

1

2 **Mammary epithelial cells have lineage-restricted metabolic identities**

3 Mathepan Mahendralingam^{1,2}, Kazeera Aliar¹, Alison Elisabeth Casey¹, Davide
4 Pellacani⁴, Hyeyeon Kim^{1,2}, Vladimir Ignatchenko¹, Mar Garcia Valero³, Luis Palomero³,
5 Ankit Sinha^{1,2}, Vid Stambolic¹, Mina Alam¹, Aaron Schimmer^{1,2}, Hal Berman¹, Miquel
6 Angel Pujana³, Connie Eaves⁴, Thomas Kislinger^{1,2*}, Rama Khokha^{1,2,5*}

7

8 Affiliations:

- 9 1. Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, M5G
10 2C1, Canada
- 11 2. Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M5G 2C1,
12 Canada
- 13 3. Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research,
14 Barcelona, Catalonia, 08908, Spain
- 15 4. Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, V5Z 1L3, Canada
- 16 5. Lead contact

17

*Correspondence: rama.khokha@uhnresearch.ca, thomas.kislinger@utoronto.ca

18

19

1 **ABSTRACT**

2 Cancer metabolism adapts the metabolic network of its tissue-of-origin. However, breast
3 cancer is not a disease of a singular origin. Multiple epithelial populations serve as the
4 culprit cell-of-origin for specific breast cancer subtypes, yet knowledge surrounding the
5 metabolic network of normal mammary epithelial cells is limited. Here, we show that
6 mammary populations have cell type-specific metabolic programs. Primary human breast
7 cell proteomes of basal, luminal progenitor, and mature luminal populations revealed their
8 unique enrichment of metabolic proteins. Luminal progenitors had higher abundance of
9 electron transport chain subunits and capacity for oxidative phosphorylation, whereas
10 basal cells were more glycolytic. Targeting oxidative phosphorylation and glycolysis with
11 inhibitors exposed distinct metabolic vulnerabilities of the mammary lineages.
12 Computational analysis indicated that breast cancer subtypes retain metabolic features
13 of their putative cell-of-origin. Lineage-restricted metabolic identities of normal mammary
14 cells partly explain breast cancer metabolic heterogeneity and rationalize targeting
15 subtype-specific metabolic vulnerabilities to advance breast cancer therapy.

16

17

18

1 INTRODUCTION

2 Molecular classification of breast cancers using the PAM50 classifier has identified 5 main
3 patient groups (Luminal A, Luminal B, HER2, Claudin-low, Basal-like) with distinct
4 transcriptional programs, survival outcomes and susceptibilities to anti-cancer regimens^{1–}
5 ³. Metabolomics on primary tumors report distinct metabolic phenotypes for each of the
6 breast cancer subtypes^{4–8}. Investigation of subtype-specific metabolic features have
7 focused on the effects of estrogen receptor⁹, *HER2* amplification¹⁰ and/or driver mutations
8 (*Tp53*, *Pik3ca*)¹¹. However, several of these markers and mutations are shared amongst
9 breast cancer subtypes and can only partly explain the metabolic heterogeneity¹². Tissue-
10 of-origin has emerged as an important intrinsic determinant of cellular metabolism¹³. This
11 is based on studies showing cancers use the metabolic network of their normal
12 counterpart as a backbone for aberrant proliferation^{14–16}. This poses a challenge in the
13 context of breast cancer as there are multiple cell(s)-of-origin^{17–19}, each postulated to give
14 rise to a specific breast cancer subtype. Whether individual precursor cells from a single
15 tissue have intrinsic differences in their metabolism remains unknown.

16 The mammary gland is composed of two epithelial lineages, the basal and luminal
17 lineages. Milk-producing luminal cells and contractile basal cells operate in unison to carry
18 out the overall function of the breast²⁰. The luminal lineage can be segregated into luminal
19 progenitors and mature luminal (more differentiated) populations, whereas markers to
20 segregate subpopulations within the basal lineage have not been conclusively defined
21 (Figure 1A). Each of these three normal mammary epithelial cell (MEC) types serve as
22 the putative cell-of-origin for distinct breast cancer subtypes. Expression analyses have
23 projected that basal cells give rise to the Claudin-low subtype, mature luminal cells to
24 Luminal A & B and luminal progenitors transform to the aggressive Basal-like subtype²¹.
25 Mouse models with lineage-specific promoters also support the observation that the same
26 mutational event results in different breast cancers depending upon the cell-of-origin^{22–24}.
27 Transcriptomic and epigenomic profiling of both human and mouse normal MECs have
28 revealed lineage-specific regulatory networks^{25–27}. With respect to metabolism, fetal
29 mammary stem cells had high transcript levels of glycolysis enzymes²⁸ and luminal
30 progenitors were shown to have a greater capacity to handle reactive oxygen species
31 (ROS) than basal cells²⁹. Nevertheless, the metabolic networks of normal mammary cell
32 types have yet to be resolved and whether breast cancer subtypes retain these metabolic
33 features from their distinct cells-of-origin remains unknown.

34 Here, we uncover the distinct metabolic identities of the three normal mammary
35 epithelial cell types by using a combination of proteomics, characterization of the
36 mitochondria and pharmacological inhibition. In addition, their distinct metabolic networks
37 not only underlie the differential dependencies of mammary progenitors to metabolic
38 inhibitors, but are also inherited by the specific breast cancer subtypes.

39

1 RESULTS

2 Proteomes of Human Mammary Cells Expose Differential Metabolic Protein 3 Abundance

4 To discover protein distinctions of primary human MEC populations we generated their
5 global proteomes. We performed mass spectrometry-based shotgun proteomics on
6 equivalent numbers of FACS-purified basal (CD45⁻CD31⁻CD49^{hi}EpCAM^{lo-med}; color-
7 coded as red in all figures), luminal progenitor (CD45⁻CD31⁻CD49^{lo}EpCAM^{med}; light
8 blue), and mature luminal (CD45⁻CD31⁻CD49^{hi}EpCAM^{lo}; dark blue) cells from 10 normal
9 human breast samples obtained from reduction mammoplasties (Figure 1B, S1A). Our
10 patient cohort represented diverse physiologies, covering a wide age range (28-67 years
11 old) and sex hormone status (3 luteal, 3 follicular, 4 post-menopausal). We detected 6034
12 unique proteins (Figure 1B). Expression of known markers for each mammary cell type
13 was accurately captured by our proteomics data (Figure S1B); higher abundance of
14 Vimentin and ITGA6 (Integrin α 6, CD49f) was seen in basal cells, higher KIT and
15 ALDH1A3 levels in luminal progenitors, and higher GATA3, FOXA1 and KRT8/18
16 (Cytokeratin 8/18) in the mature luminal. Principal component analysis highlighted the
17 distinct proteomes of mammary cells; the dominant clustering feature was mammary cell
18 type with a minor segregation of post-menopausal samples within each cluster (Figure
19 1C). Out of the 6034 proteins, 5881 were detected in all three cell types (Figure S1C).
20 MEC-specific proteomes separated into deciles based on median intensity were enriched
21 for specific functional classes of proteins (Figure S1D and S1E)³⁰. For instance, the GO
22 biological process in the first decile for basal cells and luminal progenitors was translation
23 of membrane proteins, whereas mature luminal demonstrated abundance for proteins
24 with pleiotropic functions that were annotated as neutrophil terms (Figure S1D and S1E).

25 A metabolic network is defined as the core set of metabolic proteins essential for
26 the structure and function of a cell¹³. We first filtered the total proteomes for metabolic
27 proteins using a curated list of 2753 metabolic enzymes, transporters and subunits³¹. One
28 sixth (1020/6034) of our global mammary proteomic dataset was classified with this
29 annotation (Figure 1B). Unsupervised hierarchical clustering of the metabolic proteome
30 clustered based on mammary lineages (Figure 1D). To determine the metabolic network
31 functioning within each MEC, we sought proteins that were significantly abundant in one
32 population versus the other two (One-way ANOVA in conjunction with a Tukey's test,
33 $P < 0.05$). The resulting metabolic networks for basal, mature luminal and luminal
34 progenitor were composed of 45, 123, 179 metabolic proteins, respectively (significant
35 hits bar in Figure 1D). Pathway analysis using Enrichr^{32,33} revealed unique GO Biological
36 Processes enriched in each metabolic network as found in Figure 1E. We also
37 constructed a global map of MEC metabolism (Figure S2) using a published template³⁴
38 and one focused on glucose metabolism (Figure 1F), where we color-coded proteins to
39 show their corresponding MEC-specificity.

1 Basal cells were enriched for GO terms relating to glycolysis (Figure 1E) and
2 displayed an abundance of glycolytic enzymes (PFKM, ALDOC, GAPDH and PKM);
3 PFKM and PKM perform two of three key irreversible phosphorylation events in glycolysis
4 (Figure 1F). In the mature luminal metabolic network, we noted diverse pathways relating
5 to neutrophil activity, glutamine and glutathione (Figure 1E). It was also enriched for
6 enzymes in hexose and fructose metabolism such as FBP1, ALDOA and LDHB; LDHB
7 diverts pyruvate from the TCA cycle by converting it to lactate (Figure 1F). It was striking
8 that the top 10 pathways in luminal progenitors related to oxidative phosphorylation
9 (OXPHOS; Figure 1E), demonstrating greater abundance of the majority of electron
10 transport chain (ETC) subunits as well as nearly all enzymes in the TCA cycle (Figure
11 1F). MECs had isozyme-specific expression of IDH (mitochondrial IDH3 in basal cells,
12 IDH2 in luminal progenitors and IDH1 in mature luminal), possibly due to different levels
13 of (NAD(P)H) and demand of that in particular cell type³⁵. Pyruvate generated from the
14 carbons of glucose is considered a major contributor to the TCA cycle, however luminal
15 progenitors did not display any enrichment of glycolytic enzymes. Interrogation of our
16 MEC metabolism map (Figure S2) revealed numerous non-glycolytic mechanisms to
17 generate TCA cycle intermediates in luminal progenitors. This was the only population to
18 exhibit enrichment for enzymes involved in branched-chain amino acid catabolism
19 (modifying isoleucine, leucine and valine into acetyl-CoA) and was also strongly enriched
20 for proteins involved in β -oxidation (fatty acids into acetyl-CoA) (Figure S2). Luminal
21 progenitors also had high level of PHGDH, a key enzyme in serine biosynthesis, shown
22 to contribute ~50% of the anaplerotic flux into the TCA cycle³¹ (Figure S2). This
23 engagement of diverse metabolic pathways that break down nutrients to feed the TCA
24 cycle underscores the strong preference of OXPHOS in this cell type. Altogether,
25 proteomes revealed that each mammary cell type has a distinct metabolic network, which
26 may represent the core set of metabolic proteins necessary for its structure and function.
27

28 **Mitochondria Structure and Function is Mammary Cell Type-Specific**

29 Next, we interrogated our published mouse mammary proteomic dataset²⁵ derived from
30 analogous MEC populations and found that metabolic proteomes clustered based on
31 mammary cell types (Figure S3A and S3B), similar to human MECs (Figure 1D). Since
32 human luminal progenitors were endowed with TCA cycle and ETC proteins, we utilized
33 murine MECs cultured as a monolayer to examine their capacity to undergo OXPHOS as
34 measured by the Seahorse bioanalyzer (Figure 2A and 2B). Specifically, we performed
35 the standard mitochondrial stress test, which quantifies oxygen consumption rate (OCR),
36 a readout for mitochondrial respiration, while exposing cells to inhibitors (Oligomycin,
37 Antimycin A) or enhancers (FCCP) of this process. At baseline respiration, basal cells
38 had the lowest level of OCR compared to either of the luminal populations (Figure 2B).
39 Even with the addition of FCCP, which boosts OCR, basal cells had OCR levels
40 comparable to or less than the baseline OCR of the two luminal cell types. Luminal

1 progenitors and mature luminal cells had similar OCR profiles, except for maximal
2 respiration, which was significantly higher in luminal progenitors (Figure 2B). Thus,
3 mammary cell types have distinct capacities for mitochondrial respiration, with luminal
4 progenitors showing the highest OXPHOS capacity.

5 We next examined mitochondrial morphologies by transmission electron
6 microscopy (TEM) of pelleted FACS purified mouse MEC (Figure 2C and S3C). Basal
7 cells tended to have several small circular mitochondria with glossy cristae. This is similar
8 to the morphology of mitochondria in hematopoietic and embryonic stem cells³⁶. In
9 contrast, luminal progenitors had long, tubular mitochondria with elaborate cristae. The
10 larger size of the mitochondria and higher cristae density are thought to be efficient in
11 supporting OXPHOS³⁷, consistent with our data on OCR (Figure 2B). Mature luminal cells
12 surprisingly had indiscernible mitochondria, possibly due to the shearing stress
13 experienced by these larger cells during FACS. We performed intracellular flow cytometry
14 to further characterize the mitochondria (Figure 2D), using MitoTracker Green (MTG; total
15 level of mitochondria) and MitoTracker Red (MTR; mitochondrial activity). Both dyes
16 showed no significant differences among MECs (Figure 2D). CellROX and MitoSOX
17 measure cellular and mitochondrial reactive oxygen species (mROS), respectively.
18 Although total cellular ROS showed minimal differences, mROS levels varied significantly
19 (Figure 2D). Basal and mature luminal cells had equivalent high levels of mROS, whereas
20 luminal progenitors had the least amount despite having high mitochondrial respiration
21 (Figure 2D). These observations can be explained by the multiple antioxidant
22 mechanisms previously reported in luminal progenitors but not in basal cells²⁹.

23 The high mROS levels in basal cells were intriguing, as they did not have high
24 OCR. This led us to ask whether the mitochondria had an alternative role in this population
25 beyond bioenergetics. The non-ATP functions of the mitochondria are becoming more
26 appreciated. For example, mitochondrial membrane potential has been linked to stem cell
27 capacity^{38,39}. We filtered our mouse and human MEC proteomes using MitoCarta, a
28 curated list of mitochondrial proteins⁴⁰, and observed cell type-based clusters by
29 unsupervised hierarchical clustering (Figure 2E). Basal and luminal lineages are each
30 enriched for their own progenitors. We segregated MECs based on high (MTG^{hi}MTR^{hi}) or
31 low (MTG^{hi}MTR^{lo}) mitochondrial activity⁴¹, and enumerated luminal and basal progenitor
32 capacity using the colony-forming cell (CFC) assay (Figure 2F). Cells with high
33 mitochondrial activity had significantly greater CFC number than those with low
34 mitochondrial activity in both mammary lineages (Figure 2E). Basal, but not luminal, cells
35 with high mitochondrial activity also displayed enrichment of CFC capacity when
36 compared to their total unfractionated control. The EpCAM⁺CD49^{hi} basal cell
37 compartment contains mammary stem cells, basal progenitors and differentiated cells.
38 There is avid interest in teasing out new markers for progenitor-enriched basal subsets
39 and our data show that mitochondrial activity may serve such a role. Overall, our findings
40 also show that mitochondrial morphology and function varies with mammary cell type.

1

2 **Mammary Lineages Exhibit Differential Metabolic Vulnerabilities**

3 We determined if metabolic distinctions of mammary lineages manifested as differential
4 sensitivity to various metabolic drugs using the CFC assay (Figure 3A). To inhibit
5 OXPHOS, we used complex-specific (Rotenone→Complex I; Atpenin A5→Complex II;
6 Antimycin A→Complex III; Oligomycin→Complex V) and non-ETC inhibitors
7 (Tigecycline→mitochondrial ribosomes; UK5099→Mitochondrial pyruvate carrier).
8 Furthermore, inhibition of glycolysis was achieved at multiple levels (BAY-876→Glucose
9 transporter 1; 2-Deoxy-D-glucose→Hexokinase; Galloflavin→Lactate dehydrogenase;
10 Dichloroacetate→Pyruvate dehydrogenase kinase). Most metabolic inhibitors resulted in
11 a potent dose-dependent reduction in progenitor capacity of both lineages, as
12 enumerated by absolute CFCs (Figure S4A & S4B). Relative CFC counts allow
13 comparison of the selective vulnerability of luminal and basal progenitors to metabolic
14 inhibitors (Figure 3B and 3C). Analyses of the dependencies show that the two mammary
15 lineages require specific ETC complexes for their progenitor capacity. We observed that
16 inhibition of Complex I preferentially decreased luminal CFCs (Figure 3B) whereas basal
17 CFCs were significantly more sensitive to Complex II or III inhibition. Complex V inhibition
18 showed no selective effect (Figure 3B). Tigecycline treatment abrogated the progenitor
19 capacity of basal over luminal CFCs (Figure 3B), supporting our earlier data that
20 mitochondria strongly influence basal progenitor activity (Figure 2F). UK5099 prevents
21 entry of pyruvate into mitochondria, but had a minimal effect on CFC capacity (Figure
22 3B), in line with our observation that luminal progenitors may not rely on cytosolic pyruvate
23 (Figure 1E and S2). Pathway analyses of the basal metabolic network had highlighted
24 glycolysis as the most significant term (Figure 1E). Our series of glycolytic drugs
25 demonstrated that basal CFCs were far more sensitive than luminal CFCs to all 4
26 compounds (Figure 3C). Collectively, this set of experiments demonstrates the lineage-
27 specific metabolic vulnerabilities of mammary cells.

28 **Breast Cancer Subtypes Retain Metabolic Features of their Putative Cell-of-Origin**

29 To interrogated whether any of the PAM50 breast cancer subtypes had significant
30 enrichment of our MEC-specific metabolic network (Figure 4A) we performed single
31 sample gene set enrichment analysis (ssGSEA)⁴² on breast cancer patients from the
32 METABRIC database⁴³. We observed striking relationships between the metabolic
33 preferences of normal MEC and breast cancer subtypes (Figure 4A). Specifically, the
34 highly mesenchymal Claudin-low subtype was most enriched for the basal cell metabolic
35 network. The aggressive basal-like breast cancer was most significantly correlated to the
36 luminal progenitor metabolic network (Figure 4A). Luminal A and B subtypes showed
37 significant enrichment for the mature luminal metabolic network (Figure 4A). Thus, breast
38 cancer subtypes displayed strong activity for metabolic cluster of their proposed cell-of-

1 origin, suggesting that breast cancers co-opt the metabolic network of their precursor
2 cells.

3 Recent studies have reported the successful targeting of metabolic vulnerabilities
4 which are specific to the tissue-of-origin^{16,44} or stem from chromosomal abnormalities^{45,46}.
5 We therefore used cBioportal^{47,48} to determine copy number amplifications in metabolic
6 genes within our MEC-specific networks in order to identify novel subtype-specific
7 metabolic targets. Interrogation of our luminal progenitor metabolic network revealed
8 *PHGDH*, a known amplified gene and selective vulnerability in ER- basal-like breast
9 cancers^{31,49}, the subtype thought to originate from luminal progenitors. We identified 5
10 other highly abundant proteins in our mature luminal metabolic network, namely EPHX1,
11 NIT1, CYB5R1, GALNT2 and KMO, whose genes were amplified in ER+, PR+ as well as
12 most consistently in Luminal A and B breast cancers (Figure 4B). These 5 proteins do not
13 participate in the same metabolic pathway but are all found on chromosome 1q. Whole-
14 arm amplification of 1q together with 16q loss (+/-) is a hallmark chromosomal event in
15 ER+ breast cancers^{43,50,51}. The fact that we find metabolic network specific proteins being
16 amplified at the chromosome level in the respective breast cancer subtypes points to
17 these targets as possible cell-of-origin-specific metabolic vulnerabilities, which require
18 further investigation.

19 **DISCUSSION**

20 Using a combination of proteomics, characterization of the mitochondria and
21 pharmacological inhibition, we uncovered distinct metabolic identities of the three normal
22 mammary epithelial cell types (Figure 4C). This highlights a previously underappreciated
23 metabolic heterogeneity present in the epithelial compartment of the normal human and
24 mouse mammary gland. The observed lineage-drive metabolic programs may be intrinsic
25 to cell identity or a reflection of cellular adaptations to distinct mammary
26 microenvironments. Basal cells are in contact with the basement membrane which
27 separates the epithelial layers from a complex mammary stroma composed of immune
28 cells⁵², adipocytes⁵³ and fibroblasts⁵⁴. Conversely, luminal cells are exposed apically to the
29 lumen of the mammary ductal tree. Since all our analyses were performed *ex vivo* on
30 purified mammary cells, we reason that metabolic distinctions are hardwired and likely
31 necessary to facilitate unique form and function of each mammary cell type. Our MEC-
32 specific metabolic networks will enable further study into the influence of normal cells on
33 the metabolic phenotype of known breast cancer subtypes. In addition, global proteomes
34 of primary FACS-purified human and analogous mouse mammary cell types provide a
35 valuable resource to further understand the regulatory networks that define these different
36 epithelial lineages.

37 The metabolic phenotype of a cancer cell is dependent upon integrating multiple
38 intrinsic and extrinsic cues⁵⁵. The importance of the tissue-of-origin in tumor metabolism
39 has now been established¹³. It has also been postulated that tumors located in the same
40 tissue but derived from different cell(s)-of-origin would display different metabolic

1 properties, however this has never been experimentally shown⁵⁶. Our work demonstrates
2 that part of the metabolic heterogeneity observed in breast cancers is instructed by the
3 diverse cellular origins of these cancers (Figure 4C). For instance, Claudin-low, Basal-
4 like, Luminal A & B appear to inherit metabolic features of basal cell, luminal progenitor
5 and mature luminal populations, respectively. Arguably, cell lineage could be one of the
6 most important determinants of cellular metabolism, as all perturbations (mutational or
7 microenvironmental) will hijack the pre-existing metabolic network of the cell-of-origin as
8 a backbone. Thus, in addition to mutational events, other characteristics of the tumor such
9 as the cell-of-origin need to be considered in order to maximize the success of
10 personalized cancer medicine. Our study lays the foundation for rationalized targeting of
11 subtype-specific metabolic vulnerabilities, as informed by the metabolic networks of
12 mammary epithelial cells.

13
14
15
16

17 **AUTHOR CONTRIBUTIONS**

18 Conceptualization: MM, AEC, RK.
19 Methodology: MM, AEC, DP, CE, AS, VI, HK, VS.
20 Formal analysis: MM, AEC, KA, LP, MGV.
21 Resources: AS, TK, HB, CE, RK.
22 Writing: MM, TK, MAP, RK.
23 Visualization: MM, MA, KA.
24 Funding acquisition: TK, CE, RK.
25 Supervision: TK, MAP, RK.

26 **ACKNOWLEDGMENTS**

27 We thank members of the Khokha Lab for reviewing this manuscript and helpful
28 discussions. We are grateful to the Nanoscale Biomedical Imaging Facility at SickKids
29 and the Princess Margaret and SickKids flow cytometry core facilities. This work is
30 supported by funding from Canadian Institutes of Health Research (CIHR), Canadian
31 Breast Cancer Foundation, and Canadian Cancer Society Research Institute. MM
32 received a CIHR Masters Award.

33
34
35
36

1 METHODS

2 **Human patient samples:** All human tissue was acquired with patient consent and approval by
3 the Institutional Research Ethics Board of the University of British Columbia (UBC; Vancouver,
4 BC) and University Health Network (Toronto, ON). Hormonal status (premenopausal, follicular
5 and luteal) was determined by a pathologist examining breast specimens at UBC (Ramakrishnan
6 et al., 2002). Reduction mammoplasty specimens were minced and enzymatically dissociated in
7 DMEM:F12 1:1 media with 15 mM HEPES plus 2% BSA, 1% penicillin-streptomycin, 5 µg/ml
8 insulin, 300 U/ml collagenase (Sigma, C9891) and 100 U/ml hyaluronidase (Sigma, H3506)
9 shaking gently at 37°C, overnight or for 16-18 hours. Epithelial organoids were harvested by
10 centrifugation at 80g for 30 seconds and viably cryopreserved, as described previously (Labarge
11 et al., 2013).

12 **Human breast single cell suspensions:** Human breast tissue organoids were thawed and
13 dissociated into single cell suspensions as reported previously (Eirew et al., 2010) . Briefly,
14 organoids were triturated in 0.25% trypsin-EDTA (Stem Cell Technologies, 07901) followed by 5
15 U/ml dispase (Stem Cell Technologies, 07913) and 50 µg/ml DNase I (Sigma, D4513) as
16 described above for mouse samples, but for 5 minutes each. Cells were then washed in between
17 steps with HBBS + 2% FBS and filtered using a 40 µm cell strainer.

18 **Human breast FACS staining:** For FACS staining, antibodies against CD45 (PECy7), CD31
19 (PECy7), EpCAM (APCCy7) and CD49f (FITC) were used. Lineage (Lin) positive cells were
20 defined as CD31⁺CD45⁺. Human mammary cell subpopulations were defined as: basal (Lin⁻
21 EpCAM^{lo-med}CD49f^{hi}); luminal progenitor (Lin⁻EpCAM^{hi}CD49f^{med}); mature luminal (Lin⁻
22 EpCAM^{hi}CD49f^{lo}). Dead cells were excluded following doublet exclusion using DAPI.

23 **Mice:** All experiments were performed using 8-12 weeks old virgin female FVB wild-type mice
24 (The Jackson Laboratory or Charles River). Mice were ovariectomized bilaterally, then allowed
25 one week to recover. A slow-release 0.14 mg 17-β estradiol plus 14 mg progesterone pellet
26 (Innovative Research of America) was then placed subcutaneously near the thoracic mammary
27 gland for 2 weeks. This was done to obtain large quantities of viable mammary stem/progenitor
28 cells for subsequent analysis, as previously reported (Casey et al., 2018; Shiah et al., 2015). All
29 mice were cared for according to guidelines established by the Canadian Council for Animal Care
30 under protocols approved by the Animal Care Committee of the Ontario Cancer Institute.

31 **Mouse mammary single cell suspensions:** Harvested mammary glands were manually minced
32 with scissors for 2 minutes, and then enzymatically dissociated using 750 U/ml collagenase and
33 250 U/ml hyaluronidase (Stem Cell Technologies, 07912) and diluted in DMEM:F12 for 1.5 hours.
34 Samples were vortexed at the 1- and 1.5-hour mark. Red blood cells were lysed using ammonium
35 chloride (Stem Cell Technologies, 07850). Cells were then mixed in trypsin-EDTA (0.25%, Stem
36 Cell Technologies, 07901) that had been pre-warmed to 37°C using a 1mL pipette for 2 minutes.
37 Next, they were washed in Hanks Balanced Salt Solution (HBSS) without calcium or magnesium
38 plus 2% FBS and centrifuged at 350g. Finally, cells were mixed in dispase 5 U/ml (Stem Cell
39 Technologies, 07913) plus 50 µg/ml DNase I (Sigma, D4513) for 2 minutes, washed in HBBS +
40 2% FBS and filtered using a 40 µm cell strainer to obtain single cells.

1 **Mouse mammary FACS staining:** Dead cells were excluded following doublet exclusion using
2 DAPI or Zombie UV Fixable Viability Kit (BioLegend) according to manufacturer's instructions.
3 For FACS staining, antibodies against TER119 (PECy7 or eFluor450), CD31 (PECy7 or
4 eFluor450), CD45 (PECy7 or eFluor450), EpCAM (APCCy7), CD49f (FITC or PECy7), CD49b
5 (PE) and Sca-1 (APC or Brilliant Violet 711) were used. Lineage (Lin) positive cells were defined
6 as Ter119⁺CD31⁺CD45⁺. Mouse mammary cell subpopulations were defined as: total basal (Lin⁻
7 EpCAM^{lo-med}CD49f^{hi}); total luminal (Lin⁻EpCAM^{hi}CD49f^{lo}); luminal progenitor (Lin⁻
8 EpCAM^{hi}CD49f^{lo}CD49b⁺Sca-1⁻); mature luminal (Lin⁻EpCAM^{hi}CD49f^{lo}CD49b^{-/+}Sca-1^{+/+}). High and
9 low mitochondrial activity populations were defined as MitoTracker Red^{Hi}MitoTracker Green^{hi} and
10 MitoTracker Red^{lo}MitoTracker Green^{hi}, respectively, and applied after gating for total luminal and
11 basal populations. Fluorophores are specifically mentioned in individual figures. Cell sorting was
12 performed on a BD FACSAria™ II.

13 **Mouse CFC assay:** 350 cells of the specified FACS-purified population were seeded together
14 with 20,000 irradiated NIH 3T3 cells in a 6-well plate. Cells were cultured for 7 days at 5% oxygen
15 in EpiCult-B mouse medium (Stem Cell Technologies, 05610) supplemented with 5% FBS, 10
16 ng/ml EGF, 20 ng/ml basic FGF, 4 µg/ml heparin, and 5 µM ROCK inhibitor (Millipore). Cells were
17 allowed to adhere for 24 hours, and then either vehicle control (0.1% DMSO) or the indicated
18 concentrations of inhibitors were added for the remaining six days.

19 **Mammary cell intracellular flow cytometry:** All intracellular dyes were used to stain cells prior
20 to cell surface marker staining protocol. Staining for total mitochondria (50 nM MitoTracker Green
21 FM, Thermo Fisher, M7514), mitochondrial activity (250 nM MitoTracker Red CMXRos, Thermo
22 Fisher, M7513), mitochondrial ROS (5 µM MitoSOX, Thermo Fisher, M36008), and cytosolic ROS
23 (5 µM CellROX Green, Thermo Fisher, C10492) was performed by incubating cells at 37°C for
24 20-30 minutes following the manufacturer's protocols and directly analysed without fixing. Cell
25 analysis was performed in BD Biosciences Fortessa. Median fluorescent intensity (MFI) refers to
26 the fluorescence intensity of each event (on average) of the selected cell population, in the chosen
27 fluorescence channel (PE Texas Red or FITC) and was determined by using the flow cytometry
28 analysis software FlowJo.

29 **Metabolic inhibitors used *in vitro*:** Vehicle and drugs were added such that the final
30 concentration of DMSO did not exceed 0.1% (vol/vol). The following drugs were used in this study:
31 2-Deoxy-D-glucose (Sigma; D8375), dichloroacetate (Sigma; 347795), BAY-876 (Structural
32 Genomics Consortium), rotenone (Sigma; R8875), tigecycline (CarboSynth, 220620-09-7),
33 antimycin A (Sigma, A8674), oligomycin (Sigma, 75351), atpenin A5 (Cayman Chemicals,
34 11898), UK-5099 (Sigma, PZ0160), galloflavin (Sigma; SML0776).

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

1 **Seahorse:** MEC subpopulations (Luminal progenitor, mature luminal and basal cell) were FACS-
2 purified from unstaged mice and 10,000 cells were plated into each well of collagen pre-coated
3 Seahorse plates. The cells were cultured in the 5% O₂ incubator for 6 days to reach at least 80-
4 90% confluence. On the 7th days, cells were switched to DMEM:HAM's F12 with no bicarbonate
5 containing 5% FBS, insulin (Thermo Fisher, 12585014), EGF (STEMCELL Technologies;
6 78006.1), bFGF (STEMCELL Technologies), hydrocortisone (STEMCELL Technologies,
7 78003.1), Rock inhibitor (Millipore, SCM075) in 5% oxygen conditions. Then the plate was allowed
8 to equilibrate for 1 hour in the Seahorse incubator. Inhibitors used for the assay include oligomycin
9 (2 µM), FCCP (1 µM, Sigma, C2920) and antimycin A (1 µM). After the assay, cell viability was
10 determined using the CyQUANT nuclear dye (Thermo Fisher, C35007). Data was analyzed on
11 the WAVE platform and normalized to the number of live cells determined after the viability assay.

12 **Proteomics on FACS-purified human mammary subpopulations:** For Liquid
13 Chromatography-Mass Spectrometry (LC-MS) of human mammary subpopulations, 100,000 cells
14 from each population were isolated from each patient, as described (Casey et al., 2018). After
15 FACS purification, cells were washed in ice-cold PBS and pelleted. Pellets were then
16 resuspended in 50% (vol/vol) 2, 2, 2-trifluoroethanol in PBS and disrupted into cellular lysates
17 sequentially by repeated probe sonication, followed by six freeze-thaw cycles. Proteins in cellular
18 lysates were denatured by incubation at 60°C for 2 h, oxidized cysteines reduced using 5 mM
19 dithiothreitol for 30 min at 60°C and alkylated through reaction with 25 mM iodoacetamide for 30
20 min at room temperature in the dark. Each sample was diluted five times using 100 mM
21 ammonium bicarbonate, pH 8.0. Proteins were digested into peptides through addition of 5 µg of
22 MS-grade trypsin (Promega). The digestion was performed overnight at 37°C and subsequently
23 desalted using OMIX C18 pipette tips (Agilent). Peptides were semidried through vacuum
24 centrifugation and resuspended in water with 0.1% formic acid. Subsequently, all samples were
25 analyzed using an Easy-LC1000 (Thermo Fisher Scientific) coupled to the Orbitrap Fusion
26 tandem mass spectrometer (Thermo Fisher Scientific). Peptides were separated on an ES803
27 (Thermo Fisher Scientific) nano-flow column heated to 50°C using a 4-h reverse-phase gradient.

28

29 [Bioinformatics Analysis of human mammary subpopulation proteomes](#)

30 **Proteomics Processing:** Mass spectrometric data was analyzed using the MaxQuant
31 quantitative proteomics software (version 1.5.8.3) and a Human UniProt sequences FASTA
32 database (complete human proteome; release 2015-01, 42,041 sequences).
33 Carbamidomethylation of cysteine was specified as a fixed modification and oxidation of
34 methionine was specified as a variable modification. Proteins were identified with a minimum of
35 two razor+unique peptides, the maximum false peptide discovery rate was specified as 1%, and
36 “match between runs” was enabled. The distribution of intensity-based absolute quantification
37 (iBAQ) values was adjusted to the distribution of label-free quantification (LFQ) values based on
38 the median for each sample. This allowed for imputation of missing LFQ values with iBAQ
39 values (Wojtowicz et al., 2016). Non-zero values were log₂-transformed. The final list consisted of
40 6034 unique protein groups detected in at least one of the samples. Further data processing was
41 performed using the R statistical environment (version 3.5.2) (Bunn and Korpela). For protein
42 groups in which both LFQ and iBAQ values were missing, the 0 values were imputed with a

1 random value between 1 and 1.5. Imputation was performed as a precautionary measure for
2 further statistical analysis. As four samples were run on a separate day, intensity values were
3 then adjusted for sample batch effects using the ComBat method in the surrogate variable
4 analysis “sva” R package (version 3.30.1) (Johnson et al., 2007; Leek et al.).

5
6 **Total Proteome Bioinformatics:** Non-imputed ComBat-modified iBAQ-adjusted LFQ values
7 were used to discover uniquely expressed proteins in each cell type. Averages across samples
8 in each cell type were taken, resulting in one mean expression value for each protein in each cell
9 type ($n_{BC} = 9$, $n_{LP} = 10$, $n_{ML} = 10$). Next, the values of zero for each cell type and associated
10 proteins were excluded from the analysis. Number of proteins expressed in each cell type were
11 summarized in a Venn diagram, created using the “VennDiagram” R package (version 1.6.20)
12 (Chen and Boutros, 2011). Gene set enrichment was conducted on the same values. Mean
13 expression values for each protein in each of the cell types were ranked according to descending
14 \log_2 median intensities and grouped into deciles. The protein with the highest intensity received
15 a rank of 1 and thus, was placed in the first decile. Meanwhile, the protein with the lowest mean
16 intensity received a rank of y and was placed in the tenth decile, where y represents the total
17 number of proteins detected in a particular cell type. Pathway analysis via the “enrichR” R
18 package (version 1.0) was conducted on the proteins in each decile (Chen et al., 2013; Kuleshov
19 et al., 2016).

20
21 Principal component analysis (PCA) was performed by calculating Euclidean distances of scaled
22 expression values. PCA scores were plotted in a plane defined by the first two components (that
23 is, PC1 and PC2) using the “ggbiplot” R package (Vu, 2019). Ellipses were drawn around cell
24 type clusters, where centroids were the barycentre of each cluster and the diameter represented
25 the maximum variance.

26
27 Heat maps depicted z-scores of protein expression values (x) computed using the formula: $(x -$
28 $\text{mean}(x))/\text{standard_deviation}(x)$. Divisive hierarchal clustering dendrograms of Pearson distance
29 matrices for samples and proteins were created using DIANA (DIvisive ANALysis Clustering)
30 method in the “cluster” R package (version 2.0.7-1) (Maechler et al.). Heat maps were plotted
31 using the “pheatmap” (version 1.0.12) and “RColorBrewer” (version 1.1-2) R packages (Kolde,
32 2019; Neuwirth, 2014).

33
34 **Metabolic Cluster Derivation and Pathway Analysis:** A metabolic proteome was obtained by
35 filtering the total proteome using a curated list of 2753 genes that encompasses all known human
36 metabolic enzymes and transporters (Possemato et al., 2011). Based on matching by gene
37 symbols, 1020 proteins related to metabolism were found in the total proteome of 6034 proteins,
38 including “PKM” which was not identified in the curated list. As multiple protein groups in the
39 proteome shared the same gene symbols, duplicates were included in the analysis. Metabolic
40 signatures were acquired by looking at proteins in which mean expression met the fold-change
41 and statistical change cut-offs in each cell type compared to the other two cell types ($n_{BC} = 9$, n_{LP}
42 $= 10$, $n_{ML} = 10$). The \log_2 fold-change (FC) cut-off was greater than 0 and the statistical
43 significance cut-off was $P < 0.05$ in a one-way ANOVA and Tukey’s multiple comparisons test.
44 Pathway analysis of metabolic clusters was conducted using Enrichr

1 (<https://amp.pharm.mssm.edu/Enrichr/>). Enrichr is a comprehensive gene set enrichment tool that
2 is available both as a web interface (Chen et al., 2013) and an R package (Kuleshov et al., 2016).
3 It queries a list of gene symbols and returns commonly annotated pathways by searching large
4 gene set libraries. The gene set library selected for our analysis was Gene Ontology Biological
5 Process (GOBP) 2018. For each cell-type signature, the top ten GOBP terms enriched by gene
6 sets were sorted by lowest to highest combined score ($\ln(p\text{-value}) * z\text{-score}$), a metric used by
7 Enrichr to find the best ranking terms compared to other methods.

8
9 **Correlations to PAM50 Breast Cancer Subtypes:** Gene expression for PAM50 breast cancer
10 subtypes (Her2, Luminal A, Luminal B, Basal-like, and Claudin-Low) and clinical annotations was
11 performed in the METABRIC cohort (Curtis et al., 2012) and was obtained from cBioPortal
12 (Cerami et al., 2012; Gao et al., 2013). It provided gene expression profiles and classified breast
13 cancer subtypes for 1980 patients. The gene expression profiles for the breast cancer subtypes
14 (were correlated to our metabolic signatures via single-sample Gene Set Expression Analysis
15 (ssGSEA) using the “GSVA” R package (version 1.30.0) (Hänzelmann et al., 2013). ssGSEA
16 scores for each signature in the breast cancer subtypes were assessed for significance using a
17 one-way ANOVA and student’s t-test.

18
19 **Statistical Analysis and Reproducibility:** All details pertaining to biological “n” numbers or error
20 bars can be found in the relevant figure legends. Details pertaining to the statistical analysis of
21 global and metabolic proteome can be found in the relevant methods section detailing
22 bioinformatics analyses. Statistically significant differences are indicated by asterisks, which
23 denote size of significance levels (p-values: ns $P > 0.05$; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$;
24 **** $P \leq 0.0001$.) For intracellular flow cytometry analysis statistical significance was calculated
25 using two-way ANOVA and Tukey’s multiple comparisons test. For *in vitro* clonogenic assays
26 comparing high and low mitochondrial mammary cells, a two-way ANOVA and Bonferonni’s
27 multiple comparison test was used. For *in vitro* clonogenic assays, statistical significance for all
28 drug testing comparisons was calculated using two-way ANOVA and Sidak’s multiple
29 comparisons test.

30 **Data Availability:** The mass spectrometry data associated with this manuscript will be submitted
31 to a public repository (the Mass spectrometry Interactive Virtual Environment;
32 <http://massive.ucsd.edu>). These data are associated with the identifier _____ at FTP
33 download site: _____. The mouse mammary proteome data (used in Figure 2E, S3B) is
34 published (Casey et al., 2018) and can be downloaded from the FTP download
35 site: <ftp://MSV000079330@massive.ucsd.edu> with the identifier MSV000079330

36
37
38
39
40
41
42

1 REFERENCE

- 2 1. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- 3 2. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of
4 breast cancer. *Breast Cancer Res.* **12**, R68 (2010).
- 5 3. Sørliie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with
6 clinical implications. *Proc. Natl. Acad. Sci.* **98**, 10869–10874 (2001).
- 7 4. Brauer, H. A. *et al.* Impact of Tumor Microenvironment and Epithelial Phenotypes on Metabolism in
8 Breast Cancer. *Clin. Cancer Res.* **19**, 571–585 (2013).
- 9 5. Budczies, J. *et al.* Comparative metabolomics of estrogen receptor positive and estrogen receptor
10 negative breast cancer: alterations in glutamine and beta-alanine metabolism. *J. Proteomics* **94**,
11 279–288 (2013).
- 12 6. Cappelletti, V. *et al.* Metabolic Footprints and Molecular Subtypes in Breast Cancer. *Dis. Markers*
13 **2017**, 1–19 (2017).
- 14 7. Tang, X. *et al.* A joint analysis of metabolomics and genetics of breast cancer. *Breast Cancer Res. BCR*
15 **16**, 415 (2014).
- 16 8. Terunuma, A. *et al.* MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer
17 prognosis. *J. Clin. Invest.* **124**, 398–412 (2014).
- 18 9. Kulkoyluoglu-Cotul, E., Arca, A. & Madak-Erdogan, Z. Crosstalk between Estrogen Signaling and
19 Breast Cancer Metabolism. *Trends Endocrinol. Metab.* **30**, 25–38 (2019).
- 20 10. Zhang, D. *et al.* Proteomic study reveals that proteins involved in metabolic and detoxification
21 pathways are highly expressed in HER-2/neu-positive breast cancer. *Mol. Cell. Proteomics MCP* **4**,
22 1686–1696 (2005).
- 23 11. DeBerardinis, R. J. & Chandel, N. S. Fundamentals of cancer metabolism. *Sci. Adv.* **2**, e1600200
24 (2016).

- 1 12. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours.
2 *Nature* **490**, 61–70 (2012).
- 3 13. Mayers, J. R. & Vander Heiden, M. G. Nature and Nurture: What Determines Tumor Metabolic
4 Phenotypes? *Cancer Res.* **77**, 3131–3134 (2017).
- 5 14. Yuneva, M. O. *et al.* The metabolic profile of tumors depends on both the responsible genetic lesion
6 and tissue type. *Cell Metab.* **15**, 157–170 (2012).
- 7 15. Hu, J. *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic
8 network. *Nat. Biotechnol.* **31**, 522–529 (2013).
- 9 16. Mayers, J. R. *et al.* Tissue of origin dictates branched-chain amino acid metabolism in mutant Kras-
10 driven cancers. *Science* **353**, 1161–1165 (2016).
- 11 17. Oakes, S. R., Gallego-Ortega, D. & Ormandy, C. J. The mammary cellular hierarchy and breast cancer.
12 *Cell. Mol. Life Sci.* **71**, 4301–4324 (2014).
- 13 18. Tharmapalan, P., Mahendralingam, M., Berman, H. K. & Khokha, R. Mammary stem cells and
14 progenitors: targeting the roots of breast cancer for prevention. *EMBO J.* **38**, e100852 (2019).
- 15 19. Visvader, J. E. & Stingl, J. Mammary stem cells and the differentiation hierarchy: current status and
16 perspectives. *Genes Dev.* **28**, 1143–1158 (2014).
- 17 20. Inman, J. L., Robertson, C., Mott, J. D. & Bissell, M. J. Mammary gland development: cell fate
18 specification, stem cells and the microenvironment. *Development* **142**, 1028–1042 (2015).
- 19 21. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor
20 development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).
- 21 22. Koren, S. *et al.* PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nature*
22 **525**, 114–118 (2015).
- 23 23. Molyneux, G. *et al.* BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and
24 not from basal stem cells. *Cell Stem Cell* **7**, 403–417 (2010).

- 1 24. Van Keymeulen, A. *et al.* Reactivation of multipotency by oncogenic PIK3CA induces breast tumour
2 heterogeneity. *Nature* **525**, 119–123 (2015).
- 3 25. Casey, A. E. *et al.* Mammary molecular portraits reveal lineage-specific features and progenitor cell
4 vulnerabilities. *J. Cell Biol.* **217**, 2951–2974 (2018).
- 5 26. Pellacani, D. *et al.* Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active
6 Enhancer States and Associated Transcription Factor Networks. *Cell Rep.* **17**, 2060–2074 (2016).
- 7 27. Shiah, Y.-J. *et al.* A Progesterone-CXCR4 Axis Controls Mammary Progenitor Cell Fate in the Adult
8 Gland. *Stem Cell Rep.* **4**, 313–322 (2015).
- 9 28. Girardi, R. R. *et al.* Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage
10 Specification Programs in Early Mammary Gland Development. *Cell Rep.* **24**, 1653-1666.e7 (2018).
- 11 29. Kannan, N. *et al.* Glutathione-dependent and -independent oxidative stress-control mechanisms
12 distinguish normal human mammary epithelial cell subsets. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7789–
13 7794 (2014).
- 14 30. Sinha, A. *et al.* The Proteogenomic Landscape of Curable Prostate Cancer. *Cancer Cell* **35**, 414-
15 427.e6 (2019).
- 16 31. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in
17 breast cancer. *Nature* **476**, 346–350 (2011).
- 18 32. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.
19 *BMC Bioinformatics* **14**, 128 (2013).
- 20 33. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016
21 update. *Nucleic Acids Res.* **44**, W90-97 (2016).
- 22 34. Lin, K. H. *et al.* Systematic Dissection of the Metabolic-Apoptotic Interface in AML Reveals Heme
23 Biosynthesis to Be a Regulator of Drug Sensitivity. *Cell Metab.* **29**, 1217-1231.e7 (2019).

- 1 35. Losman, J.-A. & Kaelin, W. G. What a difference a hydroxyl makes: mutant IDH, (R)-2-
2 hydroxyglutarate, and cancer. *Genes Dev.* **27**, 836–852 (2013).
- 3 36. Folmes, C. D. L., Dzeja, P. P., Nelson, T. J. & Terzic, A. Metabolic plasticity in stem cell homeostasis
4 and differentiation. *Cell Stem Cell* **11**, 596–606 (2012).
- 5 37. Cogliati, S. *et al.* Mitochondrial Cristae Shape Determines Respiratory Chain Supercomplexes
6 Assembly and Respiratory Efficiency. *Cell* **155**, 160–171 (2013).
- 7 38. Katajisto, P. *et al.* Asymmetric apportioning of aged mitochondria between daughter cells is
8 required for stemness. *Science* **348**, 340–343 (2015).
- 9 39. Wu, M.-J. *et al.* Epithelial-Mesenchymal Transition Directs Stem Cell Polarity via Regulation of
10 Mitofusin. *Cell Metab.* (2018). doi:10.1016/j.cmet.2018.11.004
- 11 40. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian
12 mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).
- 13 41. Tal, M. C. *et al.* Absence of autophagy results in reactive oxygen species-dependent amplification of
14 RLR signaling. *Proc. Natl. Acad. Sci.* **106**, 2770–2775 (2009).
- 15 42. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-
16 seq data. *BMC Bioinformatics* **14**, 7 (2013).
- 17 43. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel
18 subgroups. *Nature* **486**, 346–352 (2012).
- 19 44. Chowdhry, S. *et al.* NAD metabolic dependency in cancer is shaped by gene amplification and
20 enhancer remodelling. *Nature* (2019). doi:10.1038/s41586-019-1150-2
- 21 45. Dey, P. *et al.* Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer.
22 *Nature* **542**, 119–123 (2017).
- 23 46. Tang, Y.-C. *et al.* Aneuploid Cell Survival Relies upon Sphingolipid Homeostasis. *Cancer Res.* **77**,
24 5272–5286 (2017).

- 1 47. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional
2 cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
- 3 48. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the
4 cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
- 5 49. Locasale, J. W. *et al.* Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to
6 oncogenesis. *Nat. Genet.* **43**, 869–874 (2011).
- 7 50. Russnes, H. G. *et al.* Genomic architecture characterizes tumor progression paths and fate in breast
8 cancer patients. *Sci. Transl. Med.* **2**, 38ra47 (2010).
- 9 51. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours.
10 *Nature* **490**, 61–70 (2012).
- 11 52. Chakrabarti, R. *et al.* Notch ligand Dll1 mediates cross-talk between mammary stem cells and the
12 macrophageal niche. *Science* (2018). doi:10.1126/science.aan4153
- 13 53. Joshi, P. A. *et al.* PDGFR α + stromal adipocyte progenitors transition into epithelial cells during
14 lobulo-alveologenesis in the murine mammary gland. *Nat. Commun.* **10**, 1760 (2019).
- 15 54. Morsing, M. *et al.* Evidence of two distinct functionally specialized fibroblast lineages in breast
16 stroma. *Breast Cancer Res. BCR* **18**, 108 (2016).
- 17 55. Vander Heiden, M. G. & DeBerardinis, R. J. Understanding the Intersections between Metabolism
18 and Cancer Biology. *Cell* **168**, 657–669 (2017).
- 19 56. Kim, J. & DeBerardinis, R. J. Mechanisms and Implications of Metabolic Heterogeneity in Cancer. *Cell*
20 *Metab.* **30**, 434–446 (2019).

21

22

23

24

25

1 **FIGURE LEGENDS**

2 **Figure 1: Proteomics illustrate distinct metabolic networks of human MECs.**

- 3 A. Mammary epithelial cell (MEC) hierarchy depicting the basal and luminal lineages
4 and cell surface markers used to FACS-purify basal, luminal progenitor and mature
5 luminal cells.
- 6 B. Schematic depicting workflow on how human breast samples (n=10) were processed
7 to single cells, the FACS gating strategy used to segregate mature luminal (ML),
8 luminal progenitor (LP) and basal cells (BC) populations. Purified fractions were then
9 prepared for liquid chromatography-mass spectrometry (LC-MS). Proteomics yielded
10 6034 uniquely detected proteins, whose abundance was corrected for batch effects,
11 and missing values were imputed prior to downstream analyses. Total proteomes
12 were filtered down to 1020 metabolic proteins.
- 13 C. Principal component analysis of total proteome from human BC, LP and ML. Dot
14 colour represents a mammary cell type, dot-shape represents hormone status
15 (follicular, luteal or postmenopause) and ellipses represents clusters of sample types.
- 16 D. Heatmap showing unsupervised hierarchical clustering and enrichment of the 1020
17 metabolic proteins in human MECs. Patient covariates (cell type, hormone status,
18 age) are shown in the bars aligning the heatmap. Each line found in the “Significant
19 Hits” bar is a metabolic protein whose expression was significantly enriched in only
20 one cell type using a one-way ANOVA in conjunction with Tukey’s test ($p < 0.05$) and
21 colour-coded for that cell type.

1 E. Bar graphs summarize the top 10 most significant GO biological processes according
2 to Enrichr for each cell type's metabolic network. Enrichr calculates a combined
3 score, a metric used to find the best terms.

4 F. Metabolic proteins participating in glycolysis, TCA cycle and ETC found in our
5 proteomes are illustrated. Proteins that were significantly enriched in MEC-specific
6 metabolic network appear in bold and are colour-coded to signify cell type.

7 **Figure 2: Mitochondrial structure and function varies with mammary lineage.**

8 A. FACS gating strategy to purify analogous mouse MEC populations.

9 B. Oxygen consumption rate, determined by Seahorse Bioanalyzer, of mouse MECs at
10 baseline, after exposure to Oligomycin (2 μ M), FCCP (1 μ M) and Antimycin A (1 μ M).

11 The left panel depicts the kinetic view of the data, which is quantified in the right panel
12 (n = 3 mice; 4 technical replicates per n). All data are mean \pm SEM. * P \leq 0.05; **
13 P \leq 0.01; *** P \leq 0.001; ****P \leq 0.0001.

14 C. Representative transmission electron micrographs of FACS-sorted mammary cell
15 pellets. Arrows indicate mitochondria. Magnifications are specified in each image.

16 D. Flow plots and quantification of median fluorescent intensity (MFI) for MitoTracker
17 Red (mitochondrial activity), MitoTracker Green (total mitochondria), MitoSOX
18 (mitochondrial ROS) and CellROX (total ROS). Each dot represents a biological
19 replicate (n=3-4 mice).

20 E. Heatmap showing unsupervised hierarchical clustering and z-scores of mitochondrial
21 protein abundance in mouse and human mammary proteomes with defined sex
22 hormone status and patient characteristics. MitoCarta⁴⁰, a curated list of

1 mitochondrial proteins, was used to filter our total MEC-specific proteomes. The
2 mouse proteome was obtained from a recent publication ²⁵.

3 F. Gating strategy used to sort basal and luminal cells with high and low mitochondrial
4 activity (left) that were subsequently cultured in the colony forming cell (CFC) assay.
5 Representative colony images and quantification of colonies formed from culturing
6 luminal and basal cells with either high or low mitochondrial activity in CFC assay.
7 Each dot represents a biological replicate (n=3 mice).

8 **Figure 3: Metabolic inhibitors expose lineage-restricted vulnerabilities of MECs.**

9 A. FACS gating strategy and pictorial summary of metabolic inhibitors and their
10 respective targets used to measure effects on mammary progenitor activity using the
11 CFC assay.

12 B. Dose-dependent effects of oxidative phosphorylation (OXPHOS) inhibitors on mouse
13 mammary CFCs. Colony counts were normalized to their respective basal or luminal
14 vehicle control;. Number of biological replicates per drug is shown in brackets.
15 Matched-pairwise analysis. Data are mean \pm SEM. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$;
16 **** $P \leq 0.0001$.

17 C. CFC enumeration after treatment after treatment with glycolysis inhibitors

18 D. Representative images of CFC plates after 6 days of culture, inhibitor concentration
19 and colony type are indicated.

20 **Figure 4: Breast cancer subtypes retain metabolic features of specific primary** 21 **MECs.**

22 A. Violin plots of single sample gene set enrichment analysis (ssGSEA) scores
23 comparing the metabolic network of basal, mature luminal and luminal progenitor

1 cells to the PAM50 subtypes of breast cancer. Each dot represents a patient from
2 the METABRIC study. One-way ANOVA in conjunction with a Tukey's test was
3 performed to determine the statistical significance of the differences in median
4 ssGSEA scores for different breast cancer subtypes, all compared to the subtype
5 with the highest median. *** $P \leq 0.001$; **** $P \leq 0.0001$.

6 B. Copy-number amplification frequency of *KMO*, *EPHX1*, *GALNT2*, *CYB5R1* and
7 *NIT* in the breast cancer patients from the METABRIC cohort ⁴³, grouped based
8 on PAM50 classifier, Estrogen receptor (ER) and Progesterone receptor (PR)
9 status. *KMO*, *EPHX1*, *GALNT2*, *CYB5R1* and *NIT* were found all highly abundant
10 in the mature luminal metabolic network.

11 C. Graphical abstract of the key findings in this study. Mammary epithelial cell types
12 have lineage-restricted metabolic identities, as found by metabolic protein
13 abundance, characterization of the mitochondria and drug effects on progenitor
14 capacity. This is visualized by the distinct coloured nodes in each normal MEC
15 population. The large amount of heterogeneity in breast cancer metabolism is
16 represented by the unique colours for the nodes in each subtype. Part of this
17 heterogeneity can be explained by the diverse cellular origin of breast cancer
18 subtypes, where they inherit metabolic features of their cell-of-origin, projected
19 here by the overlapping nodes with the same node colours as their primary MECs.

20 **Figure S1: Characterization of proteomic datasets of primary FACS-purified human**
21 **mammary epithelial cells.**

22 A. Gating strategy for FACS-purifying human mammary epithelial cells. Total cells
23 from dissected human breast tissue are gated to exclude debris. Doublet, dead

1 cell and Lineage (Lin+) exclusion ensures sorting of single, live and non-immune
2 cells.

3 B. Heatmap shows unsupervised hierarchical clustering and abundance of a set of
4 known marker proteins well established for distinguishing mammary epithelial cell
5 types.

6 C. Venn diagram summarizing the distribution of the 6040 detected proteins among
7 mammary populations. The numbers in brackets are the total number of proteins
8 detected in that cell type.

9 D. Pathway analysis using Enrichr was performed on each decile for each MEC type.
10 The top 2 GO Biological Processes per each decile are summarized with its
11 associated adjusted p-value in brackets.

12 **Figure S2: Map of human mammary epithelial cell metabolism**

13 A. Metabolic map is adapted from a previously published template ³⁴. Proteins are
14 coloured-coded to denote which mammary cell-type specific metabolic network
15 they demonstrated their highest expression level (Black = not significant or not
16 detected, Light blue = luminal progenitors, Dark blue = mature luminal and red =
17 basal).

18 **Figure S3: Characterization of the mouse mammary mitochondria.**

19 A. Gating strategy for FACS-purifying mouse mammary epithelial cells. Total cells
20 from dissected mouse mammary gland are gated to exclude debris. Doublet, dead
21 cell and Lineage (Lin+) exclusion ensures single, live and non-immune cells are
22 analyzed.

1 B. Heatmap showing unsupervised hierarchical clustering and z-scores of only the
2 metabolic proteins, determined by a curated list ³¹, from our previously published
3 mouse mammary proteomic dataset ²⁵.

4 C. Representative transmission electron micrographs of FACS-sorted mammary cell
5 pellets. Arrows indicate mitochondria. Magnifications are specified in each image.

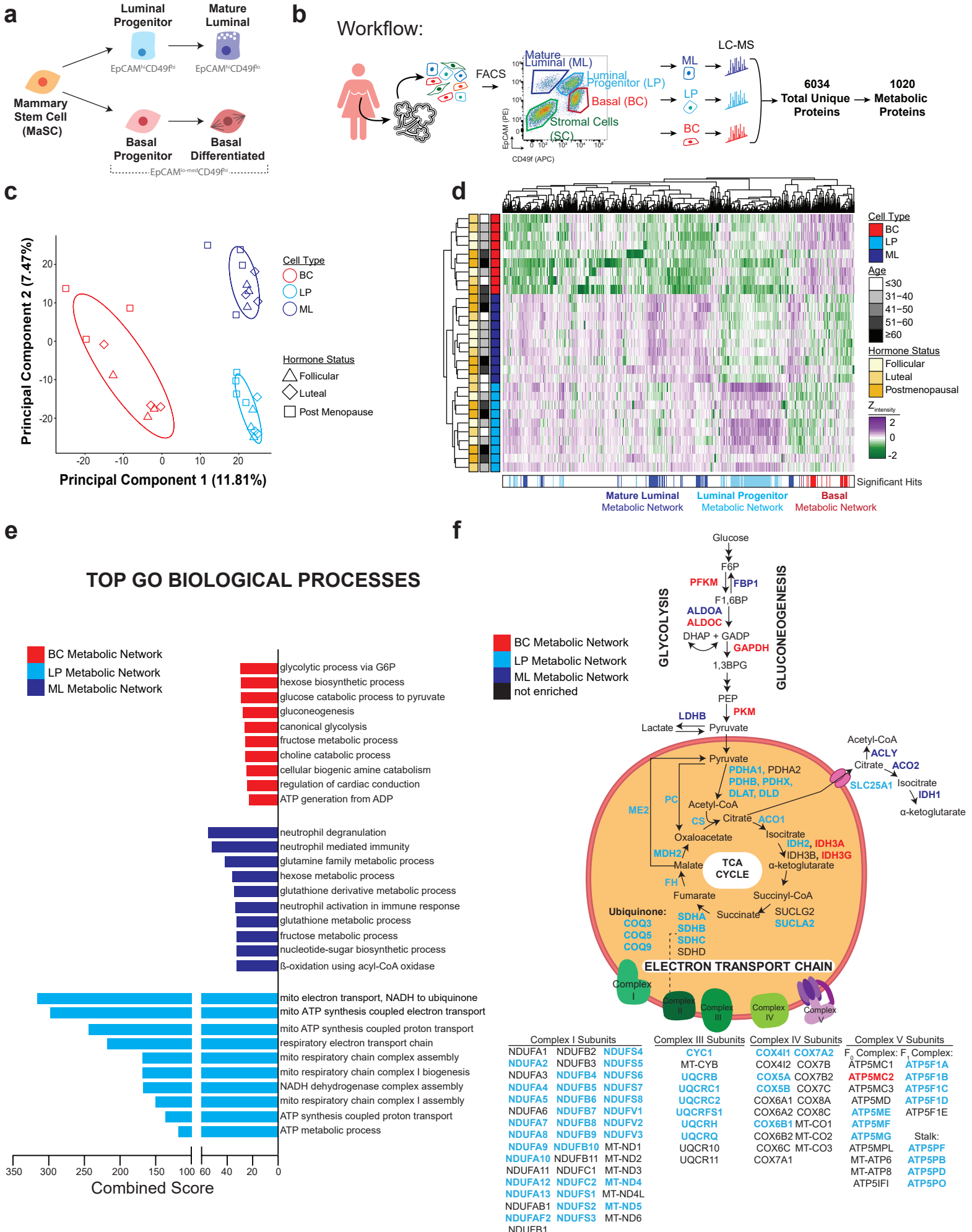
6 **Figure S4: Absolute CFCs of mammary lineages following treatment with metabolic**
7 **inhibitors.**

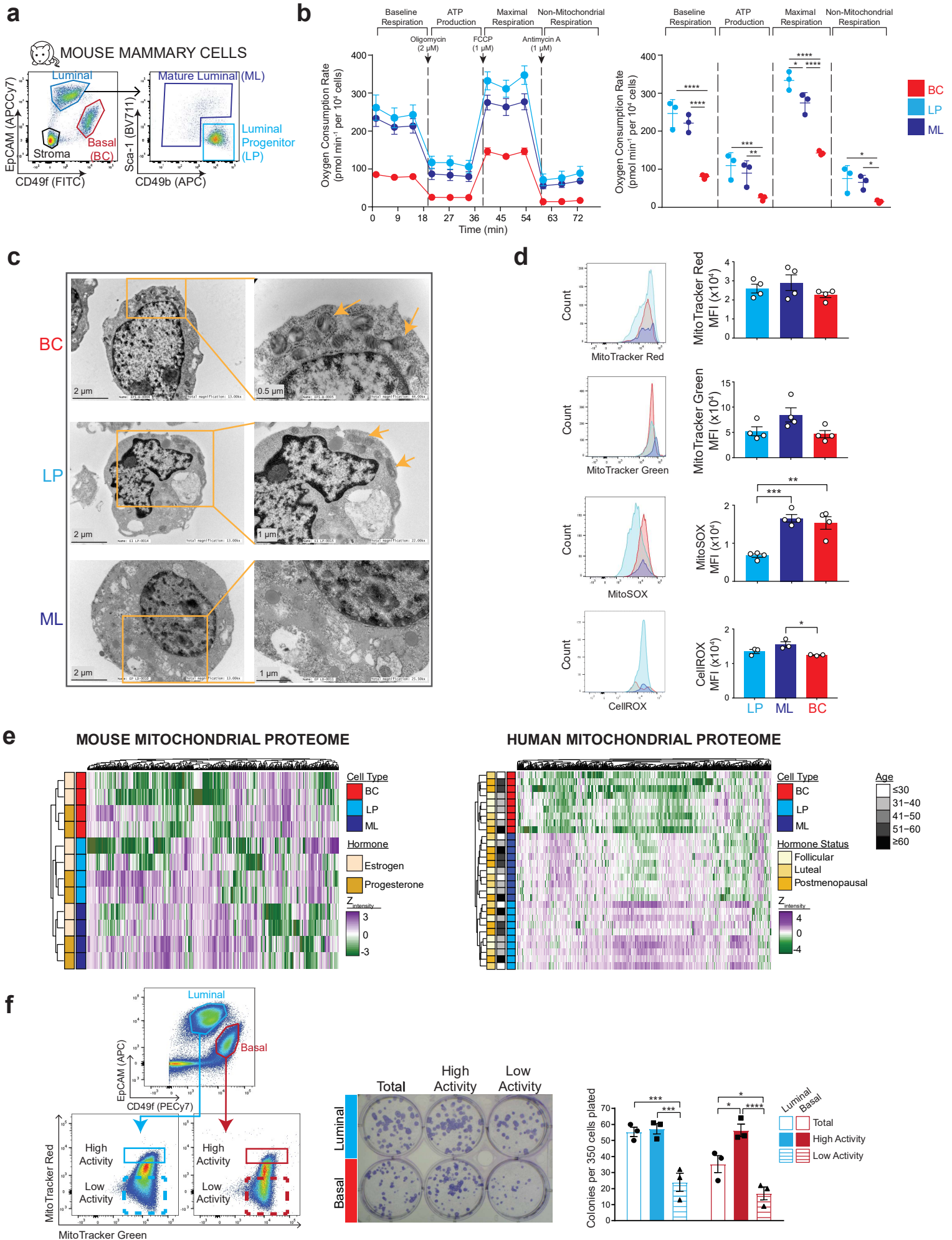
8 A. Quantification of absolute CFC counts at various concentrations of the specified
9 OXPHOS inhibitor. Basal colonies are red and luminal colonies are blue. Each dot
10 represents a mouse and number of biological replicates per drug is shown in
11 brackets. Data are mean \pm SEM. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

12 B. Quantification of absolute CFCs after treatment with glycolysis inhibitors.

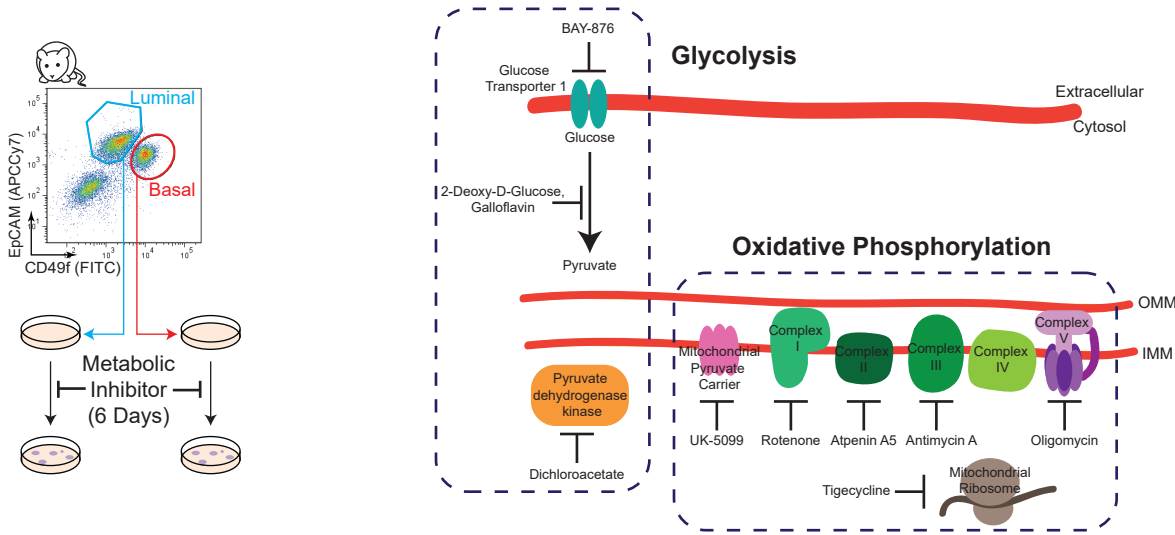
13 C. Representative images of CFC plates at the end of 6-day treatment. Inhibitors,
14 concentrations and the mammary cell types are indicated.

15

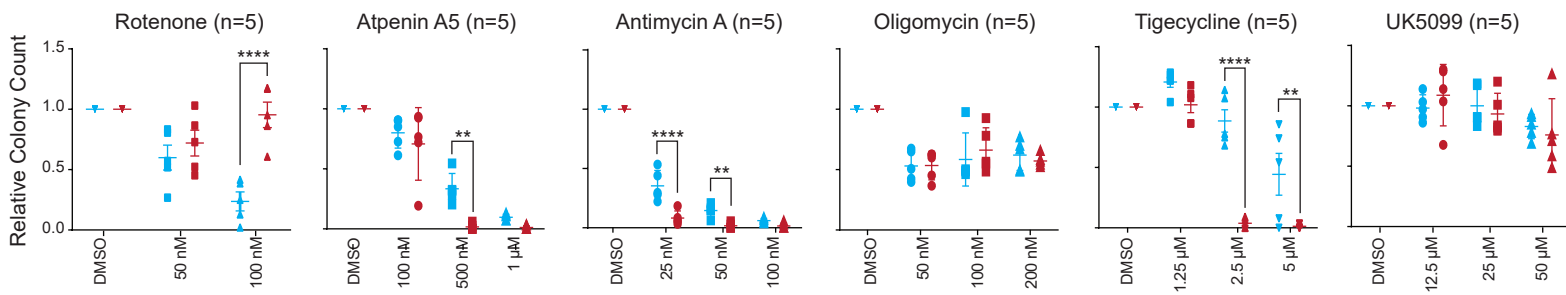




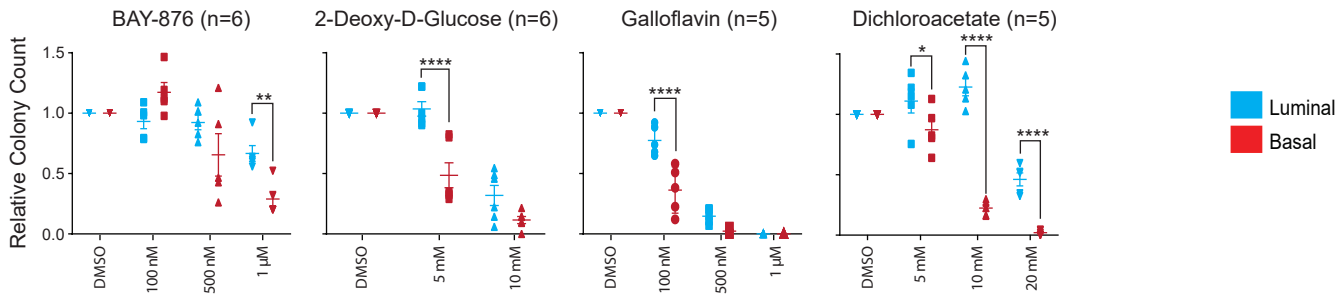
a



b



c



d

