

Genome-wide Metabolic Modeling of Human Ovarian Follicle Development

BY

Ashwin Lakshman Koppayi
B.E. University of Mumbai, India, 2018

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Science in Bioengineering
in the Graduate College of the
University of Illinois at Chicago, 2021

Chicago, Illinois

Defense Committee:

Dr. Beatriz Peñalver Bernabé, Chair and Advisor
Dr. Salman Khetani, Department of Biomedical Engineering
Dr. Zhangli Peng, Department of Biomedical Engineering

Dedicated to my family, friends, and my country.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation and a special thanks to my research advisor Dr. Beatriz Peñalver Bernabé, for being a tremendous mentor for me. I thank her for her constant support and encouragement in achieving my goals.

I would also like to extend my heartfelt thanks to my other committee members Dr. Salman Khetani and Dr. Zhangli Peng for their assistance in serving as my committee members.

It would have been impossible without the BeaLab team who provided feedback and motivation to improve my presentation and research. Special thanks to my friends who were always there for me.

Lastly, I would like to express my love and sincere gratitude to my dearest parents Lakshmanan Koppayi and Lakshmi Lakshman and to my dearest brother Akhil and sister in-law Ramya for providing valuable suggestions in completing my thesis successfully. I am grateful for their utmost support and encouragement throughout my Master's program and for always being there for me.

TABLE OF CONTENTS

CHAPTER 1	1
INTRODUCTION	1
1.1. BACKGROUND	1
1.2. OVARIAN FOLLICLE DEVELOPMENT IN HUMANS	3
1.3. METABOLISM DURING OVARIAN FOLLICLE DEVELOPMENT	6
1.4. GENOME-SCALE METABOLIC MODEL OF HUMAN	8
1.5. CONTEXT SPECIFIC METABOLIC MODEL	11
CHAPTER 2	13
METHODS	13
2.1 IDENTIFICATION OF METABOLIC ENZYMES PRESENT IN THE DISTINCT OVARIAN FOLLICLE CELLS DURING OVARIAN FOLLICLE DEVELOPMENT USING SINGLE-CELL TRANSCRIPTOMIC	13
2.2 GENERATION OF CONTEXT-SPECIFIC GENOME-WIDE METABOLIC MODELS USING OVARIAN FOLLICLE DEVELOPMENT IN HUMANS USING FASTCORE	16
2.3 STAGE-SPECIFIC MODEL OF OOCYTE AND GRANULOSA CELLS	17
2.4 COMMUNITY DETECTION OF FOLLICLE MODEL	18
2.5 SIGNIFICANT METABOLITES, GENES, AND METABOLIC PATHWAYS DURING OVARIAN FOLLICLE DEVELOPMENT	19
CHAPTER 3	20
RESULT AND DISCUSSION	20
3.1 OVARIAN FOLLICLE TRANSCRIPTION ACTIVITY	20
3.1.1 <i>Highly variable genes during follicle development</i>	20
3.1.2 <i>Principal Component Analysis</i>	23
3.1.3 <i>Differentially expressed genes during follicle development</i>	25
3.2 OVARIAN FOLLICLE MODEL.....	29
3.3 BI-PARTITE GRAPH OF ENZYME AND METABOLITES	30
3.4 METABOLITES PRESENT IN STAGE SPECIFIC MODELS.....	31
3.5 TOP METABOLITES BASED ON METABOLITE FLOW INTENSITY	33
3.6 TOP ENZYMES BASED ON THE ENZYME FLOW INTENSITY	35
3.7 ENRICHED PATHWAY IN STAGE SPECIFIC MODEL	37
CHAPTER 4	39
CONCLUSION	39
4.1 FUTURE PERSPECTIVES	40
APPENDIX	47
VITA	48

LIST OF TABLES

TABLE		PAGE
1	Human Metabolic model properties	10
2	Reason for Ovariectomy in donors	13
3	Donor's biodata and number of follicle isolated	14
4	Differentially expressed genes during follicle development in Oocyte	26
5	Differentially expressed genes during follicle development in Granulosa	28

LIST OF FIGURES

FIGURE	PAGE
1 Hypothalamic -Pituitary -Ovarian Axis that regulates follicle development.	4
2 Human Ovarian Follicle Development Stages	5
3 Pipeline to process single cell RNA sequencing data	15
4 Generation of Context specific model using FASTCORE	16
5 Heatmap of the 100 most variable genes across the Oocyte Samples	20
6 Heatmap of variable genes across the samples in Granulosa Sample	21
7 Principal Component Analysis of Samples based on A) Cell Type B) Stage of follicle development	23
8 Expression of markers in the samples: A) Oocyte specific markers B) Granulosa specific markers	24
9 Differential gene expression between the subsequent stages of follicle development in oocyte cells	26
10 Differential gene expression between the subsequent stages of follicle development in granulosa cells	27
11 Enzyme-Metabolite Bipartite network of Recon3D and Human Ovarian Follicle Model.	30
12 Venn diagram of Metabolites obtained from at different stages in Oocyte Models.	31
13 Venn diagram of Metabolites obtained from at different stages in Granulosa Models.	32
14 Top Metabolites in stage specific models based on Z score of normalized flow intensity through the metabolites.	33
15 Top Enzymes in stage specific models based on Z score of normalized flow intensity through the enzymes.	35
16 Metabolic Pathways in each stage specific models based on Z score of Normalized flow intensity through a pathway	38

LIST OF ABBREVIATIONS

GnRH: Gonadotrophin Releasing Hormone

FSH: Follicle Stimulating Hormone

FF: Follicular Fluid

LH: Luteinizing Hormone

PCOS: Polycystic Ovarian Syndrome

IVM: In vitro Maturation

IVF: In vitro Fertilization

GEMS: Genome-scale Metabolic Models

FBA: Flux Balance Analysis

LP: Linear Programming

scRNA seq: Single Cell RNA Sequencing

UP1: Universal Primer 1

UP2: Universal Primer 2

polyA: Poly Adenylated Tail.

OPM: Oocyte at Primordial Stage of Follicle Development

OPR: Oocyte at Primary Stage of Follicle Development

OSC: Oocyte at Secondary Stage of Follicle Development

OAT: Oocyte at Antral Stage of Follicle Development

OPO: Oocyte at Preovulatory Stage of Follicle Development

SPM: Granulosa at Primordial Stage of Follicle Development

SPR: Granulosa at Primary Stage of Follicle Development

SSC: Granulosa at Secondary Stage of Follicle Development

SAT: Granulosa at Antral Stage of Follicle Development

SPO: Granulosa at Preovulatory Stage of Follicle Development

c -Cytoplasm

m -Mitochondria

i -Inner Mitochondria

e -Extracellular space

g -Golgi apparatus

n -Nucleus

l -Lysozyme

r -Endoplasmic Reticulum

x-Peroxisome

PI3K-Phosphoinositide-Kinase

Akt-Serine Threonine -Protein Kinase

FOXO3- Forkhead Box O3

NOBOX- Newborn Ovarian Homeobox

PTEN- Phosphate and Tensin Homolog

SOHLH-Spermatogenesis and oogenesis helix loop helix family

LDHA-Lactate Dehydrogenase A

GADPH- Glyceraldehyde-3-Phosphate Dehydrogenase

ESR-Estrogen Receptor

PGR-Progesterone Receptor

HSD -3-beta (β)-hydroxysteroid dehydrogenase

ZP -Zona Pellucida

MT-Mitochondrial

SUMMARY

Ovarian follicle development involves intracellular and intercellular metabolic communication between the female germ cells, oocytes, and the neighboring somatic cells, granulosa, and theca cells. The development competence of the oocyte that can render a successful pregnancy depends on the accumulation of the required metabolites and other materials (e.g., proteins, transcripts) essential for later stages, such as fertilization and subsequent embryo pre-implantation. The complex dynamic and bi-directional communication between the oocyte and granulosa cells to achieve a competent oocyte warrants the use of systems biology approaches. Here we created a context specific model of the human ovarian follicle by overlaying human single-cell transcriptomic data from different stages of follicle development to the latest genome-scale human metabolic model (Recon3D) using FASTCORE. Analysis of single cell RNA sequencing data revealed that there were 18,741 actively transcribed genes in the oocyte samples and 17,092 actively transcribed genes in granulosa samples at the different stages of follicle development. Using an unsupervised method, we identified five different clusters that were indeed associated with the different follicular stages (i.e., primordial, primary, secondary, antral, and pre-ovulatory) and cell-type (oocyte and granulosa cells). The Follicle metabolic model contains 10,538 reactions, 3,484 metabolites and 2,954 genes that encode enzymes. As expected, the follicle metabolic model included several key follicle metabolic pathways during follicle development *in vivo* (such as, pyruvate metabolism, steroid metabolism, androgen and estrogen synthesis and metabolism). In summary, we have developed a human follicle metabolic model for the first time that could help to better understand the biology of human ovarian follicle development and thus serve to develop novel treatments for reproductive diseases in the future, such as polycystic ovarian syndrome (PCOS), endometriosis, or improve *in vitro* fertilization.

Chapter 1

Introduction

1.1. Background

Ovary in the female reproductive system can carry out the function of producing a mature female germ cell that is necessary for fertilization. The ovarian tissue is highly organized and comprises different cell types like germ cell and somatic cells [1]. These different cell types interact with each other and form a functional unit called ovarian follicle. A human ovarian follicle has a germ cell (oocyte) at the center, and it is surrounded by somatic cells (granulosa, and theca cells) [1,5]. Within ovarian tissue and during reproductive age, resting and maturing follicles coexist. During each menstrual cycle, several resting follicles are activated by local signaling pathways in response to Follicle Stimulating Hormones (FSH) produced by pituitary glands [23]. Upon binding to its receptor (FSHR), FSH activates several signaling pathways, such as Phosphoinositide 3-kinase-Serine Threonine Protein Kinase -Forkhead Box O3 (PI3K-Akt-FOXO3) that regulated the activation of dormant follicles and deactivates several inhibitory and maintenance signaling pathways like Phosphate and Tensin Homolog (PTEN) and Hippo signaling pathway that plays a role in maintaining the dormancy of the follicle [1,4]. One of the activated follicles becomes dominant and ovulates for subsequent fertilization. Rest of the follicle degenerates and dies (atresia) at different stages of follicle development.

The development competence or quality of the oocyte depends on the synthesis and utilization of metabolites and proteins required for growth, fertilization, and embryo formation. Oocyte and granulosa cells communicate with among themselves and with other cells through juxtacrine communications, via gap junctions between cells, and autocrine and paracrine factors to orchestrate the maturation of the

ovarian follicle. Oocyte and granulosa cell exchange metabolites, and metabolites produced by one cell can be substrates for other ovarian cell types. For instance, the energy coin of the oocyte is pyruvate, which is produced by the surrounding granulosa cells, and androgen synthesized by the theca cells diffuses to the nearby granulosa cells where it is converted to estrogen. This bidirectional metabolic crosstalk between the oocyte with the surrounding somatic cells is still unexplored. Disturbance in this communication can be a result of female infertility such as polycystic ovarian syndrome (PCOS) or premature ovarian failure (POF), chemotherapy-induced infertility [5,18,27]. Women with infertility may choose to undergo some form of Assisted Reproductive Technology like *in vitro* Fertilization (IVF) and *in vitro* Maturation (IVM). IVF involves multiple hormonal injections to retrieve a mature follicle from the ovary, while in IVM only non-maturate follicles (secondary/antral) are collected. These immature late secondary and antral follicles are matured in culture which is supplemented with gonadotropin hormones like Follicle Stimulating Hormone (FSH), Luteinizing Hormone (LH) and other nutrients. While IVM has less side-effects as the hormones are given to the follicles outside the body, IVM success rate is still quite low, with the lack of understanding ovarian follicle development metabolism as one of the leading reasons. Also, the earlier stages of follicle development (primordial and primary) which are most abundant, is incapable for *in vitro* maturation yet. This is due to lack of studies on metabolism in the initial stages of follicle development which is in turn due to lack of availability of ovarian tissue samples. To address these challenges the cell-specific genome-wide metabolic models could be a great tool to increase our understanding of the intercellular and intracellular metabolism during ovarian follicle development and could also generate new data-driven hypotheses for further studies.

1.2. Ovarian Follicle Development in Humans

Ovarian follicle development or folliculogenesis in humans is a complex, dynamic and highly selective process. The ovarian tissue, where the follicles reside, can be generally divided into two regions: - the inner region containing blood vessel and nerves called medulla and the outer peripheral region, which has numerous follicles forming the ovarian reserve, is called cortex.[4]. The ovarian reserve is formed by mitosis of female germ cells while in utero by a process called oogenesis. There are up to 6 million germ cells in the future fetal ovary when the mitosis is completed. These germ cells then enter meiosis and are arrested in the prophase I. While many germ cells die due apoptosis during this stage, the remaining ones become surrounded by pre granulosa cells to form primordial follicles. At birth, it is estimated that there are up to 1 million follicles in the ovarian reserve and does not increase after birth. The number of follicles decreases bi-exponentially by age and approximately 400,000 follicles remain in the ovarian reserve at the time of puberty. During the reproductive age of women, up to 1,000 follicles enter folliculogenesis in each menstrual cycle [10]. In a highly selective process, only one follicle usually in humans, the most responsive to the presence of the follicle stimulated hormone (FSH), becomes the dominant follicle, and will ovulate and the same time that secretes components to trigger atresia in the rest of the activated follicles. [5].

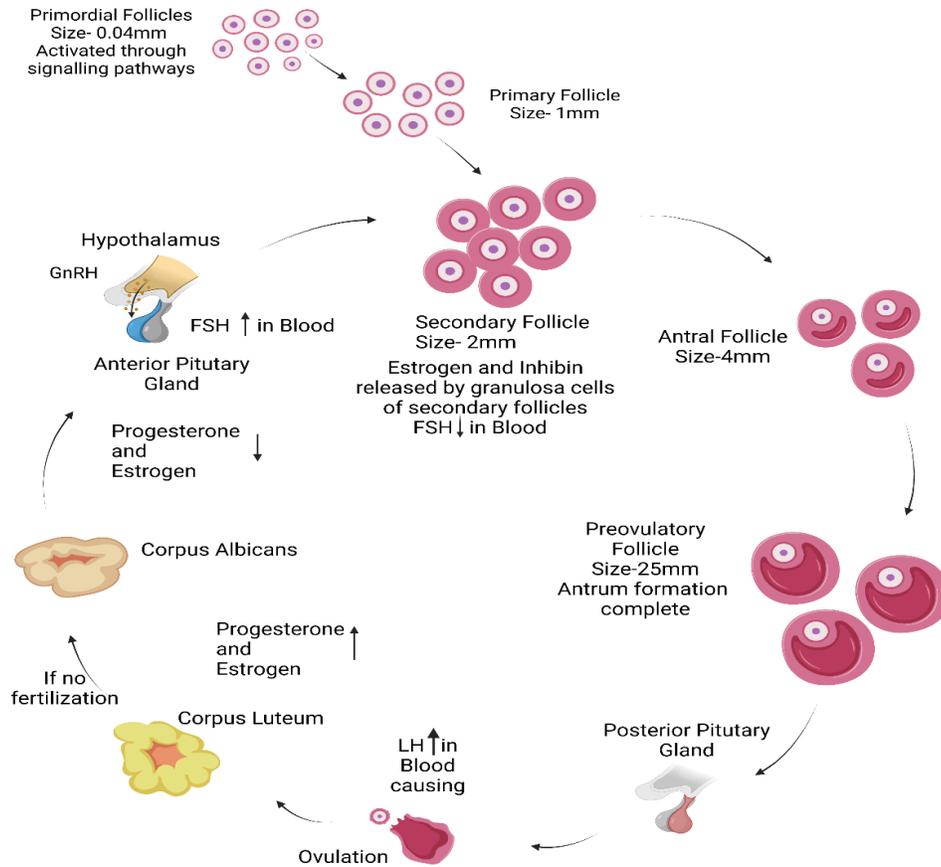


Figure 1: Hypothalamic -Pituitary -Ovarian Axis that regulates follicle development.

Human follicle development is a rhythmic cycle which begins when the hypothalamus of the brain releases Gonadotropin regulating hormone (GnRH) every month. GnRH diffuses to the anterior pituitary gland and binds to 7-transmembrane G-protein receptor and activates secretion of Follicle Stimulating Hormone (FSH) in blood. FSH is transported to the ovary where it binds to its receptor in the follicle (FSHR) and activates the dormant follicle indirectly through a PI3K-Akt-FOXO3, Akt, calcium-dependent signaling pathway. Once the dormant follicle is activated the follicular phase of the ovarian follicle development begins [23].

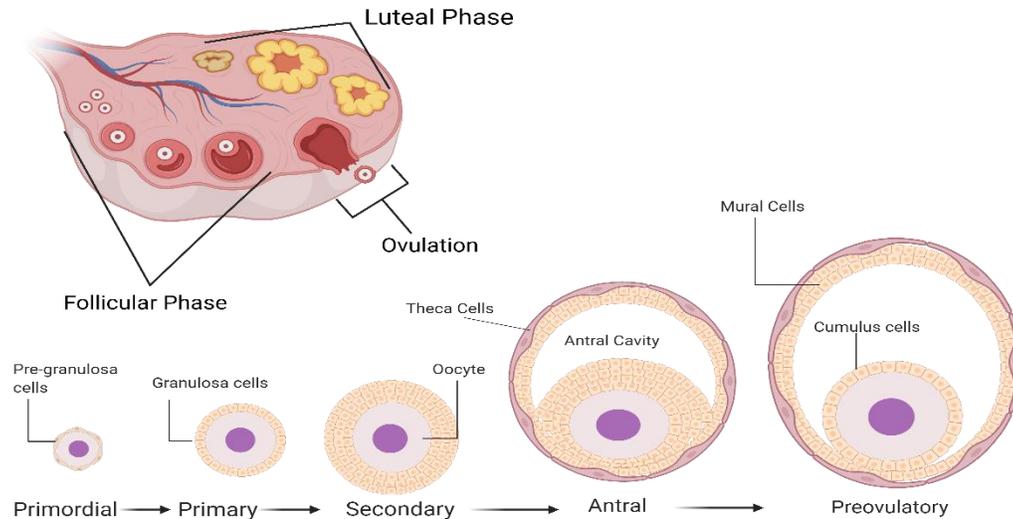


Figure 2: Human Ovarian Follicle Development Stages

Follicular phase can be divided into five major developmental stages based on the size and morphology of the follicle and the number of granulosa cells [1]: primordial, primary, secondary, antral, and preovulatory follicles (**Fig. 2**). These developmental stages are tightly regulated by the positive and negative feedback of hormones in the hypothalamus-pituitary-ovarian axis which is known as the female reproductive axis [6]. Primordial follicles are in a resting, yet metabolically active, stage in the cortex region of the ovarian tissue,[36] and their oocyte is surrounded by squamous granulosa cells. Autocrine and paracrine factors, such as Newborn Ovarian Homeobox (NOBOX) and Spermatogenesis and Oogenesis Helix Loop Helix family (SOHLH1, SOHLH2) also influence the activation of the primordial follicle [21,31]. Upon activation, the oocyte grows and squamous granulosa cells surrounding the oocyte start proliferating and differentiate into cuboidal cells, rendering the primary follicle. The granulosa cells of primary follicle express FSH receptor to which FSH can bind promoting further granulosa cell proliferation and later estradiol production.[4] Primary follicles grow further in size as the mitotic rate of granulosa cell increases, leading to the next developmental: secondary follicle. At this point, theca cells, which are responsible for the conversion of cholesterol into androgens, start surrounding the granulosa cells and trigger the formation of blood vessels and lymph networks around the follicle. The androgen synthesized in the theca cells diffuse to the granulosa cells where it is converted to estrogen metabolites [22]. The secondary follicle releases Inhibin that along with the estrogen provides a negative

feedback to the FSH production by anterior pituitary. Subsequently, the antral cavity starts forming and it is filled with antral fluid that contains granulosa cell debris and is enriched in hyaluronic and taurine acids produced by the granulosa cells. The formation of the antral cavity is essential to maintain the required oxygen flow to the oocyte. The last stage of follicle development is the preovulatory stage during which the oocyte becomes surrounded by a differentiated type of granulosa cells, called cumulus cells, forming what is called cumulus oocyte complex (COC). The increase in estrogen causes a positive feedback loop to produce Luteinizing Hormone (LH) from the posterior pituitary gland. Surge of LH causes the follicle to rupture (Ovulation) and the COC migrates towards the fallopian tubes. Ovulation marks the beginning of the luteal phase during which the ruptured follicle forms the Corpus Luteum. The Corpus Luteum produces progesterone which is required to maintain pregnancy. Progesterone stimulates the endometrium and prepares the uterus for implantation. The rise in progesterone causes a negative feedback loop and decreases the release of FSH and LH from pituitary hormone. If the fertilized egg is implanted in the endometrium (pregnancy), the corpus luteum remains and maintains the level of hormones essential for embryo development (progesterone and estrogen). Otherwise, the corpus luteum regresses into corpus albicans and progesterone and estrogen levels fall. This fall represents the end of the menstrual cycle and beginning of new cycle is caused due to feedback loop which produces FSH for the next cycle [13,21,23,26,36].

1.3. Metabolism during Ovarian follicle development

Oocyte metabolism changes during ovarian folliculogenesis and thus it is plausible that it could be employed to predictor the maturation and competence of the oocyte for a successful pregnancy. Consumption profiles of human oocyte *in vitro* culture has revealed that several amino acids like glutamate, glutamine, arginine, and valine were different in maturing oocytes as compared to degenerating oocytes (Atretic) [16,19]. Although these studies have identified the metabolic profile of oocytes in *in vitro* culture, the metabolic profile during each stage of follicle development still needs to

be further explored in humans. Hence, we will be using the literature on metabolism in animal models when no information is available in human. There might be species specific differences which need to be considered while using these previous studies.

Major pathways for the metabolism of carbohydrates in follicle is Glycolysis. Glucose is the primary source of energy in the granulosa cells due to presence of SLC2A1 and SLC2A4 which are the genes that encode for the Glucose Transporter. Also, the enzymes that are linked to glucose metabolism like Phospho-fructo-kinase have high activity in the granulosa cells [4]. Whereas, in oocyte the activity of these enzymes is low due to which the oocyte has less glycolytic activity. The pyruvate produced by the granulosa cells is transported to the oocyte. In the oocyte, pyruvate is metabolized through the citric acid cycle followed by oxidative phosphorylation to produce energy. Other pathways that are known to metabolize glucose in the follicle are pentose phosphate pathway (PPP), hexosamine biosynthetic pathway (HBP), which involves glutamine and glucose for glycosylation of proteins, hyaluronic acid synthesis and polyol pathway that produces sorbitol and fructose [9]. In mice, oocyte from primordial stage secretes lactate and consumes pyruvate at twice the rate of lactate indicating its dependence on pyruvate as a primary source of energy. Also, the rate of consumption of pyruvate and oxygen increase in mice oocytes from the primary follicle till the antral stage. After the antrum formation, the follicular fluid acts a source of oxygen for the oocyte and the lactate, which is present in follicular fluid, is converted to pyruvate by the enzyme Lactate Dehydrogenase (LDH). In the final stage of follicle development (preovulatory), there is increase in glucose consumption and lactate production which is mediated by the glucose transporters and glycolytic enzymes [9].

1.4. Genome-scale Metabolic Model of Human

Genome-wide metabolic models (GEMS) are *in silico* tools that can be employed to study cellular metabolism of different organisms, from unicellular (e.g., bacteria) to multicellular (e.g., human). Genome-wide metabolic models are mathematical representations of cellular metabolism in the form of a set of linear equations each corresponding to a biochemical reaction known to be present in a given organism. The model also includes the enzymes that catalyze each biochemical reaction as well as the associated genes that encode for them, which is useful for integration of different types of omics data (e.g., transcriptomics, proteomics) [2]. With advancement of sequencing techniques and increase computational power, generation and simulations of mathematical metabolic models is much easier and faster than manual curation by literature mining [14].

The GEM models can be used for a various application to: a) predict the active metabolic reactions (called metabolic flux); b) optimize culture media in biotechnology applications (e.g., degradation of biomass, production of natural products); c) study the effect of genetic mutations on metabolism; or d) analyze drug metabolism, just to name a few [29]. As the problem is ill-posed multiple solutions are possible, optimization techniques are employed to identify the set of conditions that enable the desired solution, with linear programming one of the common ones (Flux Balance Analysis).

Flux Balance Analysis (FBA) uses linear programming (LP) to calculate the activity or metabolic flux through each reaction. It is a constraint-based optimization method that uses stoichiometric constraints as well as inequalities that bound the metabolic reconstruction to maximize biological objective functions, such as biomass production, or secretion or consumption of a metabolite [30]. The assumption of the flux balance analysis is that the system is at steady state. Steady state of Flux balance Analysis is an ideal state and adding kinetic parameter would be logical for accurate predictions. Although FBA cannot handle dynamics, one of its main advantages is that it can compute metabolic fluxes in large metabolic networks quickly, thus allowing the prediction of key endo and exo-metabolites.

Latest genome-wide reconstruction of human metabolism was a large international community effort and was called Recon3D [3]. Recon3D consists of 13,543 reactions, 4,140 unique metabolites, 3,697 genes encoding for enzymes and 111 metabolic subsystems and contained 9 compartments, including cytosol, nucleus, mitochondria, inner mitochondria, lysozyme, endoplasmic reticulum, golgi apparatus, peroxisome, extracellular matrix. The reactions and the metabolites in the Recon3D is attached in supplementary File B.

Stoichiometric matrix (S) is the representation of biochemical pathways, in which the columns of stoichiometric matrix represent a biochemical reaction ($n=13,543$ in the case of Recon3D) and in rows represent a given metabolite ($m= 8,399$ for Recon3D). A product of a given reaction is represented by a positive integer with value equal to the stoichiometric coefficient in the biochemical reaction. A substrate (consumed metabolite) is depicted by a negative integer with value equal to the stoichiometric coefficient in the biochemical reaction. If a metabolite (i) does not participate in a given reaction (j), the value of S is set to 0 ($S [j, i] = 0$). Recon3D model also includes a biomass reaction, which depicts the metabolites needed for a given cell or to maintain a given structure. S also includes spontaneous reactions—those that do not require enzymes to occur, and transport reactions between the cellular compartments (e.g., media to the cytoplasm, cytoplasm to the nucleus or mitochondria). Other than these metabolic reactions, there are three more types of reactions in metabolic reconstruction that use and recycle accumulated metabolites, or produce the required metabolites:

1. *Exchange reactions* - are reactions that move metabolites across *in silico cellular* compartments that do not require a transporter. These compartments are intra- and inter- cellular membranes present in the cell.
2. *Sink reactions* - As there are still multiple metabolic pathways that are not known, some metabolites can accumulate in the system. Thus, sink reactions are a mathematical artifact to prevent metabolites from accumulating inside the cell and avoid blocking reactions. Sink reactions are reversible.

3. *Demand reactions* - Like sink reactions, demand reactions are also a mathematical artifact to add metabolites that need to be consumed but do not have any reaction that allows them to access the given compartment where they are needed.

Each reaction present in the system can be constrained by adding limits to the flux through the reaction. For instance, irreversible reactions can be modelled by adding a lower limit of 0 and a positive upper limit. This is useful to simulate conditions like anaerobic systems or limiting cellular growth etc [3].

Table 1: Human Metabolic Model (Recon3D) Properties

Model	Recon3D	GapFilled Recon3D
Reactions in Model	13543	14083
Metabolites in Model	4140	4289
Genes	3697	3697
Subsystem	111	111
Metabolites in Cytoplasm	3346	3417
Metabolites in Mitochondria	966	996
Metabolites in Peroxisome	514	529
Metabolites in Extracellular space	1646	1657
Metabolites in Lysosome	453	494
Metabolites in Endoplasmic Reticulum	955	995
Metabolites in Nucleus	182	204
Metabolites in Golgi apparatus	335	390
Metabolites in inner mitochondrial chamber	2	24
Blocked Reactions	1582	1112

1.5. Context specific Metabolic Model

Genome scale model aims to include all the reaction that are present in an organism which has wide range of application from understanding the physiology to simulate the organism in different conditions. Although these models represent all the plausible metabolic reactions that can occur in an organism. However, in the case of multicellular systems, such as humans, not all the genes can be transcribed in all the cell types, thus their metabolism is distinct. Cell specific models have been proven to be a better predictor of the cellular metabolism of a given entity than using the entire set of plausible reactions. Context specific models are subset of the genome scale model which contains all the active reactions occurring in a specific context (Cell type, diseased stage, developmental stages) [33]. There are different algorithms that allow to generate context specific metabolic models by integrating different types of omics data with the entire organism metabolic model.

Context specific models provide a framework for *in silico* experiments involving simulations of cellular metabolism. These models can be used to predict the outcome of experiments, saving time and resources. Several methods have been proposed to create context-specific models and each method has different assumptions and requirements. These methods can be classified into three main categories based on their algorithm which are: 1) Group Iterative Multiple Model Estimation(GIMME)-like: This algorithms involve minimizing flux through lowly expressed genes; 2) Integrative Metabolic Analysis Tools(iMAT)-like: This algorithms finds optimal trade-off between the removing reactions that have low gene expression and keeping reaction with high gene expression; 3) Model Building Algorithm(MBA)-like: This algorithm takes core reactions that are retained and active while adding only a minimal set of additional reaction. Different algorithms can be used to integrate specific types of omics data (metabolomics, proteomics, transcriptomics).[35]

FASTCORE belongs to the MBA family of algorithms. It is a fast and efficient method that requires a core set of reactions to reconstruct a large context specific model. Different types of omics data can be integrated using FASTCORE.

Chapter 2

Methods

2.1 Identification of metabolic enzymes present in the distinct ovarian follicle cells during ovarian follicle development using single-cell transcriptomic.

Single cell RNA sequencing data was retrieved from the European Bioinformatics Institute database, GSE107746 [41]. Follicles were isolated from 7 donor patients who underwent ovariectomy as described in Table 2 and 3. The follicles were selected from different stage of follicle development based on their morphology. The median age of donor was 28. The dataset comprises 80 samples from oocyte cells and 71 samples from granulosa cells isolated from different stages of follicle development. The follicle was further split into oocyte from 5 stages (17 Primordial, 25 Primary, 12 Secondary, 23 Antral, 3 Preovulatory) and granulosa (8 Primordial, 15 Primary, 6 Secondary, 24 Antral, 18 Preovulatory).

Table 2: Reason for ovariectomy in donors.

Reason for Ovariectomy	Number of Donor
Sex Reassignment	1
Fertility Preservation for Cervical Cancer	1
Fertility Preservation for Endometrial Cancer	2
Fertility Preservation for Benign Ovarian Mass	2
Fertility Preservation for Lymphoma	1

Table 3: Donor's biodata and number of follicle isolated [41]

Donor	Follicle number obtained	Age	The day of menstrual cycle
Donor 1	16	25	Day 10
Donor 2	13	26	Day 13
Donor 3	26	24	Day 15
Donor 4	4	31	Day 8
Donor 5	4	34	Day 14
Donor 6	11	26	Day 9
Donor 7	3	28	Day 11

The raw data was processed as described in the flowchart in Figure (3). Briefly, low quality reads and artificial sequences like Illumina adapter, UP1, UP2 and polyA tails were removed from each sample using Cutadapt [28]. The trimmed sequences were mapped to the human latest available reference genome using splice aware mapping tool: STAR [7]. The quality of the reads was checked and only samples which had greater than 60% of uniquely mapped reads were used for further analysis. Final dataset had 73 cells from oocyte and 64 cells from granulosa cells. The mapped reads were quantified using HTSeq [37]. Expressed genes whose total counts were less than 10 or were only expressed in less than 30% of the samples at each stage of both the cell types were further removed. Principal component analysis (PCA) was employed to understand the variance between the samples. Differentially expressed genes between subsequent stages for each cell type were identified using DESeq2 [24] to detect the dynamic expression patterns during follicle development. Genes which had Benjamin Hochberg adjusted p-value less than 0.05 and log fold change greater than or equal to 0.5 were considered significant. The prevalence of each gene was calculated

for each stage of follicle development and cell type and was then used to create the context specific follicle metabolic model for each cell and each developmental stage.

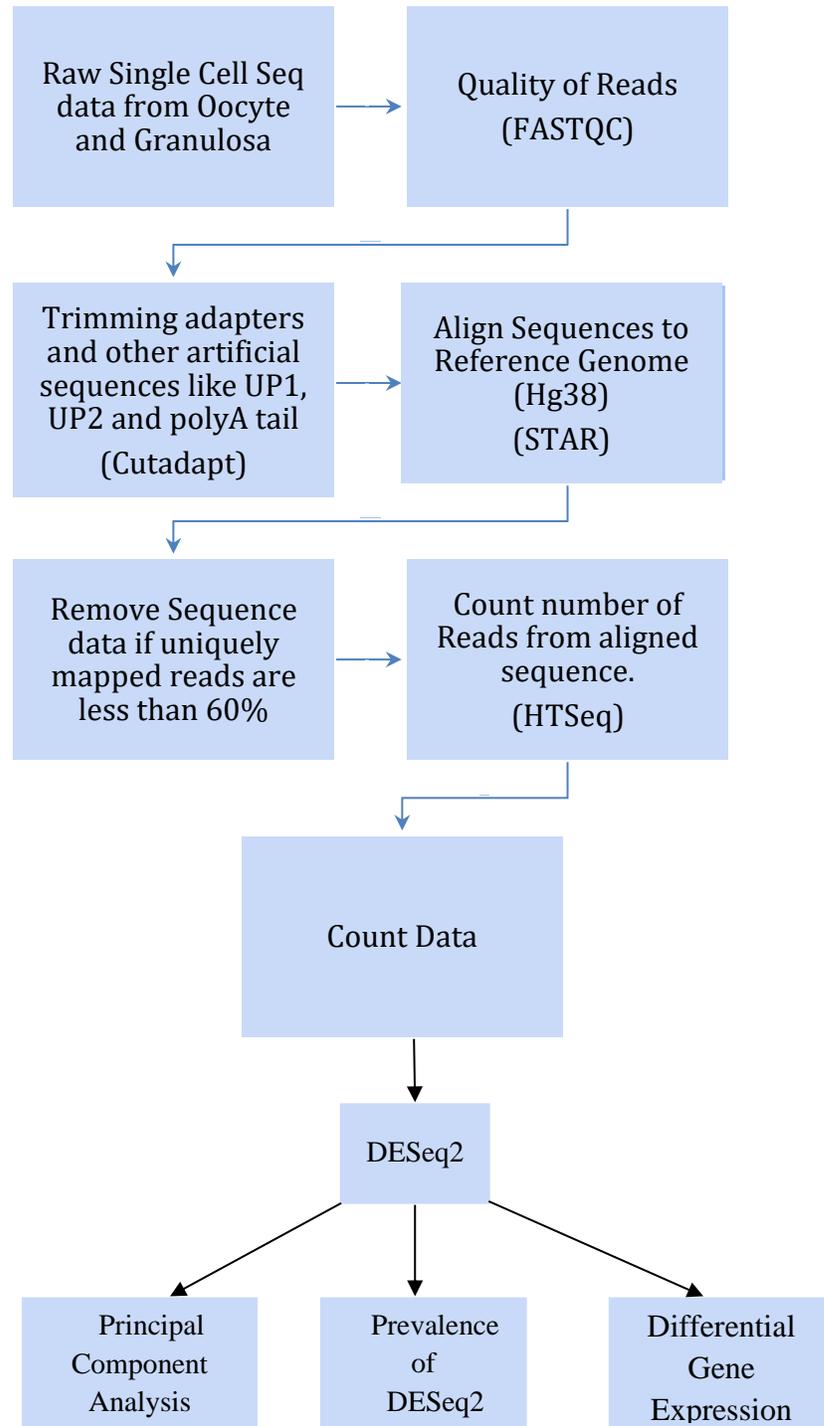


Figure 3: Pipeline to process single cell RNA sequencing data

2.2 Generation of context-specific genome-wide metabolic models using ovarian follicle development in humans using FASTCORE.

Context-specific metabolic model of human ovarian follicle cell was created by integrating the processed single cell-transcriptomic data from different follicular developmental stages to the latest available reconstruction of the human metabolic model (Recon3D) [4]. The workflow of creating a context-specific model is represented in Figure (4).

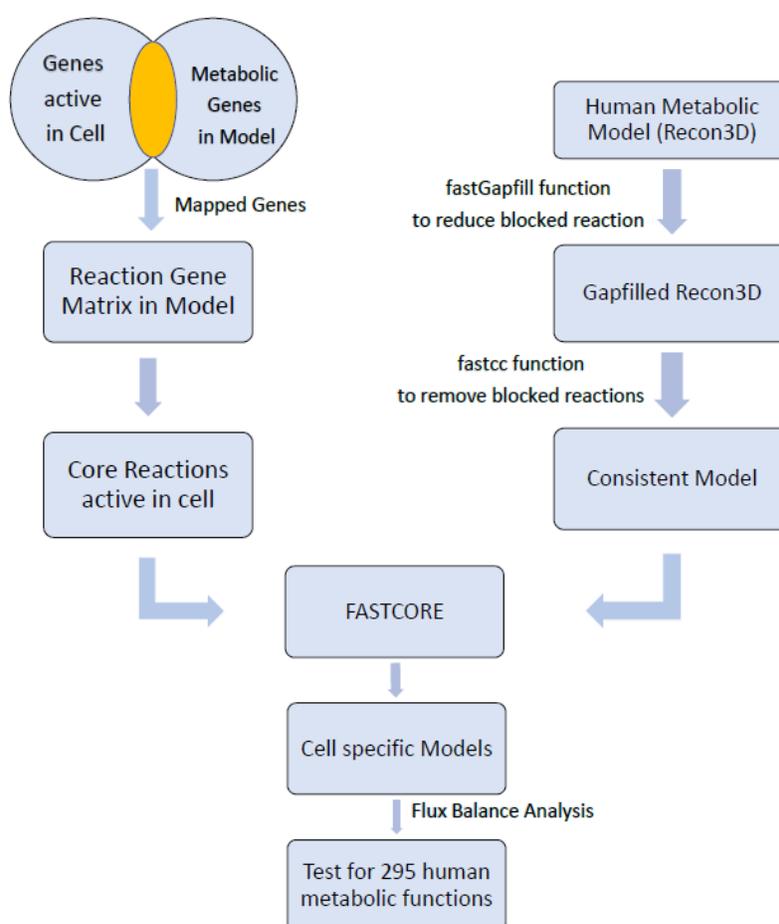


Figure 4: Generation of Context specific model using FASTCORE. The genes that are expressed during a particular stage of follicle development were mapped onto the metabolic genes present in the genome scale human metabolic model (Recon3D). Mapped genes or common genes was overlaid on the reaction gene matrix to obtain core set of reactions. Recon3D was subjected to Gap filling algorithm(fastGapFill) which reduces the number of blocked reactions (Reactions with zero metabolic flux). The obtained Gapfilled model still had blocked reactions which was removed using fastcc function. The Consistent model (Model without any blocked reaction)

Reconstruction of context-specific genome-wide metabolic models for ovarian follicles was generated using the FASTCORE algorithm implemented in the Constraint based Reconstruction and Analysis (COBRA)Toolbox [15]. The input for the FASTCORE algorithm is the global consistent model, in this case Recon3D, and core reactions which are active in a particular context of interest, using as a proxy the genes that are transcribed in the oocyte and granulosa cells each stage during ovarian follicle development. FASTCORE identified the subset of core reactions that includes those reactions for which the genes that encode their associated enzymes were present during follicle development and a subset of reactions that allow to reconstruct a flux consistent subnetwork of core reactions from the global consistent model.

2.3 Stage-specific model of Oocyte and Granulosa Cells

We constructed cell specific (oocyte, O; and granulosa cells, S) and stage specific models (primordial, PR; primary, PR; secondary, SC; antral, AT; and preovulatory, PO) using FASTCORE. Each model was constrained to plasma concentrations in females obtained from the human metabolome database (<https://hmdb.ca/>) and then tested for 295 metabolic function reactions of humans provided in the supplementary File B.

2.4 Community detection of Follicle Model

Enzyme-metabolite network (bipartite) and enzyme network of the ovarian follicle model were constructed. The bipartite graph consists of connections between enzymes and metabolites they catalyze. The enzyme network consists of the edges connecting enzymes that catalyze common metabolites. Infomap, a well-known community detection algorithm, was employed to identify the communities or modules in which the enzyme network could be split. Infomap decomposes the enzyme network into communities based on the amount of information flow, through the elements in the network. Infomap algorithm calculates the flow of information through the network using the map equation [34]. Closely related enzymes that have common metabolites have higher information flow through them as compared to the enzymes which do not have common metabolites. This flow was used to calculate normalized flow through communities, pathways, metabolites, and enzymes represented as follows:

$$f_{N_i} = \frac{\sum_{k=l}^{k=n_i} w_k f_k C_k}{\sqrt{\sum_{k=l}^{k=n_i} w_k}} \quad (\text{eq.1})$$

where f_{N_i} is the normalized intensity flow of the community [32] (or pathway, or metabolite) i , f_k is the flow calculated with Infomap for element k in the community i , w_k is the number of metabolites that are catalyzed by the enzyme k according to the enzyme-metabolite bi-partite graph, C_k is the normalized mean count for the given gene in the transcriptomic data. If the gene is not present, not transcribed, then C_k was set to zero. These normalized flow intensity calculations account for the differences in the community size as well as dynamic expression of the genes. Infomap was run with default parameters and setting the number of bootstraps to 1,000. The normalized flow intensity was used to calculate the enriched pathways, metabolites, and enzymes.

2.5 Significant metabolites, genes, and metabolic pathways during ovarian follicle development

Metabolic pathway enrichment using the subsystem definition in Recon3D, f_p was calculated as the normalized intensity flow of all the genes or nodes that belongs to a specific pathway (subsystem) within the follicle model (e.g., pyruvate metabolism), and it was compared with the background normalized intensity flow of the equal number of randomly selected genes for the entire network using the enzyme metabolite bigraph [30].

$$Z\ Score_{(Pathway, Metabolites, Enzymes)} = \frac{f_p - \mu}{\sigma} \quad \text{eq.2}$$

These values were employed to calculate a Z-score for each metabolic pathway in the ovarian follicle metabolic model. Similarly, enzyme and metabolite Z-scores were established. In the case of enzymes, the flow was determined accounting for all the metabolites that are catalyzed by the given enzyme in the entire network, w_k . Similarly, for metabolites, the flow was determined based on the flow of the enzymes that catalyze the reactions in which the metabolite participates, setting w_k to 1. Common metabolites, such as water, oxygen, or ATP, for instance, were removed from the list.

Chapter 3

Result and Discussion

3.1 Ovarian follicle transcription activity

3.1.1 Highly variable genes during follicle development

Transcriptomic analysis revealed the dynamic expression pattern at each stage of the follicle development. The count data was normalized using DESeq2 method. i.e., Median of ratio's method and then the samples were further filtered stage wise to reduce the background noise or false expression. After processing it was observed that the oocytes expressed 18,741 genes while the granulosa cells expressed 17,092 genes.

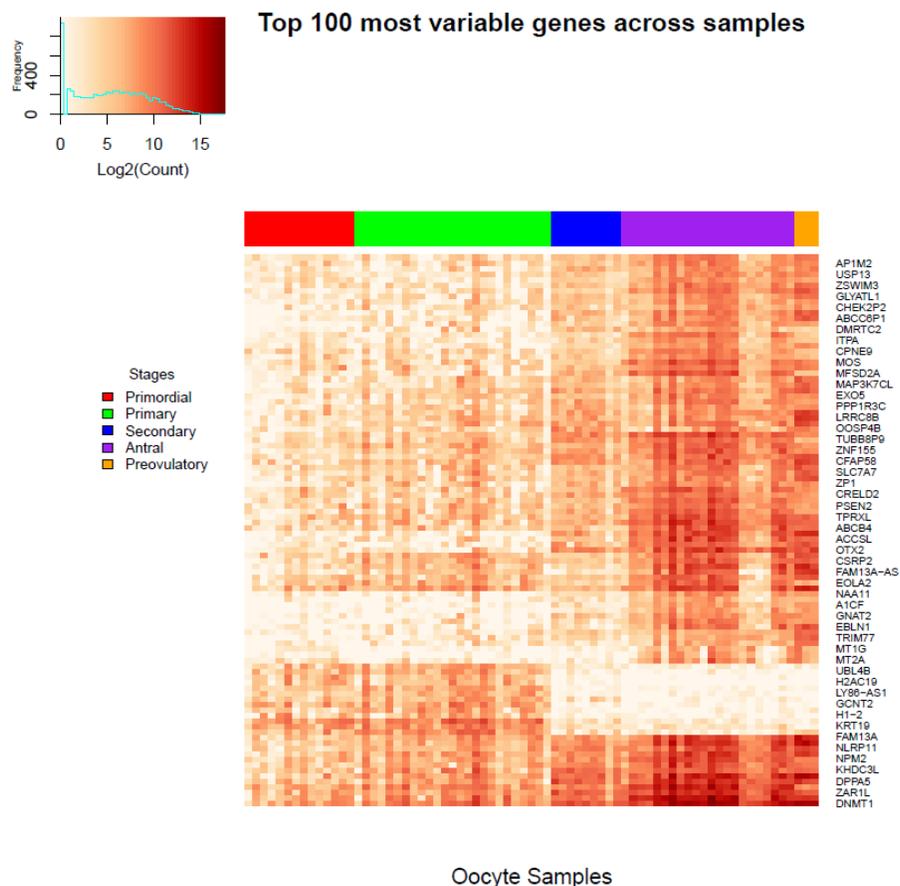


Figure 5: Heatmap of the 100 most variable genes across the Oocyte Samples.

The oocyte transcriptional profile was dynamic over the entire folliculogenesis process, with the primary and primordial oocytes having similar levels of transcription compared with secondary follicles onwards (Fig. 5). Oocyte specific genes such as *ZP1*, *OOSP4B*, *OTX2* were among the most variable transcripts in the oocyte. Most of the oocyte transcripts increased their abundance as folliculogenesis took place, except for a small cluster from *UBLAB*, *H2AC19*, *KRT19*. Of note, two antral samples from oocytes presented a different transcriptional profile pattern compared with the rest of the oocyte antral samples. While we cannot conclude the reason for this deviation, one plausible explanation could be that the quality of the antral oocytes was not adequate and were in a degenerative stage due to atresia.

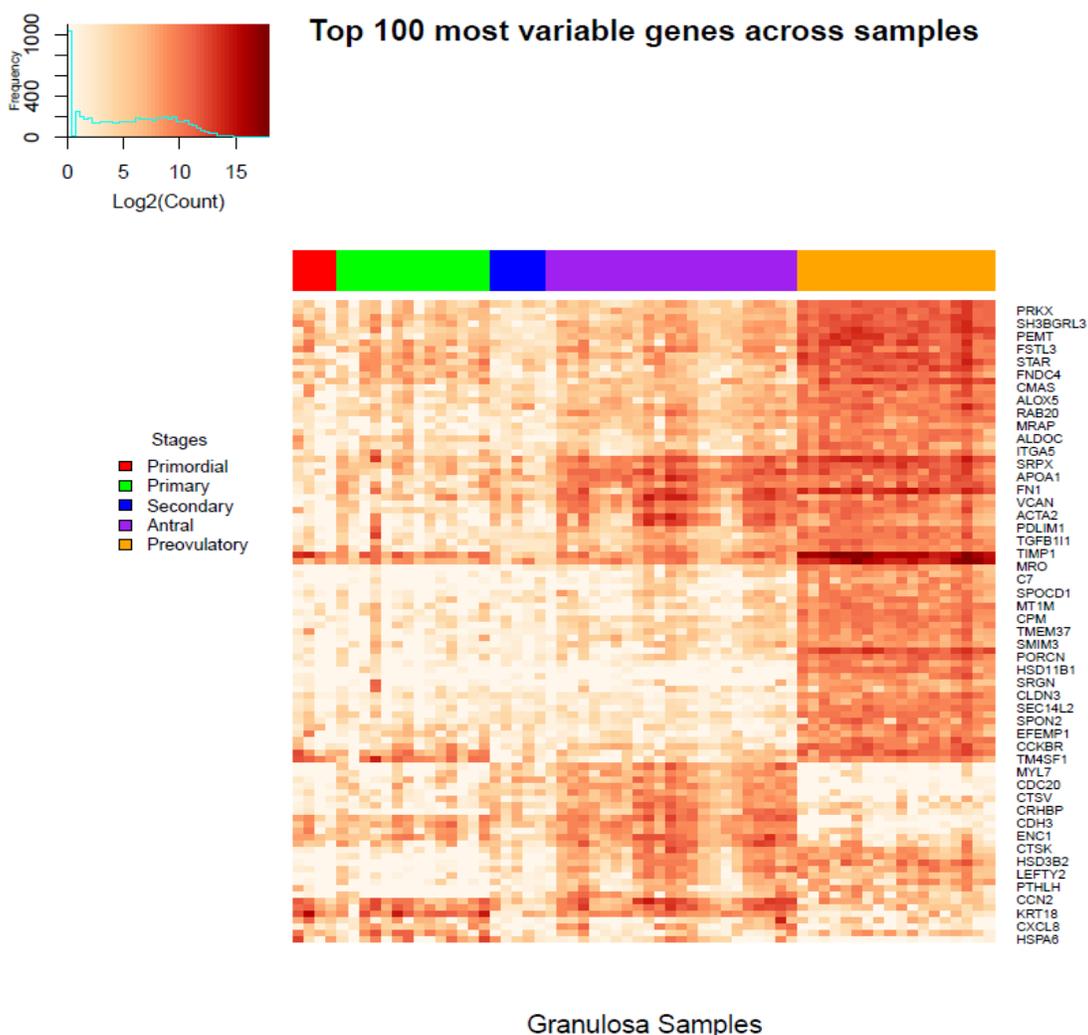


Figure 6: Heatmap of variable genes across the samples in Granulosa Sample

From the expression pattern of the most variable genes in the granulosa samples, the expression patterns during primordial and primary stages were the most similar as compared to the rest of the stages of follicle development—as it was in the oocyte profiles. Some of the most notable genes which appeared in the heatmap are *HSD11A1*, *STAR* and *KRT18*.

HSD11B1 increased at the last stage of follicle development, an enzyme involved in the conversion of glucocorticoid to cortisol [38]. *STAR* is granulosa specific gene with high expression in the later stage of follicle development, which was also observed in the principal component analysis in **Figure 8B**. *KRT18* had higher expression in the antral stage of follicle development which have been identified as a potential gene that might be associated with maintaining the ovarian reserve cause atresia in follicle when the expression of *KRT18* decreases [11]. The expression level of *TIMP1* and *MRO* was notably higher in all the stages and maximum in the preovulatory stage. *TIMP1* is an inhibitor of collagenases and can inhibit ovulation [12]. *MRO* gene expression changes in PCOS pathophysiology. [20] *ALDOC* gene which encodes for the Aldolase enzyme was observed in granulosa sample.

3.1.2 Principal Component Analysis

Using Principal Component Analysis (PCA), we observed that the sequencing data separated clearly by cell type and ovarian follicle cellular stage, indicating that indeed there is a unique transcriptional program for oocytes and for granulosa cells during folliculogenesis. Also, the variability in the oocyte cluster was greater than the variability within the granulosa transcriptional pattern, which indicates that the oocyte transcriptional pattern is more dynamic and diverse than the granulosa transcription during folliculogenesis. Also, the PCA separated the folliculogenesis stages within a given cell type in the same temporal order as during the developmental process (e.g., primordial cells were closer to primary and primary to secondary).

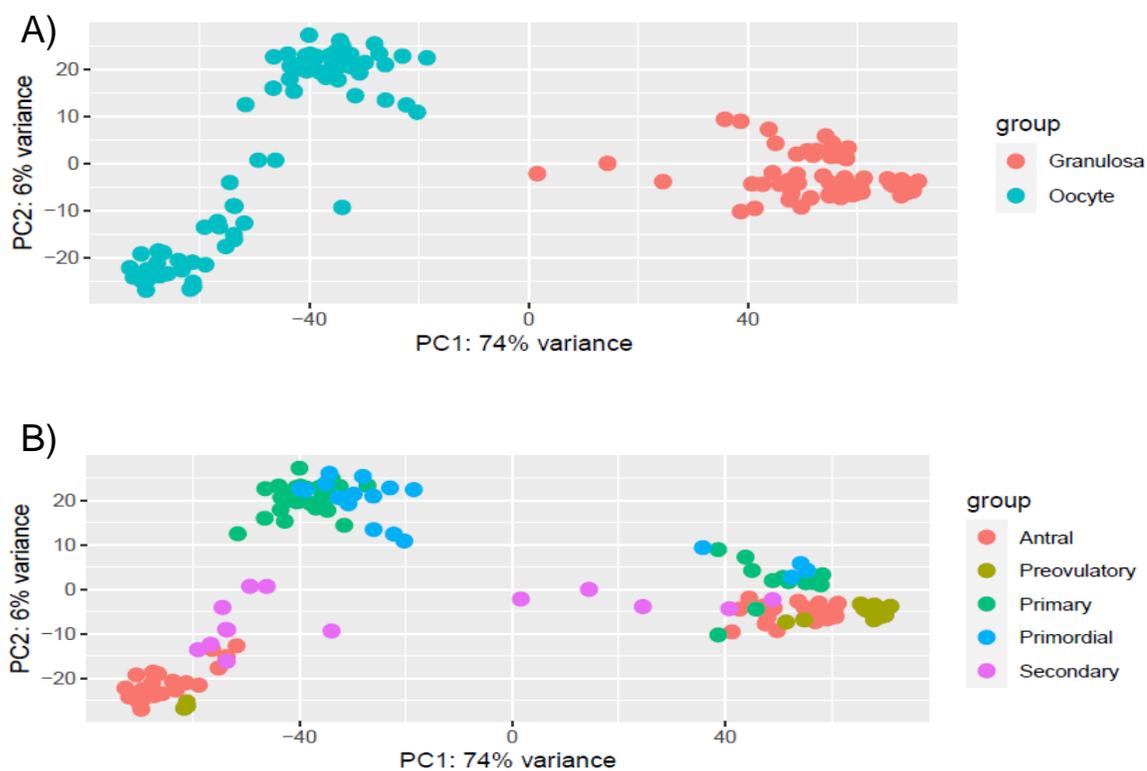


Figure 7: Principal Component Analysis of Samples based on A) Cell Type B) Stage of follicle development.

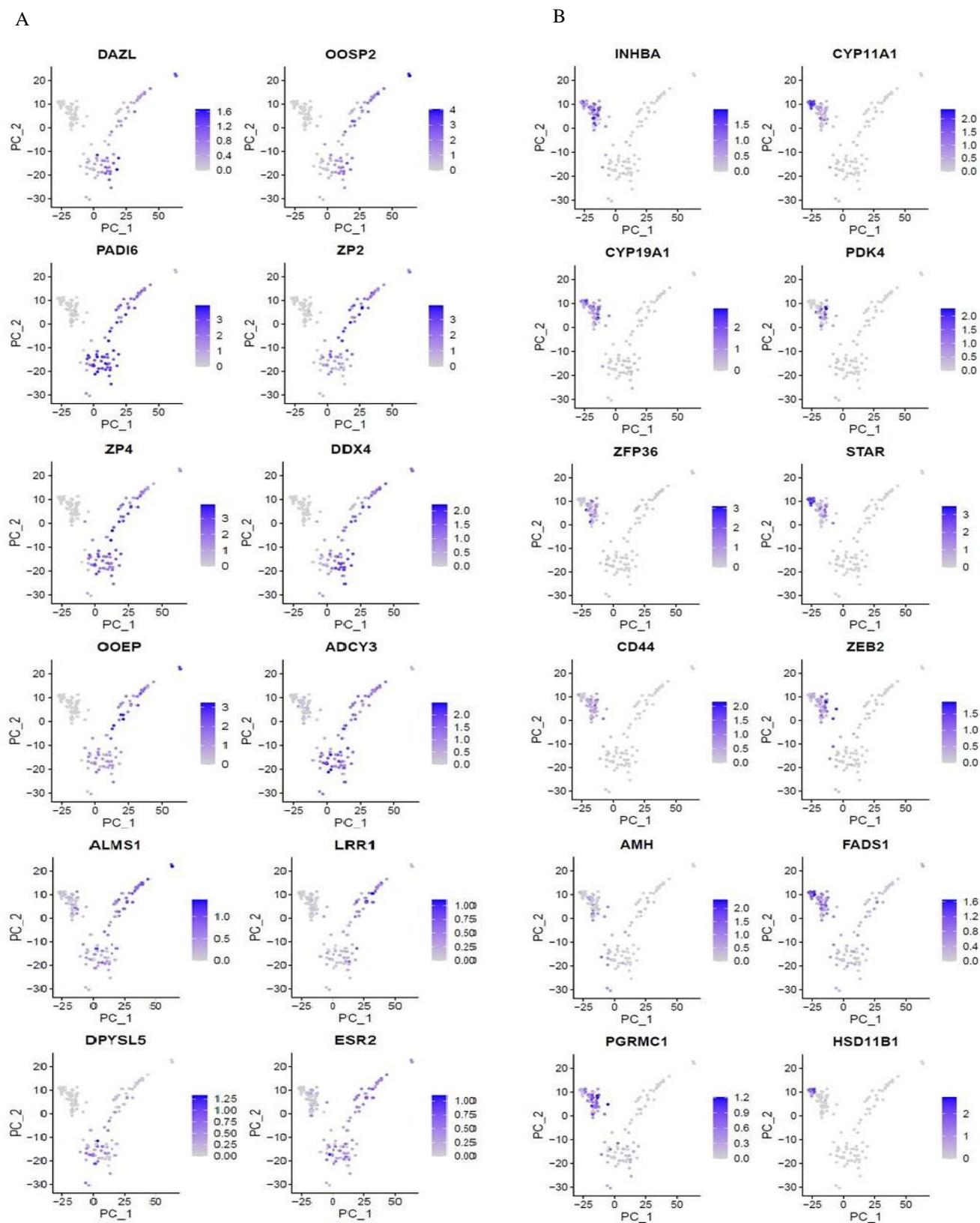


Figure 8: Expression of A) Oocyte Specific Markers, B) Granulosa Specific Markers in samples. X-axis represents Principal Component 1(PC₁) and Y-axis is Principal Component 2(PC₂). The scale represents normalized count.

We further validated the PCA results with several cell specific markers of oocyte and granulosa cells from the ovarian kaleidoscope database. We selected 12 genes which was localized in oocyte like *DAZL*, *OOSP2*, *PADI6*, *ZP2*, *ZP4*, *DDX4*, *OOEP*, *ADCY3*, *ALMS1*, *LRR1*, *DPYSL5*, *ESR2*. Out of these 12 genes 3 were previously observed in the previously published result (*DAZL*, *ZP2*, *DDX4*). The genes which code for enzyme in oocyte are *ADCY3* and *PADI6* involved in oocyte maturation. Similarly, 12 genes localized in granulosa cells were selected which includes *INBHA*, *CYP11A1*, *CYP19A1*, *PDK4*, *ZFP36*, *STAR*, *CD44*, *ZEB2*, *AMH*, *PGRMC1*, *FADS1*, *HSD11B1*. Out of these 12 genes, 6 genes were previously observed (*INBHA*, *AMH*, *STAR*, *ZEB2*, *CD44*, *CYP11A1*). The genes which code for enzymes in granulosa are *HSD17A1* and *CYP19A1* involved in steroid metabolism. The genes like *STAR* and *CYP11A1* are expressed only from secondary stage of follicle development which could be observed in the plot in Figure 8B.

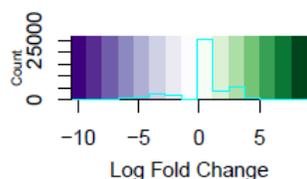
3.1.3 Differentially expressed genes during follicle development

The samples were separated into their cell types and differentially expressed genes between the subsequent stage was estimated using DESeq2. Genes were considered as differentially expressed if the FDR corrected p value was less than 0.05 and the log fold change was greater than 0.5. The heatmaps (Figure 9 and 10) were generated by combining the log fold change of all the differentially expressed genes during the follicle development stages. If the gene was not significant the fold change was set to 0.

There were 9,996 genes (**Table 4**) which were significant gene in oocyte ($p < 0.05$, $LFC > 0.5$) during the follicle development. Differential Expression of the genes that encode the Zona Pellucida family (*ZP1*, *ZP2*, *ZP3*, *ZP4*), which is essential in primordial and primary follicles, was localized to the initial stage of follicular development and were not significant in later stages, although the ZP transcripts were present in the oocyte during all the stages of follicle development.

Table 4: Differentially Expressed genes between subsequent stages of follicle development in Oocyte.

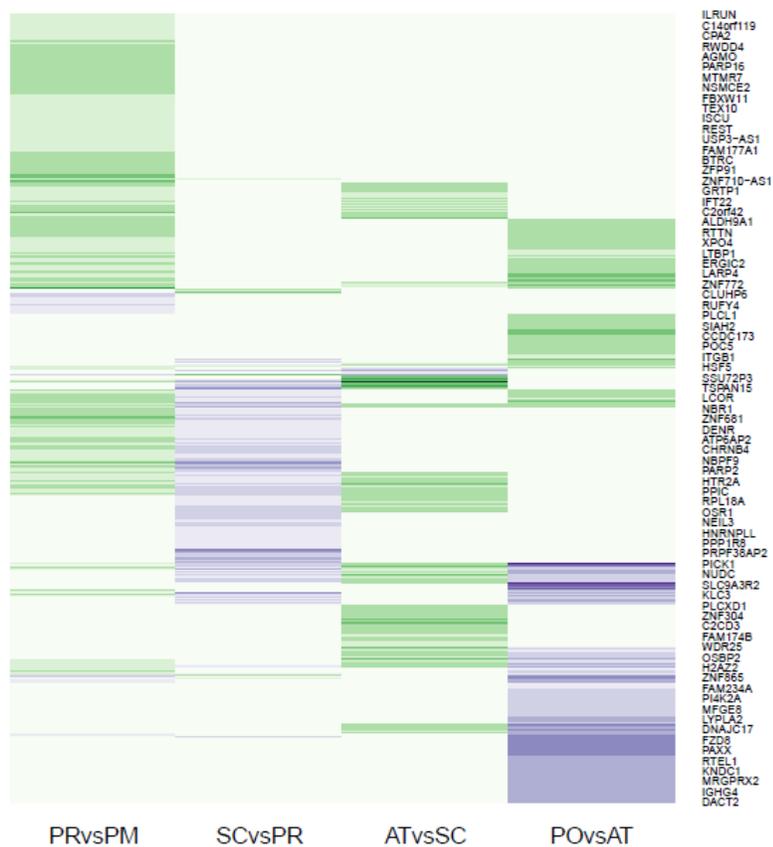
Comparison	Total Significant Genes (FDR<0.05 & LFC >0.5)	Upregulated (LFC>0.5)	Downregulated (LFC<-0.5)
Primary vs Primordial	6039	5595	444
Secondary vs Primary	3212	156	3056
Antral vs Secondary	2741	2606	135
Preovulatory vs Antral	4551	1896	2655

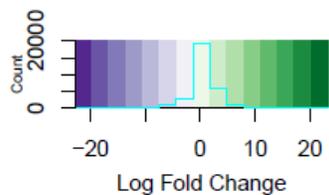


Oocyte Sample

$p < 0.05$
LFC > abs (0.5)

PRvsPM- Primary vs Primordial
SCvsPR-Secondary vs Primary
ATvsSC-Antral vs Secondary
POvsAT-Preovulatory vs Antral

**Figure 9: Differential gene expression between the subsequent stages of follicle development in oocyte cells**



Granulosa Sample

$p < 0.05$
 $LFC > \text{abs}(0.5)$

PRvsPM- Primary vs Primordial
 SCvsPR-Secondary vs Primary
 ATvsSC-Antral vs Secondary
 POvsAT-Preovulatory vs Antral



Figure 10: Differential gene expression between the subsequent stages of follicle development in granulosa cells

There are 7,931(**Table 5**) genes which were differentially expressed ($p < 0.05$, $LFC > 0.5$) in granulosa during the follicle development, with a smaller number of significant genes between the primary and primordial stage. The expression levels in primary granulosa cells were higher as compared to the secondary granulosa samples with 1,662 genes downregulated which means the genes had higher expression in the primary stage of follicle development. Steroidogenic enzymes like CYP19A1, CYP11A1, HSD17B1 were not significant between secondary and primordial but were highly upregulated between Antral and Preovulatory stages. Genes which encode for hormone receptor like AR, ESR1 were only significantly expressed in preovulatory and antral stages. While ESR2 and PGR receptors for estrogen and progesterone was not significant in the samples.

Table 5: Differentially expressed genes between subsequent stages of follicle development in Granulosa Cell.

Comparison	Total Significant Genes (FDR<0.05 & LFC >0.5)	Upregulated (LFC>0.5)	Downregulated (LFC<-0.5)
Primary vs Primordial	1016	902	114
Secondary vs Primary	1729	67	1662
Antral vs Secondary	4736	4702	34
Preovulatory vs Antral	3139	1177	1962

3.2 Ovarian Follicle Model

We generated a follicle-specific genome-wide metabolic model by overlaying the genes that were transcribed in the oocyte and granulosa cell during folliculogenesis into Recon3D using FASTCORE [39].

FASTCORE works the following:

- 1) The algorithm first identifies all the reactions that are blocked. Blocked reactions are those that their metabolites are not produced by another reaction and/or their products are not consumed by other reactions. Blocked reactions are common as our knowledge of human metabolism is not complete. Removal of blocked reactions in Recon3D rendered a consistent model. The consistent part of the Recon 3D model has 12,971 reactions, 4,389 unique metabolites, and 3,013 genes.
- 2) Then the set of follicle biochemical reactions that could be active during folliculogenesis were overlaid over the consistent Recon3D model.
- 3) Finally, FASTCORE identified the minimum set of reactions from the consistent part of Recon3D that minimize the number of block reactions in the context-specific ovarian follicle.

The consistent ovarian follicle model has 10,538 reactions, 3,484 unique metabolites and 2,954 metabolic genes that code for enzymes. Then, we determined whether the model was able to carry out all the reactions that are well-known to take place in human cells (e.g., production of ATP) as well as the functions specific from ovarian cells (e.g., estrogen and progesterone production). The ovarian follicle model was able to model 276 metabolic functions out of 295, including the secretion of estradiol and progesterone by the granulosa cells; the consumption of pyruvate by the oocyte (Attached in supplementary file).

3.3 Bi-partite Graph of Enzyme and Metabolites

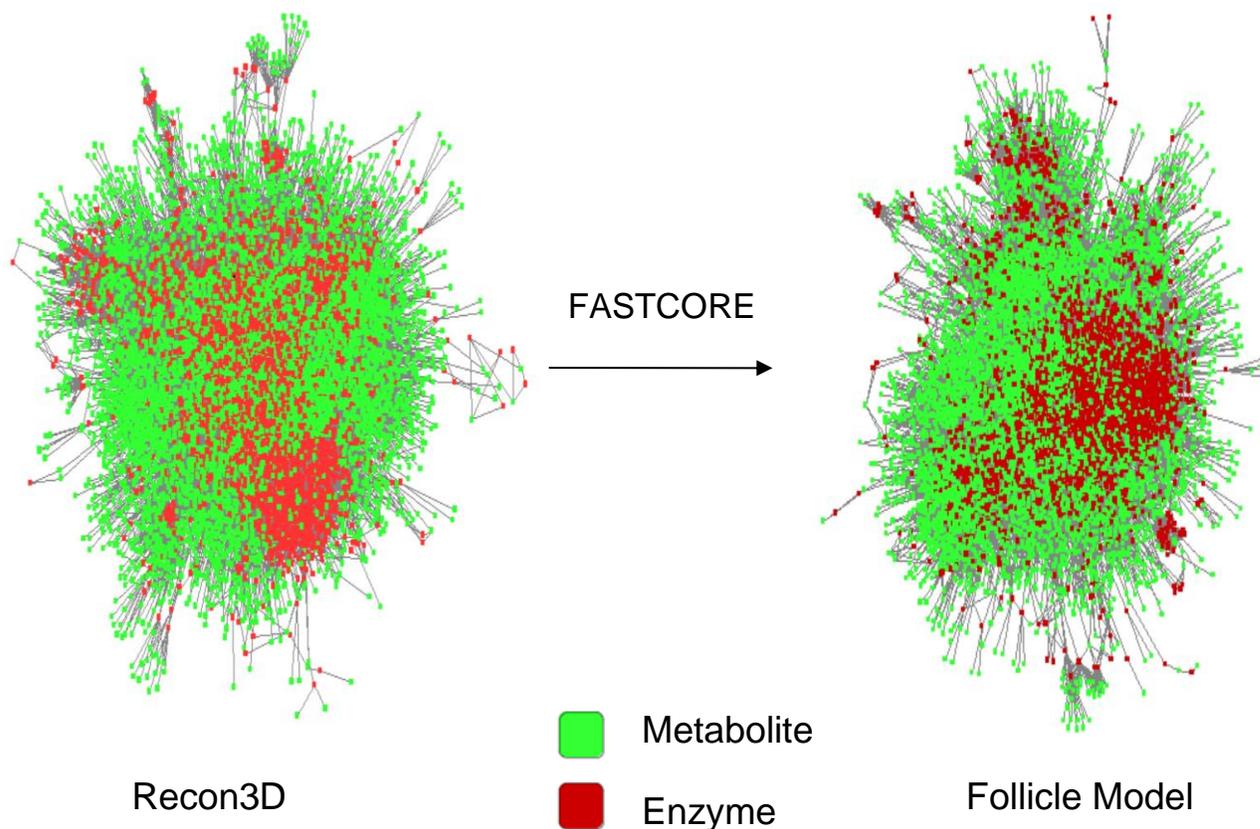


Figure 11: Enzyme-Metabolite Bipartite network of Recon3D(Left) and Human Ovarian Follicle Model (Right).

The bipartite graph is a connection between the enzyme and metabolite that is catalyzed by the enzyme. The green node indicates the metabolites, and the red node indicates the enzyme. There were 8824 nodes and 94054 edges in Recon3D whereas the follicle model has 7953 nodes and 70138 edges.

Enzyme network was created by finding the connection between two enzymes that catalyzes same metabolites. It is a highly connected network and requires community detection algorithms like Infomap to identify the hotspots in these networks. We obtained 31 communities or hotspots from the enzyme network and calculated the enrichment scores for pathways, metabolites, and enzymes.

3.4 Metabolites present in stage specific models.

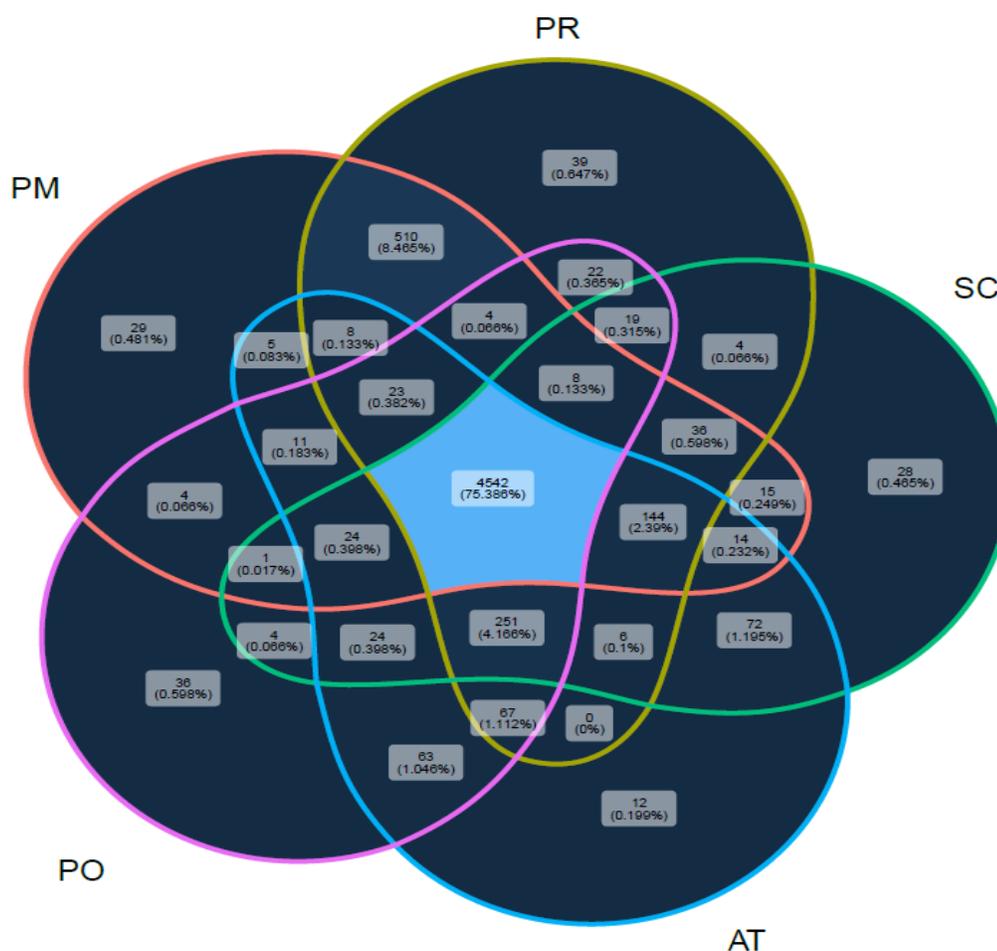


Figure 12: Venn diagram Metabolites obtained from at different stages in Oocyte Models. (PM: Primordial, PR: Primary, SC: Secondary, AT: Antral, PO: Preovulatory)

There are 75% of metabolites which are common in all the stages of follicle metabolites in oocyte. Metabolite f1Alpha which is involved in the keratan sulfate pathway was unique only oocytes from primordial follicles. Glycolipids and 4-hydroxy benzoic acid, which is involved in phenylalanine metabolism, were unique in the oocytes of primary follicles. Oocytes from secondary follicles presented methyl histamine, part of histidine metabolism and dihydroxyacetone phosphate in oocytes of antral follicles, part of the triacylglycerol pathway. Dimethylallylpyrophosphate involved in cholesterol metabolism was the unique metabolite in preovulatory follicle model.

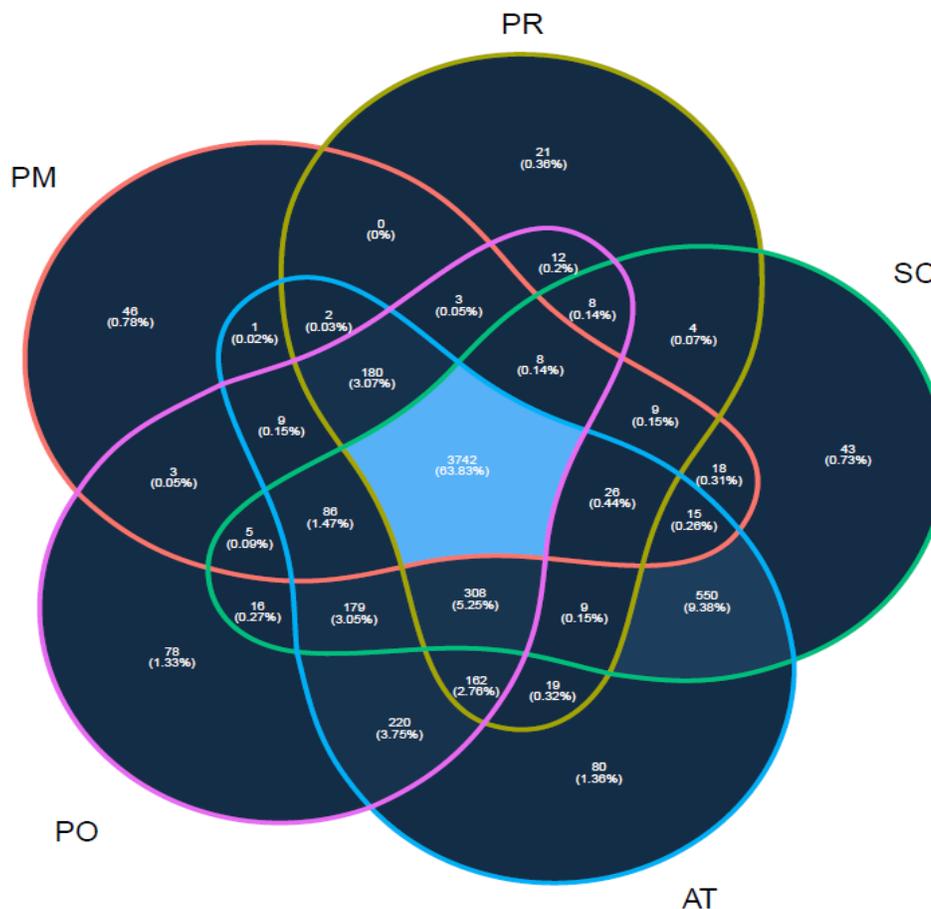


Figure 13: Metabolites obtained from at different stages in Granulosa Models.
(PM: Primordial, PR: Primary, SC: Secondary, AT: Antral, PO: Preovulatory)

There are 63% metabolites that were common in all stages of follicle development in granulosa cells. The granulosa cells from the primordial stage had hydroxybutyric acid in the mitochondrial compartment as unique metabolite which is involved in butanote pathway. The granulosa cells from the primary stage had cysteinyl glycine as unique metabolite that is involved in the glutathione pathway. The granulosa cells from the secondary stage had formaldehyde as unique metabolite that is involved in the various amino acid metabolic pathways like glycine, serine, threonine, tyrosine, and tryptophan pathway. The granulosa cells from the antral stage had creatine as unique metabolite that is involved in one of the central reactions in urea metabolism and the granulosa cells from the preovulatory stage had folic acid as unique metabolites involved in folate metabolism. These metabolites maybe used as identifiers after proper of cell stages in oocyte and granulosa.

3.5 Top Metabolites based on Metabolite Flow Intensity

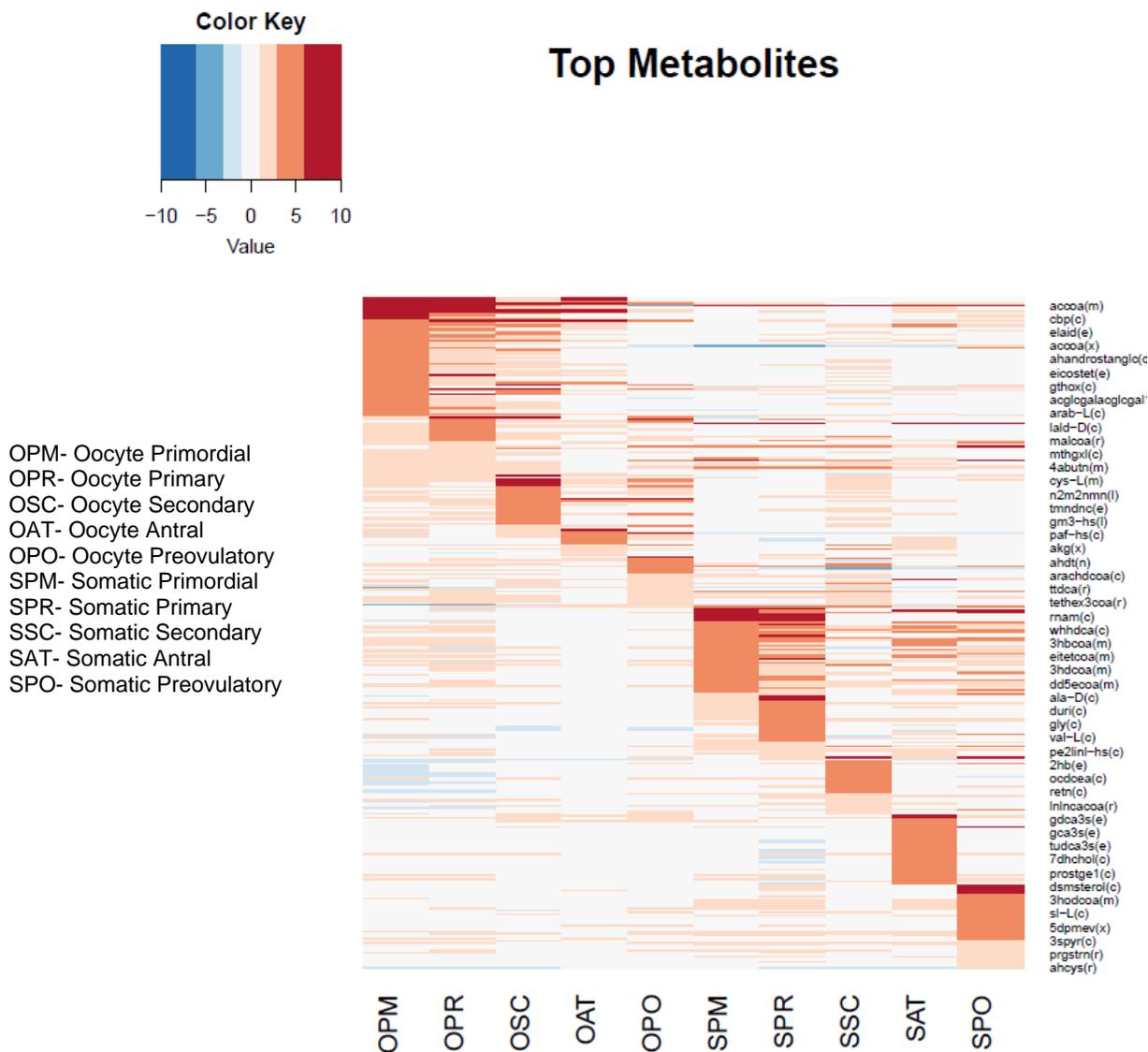


Figure 14: Top Metabolites in stage specific models based on Z score of normalized flow intensity through the metabolites.

The heatmap of Z score normalized flow intensity through metabolites revealed stage specific clustering of metabolites. The metabolites enriched in the primary and primordial models of both cell types were similar (Figure 14). The normalized flow intensity of the metabolites was lower in the oocyte antral and preovulatory stages which indicates that the initial stage of follicle presents higher transcriptional activity, whereas in number of metabolites in the granulosa cells were enriched throughout the follicle stages.

Oocytes are transcriptionally silent after the primordial primary stage, while the granulosa cells are reproducing during most of the stages of ovarian follicle development.

Coenzymes like Acyl-CoA (accoa), Malonyl-CoA (malcoa), DodecylCoA (dd5ecoa), Hydroxy-CoA (hbcoa), Eicosatetraenoyl-CoA (eitetcoa), Hydroxydodecyl-CoA (hdcoa), Eicosanoyl-CoA (arachdcoa), Tetracosahexaenoyl-CoA (tethexcoa), which are involved in the fatty acid oxidation, were highly enriched on the initial stages of follicle development in the oocyte. During maturation, lipids get accumulated in the form of droplets in oocyte which are later become an important and energy source after ovulation [9]. The fatty acid oxidation has been previously linked to development competence of oocyte in mice. Supplementing the culturing system with l-carnitine has increased the lipid metabolism and improved the development competence in oocyte [8]. Role of Co-enzymes in development competence enzymes has not been explored and could be an interesting area because many co-enzymes are involved in the fatty acid oxidation.

Amino acids like alanine, valine, glycine and cysteine were enriched in the granulosa metabolic model in the initial primordial and primary stages. Out of well-known steroid hormones that participate in ovarian follicle development, such as estrogen, androgen, testosterone, and progesterone, only progesterone was amount the top metabolites in the granulosa preovulatory model, which agrees with the literature. Estrogen, androgen, and testosterone were indeed produced by the granulosa cells but were not among the top metabolites from any stage.

3.6 Top Enzymes based on the Enzyme flow intensity.

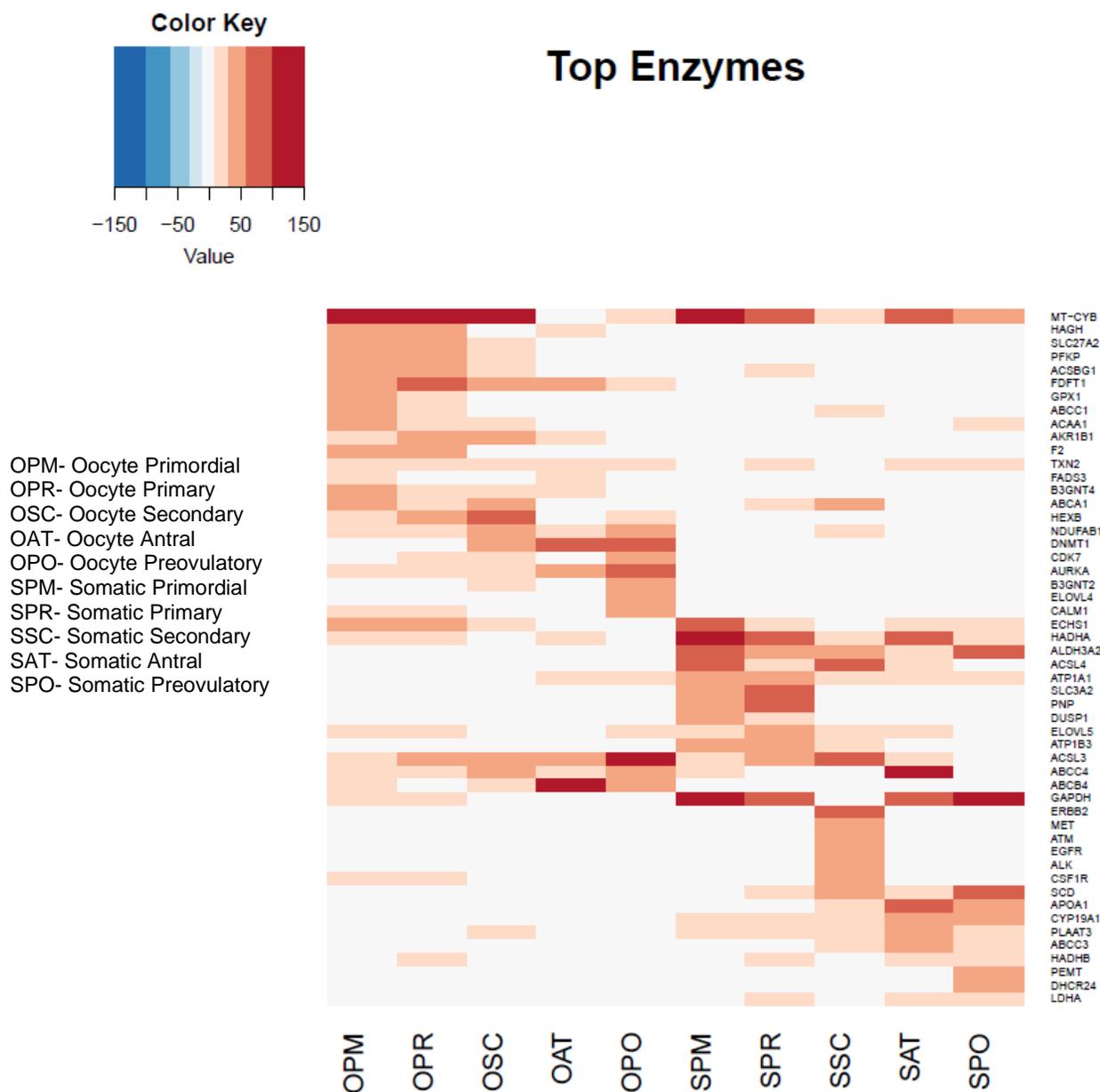


Figure 15: Top Metabolites in stage specific models based on Z score of normalized flow intensity through the metabolites.

MT-CYB gene encodes a mitochondrial enzyme which was prominent in all the stages and cell types except for the oocyte antral stage of follicle development. *MT-CYB* expression level is an indicator of oocyte quality. Compromised oocytes tend to have decreased levels of mitochondrial genes, [17] which may indicate that the quality of some antral oocytes from the transcriptomic data might be compromised as

observed in the heatmap in **Figure 5**. *CYP19A1* gene, which encodes for an aromatase enzyme essential for the productions of estrogen, had the highest elevated normalized flow at the later stages of folliculogenesis. *ALDH3A2*, which codes for aldehyde dehydrogenase that reduces toxic effect of acetaldehyde on cell viability by breaking down acetaldehyde generated during steroidogenesis. *ALDH3A2* was localized in granulosa cells. *LDHA* encoding for lactate dehydrogenase A enzyme and *GAPDH* encoding for the glyceraldehyde 3 phosphate was localized to granulosa cells. Both are enzyme involved in glycolysis. Genes like *NDUFAB1* and *PLAAT3* involved in fatty acid pathways are less explored in follicle development in human. *NDUFAB1* which is NADH: Ubiquinone Oxidoreductase enzyme which is a acyl carrier protein involved in the mitochondrial fatty acid oxidation was significant in all the oocyte model and has been previously explored in Chicken ovarian follicle but not in human. *PLAAT3* which codes for Phospholipase and acyltransferase 3 which is significant in Granulosa models.

3.7 Enriched Pathway in stage specific model

Next, we checked which pathways were significant at any given stage based on the normalized flow intensity through the genome-wide metabolic models. For instance, the oocyte had high normalized flow intensity through very well-known pathways in the oocyte like pyruvate metabolism, glycolysis, and pentose phosphate pathways. In granulosa cells, glycolysis pathway was the most active one compared to the other stage specific models. The granulosa cells in preovulatory metabolic model had high flow intensity through the glycolysis pathway as expected, because of the surge in the FSH before ovulation which increases the uptake of glucose and production of lactate. The presence of *LDHA* and *GADPH* enzymes in the most significant enzymes localized to granulosa cells further validates the possibility of glycolysis as a major energy in granulosa cells (**Figure 15**).

Fatty acid oxidation was higher in all the cell specific models except in oocyte of preovulatory follicle due to the high flow intensity (enrichment) of many Coenzymes A during follicle development (**Figure 14**).

Granulosa cells in the secondary follicles had consistently low normalized flow intensity. The glycolytic pathway had higher enrichment as compared to the oocyte model which is what was expected because granulosa cells are glycolytic function. Similarly, the major energy source in the oocyte model was pyruvate.

The folate metabolism is prominent in the initial stages of follicle development of mice and the same was observed in the follicle models of oocyte in human interestingly the granulosa model also showed significance in the folate metabolism which might be due to species specific difference in metabolism in mice [32].

Of note, none of the common active steroid pathways in the granulosa cells were enriched, yet the granulosa cells were able to produce them. These apparently contradictory observations are not, as the canonical metabolic pathways we commonly employed to describe cellular metabolism are a human-made division of the complex cellular metabolic pathways.

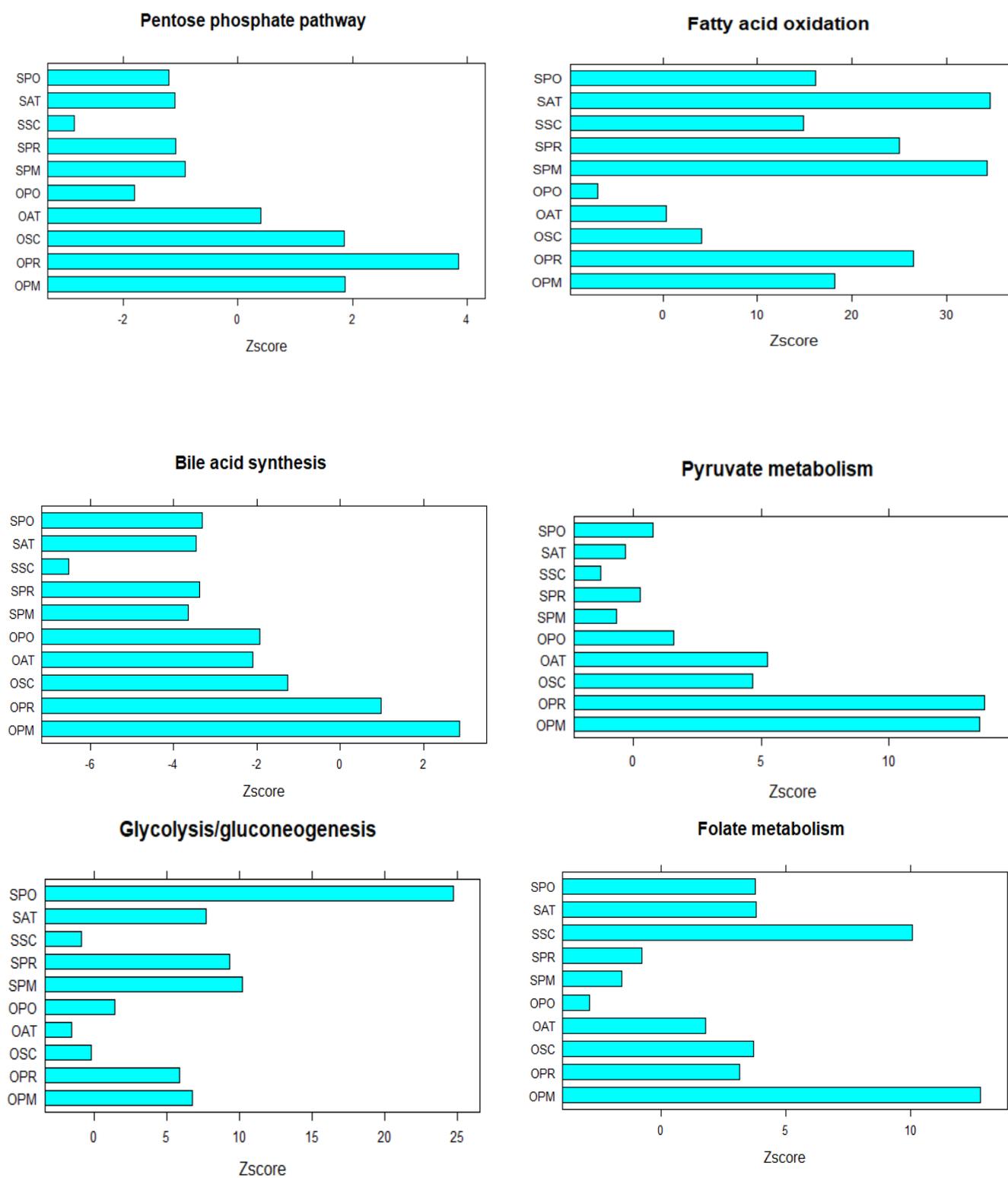


Figure 16: Some known pathways which were significant based on the Z scores of normalized flows intensity.

Chapter 4

Conclusion

Ovarian follicle development is a complex and dynamic process which is regulated at different cellular levels. While the follicle development has been studied at the transcriptomic level, there is limited information at the metabolic level. This may be due to the costs of metabolomics analysis, the difficulty to understand non-targeted metabolomics, the lack of relevant and significant biological samples available or lack of technology that allows to simulate the condition inside the body. Genome-wide scale metabolic models could be used to overcome such hurdles and make use of the available omics data to generate new data-driven hypothesis that can be tested in the laboratory to develop better technology. Although the global genome-wide scale models can be used to study and understand the metabolism in humans, specific difference in cellular function might be masked. Genome-wide scale context specific metabolic modelling has led to the fast and easy re-construction of cell specific metabolic models with the integration of available omics data such as transcriptomics, proteomics, or metabolomics.

In this study, we re-analyzed publicly available, single cell transcriptomic data from oocyte and granulosa cell from five different stages of follicle development to create cell-specific genome-wide metabolic models of oocyte and granulosa cells at each stage during ovarian follicle development. The most comprehensive and latest reconstruction of human (Recon3D) was employed to model the metabolism in the cell types. We have created the first human follicle model that consist of all the plausible metabolic reactions that could occur during follicle development. We also created stage specific metabolic models of oocyte and granulosa to study the difference in metabolism at different stages (primordial, primary, secondary, antral and preovulatory) and cell types during follicle development. Our ovarian follicle model is relevant as it could simulate 276 essential human metabolic functions for survival and growth (biomass reactions, ATP production from glucose) and reactions specific to the follicle development like secretion of progesterone

and estrogen, production of androgen, to name few. To study the difference in the metabolism at every stage of follicle development, we used stage specific models to identify the differences in the metabolites present at each stage of follicle development. We could observe that there were lot of metabolites that were unique to specific stages. Using a community detection algorithm, we obtained the flow of information through the network of metabolites and enzymes. There were several known pathways for energy which were significant based on the normalized flow intensity in the models like pyruvate metabolism, fatty acid oxidation, glycolysis, and Pentose phosphate pathway. The fatty acid oxidation and synthesis were highly active in all the models and since the lipid metabolism is still unknown during each stage of follicle metabolism makes it a good place to start for future studies. We also obtained top metabolites and enzymes that are active in each stage of follicle development. Many of the top metabolites obtained were Co-enzymes which again points towards the lipid metabolism. The enzymes like *LDHA* and *GADPH* can be used for validation of future in silico models of follicle development with different datasets as it has been studied previously in the mammals [9].

4.1 Future perspectives

Genome-wide context specific metabolic models have great potential to understand metabolism at the cellular level. The advancement of high throughput analysis combined with efficient algorithms to integrate multiomics data with the genome scale metabolic models has led to easy and fast reconstruction of metabolic relevant models. These models can be employed to shed light into the metabolomic behavior of the cells in different condition and can be employed to design new culture media according to the consumption pattern in the cell.

There are several limitations in our study. For instance, although all the ovaries appeared histologically healthy, the donors were preserving their tissue for fertility preservation due to cancer, tumor, or sex reassignment (which required testosterone treatment). Thus, it is plausible that some of the samples do not represent healthy follicles. It would have been ideal to have samples from healthy donors. The model

generated from these data could simulate some of the well-known metabolic pathways during ovarian follicle development like glycolysis, pyruvate metabolism, fatty acid oxidation, but integration of proteomics and metabolomic data will enhance the model further in terms of prediction of key metabolites consumed and secreted. Also, other cell types available in the ovarian tissue needs to be added in the model to make a complete follicle model at each stage of follicle development, including theca cells, mural and cumulus granulosa cells.

Genome-wide models of ovarian follicle development can be used to study the diseased state like PCOS in which genes involved in steroid hormones are affected which leads to excess production of androgens. These conditions can be simulated by knocking out genes responsible for PCOS and checking its effect on the pathways and members of our laboratory are currently exploring this application.

In future, the use of machine learning and artificial intelligence combined with high-throughput sequencing could pave the way to build efficient and personalized culture systems for IVF and IVM and increase the success rate of pregnancy in women with infertility.

References:

1. Adhikari, Deepak, and Kui Liu. "Regulation of Quiescence and Activation of Oocyte Growth in Primordial Follicles." *Oogenesis*, Springer London, 2012, pp. 49–62, doi:10.1007/978-0-85729-826-3_4.
2. Angione, C. (2019). Human Systems Biology and Metabolic Modelling: A Review—From Disease Metabolism to Precision Medicine. *BioMed Research International*, 2019, 1-16. doi:10.1155/2019/8304260
3. Brunk, E., Sahoo, S., Zielinski, D. et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* 36, 272–281 (2018). <https://doi.org/10.1038/nbt.4072>
4. Chen, Yao, et al. "The Factors and Pathways Regulating the Activation of Mammalian Primordial Follicles in Vivo." *Frontiers in Cell and Developmental Biology*, vol. 8, Frontiers Media S.A, 2020, pp. 575706–575706, doi:10.3389/fcell.2020.575706.
5. Cordeiro, Marilia Henriques, et al. "Chapter 1 - Ovarian Follicle Biology and the Basis for Gonadotoxicity." *Cancer Treatment and the Ovary*, Elsevier Inc, 2016, pp. 3–20, doi:10.1016/B978-0-12-801591-9.00001-1.
6. Davis H.C., Hackney A.C. (2017) The Hypothalamic–Pituitary–Ovarian Axis and Oral Contraceptives: Regulation and Function. In: Hackney A. (eds) *Sex Hormones, Exercise and Women*. Springer, Cham. https://doi.org/10.1007/978-3-319-44558-8_1
7. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PMID: 23104886; PMCID: PMC3530905.
8. Dunning, K. R., Akison, L. K., Russell, D. L., Norman, R. J., & Robker, R. L. (2011). Increased beta-oxidation and improved oocyte developmental competence in response to l-carnitine during ovarian in vitro follicle development in mice. *Biology of reproduction*, 85(3), 548–555. <https://doi.org/10.1095/biolreprod.110.090415>

9. Collado-Fernandez, E., Picton, H. M., & Dumollard, R. (2012). Metabolism throughout follicle and oocyte development in mammals. *The International journal of developmental biology*, 56(10-12), 799–808. <https://doi.org/10.1387/ijdb.120140ec>
10. Faddy MJ, Gosden RG. A model conforming the decline in follicle numbers to the age of menopause in women. *Hum Reprod*. 1996 Jul;11(7):1484-6. doi: 10.1093/oxfordjournals.humrep.a019422. PMID: 8671489.
11. Gaytan, F., Morales, C., Roa, J., & Tena-Sempere, M. (2018). Changes in keratin 8/18 expression in human granulosa cell lineage are associated to cell death/survival events: potential implications for the maintenance of the ovarian reserve. *Human reproduction (Oxford, England)*, 33(4), 680–689. <https://doi.org/10.1093/humrep/dey010>
12. Goldman S, Shalev E. MMPS and TIMPS in ovarian physiology and pathophysiology. *Front Biosci*. 2004 Sep 1;9:2474-83. doi: 10.2741/1409. PMID: 15353300.
13. Griffin J, Emery BR, Huang I, Peterson CM, Carrell DT. Comparative analysis of follicle morphology and oocyte diameter in four mammalian species (mouse, hamster, pig, and human). *J Exp Clin Assist Reprod*. 2006 Mar 1;3:2. doi: 10.1186/1743-1050-3-2. PMID: 16509981; PMCID: PMC1413548.
14. Gu, C., Kim, G.B., Kim, W.J. et al. Current status and applications of genome-scale metabolic models. *Genome Biol* 20, 121 (2019). <https://doi.org/10.1186/s13059-019-1730-3>
15. Heirendt, L., Arreckx, S., Pfau, T. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 14, 639–702 (2019). <https://doi.org/10.1038/s41596-018-0098-2>
16. Hemmings, K. E., Maruthini, D., Vyjayanthi, S., Hogg, J. E., Balen, A. H., Campbell, B. K., Leese, H. J., & Picton, H. M. (2013). Amino acid turnover by human oocytes is influenced by gamete developmental competence, patient characteristics and gonadotrophin treatment. *Human reproduction (Oxford, England)*, 28(4), 1031–1044. <https://doi.org/10.1093/humrep/des458>

17. Hsieh RH, Au HK, Yeh TS, Chang SJ, Cheng YF, Tzeng CR. Decreased expression of mitochondrial genes in human unfertilized oocytes and arrested embryos. *Fertil Steril*. 2004 Mar;81 Suppl 1:912-8. doi: 10.1016/j.fertnstert.2003.11.013. PMID: 15019829.
18. Jankowska K. (2017). Premature ovarian failure. *Przegląd menopauzalny = Menopause review*, 16(2), 51–56. <https://doi.org/10.5114/pm.2017.68592>
19. Karen E. Hemmings, Henry J. Leese, Helen M. Picton, Amino Acid Turnover by Bovine Oocytes Provides an Index of Oocyte Developmental Competence In Vitro, *Biology of Reproduction*, Volume 86, Issue 5, 1 May 2012, 165, 1–12, <https://doi.org/10.1095/biolreprod.111.092585>
20. Kenigsberg, S., Lima, P. D., Maghen, L., Wyse, B. A., Lackan, C., Cheung, A. N., Tsang, B. K., & Librach, C. L. (2017). The elusive MAESTRO gene: Its human reproductive tissue-specific expression pattern. *PloS one*, 12(4), e0174873. <https://doi.org/10.1371/journal.pone.0174873>
21. Kidder GM, Vanderhyden BC. Bidirectional communication between oocytes and follicle cells: ensuring oocyte developmental competence. *Can J Physiol Pharmacol*. 2010 Apr;88(4):399-413. doi: 10.1139/y10-009. PMID: 20555408; PMCID: PMC3025001.
22. Kishi, H, Kitahara, Y, Imai, F, Nakao, K, Suwa, H. Expression of the gonadotropin receptors during follicular development. *Reprod Med Biol*. 2018; 17: 11– 19. <https://doi.org/10.1002/rmb2.12075>
23. Larose H, Shami AN, Abbott H, Manske G, Lei L, Hammoud SS. Gametogenesis: A journey from inception to conception. *Curr Top Dev Biol*. 2019; 132:257-310. doi: 10.1016/bs.ctdb.2018.12.006. Epub 2019 Jan 8. PMID: 30797511; PMCID: PMC7133493.
24. Love, M.I., Huber, W., Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15:550. 10.1186/s13059-014-0550-8.
25. Laurent Heirendt & Sylvain Arreckx, , Costas D. Maranas, Nathan E. Lewis, Thomas Sauter, Bernhard Ø. Palsson, Ines Thiele, Ronan M.T. Fleming, **Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0**, *Nature Protocols*, volume 14, pages 639–702, 2019 doi.org/10.1038/s41596-018-0098-2.

26. Macklon NS, Fauser BC. Follicle-stimulating hormone and advanced follicle development in the human. *Arch Med Res.* 2001 Nov-Dec;32(6):595-600. doi: 10.1016/s0188-4409(01)00327-7. PMID: 11750735.
27. Mikhael S, Punjala-Patel A, Gavrilova-Jordan L. Hypothalamic-Pituitary-Ovarian Axis Disorders Impacting Female Fertility. *Biomedicines.* 2019;7(1):5. Published 2019 Jan 4. doi:10.3390/biomedicines7010005.
28. Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10-12, May 2011. doi: <http://dx.doi.org/10.14806/ej.17.1.200>.
29. Nielsen, Jens. “Genome-Scale Metabolic Models Applied to Human Health and Disease.” *Wiley Interdisciplinary Reviews.*, vol. 9, no. 6, John Wiley & Sons, 2017, doi:10.1002/wsbm.1393.
30. Palsson, Bernhard Ø., et al. “What Is Flux Balance Analysis?” *Nature Biotechnology*, vol. 28, no. 3, Nature Publishing Group, 2010, pp. 245–48, doi:10.1038/nbt.1614.
31. Pangas SA, Choi Y, Ballow DJ, Zhao Y, Westphal H, Matzuk MM, et al. Oogenesis requires germ cell-specific transcriptional regulators *Sohlh1* and *Lhx8*. *Proc Natl Acad Sci.* 2006;103:8090–8095. doi: 10.1073/pnas.0601083103.
32. Peñalver Bernabé, B., Thiele, I., Galdones, E., Siletz, A., Chandrasekaran, S., et al (2019). Dynamic genome-scale cell-specific metabolic models reveal novel inter-cellular and intra-cellular metabolic communications during ovarian follicle development. *BMC bioinformatics*, 20(1), 307. <https://doi.org/10.1186/s12859-019-2825-2>
33. Robaina Estévez, Semidán, and Zoran Nikoloski. “Generalized Framework for Context-Specific Metabolic Model Extraction Methods.” *Frontiers in Plant Science*, vol. 5, Frontiers Media S.A, 2014, pp. 491–491, doi:10.3389/fpls.2014.00491.
34. Rosvall, M., Axelsson, D. & Bergstrom, C. The map equation. *Eur. Phys. J. Spec. Top.* 178, 13–23 (2009). <https://doi.org/10.1140/epjst/e2010-01179-1>

35. Richelle, Anne, et al. “Increasing Consensus of Context-Specific Metabolic Models by Integrating Data-Inferred Cell Functions.” *PLoS Computational Biology*, vol. 15, no. 4, Public Library of Science, 2019, pp. e1006867–e1006867, doi:10.1371/journal.pcbi.1006867.
36. Richards, Joanne S., and Stephanie A. Pangas. “The Ovary: Basic Biology and Clinical Implications.” *The Journal of Clinical Investigation*, vol. 120, no. 4, American Society for Clinical Investigation, 2010, pp. 963–72, doi:10.1172/JCI41350.
37. Simon Anders, Paul Theodor Pyl, Wolfgang Huber HTSeq — A Python framework to work with high-throughput sequencing data *Bioinformatics* (2014), in print, online at doi:10.1093/bioinformatics/btu638
38. Tetsuka, M., & Tanakadate, M. (2019). Activation of HSD11B1 in the bovine cumulus-oocyte complex during IVM and IVF. *Endocrine connections*, 8(7), 1029–1039. <https://doi.org/10.1530/EC-19-0188>
39. Vlassis, Nikos, et al. “Fast Reconstruction of Compact Context-Specific Metabolic Network Models.” *PLoS Computational Biology*, vol. 10, no. 1, Public Library of Science, 2014, pp. e1003424–e1003424, doi:10.1371/journal.pcbi.1003424.
40. Zhang, Cheng, and Qiang Hua. “Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine: Application of GEMs.” *Frontiers in Physiology*, vol. 6, no. January 2016, doi:10.3389/fphys.2015.00413.
41. Zhang Y, Yan Z, et al. Transcriptome Landscape of Human Folliculogenesis Reveals Oocyte and Granulosa Cell Interactions. *Mol Cell*. 2018 Dec 20;72(6):1021-1034.e4. doi: 10.1016/j.molcel.2018.10.029. Epub 2018 Nov 21. PMID: 30472193.

APPENDIX

File A: Human Metabolic Reactions and Metabolites

File B: Test for 295 Human Metabolic Functions

VITA

Ashwin Koppayi

EDUCATION	Master of Science in Bioengineering University of Illinois at Chicago	2019-2021
	Bachelor of Engineering in Biotechnology University of Mumbai	2014-2018
BIOINFORMATICS SKILLS	RNA Sequence Analysis, Metagenomic Sequence Analysis, Machine Learning, Biostatistics, Mathematical Modeling	
PROGRAMMING SKILLS	R, MATLAB, Python, SQL	
TOOLS	FastQC, Cutadapt, STAR, TopHat, HTSeq, Deseq2, GSEA, CobraToolBox, IGV	