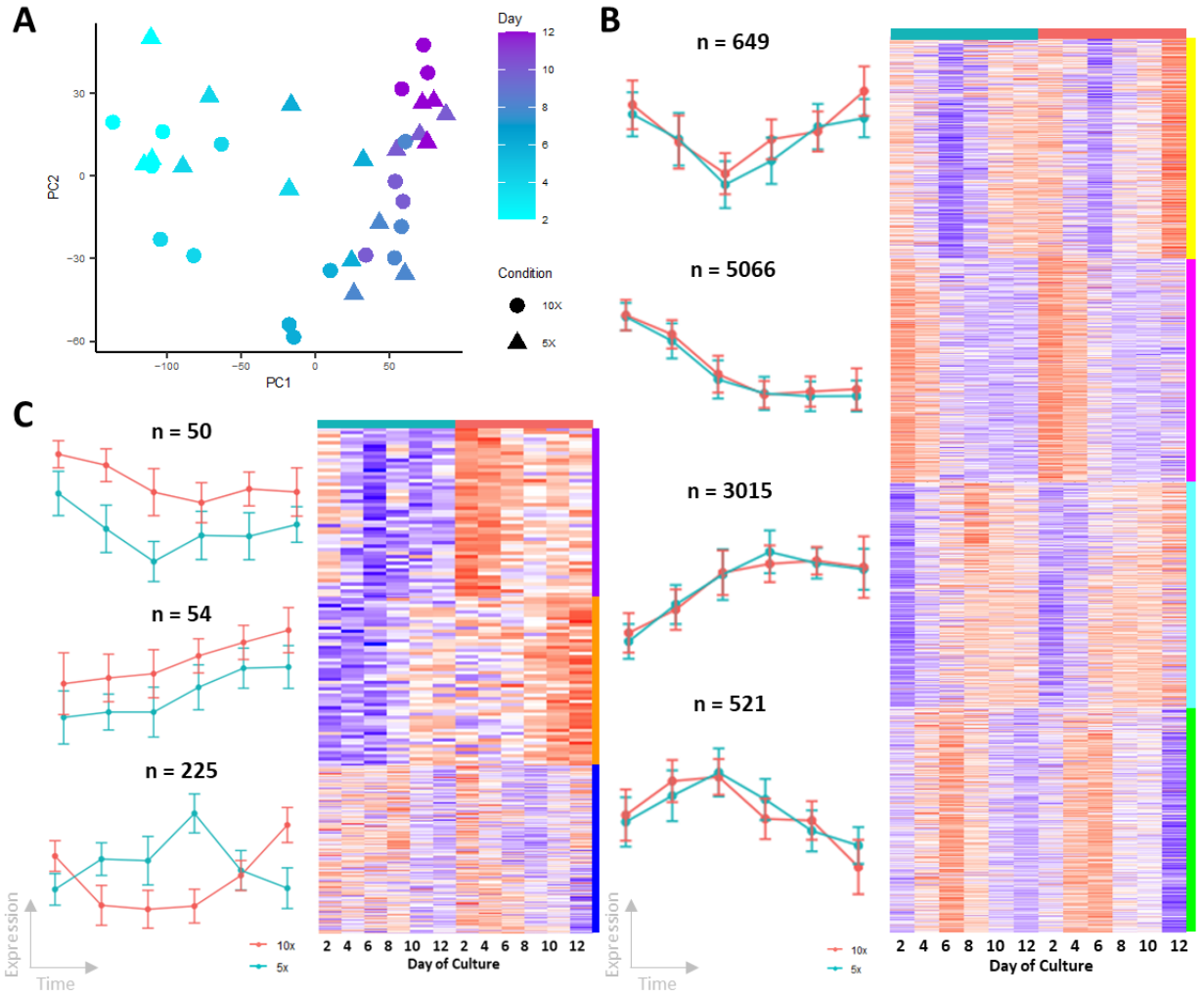


Figure 3.2: Profiling of the follicular somatic cell transcriptome over time *in vitro*. A) Principal component plot of the 36 somatic cell RNA samples, B) heatmap of standardized expression level for genes in four clusters obtained from SOM analysis of gene expression over time, and C) heatmap of standardized expression level for genes in three clusters obtained from SOM analysis of genes differing in 5X vs. 10X conditions.

molecules in folliculogenesis and hormone production are widely reported, and pro-angiogenic molecules like VEGFs have previously been suggested as exogenous growth factors for follicle development[15,63,66]. The third cluster (cyan, n=3015 genes) contained genes that increased in expression steadily as culture progressed, and plateaued from Days 10 to 12 (Figure 3.2B). This group was enriched for genes related to chromosome condensation (*Ncap2d/3d, Ncaph2*) and meiosis (*Hspa2, Syce2, Tex15*). While the genes in this cluster have not been studied in granulosa cells specifically, the role of granulosa cell-derived signals for supporting oocyte meiotic

progression is well-documented[67–70]. The final cluster (green, n=521) contained genes that peaked in expression on Day 6, then decreased as culture progressed (Figure 3.2B). This cluster was enriched for genes related to DNA replication (*Brca2, Pola1, Mcm6*) and mitosis (*Nasp, Ccne1, Cdt1*). The expression pattern of proliferation-related genes in this cluster likely reflects the rapid granulosa cell proliferation observed during early- and mid-culture and agrees with a similar pattern of mitosis-related genes from our previous study[63].

We next evaluated differences between the 5X and 10X conditions over the duration of culture. Follicles in both conditions showed similar growth kinetics until ~Day 8 of culture, at which point 5X growth slowed down when compared to 10X (Figure 3.1C)[63]. Data was filtered to select 1) genes with differing expression across conditions and 2) genes with differing expression patterns across conditions (Experimental Methods). 199 genes and 782 genes were retained from these respective filtering approaches for downstream analysis. Each gene subset was sorted into four groups using self-organizing maps, then DE gene lists were submitted to *PANTHER* for gene ontology enrichment analysis[64]. Three of the eight gene groups were of particular biological interest and are represented in Figure 3.2C. The first group (purple, n=50 genes) and second group (orange, n=54 genes) contained genes with consistently higher expression in 10X over the period of culture, with either declining (purple) or increasing (orange) expression over time (Figure 3.2C). The last group contains genes with differing, temporally dynamic expression patterns in 5X and 10X (Figure 3.2C, blue). The purple cluster was comprised of genes involved in mRNA splice site selection (*Prpf39, Luc7l3, Khdc4*). mRNA splicing is the process by which introns are removed from a transcript to yield mature mRNA for translation. While the importance of mRNA splicing for follicle development has not been clearly described, this process is intrinsic to the dramatic cell proliferation in the granulosa cells of growing follicles. The orange cluster was enriched for genes

involved in angiogenesis and VEGF signaling (*Vegfa, Egln1, Mif*). The importance of angiogenic factors for folliculogenesis is well documented[71–74] and our previous studies on follicles cultured *in vitro* indicated differential expression of proangiogenic factors in 10X vs. 5X[15,63]. Finally, the blue cluster contained genes related to exocytosis and vesicle-mediated signaling (*Rab3p, Syt1, Cspg5*). As described previously, EV-mediated signaling is crucial to folliculogenesis, especially after antrum formation.

Finally, we reviewed the somatic cell transcriptome in light of recent data on the single-cell transcriptome of the mouse ovary across the estrous cycle[75]. The authors reported marker genes for follicular cells corresponding with preantral, antral, and atretic follicles, which we compared to our *in vitro* time course. We observed decreasing expression of markers for preantral granulosa cells over time (*Gatm, Slc18a2, Kctd14, Fndc5, Igfbp5, Wt1, Col18a1, Tmem184a, Pcsk6*), which suggests the differentiation of granulosa cells during culture as folliculogenesis proceeds (Appendix Figure A.6A). There was a concomitant increase in expression of genes associated with the granulosa cells of antral follicles (*Inha, Slc26a7, Nppc, Grem1, Cyp19a1, Nap1l5, Tom1l1, Hsd17b1*) (Appendix Figure A.6B), genes associated with mitotic granulosa cells, which peaked on Days 6-8 of culture in our data (*Top2a, Ccnb2, Ube2c, Hmgb2, Racgap1, Cdca8, Birc5, Ccna2, Prc1*) (Appendix Figure A.6C) and increased expression of five genes associated with luteinization of granulosa cells (*Adamts1, Loxl2, Mrap, Cemip, Cdkn1a*) (Appendix Figure A.6D). Importantly, the somatic cell transcriptome had dynamic expression of theca-related genes throughout culture. Four of the top ten reported marker genes for early theca cells decreased over time *in vitro* (*Adcy7, Thbd, Mest, Lamc3*) and seven of the top ten markers for steroidogenic theca increased over time (*Mgarp, Cyp17a1, Fdx1, Hao2, Folr1, Acsbg1, Serpina5*), suggesting further maturation and differentiation of the theca layer in cultured follicles

(Appendix Figure A.6E-F). We also evaluated follicular somatic cell expression of the four estrous cycle biomarkers identified by Morris et al. (*Prss35, Inhba, Tinagl1, Nppc*)[75]. We observed increasing expression of *Nppc* and *Inhba*, and conversely a decrease in *Prss35* and *Tinagl1* over time *in vitro* (Appendix Figure A.6G). The last 4 to 6 days of folliculogenesis, when follicles progress from preantral to antral stages, coincide with the estrous cycle in the mouse. Strikingly, the follicular profile of these genes whose protein products hold promise as secreted biomarkers are mirrored in the *in vitro* system.

3.4.3 Trajectory of the Oocyte Transcriptome during Culture

We next investigated transcriptional changes in the oocyte during culture *in vitro* and across our conditions. Like the somatic cells, oocyte samples traveled first along PC-1, then along PC-2 towards the end of culture (Figure 3.3A). Using an approach analogous to the somatic cells, we first evaluated genes with temporally dynamic expression, then those with differing expression in the 5X vs. 10X conditions. After removing genes with no significant temporal changes, we sorted the remaining 10,878 genes using self-organizing maps without accounting for the follicle culture condition (5X or 10X). The first cluster (yellow, n=2734 genes) contained genes that increased steadily in expression over the period of culture (Figure 3.3B). This group was enriched for genes related to spindle assembly (*Spire1, Cenpa, Mad2l1*), RNA polymerase II assembly (*Polr2g, Gtf2a2, Med21*), and chromatin organization (*Rad21, Pcid2, Ccnb1/2*). These biological processes are all key to meiotic resumption and the increase in gene transcription that occurs during maturation. Notably, genes *Ccnb1/2* code for proteins cyclin B1 and B2 which are involved in the assembly of maturation promoting factor (MFP), a protein complex essential to meiotic resumption. The second cluster (pink, n=1163 genes) contained genes with high initial expression that dropped leading up to Day 8 of culture, then increased again (Figure 3.3B). This cluster

contained genes related to fatty acid oxidation (*Echdc2*, *Pck2*, *Dbp*). Interestingly, fatty acid oxidation (FAO) is an important metabolic pathway for oocytes to generate ATP and aberrant FAO

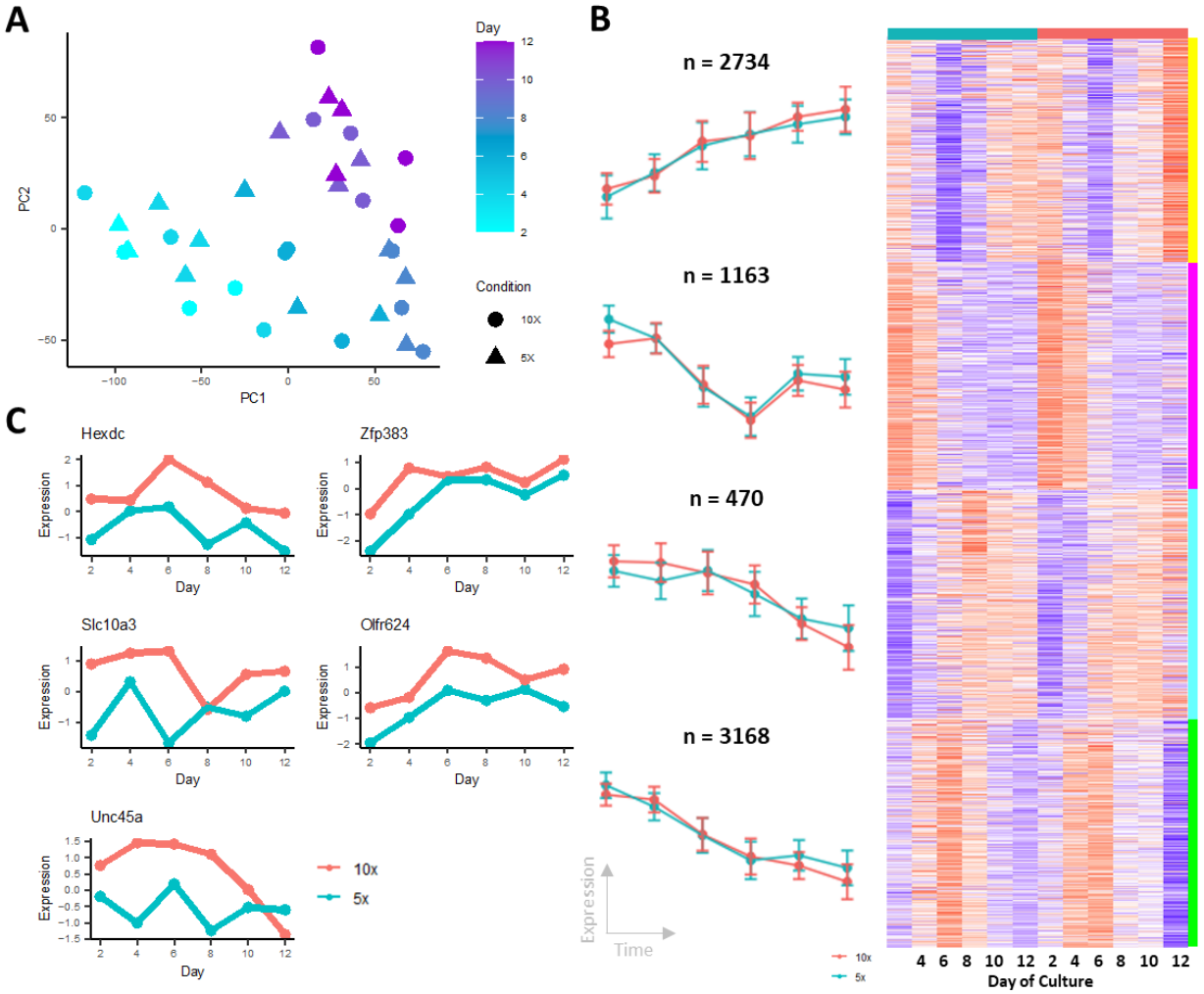


Figure 3.3: Profiling of the oocyte transcriptome over time *in vitro*. A) Principal component plot of the 36 oocyte RNA samples, B) heatmap of standardized expression level for genes in four clusters obtained from SOM analysis of gene expression over time, and C) expression plots for five genes with differing expression profiles in 5X vs. 10X conditions.

in *in vitro*-matured oocytes leads to poor embryo development[76]. The third cluster (cyan, n=470 genes) contained genes that decreased steadily as culture progressed (Figure 3.3B). This group was enriched for genes related to microtubule and organelle organization (*Dnah7b/9/11*, *Cdc14b*, *Sybu*, *Stk36*). Interestingly, *Dnah7b/9/11* code for dynein heavy chain proteins, which assemble into complexes important for microtubule-based transport. Additionally, organelle re-organization is

an important part of oocyte cytoplasmic maturation and is essential for fertilization. The final cluster (green, n=3168) contained genes with a more dramatic decrease in expression over the period of culture compared to the cyan cluster (Figure 3.3B). This cluster was enriched for genes related to PI3K/AKT signaling (*Pten, Pik3r2/3, Pigk, Impa2*). The PI3K/AKT pathway has been widely studied in the context oocyte survival and apoptosis, and it is generally accepted that disruption of this signaling cascade leads to decreased oocyte quality and aberrant DNA damage repair in a variety of animal models[77]. However, the genes and proteins from the cascade that impact oocyte survival directly have not been identified. Our results suggest decreasing expression of PI3K inhibitor *Pten* in the oocyte as maturation progresses *in vitro*, and decreasing expression of *Impa2*, which codes for inositol monophosphatase 2. While not well-characterized in the ovary, *Impa2* modulates mTORC1, which directly impacts ribosome biogenesis, protein synthesis, and metabolism in the oocyte[77].

We next reviewed genes with differing expression in 5X versus 10X in the oocyte samples. Remarkably, we identified only 14 genes with differential expression across conditions, and only five of the 14 were genes with biological annotation: *Unc45a, Hexdc, Slc10a3, Olfr624,* and *Zfp383* (Figure 3.3C). Of these, only *Hexdc* has known biological significance in the oocyte: it codes for hexosaminidase D, one of a family of proteins important for sperm-egg interaction during fertilization[78]. *Unc45a*, which codes for a protein involved in actomyosin complex formation, *Slc10a3*, which codes for a member of the solute carrier protein family, *Olfr624*, which codes for a member of the olfactory receptor family, and *Zpf383*, which codes for a member of the zinc finger protein family, are not characterized in the ovary and have no known role in folliculogenesis or oogenesis.

3.4.4 Ligand-receptor analysis of the follicular somatic cell and oocyte compartments

We next used ligand-receptor analysis to identify bidirectional signaling between oocytes and somatic cells. We used the Lewis Lab ligand-receptor pair database and identified 28 ligands and 106 receptors in the oocyte and somatic cell gene lists, culminating in 32 ligand-receptor pairs between datasets. From these pairs we identified 6 with strong oocyte-somatic cell interaction (Figure 3.4). Three of the six pairs had high oocyte expression of the receptor and high somatic cell expression of the ligand (Figure 3.4A).

The first pair displayed in Figure 3.4A, *Oprd1* and *Penk*, which code for opioid receptor delta 1 and proenkephalin, are part of the opioid peptide signaling cascade. Endogenous opioid peptides such as β-endorphin and α-neoendorphin from the cumulus granulosa mediate oocyte maturation, but the endogenous peptide proenkephalin has yet to be characterized in the context of folliculogenesis[79]. The next pair, *Ngln1* and *Nrxn1*, code for neuroligin-1 and neurexin-1. Neuroglin-1 is a transmembrane protein from the NLGN protein family, which binds to transmembrane neurexin proteins to mediate cell-cell adhesion. Interestingly, the cell-cell adhesion mediated by these two protein families is well-documented in neurons[80] but has not previously been reported in ovarian follicles. The third pair, *Ntsr1* and *Nts*, code for neurotensin receptor 1 and the ligand neurotensin. The mechanisms of action by which neurotensin impact late follicle development have not been elucidated, but the presence of neurotensin has been reported in late-stage rat and human follicles and was modulated *in vitro* by exogenous hCG or PI3K/MAPK signaling pathway antagonists[81]. It has been proposed that neurotensin specifically modulates inflammation, cell migration, and vascular permeability prior to ovulation[81].

The other three pairs shared the converse relationship (high somatic cell receptor expression, high oocyte ligand expression, Figure 3.4B). The first pair in Figure 3.4B, *Tacr3* and *Tac2*, code for tachykinin receptor 3 and tachykinin 2. Tachykinins (TKs) are extensively

characterized in the brain and gut, but more recently TKs have been identified in follicular fluid and TK-related mRNA has been observed in oocytes and granulosa cells[82]. While their mechanism

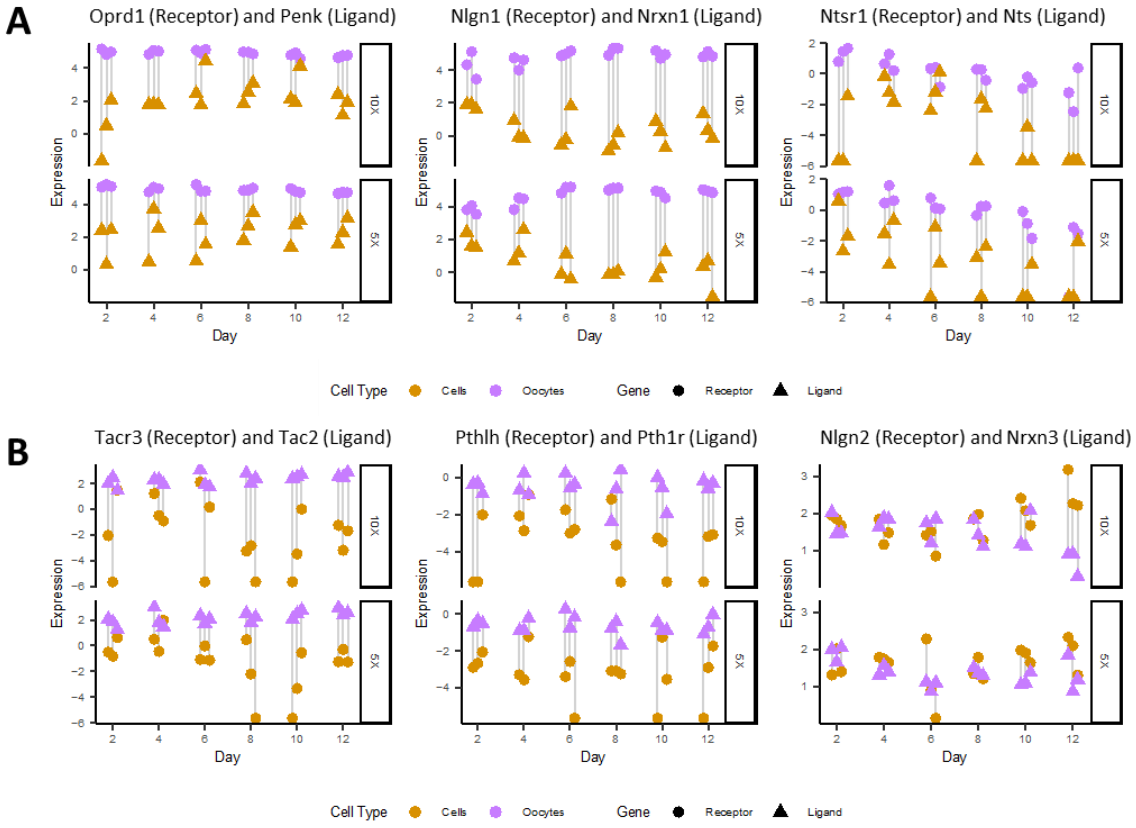


Figure 3.4: Ligand-receptor analysis of intrafollicular signaling. A) Ligand-receptor pairs where oocytes showed high receptor expression (purple circles) and somatic cells showed high ligand expression (gold triangles) and B) Pairs where oocytes showed high ligand expression (purple triangles) and somatic cells showed high receptor expression (gold circles).

of action in mammals has not been verified, work in ascidians suggests that TKs facilitate oocyte maturation via activation of key proteases[82]. The second pair, *Pth1r* and *Pthlh*, code for parathyroid hormone 1 receptor and parathyroid hormone like hormone (PTHLH). PTHLH has been identified in many tissues and regulates a diverse range of biological processes in cell migration, differentiation, and survival. In mouse oocytes, it has been reported that PTHLH has limited impacts on oocyte maturation but is essential for preimplantation embryonic development[83]. The final pair, *Nlgn2* and *Nrxn3*, code for neuroligin-2 and neurexin-3. Notably, the ligand-receptor

relationship for these genes is reversed from the *Ngln1-Nrxn1* pair shown in Figure 3.4A, demonstrating transcription of genes from both protein families in oocytes and follicular somatic cells.

3.4.5 Proteomic Profiling of Follicles Co-Cultured in vitro

To validate genes and pathways of interest, we performed shotgun proteomics on follicles cultured *in vitro* in 5X and 10X. We profiled the follicular proteome on Days 4, 8, and 12 of culture. Comparative analysis of 5X vs. 10X samples identified 161, 234, and 509 proteins with differential abundance ($p < 0.05$) at Days 4, 8, and 12 respectively. We first reviewed proteins with profiles concordant to their respective genes in the RNAseq dataset. Among the gene-protein pairs were metabolism-related genes *Fads1, Fabp5, Lpcat3, Atp5a1/5b,* and *Cox4i1/2* (Figure 3.5A). *Fads1* and *Fabp5* code for fatty acid desaturase 1 and fatty acid binding protein 5 respectively, while *Lpcat3* codes for an acetyltransferase involved in fatty acid breakdown. All three genes showed progressively higher transcription during the culture period and their respective proteins increased in abundance across the sampled time points (Figure 3.5A). Interestingly, these genes were not among the most statistically significant in our somatic cell RNAseq analysis, but their proteins show differing abundance in 5X vs. 10X. These findings also suggest FAO is not confined to the oocyte compartment of the follicle. The remaining metabolism-related genes, *Atp5b, Atp5a1,* and *Cox4i1/2* are involved in ATP synthesis within the mitochondria. Expression of these genes and abundance of their proteins peaked mid-culture with slightly higher protein abundance for ATP5A1 and ATP5B in the 10X condition (Figure 3.5B). We also observed concordant expression/abundance of vesicle signaling-related genes *Arf5*, *Napa*, and *Ykt6* (Figure 3.5C), which all increased in expression over the period of culture. Protein abundance generally increased over culture, with greater abundance in 10X versus 5X (Figure 3.5C). These findings augment the

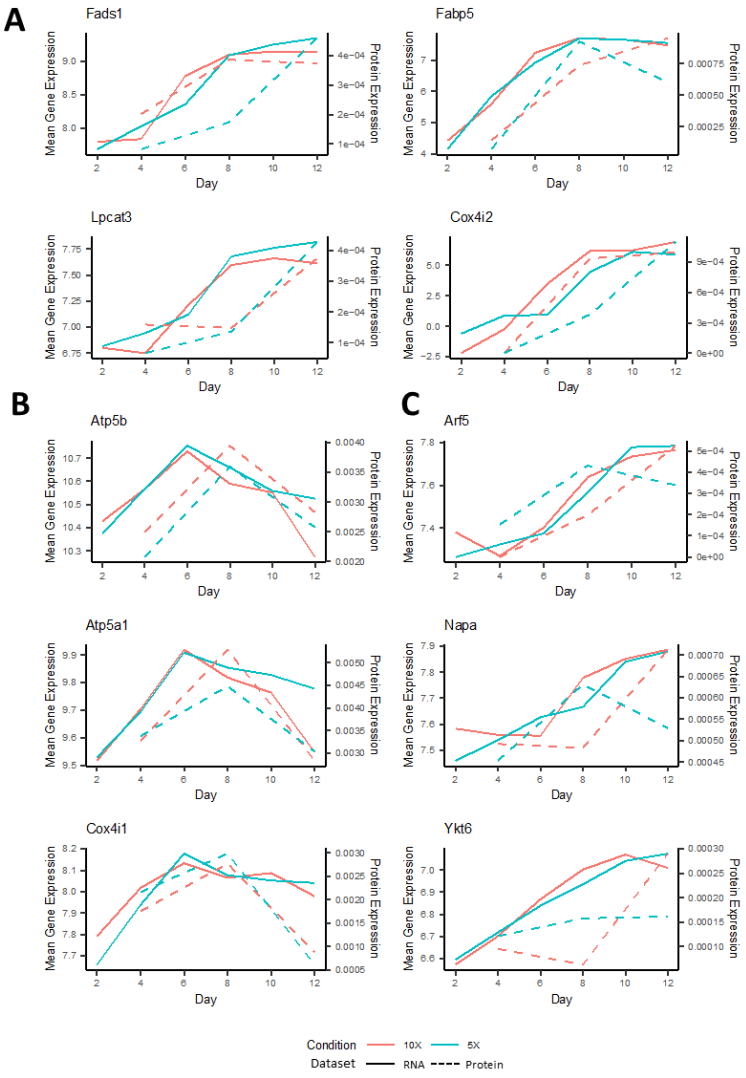


Figure 3.5: Proteomic profiling of follicles cultured *in vitro*. RNA (left axis) and protein (right axis) abundance over the culture period for genes involved in A) fatty acid oxidation, B) ATP synthesis, and C) vesicle-mediated signaling.

earlier RNAseq results presented in Figure 3.2A and provide further evidence for the importance and dynamic expression of vesicle signaling-related genes in follicle culture *in vitro*.

We also reviewed the proteomics data independent of genes from RNAseq to identify additional biological processes with differential activity in 5X vs. 10X. Pathways analysis revealed 64, 58, and 84 biological processes enriched in 10X follicles ($p < 0.05$, see Experimental Methods for details) on Days 4, 8, and 12 respectively by mapping protein IDs back to their genes.

The majority of gene ontology terms were shared across at least two time points, with numerous cell cycle-related proteins differing on all three days sampled (e.g. TOP2B, GNAI3, RPL23*;* Appendix Figure A.7A). The differential abundance of cell cycle-related proteins agrees with our previous findings in follicle growth kinetics (Figure 3.1C) and RNAseq analysis (Figure 3.2B). On Day 8, we identified differing patterns in 5X vs. 10X follicles for proteins related to cell proliferation (e.g. IGF1R, SHMT2, CALR; Appendix Figure A.7B). In 10X, proliferation-related proteins peaked in 10X on Day 8 of culture, while their relative abundance was dysregulated in 5X. Finally, the Day 12 comparison alone was enriched for a variety of proteins related to MAPK signaling (e.g. PEBP1, RAP1B, YWHAZ; Appendix Figure A.7C). MAPK signaling-related proteins generally increased over time in 10X, but peaked on Day 8 then decreased in 5X (Appendix Figure A.7C). The MAPK signaling cascade has been identified as an effector of numerous follicle-related processes, primarily granulosa cell proliferation and differentiation[84]. The Day 12 comparison also included numerous proteins related to negative regulation of apoptosis and cell survival (e.g. PAM16, PRKDC, MYBBPR1A; Appendix Figure A.7D). Interestingly, in 10X these pro-survival proteins maintained comparatively high abundance levels or increased over culture, while abundance in the 5X condition differed significantly ($p < 0.05$) on Days 8 and 12 (Appendix Figure A.7D).

3.5 **Discussion**

Here we have presented detailed transcriptomic and proteomic analysis of the group follicle culture system previously studied by our group and others[14,15,63]. The 5X/10X follicle culture system serves as a highly controlled, reproducible model for studying early follicle development *in vitro*, where individual primary follicles will not survive without paracrine support. This study builds on previous findings suggesting a synergistic profile in 10X follicles, which consistently

reach a sufficient terminal size of ~300 µm and produce mature oocytes, when compared to the smaller groups of 5X[15,63]. In our prior transcriptomic study using microarray, we identified a myriad of biological processes with differential activity in 10X versus 5X, constituting a group of genes and processes differentiating the superior and inferior culture conditions. Here, we have built on those findings with a more robust transcriptomic analysis platform (RNAseq) used to profile the follicular somatic cells in culture, and for the first time we have profiled the oocytes of follicles developing *in vitro* from the primary stage. We then validated a number of transcriptome-level findings using shotgun proteomics of follicles cultured in 5X and 10X.

Our RNAseq-based profiling of follicular somatic cells agreed with our previous microarray findings on many points and revealed new genes and processes with dynamic temporal and 5X vs.10X transcription. In our analysis of transcriptome changes over time (Figure 3.2B) we observed enrichment for genes related to angiogenesis (e.g. *Angpt1, Ptpn1, Nedd4*; Figure 3.2B) which agreed with similar findings in our microarray analysis. We also found patterns of gene expression related to cell proliferation and Wnt signaling that echoed our previous study (Figure 3.2B). Importantly, our analysis of 5X versus 10X somatic cell profiles identified a cluster of 54 angiogenesis- and VEGF signaling-related genes (e.g. *Vegfa, Egln1, Mif*; Figure 3.2C) that were more highly expressed in 10X than 5X throughout culture. This supports our previous reports on the synergistic production of angiogenesis-related cytokines in 10X culture[15] and differing expression of angiogenesis-related genes in 10X culture[63], suggesting that pro-angiogenic cytokines may serve as key exogenous factors for developing a well-defined, translational follicle culture system in the future. The somatic cell analysis also revealed new, additional genes and pathways, including expression of genes involved in vesicle-mediated signaling (e.g. *Rab3p, Syt1, Cspg5*; Figure 3.2B-C) and meiosis (e.g. *Hspa2, Syce2, Tex15*; Figure 3.2B). Vesicle-mediated

signaling has recently received more attention in the study of folliculogenesis, especially given the importance of signaling via secreted factors in the follicular fluid of antral follicles[85–87]. The dynamic expression of meiosis-related genes in follicular somatic cells may reflect the somatic compartment's role in supporting oogenesis[67–70]. These findings together imply that pro-angiogenic factors are crucial to successful follicle development *in vitro*, and constitute a group of new genes and processes important to folliculogenesis identified through modern sequencing techniques.

We also compared our somatic cell data to a recent report on the single-cell atlas of the cycling mouse ovary[75]. The authors reported marker genes for granulosa and theca cells across folliculogenesis, which we compared to our dataset. Over the period of culture, we observed decreasing expression of preantral granulosa cell markers and a coinciding increase in markers for antral granulosa cells (Appendix Figure A.6A-B). We also found dynamic expression of mitotic granulosa cell markers, which peaked on Days 6 and 8 of culture and sometimes lagged in the 5X condition, although the difference between 5X and 10X was not statistically significant (Appendix Figure A.6C). Interestingly, we observed dynamic expression of marker genes for early theca (Appendix Figure A.6E) and steroidogenic theca (Appendix Figure A.6F), which decrease and increased over time *in vitro*, respectively. Expression of these theca-specific genes, paired with confirmed production of androstenedione in cultured follicles[15], suggests that follicles isolated from the primary stage contain pre-theca cells that differentiate into steroidogenic theca cells *in vitro*.

We next turned to the transcriptome of oocytes from the group culture system. Numerous transcriptomic studies have been performed on mature oocytes obtained from mice or human donors via egg retrieval protocols, but our dataset for the first time reports the profile of oocytes

developing from the primary stage *in vitro*. We observed a number of dynamic temporal changes in the oocyte transcriptome, including changes in expression of FAO-related genes (e.g. *Echdc2, Pck2, Dbp*; Figure 3.3B). While FAO activity has previously been profiled in late-stage oocytes[76], our gene expression data suggests dynamic expression of FAO-related genes throughout oogenesis *in vitro*. We also identified temporal changes in genes related to organelle and microtubule organization (e.g. *Dnah7b/9/11, Cdc15b, Sybu, Stk36*; Figure 3.3B). Interestingly, these genes decreased in expression as culture progressed, prompting further inquiries into the timeline for oocyte cytoplasmic maturation leading up to the resumption of meiosis. Finally, our temporal analysis of oocytes revealed decreasing expression of genes related to PI3K/AKT signaling as culture progressed (e.g. *Pten Pik3r2/3, Pigk, Impa2*; Figure 3.3B), which supports previous studies on this pathway's importance for oocyte maturation [77]. Our comparison of 5X and 10X oocyte profiles yielded a surprisingly small number of significant genes (n=14) with few biological annotations to enlighten us to their significance (Figure 3.3C). The lack of transcriptomic differences between 5X and 10X oocytes suggests that the phenotypic differences observed in vitro, and the differential oocyte maturation outcomes, is driven by the somatic cell compartment of the follicle. Furthermore, future efforts to develop a well-defined culture system for individual early-stage follicles using exogenous supplements may be entirely focused on supporting proper somatic cell development, which in turn supports oogenesis[67–69].

To augment these RNAseq findings, we performed ligand-receptor analysis of the somatic cell and oocyte datasets. We identified 32 ligand-receptor pairs and highlighted 6 with strong reciprocal relationships between somatic cells and oocytes (Figure 3.4). Interestingly, all identified pairs were from signaling pathways with little biological annotation in follicles, including endogenous opioid receptor signaling, neuroligin-neurexin cell-cell adhesion, and neurotensin

signaling. All of these findings require further investigation using quantitative proteomic approaches and may serve as additional targets for improving follicle culture methods. Importantly, transcriptome-based ligand-receptor analysis does not encompass the many non-protein ligands and cues that drive follicle development, such as lipids and steroid hormones. These require further investigation using other methods.

We concluded our study with comparative proteomics of cells isolated from 5X and 10X follicles. Progress in profiling the proteome of ovarian follicles has been hampered in part by the delicate nature of follicle culture, which requires significant supplemental with serums such as fetuin. Here for the first time we have captured the proteomic profile of follicles cultured *in vitro* through isolation of thousands of follicles and proteomic analysis of cell pellet lysates. Future iterations of this work might focus on analysis of the secretome alone by profiling the spent media of follicles cultured in vitro, but the challenge of removing serum from culture media remains to be solved. In our analysis, we identified numerous points of validation for our RNAseq data, especially for genes and proteins related to FAO (e.g. FADS1, FABP5, LPCAT3) and vesicle-mediated signaling (e.g. ARF5, NAPA, YKT6) (Figure 3.5A,C). We also identified differential abundance in 5X versus 10X for proteins involved in MAPK signaling (e.g. PEBP1, RAP1B, YWHAZ) and cell survival (e.g. PAM16, PRKDC, MYBBPR1A) (Appendix Figure A.7), which were not evident on the RNA-level. These biological processes and specific proteins may also be used as targets for exogenous modulation to support follicle development *in vitro*.

3.6 **Conclusions**

In this study, we have reported a variety of genes, proteins, and biological processes in which they participate that undergo dynamic changes during follicle development *in vitro*. These transcriptome and proteome datasets, while mined here for the purposes of comparing the 5X and

10X conditions, may also serve as a reference for groups exploring the behavior of specific pathways or processes during follicle development. The results presented here also serve as a group of genes and proteins that are associated with successful follicle development *in vitro*, and thus may serve as candidates for exogenous supplementation in the development of a well-defined translational follicle culture system, which is urgently needed as a new assisted reproductive technology. Future work in this area should also focus on further investigation of specific processes not previously identified in the follicle, such as new cell-cell adhesion relationships (e.g. neuroligins and neurexins) and endogenous opioid peptide signaling, which may be crucial facets of folliculogenesis yet to be explored.

Chapter 4 Cellular Atlas of the Human Ovary Using Morphologically Guided Spatial Transcriptomics and Single Cell Sequencing

4.1 Preface

While numerous parallels have been drawn between murine and human folliculogenesis, the composition of the human ovary and a number of human-specific processes differentiate the two species. Progress in understanding human folliculogenesis, especially its early stages, has been stunted by the lack of human ovarian tissue for research. In this chapter, we use ST and scRNAseq of rare human ovary samples to uncover intra- and extrafollicular genes and biological processes that support follicle development and other ovarian functions like steroid hormone production. Our analysis uncovered a variety of processes that may support follicle development, including cues provided by the stroma around early-stage follicles, pathways upregulated in granulosa and theca cells from antral follicles, and spatially coordinated gene expression in the cumulus cells around the oocyte. We also confirm previous findings using single-cell RNA sequencing and provide robust cell-type specific marker panels for future studies. This chapter not only expands our understanding of ovarian biology, but charts a new course for future work aimed at deciphering the cues and processes that drive ovarian processes.

4.2 Introduction

The development, differentiation, and spatial organization of the many cell populations in the human ovary are essential for its reproductive and endocrine functions. The outer ~1 mm layer of the organ, the ovarian cortex, contains quiescent primordial follicles, which constitute the

follicular reserve, and transitioning and primary follicles. The inner part of the organ, the medulla, contains the growing secondary and antral follicles, as well as corpora lutea. Both regions are rich in vasculature and immune cells residing within the stroma, maintained by constant interactions among the stromal, immune, endothelial cells and other, rarer cell types. The functional unit of the ovary, the ovarian follicle, contains an oocyte in the center, surrounded by specialized somatic cells that secrete hormones and support the many steps of oocyte maturation. In women of reproductive age, a small portion of quiescent primordial follicles are activated in each menstrual cycle to join the pool of growing follicles. The growing follicles expand across both the cortex and medulla, each taking on a multilayered three-dimensional architecture containing the oocyte surrounded by cumulus granulosa, and outer layers of mural granulosa, and theca cells, which are separated from the cumulus-oocyte complex by a fluid-filled antrum. Paracrine crosstalk among the follicular cells and endocrine signals from the hypothalamic-pituitary-ovarian axis work together to trigger ovulation, which produces mature eggs for fertilization. The cellular diversity of the ovary and the complex spatially defined structures in growing follicles have been difficult to study, largely due to the scarcity of tissue from healthy women of reproductive age, the lack of unbiased functional profiling, and the technical limitations of spatial transcriptomics (ST). Past studies using animal models have identified major classes of ovarian structures and their putative functions. However, human follicle development and the role of non-follicular cells in this process remain poorly understood. A ST-based characterization of the human ovary is critical to understanding follicle development, hormone production, and ovarian aging.

The ovarian cortex is of particular research interest as the home of the follicular reserve and the microenvironment in which follicle activation occurs. In recent years a few single-cell sequencing studies have reported that the cortex and the medulla contain the same general cell

types[9], but the two regions are histologically distinct, e.g., having different extracellular matrices tailored to their functions, and different gradients of growth factors[88]. The mechanisms driving follicle activation, either within the follicular structure or from the microenvironment outside the follicles, are poorly understood. Studies of the oocyte compartment of early-stage follicles have aimed to identify markers of quiescence and activation; but isolation of oocytes for sequencing may alter their transcriptome or lead to spontaneous activation[89,90], underscoring the need to collect data from intact tissues to understand what defines these early stages of follicle development. With the newly emerged ST technology, early-stage follicles and the surrounding cortical stroma can be sampled with minimal perturbation to uncover cortex-specific gene activities related to early folliculogenesis. Likewise, ST presents a previously unattainable opportunity to profile the gene activities in diverse types of follicles recognized by their morphology and spatial context. Previous studies have reported on the transcriptional profiles of granulosa and theca cells[8,9], but hypotheses surrounding oocyte-cumulus granulosa cell bidirectional crosstalk and factors influencing the follicular cell phenotypes have not been validated using a spatial approach.

In this study we compiled a new, functionally targeted ST dataset of the human ovary, comprised of 257 regional samples collected using Nanostring's *GeoMx™* platform from the ovaries of two reproductive-age, premenopausal donors. We also collected single-cell RNA sequencing data of 21,198 dissociated cells from three additional donors. Using these data we uncovered transcriptional profiles of early-stage oocytes sampled from their native microenvironment, gene signatures of theca and granulosa cells, spatially defined patterns of gene expression across the medulla and cortex, and major stromal and immune cell types of the human ovary. Comparisons between our ST and single cell RNA sequencing (scRNAseq) datasets confirmed the robustness of the *GeoMx*™ platform for sampling specific local regions of small

size and distinct function. This integrated analysis led to a new cell atlas of the healthy reproductive-age human ovary, revealing spatially defined transcriptional patterns and adding to our knowledge on the cellular diversity of this organ.

4.3 Experimental Methods

4.3.1 Ovarian Donor Tissue Processing

This study used tissue samples from five de-identified deceased donors procured through the International Institute for the Advancement of Medicine (IIAM) and the associated Organ Procurement Organization (OPO) involved in the harvest. All five donors were pre-menopausal and examination of provided medical records indicated no pathological conditions impacting ovarian function (for their age, BMI, recorded "race" and Cold Ischemic Time, see Extended Data Fig. 1A). Before cross-clamp, the organs were perfused with Belzer University of Wisconsin® Cold Storage Solution (Bridge of Life, SC, USA), Custodiol® HTK (Histidine-Tryptophan-Ketoglutarate) Solution (Essential Pharmaceuticals, NC, USA), or SPS-1 Static Preservation Solution (Organ Recover Systems, IL, USA).

4.3.2 GeoMx™ Spatial Transcriptomics Experiments

4.3.2.1 Slide Preparation and ROI Sample Collection

The cross sections of fresh ovarian tissue from Donors 1 and 2 were resected from the central region of the organ along the short axis, then fixed in 4% paraformaldehyde (Fisher Scientific) for 24 hours at 4°C and rinsed five times for 24 hours each in $DPBS^{-/-}$ at 4°C. The tissue was then embedded in paraffin for sectioning. Every other 5 µm section was stained with hematoxylin and eosin (Fisher Scientific) to select an unstained slide for spatial analysis. The selected slides from Donors 1 and 2 were deparaffinized and rehydrated, followed by antigen

retrieval for 15 minutes at 90°C in Tris-EDTA (Fisher Scientific). RNA targets were exposed using proteinase K (1 μg/mL) in phosphate-buffered saline (PBS) for 15 minutes at 37°C. For *in situ* hybridization, *GeoMx*™ reagents and protocols were used according to the manufacturer's instructions. The slide was stained with morphology markers prior to RNA collection. For Donor 1, we used SYTO82 (a nuclei marker, ThermoFisher, US) and 3 protein markers: anti-smooth muscle actin to visualize pericytes and smooth muscle cells (SMA, Alpha-Smooth Muscle Actin Monoclonal Antibody conjugated to Alexa Fluor 488, ThermoFisher, US), anti-Ki67 to visualize proliferating cells (rabbit monoclonal 9027S, Cell Signaling Technologies, US), and anti-CD68 (sc-20060 AF594, Santa Cruz Biotechnology, US). For Donor 2, we again used SYTO82 and anti-CD68, along with anti-DAZL to visualize oocytes (MCA2336, BioRad, US) and anti-PGR to visualize cells expressing progesterone receptor such as luteinized follicular cells (ab63605, Abcam, US).

4.3.2.2 Nanostring Library Preparation and Sequencing

The prepared slide was covered with Buffer S (NanoString, US) and loaded into the NanoString *GeoMx*™ DSP instrument, which uses digital images of a slide-mounted tissue section, with up to 4 fluorescently labeled markers, to guide the selection of Regions-of-Interests (ROIs). ROIs are defined by the user based on cell morphology and related tissue features, using the "polygon", "contour", and "segmentation" tools provided in GeoMx DSP. Target-enrichment ROIs were segmented into areas of interest (AOIs) based on staining for CD68, Ki-67, SMA, DAZL, and PGR. To collect concentric ROIs from ovarian follicles, a mask was designed in ImageJ(Schneider et al., 2012)4747 and imposed on the follicle's region in the tissue.

In this experiment, RNA molecules are affixed by photocleavable linkers; then programmable UV light is used to release RNA from each ROI. RNA samples are collected by

microcapillary aspiration into microtiter wells for sequencing-based analysis. In some cases, the selection was aided by histological images of one or both of the flanking tissue sections, which were not used for sample collection. Some of the ROIs are serial subdivisions of the same morphological area to capture local spatial changes. Some of the ROIs produced 2-3 AOIs/samples each due to targeted enrichment using 1 or 2 of the protein markers. Each sample is associated with additional information about its location, surface area (μm^2), and number of nuclei (an estimate of cell number).

4.3.2.3 Nanostring Data Processing

We used the *GeoMx*^{TM} analysis software to process the NanoString data, including steps for read decoding, trimming, and deduplication, and obtained the gene-by-sample read count data for 18,676 genes passing *GeoMx*'s default quality control (QC) criteria. This is after removing gene #1309, "NegProbe-WTX", from the original data matrix, For Donor 1, we initially collected 94 samples; but two yielded fewer than 100,000 reads and were removed, leaving 92 samples, which came from 74 ROIs. They represent 14 "ROI Types", as described in Figure 4.1F. The 92 samples had library size (i.e., total read count) ranged in 146,463-4,464,464 (mean = 805,579), surface area ranged in 7,619-359,037 $\square m^2$ (mean = 60,591 $\square m^2$), and nuclei counts ranged in 0-2,544 (mean = 424). Raw counts, which were already floored at 1 rather than 0 by NanoString, were normalized into counts per million (CPM), then log-transformed with base 2, and quantile normalized. These 92-sample normalized data were used in the analyses shown in Figure 3A. For all other analyses presented, we condensed the target-enriched AOIs from the same ROI into a single sample by summing up the reads over the AOIs, leading to the reduction of 29 AOIs to 11 ROIs, and the total sample number from 92 to 74. These 74 samples were similarly normalized to CPM, log-2

transformed, quantile normalized, and subjected to principal component analysis (PCA) and k-means clustering (k=6) shown in Figure 1D.

For Donor 2, the 165 samples from 147 ROIs were normalized to CPM, log-2 transformed, quantile normalized, with PCA and k-means clustering (k=9) shown in Figure 1E and Extended Data Fig. 1E, 2B, 4B. The 165 samples from Donor 2 had library sizes ranged 24,862-3,404,417 (mean = 506,800, surface area 577-113,205 μm^2 (mean = 30,152 μm^2), and nuclei counts 0-1,497 (mean = 240). ROI selection for Donor 2 covered 12 ROI Types, including seven of the 14 ROI Types analyzed in Donor 1 and five new ROI Types, bringing the total to 19.

While the analyses was based on NanoString's floor=1 raw counts data, we examined an alternative method: minus 1 on all counts data and re-calculate CPM, then obtain log2(CPM+1). We found that the difference is minor in terms of the PCA projection of the 92 samples, thus we opted to stay with the log2(CPM) method on the original, floor=1 data. The quantile-normalized data for 74 ROI-based samples for Donor 1 and the 165 AOI-based samples for Donor 2, were shown in Figures 2D and 3C, for the marker genes for oocyte, theca cells, and granulosa cells, as discussed in the main text. These quantile-normalized data were also used in PCA for subsets of samples, shown in Figures 3B, 4D and Extended Data Fig. 4B-D.

4.3.2.4 Cluster Analysis of ST Data

For Donor 1, PC1-PC2 in Figure 1D revealed the separation of 5 clusters (C1-C5). The sole oocyte sample separated itself in PC3 (Extended Data Fig. 1C), and was designated Cluster 6 (C6). For Donor 2, clustering was performed on all 165 samples (without consolidation of target-enriched samples) to highlight the unique signature of Early Stage Oocytes (ROI Type #4) captured by DAZL, with k-means clustering (k=9) shown in Figure 1E.

4.3.2.5 Differential Expression Analysis

We performed 17 series of DE analysis for Donor 1 using the quantile normalized log2(CPM) data for 74 ROIs. Each DE analysis compares two groups of ROIs, as described below, and produced four statistics: fold-change, t-score, raw p-value, and adjusted p-value, i.e., the Benjamini-Hochberg False Discovery Rate. Fold change is reported on the log2 scale, and the t-score is from t-test without assuming equal variance in the two groups being compared.

The 17 DE series for Donor 1 are for comparing samples in each cluster with those in all other clusters (*cluster1vsAll, cluster2vsAll, cluster3vsAll, cluster4vsAll, cluster5vsAll*), comparing one cluster with its "adjacent" cluster (*cluster1vs2, cluster2vs3, cluster3vs4, cluster4vs5*), comparing target-enriched areas against their matched unenriched AOIs using paired-t test: *SMA* (SMA+ areas against their matched, untargeted area), *CD68* (CD68+ areas against their matched, untargeted area), *Ki67* (Ki67+ areas against their matched, untargeted area), gradient in the *Cortex surface* (linear regression across the 11 consecutive cortex surface layers, Figure 4B), in *Ring34* (regression across the 4 consecutive cumulus cell rings around Antral Follicle 1, shown in Figure 3D-E), *Ring52* (regression across the 5 consecutive rings around the secondary follicle), *cluster6vsAll* (only fold-change, as there is only one sample in Cluster 6), and *cluster5vs6* (only fold change). The five sets of one-cluster-versus-all-other comparison (such as *cluster1vsAll*) are based on the 73 ROIs, without using the oocyte sample. Positive FC and t-score values indicate that the expression is higher in this one cluster. In the four sets of one-cluster-versus-the-adjacent comparison (such as *cluster1vs2*), positive FC and t-score values indicate that the expression is higher in the second, higher numbered cluster. The three DE results for targeted AOIs are based on paired t-tests for SMA (8 pairs), CD68 (3 pairs), and Ki67 (7 pairs). Positive values mean higher expression in the enriched AOIs. This analysis relied on the 29 targeted samples from the 11 ROIs,

using quantile normalization among the 29 rather than the quantile normalized data across the entire set of 92 or 74 samples. In the *Cortex Layer* gradient analysis, positive values mean higher expression towards the surface. Fold change is the fitted slope over the series of layer numbers, 1 to 11. Since all fold changes are in the log2 scale, a slope of 1/11 means a FC of 0.0909 per layer, or a FC of 1.0 overall, or 2 fold increase after 11 layers, spanning about 330 □m. In Antral Follicle 2, the oocyte is not present in the sampled section, although it is discernible in the H&E image of the adjacent section. For the regression result of the gradient across the 4 rings in "Ring34" and 5 rings in "Ring52", positive values mean higher expression towards the outside of the follicle. Fold-change is the slope, in the unit of per-ring. For instance, in ring34, a FC of 1/4 would mean a FC of 1 overall, or two-fold change between the first and the 4th ring. Note that the center ROI of follicle, for both ring34 and ring52, was not used in the regression.

For Donor 2, we performed 12 series of DE analysis using the quantile normalized log2(CPM) data for all 165 ROIs. Each DE analysis compares two groups of ROIs, as described below, and produced the same four statistics: fold-change, t-score, p-value, and FDR-adjusted p-value. The 12 DE series are: *Cumulus vs. Non-Cumulus Granulosa* (ROI Type #13 vs. Type #12, denoted d13.12), *Theca vs. Granulosa* (ROI Type #14 vs. Types #12-13, denoted d14.1312), *oocyte vs. other* (ROI Type #4, n= 11, vs. the other 154 ROIs, denoted oocyte.other), *Medulla vs. similar Cortex* (Cluster 5, n=38, vs. Cluster 2, n=28, denoted d5.2), *Granulosa vs. other* (ROI Type #12-13, n=18, vs. all other ROIs minus the 11 samples for early-stage oocytes, n =136, designed to be comparable to Donor 1 granulosa DE results, denoted gc.other), *Theca vs. other* (ROI Type #14, n=7, vs. all other ROIs minus early-stage oocytes, n=118, designed to be comparable to Donor 1 theca DE results, denoted tc.other), *Ovarian cyst vs. other cortex* (Cluster 3, for ovarian cyst, n=21 vs. Cluster 1, n = 16, denoted d3.1), *Within-cortex comparison* (Cluster 2 cortex samples,

n=28, vs. Clusters 1&3, n = 37, denoted d2.13), *CD68-enriched* (n=6) vs. all other ROIs minus early-stage oocytes (n=130) (designed to be comparable to Donor 1 CD68-enriched DE results, denoted cd68.other), *Medulla vs. proximal cortex* (Cluster 5 Medulla samples, n=38, vs. Clusters 1&3 cortex samples, n=37, denoted d5.13), *Medulla vs. all cortex* (Cluster 5 Medulla samples, n=38, vs. Clusters 1-3 cortex samples, n=65, denoted d5.123), and *follicle rich vs. follicle null* (the nine DAZL+ AOIs vs. surface-most samples from the three cortex series, ROIs #61, 67, 95, denoted richvnull, shown in Figure 4E).

4.3.2.6 Pathway Analysis of DE Results

We used *LRpath*[27] to examine the enrichment pattern of KEGG and Gene Ontology terms in each of the DE results. We converted gene symbols to Entrez IDs and uploaded the p value (for ranking significance) and fold-change (for direction of change) as input data. We selected "directional analysis" to separate enrichment signals for up- and down-regulation. For all 8,423 GO and KEGG terms and each DE series we collected odds ratio (OR) of enrichment, p value, and FDR. When OR >1, the genes in this pathway are enriched among those with positive t scores in the gene-level analysis. For the two DE analyses involving Donor 1 Cluster 6, since there is only one oocyte sample and only fold-change (with t-score or p-value), we created a mock p value based on fold-change so that both significance rank and direction are available in the input file. For some of the comparisons, we only evaluated pathways with positive t scores (OR>1), as the other direction is the “background” (e.g., one cluster vs. all other). Note that pathway terms with too few genes tend to be noisy and have less significant P values. Those with too many genes tend to have significant p values even when the enrichment ratio is moderate.

4.3.2.7 Identification of Marker Genes for Oocyte, Theca Cells, and Granulosa Cells

The 76 oocyte marker genes (Figure 3D) came from the *oocyte vs. other* FC>2. The 45 theca cell marker genes (Figure 4C) came from *Theca vs. Granulosa* FC>1.2 and *Theca vs. other* FC>1, minus one gene that also appeared in the 76 oocyte markers. The 74 granulosa cell marker genes (Figure 4C) came from *Theca vs. Granulosa* FC<(-1) and *Granulosa vs. other* FC>1.5, minus two genes that also appeared in the 76 oocyte markers.

4.3.3 Single-Cell RNA Sequencing Experiments using 10X Chromium

4.3.3.1 Tissue Dissociation, Cell Sorting, and Sequencing

We collected scRNAseq data from three additional donors (Extended Data Fig. 1A). Upon arrival, ovarian tissue (cortex and medulla) from Donors 3-5 was chopped into ~1 mm cubes at room temperature in Quinn's Advantage Medium with HEPES (CooperSurgical, Denmark) with 10% Quinn's Advantage Serum Protein Substitute (CooperSurgical). These tissue cubes were rinsed twice with Dulbecco's Phosphate Buffered Saline without calcium or magnesium ($DPBS^{-/-}$) (Fisher Scientific, US), then transferred to a digestion solution containing 0.5 mg/mL Collagenase IA (Sigma Aldrich, Germany) and 0.01 mg/mL DNase I (Worthington Biochemical, US) in DPBS with calcium and magnesium ($DPBS^{+/+}$, Fisher Scientific). Tissue was transferred to a shaker of 150 rpm to digest for 30 minutes at 37°C, then the tissue was strained through a 70 μm strainer (Fisher Scientific), followed by a 30 μm strainer (PluriSelect, US). The cell suspension was quenched with ice cold 10% fetal bovine serum (FBS, Fisher Scientific) in $DBPS^{-/-}$. The remaining tissue with undissociated cells was placed in fresh digestion solution and shaken for an additional 30 minutes before straining and quenching. Cell suspensions were stored on ice, treated with red blood cell lysis buffer (BioLegend, US) according to the manufacturer's instructions. Suspensions were then rinsed in 1% bovine serum albumin (BSA, Fisher Scientific) in $DPBS^{-/-}$

followed by 0.04% BSA in $DPBS^{-/-}$ to remove debris. Cells were pelleted at 100g for 5 minutes between all rinsing steps. Cortex and medulla were enzymatically digested separately, then combined into a single cell suspension at a 1:1 ratio. Combined cortex/medulla single-cell suspensions from tissue dissociation were incubated with 3 μM 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI, Fisher Scientific) per 1 million cells for 30 minutes on ice in preparation for fluorescence-activated cell sorting (FACS) at the University of Michigan's Flow Cytometry Core. Cells were sorted on a MoFlow Astrios (Beckman Coulter, US) to remove dead cells (DAPI+) and collect live cells (DAPI-). FlowJo v10 software was used for gating and cell counting. We performed three runs of scRNAseq using dissociated cells from Donors 3, 4, and 5, respectively, thus creating three experimental batches. Freshly sorted cells were submitted to the Advanced Genomics Core at the University of Michigan, loaded to the 10X Genomics Chromium controller for droplet-based single-cell capture. Cell lysis and RNAseq library preparation used the Chromium Next GEM Single Cell 3' LT Kit according to the manufacturer's instructions. Libraries were sequenced on NovaSeq F0 using an S4 flow cell. The Core performed basic quality control (QC) and read alignment using standard procedures, and provided unique molecular identifier (UMI) counts data in a cell-by-gene matrix for each sample.

4.3.3.2 Single-Cell RNA Sequencing Data Analysis

scRNA data from the three donors were initially processed by the U-M Advanced Genomics Core using Cell Ranger v4.0.0. The main steps include the extraction of cell-barcodes and the Unique Molecular Identifiers (UMIs) from the raw paired-end sequencing reads, alignment to human Ensembl genes, and UMI-based deduplication, leading to a cell-by-gene UMI count table, represented by the "filtered_feature_bc_matrix" for 14,322, 13,901, and 9,149 cell for the three samples, respectively, for 20,886 genes.

Downstream analysis used a combination of Seurat and custom codes in R. Cell filtering used (1) minimal number of UMI (called "nCount" in Seurat) and (2) % of transcripts corresponding to mitochondria-encoded genes ("% MT"). The cutoff values varied for the three samples: nCount: 2,300, 1,275, and 1,096; % MT: 15,12, and 15, respectively. The cutoffs were chosen based on each sample's distribution of nCount and % MT (not shown). After filtering, there are 7,571, 7,228, and 6,339 cells left for further analysis. The average of nCount, # of genes detected ("nFeature"), and % MT for these cells are shown in Extended Figure 1A.

4.3.3.3 Identification of Major Cell Types

We normalized the counts data to CPM, and analyzed the log2(CPM+1) data for the three donors separately. Initial analyses showed that Donor 5 contained the greatest proportion of non-stromal cells, and we analyzed this donor's 6,339 cells first. We selected 2,034 genes of high expression and high variability for PCA, and used the top 50 PCs for k-means cluster (k=14) and t-SNE projection (Figure 5A). The 14 clusters can be merged to five main clusters (Figure 5A), corresponding to Stroma cells, two apparent subtypes of Immune cells, Pericyte, and Endothelial cells, as annotated by marker gene expression (Figure 5C-D). Of the 6,339 cells, 5,399 (85.2%) were stromal cells.

For Donors 3 and 4, since the stromal cells dominated more strongly than in Donor 5, k-means clustering could not readily separate the non-stromal cells even though they appeared in distinct clusters in t-SNE (Extended Data Fig. 5A-B). We used the centroids of the five main cell types in Donor 5 to perform supervised annotation for cells in Donors 3 and 4 (i.e., a cell is assigned to a cell type by the maximal correlation among the 5 centroids), with the resulting cell label shown in Extended Data Fig. 5A-B. For Donors 3 and 4, the analyses proceeded by the selection of 3,702 and 2,251 genes, respectively, that were highly expressed and highly variable, followed by PCA

and t-SNE projection using the top 50 PCs. After supervised cell annotation, 7,438 of the 7,571 cells (98.2%) in Donor 3, and 7,052 of the 7,228 cells (97.6%) in Donor 4, were stromal cells. Marker genes shown in Figure 5D came from a series of DE analyses comparing one cell type against the other, as was done for the immune cell types described below.

4.3.3.4 Identification of Four Immune Cell Types

For immune cells, we combined the 59, 133, and 671 immune cells from Donors 3,4, and 5, respectively, and re-ran PCA for these 863 cells. UMAP projection using the top 20 PCs revealed four immune cell subtypes (Figure 5B). Marker-based annotation (Figure 5C-D) identified them as NK cells (n=105), T cells (n=409), Macrophage (n=319), and Mast cells (n=30).

4.3.3.5 Lack of Heterogeneity Among Stromal Cells

We attempted to identify subtypes of stromal cells, or gradients among them that could reflect shifting regulatory states. We observed that the stromal cells in our datasets were strikingly uniform in their expression profile. The donor-to-donor variation (i.e., the "batch effect"), while not a prominent factor in the identification of major cell types or immune cell subtypes (as shown in Figure 5), became dominant when characterizing the variation among stromal cells. Further, slight changes in gene selection, or in the algorithm dealing with the 0-counts (e.g., using CPM or CPK - counts per thousand), dramatically altered PCA, subsequent UMAP or t-SNE projection, or the k-means clustering results. Most projections and clustering solutions were driven by the cells' "library size", hence by their length of 0- and 1-counts, suggesting that there is no discernible heterogeneity among the stromal cells.

For instance, among the 5,399 stromal cells in Donor 5, we selected 4,881 cells of the largest library size, then selected 3,473 most highly expressed genes for use in PCA, and in t-SNE projection using the top 50 PCs. In parallel, we re-ran quantile normalization using the 4881-by-

3473 data, and subsequent PCA and t-SNE. In the third option, we "imputed" 0 CPM values in each sample with the half of the minimal non-zero CPM value, calculated CPM again, and ran quantile normalization of log2(new-CPM), followed by PCA and UMAP and t-SNE projections. In additional tries, we ran analysis on a more stringent set of 2,122 cells, and repeated the procedure on Donor 3 and Donor 4 stromal cells with multiple selections of the strongest cells and strongest genes. In all case, the projections failed to reveal separate groups of cells (not shown), and the persistent unimodal distribution was driven by gradients of library size.

In an alternative, supervised approach, we took advantage of the robust gene sets from the NanoString data for oocytes, theca cells, and granulosa cells (Figures 2D, 3C) to calculate an oocyte score, a theca score, and a granulosa score for each stromal cell. This is essentially an attempt to identify the rare subsets of stromal cells that could have been oocytes or theca or granulosa cells that were difficult to uncover using unsupervised approaches. For each cell, we summed the expression values over the 76 oocyte marker genes, 45 theca markers, and 94 granulosa markers, in either the linear CPM space or the log(CPM+1) space (or CPM imputed with half-minimal values), with or without normalizing each cell's score by its library size, in either the linear or the log space. Of all the combinatorial options of constructing such scores, a few showed that some cells seemed to have a theca-like character and were near each other in UMAP, and some of them also showed a granulosa-like character, An example is shown in Extended Data 5B. However, they do not appear as a distinct subtype of stromal cells. There is no evidence that cells at the peripheral of the UMAP projection were hybrid cells with non-stromal cells (using markers for major cell types; not shown).

4.4 **Results**

4.4.1 Spatial transcriptomics of the human ovary

We used the *GeoMx*^{TM} technology to profile samples from 74 regions of interest (ROIs) in a whole-ovary tissue section from a healthy donor, followed by samples from 147 ROIs collected from a second donor. The donors were post-pubescent and pre-menopausal (18 and 27 years old) and had no known reproductive disorders (for complete donor information, see Appendix Figure A.8A). The ROIs were individually selected by histological features in different tissue locations to represent known functional units in the ovary (Figure 4.1A). The 92 samples in Donor 1 represented 14 ROI types; and the sampling in Donor 2 was designed to repeat some functional types while extending to additional types based on the analysis of Donor 1 data. In all, we analyzed 19 pre-specified ROI Types (see additional details in Appendix Figure A.8B and Methods). Some ROIs produced multiple split samples from the use of antibody-guided collection from sub-regions within the ROI (called Area-of-Interest, AOIs), leading to 92 and 165 samples respectively from the two donors. A brief description of the ROI types is as follows. Types 1-5 represent regions in the ovarian cortex: *Ovarian Surface*, *Cortex*, *Cortex Layers*, *Early-Stage Follicles*, *and Primary Follicle Rings* (Figure 4.1A). Types 6-9 represent regions in the medulla: *Medulla*, *Follicle-Adjacent Medulla*, *Medulla Cross Section*, and *Vascular Regions*. Types 10-14 were sampled from growing ovarian follicles (Figure 4.1B-C): *Secondary Follicle*, *Late-Stage Oocyte*, *Cumulus Granulosa*, mural *Granulosa (referred to from now on as granulosa)*, and *Theca* (Figure 4.1A). Types 15-16 were sampled from an ovarian cyst in Donor 2: *Cyst-Inner Layer* and *Cyst-Outer Layer* (Figure 4.1A,C). The remaining types were sampled using antibody-guided selection of AOIs enriched for specific cell types: *CD68+* (Type 17) for macrophages, *SMA+* (Type 18) for myofibroblasts and smooth muscle cells, and *PGR+* (Type 19) to identify cells expressing progesterone receptor (PGR) (Figure 4.1A).

We removed ROIs collected using an anti-Ki67 (targeting proliferating cells) from analysis due to poor antibody localization (see Methods for details). The ROIs adopted different sizes and shapes and varied by cell counts and sequencing depth. For example, the 74 ROIs from Donor 1 covered an average area of ~60,000 μm^2 (range: 7,619-359,038 μm^2), contained a monolayer of ~400 cells (range: 0-2,544, based on nuclei counts), and yielded ~806K identified transcripts (range: 57,018-14,388,977) among the 18,676 unique protein-coding genes analyzed by the *GeoMx*™ Human Whole Transcriptome Atlas panel. Likewise, the 147 ROIs from Donor 2 covered an average area of ~30,000 μm^2 (range: 576-113,204 μm^2), contained an average of 240 cells (range: 0-1,497), and yielded an average of ~507K transcripts (range: 24,863-3,404,489).

Cluster analysis of the 74 ROI-based samples in Donor 1 and the 165 AOI-based samples in Donor 2 (Methods) revealed 6 and 9 clusters, respectively (Figures 4.1D-E). Strikingly, even though each region contained a mixture of a few hundred cells, their observed clusters mapped to the pre-recognized functional types of ROIs from the cortex, medulla, and follicle-enriched sites, and were consistent between the two donors (highlighted by color-coded cross-tabulation in Figure 4.1F). For example, Cluster 1 in Donor 1 contained 21 of the 22 ROIs from the cortex (ROI Type 1-4); and Clusters 2 and 3 corresponded to the five ROI types (6, 7, 9, 17, and 18) from the cortex to the medulla (Figure 4.1F). The ROIs enriched for theca cells and those enriched for cumulus granulosa (Type 12) and granulosa cells (Type 13) aggregated into well-separated Cluster 4 and Cluster 5, respectively, in perfect consistency between the two independently sampled follicles

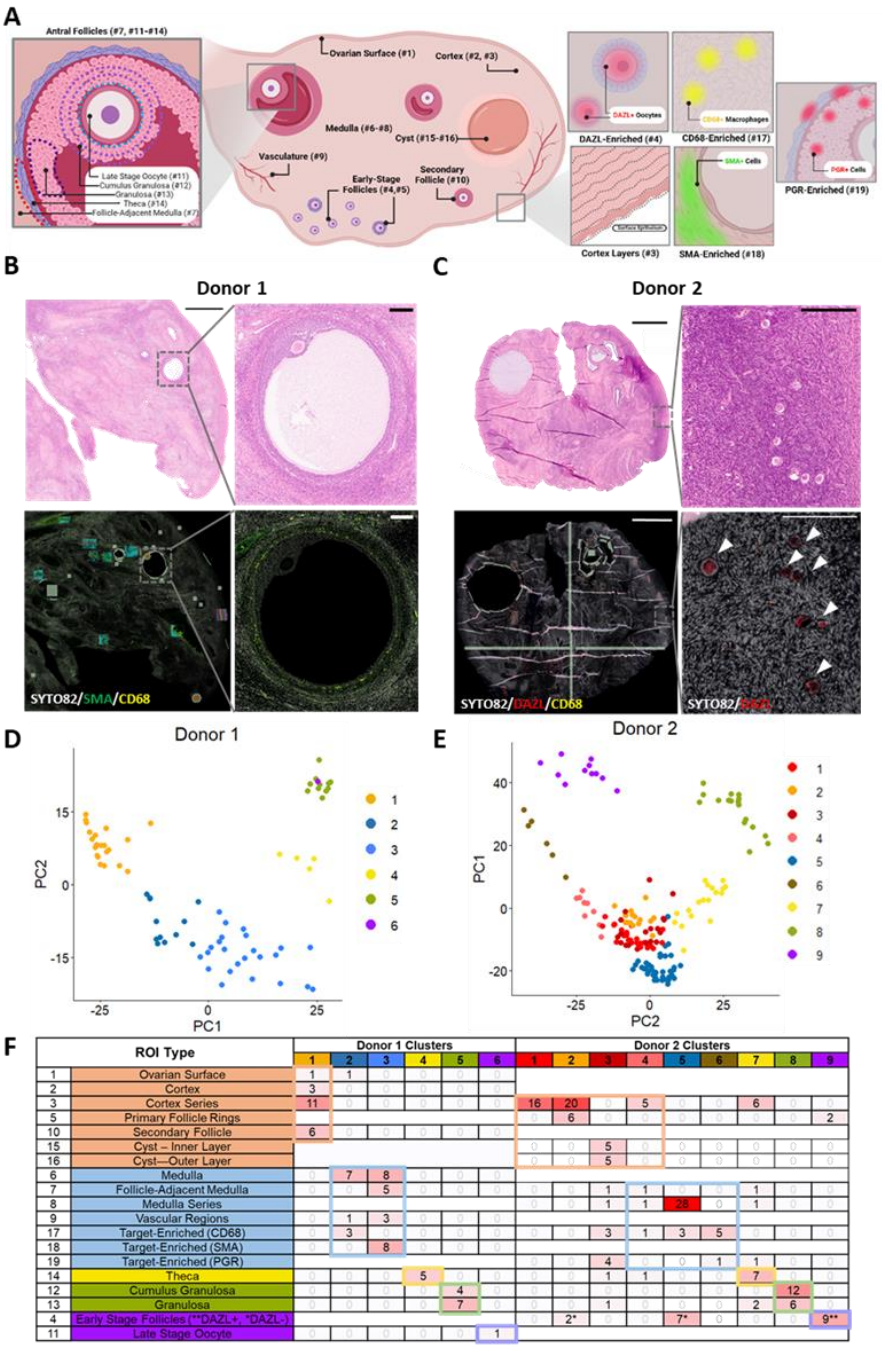


Figure 4.1: Spatial transcriptomics analysis of the human ovary. A. An illustrated summary of ROI types (all diagrams created with biorender.com). We used the histological image of an adjacent whole-ovary tissue section to select ROIs to represent 19 functional types of local cell communities. B. H&E (top panels) and immunofluorescent (IF) images (bottom panels) of the fixed tissue section from Donor 1. Left panels: whole tissue section. Right panels: zoomed-in view of the antral follicle. Gray, green, and yellow colors indicate SYTO82, SMA, CD68 staining signals, respectively. Scale bars: left - 2.5 mm; right - 200 µm. C. H&E and IF images of the fixed tissue section from Donor 2, in the same scale and layout as in B, except that SMA was not used, and red indicates DAZL signal. D,E. Principal component plots (PC1-PC2) of the 74 ROIs from Donor 1 (D) and 165 ROIs from Donor 2 (E), colored by the clusters identified. F. Annotation of the observed clusters by their mapping to the pre-defined ROI types. Shown is sample number cross-tabulation between the 19 ROI types (rows) and the 6 and 9 clusters (columns) in the two donors. The major categories of ROI types and their mapped clusters are highlight by color: orange = cortex, blue = medulla, yellow = theca, green = granulosa, purple = oocytes.

(Figure 4.1D,F). The theca- and granulosa-enriched ROIs were spatially adjacent (Figure 4.1B insets), thus the fact that their expression profiles were distinguishable (as Clusters 4 and 5) confirms the location-specificity of the technology. Cluster 6 is formed by the single ROI taken from the oocyte of an antral follicle (Figure 4.1F) and it was clearly separated in PC3 (Appendix Figure A.8C). Similarly, in Donor 2, Clusters 1 and 2 corresponded to the surface regions of the cortex, while Cluster 3, 4, 5, 6 mainly contained ROIs sampled from cortex and medulla to represent local characteristics, such as the ovarian cyst and antibody-targeted subareas enriched for CD68+ and PGR+ cells (Methods). Donor 2 data confirmed and extended these patterns. For example, ROIs intended to capture theca cells and granulosa cells arose in cluster analysis as distinguishable groups: Cluster 7 for theca and 8 for granulosa. Further, nine ROIs in Donor 2 had a DAZL+ subarea analyzed; and they stood out as a new Cluster 9, along with two additional oocytes from early-stage follicles (Figure 4.1E,F). In short, even though each of the locally sampled regions contained a mixture of cells, we were able to reproducibly identify robust gene activity signatures for the oocytes (Cluster 6 in Donor 1 and Cluster 9 in Donor 2), the theca cells (Clusters 4 and 7 in the two donors), and the granulosa cells (Clusters 5 and 8), along with local characteristics in pre-selected ROI types across cortex and medulla. In the sections below we will describe these findings in more detail.

4.4.2 Gene expression profile of early-stage oocytes

As described above, we profiled an oocyte sample from an antral follicle in Donor 1 and it had a distinct gene signature (as Cluster 6). To increase sample size, we selected 9 ROIs in Donor 2 from cortex areas enriched for primordial and primary follicles and used the antibody against the oocyte-specific protein DAZL to target the oocyte subarea within each ROI (Methods; Figure 4.2A), leading to 9 pairs of DAZL+ and DAZL- samples. In addition, we selected two ROIs

without antibody guidance from the center of two primary follicles (Appendix Figure A.9A). The 9 DAZL+ samples, along with the two primary follicle samples, formed Cluster 9 (Figure 4.1E-F), while the 9 DAZL- samples were similar to the medulla and cortex samples (Fig. 4.1F, Appendix Figure A.9B). The eleven Cluster 9 samples, when compared with the other 8 clusters, showed higher expression of previously reported oocyte markers such as *TUBB8*, *ZP3*, *LHX8*, and *OOSP2*[9] (Figure 4.2B), and uncovered new oocyte markers such as *PADI6, UCHL1, ZFAND2A, and REC114* (Figure 4.2C). *PADI6*, which codes for a member of the peptidyl arginine deiminase family, localizes to cytoplasmic lattices in mouse oocytes and has been shown to be essential for early embryonic development[92] but has not previously been reported as oocyte-specific among ovarian cells. Likewise, *UCHL1* codes for a member of the ubiquitin C-terminal hydrolase family of proteins, whose activity contributes to fertilization and embryogenesis in mice[93]. *REC114* codes a protein involved in the programmed formation of DNA double-strand breaks during meiosis but its functional significance for oogenesis has not been extensively characterized in mice or humans[94]. Finally, ZFAND2A codes for a protein involved in zinc ion binding activity, and although not well-characterized in humans, has been previously reported as upregulated in human oocytes[95]. These studies suggest that these genes may have biological importance for oocyte development and here we have shown for the first time their oocyte-specific expression compared to all other ROIs. Importantly, the eleven Cluster 9 samples from Donor 2 recapitulated the pattern seen in the Cluster 6 sample from Donor 1. In all, we identified 76 markers for human early-stage oocytes (Figure 4.2D), many of which are involved in meiosis (*HSP90A1, UBB, RALB, ESRP1, EIF4ENIF1, CDK7, ZAR1*).

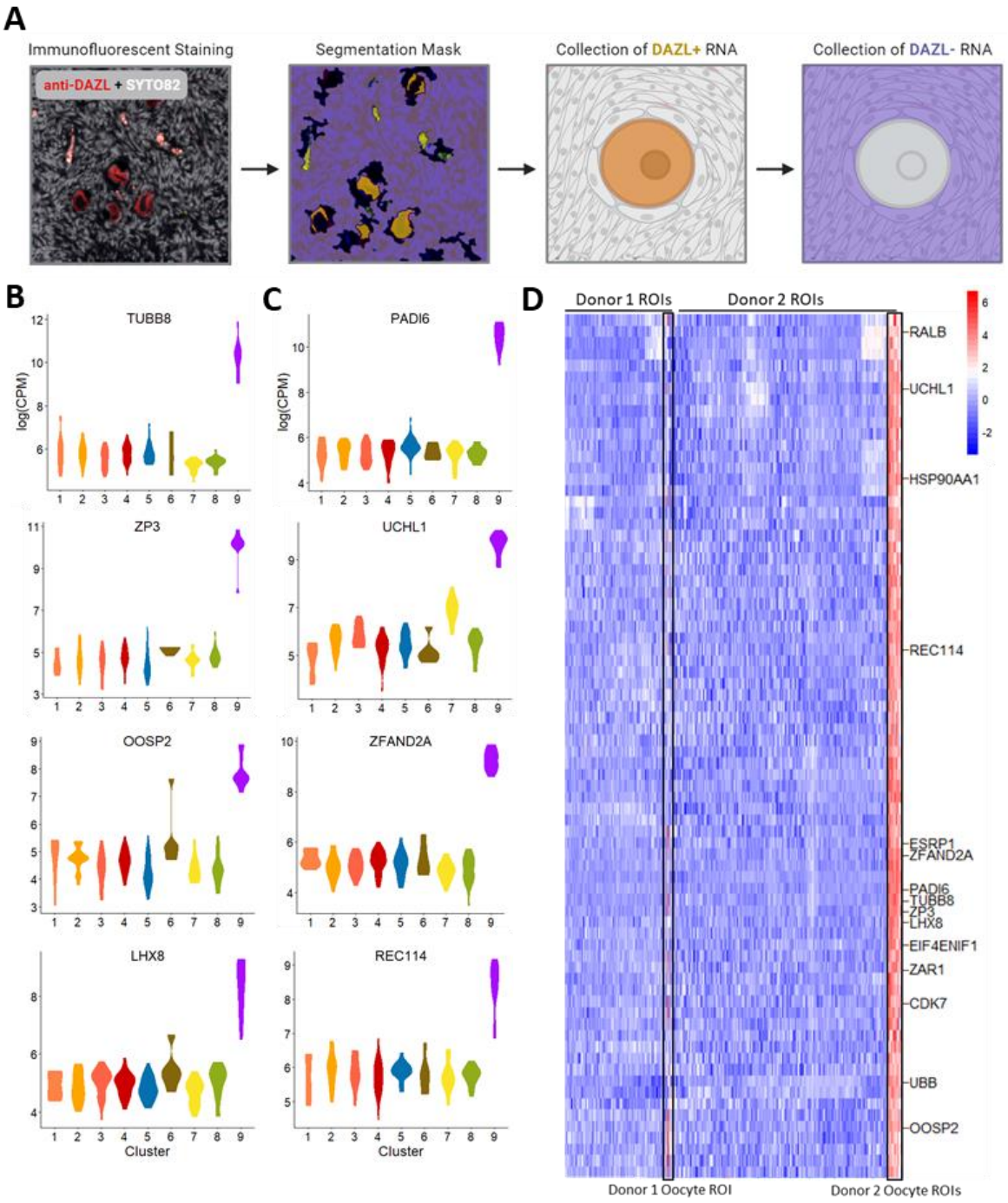


Figure 4.2: Transcriptional signature of human oocytes. A. The process of antibody-guided sample collection from subareas of an ROI. In this example, after the sample in DAZL+ area (shown in red) is collected by photo-activated cleavage, the sample in the remaining areas of the ROI (DAZL-, shown in blue) is collected in a second step, yielding a pair of samples from a single ROI. B. Expression pattern of four previously known canonical oocyte marker genes across the nine clusters in Donor 2. C. Four of the newly identified marker genes. D. Expression specificity of the 76 oocyte marker genes (in rows) compared across, from left to right, 74 samples from Donor 1 ordered by the 6 clusters, and 165 samples from Donor 2 ordered by the 9 clusters. Genes shown in B and C, along with additional genes of interest, are indicated on the heatmap. Color scale explained in Methods.

4.4.3 Gene signatures of theca and granulosa cells

Among the 74 ROIs in Donor 1, the 5 theca and the 11 granulosa ROIs separated into two clusters (Clusters 4 and 5, respectively, Figure 4.1D, F). Focused PCA of these samples revealed a three-way separation that also distinguished the 7 mural granulosa samples from the 4 cumulous granulosa samples (Figure 4.3A). These distinct profiles were reproduced in Donor 2 (Figure 4.1E-F), where focused PCA separated the 9, 12, and 11 ROIs enriched for the theca, cumulus granulosa, and granulosa cells into 3 groups (Figure 4.3B) even though they were collected from five antral follicles. Differential expression analysis identified 96 granulosa-specific genes and 46 theca-specific genes (Methods; Figure 4.3C). The stark contrast in expression of these genes across the two cell types exemplifies their distinct functions. In addition, hundreds of other genes were high in both theca and granulosa cells, but not strongly different between them (not shown).

The list of theca-specific genes included canonical theca cell markers *CYP17A1, PTCH2, APOE, DHCR24, INSL3, BGN* and *CYP11A1*[96,97] and the more recently reported *ANPEP*[98], as well as those not reported for theca, but having notable biological relevance, such as *S100A13, ALAS1, FDX1,* and *DLK1*. *S100A13* codes for a calcium-binding protein and has been reported as enriched in Leydig cells in a comparison of cell lines from different organ systems [99], but its biological significance in Leydig cells or their ovarian counterpart, theca cells, has yet to be elucidated[99]. Notably, the list included *FDX1* and *ALAS1*, which play important roles in cholesterol acquisition and heme biosynthesis, respectively, both of which are important for the conversion of cholesterol into androgens by the P50 steroidogenic enzymes in theca cells[100,101]. Finally, *DLK1* codes for a transmembrane protein that has been implicated in growth hormone signaling, which is essential for theca cell differentiation[102,103].

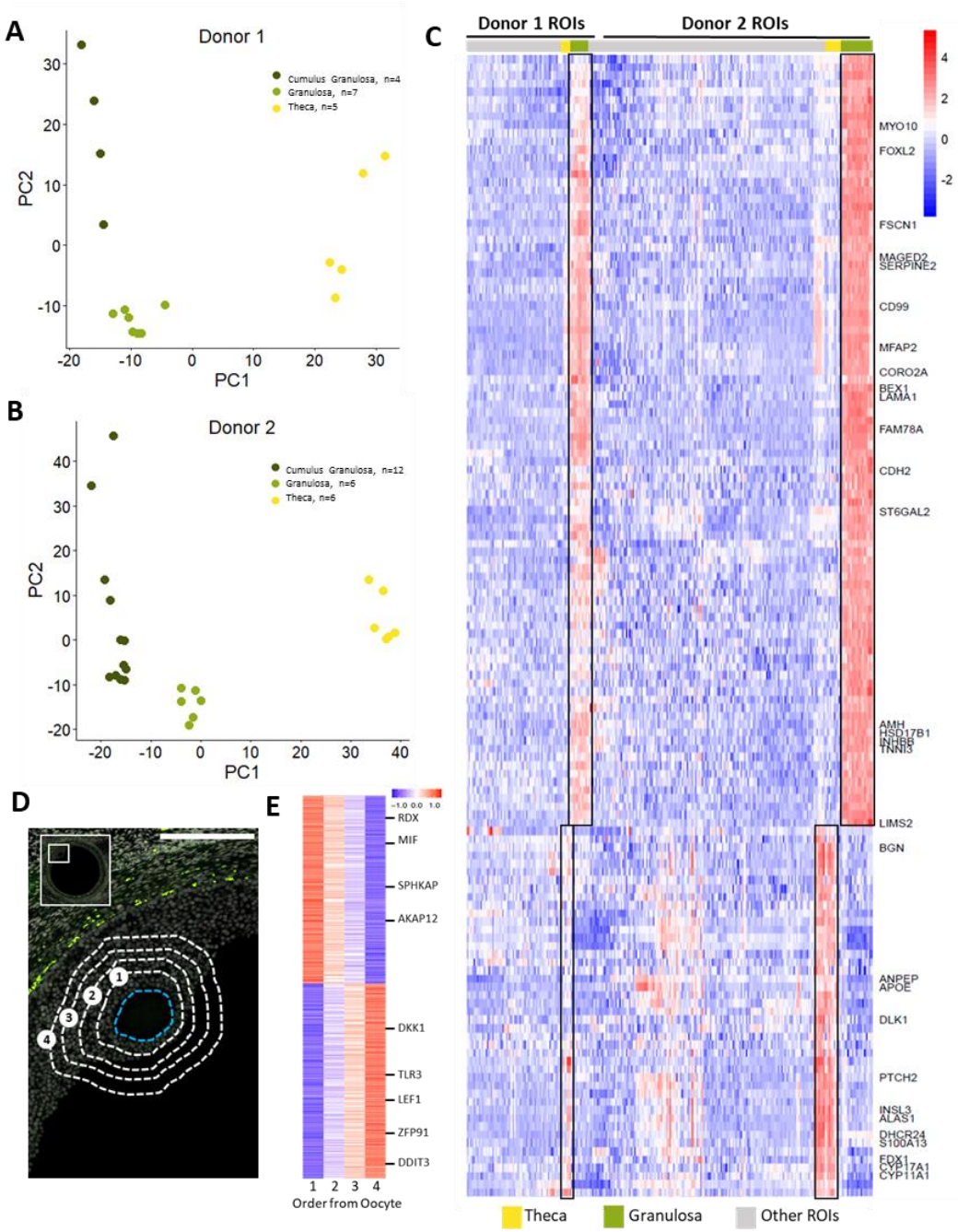


Figure 4.3: Robust signatures of granulosa and theca cells. A. Focused principal component analysis (PCA) of the 5 theca-enriched ROIs and 11 granulosa enriched ROIs from Donor 1, showing their separation into two clusters. B. Similar PCA plot for 6 theca and 18 granulosa samples from Donor 2. C. Expression specificity of the 94 granulosa cell maker genes (top) and the 45 theca cell marker genes (bottom), compared across the 74 and 165 samples from the two donors, ordered by clusters in the same way as in Figure 4.2D. D. Four concentric ring-shaped ROIs, for cumulus granulosa regions surrounding the oocyte in a follicle from Donor 1 (scale bar = 100 µm). E. Regression analysis over the ordered series of 4 ROIs in D identified 1,407 DE genes ($p < 0.05$), shown as a heatmap of genes (in rows) across the 4 samples. Color scale explained in Methods.

Likewise, the list of granulosa-specific genes included canonical marker genes *HSD17B1, INHBB, FOXL2, AMH,* and the recently reported *TNNI3, MAGED2, CD99, SERPINE2, CDH2, and BEX1*[8,9], along with many histone protein-coding genes, which have not been reported for granulosa cells. However, numerous studies have indeed suggested that proliferating granulosa cells undergo large-scale epigenetic change during late folliculogenesis, partly driven by gonadotrophins such as follicle-stimulating hormone (FSH)[104–107]. The list also included genes involved in granulosa cell-oocyte signaling, such as the transzonal projection-related genes *FSCN1* and *MYO10,* and the tetraspanin protein gene *TSPAN7*, which is involved in extracellular-vesicle mediated signaling between granulosa cells and the oocyte[85,108]. The list confirmed many of the genes previously reported as upregulated in granulosa cells in RNA-based comparisons: *LIMS2, CORO2A, LAMA1, FAM78A, ST6GAL2,* and *MFAP2*[109–111].

In Donor 1, we also sampled an inside-out series of four cumulus granulosa ROIs, each 30 µm wide, corresponding to four concentric rings surrounding the antral oocyte (Figure 4.3C), which formed Cluster 6. We performed a regression analysis across the four ordered series of samples to identify potential gene expression changes in the vicinity of the oocyte. Moving outward from the oocyte we observed increased expression of Wnt signaling genes (*DDIT3, DKK1, LEF1, TLR3, ZFP91*) (Figure 4.3D). FSH-mediated downregulation of this pathway in granulosa cells is important for progesterone production and cumulus cell expansion prior to ovulation[112]. Conversely, there was increased expression of genes related to protein kinase A (PKA) signaling (*AKAP12, MIF, RDX, SPHKAP*) moving inward to the oocyte. PKA signaling in granulosa cells has been well-documented in the literature; it modulates FSH-induced granulosa cell differentiation and oocyte maturation as follicle maturation proceeds[102,113].

4.4.4 Transcriptional gradient in the ovarian cortex and functional heterogeneity across cortex and medulla

To investigate transcriptome signatures in the outer surface of the ovarian cortex we selected 11 consecutive regions, each 30 µm deep and 350 µm wide, covering the outer 0.33 mm layer of the Donor 1 tissue (Figure 4.4A). Strikingly, even though adjacent regions are only 30 µm apart, the gene expression profiles of this linear sample series showed a 11-step, ordered transition (Appendix Figure A.10A). Linear regression analysis by layer number revealed 313 genes with a significant (FDR<0.05) depth-related expression gradient in the outer cortex (Figure 4.4B). Genes with increasing expression moving inwards from the surface epithelium included those involved in hormone signaling (*NR4A1*, *CEBPD*, *STAR*, and *ADAMTS4*), insulin-like growth factor binding protein (IGFBP) signaling (*IGFBP2*, *IGFBP3*, *IGFBP4*, *IGFBP6*, and *IGFBP7*), and ECM remodeling (*VIM*, *COL1A2*, *TIMP1*, *TIMP2*, *MMP2*, *COL8A1*, *COL12A1*, *COL14A1*, *COL16A1*, and *COL18A1*) (Figure 4.4B). Both *NR4A1* and *CEBPD* encode transcription factors related to hormone synthesis. *NR4A1* stimulates the expression of *STAR* and other genes involved in androgen production via the cAMP/PKA pathway[114]. Luteinizing hormone (LH) stimulates the expression of *CEBPD*; and the resulting transcription factor, CCAAT/enhancer-binding protein delta, has been implicated in LH-triggered events in mature granulosa and theca cells[115,116]. These results showed for the first time a high-resolution, spatially resolved gradient in the outer cortex, with progressively higher levels of hormone signaling when moving deeper into the cortex layers and more proximal to the growing follicles and the medulla.

To expand this approach when analyzing the Donor 2 tissue, we sampled five linear series of consecutive regions: two "medullary series" (M1 and M2) traversing the short and long axes of the tissue section, and three "cortex series" (C1-C3) covering 1.5 mm of the cortex (Figure 4.4C).

Each series contained 15 or 16 consecutive straight-line samples. The cortex series started at the surface epithelium from three sides, with sample width of 150 µm and depth of 100 µm. The medulla series had the same width (150 µm) and depth of 500 µm, and they connected with the cortex series at the end. In the overall clustering of all Donor 2 samples, most of the ROIs from the three cortex series fell into Cluster 1 (n=16) and 2 (n=20), which are cortex sample clusters, while a few others were mixed into other clusters, i.e., 5 in Cluster 4, which "tracks" towards Cluster 6 - the CD68+ samples (Figure 4.1E); and 6 in Cluster 7 - mainly containing theca samples (Figure 4.1F). The medullary series were more homogenous, with 28 of the 31 samples forming Cluster 5, which also contained 7 of the 9 $DAZL^-$ samples surrounding early-stage follicles (Figure 4.1F). Focused PCA of the five series and the 9 $DAZL^-$ samples showed these patterns in more detail (PC1-3 in Figure 4.4D, PC1-2 in Appendix Figure A.10B). Samples in each of the three cortex series covered a wide PC1-PC3 space, yet they did not follow a linear gradient and did not clearly replicate each other (Appendix Figure A.10C), reflecting diverse functional characteristics being sampled. For example, the cortex series drove Clusters 1-2, yet they came near Cluster 3 (Figure 4.1E), which had all 10 samples for ovarian cyst, along with 4 of 5 of PGR^+ and 3 of 12 $CD68^+$ samples (Figure 4.1F). Other cortex samples may have been near blood vessels, regions of atresia, or medulla samples with high CD68 expression. In general, the three cortex and the two medulla series are separable in PC3 (Figure 4.4D), with one of the cortex series resembled the two medullary series in PC1 (Appendix Figure A.10B), while the 9 $DAZL^-$ samples, which were from cortex regions dense with early-stage follicles (Figure 4.2A), appeared to "bridge" the space between the cortex series and the medullary series (Figure 4.4D). In sum, while the 11-sample series in the outer cortex of Donor 1 showed a clean, linear gradient, the three cortex series of 1.5 mm tissue depth

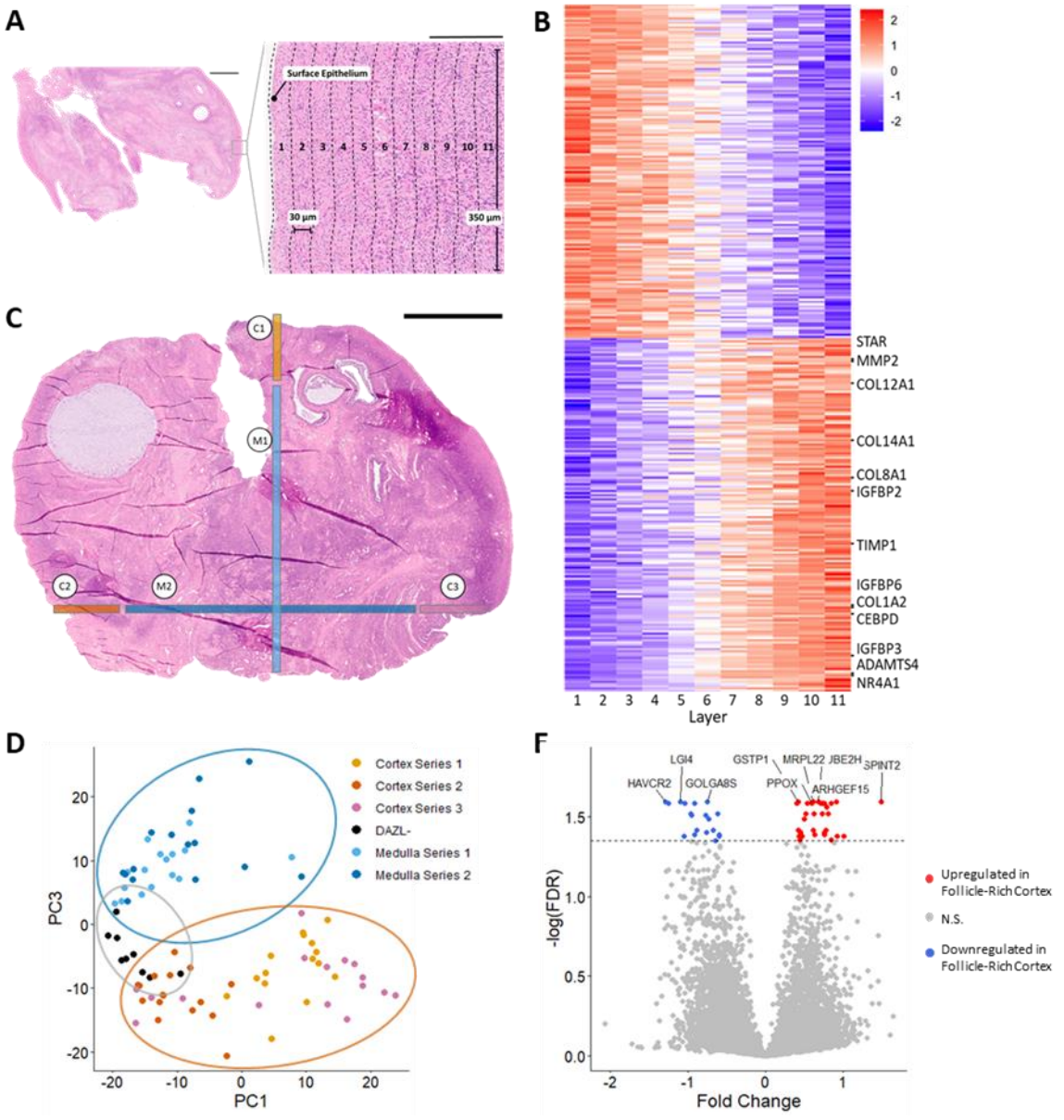


Figure 4.4: Transcriptional heterogeneity across cortex and medulla areas. A. In Donor 1, a series of 11 consecutive tissue layers were sampled at the surface of the cortex, each with 30 µm depth and 350 µm width (scale bar = 2.5mm; scale bar for insert = 100 µm). Shown is the H&E image. B. Linear regression analysis over the 11 ordered samples identified 313 differentially expressed genes ($p < 0.05$), shown as a heatmap, with notable genes indicated to the right. C. Five straight-line series of 15-16 consecutive samples for Donor 2, three for cortex regions (3 warm colors) and two traversing the medulla (2 blue colors), indicated on the H&E image (scale bar = 2.5 mm). D. PC1-3 plot of the 3 cortex and 2 medulla series, each with 15-16 samples, plus the 9 DAZL- subareas near the follicles. The six series are shown in different colors. E. Comparison between the 9 DAZL- near-follicle ROIs and the 3 surface-most cortex ROIs, one from each of the 3 cortex series, identified genes that are significantly (FDR<0.05) higher (red, n=33) and lower (blue, n=21) in follicle-rich cortex vs. cortex without follicles. Shown is the "volcano plot" of fold-change (x-axis) and logged false-discovery rate (y-axis).

in Donor 2 revealed more heterogeneity, while the two medulla series showed a relatively homogeneous profile (Cluster 5) for samples in the interior of the ovary.

We performed a series of DE analysis among these clusters to elucidate their functional characteristics (Methods). The medulla samples (Cluster 5), when compared to cortex samples (Clusters 1-3), showed higher expression of *WFIKKN2*, *GSTM5*, and *IGFBP5* and enrichment for terms related to protein targeting to the endoplasmic reticulum and RNA translation ("DE11"). Conversely, the cortex samples showed higher expression of *APOC1*, *DNAI3*, *CYP17A1*, and *DHCR24* ("DE11"). Within the cortex samples, Clusters 1-4 represented several characteristics. Cluster 4 is connected with CD68+ samples in Cluster 6 (Figure 4.1E), suggesting that they carry local inflammation signatures. Cluster 3 (dark red in Figure 4.1E) contained all 10 cyst samples and likely represent signatures of dysregulated follicle development or regressing follicles. Cluster 2 contained all 6 "rings" around the primary follicles, while Cluster 1 (bright red), consisting entirely of 16 samples from the cortex series without drawing any other ROI types, is closest to the medulla profile in Cluster 5 (Figure 4.1F). We compared Clusters 1/3 against Cluster 2 (DE8) and Cluster 3 against Cluster 1 (DE7), for increasingly finer distinctions among different types of cortex regions. For example, Clusters 1/3 showed higher expression than Cluster 2 of *DHCR24*, *CYP17A1*, *APOC1*, and *GSTA1* and enrichment for gene ontology terms related to cellular response to steroid synthesis and cholesterol metabolism. Conversely, Cluster 2 showed higher expression of *WFIKKN2*, *GSTM5*, and *IGFBP5* and enrichment for terms related to protein targeting to the endoplasmic reticulum and RNA translation.

We also examined transcriptional differences between cortex regions dense with follicles: the nine DAZL- samples, and the three cortex surface ROIs, which contained no follicles, one from each of the three cortex series. DE analysis revealed 38 genes with significantly higher expression

in the follicle-rich cortex, including *SPINT2* and *ARHGEF15*, and 26 genes with significantly lower expression (Figure 4.4E). *SPINT2* codes for the protein HAI2 and was highly expressed in the granulosa cells previously[8]. HAI2 is a key inhibitor in the hepatocyte growth factor (HGF) signaling pathway, which has been shown to modulate the remodeling of the ovarian surface epithelium and cortex after the cyclic disruption caused by ovulation[117]. *ARHGEF15* was recently identified in Sertoli cells and shown to be essential to spermatogenesis[118]. The identification of these genes in follicle-rich cortex may aid the understanding of how the cortex maintains a microenvironment conducive to follicle development, from quiescence to activation.

4.4.5 Single-cell RNA sequencing provides a human ovarian cell atlas

Other than the two samples analyzed for ST we collected scRNAseq data from three other donors (Donors 3-5, ages 28, 37, and 31, Appendix Figure A.8A). In all, we measured 37,047 cells, of which 21,198 passed quality filtering (Methods), with an average of 6,950 unique transcripts and 2,066 detected genes per cell. Clustering analysis for each donor consistently revealed 4 major cell types (Donor 5 in Figure 4.5A, Donors 3-4 in Appendix Figure A.11A-B): stromal cells, immune cells (appearing as two clusters in the joint analyses of all cells), endothelial cells, and pericytes, and they were identified by known marker genes (Figure 4.5C). Endothelial cells showed high expression of their canonical markers *VWF, PECAM1, CD34,* and *CLDN5*[1,2], as well as *NOTCH4*, which is typically confined to arterial and sprouting endothelial cells[66]. Pericytes were identified by their characteristic expression of *RGS5, NOTCH3, ACTA2,* and *MUSTN1* (Figure 4.5C-D). Interestingly, pericytes in our data showed high expression of *PDGFRB*, which has been reported in other organ systems for platelet-derived growth factor (PDGF) signaling between pericytes and endothelial cells, and it is also essential to proper theca cell development and steroidogenesis[66].

Focused clustering of the immune cells from all 3 donors revealed four immune cell subtypes (Figure 45B), which were annotated as T cells (*CD3D, CD3G, IL7R*)[8,9], natural killer cells (*KLRD1, TRCD, GNLY*)[115,119], macrophages (CD68, CD14, FOLR2)[8,9], and mast cells (*KIT, TPSB2, TPSAB1*)[120,121] (Figure 4.5C). Additional genes that are specific for the three major cell types and the four immune cell subtypes (Figure 4.5D) add to the expanding cell atlas for the healthy human ovary.

Stromal cells had by far the largest number, reflecting the dominance of this cell population in the ovary. As a group, they were characterized by mesenchymal cell markers *DCN* and *PDGFRA*, and the more recently reported ovarian stromal cell markers *APOE, LUM, ARX*, and *FHL2*[1,2] (Figure 4.5D). Focused analyses of the stromal cells did not reveal clear subtypes (see Methods for details of the many attempts). Since the spatial analyses have generated robust gene lists for the oocyte, theca cells, and the granulosa cells, we used these marker panels to score each stromal cell for its chance to be a theca-like or granulosa-like stromal cell and, in even rarer cases, be an oocyte. After carefully selecting the subset of stromal cells with high sequencing depth and using genes with high detection rate, we projected the stromal cells for each of the three donors separately, and then asked if the theca-, granulosa- and oocyte-scores tend to aggregate in some regions of the projection. Of the multiple scoring algorithms examined, some showed that some stromal cells indeed scored high for the theca gene panel, and many of them also scored high for the granulosa panel, although to a lesser degree (Appendix Figure A.11B). Nonetheless, these stromal cells did not clearly separate as a distinct subtype. The oocyte scores did not reveal any clearcut rare population to be potential oocytes (Appendix Figure A.11B). This is not unexpected,

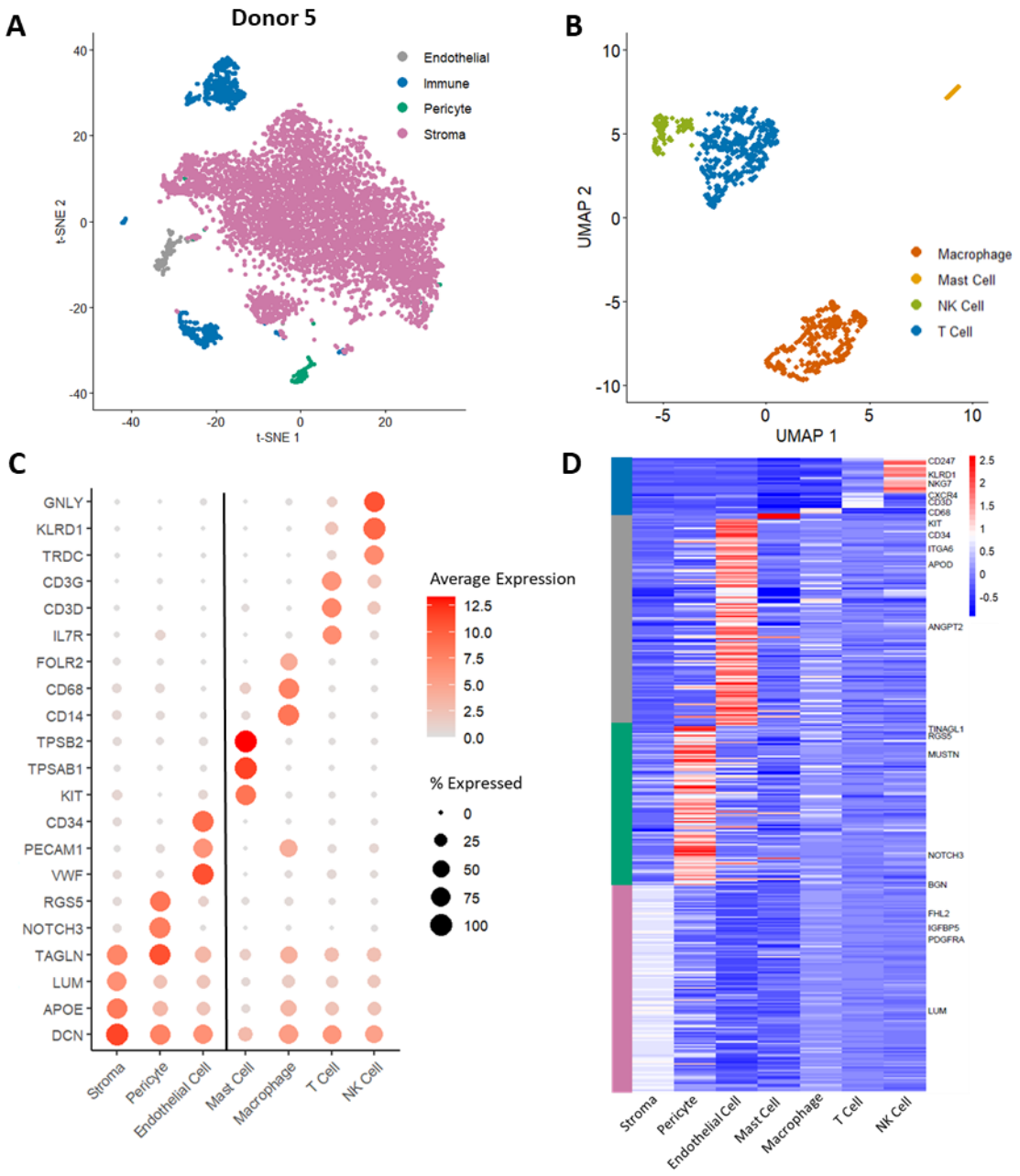


Figure 4.5: Cell types in the human ovary identified by scRNAseq analysis. A. t-SNE projection of 6,339 cells from Donor 5, colored by the four major cell types identified. B. UMAP projection of immune cells (n=863 from 3 donors), colored by the four immune cell subtypes identified. C. Known marker genes used to annotate the 3 major cell types and 4 immune cell subtypes. Color indicates expression level, while symbol size indicates detection rate. E. Centroid data for marker genes for the 3 major cell types and 4 immune cell subtypes. Number of genes displayed: stroma - 119, pericyte - 92, endothelial - 120, mast cell - 3, macrophage - 3, T cell - 9 , NK cell - 19. Additional literature-based marker genes are indicated on the right.

as oocytes are extremely rare in the dissociated cells from the tissue and there is a ~30 µm size limit in the cell-capturing apparatus. In sum, stromal cells were the most abundant in our data and they appeared homogeneous, with some evidence for a rare subset to be theca-like.

4.5 **Discussion**

In this study we systematically analyzed the spatial and cellular heterogeneity of the human ovary, using samples from donors without a history of cancer, known diseases that affect ovarian function or prior androgen therapy. In contrast to biopsies taken during surgery, the whole cadaveric ovaries we used allowed us to select functional regions throughout the tissue for spatial analyses, and to isolate single cells from both cortex and medulla. In the past, studies of ovarian function have faced at least three challenges. First, dissociated cells are dominated by the most abundant cell type: the stromal cells; while the most functionally important cells - those around the follicles - are exceedingly rare, even in the ovarian cortex. Second, cell sorting efforts to enrich for oocytes, theca cells, or granulosa cells, and attempts to visualize them in the intact tissue, relied on a few known markers, without characterizing the activities and spatial patterns of other genes in an unbiased fashion. Third, some of the earliest and most popular spatial transcriptomic technologies are only able to analyze coarse regions at pre-specified grid points (e.g., Visium's spots are 55 μm in diameter, with a spot-to-spot distance of 100 μm), yet the functional units of the human ovary, especially the early-stage follicles, are in the size range of 20-50 μm. We addressed these challenges by adopting NanoString's *GeoMx*™ technology, which is unbiased (i.e., it profiles nearly all coding genes) and allows focal analyses of individually selected tissue areas, or even subareas targeted by specific antibodies. In all, we analyzed 92 samples from 74 ROIs in one donor, followed by validation with 165 samples from 147 ROI from the second donor, covering a wide variety of functional regions across the whole ovary. This strategy allowed us to avoid spending most of the resources on collecting data from the stromal cells and yielded local transcriptomic profiles around both early-stage and antral follicles. We then complemented this

powerful series of spatial data with scRNAseq, identifying major cell types such as stromal cells, pericytes, endothelial cells, and four subtypes of immune cells.

For oocytes, we profiled a late-stage oocyte from an antral follicle in Donor 1 and 11 samples in Donor 2 representing early-stage oocytes. These samples cluster well among themselves and are clearly enriched for canonical oocyte markers (Figure 4.2B), confirming that the *GeoMx*™ technology, including the use of DAZL antibody-guided sample collection, can indeed profile oocytes in tissue sections. These data led to a list of 76 oocyte-specific genes, including those related to meiosis (*REC114*) and embryonic development (*PADI6* and *UCHL1*) (Figure 4.2C-D). As a community resource, this list provides many more gene candidates than before for developing new reagents to study oocytes of different stages.

Similarly, we took advantage of the custom ROI selection capability of *GeoMx*™ to identify dozens of genes specific for theca and granulosa cells (Figure 4.3C). These two cell types reside in adjacent areas around developing follicles and are difficult to dissect into separate samples. Thus, the fact that they arose in our data as distinct clusters (Figure 4.1D-E) and were stably reproduced across follicles and across donors, confirms the functional identity of these cells as well as the location specificity of the technology. Further analyses revealed the finer distinction between cumulus granulosa and the other granulosa layers (Figure 4.3A-B), and a gene activity gradient across concentric layers of cumulus granulosa around the oocyte in an antral follicle (Figure 4.3D-E). More detailed analyses of such gradient are needed to examine the longstanding hypothesis of bidirectional crosstalk between the oocyte and the cumulus cells during follicle development[1–3].

We collected nine sample pairs in cortex regions dense with early-stage follicles, where the $DAZL^+$ sample of the pair corresponds to the oocyte, while the $DAZL^-$ sample represents cortex

cells surrounding the primordial, transitional, and primary follicles (Figure 4.2A). The DAZL$^-$, “follicle-rich" cortex regions gave us the opportunity to compare them with other, “follicle-null” outer cortex samples (Extended Data Fig. 4D) to identify genes that may be involved in the development of early-stage follicles through the emergence of theca and granulosa cells and their spatial organization around the maturing oocyte. This analysis highlighted genes related to steroidogenesis (*ARHGEF15, GSTP1*) and tissue remodeling (*SPINT2*) (Figure 4.4E), and they need further validation to understand how the ovarian stroma influences follicle quiescence and early development.

Since it is impractical to cover the entire tissue section with ROIs we systematically profiled serial samples across several cortex and medulla regions. In Donor 1, we used 11 consecutive samples, at 30 µm depth increment, to study the outer cortex, and observed a graded change in genes related to hormone signaling (*NR4A1*, *STAR*, *ADAMTS4*) and ECM remodeling (*VIM, COL8/12/14/16/18*) (Figure 4.4A-B). Using a similar approach in Donor 2, we sampled three straight-line, linear series, with 100 µm increment, to span the full 1.5 mm depth of the cortex, and two series of medulla samples to traverse the entire long and short axis of the tissue. These cortex and medulla series did not show clean gradients (Figure 4.4C), rather they confirmed the medulla-cortex difference (Figure 4.4D) we had seen in Donor 1, and revealed transcriptional variation across the surface and the interior of the ovary. For example, the cortex series, from the very surface to 1.5 mm deep, did not always "start" in the same transcriptomic state (Extended Data Fig. 4D), nor ended in similar profiles; rather the 15 or 16 samples in each series "visited" multiple functional characteristics (Extended Data Fig. 4C) that may correspond to areas of atresia, vasculature, or immune activity, which were not evident at the time of sample selection. Nonetheless, some broad patterns did emerge: the two medulla series are similar and clustered

away from the three cortex series (Figure 4.4D), and one of the three cortex series is more medulla-like (Appendix Figure A.10B). Within the cortex samples, we found a steroidogenic profile in Clusters 1 and 3, with high expression of *DHCR24, CYP17A1*, and *GSTA1*, while Cluster 2 showed high expression of *WFIKKN2, GSTM5*, and *IGFBP5* (Figure 4.4C-D). Future studies will need denser sampling and more detailed local phenotyping to fully understand the functional changes that accompany follicle differentiation in various regions of the medulla and cortex.

Our scRNAseq data revealed four major cell types, which are mostly in agreement with those reported by two previous reports. In one, Fan et al. analyzed 20,676 cells isolated from antral follicles from 5 donors undergoing fertility preservation prior to anti-cancer treatment and identified five major types: theca/stroma (Th/S), granulosa (G), immune (Imm), endothelial (E), or smooth muscle (SmMusc)[1]. Based on the top marker genes reported for each cluster, their theca/stroma clusters likely matched our stromal cells (*TCF21, DCN, LUM, IGFBP5, APOE*); two of their immune clusters (4 and 12) corresponded to our NK cells (*GNLY, KLRD1, CTSW*) and T cells (*CD3D, CXCR4*) respectively. Their endothelial clusters mapped to our endothelial cells (*VWF, A2M, CLDN5*), while their smooth muscle clusters corresponded to our pericytes (*RGS5, ACTA2, TAGLN*). We did not enrich for rare follicular cells through follicle resection or marker-guided enrichment, and as a result we did not identify granulosa and theca cell clusters. However, the granulosa- and theca-enriched ROIs from our spatial analysis emerged as distinct clusters and they correlated specifically with the theca/stroma and granulosa clusters, respectively, from the Fan et al. study (not shown). In a second study, Wagner et al. profiled 12,160 cells isolated from the ovarian cortex of 4 donors undergoing caesarian section or gender reassignment surgery[9]. They identified five types of somatic cells: immune, endothelial, granulosa, stromal, and smooth muscle cells. They reported high expression of *PDGFRA, DCN, COL1A1*, and *COL6A1* in their stromal

cells, which matched our stroma cluster. Likewise, they reported high expression of *MYH11*, *MCAM, RGS5,* and *TAGLN* in smooth muscle cells, which corresponded to our pericytes. Their endothelial cells had high expression of *VWF* and *CDH5*, matching our endothelial cell cluster. For immune cells, they identified a coarse group of cells expressing markers for T cells (*CD2*, *CD3G, CD8A*) and antigen-presenting cells (*CD14*, *B2M*), while we identified four specific immune cell types. As explained above, we did not find granulosa cells in our scRNAseq data. Rather it was our spatial data that led to robust marker genes for each follicular cell type, which included canonical markers like *FOXL2, AMH*, and *INHBB* for granulosa, *CYP17A1, CYP11A1*, and *PTHC2* for theca, and dozens of new markers (Figure 4.3C). These gene lists confirm many of the granulosa and theca markers reported by Fan et al. and Wagner et al[1,2].

4.6 **Conclusions**

The RNA markers reported for the oocyte, theca cells, and granulosa cells in the ovary are among the strongest contributions of this study. Some of these markers can be further developed into probes for isolating specific cell population for in-depth experimentation, such as being used to improve *in vitro* follicle culture, to promote the differentiation and maturation of stem cell-derived oocytes, instead of using fetal somatic cells[122,123]. Or they can be used in the isolation of steroidogenic cells expressing genes such as *STAR* and *CYP17A1* for hormone production *in vitro*. The markers may also be used for higher-resolution spatial analysis and lineage tracing, or as targets of perturbation to understand the wide range of disorders affecting the female reproductive system, such as female infertility, ovarian aging, or somatic aberrations that lead to cancers.

This study also underscored the limitations of the current spatial transcriptomic technologies. For small functional units like the ovarian follicles, single-cell or even subcellular resolution is needed to fully elucidate the local crosstalk between the oocyte and its surrounding

cell community. Highly multiplexed *in situ* imaging methods such as MERFISH can measure the distribution of single RNA molecules for a few hundred RNA markers[124]; and our lists of oocyte-, theca- and granulosa-specific genes are essential for the optimal design of such gene panels. We expect that the rapidly improving resolution and multiplexity will soon allow us to study all stages of oocyte development, from the smallest primordial follicles, to transitional and primary follicles, and onward to late-stage oocytes, the corpus luteum, and subsequent events in follicle maturation and degeneration. The single-cell and spatial analyses we report here serve as a key step towards these more powerful data series for understanding healthy ovarian function across women's reproductive lifespan, as well as disorders such as polycystic ovarian syndrome, premature ovarian aging, and ovarian cancers.

Chapter 5 Conclusions and Future Directions

5.1 Summary of Findings

The aim of this doctoral thesis was to identify factors driving follicle maturation using omics techniques for development of a translational ovarian follicle culture system for fertility preservation. Ovarian follicle development is driven by a complex milieu of soluble signals across the cellular compartments of the follicle, the extrafollicular microenvironment in the ovary, and systemic circulation. Our understanding of early follicle development in humans and even in mice is limited, leading to challenges culturing early-stage follicles *in vitro*. We utilized unbiased, comprehensive transcriptomic and proteomic techniques to investigate the intra- and extra-follicular factors driving early murine folliculogenesis. We profiled murine follicles cultured *in vitro* and uncovered genes, pathways, and proteins that differentiate between successful and suboptimal culture conditions. In parallel, to advance our knowledge of human folliculogenesis, we analyzed the transcriptome of ovarian cells using rare cadaveric tissue samples, discovering a myriad of extra-follicular processes that may regulate follicle activation and early development. This work has illuminated a variety of factors driving early folliculogenesis and serves as a data-driven reference list of cues and signaling pathways that may be manipulated in the future to improve follicle outcomes *in vitro*.

In the first aim, we used microarray analysis of follicular somatic cells to identify genes and pathways associated with successful follicle development *in vitro*. We employed a group culture system for primary ovarian follicles, where follicles were co-encapsulated in alginate

hydrogels in groups of five (5X) or ten (10X). This system provides the paracrine signals essential for early-stage follicles, allowing us to profile cells across development and compare the suboptimal 5X condition to 10X, which consistently produces mature oocytes. By comparing these conditions transcriptionally, we identified calcineurin signaling, Wnt signaling, prolactin signaling, and angiogenic signaling as upregulated in 10X follicles. While calcineurin and prolactin signaling are not well-characterized in the ovary, Wnt ligands and angiogenic factors like VEGFs are widely reported in the literature as signals essential for follicle development[66,72]. Exogenous supplementation with these factors may improve growth and survival of individual primary follicles *in vitro*, thereby advancing our progress towards establishing a well-defined, translational culture system for the clinic.

Comparing the 5X and 10X conditions using the microarray technology suggested meaningful transcriptional differences between 5X and 10X follicles that could be used to identify candidate factors and pathways for modulating folliculogenesis *in vitro*. However, the transcriptional activity of the oocytes of the cultured follicles, which participates in bidirectional signaling with somatic cells to accomplish maturation[67–69], was not yet reported. In our follow up study, we profiled 5X and 10X follicles by analyzing somatic cells and oocytes separately using RNA sequencing, then profiled proteome-level differences between the conditions using shotgun proteomics. In this study, we validated a number of our findings from the microarray, where angiogenic and Wnt signaling pathways were again associated with the 10X condition in the somatic cell dataset. Our analysis of the oocytes yielded entirely new information on the transcriptional changes marking oogenesis from the primary follicle stage, including changes in transcription of genes related to fatty acid oxidation and organelle organization. Surprisingly, we observed very little difference (14 genes) in the transcriptome of oocytes when comparing 5X and

10X, suggesting that the differing outcomes in these conditions (i.e. terminal diameter and oocyte maturation) are driven by the somatic cells. We also explored somatic-oocyte crosstalk using ligand-receptor analysis, uncovering a number of signaling pathways not previously identified in the follicle. Finally, a comparative proteomics analysis of 5X versus 10X follicles confirmed a number of transcriptome-level findings (fatty acid oxidation, vesicle-mediated signaling) and identified additional proteins and pathways of interest that may contribute to successful follicle development *in vitro* (pro-survival proteins PAM16, PRKDC, MYBBPR1A; MAPK proteins REBP1, RAP1B, YWHAZ). When viewed in light of the previous microarray study, this work suggests that pro-angiogenic factors may support early follicle development in future culture systems, given their importance across all our methodological approaches. Modulation of Wnt signaling, metabolic pathways, and cell survival pathways may also improve culture outcomes.

In our final study, we investigated transcriptional heterogeneity in the human ovary, which has been largely uncharacterized. Using ST and scRNAseq, we profiled the cellular constituents of the ovary and discovered gradients of gene expression across various compartments of the organ. We used antibody-guided ST sample collection to selectively isolate oocyte RNA from early-stage follicles, developing a robust reference dataset of these cells without the transcriptional perturbation caused by tissue dissociation and mechanical cell isolation. We analyzed the granulosa and theca compartments of antral follicles and developed comprehensive marker gene sets for each cell type, discovering new genes and biological processes not previously implicated in follicular cell function. We also confirmed gradients of gene expression across the cumulus granulosa compartment using ST. In our analysis of the nonfollicular compartments of the ovary, we observed graded changes in gene expression related to hormone signaling and ECM remodeling in the ovarian cortex. We also identified transcriptional differences between stromal cells adjacent

to early-stage follicles versus regions with no follicles, suggesting that the ovarian stromal microenvironment may influence early follicle development or vice versa, early-stage follicles may influence their neighboring stromal cells. Our scRNAseq findings agreed with previous reports of the cellular makeup of the human ovary[8,9], which is comprised primarily of ovarian stromal cells with smaller populations of endothelial, pericyte, and immune cells. The lack of granulosa, theca, and luteinized cells in this dataset was expected and suggests that enrichment methods such as antral follicle resection or fluorescence-activated cell sorting (FACS) are essential for meaningful analysis of these less abundant cell types. Our findings together addressed a number of unresolved questions in our understanding of the human ovary and serve as a basis for further explorations into the intra- and extrafollicular factors driving follicle development.

5.2 **Future Directions**

5.2.1 Standardized Ovarian Follicle Culture for Assisted Reproduction

A well-defined, universal ovarian follicle culture system would serve as a broad fertility preservation option for women and girls who face gonadotoxic anti-cancer treatment. The studies detailed in this dissertation have deciphered intra- and extrafollicular factors driving follicle development through unbiased transcriptomic and proteomic analysis. In our microarray study, 10X follicles showed high expression of Wnt signaling factors *Wnt3a/7a* early in culture with subsequent downregulation, as well as early expression of angiogenin genes *Ang3/5*. Later in culture, 10X follicles showed increasing expression of steroidogenesis-related genes *Star* and *Cyp11b1*. Conversely, all these genes and pathways were dysregulated in 5X follicles and thus their temporal expression patterns in 10X are associated with successful follicle development *in vitro*.

Wnt factors play important roles in folliculogenesis, particularly in steroid hormone production[66]. In rats, granulosa cells exposed to WNT3A showed suppressed steroidogenic capacity, while treatment with Wnt inhibitor IWR-1 increased media concentrations of estradiol in granulosa cell cultures[125]. This suggests that Wnt inhibitors may rescue dysregulated steroidogenesis observed in 5X follicles through upregulation of steroidogenic enzymes and downstream hormone production. The roles of WNT7A and other non-canonical Wnts in folliculogenesis are less understood, but recent studies on single WNT proteins suggest that each ligand has unique effects on cell functions, underscoring the importance of future studies evaluating WNT7A and other ligands in the context of folliculogenesis specifically[66]. Likewise, angiogenin is not well-characterized in the ovary, although other classes of angiogenic factors play pivotal roles in folliculogenesis. Angiogenin is a potent pro-angiogenic factor and studies in cattle suggest its effects on follicle development are stage-dependent[126]. In small antral follicles, angiogenin supplementation inhibited progesterone and androstenedione production, while the same treatment in later stage follicles promoted granulosa cell proliferation and production of estradiol[126]. In our culture system, early 10X follicle expression of angiogenin genes and subsequent downregulation suggests that low angiogenin concentrations late in follicle development may support follicle maturation, but future studies specifically targeting the mechanism of action for angiogenin in follicles are needed.

Our subsequent RNA sequencing study supported and augmented microarray findings, where we again observed dynamic expression of pro-angiogenic genes including angiopoietin-like proteins (*Angptl3/4/6*), angiopoietins (*Angpt1/2/4*), and VEGFs (*Vegfa/b*) during culture. Both angiopoietin and VEGF genes showed increasing expression as culture proceeded in 10X, which agrees with previous secretome-level findings from Zhou et al. 2018[15]. Angiopoietin expression

in follicular cells is modulated by steroid hormones, and in the presence of VEGF the angiopoietins stimulate pericyte-mediated extracellular matrix degradation and endothelial cell infiltration, leading to vascularization of the theca layer[127]. Preliminary individual follicle rescue experiments (Appendix B) using a supplementation schedule for four cytokines of interest (VEGFA, MCP-1, KC, and follistatin) showed improved growth of individual primary follicles when compared to follicles cultured in control media. However, cytokines were not evaluated individually, making it difficult to determine the impact of VEGFA alone. Based on evidence in the literature for synergistic activities of VEGFs and angiopoietins, supplementation with both pro-angiogenic factors may improve individual follicle development *in vitro*. We also identified a number of genes related to vesicle-mediated signaling in our 5X vs. 10X analysis (*Rab3p, Syt1, Cspg5*), which all showed increasing expression over time in 10X but dysregulated, oscillatory expression in 5X. Vesicle-mediated signaling is an important mode of communication for granulosa cells in antral follicles, but literature on the modulation of this biological process in ovarian follicles or other cell types is limited. Further investigation of vesicle-mediated signaling in the follicle is needed to determine whether the dysregulation of vesicle-mediated signaling in 5X follicles is causative or simply a result of impaired follicle development and lower rates of antrum formation when compared to 10X follicles. Nonetheless, here we have identified a number of biological pathways and candidate factors, especially pro-angiogenic and Wnt signaling factors, that are associated with successful follicle development *in vitro*. These factors may now be used to modulate follicle development *in vitro* in a biomimetic culture system by recapitulating cues essential for folliculogenesis *in vivo*.

After development of a standardized follicle culture system for murine follicles, translation to human follicles will necessitate overcoming additional challenges. Namely, human follicles

reach much larger terminal sizes (~2 cm in diameter, as compared to 350 μm for mice) and folliculogenesis takes much longer in humans, on the order of several months. Follicle culture in a degradable, tunable hydrogel such as poly(ethylene glycol) (PEG)[128] that allows for dramatic volumetric expansion of follicles, paired with a standardized cocktail of exogenous factors that recapitulate the soluble signals experienced *in vivo*, has potential to support human follicle development from its early stages. Factors specific to human folliculogenesis such as emergence of the dominant follicle must also be addressed to support oocyte maturation and competence in a biomimetic culture system.

The cues and pathways identified in these studies may also serve as a means to generate induced pluripotent stem cell (iPSC)-derived germ cells and/or follicular somatic cells[129]. These cells are derived through genetic reprogramming of iPSCs to overexpress key transcription factors. In mice, live births have been achieved from oocytes derived from iPSC-derived oocytes, but challenges remain with understanding the mechanisms of development and improving efficacy of this approach[129]. iPSC-derived germ cells hold promise as a form of assisted reproductive technology that addresses currently unmet reproductive needs and may serve as a revolutionary research platform for developmental biologists. Additionally, iPSC-derived follicular somatic cells may support oogenesis and/or serve as steroid hormone-producing cells for a variety of applications.

5.2.2 Questions Regarding Ovarian Follicle Development

A number of important biological questions remain regarding the process of folliculogenesis *in vitro* and how it compares to follicles maturing *in vivo*. Prior hormone analysis of ovarian follicles co-cultured *in vitro* showed production of androstenedione that increased with follicle maturation[15], suggesting the presence of androgen-producing cells in follicles isolated at

the primary stage. Our prevailing hypothesis has been that in a group culture system, there are sufficient pre-theca cells present to sustain androgen production, but an alternative hypothesis is that bipotential pre-cursor cells or follicular cells exhibiting phenotypic plasticity present in primary follicles potentially differentiate into granulosa and theca-like phenotypes, allowing for androstenedione and downstream estradiol production. This concept of bipotential or plastic follicular somatic cells has recently been proposed with the advent of iPSC research mechanistic studies on granulosa cell differentiation[130–132]. Future work in this area should focus on lineage tracing for follicular cells in *in vitro* culture systems, either through immunofluorescent staining for granulosa- and theca-specific markers throughout culture or using a more sophisticated Cre/Lox reporter system for the expression of cell type-specific markers during culture. Single-cell profiling of ovarian follicles cultured *in vitro* would also reveal changes in the cellular makeup of the structure as maturation proceeds, and these findings could be benchmarked against data from follicles maturing *in vivo* such as the work recently published by Morris et al. 2022[75].

Our work on the transcriptome and proteome of follicles cultured *in vitro* also brings to light a number of questions on how *in vitro* maturation differs from folliculogenesis *in vivo*, and the impact of any differences on follicle maturation and oocyte competence. In our microarray study, we made high-level comparisons to microarray data from *in vivo* follicles published by Bernabe et al. 2019[33], identifying a number of unsurprising metabolism-related differences in *in vitro* vs. *in vivo* follicles. However, more clear conclusions may be drawn using a single-cell sequencing approach, where *in vitro* follicles could be compared to *in vivo* follicles of the same size and stage in a pairwise experimental design. This type of single cell analysis may also reveal how the process of hydrogel encapsulation impacts follicular cells, where various biomaterial platforms such as alginate and PEG could be compared. With the increasing accessibility of single-

cell technology, this type of in-depth analysis of *in vitro* follicles is the next step towards benchmarking a translational, standardized *in vitro* culture system against healthy follicles developing *in vivo.*

5.2.3 Conclusions

In conclusion, this work marks a significant step forward in the design of a well-defined universal follicle culture system for fertility preservation. By unraveling the complexities of folliculogenesis through comprehensive transcriptomic and proteomic analysis, these findings serve as foundational knowledge for designing a biomimetic microenvironment to facilitate folliculogenesis *in vitro* through delivery of exogenous cues. The analysis of the human ovary and its findings will aid in the transition from murine to human follicle culture, where identification of human-specific follicle maturation cues will compliment advances in biomaterials that accommodate the expansion of human follicles. This dissertation also strengthens our knowledge of folliculogenesis and reproductive biology, which will inspire future discoveries that shape the future of women's health and assisted reproduction.

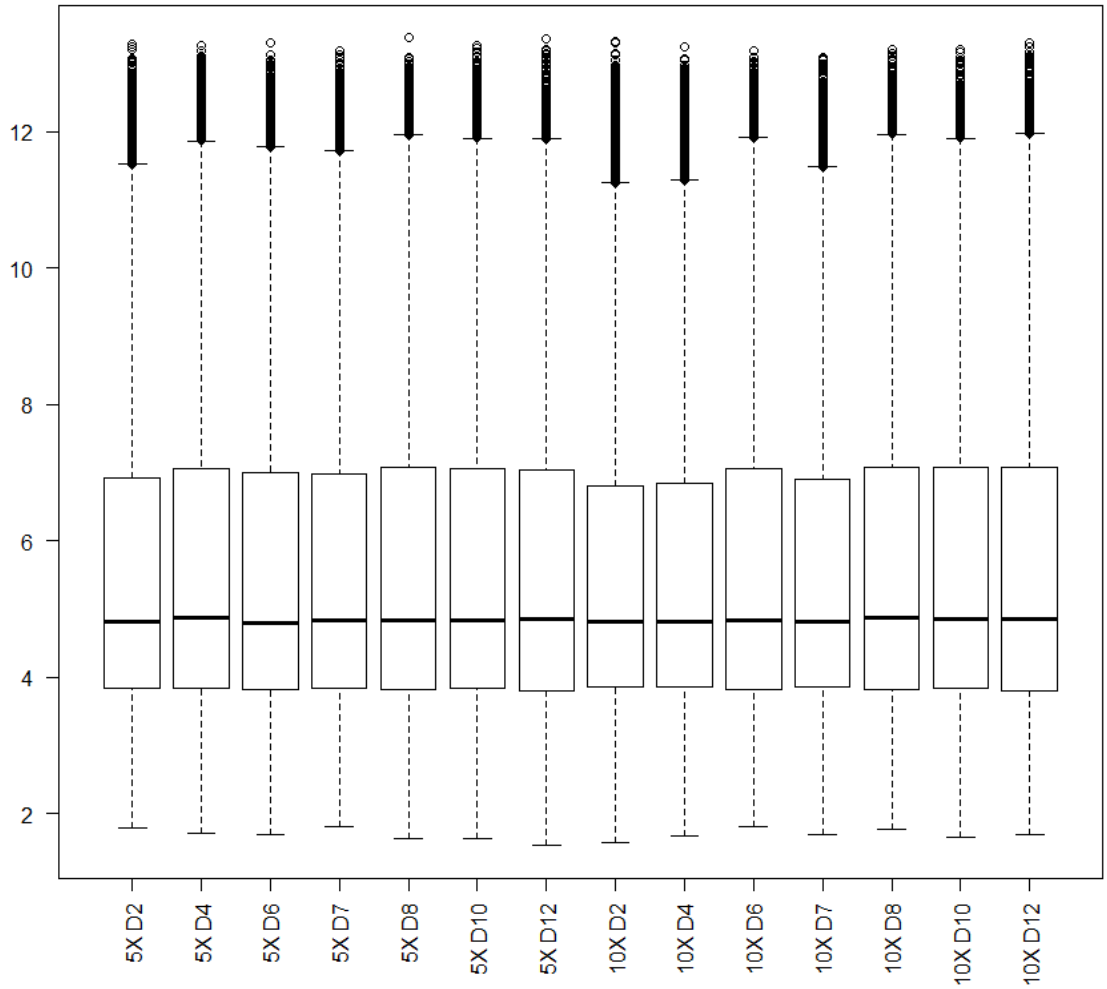
Appendices

Appendix A: Supplemental Tables and Figures

Microarray Sample Collection: Number of Follicles Isolated										
Day of Cell Isolation										
	Day 2		Day 4	Day 6	Day 7	Day 8		Day 10	Day 12	Total
5X	40	30	60	45	45	30	40	50	40	380
10X	50	40	60	60	50	50		60	60	430

Shading Denotes Experiment Number								
1	2	3	4	5	6	7	8	9

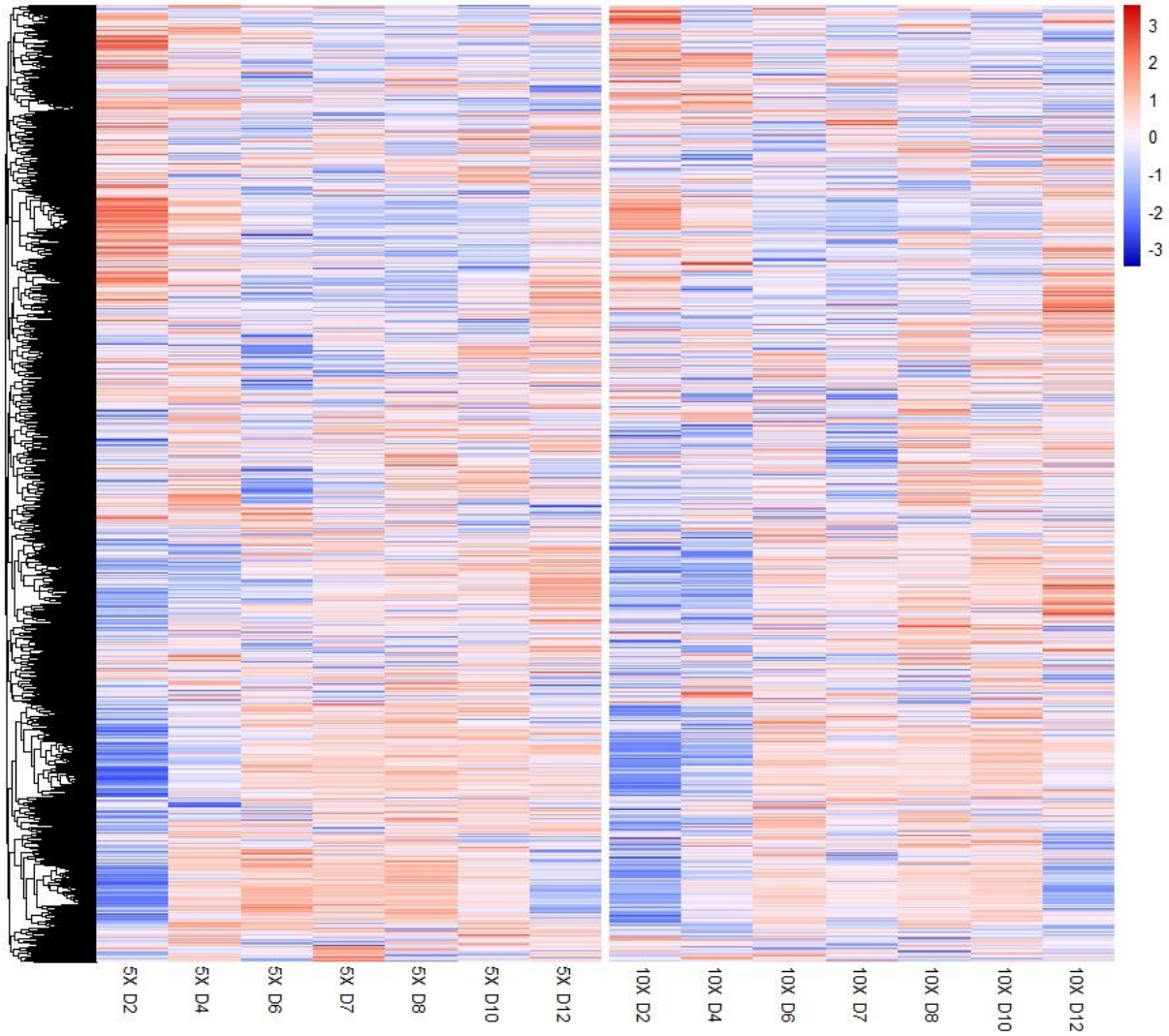
Appendix Table A.1: N values for each of the 14 microarray samples.



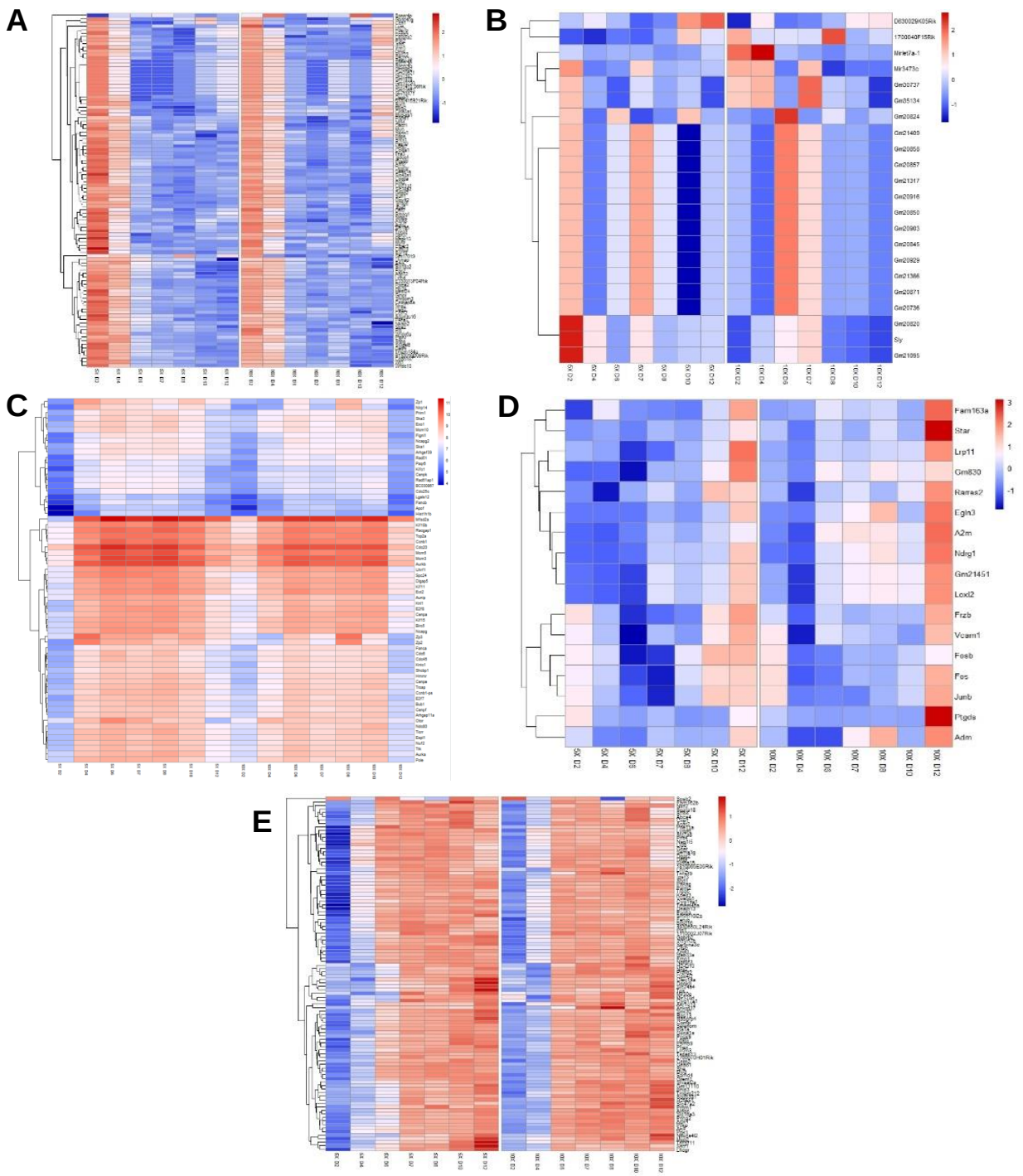
Appendix Figure A.1: Boxplot of all microarray samples after normalization. Y axis = raw intensity.

Cluster Color	High	Low
Magenta	5X Days 6 to 12, 10X Days 4 to 12	5X Days 2 and 4, 10X Day 2
Yellow	5X Days 2, 4, 8, 10, 12 and 10X Days 2, 4, 6, 8, 10, 12	5X Days 6 and 7, 10X Day 7
Purple	5X Days 4 to 12, 10X Days 6 to 12	5X Day 2, 10X Days 2 and 4
Blue	5X Days 4 to 10, 10X Days 4 to 10	5X Days 2 and 12, 10X Days 2 and 12
Orange	5X Days 6 to 12, 10X Days 6 to 12	5X Days 2 and 4, 10X Days 2 and 4
Cyan	5X Days 10 and 12, 10X Day 12	5X Days 2 to 8, 10X Days 2 to 10

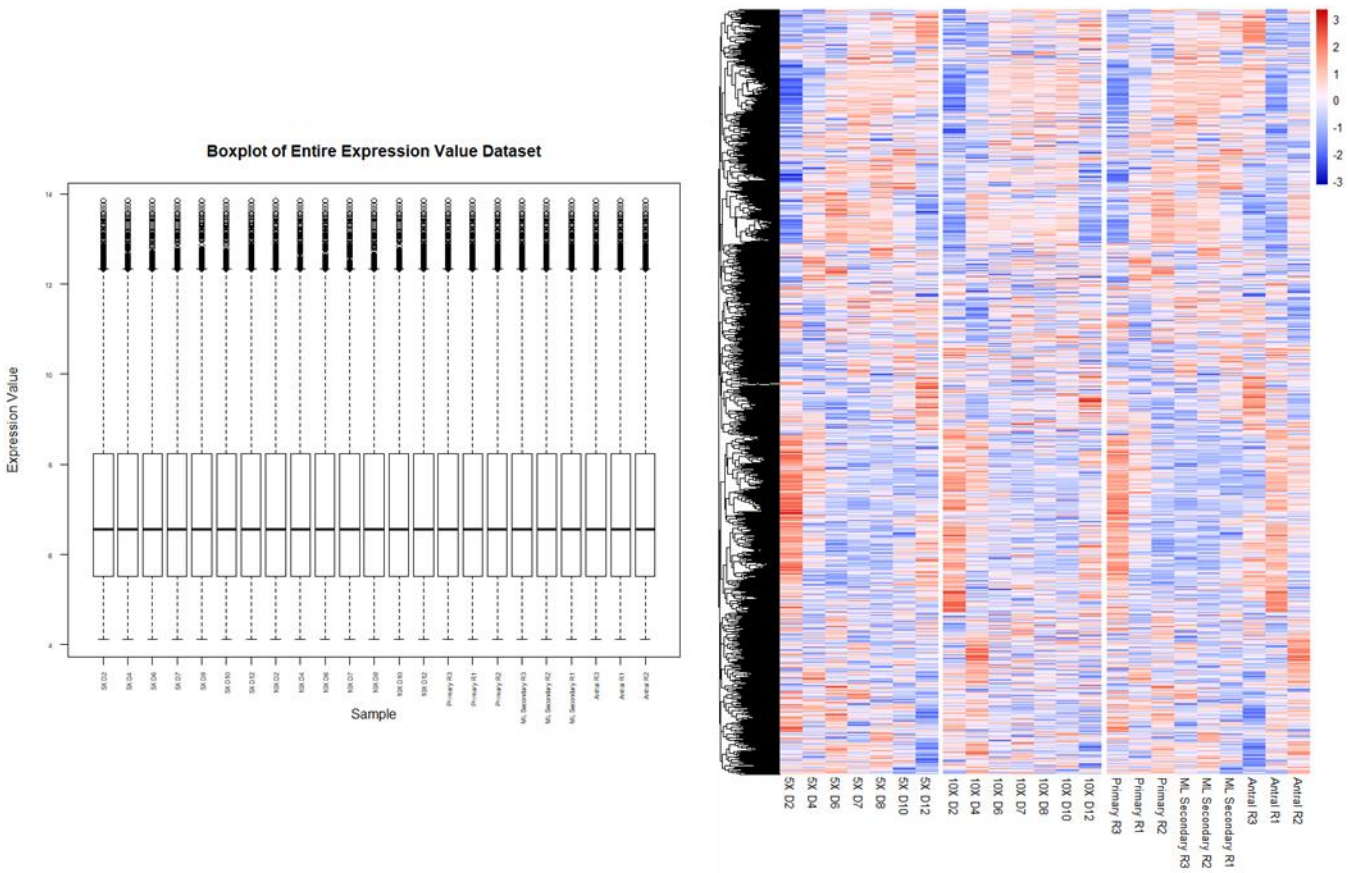
Appendix Table A.2: Table outlining experimental design for limma differential gene expression analysis in Figure 2; all regressions were performed high~low.



Appendix Figure A.2: High resolution heatmap of all normalized gene expression (n = 13,313 genes) across all 14 microarray samples. Plot was generated using pheatmap (version 1.0.12)59 in R (version 3.6.9, https://www.r-project.org/)52.



Appendix Figure A.3: Top differentially expressed genes from LR path analysis as reported in Figure 2. a) Top 100 genes with logFC > 1.5 for the magenta cluster, b) top 17 genes with logFC > 1.5 for the purple cluster, c) top 65 genes with logFC > 1.5 for the blue cluster, d) top 100 genes with logFC > 1.5 for the orange cluster, and e) top 17 genes with logFC > 1.5 for the cyan cluster. A maximum of 100 genes were plotted for each cluster; all genes with logFC > 1.5 were plotted for clusters with less than 100 genes meeting these criteria. Plots were generated using pheatmap (version 1.0.12)59 in R (version 3.6.9, https://www.r-project.org/)52.



Appendix Figure A.4: Boxplot and heatmap of in vitro microarray samples with in vivo microarray samples, normalized together as outlined in the methods section. The heatmap shows all 11,424 genes shared between the datasets. Plots were generated using R (version 3.6.9, https://www.r-project.org/)52, with the heatmap on the right also using pheatmap (version 1.0.12)59.

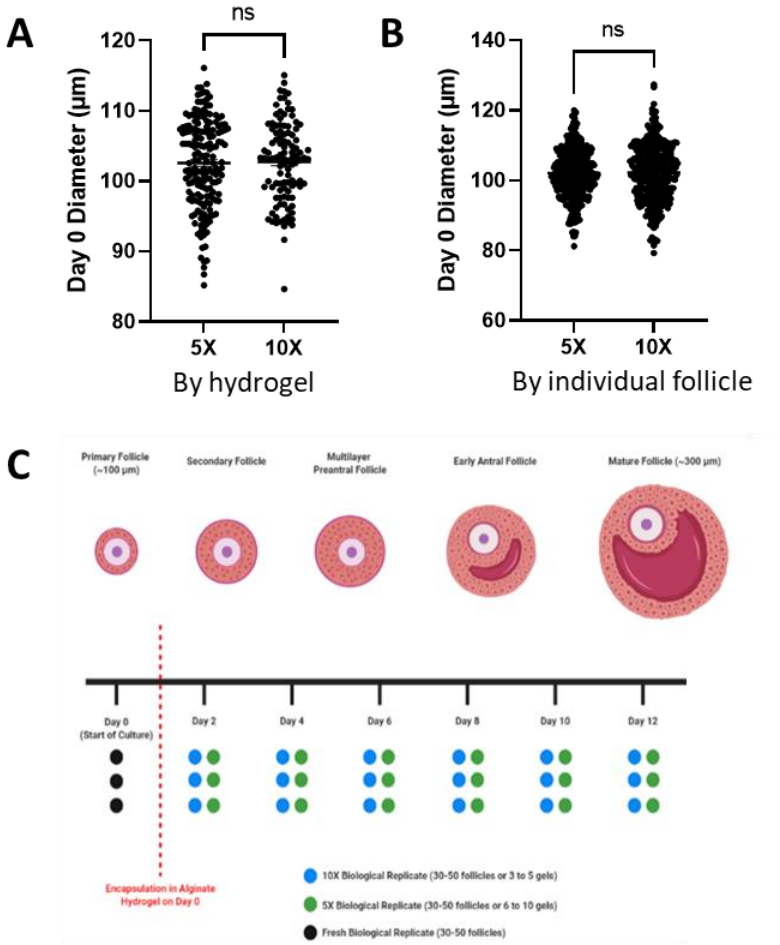
RNAseq Sample Collection: Number of Follicles Isolated			
Day of Culture	Replicate	5X	10X
Fresh Control (D0)	1	100	
	2	100	
	3	100	
	4	100	
	5	100	
	6	100	
	7	100	
	8	100	
	9	100	
2	1	50	60
	2	65	70
	3	45	60
4	1	40	60
	2	60	70
	3	50	70
6	1	50	50
	2	70	60
	3	40	70
8	1	45	70
	2	70	60
	3	35	60
10	1	55	60
	2	50	70
	3	45	60
12	1	55	60
	2	55	50
	3	30	50

Shading Denotes Experiment Number								
1	2	3	4	5	6	7	8	9

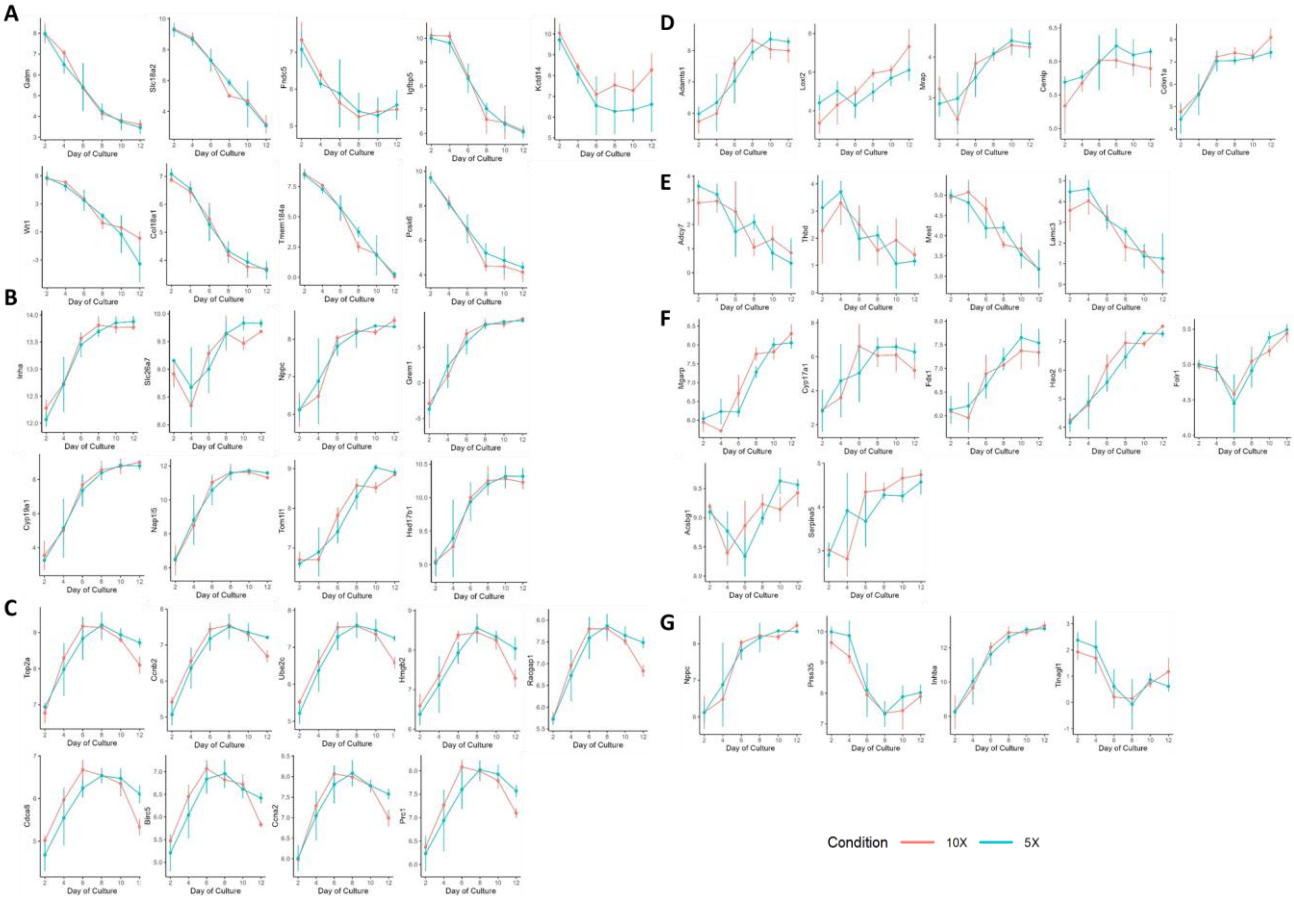
Appendix Table A.3: Follicle numbers for each RNA sample for RNA sequencing. Color denotes experiment number.

Day of Culture	Follicle Diameter (µm)	Follicle Volume (µm^3)	Antrum Diameter (µm)	Antrum Volume (µm^3)	Volume for Granulosa Cells (minus antrum + oocyte)	Number of Granulosa Cells in Follicle
0	100	5.23×10^5	NA	NA	4.58×10^5	875
2	125	1.022×10^6	NA	NA	9.565×10^5	1,829
4	150	1.767×10^6	NA	NA	1.7015×10^6	3,253
6	200	4.189×10^6	NA	NA	4.1235×10^6	7,884
8	225	5.964×10^6	50	6.55×10^4	5.833×10^6	11,153
10	280	1.149×10^7	75	2.208×10^5	1.1203×10^7	21,420
12	300	1.413×10^7	100	5.23×10^5	1.354×10^7	25,890

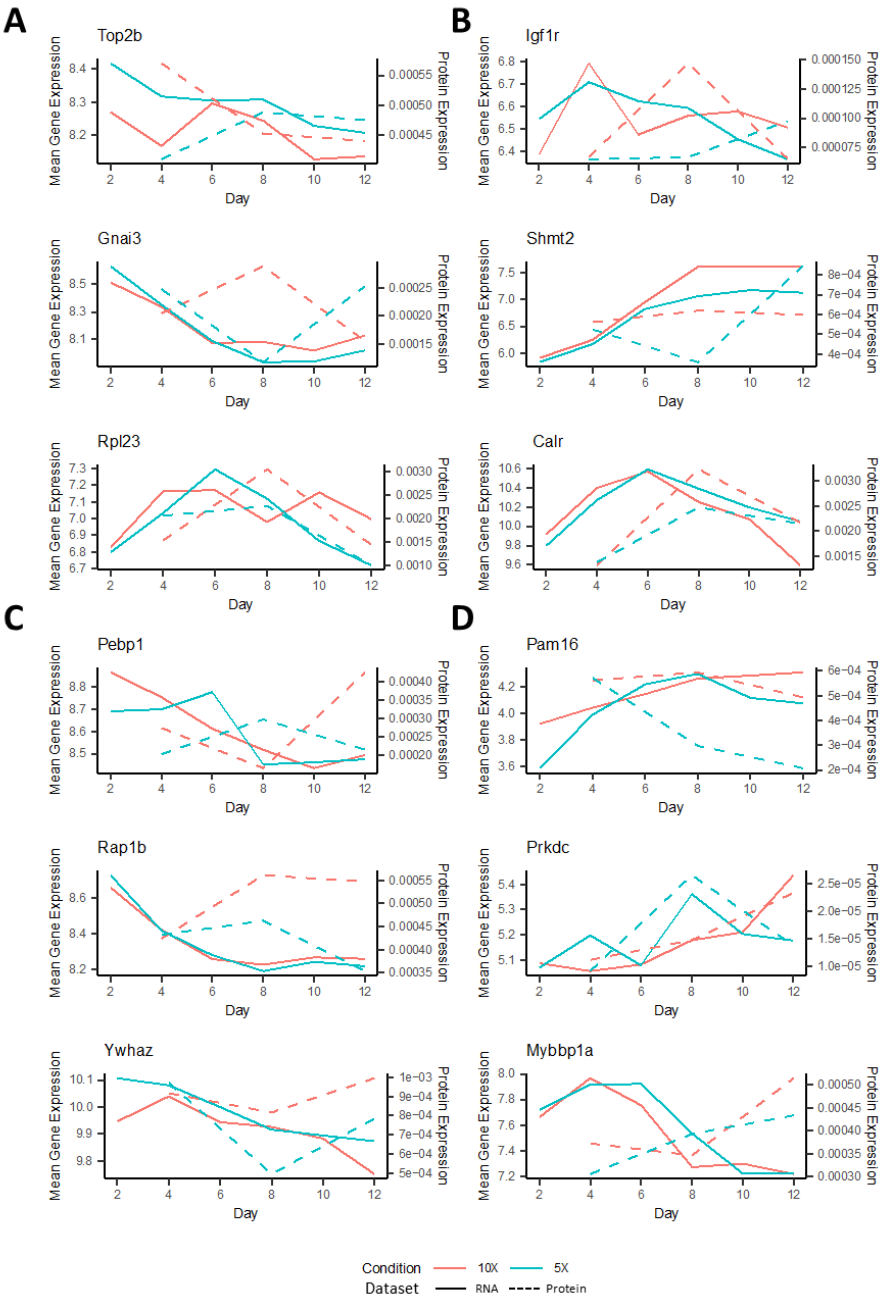
Appendix Table A.4: Table with theoretical volume for follicles cultured *in vitro* over time, calculated based on average diameter at each time point. Assumptions: oocyte is 50 µm in diameter across all time points; granulosa cells are 10 µm in diameter across all time points. Based on these assumptions and using the equation for volume of a sphere $V = \frac{4}{3}\pi r^3$, the volume of a granulosa cell is 523 µm^3 and the volume of an oocyte is 65,500 µm^3.



Appendix Figure A.5: A) Initial diameter of ovarian follicles for RNA sequencing, by hydrogel (5X or 10X), visualizing no statistical difference in starting diameters across conditions. B) Initial diameter of ovarian follicles for RNA sequencing by individual follicle. C) Schematic visualizing experimental design for RNA sequencing sample collection, where each dot is an RNA sample. The scheme was completed for somatic cells (n=45 samples) and oocytes (n=45 samples).



Appendix Figure A.6: Somatic cell transcriptome expression of marker genes for A) preantral granulosa, B) antral granulosa, C) mitotic granulosa, D) luteinizing granulosa, E) early theca, and F) steroidogenic theca.



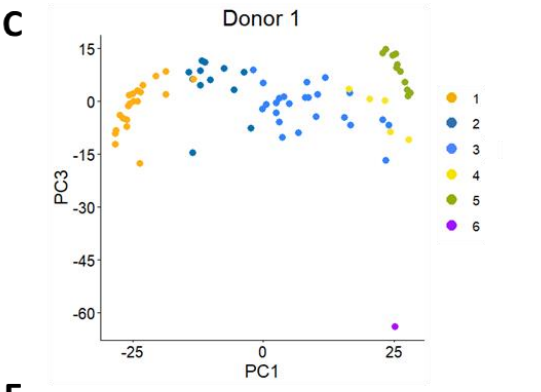
Appendix Figure A.7: RNA and protein expression of genes/proteins related to A) cell cycle, B) cell proliferation, C) MAPK signaling, and D) cell survival.

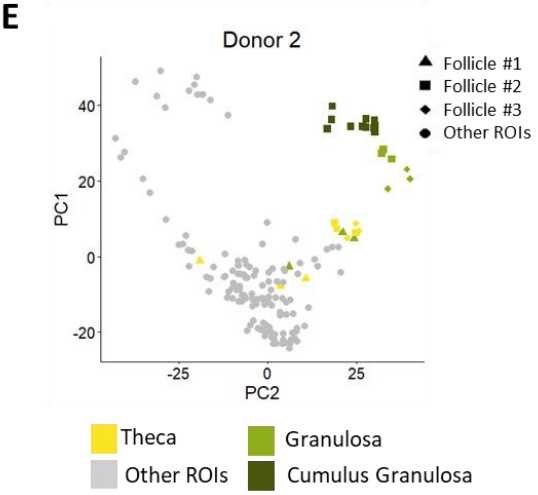
A

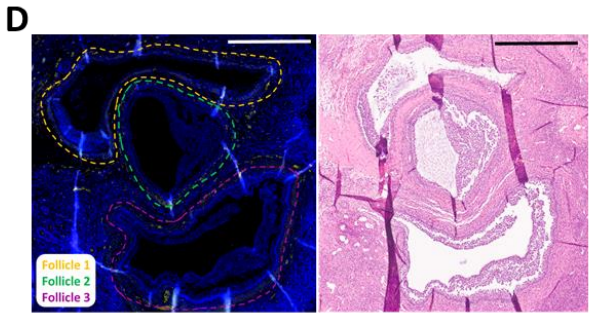
Sample	Age	BMI	Race	Cold Ischemic Time (hr)	Total Cells Sequenced	Cells Retained after Filtering	Average nCount_RNA	Average nFeature_RNA	Average % MT
1	27	28.902	Black	21.2	N/A	N/A	N/A	N/A	N/A
2	18	33.498	White	7.4	N/A	N/A	N/A	N/A	N/A
3	28	26.402	Asian	25.4	14,322	7,571	4,741	1,559	7.73
4	37	30.45	Black	26.6	13,901	7,288	8,494	2,441	5.68
5	31	31.608	White	14.7	9,149	6,339	7,182	2,242	6.47

B

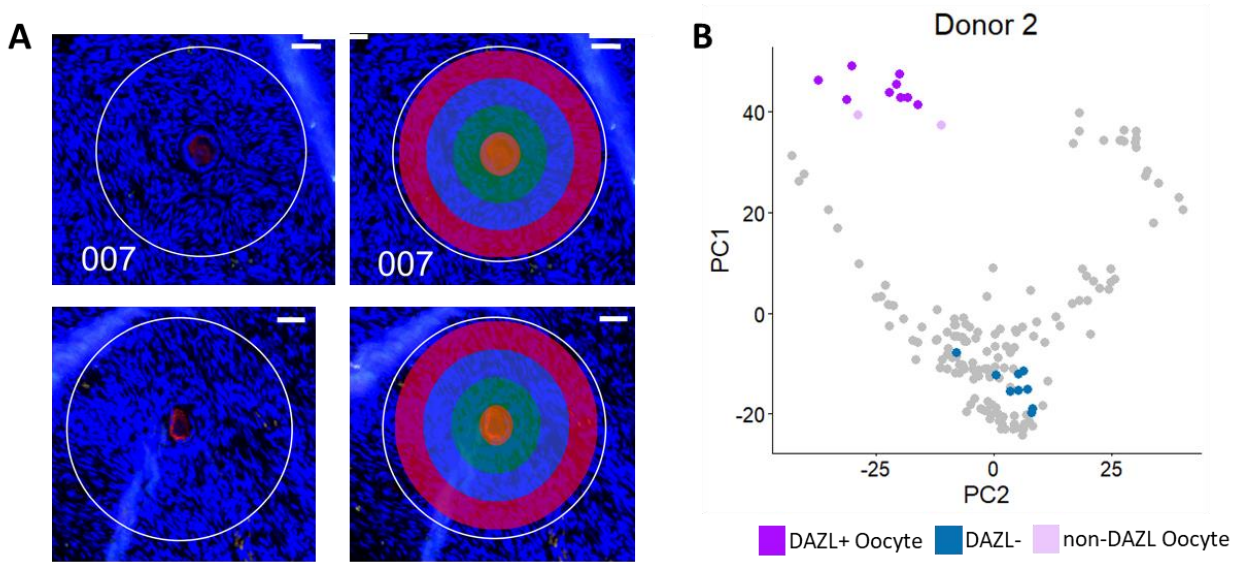
ROI Type #	ROI Type Description	Tissue Region	Donor 1, #ROIs	Donor 2, #ROIs
1	Ovarian Surface	Cortex	2	0
2	Cortex		3	0
3	Cortex Series		11	47
5	Primary Follicle Rings		0	6
10	Secondary Follicle		5	0
6	Medulla	Medulla	15	0
7	Follicle-Adjacent Medulla		5	3
8	Medulla Series		0	31
9	Vascular Regions		4	0
15	Cyst – Inner Layer		0	5
16	Cyst—Outer Layer		0	5
19	Target-Enriched (PGR)		0	6
17	Target-Enriched (CD68)		3	6
18	Target-Enriched (SMA)		8	0
4	Early Stage Follicles	Follicle	0	11
11	Late Stage Oocyte		1	0
12	Cumulus Granulosa		4	12
13	Granulosa		7	9
14	Theca		5	9



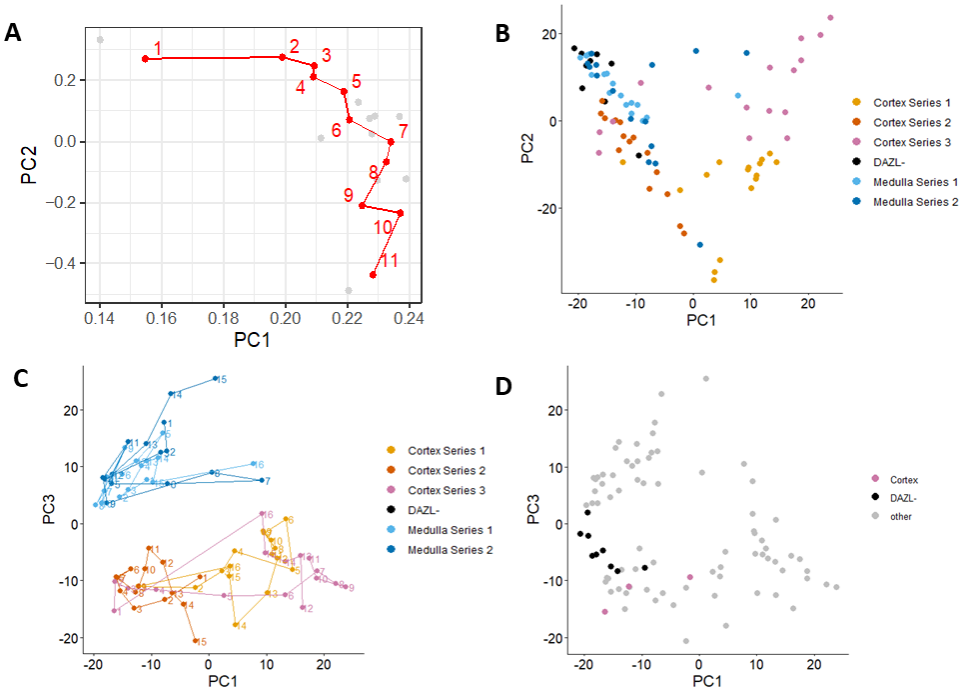




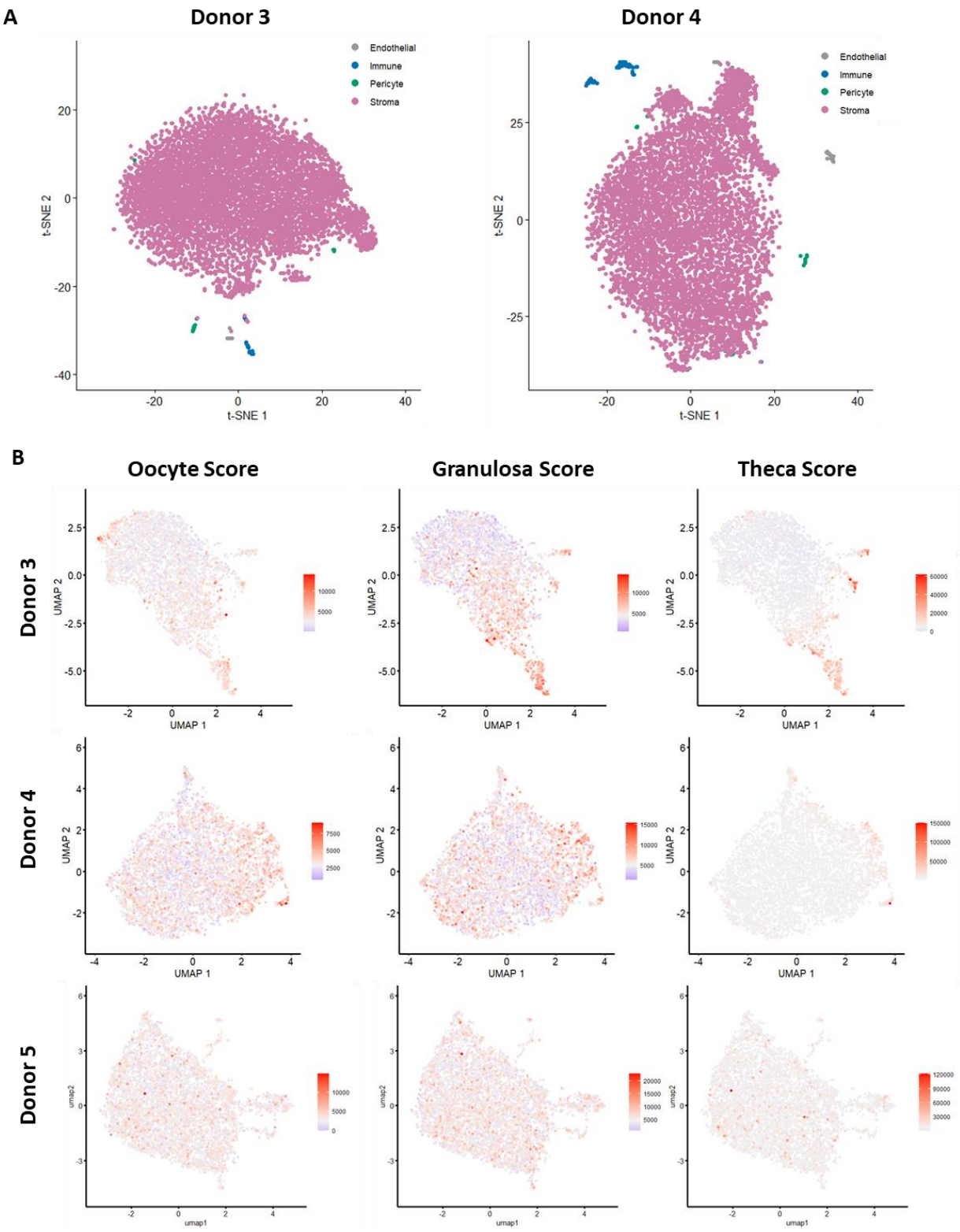
Appendix Figure A.8: A. Table of tissue donor information. B. Table of ROI types, descriptions, and number of ROIs for Donors 1 and 2 used for ST analysis. C. PC1-3 plot for Donor 1 ST clusters, highlighting the individual oocyte from Donor 1 in purple. D. IF (left) and H&E (right) images of Follicles 1-3 from Donor 2 (scale bar = 700 µm). E. PC1-2 plot highlighting ROIs from Donor 2 Follicles 1-3, where Follicle #1 ROIs shifted from the clustering of Follicle #2-3 ROIs.



Appendix Figure A.9: A. Images of early-stage oocytes sampled using contour modality, with ROI #7 on top and ROI #10 on bottom, without (left) and with (right) overlay of ROI segments, scale bar = 30 µm. B. PC1-2 plot of Donor 2 ROIs highlighting the transcriptional difference between DAZL+ segments (purple) and DAZL- segments (blue) from ROI Type #4.



Appendix Figure A.10: A. PC1-2 plot of 11 cortex layer ROIs visualizing linear progression of transcriptome profiles. B. PC1-2 plot of the 3 cortex and 2 medulla series plus the 9 DAZL- subareas highlighting the transcriptional similarity of Cortex Series 2 to the two Medulla Series. The six series are shown in different colors. C. PC1-3 plot of the 3 cortex and 2 medulla series with lines and numbers labeling layers 1-15/16, highlighting the non-linear transcriptome of each series.



Appendix Figure A.11: A. t-SNE projections for Donors 3 (left, n = 7,571 cells) and 4 (right, n = 7,288 cells) for the four major cell types identified. B. Focused reprojection of stromal cells from Donors 3, 4, and 5 (rows), each scored for expression of oocyte (n=76), granulosa (n=96), and theca (n=46) cell marker genes identified from ST analysis (columns).

Appendix B: Pilot Experiments

Preface: This appendix includes pilot experiments I performed during my doctoral studies. Each section includes my motivating hypothesis for the work, the methods employed, preliminary results, limitations or challenges encountered, and recommendations for future work in each area.

I. **Individual Primary Ovarian Follicle Rescue**

Motivation and Hypothesis

My work using the murine follicle group culture system as a model for understanding early-stage folliculogenesis *in vitro* is predicated on the idea of developing a well-defined, universal, translational follicle culture system through the use of exogenous factors such as growth factors and cytokines that will recapitulate the paracrine signaling experienced by follicles *in vivo*. Previous studies using conditioned media to "rescue" individual early-stage follicles *in vitro* support this hypothesis[133] and I sought to add to the growing body of work in this area using the 5X/10X group culture system specifically, alongside a preliminary cocktail of exogenous factors identified by Zhou et al.[15] using Luminex® cytokine analysis. We hypothesized that the preliminary cocktail, or "supplemented media", would improve growth and survival of individually-encapsulated and cultured primary follicles *in vitro* compared to control follicles cultured in standard growth media. Likewise, we expected that conditioned media from 10X follicles, which contains a wealth of factors secreted by the follicles during culture, would rescue development of individual primary follicles through unidirectional delivery of the paracrine cues from 10X culture.

Methods

Mouse primary ovarian follicles (90 – 110 μm) were enzymatically isolated from 10-12 days old B6CBAF1 pups and individually encapsulated in 0.3% (w/v) alginate. Follicles were cultured individually three different conditions for 12 days. The control condition was standard growth media (GM) (α-MEM supplemented with 3 mg/mL bovine serum albumin, 1 mg/mL bovine fetuin, 5 μg/mL insulin, 5 μg/mL transferrin, 5 ng/mL selenium, and 10 mIU/mL highly purified, human-derived, follicle-stimulating hormone (FSH); 2) The supplemented media (SM)

condition was GM supplemented with VEGF-A, MCP-1, KC, and follistatin, which were most associated with 10X follicles in a previous cytokine analysis[15], and 3) conditioned media (CM) from follicles cultured in 10X, where follicles received conditioned media starting on Day 0 (i.e. on Day 0, individual follicles received CM from Day 2 10X follicles; on Day 2, individual follicles received CM from Day 4 10X follicles, and so on). The rescue cocktail components were supplemented in concentrations derived directly from cytokine analysis, as shown below in Appendix Table B.1.

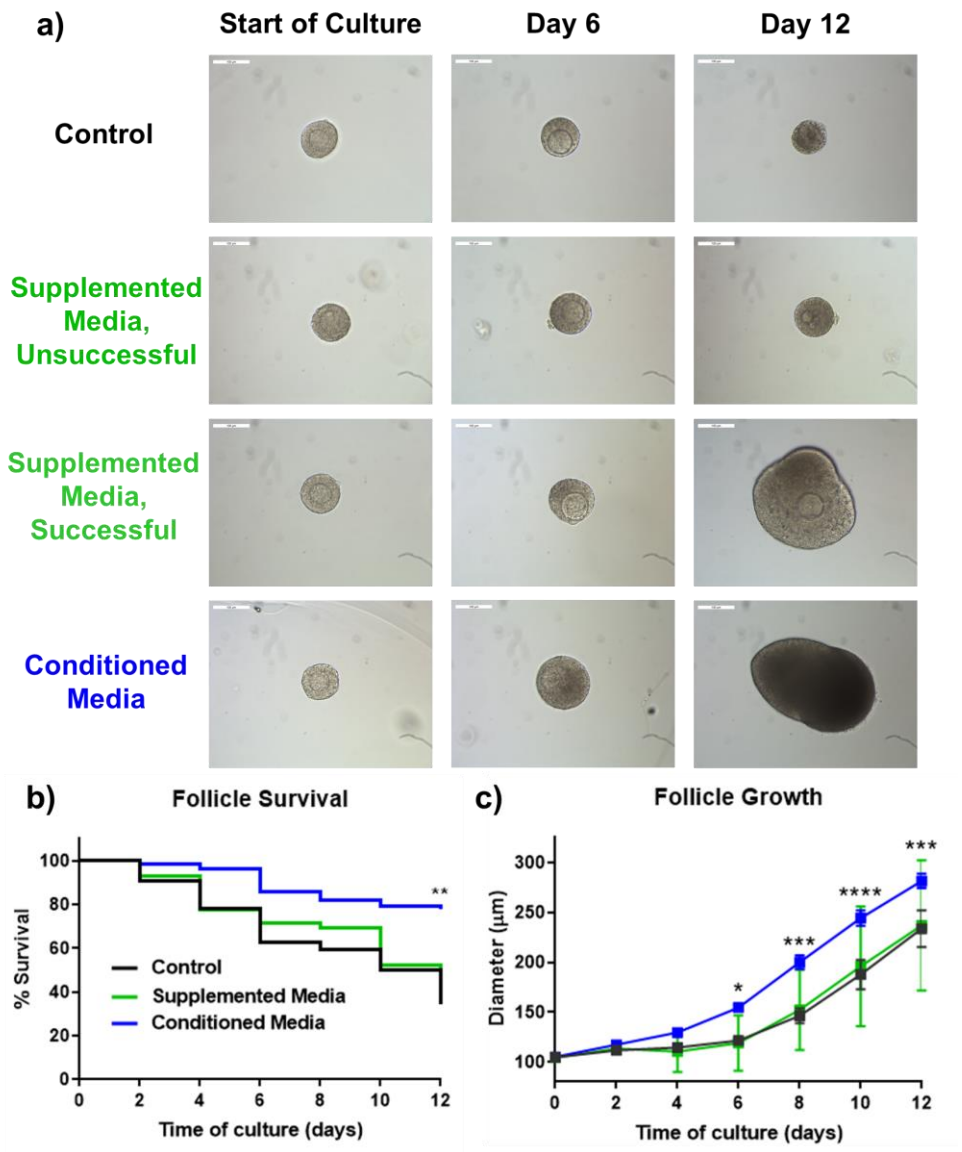
Day of Culture	VEGF-A (pg/mL)	MCP-1 (pg/mL)	KC (pg/mL)	Follistatin (pg/mL)
4	0.25	50	500	50
6	1.75	75	1000	275
8	20	100	1750	2500
10	100	100	2000	7500
12	250	125	3500	1500

Appendix Table B.1: Concentrations of each cytokine added to the supplemented media cocktail, derived from actual concentrations present in cultures from previous studies.

Follicle growth and survival were tracked over the duration of culture. Media from the culture experiments was collected every 2 days and subjected to Luminex cytokine analysis to identify factors differentiating the culture conditions and their outcomes.

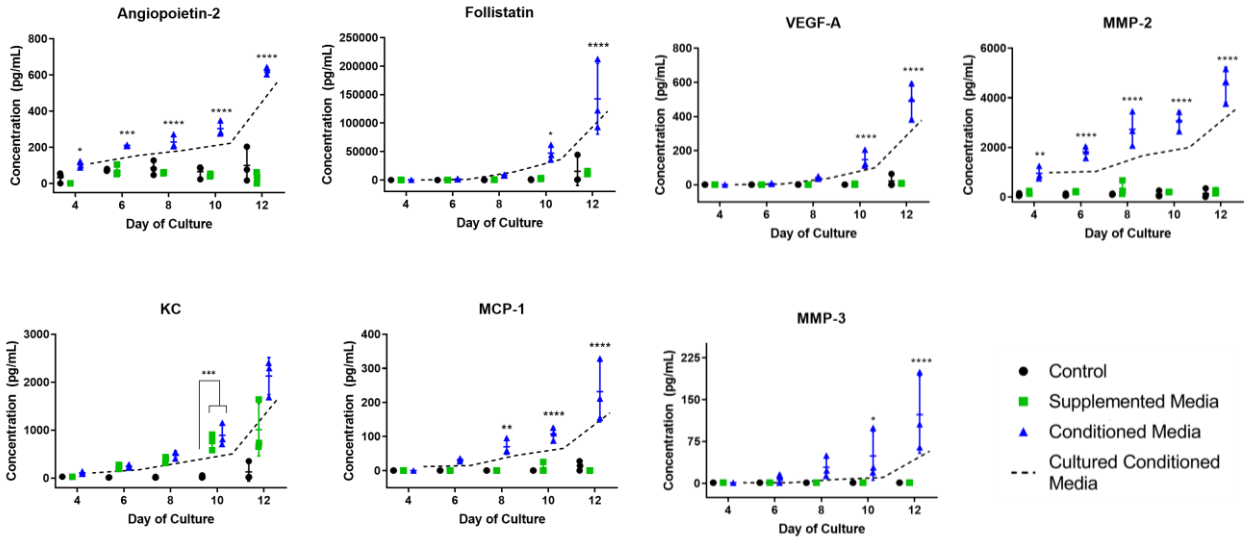
Results

Compared to control (n = 21 follicles) and SM (n = 58 follicles) (34% and 40% survival, respectively, after 12 days of culture), CM (n = 74 follicles) significantly ($p<0.01$) improved follicle survival (78%) as shown in Appendix Figure B.1. Some follicles in the SM group reached sizes up to ~300 μm, but the majority of the cohort did not survive to the end of culture. Follicles in the CM group grew to 281 ± 7.28 μm (mean ± SEM), which was significantly ($p<0.0001$) larger than the SM group, which only reached 237 ±



Appendix Figure B.1: a) Primary follicles cultured in experimental media conditions over 12 days of culture, including successful and unsuccessful follicles cultured in supplemented media (scale bar = 100 µm), b) follicles cultured in conditioned media have significantly higher survival (78%) over 12 days of culture compared to control and supplemented media conditions (34% and 40% respectively), and c) follicles cultured in conditioned media reached an average of 281 ± 7.28 µm (mean ± SEM) compared to supplemented media (237 ± 9.61 µm) and control groups over 12 days of culture. *p < 0.05, **p < 0.01, ***p < 0.001, ****p< 0.0001

9.61 µm at the end of culture. The lack of success in the SM group suggests that the added 4 cytokines (follistatin, KC, MCP-1, and VEGF-A) provided an incomplete milieu of factors required to promote folliculogenesis and normally present *in vivo* or secreted when follicles are cultured in groups. Cytokine analysis of conditioned media from control, SM, and CM follicles showed vastly higher concentrations of cytokines of interest in the CM group, suggesting either a)



Appendix Figure B.2: Primary follicles cultured in experimental media conditions over 12 days of culture, including successful and unsuccessful follicles cultured in supplemented media (scale bar = 100 µm), b) follicles cultured in conditioned media have significantly higher survival (78%) over 12 days of culture compared to control and supplemented media conditions (34% and 40% respectively), and c) follicles cultured in conditioned media reach an average of 281 ± 7.28 µm (mean ± SEM) compared to supplemented media (237 ± 9.61 µm) and control groups over 12 days of culture. *p < 0.05, **p < 0.01, ***p < 0.001, ****p< 0.0001

production of these cytokines by the individual follicles or, less likely b) unused cytokines in the growth media, unlikely because of the relatively short half-life of most such factors in culture (Appendix Figure B.2).

Limitations and Challenges

One significant limitation of this preliminary study was that cytokine analysis was performed in-house using methods deviating from the manufacturer's instructions, which may have skewed the measurements obtained. I also did not perform *in vitro* maturation on oocytes from SM and CM follicles which might have informed us whether the different culture conditions negatively affected the quality of the oocytes resulting in poor oocyte quality despite overall growth and survival of the follicles. Future work in development of a rescue cocktail for individual follicles should undoubtedly include IVM and eventually live-birth studies to confirm the sufficiency of the cocktail for proper oocyte maturation. However, the growth and survival data presented in this section, along with the normal morphology observed during culture, suggest that IVM outcomes may be favorable.

Directions for Future Work

Our cocktail of four factors was insufficient to rescue individual primary follicles *in vitro*, but CM from follicles cultured in groups promoted follicle growth and survival, which agrees with previous work using unidirectional signaling via CM for ovarian follicle culture. Follicles cultured in CM had a distinct secretory profile compared to control and SM groups, with the caveat that cytokine analysis results should be viewed in light of the limitations outlined above. Future cytokine studies should be conducted following the manufacturer's instructions, and custom-made, broader cytokine arrays that include more targets, should be used in a data-driven approach using results from transcriptome and proteome profiling. Further work in this area should build on the work outlined in Chapters 2 and 3, which suggest numerous signaling pathways and factors that may be suitable for exogenous supplementation and rescue of individual primary follicles in well-defined culture system.

II. Androstenedione Supplementation for Follicle Culture

Motivation and Hypothesis

Previous work using the 5X/10X culture system profiled hormones present in conditioned media from follicles in group culture and reported higher concentrations of androstenedione, estradiol, and progesterone in 10X follicles. The presence of androstenedione, which is an essential precursor for estradiol, was of interest in this study, since ovarian follicles isolated at the primary stage and cultured *in vitro* do not develop a well-defined theca cell layer (the cells responsible for androstenedione production) analogous to that observed *in vivo*. Presence of androstenedione in the culture media suggests the presence of theca cells, but we hypothesized that in group culture, cumulatively there were sufficient theca cells for androgen production whereas individual primary follicles *in vitro* did not contain sufficient pre-theca cells to support androstenedione production. Accordingly, I investigated if androstenedione supplementation could rescue individual primary follicles *in vitro* by providing this crucial hormone exogenously.

Methods

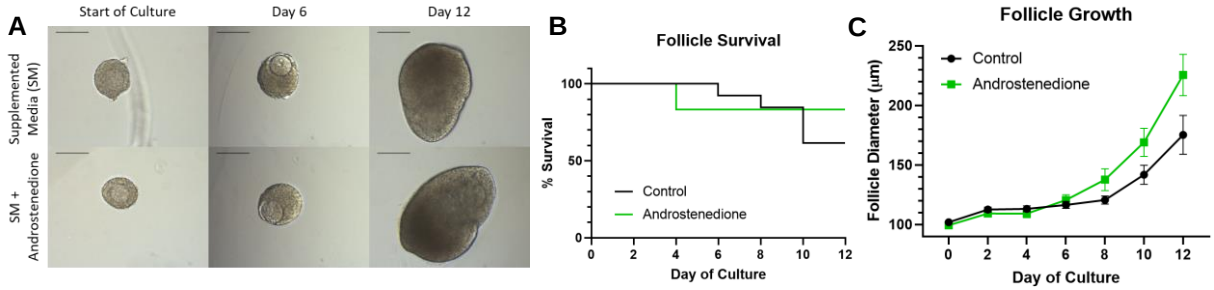
Mouse primary ovarian follicles (90 – 110 μm) were enzymatically isolated from 10-12 days old B6CBAF1 pups and individually encapsulated in 0.3% (w/v) alginate. Follicles were cultured in SM alone, following the same schedule and cytokine concentrations outlined in the previous section, or SM plus androstenedione, delivered in concentrations measured in a previous study[15] and outlined in Appendix Table B. Follicle growth and survival were tracked over the 12-day duration of culture.

Day of Culture	Androstenedione (ng/mL)
4	0.2
6	0.2
8	0.2
10	0.4
12*	0.5

Appendix Table B.2: Concentrations of each cytokine added to the supplemented media cocktail, derived from actual concentrations present in cultures from previous studies.

Results

Androstenedione + SM showed some improvement over SM alone, but the experiment was only performed twice times (n = 32 for control, n = 32 for androstenedione + SM) so it is difficult to determine whether the effect of androstenedione supplementation would significantly improve follicle outcomes.



Appendix Figure B.3: a) Primary follicles cultured in SM or SM + androstenedione over 12 days of culture (scale bar = 100 µm), follicles cultured in the androstenedione condition showed modest improvement in survival (b) and growth (c), but differences were not statistically significant.

Limitations and Challenges

These experiments were performed three times with no statistically significant difference between the conditions, in follicle survival or growth. It is possible that with a larger sample size, statistically significant differences could emerge. However, it is also possible that the supplementation of this one hormone alone, added to a preliminary rescue cocktail with limited success, is not sufficient for individual primary follicle rescue *in vitro*.

Directions for Future Work

Future studies in this area should focus on a data-driven selection of the targets most likely to improve follicle development *in vitro.* Much like the recommendations made in the previous

section, pathways and factors identified in the RNAseq and proteomics analyses in Chapters 2 and 3 are ideal candidates for improving the supplemented media cocktail. Hormone production in the follicle culture system is still of extreme importance and should be evaluated in future iterations of the individual primary follicle culture system. Additionally, the presence of theca cells in follicles cultured from the primary stage has not been confirmed outside of hormone analysis, and staining for theca precursor markers like *ANPEP* and steroidogenic theca cell markers such as *CYP17A1* would indicate what theca-like cells are present at the start of culture and when cells become steroidogenic.

Appendix C: Protocols

Preface: This appendix documents valuable methods and protocols I developed during my doctoral studies. These include human ovarian tissue digestion for single-cell sequencing, human ovarian tissue digestion for single-oocyte collection, and human ovarian tissue digestion for isolation of live follicles. All three protocols are derived from a set of core methods for enzymatic tissue digestion but vary in the concentration and composition of the enzymatic digest solution, the time required for cell isolation, and the viability of desired cell populations vs. those to be discarded. These protocols are significant in that they are the first such protocols developed and shared publicly for handling ovarian tissue and have been validated by numerous individuals in the Shikanov Lab as well as collaborators around the world.

I. Protocol for Human Ovarian Tissue Digestion for Single-Cell Sequencing

Abstract

This protocol outlines steps to obtain a viable single-cell suspension from fresh or cryopreserved human ovarian cortex and medulla. Here, we do not outline steps for removing the ovarian cortex from the organ and slicing 1 mm thick sections, nor do we outline cryopreservation and tissue thawing methods. Tissue is digested enzymatically using Collagenase IA and DNase I then strained to obtain a single-cell suspension with little debris. Viability can be tested before proceeding to cell sorting and sequencing. Note that RBC lysis step may not be necessary if organ was perfused prior to tissue collection.

Materials

A. Supplies for tissue preparation: glass petri dish, disposable #10 surgical blade, Dulbecco's Phosphate Buffered Saline without calcium chloride or magnesium chloride (DPBS -/-) (optional: tissue chopping device for mincing tissue).

B. Supplies for digestion: Small spatulas, 70 and 30 μm strainers, 30 and 60 mm petri dishes, conical tubes (1.5, 5, and 15 mL), sterile Dulbecco's Phosphate Buffered Saline with calcium chloride and magnesium chloride (DPBS +/+) and -/-, BSA (Fisher BP9706), FBS (Fisher 10-082-147), Collagenase IA (Sigma C2674), DNase I (Sigma D4138).

Preparations

1. Prepare enzymes: Dilute DNase I stock (2% w/v) to 2 mg/mL by combining 10 μL stock with 90 μL DPBS +/+. Mix well and store on ice for the duration of cell isolation. Collagenase stock at 100 mg/mL should be stored on ice for the duration of cell isolation.

2. Pre-weigh an empty 30 mm dish. This will be used to measure the mass of tissue prior to digestion.

3. Prepare solutions for digest.

3.1 Prepare Inactivation Solution (15 mL DPBS -/- with 10% FBS). Keep solution on ice for duration of experiment.

3.2 Aliquot 15 mL DPBS +/+ for digest solution. It is crucial that DPBS +/+ be used, as the magnesium and calcium aid in enzymatic digestion. Keep in a bead bath for duration of experiment to keep solution warm.

3.3 Prepare red blood cell (RBC) lysis buffer. Dilute 1 mL RBC lysis buffer stock into 9 mL sterile DI water. Keep solution warm in bead bath for duration of experiment.

3.4 Prepare 1% BSA in DPBS -/- (15 mL total) and refrigerate.

3.5 Prepare 0.04% BSA in DPBS -/- (15 mL total) and refrigerate.

Tissue Cutting and Digest

4. Starting with 1 mm thick pieces of ovarian cortex or medulla, cut all tissue into ~1mm^3 pieces either a) mechanically using a #10 blade in a glass petri dish or b) using a tissue chopper.

5. After cutting, move tissue to a cell strainer (any pore size) and rinse tissue with DPBS -/- twice to wash off any media or previous solutions.

6. Using a spatula, transfer the cut tissue to the pre-weighed 30 mm petri dish. Weigh tissue in pre-weighed dish to obtain tissue mass. Add 2 mL DPBS +/+ to the dish, then add 20 µL Collagenase IA (stock = 100 mg/mL) and 20 µL diluted DNase I for a solution of 1 mg/mL Collagenase IA and 0.02 mg/mL DNase I in DPBS +/+.

7. Place 30 mm dish with tissue in a secondary container and tape down if necessary. Tape dish to shaker and digest for 30 minutes at 150 rpm and 37ºC. Tissue pieces should be freely moving in the digest solution to obtain proper enzymatic digestion and leverage mechanical disruption for cell isolation.

8. After 30 minutes, strain tissue and supernatant through a 70 μm strainer over a 60 mm dish. Return undigested tissue to the original digestion dish using a spatula, and add fresh enzymatic solution (2 mL) to the tissue, then return it to the shaker for 30 additional minutes.

9. Meanwhile, add 2 mL ice-cold Inactivation Solution to the 60 mm dish of digested tissue, then strain through a 30 μm strainer into a clean 60 mm dish. Transfer the inactivated cell suspension to a 5 mL tube and store in ice.

10. Repeat Steps 8-9 with the digest dish containing the remaining tissue, adding the second inactivated cell suspension to the first.

11. Aliquot supernatant into 1.5 mL tubes and spin at 100g for 5 minutes.

12. Resuspend pellets in 3 mL 1X RBC lysis solution and incubate at RT for 2 minutes. Centrifuge at 100g for 5 minutes, aspirate supernatant and resuspend in 1 mL cold PBS.

13. Rinse 2x in PBS w/ 1% BSA to remove cell debris, centrifuging at 100g for 5 minutes after each rinse. Meanwhile prepare small HOT water bath to kill control cells – put water on hot plate to warm.

14. Aspirate supernatant and resuspend pellet in 1mL cold DPBS -/- w/ 0.04% BSA

15. Take a small aliquot of the suspension for unstained control and take a small aliquot to kill and use as positive control (dunk in very hot water for ~30 seconds)

16. Incubate suspension with DAPI at 3 uM per 1 million cells for 30 minutes on ice.

    a. Stock is 14.3 mM, so to make 5 mL working concentration, combine 5 mL solution with 1.05 uL DAPI.

    b. Incubate positive control with stain as well.

17. Use DAPI-stained cell suspension for fluorescence-activated cell sorting (FACS), where DAPI+ cells (with compromised cell membranes) can be removed to obtain a cell suspension with

high viability appropriate for single-cell sequencing. The positive control can be used to confirm that DAPI staining worked.

Comments and Reflections

This protocol maximizes cell viability across the board in the least amount of time possible. This makes it an ideal procedure for scRNAseq preparation of the whole ovary but does not favor the viability of follicular cells or other cell types that may be of interest. To obtain a cell suspension enriched specific or rare for cell type(s) of interest, digestion of large volumes of tissue and subsequent fluorescence-activated cell sorting (FACS) may be appropriate. Cortex and medulla may be digested together or separate, but separate digests for the two ovarian tissue types is advantageous in that it gives the experimentalist control over the contributions of each tissue to the final cell suspension. Future improvements to this protocol could include the introduction of different enzymes for tissue digestion or modifications that selectively protect the viability of certain cell types.

II. Protocol for Human Ovarian Tissue Digestion for Single-Oocyte Collection and RNA Sequencing

Abstract

This protocol outlines steps to obtain live denuded oocytes from fresh or cryopreserved human ovarian cortex for single oocyte RNA sequencing. Here, we do not outline steps for removing the ovarian cortex from the organ and slicing 1 mm thick sections, nor do we outline cryopreservation and tissue thawing methods. Tissue is digested enzymatically using Collagenase IA and DNase I then follicles are isolated from the digested tissue first using cell strainers and then by individual collection from the suspension. Follicles are further digested enzymatically until liberated oocytes can be collected by hand. Note that RBC lysis step may not be necessary if organ was perfused prior to tissue collection.

Materials

A. Supplies for tissue preparation: glass petri dish, disposable #10 surgical blade, DPBS -/- (optional: tissue chopping device for mincing tissue).

B. Supplies for digestion: Small spatulas, 70 and 30 μm strainers, 30 and 60 mm petri dishes, conical tubes (1.5, 5, and 15 mL), sterile DPBS +/+ and -/-, BSA (Fisher BP9706), FBS (Fisher 10-082-147), Collagenase IA (Sigma C2674), DNase I (Sigma D4138), hyaluronidase (Sigma H4272), 4-well plates, L-15 medium, Ovoil, stripper pipette with 75 μm tips, dissection microscope, wide-bore pipette tips, incubating shaker.

Preparations

1. Prepare enzymes: Dilute DNase I stock (2% w/v) to 2 mg/mL by combining 10 μL stock with 90 μL DPBS +/+. Mix well and store on ice for the duration of cell isolation. Collagenase stock at 100 mg/mL should be stored on ice for the duration of cell isolation.

2. Pre-weigh an empty 30 mm dish. This will be used to measure the mass of tissue prior to digestion.

3. Prepare solutions for digest.

3.1 Prepare Inactivation Solution (15 mL DPBS -/- with 10% FBS). Keep solution on ice for duration of experiment.

3.2 Aliquot 15 mL DPBS +/+ for digest solution. It is crucial that DPBS +/+ be used, as the magnesium and calcium aid in enzymatic digestion. Keep in a bead bath for duration of experiment to keep solution warm.

3.3 Prepare red blood cell (RBC) lysis buffer. Dilute 1 mL RBC lysis buffer stock into 9 mL sterile DI water. Keep solution warm in bead bath for duration of experiment.

Tissue Cutting and First Digest

4. Starting with 1 mm thick pieces of ovarian cortex, cut all tissue into ~1mm3 pieces either a) mechanically using a #10 blade in a glass petri dish or b) using a tissue chopper.

5. After cutting, move tissue to a cell strainer (any pore size) and rinse tissue with DPBS -/- twice to wash off any media or previous solutions.

6. Using a spatula, transfer the cut tissue to the pre-weighed 30 mm petri dish. Weigh tissue in pre-weighed dish to obtain tissue mass. Add 2 mL DPBS +/+ to the dish, then add 20 µL Collagenase IA (stock = 100 mg/mL) and 20 µL diluted DNase I for a solution of 1 mg/mL Collagenase IA and 0.02 mg/mL DNase I in DPBS +/+.

7. Place 30 mm dish with tissue in a secondary container and tape down if necessary. Tape dish to shaker and digest for 2 hours at 150 rpm and 37ºC. Tissue pieces should be freely moving in the digest solution to obtain proper enzymatic digestion and leverage mechanical disruption for cell isolation.

8. After 2 hours, strain tissue and supernatant through a 70 μm strainer over a 60 mm dish. A 30 μm strainer can also be used to "catch" follicles with diameters ranging from 30-70 μm in the top of the 30 μm cell strainer - backwash the 30 μm strainer with DPBS -/- to collect these follicles. To collect follicles with diameters > 70 μm, you may consider implementing a larger cell strainer (150 μm) and backwashing the 70 μm cell strainer, however this will also yield 70-150 μm tissue fragments remaining from the digest.

9. Add 2 mL ice-cold Inactivation Solution to the 60 mm dish of backwashed follicles to halt enzyme activity. Keep follicles in solution on ice and proceed with mechanical follicle collection using a dissection microscope and a stripper with a 75 μm tip. Follicles are best collected by hand with a stripper, then deposited in a 60 mm dish with L-15 drops (35 μL/ea) under Ovoil.

10. Optional RBC Lysis step: If tissue is bloody prior to digest, consider spinning down cell suspension for 5 minutes at 100g at the end of the first digest and resuspending in 1X RBC lysis buffer, then incubating at room temperature for 2 minutes to lyse red blood cells.

Second Digest and Oocyte Isolation

11. Collect all primordial follicles from initial digest and distribute among the wells of a 4-well plate with 0.5 mL DPBS +/+ per well. To each well, add a fresh 5 μL aliquot of Collagenase IA and 5 μL diluted DNase I. Use fresh enzyme aliquot directly from the freezer.

12. Place 4-well plate with follicles in a secondary container and tape down if necessary. Tape dish to shaker and digest for 1 hour at 150 rpm and 37ºC. After 1 hour, add 0.5 mL Inactivation Solution directly to each well of the 4-well plate to halt enzymatic digestion.

13. From the 4-well plate of digested follicles, isolate completely denuded oocytes by hand. To denude oocytes from semi-intact follicles, transfer individual follicles to 35 μL droplets of L-15

with 10 IU hyaluronidase (1.3 µL stock into 35 µL) and pipette up and down with 75 µm stripper tip to remove adhering cells.

14. Isolated oocytes can be imaged in the plate under oil, or transferred individually to a PCR plate pre-loaded with cell lysis buffer for direct cell lysis sequencing.

Comments and Reflections

This protocol sacrifices viability of all other cell types in favor of higher yield of viable oocytes in the shortest amount of time possible. The final steps of handling and moving individual follicles and oocytes turn translucent when freed from the granulosa cells. Working with the oocytes in small droplets of media under oil is essential to tracking and retaining the majority of isolated oocytes. In development of this protocol, we trialed numerous stripper tips with diameters ranging 30-150 µm, thinking that smaller diameter tips would allow for finer control of oocytes. In reality, the 30 and 50 µm tips were difficult to work with because they would not stay perfused, causing oocytes to get stuck. We concluded that 75 µm tips were most helpful for all steps in handling follicles and denuded oocytes. Future improvements to this protocol should focus on obtaining more oocytes or introducing methods that alleviate some of the mechanical manipulation involved, as this manipulation is how most oocytes are lost in the isolation process.

III. Protocol for Human Ovarian Tissue Digestion for Live Follicle Isolation for Artificial Ovary Hydrogels or Isolated Follicle Culture

Abstract

This protocol outlines steps to obtain live primordial/primary follicles from fresh or cryopreserved human ovarian cortex. Here, we do not outline steps for removing the ovarian cortex from the organ and slicing 1 mm thick sections, nor do we outline cryopreservation and tissue thawing methods. Tissue is digested enzymatically using Collagenase IA and DNase I then follicles are isolated from the digested tissue first using cell strainers and then by individual collection from the suspension. Note that RBC lysis step may not be necessary if organ was perfused prior to tissue collection.

Materials

A. Supplies for tissue preparation: glass petri dish, disposable #10 surgical blade, DPBS -/- (optional: tissue chopping device for mincing tissue).

B. Supplies for digestion: Small spatulas, 70 and 30 μm strainers, 30 and 60 mm petri dishes, conical tubes (1.5, 5, and 15 mL), sterile DPBS +/+ and -/-, BSA (Fisher BP9706), FBS (Fisher 10-082-147), Collagenase IA (Sigma C2674), DNase I (Sigma D4138), wide-bore pipette tips, incubating shaker.

Preparations

1. Prepare enzymes: Dilute DNase I stock (2% w/v) to 2 mg/mL by combining 10 μL stock with 90 μL DPBS +/+. Mix well and store on ice for the duration of cell isolation. Collagenase stock at 100 mg/mL should be stored on ice for the duration of cell isolation.

2. Pre-weigh an empty 30 mm dish. This will be used to measure the mass of tissue prior to digestion.

3. Prepare solutions for digest.

3.1 Prepare Inactivation Solution (15 mL DPBS -/- with 10% FBS). Keep solution on ice for duration of experiment.

3.2Aliquot 15 mL DPBS +/+ for digest solution. It is crucial that DPBS +/+ be used, as the magnesium and calcium aid in enzymatic digestion. Keep in a bead bath for duration of experiment to keep solution warm.

3.3 Prepare red blood cell (RBC) lysis buffer. Dilute 1 mL RBC lysis buffer stock into 9 mL sterile DI water. Keep solution warm in bead bath for duration of experiment.

Tissue Cutting and Digest

4. Starting with 1 mm thick pieces of ovarian cortex, cut all tissue into ~1mm3 pieces either a) mechanically using a #10 blade in a glass petri dish or b) using a tissue chopper.

5. After cutting, move tissue to a cell strainer (any pore size) and rinse tissue with DPBS -/- twice to wash off any media or previous solutions.

6. Using a spatula, transfer the cut tissue to the pre-weighed 30 mm petri dish. Weigh tissue in pre-weighed dish to obtain tissue mass. Add 2 mL DPBS +/+ to the dish, then add 20 µL Collagenase IA (stock = 100 mg/mL) and 20 µL diluted DNase I for a solution of 1 mg/mL Collagenase IA and 0.02 mg/mL DNase I in DPBS +/+.

7. Place 30 mm dish with tissue in a secondary container and tape down if necessary. Tape dish to shaker and digest for 2 hours at 150 rpm and 37ºC. Tissue pieces should be freely moving in the digest solution to obtain proper enzymatic digestion and leverage mechanical disruption for cell isolation.

8. After 2 hours, strain tissue and supernatant through a 70 µm strainer over a 60 mm dish. A 30 µm strainer can also be used to “catch” follicles with diameters ranging from 30-70 µm in the top

of the 30 μm cell strainer—backwash the 30 μm strainer with DPBS -/- to collect these follicles. To collect follicles with diameters > 70 μm, you may consider implementing a larger cell strainer (150 μm) and backwashing the 70 μm cell strainer, however this will also yield 70-150 μm tissue fragments remaining from the digest.

9. Add 2 mL ice-cold Inactivation Solution to the 60 mm dish of backwashed follicles to halt enzyme activity. Keep follicles in solution on ice and proceed with imaging of follicles or encapsulation in hydrogels for culture.

10. Optional RBC Lysis step: If tissue is bloody prior to digest, consider spinning down cell suspension for 5 minutes at 100g at the end of the digest and resuspending in 1X RBC lysis buffer, then incubating at room temperature for 2 minutes to lyse red blood cells.

Comments and Reflections

This protocol was developed through a series of empirical modifications to the protocol developed by Wagner et al.[9] and is optimized for follicle viability, at the cost of damage to non-follicular cells. Adherence to the protocol as written, paired with careful preparation and organization of the workspace, yields reproducible results, but failure to prepare for the procedure properly will yield subpar follicle viability due to the sensitivity of the follicle suspension to temperature—there is no room for retrieving extra dishes, mixing solutions mid-procedure, etc. Future improvements to this protocol could focus on enzymatic digest re-formulations that provide even higher follicle viability or minimize the time required to obtain isolated follicles. This protocol may serve as a basis for future work to obtain isolated follicles for culture in a clinical setting.

Bibliography

1. Hashim, D. *et al.* The global decrease in cancer mortality: Trends and disparities. *Annals of Oncology* **27**, 926–933 (2016).
2. Kinahan, K. E. *et al.* Adult Survivors of Childhood Cancer and Their Parents. *J Pediatr Hematol Oncol* **30**, 651–658 (2008).
3. Jeruss, J. S. & Woodruff, T. K. Preservation of Fertility in Patients with Cancer. *New England Journal of Medicine* **360**, 902–911 (2009).
4. Oktay, K., Bedoschi, G., Pacheco, F., Turan, V. & Emirdar, V. First pregnancies, live birth, and in vitro fertilization outcomes after transplantation of frozen-banked ovarian tissue with a human extracellular matrix scaffold using robot-assisted minimally invasive surgery. *Am J Obstet Gynecol* **214**, 94.e1-94.e9 (2016).
5. Jones, A. S. K. & Shikanov, A. Follicle development as an orchestrated signaling network in a 3D organoid. *J Biol Eng* **13**, 2 (2019).
6. Wigglesworth, K., Lee, K.-B., Emori, C., Sugiura, K. & Eppig, J. J. Transcriptomic diversification of developing cumulus and mural granulosa cells in mouse ovarian follicles. *Biol Reprod* **92**, 23, 1–14 (2014).
7. Knight, P. G. & Glister, C. TGF-β superfamily members and ovarian follicle development. *Reproduction* **132**, 191–206 (2006).
8. Fan, X. *et al.* Single-cell reconstruction of follicular remodeling in the human adult ovary. *Nat Commun* **10**, 3164 (2019).

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

10. Kim, C.-H. *et al.* The effect of fibroblast co-culture on in vitro maturation of mouse preantral follicles. *Development & Reproduciton* **17**, 269–274 (2013).

11. Tingen, C. M. *et al.* A macrophage and theca cell-enriched stromal cell population influences growth and survival of immature murine follicles in vitro. *Reproduction* **141**, 809–820 (2011).

12. Gupta, P. S. P., Ramesh, H. S., Manjunatha, B. M., Nandi, S. & Ravindra, J. P. Production of buffalo embryos using oocytes from in vitro grown preantral follicles. *Zygote* **16**, 57–63 (2008).

13. Tagler, D. *et al.* Promoting extracellular matrix remodeling via ascorbic acid enhances the survival of primary ovarian follicles encapsulated in alginate hydrogels. *Biotechnol Bioeng* **111**, 1417–1429 (2014).

14. Hornick, J. E., Duncan, F. E., Shea, L. D. & Woodruff, T. K. Multiple follicle culture supports primary follicle growth through paracrine-acting signals. *Reproduction* **145**, 19–32 (2013).

15. Zhou, H. *et al.* Synergy of paracrine signaling during early-stage mouse ovarian follicle development in vitro. *Cell Mol Bioeng* **11**, 435–450 (2018).

16. Pan, H., O'Brien, M. J., Wigglesworth, K., Eppig, J. J. & Schultz, R. M. Transcript profiling during mouse oocyte development and the effect of gonadotropin priming and development in vitro. *Dev Biol* **286**, 493–506 (2005).

17. Skory, R. M. *et al.* Microarray analysis identifies COMP as the most differentially regulated transcript throughout in vitro follicle growth. *Mol Reprod Dev* **80**, 132–144 (2013).

18. Green, L. J., Zhou, H., Padmanabhan, V. & Shikanov, A. Adipose-derived stem cells promote survival, growth, and maturation of early-stage murine follicles. *Stem Cell Res Ther* **10**, 102 (2019).

19. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* Preprint at https://www.r-project.org/ (2019).

20. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).

21. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

22. Gentleman, R, Carey, V, Huber, W, and Hahne, F. genefilter: Methods for filtering genes from high-throughput experiments. Preprint at (2019).

23. Pages, H, Carlson, M, Falcon, S, and Li, N. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. Preprint at (2019).

24. Wickham, H. stringr: Simple, Consistent Wrappers for Common String Operations. Preprint at https://cran.r-project.org/package=stringr (2019).

25. Wickham, H, Francois, R, Henry, L, and Muller, K. dplyr: A Grammar of Data Manipulation. Preprint at https://cran.r-project.org/package=dplyr (2019).

26. Kolde, R. pheatmap: Pretty Heatmaps. Preprint at https://cran.r-project.org/package=pheatmap (2019).

27. Kim, J. H. *et al.* LRpath analysis reveals common pathways dysregulated via DNA methylation across cancer types. *BMC Genomics* **13**, 1 (2012).

28. Lee, C., Patil, S. & Sartor, M. A. RNA-Enrich: A cut-off free functional enrichment testing method for RNA-seq with improved detection power. *Bioinformatics* (2016) doi:10.1093/bioinformatics/btv694.

29. Newton, M. A., Quintana, F. A., den Boon, J. A., Sengupta, S. & Ahlquist, P. Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann Appl Stat* (2007) doi:10.1214/07-aoas104.

30. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).

31. Sean, D. & Meltzer, P. S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).

32. Yuan, J. som: Self-Organizing Map. Preprint at https://cran.r-project.org/package=som (2016).

33. Peñalver Bernabé, B. *et al.* Dynamic genome-scale cell-specific metabolic models reveal novel inter-cellular and intra-cellular metabolic communications during ovarian follicle development. *BMC Bioinformatics* **20**, 307 (2019).

34. Bolstad, B. preprocessCore: A collection of pre-processing functions. R package version 1.46.0 Preprint at https://github.com/bmbolstad/preprocessCore (2019).

35. Nilsson, E. E., Schindler, R., Savenkova, M. I. & Skinner, M. K. Inhibitory actions of Anti-Müllerian Hormone (AMH) on ovarian primordial follicle assembly. *PLoS One* **6**, (2011).

36. Mazerbourg, S. & Monget, P. Insulin-like growth factor binding proteins and IGFBP proteases: A dynamic system regulating the ovarian folliculogenesis. *Frontiers in Endocrinology* vol. 9 1 Preprint at https://doi.org/10.3389/fendo.2018.00134 (2018).

37. Hatzirodos, N., Irving-Rodgers, H. F., Hummitzsch, K. & Rodgers, R. J. Transcriptome profiling of the theca interna from bovine ovarian follicles during atresia. *PLoS One* **9**, 99706 (2014).

38. Lim, E. J. & Choi, Y. Transcription factors in the maintenance and survival of primordial follicles. *Clinical and Experimental Reproductive Medicine* vol. 39 127–131 Preprint at https://doi.org/10.5653/cerm.2012.39.4.127 (2012).

39. Thompson, J. A new window into ovarian follicle development. *Biology of Reproduction* vol. 95 1–2 Preprint at https://doi.org/10.1095/biolreprod.116.146647 (2016).

40. Del Peso, L. *et al.* The von Hippel Lindau/hypoxia-inducible factor (HIF) pathway regulates the transcription of the HIF-proline hydroxylase genes in response to low oxygen. *Journal of Biological Chemistry* **278**, 48690–48695 (2003).

41. Liu, X. *et al.* Paracrine Regulation of Steroidogenesis in Theca Cells by Granulosa Cells Derived from Mouse Preantral Follicles. *Biomed Res Int* **2015**, 1–8 (2015).

42. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res* **41**, D991–D995 (2013).

43. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* (2002) doi:10.1093/nar/30.1.207.

44. Heinken, A., Sahoo, S., Fleming, R. M. T. & Thiele, I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* **4**, 28–40 (2013).

45. Carlsson, I. B. Anti-Mullerian hormone inhibits initiation of growth of human primordial ovarian follicles in vitro. *Human Reproduction* **21**, 2223–2227 (2006).

46. Elnashar, A. M. The role of Anti-Mullerian hormone in assisted reproduction. *Middle East Fertil Soc J* **17**, 160–162 (2012).

47. Xu, J., Bishop, C. V., Lawson, M. S., Park, B. S. & Xu, F. Anti-Müllerian hormone promotes pre-antral follicle growth, but inhibits antral follicle maturation and dominant follicle selection in primates. *Human Reproduction* **31**, 1522–1530 (2016).

48. Prasasya, R. D. & Mayo, K. E. Regulation of Follicle Formation and Development by Ovarian Signaling Pathways. in *The Ovary* 23–49 (Elsevier, 2019). doi:10.1016/B978-0-12-813209-8.00002-9.

49. Hoang, Y. D., McTavish, K. J., Chang, R. J. & Shimasaki, S. Paracrine regulation of theca androgen production by granulosa cells in the ovary. *Fertil Steril* **100**, 561–567 (2013).

50. Sanfins, A., Rodrigues, P. & Albertini, D. F. GDF-9 and BMP-15 direct the follicle symphony. *J Assist Reprod Genet* **35**, 1741–1750 (2018).

51. Xiao, S. *et al.* A microfluidic culture model of the human reproductive tract and 28-day menstrual cycle. *Nat Commun* **8**, (2017).

52. Winkler-Crepaz, K. *et al.* Novel dynamic culture system to support initiation of primordial follicle growth in prepubertal mouse ovaries. *Fertil Steril* **102**, (2014).

53. Araújo, V. R., Duarte, A. B. G., Bruno, J. B., Pinho Lopes, C. A. & de Figueiredo, J. R. Importance of vascular endothelial growth factor (VEGF) in ovarian physiology of mammals. *Zygote* **21**, 295–304 (2013).

54. Gonçalves, P. B., Ferreira, R., Gasperin, B. & Oliveira, J. F. Role of angiotensin in ovarian follicular development and ovulation in mammals: A review of recent advances. *Reproduction* **143**, 11–20 (2012).

55. Lai, W. A. *et al.* Calcineurin and CRTC2 mediate FSH and TGFβ1 upregulation of Cyp19a1 and Nr5a in ovary granulosa cells. *J Mol Endocrinol* **53**, 259–270 (2014).

56. Emori, C. & Sugiura, K. Role of oocyte-derived paracrine factors in follicular development. *Animal Science Journal* **85**, 627–633 (2014).

57. Castilla, A. *et al.* Prolactin in Ovarian Follicular Fluid Stimulates Endothelial Cell Proliferation. *J Vasc Res* **47**, 45–53 (2009).

58. Blanchette Porter, M., Brumsted, J. R. & Sites, C. K. Effect of prolactin on follicle-stimulating hormone receptor binding and progesterone production in cultured porcine granulosa cells. *Fertil Steril* **73**, 99–105 (2000).

59. Buccione, R., Schroeder, A. C. & Eppig, J. J. Interactions between somatic cells and germ cells throughout mammalian oogenesis. *Biol Reprod* **43**, 543–547 (1990).

60. Sirotkin, A. V. Growth factors controlling ovarian functions. *J Cell Physiol* **226**, 2222–2225 (2011).

61. Field, S. L., Dasgupta, T., Cummings, M. & Orsi, N. M. Cytokines in ovarian folliculogenesis, oocyte maturation and luteinisation. *Mol Reprod Dev* **81**, 284–314 (2014).

62. Kehoe, S., Jewgenow, K., Johnston, P. R., Mbedi, S. & Braun, B. C. Signalling pathways and mechanistic cues highlighted by transcriptomic analysis of primordial, primary, and secondary ovarian follicles in domestic cat. *Sci Rep* **11**, 1–17 (2021).

63. Jones, A., Bernabé, B. P., Padmanabhan, V., Li, J. & Shikanov, A. Capitalizing on transcriptome profiling to optimize and identify targets for promoting early murine folliculogenesis in vitro. *Sci Rep* **11**, 1–14 (2021).

64. Thomas, P. D. *et al.* PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res* **13**, 2129–2141 (2003).

65. Shao, T. *et al.* Autophagy regulates differentiation of ovarian granulosa cells through degradation of WT1. (2022) doi:10.1080/15548627.2021.2005415.

66. Hernandez Gifford, J. A. The role of WNT signaling in adult ovarian folliculogenesis. *Reproduction* **150**, E137-R148 (2015).

67. Kidder, G. M. & Vanderhyden, B. C. Bidirectional communication between oocytes and follicle cells: ensuring oocyte developmental competence. *Can J Physiol Pharmacol* **88**, 399–413 (2010).

68. Russell, D. L., Gilchrist, R. B., Brown, H. M. & Thompson, J. G. Bidirectional communication between cumulus cells and the oocyte: Old hands and new players? *Theriogenology* **86**, 62–68 (2016).

69. Wigglesworth, K. *et al.* Bidirectional communication between oocytes and ovarian follicular somatic cells is required for meiotic arrest of mammalian oocytes. *Proceedings of the National Academy of Sciences* **110**, E3723–E3729 (2013).

70. Yeo, C. X., Gilchrist, R. B. & Lane, M. Disruption of Bidirectional Oocyte-Cumulus Paracrine Signaling During In Vitro Maturation Reduces Subsequent Mouse Oocyte Developmental Competence1. *Biol Reprod* **80**, 1072–1080 (2009).

71. Hazzard, T. M. & Stouffer, R. L. Angiogenesis in ovarian follicular and luteal development. *Best Practice and Research: Clinical Obstetrics and Gynaecology* vol. 14 883–900 Preprint at https://doi.org/10.1053/beog.2000.0133 (2000).

72. Shin, S.-Y., Lee, H.-J., Ko, D.-S., Lee, H.-C. & Park, W. Il. The regulators of VEGF expression in mouse ovaries. *Yonsei Med J* **46**, 679 (2005).

73. Ryan, K. J. & Makris, A. Significance of Angiogenic and Growth Factors in Ovarian Follicular Development. in *Regulation of Ovarian and Testicular Function* 203–209 (1987). doi:10.1007/978-1-4684-5395-9_10.

74. Man, L. *et al.* Engineered endothelium provides angiogenic and paracrine stimulus to grafted human ovarian tissue. *Sci Rep* **7**, 8203 (2017).

75. Morris, M. E. *et al.* A single-cell atlas of the cycling murine ovary. *Elife* **11**, (2022).

76. Dunning, K. R., Anastasi, M. R., Zhang, V. J., Russell, D. L. & Robker, R. L. Regulation of fatty acid oxidation in mouse cumulus-oocyte complexes during maturation and modulation by PPAR agonists. *PLoS One* **9**, (2014).

77. De Felici, M. & Klinger, F. G. Pi3k/pten/akt signaling pathways in germ cell development and their involvement in germ cell tumors and ovarian dysfunctions. *International Journal of Molecular Sciences* vol. 22 Preprint at https://doi.org/10.3390/ijms22189838 (2021).

78. Miranda, P. V, Gonzá Lez-Echeverría, F., Blaquier, J. A., Mahuran, D. J. & Tezó, J. G. *Evidence for the participation of β-hexosaminidase in human sperm-zona pellucida interaction in vitro*. *Molecular Human Reproduction* vol. 6 (2000).

79. Agirregoitia, E. *et al.* Expression and localization of opioid receptors during the maturation of human oocytes. *Reprod Biomed Online* **24**, 550–557 (2012).

80. Südhof, T. C. *Neuroligins and Neurexins Link Synaptic Function to Cognitive Disease*. vol. 455 (2008).

81. Al-Alem, L. *et al.* Neurotensin: A neuropeptide induced by hCG in the human and rat ovary during the periovulatory period. *Biol Reprod* **104**, 1337–1346 (2021).

82. Aoyama, M. *et al.* A novel biological role of tachykinins as an up-regulator of oocyte growth: Identification of an evolutionary origin of tachykininergic functions in the ovary of the ascidian, Ciona intestinalis. *Endocrinology* **149**, 4346–4356 (2008).

83. Han, Z., Guo, L., Li, H. & Sun, Q.-Y. Effect of Parathyroid Hormone-Like Hormone on Mouse Oocyte Maturation. *Biol Reprod* **87**, 315 (2012).

84. Candelaria, J. I., Rabaglino, M. B. & Denicol, A. C. Ovarian preantral follicles are responsive to FSH as early as the primary stage of development. *Journal of Endocrinology* **247**, 153–168 (2020).

85. Jankovičová, J., Sečová, P., Michalková, K. & Antalíková, J. Tetraspanins, more than markers of extracellular vesicles in reproduction. *International Journal of Molecular Sciences* vol. 21 1–30 Preprint at https://doi.org/10.3390/ijms21207568 (2020).

86. Tesfaye, D. *et al.* Extracellular vesicle mediated molecular signaling in ovarian follicle: Implication for oocyte developmental competence. *Theriogenology* **150**, 70–74 (2020).

87. Tesfaye, D. *et al.* Extracellular vesicle mediated molecular signaling in ovarian follicle: Implication for oocyte developmental competence. *Theriogenology* (2020) doi:10.1016/j.theriogenology.2020.01.075.

88. Ouni, E., Vertommen, D., Chiti, M. C., Dolmans, M.-M. & Amorim, C. A. A Draft Map of the Human Ovarian Proteome for Tissue Engineering and Clinical Applications. *Molecular & Cellular Proteomics* **18**, S159–S173 (2019).

89. Hovatta, O., Wright, C., Krausz, T., Hardy, K. & Winston, R. M. L. Human primordial, primary and secondary ovarian follicles in long-term culture: Effect of partial isolation. *Human Reproduction* **14**, 2519–2524 (1999).

90. Telfer, E. E., McLaughlin, M., Ding, C. & Thong, K. J. A two-step serum-free culture system supports development of human oocytes from primordial follicles in the presence of activin. *Human Reproduction* **23**, 1151–1158 (2008).

91. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* vol. 9 671–675 Preprint at https://doi.org/10.1038/nmeth.2089 (2012).

92. Yurttas, P. *et al.* Role for PADI6 and the cytoplasmic lattices in ribosomal storage in oocytes and translational control in the early mouse embryo. *Development* **135**, 2627–2636 (2008).

93. Mtango, N. R. *et al.* Essential role of maternal UCHL1 and UCHL3 in fertilization and preimplantation embryo development. *J Cell Physiol* **227**, 1592–1603 (2012).

94. Kumar, R. *et al.* Mouse REC114 is essential for meiotic DNA double-strand break formation and forms a complex with MEI4. *Life Sci Alliance* **1**, (2018).

95. Llonch, S. *et al.* Single human oocyte transcriptome analysis reveals distinct maturation stage-dependent pathways impacted by age. *Aging Cell* **20**, (2021).

96. Asiabi, P., Dolmans, M. M., Ambroise, J., Camboni, A. & Amorim, C. A. In vitro differentiation of theca cells from ovarian cells isolated from postmenopausal women . *Human Reproduction* **35**, 2793–2807 (2020).

97. Richards, J. S., Ren, Y. A., Candelaria, N., Adams, J. E. & Rajkovic, A. Ovarian Follicular Theca Cell Recruitment, Differentiation, and Impact on Fertility: 2017 Update. *Endocr Rev* **39**, 1–20 (2018).

98. Guahmich, N. L. *et al.* Human theca arises from ovarian stroma and is comprised of three discrete subtypes. *Communications Biology 2023 6:1* **6**, 1–15 (2023).

99. Ridinger, K., Schä Fer ‡, B. W., Durussel, I., Cox, J. A. & Heizmann, C. W. S100A13 BIOCHEMICAL CHARACTERIZATION AND SUBCELLULAR LOCALIZATION IN DIFFERENT CELL LINES*. *Journal of Biological Chemistry* **275**, 8686–8694 (2000).

100. Harris, R. A., McAllister, J. M. & Strauss, J. F. Single-Cell RNA-Seq Identifies Pathways and Genes Contributing to the Hyperandrogenemia Associated with Polycystic Ovary Syndrome. *Int J Mol Sci* **24**, 10611 (2023).

101. Yazawa, T. *et al.* Transcriptional Regulation of Ovarian Steroidogenic Genes: Recent Findings Obtained from Stem Cell-Derived Steroidogenic Cells. *BioMed Research International* vol. 2019 Preprint at https://doi.org/10.1155/2019/8973076 (2019).

102. Ansell, P. J. *et al. Regulation of Growth Hormone Expression by Delta-Like Protein 1 (Dlk1)*.

103. Liu, T., Qin, Q. Y., Qu, J. X., Wang, H. Y. & Yan, J. Where are the theca cells from: the mechanism of theca cells derivation and differentiation. *Chinese Medical Journal* vol. 133 1711–1718 Preprint at https://doi.org/10.1097/CM9.0000000000000850 (2020).

104. Jin, J. *et al.* Ovulatory signal-triggered chromatin remodeling in ovarian granulosa cells by HDAC2 phosphorylation activation-mediated histone deacetylation. *Epigenetics Chromatin* **16**, (2023).

105. Madogwe, E. *et al.* Global analysis of FSH-regulated gene expression and histone modification in mouse granulosa cells. *Mol Reprod Dev* **87**, 1082–1096 (2020).

106. Kawai, T., Richards, J. A. S. & Shimada, M. Large-scale DNA demethylation occurs in proliferating ovarian granulosa cells during mouse follicular development. *Commun Biol* **4**, (2021).

107. Demanno, D. A. *et al. Follicle-Stimulating Hormone Promotes Histone H3 Phosphorylation on Serine-10*. https://academic.oup.com/mend/article/13/1/91/2741657 (1999).

108. Alberico, H. C. & Woods, D. C. Role of Granulosa Cells in the Aging Ovarian Landscape: A Focus on Mitochondrial and Metabolic Function. *Frontiers in Physiology* vol. 12 Preprint at https://doi.org/10.3389/fphys.2021.800739 (2022).

109. Papler, T. B., Bokal, E. V., Maver, A., Kopitar, A. N. & Lovrečić, L. Transcriptomic analysis and meta-analysis of human granulosa and cumulus cells. *PLoS One* **10**, (2015).

110. Hatzirodos, N. *et al. Transcriptome profiling of granulosa cells of bovine ovarian follicles during growth from small to large antral sizes*. http://www.biomedcentral.com/1471-2164/15/24 (2014).

111. Hatzirodos, N., Hummitzsch, K., Irving-Rodgers, H. F. & Rodgers, R. J. Transcriptome comparisons identify new cell markers for theca interna and granulosa cells from small and large antral ovarian follicles. *PLoS One* **10**, (2015).

112. Abedini, A. *et al.* WNT5a is required for normal ovarian follicle development and antagonizes gonadotropin responsiveness in granulosa cells by suppressing canonical WNT signaling. *FASEB Journal* **30**, 1534–1547 (2016).

113. Puri, P. *et al.* Protein Kinase A: A Master Kinase of Granulosa Cell Differentiation. *Sci Rep* **6**, (2016).

114. Ouni, E. *et al.* Spatiotemporal changes in mechanical matrisome components of the human ovary from prepuberty to menopause. *Human Reproduction* **35**, 1391–1410 (2020).

115. Bagalad, B., Mohan Kumar, K. & Puneeth, H. Myofibroblasts: Master of disguise. *Journal of Oral and Maxillofacial Pathology* vol. 21 462–463 Preprint at https://doi.org/10.4103/jomfp.JOMFP_146_15 (2017).

116. Hewitt, S. C. & Korach, K. S. Estrogen Receptors: New Directions in the New Millennium. *Endocr Rev* **39**, 664–675 (2018).

117. Zhou, H., Pon, Y. & Wong, A. HGF/MET signaling in ovarian cancer. *Curr Mol Med* **8**, 469–480 (2008).

118. Chen, F. *et al.* ARHGEF15 in Sertoli cells contributes to germ cell development and testicular immune privilege†. *Biol Reprod* **107**, 1565–1579 (2022).

119. Naba, A. *et al.* The matrisome: In silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular and Cellular Proteomics* **11**, (2012).

120. Ni, Y. *et al.* A Novel pro-adipogenesis factor abundant in adipose tissues and over-expressed in obesity acts upstream of PPARγ and C/EBPα. *J Bioenerg Biomembr* **45**, 219–228 (2013).

121. Espey, L. L. *et al.* Induction of early growth response protein-1 gene expression in the rat ovary in response to an ovulatory dose of human chorionic gonadotropin. *Endocrinology* **141**, 2385–2391 (2000).

122. Lin, I. Y. *et al.* Suppression of the SOX2 neural effector gene by PRDM1 promotes human germ cell fate in embryonic stem cells. *Stem Cell Reports* **2**, 189–204 (2014).

123. Wang, H. *et al.* Induction of meiosis by embryonic gonadal somatic cells differentiated from pluripotent stem cells. *Stem Cell Res Ther* **12**, 607 (2021).

124. Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).

125. Li, L. *et al.* Wnt/β-catenin signaling regulates follicular development by modulating the expression of Foxo3a signaling components. *Mol Cell Endocrinol* **382**, 915–925 (2014).

126. Dentis, J. L., Schreiber, N. B., Burress, A. M. & Spicer, L. J. Effects of angiogenin on granulosa and theca cell function in cattle. doi:10.1017/S1751731116002044.

127. Robinson, R. S. *et al.* Angiogenesis and vascular function in the ovary. *Reproduction* **138**, 869–881 (2009).

128. Tomaszewski, C. E., DiLillo, K. M., Baker, B. M., Arnold, K. B. & Shikanov, A. Sequestered cell-secreted extracellular matrix proteins improve murine folliculogenesis and oocyte maturation for fertility preservation. *Acta Biomater* **132**, 313–324 (2021).

129. Offspring from Oocytes Derived from in Vitro Primordial Germ Cell–like Cells in Mice.

130. Babayev, E., Xu, M., Shea, L. D., Woodruff, T. K. & Duncan, F. E. Follicle isolation methods reveal plasticity of granulosa cell steroidogenic capacity during mouse in vitro follicle growth. *Mol Hum Reprod* (2022) doi:10.1093/MOLEHR/GAAC033.

131. Dzafic, E., Stimpfel, M. & Virant-Klun, I. Plasticity of granulosa cells: on the crossroad of stemness and transdifferentiation potential. doi:10.1007/s10815-013-0068-0.

132. Wu, G. M. J., Chen, A. C. H., Yeung, W. S. B. & Lee, Y. L. Current progress on in vitro differentiation of ovarian follicles from pluripotent stem cells. *Front Cell Dev Biol* **11**, (2023).

133. Tomaszewski, C. E. *et al.* Adipose-derived stem cell-secreted factors promote early stage follicle development in a biomimetic matrix. *Biomater Sci* **7**, 571–580 (2019).