

# **Lineage-restricted metabolic identities of mammary epithelial cells**

By

Mathepan Mahendralingam

A thesis submitted in conformity with the requirements  
for the degree of Master of Science

Department of Medical Biophysics

University of Toronto

© Copyright by Mathepan Mahendralingam 2019

# **Lineage-restricted metabolic identities of mammary epithelial cells**

Mathepan Mahendralingam

Master of Science

Department of Medical Biophysics  
University of Toronto

2019

## **Abstract**

Cancer metabolism has been proposed to adapt the metabolic network from its tissue-of-origin. However, breast cancer is not a disease of a singular origin. There are multiple epithelial subpopulations that serve as the culprit cell-of-origin, giving rise to distinct breast cancer subtypes. Knowledge surrounding the metabolic identities of normal mammary epithelial cells (MEC) is limited. Proteomic profiling of primary FACS-purified human MECs revealed that each subpopulation possesses distinct metabolic networks. Luminal progenitors were enriched for electron transport chain subunits, had an enhanced capacity to undergo oxidative phosphorylation and were vulnerable to Complex I inhibition. Basal cells were more glycolytic, but their progenitor capacity was still dependent on mitochondrial activity. Targeting these pathways with inhibitors exposed lineage-restricted metabolic vulnerabilities. Furthermore, breast cancer subtypes demonstrated significant enrichment for the metabolic cluster of specific MEC. My work demonstrates that normal MEC have lineage-restricted metabolic identities, which can partly explain the metabolic heterogeneity observed in breast cancer.

## Acknowledgments

In the past three years, I have met several incredible people who have taught me many invaluable lessons that helped me develop personally and as a scientist. I want to thank Dr. Rama Khokha for giving me the chance to join her lab. I'm grateful that she let me pursue whatever avenue of metabolism research I wanted. Looking back, I wish I realized earlier that it was a stupid plan to think I could answer all the unsolved questions in mammary metabolism. But I'm glad Rama let me try, since I was free to be curious and further mature as a scientist. I'm forever grateful that she let me volunteer at PMH and has continually supported my goals of becoming a clinician-scientist. I also want to thank Dr. Sean Egan and Dr. Aaron Schimmer for providing me their guidance and insight, which has helped tremendously throughout my Master's degree.

I want to thank everyone in the Khokha Lab. Especially, Dr. Alison Casey for teaching me everything and being my mentor. I always appreciate you cheering me up whenever I messed up my experiments for the billionth time. In addition, I want to thank Pirashaanthy for making me more critical of scientific literature. Hyeyeon for putting up with all my terrible jokes, my dumb questions about Illustrator and coming with me to set up my flow experiments. My Master's degree would not have been possible without the help of these three incredibly intelligent and talented individuals. Lastly, I need to thank Kazeera Aliar for her brilliant bioinformatics skills. Grad school would not have been a memorable experience without all the amazing friends that I have made. Thank you for all the laughs and listening to me marvel over the cell-of-origin and mitochondria.

# Table of Contents

|  |           |
|--|-----------|
| Acknowledgments.....   | iii       |
| Table of Contents .....  | iv        |
| List of Tables .....   | vii       |
| List of Figures .....  | viii      |
| Abbreviations .....  | x         |
| <b>Chapter 1 : Background.....</b>                               | <b>1</b>  |
| 1.1 Mammary Gland Biology .....                                  | 1         |
| 1.1.1 Mammary Epithelial Hierarchy .....                         | 1         |
| 1.1.2 Inherent properties of mammary epithelial cells.....       | 3         |
| 1.1.3 Cell of origin for breast cancer .....                     | 5         |
| 1.1.4 Influence of Normal on Tumorigenesis .....                 | 7         |
| 1.1.5 Breast Cancer Subtypes .....                               | 8         |
| 1.2 Basic Metabolism and Relevance to Cancer.....                | 9         |
| 1.2.1 Glycolysis and Warburg Effect.....                         | 10        |
| 1.2.2 Tricarboxylic Acid (TCA) Cycle and Cancer Initiation ..... | 11        |
| 1.2.3 Electron Transport Chain (ETC).....                        | 14        |
| 1.2.4 Normal Mammary Epithelial Cell Metabolism .....            | 18        |
| 1.2.5 Breast Cancer Metabolism.....                              | 18        |
| 1.2.6 Influence of normal on cancer metabolism.....              | 19        |
| 1.3 Thesis Outline .....   | 21        |
| 1.3.1 Rationale and hypothesis .....                             | 21        |
| 1.3.2 Objectives.....  | 21        |
| <b>Chapter 2 .....</b>   | <b>22</b> |
| 2 Methodology .....  | 22        |

|                  |  |           |
|------------------|--|-----------|
| 2.1              | Human mammary sample preparation, single cell preparation, and FACS.....               | 22        |
| 2.2              | Proteomics on FACS-purified human mammary subpopulations .....                         | 23        |
| 2.3              | Bioinformatics Analysis of human mammary subpopulation proteomes.....                  | 24        |
| 2.3.1            | Total Proteome Analyses .....  | 24        |
| 2.3.2            | Generating Metabolic Signatures.....   | 25        |
| 2.3.3            | Pathway Analysis using Enrichr .....   | 27        |
| 2.3.4            | Correlations to PAM50 Breast Cancer Subtypes .....                                     | 27        |
| 2.4              | Mouse Strains.....   | 28        |
| 2.5              | Mouse Mammary gland dissociation, single-cell preparation and FACS analysis .....      | 28        |
| 2.6              | Measuring Oxygen Consumption Rate .....  | 29        |
| 2.7              | Intracellular flow cytometry.....  | 30        |
| 2.8              | Transmission electron microscopy (TEM) .....   | 30        |
| 2.9              | In vitro colony forming cell (CFC) assay .....   | 30        |
| 2.10             | Drug testing in vitro .....  | 31        |
| 2.11             | Statistical Analysis .....   | 31        |
| <b>Chapter 3</b> | .....  | <b>32</b> |
| <b>3</b>         | <b>Results .....</b>   | <b>32</b> |
| 3.1              | Proteomics reveals differential expression of metabolic proteins .....                 | 33        |
| 3.2              | Mammary subsets have distinct metabolic preferences .....                              | 36        |
| 3.3              | Mouse LP have enhanced capacity to undergo OXPHOS .....                                | 42        |
| 3.4              | Mitochondrial structure and function are MEC-specific .....                            | 43        |
| 3.5              | BC progenitor activity is dependent upon mitochondrial activity .....                  | 45        |
| 3.6              | MEC demonstrate ETC complex specific vulnerabilities.....                              | 46        |
| 3.7              | BC are vulnerable to glycolytic inhibitors .....                                       | 48        |
| 3.8              | Metabolic profiles of mammary progenitors are mirrored in breast cancer subtypes ..... | 49        |
| <b>Chapter 4</b> | .....  | <b>53</b> |

|                                     |    |
|-------------------------------------|----|
| 4 <b>Conclusion</b> .....           | 53 |
| 4.1 Discussion & Significance ..... | 53 |
| 4.2 Future Directions.....          | 54 |
| 4.3 Limitations .....               | 55 |
| References.....                     | 57 |
| Appendix 1 .....                    | 68 |

## List of Tables

|  |    |
|--|----|
| Table 2.1 – Summary of clinical covariates for the cohort of patients used in this study. .... | 22 |
| Table 2.2 – Mature Luminal Metabolic Cluster .....   | 26 |
| Table 2.3 – Basal Metabolic Cluster .....  | 26 |
| Table 2.4 – Luminal Progenitor Metabolic Cluster .....   | 27 |

## List of Figures

|   |    |
|---|----|
| Figure 1. 1 – Gross morphology and cellular composition of the human mammary gland. ....  | 2  |
| Figure 1. 2 – The mammary epithelial hierarchy and its relationship to putative cells-of-origin for breast cancers .....                                    | 6  |
| Figure 1. 3 - Schematic for Glycolysis and Tricarboxylic Acid (TCA) Cycle. ....   | 13 |
| Figure 1. 4 – Overview of oxidative phosphorylation (OXPHOS) .....  | 16 |
| Figure 2.1 - Gating strategy for FACS-purifying human mammary epithelial cells .....  | 23 |
| Figure 2.2 - Gating strategy for FACS-purifying mouse mammary epithelial cells .....  | 29 |
| Figure 3.1 – Workflow summarizing tissue preparation, FACS-purification, and summary of our proteomics done on primary human mammary epithelial cells ..... | 33 |
| Figure 3.2 – Characterization of proteomic datasets of primary FACS-purified human mammary epithelial cells .....   | 34 |
| Figure 3.3 – Protein distribution plots demonstrate MEC are enriched for unique functional classes of proteins .....  | 35 |
| Figure 3.4 – Human MEC have lineage-restricted metabolic networks .....   | 37 |
| Figure 3.5 – LP show enrichment for TCA Cycle and ETC proteins, whereas ML and BC are enriched more for glycolysis.....                                     | 38 |
| Figure 3.6 –Map of mammary epithelial cell metabolism .....   | 41 |
| Figure 3.7 – Luminal populations have an enhanced capacity to undergo oxygen consumption. ....  | 42 |
| Figure 3.8- Mitochondrial structure and function is mammary cell-type specific .....  | 44 |
| Figure 3.9 – Mammary progenitor activity is dependent on mitochondrial activity .....   | 45 |

Figure 3.10 – Mammary progenitors display lineage-restricted vulnerabilities to OXPHOS inhibitors ..... 47

Figure 3.11 – BC are more sensitive to glycolytic inhibitors ..... 48

Figure 3.12 – Breast cancer subtypes demonstrate selective activity of their respective cell-of-origin’s metabolic cluster..... 51

Figure 3.13 – Copy-number alterations of potential cell-of-origin specific metabolic targets..... 52

Figure 4.1 – Lactate crosstalk and progesterone induced autophagy ..... 68

## Abbreviations

|  |   |
|--|---|
| <b>2-HG:</b> 2-hydroxyglutarate  | <b>FCCP:</b> Carbonyl cyanide-p-trifluoromethoxyphenylhydrazone |
| <b>3PG:</b> 3-phosphoglycerate   | <b>FGF:</b> Fibroblast Growth Factor                            |
| <b>5mC:</b> 5-methylcytosine   | <b>FH:</b> fumarate hydratase                                   |
| <b>A5:</b> Atpenin A5  | <b>FMN:</b> Flavin mononucleotide                               |
| <b>AA:</b> Antimycin A   | <b>GALNT2:</b> Polypeptide N-Acetylgalactosaminyltransferase 2  |
| <b>ADP:</b> Adenosine diphosphate  | <b>G3P:</b> glycerol 3-phosphate                                |
| <b>ATP:</b> Adenosine triphosphate   | <b>G6P:</b> Glucose-6-phosphate                                 |
| <b>BC:</b> Basal Cell  | <b>GADP:</b> Glyceraldehyde 3-phosphate                         |
| <b>CI:</b> Complex I   | <b>GDP:</b> Guanosine diphosphate                               |
| <b>CII:</b> Complex II   | <b>GTP:</b> Guanosine triphosphate                              |
| <b>CIII:</b> Complex II  | <b>HER2:</b> Human Epidermal Growth Factor Receptor 2           |
| <b>CIV:</b> Complex IV   | <b>HR:</b> Hormone Receptor                                     |
| <b>CoQ:</b> Coenzyme Q <sub>10</sub>   | <b>HSC:</b> Hematopoietic Stem Cell                             |
| <b>CS:</b> Citrate Synthase  | <b>IDH:</b> Isocitrate dehydrogenase                            |
| <b>CV:</b> Complex V   | <b>IF:</b> Immunofluorescence                                   |
| <b>CytoC:</b> Cytochrome C   | <b>IMM:</b> Inner mitochondrial membrane                        |
| <b>CYB5R1:</b> Cytochrome B5 Reductase 1                                     | <b>ISC:</b> Intestinal Stem cell                                |
| <b>E:</b> 17 $\beta$ -estradiol  | <b>K14:</b> Keratin 14 (Basal)                                  |
| <b>EGF:</b> Epidermal Growth Factor  | <b>K18:</b> Keratin 18 (Luminal)                                |
| <b>EP:</b> 17 $\beta$ -estradiol + progesterone                              | <b>K5:</b> Keratin 5 (Basal)                                    |
| <b>EPHX1:</b> Epoxide Hydrolase 1  | <b>K8:</b> Keratin 8 (Luminal)                                  |
| <b>ER:</b> Estrogen Receptor   | <b>KMO:</b> Kynurenine 3-Monooxygenase                          |
| <b>ESC:</b> Embryonic Stem cell  | <b>LDH:</b> Lactate Dehydrogenase                               |
| <b>ETC:</b> Electron Transport Chain   | <b>LP:</b> Luminal Progenitor                                   |
| <b>F16BP:</b> Fructose 1,6-bisphosphate                                      | <b>LTR:</b> LysoTracker Red                                     |
| <b>F6P:</b> Fructose 6-phosphate   | <b>MaSC:</b> Mammary Stem Cell                                  |
| <b>FACS:</b> Fluorescent-Activated Cell Sorting                              | <b>MEC:</b> Mammary Epithelial Cell                             |
| <b>FADH/FADH<sub>2</sub> (oxidized/reduced):</b> Flavin Adenine Dinucleotide |   |

**ML:** Mature Luminal

**mROS:** Mitochondrial Reactive Oxygen Species

**mtDNA:** Mitochondrial Deoxy ribonucleotide

**MTG:** MitoTracker Green

**MTR:** MitoTracker Red

**NAD<sup>+</sup>/NADH (oxidized/reduced):**

Nicotinamide Adenine Dinucleotide

**NADP<sup>+</sup>/NADPH (oxidized/reduced):**

Nicotinamide Adenine Dinucleotide

Phosphate

**NIT1:** Nitrilase-like protein 1

**OL:** Oligomycin

**OMM:** Outer Mitochondrial Membrane

**OXPPOS:** Oxidative Phosphorylation

**PEP:** Phosphoenolpyruvate

**PFK:** Phosphofructokinase

**PHGDH:** Phosphoglycerate Dehydrogenase

**P<sub>i</sub>:** Inorganic Phosphate

**PK:** Pyruvate Kinase

**PPP:** Pentose Phosphate Pathway

**PR:** Progesterone Receptor

**RISP:** Rieske Iron-Sulfur Protein

**Rot:** Rotenone

**SDH:** Succinate Dehydrogenase

**TCA:** Tricarboxylic

**TDLU:** Terminal Ductal Lobular Unit

**TEB:** Terminal End Bud

**TEM:** Transmission Electron Microscopy

**TET:** Ten-Eleven Translocation

**UPLC-MS:** Ultra-high Pressure Liquid Chromatography/Mass Spectrometry

**$\alpha$ KG:**  $\alpha$ -Ketoglutarate

# Chapter 1 : Background

**This chapter contains a modified figure and excerpts from a published review:**

Tharmapalan, P., Mahendralingam, M., Berman, H. K., & Khokha, R. (2019). Mammary stem cells and progenitors: targeting the roots of breast cancer for prevention. *The EMBO Journal*.

## 1.1 Mammary Gland Biology

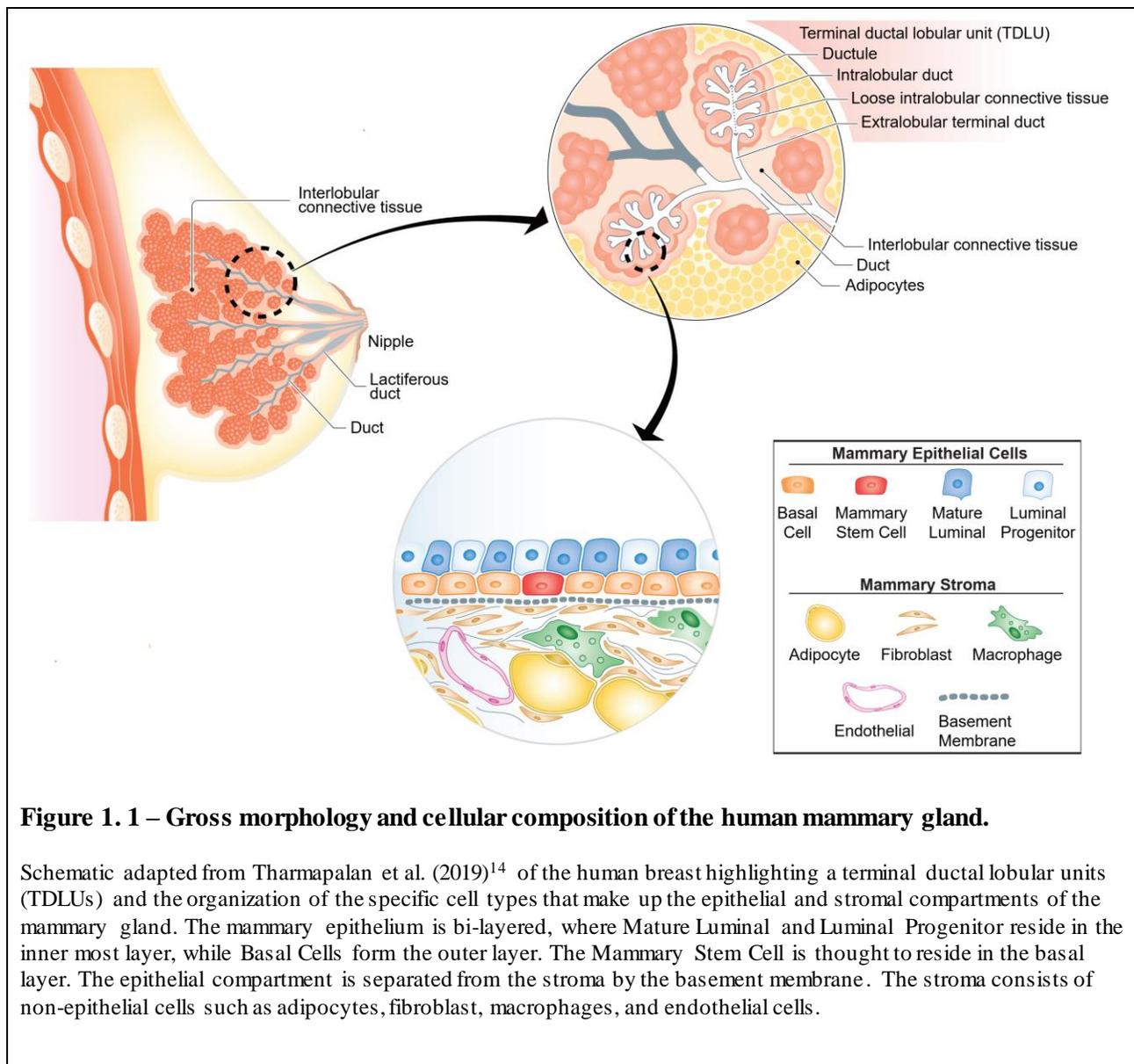
The mammary gland's essential function is to produce milk for newborn infants. The mammary gland has a unique developmental timeline. It sprouts and invades into the mammary fat pad during early embryogenesis, but then halts its growth till the onset of puberty<sup>1,2</sup>. Under the influence of estrogen, it undergoes ductal elongation and branching to fill the fat pad<sup>1,2</sup>. Repetitive cycles of fluctuating hormones, seen in key physiological events like the reproductive cycle and pregnancy, induce expansion and regression of the mammary gland<sup>3,4</sup>. The mammary gland forms a highly organized ductal tree<sup>5</sup>. In humans, radially branching ducts end in functional pyramidal lobules termed "terminal ductal lobulo-alveolar units" (TDLUs), which ultimately differentiate into milk-secreting acini during lactation. TDLUs are thought of as the site of origin for the majority of human breast cancers<sup>6,7</sup> and contain proliferative stem and progenitor cells<sup>8</sup>. The mouse mammary ductal tree is similar with their own lobuloalveolar structures. Thus, making the mouse a suitable model to study breast biology that is applicable to its human counterpart<sup>5</sup>.

### 1.1.1 Mammary Epithelial Hierarchy

The bilayered mammary gland is composed of two epithelial lineages: basal and luminal<sup>1,2,9</sup>. Both mammary epithelial cells (MEC) function in unison in order to carry out the overall function of the breast. The inner layer facing the lumen is composed of luminal cells, which are responsible for secreting milk<sup>1,2,9</sup>. The outer layer consists of basal cells (BC), which are contractile<sup>1,2,9</sup>. The high regenerative potential of the mammary gland is owed to its diverse pools of stem and progenitor populations found within each lineage<sup>10</sup>. The luminal lineage is composed of two cell types: mature luminal (ML) and luminal progenitor (LP) cells<sup>11</sup>. Despite being highly responsive to hormones, reports suggest that only 30-50% of mammary epithelial

cells express the hormone receptors (HR), estrogen receptor (ER) or progesterone receptor (PR)<sup>11,12</sup>. Mature Luminal (ML) are the only HR+ cells in the gland<sup>11</sup>. This population responds to hormonal cues and relays potent paracrine effectors to the HR- stem/progenitor populations to cause them to proliferate<sup>13,14</sup>. The basal lineage is largely thought to be devoid of HR.

The list of cell surface markers used for fluorescent-activated cell sorting (FACS) have not substantially changed since 2006<sup>10,15</sup>. Human BC can be delineated by EpCAM<sup>-</sup>CD49<sup>hi</sup>, EpCAM<sup>+</sup>CD49<sup>lo</sup> for ML, and EpCAM<sup>+</sup>CD49<sup>hi</sup> for LP after excluding lineage (Lin) positive (CD31<sup>+</sup>Ter119<sup>+</sup>CD45<sup>+</sup>) and dead cells (Figure 2.1). Gating of mouse MEC is similar, except



there is little separation achieved between the two luminal populations when EpCAM and CD49f are used (Figure 2.2). CD49b and Sca1 are therefore used in the mouse to distinguish LP (CD49b<sup>+</sup>Sca1<sup>-</sup>) and ML (CD49b<sup>+</sup>Sca1<sup>+</sup>, CD49b<sup>-</sup>Sca1<sup>+</sup> or CD49b<sup>-</sup>Sca1<sup>-</sup>)<sup>11</sup>. In addition to these FACS markers, there are defined markers for the mammary subpopulations that demonstrate bimodal expression in the gland<sup>6</sup>. Luminal cells express cytokeratin 8 (K8) and K18. The LPs can also be identified by KIT and ALDH expression<sup>11</sup>. In addition to expressing K5, K14, and  $\alpha$ SMA, BC are known to express mesenchymal markers such as vimentin<sup>16</sup>.

Following orthotopic transplantation *in vivo*, only BC possess the functional capacity to generate a full ductal tree, with the ability to both self-renew and contribute to all subsequent MEC lineages<sup>10,15</sup>. For this reason, it is believed that the mammary stem cell (MaSC) resides in the basal compartment. This has led to labeling of the EpCAM<sup>-</sup>CD49f<sup>hi</sup> population as “mammary stem cell-enriched,” even though this population also contains basal progenitors and differentiated basal cells<sup>16</sup>. To resolve this heterogeneity, several markers have been proposed to purify rare MaSC populations such as Lgr5<sup>17-19</sup>, Procr<sup>20</sup>, Tspan8<sup>21</sup>, Bcl11b<sup>22</sup> and Dll1<sup>23</sup>. Although each marker enriches for refined populations with enhanced transplantation efficiency, an exclusive MaSC signature remains elusive.

In addition, the organization of the mammary epithelial hierarchy is under constant scrutiny. There are two prevailing viewpoints. One is that there is a bipotent stem cell that gives rise to both luminal and basal lineages. The other side argues that each lineage has its own stem cell and thus is unipotent<sup>19</sup>. As of now, the most recent lineage tracing reports argue for a bipotent MaSC at the apex of the hierarchy<sup>20,24</sup>. Single cell RNA-sequencing (scRNA-seq) of primary human breast epithelial cells found a continuous lineage hierarchy that connected the basal lineage to two differentiated luminal branches<sup>25</sup>.

### **1.1.2 Inherent properties of mammary epithelial cells**

Recent work has demonstrated that the mammary subpopulations are equipped with different systems to respond to cellular and genomic damage.

Reactive oxygen species (ROS), such as hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), superoxide (O<sub>2</sub><sup>•-</sup>) and hydroxyl anions, are by-products of cellular processes that consume oxygen<sup>26</sup>. They arise during normal

cellular activity but high levels can damage DNA, proteins, and lipids<sup>26</sup>. LP have higher levels of ROS and consume more oxygen, but are equipped with higher levels of the superoxide dismutase (SOD1-3) enzymes that catalyze  $O_2^\circ$  to  $O_2$  and  $H_2O_2$  and high levels of glutathione peroxidases<sup>27</sup>. Human BC had only glutathione-dependent mechanisms to handle ROS<sup>27</sup>. The limited antioxidant mechanisms present in BC was targeted by treating them with buthionine sulfoximine (BSO), an inhibitor of glutathione synthesis<sup>27</sup>. Indeed, LP were more resistant to BSO treatment than BC due to their additional glutathione-independent antioxidant mechanisms<sup>27</sup>.

Cells undergoing active transcription require efficient DNA damage repair (DDR) capacity to prevent mutations and ensure genomic integrity. Molecular profiling indicates that LP have high levels of transcription with twice the number of hypomethylated regions and four times the total RNA content compared to BC<sup>28</sup>. R loops, three-stranded nucleic acid structures comprised of RNA:DNA hybrid plus a displaced DNA strand, are naturally occurring by-products of transcription and linked to increased replication stress and DNA damage<sup>29</sup>. LP had significantly more R loops compared to BC, especially at genes responsible for luminal fate determination<sup>30</sup>. Telomeres form at the distal ends of chromosomes, maintenance of which is necessary for adequate genomic stability. Despite having the shortest telomere lengths (<3kb), LP were the only subset to express human telomerase (hTERT) and expressed higher levels of DDR genes<sup>31</sup>. BC alternatively had a telomere length of 6-8 kb<sup>31</sup>. When transduced with a lentiviral vector expressing either Cyclin E or H-RAS G12V, two oncogenes that induce DNA-replication stress, MEC demonstrate divergent DNA damage response. Luminal subsets show extensive DNA damage and widespread repair activation (more  $\gamma$ H2AX & 53BP1 foci), whereas BC saw little to none<sup>32</sup>.

Collectively, these works demonstrate that the different mammary lineages are intrinsically unique in how they protect themselves from cellular and genomic stress. LP have a wealth of mechanisms to prevent ROS-associated damage and genomic insults from DNA damage which were not apparent in the BC. These protective mechanisms may underlie the suitability of these cells to undergo the strenuous route of tumorigenesis.

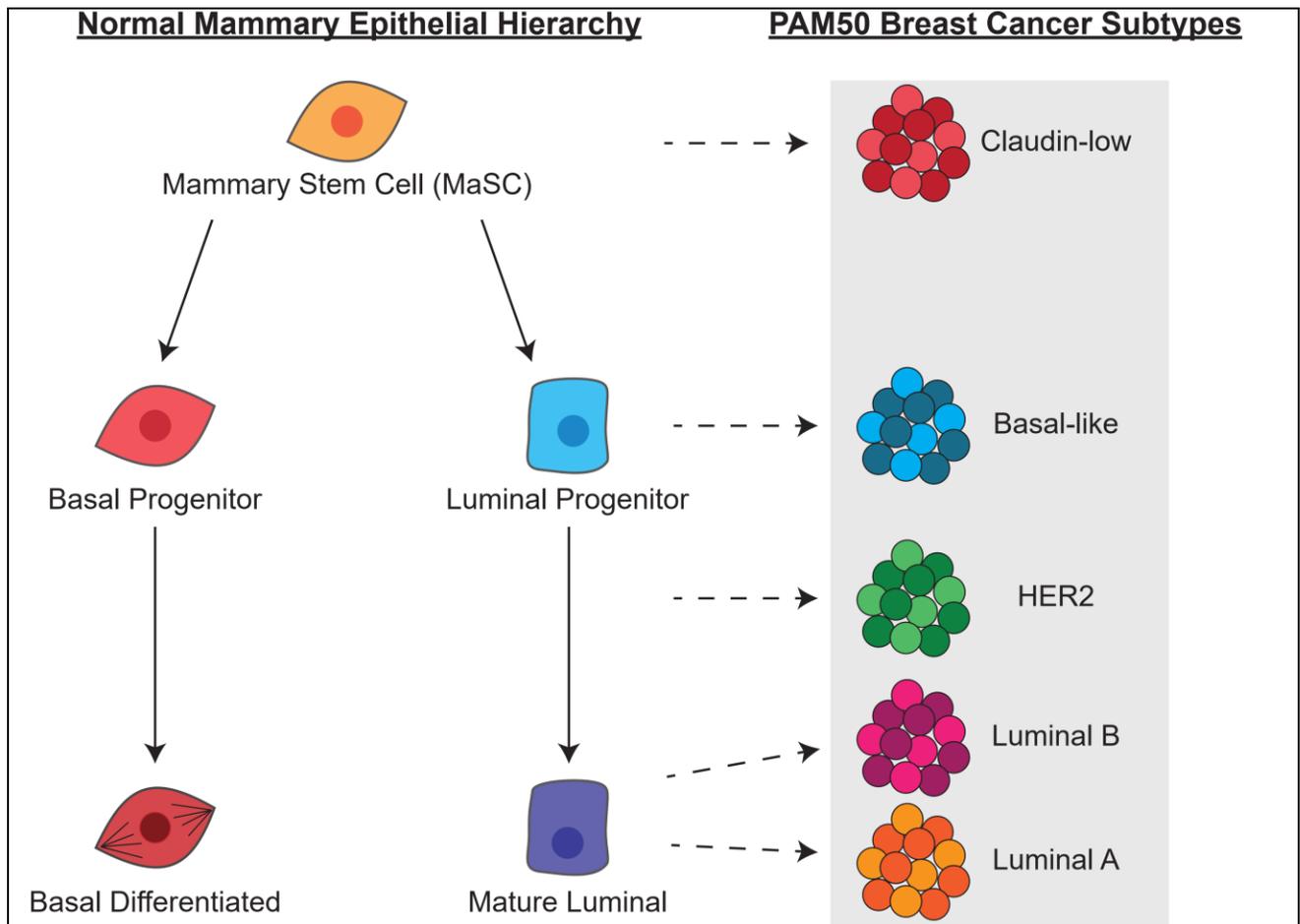
### 1.1.3 Cell of origin for breast cancer

The “cell-of-origin” refers to the cell that acquires mutations leading to cancer<sup>33,34</sup>. However, not every cell in the human body can become malignant, only specific populations are suitable candidates for transformation. A large body of evidence suggests that stem/progenitor populations are the cell-of-origin for many cancer types<sup>33–35</sup>. This is most likely due to their replicative potential and long lifespan that enables them to accumulate mutations<sup>36</sup>. In fact, it has been proposed that variations in cancer risk among different tissues can be explained by the number of stem cell divisions that occurred in that tissue<sup>37,38</sup>. The study’s meta-analysis of relationships between stem cell divisions from different tissue types and the risk of 17 different cancer types across 69 countries supports the concept that lifetime risk of a particular cancer strongly correlates with the total number of divisions of normal self-renewing stem cells in that tissue<sup>38</sup>.

This concept is especially relevant to the breast as this tissue contains many distinct progenitor populations each with the potential to serve as putative cell(s)-of-origin for breast cancer<sup>9,14</sup>. Comparison of gene expression in normal mammary subpopulations and breast cancer subtypes have shed light on the diverse cellular origins of breast cancer. It is proposed that BC give rise to Claudin-low subtype, ML to Luminal A/B and LP transform to give rise to the highly aggressive Basal-like subtype tumours<sup>39,40</sup>. The importance of the cell-of-origin has also been demonstrated in mouse models with lineage-specific promoters. Mice with deletion of *Brca1* and *p53* only in cells of the luminal lineage develop mammary cancers that resemble human BRCA1 tumors histologically<sup>41</sup>. In fact, flow cytometry analysis of breast tissue from BRCA1 mutation carriers have demonstrated an expanded LP population<sup>39</sup>. It has also been shown that LP yield Basal-like breast cancers following oncogenic insults, irrespective of BRCA1<sup>42,43</sup>. The importance of cell-of-origin has also been shown using human MEC. Transformation of normal human CD10<sup>+</sup> BC was able to generate Claudin-low breast cancer<sup>44</sup>. However, transformation of EPCAM<sup>+</sup>CD10<sup>-</sup> MEC yielded a mix of both ER positive and negative cancers. A recent study with 37 epithelial markers used to characterize >15,000 normal breast cells detected 11 differentiation states for luminal cells and 2 for BC<sup>6</sup>. All human breast tumors were similar to at least one of the 11 luminal differentiation states. However, the triple negative breast cancers (TNBC) showed a mixed phenotype consisting of both luminal and BC markers<sup>6</sup>. This highlights the heterogeneous

nature of these mammary subpopulation and the need for more refined cell surface signatures to accurately identify the putative cell(s)-of-origin.

Gene expression programs operating in normal mammary cells are hijacked in other aspects of tumorigenesis such as metastasis. Aggressive metastatic TNBC also show gene expression similarities to fetal (embryonic days 16 and 18) mouse MaSC, a stage in development with high stem cell capacity<sup>45,46</sup>. In addition, metastatic cells use distinct mammary cell programs that



**Figure 1. 2 – The mammary epithelial hierarchy and its relationship to putative cells-of-origin for breast cancers**

Figure adapted from Stingl & Visvader (2014)<sup>9</sup>. On the left is the mammary epithelial hierarchy, where the bipotent mammary stem cell (MaSC) sits at the apex and gives rise to basal and luminal lineages, which each containing progenitor and differentiated populations. On the right are the PAM50 subtypes of breast cancer. Based on gene expression correlations and mouse models of breast cancer, the putative cell-of-origin for each of the subtypes is delineated by the dash line.

affect their metastatic patterns<sup>47</sup>. scRNA-Seq data on metastatic cells derived from TNBC PDXs revealed that advanced-stage metastatic cells have higher expression levels of luminal differentiation genes (CDH1, MUC1, CD24) and were similar to the primary tumor<sup>47</sup>. Whereas earlier-stage metastatic disease (low burden) were distinct from the primary and demonstrated higher expression of BC (LGR5, BMI1), pluripotency genes (OCT4 and SOX2) and epithelial-to-mesenchymal transcription factors such as SNAI2 (SLUG)<sup>47</sup>. This work demonstrates that normal developmental programs found in MEC are co-opted by breast cancers and dictate their phenotype.

#### **1.1.4 Influence of Normal on Tumorigenesis**

Cancer has long been described as caricatures of normal tissue renewal<sup>33</sup>. This suggests that the tissue-of-origin dictates several key features of the tumorigenic process.

Hoadley *et al* used molecular clustering to show that chromosome-arm-level aneuploidy, DNA hypermethylation, mRNA and miRNA and reverse-phase protein array data clustered on the basis of histology or tissue-of-origin<sup>48</sup>. Cancers retain characteristics of normal tissues despite acquiring countless mutations, which is suggestive of the fact that tumorigenesis co-opts normal regulatory networks. One example relates to how cancer cells retain epigenetic features from their tissue-of-origin. Epigenomic features can explain the rate of somatic mutation that differs across the genome. Polak *et al.* analyzed mutations from 173 cancer genomes derived from 8 different tissue types and used 424 epigenetic features from more than 100 normal cell types<sup>49</sup>. This revealed that the density and distribution of somatic mutations across the genome of a particular cancer is strongly linked to chromatin marks specific to that cancer's cell-of-origin<sup>49</sup>. For example, density of somatic mutations in liver cancer followed levels of H3K4me1 in normal hepatocytes but not H3K4me1 in melanocytes<sup>49</sup>. The opposite was also true where melanoma mutational density correlated with levels of H3K4me1 in melanocytes, not hepatocytes<sup>49</sup>. In addition, mechanisms of proliferation occur in a tissue-specific manner<sup>50</sup>. A gain-of-function screen on human mammary epithelial cells (HMEC) and human pancreatic nestin-expressing epithelial cells (HPNEs)<sup>50</sup> revealed that 80-90% of proliferation genes were highly tissue-specific. These tissue-specific proliferation genes show much stronger enrichment with somatic copy number changes (SCNAs) and also provide a better prediction of whole chromosome-arm aneuploidy in their respective cancer than in cancers from different tissues<sup>50</sup>.

Collectively, the aforementioned examples demonstrate how several features of a tumor are partly determined by the normal tissue that the cancer is derived from. Therefore, the tissue/cell-of-origin is simply not a vessel for mutations to accumulate within, but in actuality dictates the outcomes of such mutations.

### **1.1.5 Breast Cancer Subtypes**

Breast cancer is a heterogeneous disease which makes matching patients with the best therapy a major challenge. Classical diagnostic markers used to stratify patients are ER, PR and human epidermal growth factor receptor 2 (HER2). However, molecular classification of breast cancers based on the Prediction Analysis of Microarray (PAM) 50 gene signature has led to the identification of patient groups with distinct transcriptional programs and clinical characteristics. The four “intrinsic” subtypes originally identified include Luminal A, Luminal B, Basal-like and HER2-enriched<sup>51,52</sup> and later on, the Claudin-low subtype was added<sup>40</sup>.

Each of the breast cancer subtypes have distinct mutational profiles and expression of diagnostic markers. Luminal A breast cancers are the most common and are largely HR+. Common mutations seen with this subtype are PIK3CA (49%), GATA3 (14%), MAP3K1 (14%) and TP53 (12%)<sup>53</sup>. Luminal A have the lowest mutation rate, and have been described to have quiet genomes<sup>53</sup>. Luminal B tumours are similar to Luminal A, as they also express HR+ and share common mutations like TP53 (32%), PIK3CA (32%), GATA3 (15%) and MAP3K1 (5%)<sup>53</sup>. However, the distinguishing feature that separates the two is that Luminal B breast cancers express proliferation genes at a higher level<sup>53</sup>. The HER2-enriched subtype is defined by expression of genes within the HER2 amplicon. This subtype is quite heterogeneous, including a mix of HR+ and HR- tumours<sup>53</sup>. Common mutations include TP53 (84%) and PIK3CA (7%)<sup>53</sup>. TNBC refers to a broad range of tumours that do not express any HR. Basal-like and Claudin-low tumours are TNBCs. Commonly mutated genes in Basal-like breast include TP53 (84%), PTEN (35%) and PIK3CA (7%)<sup>53</sup>. Of particular note, Basal-like breast cancers show elevated expression of DNA repair proteins, consistent with the highest mutation rate<sup>53</sup>. This cancer is commonly seen in BRCA1 mutation carriers. Claudin-low tumours show characteristically low expression of tight junction (claudin 3, 4, 7, cingulin, occludin) and cell-cell adhesion (E-cadherin) proteins<sup>40,54</sup>. Transcriptionally, they show low levels of proliferation genes and high levels of mesenchymal markers<sup>40,54</sup>.

Breast cancer subtypes have unique intrinsic vulnerabilities as they are susceptible to specific anti-cancer therapies. Patients with Luminal A and B subtypes tend to receive endocrine therapies<sup>55</sup>, either tamoxifen or aromatase inhibitors<sup>56,57</sup>. ER+ breast cancers receiving endocrine therapies have a 50% lower chance of recurrence than ER+ patients not receiving endocrine therapy. However, despite being disease-free after 5 years of endocrine therapy, the risk of recurrence up to 20 years after diagnosis can range from 10 to 41% depending on characteristics of the original tumor such as grade and node status<sup>58</sup>. Paclitaxel seemed to benefit patients with a low PAM50 proliferation score, which is typically seen in the Luminal A subtype<sup>59</sup>. However, Luminal B tumours are vulnerable to a combination of docetaxel, doxorubicin and cyclophosphamide, which was not seen in Luminal A tumours<sup>60</sup>. Chemotherapy agents like taxanes, anthracyclines and alkylating agent are the first-line therapy for TNBC<sup>55</sup>. Basal-like tumors have demonstrated sensitivity to docetaxel compared to doxorubicin, which was not seen for the other subtypes<sup>61</sup>. Patients with HER2-enriched breast cancers have benefited from Trastuzumab<sup>55</sup> and combination with lapatinib<sup>62</sup>. HER2 cancers co-expressing HR may also benefit from endocrine therapies<sup>56,62</sup>. Claudin-low has demonstrated intermediate response to anthracycline/taxane-based chemotherapy<sup>40</sup>.

The above highlights the distinct transcriptional programs operating in each subtype and how this may underlie their unique susceptibility to specific anti-cancer regimens. However, risk of recurrence and poor long-term survival are common, thus highlighting a need for better targets for treatment.

## **1.2 Basic Metabolism and Relevance to Cancer**

Metabolism involves either creating (anabolism) or breaking down (catabolism) macromolecules to the 4 building blocks of life (amino acids, nucleotides, sugars and lipids). The following section describes basic background and relevance to cancer of glycolysis, tricarboxylic acid (TCA) cycle and oxidative phosphorylation (OXPHOS). In addition, I briefly discuss the metabolic heterogeneity of breast cancer and determinants of cellular metabolism with a focus on the role of the cell-of-origin.

### 1.2.1 Glycolysis and Warburg Effect

Glycolysis is the main metabolic pathway responsible for breaking down glucose. It entails a series of 10 catabolic reactions that yields a net production of 2 pyruvate, 2 ATP and 2 NADH (Figure 1.3). All the steps in glycolysis are reversible, which is necessary to allow for gluconeogenesis, except for 3 key phosphorylation events. Hexokinase is the first enzyme in glycolysis that phosphorylates Glucose to Glucose-6-phosphate (G6P)<sup>63,64</sup>. Phosphofructokinase (PFK) is the rate-limiting enzyme in glycolysis, which produces fructose 1,6-bisphosphate (F16BP) from fructose 6-phosphate (F6P)<sup>63,64</sup>. Lastly, pyruvate kinase (PK) phosphorylates phosphoenolpyruvate (PEP) to pyruvate, the last step of glycolysis<sup>63,64</sup>. All three irreversible steps are inhibited by ATP, demonstrating the necessary negative feedback inhibition to prevent excessive ATP production<sup>63,64</sup>. Pyruvate, the final product of glycolysis, can either enter mitochondria to participate in the Tricarboxylic Acid (TCA) Cycle or is converted into lactate by lactate dehydrogenase (LDH) and exported out of the cell<sup>63,64</sup>.

The first studies to document the role of altered glucose metabolism in cancer were performed by Nobel Laureate Otto Warburg<sup>65,66</sup>. He saw that ascites cells consume glucose and produce large quantities of lactate regardless of oxygen levels. He postulated that mitochondria are dysfunctional in cancer cells and called this phenomenon the “Warburg Effect” or aerobic glycolysis. Although most cancer cells show high glycolytic flux, Warburg’s conclusions were not quite right<sup>67</sup>. Mitochondria are been demonstrated to be essential for tumorigenesis and glycolysis may play additional roles beyond ATP production in cancer<sup>68</sup>.

Glycolysis generates a net production of 2 ATP via the last phosphorylation event by PK. However, cancer cells express PKM2, an isoform of pyruvate kinase with less activity<sup>69</sup>. Knockdown of PKM2 and overexpression of the PKM1 isoform halts tumor formation and also suppression of aerobic glycolysis<sup>70</sup>. Thus, cancers may not need ATP from glycolysis. In addition, recent work with radiolabeled <sup>13</sup>C-lactate demonstrated that many of the carbons that make up TCA Cycle intermediates are derived from lactate, not glucose<sup>71,72</sup>. Cells actively import lactate and incorporate its carbons into the TCA cycle since the radiolabeled isotopes were infused intravenously<sup>71,72</sup>. These data argue that glycolysis and the TCA cycle may not be so connected in normal and cancer cells.

The prevailing notion about glycolysis in cancer is that its main function may be to contribute to macromolecule production<sup>73</sup>. Glycolysis is often referred to as “central carbon metabolism,” since several of the metabolites formed during glucose catabolism are shunted off to other essential metabolic pathways (Figure 1.3). For example, G6P and F6P can enter the oxidative and non-oxidative branches of the pentose phosphate pathway (PPP), respectively, which is necessary to make nucleotides<sup>73</sup>. In addition, F6P can enter the hexosamine biosynthetic pathway necessary to produce UDP-*N*-Acetylglucosamine, the precursor for glycosylation-related post-translation modifications of proteins<sup>73</sup>. Glyceraldehyde 3-phosphate (GADP) can be converted to glycerol 3-phosphate, which then feeds into lipid synthesis<sup>73</sup>. Phosphoglycerate dehydrogenase (PHGDH) catalyzes the conversion of 3-phosphoglycerate (3PG) to 3-phosphohydroxypyruvate (3PHP), which is further modified to produce serine<sup>73</sup>. Glycolysis does more than generate pyruvate or ATP, it serves as an essential generator of precursors for other metabolic pathways.

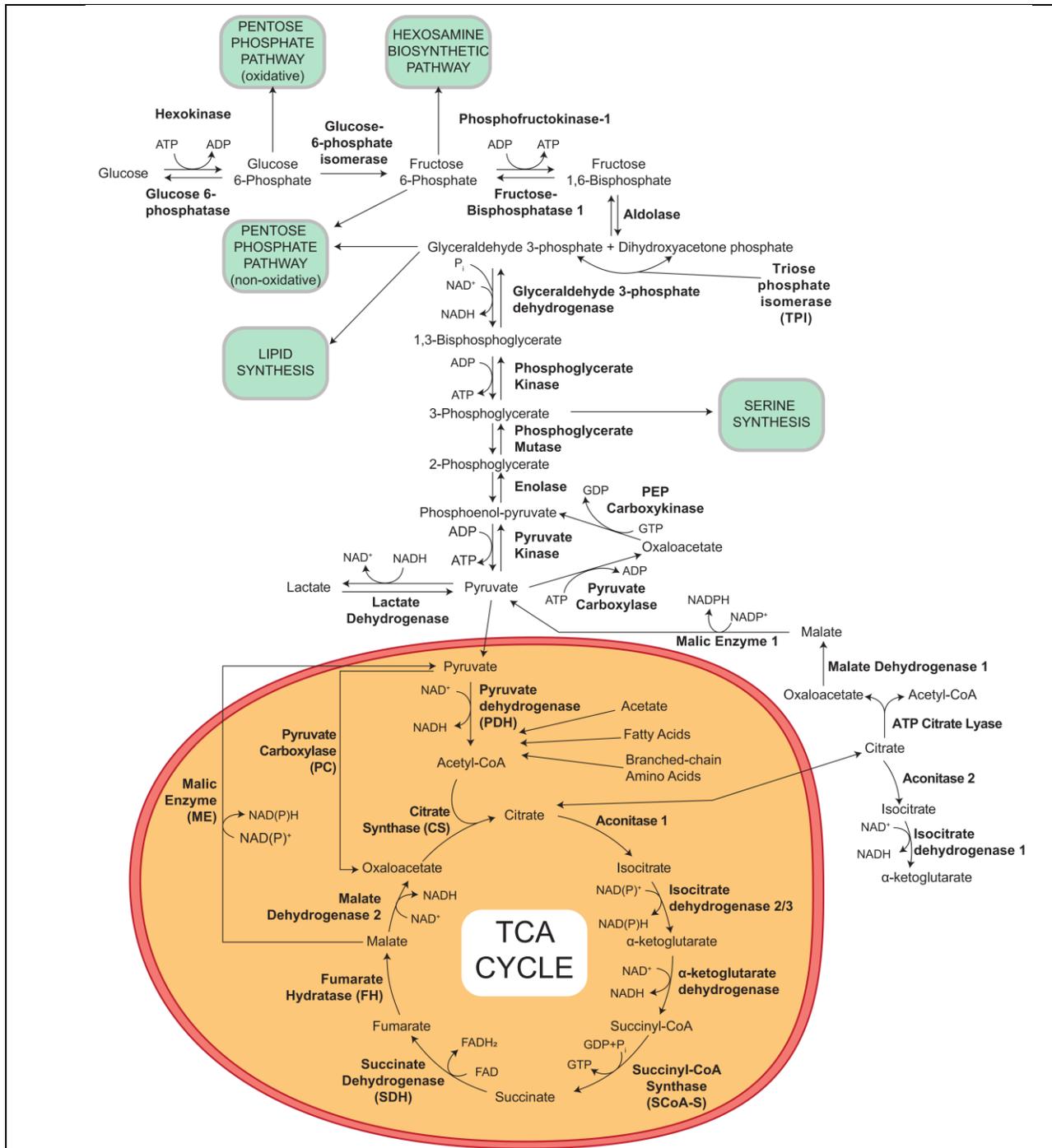
Several inhibitors of the critical steps in glycolysis have been identified and demonstrated promising anticancer activity<sup>74,75</sup>. However, many proved toxic due to adverse effects like hypoglycemia<sup>74</sup>. Glucose catabolism is an essential process for all cells and perhaps better therapeutic success will be obtained by targeting off-shoot pathways of glycolysis, rather than the main pathway. Although the function of glycolysis in cancer is constantly evolving, it still remains a defining feature of reprogrammed cellular metabolism<sup>76</sup>.

### **1.2.2 Tricarboxylic Acid (TCA) Cycle and Cancer Initiation**

The TCA cycle functions to generate the electron carriers, NADH and FADH<sub>2</sub>. It begins with the incorporation of Acetyl-CoA into citrate, which is mediated by citrate synthase (CS). Due to the cyclic nature of this pathway, it can continuously generate electron carriers, but requires precursors to keep the cycle going (Figure 1.3). The TCA cycle accomplishes this by acting as a funnel receiving metabolites from several different mitochondrial metabolism pathways. For instance, Acetyl-CoA can be generated through the conversion of carnitines, breakdown of acetate via acyl-CoA synthetase short-chain family, member 1 (ACSS1), breakdown of branched chain amino acids and breakdown of fatty acids via  $\beta$ -oxidation (Figure 1.3).

Mutations in TCA cycle enzymes that lead to cancer have been well documented. The most famous of these is isocitrate dehydrogenase (IDH). IDH1/2 mutations are common in

glioblastoma and acute myeloid leukemia<sup>77,78</sup>. Mutant IDH1 and 2 gain the ability to produce an “oncometabolite” 2-hydroxyglutarate (2-HG), instead of  $\alpha$ -ketoglutarate ( $\alpha$ KG)<sup>79–81</sup>. Loss-of-function mutations in succinate dehydrogenase (SDH) and fumarate hydratase (FH) have been shown to predispose individuals to cancers<sup>82</sup>. Germline SDH mutations leads to paragangliomas and pheochromocytomas, whereas germline FH mutations lead to leiomyomas and renal cell cancer<sup>82</sup>. These mutations do not produce oncometabolites, but instead lead to a buildup of the reactant, succinate for SDH deficiency and Fumarate for FH deficiency. Mechanistically, succinate, fumarate and 2-HG all have similar effects on epigenetic machinery as they inhibit members of the  $\alpha$ -KG-dependent dioxygenase family. These metabolites can inhibit the ten-eleven translocation (TET) family of 5-methylcytosine (5mC) hydroxylases, which are responsible for DNA demethylation and the JmjC domain-containing histone demethylases (KDMs)<sup>80,81,83</sup>. Inhibition of these demethylases leads to an increase in methylation of CpG islands and also histones, which alters the epigenome in such a way as to block differentiation leading to acquisition of stem-like properties<sup>80,84</sup>. TCA Cycle intermediates play an essential role in cancer initiation and also demonstrates diverse functions beyond participating in their canonical metabolic pathway. In fact, these metabolites have documented pleotropic effects important for tumorigenesis. 2-HG inhibits collagen maturation by inhibiting prolyl hydroxylase domain-containing proteins (PHD), an  $\alpha$ -KG-dependent enzyme, that is responsible for prolyl-hydroxylation of collagen<sup>85</sup>. In addition, succinate and fumarate can inhibit homologous-recombination<sup>86,87</sup>. This was exploited as a “metabolic” synthetic lethality, as cancer cells with SDH and FH mutations were more susceptible to poly(ADP)-ribose polymerase (PARP) inhibitors<sup>87</sup>.



**Figure 1. 3 - Schematic for Glycolysis and Tricarboxylic Acid (TCA) Cycle.**

Schematic summarizes all the major reactions occurring in glycolysis and the Tricarboxylic Acid Cycle. In addition, offshoot pathways from glycolysis are listed in the teal textbox. Abbreviations include the oxidized/reduced forms of electron carriers include Nicotinamide adenine dinucleotide (NAD<sup>+</sup>/NADH), Nicotinamide adenine dinucleotide phosphate (NADP<sup>+</sup>/NADPH), Flavin adenine dinucleotide (FADH/FADH<sub>2</sub>). Other abbreviations: Adenosine triphosphate (ATP), Adenosine diphosphate (ADP), Guanosine triphosphate (GTP), Guanosine diphosphate (GDP), inorganic phosphate (Pi).

### 1.2.3 Electron Transport Chain (ETC)

At the heart of mitochondria is the electron transport chain (ETC), which drives ATP production and has given this organelle its popularized function as “powerhouse of the cell.” The ETC is composed of 4 distinct multi-subunit complexes (CI-IV) (Figure 1.4). The complexes are responsible for transferring electrons along the inner mitochondrial membrane (IMM) to the final electron acceptor oxygen<sup>63,64</sup>. Throughout this process, each complex, except CII, pump protons into the intermembrane space<sup>63,64</sup>. This creates a proton gradient that is essential to drive the unfavourable reaction of ATP production by ATP Synthase (also referred as Complex V)<sup>63,64</sup>. The ETC and CV are located specifically within cristae, folding of the inner mitochondria membrane (IMM) that protrude into the mitochondrial matrix<sup>88</sup>. The structure of the mitochondria (surface area, shape and density of cristae) has been proposed to be a key determinant of ETC efficiency<sup>64,89</sup>. Each mitochondrial complex has a distinct structure and function (Figure 1.4)<sup>63,64</sup>.

Complex I (CI), also known as NADH dehydrogenase, is the largest. As the name suggests, it is responsible for receiving electrons specifically from NADH<sup>63,64</sup>. It is a massive L-shaped structure composed of 45 subunits and is nearly one million Daltons in total<sup>63,64</sup>. Of the 45 subunits, 39 are derived from the nuclear genome and 7 are derived from mitochondrial DNA<sup>63</sup>. A pair of electrons traversing through Complex I is said to pump 4 protons into the inner membrane space<sup>63,64</sup>. Mitochondrial reactive oxygen species (mROS) can be produced by CI in the mitochondrial matrix<sup>90,91</sup>.

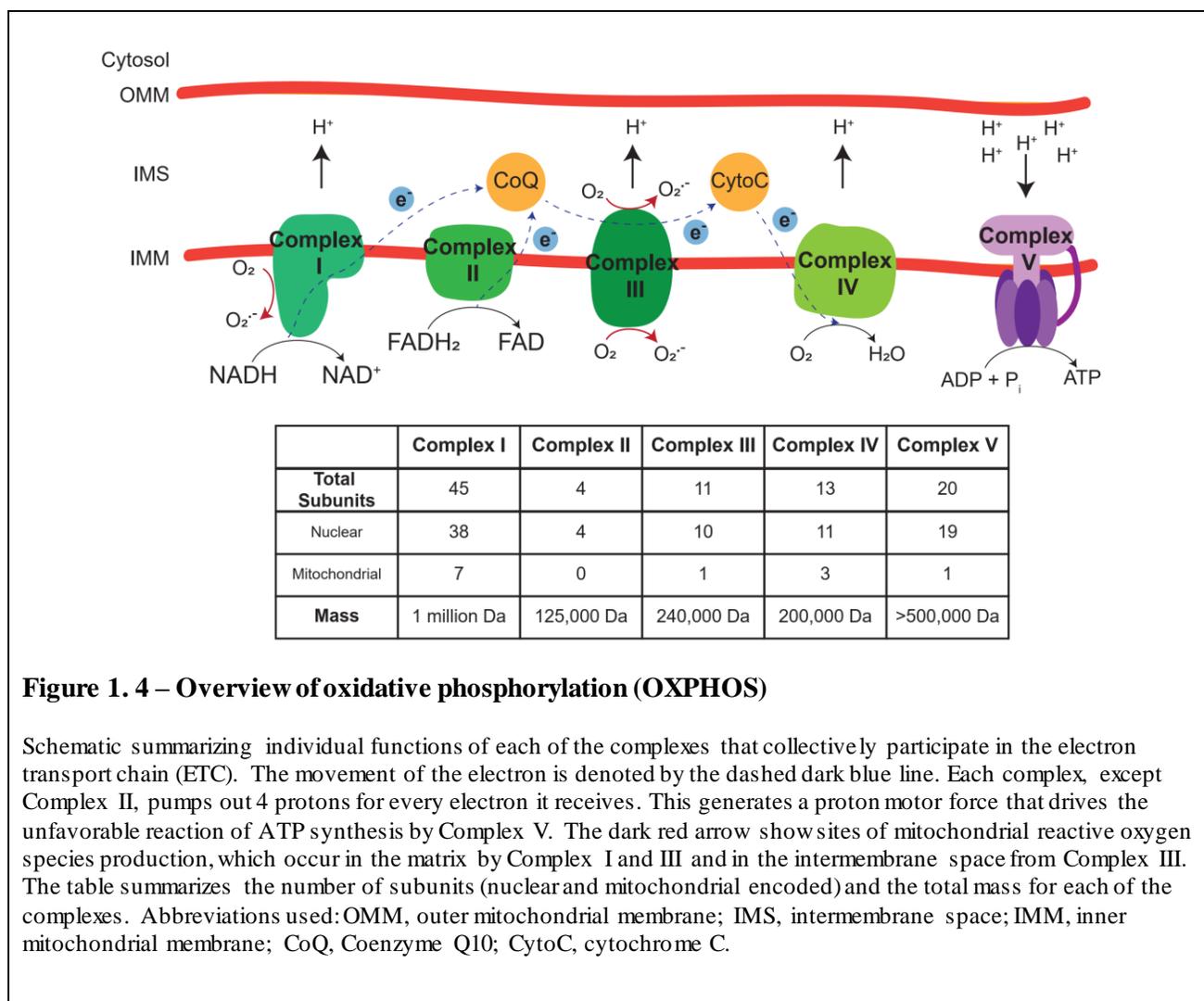
Complex II (CII) is succinate dehydrogenase (SDH), the same enzyme from the TCA Cycle. It specifically receives electrons from FADH<sub>2</sub>, but unlike the other complexes it does not pump out protons (Karp, 2009). This 125,000 Daltons complex is composed of 4 nuclear-encoded subunits<sup>63</sup>.

Electrons that enter the ETC through CI (via NADH) or CII (via FADH<sub>2</sub>) are received by Coenzyme Q<sub>10</sub> (CoQ), a hydrophobic molecule that shuttles electrons to Complex III (CIII)<sup>63,64</sup>. The electrons are then passed on to Cytochrome C (CytoC), which results in 4 protons being pumped into the IMS. CIII is a 11 subunit complex (with 10 nuclear encoded subunits, and 1

mitochondrial DNA derived subunit) that weights around 240,000 Daltons<sup>63</sup>. CIII is able to produce mROS in both the mitochondrial matrix and also the intermembrane space<sup>90,91</sup>.

Complex IV (CIV), also known as cytochrome c oxidase, is the fourth complex of the ETC. This complex catalyzes the final step in the ETC, which involves transferring electron from reduced CytoC to molecular oxygen<sup>63</sup>. Each oxygen allows for translocation of 10 protons, 4 of which are used to make H<sub>2</sub>O<sup>63</sup>. Structurally, CIV is a 13-subunit complex protein that weights around 200,000 Daltons. CIV is made by 10 nuclear encoded genes and 3 mitochondrial DNA encoded subunits<sup>63</sup>.

Complex V (CV) or ATP synthase is a multi-protein complex that drives ATP production in the mitochondria<sup>63,92</sup>. It is composed of 2 major subcomplexes, F<sub>1</sub> and F<sub>0</sub><sup>63,92</sup>. The F<sub>1</sub> portion is made of several subunits, 3 $\alpha$ , 3 $\beta$ , 1 $\delta$ , 1 $\gamma$ , 1 $\epsilon$ <sup>63,92</sup>. Alternate  $\alpha$  and  $\beta$  subunits form the F<sub>1</sub> head portion, where each  $\beta$  subunit serves as a catalytic site for ATP synthesis<sup>63</sup>. The F<sub>0</sub> subunit is embedded in the mitochondrial membrane and is composed of 1a, 2b and 10-14c subunits<sup>63,92</sup>. The F<sub>0</sub> subunit contains a channel through which protons flow from the intermembrane space into the mitochondrial matrix<sup>63,92</sup>. This initiates a rotary motion of both subunits and drives ATP synthesis from ADP and inorganic phosphate<sup>92</sup>.



**Figure 1. 4 – Overview of oxidative phosphorylation (OXPHOS)**

Schematic summarizing individual functions of each of the complexes that collectively participate in the electron transport chain (ETC). The movement of the electron is denoted by the dashed dark blue line. Each complex, except Complex II, pumps out 4 protons for every electron it receives. This generates a proton motor force that drives the unfavorable reaction of ATP synthesis by Complex V. The dark red arrow shows sites of mitochondrial reactive oxygen species production, which occur in the matrix by Complex I and III and in the intermembrane space from Complex III. The table summarizes the number of subunits (nuclear and mitochondrial encoded) and the total mass for each of the complexes. Abbreviations used: OMM, outer mitochondrial membrane; IMS, intermembrane space; IMM, inner mitochondrial membrane; CoQ, Coenzyme Q10; CytoC, cytochrome C.

Targeting mitochondrial respiratory chain complexes with inhibitors has long been an active area of research due to strong role the ETC plays in multiple disease states, especially cancer<sup>93</sup>. Mechanistically, it has been presumed that inhibition of any ETC components will yield the same effect (decreased oxygen consumption and ATP production), since each complex collectively participates in OXPHOS. However, the differential metabolic, epigenetic and transcriptional alterations are specific to the complex being inhibited.

Inhibition of each complex will lead to specific changes in metabolite concentrations. For example, since CI and CII oxidize NADH and FADH<sub>2</sub>, respectively, all enzymatic reactions occurring in the mitochondrial that utilize oxidized forms of these electron carriers, NAD<sup>+</sup> and FADH, will be inhibited due to decreased reserves<sup>94</sup>. Acyl-CoA dehydrogenase and β-

hydroxyacyl-CoA dehydrogenase require FADH and NAD<sup>+</sup>, respectively, for fatty acid oxidation to occur. In addition, it was recently shown that the main proliferation defect due to ETC inhibition was caused by a redox imbalance ( $\uparrow$ NADH/NAD<sup>+</sup> ratio) that leads to reduced cytosolic aspartate synthesis<sup>95,96</sup>. This proliferation defect can be reversed by culturing cells in supra-physiological concentrations of pyruvate or aspartate, which restores levels of NAD<sup>+</sup> levels and activity of the malate-aspartate shuttle<sup>95-97</sup>. Metabolomics on isolated mitochondria recently revealed unexpected complex-specific metabolite changes that occur upon ETC inhibition. For example, inhibition of CI using Piericidin led to Acetyl-CoA accumulation in the mitochondrial matrix<sup>94</sup>. This was most likely due to a high NADH/NAD<sup>+</sup> ratio that prevents activation of CS, the enzyme which incorporates Acetyl-CoA into the TCA cycle. Inhibition of CI led to a >250-fold increase in the acetyl carnitine/carnitine ratio<sup>94</sup>. Inhibition of CIII alternatively led to high levels of choline, carbamoyl aspartate, and succinate and decreased levels of betaine<sup>94</sup>. Inhibition of CV via oligomycin lead to accumulation of  $\alpha$ -KG and malate<sup>94</sup>. These mitochondrial specific metabolite changes were not seen when the same analysis was done on whole-cell lysates, thus the mechanisms and significance of these changes have yet to be determined.

As epigenetic enzymes require a specific metabolite as a cofactor, the aforementioned complex-specific changes should alter the epigenome with subsequent changes to transcriptional output of the cell. The lab of Dr. Navdeep Chandel knocked out the Rieske iron-sulfur protein (RISP), an essential subunit of CIII, in the fetal hematopoietic stem cell (fHSC)<sup>98</sup>, CD4<sup>+</sup> T cells<sup>99</sup> and also regulatory T (Treg) cells<sup>100</sup>. CIII deficient fHSC were able to proliferate and self-renewal, but experienced a block in differentiation capacity and could not produce progenitors<sup>98</sup>. CIII deficient CD4<sup>+</sup> T cells had compromised immune function as they had diminished IL-2 production and failed to undergo antigen-specific expansion of CD4<sup>+</sup> and CD8<sup>+</sup> T cells in response to a viral infection<sup>99</sup>. The same CIII knockout in Treg led to an inability of these cells to perform immunosuppressive functions<sup>100</sup>. These works collectively demonstrate that the same complex is linked to distinct cellular functions depending upon the cell type.

In addition, each of the mitochondrial complexes may participate in specific functions within a single cell type. Treatment of regulatory T (Treg) cells with ETC inhibitors for 24 hours caused

complex-specific changes in metabolite accumulation and gene expression<sup>100</sup>. CI is important for their proliferation, however CII is responsible for IFN $\gamma$  and cytokine production<sup>101</sup>. These studies demonstrate that each of the mitochondrial complexes are unique entities that are tied to specific functions within a single cell-type.

#### **1.2.4 Normal Mammary Epithelial Cell Metabolism**

There literature on metabolism in MECs is limited. As mentioned above, human LPs are endowed with a higher antioxidant capacity<sup>27</sup>. However, 6 weeks old mouse mammary LP had higher mROS and total ROS levels as compared to a MaSC-enriched population<sup>102</sup>. Instead of gating the entire basal population, they selectively analyzed the tip of the BC population with the highest CD49f and CD24 expression<sup>102</sup>, which has been proposed to be enriched for a MaSC population due to its high transplantation efficiency<sup>10</sup>. In addition, fetal MaSC (embryonic day 16 and 18) have high transcript levels of glycolytic enzymes<sup>46</sup>. This is in accordance with other stem cell populations, where the hematopoietic and embryonic stem cells are highly glycolytic<sup>103</sup>. However, the metabolic identities of normal mammary subpopulations remain understudied.

#### **1.2.5 Breast Cancer Metabolism**

Steady-state metabolomics of different breast cancer subtypes has revealed striking metabolic heterogeneity between subtypes<sup>104–108</sup>. Metabolomes of Luminal A and B tumors consistently cluster separately from the more aggressive subtypes (Basal-like and HER2) in several reports<sup>104,105,107–109</sup>. Basal-like and HER2 tumours have higher abundance of glycolytic and TCA Cycle metabolites<sup>105,107,109</sup>. In addition, these highly aggressive breast cancers also demonstrate high abundance of PPP metabolites, indicating potential diversion of glycolytic metabolites to other biosynthetic pathways<sup>105,109</sup>. Despite finding no IDH mutations in their patient cohorts, Basal-like and ER- breast cancers have high abundance of 2-HG<sup>104,105,107,108</sup>. When compared to normal breast, 2-HG levels was elevated more than 20-fold in ER+ and 200-fold in ER- breast cancers<sup>107</sup>. Mechanistically, these tumours produce 2-HG from the high abundance levels of glutamine and glutamate, which can be converted to  $\alpha$ KG, the substrate needed for this reaction<sup>104,108</sup>.

Even within a single breast cancer subtype, there has been reported large inter-tumor metabolic heterogeneity. One study found three distinct metabolome clusters for Luminal A tumours<sup>110</sup>. Clustering of ER- or TNBC metabolomes grouped largely based on ancestry (African-American of Western African ancestry or European-American ancestry)<sup>108</sup>. African-American women with ER- breast compared to European-Americans cancers tended to have higher MYC activation, which enhanced glutamine incorporation into 2-HG leading to hypermethylation and overall poorer prognosis compared to ER- breast cancers of European Americans<sup>108</sup>.

### 1.2.6 Influence of normal on cancer metabolism

As mentioned above, several aspects of the tumorigenic phenotype are linked to cell-of-origin (Section 1.14). In fact, the metabolism of cancer has been proposed to be more similar to the tissue it originated from.

The simplest example comes from clustering 225 metabolites identified in 928 cell lines from the Cancer Cell Line Encyclopedia (CCLE), which clustered primarily based on lineage<sup>111</sup>. The hematopoietic cell lines had a distinct metabolome as compared to non-hematopoietic lines<sup>111</sup>. Computational analysis of metabolic gene expression data has also shown striking similarities of a cancer to its tissue-of-origin<sup>112,113</sup>. However, the importance of the tissue-of-origin was made clear by observing metabolic changes that occur when the same mutational event exerted its effects in different tissues. MYC-induced liver tumors relied more upon glutamine catabolism, whereas MYC-induced lung tumors relied more on glutamine anabolism from glucose<sup>114</sup>. Thus, metabolic alterations and their connection to cancer gene mutations are largely dependent upon tissue-of-origin. In addition, Mayer *et al.* demonstrated that lung and pancreatic cancers, despite having the same initiating event ( $Kras^{G12D/+};Trp53^{lox/lox}$ ), have different dependencies for branched amino acids that reflects the metabolic preferences of the tissue-of-origin<sup>115</sup>. KRAS is well-known to induce macropinocytosis, a nutrient acquisition method where cells engulf extracellular sources of protein like albumin and break it down as a source of amino acids<sup>116</sup>. However, KRAS-induced macropinocytosis only occurred in pancreatic, not lung cancer<sup>115</sup>. Knockdown of BCAT, the enzyme responsible for the first step of branched-chain amino acid catabolism, was only a liability in lung and not pancreatic  $Kras^{G12D/+};Trp53^{lox/lox}$  cancer<sup>115</sup>. This study shows the potential clinical success of targeting tissue-of-origin specific metabolic vulnerabilities. In fact, it was recently shown that successful targeting of NAD metabolism was

dependent upon the tissue-of-origin<sup>117</sup>. A cell can make NAD through one of many distinct pathways<sup>118</sup>. The “Preiss-Handler Pathway (PHP)” utilizes nicotinic acid to generate NAD, whereas the Salvage Pathway generates NAD by using nicotinamide as starting material<sup>118</sup>. Chowdhry et al demonstrated that if the tissue-of-origin had high levels of *NAPRT*, a key enzyme in the PHP pathway, then cancers derived from that tissue would exclusively rely on this branch of NAD synthesis and also have higher amplifications of PHP genes such as *NAPRT* or *NADSYN1*<sup>117</sup>. Tissues that had low expression of PHP genes relied more on the Salvage pathway in their cancer counter parts<sup>117</sup>. Inhibition of NAMPT, a key Salvage pathway enzyme, with FK-866 was only a liability in non-PHP amplified cancers<sup>117</sup>.

These works collectively suggest that cancer adapts the metabolic network of the tissue it originated from. In addition, targeting tissue-of-origin specific metabolic vulnerabilities may represent a viable treatment option. To date, there have been no studies demonstrating the influence of the cell-of-origin on cancer metabolism.

## **1.3 Thesis Outline**

### **1.3.1 Rationale and hypothesis**

Cancers are constrained by the metabolic network of their normal counterpart and use this as a backbone for aberrant proliferation. Thus, the metabolism of a particular cancer is more similar to its tissue-of-origin than to a cancer that arose from a different tissue<sup>113,119</sup>. This poses a challenge in the context of breast cancer, as there are multiple cell(s)-of-origin in the mammary gland, each with the unique potential to give rise to a specific breast cancer subtype<sup>9,14</sup>. To date, limited work has explored the metabolic identities of normal MEC. In addition, we do not know if these metabolic programs are adapted by their respective breast cancer subtypes. By studying normal MEC metabolism, we can potentially uncover reasons behind at least some of the metabolic heterogeneity seen across PAM50 breast cancer subtypes. I hypothesize that normal MEC have lineage-restricted metabolic networks, which may be necessary for the structure and function of that cell type. In addition, I hypothesize that the metabolism of breast cancer subtypes may be more similar to their cell-of-origin.

### **1.3.2 Objectives**

- 1) Determine if mammary subpopulations possess distinct metabolic networks
- 2) Functionally evaluate the metabolic capacities and vulnerabilities of mammary subpopulations
- 3) Delineate if breast cancer subtypes share similar metabolic gene expression to the cell-of-origin.

## Chapter 2

### 2 Methodology

#### 2.1 Human mammary sample preparation, single cell preparation, and FACS

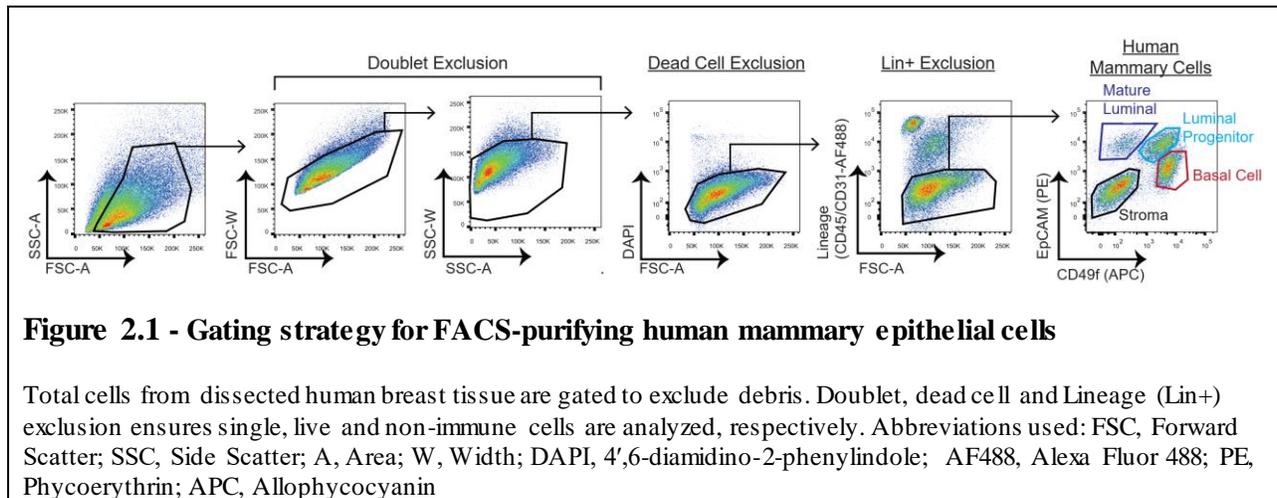
Human mammary tissues, obtained from reduction mammoplasties, were acquired with patient consent and approval by the Institutional Research Ethics Board of the University of British Columbia (Vancouver, BC). Hormonal status (premenopausal, follicular and luteal) was determined by pathologist examining breast specimens at UBC (Ramakrishnan et al., 2002). All pertinent clinical information regarding our patient cohort can be found in Table 2.1.

**Table 2.1 – Summary of clinical covariates for the cohort of patients used in this study.**

| Patient Number | Hormone Status  | Patient Age | Patient Age Group |
|----------------|-----------------|-------------|-------------------|
| 06.16          | Luteal          | 30          | ≤30               |
| 14.16          | Post-Menopausal | 59          | 51-60             |
| 19.15          | Follicular      | 33          | 31-40             |
| 21.16          | Luteal          | 28          | ≤30               |
| 31.12          | Post-Menopausal | 60          | 51-60             |
| 38.14          | Follicular      | 32          | 31-40             |
| 45.14          | Post-Menopausal | 67          | >60               |
| 46.15          | Follicular      | 39          | 31-40             |
| 49.15          | Post-Menopausal | 61          | >60               |
| 55.15          | Luteal          | 38          | 31-40             |

Reduction mammoplasty specimens were minced and enzymatically dissociated in DMEM:F12 1:1 media with 15 mM HEPES plus 2% BSA, 1% Penicillin-Streptomycin, 5 µg/ml insulin, 300 U/ml Collagenase (Sigma, C9891) and 100 U/ml Hyaluronidase (Sigma, H3506) shaking gently at 37°C, overnight or for 16-18 hours. Epithelial organoids were harvested spinning at 80g for 30 seconds and viably cryopreserved, as described previously<sup>120</sup>. Human breast tissue organoids were thawed and dissociated into single cell suspensions as reported previously<sup>121</sup>. Briefly, organoids were triturated in 0.25% trypsin-EDTA (Stem Cell Technologies, 07901) followed by 5 U/mL dispase (Stem Cell Technologies, 07913) and 50 µg/ml DNase I (Sigma, D4513) as described above for mouse samples, but for 5 minutes each.

Cells were then washed in between steps with HBBS + 2% FBS and filtered using a 40  $\mu\text{m}$  cell strainer. For FACS staining antibodies against CD45 (PECy7), CD31 (PECy7), EpCAM (APCCy7) and CD49f (FITC) were used. Lineage (Lin) positive cells were defined as CD31<sup>+</sup>CD45<sup>+</sup>. Human mammary cell subpopulations were defined as: basal (Lin<sup>-</sup>EpCAM<sup>lo-</sup>medCD49f<sup>hi</sup>); and luminal progenitor (Lin<sup>-</sup>EpCAM<sup>hi</sup>CD49f<sup>med</sup>); mature luminal (Lin<sup>-</sup>EpCAM<sup>hi</sup>CD49f<sup>lo</sup>). Dead cells were excluded following doublet exclusion using DAPI. An example is provided in Figure 2.1.



## 2.2 Proteomics on FACS-purified human mammary subpopulations

For ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) of human mammary subpopulations, 100,000 cells from each population were isolated sorted from each patient, as described<sup>122</sup>. After FACS purification, cells were washed in ice-cold PBS and pelleted. Pellets were then resuspended in 50% (vol/vol) 2, 2, 2-trifluoroethanol in PBS and disrupted into cellular lysates sequentially by repeated probe sonication, followed by six freeze-thaw cycles. Proteins in cellular lysate were denatured by incubation at 60°C for 2 h, oxidized cysteines reduced using 5 mM dithiothreitol for 30 min at 60°C and alkylated through reaction with 25 mM iodoacetamide for 30 min at room temperature in the dark. Each sample was diluted five times using 100 mM ammonium bicarbonate, pH 8.0. Proteins in lysates were digested into peptides through addition of 5  $\mu\text{g}$  of MS-grade trypsin (Promega). The digestion was performed overnight at 37°C and subsequently desalted using OMIX C18 pipette tips (Agilent). Peptides were semidried through vacuum centrifugation and resuspended in water with 0.1% formic acid.

Subsequently, all samples were analyzed using an Easy-LC1000 (Thermo Fisher Scientific) coupled to a QExactive tandem mass spectrometer (Thermo Fisher Scientific). Peptides were separated on an ES803 (Thermo Fisher Scientific) nano-flow column heated to 50°C using a 4-h reverse-phase gradient.

## **2.3 Bioinformatics Analysis of human mammary subpopulation proteomes**

Mass spectrometric data was analyzed using the MaxQuant quantitative proteomics software (version 1.5.3.8) and a Human UniProt sequences FASTA database. Carbamidomethylation of cysteine was specified as a fixed modification and oxidation of methionine was specified as a variable modification. Proteins were identified with a minimum of two unique peptides, the maximum false peptide discovery rate was specified as 1%, and “match between runs” was enabled. The distribution of intensity-based absolute quantification (iBAQ) values was adjusted to the distribution of label-free quantification (LFQ) values based on the median for each sample. This allowed for imputation of missing LFQ values with iBAQ values. Non-zero values were log<sub>2</sub>-transformed. The final list consisted of 6034 unique protein groups detected in at least one of the samples. Intensity values represent expression of proteins.

Further data processing was performed using the R statistical environment (version 3.5.2)<sup>123</sup>. For protein groups in which both LFQ and iBAQ values were missing, the 0 values were imputed with a random value between 1 and 1.5. Imputation was performed as a precautionary measure for further statistical analysis. For example, it ensured that constants did not occur when calculating standard deviation. One BC sample was excluded from the analysis as a technical error occurred with the instrument during its run. As four samples were run on a separate day, intensity values were then adjusted for sample batch effects using the ComBat method in the surrogate variable analysis “sva” R package (version 3.30.1)<sup>124,125</sup>.

### **2.3.1 Total Proteome Analyses**

Non-imputed ComBat-modified iBAQ-adjusted LFQ values were used to discover uniquely expressed proteins in each cell type. Averages across samples in each cell type were taken, resulting in one mean expression value for each protein in each cell type ( $n_{BC} = 9$ ,  $n_{LP} = 10$ ,  $n_{ML}$

= 10). Next, the values of zero for each cell type and associated proteins were excluded from the analysis. Number of protein expressed in each cell type were summarized in a triple Venn diagram, created using the “VennDiagram” R package (version 1.6.20)<sup>126</sup>. Mean expression values for each protein in each of the cell types were ranked according to descending log<sub>2</sub> median intensities and grouped into deciles. The protein with the highest intensity received a rank of 1 and thus, was placed in the first decile. Meanwhile, the protein with the lowest mean intensity received a rank of  $y$  and was placed in the tenth decile, where  $y$  represents the total number of proteins detected in a particular cell type. Pathway analysis via the “enrichR” R package (version 1.0) was conducted on the proteins in each decile<sup>127</sup>.

Principal component analysis (PCA) was performed by calculating Euclidean distances of scaled expression values. PCA scores were plotted in a plane defined by the first two components (that is, PC1 and PC2) using the “ggbiplot” R package<sup>128</sup>. Ellipses were drawn around cell type clusters, where centroids were the barycentre of each cluster and the diameter represented the maximum variance.

Heat maps depicted z-scores of protein expression values ( $x$ ) computed using the formula:  $(x - \text{mean}(x))/\text{standard\_deviation}(x)$ . Divisive hierarchal clustering dendograms of Pearson distance matrices for samples and proteins were created using DIANA (DIVISIVE ANALYSIS CLUSTERING) method in the “cluster” R package (version 2.0.7-1)<sup>129</sup>. Heat maps were plotted using the “pheatmap” (version 1.0.12) and “RColorBrewer” (version 1.1-2) R packages<sup>130,131</sup>.

### **2.3.2 Generating Metabolic Signatures**

A metabolic proteome was obtained by filtering the total proteome using a curated list of 2753 genes that encompasses all known human metabolic enzymes and transporters<sup>132</sup>. Based on matching by gene symbols, 1020 proteins related to metabolism were found in the total proteome of 6034 proteins, including “PKM” which was not identified in the curated list. As multiple protein groups in the proteome shared the same gene symbols, duplicates were included in the analysis. Metabolic signatures were acquired by looking at proteins in which mean expression met the fold-change and statistical change cut-offs in each cell type compared to the other two cell types ( $n_{BC} = 9$ ,  $n_{LP} = 10$ ,  $n_{ML} = 10$ ). The log<sub>2</sub> fold-change (FC) cut-off was greater than 0

and the statistical significance cut-off was  $P < 0.05$  in a one-way ANOVA and Tukey's multiple comparisons test. For example, for a protein in the metabolic signature for BC, where  $x$  represents the intensity values for the protein and the subscripts represent the cell-type to which the values belong, the logical expression is:

$$\log_2(\bar{x}_{BC}) - \log_2(\bar{x}_{ML}) > 0 \ \& \ \log_2(\bar{x}_{BC}) - \log_2(\bar{x}_{LP}) \ \&$$

Tukey's P-value  $x_{BC}-x_{LP} < 0.05$  & Tukey's P-value  $x_{BC}-x_{LP} < 0.05$

**Table 2.2 – Mature Luminal Metabolic Cluster**

|        |         |         |        |         |        |        |         |         |
|--------|---------|---------|--------|---------|--------|--------|---------|---------|
| ABAT   | AHCY    | ATP13A1 | EPHX1  | GNMT    | IAH1   | MAOA   | PGAM1   | SIAE    |
| ABCD3  | AKR1A1  | BLVRA   | FAH    | GNS     | IDH1   | MBOAT7 | PGAM5   | SLC1A4  |
| ABHD12 | AKR7A2  | BLVRB   | FASN   | GOT1    | IMPAD1 | MOCS1  | PGD     | SLC44A1 |
| ACAA1  | ALAD    | CLIC1   | FBP1   | GSS     | INPP4B | MPI    | PGK1    | SMPD3   |
| ACADS  | ALDH2   | CLIC6   | FUT8   | GSTM2   | INPP5J | NAGK   | PGLS    | SORD    |
| ACADSB | ALDH4A1 | CMBL    | G6PD   | GSTM3   | INPP5K | NANS   | PGM2    | SPR     |
| ACBD3  | ALDOA   | CNDP2   | GAA    | GSTP1   | KMO    | NIT1   | PHYHD1  | SQRDL   |
| ACLY   | ALOX15B | CP      | GALNT2 | GSTZ1   | LDHB   | NIT2   | PIP4K2C | STARD10 |
| ACO1   | APEH    | CYB5A   | GCLC   | GUSB    | LSS    | NQO1   | PKM     | STARD5  |
| ACOX3  | ARSD    | CYB5R1  | GFPT1  | GYG1    | LTA4H  | NSDHL  | POFUT1  | TYMP    |
| ACSF2  | ASAHI   | DAK     | GLA    | HGD     | LYPLA1 | NT5C   | PRODH   | UGDH    |
| ACSS3  | ASL     | DDT     | GLCE   | HNMT    | LYPLA2 | NUDT18 | PTER    |         |
| AGA    | ASNA1   | DEGS2   | GLO1   | HSD17B4 | MAN2A1 | NUDT5  | PTGR2   |         |
| AGL    | ATIC    | DHRS7   | GLUD1  | HSD3B7  | MAN2B1 | P2RX4  | SCP2    |         |

**Table 2.3 – Basal Metabolic Cluster**

|         |         |        |        |        |        |       |          |         |
|---------|---------|--------|--------|--------|--------|-------|----------|---------|
| ABCB1   | ALDH3A1 | ATP5G2 | DPYSL3 | IDH3A  | MAOB   | NPC1  | PKM      | SLC27A6 |
| ACOX2   | ALDH7A1 | CBR1   | GAPDH  | IDH3G  | MICAL3 | PDE4D | SARDH    | SLC2A1  |
| ADH5    | ALDOC   | CKB    | GCLM   | ISYNA1 | MME    | PDE6D | SLC12A4  | SLC3A2  |
| AKR1B1  | ATP2B1  | CRYM   | GNPDA1 | ITPR1  | MSRB3  | PFKM  | SLC25A12 | SRM     |
| ALDH1L2 | ATP2B4  | DPYSL2 | HAAO   | LPCAT2 | NNMT   | PGM1  | SLC27A1  | TF      |

**Table 2.4 – Luminal Progenitor Metabolic Cluster**

|         |          |          |        |         |         |        |          |         |
|---------|----------|----------|--------|---------|---------|--------|----------|---------|
| ABCE1   | ALDH9A1  | ATP6V1G1 | DBT    | HADH    | MTHFD1L | NDUFS2 | PDXK     | SLC2A4  |
| ABCF1   | ASRGL1   | ATP6V1H  | DDOST  | HADHA   | NDUFA10 | NDUFS3 | PHGDH    | SLC34A2 |
| ABHD11  | ASS1     | B3GNT3   | DHODH  | HADHB   | NDUFA12 | NDUFS4 | PPA2     | SLC41A3 |
| ACAA2   | ATP1A1   | BCAT2    | DLAT   | HCCS    | NDUFA13 | NDUFS5 | PPAT     | SLC6A14 |
| ACAD8   | ATP1B1   | BPNT1    | DLD    | HIBADH  | NDUFA2  | NDUFS6 | PPOX     | SOD2    |
| ACAD9   | ATP5A1   | CA3      | DUT    | HIBCH   | NDUFA4  | NDUFS7 | PTGES2   | SUCLA2  |
| ACADVL  | ATP5B    | CAD      | ECHDC1 | HSD11B1 | NDUFA5  | NDUFS8 | RDH14    | SUCLG1  |
| ACAT1   | ATP5C1   | CHKA     | ECHDC3 | IDH2    | NDUFA7  | NDUFV1 | SDHA     | THNSL1  |
| ACO2    | ATP5D    | CMAS     | ECHS1  | INPP5D  | NDUFA8  | NDUFV2 | SDHB     | TMLHE   |
| ACSF3   | ATP5F1   | COQ5     | FAHD2A | ITPR2   | NDUFA9  | NDUFV3 | SDHC     | TXNRD1  |
| ACSL1   | ATP5H    | COQ9     | FDFT1  | ITPR3   | NDUFAF2 | NME2   | SDR39U1  | UGT8    |
| ACSS1   | ATP5I    | COX4I1   | FECH   | KCNN4   | NDUFB10 | NOS2   | SFXN1    | UQCRB   |
| ADSL    | ATP5J    | COX5A    | FH     | KCTD14  | NDUFB4  | NUDT19 | SHMT1    | UQCRC1  |
| AGK     | ATP5J2   | COX5B    | FHIT   | MCCC1   | NDUFB5  | OAT    | SHMT2    | UQCRC2  |
| AGPAT9  | ATP5L    | COX6B1   | FOLH1  | MDH2    | NDUFB6  | OXCT1  | SLC12A2  | UQCRFS1 |
| AHCYL1  | ATP5O    | COX7A2   | GALK1  | ME2     | NDUFB7  | OXSM   | SLC25A1  | UQCRH   |
| AK2     | ATP6V1A  | CPT2     | GCDH   | MECR    | NDUFB8  | PC     | SLC25A10 | UQCRQ   |
| AKR1C3  | ATP6V1B1 | CS       | GMPS   | MT-ND4  | NDUFB9  | PDHA1  | SLC25A13 | VDAC1   |
| ALDH1A3 | ATP6V1B2 | CYC1     | GPD1   | MT-ND5  | NDUFC2  | PDHB   | SLC25A22 | XDH     |
| ALDH1B1 | ATP6V1E1 | CYCS     | GPHN   | MTAP    | NDUFS1  | PDHX   | SLC25A5  |         |

### 2.3.3 Pathway Analysis using Enrichr

Pathway analysis of metabolic clusters was conducted using Enrichr (<https://amp.pharm.mssm.edu/Enrichr/>). Enrichr is a comprehensive gene set enrichment tool that is available both as a web interface<sup>133</sup> and an R package<sup>127</sup>. It queries a list of gene symbols and returns commonly annotated pathways by searching large gene set libraries. The gene set library selected for our analysis was Gene Ontology Biological Process (GOBP) 2018. For each cell-type signature, the top 10 GOBP terms enriched by gene sets were sorted by lowest to highest combined score ( $\ln(\text{p-value}) * z\text{-score}$ ), a metric used by Enrichr to find the best ranking terms compared to other methods.

### 2.3.4 Correlations to PAM50 Breast Cancer Subtypes

Gene expression for PAM50 and Claudin-low breast cancer subtypes and clinical annotations was completed in the METABRIC study<sup>134</sup> and obtained from cBioPortal<sup>135,136</sup>. It provided gene expression profiles and classified breast cancer subtypes for 1980 patients. The gene expression

profiles for the breast cancer subtypes (specifically, for Her2, Luminal A, Luminal B, Basal-like, and Claudin-low) were correlated to our metabolic signatures via single-sample Gene Set Expression Analysis (ssGSEA) using the “GSVA” R package (version 1.30.0). ssGSEA scores for each signature in the breast cancer subtypes were assessed for significance using a one-way ANOVA and Tukey’s multiple comparisons test<sup>137</sup>.

## **2.4 Mouse Strains**

Virgin female FVB wild-type mice, 8-12 weeks of age were purchased either from The Jackson Laboratory or Charles River. Mice were ovariectomized bilaterally, then allowed one week to recover. A slow-release 0.14 mg 17- $\beta$  estradiol plus 14 mg progesterone pellet (Innovative Research of America) was then placed subcutaneously near the thoracic mammary gland for 2 weeks. This was done to obtain large quantities of viable mammary stem/progenitor cells for subsequent analysis, as previously reported<sup>122,138</sup>. All mice were cared for according to guidelines established by the Canadian Council for Animal Care under protocols approved by the Animal Care Committee of the Ontario Cancer Institute.

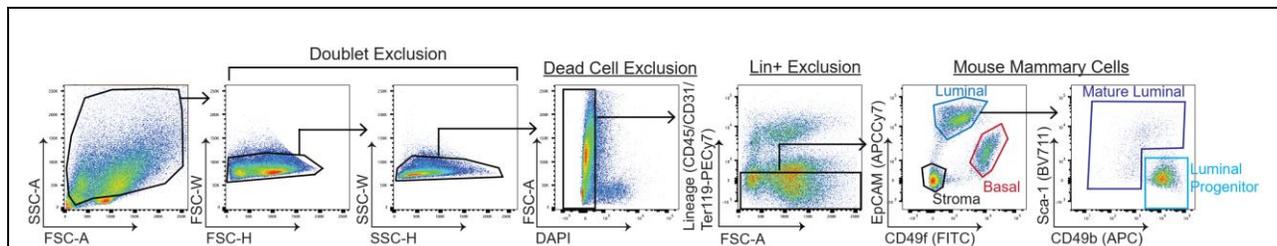
## **2.5 Mouse Mammary gland dissociation, single-cell preparation and FACS analysis**

Harvested mammary glands were manually minced with scissors for 2 minutes, and then enzymatically dissociated using 750 U/ml collagenase, 250 U/ml hyaluronidase (Stem Cell Technologies, 07912) and diluted in DMEM:F12 for 1.5 hours. Samples were vortexed at the 1- and 1.5-hour mark. Red blood cells were lysed using an ammonium chloride solution (Stem Cell Technologies, 07850). Cells were then mixed in trypsin-EDTA (0.25%, Stem Cell Technologies, 07901) that had been pre-warmed to 37°C using a P1000 pipette for 2 minutes. Cells were then washed in Hanks Balanced Salt Solution (HBSS) without calcium or magnesium plus 2% FBS and centrifuged. Cells were then similarly mixed in dispase 5 U/ml (Stem Cell Technologies, 07913) plus 50  $\mu$ g/ml DNase I (Sigma, D4513) for 2 minutes, washed in HBBS + 2% FBS and filtered using a 40  $\mu$ m cell strainer to obtain single cells.

For FACS staining antibodies were against the following antigens were used: TER119 (PECy7 or eFluor450), CD31 (PECy7 or eFluor450), CD45 (PECy7 or eFluor450), EpCAM (APCCy7), CD49f (FITC or PECy7), CD49b (PE) and Sca-1 (APC or Brilliant Violet 711). Lineage (Lin) positive cells were defined as Ter119<sup>+</sup>CD31<sup>+</sup>CD45<sup>+</sup>. Mouse mammary cell subpopulations were defined as: total basal (Lin<sup>-</sup>EpCAM<sup>lo-med</sup>CD49f<sup>hi</sup>); total luminal (Lin<sup>-</sup>EpCAM<sup>hi</sup>CD49f<sup>lo</sup>); luminal progenitor (Lin<sup>-</sup>EpCAM<sup>hi</sup>CD49f<sup>lo</sup>CD49b<sup>+</sup>Sca-1<sup>-</sup>); and mature luminal (Lin<sup>-</sup>EpCAM<sup>hi</sup>CD49f<sup>lo</sup>CD49b<sup>-/+</sup>Sca-1<sup>-/+</sup>). An example is given in Figure 2.2. High and low mitochondrial activity populations were defined as MitoTracker Red<sup>Hi</sup>MitoTracker Green<sup>hi</sup> and MitoTracker Red<sup>lo</sup>MitoTracker Green<sup>hi</sup>, respectively, and applied after gating for total luminal and basal. Dead cells were excluded following doublet exclusion using DAPI or Zombie UV Fixable Viability Kit (BioLegend) according to manufacturer's instructions. Fluorophores are specifically mentioned in figures. Cell sorting was performed on a BD FACSAria™ II.

## 2.6 Measuring Oxygen Consumption Rate

Luminal progenitor, mature luminal and basal cell populations were FACS-purified from unstaged mice and 10,000 cells plated into each well of collagen pre-coated Seahorse plates. The cells were then grown in a 5% oxygen incubator for 6 days to reach at least 80-90% confluence. On the 7<sup>th</sup> days, cells were switched to DMEM:HAM's F12 with no bicarbonate containing 5% FBS, insulin (Thermo Fisher, 12585014), EGF (STEMCELL Technologies; 78006.1), bFGF (STEMCELL Technologies), hydrocortisone (STEMCELL Technologies, 78003.1), and Rock inhibitor (Millipore, SCM075) under 5% oxygen conditions. The plate was allowed to equilibrate for 1 hour in the Seahorse incubator. Inhibitors used for the assay include Oligomycin (2 μM),



**Figure 2.2 - Gating strategy for FACS-purifying mouse mammary epithelial cells**

Total cells from dissected mouse mammary glands are gated to exclude debris. Doublet, dead cell and Lin+ exclusion ensures single, live and non-immune cells are analyzed, respectively. Abbreviations used: FSC, Forward Scatter; SSC, Side Scatter; A, Area; W, Width; H, Height; DAPI, 4',6-diamidino-2-phenylindole; PE, Phycoerythrin; Cy7, Cyanine7; APC, Allophycocyanin; FITC, Fluorescein isothiocyanate; BV711, Brilliant Violet™ 711.

FCCP (1  $\mu$ M, Sigma, C2920) and Antimycin A (1  $\mu$ M). After the assay, cell viability was determined using CyQUANT nuclear dye (Thermo Fisher, C35007). Data was analyzed on the WAVE platform and normalized to number of live cells determined after the viability assay.

## **2.7 Intracellular flow cytometry**

All intracellular dyes were used to stain cells prior to cell surface marker staining protocol. Staining for total mitochondria (50 nM MitoTracker Green FM, Thermo Fisher, M7514), mitochondrial activity (250 nM MitoTracker Red CMXRos, Thermo Fisher, M7513), mitochondrial ROS (5  $\mu$ M MitoSOX, Thermo Fisher, M36008), and cytosolic ROS (5  $\mu$ M CellROX Green, Thermo Fisher, C10492) was performed by incubating cells at 37°C for 20-30 minutes following manufacturer's instructions and directly analysed without fixing. Cell analysis was performed in BD Biosciences Fortessa. Median fluorescent intensity (MFI) refers to the fluorescence intensity of each event (on average) of the selected cell population, in the chosen fluorescence channel (PE Texas Red or FITC) and was determined using FlowJo, a flow cytometry analysis software.

## **2.8 Transmission electron microscopy (TEM)**

Mammary epithelial cells were FACS-purified from 3 EP-treated ovariectomized 8-12 week-old mice. Cells were pooled together to get a greater yield and then pelleted for 5 mins at 4°C (at max speed). Supernatant was removed and the pellet fixed with 2% glutaraldehyde in 0.1M sodium cacodylate buffer pH 7.3 without disturbing the pellet. Samples were processed by the Nanoscale Biomedical Imaging Facility (SickKids, Toronto, ON). Images were acquired using the FEI Technai 20 transmission electron microscope. Scale bars are specific to images.

## **2.9 In vitro colony forming cell (CFC) assay**

350 cells of specified FACS-purified populations were seeded together with 20,000 irradiated NIH 3T3 cells in a 6-well plate. Cells were cultured for 7 days in EpiCult-B mouse medium (Stem Cell Technologies, 05610) supplemented with 5% FBS, 10 ng/ml EGF, 20 ng/ml basic FGF, 4  $\mu$ g/ml heparin, and 5  $\mu$ M ROCK inhibitor (Millipore). Cells were allowed to adhere for 24 hours, then either vehicle control (0.1% DMSO) or the indicated concentrations of inhibitors were added for the remaining six days.

## 2.10 Drug testing in vitro

Vehicle or drugs were added such that the final concentration of DMSO did not exceed 0.1% (vol/vol). The following drugs were purchased from the companies in the brackets: 2-Deoxy-D-glucose (Sigma; D8375), Dichloroacetate (Sigma; 347795), BAY-876 (Structural Genomics Consortium), rotenone (Sigma; R8875), tigecycline (CarboSynth, 220620-09-7), Antimycin A (Sigma, A8674), Oligomycin (Sigma, 75351), Atpenin A5 (Cayman Chemicals, 11898), UK-5099 (Sigma, PZ0160), Galloflavin (Sigma; SML0776). Vehicle or drugs were added such that the final concentration of DMSO did not exceed 0.1% (v/v).

## 2.11 Statistical Analysis

All data is reported as mean  $\pm$  SEM. Calculations were completed using GraphPad Prism software (v7.00). Comparisons were made using one-way ANOVA with Tukey's multiple comparison test, two-way ANOVA with Sidak's multiple comparison test, or a two-tailed student's *t*-test. A two-way ANOVA was used whenever an experimental plan included a comparison between more than two groups followed by Bonferroni post-tests between groups of interest. Student *t*-tests were utilized wherever comparisons between only two groups were made.

## Chapter 3

### 3 Results

The data in this chapter is in part contributed by the following:

**Davide Pellacani & Connie Eaves** – Collecting and FACS-purifying the 10 human reduction mammoplasty samples into BC, LP and ML.

**Alison Casey, Vladimir Ignatchenko, Ankit Sinha & Thomas Kislinger** – Responsible for running UPLC-MS with subsequent processing of the proteomic dataset.

**Luis Palomero, Mar Garcia Valero & Miquel Pujana** – for providing the code, gene expression and copy-number data to perform the correlational analysis to breast cancer subtypes.

**Kazeera Aliar** - for performing all the in-house bioinformatics and statistical testing using the human proteomics.

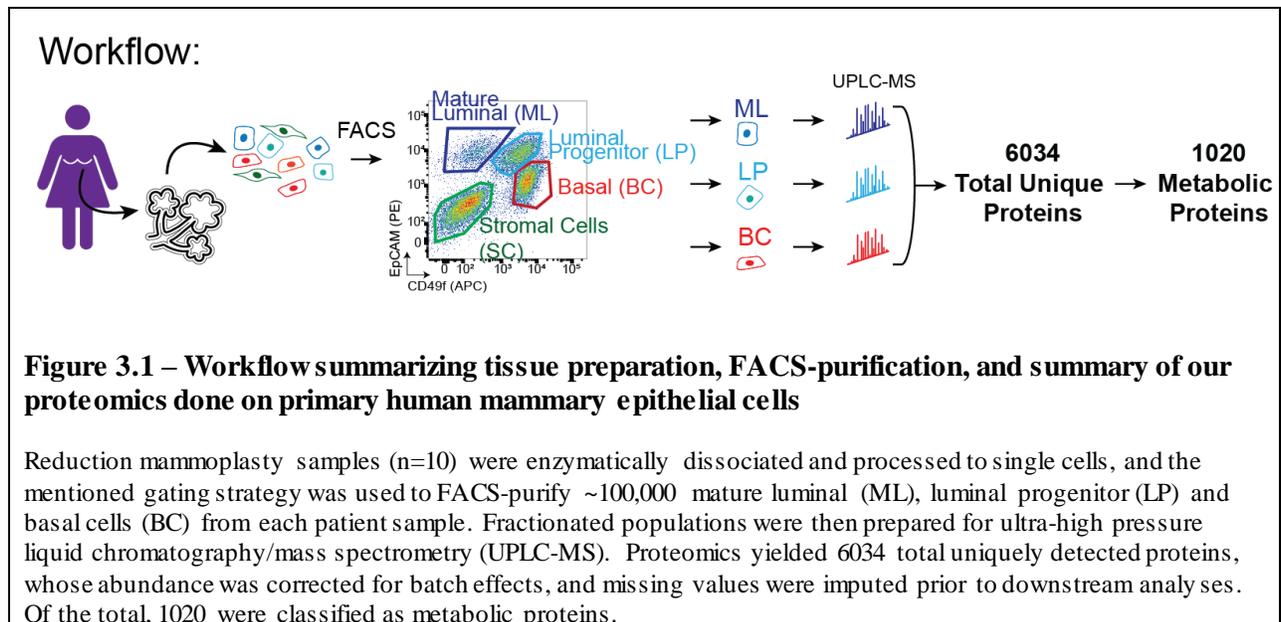
**Mina Alam** – for providing her artistic touch on the metabolic map presented in Figure 3.6.

**Alison Casey, Hyeyeon Kim, Pirashaanthi Tharmapalan & Swami Narala** - for helping with the FACS experiments and design of flow cytometry experiments.

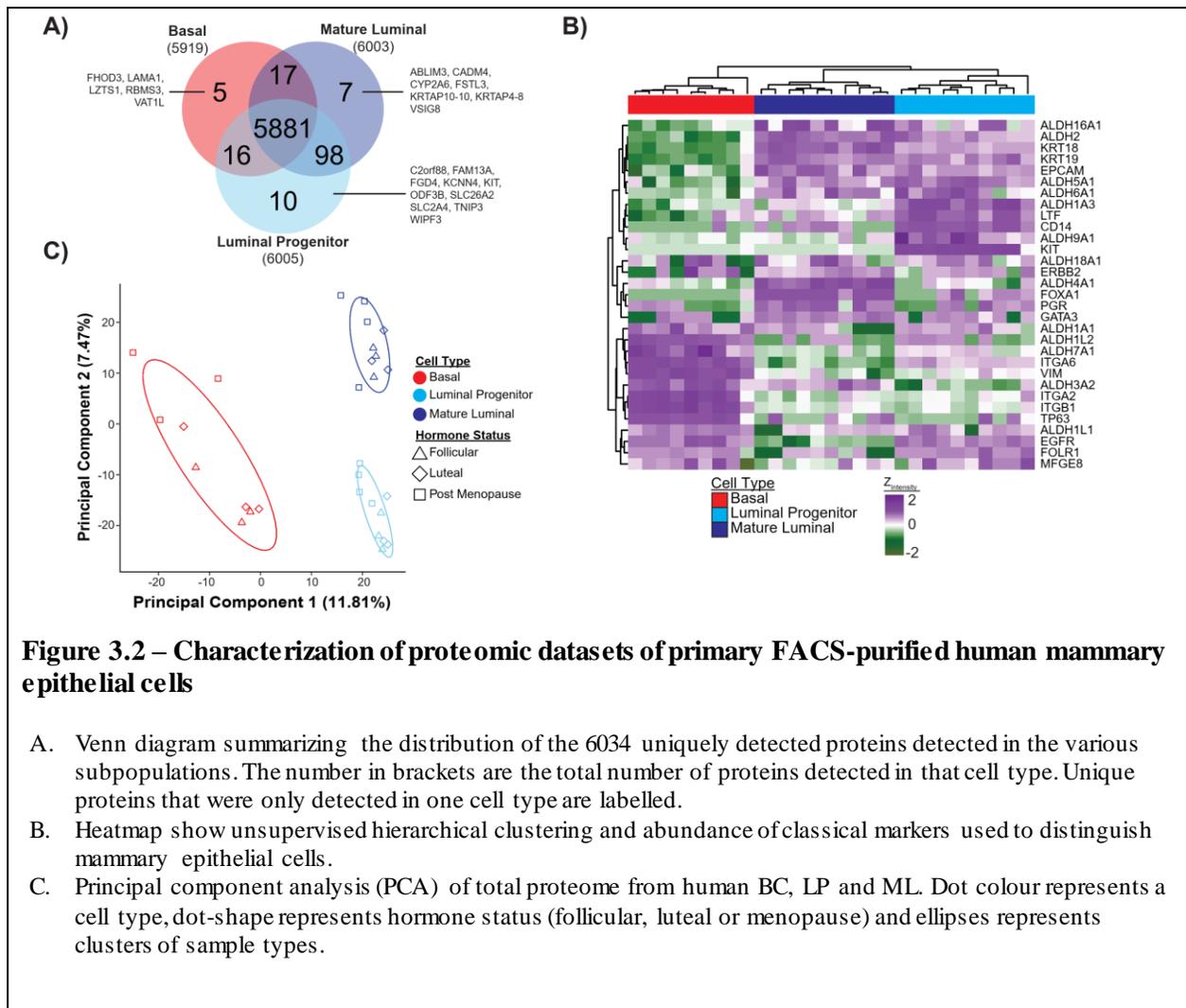
### 3.1 Proteomics reveals differential expression of metabolic proteins

To determine the metabolic network functioning within mammary subpopulations, we performed mass spectrometry-based shotgun proteomics on equivalent numbers of FACS-purified mammary subpopulations (BC, LP and ML) from 10 normal human breast samples (Figure 3.1). Our patient cohort represented diverse physiologies, covering a wide age range (28 to 67 years old) and hormone statuses (3 luteal, 3 follicular and 4 postmenopausal) (Table 2.1). There were 6034 uniquely detected proteins, 5881 of which were detected in all three subpopulations (Figure 3.2A).

Expression of known markers for each mammary subpopulation was accurately captured by our proteomics data (Figure 3.2B), similar to before<sup>122</sup>. BC showed higher abundance for Vimentin and ITGA6 (Integrin  $\alpha 6$ , CD49f), LP had higher KIT and ALDH1A3<sup>11</sup>, and ML demonstrated higher GATA3, FOXA1 and KRT8/18 (Cytokeratin 8/18)<sup>139</sup>. Principal component analysis (PCA) demonstrated that mammary subpopulations have distinct proteomes, where the first and second components explained 11.81% and 7.47% of the total variation, respectively (Figure 3.2C). There is some minor segregation of post-menopausal samples away from premenopausal samples. However, the dominant clustering feature was cell type, not hormone status (Figure 3.2C). We separated the total proteomes of each subpopulation into deciles based on median intensity (Figure 3.3A). Then we performed pathway analysis using Enrichr for each cell type to

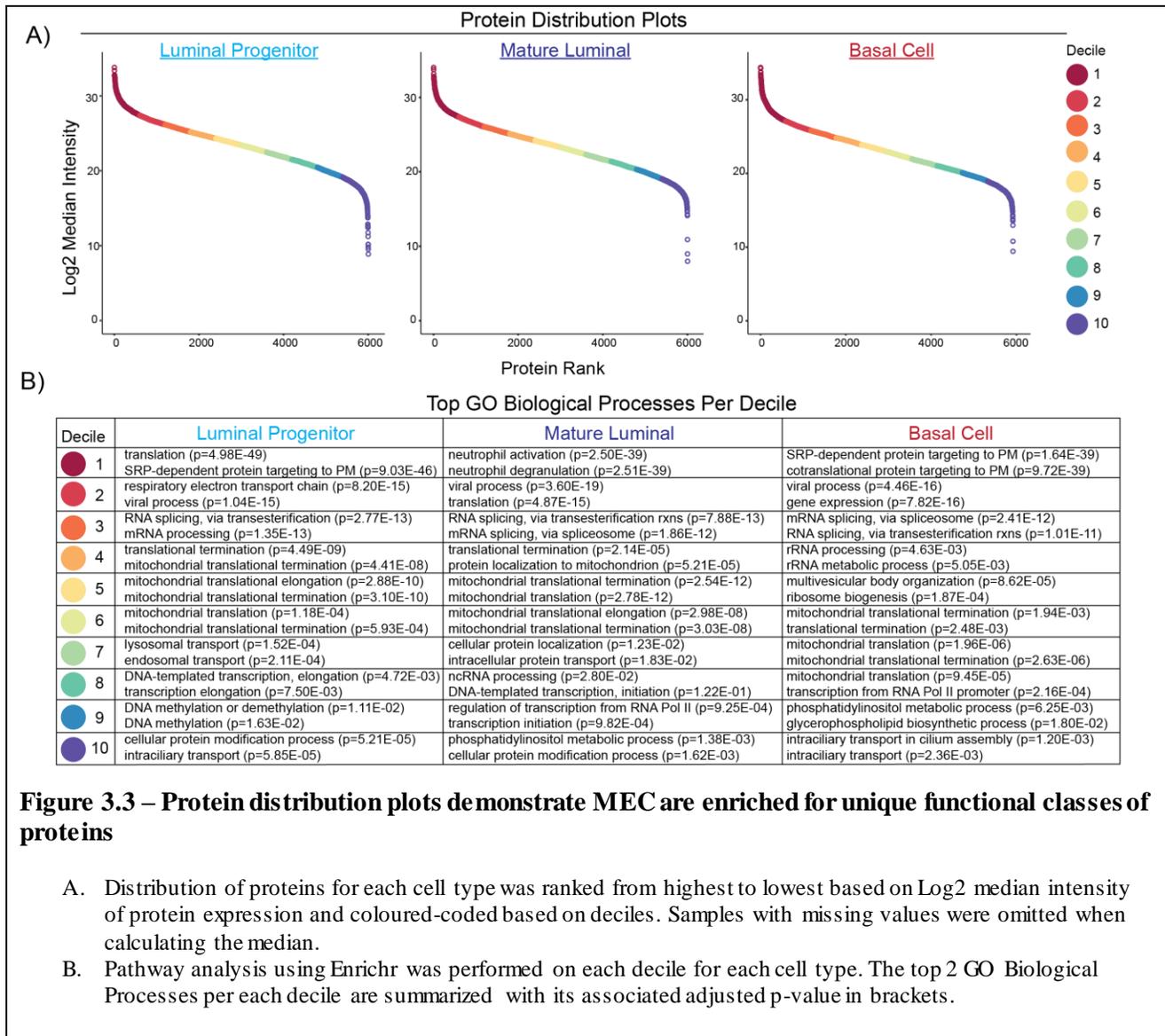


identify pathways that were enriched in each decile (Figure 3.3B). Common terms enriched in the top decile for BC and LP were those relating to translation such as “SRP-dependent protein targeting to the plasma membrane (PM).” However, there were several unique terms that appeared in top deciles of only one subpopulation. LPs were the only cell type to demonstrate enrichment for “respiratory electron transport chain,” which showed up in the second decile. In addition, the top decile of ML enriched for pathways relating to neutrophil activity such as “neutrophil activation” and “neutrophil degranulation.”



**Figure 3.2 – Characterization of proteomic datasets of primary FACS-purified human mammary epithelial cells**

- Venn diagram summarizing the distribution of the 6034 uniquely detected proteins detected in the various subpopulations. The number in brackets are the total number of proteins detected in that cell type. Unique proteins that were only detected in one cell type are labelled.
- Heatmap show unsupervised hierarchical clustering and abundance of classical markers used to distinguish mammary epithelial cells.
- Principal component analysis (PCA) of total proteome from human BC, LP and ML. Dot colour represents a cell type, dot-shape represents hormone status (follicular, luteal or menopause) and ellipses represents clusters of sample types.



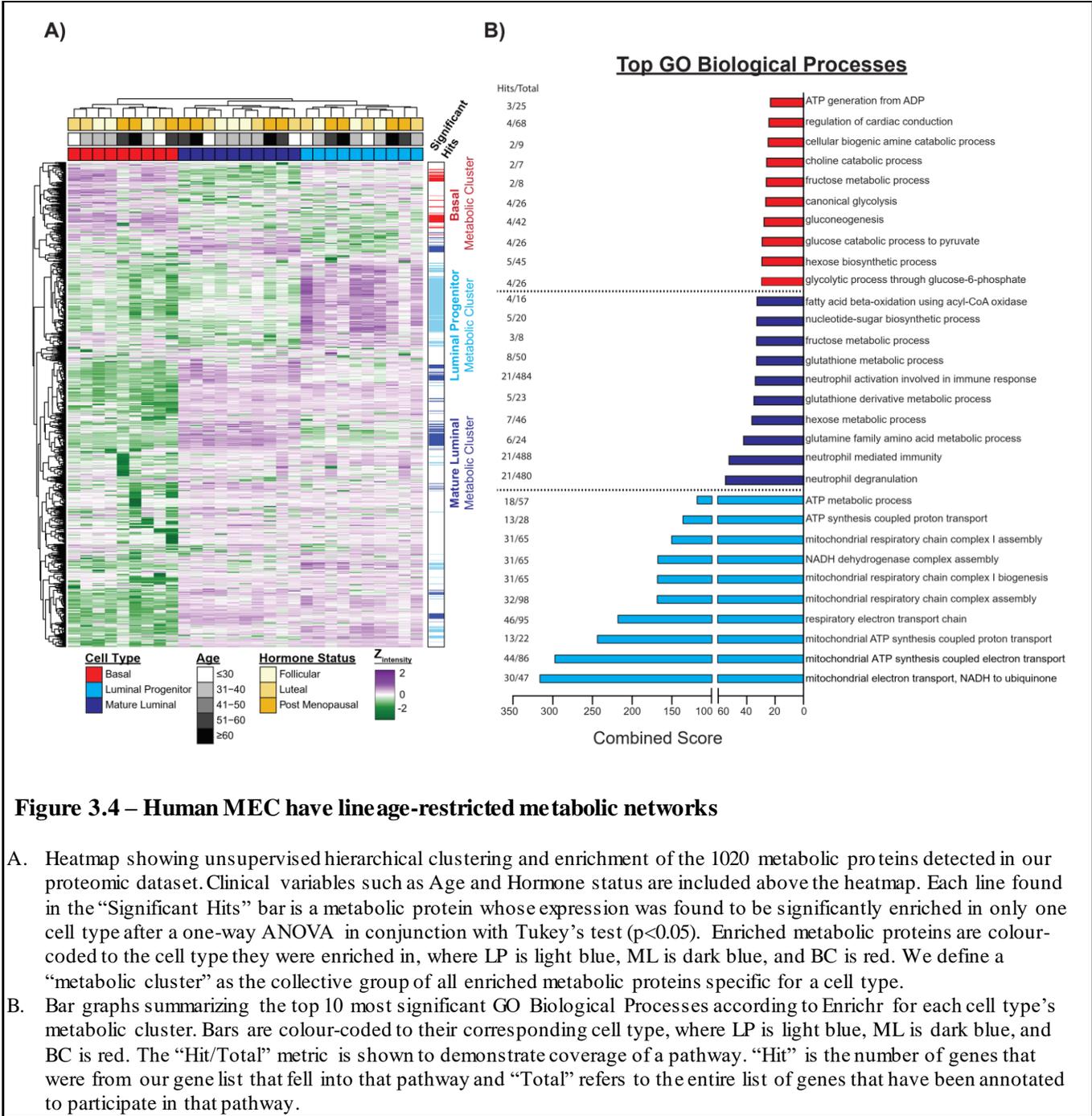
**Figure 3.3 – Protein distribution plots demonstrate MEC are enriched for unique functional classes of proteins**

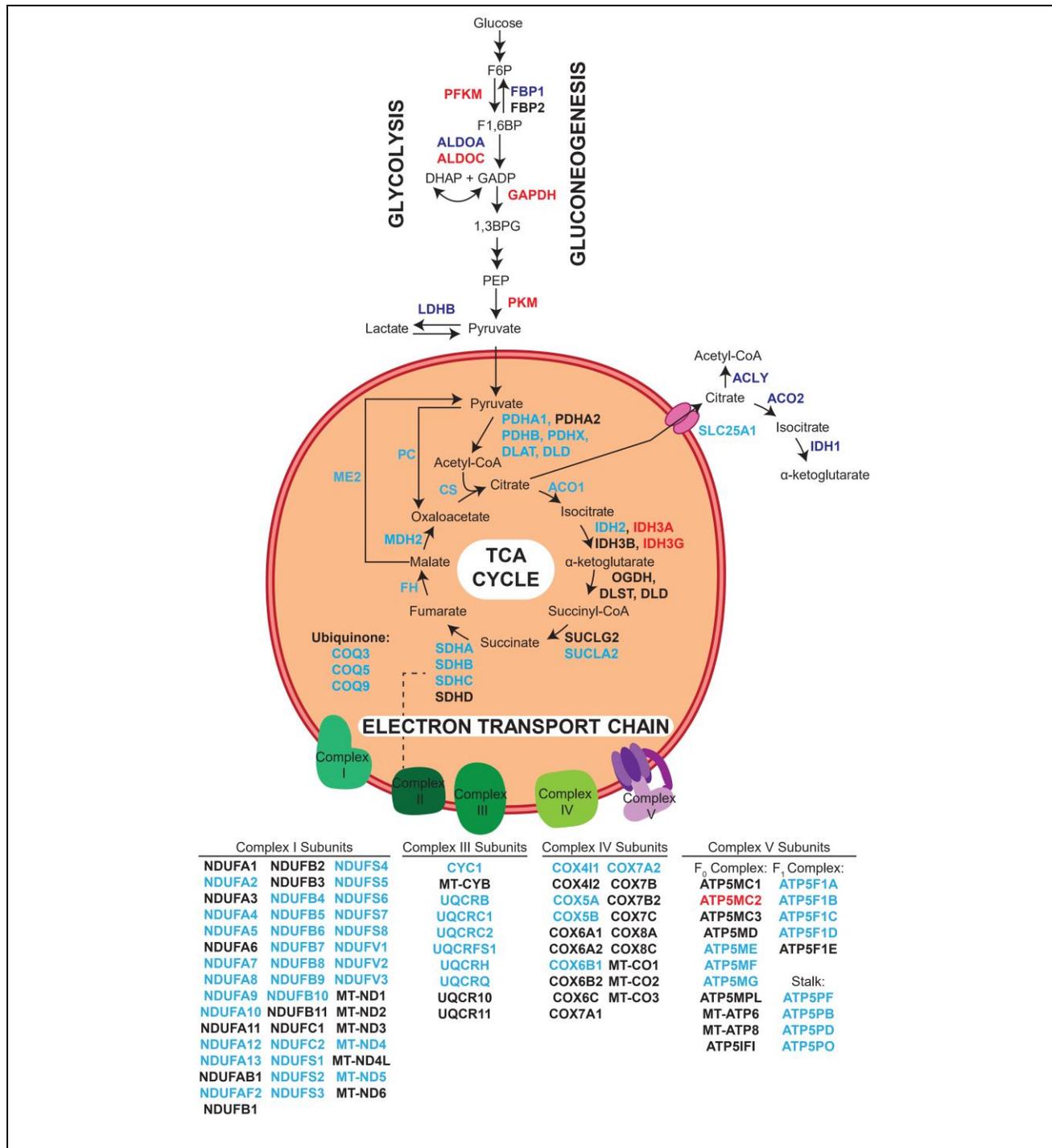
- Distribution of proteins for each cell type was ranked from highest to lowest based on Log2 median intensity of protein expression and coloured-coded based on deciles. Samples with missing values were omitted when calculating the median.
- Pathway analysis using Enrichr was performed on each decile for each cell type. The top 2 GO Biological Processes per each decile are summarized with its associated adjusted p-value in brackets.

## 3.2 Mammary subsets have distinct metabolic preferences

A metabolic network is defined as the core set of metabolic enzymes that is essential to support the structure and function of a cell<sup>19</sup>. To determine whether mammary epithelial cells were driven by distinct metabolic networks, we filtered total proteomes to look solely at metabolic proteins using a curated list of over 2753 annotated metabolic genes<sup>132</sup>. One sixth (1020/6034) of the total uniquely detected proteins were classified as those related to some metabolic function (Figure 3.1). Despite the heterogeneity within our patient cohort, unsupervised hierarchical clustering of 1020 metabolic proteins demonstrated clusters primarily based on cell type (Figure 3.4A), similar to our total proteome analysis (Figure 3.2C). This suggests that each mammary cell type operates on a unique hardwired metabolic network. We saw that each of our cell types demonstrated unique abundance for metabolic proteins that were not seen in the other cell types. To identify metabolic proteins that were uniquely enriched in one cell type versus the other two, we performed a one-way ANOVA in conjunction with a Tukey's test ( $P < 0.05$ ). In this way, we defined the set of all significantly enriched metabolic proteins for a particular subpopulation as that cell type's metabolic cluster. ML, BC and LP metabolic clusters were composed of 123, 45 and 179 metabolic proteins, respectively (Table 2.2, 2.3 & 2.4).

Next, we performed pathway analysis using Enrichr to determine if proteins in each metabolic cluster enriched for a particular pathway. The top 10 enriched GO Biological Processes for each subset are shown in Figure 3.4B. Glycolysis was the most prominent pathway to show up in BC and also pathways such as "regulation of cardiac conduction" which may underlie the contractile function of this cell type (Figure 3.4B). Our results are consistent with a previous report on high transcript levels of glycolytic enzymes in fetal MaSCs<sup>46</sup>. Though our input was limited to metabolic proteins, several neutrophil related pathways were enriched in the ML metabolic cluster (Figure 3.4B). ML also showed enrichment for pathways relating to glutamine and sugars such as hexose and fructose. The most prominent result from our pathway analysis was that the top 10 pathways for LP involved terms related to with OXPHOS (Figure 3.4B).





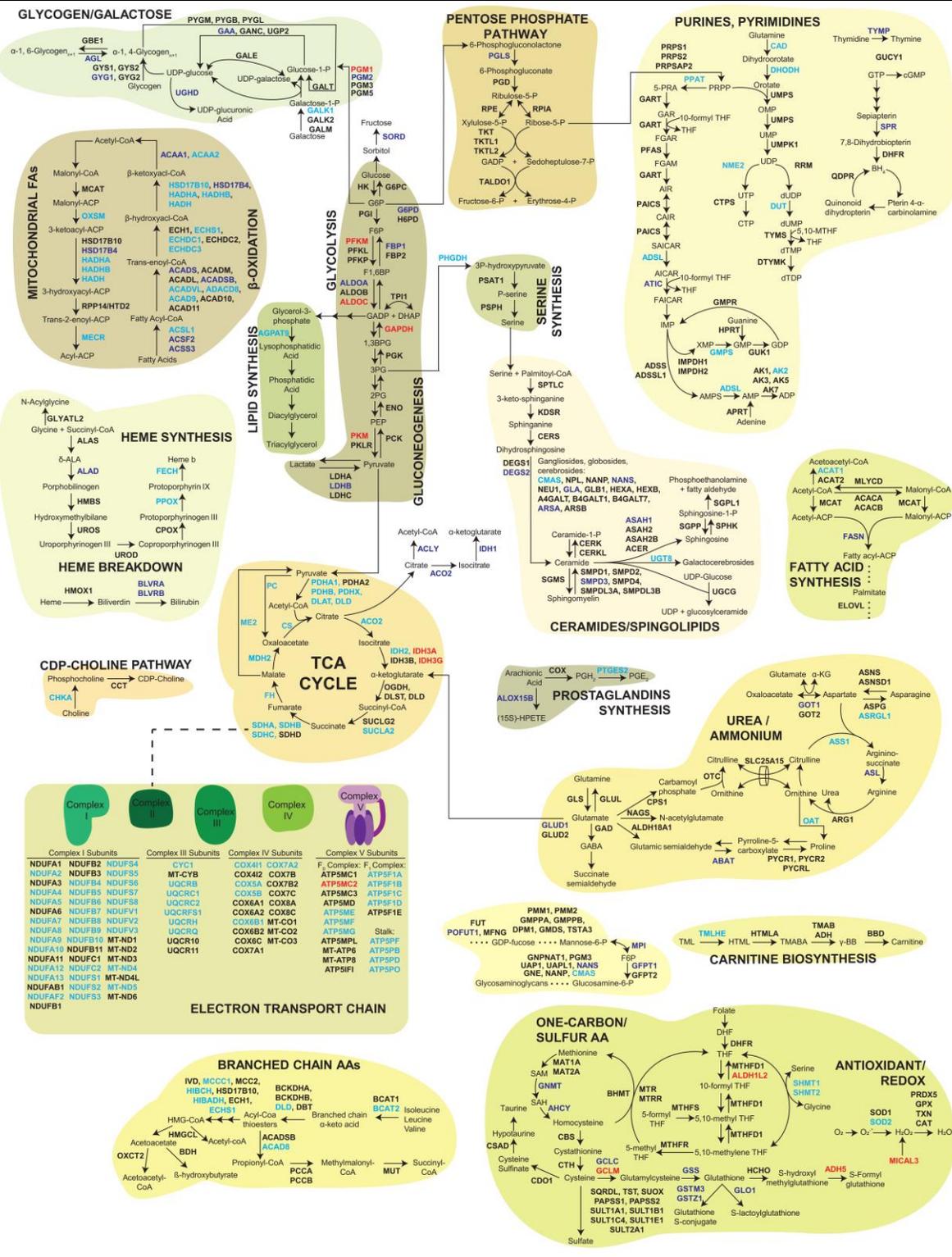
**Figure 3.5 – LP show enrichment for TCA Cycle and ETC proteins, whereas ML and BC are enriched more for glycolysis.**

Schematic representation of glycolysis, TCA cycle and ETC. We highlighted proteins that were found in our mammary cell-type specific metabolic cluster. Proteins are colour-coded to signify which metabolic cluster they were enriched in (LP, light blue; ML, dark blue; BC, red; not significant, black). Not all the reactions in Glycolysis are visualized. Only enzymes that were found in any of the metabolic clusters is included.

To gain a better understanding of this data, we colour-coded glycolysis, the TCA cycle and ETC subunits to signify which mammary cell-type specific metabolic cluster they were found in (Figure 3.5). LP demonstrated significant enrichment for the majority of ETC complex subunits as well as nearly all enzymes in the TCA cycle (Figure 3.5). Although glycolysis is presumed to be a major pathway that produces pyruvate, which subsequently enters the TCA cycle in the form of Acetyl-CoA, LP demonstrated no enrichment for any enzyme of this pathway. Conversely, glycolytic enzymes were enriched in BC (PFKM, ALDOC, GAPDH and PKM) and ML (FBP1, ALDOA and LDHB). LP show enrichment for enzymes in the TCA cycle, such as Malic Enzyme 2 (ME2), which produces pyruvate from malate, and Pyruvate carboxylase (PC), which makes oxaloacetate from pyruvate. Of particular note, mammary subpopulations demonstrated isoform-specific expression of IDH. Mitochondrial IDH3, which uses NADH as a cofactor, was higher in BC. LP were higher for IDH2, which is mitochondrial and uses NADPH as a cofactor<sup>77</sup>. Interestingly, the enzymes responsible for oxidation of citrate in the cytosol (ACO1, IDH1, ACLY) were more abundant in ML. The specific engagement of IDH isoforms suggests that each cell type has different levels of NAD(P)H and also different purposes for these enzymes.

To determine if LP were able to generate TCA intermediates through non-glycolytic means, we mapped out other proteins found in any of the clusters using a recently published metabolic map as a template<sup>140</sup> (Figure 3.6). Here, we were able to appreciate the numerous means by which LP cells are able to obtain mitochondrial Acetyl-CoA necessary for entry into the TCA Cycle. The LP metabolic cluster revealed strong enrichment for proteins involved in  $\beta$ -oxidation, which produces Acetyl-CoA from fatty acids. LP were the only cell type to demonstrate enrichment for enzymes involved in branched-chain amino acid catabolism, which entails modifying Isoleucine, Leucine and Valine into Acetyl-CoA (Figure 3.6). LP showed high expression of PHGDH, which is a key enzyme in serine biosynthesis. It was recently shown that serine produced by PHGDH gets converted to glutamine and contributes to almost 50% of the anaplerotic flux into the TCA Cycle<sup>132</sup>. This engagement in diverse metabolic pathways that break down different nutrients to enable continuation of the TCA Cycle highlights the importance of OXPHOS in LP. However, radiolabelled isotope tracing of specific nutrients is required to confidently conclude this. Altogether, proteomics demonstrated that each of the mammary lineages are uniquely

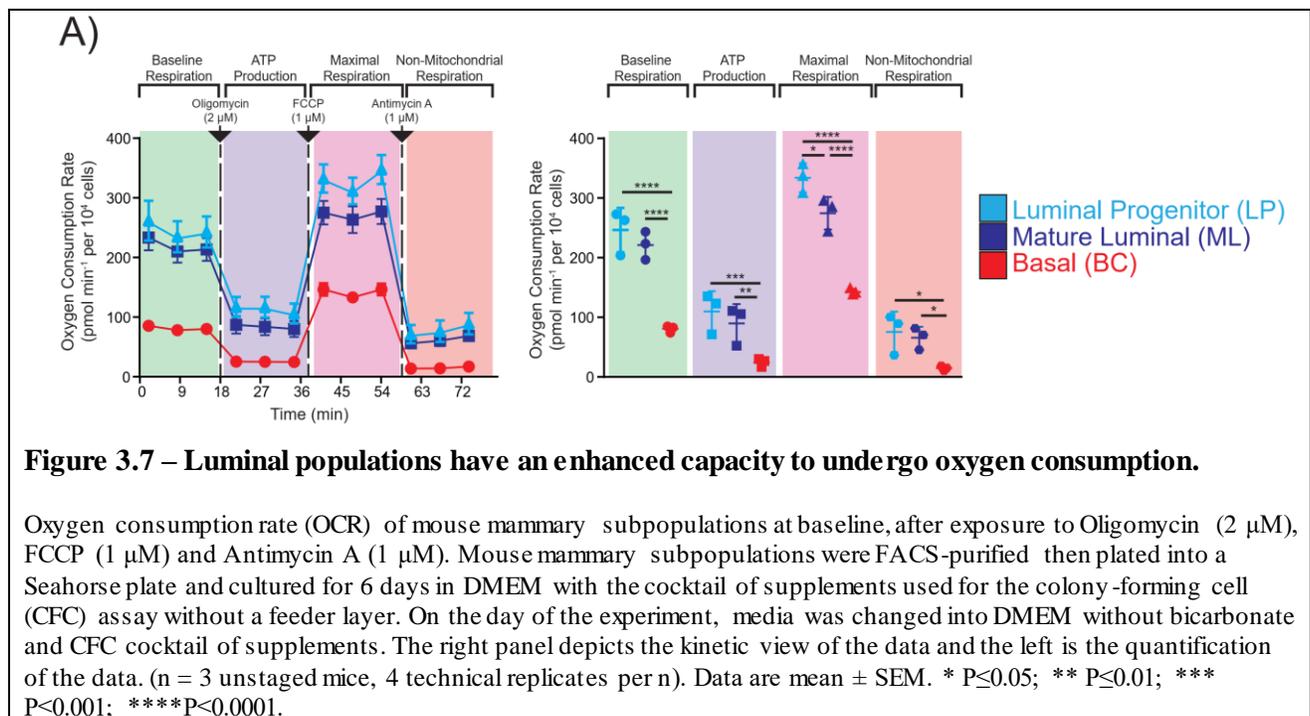
enriched for metabolic proteins that may be necessary for the structure and function of that cell. This ultimately suggests that mammary subpopulations possess lineage-restricted metabolic networks.



**Figure 3.6 –Map of mammary epithelial cell metabolism**  
 Metabolic map is adapted from Lin et al (2019)<sup>140</sup>. Major metabolic pathways are grouped together in the large circle. Proteins are colour-coded to denote which mammary cell-type specific cluster they demonstrate the highest expression level (Black = not significant or not tested, Light blue = LP, Dark blue = ML and red = Basal).

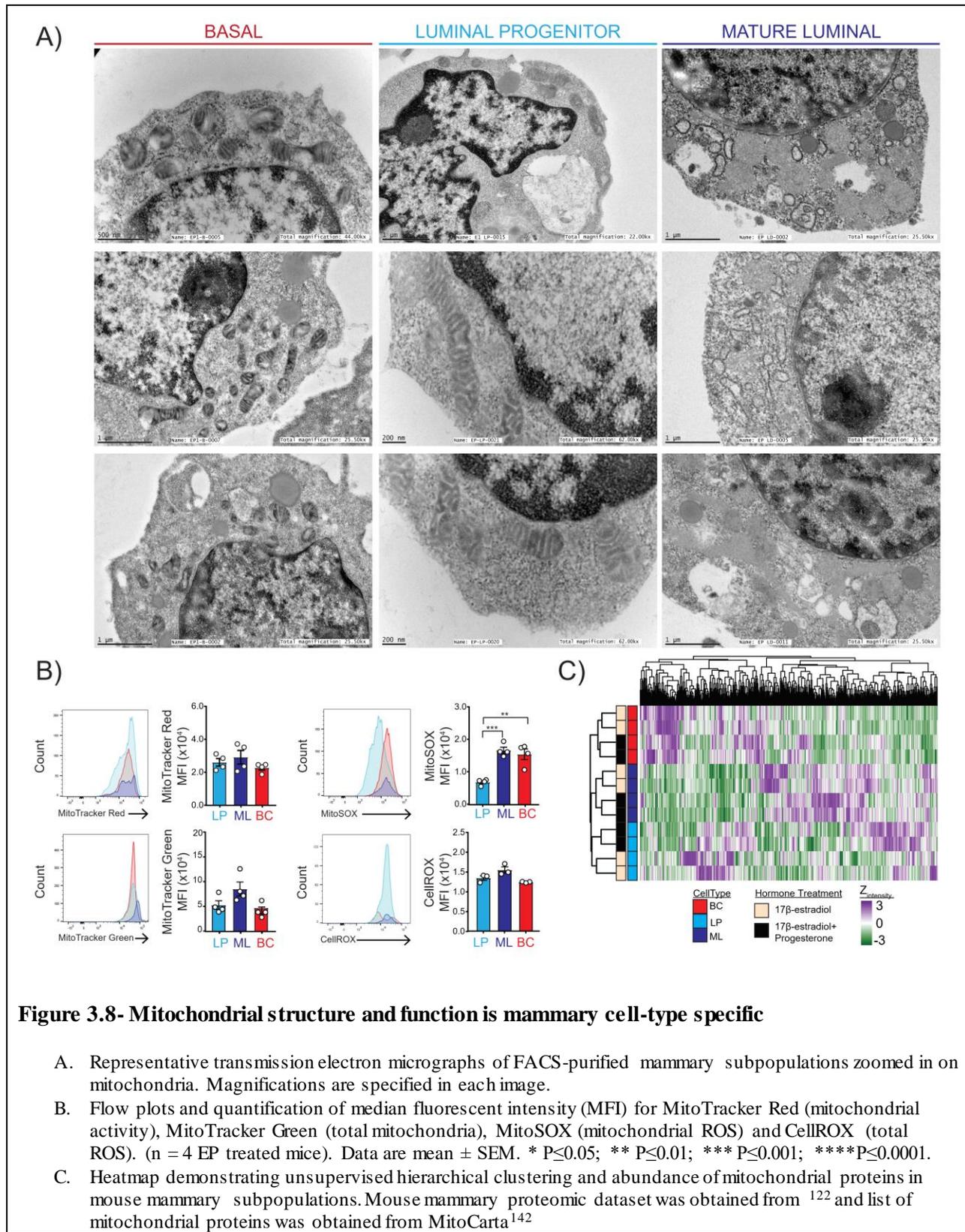
### 3.3 Mouse LP have enhanced capacity to undergo OXPHOS

As LP have high abundance for proteins participating in the ETC and TCA Cycle, we next asked whether this cell type showed an enhanced capacity to undergo OXPHOS. To this end, we tested for differences in metabolic capacity of mouse mammary subpopulations using the Seahorse bioanalyzer (Figure 3.7). Specifically, we performed a mitochondrial stress test, which entails measuring oxygen consumption rate (OCR), a readout for mitochondrial respiration, while exposing cells to inhibitors (Oligomycin, Antimycin A) or enhancers (FCCP) of this process. At baseline respiration, BC demonstrated the lowest level of OCR compared to either luminal populations (Figure 3.7). Even with the addition of FCCP, which dramatically boosts OCR, BC showed OCR levels comparable to, or less than, baseline OCR in luminal populations. LP and ML demonstrated similar mitochondrial respiration profiles, except for maximal respiration, which was significantly higher in LP. Collectively, this suggest that mammary subpopulations have distinct capacities for mitochondrial respiration, with LP showing the highest propensity to do so.



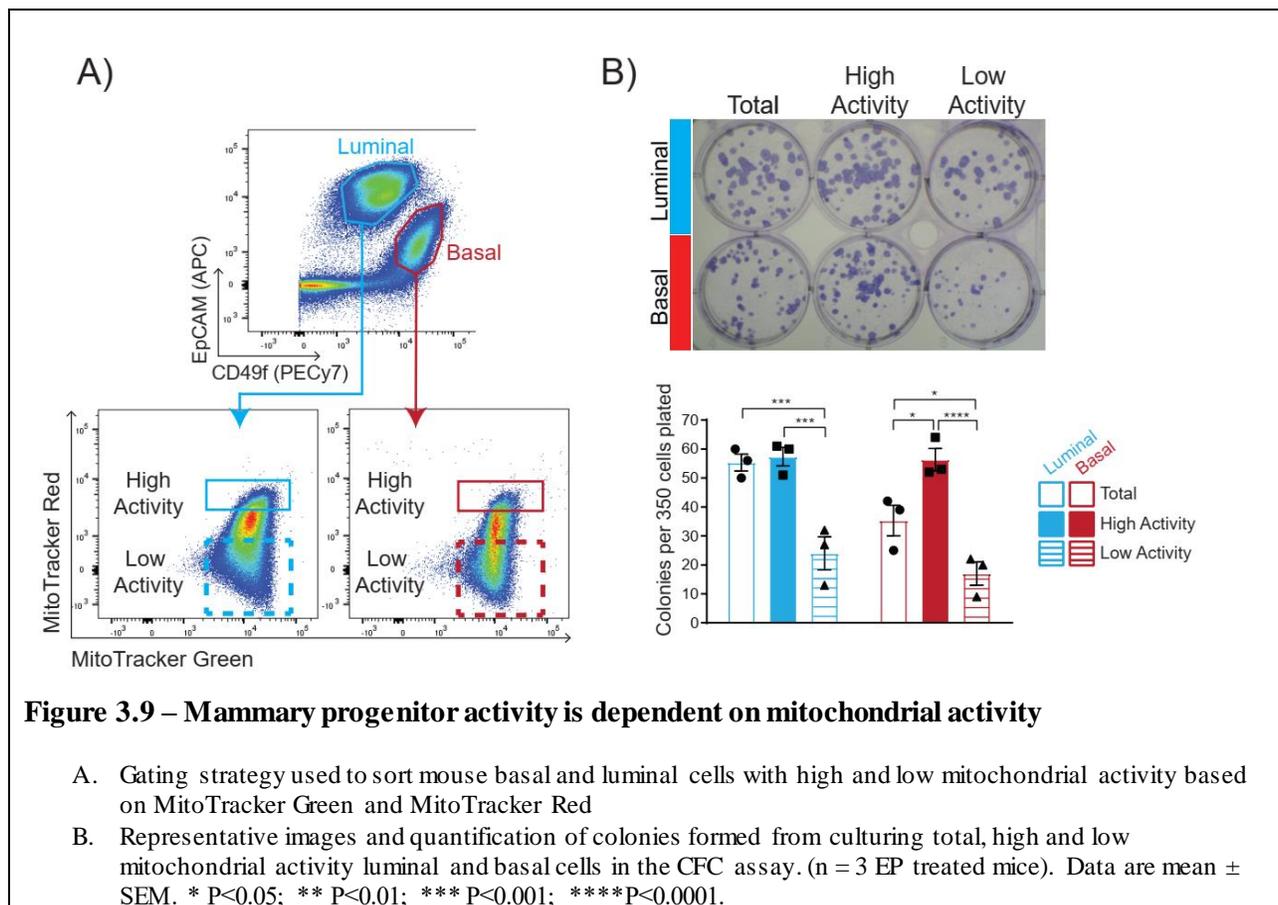
### 3.4 Mitochondrial structure and function are MEC-specific

Drastic differences in mammary cells to undergo OXPHOS led us to ask whether mitochondria showed mammary cell-type specific morphologies by transmission electron microscopy (TEM) (Figure 3.8A). BC tended to have several small circular mitochondria with a glossy cristae. This is similar to the morphology of mitochondria in hematopoietic and embryonic stem cells<sup>141</sup> (Figure 3.8A). LP, in contrast, had long, tubular mitochondria with elaborate cristae (Figure 3.8A). This structure is thought to be efficient to support OXPHOS<sup>89</sup>, consistent with our data on OCR (Figure 3.7). Next, we used intracellular flow cytometry to further characterize the mitochondria (Figure 3.8B). MitoTracker Green (MTG; measures total level of mitochondria) and MitoTracker Red (MTR; measures mitochondrial activity) revealed no significant differences between cell types (Figure 3.8B). We next used MitoSOX to measure mitochondrial reactive oxygen species (mROS). Surprisingly, LP, cells with high mitochondrial respiration, demonstrated the least amount of mROS (Figure 3.8B). This is most likely attributable to high level expression of multiple antioxidant mechanisms in these cells<sup>27</sup>. In contrast, BC and ML showed equivalently high levels of high mROS. These ROS phenotypes should be specific to mitochondria, as CellROX, readout for total cellular ROS, showed no differences among MEC (Figure 3.8B). The high levels of mROS in BC was surprising and indicated that perhaps the function of mitochondria differ based on the mammary lineages. We filtered our lab's previously generated dataset of mouse MEC proteomes<sup>122</sup> for proteins found in MitoCarta, a curated list of mitochondrial proteins<sup>142</sup>. As unsupervised hierarchical clustering revealed groupings based on cell type, it suggests that the mitochondria operate in a mammary cell-type specific manner (Figure 3.8). Thus, normal mammary epithelial cells are built and equipped with different mitochondrial machinery that leads to mammary cell-type specific mitochondrial morphology and function.



### 3.5 BC progenitor activity is dependent upon mitochondrial activity

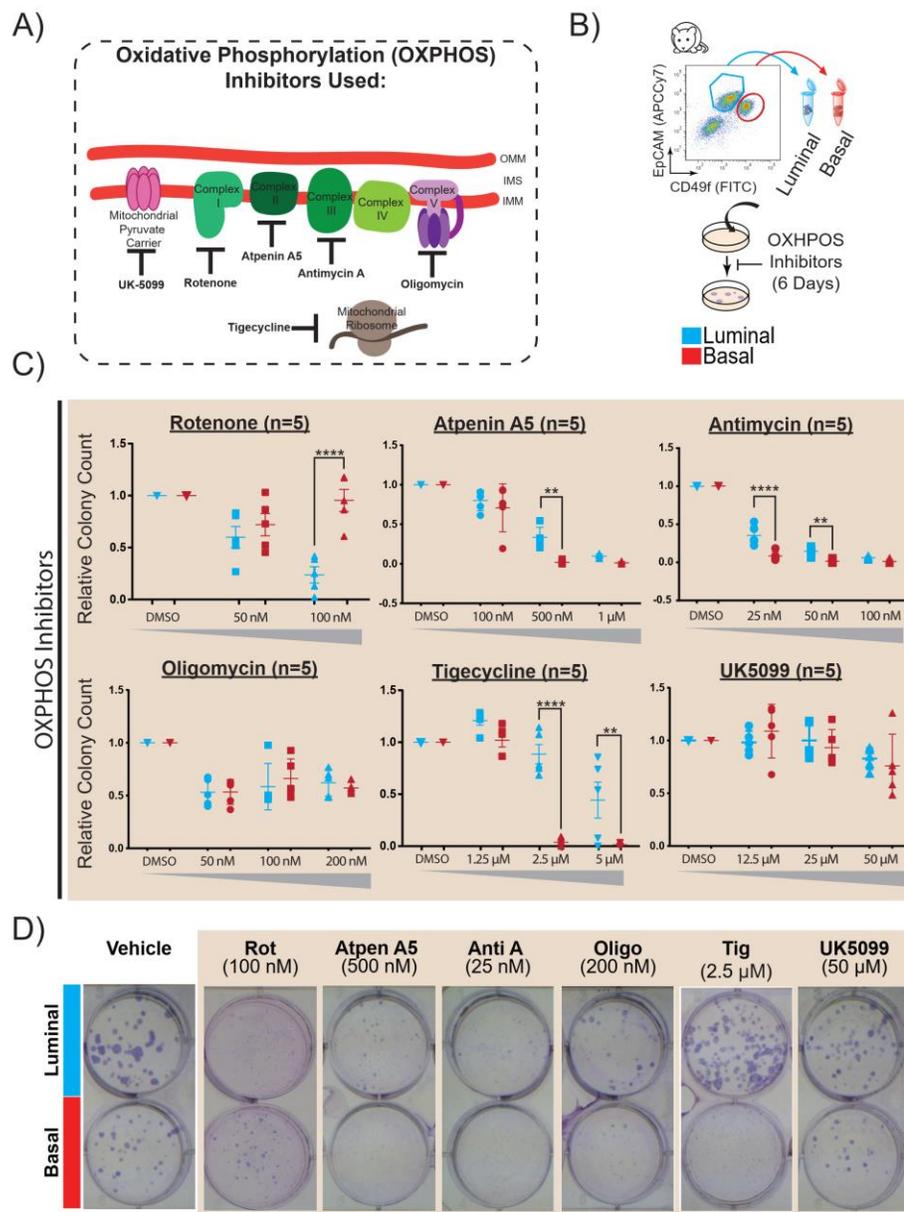
Non-ATP generating functions of mitochondria, such as maintenance of progenitor activity<sup>143,144</sup>, have been identified. Mechanistically, the metabolites produced in the mitochondria alter the fate of stem cells due to alterations of the epigenome<sup>98,145,146</sup>. To test if mitochondrial activity was necessary for mammary progenitor activity, we plated equivalent numbers of luminal and basal cells with high or low mitochondrial activity in the colony-forming cell (CFC) assay (Figure 3.9A). We denoted the MTG<sup>hi</sup>MTR<sup>hi</sup> as “High Mitochondrial Activity” and the MTG<sup>hi</sup>MTR<sup>lo</sup> population as “Low Mitochondrial Activity” (Figure 3.9A)<sup>147</sup>. Mammary cells with low activity formed fewer colonies as compared to total (Figure 3.9B). As we gated on Live cells (Figure 2.2), these low mitochondrial activity cells are probably not pre-apoptotic. Indeed, this population most likely represents the differentiated cells<sup>143</sup>. To our surprise, “High Activity” basal cells showed high progenitor activity, as they formed more colonies compared to total



basal (Figure 3.8B). Though mitochondria in these cells may not be participating in respiration, BC may require high mitochondrial activity for their progenitor activity. The underlying mechanism, whether due to ROS or a specific metabolite, has yet to be determined.

### **3.6 MEC demonstrate ETC complex specific vulnerabilities**

We next asked if metabolic distinctions could manifest as sensitivity to various drug treatments. Using the colony forming cell (CFC) assay, which is used as a measure of progenitor activity<sup>122</sup>, we evaluated the effect of OXPHOS inhibitors on progenitor capacity within each population (Figure 3.10). To inhibit OXPHOS, we used complex-specific inhibitors such as Rotenone (Complex I), Atpenin A5 (Complex II), Antimycin A (Complex III) and Oligomycin (Complex V) (Figure 3. 10a). Inhibition of Complex I using rotenone was selective at decreasing the number of LP colonies, whereas BC was unaffected. Despite LP demonstrating enhanced OXPHOS and higher abundance of ETC subunits, BC were significantly more sensitive to inhibition of Complex II or III. Both cell types demonstrated equivalent sensitivity to Complex V inhibition. It was recently shown that inhibition of ETC leads to complex-specific differential metabolite accumulation and gene expression changes<sup>94,100,101</sup>. Our results suggest that mammary lineages may link different cellular processes to ETC complexes, where CI is necessary for LP progenitor capacity but not BC. Non-ETC specific inhibition of OXPHOS was achieved using Tigecycline (inhibits mitochondrial ribosomes) and UK5099 (inhibits mitochondrial pyruvate carrier) (Figure 3.10A). Tigecycline prevents the translation of proteins encoded by mitochondrial DNA (mtDNA). The mitochondria has its own genome, which is a 16.6 kilobase, double-stranded, circular and contains no introns<sup>148,149</sup>. It encodes 2 rRNAs, 22 t-RNAs and 13 of the 90 proteins in the mitochondrial respiratory chain (Figure 1.4)<sup>148,149</sup>. Preventing translation of mitochondrial derived ETC subunits using Tigecycline completely abrogates the progenitor capacity of BC compared to LP. This data suggests that the mitochondrial respiratory chain in BC helps facilitate progenitor capacity. UK5099 is an inhibitor of the mitochondrial pyruvate carrier (MPC) and thus prevents entry of pyruvate into mitochondria. Both mammary subpopulations were relatively resistant to UK5099 treatment. This provides evidence for the previously mentioned claim that LP do not rely on cytosolic sources of pyruvate (Section 3.2, Figure 3.5 &3.6).

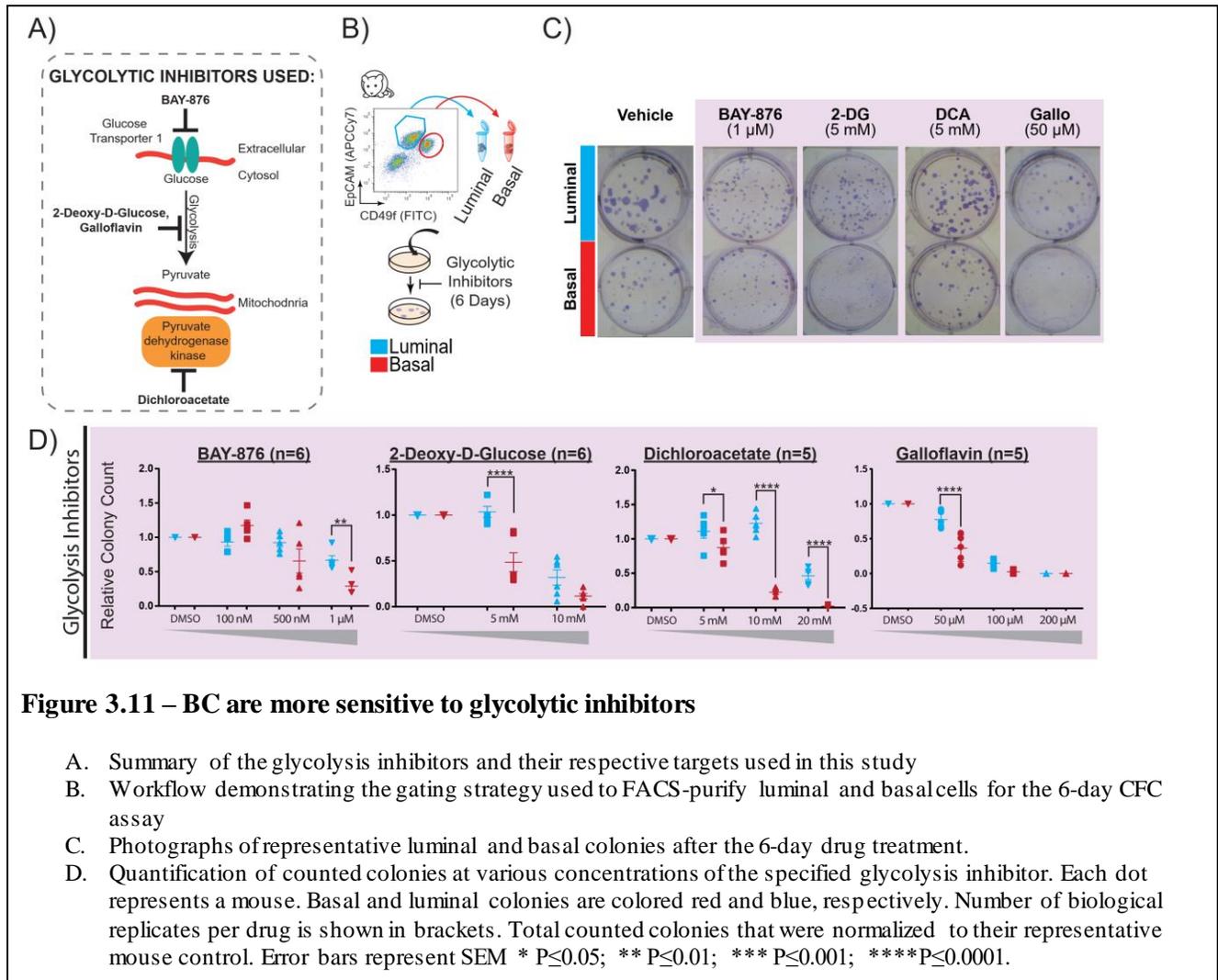


**Figure 3.10 – Mammary progenitors display lineage-restricted vulnerabilities to OXPHOS inhibitors**

- Summary of the OXPHOS inhibitors and their respective targets used in this study
- Workflow demonstrating the gating strategy used to FACS-purify luminal and basal cells for the 6-day CFC assay
- Quantification of counted colonies at various concentrations of the specified OXPHOS inhibitor. Each dot represents a mouse. Basal and luminal colonies are colored red and blue, respectively. Number of biological replicates per drug is shown in brackets. Total counted colonies were normalized to their representative mouse control. Error bars represent SEM \* P<0.05; \*\* P<0.01; \*\*\* P<0.001; \*\*\*\* P<0.0001.
- Photographs of representative luminal and basal colonies after the 6-day drug treatment.

### 3.7 BC are vulnerable to glycolytic inhibitors

To further corroborate results from pathway analysis (Figure 3.4), we tested for glycolysis as a BC-specific metabolic vulnerability (Figure 3.11). To this end, inhibition of glycolysis at multiple levels was achieved using BAY-876 (Glucose transporter 1), 2-Deoxy-D-glucose (Hexokinase), Galloflavin (Lactate Dehydrogenase) and Dichloroacetate (Pyruvate dehydrogenase kinase) (Figure 3.11A). BCs were far more sensitive to all four glycolytic drugs than LP (Figure 3.11C, D). However, at higher doses, LP colonies were reduced. This is not surprising as glycolytic inhibitors show significant toxicity in the clinic and have been largely unsuccessful in clinical trials<sup>74</sup>. Our work demonstrates that mammary subpopulations possess lineage-restricted metabolic vulnerabilities to glycolytic and OXPHOS inhibition.



**Figure 3.11 – BC are more sensitive to glycolytic inhibitors**

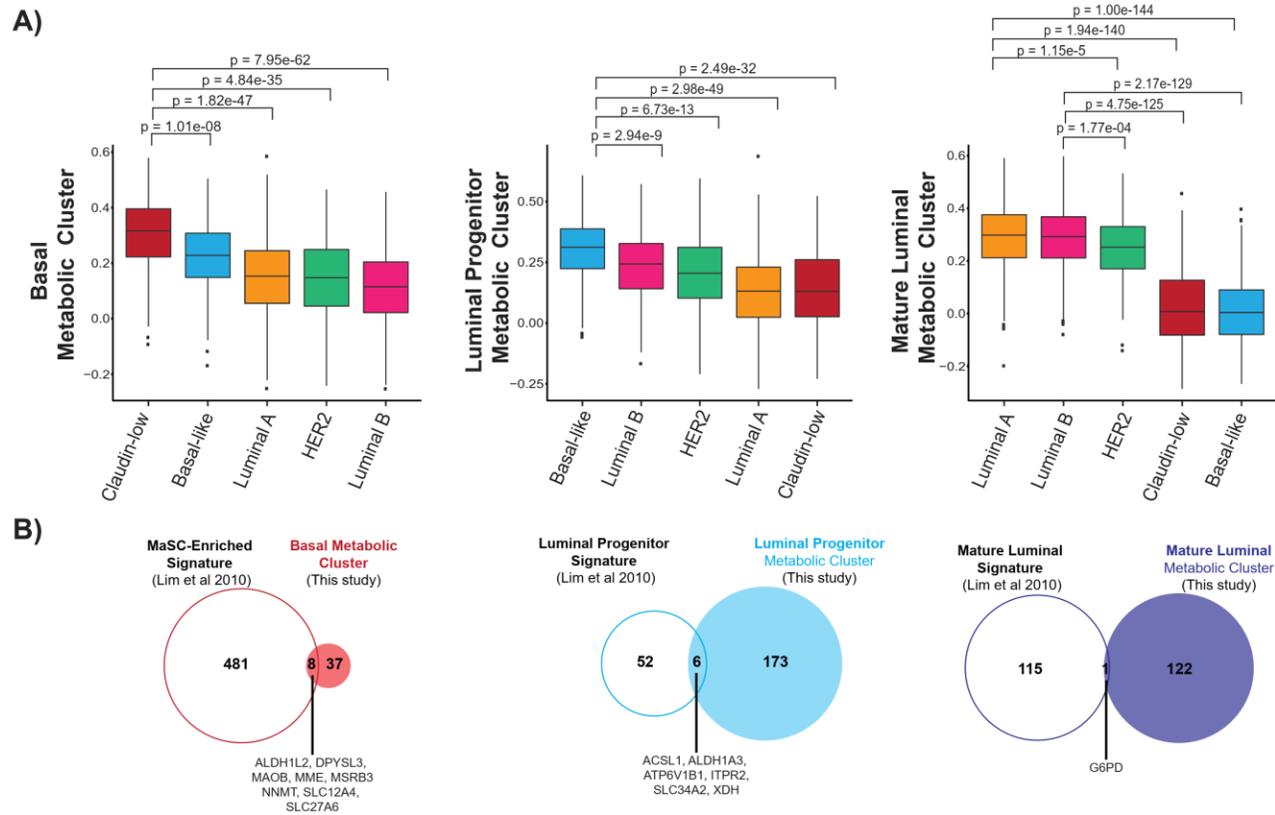
- Summary of the glycolysis inhibitors and their respective targets used in this study
- Workflow demonstrating the gating strategy used to FACS-purify luminal and basal cells for the 6-day CFC assay
- Photographs of representative luminal and basal colonies after the 6-day drug treatment.
- Quantification of counted colonies at various concentrations of the specified glycolysis inhibitor. Each dot represents a mouse. Basal and luminal colonies are colored red and blue, respectively. Number of biological replicates per drug is shown in brackets. Total counted colonies that were normalized to their representative mouse control. Error bars represent SEM \*  $P \leq 0.05$ ; \*\*  $P \leq 0.01$ ; \*\*\*  $P \leq 0.001$ ; \*\*\*\*  $P \leq 0.0001$ .

### 3.8 Metabolic profiles of mammary progenitors are mirrored in breast cancer subtypes

Thus far, we observed that mammary subpopulations have lineage-restricted metabolic identities and vulnerabilities. We wanted to determine if any of the PAM50 breast cancer subtypes demonstrated significant enrichment of their cell-of-origin's metabolic clusters (Figure 3.12A). To do this, we performed single sample gene set enrichment analysis (ssGSEA), an unsupervised and non-parametric method that calculates an enrichment score for a gene set across a collection of samples<sup>137</sup>. The enrichment score generated from this algorithm represents activity level of the gene set in every sample of a specified population relative to the other populations<sup>137</sup>. In our study, the gene sets are the mammary cell-type specific metabolic clusters and the populations are the breast cancer patients from the METABRIC database<sup>134</sup>. The BC metabolic cluster was most enriched in the highly mesenchymal Claudin-low subtype (Figure 3.12A). Luminal A and B subtypes showed significant enrichment for the ML metabolic cluster as compared to other subtypes (Figure 3.12A). The most significant correlation to the LP metabolic cluster was the highly aggressive Basal-like breast cancer (Figure 3.12A). These analyses suggests that breast cancer subtypes retain metabolic features of specific normal MEC populations. We performed similar analysis to the one published by the Visvader and Lindeman group<sup>39</sup>. Comparison of the genes that comprised our cell type specific metabolic cluster and their signatures showed very little overlap (Figure 3.12B). Their signatures were derived from performing a microarray on FACS-purified human mammary subpopulation and consisted of all upregulated genes in that subpopulation<sup>39</sup>. Our study employed the use of proteomics and our clusters were solely derived of significantly enriched metabolic proteins. The discrepancy between our gene lists highlights the potential of proteomics to capture important mammary gland biology that was not seen with microarray technology.

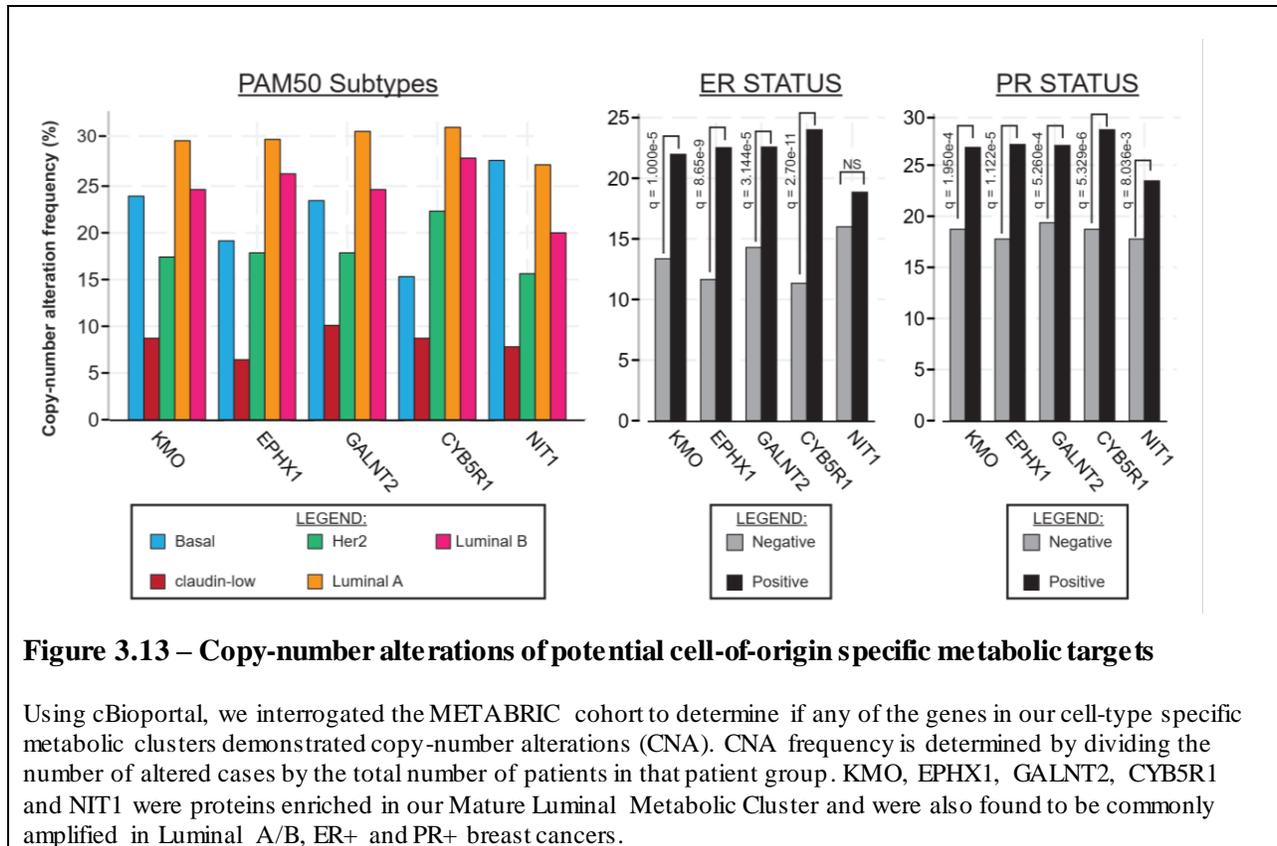
Recent studies have demonstrated success in targeting metabolic vulnerabilities that are specific to the tissue-of-origin<sup>115,117</sup> and also to chromosomal abnormalities<sup>150,151</sup>. We therefore looked for copy number alterations (CNA) in metabolic genes linked to specific mammary cell types using cBioportal<sup>135,136</sup>. These amplified metabolic genes could then represent potential cell-of-origin specific metabolic vulnerabilities. Interrogation of the LP metabolic cluster revealed an already well-studied amplified metabolic gene, *PHGDH*. This enzyme plays an important role in

serine biosynthesis and is a selective vulnerability in ER- Basal-like breast cancer<sup>132,152</sup>, the breast cancer subtype LP supposedly give rise to. Strikingly, 5 metabolic proteins from the ML metabolic cluster (Epoxide hydrolase 1 (EPHX1), nitrilase 1 (NIT1), cytochrome b5 reductase 1 (CYB5R1), polypeptide N-acetylgalactosaminyltransferase 2 (GALNT2) and kynurenine 3-monooxygenase (KMO)), were encoded by genes that were highly amplified in ER+, PR+ and for the most part Luminal A & B breast cancers from the METABRIC cohort (Figure 3.13). All 5 of these proteins do not participate in the same metabolic pathway but are all found on the q arm of chromosome 1. Whole-arm amplification of 1q together with 16q loss (+/-) is a hallmark chromosomal event in ER+ breast cancers<sup>53,134,153</sup>. The fact that we see proteins from our metabolic clusters being amplified at the gene level in the respective breast cancer subtypes suggests that these targets may represent actionable cell-of-origin specific metabolic vulnerabilities. This has already been proven successful with the *PHGDH* example, however further investigation is required to validate the ML-derived metabolic candidates.



**Figure 3.12 – Breast cancer subtypes demonstrate selective activity of their respective cell-of-origin’s metabolic cluster**

- A. Boxplots of single sample gene set enrichment analysis (ssGSEA) scores comparing the metabolic cluster of BC, ML and LP to the PAM50 subtypes of breast cancer (Luminal A, Luminal B, Claudin-low, Basal-like, HER-2) from the METABRIC cohort. Boxplots are organized left to right, from highest to least median score. Student’s t-test was performed to determine if medians were different between breast cancer subtypes and the p-values are written in the figure for the most significant subtype.
- B. Venn diagrams comparing our mammary cell-type specific metabolic clusters to the previously published mammary subpopulation signatures from Lim et al 2010. The genes that intersect are labelled.



## Chapter 4

### 4 Conclusion

This study is the first to demonstrate lineage-restricted metabolic identities in the mammary gland. We performed proteomic profiling of primary FACS-purified human mammary subpopulations, which revealed unique metabolic preferences for each cell type. We generated mammary cell-type specific metabolic clusters that represents the core set of metabolic proteins that may be necessary for the structure and function of that cell type. LP demonstrated enhanced OCR due to their enrichment for TCA Cycle and ETC proteins and optimal mitochondrial structure. Both lineages demonstrated vulnerability to specific ETC inhibitors, where BC were more sensitive to CII and III inhibition and LP were more sensitive to CI inhibition. BC and ML demonstrated higher abundance of glycolytic enzyme and were vulnerable to glycolytic inhibition. However, BC progenitor activity was dependent on mitochondrial activity, suggesting non-bioenergetics functions of mitochondria in BC. Finally, by comparing our mammary cell-type specific metabolic clusters to the PAM50 subtypes of breast cancer, we demonstrate that the metabolism of breast cancer subtypes may be more similar to the putative cell-of-origin.

#### 4.1 Discussion & Significance

Our work highlights a previously underappreciated metabolic heterogeneity present in the epithelial compartment of the breast. It would be interesting to know whether the lineage-driven metabolic programs are intrinsic to cell identity or a reflection of different adaptations to distinct mammary microenvironments. BC are in contact with the basement membrane, whereas luminal cells are exposed apically to the lumen. However, our measurements of mitochondrial respiration and drug sensitivity were performed *ex vivo* on purified populations. This suggests that these metabolic distinctions are hardwired, perhaps necessary to facilitate unique form and function, in each mammary cell type.

The metabolic phenotype of a cancer cell is highly dependent upon integrating intrinsic and extrinsic cues. The importance of the tissue-of-origin has only just recently been determined. The tissue-of-origin ignores the epithelial heterogeneity present in tissues like the mammary gland. Arguably, cell lineage could be one of the most important determinants of cellular metabolism,

as all perturbations (mutational or microenvironmental) will disrupt a pre-existing metabolic network present in the cell-of-origin. Strong correlations between breast cancers subtypes and normal mammary subpopulations identified in this thesis do in fact suggest that breast cancers adapt the metabolic network of their cell-of-origin.

Precision oncology has largely been guided by mutations to effectively treat patients. However, our work suggests that certain metabolic features of breast cancers can be explained partly by the cell-of-origin rather than by mutations or microenvironmental factors. Therefore, to target metabolic vulnerabilities of breast cancer, one needs to account for the complement of mutations in conjunction with the cell-of-origin to allow for successful subtype-specific treatments.

## 4.2 Future Directions

There are several unexplored and exciting topics that have yet to be answered, which can provide more insight into MEC metabolism and potentially breast cancer risk.

As mentioned in the introduction (Section 1.2.5), clustering of metabolomes in ER- breast cancers can be based on ethnicity<sup>108</sup>. African American women are at a higher risk of developing Basal-like breast cancers. It is not known whether breast epithelial composition is different in these individuals than in those of European ancestry, or if there is a potential role of different diets. It would be interesting to determine the effects of different diets (ketogenic, high-fat or caloric restriction) on mammary subpopulations frequency and function.

Several other extrinsic determinants of metabolism have yet to be studied in the context of the normal mammary gland. The mammary gland is a heterocellular organ that relies heavily on paracrine signaling between adjacent epithelial cells<sup>14</sup>. We interrogated our previous generated mouse proteomic dataset to look for nutrient receptor expression among subpopulations<sup>122</sup> (Appendix, Figure 4.1A). There was differential expression of the glucose transporters (GLUTs), where BC and ML have high GLUT1 expression and GLUT4 is exclusively found on BC. Conversely, LPs had low levels of GLUTs, but high expression of monocarboxylate transporter 1 (MCT1), which are involved in lactate import. These results match data obtained from a microarray-based study published by our lab<sup>138</sup> (Appendix, Figure 4.1B). The expression and

localization of MCT4 was validated by staining frozen tissue sections (Appendix, Figure 4.1C). We saw strikingly co-localization of MCT4 with K14, a basal marker. The close spatial location of these cells and inverse relationship of MCTs suggests potential lactate crosstalk across mammary epithelium. The lactate shuttle is ubiquitous throughout the human body, such as from Astrocytes to neurons<sup>154</sup> and Paneth cells to intestinal stem cells<sup>155</sup>. Though each cell is built with a unique metabolic program, perhaps they work together to support each other's metabolism *in situ*.

In addition, hormones have long been known to have an effect on cellular metabolism<sup>156</sup>, but most of the research has focused on estrogen<sup>106</sup>. Progesterone is a potent inducer of proliferation for stem and progenitor populations during key physiological events like pregnancy and the luteal phase of the reproductive cycle<sup>4</sup>. Progesterone is able to induce proliferation of HR-progenitor populations by causing HR+ ML cells to release proliferative paracrine effectors like RANKL and WNT4<sup>4,13</sup>. To accompany this massive expansion, metabolic reprogramming may be necessary as there needs to be new building blocks for growth. In fact, interrogation of our lab's mouse proteomic dataset demonstrates that metabolic proteins cluster base on cell type, similar to our study (Appendix, Figure 4.1D). But within each cell type cluster, metabolic proteins cluster based on hormone treatment. However, to date there has been no documented metabolic effector linked to this progesterone-induced expansion. One important pathway that appears to be upregulated in response to progesterone treatment is the catabolic pathway known as autophagy. In response to progesterone, the number of lysosomes increases significantly in each cell type, as seen by TEM and LysoTracker (Appendix, Figure 4.1E&F).

### **4.3 Limitations**

Metabolomics or radiolabelled isotope tracing is the gold-standard to determine if metabolic flux differs between distinct cell populations. However, in our hands, we were unable to detect any metabolites above the internal standards. Metabolite detection involves rapid isolation of tissue. However, to obtain purified mammary subpopulations, mechanical and chemical dissociation steps involving a series of temperature changes, washes and sorting are required. This 7-hour protocol is the only way to obtain purified populations, which we find is not compatible for

subsequent detection of metabolites. Future work should involve development of a more rapid isolation technique to allow for analysis of metabolic flux in rare stem/progenitor populations.

## References

1. Inman, J. L., Robertson, C., Mott, J. D. & Bissell, M. J. Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**, 1028–1042 (2015).
2. Watson, C. J. & Khaled, W. T. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**, 995–1003 (2008).
3. Asselin-Labat, M.-L. *et al.* Control of mammary stem cell function by steroid hormone signalling. *Nature* **465**, 798–802 (2010).
4. Joshi, P. A. *et al.* Progesterone induces adult mammary stem cell expansion. *Nature* **465**, 803–807 (2010).
5. Cardiff, R. D. & Wellings, S. R. The comparative pathology of human and mouse mammary glands. *J. Mammary Gland Biol. Neoplasia* **4**, 105–122 (1999).
6. Santagata, S. *et al.* Taxonomy of breast cancer based on normal cell phenotype predicts outcome. *J. Clin. Invest.* **124**, 859–70 (2014).
7. Wellings, S. R. A hypothesis of the origin of human breast cancer from the terminal ductal lobular unit. *Pathol. Res. Pract.* **166**, 515–35 (1980).
8. Wellings, S. R., Jensen, H. M. & Marcum, R. G. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. *J. Natl. Cancer Inst.* **55**, 231–73 (1975).
9. Visvader, J. E. & Stingl, J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev.* **28**, 1143–1158 (2014).
10. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. **439**, 993–997 (2006).
11. Shehata, M. *et al.* Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res. BCR* **14**, R134 (2012).
12. Briskin, C. & Duss, S. Stem cells and the stem cell niche in the breast: an integrated hormonal and developmental perspective. *Stem Cell Rev.* **3**, 147–156 (2007).
13. Joshi, P. A. *et al.* RANK Signaling Amplifies WNT-Responsive Mammary Progenitors through R-SPONDIN1. *Stem Cell Rep.* **5**, 31–44 (2015).
14. Tharmapalan, P., Mahendralingam, M., Berman, H. K. & Khokha, R. Mammary stem cells and progenitors: targeting the roots of breast cancer for prevention. *EMBO J.* **38**, e100852 (2019).
15. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).

16. Prater, M. D. *et al.* Mammary stem cells have myoepithelial cell properties. *Nat. Cell Biol.* **16**, 942–950, 1–7 (2014).
17. Plaks, V. *et al.* Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis. *Cell Rep* **3**, 70–78 (2013).
18. de Visser, K. E. *et al.* Developmental stage-specific contribution of LGR5(+) cells to basal and luminal epithelial lineages in the postnatal mammary gland. *J. Pathol.* **228**, 300–9 (2012).
19. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189–193 (2011).
20. Wang, D. *et al.* Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81–4 (2015).
21. Fu, N. Y. *et al.* Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. (2017). doi:10.1038/ncb3471
22. Cai, S. *et al.* A Quiescent Bcl11b High Stem Cell Population Is Required for Maintenance of the Mammary Gland. *Cell Stem Cell* **20**, 247-260.e5 (2017).
23. Chakrabarti, R. *et al.* Notch ligand Dll1 mediates cross-talk between mammary stem cells and the macrophageal niche. *Science* **360**, ean4153 (2018).
24. Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. In situ identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322–327 (2014).
25. Nguyen, Q. H. *et al.* Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat. Commun.* (2018). doi:10.1038/s41467-018-04334-1
26. Gorrini, C., Harris, I. S. & Mak, T. W. Modulation of oxidative stress as an anticancer strategy. *Nat. Rev. Drug Discov.* **12**, 931–947 (2013).
27. Kannan, N. *et al.* Glutathione-dependent and -independent oxidative stress-control mechanisms distinguish normal human mammary epithelial cell subsets. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7789–7794 (2014).
28. Gascard, P. *et al.* Epigenetic and transcriptional determinants of the human breast. *Nat. Commun.* **6**, 6351 (2015).
29. Aguilera, A. & García-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
30. Zhang, X. *et al.* Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. *Nat. Commun.* **8**, 15908 (2017).

31. Kannan, N. *et al.* The luminal progenitor compartment of the normal human mammary gland constitutes a unique site of telomere dysfunction. *Stem Cell Rep.* **1**, 28–37 (2013).
32. Morel, A.-P. *et al.* A stemness-related ZEB1–MSRB3 axis governs cellular plasticity and breast cancer genome stability. *Nat. Med.* **23**, 568–578 (2017).
33. Pierce, G. B. & Speers, W. C. Tumors as caricatures of the process of tissue renewal: prospects for therapy by directing differentiation. *Cancer Res.* **48**, 1996–2004 (1988).
34. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
35. Kreso, A. & Dick, J. E. Evolution of the Cancer Stem Cell Model. *Cell Stem Cell* **14**, 275–291 (2014).
36. Chaffer, C. L. & Weinberg, R. A. How does multistep tumorigenesis really proceed? *Cancer Discov.* **5**, 22–24 (2015).
37. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
38. Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330–1334 (2017).
39. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).
40. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68 (2010).
41. Molyneux, G. *et al.* BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* **7**, 403–417 (2010).
42. Koren, S. *et al.* PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nature* **525**, 114–118 (2015).
43. Van Keymeulen, A. *et al.* Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* **525**, 119–123 (2015).
44. Keller, P. J. *et al.* Defining the cellular precursors to human breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2772–2777 (2012).
45. Spike, B. T. *et al.* A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* **10**, 183–97 (2012).
46. Girardi, R. R. *et al.* Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. *Cell Rep.* **24**, 1653–1666 (2018).

47. Lawson, D. A. *et al.* Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
48. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291–304.e6 (2018).
49. Polak, P. *et al.* Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364 (2015).
50. Sack, L. M. *et al.* Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell* **173**, 499–514.e23 (2018).
51. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
52. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10869–10874 (2001).
53. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
54. Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8**, R76 (2007).
55. Goldhirsch, A. *et al.* Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **22**, 1736–1747 (2011).
56. Durbier, A. K. *et al.* Association between breast cancer subtypes and response to neoadjuvant anastrozole. *Steroids* **76**, 736–740 (2011).
57. Ellis, M. J. *et al.* Whole Genome Analysis Informs Breast Cancer Response to Aromatase Inhibition. *Nature* **486**, 353–360 (2012).
58. Pan, H. *et al.* 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N. Engl. J. Med.* **377**, 1836–1846 (2017).
59. Martín, M. *et al.* PAM50 proliferation score as a predictor of weekly paclitaxel benefit in breast cancer. *Breast Cancer Res. Treat.* **138**, 457–466 (2013).
60. Hugh, J. *et al.* Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **27**, 1168–1176 (2009).
61. Martin, M. *et al.* Genomic predictors of response to doxorubicin versus docetaxel in primary breast cancer. *Breast Cancer Res. Treat.* **128**, 127–136 (2011).

62. Llombart-Cussac, A. *et al.* HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. *Lancet Oncol.* **18**, 545–554 (2017).
63. Karp, G. *Cell and Molecular Biology: Concepts and Experiments.* (John Wiley & Sons, 2009).
64. Lehninger, A. L., Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry* (7th Edition). (2017).
65. Warburg, O. On respiratory impairment in cancer cells. *Science* **124**, 269–270 (1956).
66. Warburg, O., Wind, F. & Negelein, E. THE METABOLISM OF TUMORS IN THE BODY. *J. Gen. Physiol.* **8**, 519–530 (1927).
67. Vander Heiden, M. G. & DeBerardinis, R. J. Understanding the Intersections between Metabolism and Cancer Biology. *Cell* **168**, 657–669 (2017).
68. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324**, 1029–1033 (2009).
69. Vander Heiden, M. G. *et al.* Evidence for an alternative glycolytic pathway in rapidly proliferating cells. *Science* **329**, 1492–1499 (2010).
70. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).
71. Faubert, B. *et al.* Lactate Metabolism in Human Lung Tumors. *Cell* **171**, 358–371.e9 (2017).
72. Hui, S. *et al.* Glucose feeds the TCA cycle via circulating lactate. *Nature* **551**, 115–118 (2017).
73. Lunt, S. Y. & Vander Heiden, M. G. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu. Rev. Cell Dev. Biol.* **27**, 441–464 (2011).
74. Luengo, A., Gui, D. Y. & Vander Heiden, M. G. Targeting Metabolism for Cancer Therapy. *Cell Chem. Biol.* **24**, 1161–1180 (2017).
75. Martinez-Outschoorn, U. E., Peiris-Pagés, M., Pestell, R. G., Sotgia, F. & Lisanti, M. P. Cancer metabolism: a therapeutic perspective. *Nat. Rev. Clin. Oncol.* **14**, 11–31 (2017).
76. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
77. Losman, J.-A. & Kaelin, W. G. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev.* **27**, 836–852 (2013).

78. Yang, M., Soga, T. & Pollard, P. J. Oncometabolites: linking altered metabolism with cancer. *J. Clin. Invest.* **123**, 3652–3658 (2013).
79. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
80. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **18**, 553–567 (2010).
81. Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).
82. Pollard, P. J. *et al.* Accumulation of Krebs cycle intermediates and over-expression of HIF1alpha in tumours which result from germline FH and SDH mutations. *Hum. Mol. Genet.* **14**, 2231–2239 (2005).
83. Xiao, M. *et al.* Inhibition of  $\alpha$ -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).
84. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).
85. Sasaki, M. *et al.* D-2-hydroxyglutarate produced by mutant IDH1 perturbs collagen maturation and basement membrane function. *Genes Dev.* **26**, 2038–2049 (2012).
86. Jiang, Y. *et al.* Local generation of fumarate promotes DNA repair through inhibition of histone H3 demethylation. *Nat. Cell Biol.* **17**, 1158–1168 (2015).
87. Sulkowski, P. L. *et al.* Krebs-cycle-deficient hereditary cancer syndromes are defined by defects in homologous-recombination DNA repair. *Nat. Genet.* **50**, 1086–1092 (2018).
88. Frey, T. G. & Mannella, C. A. The internal structure of mitochondria. *Trends Biochem. Sci.* **25**, 319–324 (2000).
89. Cogliati, S. *et al.* Mitochondrial Cristae Shape Determines Respiratory Chain Supercomplexes Assembly and Respiratory Efficiency. *Cell* **155**, 160–171 (2013).
90. Brand, M. D. The sites and topology of mitochondrial superoxide production. *Exp. Gerontol.* **45**, 466–472 (2010).
91. Sena, L. A. & Chandel, N. S. Physiological roles of mitochondrial reactive oxygen species. *Mol. Cell* **48**, 158–167 (2012).

92. Yoshida, M., Muneyuki, E. & Hisabori, T. ATP synthase — a marvellous rotary engine of the cell. *Nat. Rev. Mol. Cell Biol.* **2**, 669 (2001).
93. Vyas, S., Zaganjor, E. & Haigis, M. C. Mitochondria and Cancer. *Cell* **166**, 555–566 (2016).
94. Chen, W. W., Freinkman, E., Wang, T., Birsoy, K. & Sabatini, D. M. Absolute Quantification of Matrix Metabolites Reveals the Dynamics of Mitochondrial Metabolism. *Cell* **166**, 1324–1337.e11 (2016).
95. Birsoy, K. *et al.* An Essential Role of the Mitochondrial Electron Transport Chain in Cell Proliferation Is to Enable Aspartate Synthesis. *Cell* **162**, 540–551 (2015).
96. Sullivan, L. B. *et al.* Supporting Aspartate Biosynthesis Is an Essential Function of Respiration in Proliferating Cells. *Cell* **162**, 552–563 (2015).
97. King, M. P. & Attardi, G. Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation. *Science* **246**, 500–503 (1989).
98. Ansó, E. *et al.* The mitochondrial respiratory chain is essential for haematopoietic stem cell function. *Nat. Cell Biol.* **19**, 614–625 (2017).
99. Sena, L. A. *et al.* Mitochondria Are Required for Antigen-Specific T Cell Activation through Reactive Oxygen Species Signaling. *Immunity* **38**, 225–236 (2013).
100. Weinberg, S. E. *et al.* Mitochondrial complex III is essential for suppressive function of regulatory T cells. *Nature* **565**, 495–499 (2019).
101. Bailis, W. *et al.* Distinct modes of mitochondrial metabolism uncouple T cell differentiation and function. *Nature* **571**, 403 (2019).
102. Diehn, M. *et al.* Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* **458**, 780–783 (2009).
103. Chandel, N. S., Jasper, H., Ho, T. T. & Passequé, E. Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. *Nat. Cell Biol.* **18**, 823–832 (2016).
104. Budczies, J. *et al.* Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. *J. Proteomics* **94**, 279–288 (2013).
105. Cappelletti, V. *et al.* Metabolic Footprints and Molecular Subtypes in Breast Cancer. *Dis. Markers* **2017**, 1–19 (2017).
106. Kulkoyluoglu-Cotul, E., Arca, A. & Madak-Erdogan, Z. Crosstalk between Estrogen Signaling and Breast Cancer Metabolism. *Trends Endocrinol. Metab. TEM* **30**, 25–38 (2019).

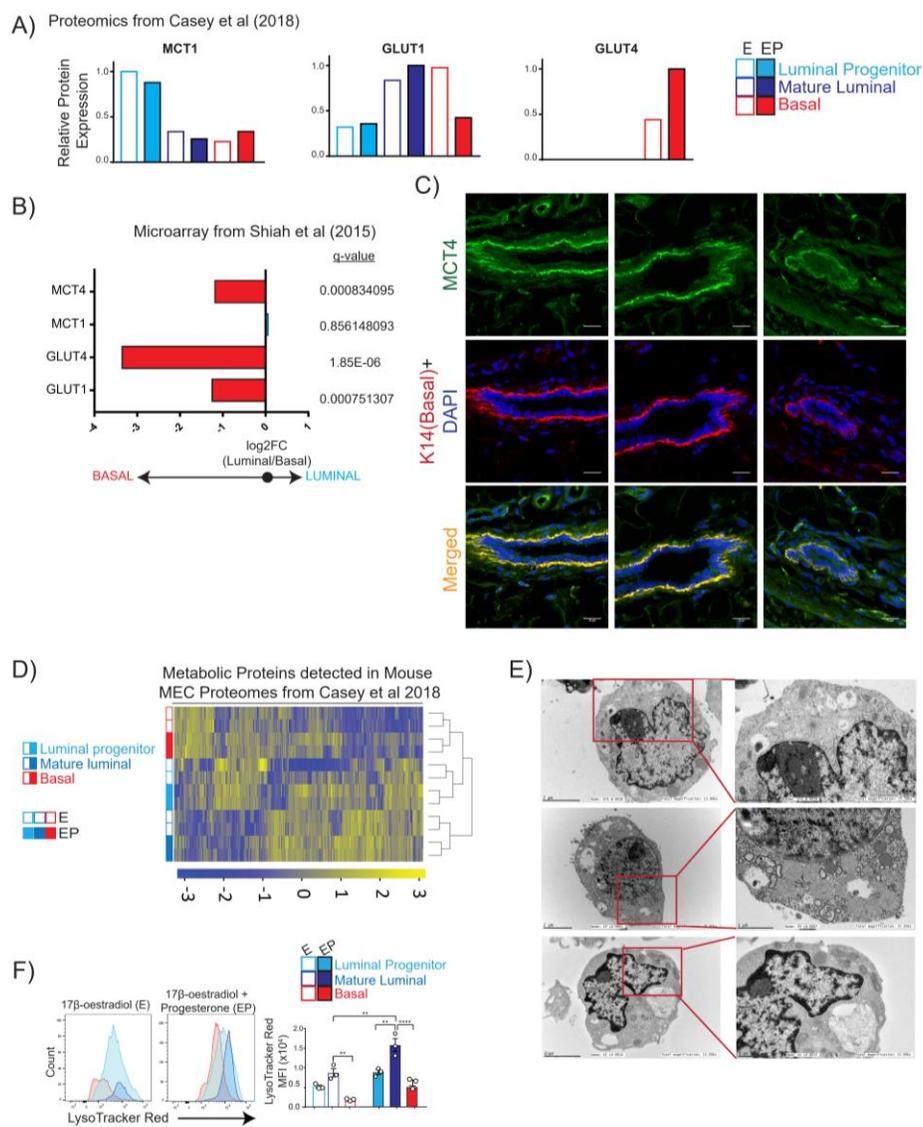
107. Tang, X. *et al.* A joint analysis of metabolomics and genetics of breast cancer. *Breast Cancer Res.* **16**, 415 (2014).
108. Terunuma, A. *et al.* MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J. Clin. Invest.* **124**, 398–412 (2014).
109. Brauer, H. A. *et al.* Impact of Tumor Microenvironment and Epithelial Phenotypes on Metabolism in Breast Cancer. *Clin. Cancer Res.* **19**, 571–585 (2013).
110. Borgan, E. *et al.* Merging transcriptomics and metabolomics - advances in breast cancer profiling. *BMC Cancer* **10**, 628 (2010).
111. Li, H. *et al.* The landscape of cancer cell line metabolism. *Nat. Med.* **25**, 850–860 (2019).
112. Gaude, E. & Frezza, C. Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nat. Commun.* **7**, (2016).
113. Hu, J. *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat. Biotechnol.* **31**, 522–529 (2013).
114. Yuneva, M. O. *et al.* The Metabolic Profile of Tumors Depends on Both the Responsible Genetic Lesion and Tissue Type. *Cell Metab.* **15**, 157–170 (2012).
115. Mayers, J. R. *et al.* Tissue of origin dictates branched-chain amino acid metabolism in mutant Kras-driven cancers. *Science* **353**, 1161–1165 (2016).
116. Commisso, C. *et al.* Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature* **497**, 633–637 (2013).
117. Chowdhry, S. *et al.* NAD metabolic dependency in cancer is shaped by gene amplification and enhancer remodelling. *Nature* (2019). doi:10.1038/s41586-019-1150-2
118. Verdin, E. NAD<sup>+</sup> in aging, metabolism, and neurodegeneration. *Science* **350**, 1208–1213 (2015).
119. Mayers, J. R. & Vander Heiden, M. G. Nature and Nurture: What Determines Tumor Metabolic Phenotypes? *Cancer Res.* **77**, 3131–3134 (2017).
120. Labarge, M. A., Garbe, J. C. & Stampfer, M. R. Processing of human reduction mammoplasty and mastectomy tissues for cell culture. *J. Vis. Exp. JoVE* 1–7 (2013). doi:10.3791/50011
121. Eirew, P., Stingl, J. & Eaves, C. J. Quantitation of human mammary epithelial stem cells with in vivo regenerative properties using a subrenal capsule xenotransplantation assay. *Nat. Protoc.* **5**, 1945–1956 (2010).

122. Casey, A. E. *et al.* Mammary molecular portraits reveal lineage-specific features and progenitor cell vulnerabilities. *J. Cell Biol.* **217**, 2951–2974 (2018).
123. Bunn, A. & Korpela, M. An Introduction to dplR. 16
124. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
125. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. sva: Surrogate Variable Analysis R package version 3.10.0. *Bioconductor* Available at: <http://bioconductor.org/packages/sva/>. (Accessed: 28th July 2019)
126. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**, 35 (2011).
127. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-97 (2016).
128. Vu, V. *A biplot based on ggplot2. Contribute to vqv/ggbiplot development by creating an account on GitHub.* (2019).
129. Maechler, M., Rousseeux, P., Struyf, A., Hubert, M. & Hornik, K. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1.
130. Kolde, R. *heatmap: Pretty Heatmaps.* (2019).
131. Neuwirth, E. *RColorBrewer: ColorBrewer Palettes.* (2014).
132. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–350 (2011).
133. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
134. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
135. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
136. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pii (2013).
137. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

138. Shiah, Y.-J. *et al.* A Progesterone-CXCR4 Axis Controls Mammary Progenitor Cell Fate in the Adult Gland. *Stem Cell Rep.* **4**, 313–322 (2015).
139. Asselin-Labat, M.-L. *et al.* Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat. Cell Biol.* **9**, 201–209 (2007).
140. Lin, K. H. *et al.* Systematic Dissection of the Metabolic-Apoptotic Interface in AML Reveals Heme Biosynthesis to Be a Regulator of Drug Sensitivity. *Cell Metab.* **29**, 1217–1231.e7 (2019).
141. Folmes, C. D. L., Dzeja, P. P., Nelson, T. J. & Terzic, A. Metabolic plasticity in stem cell homeostasis and differentiation. *Cell Stem Cell* **11**, 596–606 (2012).
142. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).
143. Katajisto, P. *et al.* Asymmetric apportioning of aged mitochondria between daughter cells is required for stemness. *Science* **348**, 340–343 (2015).
144. Wu, M.-J. *et al.* Epithelial-Mesenchymal Transition Directs Stem Cell Polarity via Regulation of Mitofusin. *Cell Metab.* (2018). doi:10.1016/j.cmet.2018.11.004
145. Schell, J. C. *et al.* Control of intestinal stem cell function and proliferation by mitochondrial pyruvate metabolism. *Nat. Cell Biol.* **19**, 1027–1036 (2017).
146. Flores, A. *et al.* Lactate dehydrogenase activity drives hair follicle stem cell activation. *Nat. Cell Biol.* **19**, 1017–1026 (2017).
147. Tal, M. C. *et al.* Absence of autophagy results in reactive oxygen species-dependent amplification of RLR signaling. *Proc. Natl. Acad. Sci.* **106**, 2770–2775 (2009).
148. Verma, M., Kagan, J., Sidransky, D. & Srivastava, S. Proteomic analysis of cancer-cell mitochondria. *Nat. Rev. Cancer* **3**, 789–795 (2003).
149. Chatterjee, A., Mambo, E. & Sidransky, D. Mitochondrial DNA mutations in human cancer. *Oncogene* **25**, 4663–4674 (2006).
150. Dey, P. *et al.* Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. *Nature* **542**, 119–123 (2017).
151. Tang, Y.-C. *et al.* Aneuploid Cell Survival Relies upon Sphingolipid Homeostasis. *Cancer Res.* **77**, 5272–5286 (2017).
152. Locasale, J. W. *et al.* Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat. Genet.* **43**, 869–874 (2011).
153. Russnes, H. G. *et al.* Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci. Transl. Med.* **2**, 38ra47 (2010).

154. Suzuki, A. *et al.* Astrocyte-Neuron Lactate Transport Is Required for Long-Term Memory Formation. *Cell* **144**, 810–823 (2011).
155. Rodríguez-Colman, M. J. *et al.* Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature* **543**, 424–427 (2017).
156. Randle, P. J. Endocrine Control of Metabolism. *Annu. Rev. Physiol.* **25**, 291–324 (1963).

## Appendix 1



**Figure 4.1 – Lactate crosstalk and progesterone induced autophagy**

- Protein express of MCT1, GLUT1 and GLUT4 from a mouse proteomic dataset.
- Gene expression of MCT1, MCT4, GLUT1, GLUT4 from a mouse microarray. Data is presented as Log<sub>2</sub>FC and accompanying q-value is also labelled
- Confocal images of frozen mammary tissue sections (MCT4 (Green), K14 (red) and DAPI (blue))
- Unsupervised hierarchical clustering of all the detected metabolic proteins from a mouse proteomic dataset.
- Transmission electron microscopy images demonstrating accumulation of lysosomes in EP treated MEC
- Flow plots and quantification of median fluorescent intensity (MFI) of LysoTracker Red