A cellular and transcriptomic dissection of the human breast for studying mechanisms of cell and tissue function

Katelyn Del Toro[1,3], Rosalyn W. Sayaman[2,3,4], Kate Thi[2], Yamhilette Licon-Munoz[1], and William C. Hines[1]

[1]*Department of Biochemistry and Molecular Biology; University of New Mexico School of Medicine, 1 University of New Mexico MSC08 4670, Albuquerque NM 87131, USA*

[2]*Life Sciences Division, Lawrence Berkeley National Laboratory; Mailstop 977R225A, 1 Cyclotron Road; Berkeley, CA 94720, USA.*

[3]Equal contributions to work

Current affiliations:

[4]Present address: Department of Laboratory Medicine, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA.

Correspondence and contact for reagent and resource sharing should be addressed to Curt Hines (e-mail: wchines@salud.unm.edu), www.breastcancerlab.com.

Keywords: Breast, flow cytometry, FACS, RNA-sequencing (RNA-seq), human mammary epithelial cells (HMEC), Gene Set Enrichment Analysis (GSEA), Cytoscape, Enrichment Map, luminal epithelial cells, myoepithelial, vascular smooth muscle, pericytes, endothelial, lymphatic, fibroblasts, adipocyte-derived mesenchymal cells, adipocytes, leukocytes, erythrocytes.

Abstract

A fundamental question in biology, central to our understanding of cancer and other pathologies, is determining how different cell types coordinate to form and maintain tissues. Recognizing the distinct features and capabilities of the cells that compose these tissues is critical. Unfortunately, the complexity of tissues often hinders our ability to distinguish between neighboring cell types and, in turn, scrutinize their transcriptomes and generate reliable and tractable cell models for studying their inherently different biologies. In a companion article, we introduced a novel method that permits the identification and purification of the twelve cell types that compose the human breast—nearly all of which could be reliably propagated in the laboratory. Here, we explore the nature of these cell types. We sequence mRNAs from each purified population and investigate transcriptional patterns that reveal their distinguishing features. We describe the differentially expressed genes and enriched biological pathways that capture the essence of each cell type, and we highlight transcripts that display intriguing expression patterns. These data, analytic tools, and transcriptional analyses form a rich resource whose exploration provides remarkable insights into the inner workings of the cell types composing the breast, thus furthering our understanding of the rules governing normal cell and tissue function.

Introduction

Understanding the biology of tissues and ultimately developing effective cancer therapies requires us to recognize the symphony of interactions in tissues that occur among its many different cell types—and the myriad of molecules they synthesize. Determining how cells communicate with each other and their microenvironments is critical, as is defining the roles of stromal cells in regulating epithelial cell behavior, shaping tissues, and promoting or preventing overt cancer[1-4]. To this puzzle, there are many pieces. Clarifying and appreciating the roles of each cell type requires a substantial foundation in our ability to identify and distinguish among these distinct tissue elements so we may learn how they fit together to form a functioning organ. Tumors are caricatures of normal tissues[5], yet the intricate cellular composition and arrangement of cells in many tissues, including the breast, have yet to be adequately dissected, leaving many unanswered questions about these cells' unique and essential qualities.

The human dissections and inspiring observations made by Andreas Vesalius, William Harvey, and other Renaissance anatomist-physicians produced astonishing breakthroughs in medical science[6,7]. In a companion article and with a similar passion for anatomical dissection, we applied modern tools of histology and cell and molecular biology to examine the fabric of the human breast—to learn its cellular composition and study its mechanistic underpinnings. Using immunofluorescence and advanced flow cytometry, we developed methods to resolve, isolate, quantify, and culture each breast cell type, reconciling every cell in the tissue. The product of this labor was a fluorescence-activated cell sorting (FACS) strategy that used redundant markers and a rigorous gating strategy to identify the twelve distinct breast cell types (Figure 1a, Figure 1—figure supplement 1)]. These populations included two separate luminal epithelial cell types, myoepithelial cells, adipocytes, leukocytes, pericytes, vascular smooth muscle cells, erythrocytes, adipose-derived mesenchymal stem cells (fibroblasts), lymphatic and vascular endothelial cells, and a newly identified rare epithelial cell population with unknown function.

Although this FACS method resolved leukocytes as a single CD45^{Pos} fraction, the strategy is adaptable and can be further refined to scrutinize the array of leukocyte populations in the tissue. We now routinely sort and culture nine of the twelve breast cell types (excluded are red blood cells, leukocytes, and vascular smooth muscle cells). Some of these primary models are the first of their kind from this tissue, including breast pericytes, lymphatic endothelial cells, and our newly defined epithelial cell population. Their isolation and culture add to the repertoire of available models and permit functional studies that were not previously possible[8].

Here, we asked the critical question of what defines each breast cell type—and sought to identify their characteristic features and functions. Unsurprisingly, hierarchical clustering of transcriptional profiles echoed FACS results by cleanly separating the eleven sequenced cell types—revealing remarkable correlations among biological replicates (of each cell type, isolated from random individuals). Principal component analysis (PCA) established genes that heavily contribute to cellular diversity. We used PCA-derived gene lists to identify the distinguishing features of epithelial, mesenchymal, endothelial, and leukocyte cell lineages—each being a group of cell types with similar origins and characteristics. These included well-recognized features, such as keratinization of epithelial cells, muscle contraction by mesenchymal cells, vasculogenesis by endothelial cells, and more. While these results validated our analytic approach, they revealed many other processes we had not previously associated with a specific cell lineage. Shifting to a more gene-centric approach, we investigated what families of genes contributed to cell-type-specific functions by identifying the gene families that varied most across the spectrum of cell types. Many gene families showed little to no change in expression (e.g., exportins), whereas others were extensively differentially expressed (e.g., laminins, integrins, Rho GTPase activating proteins, and more). Differences in these genes' expression, and the biological processes in which they are involved, were further solidified—and others identified—via comparisons between cell-type pairs and application of gene set enrichment and pathway analyses to reveal the 'essence' of each cell type. We highlighted six comparisons that we believed to be of the highest interest as they reveal features of every major cell type and identify those distinguishing closely related cell types. Moreover, we have performed all 55 possible pairwise cell-type comparisons and provided the results in a user-friendly analysis tool (Pathway Analyzer[9]) and web application (Breast Cell Atlas: Transcript Finder[10]). These tools allow one to thoroughly engage with the dataset and actively explore each cell type's expressed genes and enriched biological pathways.

Results and Discussion

Identification of cell lineages and analytic pipeline

The breast is a compound tubuloalveolar gland whose subcutaneous tissue comprises, at a minimum, 12 major cell types. In an adjoining article, we describe the development and validation of an antibody panel and gating strategy capable of isolating these cell types[8], cell populations we numbered from one to twelve (which refers to their respective FACS gate using the isolation strategy outlined in Figure 1a and Figure 1—figure supplement 1). These cell populations are described as follows: ER^{Pos} (Pop1/Lum1) and ER^{Neg} (Pop2/Lum2) luminal epithelial cells, myoepithelial cells (Pop3), a rare epithelial cell population (Pop4), vascular

82 smooth muscle cells (vSMCs, Pop5), pericytes (Pop6), lymphatic endothelial cells (Pop7),
83 fibroblasts (Pop8), vascular endothelial cells (Pop9), $CD45^{Pos}$ leukocytes (Pop10), erythrocytes
84 (Pop11), and adipocytes (Pop12). These different cell types are often broadly classified as
85 members of epithelial, endothelial, mesenchymal, or hematopoietic lineages (Figure 2b).

86 In the early days of this study, we had not yet completed identifying all populations revealed by
87 FACS, and several remained a mystery even after finalizing our isolation strategy. Establishing
88 their identities was particularly challenging. We were aided by several helpful clues, however.
89 One was the distinct morphologies the cell types displayed when cultured. These observations
90 allowed us to determine whether a particular population appeared epithelial-like, e.g., Pop4
91 epithelial cells, or presented with more mesenchymal-like characteristics, e.g., Pop6 pericytes.
92 We also backtracked and stained tissues using the same antibody combinations used in FACS
93 and hunted for cell types in the tissue that shared identical staining patterns. Ultimately, the
94 transcriptome information from RNA-sequencing provided definitive evidence that solidified
95 each cell type's identity.

96 With their identities and histological locations now solved (Figure 2a-f, Figure 2—figure
97 supplement 1-2)[11], we shifted our attention to exploring the individual biologies of these
98 different cell types by diligently dissecting their transcriptomes through a comparative lens
99 provided by the comprehensive collection of cell types we had isolated and sequenced. Our
100 comprehensive analytic pipeline is described in Figure 3 (Figure 3, Figure 3—figure supplement
101 1).

102 mRNA sequencing and validation of FACS purified cell types

103 To investigate the unique biology of the breast cell types, we FACS isolated each population to
104 sequence and analyze their transcriptomes. This process entailed processing and collagenase-
105 treating breast tissues (reduction mammoplasty tissues free of proliferative disease), followed by
106 low speed (<100g) centrifugation. The first population collected was the lipid layer that had
107 phase-separated under centrifugal force (Pop12 adipocytes). All other cell types were FACS-
108 purified from the pelleted tissue 'organoids' following their trypsinization to single-cell
109 suspensions (Figure 1 and Figure 1- figure supplement 1). While developing and testing our
110 FACS antibody panel, we typically stained between 100,000 to 500,000 cells[8]. However, many
111 more were needed for mRNA library construction and sequencing here, so we scaled up our
112 operation by increasing the number of cells stained by roughly 1000-fold.

113 FACS purifying cells for RNA-sequencing was performed on freshly collected tissues. This
114 process required relatively large amounts of tissue (>30 grams) and sort times (14 to 18
115 continuous hours for each tissue) to produce cell yields that could support deep RNA-
116 sequencing. In addition to the above lipid layer, the sorting provided 11 FACS-purified cell
117 populations from each specimen. This exercise yielded hundreds of thousands to millions of cells
118 for the more abundant cell types (Figure 3-table supplement 1). In contrast, yields were much
119 lower for the rarer types, which produced a range of only 24,000-130,000 cells. After cell
120 purification, the RNA isolated from each sample correlated with cell counts, and in pilot
121 experiments, we found it essential to isolate RNA immediately following FACS to preserve its
122 integrity. In two independent sequencing runs, we sequenced 36 samples representing eleven of

the twelve identified breast cell types (red blood cells did not produce measurable RNA and thus were not sequenced). All cell populations were isolated from four random individuals (aged 16, 23, 27, and 29 years/old), and nearly all were sequenced in triplicate (from three of the four individuals, Figure 3—table supplement 2). In only two instances did we not have enough RNA of sufficient quality to sequence in triplicate –and both were rare populations with low FACS yields (Pop5 vascular smooth muscle cells and Pop7 lymphatic endothelial cells, Figure 1b, Figure 4a). We also isolated and sequenced RNAs from snap-frozen tissues and unsorted cell suspensions, which we used as mixed cell controls. After sequencing mRNAs from the samples and aligning the FASTQ sequencing data to the human genome, we calculated the average mRNA counts across all samples to be 39.5 million paired-end reads—all with Phred[12] quality scores of 25 or more, indicating the sequencing was of high depth and quality. We found these early sequencing results encouraging and moved forward with quality control analyses.

Nearly all sequenced samples demonstrated excellent read counts. However, covariance analysis, assessment of Cook's distances, and distribution of regularized log values[13] identified five samples with reduced transcript complexity and lower correlations. All were the rarer cell types, i.e., Pop4 epithelial cells, Pop5 vSMCs, and lymphatic endothelial cells (Figure 4b,c). Nevertheless, we determined that only Pop4 epithelial cells from individual N239 needed to be removed from critical downstream analyses. We retained the four other low-complexity samples because they clustered well with their cell-type counterparts (hierarchical clustering and PCA) and provided meaningful insights for many transcripts (Figure 4b, Figure 3—table supplement 2). These samples were: Pop4 epithelial cells (from individuals N225 and N228), Pop5 vSMCs (from N228), and Pop7 lymphatic endothelial cells (from individual N228). Although outlier genes existed in these samples, concordance between replicates was high for most others. Despite these minor issues, the sequencing data across the board was impressively robust. We did not detect sequencing artifacts, such as correlations with sequencing run or individuals from which the cells were derived, as all samples clustered tightly based on their inherent cell type (Figure 4b,d). The substantial number of reads detected per sample, the highly correlated replicates, and the large number of unique transcripts detected in these specimens (21,559) indicated that the dataset was rich and of excellent quality. Indeed, these data were immediately helpful in determining the identities of several obscure cell populations.

Before we had sequenced the mRNAs from the cell types, there were several FACS populations whose identities were a mystery (as discussed in the adjoining article)[8]. Most notably were populations #4, 5, and 6, which we ultimately identified as a rare epithelial cell type, vascular smooth muscle cells, and pericytes. RNA-sequencing provided conclusive evidence that helped identify these cell types by revealing their relationships to other populations and supplying information on expressed markers. After identifying these three populations, we explored all remaining cell types. Reassuringly, transcript levels of genes encoding cell surface markers used to sort each cell type tracked nicely with their respective FACS-staining levels (Figure 4—figure supplement 1a,b; Figure 1a). And GAPDH, often used as a uniformly-expressed internal control gene, was indeed expressed equally among the cell types, as were the transcripts of other widely-used internal control genes (Figure 4—figure supplement 1c, Figure 3—table supplement 2). In addition, analysis of hundreds of other genes, such as those uniquely expressed by each cell type

165 or those used as cell type markers, i.e., genes encoding CD antigens, validated and helped us
166 build confidence in our dataset (Figure 5, Figure 5—figure supplement 1). To facilitate the
167 exploration of these RNA-sequencing data, we have created a publicly available tool for
168 visualizing transcript levels in each cell type (http://breastcancerlab.com/breast-atlas/).

169 Identification of cell-type specific transcriptional patterns

170 One of our early goals was to identify the distinct transcriptional patterns of the different breast
171 cell types. To address this question and further explore the dataset, we clustered the samples'
172 transcriptomes using a standard Euclidean distance metric. Hierarchical clustering recapitulated
173 the distinct nature of the different cell types and mixed cell controls (Figure 6). Clustering was
174 robust, especially for the two luminal cell types, myoepithelial cells, pericytes, fibroblasts,
175 vascular endothelial cells, and adipocyte samples (all of which displayed approximately unbiased
176 (AU) *p*-values >0.95 when evaluated by multiscale bootstrap resampling[14]). Notably, biological
177 replicates of each cell type clustered tightly together, and similar cell types formed adjacent
178 clusters. These related pairs included: a) vascular and lymphatic endothelial cells (FACS Pops
179 7&9), b) vSMCs and pericytes (Pops5&6), c) ER^{Pos} and ER^{Neg} luminal epithelial cells (Pops
180 1&2), and d) myoepithelial cells and related Pop4 epithelial cells (Pops3&4). As expected,
181 fibroblasts (Pop8) clustered with the other mesenchymal cell types (vSMCs and pericytes)—
182 albeit at a greater distance— reflecting their inherently distinct features. In contrast, the most
183 distant samples that stood far from the others were the $CD45^{Pos}$ leukocytes (Pop10)—a
184 separation that we anticipated, given their unique biology. The clustering confirmed that the
185 eleven sorted cell types were indeed distinct and captured the underlying differences—and
186 similarities— in their gene expression patterns.

187 Several features emerged after closely examining the hierarchical clustering of genes. Gene
188 expression trends, in particular, were apparent across populations, and many clusters contained
189 genes encoding prominent cell markers, which helped us identify or confirm each cell type's
190 identity (Figure 6). For example, p-selectin (SELP) and HOXD1 were genes in a cluster that
191 differentiated endothelial cells. Epithelial cell adhesion molecule (EPCAM), zona occludens 3
192 (TJP3), and epidermal growth factor (EGF) were among the many genes in a cluster highly
193 expressed by luminal epithelial cells. Calmodulin-like 3 (CALML3) and G-protein-coupled
194 receptor 87 (GPR87) were highly expressed by Pop4 epithelial and myoepithelial cells, along
195 with the chloride channel regulator CLCA2, which is regulated by the myoepithelial lineage
196 marker p63 and is involved in basal cell adhesion. Finally, the perivascular cell cluster contained
197 angiopoietin 1 (ANGPT1) and connexin 40 (GJA5), which are essential for blood vessel
198 development and maintenance. The clustering thus demonstrated high agreement among the
199 biological replicates and identified clusters of shared genes that provided substantial evidence
200 confirming the identity of each cell population (or pointing to its identity).

201 The hierarchical clustering used to compare cell types was also notably stable. Whereas Figure 6
202 shows the clustering of the 2000 most differentially expressed genes, the pattern did not
203 significantly change when we repeated the analysis with different inclusive thresholds, such as
204 when we used the top 100, 200, 500, 1000, or 3000 genes that were most differentially
205 expressed. Although the cellular relationships that emerged from these analyses were largely

206 confirmatory, they helped us identify and characterize several populations early on. For example,
207 before we had ascertained the identity of FACS Pop5 as vSMCs, its close clustering with related
208 Pop6 pericytes here provided a critical clue[8]. The clustering also revealed similarities between
209 the rare Pop4 epithelial cells and myoepithelial cells. However, it also showed that Pop4's
210 expression of some genes more resembled those found in luminal epithelial cells that
211 distinguished this cell population from myoepithelial cells[8]. The molecular patterns that emerged
212 provided information on each population and objective evidence of the relatedness among cell
213 types. To explore these relationships—and features fundamental to each cell type—we needed to
214 decipher their transcriptomes in finer detail.

215 Distinguishing the features of cell lineage

216 Differentiated cells are often classified by their developmental history and the lineage from
217 which they are derived. The different breast cells, for example, are often described as being
218 mesenchymal (fibroblasts, pericytes, vSMC, and adipocytes), hematopoietic/blood (leukocytes,
219 erythrocytes), endothelial (vascular and lymphatic), or epithelial derived (ER^{Pos} and ER^{Neg}
220 luminal, myoepithelial, and Pop4 epithelial, Figure 2b). Although our primary goal was to
221 identify the qualities that make each cell type unique, our sequencing of every breast cell type
222 presented an unprecedented opportunity to explore the broad attributes of cell lineage through
223 the comparative lens afforded by our dataset's comprehensive nature. We accomplished this
224 through principal component and pathway analysis, as this approach allowed us to analyze the
225 samples en masse—and not focus on any particular pairwise comparison prematurely. As a
226 result, these analyses produced an astonishing amount of data. Unfortunately, we can present
227 only a fraction of it here. However, we have provided additional tools and tables to aid the
228 exploration of all data and results presented herein (Figure 3—table supplements, Pathway
229 Analyzer[9], and the Breast Cell Atlas: Transcript Finder App[10]). Below are our analyses of the
230 first four principal components that revealed the distinguishing features of the four broad cell
231 lineages composing breast tissues.

232 Principal component analysis (PCA) of the transcripts from all cell types showed gene
233 expression variances were distributed across multiple principal components (PCs), with the bulk
234 of the variation (71%) being captured by the first four components (Figure 7—figure supplement
235 1). The projection of principal component 1 (PC1) and principal component 2 (PC2) markedly
236 separated the four cell lineages in two-dimensional space. Genes contributing to PC1 were
237 primarily responsible for distinguishing the mesenchymal and epithelial cell types, as these
238 lineages projected at PC1's opposing ends (Figure 7a). Interestingly, endothelial cells—which
239 derive from mesenchyme but are classically regarded as a specialized epithelial cell type[15,16]—
240 projected as an independent group that aligned more closely with the mesenchymal cell types
241 along PC1. In contrast, the variance explained by genes composing the second principal
242 component (PC2) emerged from leukocytes, as these cells formed a single group at its distal end,
243 far apart from the other cell types (Figure 7a). A similar analysis of principal components 3 and 4
244 resolved the endothelial cells and two subtypes of the mesenchymal and epithelial lineages, i.e.,
245 perivascular cells and myoepithelial/Pop4 epithelial cells (Figure 8a).

To gain further insights into the distinguishing attributes shared by members of each cell lineage, we examined the genes impacting their PCA groupings, i.e., the loading factors with the highest influence on each principal component (Figure 3—table supplement 3). We then used these unique gene lists to probe functionally grouped Gene Ontology[17] and Reactome[18] pathway annotation networks[19].

Epithelial Lineage *(ER^{Pos} & ER^{Neg} Luminal cells, Myoepithelial cells, Pop4 Epithelial cells).* We first analyzed the cluster defined by the epithelial lineages: the ER^{Pos} and ER^{Neg} luminal myoepithelial and Pop4 epithelial cell types (Figure 2b). Based on their positioning on PC1, we expected the genes with the highest contributions to PC1 to reflect those that distinguish the epithelial and mesenchymal lineages, and in particular, highlight genes associated with luminal epithelial cells—as the ER^{Pos} and ER^{Neg} luminal cells projected the farthest along on PC1. Analysis of gene expression in each cell type showed this was indeed the case (Figure 7b). Many of PC1's top genes (at the negative epithelial end) were recognizable, as they are often used as markers of the epithelial lineage. They included keratins 6b, 8, 18, and 19, P- and E- cadherins, claudin 4, and Her2 (Figure 7b, Figure 9a-e; Figure 3—table supplement 3). Other familiar genes included integrin beta 6, the transcription factors ELF3, EHF, and TFAP2A; peptidase inhibitors SERPINA3 and SPINT2; serine protease ST14, cytoskeletal and associated proteins SHROOM3, MYO5D, and MYH14; the calcium-binding protein S100A14, among others (Figure 3—table supplement 3). These genes were among the 200 extracted from the negative epithelial end of PC1—a gene list we used to identify the biological pathways enriched within the epithelial cell lineage.

Using the epithelial gene set, we performed a two-sided hypergeometric (enrichment/depletion) test to determine if particular molecular functions or biological pathways were disproportionately represented (Gene Ontology[20], Reactome Pathway database[18]). We found many were, and we aggregated the results to form a functionally organized pathway network, which grouped similar biological processes based on their shared genes (ClueGO[21]). The 124 enriched gene sets aggregated into twenty-four distinct groups related to cell adhesion, peptidase activity, keratinization, estrogen signaling, AP-2 transcription factor regulation, mucins, and innate immune system factors (Figure 10a, Figure 3—table supplement 4: PC1 Epi tab). The most significant group (p-value $=1.4\text{x}10^{-8}$) contained the *cell adhesion molecule binding* and *cadherin binding* Gene Ontology terms. These gene sets represent different levels of the same ontology hierarchy and are highly correlated. For example, they both contain components of adherens junctions, i.e., epithelial- and placental-cadherins (CDH1 and CDH3), cingulin (CGN), and 348 other genes. By comparison, the *cell adhesion* gene set contains an additional 196 genes related to other aspects of cellular adhesion, such as desmosome-associated desmoplakin (DSP) and desmoglein 2 (DSG2, Figure 3—table supplement 3, sorted by PC1). The algorithm's ability to cluster overlapping gene sets like these, and assemble them into related clusters, allowed us to extract the common biological themes obscured within the volume and complexity of the data. Reassuringly, when we examined the expression of the genes in these pathways, we found they were indeed elevated in the four epithelial cell types (Figure 10b-e, (Figure 3—table supplement 4: PC1 Epi tab). These enriched pathways and themes describe general features of the epithelial lineage. After exploring the remaining cell lineages' shared attributes (defined by each PC), we will revisit these cells and explore their individual traits.

Leukocytes *($CD45^{Pos}$ Cells).* The second principal component separated the leukocytes ($CD45^{Pos}$ FACS population), and analysis of PC2's loading factors revealed many well-known leukocyte

markers (Figure 7a,c). For example, 'protein tyrosine phosphatase receptor type C' (PTPRC), the gene encoding the pan-leukocyte marker CD45—which we used to FACS isolate the leukocytes—was notably the first gene on the list. Other top genes included the CD antigens: CD2, CD3E, CD6, CD7, CD8A, and others. Also present was the transcription factor RUNX3, which is critical for T-cell differentiation, as was selectin P ligand (SELPLG), which is crucial for leukocyte trafficking. Using the same technique described above (for PC1 and the epithelial cells), we extracted the top genes from the leukocyte-defining end of PC2 and performed pathway enrichment analysis. We identified 453 enriched gene sets that clustered into 62 distinct thematic groups (Figure 3—table supplement 4: PC2 Leukocyte tab). Representative pathways from the ten most significant groups were: *adaptive immune response*, *activation of immune response*, *leukocyte homeostasis, T-cell activation, leukocyte chemotaxis, phagocytosis, IL-6 production, acute inflammatory response, regulation of GTPase activity, mast cell degranulation, and monocyte differentiation* (*p*-values ranged between 6.07×10^{-46} – 9.53×10^{-11}, Figure 10—figure supplement 1a). Although most leukocytes in the breast are macrophages[8], the sequenced mRNAs from the $CD45^{Pos}$ FACS gate represent a diverse population of white blood cells. Therefore, finding a broad range of enriched pathways attributable to different leukocyte subpopulations was expected. When we compared the expression of enriched genes in each cell type, we found they were indeed highly expressed by the Pop10 leukocytes (Figure 10—figure supplement 1b-d).

Mesenchymal Lineage *(vSMCs, Pericytes, Fibroblasts, Adipocytes).* Similar gene expression patterns within the four mesenchymal cell types positioned this cell group toward the positive end of PC1's axis in the 2D projection of PC1 and PC2 (Figure 7a). However, endothelial cells were also located in this region, indicating they and mesenchymal cells share the distinct expression of many genes. The relatively similar projection of endothelial and mesenchymal cells along PC1 prevented extraction of loading factors that could be solely attributable to either cell class. As we were primarily interested in identifying the defining characteristics of mesenchymal cells at this stage, we digitally removed the endothelial cell data and repeated the principal component analysis. The new PCA projection resolved the mesenchymal cells, situating them at the positive end of the first principal component in this ancillary graph ($PC1^{No\ ENDO}$, Figure 7—figure supplement 2a). Exploring the loading factors of this new principal component, we confirmed that these genes were indeed highly expressed by the mesenchymal cells. For example, vimentin (VIM), an intermediate filament protein and classical mesenchymal marker (Figure 9f), was near the top of the list, at position 192 (out of 21,560 genes, Figure 7—figure supplement 2b). The top loading factor was the non-selective receptor for endothelins (EDNRB). Endothelins are potent vasoactive peptides, and EDNRB expression was notably highest in the perivascular cell types (vSMC and pericytes) –these cells surround blood vessels and regulate vessel diameter by influencing vasoconstriction and dilation. Genes highly contributing to PC1 included many collagen family members (e.g., COL4A1, COL1A2, COL6A3, COL4A2, COL3A1, COL15A1, COL5A3, and others). Also present were ADAM metalloproteinases, the transcription factors ZEB1and ZEB2, platelet-derived growth factor receptor beta (PDGFRB), insulin-like growth factor II, laminin subunit alpha-4 (LAMA4), interleukin 24 (IL24), chondroitin sulfate proteoglycan 4 (CSPG4), along with many other notable genes (Figure 7— figure supplement 2b, Figure 3—table supplement 3: $PC1^{NoEndo}$). To

understand how these genes collectively contribute to mesenchymal cells, we performed pathway analysis using the loading factors from the mesenchymal-defining end of $PC1^{No\ ENDO}$.

Pathway enrichment using 200 of PC1's most critical mesenchymal genes identified 337 enriched gene sets from the Reactome and Gene Ontology databases (*p*-values< 0.05). These pathways assembled into a network of 96 distinct thematic groups (Figure 3—table supplement 4: PC1 Mesenchymal tab). The most significant group (*p*-value $=5x10^{-33}$) contained 94 different aggregated gene sets/pathways that dealt mainly with: cardiac development (chamber, muscle, septum), cartilage morphogenesis, smooth muscle proliferation, bone development, and chondrocyte differentiation—pathways that are consistent with our primary cultures showing the isolated breast fibroblasts have multi-lineage differentiation potential[8]. The defining genes in this group included TBX2 and HEYL transcription factors, the type III TGFβ receptor TGFBR3, fibroblast growth factor FGF2, ADAM metallopeptidase with thrombospondin type 1 motif 1a and 9 (ADAMTS1, ADAMTS9); platelet-derived growth factor receptor beta (PDGFRB), and others (Figure 11a). Representative pathways from the other top significant groups included: *Fibroblast migration, GAG metabolic process, chondrocyte development, bone development, smooth muscle contraction, negative regulation of calcium ion transport, regulation of fatty acid biosynthetic process, and dermatan sulfate biosynthesis,* with *p*-values ranging between $5.00x10^{-33}$ to $4.01x10^{-08}$ (Figure 11a, Figure 3—table supplement 4: PC1 Mesenchymal tab).

While the four mesenchymal cell types highly expressed genes in the enriched pathways, transcript levels were not always uniform across these cell types (Figure 11b-d). For example, in the *cardiac chamber development* gene set, expression of the pathway's genes was consistently higher in the perivascular cell types. The heart is composed of mesenchymal cardiac muscle cells and is a highly vascularized tissue that requires perivascular support, which may explain the frequent enrichment of cardiac-related pathways within the mesenchymal gene set (Figure 11—cardiac chamber development). The enrichment with these and other tissue-related gene sets (cartilage, bone, muscle, and adipose) indicates a deeply-rooted bond among mesenchymal cells and common dependence on a core set of genes. Later, we will use specific pairwise comparisons to explore the distinguishing features of the different mesenchymal cell types.

Endothelial Lineage *(Lymphatic and Vascular Endothelial Cells).* To identify the genes and processes responsible for shaping the unique characteristics of endothelial cells, we explored the third and fourth principal components. These components resolved the vascular and lymphatic endothelial cells as a distinct group near the positive end of PC3 (Figure 8a). Many genes contributing to the endothelial end of PC3 were familiar, such as vascular-endothelial cadherin (CDH5)—a prominent endothelial adhesion molecule (Figure 8b). We also observed von Willebrand factor (VWF); the angiopoietin regulator tie1 (TIE1); the endothelial marker CD31 (PECAM1); endothelial selectin (SELE); angiopoietin 2 (ANGPT2); as well as vascular endothelial growth factor receptors KDR and FLT4, and many other notable genes. The top gene on the list was 'podocalyxin like' (PODXL), which binds leukocyte selectin and is vital to lymphocyte-endothelial adhesion. Also present was Claudin 5, which is crucial for endothelial cell junctions. Finding so many genes with known importance to endothelial cells corroborated these genes' notable contributions to endothelial cells and showed they are characteristics of both vascular and lymphatic cell types. At the same time, the analysis identified other influential

genes that are perhaps less appreciated or whose importance to endothelial cell function are not as well characterized (listed in Figure 9, Figure 3—table supplement 3). To explore the biological contribution of these genes to endothelial cell biology, we performed pathway enrichment analysis.

Hypergeometric analysis of the endothelial genes from PC3 identified 650 enriched biological pathways (*p*-values< 0.05, Figure 3—table supplement 4: PC3 Endothelial tab). These pathways assembled into 61 groups based on the genes they shared. Not surprisingly, most groups pertained to blood vessel and circulatory system biology (Figure 12a). The most significant group (*p*-value $=6.27\text{x}10^{-19}$) involved cell motility and chemotaxis and contained *ameboidal-type cell migration*, *angiogenesis*, and *signaling by VEGF* gene sets (Figure 12a,b). As expected, PC3-identified genes within these enriched pathways were highly and differentially expressed by the lymphatic and endothelial cell types (Figure 12b-d). Representative pathways from these groups included: *cell-matrix adhesion, endothelial- and muscle- cell differentiation, lymph vessel morphogenesis, regulation of establishment of endothelial barrier, regulation of coagulation, vasculogenesis, and monocyte chemotaxis.* Many notable genes were present in these pathways, including podoplanin (PDPN) —a marker we used to sort lymphatic endothelial cells— ephrins A1 and B2 (EFNA1 and EFNB2), activin a receptor like type 1 (ACVRL1), platelet-derived growth factor subunit b (PDGFB), along with the aforementioned cadherin 5, claudin 5 and the VEGF receptors (KDR and FLT4). Finding vascular-related pathways enriched in the endothelial data set was not surprising, but identifying these pathways was reassuring, as it validated the approach and indicated these processes were occurring in both lymphatic and vascular endothelial cell types found in the breast.

Outside of the anticipated pathways, enrichment analysis of the endothelial genes also uncovered: *Signaling by SCF-KIT* (stem cell factor/c-kit)*, adherens junction organization, regulation of peptidyl-tyrosine phosphorylation* (JAK/STAT)*, semaphorin receptor binding, interleukin-6 secretion* (and IL-10, IL-12 cytokines)*, aminoglycan biosynthetic process, hyaluronic acid binding,* along with roughly 500 other enriched processes and pathways (*p*-values ranging between 1.64E-06 - 2.08E-02, Figure 3—table supplement 4: PC3 Endothelial tab). Enrichment of '*Signaling by SCF-KIT'* pathway was driven by elevated endothelial expression of the kit tyrosine kinase receptor (KIT), the GRB2-Related Adapter Protein (GRAP), and LYN Proto-Oncogene (LYN). Notably, Kit's ligand, stem cell factor (KITLG), was differentially expressed between endothelial cell types, as vascular endothelial cells expressed it at a much higher level (40-fold higher) than lymphatic endothelial cells (Figure 12—figure supplement 1).

Endothelial semaphorin expression was intriguing. Semaphorins are a large family of signaling ligands that exist as either secreted or membrane-associated proteins which provide movement and guidance cues[22]. The *Semaphorin receptor binding* pathway was enriched predominantly because of endothelial expression of semaphorin 3a, 3d and 6b (SEMA3A, SEMA3D, SEMA6B, *p*-value$=1.47\text{x}10^{-3}$; (Figure 12—figure supplement 2, Figure 3—table supplement 4: PC3 Endothelial tab). The class 3 semaphorins are the only semaphorins in humans that are secreted, and we found that both SEMA3A and SEMA3D were expressed rather exclusively by lymphatic endothelial cells (Figure 12—figure supplement 2a). Interestingly, the co-receptor for sema3

proteins, neuropilin1 and 2 (NRP1 and NRP2), which bind VEGF family members, were highly expressed by the endothelial cell types (*p*-values of 3.13×10^{-27} and 1.23×10^{-53}, respectively), with vascular endothelial cells (and pericytes) differentially expressing the highest levels of NRP1. In contrast, the highest levels of NRP2 occurred in lymphatic endothelial cells, which was 2-fold higher than in their vascular counterparts (Figure 12—figure supplement 2b,c). Others have implicated NRP2 as a critical factor in lymphangiogenesis[23,24], and our results are consistent with it fulfilling such a role for breast lymphatic cells.

The above analysis has identified many features shared between lymphatic and vascular endothelial cells. Later, we will directly compare these two endothelial cell types and explore their individual traits.

Perivascular sub-lineage *(vSMC and Pericytes).* After analyzing the endothelial cells, we were left with two other cell groups to explore that were resolved by PC3 and PC4. Perivascular cells were the first of these (Figure 8a). This mesenchymal cell subclass includes pericytes and vascular smooth muscle cells (vSMCs), and the group projected toward the negative end of the third principal component, opposite to the endothelial cells. Like fibroblasts and adipocytes, perivascular cells are mesenchymal in nature; but perform critical functions to the blood vessels they envelop. PC3's loading factors (at the perivascular end) included genes with known pericyte and vSMC significance (Figure 8c), such as smooth muscle actin (ACTA2, Figure 9g), smooth muscle calponin (CNN1), glutamyl aminopeptidase (ENPEP), regulator of G-protein signaling 5 (RGS5), and platelet-derived growth factor receptor beta (PDGFRB). At the top of the list was the atypical cadherin FAT1, along with phosphodiesterase 3A (PDE3A), peptidase inhibitor 15 (PI15), and carboxypeptidase E (CPE, Figure 8c). Both perivascular cell types differentially expressed AXL receptor tyrosine kinase, which was sixth on the list and, in addition to its other physiological functions, serves as a receptor/co-receptor for many viruses, including Ebola, Lassa, Marburg[25], SARS-CoV-2[26], and Zika viruses[27] –which may help explain the vessel involvement often observed in infected individuals. The perivascular cells also notably and differentially expressed members of the Notch and Hedgehog signaling pathways, such as NOTCH3, HEYL, GLI2, and GLI3, suggesting these cell types fulfill fundamental roles during tissue morphogenesis and maintenance. As before, we used the top 200 genes from the defining principal component for pathway enrichment analysis.
Hypergeometric analysis of the perivascular genes identified 456 enriched processes and pathways from the Gene Ontology and Reactome databases (*p*-values< 0.05, Figure 3—table supplement 4: PC3 Perivascular tab). These aggregated into 24 distinct groups (Figure 13a). Many groups reflected vascular-related biology in line with known perivascular cell functions. Among these were *regulation of blood vessel diameter, regulation of systemic arterial blood pressure by hormone,* and *smooth muscle contraction.* Enriched pathways from other significant groups included: *Epithelial tube morphogenesis, blood vessel- and vascular- development, ECM proteoglycans, artery-* and *cardiac tissue- development, cell fate commitment, syndecan interactions,* and *mesenchymal cell differentiation* (Figure 13 b-e). These analyses identify characteristic processes of both perivascular cell types.

Myoepithelial sub-lineage *(Myoepithelial Cells and Pop4 Epithelial Cells).* The final cell group separated by PC3 and PC4 was an epithelial subclass comprised of myoepithelial cells and

related Pop4 epithelial cells (a rare cell type we identified during FACS panel development[8]). The fourth principal component resolved these cells, and when we investigated PC4's loading factors, we found that the most significant gene was none other than keratin 5 (KRT5). Finding keratin 5 was reassuring because it has a long and storied history as a marker of basal epithelial cells[28-30]. Another top gene was alpha-6 integrin (ITGA6/CD49f), which we anticipated because CD49f formed the backbone of our FACS panel –and myoepithelial cells consistently exhibited the highest CD49f staining intensities (Figure 1a, Figure 9f,h). Among the many genes contributing significantly to PC4 (Figure 8d), included the well-known myoepithelial markers: keratins 14 (KRT14, Figure 8d) and 17 (KRT17), laminin subunit alpha-3 (LAMA3), the alpha-1 chain of collagen type XVII (COL17A1), smooth muscle actin (ACTA2, Figure 9g), p63 (TP63, Figure 9i), E-cadherin (CDH1, Figure 9c,h), P-cadherin (CDH3), the transcriptional regulatory factor TRIM29, and integrin subunit beta 4 (ITGB4, Figure 8d, Figure 3—table supplement 3). Many of these myoepithelial genes notably contribute to the molecular cluster defining the basal-like breast cancer subtype—which provided the rationale for its 'basal-like' designation[31].

As with the preceding cell lineages, we extracted the top 200 loading factors from the defining end of the principal component (PC4, in this case) and used the gene list to explore potentially enriched pathways and processes. Pathway enrichment identified 333 gene sets, which aggregated into 24 groups of related biological functions (Figure 3—table supplement 4: PC4 Myoepithelial (MEP) tab). Processes related to the basement membrane, extracellular matrix organization, and laminin interactions led the way, forming the most significantly enriched group (p-value $=3.46\times10^{-16}$, Figure 14a). These gene sets were dominated by integrins (ITGA3, ITGA6, ITGB4, ITGB6, ITGB8) and laminin chains (LAMA3, LAMB3, LAMC2), which were highly expressed by myoepithelial cells (Figure 14b-c). Notably, these three laminin genes encode the a3, b3, and g2 chains that form the heterotrimeric laminin 332 molecule—an essential epithelial basement-membrane-specific variant that reinforces the cellular adhesion to the basal lamina[32].

Representative pathways from the other significantly enriched groups included: *cell junction organization, keratinization and skin development, prostate glandular acinus development, morphogenesis of a branching epithelium, epithelial tube formation, actin filament bundle assembly, smooth muscle contraction, intermediate filament organization, regulation of Wnt signaling pathway, and others.* (Figure 14a-e, Figure 3—table supplement 4: PC4 Myoepithelial tab). Among the many integrin, laminin, keratin, and collagen genes represented in these enriched pathways were cadherins CDH1 and CDH3, desmosome-related desmocollin 3 (DSC3) and desmogleins DSG2 and DSG3, stratafin (SFN, 14-3-3 sigma); wnt-modulators 'secreted frizzled-related protein 1' (SFPR1) and 'Dickkopf wnt signaling pathway inhibitor 3' (DKK3); epidermal growth factor receptor (EGFR), fibroblast growth factor 1 (FGF1), and others (Figure 3—table supplement 4: PC4 Myoepithelial tab).

Myoepithelial cells have historically been the most studied breast cell type, likely because of their prevalence in the tissue and ease of culture. Our results are consistent with prior knowledge of myoepithelial function: that these cells are basally located, help shape the basement membrane to which they attach via hemidesmosomes and are rich with smooth muscle actin and myosin that provide them with contractile ability. Furthermore, the mammary gland is formed through invagination of the skin, and the myoepithelial gene list reflects this connection, as it contains

genes associated with epidermal development. This list includes basal keratins (KRT-14, -17, -5, -6b, -6c, and -7), the laminin 332 chains, desmosomal genes, vitamin D receptor (VDR), and transcription factors SOX9 and p63. These attributes distinguish myoepithelial cells from all others in the breast. Our results pinpoint the specific factors produced by myoepithelial cells, affirming their pivotal role in ECM-mediated cellular functions.

This section highlights the broad characteristics of cell lineages composing the breast. These characteristics emerged from pathway enrichment analyses using gene lists defined by the principal components aligning and separating each cell lineage. Hundreds of pathways were revealed. Some we anticipated, whereas many others were not. Although we have here highlighted the most statistically significant pathways emerging from this analysis (Figures 10-14), many more exist and are listed in the supplementary tables (Figure 3—table supplement 4). A full review of these results will undoubtedly provide a more comprehensive perspective.

Gene families contributing to cellular identity

Above, we explored the distinguishing qualities of different cell lineages by determining the pathways enriched in PCA-defined gene lists. These analyses relied on curated gene sets from the Gene Ontology (GO) and Reactome databases—but not every human gene is represented in these archives. For example, the Reactome knowledgebase currently contains 11,291 human proteins (including known isoforms) that, by our count, are encoded by just over 9,300 genes. Because we measured over 20,000 genes in each cell type, many are not captured in the above analysis. Therefore, we have complemented the prior results by shifting to a more 'gene-centric' approach. This method entailed identifying all differentially expressed (DE) genes in the breast cell dataset (DESeq2[13], adj. p-value ≤ 0.1, Figure 3—table supplement 5). We then used this information to identify gene families with the most variably expressed genes. Due to their differential expression, these gene types must contribute to cell-specific processes and the unique features of each cell type. Their identification and patterns of expression provided countless insights.

The HUGO Gene Nomenclature Committee has organized 25,320 human genes into 1,519 manually curated gene families (HGNC gene group index[33,34]). When we explored the breast cells' transcriptomes, we identified 1,421 gene families in our dataset (Figure 3—table supplement 6). These included *vacuolar-type ATPases*, *matrix metallopeptidases*, *mucins*, *collagens*, *glutathione transferases*, *erb-B2 tyrosine kinases*, *histones*, *interleukins*, and more (Figure 3— Table supplement 6). The sizes of the families varied widely, from 1 to 831 genes ($\bar{x} = 14, \mathrm{sd} = 42.2$) with a median family size of six genes per family (Fig.15—figure supplement 1).

To determine the families with the most variably expressed genes, we examined the expression of each gene, from each family, within every breast cell type—and focused on families that had six or more genes ($\geq$ median size). We discovered 187 gene families composed entirely of differentially expressed genes (Figure 15a). Twenty of these families remained once we excluded those containing fewer than six genes. These included *Rho GTPase activating proteins*, *fibulins*, *histocompatibility complex genes*, *laminin subunits, integrin beta subunits, and more* (Figure 15b, Figure 3—Table supplement 6). The cellular processes affected by genes in these twenty

541 families are wide-ranging, as they are involved in Rho signaling, cAMP regulation, interactions
542 with the immune system, and cellular adhesion—between other cells, the extracellular matrix,
543 and basement membranes. Below, we highlight nine families that we believe are most intriguing,
544 but we have also provided a full accounting of each gene family in a table supplement—
545 indicating the number and percentage of genes differentially expressed among them (Figure 3—
546 Table supplement 6)

547 ***Laminins.*** Each gene family told a story. The family of laminins, for example, is composed of 12
548 genes that code for laminin a, b, and g chains. These chains aggregate to form an array of
549 heterotrimeric laminin molecules with distinct properties that polymerize in the extracellular
550 matrix and help shape basement membranes. Above, we found that myoepithelial cells highly
551 expressed the chains composing laminin 332 (Figure 14). Inspection of the entire laminin family
552 revealed that fibroblasts highly express four laminins (LAMA1, LAMA2, LAMB1, and
553 LAMC3); endothelial cells express the highest levels of three (LAMA4, LAMB1, and LAMC1);
554 whereas ER^{Pos} luminal, vascular endothelial, and pericytes significantly expressed LAMA5
555 (Figure 15b- laminin subunits). There were also several intriguing instances where closely
556 related cell types expressed widely disparate levels of a particular laminin chain. Notable
557 examples were the differential expression of LAMA1 between ER^{Pos} and ER^{Neg} luminal cells,
558 LAMB1 between the two endothelial cell types, and LAMC3 between pericytes and vSMCs.
559 Furthermore, no cell type meaningfully expressed LAMB4, consistent with its consideration as a
560 pseudogene[35].

561 The relationship between cells and their basement membranes is dynamic and reciprocal. Each
562 cell type contributes to the basal lamina's structural diversity by secreting basement membrane
563 components like the laminins and other extracellular molecules. These integral basement
564 membrane proteins anchor the cells via membrane-tethered receptors and transmit signals to the
565 cells, where they become integrated and influence multiple cellular processes.

566 ***Fibulins.*** One such related and intriguing ECM-related gene family was the fibulins. Fibulins are
567 RGD-containing extracellular matrix glycoproteins that bind basement membrane proteins, such
568 as tropoelastin, fibrillin, and fibronectin[36]. Whereas we found that fibroblasts highly expressed
569 most fibulin family members, fibulin 5 (FBLN5) levels were highest in the ER^{Neg} luminal
570 epithelial cells (Figure 15b- fibulins). FBLN5 was also notably expressed by lymphatic
571 endothelial cells at substantially higher levels than their vascular counterparts. Fibulin 5 is one of
572 the smallest fibulins, is secreted, has unique calcium and integrin binding properties, and is
573 essential for elastic fiber formation[36]—and thus likely contributes to the elasticity of the
574 lymphatic endothelium. We also found fibulin 3 (EFEMP1) was also differentially expressed by
575 lymphatic endothelial cells, presumably providing similar contributions.

576 The fibulins and laminins are just two examples of the 187 gene families whose members were
577 entirely differentially expressed. Of the remaining families (having fewer than 100% DE genes),
578 we identified 211 containing between 75 to 99% DE genes (Figure 15a). Over half of these
579 families were composed of six or more genes. These 126 families included *integrin alpha*
580 *subunits*, *toll-like receptors, kruppel like factors, adenylate cyclases, low-density lipoprotein*

receptors, glypicans, interleukin receptors, ephrins, and others (Figure 3—Table supplement 6). One of the larger and more familiar families was the *collagens*.

Collagens. The family of collagens consists of 45 individual genes, and, as expected, we found mesenchymal and myoepithelial cell types highly expressed most of them. There were some notable exceptions, however. For example, the vascular endothelial cells highly expressed COL15A1 and COL9A3, whereas COL22A1 was expressed uniquely by luminal epithelial cells and adipocytes (Figure 15c). Particularly enlightening were the cell types expressing the collagen IV genes.

Collagen IV is a sheet forming collagen and the principal constituent of basement membranes. It provides cellular integrin and non-integrin attachment sites through its triple-helical and non-collagenous domains[37]. There are six collagen IV isoforms, which assemble to form one of three triple-helical heterotrimers, α1α1α2, α3α4α5, or α5α5α6[37,38]. We found the expression of collagen IV chains in the breast cells quite revealing, as the most robust levels occurred in cells associated with the basal lamina, but patterns of isoform expression differed by cell type. For example, pericytes and vSMCs, which are embedded within, and help maintain the vascular basement membrane, expressed the highest levels of COL4A1, COL4A2, and COL4A4 (Figure 15c, Figure 15—figure supplement 2a-c). In contrast, myoepithelial cells, which contribute significantly to epithelial basement membranes, exhibited the highest levels of COL4A5 and COL4A6. From these results, we would predict that the dominant collagen IV heterotrimer in epithelial basement membranes to be α5α5α6— whereas our data suggests it would be α1α1α2 within the basal lamina of the vasculature.

Integrins. The family of integrins was highly differentially expressed across the breast cell types. Integrins connect the extracellular matrix (ECM) to the cytoskeleton and act as receptors that integrate extracellular signals, relaying information to the cell's interior. Each integrin heterodimer forms from one α and one β subunit, from families of 18 α and 8 β isoforms, generating the 24 observed αβ combinations. When we investigated the family of integrins in the breast cell types, we found they were all differentially expressed—except for ITGAD, which was essentially absent in every cell type (Figure 15b- integrin beta subunits, Figure 15—figure supplement 2d). Due to their diversity and cell-surface expression, sixteen integrins serve as 'cluster of differentiation' (CD) antigens. Above, we introduced that a critical marker in our FACS was CD49f, which is integrin α6 (ITGA6) —and that myoepithelial and endothelial cell types highly expressed it (Figure 4—figure supplement 2a, Figure 14b). Integrin a6 binds laminins through heterodimers formed with either b1 or b4 subunits. The a6b1 and a6b4 integrins, along with a3b1 and a7b1, are the four main laminin-binding integrins. Thus, along with α6 integrin, the main laminin-binding integrin heterodimers use a3 and a7 integrins. Investigating these transcripts' expression in breast cells, we found that myoepithelial cells and ER^{Pos} luminal epithelial cells expressed the highest levels of ITGA3. In contrast, pericytes and vSMCs expressed the highest levels of ITGA7, and fairly specifically, which is consistent with prior reports of this subunit's expression in vascular smooth muscle cells[39] (Fig.15—figure supplement 2d-e). These data suggest that myoepithelial and perivascular cells use distinct integrin heterodimers to bind laminins within their respective basement membranes.

622 Another CD antigen is ITGAM (CD11b), which is frequently used to identify myeloid cells. We
623 indeed found that within the breast cell populations, ITGAM was highly expressed by
624 leukocytes—although it was also strongly expressed by adipocytes and luminal epithelial cells
625 (Fig.15—figure supplement 2d). Moreover, leukocytes also highly expressed six other integrins:
626 ITGAE, ITGAL, ITGAX, ITGA4, ITGB7, and ITGB2 (Fig.15—figure supplement 2d). ITGB7
627 and ITGB2 were exclusively expressed by leukocytes, consistent with their recognized role in
628 trafficking leukocytes to sites of injury and infection[40]. Our data reflect this cell-type specificity,
629 and transcript levels within the different breast cell types provide a detailed accounting of these
630 and other expressed integrin isoforms, offering valuable insights into potential dimer
631 combinations used by the different cell types.

632 Along with the collagens, alpha integrins, and 206 additional gene families, families containing
633 between 75 to 99% DE genes included the *plakins* (88% genes DE), *spectrins* (86%), and
634 *receptor tyrosine kinases* (80%). By comparison, we identified 209 families whose members
635 were expressed relatively equally across all breast cell types (0-24%), such as the *exportins* (0%,
636 Figure 15e). Furthermore, we also investigated families formed by a more moderate makeup of
637 DE genes, i.e., between 25 to 74% (Figure 3—table supplement 6). Among these 814 families
638 were the *semaphorins* (70%), *proteoglycans* (71%), *mucins* (71%), *prostaglandin receptors*
639 (67%), *nuclear hormone receptors* (66%), and *claudins* (64%, Figure 3—table supplement 6).
640 Also included were the *matrix metalloproteinases* (61% DE), which are a family of proteinases
641 that degrade all extracellular matrix components—and thus help regulate the homeostatic
642 balance of ECM molecules in the breast microenvironment (Figure 15d).

643 ***Matrix Metalloproteinases (MMPs).*** Of the 23 MMP isoforms encoded within the human
644 genome, we identified 14 differentially expressed among the breast cell types. Fibroblasts highly
645 expressed over half of these MMPs, which included gelatinase A (MMP2), stromelysin 1 and 2
646 (MMP3, MMP10), macrophage elastase (MMP12), membrane type 1,3, and 4 (MMP14,
647 MMP16, MMP17), and MMP19. Predominant substrates for these MMPs include fibronectin,
648 laminins, type III, IV, and V collagens, elastin, fibrin, pro-MMP2, pro-gelatinase A, aggrecan,
649 and cartilage oligomeric matrix protein[41]. In addition, we found leukocytes (and adipocytes)
650 expressed gelatinase B (MMP9) at its highest levels, consistent with reports of it being absent in
651 fibroblasts and playing an essential role in ECM proteolysis and leukocyte migration[42,43].
652 Arguably the most notable deviation from the near-universal fibroblast expression of MMPs was
653 the differential expression of matrilysin (MMP7) by ER^{Neg} luminal epithelial cells, where we
654 found levels to be 18-fold higher than in their ER^{Pos} counterparts –and 237-fold higher than in
655 fibroblasts (Figure 15d).

656 ***Tight Junctions.*** Essential to multicellular life are the mechanisms in which cells join to form
657 the fabric of tissues and guide their shape and functions. Above, we introduced the family of
658 integrins that serve as central receptors binding cells to the ECM. Different cell lineages have
659 distinct requirements. For example, breast fibroblasts are embedded within a collagen-rich
660 matrix, whereas perivascular cells, which are also mesenchymal in nature, encircle vessels, help
661 generate the endothelial basement membrane, and collaborate with endothelial cells to form the
662 vasculature. The endothelial cells, in turn, must create tight junctions to generate a patent and
663 sealed network for transporting liquid blood—especially important for forming the blood-brain

664 barrier of the central nervous system. Epithelial cells are also unique because their glandular role
665 requires them to develop functional sheets while maintaining a semi-permeable barrier to the
666 luminal space, where they may secrete and collect milk. The lumen is also essentially connected
667 to the outside environment through the ductal network and nipple opening, so epithelial cells
668 must also establish a barrier to protect against bacteria and other pathogens—working in concert
669 with adjacent leukocytes that closely associate with the epithelium. Thus, the formation of tight
670 junctions is essential to the gland.

671 Tight junctions are the prominent cell-to-cell junctions of epithelia and endothelia, establishing
672 the selectively permeable barrier between neighboring cells[44]. Claudin and occludin proteins are
673 vital components of this cell adhesion complex. Investigating their expression in the breast cells,
674 we found the vast majority of claudins (11 of the 19 expressed) were predominantly limited to
675 the two luminal epithelial cell types, although there were some notable exceptions (Figure
676 16a,b). One was the specific expression of claudins 5 and 14 (CLDN5, CLDN14) by endothelial
677 cells—and to a lesser extent, by adipocytes. Whereas the vascular and lymphatic endothelial
678 cells both highly expressed CLDN5, CLDN14-expression was limited to the vascular endothelial
679 cells. Another exception was myoepithelial CLDN11 expression. Also, of the eleven claudins
680 expressed by the luminal epithelial cells, five were differentially expressed between the two
681 luminal types, with CLDN9, 22, and 23 transcripts being more highly expressed by the ER^{Pos}
682 cells and CLDN16 and 10 being more highly expressed by the ER^{Neg} cells (Figure 16a,b).
683 Furthermore, claudin 1 was one of the six claudins expressed by the luminal epithelial cells, and
684 tissue staining confirmed luminal cell expression, as the staining was confined to the apical
685 surface of the paracellular space, circumscribing the luminal surface (Figure 16—figure
686 supplement 1). These data are consistent with the epithelial and endothelial use of tight junctions,
687 and results pinpoint the differential isoform specificity among the five claudin-expressing cell
688 types.

689 The other critical components of tight junctions are the occludins (tight junction proteins
690 encoded by TJP1, TJP2, and TJP3 genes). When we explored their expression among the breast
691 cell types, we found that TJP1 and TJP2 (encoding zona occludens -1 and -2) shared a similar
692 pattern, as they were both expressed relatively equally among all cell types. The only exceptions
693 were that for TJP1, leukocyte levels were roughly 50-fold lower than the other cell types, and for
694 TJP2, levels were lower in both fibroblasts and leukocytes. In contrast, TJP3 expression (ZO-3)
695 was predominantly limited to the luminal epithelial cells (Figure16c-e).

696 ***Gap Junctions.*** In contrast to the occluding tight junctions that seal the paracellular space
697 between neighboring cells, gap junctions create channels between them, allowing the exchange
698 of ions and low molecular weight metabolites between neighboring cells[45]. We found that the
699 breast cell types expressed eleven of the 20 genes in the gap junction family (Figure 17 a,b).
700 Myoepithelial cells expressed two of these genes (GJA1 and GJB5, i.e., connexin 43 and 31.1),
701 whereas luminal epithelial cells expressed six (GJB2, B3, B4, B5, C3, and D3), sharing only
702 with myoepithelial cells the expression of GJB5. Furthermore, ER^{Pos} and ER^{Neg} luminal cells
703 differentially expressed three gap junction genes, with GJB3 and GJB4 being higher in ER^{Pos}
704 cells—and GJB2 being highest in the ER^{Neg} luminal cells. Another intriguing feature was the
705 fibroblast and endothelial expression of GJA1, which they shared with Pop4 epithelial and

706 myoepithelial cells. We also found that leukocytes uniquely expressed the highest levels of
707 GJB7, although we are uncertain what role it would fulfill in these cells. One final observation
708 was the perivascular expression of GJA4, GJA5, and GJC1; and the differential expression of
709 GJA5 by vSMCs, at a 5.5-fold higher level than pericytes (Figure 17 a,b). From these data, we
710 can identify the potential homomeric and heteromeric gap junction complexes used by the
711 perivascular and other cell types.

712 Whereas gap junctions facilitate the direct connection of cytoplasm between cell types and are
713 essential mediators of cell-to-cell communication, other forms of communication occur through
714 autocrine, paracrine, and endocrine signaling mechanisms. Multicellular organisms—and
715 tissues— rely on efficient and responsive communication between different cells, sometimes
716 separated by great distances. One of the most evolutionary-protected mechanisms facilitating cell
717 communication is the production of proteins, peptides, fatty acids, glycans, cholesterol
718 derivatives, gasses, and other metabolites by one cell, which are detected by another and elicit a
719 response[46]. Unfortunately, the vast number of ligand and receptor families preclude us from
720 exploring each of them here. Instead, we chose to inspect two different receptor families that
721 exemplify the heterogeneous expression patterns observed among breast cell types: *receptor*
722 *tyrosine kinases* and *nuclear hormone receptors*. The most studied and best-understood family of
723 receptors possessing inherent kinase activity are the receptor tyrosine kinases (RTKs).

724 ***Receptor Tyrosine Kinases***. The human genome encodes 58 different receptor tyrosine kinases
725 (RTK) whose ligands span across twenty subfamilies. These ligands are wide-ranging and
726 include the ephrins, glial-cell derived neurotrophic factor, and: epidermal-, nerve-, hepatocyte-,
727 fibroblast-, vascular endothelial-, and insulin-like growth factors, among others. RTKs are vital
728 regulators of critical cellular processes and are frequently involved in many cancer types. Among
729 the breast cell types, most RTKs displayed a cell-type-specific expression pattern (Figure 18a-
730 m). Notably among these was the robust mesenchymal expression of platelet-derived growth
731 factors alpha and beta (PDGFRA, PDGFRB), fibroblast growth factor receptor 1 (FGFR1), AXL
732 tyrosine kinase receptor (AXL), and orphan receptor 2 (ROR2, Figure 18a-m). In contrast, most
733 cells equally expressed insulin receptor (INSR), insulin-like growth factor 1 receptor (IGF1R),
734 receptor-like tyrosine kinase (RYK), and lemur tyrosine kinase 2 (LMTK2). As expected, the
735 endothelial cells differentially expressed angiopoietin receptors tie2 and tie1 (TEK and TIE1)
736 and VEGF receptors 1 and 3 (FLT1 and FLT4), which are known to regulate many facets of
737 vascular biology (Figure 18a,i). We have previously described, consistent with prior reports[47-49],
738 how stem cell growth factor receptor (KIT) was highly expressed by ER^{Neg} luminal cells[8], but we
739 also found it to be appreciably expressed by vascular endothelial cells and vSMCs (Figure 18a,
740 Figure 12—figure supplement 1). Her2 (ERBB2) is amplified in a subset of breast cancers and
741 has been successfully targeted for therapeutic intervention (Herceptin/trastuzumab). We found
742 that Her2, along with two other members of the ErbB family (ERBB3 and ERBB4), were highly
743 expressed by luminal epithelial cells. In contrast, the remaining family member, epidermal
744 growth factor receptor (EGFR), was expressed across a broader range of cell types, with
745 fibroblasts and myoepithelial cells displaying the highest transcript levels (Figure 18j-m).
746 Overall, RTKs display a complex pattern of expression. Most, 80%, were expressed

747 differentially across breast cell types, sometimes even within the same lineage, e.g., PDGFRB,
748 KIT, ERBB3.

749 ***Nuclear Hormone Receptors***. Nuclear hormone receptors were another receptor family
750 demonstrating high cell type specificity. These receptors act as ligand-induced transcription
751 factors and are activated upon binding to small lipophilic molecules, such as thyroid or steroid
752 hormones, vitamin D3, retinoic acids (vitamin A), ecdysone, oxysterols, bile acids, fatty acids,
753 leukotrienes, and prostaglandins[50]. The human genome has 49 nuclear hormone receptors, and
754 their regulatory effects are wide-ranging. Of this family's many prominent members, the most
755 notable for the breast field is estrogen receptor (ERα).

756 Interestingly, much of what we know about nuclear hormone receptors is rooted in the early
757 studies of estrogen and how it regulates female reproductive organs, including the breast[51]. We
758 found 44 nuclear hormone receptors expressed in breast cells—31 differentially. The androgen,
759 progesterone, and estrogen α receptors were all strongly expressed by luminal epithelial cells—
760 more so in ER^{Pos} cells than their ER^{Neg} counterparts (Figure 19a-d). Both thyroid hormone
761 receptors were similarly expressed among the cell types, except perivascular cells, which
762 expressed 5.3-fold lower levels of THRA—and endothelial cells and leukocytes, which
763 expressed 8.5-fold lower levels of THRB (Figure 19e). The cell-type specific expression
764 extended to many other nuclear hormone receptors, such as vitamin D receptor (epithelial cell
765 types), estrogen-related receptor beta (ER^{Pos} luminal cells), hepatocyte nuclear factor 4 gamma
766 (both luminal cell types), estrogen receptor beta (Pop4 epithelial and vascular endothelial cells),
767 and nuclear receptor subfamily members –5A2 (both endothelial cell types), –4A1(mesenchymal
768 cells), –4A2 and –4A3 (perivascular cells, Figure 19f-m). Nuclear hormone receptors possess
769 broad influence over many cell processes. The cell-specific expression patterns detailed here
770 identify the cell types in which these receptors likely play a prominent role.

771 Analyzing gene families has provided us with an unconventional look at how different types of
772 genes contribute to the cell types in the breast, and the results identified the kinds of genes that
773 vary most across the spectrum of cell types. Above, we meticulously described nine gene
774 families whose functions extend from the nucleus (hormone receptors) to the exterior cell
775 microenvironment (laminins). However, these represent only a limited percentage of families we
776 identified as having a high proportion of highly variable genes. Many more are listed in the
777 supplementary tables (Figure 15—Table supplement 1, Figure 3—Table supplement 6). To
778 further understand how these different genes collectively contribute to the individuality of each
779 breast cell type, we needed to identify and evaluate the cellular pathways in which these genes
780 are collectively involved—and the higher-order processes they bestow.

781 Defining the nature of each breast cell type

782 To acquire a broad perspective of each breast cell type's fundamental properties and individual
783 functions, we performed a wholesale assessment of the biological pathways and processes
784 disproportionately employed by these different cell types. Gene set enrichment analysis
785 (GSEA)[52-54] accounted for each of the 20,000+ expressed genes in these cells' transcriptomes.
786 The algorithm determined whether a particular pathway's genes were enriched within a
787 bidirectionally ranked list of genes constructed from two contrasted cell types. One advantage of

GSEA is that it doesn't require large transcriptional variations to pinpoint enriched pathways, as it can identify situations where many genes in a pathway differ in a small but coordinated fashion.

Several databases organize functionally related genes into distinct pathways. Here, we present our GSEA results that used gene sets defined by the Reactome Knowledgebase[18]. This database contains a manually curated list of 2,585 pathways cross-referenced to other scholarly resources, such as NCBI[55], ENSEMBL[56], KEGG[57], and GO[17,20]. Each Reactome pathway represents a collection of genes whose products coordinate to perform a specialized biological process. To determine which of these pathways are potentially active—or differentially enhanced—in a particular cell type relative to another, we analyzed every transcript, in each breast cell type, across every cell-cell comparison, which required 55 separate differential gene expression contrasts (Figure 20, Figure 3—Table supplement 7 [1v2]).). For each pairwise comparison, we calculated: a) the log fold change for each measured transcript, b) the *p*-value for the calculated difference (Wald test), and c) the Benjamini-Hochberg adjusted *p*-value. Using these values, we ranked all expressed genes bidirectionally, placing the most statistically significant genes at the extreme ends of the list, with one cell type represented at the top (most positive) —and the other at the bottom (most negative). After ranking the genes, we performed GSEA, which extracted pathways from the Reactome database, identified the position of each pathway's genes within our ranked list, and used a running sum algorithm to calculate a normalized enrichment score for each pathway. For each pairwise comparison, hundreds of enriched pathways were typically identified, each consisting of 15 to 200 genes.

To help us navigate the overwhelming amount of data generated from these combined analyses, we integrated our GSEA results and gene expression values into a single pathway analysis tool that we programmed within Microsoft Excel (Pathway Analyzer[9]). This facile tool efficiently summarizes results according to user inputs (e.g., cell types to contrast, FDR cutoffs to use, biological themes to investigate, specific pathways to examine, and genes to explore).

With the aid of the Pathway Analyzer tool, we found that the number of pathways enriched in each cell type varied widely, ranging between 7 to 462 ($\bar{x} = 194, \mathrm{sd} = 120$). On average, fibroblasts had the fewest enriched pathways across all comparisons ($\bar{x} = 85 \pm 59$), whereas those having the most were the vSMCs ($\bar{x} = 315 \pm 86$), pericytes ($\bar{x} = 272 \pm 121$), vascular endothelial cells ($\bar{x} = 287 \pm 126$), and lymphatic endothelial cells ($\bar{x} = 229 \pm 114$). When we compared cell types across different lineages, akin to the PCA analysis performed above, many of the same enriched pathways emerged, e.g., keratinization within the epithelial cell types, angiogenesis by endothelial cells, and more. Unfortunately, the magnitude and comprehensive nature of the GSEA results preclude us from presenting here every identified aspect. Instead, we will focus on the biology of closely related cell types since the broad features of different cell lineages were examined above. However, the results from all analyses are accessible, and one can thoroughly explore any of the 55 possible comparisons using the Pathway Analyzer tool. Below, we present the most thought-provoking findings from six cell-type comparisons; these are: a) the two luminal epithelial cell types, b) ER^{Neg} luminal epithelial cells and myoepithelial cells, c) ER^{Neg} luminal epithelial cells and vascular endothelial cells, d) the two endothelial cell types, e) fibroblasts and adipocytes, and finally f) fibroblasts and pericytes. Some enriched

830 pathways we anticipated, such as finding the enrichment of 'estrogen signaling' in ER^{Pos} luminal
831 cells. These results validated the analytic approach and helped us establish confidence in the
832 pathway enrichments.

833 ***Estrogen signaling in luminal epithelial cells: a validating example***

834 Luminal epithelial cells help shape the breast's glandular structure by forming the lumen of
835 mammary ducts. The breast's two major luminal cell types are commonly classified by their
836 divergent staining for estrogen receptor alpha (ERα), and we were able to isolate each of these
837 cell types based on their mutually exclusive staining for α6 integrin (ITGA6/CD49f; Figure1,
838 Figure 1—figure supplement 1)[8]. To determine the distinguishing characteristics of these two
839 cell types, we contrasted their expressed transcripts, exposing 1,301 differentially expressed
840 genes (BH/FDR adj. p-value < 0.1, Figure 21). ER^{Pos} luminal cells expressed higher levels of 546
841 of these, including progesterone receptor (PGR), epiregulin (EREG), fibronectin (FN1), and
842 FOXA1. In contrast, transcripts differentially expressed by the ER^{Neg} cells included epidermal
843 growth factor (EGF), matrix metallopeptidase 7 (MMP7), KIT, and 752 other genes (Figure 21,
844 Figure 3—table supplement 7 [1v2]). We ranked all expressed genes by their differential
845 expression (based on their fold-change and statistical significance) and performed pathway
846 enrichment analysis to identify the cellular functions impacted by these cell types' transcriptional
847 differences.

848 Gene set enrichment analysis (GSEA) assessed where each pathway's genes fell along our
849 ranked gene list, and the algorithm calculated a normalized enrichment score, nominal *p*-value,
850 and false discovery rate (FDR) for each Reactome pathway. The *estrogen-dependent gene*
851 *expression* set was among the top pathways enriched in the ER^{Pos} luminal cells (*p*-value <0.001,
852 *q*-value $=4.84 \times 10^{-4}$). Identifying this estrogen signaling pathway was validating and served as a
853 fine example of the enrichment algorithm (Pathway Analyzer [1v2]). Differentially expressed
854 genes contributing to this pathway's enrichment included progesterone receptor (PGR), FOXA1,
855 erb-b2 receptor tyrosine kinase 4 (ERBB4), GATA binding protein 3 (GATA3), estrogen
856 receptor a (ESR1), MYC, and others (Figure 22a). These genes were among the 65% in the
857 pathway that fell within the leading edge of the analysis (the leading-edge genes are the core
858 genes that contributed most to the GSEA enrichment score, Figure 22a). Furthermore, when we
859 analyzed transcripts from this estrogen-related pathway, we confirmed that the leading-edge
860 genes were indeed differentially expressed between the ER^{Pos} and ER^{Neg} luminal epithelial cells
861 (Figure 22b, Pathway Analyzer).

862 Apart from *estrogen-dependent gene expression,* we identified an additional 324 enriched
863 pathways between the two luminal cell types, 211 within the ER^{Pos} cells and 113 within the
864 ER^{Neg} cells (FDR ≤ 10%, Pathway Analyzer). We used a pathway enrichment and visualization
865 method to extract information rooted within these GSEA results (Enrichment Map[58]). This tool
866 analyzed the pathway's gene sets and, based on the proportion of genes they shared, formed a
867 complex network of interconnected pathways, placing redundant and biologically related
868 pathways close to each other. We connected pathways when they shared 50% or more genes, and
869 this caused distinct clusters, or 'biological themes,' to emerge in the network (Figure 23,
870 Supplemental Cytoscape file [Enrichment Map 1v2]). We circled these grouped pathways and

named their overarching theme using Reactome's hierarchy nomenclature as a guide. For example, the *estrogen-dependent gene expression* gene set clustered with the two other sets related to estrogen signaling: *estrogen-receptor mediated signaling* and *extra-nuclear estrogen signaling.* We encircled all three on the map and labeled this overall theme 'Estrogen Signaling.' (Figure 23). We adhered to this strategy for each theme in the network and often arranged related themes together. For example, we placed the enriched 'FOXO mediated transcription' theme adjacent to 'Estrogen Signaling' on the enrichment map because FOXO3 is an established ERα regulator [59,60].

To evaluate the GSEA results and biological rationale for each pathway's enrichment, we studied the genes from each pathway—and their expression levels in the different breast cell types. These details often provided valuable insights (Pathway Analyzer). For example, high-scoring genes within the *ER mediated signaling* pathway included: HSBP1, which binds to ERα and promotes membrane localization; JUND, which interacts with ERα and occupies estrogen response element sites within the promoters of target genes; progesterone receptor, which is a direct downstream target of ERα; and others, like FOXA1, AKT1, epiregulin (EREG), amphiregulin (AREG), and SRC proto-oncogene. These genes were all highly expressed by ER^{Pos} luminal cells, and their transcript levels in these cells stood in stark contrast to those we observed within ER^{Neg} luminal cells –and the same was observed for FOXO mediated transcription (Figure 23—figure supplement 1a,b). In some cases, a pathway's genes were uniquely expressed (or nearly so) by one of the contrasted cell types. We indicated these genes on the enrichment map under each theme's name. For example, the genes encoding estrogen receptor a (ESR1), amphiregulin growth factor (AREG), transcriptional factors FOXA1 and MYB; tyrosine kinase receptor ErbB4, and the trefoil factor mucosal proteins TFF1 and TFF3 were all highly and differentially expressed by ER^{Pos} luminal cells (Figure 22b, Figure 23). Collectively, these data provided strong evidence that estrogen signaling is active in this cell type—as one would anticipate in cells expressing estrogen receptor.

Identifying the association between estrogen signaling pathways and ER^{Pos} cells was validating—as was later finding additional anticipated themes in other cell-type contrasts. Moreover, these correlations assured us that the pathway analysis had indeed identified processes used disproportionately between cell types. Below, we move on from the detailed example of estrogen signaling in these luminal cells and sift through the complexity and sheer volume of remaining results.

ER^{Pos} vs. ER^{Neg} luminal epithelial cells (Pop1 vs. Pop2)

When we compared transcripts expressed by ER^{Pos} and ER^{Neg} luminal cell types (FACS populations #1 and #2, Figure 1a), we identified 325 enriched pathways at a false discovery rate of 10%. The majority of these pathways were associated with cell signaling (83 pathways), metabolism (55 pathways), and immune-related functions (42 pathways). Others involved extracellular matrix organization, developmental biology, nervous system development, cell signaling, muscle contraction, and vesicle-mediated transport (Figure 23, Pathway Analyzer). Further exploration of expressed transcripts defined by each pathway showed that the GSEA algorithm functioned well at identifying significant, concordant differences in the gene levels

912 between cell types—even if individual transcript levels only modestly diverged. There were,
913 however, pathways that contained a significant share of highly differentially expressed genes—
914 some of these being principal players in the pathway. For example, fibronectin (FN1) and
915 TGFB1 are members of the *ECM proteoglycans* gene set and were expressed at over 35- and 10-
916 fold higher levels in ER^{Pos} luminal cells than their ER^{Neg} cell counterparts (Adj. *p*-value = $3.2x10^{-}$
917 ^{8} and $3.8x10^{-7}$, Figure 23, Figure 3—table supplement 7 [1v2]). Because implications of these
918 stark differences were often patently evident, we primarily focus on pathways that contain these,
919 but details of all enriched pathways can be thoroughly examined with the provided analysis tool
920 (Pathway Analyzer).

921 **Cell Signaling & Signal Transduction (1v2).** The PI3K/Akt, ErbB4, VEGF, and HGF/MET
922 pathways are all notable receptor tyrosine kinase pathways that we found enriched within ER^{Pos}
923 cells (Figure 23, Pathway Analyzer, Supplemental Cytoscape file [Enrichment Map 1v2]). Along
924 with p53, Phosphatidylinositol 3-kinase (PIK3CA) is the most mutated gene in breast tumors
925 (36% of tumors), and genes in the phosphoinositide 3-kinase (PI3K/AKT) pathway are the most
926 frequently altered in human cancers[61]. Across the span of normal breast cell types, we found
927 PIK3CA to be evenly expressed (Figure 23—figure supplement 1c). The PI3K/AKT pathway's
928 enrichment in ER^{Pos} luminal cells was instead primarily driven by the differential expression of
929 the serine-threonine kinase AKT1, the EGFR ligand epiregulin (EREG), and the receptor
930 tyrosine kinases: fibroblast growth factor receptor 1 (FGFR1) and platelet-derived growth factor
931 receptor beta (PDGFRB). Other genes contributing to the pathway's enrichment included the
932 EGFR receptors ERRB2, ERBB3, and ERBB4 (Figure 23—figure supplement 1c). These data
933 show that the ER^{Pos} luminal cells contain different signaling elements than ER^{Neg} cells and
934 possess the potential for stimulating themselves and neighboring cell types.

935 Another prominent member of the epidermal growth factor family of receptors is ERBB4. This
936 receptor dimerizes with ERα and is involved in many facets of downstream ER signaling[62]. Not
937 surprisingly, we found ER^{Pos} luminal cells enriched with the ERBB4 pathways: *signaling by*
938 *ERBB4* and *nuclear signaling by ERBB4* (FDR=3.9% and 2.9%, Figure 23). This enrichment
939 was driven by the cell's prominent expression of ERBB4 and its ligands betacellulin (BTC) and
940 epiregulin (EREG), although ER^{Neg} luminal cells differentially expressed several pathway
941 members as well, including neuregulins 1 & 2 (NRG1, NRG2). In addition, glial fibrillary acidic
942 protein (GFAP), another pathway member, was also expressed—nearly exclusively—by ER^{Neg}
943 luminal cells. In CNS astrocytes (cells that express high levels of GFAP), GFAP promoter
944 activity is inhibited when bound by an erbb4-containing complex[63]. This inhibition perhaps
945 explains the reduced GFAP levels in ER^{Pos} cells, as these cells have high ERBB4 levels (Figure
946 23—figure supplement 1d).

947 Another theme emerging from the analysis of the two luminal cell types was VEGF signaling, as
948 we found the pathways *signaling by VEGF* and *VEGFR2 mediated vascular permeability*
949 enriched within ER^{Pos} luminal cells (FDR=0.17% and 4.2%, respectively). This result is,
950 however, an excellent example of how comprehensive information on all the breast cell types
951 provided substantial perspective. Between the two luminal cell types, the VEGF pathways were
952 indeed enriched in the ER^{Pos} cells—predominantly due to their differential expression of the
953 serine/threonine kinase AKT1, amyloid beta precursor like protein 2 (APLP2), neuropilin 1

954 (NRP1), and other pathway members. However, when we explored these transcript levels across
955 all cell types, we found their expression in ER^{Pos} cells was not dramatically different from most
956 other breast cell types (Figure 23—figure supplement 1e). It thus became clear that the
957 pathway's enrichment was not a feature of its overexpression in ER^{Pos} cells but rather its relative
958 repression in ER^{Neg} luminal epithelial cells (Pathway Analyzer). Because of the importance of
959 these comparative insights, we similarly evaluated gene expression across all cell types for each
960 pathway, using heat maps and the Pathway Analyzer tool.

961 Another pathway significantly enriched in ER^{Pos} luminal cells was HGF/MET. Hepatocyte
962 growth factor (HGF) was first identified as a fibroblast-secreted factor that caused dramatic
963 epithelial cell migration[64]. Its cognate receptor is MET receptor tyrosine kinase[65]. We found
964 three HGF/MET signaling pathways enriched within ER^{Pos} luminal cells, including *signaling by*
965 *MET* (FDR=0.6%, Figure 23, Pathway Analyzer). Enrichment was mainly driven by the
966 differential expression of fibronectin, integrin alpha 3 (ITGA3), and several laminin subunits
967 (LAMA1, -A5, and -B2). These proteins are involved in focal adhesion signaling that acts
968 through focal adhesion kinase (PTK2), which was also modestly elevated in ER^{Pos} cells (Figure
969 23—figure supplement 1f). Interestingly, when we examined HGF expression across all cell
970 types, we found, consistent with its initial discovery, that it was indeed highly expressed by
971 breast fibroblasts, although pericytes and adipocytes also expressed it. HGF's receptor, MET,
972 was expressed by vascular endothelial cells, adipocytes, and the breast's four epithelial cell types
973 (Figure 22—figure supplement 1f).

974 **Metabolism (1v2).** To evaluate metabolism-related pathways, we coded a filter in the Pathway
975 Analyzer tool, allowing us to extract pathways involved in specific biological themes. At a 10%
976 FDR threshold, we identified fifty-five enriched pathways related to metabolism—thirty within
977 ER^{Pos} luminal cells and twenty-five within ER^{Neg} luminal cells. Those related to energy
978 metabolism were enriched in ER^{Pos} luminal cells (Figure 23). Notable among these were:
979 *integration of energy metabolism* (FDR=0.550%), *glycolysis* (FDR=7.94%), *regulation of insulin*
980 *secretion* (FDR=5.51%)*, PKA activation in glucagon signaling* (FDR=4.87%)*,* and
981 *glucuronidation* (FDR=3.25%, Pathway Analyzer). Genes contributing to these pathways'
982 enrichment included: the inositol 1,4,5-trisphosphate receptor (ITPR1), which is a PKA-activated
983 calcium channel that acts downstream of glucagon and cAMP signaling; fatty acid synthase
984 (FASN), which catalyzes the synthesis of palmitic acid from acetyl-CoA and malonyl-CoA; and
985 cAMP-dependent protein kinase (PRKAR2B), which transduces cAMP signals (Figure 23—
986 figure supplement 1g). Elevated transcript levels of several glycolytic and regulatory enzymes
987 within ER^{Pos} cells also led to the enrichment of the core glycolysis pathway. These were:
988 hexokinase-1 (HK1), phosphofructokinase (PFKP), the bifunctional enzyme 6-phosphofructo-2-
989 Kinase/Fructose-2,6-Biphosphatase 1 (PFKFB1), enolase gamma (ENO2), and pyruvate kinase
990 M1/2 (PKM). These enzymes catalyze the first, last, and an intermediate step of glycolysis
991 (HK1, PKM, ENO2) –and regulate other key regulatory metabolites in the pathway
992 (PFKFB1),Figure 23, Figure 23—figure supplement 1h, Pathway Analyzer). Thus, significant
993 evidence suggests ER^{Pos} cells have increased bioenergetic demands. As such, it was not
994 surprising that we also found enrichment of pathways involved in lipid metabolism.

995 Metabolism of phospholipids, sphingolipids, and fatty acids were three themes that emerged
996 from the comparison of the two luminal cell types (Figure 23). The ER^{Pos} cells' disproportional
997 expression of many of these pathways' genes led to the enrichment *sphingolipid metabolism*
998 (FDR=0.04%), *phosphoinositol metabolism* (FDR=0.06%), *fatty acid metabolism* (FDR=9.53%),
999 and related pathways (Pathway Analyzer). Analysis of the leading-edge genes from these
1000 pathways demonstrated their elevated expression in the ER^{Pos} cells (Figure 23—figure
1001 supplement 1i-k). The most notable genes involved in phospholipid metabolism included the
1002 solute carriers SLC44A4 and SLC44A5. These membrane-embedded choline transporters
1003 provide a critical metabolic precursor for phosphatidylcholine synthesis. ER^{Pos} cells also highly
1004 expressed pleckstrin homology domain protein (PLEKHA6) and phosphatidylinositol 4,5-
1005 bisphosphate 5-phosphatase A (INPP5J). PLEKHA6 interacts with phosphatidylinositol 3,4,5-
1006 trisphosphate (PIP3), whereas INPP5J dephosphorylates PIP3. Fatty acid elongase (ELOVL5)
1007 and fatty acid synthase (FASN) are both involved in synthesizing fatty acids, were highly ranked,
1008 and contributed to the pathways' enrichment.

1009 The set of pathways involving protein metabolism covers the range of reactions that proteins
1010 encounter over their lifespan, from synthesis to degradation (including protein folding, post-
1011 translational modifications, and trafficking). We found both luminal cell types enriched with
1012 distinct protein metabolic pathways (Figure 23). In ER^{Pos} cells, two of the eleven significant
1013 pathways were *N-glycan antennae elongation in the medial trans golgi* (FDR=0.79%) and
1014 *sumoylation of intracellular receptors* (FDR=0.95%, Pathway Analyzer). Differentially
1015 expressed genes in the former pathway included carbohydrate (N-acetylgalactosamine 4-0)
1016 sulfotransferases (CHST8, CHST10), ST8 alpha-N-acetyl-neuraminide alpha-2,8-
1017 sialyltransferases (ST8SIA2, ST8SIA6), and mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-
1018 acetylglucosaminyltransferase isoenzymes (MGAT4A, MGAT4B). These enzymes are active in
1019 the medial Golgi, where N-glycan remodeling occurs. When we explored the second pathway,
1020 *sumoylation* of *intracellular receptors,* we found that the small ubiquitin-like modifiers
1021 (SUMO1, 2, 3) were surprisingly expressed evenly across cell types. The pathway's enrichment
1022 was instead driven by elevated expression of the retinoic acid, retinoid X, progesterone,
1023 androgen, and estrogen receptors, indicating the receptors that could be modified by sumoylation
1024 (Pathway Analyzer).

1025 The protein metabolism pathways enriched in ER^{Neg} cells included *amyloid fiber formation*
1026 (FDR=5.95%) and *O-linked glycosylation of mucins* (FDR=9.28%, Pathway Analyzer). We also
1027 observed that many, but not all, large ribosomal protein components were elevated in ER^{Neg}
1028 cells, leading to the enrichment of ribosome-involved pathways, *eukaryotic translation initiation*
1029 (FDR≤0.001%) and *eukaryotic translation elongation* (FDR≤0.001%). Historically, these genes
1030 have been considered invariant, but more recent evidence suggests otherwise, that levels may be
1031 linked to cell specification[66]—as we observed among the breast cell types here.

1032 **Extracellular Matrix Organization (1v2).** The extracellular matrix (ECM) is a dynamic
1033 structure consisting of a heterogeneous composition of scaffolding proteins, growth factors,
1034 proteoglycans, and remodeling enzymes. It undergoes persistent remodeling, primarily through
1035 reciprocal interactions with nearby cells in the immediate microenvironment. We found eleven
1036 pathways enriched between the two luminal cell types, with ER^{NEG} cells overexpressing genes

1037 related to *collagen synthesis* and assembly (FDR=1.58%), whereas the ER^{Pos} cells were enriched
1038 in degradative pathways *degradation of the extracellular matrix* and *collagen degradation* (FDR
1039 4.08% and 5.55%, Figure 23, Pathway Analyzer). Driving enrichment of the synthetic pathway
1040 in ER^{Neg} cells were collagens (COL4A2, 6A2, 6A1, 22A1), lysyl oxidases (LOX, LOXL4), and
1041 matrix metallopeptidases (MMP3, MMP7). In contrast, the degradative pathway in ER^{Pos} cells
1042 was driven by differential expression of tissue inhibitor of metalloproteinase 2 (TIMP2), calpains
1043 (CAPN8, CAPN2, CAPN13), ADAM metallopeptidase 10 (ADAM10), cathepsin L (CTSL),
1044 elastin (ELN), matrix metalloproteinases (MMP12, MMP13), and collagens (COL7A1, 12A1,
1045 4A5, 4A6). These data suggest that these two related—but distinct— luminal cell types have
1046 contrasting roles in the synthesis and degradation of specific ECM proteins (Figure 23—figure
1047 supplement 1 l,m).

1048 Additional enriched pathways in ER^{Pos} cell relating to the ECM included: *laminin interactions*
1049 (FDR=4.1%), *non-integrin ECM interactions* (FDR=0.18%)*, integrin cell-surface interactions*
1050 (FDR=0.47%)*, syndecan interactions* (FDR=6.47%)*,* and *ECM proteoglycans* (FDR=0.31%;
1051 Figure 23—figure supplement 1 n-r). Differentially expressed genes within these pathways
1052 included collagen IV isoforms A5 and A6 (COL4A5, COL4A6), laminin chains (LAMA5,
1053 LAMB2), fibronectin (FN1), transforming growth factor-beta -1 and -2 (TGFB1, TGFB2),
1054 platelet-derived growth factor beta (PDGFB), agrin (AGRN), synedecan 3 (SDC3), and integrins
1055 α5 and β3 (ITGA5, ITGB3), all of which had higher transcript levels in ER^{Pos} luminal cells
1056 compared to their ER^{Neg} counterparts.

1057 **Tight junctions (1v2).** Claudins are a family of genes we explored in the previous section,
1058 where we found, with few exceptions, that their transcription was mainly limited to endothelial
1059 and epithelial cells—predominantly the two luminal cell types (Figure 16a,b). Here, using
1060 GSEA, we identified significant differences between the ER^{Pos} and ER^{Neg} luminal cells (FDR =
1061 9.1%). Specifically, the ER^{Neg} cells expressed significantly higher levels of CLDN10, CLDN16,
1062 CLDN11, CLDN8, and CLDN1, which led to enrichment of the *tight junction interactions*
1063 pathway (Figure 23, Figure 16a). This was notably the only Reactome 'Cell-cell
1064 Communication' pathway enriched in either cell type (other pathways, for example, involve
1065 *adherens junctions*, *cell-ECM interactions*, and *hemidesmosome assembly*, etc.). Despite the
1066 elevated levels of five claudins in ER^{Neg} cells, CLDNs 1, 7, 3, and 12 were expressed at roughly
1067 the same level in both cell types, whereas ER^{Pos} cells had higher levels of CLDNs 23, 6, and 9.
1068 Thus, the enrichment indicated that the two luminal cell types possess unique tight junction
1069 components to support their individual physiologies. These tight junctions are critical to the
1070 cells' ability to form semi-permeable barriers, which helps retain fluid and protect against
1071 microorganisms in the outside environment –to which the ductal system is connected via the
1072 nipple opening.

1073 **Immune System (1v2).**

1074 Using GSEA, we identified forty-two enriched pathways related to immune function. The vast
1075 majority of these were in the ER^{Neg} luminal cells (33 pathways), which overlapped with the two
1076 innate immune system pathways previously identified by the PCA/overrepresentation analysis
1077 (Figure 10a). One of the larger themes was interleukin signaling, where we found the enrichment

of nine pathways related to: *interleukin-4 and interleukin-13 signaling, interleukin-10 signaling, interleukin-1 family signaling, interleukin-2 family signaling,* and *interleukin-12 family signaling* (FDR= 0.181%, 0.792%, 2.33%, 4.79%, 1.096%, respectively; Figure 23; Pathway Analyzer). Enrichment of these pathways was driven by the ER^{Neg} cell's differential expression of the interleukins 33, 1a, and 18 (IL33, IL1A, IL18), chemokine ligands 4 and 5 (CCL4, CCL5), tumor necrosis factor receptor TNFRSF1B, serum amyloid-A1 (SAA1), interleukin-2 receptor gamma (IL2RG), interleukin-1 receptor type II (IL1R2), interleukin-27 receptor alpha (IL27RA), colony-stimulated factor 2RB (CSF2RB), and lipocalin 2 (LCN2, Figure 23- figure supplement 1s,t; figure supplement 2a-c). We notably found breast macrophages concentrated near the epithelial structures and within the ducts and epithelial compartment[8], and thus were not surprised to find luminal cell enrichment of interleukin pathways that involve macrophage activation. This indicates that the ER^{Neg} cells are disproportionately responsible for regulating macrophage physiology –a crosstalk that is subverted in breast tumors[67].

Among the other enriched immune-related pathways, we discovered ER^{Neg} luminal cells expressed higher levels of genes involved in *toll-like receptor cascades* and *antimicrobial peptides* (FDR=8.208%, 4.1067%, respectively). The ER^{Neg} cells differentially expressed toll-like receptors 1 and 2 (TLR1 and TLR2), defensins B4A and B1 (DEFB4A and DEFB1), peptidase inhibitor 3 (PI3), and clusterin (CLU, Figure 23- figure supplement 2d,e). Furthermore, all ten toll-like receptors were highly expressed by luminal cells, with half being the highest in these cells compared to all other cell types (TLR1, TLR2, TLR2, TLR5, and TLR6, Figure 23—figure supplement 2d ; Figure 23—figure supplement 3a-f). Toll-like receptors play a vital role in the innate immune system by triggering an immune response to microbial agents. For example, TLR1 and TLR2 coordinate to mediate the response to bacterial lipoproteins. Although the toll-like receptors are well studied in the colonic epithelium, their contributions to the homeostasis of the breast epithelium are not as well characterized. Because luminal epithelial cells have the potential for elevated microorganism exposure, it is reasonable to expect TLR and antimicrobial peptide expression, as they are needed to provide innate immune protection.

Milk proteins (1v2). Although not part of the Reactome gene sets, we noticed several genes encoding prominent milk proteins were highly differentially expressed by ER^{Neg} luminal cells, such as lactotransferrin (LTF) and casein kappa (CSN3). Lactotransferrin is the major iron-binding protein in milk, whereas casein kappa is critical for stabilizing casein micelle formation in milk. To explore the cellular expression of other milk constituents, we created a list of genes from prior reports of human milk composition [68-70]. This list comprised 48 genes and included lactotransferrin, caseins, lactalbumin, and other milk proteins (Figure 23—figure supplement 4a). Most of these were expressed by at least one breast cell type (41/48), and nearly all were differentially expressed (37/41; BH/FDR adj. p-value < 0.05), with one of the two luminal cells most often being the cell type with the highest transcript level (25/37). After comparing expression between the luminal cell types, we found 12 genes whose expression significantly differed between the two (adj.*p*-val <0.05), with ER^{Neg} cells most often having the higher transcript levels (8/12). These eight genes were: lactotransferrin (LTF), secreted phosphoprotein 1 (SPP1), polymeric immunoglobulin receptor (PIGR), xanthine dehydrogenase (XDH), clusterin (CLU), complement component 3 (C3), haptoglobin (HP), and milk fat globule-EGF factor 8

1120 protein (MFGE8, Figure 23—figure supplement 4b-j). These results demonstrate that the luminal
1121 cells are primed for milk production despite being derived from the resting breast tissues of non-
1122 lactating women and, at rest, transcribe the genes encoding major milk proteins–with the highest
1123 levels often being in the ER^{Neg} luminal cells.

1124 ***ER^{Neg} luminal epithelial vs. Myoepithelial cells (Pop2 vs. Pop3)***

1125 Whereas the luminal cells shape the lumen of the gland's ductal structure, myoepithelial cells
1126 form the basal layer, anchoring the epithelium to the basal lamina (Figure 9 a,d,e,g,h). The ER^{Neg}
1127 luminal and myoepithelial cells (FACS populations #2 and #3) are two of the most abundant cell
1128 types in the breast (Figure 1); they share the same epithelial lineage and spatial proximity—and
1129 physically interact. Most clinical breast tumors have a luminal phenotype and presumably
1130 originate from luminal cells, with the ER^{Neg} luminal cells being the most-likely cell of origin for
1131 basal-like breast tumors[49,71], despite these tumors' basal-like features. For these reasons, we were
1132 interested in—and explored— the differences between ER^{Neg} luminal and myoepithelial cells,
1133 focusing on the biochemical and physiological processes distinguishing these two cell types.

1134 Analysis of transcripts expressed by the myoepithelial and ER^{Neg} luminal cells revealed 2,239
1135 differentially expressed genes (BH/FDR adj. *p*-value < 0.1, Figure 24, Figure 3—table
1136 supplement 7 [2v3]). ER^{Neg} luminal cells expressed the higher level of roughly half of these
1137 genes (51.6%), which included CD24, keratin 19, prominin 1 (PROM1), epidermal growth factor
1138 (EGF), and peptidase inhibitor 3 (PI3). Myoepithelial cells, in contrast, differentially expressed
1139 1,083 genes; among these were genes previously associated with this cell type, including smooth
1140 muscle actin (ACTA2), keratin 5 (KRT5), p63 (TP63), and oxytocin receptor (OXTR, Figure
1141 24). Using the *p*-values from this differential expression analysis, we created a ranked list of all
1142 expressed genes and used this list to perform pathway enrichment analysis. GSEA revealed
1143 enrichment of 426 Reactome pathways (FDR≤10%), which clustered into distinct biological
1144 themes (Figure 25). The most significant of these involved cell-cell communication, extracellular
1145 matrix organization, and metabolism (Pathway Analyzer).

1146 **Cell-cell communication (2v3).** Proper intercellular communication is essential for maintaining
1147 homeostatic balance and normal tissue function. When we contrasted the ER^{Neg} luminal and
1148 myoepithelial cells, we found enrichment of four cell-cell communication pathways: one in the
1149 ER^{Neg} luminal cells (tight junctions) and three in the myoepithelial cells (cell-junctions, cell
1150 communication, and ECM interactions). Given the previous identification of claudin expression
1151 in the luminal epithelial cells, it was not surprising to find the *tight junction interactions* pathway
1152 enriched in the ER^{Neg} luminal cells, as they—in comparison to myoepithelial cells—
1153 differentially expressed most epithelial claudins, such as CLDN1, 3, 4, 7, 8, 10, 12, and 16 (FDR
1154 =5.57%). The only significant exception was the myoepithelial expression of CLDN 11 (Figure
1155 16 a,b). These results show that luminal cells express significantly more claudins than
1156 myoepithelial cells and thus rely on the tight junction features that these different claudins
1157 provide.

1158 Although luminal cells differentially expressed claudins and were enriched with the *tight*
1159 *junction interactions* pathway, we found other cell communication pathways enriched in
1160 myoepithelial cells. These were the *cell-cell communication* pathway (FDR = 0.0919%) and its

underlying pathways: *cell junction organization* (FDR =0.0076%) and *cell-extracellular matrix interactions* (FDR =0.1682%). Notable genes shared among these three pathways include actinin and filamin proteins that promote the branching of actin filaments and anchor the filaments to proteins in the cell's membrane. Shared in the enriched pathways were also the cell adhesion molecule 3 (CADM3), keratins 5 and 14 (KRT5, KRT14), and cadherins 3, 4, and 11 (CDH3, CDH4, CDH11, Figure 25—figure supplement 1a). As we found above with the PCA-derived gene list for myoepithelial cells (myoepithelial and Pop4 epithelial cells), a general feature of myoepithelial cells is their expression of ECM and basement membrane components and formation of desmosomes involved in cell junction organization (Figure 14 a-c). Thus, we were not surprised to find myoepithelial enrichment of the *cell-extracellular matrix interactions* pathway because these cells form and help shape the epithelial basement membrane. For the same reason, we anticipated other pathways involved in extracellular matrix organization would be enriched within myoepithelial cells. We found this was indeed the case.

ECM organization (2v3). As the non-cellular component of tissues, the extracellular matrix is formed by a network of proteins, proteoglycans, growth factors, and minerals that provides structural support, elasticity, and tensile and compressive strength to the tissue[72]. The ECM and basement membrane composition is dynamic and is constantly remodeled by cells in their immediate vicinity. Notably, we found pathways associated with extracellular matrix organization as some of the most significantly enriched pathways in myoepithelial cells, as nine of the ten most statistically significant pathways in the myoepithelial cells involved ECM and basement membrane organization (Pathway analyzer). Myoepithelial cells were enriched in nearly every aspect of ECM organization, including *Elastic fiber formation* (FDR= 0.01%), *Laminin interactions* (FDR<0.001%), *non-integrin membrane ECM interactions* (FDR<0.001%), *ECM Proteoglycans* (FDR<0.001%); *activation of matrix metalloproteinases* (FDR=4.38%), *collagen formation* and *collagen degradation* pathways (FDR<0.001%). These cells differentially expressed many collagen and laminin encoding genes, including COL2A1, 4A5, 4A6, 8A2, 11A1, and 17A1; and laminin chains LAMA1, LAMA3, and LAMB1). Other genes contributing to these pathways' enrichment included elastin microfibril interfacer 1 (EMILIN1), metalloproteinase 3 (MMP3), tissue inhibitor of metalloproteinases 2 (TIMP2), actinin alpha 1(ACTN1), thrombospondin 1 (THBS1), fibromodulin (FMOD), heparan sulfate proteoglycan 2 (HSPG2), and syndecan 3 (SDC3), among others (Figure 25- figure supplement 1b-e). The myoepithelial cells—which reside on the basement membrane at the edge of the epithelial compartment— are thus a major contributor and remodeler of the basement membrane and ECM. This is a significant role that diverges from that of the luminal epithelial cells. However, the luminal cells were a major producer of matrilysin (MMP7) that, in addition to several ECM proteins, cleaves the milk protein casein, which we also found differentially expressed by ER^{Neg} luminal cells.

Muscle Contraction (2v3). A defining feature of myoepithelial cells has long been their expression of smooth muscle actin[73], which allows them to contract when stimulated by oxytocin, e.g., during lactation. It was thus reassuring to find *muscle contraction* (FDR = 0.0085%) and associated pathways enriched in myoepithelial cells, including *smooth muscle contraction, phase 0 rapid depolarization,* and *cardiac conduction*—the latter sharing similar

genes despite being a pathway in the heart (FDR= 0.0707%, 2.011%, 0.148% respectively, Figure 25, Pathway Analyzer). Myoepithelial cells highly expressed smooth muscle actins alpha-1 and gamma-2 (ACTA1, ACTG2), the voltage-gated sodium channel alpha subunit 3 (SCN3A), the calcium voltage-gated channel subunit L (CACNA1C), myosin heavy chain 11 (MYH11), myosin light chain kinase (MYLK), and leiomodin 1 (LMOD1, Figure 25- figure supplement 1f) that contribute to the contractile ability of this cell type.

Ion channel transport (2v3). Calcium is essential for myoepithelial contractility; Its intracellular availability is regulated by the transport of ions between the extracellular and intracellular spaces. This process was significantly enriched in myoepithelial cells, as we found *ion channel transport* and associated pathways enriched in myoepithelial cells (FDR=5.99%, Figure 25). The enrichment was driven predominantly by these cell's differential expression of anoctamin-1 (ANO1), chloride channel accessory-2 (CLCA2), and the intracellular calcium channels ryanodine receptors 1 and 3 (RYR1, RYR3, Figure 25- figure supplement 1g). Furthermore, myoepithelial cells differentially expressed the transient receptor potential (TRP) cation channels TRPV6, TRPC1, TRPC4, TRPM3, TRPC6, and TRPV2 (*TRP channels,* FDR=1.59%). TRP channels are calcium permeable and convert external stimuli into electrical or chemical signals[74] (Figure 25- figure supplement 1g). These pathways likely contribute to regulating myoepithelial cells' intracellular hyperpolarization/depolarization and contractility.

Cell signaling (2v3). We identified the enrichment of eighty-one cell signaling pathways between the ER^{Neg} and myoepithelial cells. Most of these (sixty-two) were enriched within the myoepithelial cells. The bulk of these pathways involved receptor tyrosine kinases (23/62), G-protein coupled receptors (13/62), Notch (9/62), Wnt (8/62), and Hedgehog pathways (6/62). Although the ER^{Neg} luminal cells differentially expressed KIT and were enriched with both *signaling by SCF Kit* (FDR=5.7%) and ERBB2 signaling pathways, the myoepithelial cells were enriched with twenty receptor tyrosine kinase pathways, most at a conservative false discovery rate of less than 1% (13/20, Pathway Analyzer). These pathways involved MET, neurotrophic tyrosine receptor kinases, platelet-derived growth factor, vascular endothelial growth factor, fibroblast growth factor, and insulin-like growth factor pathways (Figure 25). For example, the enrichment of the FGFR signaling pathways *signaling by FGFR, signaling by FGFR1,* and *signaling by FGFR2* (FDR= 0.328%, 0.822%, and 4.00%,) was driven by their differential expression of fibroblast growth factors 1, 2, and 7 (FGF1, FGF2, FGF7); FGF receptors 1, 2, and 4 (FGFR1, FGFR2, FGFR4); and the ras proto-oncogene (HRAS, Figure 25- figure supplement 1h).

The enrichment of Notch signaling pathways split between the luminal and myoepithelial cells. *Signaling by notch* and *signaling by notch4* were enriched in myoepithelial cells (FDR= 0.0268%, 0.059%), as these cells differentially expressed smooth muscle actin (ACTA2) and notch ligands, JAG1 and JAG2 (Figure 25- figure supplement 1i). They also had slightly elevated levels of many proteosome subunits that help compose this pathway. In contrast, ER^{Neg} luminal cells were enriched with Notch 1, 2, and 3 signaling pathways (FDR= 3.38%, 7.75%, 1.99%) due to their differential expression of the Notch receptors, NOTCH2 and NOTCH3; the downstream effector HEY2, delta/notch-like EGF repeat containing (DNER) and epidermal growth factor (EGF), and other pathway members (Figure 25- figure supplement 1j).

1245 Other myoepithelial enriched pathways involved G-protein coupled receptor-, Wnt-, and
1246 Hedgehog-signaling. Differentially expressed genes from the twelve enriched G-protein coupled
1247 receptor pathways included G-protein coupled receptor 176 (GPR176), dopamine receptor (DR),
1248 and adenylate cyclase 5 (ADCY5). Enrichment of the Wnt and Hedgehog signaling pathways
1249 was driven by elevated transcript levels of these pathways' downstream effectors, which
1250 included transcription factor 7 (TCF7), lymphoid enhancer-binding factor-1 (LEF1), GLI1, and
1251 GLI3 (Pathway Analyzer).

1252 **Metabolism (2v3).** Within the eighty-seven metabolism-related pathways enriched between
1253 luminal and myoepithelial cells, one of the larger emergent themes involved the metabolism of
1254 glycosaminoglycans (Figure 25). These pathways were enriched in myoepithelial cells and
1255 included *glycosaminoglycan metabolism, chondroitin sulfate dermatan sulfate metabolism,* and
1256 *heparan sulfate heparin HS-GAG metabolism* pathways (FDR= 3.76%, 5.50%, and 3.04%,
1257 Figure 25). Among the genes contributing to their enrichment included key enzymes in GAG
1258 metabolism, e.g., a) hyaluronan synthase 3 (HAS3), which catalyzes the addition of
1259 acetylglucosamine and glucuronic acid monosaccharides to hyaluronan polymers during
1260 hyaluronic acid synthesis; b) chondroitin-6 sulfotransferase-3 (CHST3), which catalyzes the
1261 sulfation of chondroitin; and c) the heparan sulfate proteoglycan glypican-3 (GPC3), an ECM
1262 glycoprotein that binds to, and modulates, several signaling pathways (Figure 25—figure
1263 supplement 1k).

1264 A characteristic feature of epithelial cells is their secretion and release of mucins from their
1265 apical surfaces[75]. Not surprisingly, we found the *O-linked glycosylation of mucins* pathway
1266 enriched in ER^{Neg} luminal epithelial cells (Figure 25). Luminal cells notably expressed mucin-1
1267 (MUC1), which is a marker in our FACS panel and was used to sort the luminal cells. We found
1268 many other mucins expressed by the luminal cells, including MUC15 and MUC16 (Figure 25-
1269 figure supplement 2 a-f). Mucins are highly glycosylated proteins and a component of milk,
1270 where they provide antimicrobial resistance in the GI tract of breastfed infants[76]. These mucins
1271 are derived from the luminal epithelial cells.

1272 The *metabolism of steroids* was another large theme in the network, which was enriched within
1273 ER^{Neg} luminal cells (FDR=0.998%, Figure 25, Pathway Analyzer). These cells differentially
1274 expressed hydroxysteroid dehydrogenases 17B2 and 17B4 (HSD17B2 and HSD17B4)—
1275 essential enzymes involved in the production of the androgen and estrogen sex hormones[77].
1276 Additionally, luminal cells highly expressed other steroid hormone regulators, HSD11B1 and
1277 steroid 5-alpha-reductase (SRD5A3), as well as the cytochrome p450s (CYP24A1 and
1278 CYP24B1), which function in Vitamin D metabolism (Figure 25- figure supplement 1l). We
1279 found these genes were highly expressed by both luminal cell types, indicating steroid
1280 metabolism is a general feature of the luminal lineage.

1281 The ER^{Neg} luminal cells were also enriched with other metabolic pathways, such as *fatty acid*
1282 *metabolism* (FDR=6.04%, Figure 25, Pathway Analyzer), as they differentially expressed acyl
1283 synthetase long-chain family members 1, 3, 5, and 6 (ACSL1, ACSL3, ACSL5, ACSL6), aldo-
1284 keto reductase family member C3 (AKR1C3), glutathione peroxidase 2 (GPX2), and carnitine O-
1285 acetyltransferase (CRAT, Figure 25- figure supplement 1m). The acyl synthetase enzymes

1286 catalyze long chain fatty acids into active fatty acyl-CoA esters, which is a critical step in beta-
1287 oxidation and lipid biosynthesis. The roughly 10-fold difference in acyl synthetase transcripts
1288 between luminal and myoepithelial cells indicates that luminal cells have different metabolic
1289 demands requiring the processing of long-chain fatty acids.

1290 In our analysis of the citric acid cycle pathway, we found most genes in this pathway were
1291 uniformly expressed across the entire spectrum of cell types. However, there were three notable
1292 exceptions. The first, which occurred in both luminal cell types, was their pronounced
1293 differential expression of both isoforms of isocitrate dehydrogenase (IDH1, IDH2), which
1294 operate in the mitochondria to regulate 2-hydroxyglutarate, alpha-ketoglutarate, and isocitrate
1295 (adj. p-values = 5.2e-5 and 2.1e-13). Another differentially expressed TCA cycle gene was
1296 pyruvate decarboxylase (PC), which was repressed in both endothelial cell types. Finally, we
1297 found phosphoenolpyruvate carboxykinase-1 (PCK1) expression was limited to pericytes and
1298 adipocytes (Figure 25- figure supplement 2f).

1299 ***ER^{Neg} luminal epithelial vs. Vascular endothelial cells (Pop2 vs. Pop9)***

1300 Histologists consider endothelial cells specialized epithelial cells because they exhibit classical
1301 epithelial features[16]. For example, vascular endothelial cells and epithelial cells both form the
1302 lumen of their respective structures (e.g., vessels, ducts), use tight junctions for cell-cell
1303 adhesion, and rest on a basal lamina. However, when we analyzed the breast cell transcriptomes
1304 by PCA (Figure 7a) we found that endothelial cells projected more closely to mesenchymal cell
1305 types than to the epithelial cells, which may reflect their mesodermal origin. We found this
1306 curious and decided to compare the features of vascular endothelial cells to the abundant ER^{Neg}
1307 luminal epithelial cell type (FACS populations #9 and #2).

1308 Comparing transcripts between vascular endothelial cells and ER^{Neg} luminal cells, we identified
1309 4,167 differentially expressed genes (Figure 26). Of those, 1,652 were differentially expressed by
1310 the endothelial cells, which included endothelial markers: platelet/endothelial cell adhesion
1311 molecule (PECAM1/CD31), CD93, TIE1, and TEK tyrosine kinase. In contrast, luminal cells
1312 differentially expressed the transcriptional regulator, SMARCD3, keratin 15 (KRT15), CD24,
1313 cadherin 1 (CDH1), and transcriptional repressor GATA binding-1 (TRPS1, Figure 26, Figure
1314 3—table supplement 7 [2v9]). We ranked all expressed genes by their degree of differential
1315 expression and used this list for pathway analysis. GSEA revealed five-hundred pathways
1316 enriched between ER^{Neg} luminal and vascular endothelial cells (FDR≤10%), which provided
1317 insights into the different biological pathways and systems used by each cell type (Pathway
1318 Analyzer).

1319 **Hemostasis (2v9).** One of the most significantly enriched pathways, *cell surface interactions at*
1320 *the vascular wall* (FDR<0.001%), was found in vascular endothelial cells—cells that indeed
1321 form the vascular wall (Figure 27). Endothelial cells differentially expressed selectins E, L, and
1322 P (SELE, SELL, SELP)—which mediate cell-to-cell adhesion; endothelial cell adhesion
1323 molecule (ESAM); junctional adhesion molecule 2 (JAM3); platelet/endothelial cell adhesion
1324 molecule (PECAM1); and the angiopoietin receptor TEK tyrosine kinase (TEK, Figure 27-
1325 figure supplement 1a). These gene products regulate leukocyte adhesion to vessel walls during
1326 an inflammatory response and facilitate leukocyte migration. In addition, endothelial cells

1327 differentially expressed von Willebrand factor (VWF)— a historical marker of endothelial cells,
1328 thrombomodulin (THBD), and coagulation factors F2R and F3 (Figure 27- figure supplement
1329 1b). These contributed to the enrichment of *formation of fibrin clot (clotting cascade,* FDR =
1330 3.04%). This is an essential step for blood clotting after injury, indicating that endothelial cells
1331 supply critical factors for this process, thus distinguishing them from other epithelial cell types.
1332 Notably, several of these pathway's genes are shared by mesenchymal perivascular cells (e.g.,
1333 ITGA5, ESAM, ANGPT); this shared expression contributes to the mesenchymal alignment of
1334 endothelial cells in the PCA projection (Figure 7a).

1335 **Keratinization (2v9).** Another divergent feature of endothelial and epithelial cells was the
1336 keratinization that occurred in ER^{Neg} luminal cells (*keratinization,* FDR<0.001%, Figure 27,
1337 Pathway Analyzer). Keratin filaments are heteropolymers of different Type I and II keratins that
1338 confer distinct biophysical properties to cells, connecting the cells' cytoskeleton to other cells
1339 and the basement membrane via desmosomes and hemidesmosomes[78]. We found ER^{Neg} luminal
1340 cells differentially expressed fifteen keratins, including keratins 7, 8, 19, 6b, 23, 81, 15, 16, 17,
1341 and 18. In addition, they also differentially expressed caspase-14 (CASP14)—which is involved
1342 in the cornification of epithelia— and desmoglein-1 (DSG1), a component of desmosomes
1343 (Figure 27- figure supplement 1c).

1344 **Cell Communication (2v9).** Among the pathways associated with cell communication, we
1345 found that ER^{Neg} luminal cells were, compared to the vascular endothelial cells, enriched with
1346 the broad *cell-cell communication* pathway (FDR= 0.954%) and others involving tight junctions
1347 (Figure 27- figure supplement 1d). These were *cell junction organization* (FDR=1.8%) and *tight*
1348 *junction interactions* pathways (FDR= 0.835%, Figure 27- figure supplement 1e). The
1349 enrichment resulted from ER^{Neg} luminal cells' differential expression of laminin beta 3
1350 (LAMB3), cadherins 1 and 3 (CDH1, CDH3), as well as the crumbs-cell-polarity complex
1351 component 3 (CRB3), which helps form tight junctions. Although the vascular endothelial cell
1352 types expressed claudins 5 and 14 rather exclusively, the ER^{Neg} luminal cells differentially
1353 expressed an overwhelming number of claudins (CLDN1, -3, -4, -7, -8, -10, -11, and -12),
1354 leading to the pathway's luminal cell enrichment (Figure 27- figure supplement 1e, Figure 16).
1355 Another cell-communication pathway is the *cell-extracellular matrix interactions* pathway,
1356 which we found enriched in vascular endothelial cells (FDR<0.001%). This pathway includes
1357 integrins that link ECM proteins to the cell's cytoskeleton –with the help of scaffolding and focal
1358 adhesion proteins. Genes in this pathway differentially expressed by the vascular endothelial
1359 cells included filamin C (FLNC)—which was notably repressed by both luminal cell types
1360 (compared to all others), integrin-linked kinase (ILK), and parvin beta (PARVB, Figure 27-
1361 figure supplement 1f, Pathway Analyzer [2v9]).

1362 **Metabolism (2v9).** We identified eighty-two enriched pathways that involve the metabolism of
1363 nucleic acids, lipids, proteins, and other metabolites. Some of the more statistically significant
1364 pathways within ER^{Neg} luminal cells included the *metabolism of vitamins and cofactors* (FDR=
1365 0.618%), *metabolism of water-soluble vitamins and cofactors* (FDR=0.71%), and *nicotinamide*
1366 *salvaging* (FDR=4.63%, Figure 27, Pathway Analyzer). Genes contributing to their enrichment
1367 included a) vanins 1 and 2 (VNN1 and VNN2), which are involved in vitamin B5 metabolism, b)
1368 transcobalamin 1 (TCN1)—a vitamin 12 binding protein, and c) pyridoxal kinase (PDXK),

which phosphorylates vitamin B6 (Figure 27- figure supplement 1g). We also found *sphingolipid metabolism* enriched in luminal cells (FDR= 4.37%), as they differentially expressed sphingosine-1-phosphate phosphatase 2 (SGPP2), sphingomyelin synthase 1 (SGMS1), and aldehyde dehydrogenase 3 family member B2 (ALDH3B2, Figure 27- figure supplement 1h). Additionally, *fatty acid metabolism* was enriched in ER^{Neg} luminal cells (FDR= 0.636). Enrichment resulted from the luminal cells' differential expression of epoxide hydrolase 2 (EPHX2), ELOVL fatty acid elongase 6 (ELOVL6), acyl-CoA synthetase long-chain family members 1, 3, and 6 (ACSL1, ACSL3, ACSL6); fatty acid-2 hydrolase (FA2H); and the mitochondrial isocitrate dehydrogenase IDH2 (Figure 27- figure supplement 1i). *Glycosaminoglycan metabolism* was another enriched pathway that contained substantial transcriptional differences (FDR=0.95%, Figure 27, Pathway Analyzer). Leading to its enrichment in vascular endothelial cells was their differential expression of carbohydrate sulfotransferase 1 (CHST1), versican (VCAN), biglycan (BGN), decorin (DCN), and heparan sulfate proteoglycan 2 (HSPG2/perlecan, Figure 27- figure supplement 1j). This is notable because endothelial surfaces and their surrounding ECM contain large quantities of heparan sulfate proteoglycans[79] that modulate vascular functions, such as sequestering growth factors (FGF, VEGF), regulating lipoprotein metabolism, and inhibiting clotting[80].

Immune system (2v9). Pathways associated with the immune system were enriched in both luminal and endothelial cells (Figure 27). Nine were enriched in luminal cells and 39 in endothelial cells. *Antimicrobial peptides* (FDR=0.058%), *defensins* (FDR=0.349%), and *beta-defensins* (FDR=1.19%) were enriched in luminal cells, which differentially expressed defensin beta 1 (DEFB1), S100 calcium-binding proteins A7A, A7, A8, and A9 (S100A7A, S100A7, S100A8, and S100A9), toll-like receptors 1 and 2 (TLR1 and TLR2), lipocalin 2 (LCN2), and bactericidal/permeability-increasing protein (BPI, Figure 27- figure supplement 1k). Luminal cells also differentially expressed protein tyrosine phosphatase 2 and 6 (PTPN1 and PTPN6), interferon-inducible guanylate binding protein 1, 2, 3, and 5 (GBP1, GBP2, GBP3, and GBP5), and interferon regulatory factor 6 (IRF6)—which define the enrichment of *interferon-gamma signaling* in luminal cells (Figure 27, Figure 27- figure supplement 1l).

Interferon signaling was enriched in vascular endothelial cells, which differentially expressed intercellular adhesion molecule 1 (ICAM1) and interferon alpha-inducible protein 27 (IFI27, FDR= 0.512%, Figure 27, Figure 27- figure supplement 1m, Pathway Analyzer). ICAM1 encodes a receptor for leukocyte adhesion protein LFA-1 (ITGAL) and is essential for leukocyte trans-endothelial migration[81]. Additionally*, toll-like receptor cascades* (FDR=1.36%) were enriched in vascular endothelial cells, that differentially expressed suppressor of cytokine signaling (SOCS1), interleukin receptor-associated kinase 2 (IRAK2), and TLR4 (Figure 27, Figure 27- figure supplement 1n). Two pathways associated with TLR cascades were enriched in luminal cells: *IRAK4 deficiency TLR2/4* and *regulation of TLR by endogenous ligand* (FDR=0.926% and 4.94%, respectively, Figure 27). Notably, we discovered that enrichment of many pathways associated immune system was due to the expression of proteasome subunits. For example, enrichment of *interferon signaling*, *TCR signaling, BCR signaling,* and *Interleukin-1 signaling* were all affected by overexpression of the proteosome subunits PSMD2, PSMB6, and PSMB5, among others (FDRs < 0.001%, Figure 27, Figure 27- figure supplement 1o-q).

1411 Proteosomes function in peptide processing for many pathways, including those involved in
1412 antigenic peptide presentation during immune responses. However, protein modification is
1413 required for many other pathways like cell cycle progression, apoptosis, and DNA repair.

1414 **Cell signaling (2v9).** As expected, *signaling by VEGF* and *VEGFR2 mediated vascular*
1415 *permeability* were enriched in vascular endothelial cells (FDR=0.005% and 6.81%, respectively),
1416 with differential expression of placental growth factor (PGF), neuropilin 1 and 2 (NRP1 and
1417 NRP2), nitric oxide synthase 3 (NOS3), and vascular endothelial growth factors B and C
1418 (VEGFB and VEGFC)—all of which are involved in VEGF-regulated angiogenesis (Figure 27,
1419 Figure 27- figure supplement 1r)

1420 Interestingly, *signaling by ERBB2* was enriched in endothelial cells (FDR=8.00%, Figure 27).
1421 The differentially expressed genes contributing to the enrichment of this pathway include FYN
1422 proto-oncogene (FYN), YES proto-oncogene (YES1), and heparin-binding EGF-like growth
1423 factor (HBEGF, Figure 27, figure supplement 1s). However, the luminal cells differentially
1424 expressed ERBB2, ERBB3, ERBB4, phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1),
1425 epidermal growth factor (EGF), and EGF receptor (EGFR), which define the enrichment of other
1426 pathways associated with ERBB2 signaling, like *PI3K events in ERBB2 signaling* and *ERBB2*
1427 *regulates cell motility* (Figure 27).

1428 One other surprising finding was, despite the higher levels of estrogen receptor α mRNA in
1429 ER^{Neg} luminal cells (so named because of their undetectable protein expression and lower
1430 mRNA levels compared to the ER^{Pos} luminal cells), we found *estrogen receptor-mediated*
1431 *signaling* and associated pathways enriched in vascular endothelial cells (FDR = 0.008%). After
1432 investigating other members of this pathway, we found vascular endothelial cells differentially
1433 expressed estrogen receptor β (ESR2), calmodulin-1 (CALM1), and the G protein subunits
1434 GNAI2, GNG2, GNG5, GNG11, GNB1, and GNB2, and endothelial nitric oxide synthetase
1435 (NOS3) that compose this pathway. (Figure 27, Figure 27, figure supplement 1t). This is
1436 noteworthy because endothelial cells respond to estrogen for nitric oxide and prostacyclin
1437 production[82]. Our data thus indicate that the bulk of estrogen's effects on breast endothelium
1438 likely results through ERβ signaling.

1439 ***Lymphatic endothelial vs. Vascular endothelial cells (Pop7 vs. Pop9)***

1440 Whereas the blood's vascular system circulates blood throughout the body, the lymphatic system
1441 collects and filters interstitial fluid, ultimately returning it to the bloodstream—a process that
1442 helps maintain fluid homeostasis. Much like capillaries of the blood system, lymphatic vessels
1443 are also formed by endothelial cells, although the two vascular architectures are distinct (Figure
1444 2 d,f). We found lymphatic and vascular endothelial cells expressed similar markers, like CD31
1445 (PECAM1) and von-Willebrand factor (VWF, Figure 8b) but diverge and expressed other genes
1446 that contribute to their individual functions. To explore the biological processes shaping these
1447 two endothelial cell types, we investigated their transcriptional differences.

1448 Lymphatic endothelial cells are a rare cell type in the breast. Because of this, acquiring enough
1449 cells required for mRNA-sequencing proved challenging. However, by scaling up the number of
1450 cells stained—and through sheer brute force (~15 hours of continuous FACS sorting for each
1451 sample), we succeeded in purifying a sufficient number of cells from two samples (24,713 cells

1452 from N228 and 29,111 cells from N239, Figure 3—table supplement 1). We detected roughly 21
1453 million and 43 million paired end reads, respectively from these purified lymphatic endothelial
1454 cells (N228 and N239). During quality control assessment of the RNA-sequencing data, we
1455 however found outliers and conflicts between the lymphatic cell replicates (Figure 4c).
1456 Nonetheless, the pair of lymphatic endothelial cell types clustered together well— in the
1457 hierarchical clustering dendrogram (Figure 6) and projection of principal components (Figure
1458 7a). We thus retained the lymphatic endothelial samples for further analyses, yet, we did expect
1459 to find discordance between replicates for some genes.

1460 When we compared transcript levels between lymphatic and vascular endothelial cell types, we
1461 identified 3,739 differentially expressed genes (Figure 28, Figure 3—table supplement 7 [7v9]).
1462 Lymphatic endothelial cells expressed the higher level of 1,462 of these, including podoplanin
1463 (PDPN), the marker we used to differentiate lymphatic cells in our FACS strategy (adj.*p*-value of
1464 1.14E-7). We also found other lymphatic markers differentially expressed, such as 'lymphatic
1465 vessel endothelial hyaluronan receptor1' (LYVE1), 'vascular endothelial growth factor receptor-
1466 3' (FLT4), and prospero homeobox 1 (PROX1, Figure 28). In contrast, the vascular endothelial
1467 cells differentially expressed 2,277 genes. Among these were the vascular endothelial growth
1468 factor C (VEGFC), cytokine-like 1 (CYTL1), lipocalin 1 (LCN1), endothelial selectin (SELE),
1469 and the endothelial marker CD93 (Figure 28). After ranking the expressed genes by degree of
1470 differential expression, we performed GSEA, and discovered 240 pathways enriched between the
1471 lymphatic and endothelial cell types (FDR≤10%, Pathway Analyzer). Of these enriched
1472 pathways, we've concentrated only on those established from genes where there was
1473 concordance between lymphatic endothelial cell replicates.

1474 **Immune system (7v9).** Both lymphatic and blood vascular systems are essential and active
1475 components of immune responses. Lymphatic vessels serve as conduits for lymphocyte
1476 migration from peripheral tissues to the lymph nodes, making them crucial to adaptive immune
1477 response. In contrast, vascular vessels are first responders to pathogen detection and danger-
1478 signal-sensing mechanisms in the bloodstream. Indeed, when we inspected the enrichment
1479 results, both cell types were enriched with immunity-related pathways. For instance, we found
1480 *interleukin-4 and interleukin-13 signaling* pathways enriched in lymphatic endothelial cells
1481 (FDR= 1.54%). Contributing to this enrichment was the lymphatic cells' differential expression
1482 of transforming growth factor-beta 1 (TGFB1) and IL13 receptors A1 and A2 (IL13RA1 and
1483 IL13RA2, Figure 29a). In contrast, the vascular endothelial cells were enriched with *the TCR*
1484 *signaling* and *interleukin-1 family signaling pathways* (FDR= 0.009% and 0.394%, Pathway
1485 Analyzer). The TCR signaling pathway's most highly differentially expressed genes included
1486 major histocompatibility complexes HLA-DRB1 and HLA-DQA1, vasodilatory-stimulated
1487 phosphoprotein (VASP), and the TCR subunit CD247 (Figure 29b), whereas differentially
1488 expressed genes of the IL-1 signaling pathway included HMGB1, which encodes the 'high
1489 mobility group box 1' protein that promotes inflammatory responses to infectious signals, and
1490 protein tyrosine phosphatase 7 (PTPN7) which regulates TCR signaling (Figure 29c).

1491 Other immune-related genes that we identified as being differentially expressed by lymphatic
1492 endothelial cells but were not associated with the above pathways included MRC1, TLR4,
1493 CCL21, and ACKR2 (Figure 29—figure supplement d-f, Figure 23- figure supplement 3g).

1494 MRC1 encodes mannose receptor 1, a cognate binding partner of L-selectin—that is expressed
1495 by lymphocytes and facilitates their binding to endothelial cells. Toll-like receptor 4 (TLR4)
1496 plays a crucial role in mediating the innate immune response to bacterial lipopolysaccharide
1497 (LPS), where it engages other levers of the immune response upon binding to bacterial LPS[83].
1498 chemokine ligand 21 (CCL21) mediates homing of lymphocytes to lymph organs, whereas the
1499 atypical chemokine receptor 2 (ACKR2) regulates cellular interactions with inflammatory
1500 leukocytes[84]. Interestingly, we found lymphatic and vascular endothelial cells both expressed the
1501 pattern recognition receptor NOD1, which detects bacterial peptidoglycan fragments and initiates
1502 immune responses to gram-negative bacteria[85] (Figure 29g). There is thus ample evidence
1503 showing the respective endothelial cell types are interacting with cells of the immune system and
1504 actively participating in immune surveillance –in distince ways.

1505 **ECM Organization (7v9).** While vascular endothelial cells are surrounded by a continuous
1506 basement membrane, the basement membrane surrounding lymphatic endothelial cells is thin and
1507 discontinuous (Figure 2d, Figure 2—figure supplement 2)[8]. We thus expected to find differences
1508 in their respective expression of extracellular matrix components. After performing pathway
1509 analysis, we indeed found enrichment of *ECM proteoglycans* in the lymphatic cells
1510 (FDR=1.09%). Contributing to the pathway's enrichment was the strong lymphatic expression of
1511 all three TGFβ ligands (TGFB1, -B2 and -B3), fibromodulin (FMOD), biglycan (BGN),
1512 fibronectin (FN1), tenascin XB (TNXB), integrin alpha 8 (ITGA8), collagen 6A2 (COL6A2),
1513 and laminin beta 2 (LAMB2, Figure 29h). Elastic fibers appear to be strongly associated with
1514 lymphatics, as we found the *elastic fiber formation* pathway highly enriched in this cell type
1515 (FDR=0.5%). Twenty-four genes in the elastic fiber pathway were elevated in lymphatic
1516 endothelial cells, including 'elastin microfibril interfacer 1' (EMILIN1) and latent transforming
1517 growth factor beta (LTBP1). In a prior section, we introduced the fibulin gene family, where we
1518 discussed their importance to elastic fibers and identified several family members (EFEMP1,
1519 FBLN1, FBLN5) differentially expressed by lymphatic endothelial cells –these genes are also
1520 members of the *elastic fiber formation* pathway (Figure 29i).

1521 The only ECM-related pathway enriched in vascular endothelial cells was c*ollagen biosynthesis*
1522 *and modifying enzymes* (FDR = 6.65%). Differentially expressed genes contributing to its
1523 enrichment included the collagens COL9A3, COL4A1, COL4A2, COL18A1, COL21A1,
1524 COL13A1, COL8A1, COL15A1; procollagen-lysine 2-oxoglutarate 5-dioxygenase 1 (PLOD1);
1525 and collagen beta (1-O) galactosyltransferase 1 (COLGALT1, Figure 29j). The type IV collagens
1526 are a major structural component of basement membranes, indicating that endothelial cells
1527 collaborate with pericytes to produce the vascular basement membrane.

1528 **Cell signaling (7v9).** VEGF signaling is a critical regulator of vascular development.
1529 Interestingly, we found *signaling by VEGF* enriched in lymphatic endothelial cells (FDR=
1530 0.563%) due to their elevated expression of 55 of the 105 genes in the pathway, despite the
1531 vascular endothelial cells' differential expression of several key pathway members, like vascular
1532 endothelial growth factor c (VEGFC), placental growth factor (PGF), and VEGF Receptor-1
1533 (VEGFR1/FLT1). The lymphatic endothelial cells, in contrast, expressed higher levels of other
1534 pathway members, like: 'p21 (Cdc42/Rac)-activating protein 3' (PAK3), cytochrome b245
1535 (CYBA), the mTORC2 target PRR5, and VEGFR3 (FLT4, Figure 29k).

1536 Distinct Notch signaling pathways were enriched in both endothelial cell types. We found the
1537 broader *Signaling by notch* pathway enriched in vascular endothelial cells (FDR=0.434%),
1538 whereas the *signaling by notch2* pathway was enriched in lymphatic endothelial cells—at a false
1539 discovery rate of 9.97% (near the 10% cut-off). Of the 174 genes in the *signaling by notch*
1540 pathway, 103 were differentially expressed by vascular endothelial cells, including genes
1541 encoding the Notch ligands jagged 1 (JAG1), jagged 2 (JAG2), and delta-like 4 (DLL4, Figure
1542 29l). Notch2 transcript levels were higher in the lymphatic cells, which, compared to all other
1543 breast cell types, appeared to stem from its repression within vascular endothelial cells (Figure
1544 29m). This contributed to *signaling by notch2* pathway's enrichment in lymphatics, along with
1545 robust lymphatic expression of other pathway members, like ADAM metallopeptidase domain
1546 10 (ADAM10) and mastermind-like transcription coactivator (MAML2).

1547 **Muscle contraction (7v9).** We found the *muscle contraction* pathway enriched in vascular
1548 endothelial cells (FDR= 4.30%, Pathway Analyzer). However, both endothelial cell types did not
1549 appreciably express skeletal actin (ACTA1) or smooth muscle actins (ACTA2 or ACTG2), so
1550 they do not appear to be contractile. Of course, this is consistent with contractile smooth muscle
1551 cells being the primary regulator of vessel diameter. Instead, we found the vascular endothelial
1552 cells' enrichment of the *muscle contraction* pathway primarily resulted from the differential
1553 expression of a) the calcium-permeable transient receptor potential cation channel (TRPC1), b)
1554 the calcium-dependent membrane protein annexin A6 (ANXA6), and c) several potassium
1555 channels, including KCNE3, KCNK5, and KCNK6. These intermediate and small conductance
1556 channels play an essential role by regulating the release of endothelial cell autacoids, like nitric
1557 oxide[86], and were noticeably absent in lymphatic endothelial cells (Figure 29n).

1558 ***Fibroblasts (ADMSCs) vs. Adipocytes (Pop8 vs. Pop12)***

1559 As epithelial cells form the parenchyma of the breast, the bulk of the tissue stroma is formed by
1560 the mesenchymal cell types (adipocytes, fibroblasts, vascular smooth muscle cells, and
1561 pericytes), with endothelial cells and a spectrum of leukocytes providing key elements[8]. We did
1562 not quantify adipocytes by FACS, but we could deduce from the RNA-sequencing data that
1563 adipocytes are one of, if not the most, abundant breast cell type. This was inferred from the
1564 covariance analysis and hierarchical clustering that grouped adipocytes with the whole tissue
1565 control samples, indicating they were disproportionately comprised of adipocytes (Figure 4b and
1566 Figure 6). The remaining cell types were assessed by FACS, and it came as no surprise that
1567 fibroblasts were also abundant (Figure 1b).

1568 Fibroblasts are spindle-shaped cells that predominantly provides the tissue's foundation—a
1569 collagenous-rich matrix of fibrillar extracellular proteins, which fibroblasts shape through the
1570 secretion of growth factors, proteoglycans, and matrix-bound proteases. Fibroblasts are the
1571 archetypical mesenchymal cell type, existing within a developmental lineage that is shared with
1572 adipocytes. In our adjoining article, we described how breast fibroblasts express mesenchymal
1573 stem cell markers (CD10, CD34, CD73, Thy1, CD44, CD105) and displayed a functional multi-
1574 lineage potential when cultured under osteogenic or adipogenic conditions[8]. For these reasons,
1575 we designated Pop8 fibroblasts as adipocyte-derived mesenchymal stem cells (ADMSCs) but
1576 commonly refer to them as fibroblasts. In our cultures, and especially within the tissue,

1577 adipocytes appeared as large, spherical cells containing a single large lipid droplet (unilocular
1578 adipocytes) that stained for neutral lipids[8] (Figure 2a,c; Figure 2—figure supplement 1a, 2a).
1579 Adipocytes notably serve as the body's chief storage depot of triglycerides and other
1580 acylglycerols and regulate energy metabolism through secretion of adipokines[87], e.g, leptin
1581 hormone (LEP) and adipsin (CFD), which we found highly expressed in the pop12 lipid
1582 (adipocyte) samples (Figure 31a,b). To establish the different cellular processes shaping
1583 adipocytes and fibroblasts, we examined their expressed transcripts.

1584 In contrast to the other breast cell types, freshly-isolated adipocytes were not FACS purified
1585 prior to RNA-sequencing, and therefore results should be interpreted differently. Adipocytes
1586 were instead isolated from the phase-separated lipid layer of the collagenase tissues digests,
1587 which we had visually inspected for contaminating cellular content (Hoechst staining). The
1588 number of paired-end mRNA fragments measured from the RNA isolated from the lipid layers
1589 was excellent (34 to 45 million fragments), and the replicates nicely clustered together (Figure
1590 4b; Figure 6; Figure 7—figure supplement 2a, Figure 8a). However, one sample did stray in the
1591 projection of PC1 and PC2 (Figure 7a), indicating some discordance between samples. We also
1592 noticed, in all three adipocyte replicates, low transcript levels for many genes (e.g., this can be
1593 observed in the many heatmaps presented herein, such as those containing the claudins and gap
1594 junction proteins, Figures 16,17). These adipocyte samples thus have a higher noise level than
1595 that observed in the FACS-purified samples, which is to be expected given the purity and
1596 excellent quality of those samples. The noise is likely derived from a fraction of contaminating
1597 cells in the samples, such as leukocytes. Therefore, we have prioritized genes and pathways in
1598 our analyses that are not heavily associated with immune response or where there is not evidence
1599 of robust adipocyte mRNA levels. With the caveat that the adipocyte samples should be
1600 interpreted differently, we explored the nature of adipocytes by comparing them with fibroblasts.

1601 Pairwise comparison of RNA transcripts from adipocytes and fibroblasts revealed 1,171
1602 differentially expressed genes. This was notably the lowest number of differentially expressed
1603 genes between any cell-type comparison, similar to the number differentially expressed between
1604 the two luminal epithelial cell types (1,301, figure 20). We anticipated fewer DE genes given the
1605 similarity and shared lineage of adipocytes and fibroblasts. Of the 1,171 differentially expressed
1606 genes we identified, 207 were more highly expressed by fibroblasts, and 964 were more highly
1607 expressed by adipocytes (Figure 30, Figure 3—table supplement 7 [8v12]).

1608 When we investigated the identity of the differentially expressed genes, we found fibroblasts
1609 expressed the ECM protein biglycan, delta-like non-canonical Notch ligand 1 (DLK1), matrix
1610 metalloproteinase 3 (MMP3), and others. In contrast, adipocytes differentially expressed leptin
1611 (LEP), thyroid hormone-responsive protein (THRSP), and glycerol-3-phosphate dehydrogenase
1612 (GPD1), along with 961 other genes (Figure 30). After performing GSEA, we identified 387
1613 pathways enriched between the two cell types (FDR ≤ 20%), with 69 pathways in fibroblasts and
1614 318 in adipocytes (Pathway Analyzer). Due to the large number of of immune-related genes and
1615 pathways in adipocyte samples, we increased the false discovery rate cutoff to 20% for this
1616 comparison, which resulted in the enrichment of more pathways related to fat storage and
1617 metabolism.

1618 **Extracellular matrix organization (8v12).** Twelve pathways associated with ECM organization
1619 were enriched in fibroblasts, including *degradation of the extracellular matrix*, *ECM*
1620 *proteoglycans*, and *elastic fiber formation* (FDR< 0.001%, < 0.001%, and =0.090%,
1621 respectively, Pathway Analyzer). Fibroblasts differentially expressed genes contributing to these
1622 pathways, including TIMP metallopeptidase inhibitor 1 and 2 (TIMP1 and TIMP2);
1623 metallopeptidases 2, 3, 13, 14, and 16 (MMP2, MMP3, MMP13, MMP14, and MMP16); fibulin-
1624 1 (FN1); fibrillin-1 (FBN1); and elastin (ELN, Figure 31c-e). There were two pathways
1625 associated with ECM organization enriched in adipocytes, including *laminin interactions* and
1626 *anchoring fibril formation* (FDR=3.23% and 9.97%, Figure 31f). Adipocytes differentially
1627 expressed integrins A6, B4, A3, and A2; laminins B3, A3, and C3; and collagens 4A5 and 4A6,
1628 which contributed to the enrichment of these pathways (Figure 31f).

1629 **Metabolism (8v12).** Included in the enriched metabolic pathways was *Glycosaminoglycan*
1630 *metabolism* that was associated with fibroblasts (FDR<0.001%). This pathway's genes included
1631 heparanase-2 (HPSE2), hyaluronan synthase-2 (HAS2), and glypican-3 (GPC3, Figure 31g). In
1632 contrast, we found adipocytes enriched in pathways associated with lipid metabolism, that
1633 included *triglyceride metabolism, triglyceride catabolism*, and *arachidonic acid metabolism*
1634 (FDR= 15.5%, 15.5%, and 4.13%, Pathway Analyzer, Figure 31h). Differential expression of
1635 fatty acid binding proteins (FABP4, FABP5, FABP7, and FABP9) were among those genes
1636 contributing to enrichment of pathways associated with triglyceride metabolism and catabolism
1637 (Figure 31h). These intracellular fatty acid-binding proteins (FABPs) exhibit high affinity for a
1638 single long-chain fatty acids and are essential for triglyceride storage; they mediate intracellular
1639 fatty acid transport and interact with PPAR nuclear receptors[88]. Enrichment of *arachidonic acid*
1640 *metabolism* in adipocytes was defined by differential expression of arachidonate 5-lipoxygenase
1641 (ALOX5) and its activating protein (ALOX5AP, Figure 31i), which catalyzes the first two steps
1642 of the biosynthesis of leukotrienes, which are potent mediators of inflammation.

1643 We also found adipocytes uniquely expressed cell death inducing DFFA like effector C (CIDEC,
1644 Figure 31j), which binds to lipid droplets and regulates their enlargement, thereby restricting
1645 lipolysis and favoring storage[89]. The adipocytes also uniquely expressed aquaporin AQP7 and
1646 the aquaporin pseudogenes AQP7P1 and AQP7P3 (Figure 31k-m), which control water and
1647 glycerol transport and promote the catabolism of lipids in adipocytes[90].

1648 **Cell signaling (8v12).** Accumulation of adipose tissue is regulated in part by estrogen
1649 signaling[91]. Therefore, it is unsurprising that *ESR-mediated signaling* and its associated pathways
1650 were enriched in adipocytes (FDR= 0.110%), which differentially express the contributing genes
1651 GATA binding protein 3 (GATA3), caveolin 2 (CAV2), epidermal growth factor (EGF), and
1652 epiregulin (EREG, Figure 31n). Adipocytes also expressed estrogen receptors 1 and 2 (ESR1 and
1653 ESR2), albeit at 10x lower levels lower than ER^{Pos} luminal epithelial cells but at a 4.3x higher
1654 level than fibroblasts (Figure31n, Pathway Analyzer).

1655 ***Pericytes vs. Fibroblasts (ADMSCs, Pop6 vs. Pop8)***

1656 Whereas mesenchymal adipocytes contribute to the storage and mobilization of energy-rich
1657 metabolites; and mesenchymal fibroblasts secrete ECM proteins and enzymes that shape the
1658 extracellular stromal environment, the mesenchymal perivascular cells influence and modulate

1659 blood vessels. In a companion article[8], we described how we identified vascular smooth muscle
1660 cells and pericytes while devising methods to isolate and culture all breast cell types, finding
1661 these perivascular cells to be an elusive set of cell types. Pericytes have long been
1662 overlooked[92,93], largely stemming from their heterogeneity that, in turn, required a combinatoric
1663 assessment of markers to identify and purify them. After discovering how to reliably isolate and
1664 propagate breast pericytes, we found their microscopic appearance fascinating, as they had a
1665 stellate astrocyte-like morphology that was entirely distinct from fibroblasts and adipocytes[8].
1666 However, it was difficult to tell the pericytes and fibroblasts apart at higher confluency. We were
1667 thus interested in exploring the biological processes defining these two similar mesenchymal cell
1668 types (i.e., FACS populations #6-pericytes and #8-ADMSC/fibroblasts).

1669 Analysis of transcripts expressed by freshly isolated pericytes and fibroblasts revealed 2,932
1670 differentially expressed genes (BH/FDR adj. p-value < 0.1, Figure 32). Interestingly, these were
1671 not split evenly between the two cell types, as fibroblasts had roughly three times the number of
1672 DE genes than pericytes. Among these 2,187 genes, were platelet-derived growth factor receptor
1673 alpha (PDGFRA), matrix metallopeptidase 2 (MMP2), and hydroxyacid oxidase 2 (HAO2). In
1674 contrast, pericytes differentially expressed 748 genes, including pyruvate dehydrogenase kinase
1675 4 (PDK4), CD36, and desmin (DES, Figure 32). Along with desmin, we found the pericytes
1676 differentially expressed other genes previously associated with pericytes, such as melanoma cell
1677 adhesion molecule (MCAM), G protein receptor RGS5, and the potassium channel KCNJ8. We
1678 also found pericytes uniquely expressed the serotonin receptor isoform HTR1F, which was
1679 distinct from the HTR1B isoform we discovered uniquely expressed by vascular smooth muscle
1680 cells that, coincidently, helped us identify the vSMC FACS population[8] (serotonin is a
1681 vasoactive peptide[94]). We ranked all expressed genes by their degree of differential expression
1682 and used this list for pathway analysis.

1683 We discovered 477 pathways enriched between pericytes and fibroblasts at a false discovery rate
1684 of 10%. We found the enrichment of many pathways resulted from small concordant expression
1685 differences between the cell types, such as those involved in transcriptional regulation, cell cycle,
1686 and immunity. Here, we focus on pathways containing significantly differentially expressed
1687 genes.

1688 **Muscle Contraction (6v8).** As one would predict, we found perivascular pericytes enriched with
1689 pathways involved with *muscle contraction* and *smooth muscle contraction* (FDR =0.361% and
1690 0.026%, respectively, Figure 33). Pericytes differentially expressed myosin light chain kinase
1691 (MYLK), myosin heavy chain 11 (MYH11), actin gamma 2 (ACTG2), and the voltage-gated
1692 sodium channel type IV (SCN4A). Also differentially expressed by pericytes was dystrophin
1693 (DMD), which anchors ECM proteins to cytoplasmic filamentous actin (and whose defects lead
1694 to Duchenne and Becker muscular dystrophies). Tropomodulin 1 (TMOD1) was also highly
1695 expressed by pericytes, which is a protein that regulates tropomyosin via blocking the elongation
1696 and depolymerization of actin filaments (Figure 33- figure supplement 1a). Interestingly, *cardiac*
1697 *conduction* was enriched in fibroblasts (FDR=5.64%), which differentially expressed T-box 5
1698 (TBX5) transcription factor; potassium channels KCNK1, KCNK2, KCNIP1, and KCNIP4; and
1699 the sodium channel SCN1B (Figure 33, Figure 33- figure supplement 1b). Fibroblasts share
1700 similar expression of voltage-gated channels to cardiac myocytes, which are responsible for

1701 cardiac conduction in the heart. Additionally, the dramatic difference in gene expression could
1702 be due to the low expression of many of these channels in pericytes, which may instead rely on
1703 endothelial cells for depolarization signals (as we found similar channels were differentially
1704 expressed by vascular endothelial cells when compared to their lymphatic counterparts, Figure
1705 29n).

1706 **Hemostasis (6v8).** We found fourteen pathways associated with hemostasis enriched in
1707 pericytes, including *platelet homeostasis* (FDR= 0.732%, Figure 33). Its enrichment was defined
1708 by differential expression of phosphodiesterases 1A, 1B, and 5A (PDE1A, PDE1B, and PDE5A),
1709 transient receptor cation channel C6 (TRPC6), cGMP-dependent protein kinase 1 (PRKG1), and
1710 the prostaglandin receptor PTGIR (Figure 33- figure supplement 1c). The phosphodiesterases are
1711 responsible for hydrolyzing cGMP and modulating the effects of nitric oxide signaling. Also
1712 related to the vasculature was the enrichment, in pericytes, of *cell surface interactions at the*
1713 *vascular wall* (FDR=1.14%), resulting from their differential expression of MER proto-oncogene
1714 (MER), thrombomodulin (THBD), and angiopoietin 1 (ANGTP1, Figure 33, Figure 33- figure
1715 supplement 1d).

1716 **Intracellular Transport (6v8).** Pericytes differentially expressed kinesins (KIF4A, KIF2A, and
1717 KIF26B) and tubulins (TUBB2A, 3, and 4B), leading to their enrichment of *kinesins* pathway
1718 (FDR<0.001%, Figure 33, Figure 33- figure supplement 1e). *Clathrin-mediated endocytosis* was
1719 also enriched in pericytes (FDR<0.001%), due to their differential expression of the endocytic
1720 adapter protein SGIP1, WNT5A ligand, the wnt receptor 'frizzled class receptor 4' (FZD4),
1721 angiotensin II receptor (AGTR1), and the endocytosis inhibitor synaptotagmin XI (SYT11,
1722 Figure 33, Figure 33- figure supplement 1f). This pathway regulates the cell surface expression
1723 and signaling of RTKs and GPCRs[95,96], and we found that most of the enriched pathways
1724 associated with RTKs or GPCRs were enriched in pericytes (Pathway Analyzer).

1725 **Cell signaling (6v8).** We identified thirty enriched pathways associated with RTK signaling in
1726 pericytes, including *signaling by NTRKs*, *signaling by VEGF*, and *signaling by PDGF*
1727 (FDR=0.011%, 0.365%, and 6.67%, respectively, Figure 33). Neurotropic tyrosine kinase
1728 receptors are activated by neurotrophins like NGF and NTF3—which were highly and
1729 differentially expressed by pericytes (Figure 33- figure supplement 1g). Pericytes also
1730 differentially express AXL receptor tyrosine kinase (AXL), caveolin-1 (CAV1), and integrin
1731 beta 3 (ITGB3), which define the enrichment of *signaling by VEGF* (Figure 33- figure
1732 supplement 1h). *Signaling by PDGF* was enriched in pericytes due to the differential expression
1733 of collagens COL4A1, COL4A2, COL4A4, COL5A3; platelet-derived growth factor alpha
1734 (PDGFA); and thrombospondin 4 (THBS4, Figure 33- figure supplement 1i).

1735 The enrichment of thirteen Notch signaling pathways in pericytes indicate their importance to
1736 pericyte function. The broader *signaling by notch* pathway (FDR<0.001%) was enriched largely
1737 because of the differential expression of NOTCH receptors 1 and 3 (NOTCH1, NOTCH3), the
1738 downstream transcription factors HEYL and HEY2, and SMAD family member 3 (SMAD3,
1739 Figure 33- figure supplement 1s).

1740 **Cell-Cell Communication (6v8).** Communication between pericytes and endothelial cells is
1741 critical to vessel development and function. We found pericytes enriched with the broad *cell-cell*

1742 *communication* and cell junction pathways (FDR=0.10% and 0.06%, respectively, Figure 33).
1743 These pathways involve tight junctions, adherins junctions, gap junctions, and signal regulatory
1744 proteins, and pericytes differentially expressed a) cadherins 3 and 6 (CDH3 and CDH6), classical
1745 calcium adhesion molecules; b) plectin (PLEC), which interacts and organizes connections to
1746 intermediate filaments and microtubules and are essential components of desmosomes and
1747 hemidesmosomes; and c) laminin subunit alpha-3 (LAMA3), which assembles into laminin
1748 heterotrimers that are essential components of the vascular basement membrane. (Figure 33-
1749 figure supplement 1j).

1750 **Extracellular Matrix Organization (6v8).** Pericytes and fibroblasts have different functional
1751 roles requiring distinct interactions with different ECM components. This was reflected in the
1752 different ECM organization pathways we found in either cell type. For example, acting in
1753 opposition, the *collagen formation* pathway was enriched in pericytes, whereas the *collagen*
1754 *degradation* and *degradation of the extracellular matrix* pathways were enriched in fibroblasts.
1755 Enrichment of the *Collagen formation* pathway in pericytes (FDR=4.28%) resulted from their
1756 differential expression of COL4A1, COL4A2, COL4A3, COL4A4, COL4A5, COL5A3,
1757 COL17A1, COL23A1, COL25A1, and COL27A1 (Figure 33, Figure 33- figure supplement 1k).
1758 As introduced in the discussion of the collagen family above, the perivascular cells expressed the
1759 highest levels of type IV collagens (basement membrane collagens, Figure 15—figure
1760 supplement 2a-c), consistent with their spatial location in the tissues—embedded within the
1761 vascular basement membranes—and role of maintaining these basement membranes. In contrast,
1762 fibroblasts exist in the extracellular matrix, which they are responsible for forming and shaping.
1763 We found fibroblasts differentially expressed collagens COL1A1, COL3A1, COL6A6,
1764 COL8A1, COL8A2, COL14A1, and COL16A1, contributing to the *degradation of the*
1765 *extracellular matrix* pathway's enrichment (FDR=0.705%, Figure 33, Figure 33- figure
1766 supplement 1l). Additional differentially expressed genes included matrix metalloproteinases
1767 MMP2, MMP3, MMP10, MMP12, MMP13, and MMP16; TIMP metallopeptidase inhibitor 1
1768 (TIMP1); calpains 5 and 6 (CAPN 5 and CAPN6); and the bone remodeling protein cathepsin
1769 K—which cleaves collagen I fibers (Figure 33- figure supplement 1l). We also found enrichment
1770 of *ECM proteoglycans* in fibroblasts (FDR=4.68%), as they differentially expressed versican
1771 (VCAN), asporin (ASPN), lumican (LUM), heparan sulfate proteoglycan 2 (HSPG2), decorin
1772 (DCN), and biglycan (BGN) proteoglycans (Figure 33, Figure 33- figure supplement 1m). Other
1773 genes in this pathway highly expressed by fibroblasts were tenascins N and C (TNN, TNC),
1774 TGFB3, and laminin alpha 1 (LAMA1, Figure 33- figure supplement 1m).

1775 **Elastic Fibers (6v8).** Elastic fibers are essential components of connective tissue and contribute
1776 to tissue elasticity. We discovered pathways related to elastic fibers were among the most
1777 significantly enriched pathways in fibroblasts, i.e., *molecules associated with elastic fibers*
1778 (FDR=0.29%) and *elastic fiber formation* (FDR=0.90%). Leading to these pathways' enrichment
1779 were the fibroblasts' robust expression of elastin (ELN), elastic microfibril interfacer 1
1780 (EMILIN1), fibrillins 1 and 2 (FBN1 and FBN2), bone morphogenic protein 7 (BMP7),
1781 fibronectin (FN1), and lysyl oxidase 1 (LOXL1, Figure 33, Figure 33- figure supplement 1n).

1782 **Metabolism (6v8).** Pericytes and fibroblasts were enriched with different metabolic pathways.
1783 There were fifteen pathways enriched in pericytes and thirteen in fibroblasts (Pathway analyzer).

One of the most significant pathways enriched in pericytes was *Integration of energy metabolism* (FDR=0.016%), stemming from the pericytes' differential expression of adrenoreceptor alpha 2A and 2C (ADRA2A and ADRA2C), thrombospondin receptor (CD36), adenylate cyclases 6 and 9 (ADCY6 and ADCY9), cysteinyl leukotriene receptor 2 (CYSLTR2), and the voltage-dependent calcium channel CACNB2 (Figure 33, Figure 33- figure supplement 1o). In contrast, fibroblasts were enriched in *fatty acid metabolism* (FDR=0.889%), owing to their differential expression of prostaglandin synthases PTGES, PTGS2, PTGIS, and PTGDS; the C2 domain-containing phospholipase PLA2G4A; and their unique expression of hydroxyacid oxidase 2 (HAO2, Figure 33, Figure 33—figure supplement 1p).

White adipocyte differentiation (6v8). We found forty-five of the eighty-four genes composing the *transcriptional regulation of white adipocyte differentiation* pathway were elevated in pericytes, leading to its enrichment (FDR=0.007% Figure 33). Genes differentially expressed by pericytes included NR2F2, which regulates apolipoprotein A-I transcription; CD36 (fatty acid translocase), which regulates the transport and uptake of long-chain fatty acids, and fatty acid binding protein-4 (FABP4), which mediates intracellular fatty acid transport (Figure 33—figure supplement 1q). One of the most notable genes in this pathway is phosphoenolpyruvate carboxykinase 1 (PCK1), which catalyzes the production of phosphoenol pyruvate from oxaloacetate. The activity of PCK1, which is regulated at the transcriptional level, commits mitochondrial pyruvate to gluconeogenesis or glyceroneogenesis[97]. This step is irreversible under physiological conditions and is the rate-limiting step controlling gluconeogenesis flux. We found PCK1 expression in breast cells was limited to the pericytes and adipocytes (Figure 33—figure supplement 1r), indicating, at the very least, that pericytes have an increased demand for cytoplasmic pyruvate.

1808 ## Summary

1809 Resolving the types and properties of cells in a tissue is vital to understanding the underlying
1810 symphony of interactions that make life possible. It is essential to overcome diseases such as
1811 cancer. Here, and in the adjoining article, we explored the fabric of the breast through a
1812 cytometric and molecular dissection of the tissue, which has provided us with insights into its
1813 inner workings. From the outset, we had questions regarding the number and types of cells
1814 composing the breast, their spatial location, their functions, and how they contributed to the
1815 tissue as a whole.

1816 To address these questions, we revisited breast tissues to explore their architecture and cellular
1817 composition through an exhaustive cytometric dissection of the tissue, which led to our
1818 development of a FACS procedure that reconciled every cell in the tissue (Figure 1). We isolated
1819 twelve cell populations from four lineages and developed techniques to culture nearly every one
1820 (Figure 2)[8]. Our system of defining breast cell types provides a comprehensive resource that can
1821 be used to explore interactions in more complex and physiologically relevant culture systems
1822 aimed at mimicking the tissue microenvironment.

1823 Here, we explored the nature of each breast cell type by dissecting their transcriptomes. These
1824 were acquired by scaling up our FACS procedure and painstakingly sorting each cell type to
1825 purity. RNA was then isolated and sequenced, leading to the acquisition of transcriptomes that
1826 were of high depth and quality. Although the rare cell types yielded fewer cells and transcripts,
1827 concordance of transcriptomic profiles between all cell replicates illustrated the reproducibility
1828 of our methods (Figure 4). Analysis of transcripts identified uniquely expressed genes that have
1829 already helped us decipher spatially resolved transcripts from breast tumors. Bioinformatic
1830 analyses of transcriptional profiles further helped validate the identity of several cell types by
1831 providing information on expressed markers and revealing their functional commonalities to
1832 related cell types (Figure 6).

1833 Over 20,000 genes were measured in each cell type, providing a wealth of information. As we
1834 explored the transcriptomes from every major cell type in the breast, there were so many exciting
1835 features that we had to come to terms with the fact that we couldn't describe all that we
1836 encountered. Our solution was to build resource-sharing tools that aid exploration of our data[9,10].
1837 To demonstrate the utility of these resources, we addressed three pressing questions: a) What
1838 features distinguish each cell lineage? b) What gene types confer cell-specific functions? and c)
1839 What biological processes define each cell type?

1840 To identify the cell lineages' distinguishing features, we investigated the cellular pathways
1841 impacted by lineage-specific genes, which we defined through principal component analysis.
1842 The pathway enrichment analyses affirmed established attributes of each cell lineage while
1843 identifying others we had not previously associated with these lineages (Figures 10-14).

1844 Focusing on the genes families that varied most among the different cell types provided us with
1845 an unconventional look at the types of genes contributing to the essential qualities of the breast
1846 cell types. We examined the expression of each gene, from each gene family, within every breast
1847 cell type, revealing the gene types that vary across cell types. We provide a full accounting of

each gene family in a supplemental table and have meticulously described nine gene families whose functions extend from the nucleus (e.g., hormone receptors) to the exterior cell microenvironment (e.g., laminins, Figures 15-19).

To address our last question and develop a broad perspective of each breast cell type's fundamental properties and individual functions, we performed a comprehensive assessment of the biological pathways and processes that are differentially enriched in each cell population. We performed fifty-five independent analyses identifying differentially expressed genes and enriched pathways between every possible cell-type comparison. These results are available as a downloadable interactive application (Pathway Analyzer[9]). We comprehensively analyzed six comparisons we believed were of the highest interest, as they revealed features of every major cell type and identified those distinguishing closely related cell types (Figures 21-33). As the field moves forward with single-cell and spatial transcriptomic approaches to explore and define the significance of cellular heterogeneity within tissues and tumors[98,99], we predict that the abundant number of transcripts measured here, within the full spectrum of FACS-purified breast cell types, will provide a strong foundation for identifying cell populations and interpreting results.

A broad wealth of information touching on many scientific disciplines is presented herein, and we recognize that developing deeper insights requires engagement and a focused exploration of the associated datasets. We offer novel tools to aid this exploration[9,10] and look forward to creating and sharing others. This body of work, presented here and in the adjoining article, presents a systematic approach for identifying, isolating, and culturing nearly every major breast cell type of epithelial, endothelial, and mesenchymal origin. We spent many years refining our FACS strategy, used multiple rounds of sorting to ensure purity, exhaustively stained tissues and primary cultures to validate and characterize cell types, and comprehensively and painstakingly analyzed each transcriptome. This rich resource, which we submit here, has provided countless insights and has already changed our perception of the cells composing this elaborate and beautiful tissue.

Methods

Tissue acquisition and processing. Breast tissues from reduction mammoplasties were obtained from the Cooperative Human Tissue Network (CHTN), a program funded by the National Cancer Institute. All specimens were collected with patient consent and were reported negative for proliferative breast disease by board-certified pathologists. The University of California at Berkeley Institutional Review Board and University of New Mexico Human Research Protections Office granted use of anonymous samples through exemption status, according to the Code of Federal Regulations 45 CFR 46.101. Upon receipt, several fragments (roughly 2cm^2) were embedded in optimal cutting temperature compound (OCT) in tissue cassettes, flash-frozen in nitrogen, and archived at -80ºC for later cryosectioning and staining. The remaining tissue, typically 20-100g, was thoroughly rinsed with phosphate-buffered saline and processed to organoids as previously described (gentle agitation method)[100]. Briefly, this included manually mincing and scoring tissues with a scalpel and incubating them overnight (12-18 hrs, 37°C) with 0.1% collagenase I (Gibco/Invitrogen) in Dulbecco's Modified Eagle Medium containing 100

1889 U/ml penicillin, 100 mg/ml streptomycin, and 100 mg/ml Normocin™ (Invivogen, San Diego,
1890 CA). The resulting divested tissue fragments (organoids) were collected by centrifugation (100g
1891 x 2 min) and either archived in liquid nitrogen (90%FBS+10% DMSO) or immediately
1892 processed to cell suspensions for FACS analysis. Tissues used for RNA-sequencing were
1893 immediately processed upon receipt and kept at 4°C during processing and FACS sorting, except
1894 for the overnight collagenase digestion at 37°C. RNA-sequenced cells were from fresh
1895 specimens and were not previously archived in nitrogen.

1896 **Antibodies.** A list of antibodies and reagents used in this study is provided in Figure 3—Table
1897 Supplement 8. The table includes antibody clone designations, conjugations, isotype, supplier
1898 product numbers, dilution factors used.

1899 **Cryosectioning and Immunostaining.** Immunofluorescence was performed on 10-80 mm
1900 cryosectioned tissue sections as described in the adjoining article[8]. Briefly, specimens were fixed
1901 with 4% paraformaldehyde for 5 minutes at 23°C, which was followed by 4% formaldehyde /
1902 0.1% saponin for 5 minutes at 23°C. Tissues were rinsed for 20 minutes in wash buffer (0.1%
1903 saponin/10% goat serum in PBS) and then incubated with primary antibodies, typically overnight
1904 at 4°C. Following the primary incubation period, samples were washed and incubated with
1905 secondary antibodies conjugated with Alexafluor 488, 568, 594, or 647 (Thermo). After 1 hour
1906 incubation at 23°C, samples were rinsed in PBS and their nuclei counterstained with 300nM
1907 DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride, Thermo). Coverslips were mounted
1908 with Fluoromount G (Southern Biotech). Images were captured using a Zeiss LSM710 confocal
1909 microscope, Zeiss Axioscope, or EVOS FL Auto imaging station. If needed, image contrast was
1910 applied to the entire image using Photoshop and were annotated with the antibodies used to stain
1911 the tissue.

1912 **Cell Preparation for Flow Cytometry and FACS analysis.** To prepare cell suspensions for
1913 FACS and flow cytometry analysis, we thoroughly rinsed the organoids in PBS, then pelleted
1914 them via centrifugation at 100g x 2 min. After removing PBS by aspiration, the organoids were
1915 suspended in 1 ml of 'Cell Dissociation Reagent' (Sigma# C5914; or Thermo# 13150016),
1916 incubated for two minutes at room temperature, upon which 3ml trypsin (used for RNA-
1917 sequenced samples) was added (0.25%, Thermo 25200072; Note that we have more recently
1918 found TrypLE better preserves CD49f staining, although it dampens CD34 resolution). Samples
1919 were incubated by hand/body temperature, thus permitting visual inspection and brief (1-3sec)
1920 pulse vortexing every 30-60 seconds. After the mixture became cloudy from the dissociating
1921 cells (about 8-10 min), the organoids/cells were gently and repeatedly pipetted through a 16-
1922 gauge needle until clumps dissipated. Afterward, we filtered cell suspensions through a 100mm
1923 cell strainer and added 3ml 0.1% w/v soybean trypsin inhibitor to stop the digest (Sigma#
1924 T9128). The suspensions were then filtered through a 40mm cell strainer, rinsed in 10-20ml
1925 PBS, and pelleted by centrifugation at high centrifugal force (400g x 5 min.—or longer, until all
1926 cells had pelleted, which we confirmed by microscopic examination of the supernatant). The
1927 cells were then rinsed in 10 ml Hanks balanced salt solution/1% BSA(w/v), counted, and again
1928 pelleted by centrifugation. After centrifugation, nearly all the Hanks/BSA was aspirated, leaving
1929 the cell pellet with roughly 60ml of Hanks/BSA. The pellet was resuspended in this small
1930 volume, and then we added FACS antibodies to the samples and incubated the samples on ice,

1931 covered, for 30 minutes (See Figure 3- table supplement 8 for antibodies). Following the
1932 incubation, the cells were rinsed in Hanks/1% BSA, centrifuged (400g x 5 min), and resuspended
1933 in Hanks/BSA with To-Pro-3 viability marker (diluted 1:4000, Thermo T3605), then filtered
1934 again through a 40mm cell strainer cap (into the FACS tube). We did not use DNase enzyme or
1935 hypotonic RBC lysis solutions on the cell preparations because we found they were unnecessary.
1936 We also wanted to avoid any potential deleterious effect these treatments may have on the cells,
1937 such as reducing cell viability or altering gene expression levels. We, again, also conscientiously
1938 chilled samples during cell preparation and the sorting procedure for similar concerns.

1939 **Flow Cytometry.** The FACS panel and gating strategy outlined herein was developed using a
1940 BD FACS Vantage 8-channel cytometer (FACSDIVA software) and has since been adapted and
1941 reproduced using a 14-channel SONY SY3200. PMT voltages were optimized on both
1942 instruments by analyzing individually stained compensation beads (AbC™ Anti-Mouse Bead
1943 Kit, Thermo # A10344) at increasing voltage intervals for each detector. Because the FACS
1944 Vantage has only eight available channels, it was necessary to use two markers (podoplanin and
1945 CD45) in a single channel (A488), which we experimentally validated. To-Pro-3 was used as the
1946 viability marker, so the corresponding 405nm channel (455/50) could be used for BV421. This
1947 channel was dedicated to CD49f to provide a maximum resolution of cell populations, which is
1948 essential to this sorting strategy. The considerable compensation typically required between the
1949 heavily overlapping dyes PE/Cy5 and To-Pro-3 was not needed, as the To-Pro-3 negative
1950 (viable) cells were used for downstream analyses, allowing us to use the PE/Cy5 channel for
1951 Thy1. Negative controls consisted of unlabeled beads and cells incubated with isotype control
1952 antibodies conjugated to each fluorophore. Sort times typically ranged 14-18 hours (50-120
1953 million cells stained— needed to yield enough cells for RNA-sequencing). Cells were chilled at
1954 the sample intake and collection chambers, and all samples were chilled on ice during the entire
1955 experiment. All media and PBS were filtered through a 0.2mm filters to prevent the FACS
1956 machine from triggering events from crystals and debris. Staining profiles of cells collected at
1957 the beginning and end of the sorts (sometimes separated by as many as 18 hours) were identical
1958 and did not show any signal loss or appreciable decrease in viability. Because our machines
1959 could sort cells into a maximum of 4 tubes, we first sorted cells into four groups, containing 2-4
1960 populations each. These mixtures were: Tube1 (CD34Pos) and contained Pops 7,8,9; Tube 2
1961 ($Thy1^{Pos}$) contained Pops 3,6; Tube 3 (Luminal) contained Pops 1&2; and Tube 4 (Null gate)
1962 contained Pops 4,5,10,11. During the first round of sorting, cells were typically sorted at
1963 approximately 8,000 cell/sec. We followed this initial enrichment with purity sorts for each cell
1964 type, using stringent sort masks, sorting each population into individual tubes at a rate of
1965 approximately 2,000 cells/sec. FACS data were analyzed using Flowjo software (version v7.6.3
1966 – v10.8), Tree Star Inc./ Becton, Dickinson & Company).

1967 **RNA Isolation and Sequencing.** Following the overnight digestion of tissues with collagenase,
1968 we centrifuged the samples at 100g x 2 min. This produced aqueous and lipid phases that
1969 separated under centrifugal force. The top lipid phase was collected, visualized microscopically
1970 for cellular content, and then processed immediately for RNA extraction using Qiagen RNA
1971 Lipid Tissue Mini Kit (Qiagen 74804), according to the manufacturer's instructions. Total RNA
1972 from the FACS-purified cells was isolated immediately after FACS sorting using Qiagen RNA

1973 Mini (>100,000 cells, Qiagen 74104) or RNA Micro kits (<100,000 cells, Qiagen 74004)
1974 according to the manufacturer's instructions. Total RNA was treated with DNAse I for 30
1975 minutes at 37°C (Ambion DNA-*free* DNA Removal kit, Thermo AM1906*)*, then treated with
1976 DNAse Inactivation beads. RNA was then concentrated with RNA Clean & Concentrator™-5
1977 columns (Zymo R1015), quantified by Nanodrop and Agilent Bioanalyzer assays, and stored at -
1978 80°C.

1979 **RNA Library construction and sequencing.** mRNA was captured using poly-T coated dishes
1980 (Clontech) and used for library construction using SMARTer technology (SMARTer Stranded
1981 RNA-seq Kit, Takara Biotech) following the manufacturer's guidelines. For low abundance
1982 samples, mRNA was captured on biotinylated LNA coated beads, magnetically captured and
1983 rinsed 3x in binding and wash buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl).
1984 mRNA was eluted from Dynabeads in 10ml volume of elution buffer, following by a 5 min.
1985 incubation at 68°C. It was then used for first-strand cDNA synthesis (SMARTer Stranded RNA-
1986 seq Kit). Sample clean-up and removal of adapter-dimers) was performed using Ampure XP
1987 beads (Beckman Coulter), according to the manufacturer's protocol. Specific attention was given
1988 to the samples to prevent under- or over-drying the beads, which we found to reduce PCR
1989 efficiency. PCR amplification of the RNA-Seq library was performed using SeqAmp DNA
1990 Polymerase (Takara Bio, cat# 638509). Each library was then purified prior to sequencing using
1991 SPRI AMPure Beads (Beckman Coulter). Sequencing was performed using Illumina Trueseq v.2
1992 technology, targeting approximately 30 million paired-end reads for each cell population
1993 (Illumina HiSeq 2x100bp).

1994 **RNA-Seq pre-processing and quality control assessment.** We processed raw count data from
1995 aligned fragments (Figure 3—table supplement 2) using R programming language (RStudio) and
1996 associated BioConductor packages. Quality control analyses (Cook's distance box plot,
1997 covariance matrix, sample correlation matrix) were performed using the DESeq2 package[13]. For
1998 data visualization (donut and heatmaps), the count matrix data was transformed by two methods
1999 used in the DESeq2 package: (i) regularized log transformation (rlog), and (ii) variance-
2000 stabilizing transformation (vst), both normalized to library size. Quality control assessment of
2001 expressed transcripts across sorted cell samples was conducted, and one sample was removed
2002 from statistical analyses as it was an extreme outlier (Epithelial Population 4 Sample N239). This
2003 was determined by: (i) distance matrix, (ii) Principal Component Analysis, and (iii)
2004 Kolmogorov-Smirnov test that compares each sample to the reference probability distribution of
2005 the pooled data. We included this sample in downstream visualizations for reference purposes,
2006 for example, the bar and donut graphs used throughout the manuscript, which display rlog
2007 transformed values on both log2 (bar graphs) and linear (donut graph) scales.

2008 **Differential Gene Expression Analysis.** All differential expression analyses were performed
2009 using DESeq2[13]. The DESEQ2 likelihood ratio test (LRT) was used to identify genes
2010 differentially expressed genes across all cell types (Figure 3—table supplement 5); These genes
2011 were then used to calculate the proportion of differentially expressed genes within each HGNC
2012 families/group in our expression dataset (https://www.genenames.org/data/genegroup).
2013 Differential expression p-values were corrected for multiple hypothesis testing using Benjamini-
2014 Hochberg method across all genes and all 55 possible pairwise contrasts. Volcano plots of the

2015 DESeq2 contrasted cell types were generated using the EnhancedVolcano R/Bioconductor
2016 package[101]

2017 **Principal component analysis.** Principal component analysis was performed on VST
2018 transformed data (subsetting without tissue and bulk cell controls) using the pcaExplorer
2019 R/Bioconductor package[102]. Loading factors from the ends of principal components 1, 2, 3, and 4
2020 were further analyzed using the ClueGo[21] Cytoscape package, referencing the Gene Ontology
2021 biological processes (GOBP) and Reactome pathway databases[18,20]. Subsequent enriched
2022 pathways were grouped automatically by likeness to obtain a condensed list of defining features
2023 for each cell class.

2024 **Gene Set Enrichment and Pathway Analysis.** Pathway enrichment analysis was performed
2025 using Broad Institute's GSEA software (version 4.3.2)[52]. We pre-ranked our individual gene lists
2026 using the DESeq2 calculated Fold changes and Benjamini-Hochberg adjusted *p*-values from all
2027 55 possible pairwise contrasts (Figure 3—table supplement 7). The rank metric was calculated
2028 by: (sign[Fold Change]∗ -$\log_{10}${*p*-value}). GSEA analysis was performed using a classic,
2029 unweighted enrichment statistic (p=0)[52] and was used to evaluate the enrichment of Reactome
2030 pathways[18] Pathways were obtained via the Molecular Signatures Database[54], and we included
2031 pathways with sizes between 15-200 genes. The resulting analysis produced a collection of
2032 enriched pathways and calculated statistics (1000 permutations to correct for multiple hypotheses
2033 testing) for each of the 55 pairwise comparisons. These data became the foundation for the
2034 Excel-based Pathway Analyzer tool.

2035 **Enrichment Map/Cytoscape.** Visualization of enriched gene sets was performed using the
2036 Enrichment Map Cytoscape plugin[52,58], with applied edge cut-offs of 0.5 and FDR cutoffs of 0.1
2037 (10%) for most comparisons, with the exception of the 8v12 comparison between fibroblasts and
2038 adipocytes, which had a FDR of 0.2 (20%). The produced nodes (biological pathways) were
2039 manually adjusted/grouped into themes for presentation purposes using the Reactome hierarchy
2040 as a reference. Annotations of select biological themes and notable genes were added manually.

2041 Supplemental Resources

2042 **Pathway Analyzer:** This supplemental tool integrates regularized log transformed transcript
2043 levels (mRNA-sequencing values) and enriched biological pathways identified by Gene Set
2044 Enrichment Analysis (GSEA). Data from every possible pairwise comparison of cell types is
2045 embedded in the tool, which is supplied as a MS Excel file. The file is approximately 300Mb and
2046 contains computationally intensive algorithms (e.g., we run the file on both Windows and MAC-
2047 based computers (Surface Laptop Studio [Core i7/3.3GHz with 32Gb RAM], Surface Book 2
2048 [Corei7/4.2GHz with 16Gb RAM], Dell Inspiron Desktop [Corei5/3.2GHz with 8Gb RAM],
2049 MacBook Pro [Corei7/2.1GHz with 16Gb RAM]), and we recommend using it on computers
2050 with ≥16Gb RAM).

2051 **Breast Cell Atlas: Transcript Finder:** http://breastcancerlab.com/breast-atlas/ . This online tool
2052 facilitates exploration of mRNAs expressed by the breast cells. When provided individual gene
2053 names (official gene symbol), the tool will display the transcript levels measured in each cell
2054 type. Regularized log transformed values are provided on a log2 and linear scale (bar and donut

graph, respectively), similar to those presented in the many figure elements contained herein. This application was coded using the R/Shiny package[103].

Author Contributions

Conceptualization, WCH; FACS & Methodology, WCH; RNA isolation, WCH; mRNA Library production, ZW; RNA-seq alignment, quality control, and count matrix formation, AC; Bioinformatics, KDT, RS, AC, and WCH; Pathway Analyzer: KDT and WCH; Breast Cell Atlas: Transcript Finder: KDT, WCH; Immunostaining, WCH, KT, KDT, YLM; Writing – Original draft, WCH and KDT; Writing – Review & Editing, WCH, KDT, RS; Funding, WCH.

Acknowledgments

We thank Mina Bissell, Irene Kuhn, Alex Bazarov, Ritu Mukhopadhyay (Lawrence Berkeley Laboratory), Sandy Borowsky (U.C. Davis), Kornelia Polyak (Dana Farber Cancer Institute), Eric Prossnitz (University of New Mexico Health Sciences Center), Curtis Hines (Sandia National Laboratory, retired) for thoughtful scientific discussions and/or for critical review of the manuscript. We thank Tristan Oper, Alice Rindestig, Kremena Karagyozova, Alyssa Cozzo, Gaelen Stanford-Moore, Maria Rojec, Anya Afasisheva, and Sun-Young Lee for their technical assistance. We also thank Dana Pe'er, Ambrose Carr, Zhenmao Wan, and the staff at Columbia University's Genomic Sequencing Core for Illumina RNA-sequencing. Much of the analyses presented herein would not have been possible without Reactome's curation of genes, pathways, and publicly available resources. We thank Reactome's Principal Investigators and the team of dedicated scientists that have contributed to the Reactome Project. We also thank Kerry Wiles and Erik Brooks of the Cooperative Human Tissue Network, and the anonymous women who graciously donated their remnant surgical tissue to science. This project depended on flow cytometry instruments provided at the University of New Mexico Flow Cytometry and High Throughput Screening Resource and we thank Cecelia Rietz and Wade Johnson for their assistance. Finally, a very special thank you to Michelle Scott, Lawrence Berkeley National Laboratory Advanced Microscopy and Flow Cytometry facility, for her expert technical guidance and support.

Funding

Research reported in this publication was supported by the University of New Mexico Cancer Center Support Grant (P30CA118100), an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103451, the U.S. Department of Defense (W81XWH12M9532), and the Breast Cancer Research Foundation.

References

(Note: update information on the adjoining article (#8 Thi.et.al.) with information provided by the editor, once it is accepted for publication)

1. Cox, T.R. The matrix in cancer. *Nature reviews. Cancer* **21**, 217-238 (2021).
2. Maman, S. & Witz, I.P. A history of exploring cancer in context. *Nature Reviews Cancer* **18**, 359-376 (2018).
3. Quail, D.F. & Joyce, J.A. Microenvironmental regulation of tumor progression and metastasis. *Nature medicine* **19**, 1423-1437 (2013).
4. Bissell, M.J. & Hines, W.C. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nature medicine* **17**, 320-329 (2011).
5. Pierce, G.B. Neoplasms, differentiations and mutations. *The American journal of pathology* **77**, 103-118 (1974).
6. Zampieri, F., ElMaghawry, M., Zanatta, A. & Thiene, G. Andreas Vesalius: celebrating 500 years of dissecting nature. *Global cardiology science & practice* **2015**(2015).
7. Androutsos, G., Karamanou, M. & Stefanadis, C. William Harvey (1578-1657): discoverer of blood circulation. *Hellenic J Cardiol* **53**, 6-9 (2012).
8. Thi, K., Del Toro, K., Licon-Munoz, Y., Sayaman, R.W. & Hines, W.C. Comprehensive identification, isolation, and culture of human breast cell types. *bioRxiv* (2022).
9. Hines, W.C. & Del Toro, K. Pathway Analyzer. Supplementary Software for: A cellular and transcriptomic dissection of the human breast for studying mechanisms of cell and tissue function. (2022).
10. Del Toro, K. & Hines, W.C. Breast Cell Atlas: Transcript Finder (R/Shiny App). Supplementary Web Application for: A cellular and transcriptomic dissection of the human breast for studying mechanisms of cell and tissue function. (2022).
11. Hines, W.C., Thi, K., Rojec, M., Stanford-Moore, G. & Bissell, M.J. Abstract B86: A cytometric atlas of the human breast: Comprehensive characterization reveals 12 distinct cell populations. *Cancer research* **75**, B86-B86 (2015).
12. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome research* **8**, 175-185 (1998).
13. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
14. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-1542 (2006).
15. Dyer, L.A. & Patterson, C. Development of the endothelium: an emphasis on heterogeneity. in *Seminars in thrombosis and hemostasis*, Vol. 36 227-235 (© Thieme Medical Publishers, 2010).
16. Mescher, A.L. Author. in *Junqueira's Basic Histology: Text and Atlas, 15e* (McGraw-Hill Education, New York, NY, 2018).
17. Gene Ontology, C. The Gene Ontology resource: enriching a GOld mine. *Nucleic acids research* **49**, D325-D334 (2021).
18. Jassal, B.*, et al.* The reactome pathway knowledgebase. *Nucleic acids research* **48**, D498-D503 (2020).
19. Bindea, G.*, et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093 (2009).
20. Ashburner, M.*, et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

2133 21. Bindea, G.*, et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology
2134 and pathway annotation networks. *Bioinformatics* **25**, 1091-1093 (2009).
2135 22. Kruger, R.P., Aurandt, J. & Guan, K.L. Semaphorins command cells to move. *Nat Rev Mol Cell Biol*
2136 **6**, 789-800 (2005).
2137 23. Yuan, L.*, et al.* Abnormal lymphatic vessel development in neuropilin 2 mutant mice.
2138 *Development* **129**, 4797-4806 (2002).
2139 24. Wang, J.*, et al.* NRP-2 in tumor lymphangiogenesis and lymphatic metastasis. *Cancer letters* **418**,
2140 176-184 (2018).
2141 25. Neil, S. & Bieniasz, P. Human immunodeficiency virus, restriction factors, and interferon. *J*
2142 *Interferon Cytokine Res* **29**, 569-580 (2009).
2143 26. Wang, S.*, et al.* AXL is a candidate receptor for SARS-CoV-2 that promotes infection of
2144 pulmonary and bronchial epithelial cells. *Cell research* **31**, 126-140 (2021).
2145 27. Liu, S., DeLalio, L.J., Isakson, B.E. & Wang, T.T. AXL-mediated productive infection of human
2146 endothelial cells by Zika virus. *Circulation research* **119**, 1183-1189 (2016).
2147 28. Purkis, P.E.*, et al.* Antibody markers of basal cells in complex epithelia. *Journal of cell science* **97 (**
2148 **Pt 1)**, 39-50 (1990).
2149 29. Dairkee, S. & Heid, H.W. Cytokeratin profile of immunomagnetically separated epithelial subsets
2150 of the human mammary gland. *In Vitro Cell Dev Biol Anim* **29A**, 427-432 (1993).
2151 30. Taylor-Papadimitriou, J.*, et al.* Keratin expression in human mammary epithelial cells cultured
2152 from normal and malignant tissue: relation to in vivo phenotypes and influence of medium.
2153 *Journal of cell science* **94 (Pt 3)**, 403-413 (1989).
2154 31. Perou, C.M.*, et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752 (2000).
2155 32. Miner, J.H. & Yurchenco, P.D. Laminin functions in tissue morphogenesis. *Annual review of cell*
2156 *and developmental biology* **20**, 255-284 (2004).
2157 33. Yates, B.*, et al.* Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic acids research*
2158 **45**, D619-D625 (2017).
2159 34. Gray, K.A., Seal, R.L., Tweedie, S., Wright, M.W. & Bruford, E.A. A review of the new HGNC gene
2160 family resource. *Hum Genomics* **10**, 6 (2016).
2161 35. Aumailley, M. The laminin family. *Cell adhesion & migration* **7**, 48-55 (2013).
2162 36. Timpl, R., Sasaki, T., Kostka, G. & Chu, M.L. Fibulins: a versatile family of extracellular matrix
2163 proteins. *Nat Rev Mol Cell Biol* **4**, 479-489 (2003).
2164 37. Khoshnoodi, J., Pedchenko, V. & Hudson, B.G. Mammalian collagen IV. *Microscopy research and*
2165 *technique* **71**, 357-370 (2008).
2166 38. Borza, D.-B.*, et al.* The NC1 domain of collagen IV encodes a novel network composed of the α1,
2167 α2, α5, and α6 chains in smooth muscle basement membranes. *Journal of Biological Chemistry*
2168 **276**, 28532-28540 (2001).
2169 39. Yao, C.C., Breuss, J., Pytela, R. & Kramer, R.H. Functional expression of the alpha 7 integrin
2170 receptor in differentiated smooth muscle cells. *Journal of cell science* **110 (Pt 13)**, 1477-1487
2171 (1997).
2172 40. Hyun, Y.-M., Lefort, C.T. & Kim, M. Leukocyte integrins and their ligand interactions.
2173 *Immunologic research* **45**, 195-208 (2009).
2174 41. Visse, R. & Nagase, H. Matrix metalloproteinases and tissue inhibitors of metalloproteinases:
2175 structure, function, and biochemistry. *Circulation research* **92**, 827-839 (2003).
2176 42. Welgus, H.G.*, et al.* Neutral metalloproteinases produced by human mononuclear phagocytes.
2177 Enzyme profile, regulation, and expression during cellular development. *The Journal of clinical*
2178 *investigation* **86**, 1496-1502 (1990).
2179 43. Lepidi, S.*, et al.* MMP9 production by human monocyte-derived macrophages is decreased on
2180 polymerized type I collagen. *Journal of vascular surgery* **34**, 1111-1118 (2001).

2181 44. Zihni, C., Mills, C., Matter, K. & Balda, M.S. Tight junctions: from simple barriers to
2182 multifunctional molecular gates. *Nat Rev Mol Cell Biol* **17**, 564-580 (2016).
2183 45. Nielsen, M.S.*, et al.* Gap junctions. *Comprehensive Physiology* **2**(2012).
2184 46. Heldin, C.H., Lu, B., Evans, R. & Gutkind, J.S. Signals and Receptors. *Cold Spring Harb Perspect*
2185 *Biol* **8**, a005900 (2016).
2186 47. Westbury, C.B.*, et al.* Genome-wide transcriptomic profiling of microdissected human breast
2187 tissue reveals differential expression of KIT (c-Kit, CD117) and oestrogen receptor-alpha
2188 (ERalpha) in response to therapeutic radiation. *The Journal of pathology* **219**, 131-140 (2009).
2189 48. Moreira, J.M.*, et al.* Tissue proteomics of the human mammary gland: towards an abridged
2190 definition of the molecular phenotypes underlying epithelial normalcy. *Mol Oncol* **4**, 539-561
2191 (2010).
2192 49. Lim, E.*, et al.* Aberrant luminal progenitors as the candidate target population for basal tumor
2193 development in BRCA1 mutation carriers. *Nature medicine* **15**, 907-913 (2009).
2194 50. Cato, A. & Shatkina, L. Nuclear Hormone Receptors: Regulators of Gene Transcription and
2195 Cellular Signaling. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*,
2196 1314-1320 (2006).
2197 51. Frigo, D.E., Bondesson, M. & Williams, C. Nuclear receptors: From molecular mechanisms to
2198 therapeutics. Vol. 65 847-856 (Portland Press Ltd., 2021).
2199 52. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for
2200 interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550
2201 (2005).
2202 53. Mootha, V.K.*, et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are
2203 coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).
2204 54. Liberzon, A.*, et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740
2205 (2011).
2206 55. Coordinators, N.R. Database resources of the National Center for Biotechnology Information.
2207 *Nucleic acids research* **44**, D7-19 (2016).
2208 56. Cunningham, F.*, et al.* Ensembl 2022. *Nucleic acids research* **50**, D988-D995 (2022).
2209 57. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*
2210 **28**, 27-30 (2000).
2211 58. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G.D. Enrichment map: a network-based
2212 method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).
2213 59. Giamas, G.*, et al.* Kinome screening for regulators of the estrogen receptor identifies LMTK3 as a
2214 new therapeutic target in breast cancer. *Nature medicine* **17**, 715-719 (2011).
2215 60. Guo, S. & Sonenshein, G.E. Forkhead box transcription factor FOXO3a regulates estrogen
2216 receptor alpha expression and is repressed by the Her-2/neu/phosphatidylinositol 3-kinase/Akt
2217 signaling pathway. *Molecular and cellular biology* **24**, 8681-8690 (2004).
2218 61. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature*
2219 **490**, 61-70 (2012).
2220 62. Zhu, Y.*, et al.* Coregulation of estrogen receptor by ERBB4/HER4 establishes a growth-promoting
2221 autocrine signal in breast tumor cells. *Cancer research* **66**, 7991-7998 (2006).
2222 63. Sardi, S.P., Murtie, J., Koirala, S., Patten, B.A. & Corfas, G. Presenilin-dependent ErbB4 nuclear
2223 signaling regulates the timing of astrogenesis in the developing brain. *Cell* **127**, 185-197 (2006).
2224 64. Stoker, M., Gherardi, E., Perryman, M. & Gray, J. Scatter factor is a fibroblast-derived modulator
2225 of epithelial cell mobility. *Nature* **327**, 239-242 (1987).
2226 65. Weidner, K.M.*, et al.* Evidence for the identity of human scatter factor and human hepatocyte
2227 growth factor. *Proc Natl Acad Sci U S A* **88**, 7001-7005 (1991).

2228 66. Guimaraes, J.C. & Zavolan, M. Patterns of ribosomal protein expression specify normal and
2229 malignant human cells. *Genome Biol* **17**, 236 (2016).
2230 67. Lin, E.Y., Gouon-Evans, V., Nguyen, A.V. & Pollard, J.W. The macrophage growth factor CSF-1 in
2231 mammary gland development and tumor progression. *Journal of mammary gland biology and*
2232 *neoplasia* **7**, 147-162 (2002).
2233 68. Andreas, N.J., Kampmann, B. & Mehring Le-Doare, K. Human breast milk: A review on its
2234 composition and bioactivity. *Early Hum Dev* **91**, 629-635 (2015).
2235 69. Lonnerdal, B. Nutritional and physiologic significance of human milk proteins. *Am J Clin Nutr* **77**,
2236 1537S-1543S (2003).
2237 70. Bawden, W.S., Passey, R.J. & Mackinlay, A.G. The genes encoding the major milk-specific
2238 proteins and their use in transgenic studies and protein engineering. *Biotechnology and Genetic*
2239 *Engineering Reviews* **12**, 89-138 (1994).
2240 71. Visvader, J.E. Cells of origin in cancer. *Nature* **469**, 314-322 (2011).
2241 72. Frantz, C., Stewart, K.M. & Weaver, V.M. The extracellular matrix at a glance. *Journal of cell*
2242 *science* **123**, 4195-4200 (2010).
2243 73. Skalli, O.*, et al.* A monoclonal antibody against alpha-smooth muscle actin: a new probe for
2244 smooth muscle differentiation. *The Journal of cell biology* **103**, 2787-2796 (1986).
2245 74. Venkatachalam, K. & Montell, C. TRP channels. *Annu Rev Biochem* **76**, 387-417 (2007).
2246 75. Kufe, D.W. Mucins in cancer: function, prognosis and therapy. *Nature reviews. Cancer* **9**, 874-
2247 885 (2009).
2248 76. Schroten, H.*, et al.* Inhibition of adhesion of S-fimbriated Escherichia coli to buccal epithelial
2249 cells by human milk fat globule membrane components: a novel aspect of the protective
2250 function of mucins in the nonimmunoglobulin fraction. *Infect Immun* **60**, 2893-2899 (1992).
2251 77. Labrie, F.*, et al.* The key role of 17β-hydroxysteroid dehydrogenases in sex steroid biology.
2252 *Steroids* **62**, 148-158 (1997).
2253 78. Fuchs, E. & Weber, K. Intermediate filaments: structure, dynamics, function, and disease. *Annu*
2254 *Rev Biochem* **63**, 345-382 (1994).
2255 79. Simionescu, M., Simionescu, N., Silbert, J.E. & Palade, G.E. Differentiated microdomains on the
2256 luminal surface of the capillary endothelium. II. Partial characterization of their anionic sites. *The*
2257 *Journal of cell biology* **90**, 614-621 (1981).
2258 80. Rosenberg, R.D., Shworak, N.W., Liu, J., Schwartz, J.J. & Zhang, L. Heparan sulfate proteoglycans
2259 of the cardiovascular system. Specific structures emerge but how is synthesis regulated? *The*
2260 *Journal of clinical investigation* **99**, 2062-2070 (1997).
2261 81. van Buul, J.D.*, et al.* RhoG regulates endothelial apical cup assembly downstream from ICAM1
2262 engagement and is involved in leukocyte trans-endothelial migration. *The Journal of cell biology*
2263 **178**, 1279-1293 (2007).
2264 82. Arnal, J.-F.*, et al.* Estrogen receptors and endothelium. *Arteriosclerosis, Thrombosis, and*
2265 *Vascular Biology* **30**, 1506-1512 (2010).
2266 83. Tatematsu, M.*, et al.* Raftlin controls lipopolysaccharide-induced TLR4 internalization and
2267 TICAM-1 signaling in a cell type–specific manner. *The Journal of Immunology* **196**, 3865-3876
2268 (2016).
2269 84. Hauser, M.A. & Legler, D.F. Common and biased signaling pathways of the chemokine receptor
2270 CCR7 elicited by its ligands CCL19 and CCL21 in leukocytes. *Journal of leukocyte biology* **99**, 869-
2271 882 (2016).
2272 85. Moreira, L.O. & Zamboni, D.S. NOD1 and NOD2 signaling in infection and inflammation. *Frontiers*
2273 *in immunology* **3**, 328 (2012).
2274 86. Jackson, W.F. Potassium channels in the peripheral microcirculation. *Microcirculation* **12**, 113-
2275 127 (2005).

87. Gupta, R.K. Adipocytes. *Curr Biol* **24**, R988-993 (2014).
88. Storch, J. & Thumser, A.E. Tissue-specific functions in the fatty acid-binding protein family. *The Journal of biological chemistry* **285**, 32679-32683 (2010).
89. Li, Y.*, et al.* Cidec differentially regulates lipid deposition and secretion through two tissue-specific isoforms. *Gene* **641**, 265-271 (2018).
90. Chen, L.*, et al.* AQP7 mediates post-menopausal lipogenesis in adipocytes through FSH-induced transcriptional crosstalk with AP-1 sites. *Reproductive BioMedicine Online* **41**, 1122-1132 (2020).
91. Blüher, M. Importance of estrogen receptors in adipose tissue function. *Molecular metabolism* **2**, 130 (2013).
92. Winkler, E.A., Sagare, A.P. & Zlokovic, B.V. The Pericyte: A Forgotten Cell Type with Important Implications for A lzheimer's Disease? *Brain pathology* **24**, 371-386 (2014).
93. Armulik, A., Genové, G. & Betsholtz, C. Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Developmental cell* **21**, 193-215 (2011).
94. Rapport, M.M. Serum vasoconstrictor (serotonin) the presence of creatinine in the complex; a proposed structure of the vasoconstrictor principle. *The Journal of biological chemistry* **180**, 961-969 (1949).
95. Goh, L.K. & Sorkin, A. Endocytosis of receptor tyrosine kinases. *Cold Spring Harb Perspect Biol* **5**, a017459 (2013).
96. Wolfe, B.L. & Trejo, J. Clathrin-dependent mechanisms of G protein-coupled receptor endocytosis. *Traffic* **8**, 462-470 (2007).
97. Beale, E.G., Harvey, B.J. & Forest, C. PCK1 and PCK2 as candidate diabetes and obesity genes. *Cell Biochem Biophys* **48**, 89-95 (2007).
98. Nguyen, Q.H.*, et al.* Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nature Communications* **9**, 1-12 (2018).
99. Wu, S.Z.*, et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics* **53**, 1334-1347 (2021).
100. Hines, W.C., Su, Y., Kuhn, I., Polyak, K. & Bissell, M.J. Sorting Out the FACS: A Devil in the Details. *Cell reports* **6**, 779-781 (2014).
101. Blighe, K., Rana, S. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. *R package version* **1**(2019).
102. Marini, F. & Binder, H. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics* **20**, 331 (2019).
103. Chang, W.*, et al.* shiny: Web Application Framework for R. R package version 1.6. 0. 2021. *URL R-project. org/package= shiny* (2021).

Figure 3—Table supplement 8. Antibodies

Tissue Staining (Immunofluorescence)				
Antibody specificity	**Clone**	**Conjugated**	**Source**	**Dilution**
Alpha-Smooth Muscle Actin	1A4	No	Mouse IgG2a, κ, Lab Vision/Neomarkers MS-113-P0[‖]	1:100
CD29	TS2/16	A488	Mouse IgG1, κ; Biolegend 303015	½ assay[¶]
CD49f	GoH3	No	Rat IgG2a, κ; Biolegend 313602	½ assay[¶]
E-Cadherin (CD324)	24E10	No	Rabbit mAb; Cell Signaling 3195S	1:100
E-Cadherin (CD324)	67A4	A488	Mouse IgG1, κ; BioLegend 324110	½ assay[¶]
erbB2/Her2 (CD340)	24D2	PE	Mouse IgG1, κ; Biolegend 324406	½ assay[¶]
Keratin 14	Polyclonal	No	Rabbit Polyclonal; Thermo/Lab Vision RB-9020-P1	1:400
Keratin 14	Polyclonal	No	Chicken Polyclonal; Biolegend 906001	1:400
Keratin 18	DC10	No	Mouse IgG1; Thermo/Lab Vision MS142P0	1:400
Keratin 19	A53-B/A2.26; same as Ks19.1	No	IgG2a/λ; Lab Vision/Neomarkers MS-198-PABX	1:400
Keratin 8		No	Mouse IgG2a κ	1:400
P63	4A4	None	Abcam ab32353	1:100
Pan-Laminin	Polyclonal	No	Rabbit Polyclonal; L9393	1:400
Pan-Laminin	Polyclonal	No	Rabbit Polyclonal; L9393	1:400
Podoplanin	NC-08	No	Rat IgG2a, λ	½ assay[¶]
Facs Panel Antibodies / Viability Dyes				
CD49f (alpha-6 integrin)	GoH3	BV421[*]	BioLegend 313624, 313612, 313607	2.5µl[§]
CD24	ML5	BV605[*]	BioLegend 311124, 311106, 311108	2.5µl[§]
CD227 (Muc1)	16A	PE	BioLegend 355604,	4µl[§]
Podoplanin	NC-08	A488[†]	BioLegend 337006, 337004	2.5µl[§]
CD45	HI30	A488[†]	BioLegend 304052, 304032	3µl[§]

CD90 (Thy1)	5E10	PE-Cy5	BioLegend 328112, 328134	3µl§
CD34	581	PE-Cy7	BioLegend 343516	1-2µl§
CD10	HI10a	APC-Cy7	BioLegend 312212, 312204	3µl§
TO-PRO™-3 Iodide	N/A	(642/661)	Thermo (T3605)	1:4000

** Brilliant Violet; † Alexa Fluor; (Biolegend);* § 1x volume per $6x10^6$ cells stained in FACS experiments (up to $36x10^6$, after which cells received 6x volumes; max stained cells $=120x10^6$); titration required and must be matched to cytometer. ¶ Percentage of manufacturer's suggested assay volume per 100ul volume (antibody is prediluted).

List of Table supplements and Data	
Figure 3—table supplement 1	Sample Treatment.xlsx
Figure 3—table supplement 2	CountMatrix_rLog_VST.xlsx
Figure 3—table supplement 3	PCA_Loading_Factors.xlsx
Figure 3—table supplement 4	GO from PCAs.xlsx
Figure 3—table supplement 5	DESeq2_LRT.xlsx
Figure 3—table supplement 6	Gene Families_DE Genes.xlsx
Figure 3—table supplement 7	DESeq2_files.xlsx
OSF; https://osf.io/gfhc3/	Cytoscape/Enrichment Map files (1v2, 2v3, 2v9, 6v8)
OSF; https://osf.io/hpb9t/	Pathway Analyzer[9]
http://breastcancerlab.com/breast-atlas	Breast Cell Atlas: Transcript Finder (R/Shiny App)[10]

2312 ## Figure Legends

2313 **Figure 1. FACS gating strategy for comprehensive cell type isolation. (a)** Prior to FACS
2314 isolation, breast tissues were enzymatically digested with collagenase. This was followed by
2315 low-speed centrifugation, which yielded an organoid pellet, supernatant, and an adipocyte-
2316 enriched lipid layer. RNA was isolated from the lipid layer (adipocytes, Pop12) and stored for
2317 later sequencing. The remaining cell types were FACS purified from the trypsin-dissociated
2318 organoid cell pellet, according to the outlined gating scheme. Each cell type was assigned a
2319 population number that corresponds to each population's isolation strategy (Pops 1-12). **(b)**
2320 Relative abundance of each cell type from 30 different individuals (proportion of total cell
2321 counts, as measured by FACS analysis of 44 independent FACS experiments). Outliers (*) are
2322 defined as points falling (1.5 x IQR) above/below the 75th and 25th quartile.

2323 **Figure 1—figure supplement 1. FACS strategy designed for an 8 channel FACS machine**
2324 **(BD FACS Vantage). (a)** FACS gating scheme used to isolate the RNA-sequenced cell types
2325 (antibody panel designed for an 8 channel FACS machine- BD FACS Vantage). Representative
2326 FACS data are of cells derived from disease-free reduction mammoplasty tissue surgically
2327 resected from a 22-year-old female (sample #N239).

2328 **Figure 2. IHC staining of cell types in the breast. (a)** Representative 4x scan of normal breast
2329 tissue (reduction mammoplasty tissue from 23-year-old female). The general location of the 12
2330 cell types are indicated by the their FACS population identification numbers, Pops1-12. **(b)**
2331 Chart of the 12 breast cell types and their assigned colors (used throughout this article). Cell
2332 types are grouped by their common histological classification: epithelial, mesenchymal, and
2333 blood cell types. Epithelial types are further subdivided into glandular and endothelial types,
2334 whereas the mesenchymal perivascular types are indicated (pericytes and vascular smooth
2335 muscle cells). **(c-f)** spatial position of the cell types in normal tissue immunostained with: **(c)**
2336 keratin 18 (green; luminal epithelial cells) and pan-laminin (red; epithelial and vascular basement
2337 membrane, and adipocytes). Staining reveals a collecting duct within several breast lobules
2338 (center) composed of epithelial (Pops 1-4), fibroblasts (Pop8), and capillaries that consist of
2339 pericytes (Pop6) and vascular endothelial cells (Pop9). Stromal adipocytes are also present
2340 (Pop12, upper-right). Scale = 1000μm. **(d)** podoplanin (green) and pan-laminin (red) staining
2341 reveal podoplanin-staining lymphatic vessels (Pop7) and vascular blood vessels (Pop9). Laminin
2342 embedded pericytes (Pop6) are also discernible on the capillaries; Scale = 200μm; **(e)** CD34
2343 (red) and pan-laminin (green) staining reveals a cross-section of three-and a-half breast lobules
2344 within a bed of $CD34^{Pos}$ fibroblasts. Endothelial cells forming capillaries and larger blood
2345 vessels also stain for CD34, both inside and outside the lobular areas, e.g., larger vessels are
2346 visible at lower right, along with pan-laminin staining adipocytes (Pop12); Scale = 200μm **(f)**
2347 CD49f (green) and pan-laminin (red) reveals a large blood vessel (center) surrounded by
2348 vascular smooth muscle cells (Pop5). Endothelial cells (Pop9) and pericytes (Pop6) of smaller
2349 capillaries are present, as are lymphatic vessels (Pop7) and adipocytes (Pop12); Scale = 200μm.

2350 **Figure 2—figure supplement 1. (a)** High-resolution image of the immunostained breast tissues
2351 provided in Fig. 2a. **(b)** magnified area indicated by white box in a (on left) containing several
2352 breast lobules (a.k.a, terminal ductal lobular units). **(c)** magnified area indicated by white box in

2353 a (on right) containing the cross section of a large lactiferous duct. The duct is surrounded by
2354 numerous blood vessels running parallel to the duct and is embedded in a collagenous-rich
2355 matrix.

2356 **Figure 2- Figure supplement 2. (a,b)** High-resolution (label-free) images of the immunostained
2357 breast tissues provided in Figs 2c,d.

2358 **Figure 2- Figure supplement 3. (a,b)** High-resolution (label-free) images of the immunostained
2359 breast tissues provided in Figs 2e.f.

2360 **Figure 3. Distinguishing the traits of each breast cell type.** Shown is a sketch of the RNA-
2361 sequencing workflow used for analyzing the eleven sequenced cell types. This began with **(a)**
2362 developing a FACS isolation method for separating each breast cell type (as detailed in Fig.1 and
2363 adjoining article). This procedure was scaled up and cells were sorted and RNA-sequenced.
2364 Analysis of the RNA-sequencing data involved: **(b)** validation and exploration of cellular
2365 relationships by hierarchical clustering, **(c)** covariance matrix and quality control analyses,
2366 identification of unique cell markers, and principal component analysis (PCA). PCA-derived
2367 gene lists were used to perform gene ontology enrichment analyses that revealed the shared
2368 characteristics of cell lineage. **(d)** Differential gene expression analysis was performed and used
2369 to identify gene families that varied most among breast cell types; Pairwise comparisons
2370 identified transcripts whose levels varied between sample pairs. Differentially expressed genes
2371 between cell pairs were ranked and analyzed by **(e)** Gene Set Enrichment Analysis, leading to the
2372 identification of key pathways and processes that differ between cell types. GSEA results were
2373 further analyzed by **(f)** Constructing and analyzing enrichment maps that identified the central
2374 biological themes and distinguishing features of each cell type. Analyzing the data required the
2375 development of software code and packages that are available for download: Pathway Analyzer
2376 (Open Science Framework, https://osf.io/), and online Transcript Finder
2377 (http://breastcancerlab.com/breast-atlas).

2378 **Figure 3—figure supplement 1. Project workflow.** Diagram of the sequential processes and
2379 analyses presented in the adjoining and current articles. The adjoining article (in green) describes
2380 the development and validation of a FACS isolation strategy for separating all twelve major
2381 breast cell types, culminating with the development of primary cell models for nine major breast
2382 cell types. The current article (in blue) explores the nature of cell type through bulk-mRNA-
2383 sequencing and analysis of transcript levels measured in each FACS-purified population.
2384 Analysis included principal component analysis, differential gene expression, and gene set
2385 enrichment analysis. Red-outlined boxes indicate the major products from each publication.

2386 **Figure 3—table supplement 1. Sample information/Meta Data.** Table of sample information,
2387 containing age/sex/race information supplied with the acquired tissues (B=black; W=white),
2388 Patient number, FACS gate, cell type, cell yields, RNA-isolation strategy/kit, elution volumes,
2389 nanodrop concentrations, DNase treatment used, Bioanalyzer results (concentration and integrity
2390 number), RNA-sequencing batch, and calculated reads in each sample.

2391 **Figure 3—table supplement 2.** RNA-sequencing count matrix.

2392 **Figure 3—table supplement 3.** PCA loading factors. Sheet 1, 'All Pops,' contains loading
2393 factors calculated from the analysis of all cell types (Pops 1-12). Sheet 2, 'Endothelial
2394 Removed,' contains loading factors calculated from the analysis of non-endothelial cell types
2395 (Pops 1-7,8,10-12).

2396 **Figure 3—table supplement 4.** Tabulated Pathway Enrichment Results
2397 (ClueGo/Reactome/Gene Ontology Biological Processes) that used the 200 loading factors
2398 associated with the four cell lineages: a) Epithelial end of PC1, b) Leukocyte end of PC2, c)
2399 Mesenchymal end of $PC1^{No\ Endo}$, d) Endothelial end of PC3, and the perivascular and
2400 myoepithelial lineage subtypes e) Perivascular cell end of PC3, and f) myoepithelial end of PC4.

2401 **Figure 3—table supplement 5.** RNA-Seq DESeq2 Likelihood Ratio Test (LRT) statistics, p-
2402 values, and Benjamini Hochberg adjusted *p*-values for differentially expressed genes with LRT
2403 adj.*p*-val<0.05. These results were used to calculate the proportion of differentially expressed
2404 genes in each HGNC gene family.

2405 **Figure 3—table supplement 6.** Table of the 1,421 HGNC Gene families, number of genes in
2406 each family, and the calculated number of differentially expressed genes in each family (using
2407 the DESeq2 Benjamini Hochberg Adj. *p*-value threshold of ≤ 0.1).

2408 **Figure 3—table supplement 7.** Excel workbook containing 55 DESeq results (on separate
2409 tabulated worksheets) from each possible pairwise contrast of the 11 seqenced breast cell types;
2410 e.g. Pop1v2 tab contains results from the contrast of Pop1 (ER^{Pos} luminal epithelial cells) and
2411 Pop2 (ER^{Neg} luminal epithelial cells). Tab 1v2 also contains the calculated rank metric used for
2412 GSEA analysis; i.e., $sign([Fold\ Change] * -log_{10}\{p\text{-}value\})$.

2413 **Figure 3—table supplement 8. Antibodies.** Antibody clones used for immunostaining and
2414 FACS.

2415 **Figure 4. RNA-seq quality assessment. (a)** Sample metadata for the mRNA-sequenced cell
2416 populations. **(b)** Covariance matrix displaying the correlation of gene expression for all pairwise
2417 combinations of samples. Samples N228-Pop4, N225-Pop4, N228-Pop5, N228-Pop7 have been
2418 flagged for lower correlations between replicates (*). These cell populations are the least
2419 abundant types in the breast and are among those that produced the lowest FACS cell yields **(c)**
2420 distribution of regularized log transcript levels in each sample. **(d)** Sample correlation matrix
2421 demonstrates absence of sample/batch effects.

2422 **Figure 4—figure supplement 1. mRNA levels correspond to FACS staining. (a)** Normalized
2423 mRNA values for genes encoding FACS markers used to purify cell types from tissues (and
2424 CD36, which was later used to purify cultured adipocytes). Transcript levels are provided on a) a
2425 log2 scale (bar graphs containing each biological replicate) and b) a linear scale (donut graphs of
2426 the replicate's median value), which are both color-coded by cell type. **(b)** Schematic indicating
2427 relative FACS staining levels of each marker in the twelve breast cell types. "Neg" indicates
2428 relative negative staining (i.e., compared to other populations, not absolute negativity), "-/+"
2429 indicates near background level staining, and the increasing number of "+" indicates greater
2430 staining intensity.[†] CD36 is a marker used to purify cultured adipocytes but was not used to

2431 purify adipocytes prior to RNA-sequencing in this analysis. **(c)** Normalized transcript levels for
2432 the commonly used internal control gene GAPDH (rlog, DEseq2) are presented on a log2 scale
2433 (bar graph) and linear scale (donut graph of median value).

2434 **Figure 5. Uniquely expressed genes.** A sliding threshold (fold-change) was used to identify the
2435 top 20 most uniquely expressed genes in each of the eleven sequenced cell types. Notable genes
2436 included: **a)** albumin (ALB) and connexin 30.3 (GJB4) in ER-positive luminal cells, **(b)**
2437 lactotransferrin (LTF) in ER-negative luminal cells, **(c)** alpha-1 chain of collagen XI (COL11A1)
2438 and notch ligand delta-like 3 (DLL3) in myoepithelial cells, **(d)** protocadherin 20 (PCDH20) in
2439 Pop4 epithelial cells, **(e)** lymphatic vessel endothelial hyaluronan receptor 1 (LYVE1) and
2440 prospero homeobox 1 (PROX1) in lymphatic endothelial cells, **(f)** endothelial selectin (SELE)
2441 and delta like canonical notch ligand 4 (DLL4) in vascular endothelial cells, **(g)** calcitonin
2442 related polypeptide alpha (CALCA) and 5-hydroxytryptamine receptor 1b (HTR1B) in vascular
2443 smooth muscle cells, **(h)** desmin (DES) and 5-hydroxytryptamine receptor 1f (HTR1F) in
2444 pericytes, **(i)** alpha-1 chain of collagen I (COL1A1) and wnt ligand family member 2 (WNT2) in
2445 fibroblasts, **(j)** the lipid-binding protein 'cell death inducing dffa-like effector c' (CIDEC) and
2446 glycerol-3-phosphate dehydrogenase 1 (GPD1) in adipocytes, and **(k)** CD7 and fas ligand
2447 (FASLG) in leukocytes.

2448 **Figure 5—figure supplement 1. CD antigen genes.** Median mRNA expression levels of genes
2449 encoding CD antigens (VST, log2 scale)**.** The genes are arranged by the cell type that expressed
2450 them at their highest level, as indicated by the background shading specific to each cell type.
2451 Genes differentially expressed across cell populations are annotated with (*). CD antigen genes
2452 in the uncolored panel were not expressed by any breast cell type.

2453 **Figure 6. Transcript levels of top 2000 most differentially expressed genes.** Heatmap with
2454 dendrogram of the top 2000 most differentially expressed genes across all FACS sorted
2455 populations, as well as unsorted cells and matched breast tissues (mixed cell controls). Clustering
2456 was performed using Euclidean distance measures and Ward clustering method on variance
2457 stabilized expression values (DESeq2; Variance Stabilized Transformation, VST). Adjusted p-
2458 values for these 2000 differentially expressed genes ranged between 4.9e-97 to 0.0017. All 2000
2459 genes symbols can be viewed in Figure 3—table supplement 5 (when the data is ordered by adj.
2460 *p*-value).

2461 **Figure 7. Principal Component Analysis (a)** projection of breast cell types on principal
2462 components 1 and 2 (PC1, PC2). PC1 separates the luminal ER^{Pos} and ER^{Neg}, Pop4 epithelial,
2463 and myoepithelial cell types (epithelial lineages) from the mesenchymal (fibroblasts, adipocytes,
2464 pericytes and vSMCs) and endothelial (lymphatic and vascular) lineages. PC2 separates the
2465 $CD45^{Pos}$ leukocytes (mesodermal lineage) from all other cell types. The cell types composing
2466 each of the epithelial, mesenchymal, endothelial, and mesodermal cell-lineages are indicated
2467 (dashed circles). **(b)** Median transcript levels of the 50 loading factors (genes) that contribute to
2468 the negative end of PC1(bottom end of the list, rlog/log2 scale). **(c)** Median mRNA expression of
2469 the 50 loading factors (genes) contributing to the negative end of PC2 (bottom end of the list)
2470 that separates the leukocyte samples (rlog, log2 scale). Stars (★) denote commonly used markers
2471 for luminal cells (blue), myoepithelial cells (red), and leukocytes (pink).

2472 **Figure 7—figure supplement 1. Variance captured by principal components.** A screeplot
2473 shows the sample variances captured by each principal component. 71% of the total variance is
2474 captured by the first four components.

2475 **Figure 7—figure supplement 2. PCA with endothelial cells removed. (a)** PCA projection of
2476 breast cell types, after removal of the lymphatic and vascular endothelial cells, resolves the
2477 mesenchymal cell types on $PC1^{NoEndo}$. **(b)** Median mRNA expression of the 50 loading factors
2478 (genes) contributing to the positive end of PC2 (top end of the list; rlog, log2 scale). Stars (★)
2479 denote common markers of pericytes (purple) and fibroblasts (yellow).

2480 **Figure 8. PCA plot of PC3 and PC4. (a)** PCA projection of breast cell types on principal
2481 components 3 and 4 (PC3, PC4). PC3 separates the endothelial (most positive) and perivascular
2482 (most negative) cell types, whereas PC4 separates the myoepithelial and myoepithelial-related
2483 cell types (myoepithelial and Pop4 epithelial cells). Median mRNA expression of the 50 loading
2484 factors contributing to the **(b)** positive end of PC3 (top of the list; rlog, log2 scale), **(c)** negative
2485 end of PC3 (bottom of the list; rlog, log2 scale), and **(d)** the positive end of PC4 (top of the list;
2486 rlog, log2 scale). Stars (★) denote common markers of endothelial cells (orange), myoepithelial
2487 cell (red), fibroblasts (yellow), pericytes (purple), and luminal cells (blue).

2488 **Figure 9. IHC validation of PCA markers.** Tissue immunostaining of normal human breast
2489 tissue with markers contained in the principal components. **(a)** keratin 19 (green) and pan-
2490 laminin (red); **(b)** keratin 8 (red); **(c)** e-cadherin (green) and alpha smooth muscle actin (red); **(d)**
2491 Her2 (green) and keratin 14 red); **(e)** CD49f (green) and vimentin (red); **(f)** alpha smooth muscle
2492 actin (green) and pan-laminin (red); **(g)** CD49f (green) and e-cadherin (red); and **(h)** p63 (green)
2493 and keratin 8 (red). Loading factors at the negative end of PC1 (epithelial end) included: keratin
2494 19/Krt19 (#5 on list), keratin 8/KRT8 (#6), E-cadherin/CDH1 (#22), Her2/ERBB2 (#85), and
2495 keratin 18/KRT18 (#138), and CD49f/ITGA6 (#5574). Loading factors at the positive end of
2496 $PC1^{No\ Endo}$ (mesenchymal end) included: smooth muscle actin/ACTA2 (#155 on the list), and
2497 Vimentin/VIM (#192). Loading factors at the negative end of PC3 (perivascular end) included:
2498 smooth muscle actin/ACTA2 (#7 on the list). Loading factors at the positive end of PC3
2499 (endothelial end) included: CD49f/ITGA6 (#344). Loading factors at the positive end of PC4
2500 (myoepithelial end) included: smooth muscle actin/ACTA2 (#7 on the list), P63/TP63 (#25) and
2501 CD49f/ITGA6 (#68).

2502 **Figure 10. Pathway analysis of epithelial genes (PC1). (a)** Grouped GO/Reactome terms
2503 overrepresented within the gene list 200 genes contributing to the negative/epithelial end of PC1
2504 (bottom of the list) are ordered by adjusted p-value. Grouping was done by the similarity of
2505 terms to evaluate broad differences in cell class. Associated genes for each group are listed in the
2506 4th column. Median mRNA expression (rlog, log2 scale) for PC1 genes associated with select
2507 pathways including **(b)** c*ell adhesion and binding*, **(c)** *peptidase activity*, **(d)** *keratinization*, and
2508 **(e)** *cytoskeletal binding*.

2509 **Figure 10—figure supplement 1. Pathway analysis of leukocyte genes (PC2). (a)** Grouped
2510 GO/Reactome terms associated with the 200 genes contributing to the positive end (bottom of
2511 list) of PC1 ordered by adjusted p-value. Median mRNA expression (rlog transformed) for PC2

genes associated with select pathways, including **(b)** a*daptive immune response*, **(c)** *immune response activation*, and **(d)** *leukocyte homeostasis*.

Figure 11. Pathway analysis of mesenchymal genes (PC1$^{\text{No Endo}}$). (a) Grouped GO/Reactome terms overrepresented within the gene list 200 genes contributing to the positive/mesenchymal end of PC1^{NoEndo}, ordered by adjusted p-value. Median mRNA expression (rlog, log2 scale) for PC1^{NoEndo} genes associated with select pathways, including **(b)** *cardiac tissue development*, **(c)** *fibroblast migration*, **(d)** *regulation of peptidyl-tyrosine phosphorylation*, and **(e)** *chondrocyte/skeletal development*.

Figure 12. Pathway analysis of Endothelial genes (PC3). (a) Grouped GO/Reactome terms associated with the 200 genes contributing to the positive end of PC3 (top of list), ordered by adjusted p-value. Median mRNA expression (rlog, log2 scale) for PC3 genes associated with select pathways including **(b)** *angiogenesis*, **(c)** *cell motility*, **(d)** c*ell-matrix adhesion*, and **(e)** *endothelial cell differentiation*.

Figure 12—figure supplement 1. Transcript levels of genes composing the *signaling by SCF-KIT* pathway. Transcript levels for each gene in the Reactome pathway s*ignaling by SCF-KIT.* VST gene levels are relative (normalized within each row of the heatmap). Boxed areas highlight expression of select genes in lymphatic and vascular endothelial, and ER^{Neg} luminal epithelial cell types.

Figure 12—figure supplement 2. Transcript levels of genes composing the *Semaphorins* gene family. (a) Transcript levels of genes in Semaphorin gene family. VST gene levels are relative (normalized within each row of the heatmap). Boxed areas highlight select genes highly expressed by the lymphatic and vascular endothelial cell types. Normalized mRNA values of **(b)** NRP1 and **(c)** NRP2 (rlog, DEseq2) are provided on log2 scale (bar graph of each biological replicate) and linear scale (donut graph of median value), which are both color-coded by cell type.

Figure 13. Pathway analysis of perivascular genes (PC3). (**a)** Grouped GO/Reactome terms associated with the 200 genes contributing to the negative end of PC3 (bottom of the list), ordered by adjusted p-value. Median mRNA expression (rlog, log2 scale) for PC3 genes associated with select pathways, including **(b)** *epithelial tube morphogenesis*, **(c)** *blood vessel development*, **(d)** *muscle contraction*, and **(e)** *extracellular matrix organization*.

Figure 14. Pathway analysis of myoepithelial-related genes (PC4). (a) Grouped GO/Reactome terms associated with the 200 genes contributing to the positive end of PC4 (top of the list), ordered by adjusted p-value. Median mRNA expression (rlog, log2 scale) for PC4 genes associated with select pathways, including **(b)** *extracellular matrix organization*, **(c)** *cell junction organization*, **(d)** *keratinization*, and **(e)** *morphogenesis of branching epithelium*.

Figure 15—table 1. Variably expressed gene types. Genes composing each gene family (HGNC) were assessed by DESeq2 (Adj.*p*-value ≤ 1, across all cell types). The number of genes in each family, along with the numer of those found to be differentially expressed are provided (for gene families ≥ 6 genes).

2551 **Figure 15. Gene types contributing to cell-specific features.** Of the 1,519 types of genes (gene
2552 families) defined by HGNC, we found 1,421 gene families (of all sizes) expressed within the
2553 breast cells transcriptomes. **(a)** Distribution of the proportions of differentially expressed (DE)
2554 genes in each gene family (BH Adj. *p*-value ≤ 0.1). The proportion of DE genes in these families
2555 ranged between 0-100%, which we grouped into different bins: 0-24% (blue, n=209), 25%-74%
2556 (green, n=814), 75% to 99% (orange, n=211), and 100% (red, n=187). **(b-e)** Median transcript
2557 levels (rlog, log2 scale) across all cell types of genes composing select families with **(b)** 100%
2558 DE genes, **(c)** 75%-99% DE genes, **(d)** 25-74% DE genes, and **(e)** 0-24% DE genes.

2559 **Fig.15—figure supplement 1. Distribution of gene family size.** The sizes of the HGNC
2560 families range from 1 to 831 genes. The distribution of these family sizes is provided, along with
2561 a box & whisker plot (top). The median family size is six genes per family.

2562 **Fig.15—figure supplement 2. Transcript levels of DE genes.** Transcript levels for select Type
2563 IV collagen chains: **(a)** COL4A1, **(b)** COL4A2, and **(c)** COL4A4. **(d)** Median mRNA expression
2564 (rlog, log2 scale) for integrin alpha subunits. The dashed circles identify notable differentially
2565 expressed genes, ITGA6, ITGAE, and ITGAM. **(e)** Transcript levels of the laminin binding
2566 integrins. The predominant laminin-binding integrins are αβ heterodimers of: α3β1, α7β1, α7β1
2567 and α6β4. These are integrins are encoded by ITGB1, ITGA3, ITGA7, ITGA6, and ITGB4.
2568 Collagen-binding integrins that also show laminin binding properties include heterodimers of:
2569 α1β1, α2β1, and α10β1 (encoded by ITGA1, ITGA2, ITGA10, and ITGB1). ITGA7 is
2570 predominantly limited to the perivascular cell types, whereas ITGA3, ITGA6, and ITGB4 are
2571 largely expressed by the epithelial cell types.

2572 **Figure 16. Tight Junctions (a)** Heatmap displaying transcript levels of genes belonging to the
2573 claudin gene family (VST, normalized by row). Dashed box indicates claudins that were not
2574 appreciably expressed by any cell type. Dark boxes highlight genes differentially expressed by
2575 luminal epithelial cells, endothelial cells, and myoepithelial cell types. **(b)** median rlog levels of
2576 the claudin family genes. Dashed circles identify genes outlined in the heat map. **(c-e)** Transcript
2577 levels for the zona occludens tight junction proteins: **(c)** TJP1, **(d)** TGP2, and **(e)** TJP3 (rlog
2578 values on log2 (bar graph) and linear (donut graph) scales).

2579 **Figure 16—figure supplement 1. IHC staining. (a)** Claudin-1 staining of normal breast tissue
2580 (reduction mammoplasty of 28-year-old female). (▼) Claudin-1 staining is found confined to the
2581 apical surface of the paracellular space of luminal epithelial cells, circumscribing the luminal
2582 surface of the cells. Scale = 50 μm.

2583 **Figure 17. Gap junction proteins. (a)** Heatmap displaying transcript levels of genes encoding
2584 gap junction proteins (VST, normalized by row). Boxed areas highlight the cell types expressing
2585 select gap junction proteins. Dashed boxes identify genes not appreciably expressed by any cell
2586 type, as evidenced by the rlog values shown in b. **(b)** Median rlog transcript levels of gap
2587 junction genes on a log2 scale; differentially expressed genes are encircled.

2588 **Figure 18. Receptor tyrosine kinases. (a)** Heatmap displaying transcript levels of genes
2589 belonging to the tyrosine kinase receptor gene family (VST, normalize by row). Boxed areas

indicate cell types expressing highest levels of select kinases. **(b-m)** Transcript levels of the select RTKs (rlog values on log2 (bar graph) and linear (donut graph) scales).

Figure 19. Transcript levels of nuclear hormone receptors. (a) Heatmap displaying transcript levels of genes belonging to the nuclear hormone receptor gene family (VST, normalized by row). Boxed areas indicate high/low gene expression in different cell types, which are highlighted below by the donut plots. Select genes include **(b-m)** mRNA values (rlog) are provided on log2 scale (bar graph of each biological replicate) and linear scale (donut graph of median value) which are both color-coded by cell type.

Figure 20. Differentially expressed genes across all cell type pairs. The number of differentially expressed genes between cell-type pairs, as reported by DESeq2; Benjamini-Hochberg adjusted *p*-values ≤ 0.1.

Figure 21. Differences between ER^{Pos} and ER^{Neg} luminal epithelial cell types. A contrast of expressed transcripts between the two luminal epithelial cell types revealed 1,301 differentially expressed genes (DeSeq2/Benjamini-Hochberg adjusted p-value ≤ 0.1). All expressed genes are displayed. The horizontal dashed lines (separating red and green dots) represents an Adj. *p*-value of 0.1. Established cell type markers (of each cell type) are denoted. mRNA expression values for ten genes (indicated by filled, color-coordinated boxes) are shown at left and right of the volcano plot. Color-coded stars in the bar graphs indicate the cell type being contrasted that had the higher level of that gene; i.e., ER^{Pos}(light blue) and ER^{Neg} luminal cells (dark blue).

Figure 22. Inspection of GSEA results and the validating example of the *estrogen dependent gene expression* pathway. (a) Gene Set Enrichment Analysis (GSEA) was performed using a list of genes expressed by ER^{Pos} and ER^{Neg} luminal epithelial cells, ranked by the statistical significance of their differential expression, $-log_{10}\{p\text{-value}\} * sign([Fold\ Change])$. The ranked list, depicted by the shaded bar, has the most statistically significant genes located at its extreme ends, with the ER^{Pos} cells' most differentially expressed genes represented at the left end and the ER^{Neg} cells' most differentially expressed genes on the right. GSEA identified the Reactome pathway *Estrogen dependent gene expression* as one of the the most statistically significant enriched pathways identified (FDR<0.001). The position of the 89 genes composing the *estrogen dependent gene expression* pathway within the ranked list of expressed genes are depicted by vertical black lines. Names and positions of several familiar genes in the pathway are indicated, e.g., progesterone receptor (PGR) is found at position 337 (out of 21,085 expressed genes). The GSEA algorithm assessed the distribution of the pathway's 89 genes within the ranked list using a running sum score (green trace), calculating statistics by comparing results to analysis of a randomized ranked list (1000 permutations). **(b)** Heat map depicting transcript levels within ER^{Pos} and ER^{Neg} luminal epithelial cells of the 89 genes belonging to *estrogen dependent gene expression* pathway, ordered by their relative position in the ranked list (VST values normalized by row). Positions of the familiar genes are indicated, including SP1 (which was near the bottom of the list and was not differentially expressed). Transcript levels of PGR, ESR1, and SP1 in each cell type (rlog values on log2 (bar graph) and linear (donut graph) scales).

2629 **Figure 23. ER^{Pos} and ER^{Neg} luminal cells' enriched pathways.** Visualization of pathways
2630 enriched among ER^{Pos} and ER^{Neg} luminal cells (Enrichment Map/Cytoscape). All enriched
2631 pathways with a FDR≤0.1 appear on the map (nodes, color coded by cell type). Pathways(nodes)
2632 are connected when they share 50% of their genes (or greater). Node size corresponds to the
2633 number of genes in the pathway, as indicated (bottom left). Biologically related pathways
2634 sharing genes formed clusters in the network (biological themes), and different themes with a
2635 common biological function (e.g., metabolism-, immune -, and ECM-related, etc.) were typically
2636 placed adjacent to each other (when possible). Select themes were encircled and annotated.
2637 Highly differentially expressed genes are indicated (above the theme when it coincided with the
2638 attributed cell type—and below if it was expressed by the contrasted cell type). The names of
2639 each pathway are available in the supplemental Cytoscape file and Pathway Analyzer tool.

2640 **Figure 23—figure supplement 1. Transcript levels from enriched pathways. (a-t)** Transcript
2641 levels of leading-edge genes from select pathways enriched in ER^{Pos} or ER^{Neg} luminal cells (from
2642 325 enriched pathways with a FDR of ≤ 10%, GSEA). VST gene levels are relative (normalized
2643 by row). Arrowheads under the heatmaps identify the cell type either over- or under-expressing
2644 the pathway's genes. Samples are color-coded. Due to space constraints and the large number of
2645 presented genes, their names (rows of the heatmap) must be inspected digitally (magnified in the
2646 PDF).

2647 **Figure 23—figure supplement 2. Transcript levels from enriched pathways. (a-e)** Transcript
2648 levels of leading-edge genes from select pathways enriched in ER^{Pos} or ER^{Neg} luminal cells (from
2649 325 enriched pathways with a FDR of ≤ 10%, GSEA). VST gene levels are relative (normalized
2650 by row). Arrowheads under the heatmaps identify the cell type either over- or under-expressing
2651 the pathway's genes. Samples are color-coded. Due to space constraints and the large number of
2652 presented genes, their names (rows of the heatmap) must be inspected digitally (magnified in the
2653 PDF).

2654 **Figure 23—figure supplement 3. Transcript levels of toll-like receptors (a)** Transcript levels
2655 of TLR gene family. **(b-f)** Normalized mRNA values (rlog, DEseq2) provided on log2 scale (bar
2656 graph of each biological replicate) and linear scale (donut graph of median value) which are both
2657 color-coded by cell type.

2658 **Figure 23—figure supplement 4. Transcript levels of milk proteins. (a)** Transcript levels of
2659 milk proteins. VST gene values are relative (normalized by row). **(b-j)** Normalized mRNA
2660 values (rlog, DEseq2) provided on log2 scale (bar graph of each biological replicate) and linear
2661 scale (donut graph of median value), which are both color-coded by cell type.

2662 **Figure 24. Differences between ER^{Neg} luminal and myoepithelial cell types.** A contrast of
2663 expressed transcripts between ER^{Neg} luminal and myoepithelial cells revealed 2,239
2664 differentially expressed genes (DeSeq2/Benjamini-Hochberg adjusted p-value ≤ 0.1). All
2665 expressed genes are displayed. The horizontal dashed lines (separating red and green dots)
2666 represents an Adj. *p*-value of 0.1. Established cell type markers (of each cell type) are denoted.
2667 mRNA expression values for ten genes (indicated by filled, color-coordinated boxes) are shown
2668 at left and right of the volcano plot. Color-coded stars in the bar graphs indicate the cell type

being contrasted that had the higher transcript level, i.e., ER^{Neg}(blue) and myoepithelial cells (red).

Figure 25. The ER^{Neg} luminal and myoepithelial cells' enriched pathways. Visualization of pathways enriched among ER^{Pos} and myoepithelial cell types (Enrichment Map/Cytoscape). All enriched pathways with a FDR≤0.1 appear on the map (nodes, color coded by cell type). Pathways(nodes) are connected when they share 50% of their genes (or greater). Node size corresponds to the number of genes in the pathway, as indicated (bottom left). Select biological themes were encircled and annotated. Highly differentially expressed genes are indicated (above the theme when it coincided with the attributed cell type—and below if it was expressed by the contrasted cell type). The names of each pathway are available in the supplemental Cytoscape file and Pathway Analyzer tool.

Figure 25—figure supplement 1. Transcript levels from enriched pathways. Transcript levels of leading-edge genes from select pathways enriched in ER^{Pos} or myoepithelial cells (from 426 enriched pathways with a FDR of ≤ 10%, GSEA). VST gene levels are relative (normalized by row). Arrowheads under the heatmaps identify the cell type, either over- or under-expressing the pathway's genes. Samples are color-coded. Due to space constraints and the large number of presented genes, their names (rows of the heatmap) must be inspected digitally (magnified in the PDF).

Figure 26. Differences between ER^{Neg} luminal and vascular endothelial cell types. A contrast of expressed transcripts between ER^{Neg} luminal and vascular endothelial cells revealed 4,167 differentially expressed genes (DeSeq2/Benjamini-Hochberg adjusted p-value ≤ 0.1). All expressed genes are displayed. The horizontal dashed lines (separating red and green dots) represent an adj. *p*-value of 0.1. Established cell type markers (of each cell type) are denoted. mRNA expression values for ten genes (indicated by filled, color-coordinated boxes) are shown at the left and right of the volcano plot. Color-coded stars in the bar graphs indicate the cell type being contrasted that had the higher transcript level, i.e., ER^{Neg}(blue) and vascular endothelial cells (orange).

Figure 27. The ER^{Neg} luminal and vascular endothelial cells' enriched pathways. Visualization of pathways enriched among ER^{Pos} luminal and vascular endothelial cell types (Enrichment Map/Cytoscape). All enriched pathways with a FDR≤0.1 appear on the map (nodes, color-coded by cell type). Pathways(nodes) are connected when they share 50% of their genes (or greater). Node size corresponds to the number of genes in the pathway, as indicated (bottom left). Select biological themes were encircled and annotated. Highly differentially expressed genes are indicated (above the theme when it coincided with the attributed cell type—and below if it was expressed by the contrasted cell type). Samples are color-coded. Due to space constraints and a large number of presented genes, their names (rows of the heatmap) must be inspected digitally (magnified in the PDF). The names of each pathway are also available in the supplemental Cytoscape file and Pathway Analyzer tool.

Figure 27—figure supplement 1. Transcript levels from enriched pathways. Transcript levels of leading-edge genes from select pathways enriched in ER^{Pos} luminal and vascular endothelial cells (from 500 enriched pathways with a FDR of ≤ 10%, GSEA). VST gene levels

are relative (normalized by row). Arrowheads under the heatmaps identify the cell type, either over- or under-expressing the pathway's genes.

Figure 28. Differences between lymphatic and vascular endothelial cell types. A contrast of expressed transcripts between lymphatic and vascular endothelial cells revealed 3,739 differentially expressed genes (DeSeq2/Benjamini-Hochberg adjusted p-value ≤ 0.1). All expressed genes are displayed. The horizontal dashed lines (separating red and green dots) represent an adj. *p*-value of 0.1. Established cell type markers (of each cell type) are denoted. mRNA expression values for ten genes (indicated by filled, color-coordinated boxes) are shown at left and right of the volcano plot. Color-coded stars in the bar graphs indicate the cell type being contrasted that had the higher transcript level, i.e., lymphatic (light orange) and vascular (dark orange) endothelial cells.

Figure 29. The lymphatic and vascular endothelial cells' enriched pathways. Visualization of pathways enriched among lymphatic and vascular endothelial cell types (Enrichment Map/Cytoscape). All enriched pathways with a FDR≤0.1 appear on the map (nodes, color-coded by cell type). Pathways(nodes) are connected when they share 50% of their genes (or greater). Node size corresponds to the number of genes in the pathway, as indicated (bottom left). Select biological themes were encircled and annotated. Highly differentially expressed genes are indicated (above the theme when it coincided with the attributed cell type—and below if it was expressed by the contrasted cell type). Samples are color-coded. Due to space constraints and a large number of presented genes, their names (rows of the heatmap) must be inspected digitally (magnified in the PDF). The names of each pathway are also available in the supplemental Cytoscape file and Pathway Analyzer tool.

Figure 29—figure supplement 1. Transcript levels from enriched pathways. Transcript levels of leading-edge genes from select pathways enriched in lymphatic and vascular endothelial cell (from 240 enriched pathways with a FDR of ≤ 10%, GSEA). VST gene levels are relative (normalized by row). Arrowheads under the heatmaps identify the cell type, either over- or under-expressing the pathway's genes.

Figure 30. Differences between fibroblasts and adipocytes. A contrast of expressed transcripts between fibroblasts and adipocytes revealed 1,171 differentially expressed genes (DeSeq2/Benjamini-Hochberg adjusted p-value ≤ 0.1). All expressed genes are displayed. The horizontal dashed lines (separating red and green dots) represent an adj. *p*-value of 0.1. Established cell type markers (of each cell type) are denoted. mRNA expression values for ten genes (indicated by filled, color-coordinated boxes) are shown at left and right of the volcano plot. Color-coded stars in the bar graphs identify the position of the fibroblasts (yellow) and adipocytes (brown).

Figure 31. The fibroblasts and adipocytes' enriched pathways. Visualization of pathways enriched among fibroblasts and adipocytes (Enrichment Map/Cytoscape). All enriched pathways with a FDR≤0.1 appear on the map (nodes, color-coded by cell type). Pathways(nodes) are connected when they share 50% of their genes (or greater). Node size corresponds to the number of genes in the pathway, as indicated (bottom left). Select biological themes were encircled and annotated. Highly differentially expressed genes are indicated (above the theme when it

2751 coincided with the attributed cell type—and below if it was expressed by the contrasted cell
2752 type). The names of each pathway are available in the supplemental Cytoscape file and Pathway
2753 Analyzer tool. Samples are color-coded. Due to space constraints and a large number of
2754 presented genes, their names (rows of the heatmap) must be inspected digitally (magnified in the
2755 PDF).

2756 **Figure 31—figure supplement 1. Transcript levels from enriched pathways.** Transcript
2757 levels of leading-edge genes from select pathways enriched in fibroblasts and adipocytes (from
2758 387 enriched pathways with a FDR of ≤ 10%, GSEA). VST gene levels are relative (normalized
2759 by row). Arrowheads under the heatmaps identify the cell type, either over- or under-expressing
2760 the pathway's genes.

2761 **Figure 32. Differences between pericytes and fibroblasts.** A contrast of expressed transcripts
2762 between pericytes and fibroblasts revealed 2,932 differentially expressed genes
2763 (DeSeq2/Benjamini-Hochberg adjusted p-value ≤ 0.1). All expressed genes are displayed. The
2764 horizontal dashed lines (separating red and green dots) represent an adj. *p*-value of 0.1.
2765 Established cell type markers (of each cell type) are denoted. mRNA expression values for ten
2766 genes (indicated by filled, color-coordinated boxes) are shown at the left and right of the volcano
2767 plot. Color-coded stars in the bar graphs identify the position of the pericytes (purple) and
2768 fibroblasts (yellow).

2769 **Figure 33. The pericytes and fibroblasts' enriched pathways.** Visualization of pathways
2770 enriched among pericytes and fibroblasts (Enrichment Map/Cytoscape). All enriched pathways
2771 with a FDR≤0.1 appear on the map (nodes, color-coded by cell type). Pathways(nodes) are
2772 connected when they share 50% of their genes (or greater). Node size corresponds to the number
2773 of genes in the pathway, as indicated (bottom left). Select biological themes were encircled and
2774 annotated. Highly differentially expressed genes are indicated (above the theme when it
2775 coincided with the attributed cell type—and below if it was expressed by the contrasted cell
2776 type). Samples are color-coded. Due to space constraints and a large number of presented genes,
2777 their names (rows of the heatmap) must be inspected digitally (magnified in the PDF). The
2778 names of each pathway are also available in the supplemental Cytoscape file and Pathway
2779 Analyzer tool.

2780 **Figure 33—figure supplement 1. Transcript levels from enriched pathways.** Transcript
2781 levels of leading-edge genes from select pathways enriched in pericytes and fibroblasts (from
2782 477 enriched pathways with a FDR of ≤ 10%, GSEA). VST gene levels are relative (normalized
2783 by row). Arrowheads under the heatmaps identify the cell type, either over- or under-expressing
2784 the pathway's genes.

A cellular and transcriptomic dissection of the human breast for studying mechanisms of cell and tissue function

Katelyn Del Toro, Rosalyn Sayaman, Kate Thi, Yamhilette Licon-Munoz, and William C. Hines

Figures

Figure 1

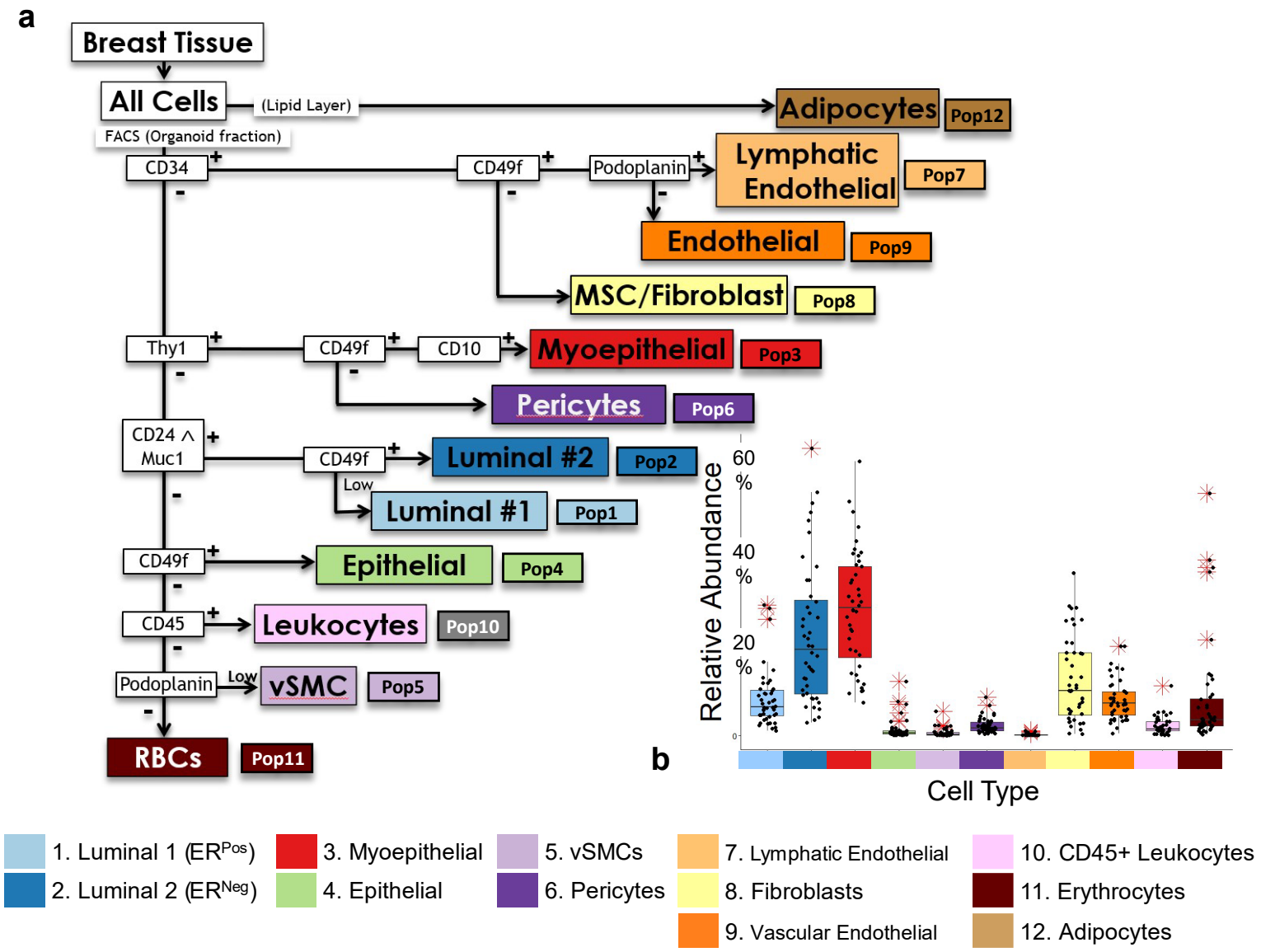
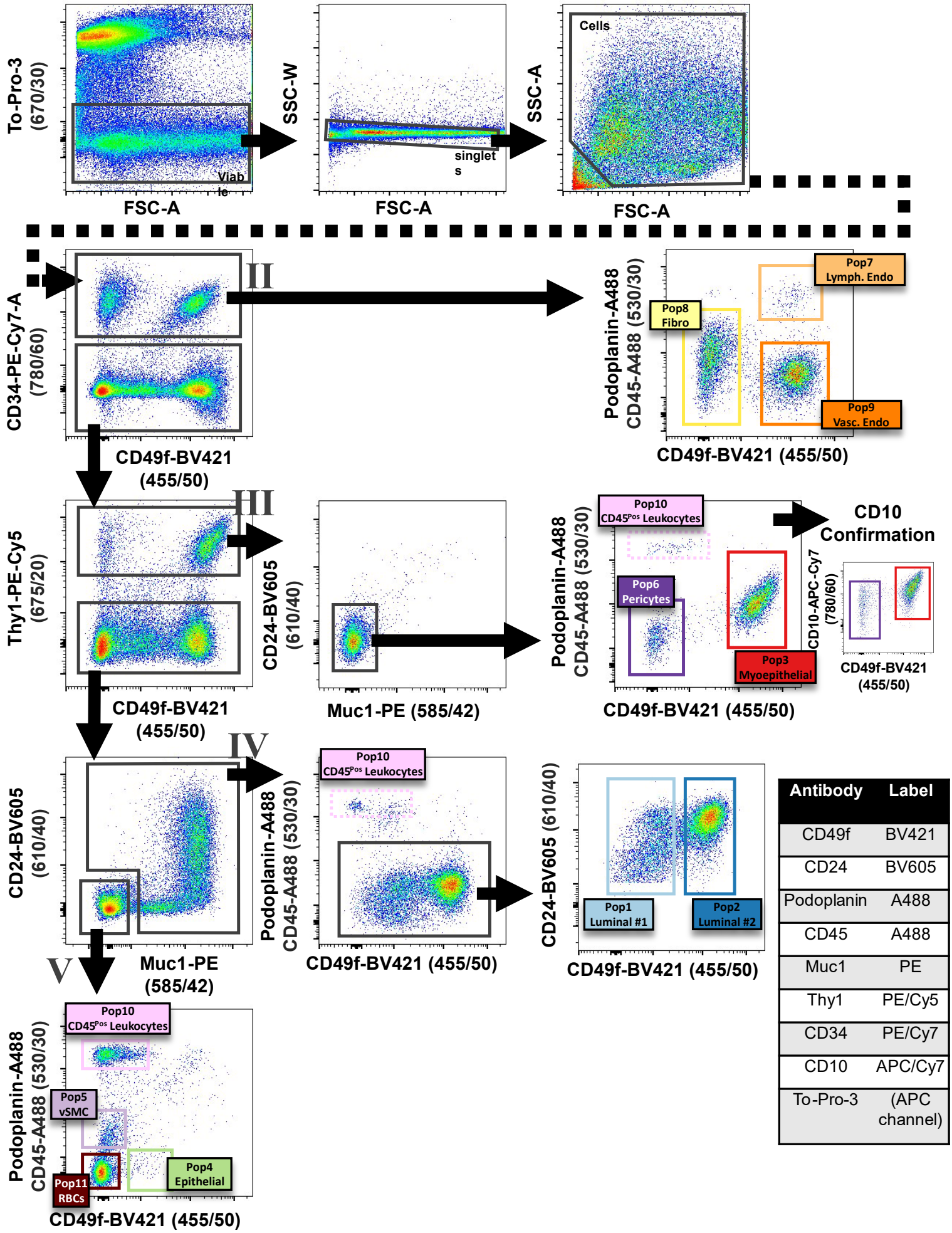

a
Breast Tissue
All Cells
(Lipid Layer)
Adipocytes
Pop12
FACS (Organoid fraction)
CD34
CD49f
Podoplanin
Lymphatic Endothelial
Pop7
Endothelial
Pop9
MSC/Fibroblast
Pop8
Thy1
CD49f
CD10
Myoepithelial
Pop3
Pericytes
Pop6
CD24 ∧ Muc1
CD49f
Luminal #2
Pop2
Low
Luminal #1
Pop1
CD49f
Epithelial
Pop4
CD45
Leukocytes
Pop10
Podoplanin
Low
vSMC
Pop5
RBCs
Pop11
b
Relative Abundance
60 %
40 %
20 %
0
Cell Type
1. Luminal 1 (ERPos)
2. Luminal 2 (ERNeg)
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. CD45+ Leukocytes
11. Erythrocytes
12. Adipocytes

Figure 1—figure supplement 1

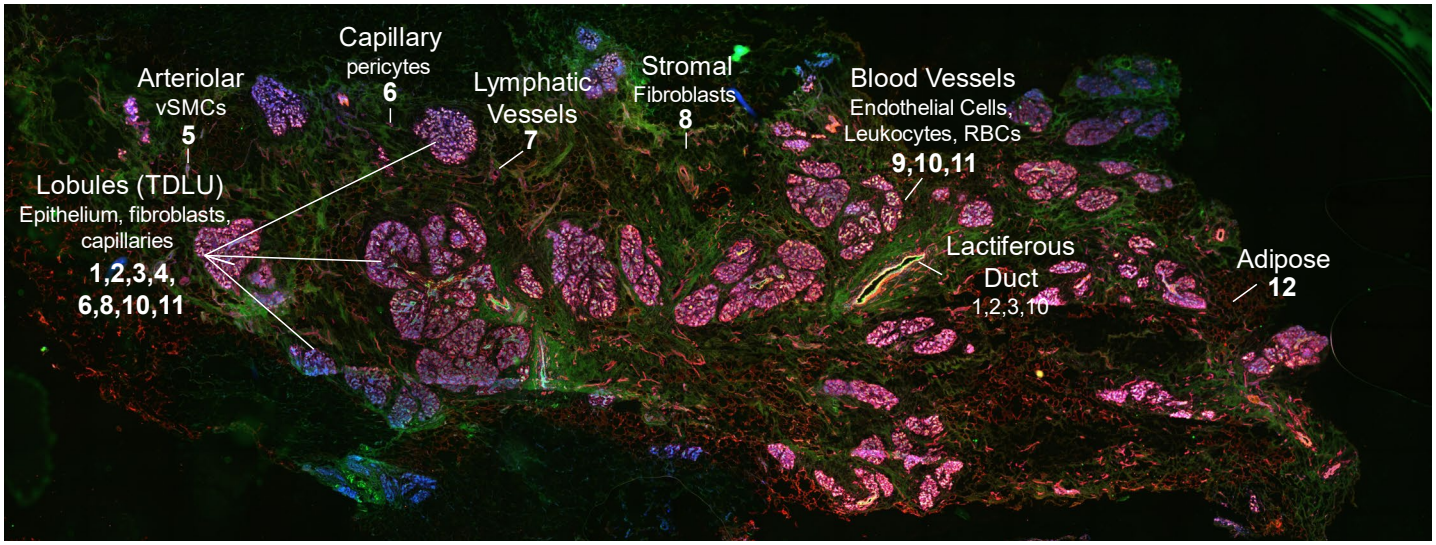
BD FACS VANTAGE
8/8 channels used



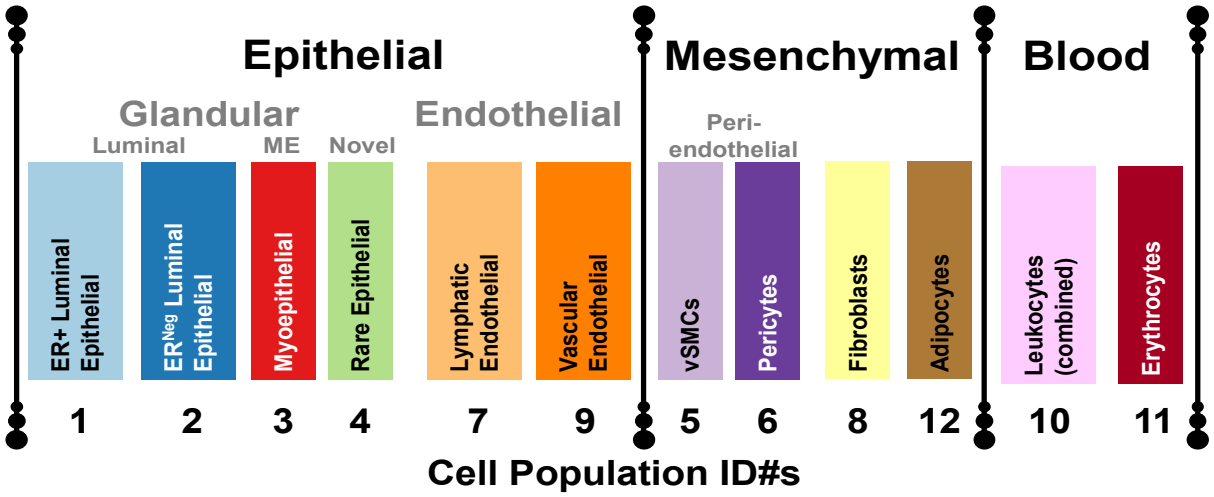
Antibody	Label
CD49f	BV421
CD24	BV605
Podoplanin	A488
CD45	A488
Muc1	PE
Thy1	PE/Cy5
CD34	PE/Cy7
CD10	APC/Cy7
To-Pro-3	(APC channel)

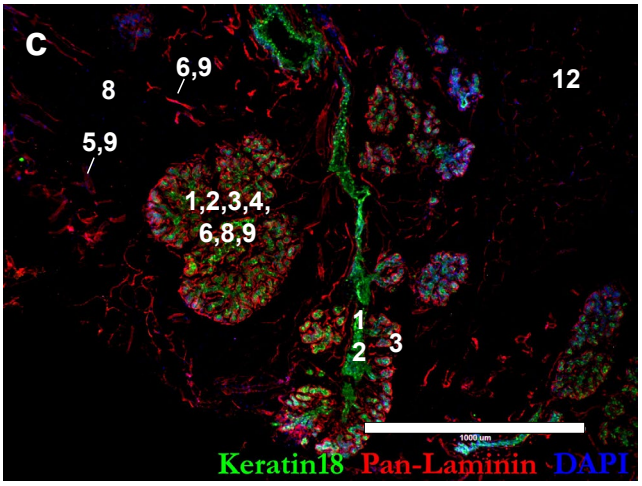
Figure 2

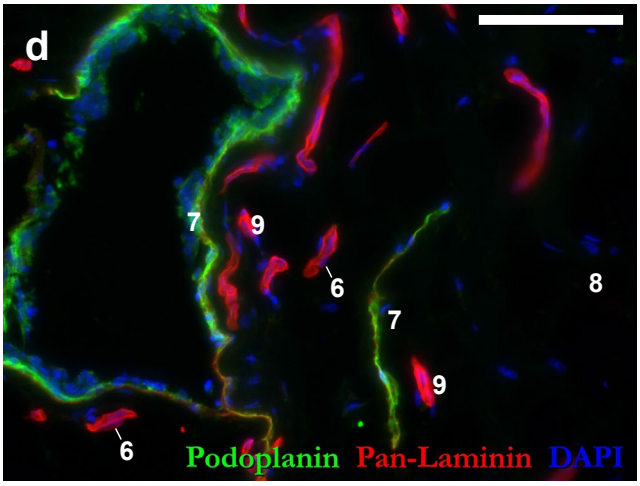
a

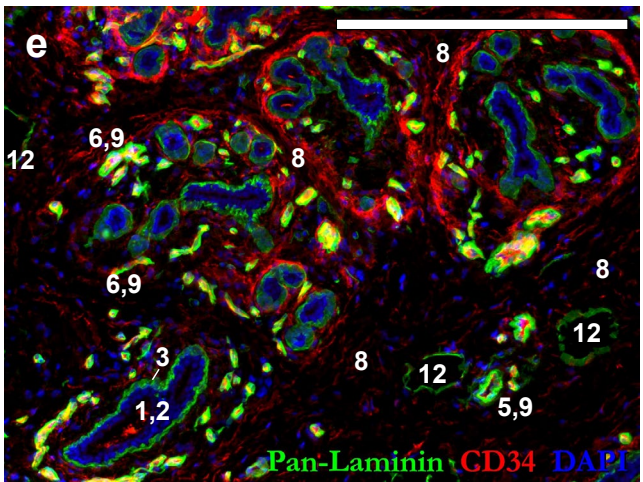

Arteriolar
vSMCs
5
Capillary
pericytes
6
Lymphatic
Vessels
7
Stromal
Fibroblasts
8
Blood Vessels
Endothelial Cells,
Leukocytes, RBCs
9,10,11
Lobules (TDLU)
Epithelium, fibroblasts,
capillaries
1,2,3,4,
6,8,10,11
Lactiferous
Duct
1,2,3,10
Adipose
12

b


Epithelial
Mesenchymal
Blood
Glandular
Endothelial
Luminal
ME
Novel
Peri-
endothelial
ER+ Luminal Epithelial
ER^Neg Luminal Epithelial
Myoepithelial
Rare Epithelial
Lymphatic Endothelial
Vascular Endothelial
vSMCs
Pericytes
Fibroblasts
Adipocytes
Leukocytes (combined)
Erythrocytes
1
2
3
4
7
9
5
6
8
12
10
11
Cell Population ID#s


c
8
6,9
12
5,9
1,2,3,4,
6,8,9
1
2
3
Keratin18 Pan-Laminin DAPI


d
7
9
6
8
7
9
6
Podoplanin Pan-Laminin DAPI


e
8
12
6,9
8
6,9
8
12
3
8
12
1,2
5,9
Pan-Laminin CD34 DAPI

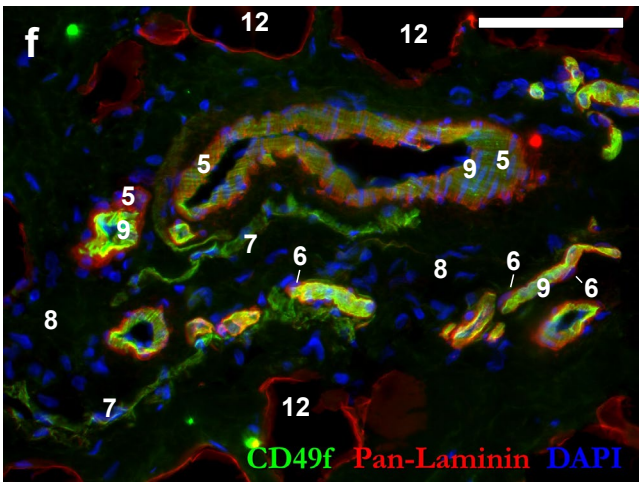

f
12
12
5
9
5
5
9
7
6
8
6
9
6
8
7
12
CD49f Pan-Laminin DAPI

Figure 2—figure supplement 1

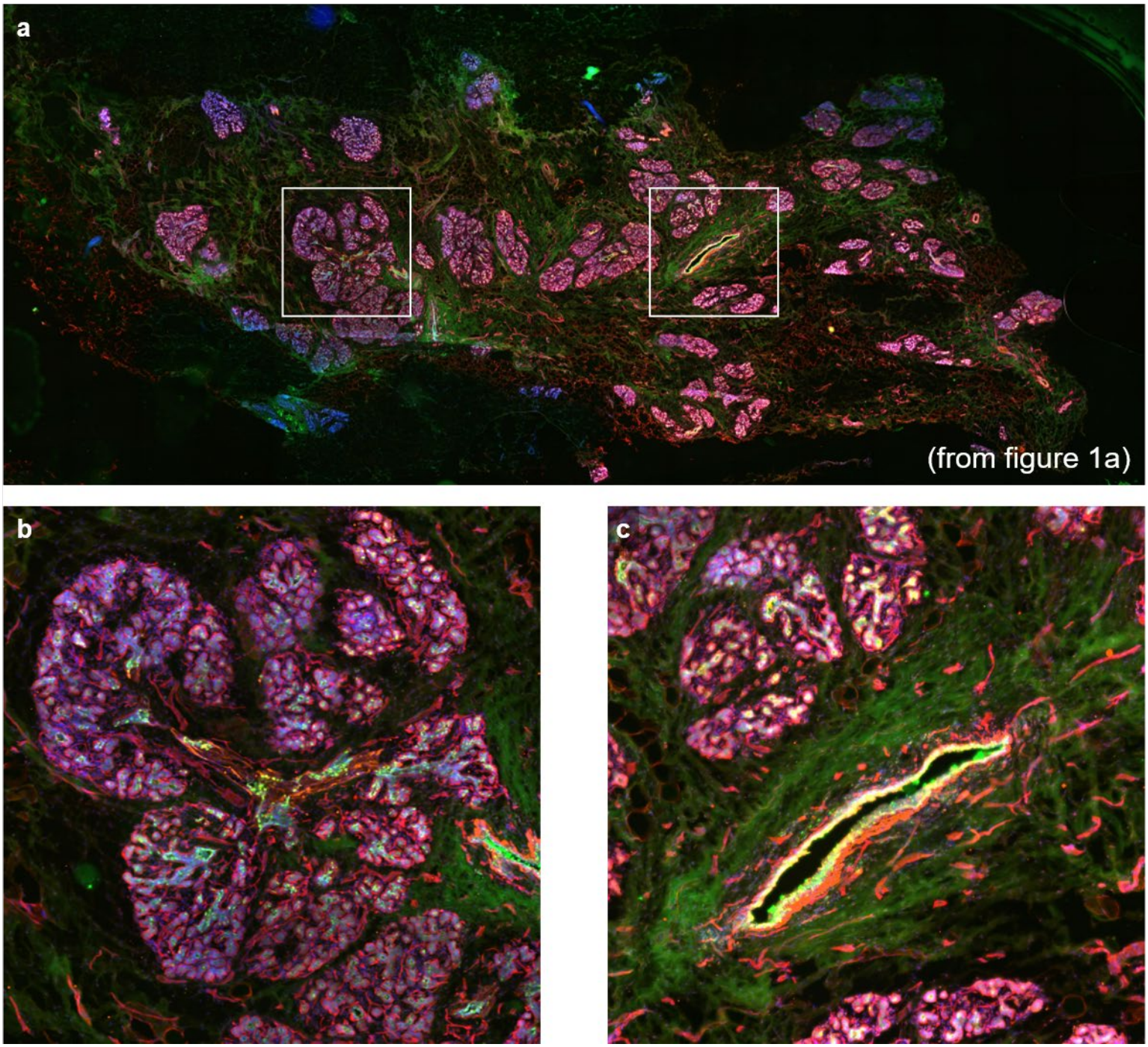
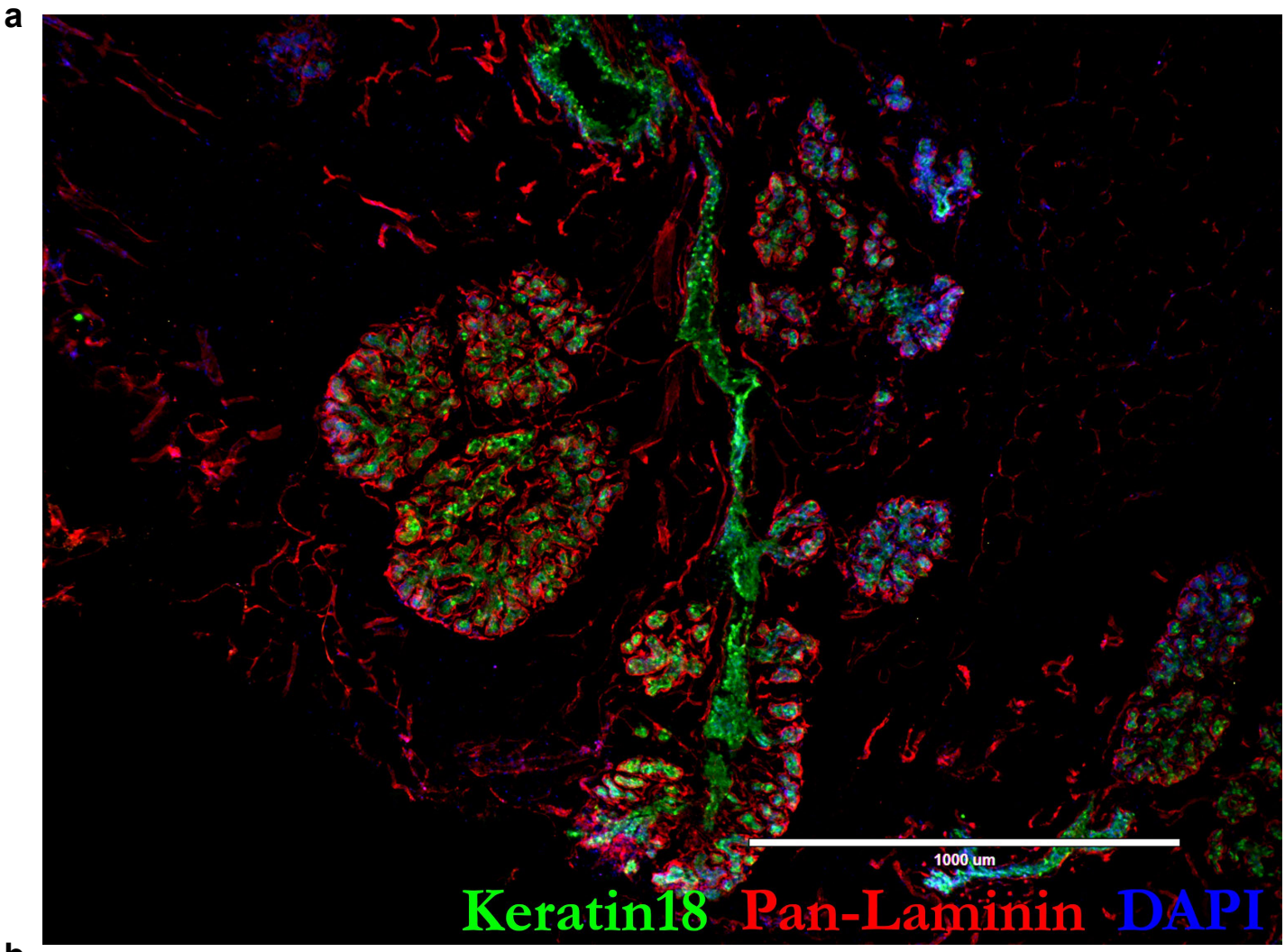


Figure 2—figure supplement 2



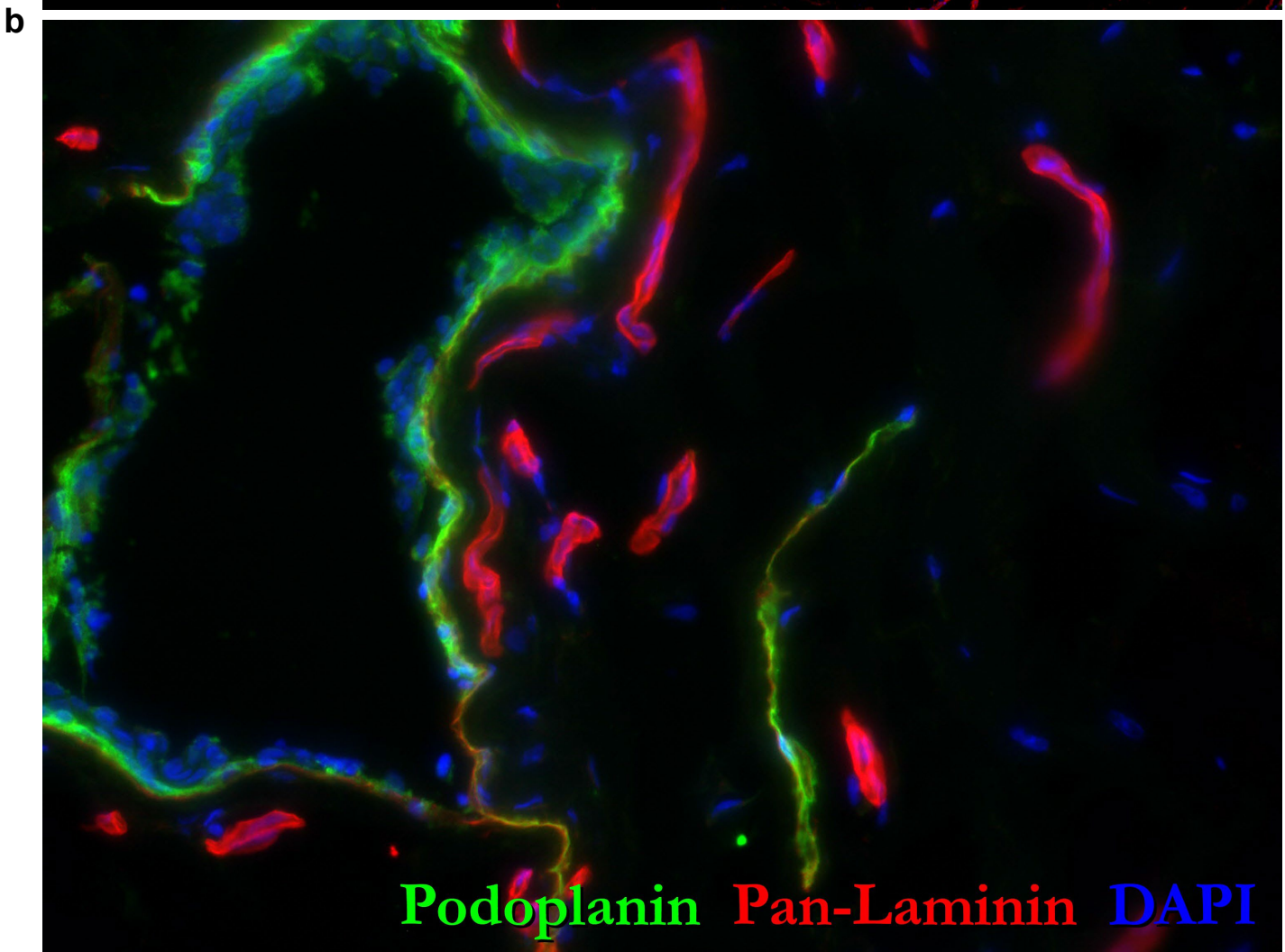


Figure 3

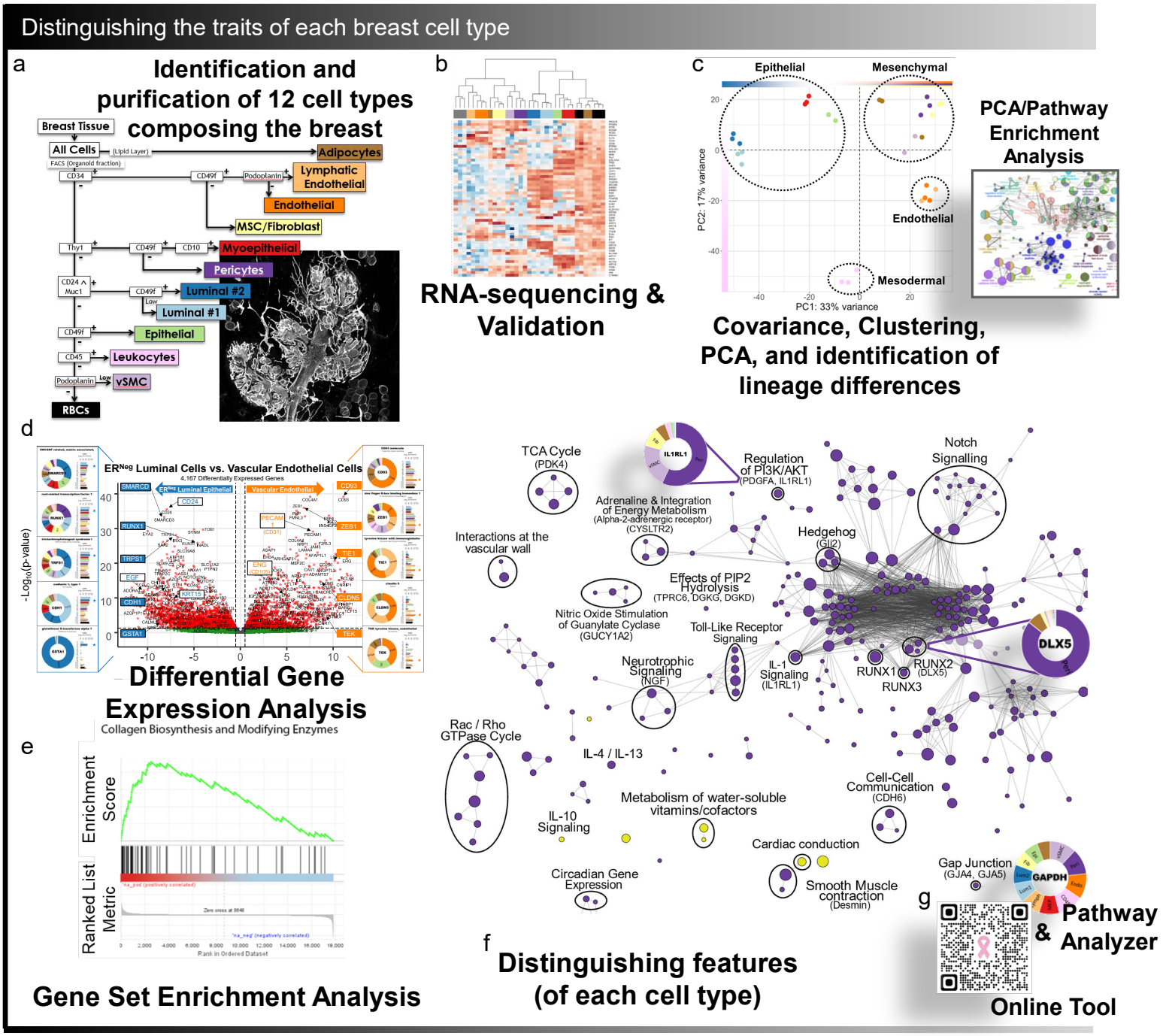

Distinguishing the traits of each breast cell type
a
Identification and purification of 12 cell types composing the breast
Breast Tissue
All Cells
(Lipid Layer)
Adipocytes
FACS (Organoid fraction)
CD34
CD49f
Podoplanin
Lymphatic Endothelial
Endothelial
MSC/Fibroblast
Thy1
CD10
Myoepithelial
Pericytes
CD24
Muc1
Luminal #2
Luminal #1
Epithelial
CD45
Leukocytes
vSMC
RBCs
b
RNA-sequencing & Validation
c
Epithelial
Mesenchymal
Endothelial
Mesodermal
PC2: 17% variance
PC1: 33% variance
PCA/Pathway Enrichment Analysis
Covariance, Clustering, PCA, and identification of lineage differences
d
ERNeg Luminal Cells vs. Vascular Endothelial Cells
4,167 Differentially Expressed Genes
ERNeg Luminal Epithelial
Vascular Endothelial
-Log10(p-value)
Differential Gene Expression Analysis
e
Collagen Biosynthesis and Modifying Enzymes
Enrichment Score
Ranked List Metric
Rank in Ordered Dataset
Gene Set Enrichment Analysis
f
TCA Cycle (PDK4)
Regulation of PI3K/AKT (PDGFA, IL1RL1)
Notch Signalling
Adrenaline & Integration of Energy Metabolism (Alpha-2-adrenergic receptor) (CYSLTR2)
Interactions at the vascular wall
Hedgehog (Gli2)
Effects of PIP2 Hydrolysis (TPRC6, DGKG, DGKD)
Nitric Oxide Stimulation of Guanylate Cyclase (GUCY1A2)
Toll-Like Receptor Signaling
Neurotrophic Signaling (NGF)
IL-1 Signaling (IL1RL1)
RUNX1
RUNX2 (DLX5)
RUNX3
DLX5
Rac / Rho GTPase Cycle
IL-4 / IL-13
Cell-Cell Communication (CDH6)
Metabolism of water-soluble vitamins/cofactors
IL-10 Signaling
Cardiac conduction
Circadian Gene Expression
Smooth Muscle contraction (Desmin)
Distinguishing features (of each cell type)
g
Gap Junction (GJA4, GJA5)
GAPDH
Pathway & Analyzer
Online Tool

Figure 3—figure supplement 1

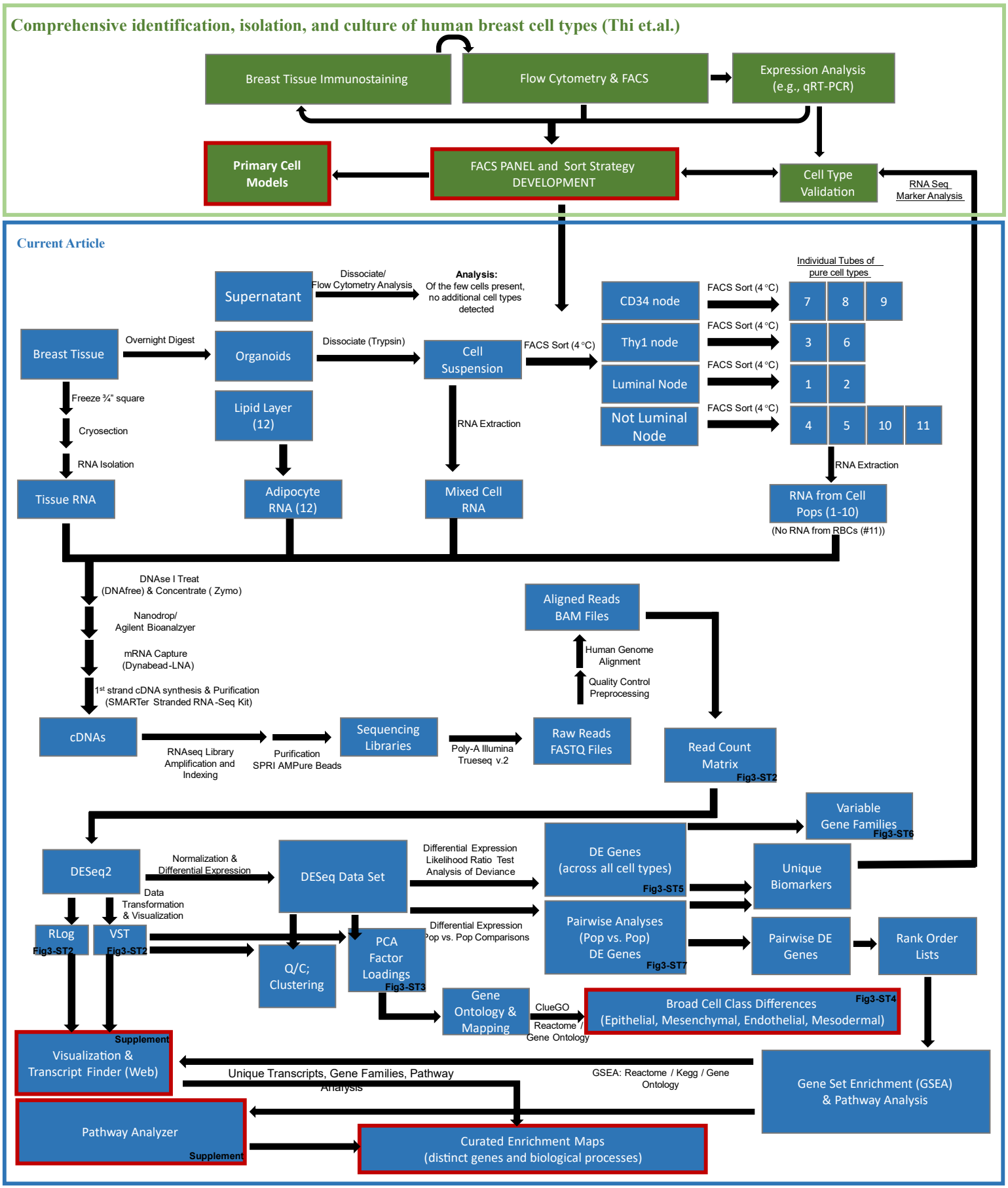


Figure 3—table supplements (supplemental files)

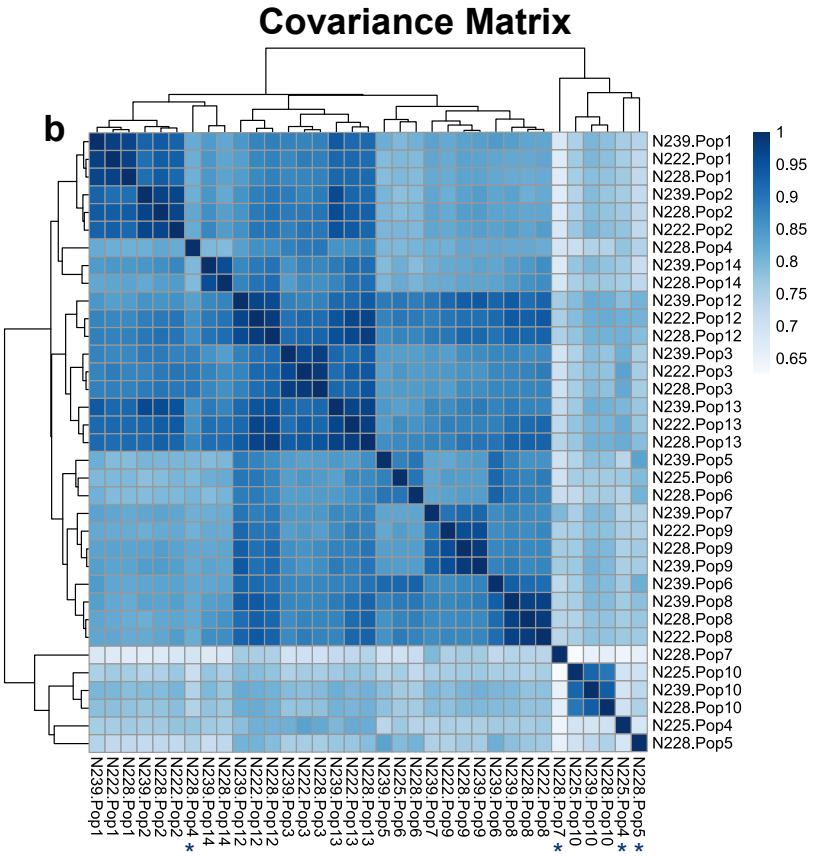
List of Table supplements and Data	
Figure 3—table supplement 1	Sample Treatment.xlsx
Figure 3—table supplement 2	CountMatrix_rLog_VST.xlsx
Figure 3—table supplement 3	PCA_Loading_Factors.xlsx
Figure 3—table supplement 4	GO from PCAs.xlsx
Figure 3—table supplement 5	DESeq2_LRT.xlsx
Figure 3—table supplement 6	Gene Families_DE Genes.xlsx
Figure 3—table supplement 7	DESeq2_files.xlsx
OSF; https://osf.io/gfhc3/	Cytoscape/Enrichment Map files (1v2, 2v3, 2v9, 6v8)
OSF; https://osf.io/hpb9t/	Pathway Analyzer[9]
http://breastcancerlab.com/breast-atlas	Breast Cell Atlas: Transcript Finder (R/Shiny App)[10]

Figure 4

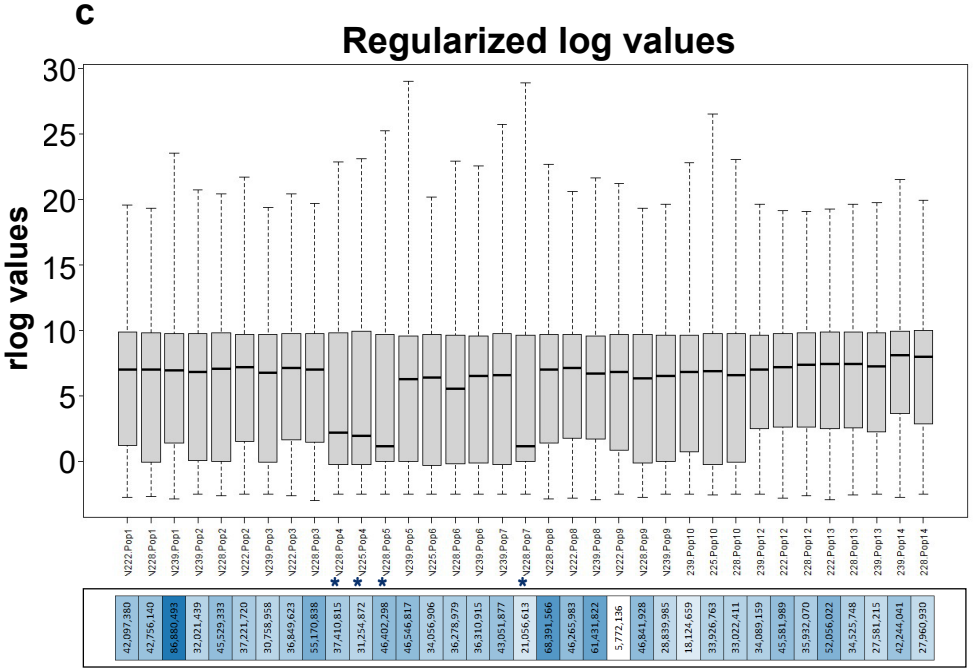
a

		Cell Populations sequenced			
		N222 29 y/o ♀	**N225** 16 y/o ♀	**N228** 27 y/o ♀	**N239** 23 y/o ♀
Pop1	Luminal 1 (ER+)	1		1	1
Pop2	Luminal 2 (ER-)	2		2	2
Pop3	Myoepithelial	3		3	3
Pop4	Epithelial		4	4	4
Pop5	Vascular Smooth Muscle Cells			5	5
Pop6	Pericytes		6	6	6
Pop7	Lymphatic Endothelial Cells			7	7
Pop8	Fibroblasts	8		8	8
Pop9	Vascular Endothelial Cells	9		9	9
Pop10	Leukocytes		10	10	10
Pop11	Red Blood Cells *no RNA / NS	-	-	-	-
Pop12	Adipocytes	12		12	12
Pop13	Unsorted Cells (dissociated, pre-FACS)	13		13	13
Pop14	Tissue			14	14

b



c



d

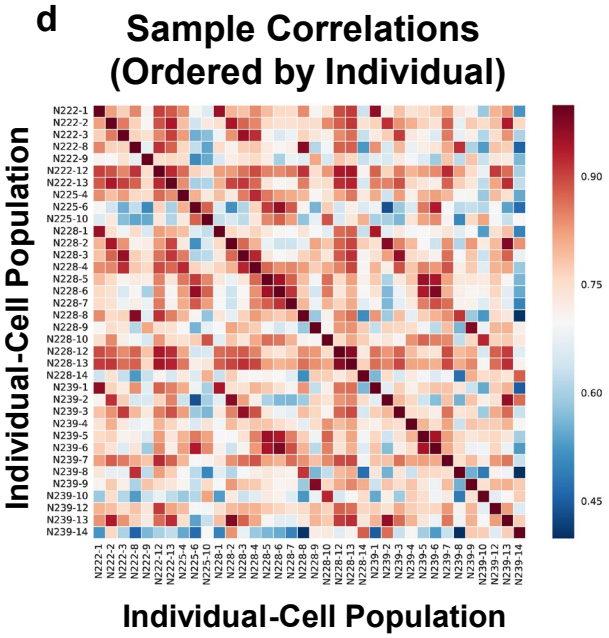
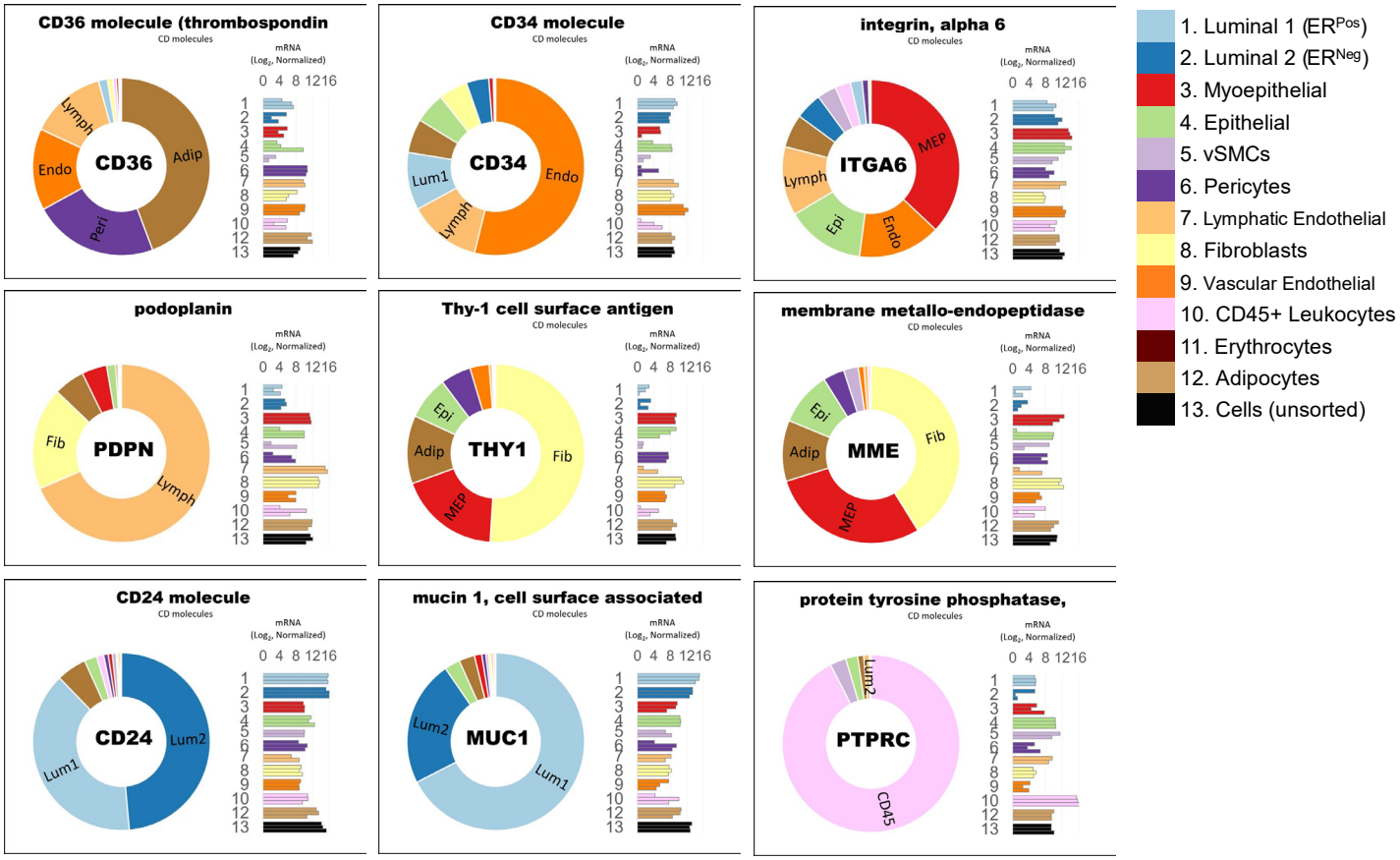


Figure 4-figure supplement 1

a



b

		ANTIBODY STAINING IN SORTING SCHEME								
Pop ID	Cell Type	CD36†	CD34	CD49f (ITGA6)	PDPN	CD90 (Thy1)	CD10 (MME)	CD24	CD227 (MUC1)	CD45 (PTPRC)
1	ER^{Pos} Luminal Epithelial		Neg	-/+	Neg	Neg	Neg	+++	+++	Neg
2	ER^{Neg} Luminal Epithelial		Neg	++	Neg	Neg	Neg	+++	+++	Neg
3	Myoepithelial		Neg	++++	++	+++	+++	Neg	Neg	Neg
4	Pop4 Epithelial		Neg	++	+	Neg	-/+	Neg	Neg	Neg
5	Vascular Smooth Muscle Cells		Neg	Neg	-/+	Neg	Neg	Neg	Neg	Neg
6	Pericytes		Neg	Neg	Neg	+++	Neg	Neg	Neg	Neg
7	Lymphatic Endothelial Cells		+++	++	+++	-/+	Neg	Neg	Neg	Neg
8	Adipocyte Derived Mesenchymal Stem Cells (Fibroblasts)		+++	Neg	+	++	+	Neg	Neg	Neg
9	Vascular Endothelial Cells		+++	++	Neg	-/+	Neg	Neg	Neg	Neg
10	Leukocytes (CD45^{Pos})		Neg	Neg	Neg	Neg	-/+	Neg	Neg	+++
11	Red Blood Cells		Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
12	Adipocytes	+++								

c

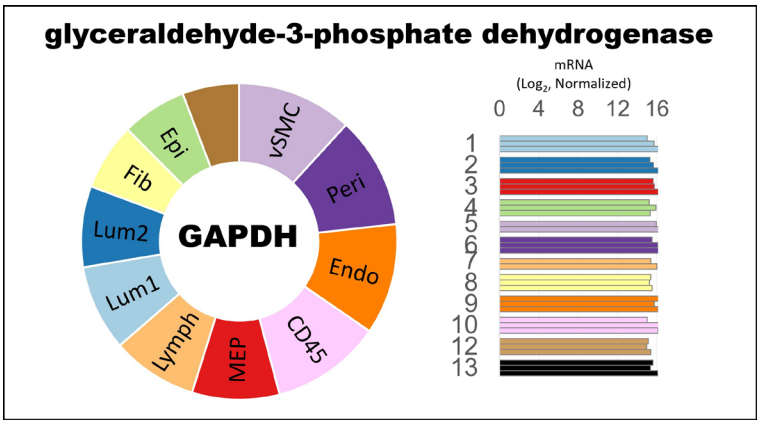


Figure 5

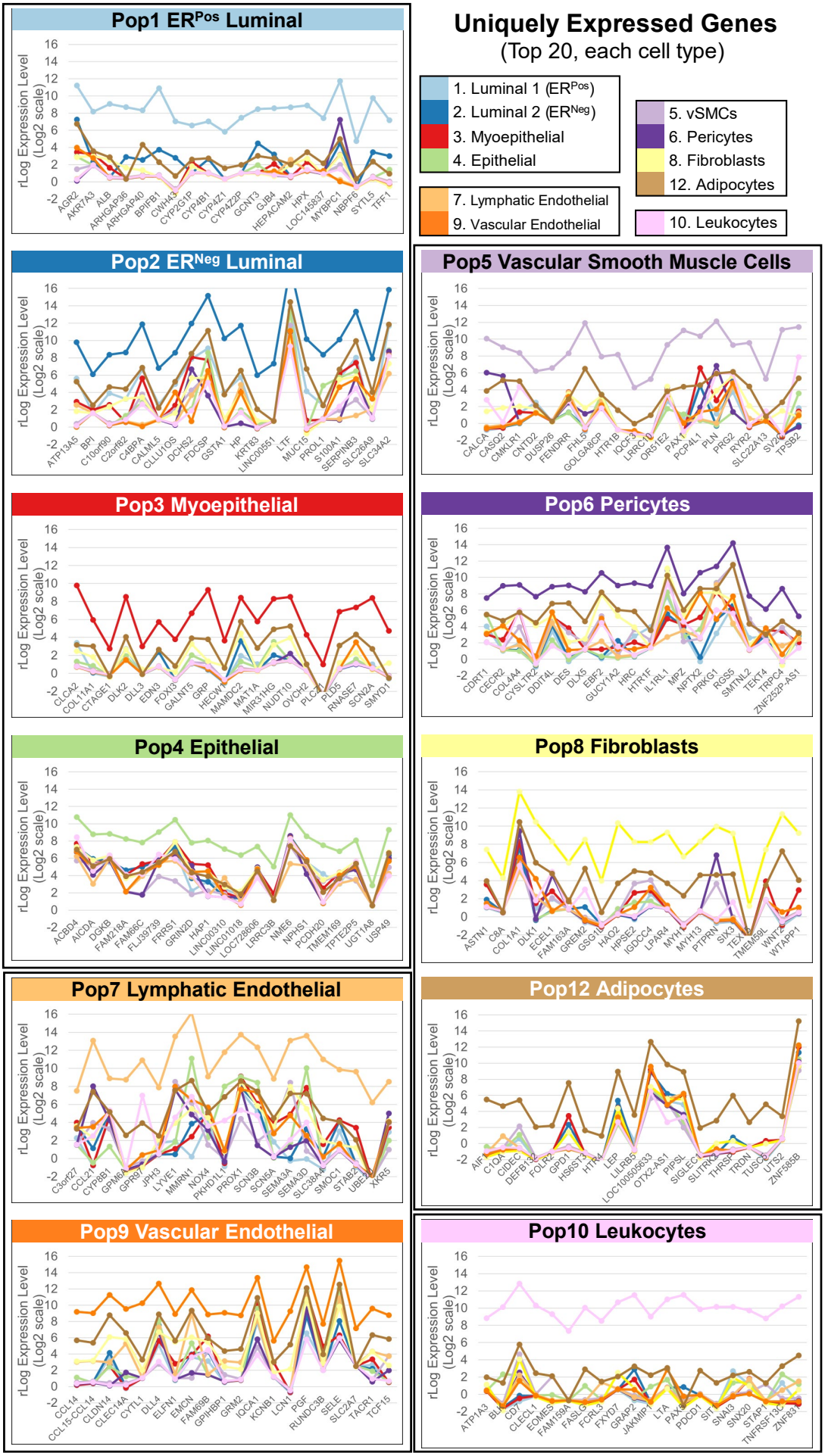

Uniquely Expressed Genes
(Top 20, each cell type)
1. Luminal 1 (ERPos)
2. Luminal 2 (ERNeg)
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
8. Fibroblasts
12. Adipocytes
7. Lymphatic Endothelial
9. Vascular Endothelial
10. Leukocytes
Pop1 ERPos Luminal
Pop2 ERNeg Luminal
Pop3 Myoepithelial
Pop4 Epithelial
Pop5 Vascular Smooth Muscle Cells
Pop6 Pericytes
Pop8 Fibroblasts
Pop12 Adipocytes
Pop7 Lymphatic Endothelial
Pop9 Vascular Endothelial
Pop10 Leukocytes
rLog Expression Level
(Log2 scale)

Figure 5-figure supplement 1

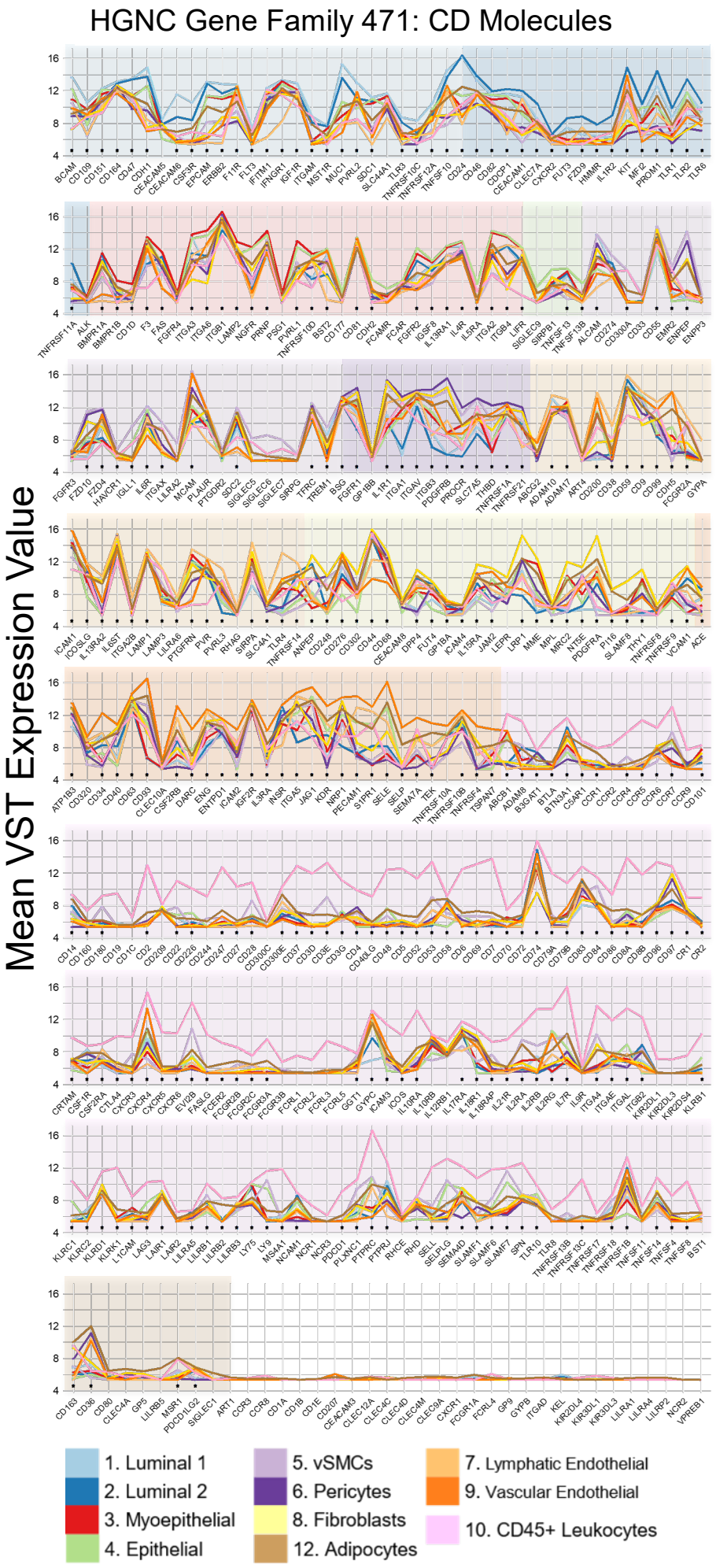


Figure 6

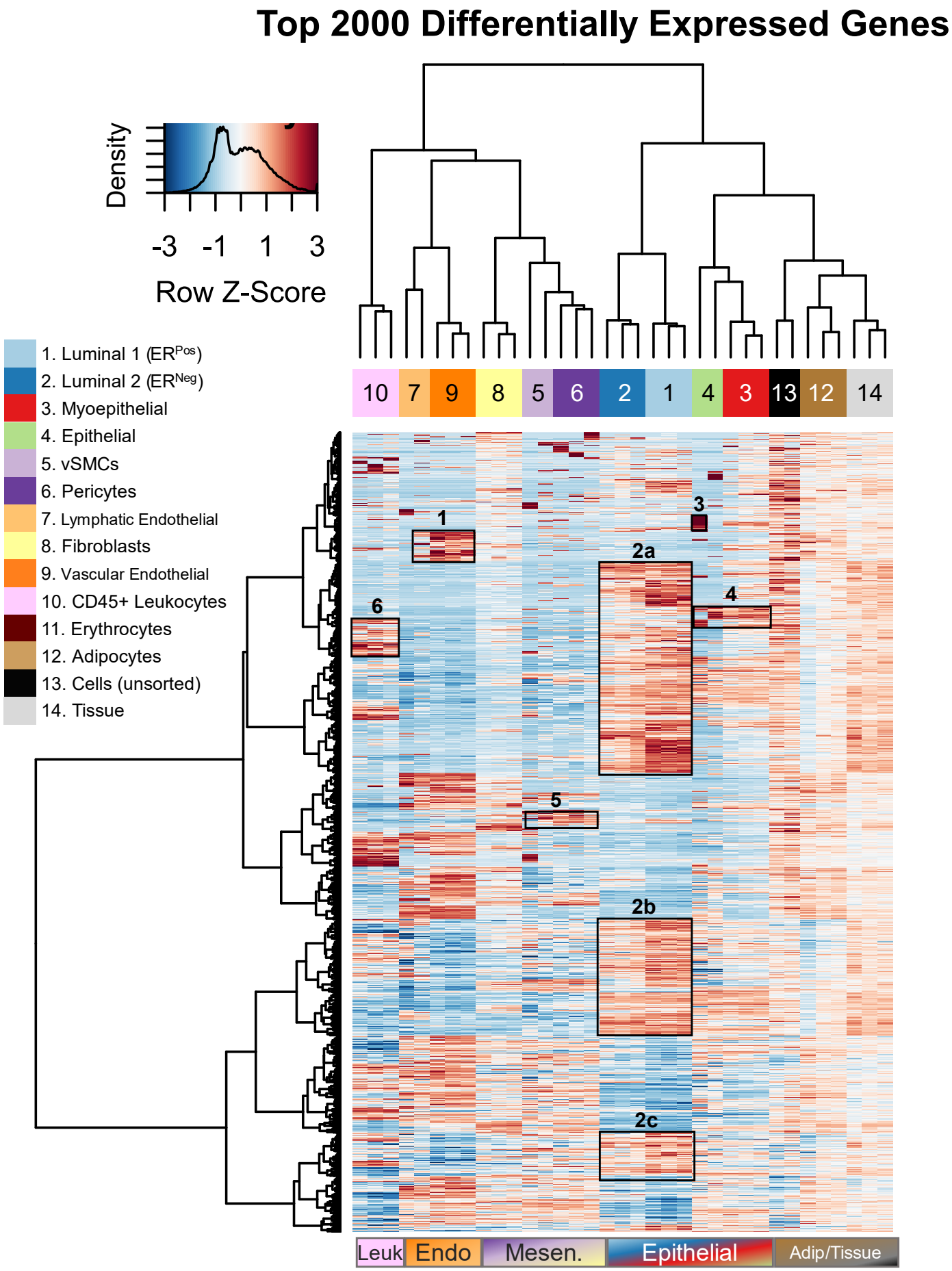

Top 2000 Differentially Expressed Genes
Density
-3 -1 1 3
Row Z-Score
1. Luminal 1 (ER^Pos)
2. Luminal 2 (ER^Neg)
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. CD45+ Leukocytes
11. Erythrocytes
12. Adipocytes
13. Cells (unsorted)
14. Tissue
10 7 9 8 5 6 2 1 4 3 13 12 14
1
3
2a
4
6
5
2b
2c
Leuk
Endo
Mesen.
Epithelial
Adip/Tissue

Figure 7

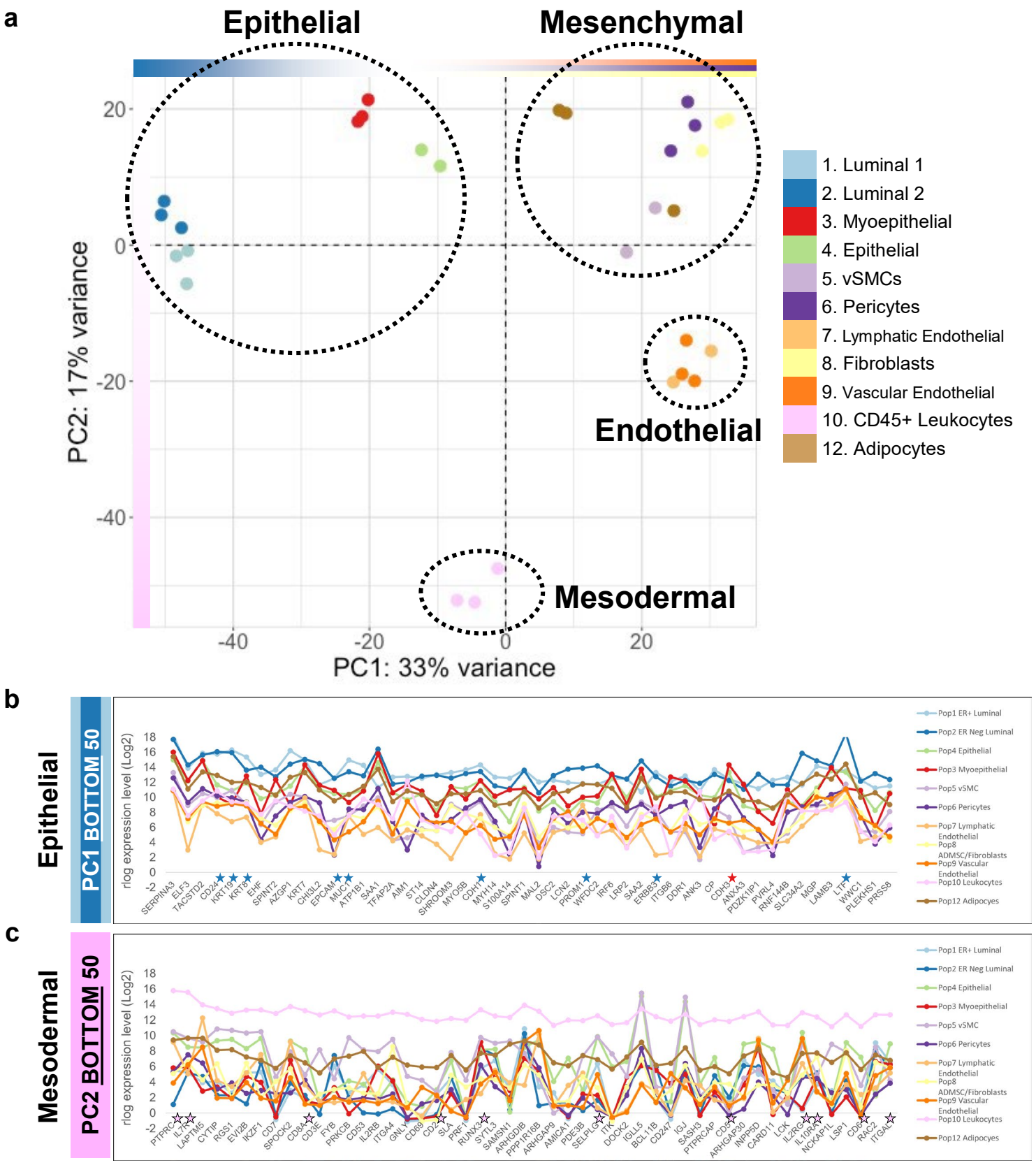

a
Epithelial
Mesenchymal
Endothelial
Mesodermal
PC2: 17% variance
PC1: 33% variance
20
0
-20
-40
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. CD45+ Leukocytes
12. Adipocytes
b
Epithelial
PC1 BOTTOM 50
rlog expression level (Log2)
SERPINA3 ELF3 TACSTD2 CD24 KRT19 KRT8 EHF SPINT2 AZGP1 KRT7 CHI3L2 EPCAM MUC1 ATP1B1 SAA1 TFAP2A AIM1 ST14 CLDN4 SHROOM3 MYO5B CDH1 MYH14 S100A14 SPINT1 MAL2 DSC2 LCN2 PROM1 WFDC2 IRF6 LRP2 SAA2 ERBB3 ITGB6 DDR1 ANK3 CP CDH3 ANXA3 PDZK1IP1 PVRL4 RNF144B SLC34A2 MGP LAMB3 LTF WWC1 PLEKHS1 PRSS8
Pop1 ER+ Luminal
Pop2 ER Neg Luminal
Pop4 Epithelial
Pop3 Myoepithelial
Pop5 vSMC
Pop6 Pericytes
Pop7 Lymphatic Endothelial
Pop8 ADMSC/Fibroblasts
Pop9 Vascular Endothelial
Pop10 Leukocytes
Pop12 Adipocyes
c
Mesodermal
PC2 BOTTOM 50
rlog expression level (Log2)
PTPRC IL7R LAPTM5 CYTIP RGS1 EVI2B IKZF1 CD7 SPOCK2 CD8A CD3E FYB PRKCB CD53 IL2RB ITGA4 GNLY CD69 CD3D SLA PRF1 RUNX3 SYTL3 SAMSN1 ARHGDIB PPP1R16B ARHGAP9 AMICA1 PDE3B SELPLG ITK DOCK2 IGLL5 BCL11B CD247 IGJ SASH3 PTPRCAP CD2 ARHGAP30 INPP5D CARD11 LCK IL2RG IL10RA NCKAP1L LSP1 CD5 RAC2 ITGAL
Pop1 ER+ Luminal
Pop2 ER Neg Luminal
Pop4 Epithelial
Pop3 Myoepithelial
Pop5 vSMC
Pop6 Pericytes
Pop7 Lymphatic Endothelial
Pop8 ADMSC/Fibroblasts
Pop9 Vascular Endothelial
Pop10 Leukocytes
Pop12 Adipocyes

Figure 7-figure supplement 1

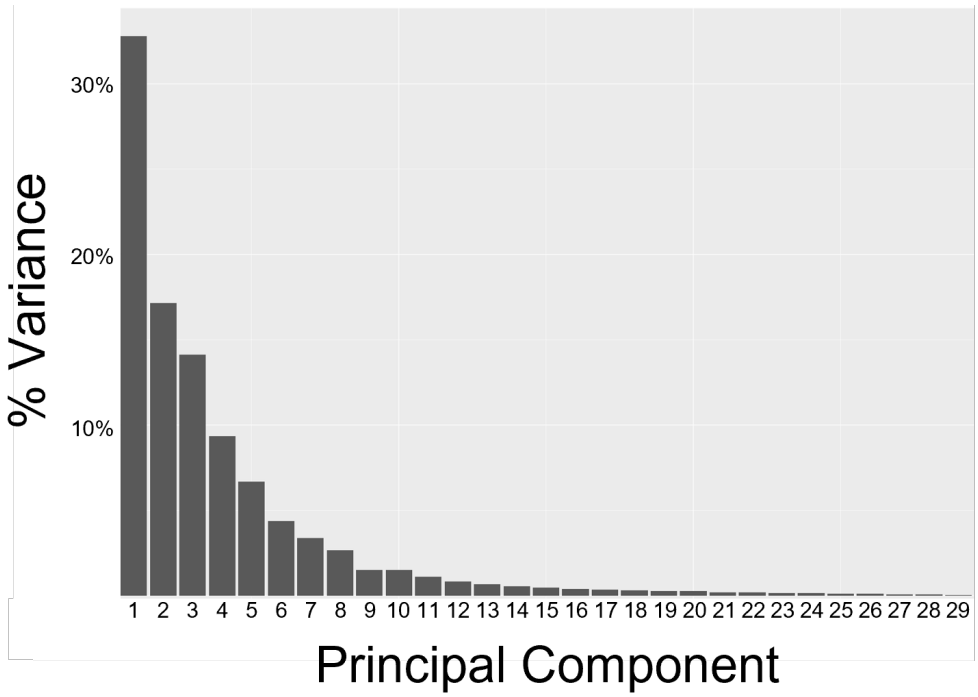


Figure 7-figure supplement 2

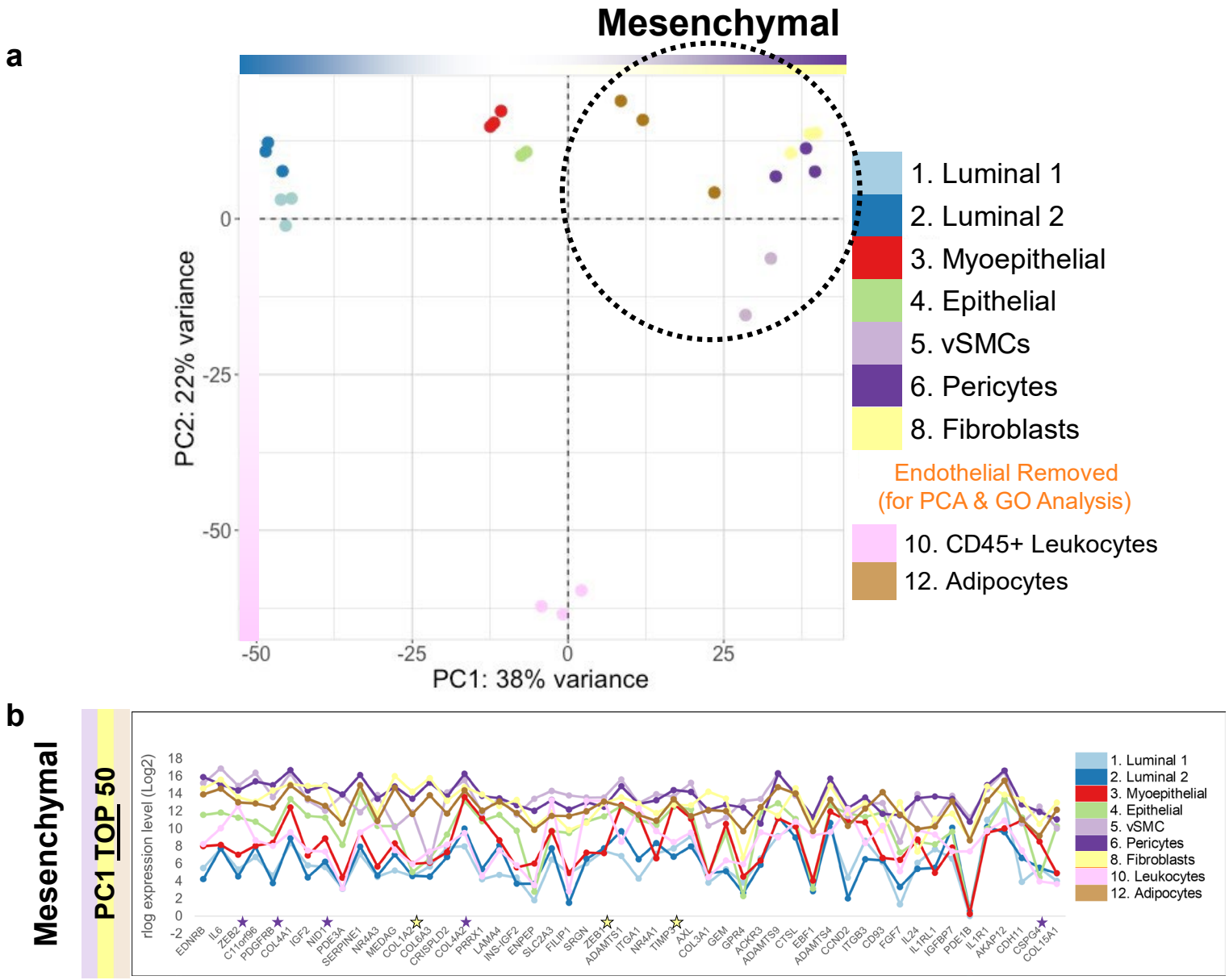


Figure 8

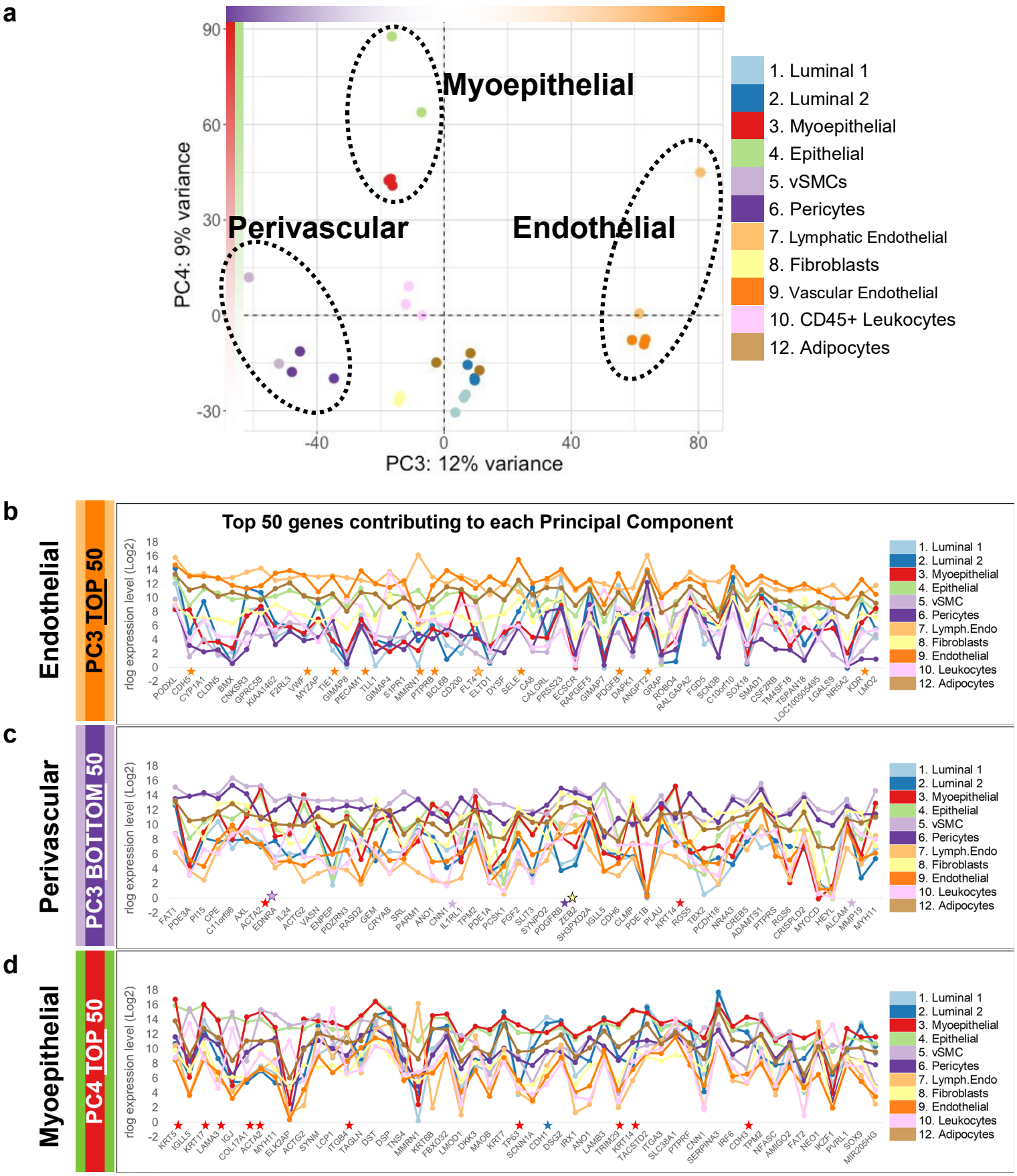

a
Myoepithelial
Perivascular
Endothelial
PC4: 9% variance
PC3: 12% variance
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. CD45+ Leukocytes
12. Adipocytes
b
Endothelial
PC3 TOP 50
Top 50 genes contributing to each Principal Component
rlog expression level (Log2)
c
Perivascular
PC3 BOTTOM 50
d
Myoepithelial
PC4 TOP 50

Figure 9

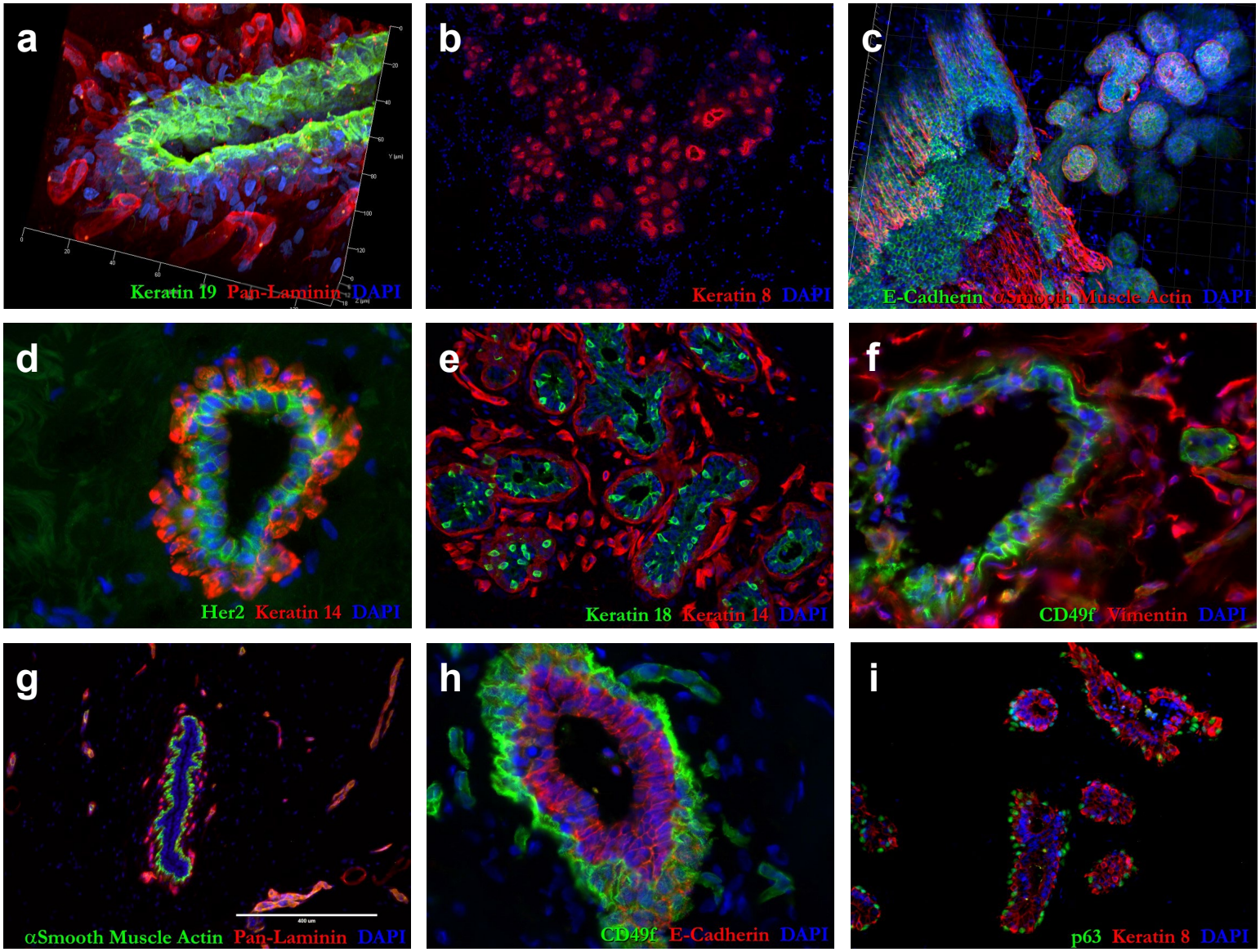

a
Keratin 19 Pan-Laminin DAPI
b
Keratin 8 DAPI
c
E-Cadherin αSmooth Muscle Actin DAPI
d
Her2 Keratin 14 DAPI
e
Keratin 18 Keratin 14 DAPI
f
CD49f Vimentin DAPI
g
αSmooth Muscle Actin Pan-Laminin DAPI
h
CD49f E-Cadherin DAPI
i
p63 Keratin 8 DAPI

Figure 10

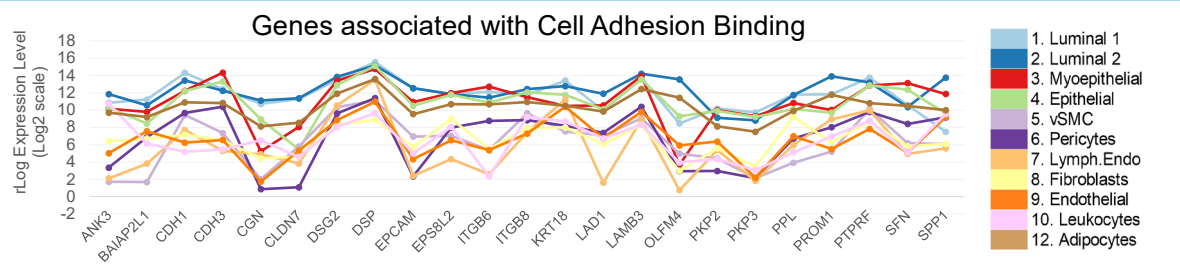
a

Principal Component 1: Epithelial Cell Types

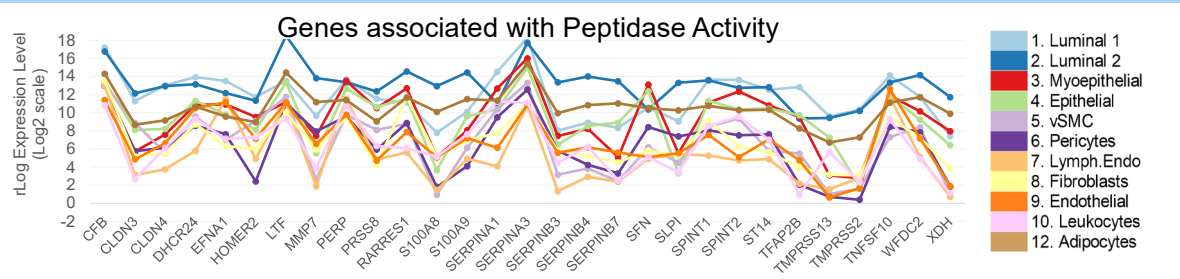
ER+ Luminal | ER- Luminal | Myoepithelial | Epithelial

(# pathways in group)	Representative Pathways	Adj. *p*-value	Associated Genes (from PC1 BOTTOM 200) Epithelial
2	cell adhesion molecule binding & cadherin binding	1.40E-08	[ANK3, BAIAP2L1, CDH1, CDH3, CGN, CLDN7, DSG2, DSP, EPCAM, EPS8L2, ITGB6, ITGB8, KRT18, LAD1, LAMB3, NECTIN4, OLFM4, PKP2 PKP3, PPL, PROM1, PTPRF, SFN, SHTN1, SPP1]
63	peptidase activity, hydrolase activity, Regulation of kinase activity…	6.05E-05	[ARHGEF5, ATP1B1, CLDN3, CLDN4, DHCR24, EFNA1, ERBB2, ESR1, LTF, PERP, PPP1R1B, PROM2, RARRES1, S100A8, S100A9, SERPINA1 SERPINA3, SERPINB3, SERPINB4, SERPINB7, SFN, SLPI, SPINT1, SPINT2, SYTL2, TFAP2B, TNFSF10, TPM1, VAV3, WFDC2, XDH]
9	keratinization, programmed cell death…	8.32E-05	[DSC2, DSG2, DSP, KRT15, KRT16, KRT18, KRT19, KRT23, KRT6B, KRT7 KRT8, PERP, PKP2, PKP3, PPL, PRSS8, ST14]
3	actin filament binding, cytoskeletal protein binding, & actin binding	2.68E-03	[ANK3, BAIAP2L1, CDH1, CDH3, CGN, CLU, COBL, EPB41L4B, EPS8L2 HOMER2, HOOK1, LIMCH1, MYH14, MYO5B, PLS1, PPP1R9A, PROM1, PRSS8, PSTPIP2, RAB27B, S100A8, S100A9, SHROOM3, SHTN1, TPM1
34	RTK signaling, ESR signaling, regulation by TFAP (AP-2) …	5.58E-03	[CYP27B1, EHF, ELF3, ESR1, FOXA1, GATA3, GRHL1, MUC1, RORC, SOX10, SOX9, TFAP2A, TFAP2B, TFAP2C, VDR]
1	calcium ion binding	6.13E-03	[ANXA3, CDH1, CDH3, CELSR2, CLU, DCHS2, DSC2, DSG2, GALNT3, LRP2, MGP, PADI2, PLS1, RASEF, S100A14, S100A8, S100A9, SYT7, TPD52]
5	Metabolism of vitamins and cofactors, ECM organization, glycosylation, vascular wall interactions, non-integrin ECM interactions	6.91E-03	[CDH1, DDR1, ITGB6, ITGB8, LAMB3, MMP7, SDC1, SDC4, SPP1]
2	Innate immune system	7.77E-03	[CFB, CHI3L1, CLU, DSP, FRK, LCN2, LTF, MGST1, MUC1, MUC4, OLFM4 PADI2, PIGR, S100A8, S100A9, SAA1, SERPINA1, SERPINA3, SERPINB SLPI, TLR2, VAMP8, VAV3, VNN1]

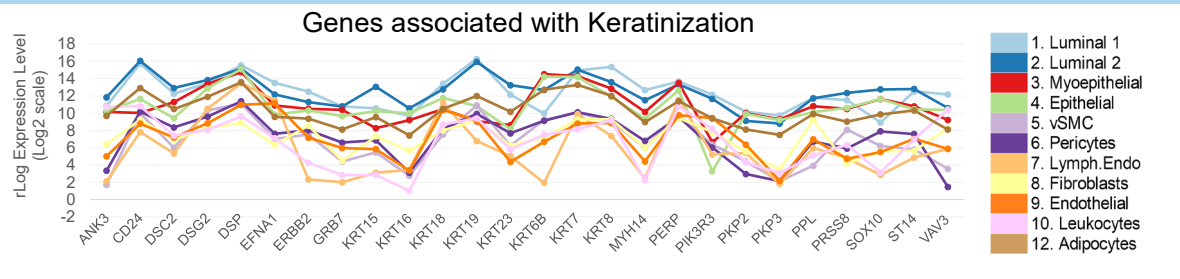
b



c



d



e

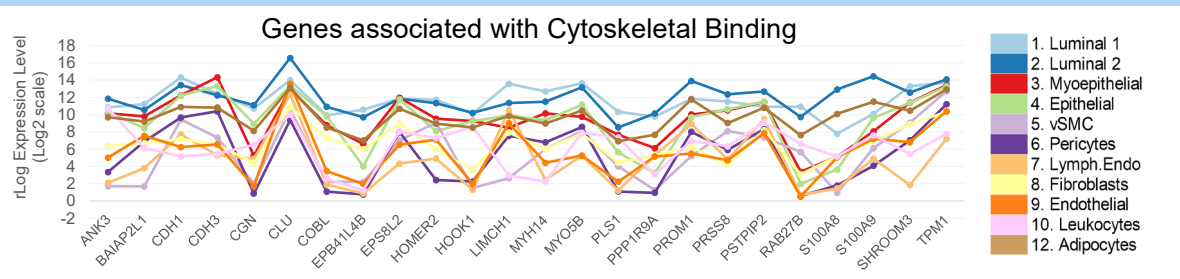


Figure 10-figure supplement 1

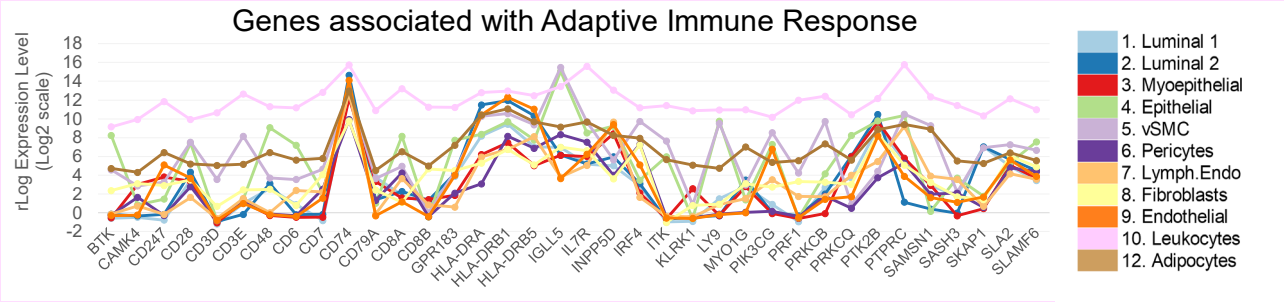
Principal Component 2: Leukocytes (Mesodermal)

Leukocytes

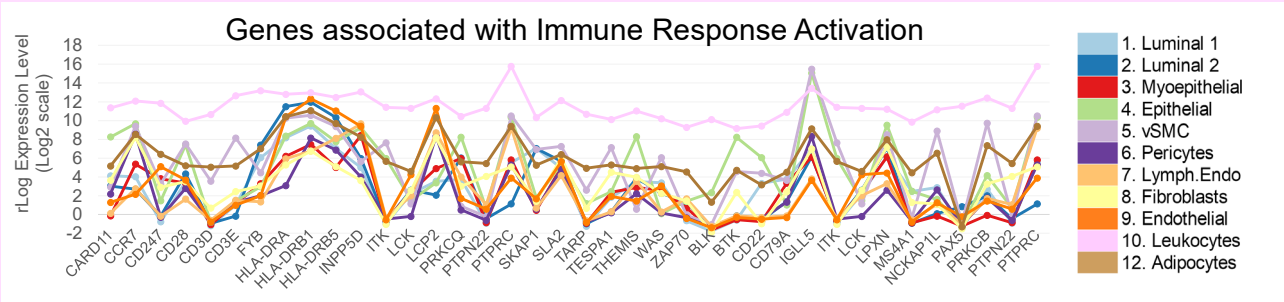
a

(# pathways in group)	Representative Pathways	*Adj. p*-value	Associated Genes (from PC2 BOTTOM 200) Leukocytes
6	adaptive immune response, lymphocyte activation, regulation of immune response …	6.07E-46	[BTK, CAMK4, CD247, CD28, CD3D, CD3E, CD48, CD6, CD7, CD74, CD79A, CD8A, CD8B, GPR183, HLADRA, HLA-DRB1, HLA-DRB5, IGLL5, IL7R, INPP5D, IRF4, ITK, JCHAIN, KLRK1, LY9, MYO1G, PIK3CG, PRF1, PRKCB, PRKCQ, PTK2B, PTPRC, SAMSN1, SASH3, SKAP1, SLA2, SLAMF6, SLAMF7, SPN, TARP, THEMIS, TNFRSF13C, WAS, ZAP70]
15	activation of immune response, T cell receptor signaling pathway, B cell receptor signaling pathway…	1.78E-41	[CARD11, CCR7, CD247, CD28, CD3D, CD3E, FYB1, HLADRA, HLA-DRB1, HLA-DRB5, INPP5D, ITK, LCK, LCP2, PRKCQ, PTPN22, PTPRC, SKAP1, SLA2, TARP, TESPA1, THEMIS, WAS, ZAP70] ; [BLK, BTK, CD22 CD79A, IGLL5, ITK, LCK, LPXN, MS4A1, NCKAP1L, PAX5, PRKCB, PTPN22, PTPRC]
141	leukocyte homeostasis, cytokine production, hemopoiesis, interleukin-10, -4, -2, -8 production, cell-cell adhesion…	2.42E-41	[CD74, CORO1A, DOCK10, DOCK11, GPR183, IL2RA, NCKAP1L, TNFRSF13C, TSC22D3]
25	T cell activation involved in immune response, Interleukin-17 production, T-helper 17 type immune response…	1.94E-26	[ADAM8, APBB1IP, ARHGAP45, ARHGAP9, ATP8A1, BIN2, BTK, CD28, CD53, CD74, CORO1A, DOCK10, DOCK11, DOCK2, GPR183, IQGAP2, IRF4, ITGAL, ITGB2, LCP1, LRMP, LY9, NCKAP1L, PIK3CG, PTGER4, PTK2B, PTPRC, RAC2, RASGRP1, SELL, SLAMF6, SPN]
53	leukocyte chemotaxis and migration, T cell chemotaxis, cellular extravasation…	5.88E-23	[ADAM8, CCL5, CCR5, CCR7, CD74, CORO1A, CXCR4, GPR183, IL16, ITGB2, JAML, KLRK1, NCKAP1L, PIK3CG, PREX1, PTK2B, RAC2, RIPOR STK4, VAV1]
6	phagocytosis, Fc-gamma receptor signaling pathway involved in phagocytosis…	8.85E-22	[BIN2, CD247, CORO1A, CYFIP2, DOCK2, IGLL5, IL2RB, IL2RG, ITGAL, ITGB2, MYO1G, NCKAP1L, PTPRC, RAC2, TARP, VAV1, WAS, WIPF1]
36	Interleukin-6 production, interleukin-1 secretion…	5.64E-17	[BANK1, CD74, IL16, INPP5D, LCP1, NCKAP1L, PTPN22]
10	acute inflammatory response…	3.69E-14	[ADAM8, ALOX5AP, BTK, CCR7, CD6, PIK3CG]
9	regulation of GTPase activity, Ras - and Rho- guanyl-nucleotide exchange factor activity…	1.29E-11	[ACAP1, ARAP2, ARHGAP15, ARHGAP30, ARHGAP45, ARHGAP9, ARHGDIB, CCL5, CCR7, DOCK10, DOCK11, DOCK2, DOCK8, FGD3, IQGAP2, KLRK1, NCKAP1L, PREX1, PTK2B, RASAL3, RASGRP1, RASGRP2, RGS1, RGS2, RHOH, SMAP2, TAGAP, VAV1]
19	mast cell degranulation, inflammatory response to antigenic stimulus…	3.44E-11	[BTK, CCR7, CD28, CD6, HLADRB1, IL2RA, RASGRP1]
6	monocyte differentiation…	9.53E-11	[EVI2B, HCLS1, INPP5D]

b



c



d

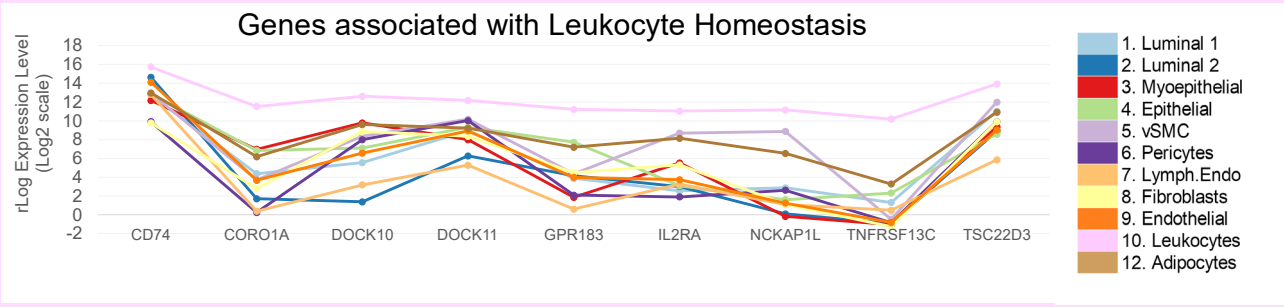


Figure 11

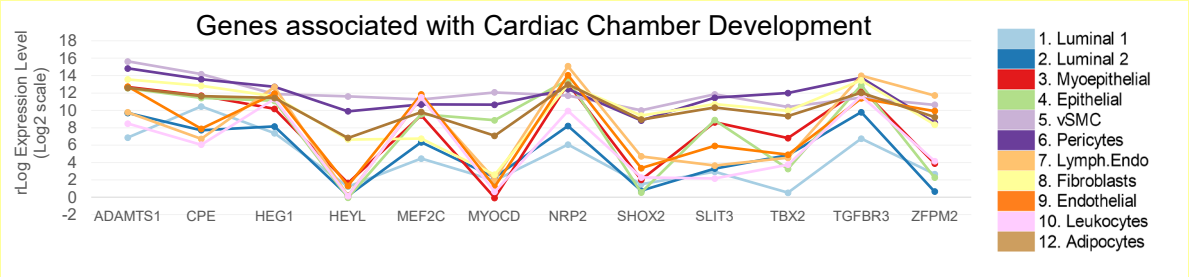
a

Principal Component 1b: Mesenchymal Cell Types

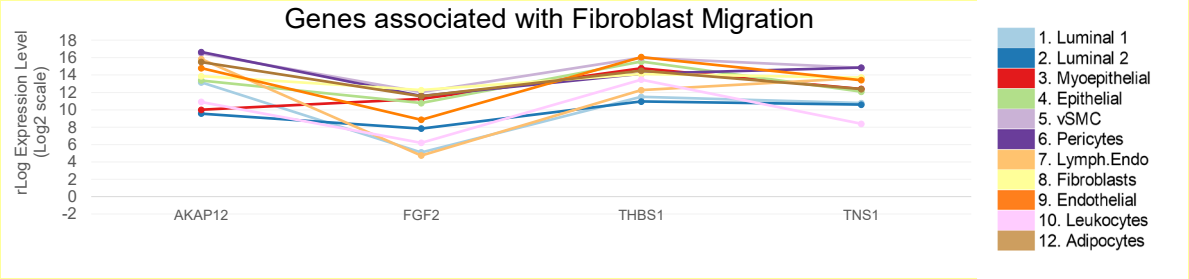
vSMCs | Pericytes | Fibroblasts | Adipocytes

(# pathways in group)	Representative Pathways	*Adj. p*-value	Associated Genes (from PC1 TOP 200) Mesenchymal
94	cardiac muscle tissue development, septum development, cartilage morphogenesis, regulation of vSMC proliferation …	5.00E-33	[ADAMTS1, CPE, HEG1, HEYL, MEF2C, MYOCD, NRP2, SHOX2, SLIT3, TBX2, TGFBR3, ZFPM2]
85	Fibroblast migration, endothelial migration, epithelial migration, regulation of angiogenesis…	5.37E-31	[ADAMTS1, ADAMTS9, ADGRA2, ANGPT2, COL4A2, CXCL8, DCN, EMILIN1, ETS1, FGF2, GPR4, HGF, HSPB6, IL6, ITGA5, PTGS2, SERPINE1, SPARC, TCF4, THBS1, THBS2, THBS4, TWIST1]
77	positive regulation of peptidyl-tyrosine phosphorylation, cardiac epithelial to mesenchymal transition, GAG metabolic process, regulation of chemokine production, …	7.40E-21	[BMP6, CSPG4, FGF7, FYN, HGF, IGF2, IL24, IL6, ITGA5, ITGB3, LIF, NR4A3, THBS4]
35	chondrocyte, cartilage, skeletal system development; regulation of mesenchymal cell proliferation …	1.67E-18	[ADAMTS12, ADAMTS4, BGN, BMP6, CDH11, CHST11, COL1A2, COL3A1 COL5A2, COL6A1, COL6A2, COL6A3, FGF2, FGFR1, GLI2, IGF2, MEF2C MMP2, PDGFRB, PRRX1, SATB2, SFRP4, SHOX2, SPARC, TBX15, TWIST1, ZEB1]
31	bone development, bone morphogenesis, regulation of histone acetylation…	4.53E-18	[BGN, BMP6, COL6A1, COL6A2, COL6A3, MEF2C, SFRP4, SHOX2, SPARC, TWIST1]
27	positive regulation of blood vessel endothelial cell migration…	1.55E-19	[ETS1, FGF2, FGFR1, PTGS2, THBS1]
25	smooth muscle contraction, vascular smooth muscle contraction…	4.58E-12	[ACTA2, CAV1, EDNRA, EDNRB, MYH11, MYOCD, PDE4D, PTGS2, SSTR2]
19	negative regulation of calcium ion transport, muscle contraction…	1.29E-11	[CAV1, GEM, KCNE4, PLN, PTGS2, SFRP4, SYT11, THBS1, TWIST1]
18	regulation of fatty acid biosynthetic process, prostaglandin metabolic process…	3.84E-12	[AVPR1A, PDK4, PTGS2]
17	glycosaminoglycan catabolic process, dermatan sulfate biosynthesis, positive chemotaxis…	4.01E-08	[ANGPT2, CXCL8, FGF2, FGF7, HGF, IL6]

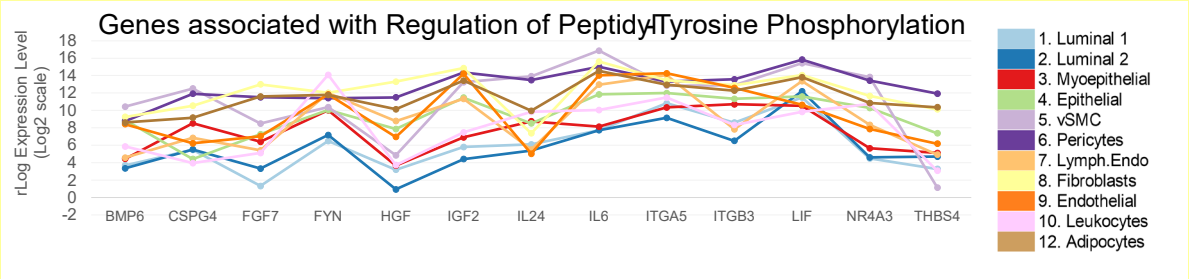
b



c



d



e

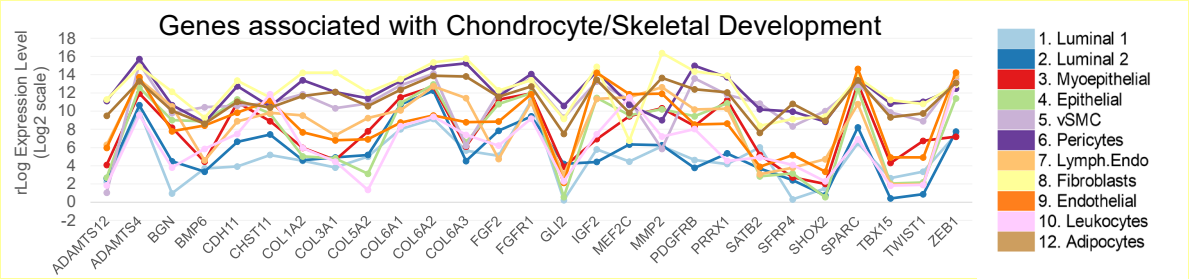


Figure 12

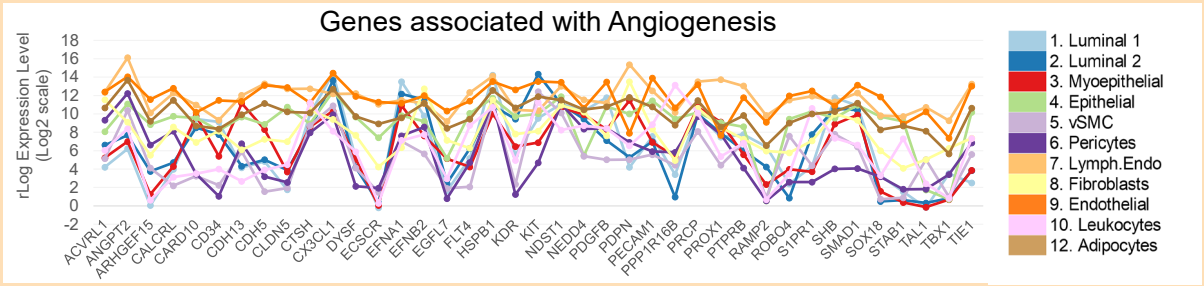
a

Principal Component 3: Endothelial Cell Types

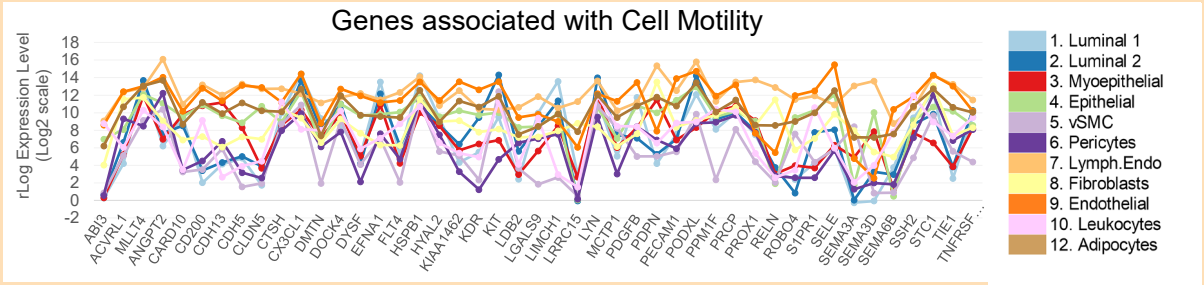
Lymphatic Endothelial | Vascular

(# pathways in group)	Representative Pathways	Adj. *p*-value	Associated Genes (from PC3TOP 200) Endothelial
65	Signaling by VEGF, endothelial cell proliferation, angiogenesis, positive regulation of vasculogenesis…	6.27E-19	[ACVRL1, ANGPT2, ARHGEF15, CALCRL, CARD10, CD34, CDH13, CDH5, CLDN5, CTSH, CX3CL1, DYSF, ECSCR, EFNA1, EFNB2, EGFL7, FLT4, HSPB1, JCAD, KDR, KIT, NDST1, NEDD4, PDGFB, PDPN, PECAM1, PPP1R16B, PRCP, PROX1, PTPRB, RAMP2, ROBO4, S1PR1, SHB, SMAD1, SOX18, STAB1, TAL1, TBX1, TIE1]
8	regulation of cell motility, regulation of cell migration…	1.33E-16	[ABI3, ACVRL1, AFDN, ANGPT2, CARD10, CD200, CDH13, CDH5, CLDN5, CTSH, CX3CL1, DMTN, DOCK4, DYSF, EFNA1, FLT4, HSPB1, HYAL2, JCAD, KDR, KIT, LDB2, LGALS9, LIMCH1, LRRC15, LYN, MCTP1, PDGFB, PDPN, PECAM1, PODXL, PPM1F, PRCP, PROX1, RELN, ROBO4, S1PR1, SELE, SEMA3A, SEMA3D, SEMA6B, SSH2, STC1, TIE1, TNFRSF14]
18	cell-matrix adhesion, regulation of chemotaxis, positive regulation of blood vessel endothelial cell migration…	5.83E-14	[ACVRL1, CD34, CDH13, CX3CL1, DMTN, KDR, LDB2, LIMCH1, LYPD5, LYVE1, PECAM1, PPM1F]
16	Endothelial-, muscle- cell differentiation, lymph vessel morphogenesis, lymphangiogenesis , aorta development…	9.18E-11	[ACVRL1, AFDN, CDH5, CLDN5, F11R, KDR, PDPN, PECAM1, PPP1R16B, PROX1, ROBO4, S1PR1, SOX18, STC1]
41	Cell-cell junction assembly, regulation of vascular permeability, tight junction organization, regulation of establishment of endothelial barrier, endothelial cell apoptotic process…	1.61E-12	[ACVRL1, AFDN, CDH13, CDH5, CLDN5, DCHS1, DMTN, EFNB2, F11R, FSCN1, KDR, LDB2, LIMCH1, MPP7, PECAM1, PPM1F, PRKCH, RAMP2, SHANK3]
12	regulation of endothelial cell migration, negative regulation of chemotaxis, negative regulation of blood vessel endothelial cell migration…	1.38E-11	[ACVRL1, ANGPT2, CARD10, EFNA1, FLT4, HSPB1, JCAD, KDR, PDGFB, PRCP, PROX1, STC1]
40	regulation of coagulation, regulation of hemostasis, regulation of platelet activation, positive regulation of phosphatidylinositol 3 -kinase activity…	9.22E-09	[CD34, DMTN, F11R, LYN, PDGFB, PDPN, TLR4] [AFDN, CNKSR3, CX3CL1, DMTN, EFNB2, F11R, LGALS9, LYN, MAGI1, PDPN, PODXL, SELE, TNFRSF14]
9	regulation of blood vessel endothelial cell migration, regulation of cell - matrix adhesion, negative regulation of angiogenesis…	1.27E-08	[ACVRL1, ANGPT2, CARD10, EFNA1, HSPB1, JCAD, KDR, PDGFB, PRCP]
15	Signaling by VEGF, regulation of vascular permeability, vasculogenesis…	3.53E-08	[CDH5, FLT4, HSPB1, KDR, MAPK11, SHB]
37	peptidyl-tyrosine modification, Signaling by SCF -KIT, regulation of lipid kinase activity, monocyte chemotaxis…	4.93E-08	[BMX, CSF2, CSF3, DMTN, EFNA1, FLT4, GPRC5B, HYAL2, KDR, KIT, LYN, PDGFB, PECAM1, RELN, STAP2, TAL1, TIE1, TNFRSF14]

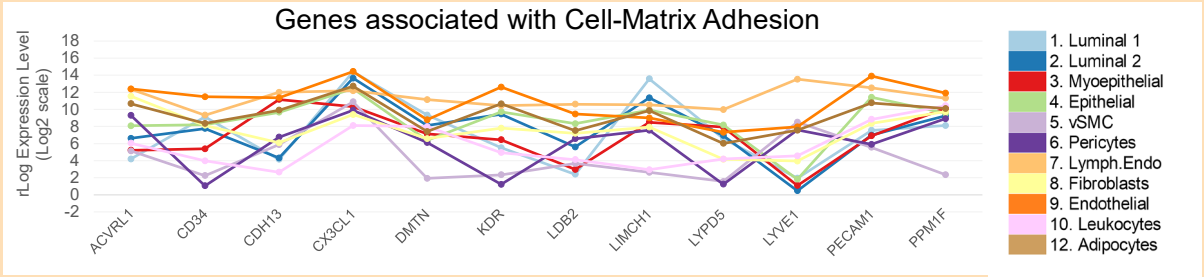
b



c



d



e

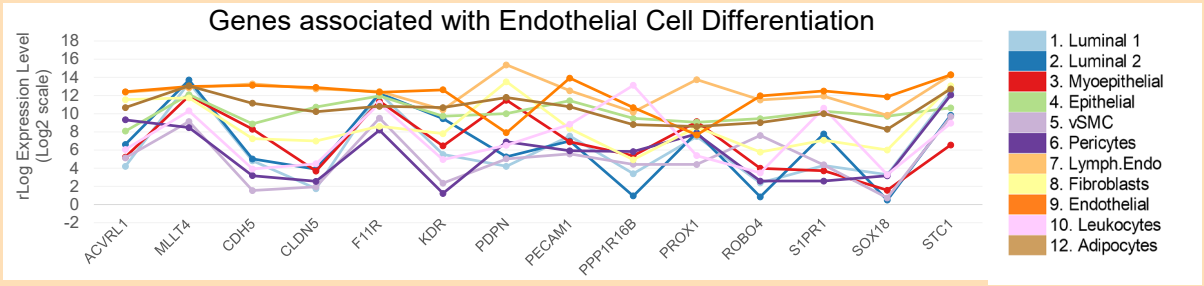


Figure 12—figure supplement 1

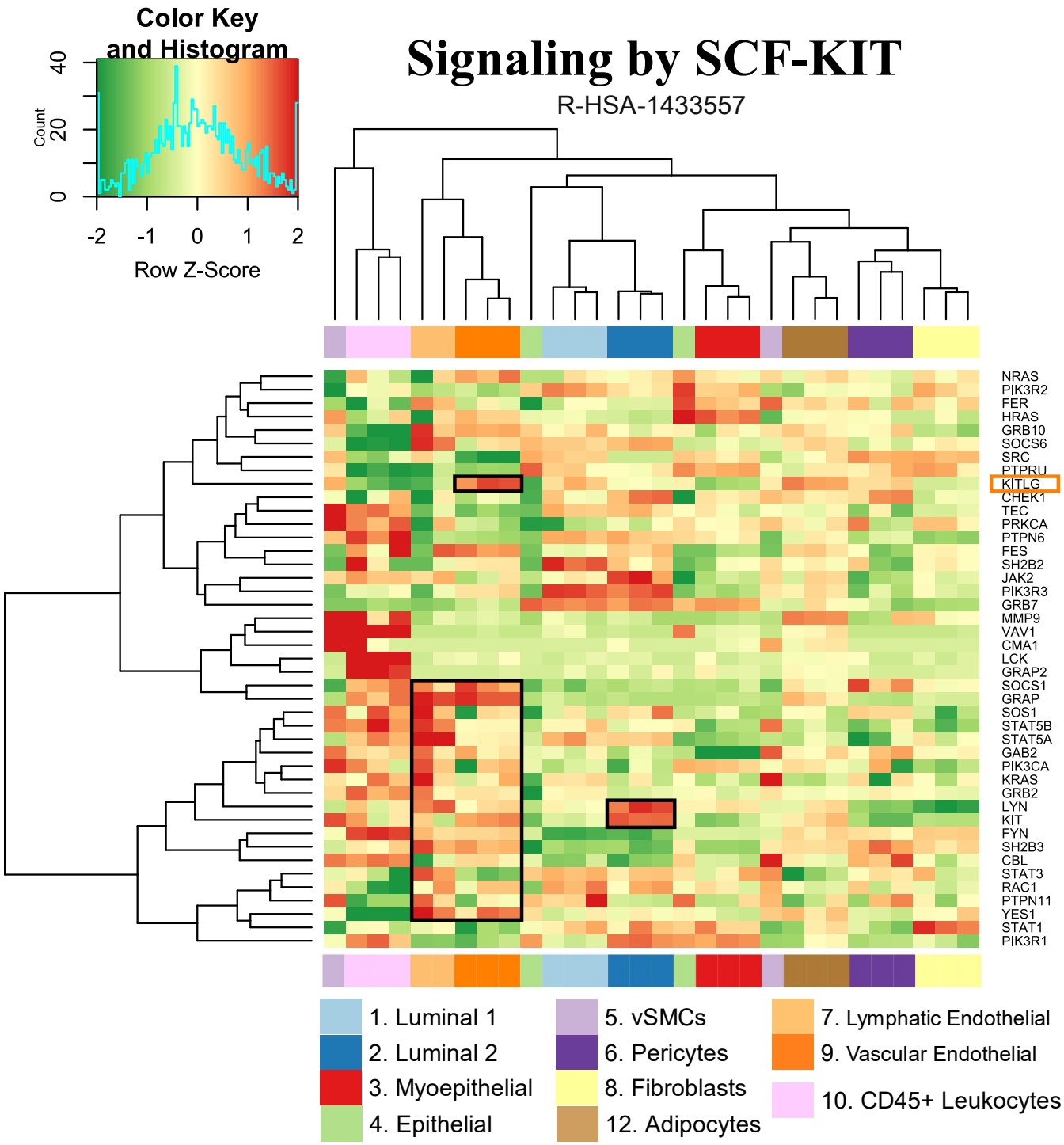


Figure 12—figure supplement 2

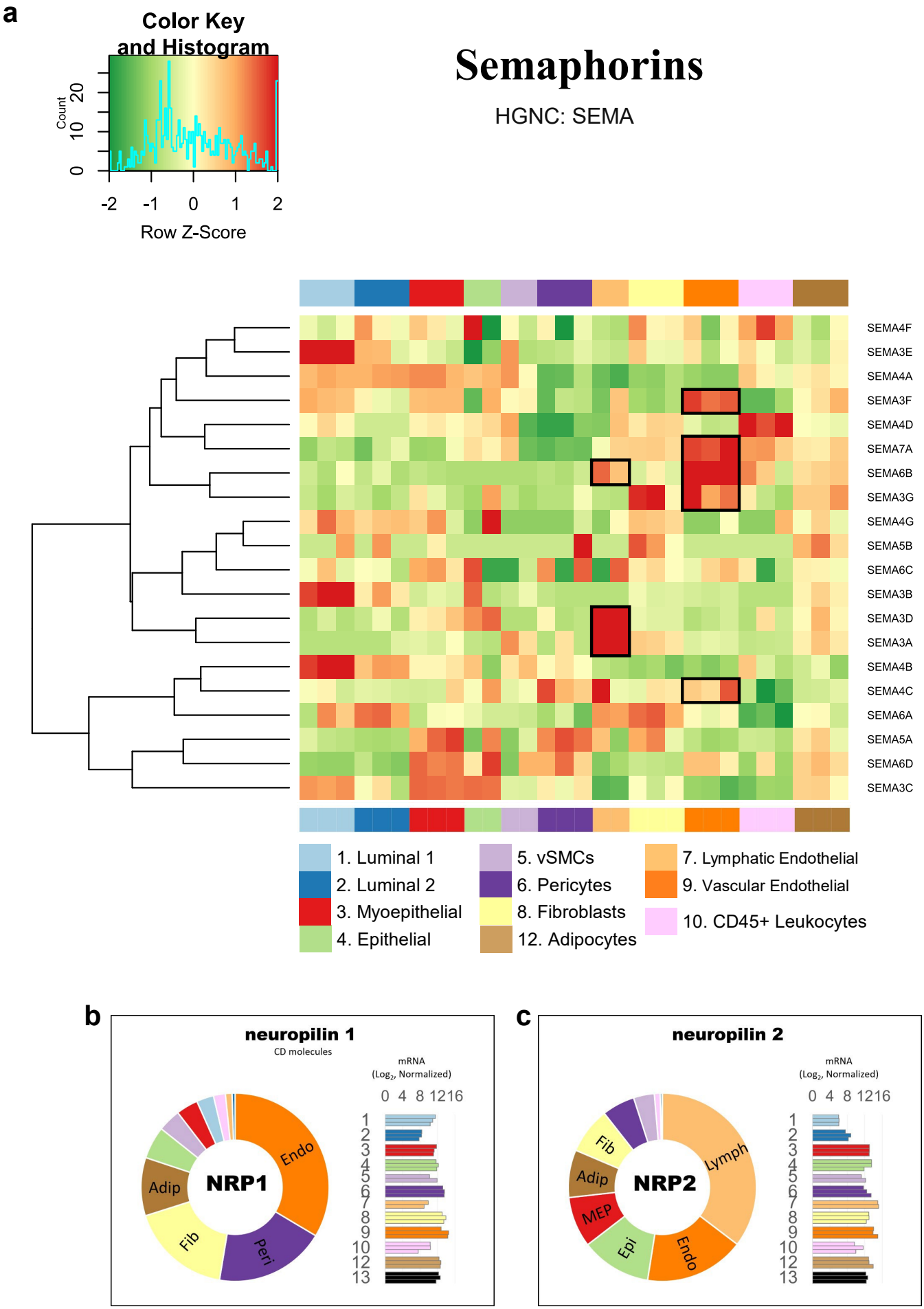


Figure 13

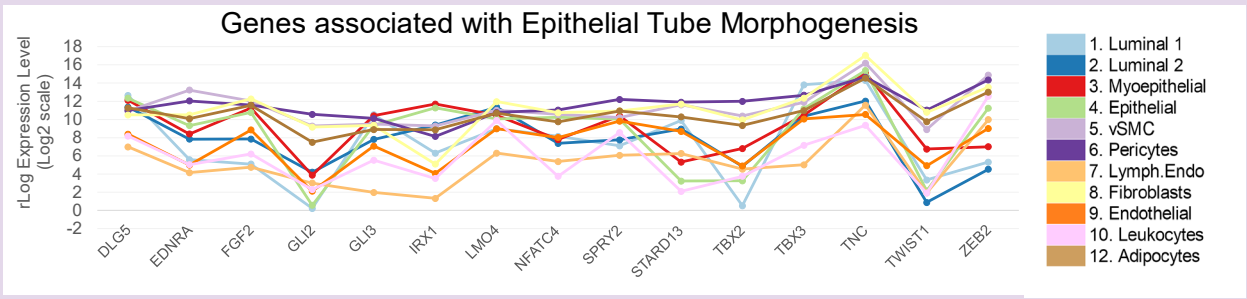
Principal Component 3: Perivascular Cell Types

a

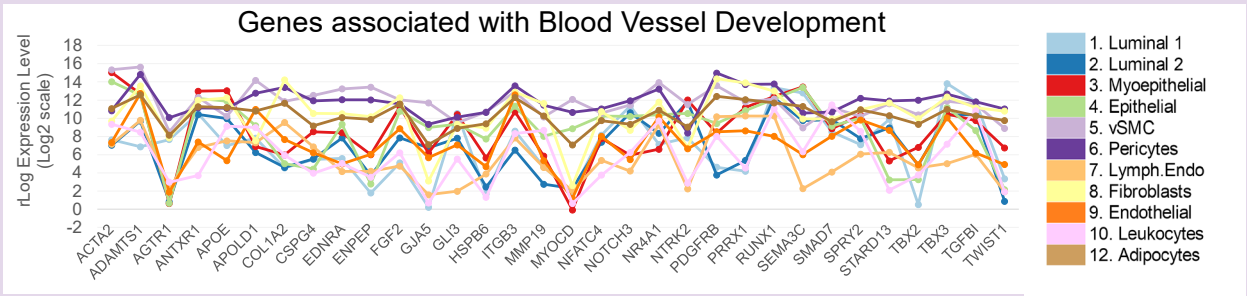
vSMCs | Pericytes

(# pathways in group)	Representative Pathways	Adj. *p*-value	Associated Genes (from PC3 BOTTOM 200) **Perivascular**
39	Epithelial tube morphogenesis, neuron fate commitment, branching morphogenesis of an epithelial tube, cell fate commitment…	2.33E-17	[DLG5, EDNRA, FGF2, GLI2, GLI3, IRX1, LMO4, NFATC4, SPRY2, STARD13, TBX2, TBX3, TNC, TWIST1, ZEB2]
5	Blood vessel development, vasculature development, urogenital system development, tissue morphogenesis, tube development	4.25E-16	[ACTA2, ADAMTS1, AGTR1, ANTXR1, APOE, APOLD1, COL1A2, CSPG4, EDNRA, ENPEP, FGF2, GJA5, GLI3, HSPB6, ITGB3, MMP19, MYOCD, NFATC4, NOTCH3, NR4A1, NTRK2, PDGFRB, PRRX1, RUNX1, SEMA3C, SMAD7, SPRY2, STARD13, TBX2, TBX3, TGFBI, TWIST1]
9	Muscle contraction, regulation of blood vessel diameter, regulation of systemic arterial blood pressure by hormone, regulation of blood pressure…	1.92E-14	[ACTA2, ACTG2, ADRA2A, CACNA1C, CACNA1G, CACNA1H, CNN1, CRYAB, EDNRA, EDNRB, GJA5, HSPB6, KCNE4, KCNJ8, LMOD1, MYH11, MYL9, MYOCD, PLN, SLMAP, SMAD7, SSTR2, TPM2]
21	Extracellular matrix organization, ECM proteoglycans, NCAM1 interactions, Degradation of the extracellular matrix…	5.74E-14	[ADAMTS1, ADAMTS5, CD44, COL14A1, COL1A2, COL5A3, COL6A1, COL6A2, FGF2, ITGA7, ITGB3, MMP19, PTPRS, SDC2, SH3PXD2A, TNC]
13	Artery development, cardiac muscle tissue development, muscle tissue development mesenchyme morphogenesis…	1.17E-10	[APOE, GJA5, GLI3, MYOCD, NOTCH3, PDGFRB, PRRX1, SMAD7, TBX2]
13	Cardiac muscle cell action potential, NCAM signaling for neurite out - growth, muscle contraction…	2.29E-10	[CACNA1C, CACNA1G, GJA5, KCNE4, KCNJ8, SLMAP]
9	Smooth Muscle Contraction, muscle contraction, regulation of muscle system process…	1.27E-10	[ACTA2, ADRA2A, CNN1, EDNRA, EDNRB, MYH11, MYOCD, SSTR2]
4	Cell fate commitment, cartilage development, neuron fate commitment	5.22E-07	[FGF2, GLI2, GLI3, LMO4, NOTCH3, PRRX1, SATB2, SPRY2, TBX15, TBX18, TBX2, TBX3, TEAD3]
7	Syndecan interactions, ECM proteoglycans, Integrin cell surface interactions…	1.06E-06	[COL1A2, COL5A3, FGF2, ITGB3, SDC2, TNC]
6	Mesenchymal cell differentiation, mesenchymal cell development, atrioventricular valve development…	3.16E-06	[DLG5, EDNRA, EDNRB, HEYL, SEMA3C, SMAD7, SPRY2, TWIST1, VASN, ZEB2]

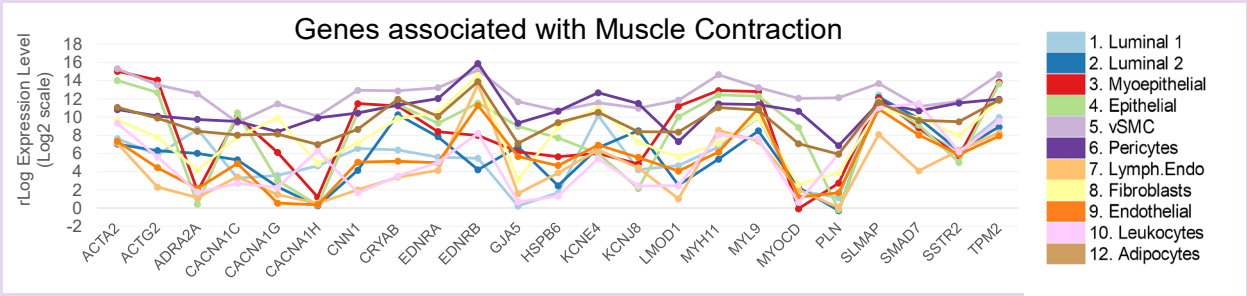
b



c



d



e

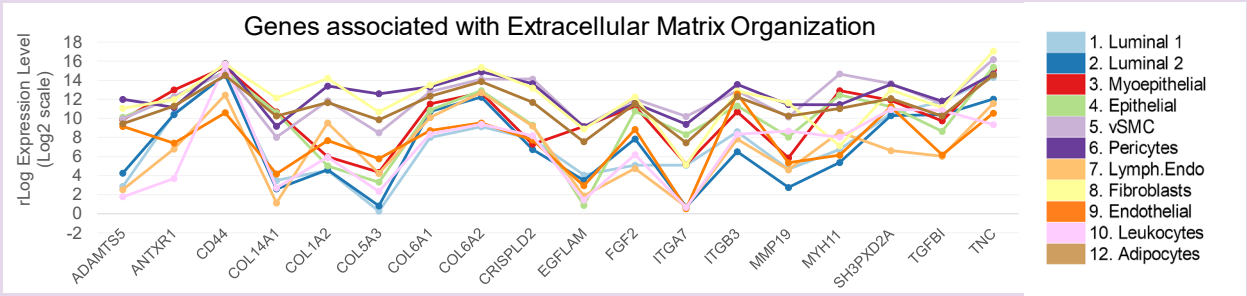


Figure 14

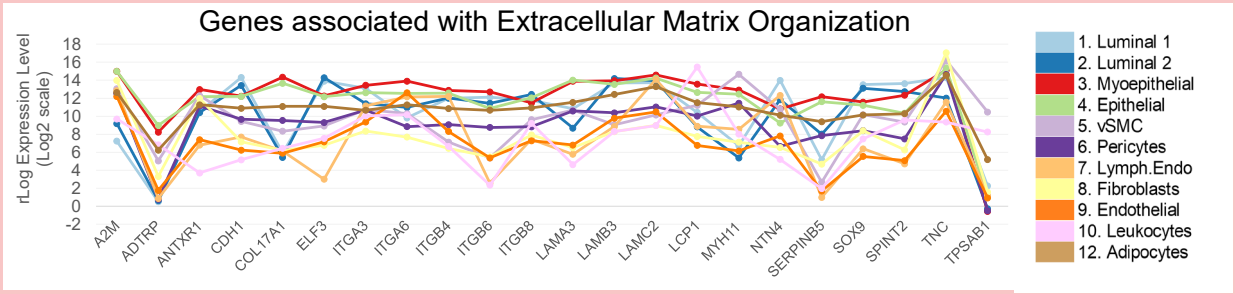
a

Principal Component 4: Myoepithelial Cells

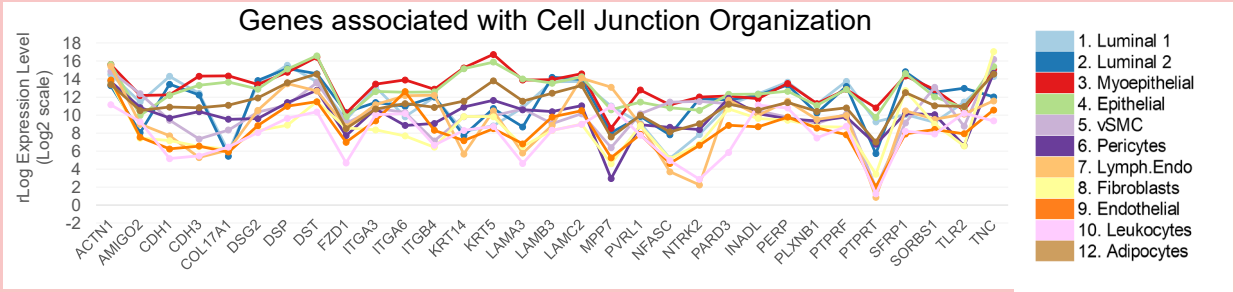
Myoepithelial Cells	Pop4 Epithelial

(# pathways in group)	Representative Pathways	Adj. *p*-value	Associated Genes (from PC4 TOP 200) Myoepithelial / Epithelial
24	Extracellular matrix organization, Laminin interactions, hemidesmosome assembly, Cell-Cell communication, Degradation of the extracellular matrix, …	3.46E-16	[A2M, ADTRP, ANTXR1, CDH1, COL17A1, ELF3, ITGA3, ITGA6, ITGB4, ITGB6, ITGB8, LAMA3, LAMB3, LAMC2, LCP1, MYH11, NTN4, SERPINB5, SOX9, SPINT2, TNC, TPSAB1]
12	Cell junction organization, desmosome organization, apoptosis,…	1.00E-14	[ACTN1, AMIGO2, CDH1, CDH3, COL17A1, DSG2, DSP, DST, FZD1, ITGA3, ITGA6, ITGB4, KRT14, KRT5, LAMA3, LAMB3, LAMC2, MPP7, NECTIN1, NFASC, NTRK2, PARD3, PATJ, PERP, PLXNB1, PTPRF, PTPRT, SFRP1, SORBS1, TLR2, TNC]
9	Keratinization, skin development, formation of the cornified envelope, …	8.47E-14	[CDH3, DSC3, DSG2, DSG3, DSP, KRT14, KRT17, KRT5, KRT6B, KRT6C, KRT7, PERP, PRSS8, SFN, ST14]
6	Epidermal cell differentiation, prostate glandular acinus development, hair follicle development, …	6.03E-11	[CDH3, DSC3, DSG2, DSG3, DSP, ESRP1, IRF6, KRT14, KRT17, KRT5, KRT6B, KRT6C, KRT7, PALLD, PERP, PRSS8, SFN, ST14, TP63, VDR]
18	Morphogenesis of a branching epithelium, stem cell differentiation, tissue morphogenesis, prostate glandular acinus development, …	2.27E-10	[FGF1, FGFR2, LGR4, NTN4, SEMA3C, SFRP1, SOX10, SOX9, SPINT2, ST14, TACSTD2, TNC, TP63, VDR]
6	Cell-cell junction organization, cell-substrate junction assembly, cell junction organization, …	8.52E-10	[CDH1, CDH3, DSG2, DSP, MPP7, NECTIN1, NFASC, PARD3, PATJ, PERP]
7	Epithelial tube formation, branching involved in labyrinthine layer morphogenesis, epithelial cell development,…	1.32E-07	[FGFR2, FZD1, IRX1, SFRP1, SHROOM3, SOX9, SPINT2, ST14]
10	Actin filament bundle assembly,...	2.07E-06	[ACTN1, CGNL1, EZR, LCP1, PIK3R1, SFRP1, SORBS1, SYNPO2, TACSTD2]
2	Intermediate filament cytoskeleton organization, intermediate filament organization	1.63E-05	[DSP, DST, EPPK1, KRT14, KRT17, KRT6C, SYNM]
1	Regulation of Wnt signaling pathway	3.16E-05	[CDH3, DIXDC1, DKK3, EGFR, FGFR2, FZD1, ITGA3, LGR4, PLEKHA4, RNF43, SFRP1, SLC25A23, SOX10, SOX9, TLR2, WIF1]
2	Smooth muscle contraction, muscle contraction	4.59E-05	[ACTA2, ACTG2, LMOD1, MYH11, MYLK, SORBS1, TPM2]

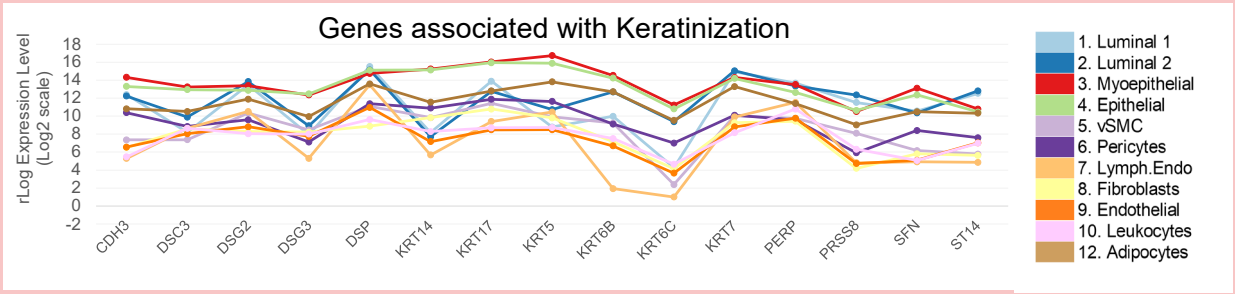
b



c



d



e

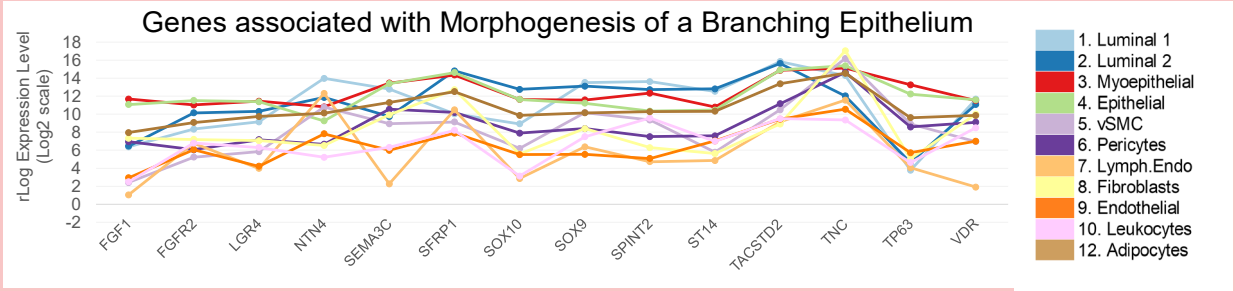


Figure 15

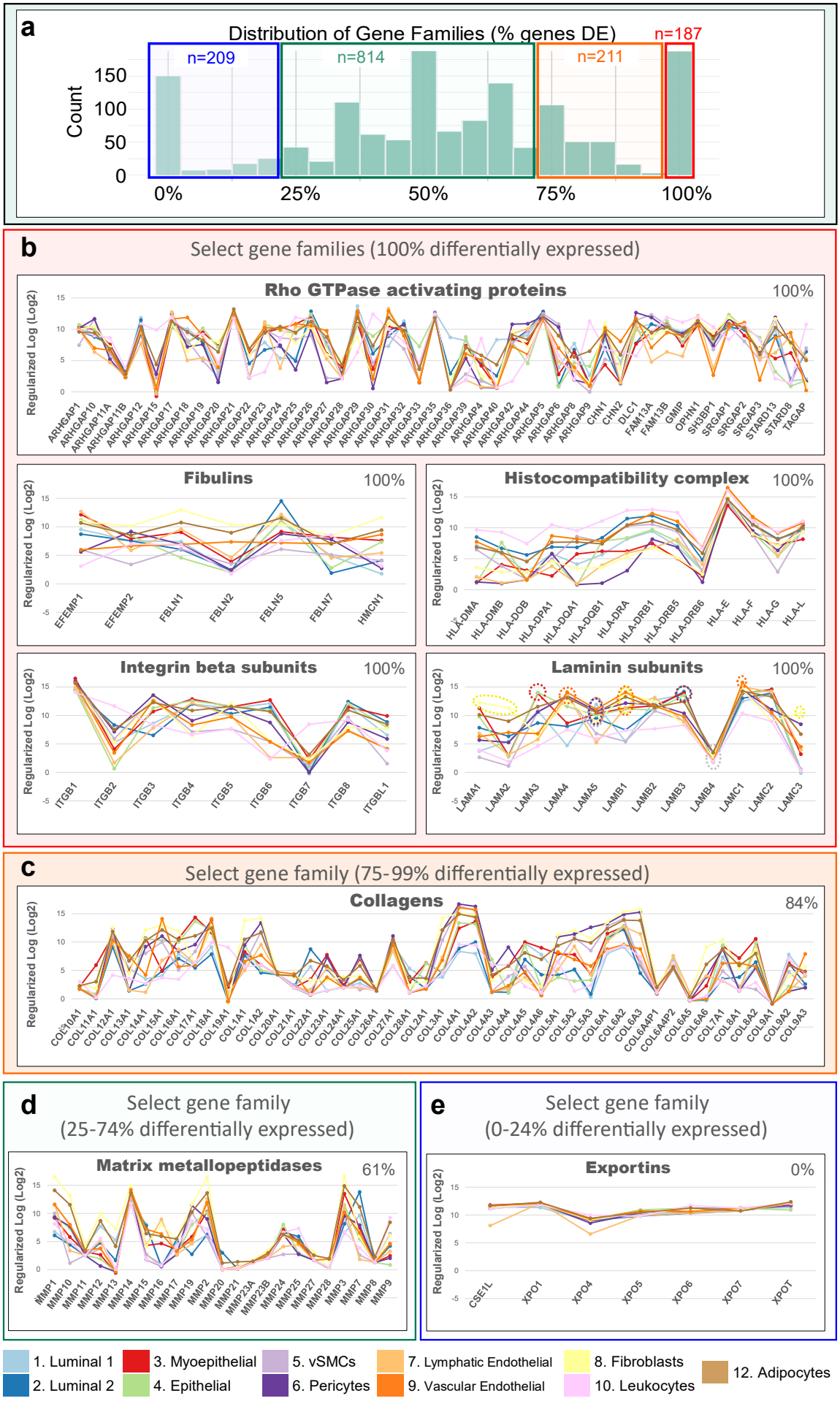

a
Distribution of Gene Families (% genes DE)
n=209
n=814
n=211
n=187
Count
150
100
50
0
0%
25%
50%
75%
100%
b
Select gene families (100% differentially expressed)
Rho GTPase activating proteins
100%
Regularized Log (Log2)
Fibulins
100%
EFEMP1
EFEMP2
FBLN1
FBLN2
FBLN5
FBLN7
HMCN1
Histocompatibility complex
100%
HLA-DMA
HLA-DMB
HLA-DOB
HLA-DPA1
HLA-DQA1
HLA-DQB1
HLA-DRA
HLA-DRB1
HLA-DRB5
HLA-DRB6
HLA-E
HLA-F
HLA-G
HLA-L
Integrin beta subunits
100%
ITGB1
ITGB2
ITGB3
ITGB4
ITGB5
ITGB6
ITGB7
ITGB8
ITGBL1
Laminin subunits
100%
LAMA1
LAMA2
LAMA3
LAMA4
LAMA5
LAMB1
LAMB2
LAMB3
LAMB4
LAMC1
LAMC2
LAMC3
c
Select gene family (75-99% differentially expressed)
Collagens
84%
d
Select gene family
(25-74% differentially expressed)
Matrix metallopeptidases
61%
MMP1
MMP10
MMP11
MMP12
MMP13
MMP14
MMP15
MMP16
MMP17
MMP19
MMP2
MMP20
MMP21
MMP23A
MMP23B
MMP24
MMP25
MMP27
MMP28
MMP3
MMP7
MMP8
MMP9
e
Select gene family
(0-24% differentially expressed)
Exportins
0%
CSE1L
XPO1
XPO4
XPO5
XPO6
XPOT
XPOT
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
9. Vascular Endothelial
8. Fibroblasts
10. Leukocytes
12. Adipocytes

Figure 15-table 1

Gene Family	DE Genes	Genes in Family
Adenylate cyclases	9	10
Aldehyde dehydrogenases	16	19
Collagens	38	45
Death inducing signaling complex	6	6
DOCK family Rho GEFs	11	11
Ephrins	6	8
Erb-b2 receptor tyrosine kinases	4	4
Fibulins	7	7
Glutathione S-transferases	15	23
GTPases, IMAP	7	7
Histocompatibility complex	14	14
Histones	37	89
Integrin alpha subunits	17	18
Integrin beta subunits	9	9
Interleukins	18	37
Interleukin receptors	34	42
Keratins, Type I	7	26
Keratins, Type II	10	26
Kruppel like factors	14	17
Laminin subunits	12	12
Lysophosphatidic acid receptors	6	6
MAP kinase phosphatases	10	11
Matrix metallopeptidases	14	23
Mucins	10	14
Plexins	9	9
Ras association domain family	10	10
Rho GTPase activating proteins	46	48
Suppressors of cytokine signaling	8	8
Tetraspanins	22	33
Toll like receptors	9	10
V-type ATPases	10	23

Fig.15—figure supplement 1

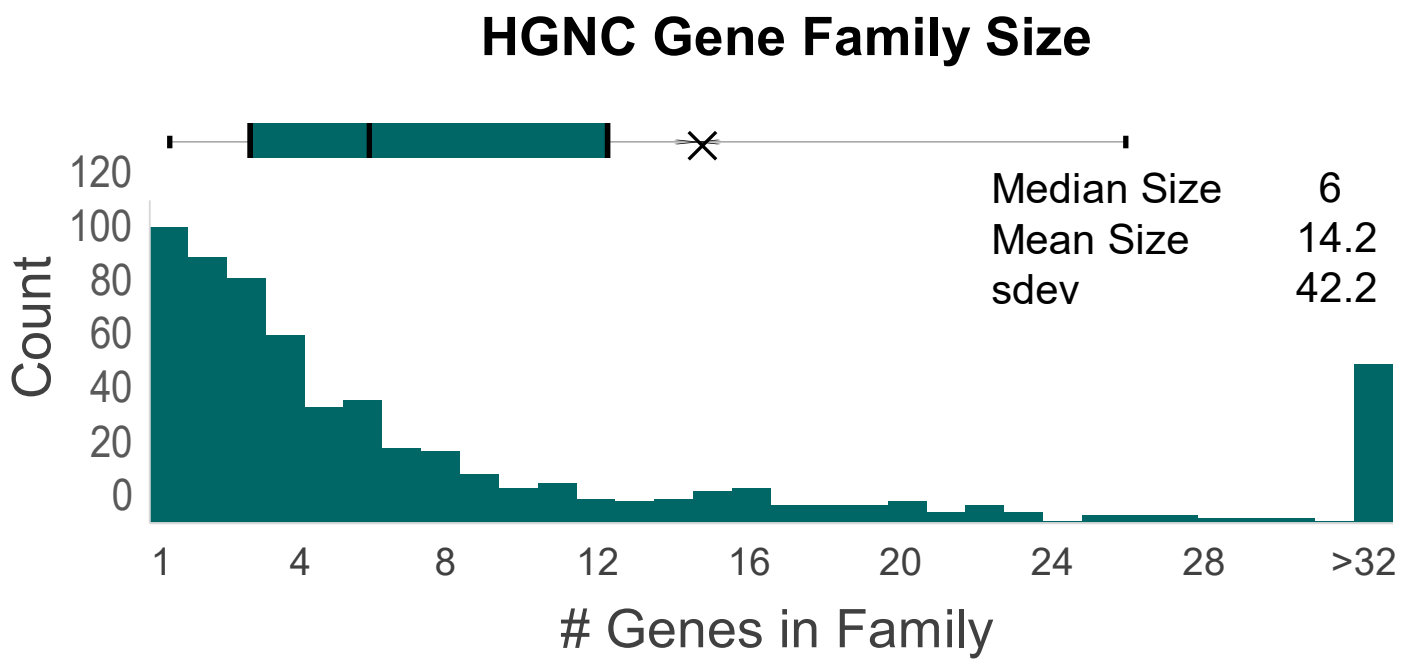
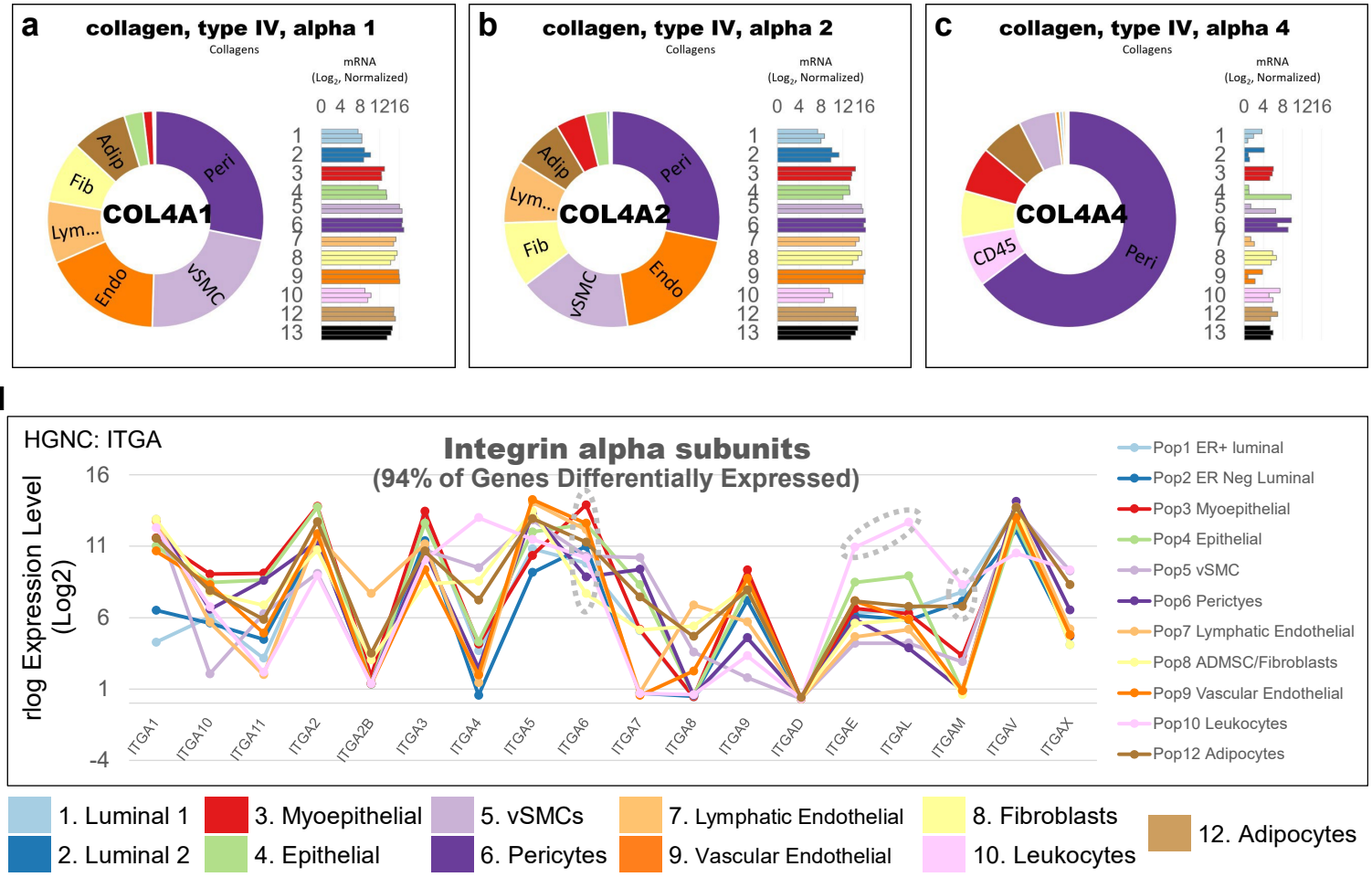


Figure 15—figure supplement 2



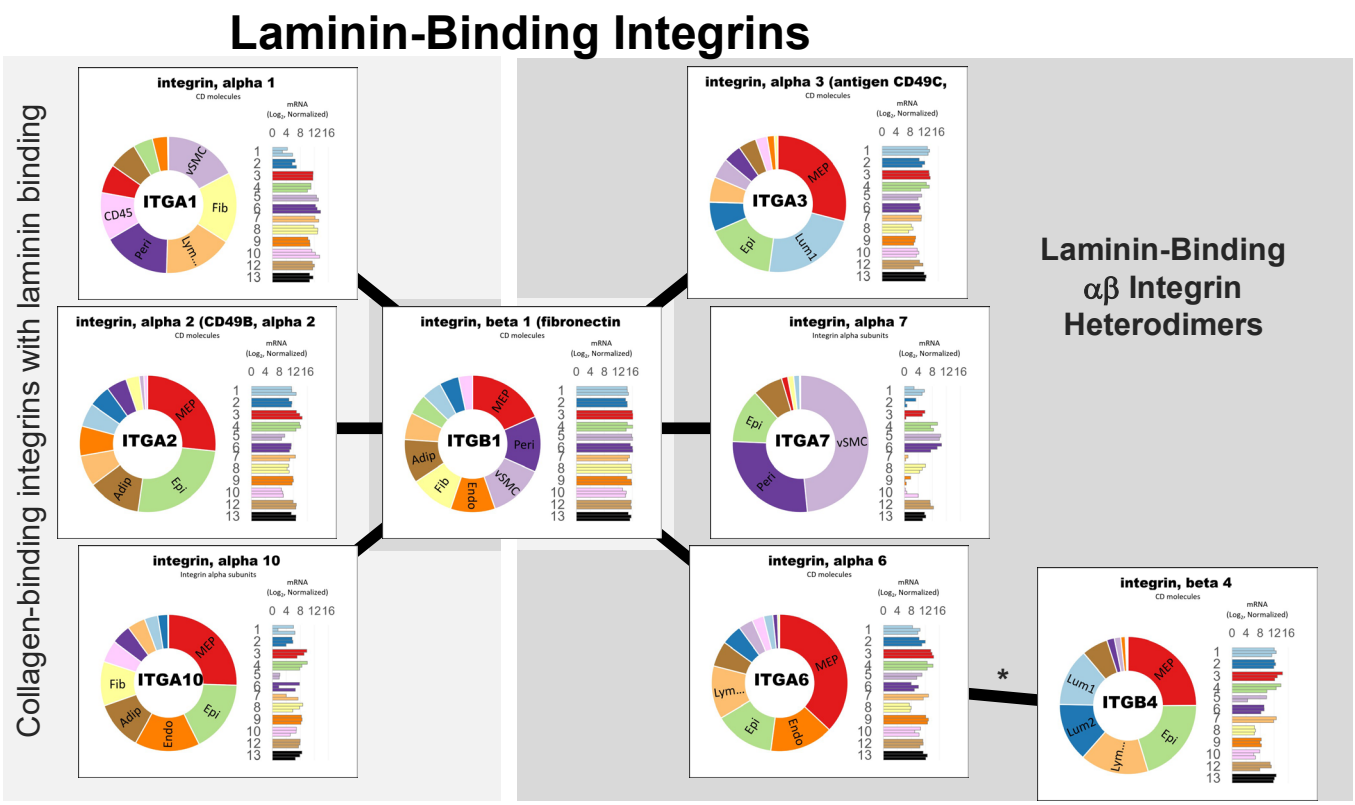
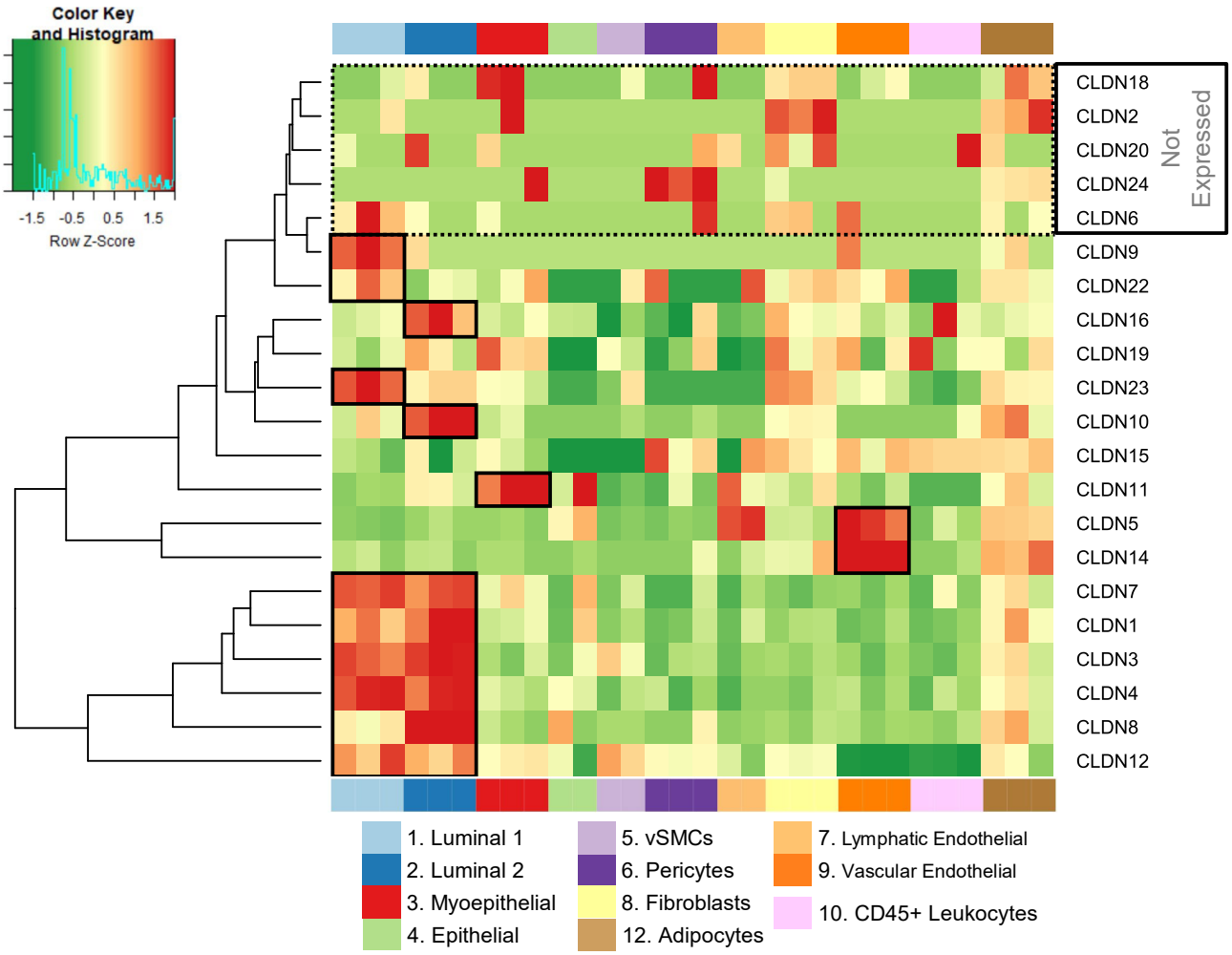


Figure 16

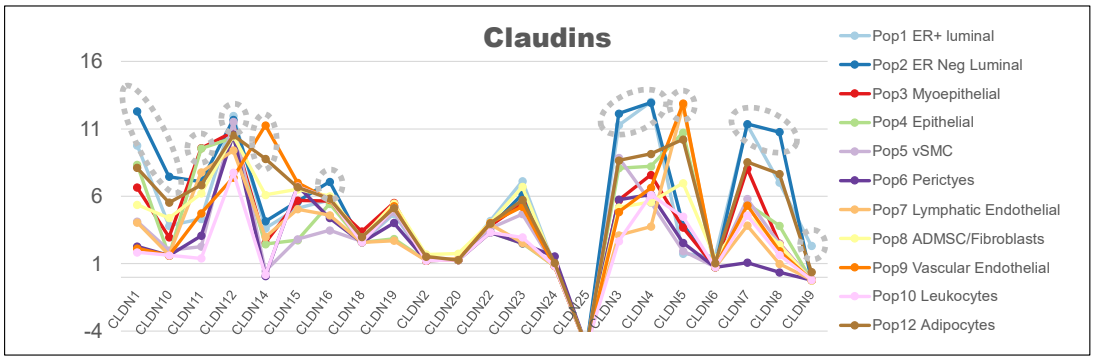
Claudins

HGNC: CLDN

a



b



Tight Junction Proteins

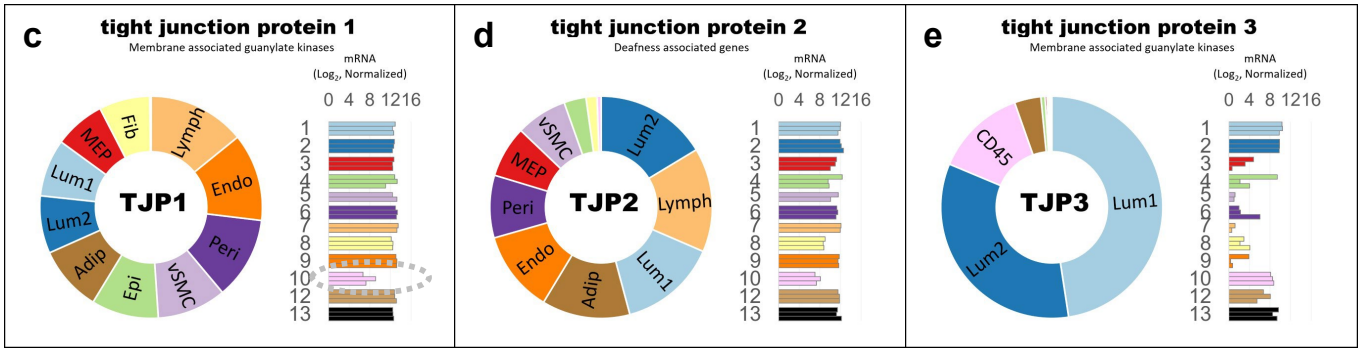


Figure 16—figure supplement 1

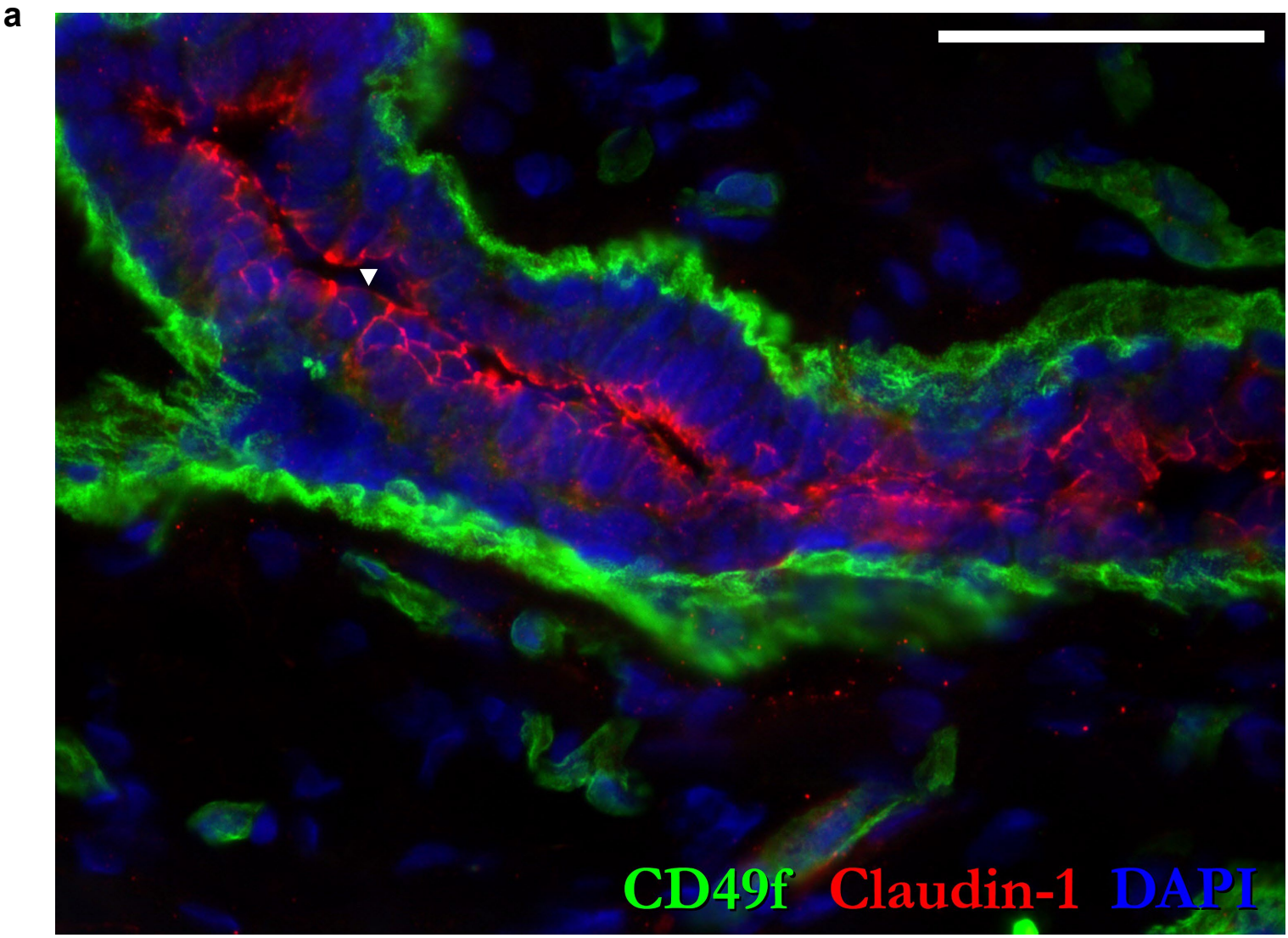
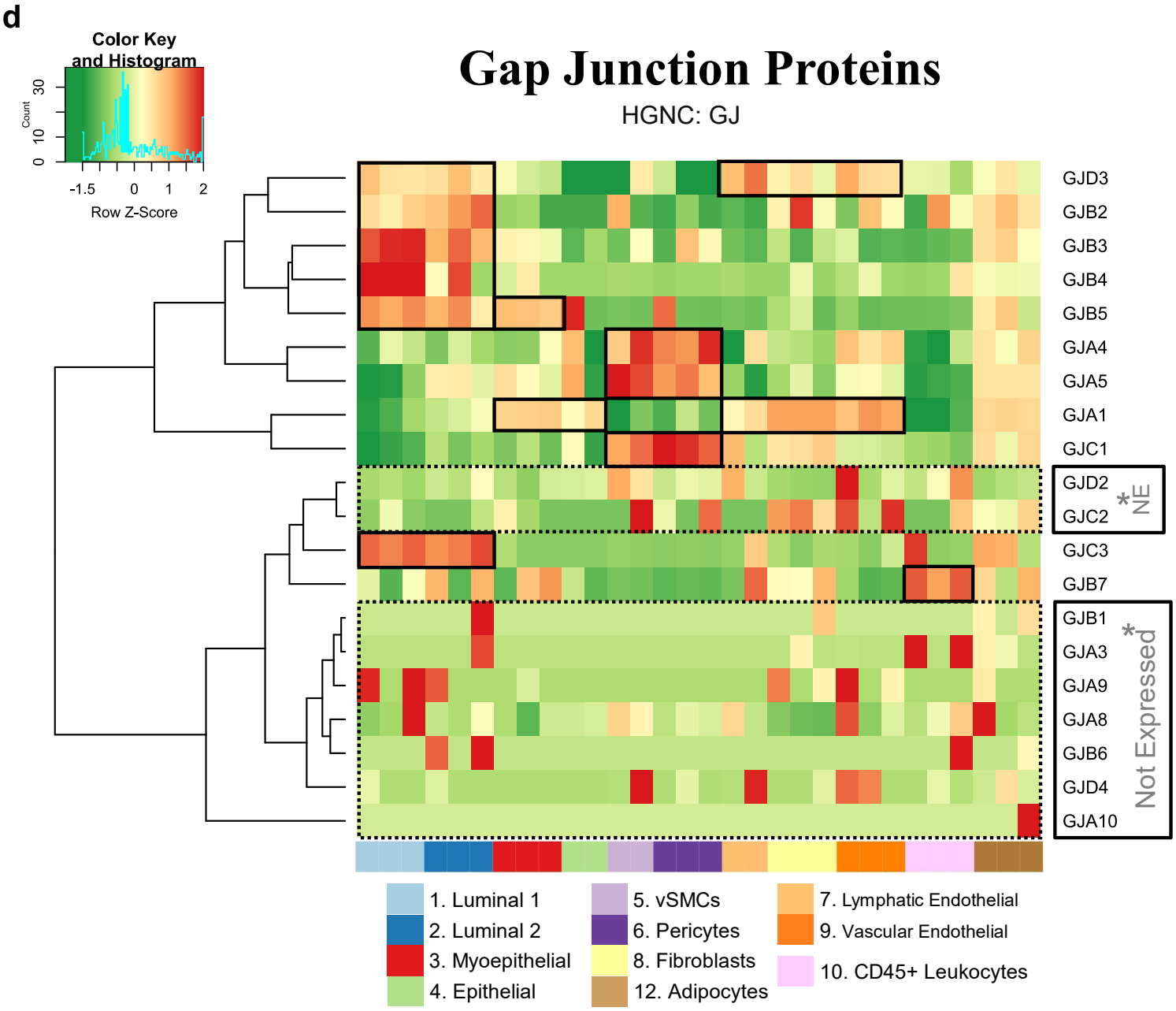


Figure 17


d
Color Key and Histogram
Count
0
10
30
-1.5
0
1
2
Row Z-Score
Gap Junction Proteins
HGNC: GJ
GJD3
GJB2
GJB3
GJB4
GJB5
GJA4
GJA5
GJA1
GJC1
GJD2
GJC2
*NE
GJC3
GJB7
GJB1
GJA3
GJA9
GJA8
GJB6
GJD4
GJA10
*Not Expressed
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
8. Fibroblasts
12. Adipocytes
7. Lymphatic Endothelial
9. Vascular Endothelial
10. CD45+ Leukocytes

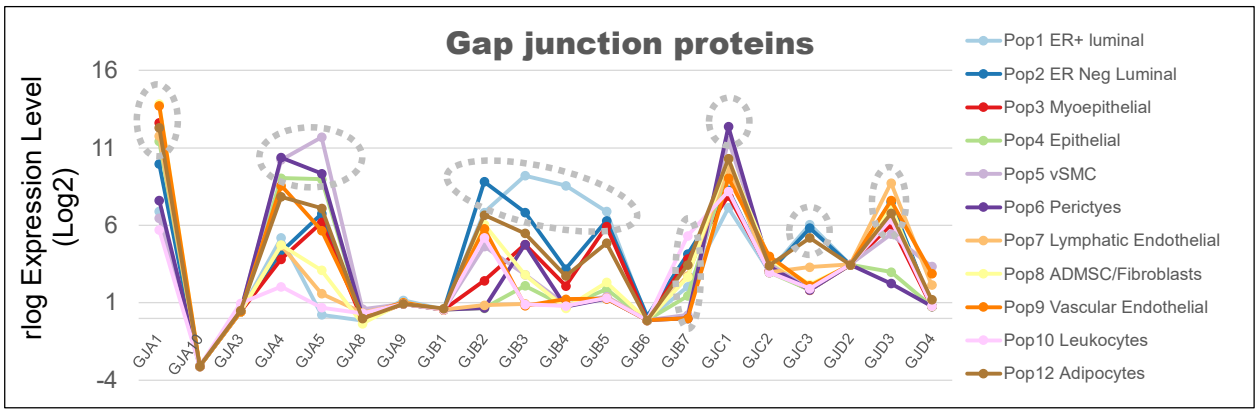

e
Gap junction proteins
rlog Expression Level (Log2)
16
11
6
1
-4
GJA1
GJA10
GJA3
GJA4
GJA5
GJA8
GJA9
GJB1
GJB2
GJB3
GJB4
GJB5
GJB6
GJB7
GJC1
GJC2
GJC3
GJD2
GJD3
GJD4
Pop1 ER+ luminal
Pop2 ER Neg Luminal
Pop3 Myoepithelial
Pop4 Epithelial
Pop5 vSMC
Pop6 Perictyes
Pop7 Lymphatic Endothelial
Pop8 ADMSC/Fibroblasts
Pop9 Vascular Endothelial
Pop10 Leukocytes
Pop12 Adipocytes

Figure 18

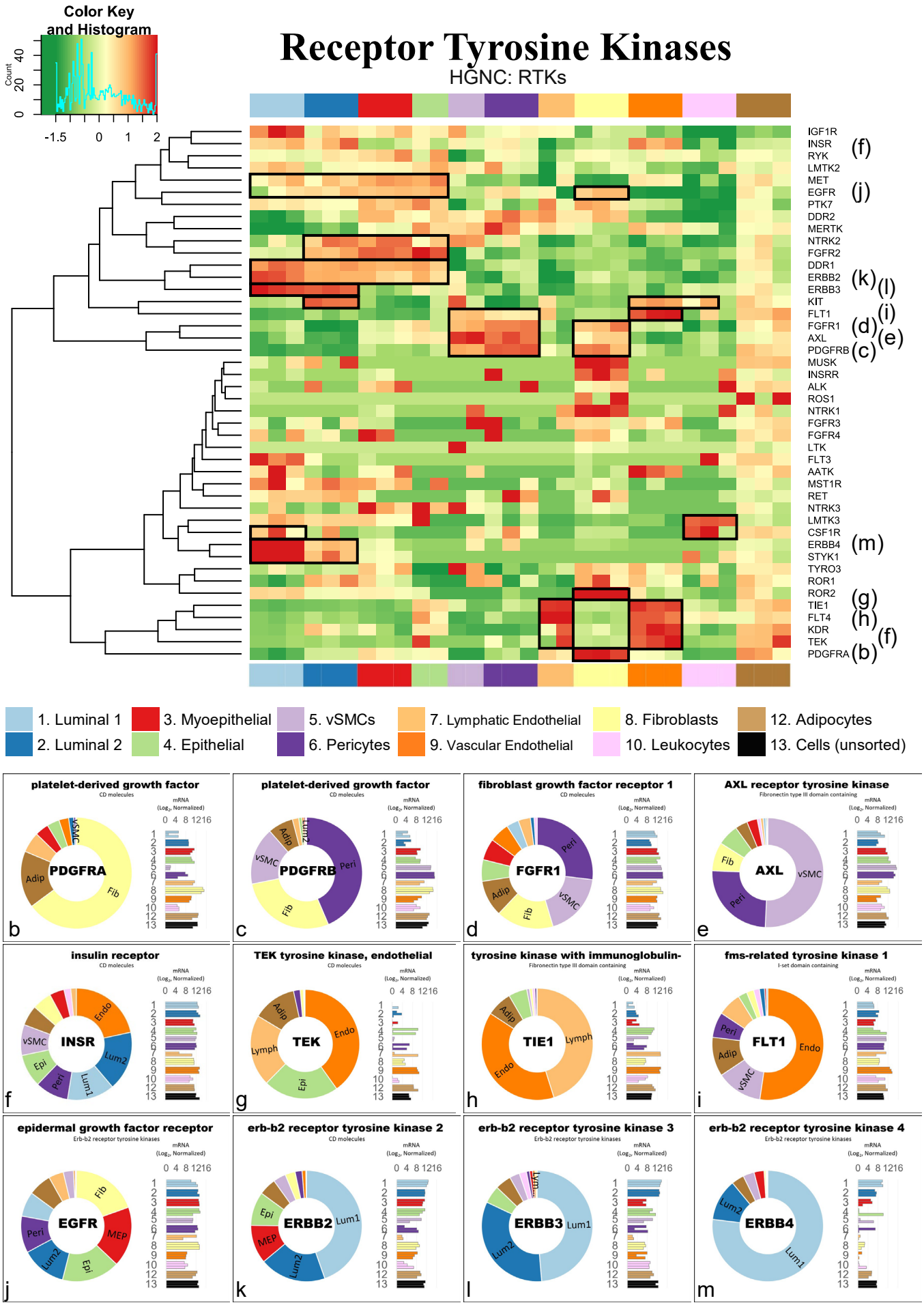

Color Key and Histogram
Count
0
20
40
-1.5
0
1
2
Receptor Tyrosine Kinases
HGNC: RTKs
IGF1R
INSR
RYK
LMTK2
MET
EGFR
PTK7
DDR2
MERTK
NTRK2
FGFR2
DDR1
ERBB2
ERBB3
KIT
FLT1
FGFR1
AXL
PDGFRB
MUSK
INSRR
ALK
ROS1
NTRK1
FGFR3
FGFR4
LTK
FLT3
AATK
MST1R
RET
NTRK3
LMTK3
CSF1R
ERBB4
STYK1
TYRO3
ROR1
ROR2
TIE1
FLT4
KDR
TEK
PDGFRA
(f)
(j)
(k)
(l)
(i)
(d)
(e)
(c)
(m)
(g)
(h)
(f)
(b)
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
9. Vascular Endothelial
8. Fibroblasts
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)
platelet-derived growth factor
PDGFRA
b
platelet-derived growth factor
PDGFRB
c
fibroblast growth factor receptor 1
FGFR1
d
AXL receptor tyrosine kinase
AXL
e
insulin receptor
INSR
f
TEK tyrosine kinase, endothelial
TEK
g
tyrosine kinase with immunoglobulin-
TIE1
h
fms-related tyrosine kinase 1
FLT1
i
epidermal growth factor receptor
EGFR
j
erb-b2 receptor tyrosine kinase 2
ERBB2
k
erb-b2 receptor tyrosine kinase 3
ERBB3
l
erb-b2 receptor tyrosine kinase 4
ERBB4
m

Figure 19

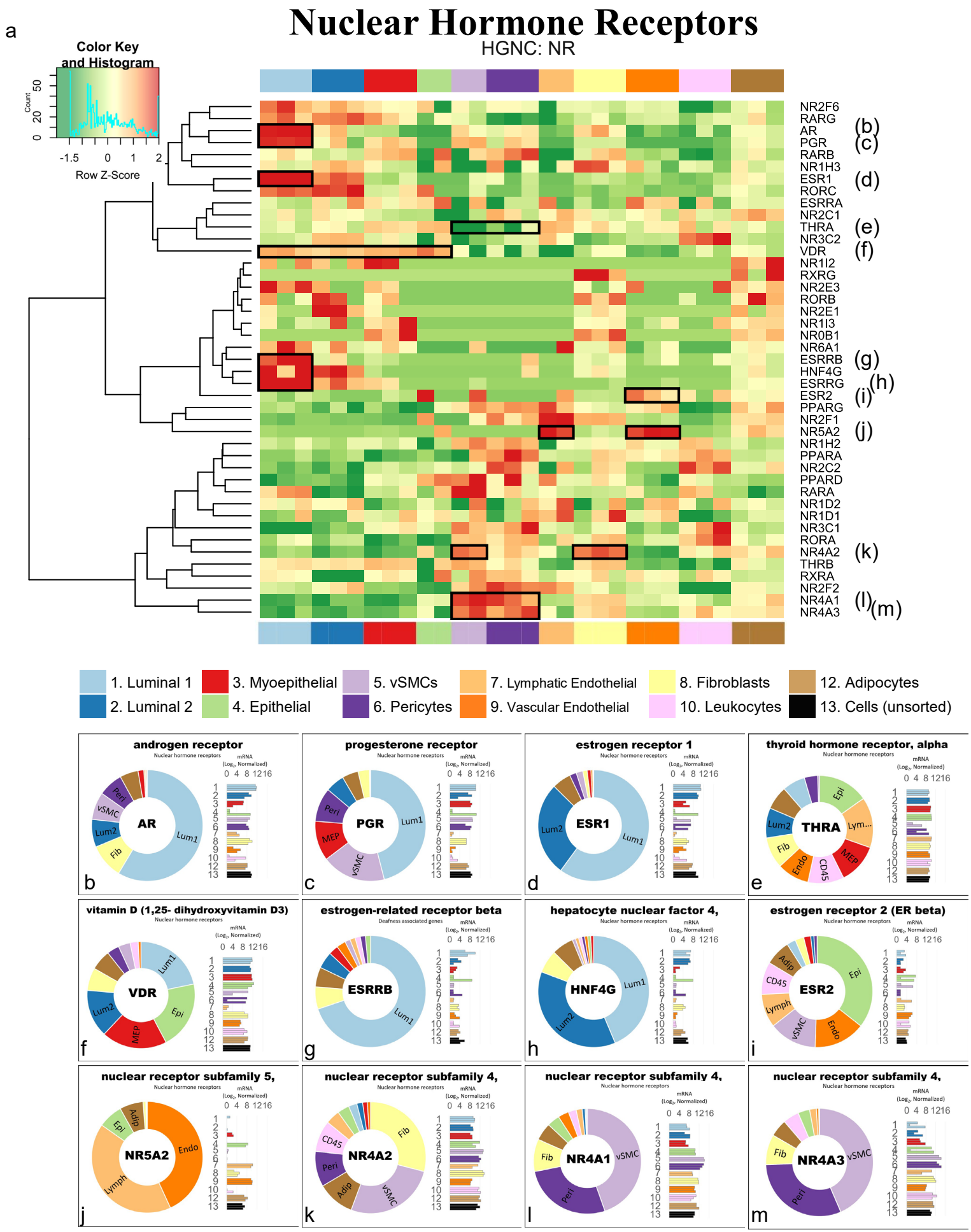

a
Nuclear Hormone Receptors
HGNC: NR
Color Key and Histogram
Count
Row Z-Score
NR2F6
RARG
AR
PGR
RARB
NR1H3
ESR1
RORC
ESRRA
NR2C1
THRA
NR3C2
VDR
NR1I2
RXRG
NR2E3
RORB
NR2E1
NR1I3
NR0B1
NR6A1
ESRRB
HNF4G
ESRRG
ESR2
PPARG
NR2F1
NR5A2
NR1H2
PPARA
NR2C2
PPARD
RARA
NR1D2
NR1D1
NR3C1
RORA
NR4A2
THRB
RXRA
NR2F2
NR4A1
NR4A3
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)
androgen receptor
AR
progesterone receptor
PGR
estrogen receptor 1
ESR1
thyroid hormone receptor, alpha
THRA
vitamin D (1,25- dihydroxyvitamin D3)
VDR
estrogen-related receptor beta
ESRRB
hepatocyte nuclear factor 4,
HNF4G
estrogen receptor 2 (ER beta)
ESR2
nuclear receptor subfamily 5,
NR5A2
nuclear receptor subfamily 4,
NR4A2
nuclear receptor subfamily 4,
NR4A1
nuclear receptor subfamily 4,
NR4A3

Figure 20

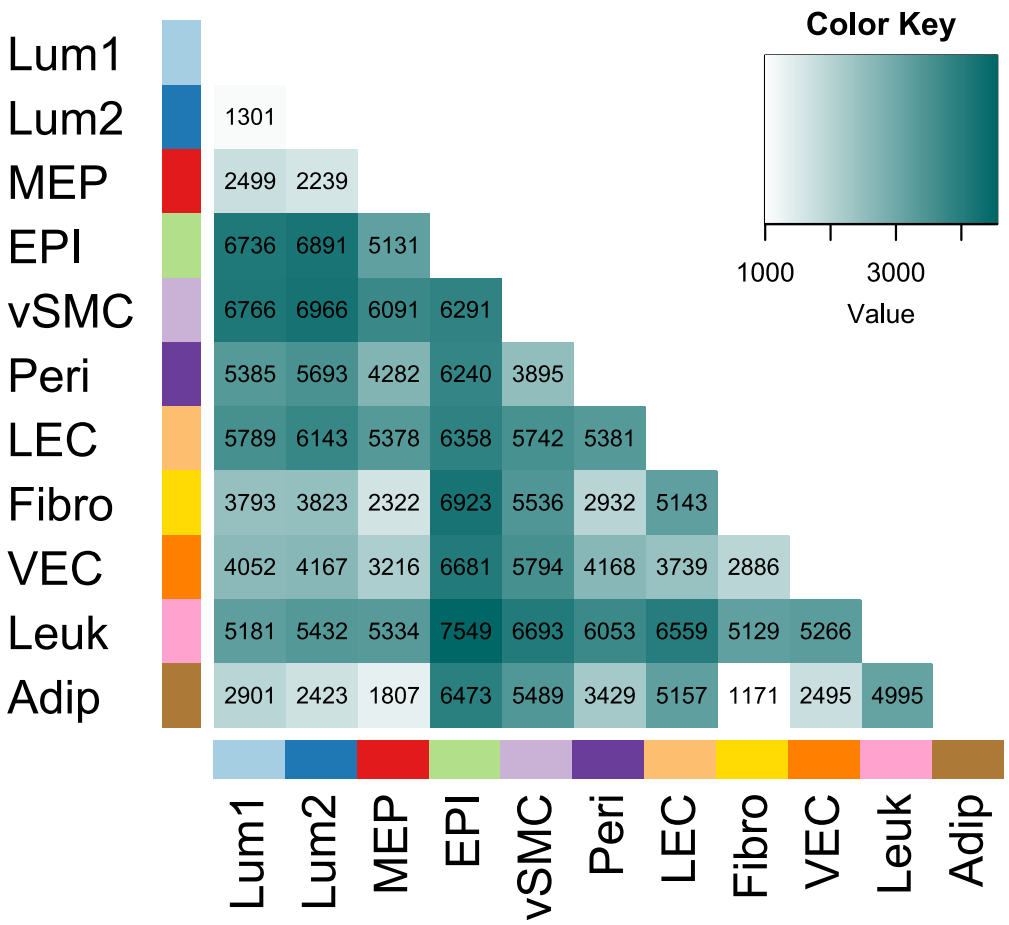

Color Key
1000
3000
Value
Lum1
Lum2
MEP
EPI
vSMC
Peri
LEC
Fibro
VEC
Leuk
Adip
1301
2499 2239
6736 6891 5131
6766 6966 6091 6291
5385 5693 4282 6240 3895
5789 6143 5378 6358 5742 5381
3793 3823 2322 6923 5536 2932 5143
4052 4167 3216 6681 5794 4168 3739 2886
5181 5432 5334 7549 6693 6053 6559 5129 5266
2901 2423 1807 6473 5489 3429 5157 1171 2495 4995

Figure 21

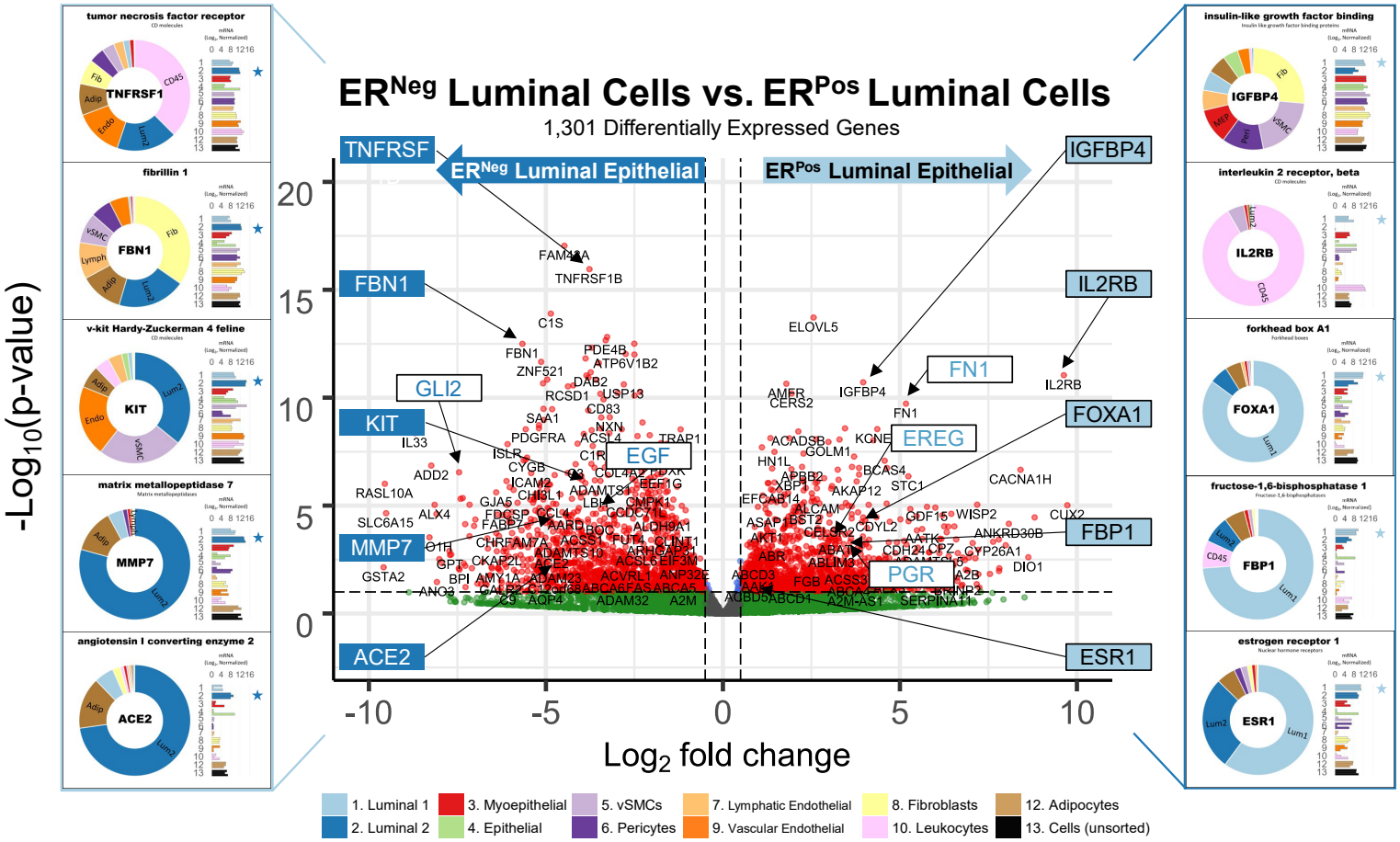

ERNeg Luminal Cells vs. ERPos Luminal Cells
1,301 Differentially Expressed Genes
ERNeg Luminal Epithelial
ERPos Luminal Epithelial
TNFRSF
FBN1
GLI2
KIT
MMP7
ACE2
IGFBP4
IL2RB
FN1
FOXA1
EREG
FBP1
PGR
ESR1
EGF
-Log10(p-value)
Log2 fold change
tumor necrosis factor receptor
fibrillin 1
v-kit Hardy-Zuckerman 4 feline
matrix metallopeptidase 7
angiotensin I converting enzyme 2
insulin-like growth factor binding
interleukin 2 receptor, beta
forkhead box A1
fructose-1,6-bisphosphatase 1
estrogen receptor 1
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
9. Vascular Endothelial
8. Fibroblasts
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)

Figure 22

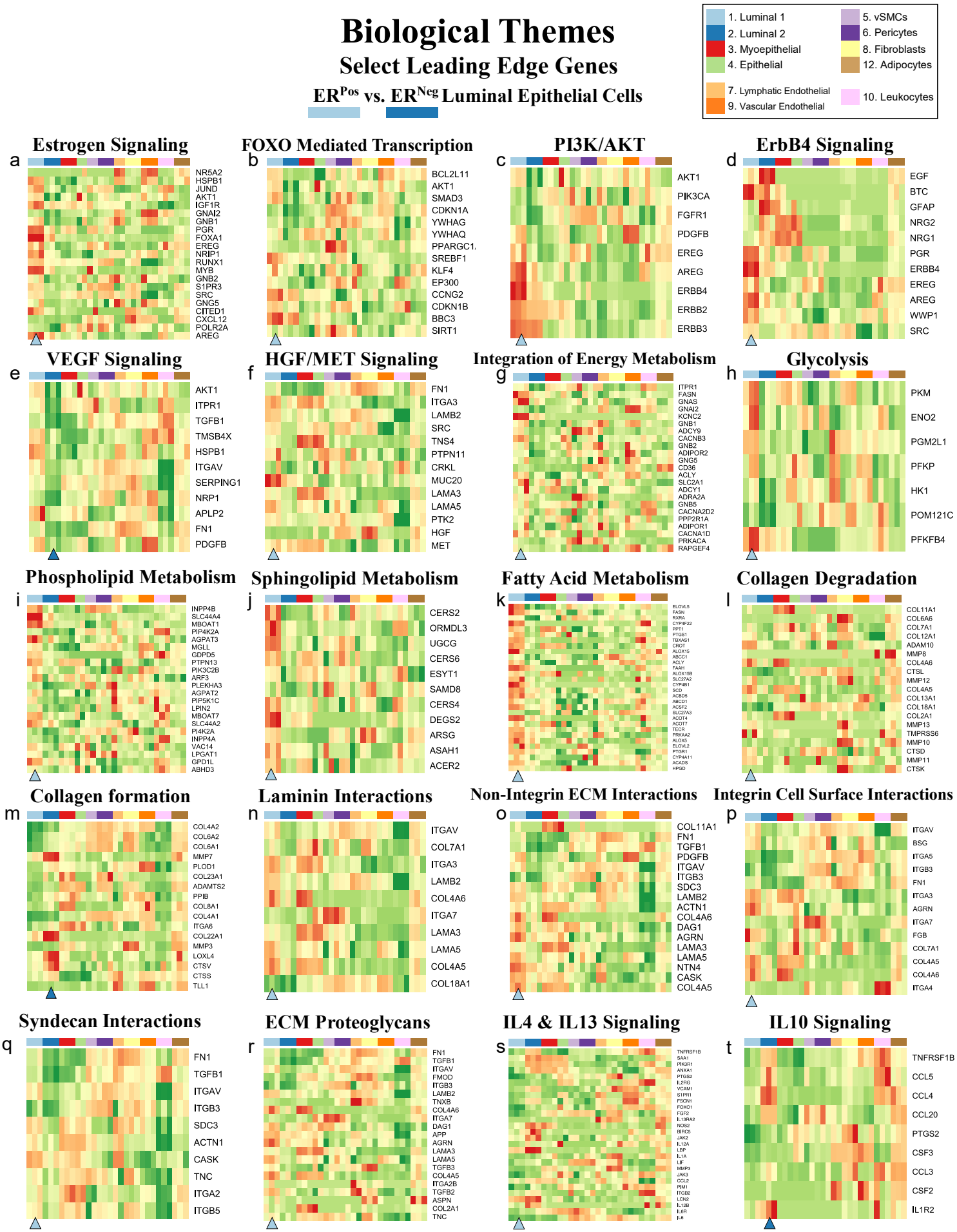

Biological Themes
Select Leading Edge Genes
ERPos vs. ERNeg Luminal Epithelial Cells
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
8. Fibroblasts
12. Adipocytes
7. Lymphatic Endothelial
9. Vascular Endothelial
10. Leukocytes
a
Estrogen Signaling
NR5A2
HSPB1
JUND
AKT1
IGF1R
GNAI2
GNB1
PGR
FOXA1
EREG
NRIP1
RUNX1
MYB
GNB2
S1PR3
SRC
GNG5
CITED1
CXCL12
POLR2A
AREG
b
FOXO Mediated Transcription
BCL2L11
AKT1
SMAD3
CDKN1A
YWHAG
YWHAQ
PPARGC1A
SREBF1
KLF4
EP300
CCNG2
CDKN1B
BBC3
SIRT1
c
PI3K/AKT
AKT1
PIK3CA
FGFR1
PDGFB
EREG
AREG
ERBB4
ERBB2
ERBB3
d
ErbB4 Signaling
EGF
BTC
GFAP
NRG2
NRG1
PGR
ERBB4
EREG
AREG
WWP1
SRC
e
VEGF Signaling
AKT1
ITPR1
TGFB1
TMSB4X
HSPB1
ITGAV
SERPING1
NRP1
APLP2
FN1
PDGFB
f
HGF/MET Signaling
FN1
ITGA3
LAMB2
SRC
TNS4
PTPN11
CRKL
MUC20
LAMA3
LAMA5
PTK2
HGF
MET
g
Integration of Energy Metabolism
ITPR1
FASN
GNAS
GNAI2
KCNC2
GNB1
ADCY9
CACNB3
GNB2
ADIPOR2
GNG5
CD36
ACLY
SLC2A1
ADCY1
ADRA2A
GNB5
CACNA2D2
PPP2R1A
ADIPOR1
CACNA1D
PRKACA
RAPGEF4
h
Glycolysis
PKM
ENO2
PGM2L1
PFKP
HK1
POM121C
PFKFB4
i
Phospholipid Metabolism
INPP4B
SLC44A4
MBOAT1
PIP4K2A
AGPAT3
MGLL
GDPD5
PTPN13
PIK3C2B
ARF3
PLEKHA3
AGPAT2
PIP5K1C
LPIN2
MBOAT7
SLC44A2
PI4K2A
INPP4A
VAC14
LPGAT1
GPD1L
ABHD3
j
Sphingolipid Metabolism
CERS2
ORMDL3
UGCG
CERS6
ESYT1
SAMD8
CERS4
DEGS2
ARSG
ASAH1
ACER2
k
Fatty Acid Metabolism
ELOVL5
FASN
RXRA
CYP4F22
PPT1
PTGS1
TBXAS1
CROT
ALOX15
ABCC1
ACLY
FAAH
ALOX15B
SLC27A2
CYP4B1
SCD
ACBD5
ABCD1
ACSF2
SLC27A3
ACOT4
ACOT7
TECR
PRKAA2
ALOX5
ELOVL2
PTGR1
CYP4A11
ACAD5
HPSD
l
Collagen Degradation
COL11A1
COL6A6
COL7A1
COL12A1
ADAM10
MMP8
COL4A6
CTSL
MMP12
COL4A5
COL13A1
COL18A1
COL2A1
MMP13
TMPRSS6
MMP10
CTSD
MMP11
CTSK
m
Collagen formation
COL4A2
COL6A2
COL6A1
MMP7
PLOD1
COL23A1
ADAMTS2
PPIB
COL8A1
COL4A1
ITGA6
COL22A1
MMP3
LOXL4
CTSV
CTSS
TLL1
n
Laminin Interactions
ITGAV
COL7A1
ITGA3
LAMB2
COL4A6
ITGA7
LAMA3
LAMA5
COL4A5
COL18A1
o
Non-Integrin ECM Interactions
COL11A1
FN1
TGFB1
PDGFB
ITGAV
ITGB3
SDC3
LAMB2
ACTN1
COL4A6
DAG1
AGRN
LAMA3
LAMA5
NTN4
CASK
COL4A5
p
Integrin Cell Surface Interactions
ITGAV
BSG
ITGA5
ITGB3
FN1
ITGA3
AGRN
ITGA7
FGB
COL7A1
COL4A5
COL4A6
ITGA4
q
Syndecan Interactions
FN1
TGFB1
ITGAV
ITGB3
SDC3
ACTN1
CASK
TNC
ITGA2
ITGB5
r
ECM Proteoglycans
FN1
TGFB1
ITGAV
FMOD
ITGB3
LAMB2
TNXB
COL4A6
ITGA7
DAG1
APP
AGRN
LAMA3
LAMA5
TGFB3
COL4A5
ITGA2B
TGFB2
ASPN
COL2A1
TNC
s
IL4 & IL13 Signaling
TNFRSF1B
SAA1
PIK3R1
ANXA1
PTGS2
IL2RG
VCAM1
S1PR1
FSCN1
FOXO1
FGF2
IL13RA2
NOS2
BIRC5
JAK2
IL12A
LBP
IL1A
LIF
MMP3
JAK3
CCL2
PIM1
ITGB2
LCN2
IL12B
IL6R
IL6
t
IL10 Signaling
TNFRSF1B
CCL5
CCL4
CCL20
PTGS2
CSF3
CCL3
CSF2
IL1R2

Figure 23

Enriched Gene Sets & Biological Themes
Luminal Epithelial Cells:
ER^{Pos} Luminal (Pop1,) vs. ER^{Neg} Luminal (Pop2,)

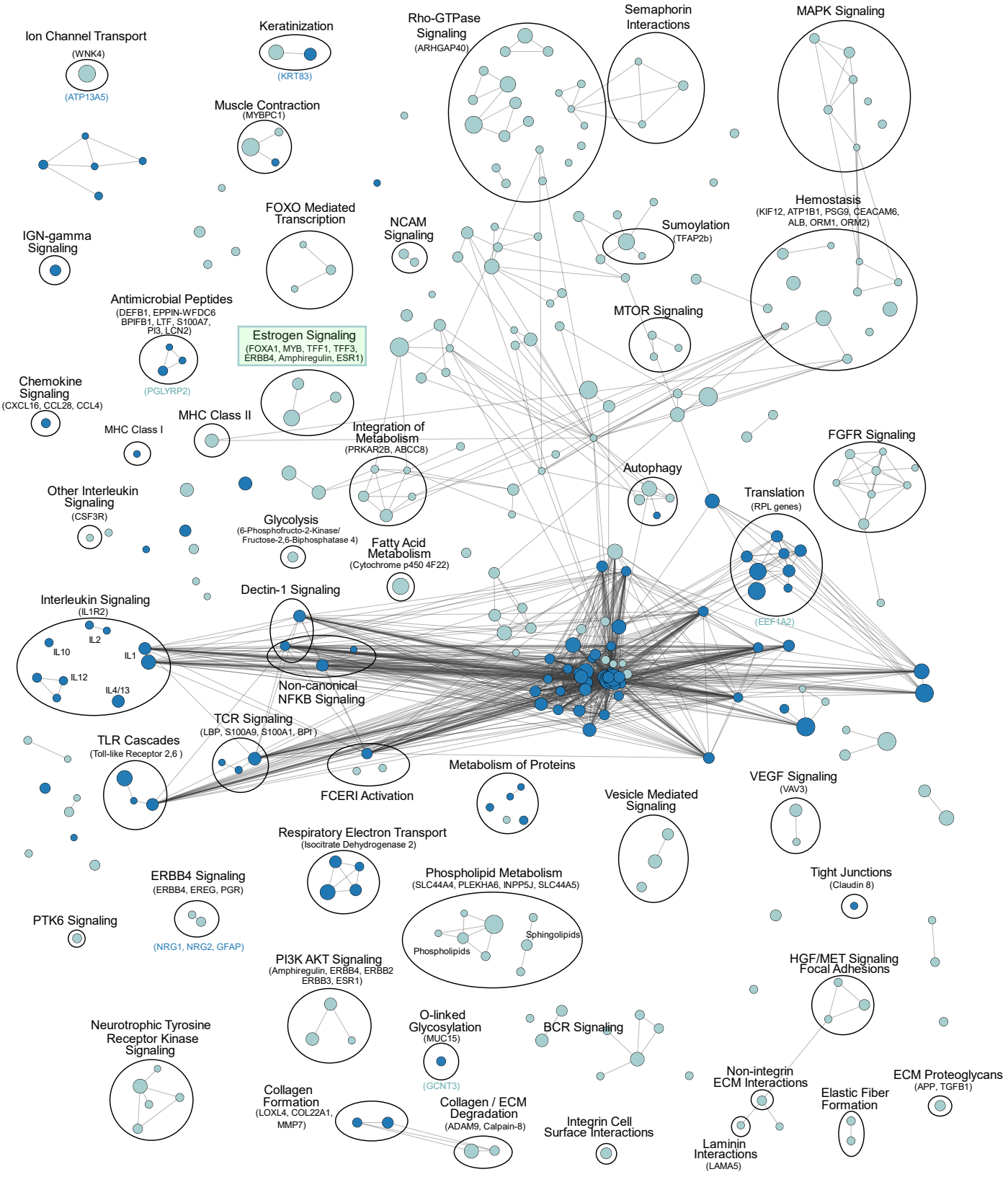


Figure 23—Supplementary Figure 1

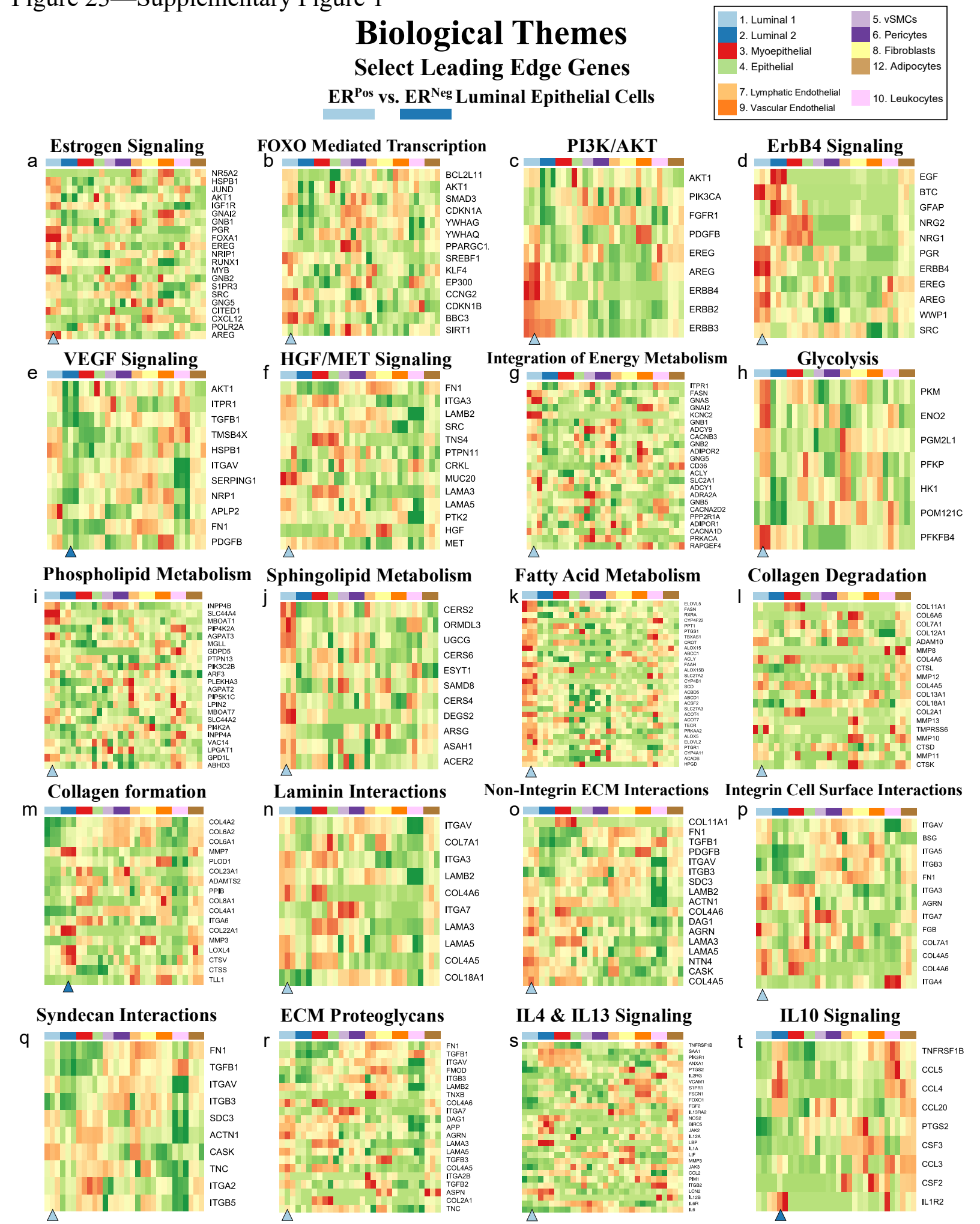


Figure 23—figure supplement 2

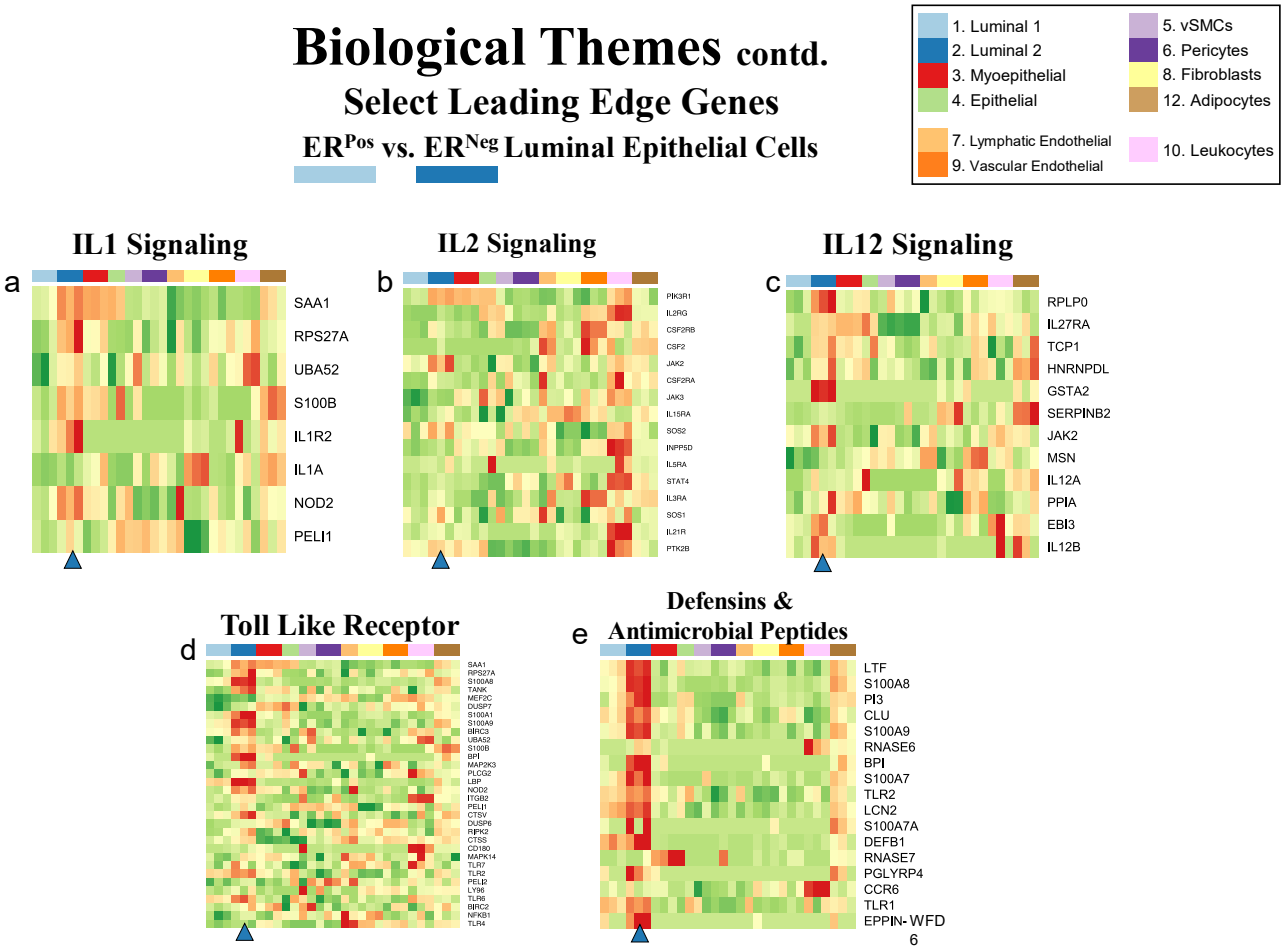


Fig.23—figure supplement 3

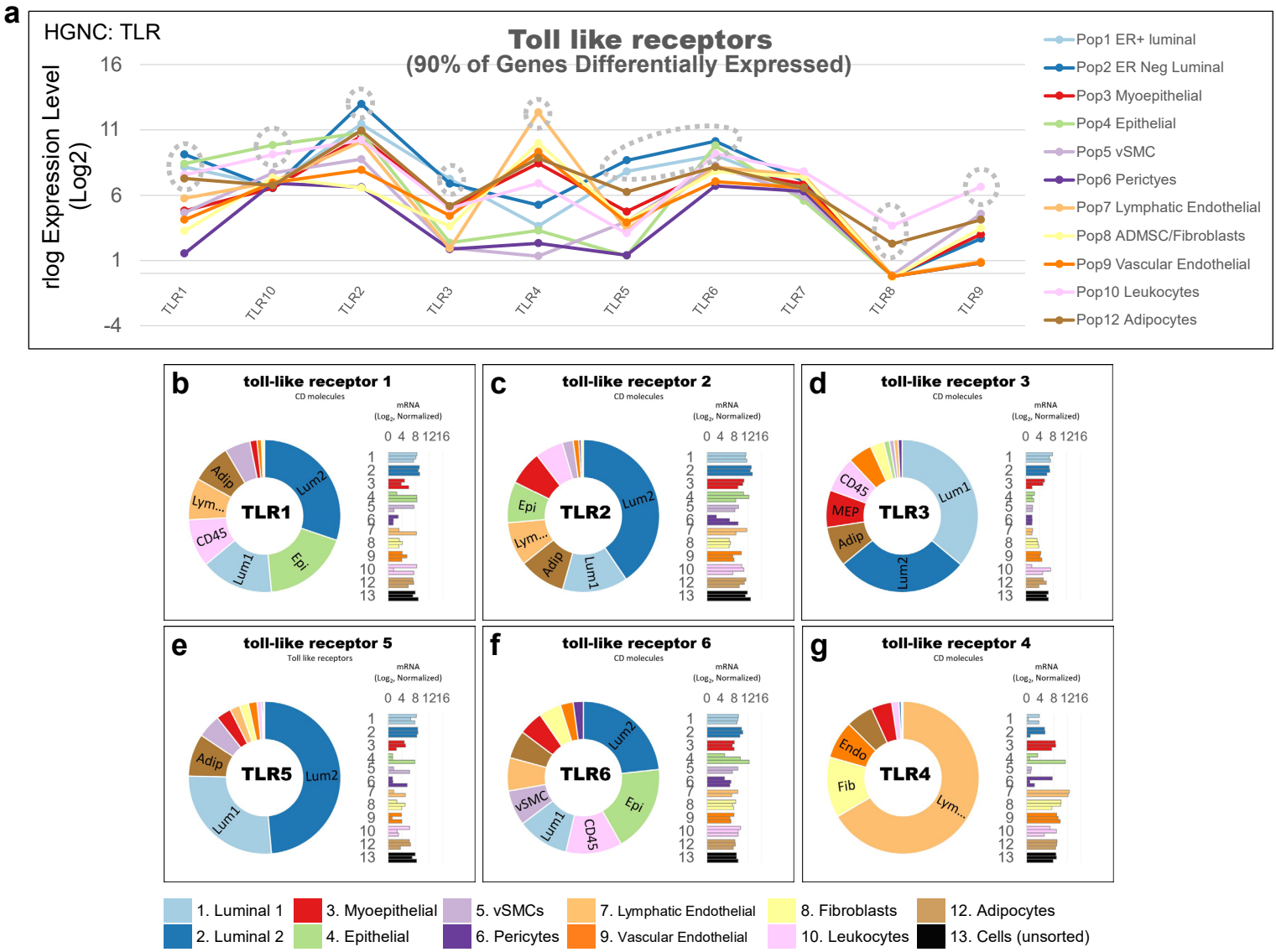


Figure 23—figure supplement 4

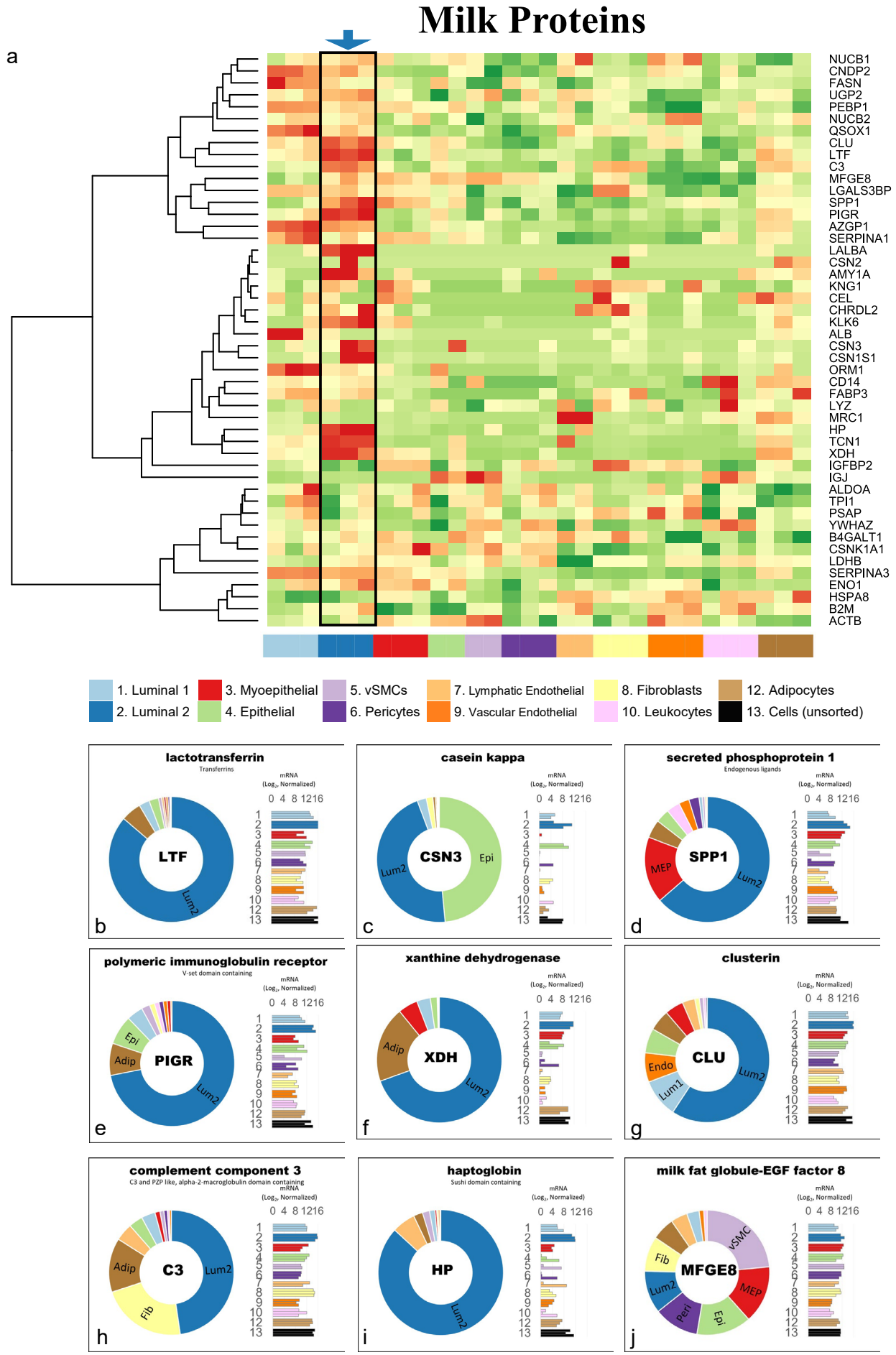


Figure 24

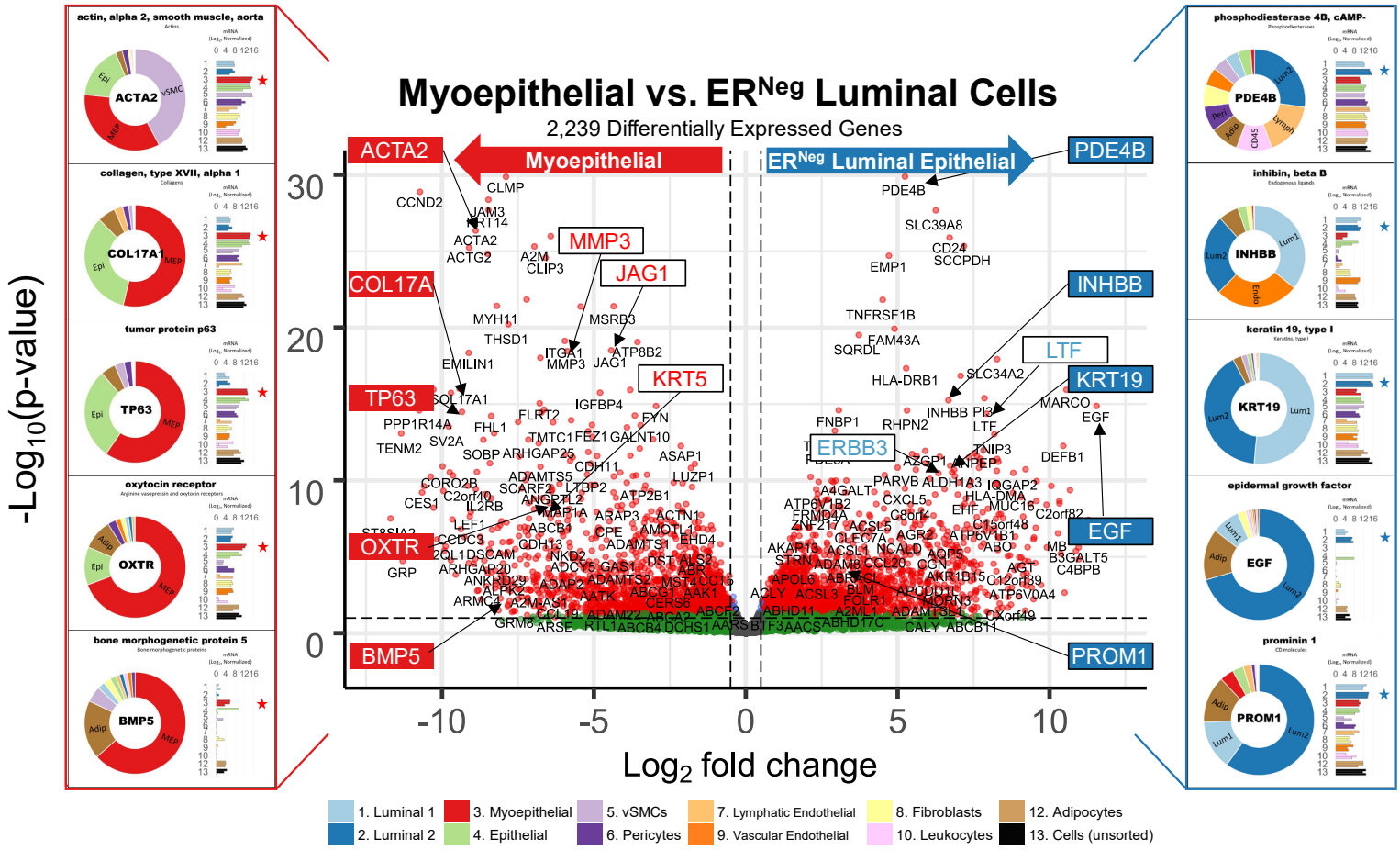

Myoepithelial vs. ERNeg Luminal Cells
2,239 Differentially Expressed Genes
Myoepithelial
ERNeg Luminal Epithelial
ACTA2
COL17A
TP63
OXTR
BMP5
MMP3
JAG1
KRT5
PDE4B
INHBB
LTF
KRT19
ERBB3
EGF
PROM1
-Log10(p-value)
Log2 fold change
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
9. Vascular Endothelial
8. Fibroblasts
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)

Figure 25

Enriched Gene Sets
Epithelial Cells:
ER^{Neg} Luminal (Pop2,●) vs. Myoepithelial (Pop3,●)

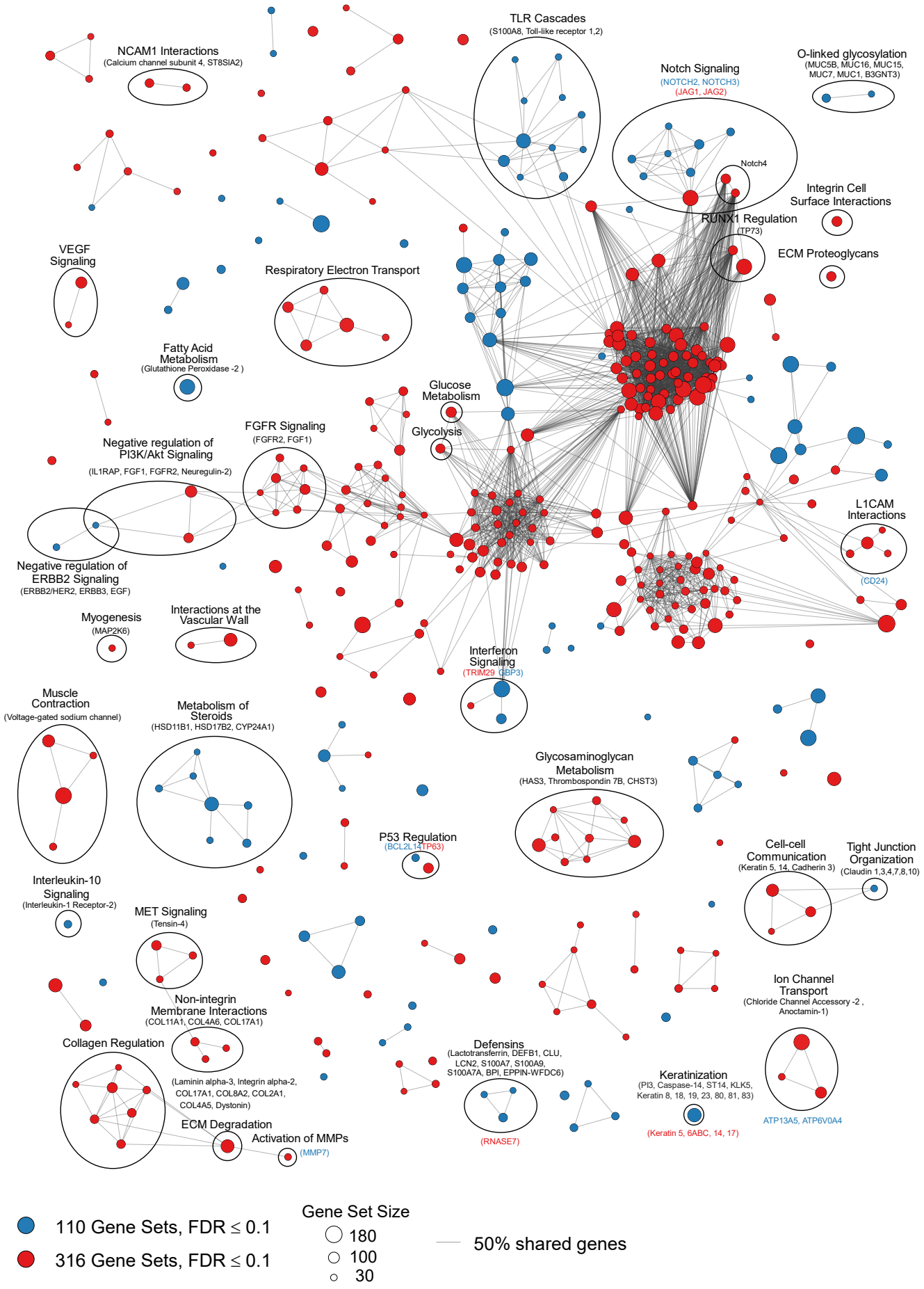
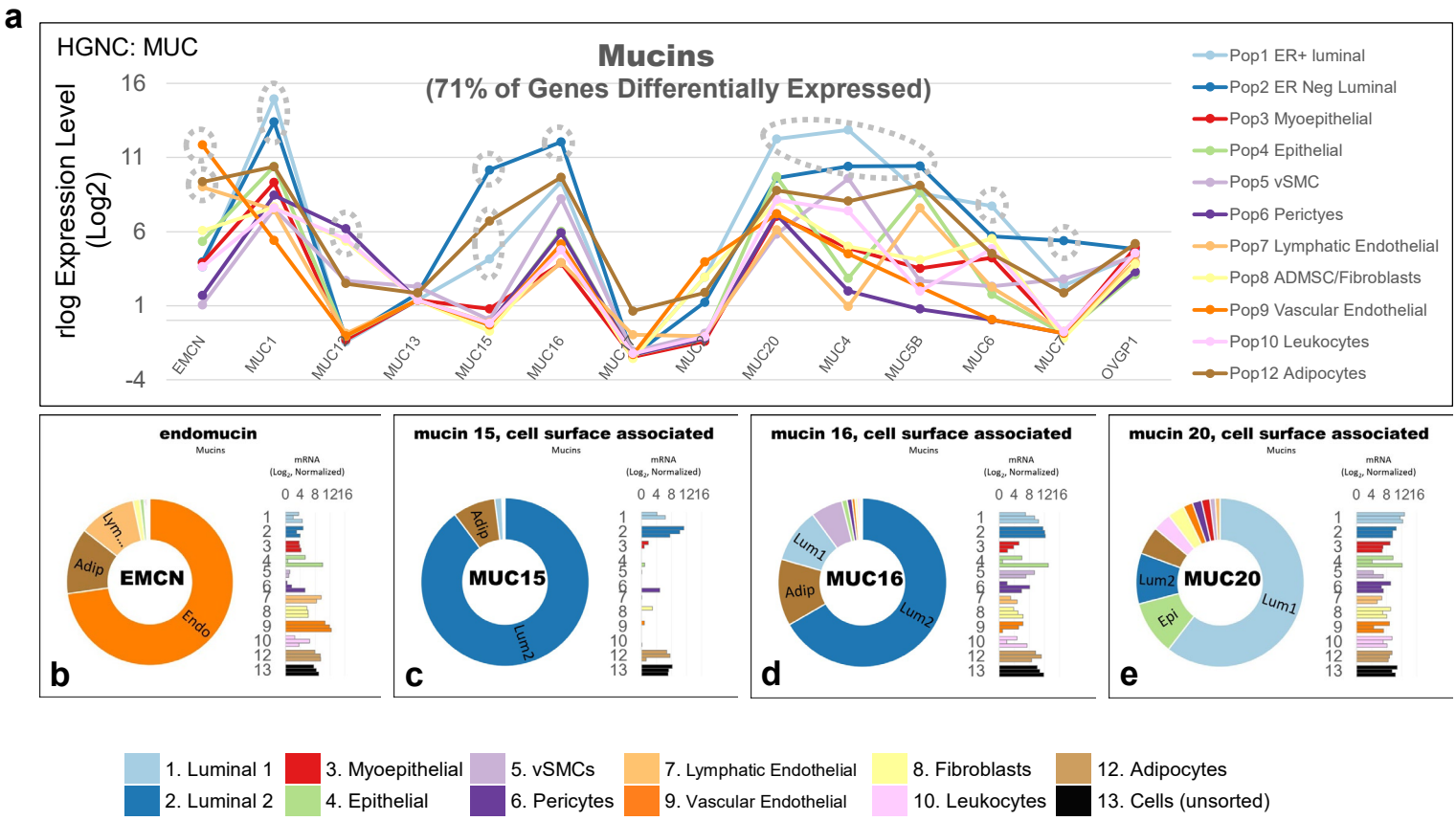


Figure 25—figure supplement 1

Fig.25—figure supplement 2



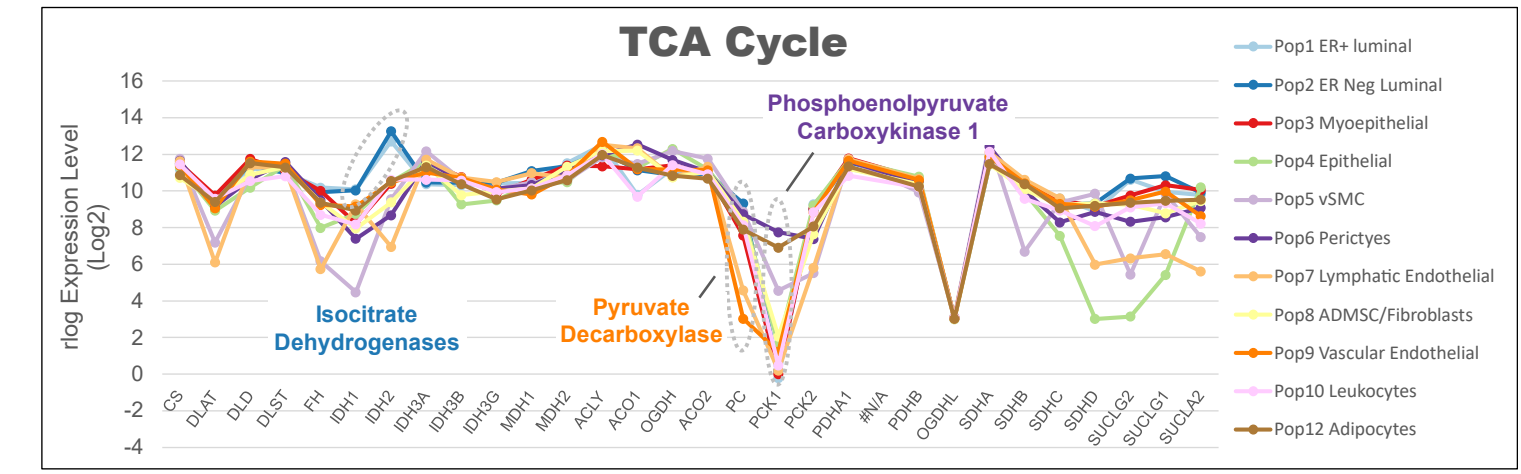


Figure 26

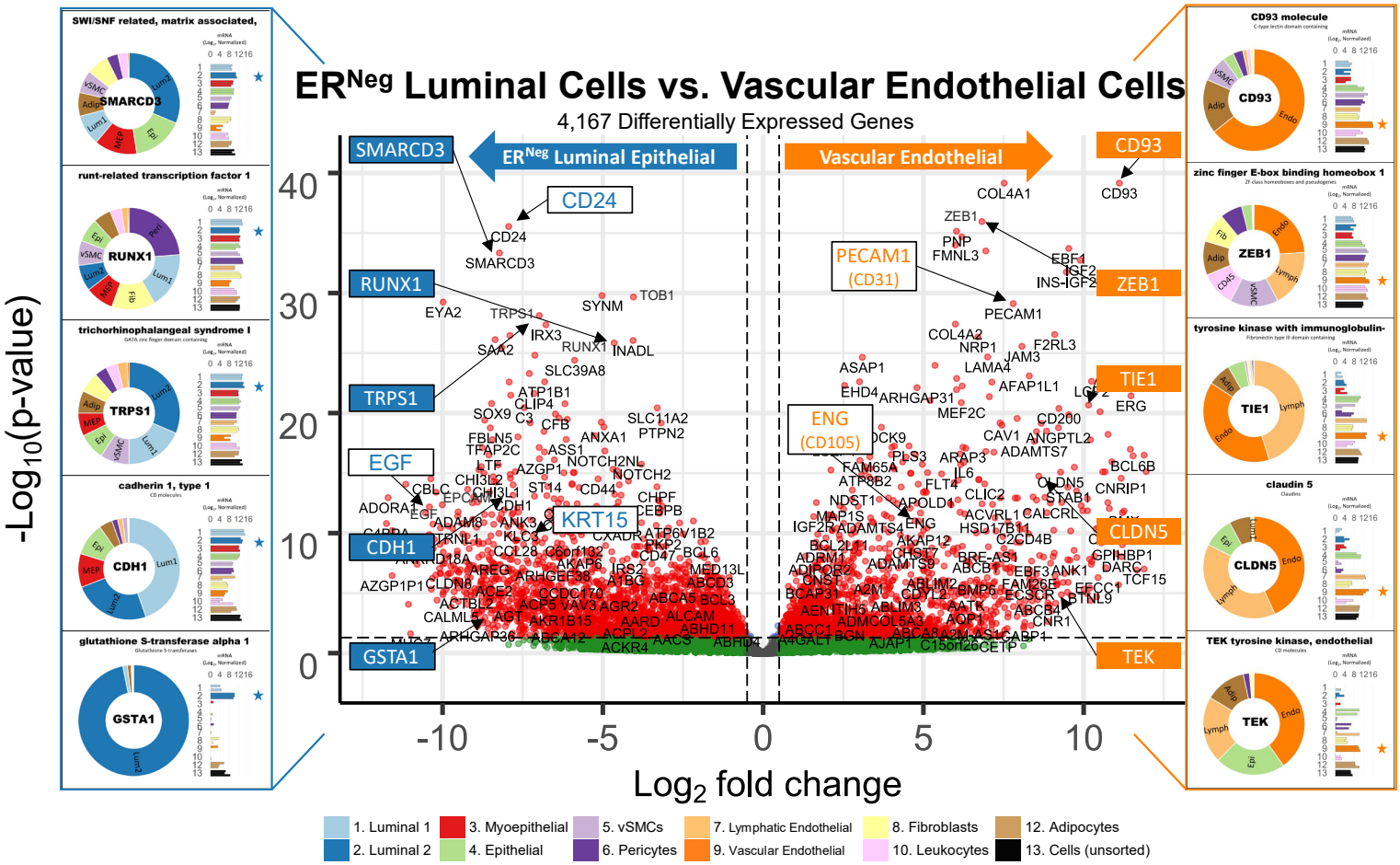

ERNeg Luminal Cells vs. Vascular Endothelial Cells
4,167 Differentially Expressed Genes
ERNeg Luminal Epithelial
Vascular Endothelial
SMARCD3
CD24
RUNX1
TRPS1
EGF
CDH1
KRT15
GSTA1
CD93
ZEB1
PECAM1 (CD31)
TIE1
ENG (CD105)
CLDN5
TEK
-Log10(p-value)
Log2 fold change
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)

Figure 27

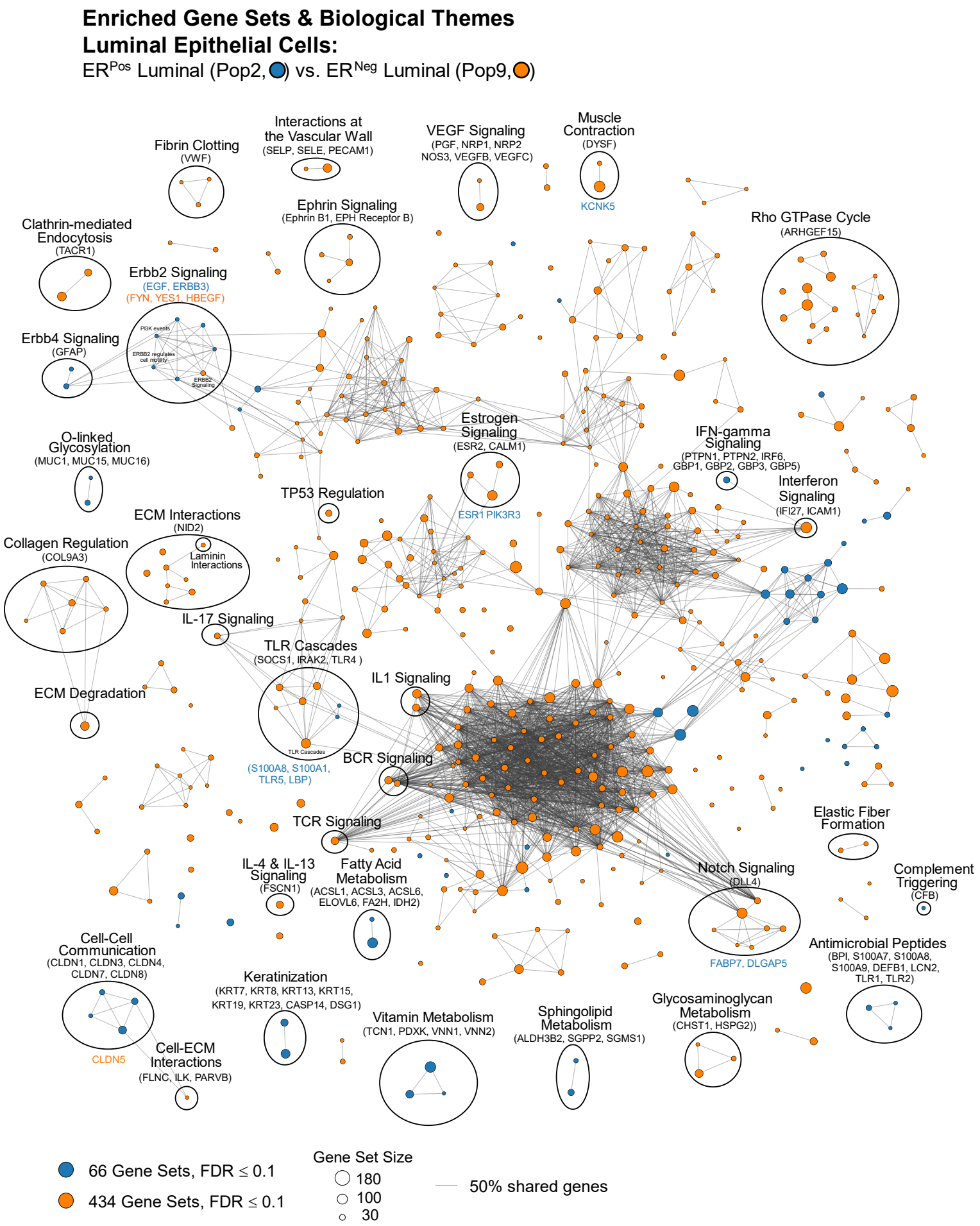

Enriched Gene Sets & Biological Themes
Luminal Epithelial Cells:
ERPos Luminal (Pop2,) vs. ERNeg Luminal (Pop9,)
Fibrin Clotting
(VWF)
Interactions at the Vascular Wall
(SELP, SELE, PECAM1)
VEGF Signaling
(PGF, NRP1, NRP2 NOS3, VEGFB, VEGFC)
Muscle Contraction
(DYSF)
KCNK5
Ephrin Signaling
(Ephrin B1, EPH Receptor B)
Clathrin-mediated Endocytosis
(TACR1)
Rho GTPase Cycle
(ARHGEF15)
Erbb2 Signaling
(EGF, ERBB3)
(FYN, YES1, HBEGF)
Erbb4 Signaling
(GFAP)
Estrogen Signaling
(ESR2, CALM1)
ESR1 PIK3R3
IFN-gamma Signaling
(PTPN1, PTPN2, IRF6, GBP1, GBP2, GBP3, GBP5)
Interferon Signaling
(IFI27, ICAM1)
O-linked Glycosylation
(MUC1, MUC15, MUC16)
TP53 Regulation
ECM Interactions
(NID2)
Laminin Interactions
Collagen Regulation
(COL9A3)
IL-17 Signaling
TLR Cascades
(SOCS1, IRAK2, TLR4)
IL1 Signaling
ECM Degradation
(S100A8, S100A1, TLR5, LBP)
BCR Signaling
TCR Signaling
Elastic Fiber Formation
IL-4 & IL-13 Signaling
(FSCN1)
Fatty Acid Metabolism
(ACSL1, ACSL3, ACSL6, ELOVL6, FA2H, IDH2)
Notch Signaling
(DLL4)
FABP7, DLGAP5
Complement Triggering
(CFB)
Cell-Cell Communication
(CLDN1, CLDN3, CLDN4, CLDN7, CLDN8)
CLDN5
Keratinization
(KRT7, KRT8, KRT13, KRT15, KRT19, KRT23, CASP14, DSG1)
Antimicrobial Peptides
(BPI, S100A7, S100A8, S100A9, DEFB1, LCN2, TLR1, TLR2)
Vitamin Metabolism
(TCN1, PDXK, VNN1, VNN2)
Sphingolipid Metabolism
(ALDH3B2, SGPP2, SGMS1)
Glycosaminoglycan Metabolism
(CHST1, HSPG2))
Cell-ECM Interactions
(FLNC, ILK, PARVB)
66 Gene Sets, FDR ≤ 0.1
434 Gene Sets, FDR ≤ 0.1
Gene Set Size
180
100
30
50% shared genes

Figure 27—figure supplement 1

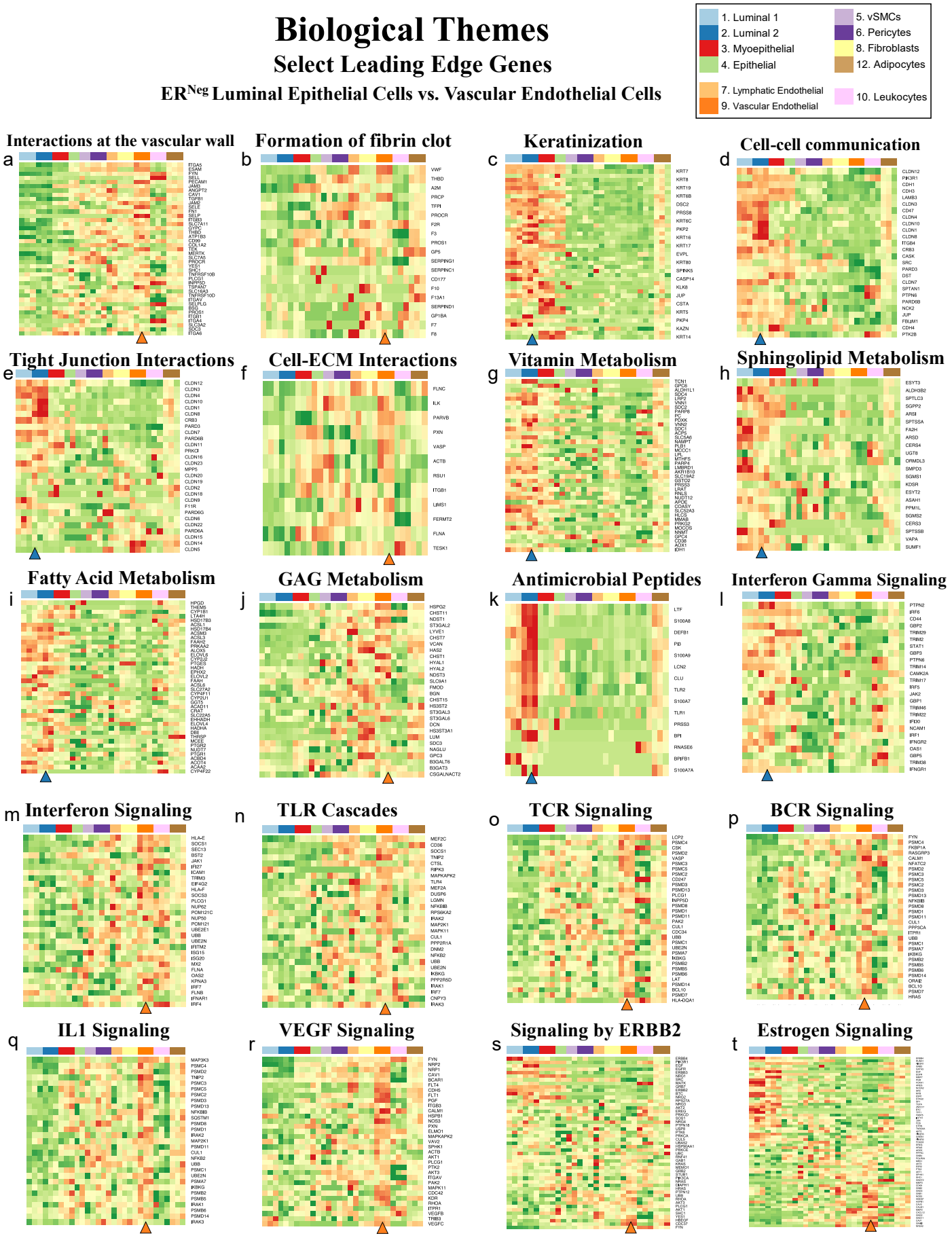


Figure 28

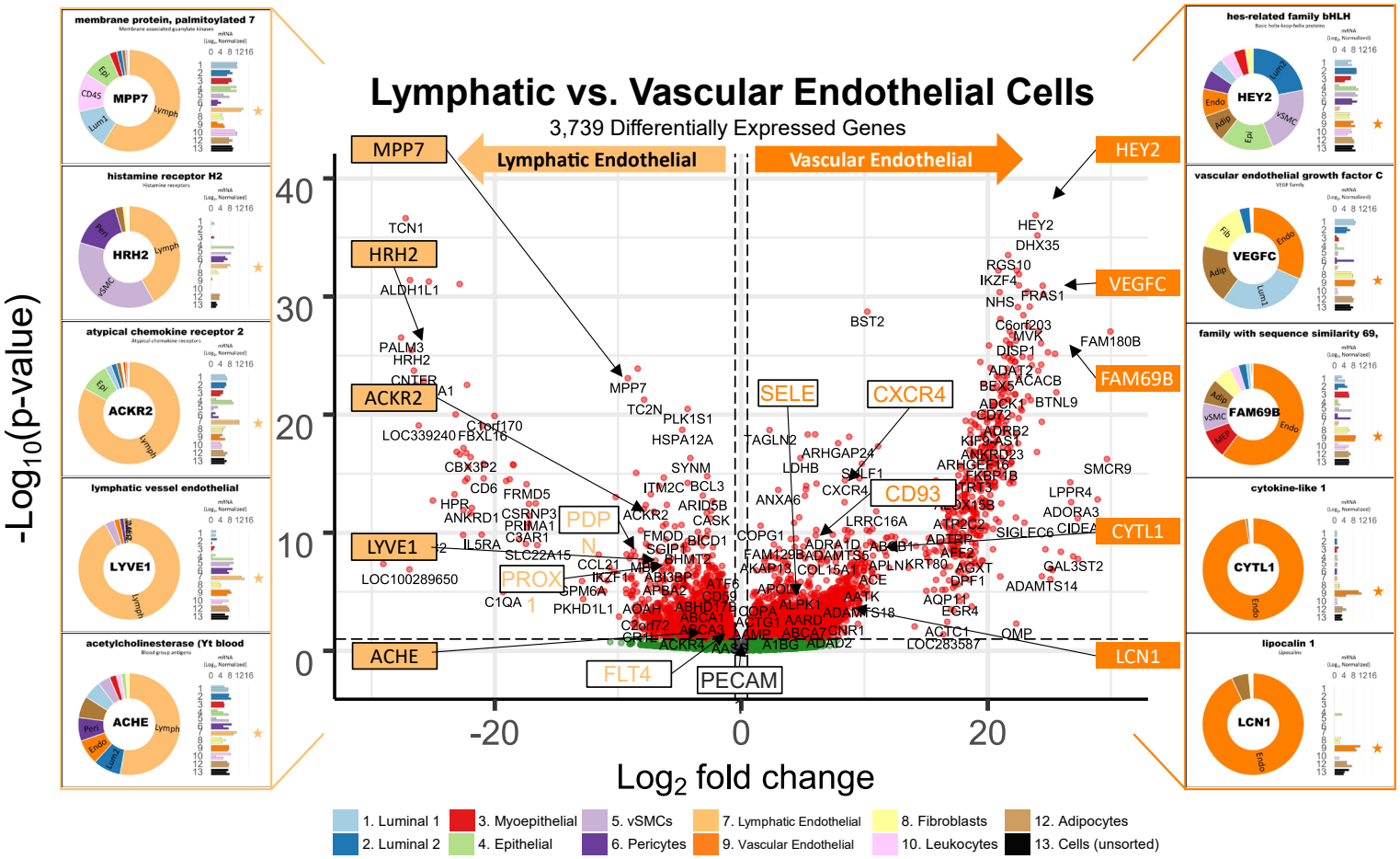

Lymphatic vs. Vascular Endothelial Cells
3,739 Differentially Expressed Genes
Lymphatic Endothelial
Vascular Endothelial
-Log10(p-value)
Log2 fold change
MPP7
HRH2
ACKR2
LYVE1
ACHE
PDPN
PROX1
FLT4
PECAM
SELE
CXCR4
CD93
HEY2
VEGFC
FAM69B
CYTL1
LCN1
membrane protein, palmitoylated 7
histamine receptor H2
atypical chemokine receptor 2
lymphatic vessel endothelial
acetylcholinesterase (Yt blood
hes-related family bHLH
vascular endothelial growth factor C
family with sequence similarity 69,
cytokine-like 1
lipocalin 1
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
9. Vascular Endothelial
8. Fibroblasts
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)

Figure 29

Biological Themes
Select Leading Edge Genes
Lymphatic vs. Vascular Endothelial Cells
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
8. Fibroblasts
12. Adipocytes
7. Lymphatic Endothelial
9. Vascular Endothelial
10. Leukocytes
IL-4 and IL-13 Signaling
a
TIMP1
IL6R
LIF
IL1B
STAT3
RHOU
MAOA
FN1
CCL22
CEBPD
F13A1
ANXA1
STAT1
TGFB1
VIM
BCL2L1
COL1A2
SOCS5
IL13RA1
LBP
IL13RA2
TCR Signaling
b
HLA-DRB1
HLA-DQA1
VASP
PSMD4
PSMC4
CD247
HLA-DRA
PSMC2
PSMD3
PSMC5
PSMA7
HLA-DRB5
PSMC3
PSMD14
HLA-DPA1
PTPRJ
PSMC1
PSMD1
PSMB2
ENAH
PSMA6
PSMC6
PSMD11
PSMD2
WAS
HLA-DQB1
IL-1 Family Signaling
c
IL18RAP
PTPN7
PSMD4
PSMC4
IL1RL1
PSMC2
PSMD3
PSMC5
PSMA7
PSMC3
SAA1
TOLLIP
PSMD14
PSMC1
PSMD1
PSMB2
PTPN13
HMGB1
PSMA6
PSMC6
PSMD11
TNIP2
PTPN23
PSMD2
IL33
d
mannose receptor, C type 1
C-type lectin domain family
mRNA (Log2, Normalized)
0 4 8 12 16
MRC1
Adip
Lymph
1 2 3 4 5 6 7 8 9 10 12 13
e
chemokine (C-C motif) ligand 21
Chemokine ligands
mRNA (Log2, Normalized)
0 4 8 12 16
CCL21
Lymph
1 2 3 4 5 6 7 8 9 10 12 13
f
atypical chemokine receptor 2
Atypical chemokine receptors
mRNA (Log2, Normalized)
0 4 8 12 16
ACKR2
Epi
Lymph
1 2 3 4 5 6 7 8 9 10 12 13
g
nucleotide-binding oligomerization
Caspase recruitment domain containing
mRNA (Log2, Normalized)
0 4 8 12 16
NOD1
Lum1
Lum2
CD45
Fib
Adip
Lymph
Endo
1 2 3 4 5 6 7 8 9 10 12 13
h
Elastic Fiber Formation
TGFB2
EFEMP1
FBLN5
MFAP4
EMILIN1
LTBP1
FBLN1
FN1
LTBP3
ITGA8
TGFB3
MFAP2
TGFB1
ITGB8
EMILIN2
ITGAV
MFAP3
FBN1
MFAP1
FBN3
LOX
LTBP4
ELN
ECM Proteoglycans
i
TGFB2
FMOD
TNXB
COL6A2
PTPRS
BGN
COL6A6
FN1
COL6A1
COL5A2
LAMB2
TNN
ITGA6
DCN
TGFB3
TGFB1
LAMA3
TNC
APP
COL6A3
ITGA3B
COL1A2
MATN1
ITGAV
LRP4
COL3A1
Collagen Biosynthesis
j
COL4A6
PAHA1
COL25A1
COL27A1
COL5A3
COL5A3
COL4A1
SERPINH1
COL1A1
PCOLCE
COL4A2
COL19A1
ADAMTS3
COL5A1
P4HB
TLL2
COL21A1
COL7A1
COL14A1
PLOD1
COL13A1
COL8A1
ADAMTS14
FP1
COL5A1
COL9A3
COL9A2
VEGF Signaling
k
VEGFA
VEGFB
VEGFC
FLT1
FLT4
NRP1
NRP2
SRC
PAK3
CYBA
ROCK2
MAPK14
PRR5
VAV3
PRKACA
MAPK13
JUP
PAK2
ITPR1
ITGAV
KRAS
NCK2
CYFIP2
AXL
PAK1
BAIAP2
WASF1
PIK3CA
CYBB
PRKCA
Notch Signaling
l
HEY2
ELF3
ATP2A3
PSMC4
WWC1
PRKCI
E2F3
PSMC5
NEURL1
PSMC3
PSMD14
PSMC1
PSMD1
PSMA6
NEURL1B
HDAC1
ST3GAL4
NOTCH3
HDAC3
HDAC6
PSMD13
PSMB7
PSMB1
DLL4
Notch2 Signaling
m
MAML2
NOTCH2
ADAM10
PSEN2
EP300
RBPJ
CREB1
FCER2
MAMLD1
MIB1
MDK
UBB
DLL1
MAML3
PSEN1
APH1B
UBC
Muscle Contraction
n
TRPC1
FKBP1B
ANXA6
FXYD1
KCNJ14
NPPC
TNNT2
DMD
KCNE3
KCNK6
MYL9
KCNE2
KCNQ1
KCNK5
ATP2A3
TCAP
CALM1
SCN4A
ITPR2
FGF12
FGF11
VCL
LMOD1
MYL5
ACTG2
ACTC1

Figure 30

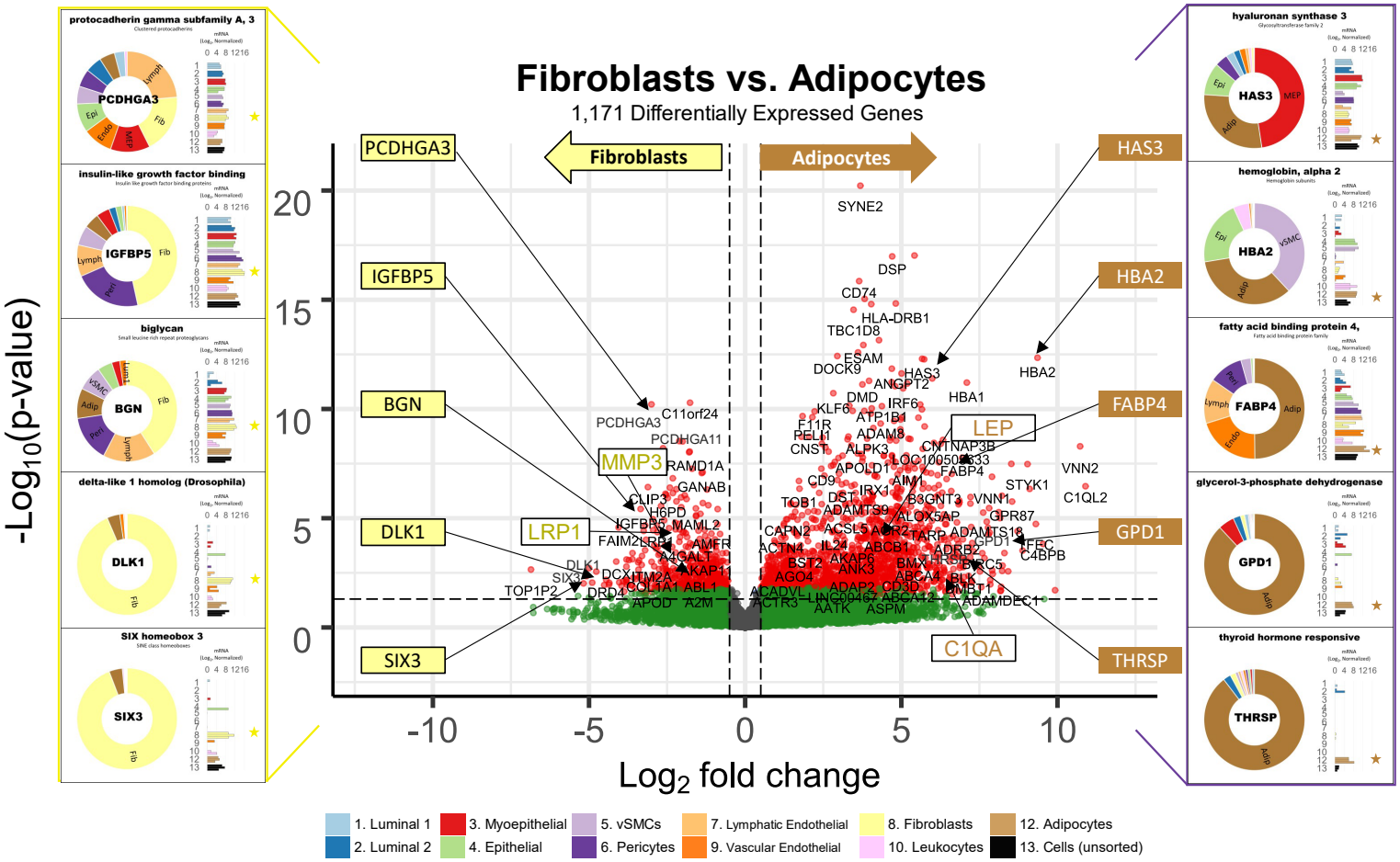

Fibroblasts vs. Adipocytes
1,171 Differentially Expressed Genes
Fibroblasts
Adipocytes
-Log10(p-value)
Log2 fold change
PCDHGA3
IGFBP5
BGN
DLK1
SIX3
HAS3
HBA2
FABP4
GPD1
THRSP
MMP3
LRP1
LEP
C1QA
protocadherin gamma subfamily A, 3
insulin-like growth factor binding
biglycan
delta-like 1 homolog (Drosophila)
SIX homeobox 3
hyaluronan synthase 3
hemoglobin, alpha 2
fatty acid binding protein 4,
glycerol-3-phosphate dehydrogenase
thyroid hormone responsive
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)

Figure 31

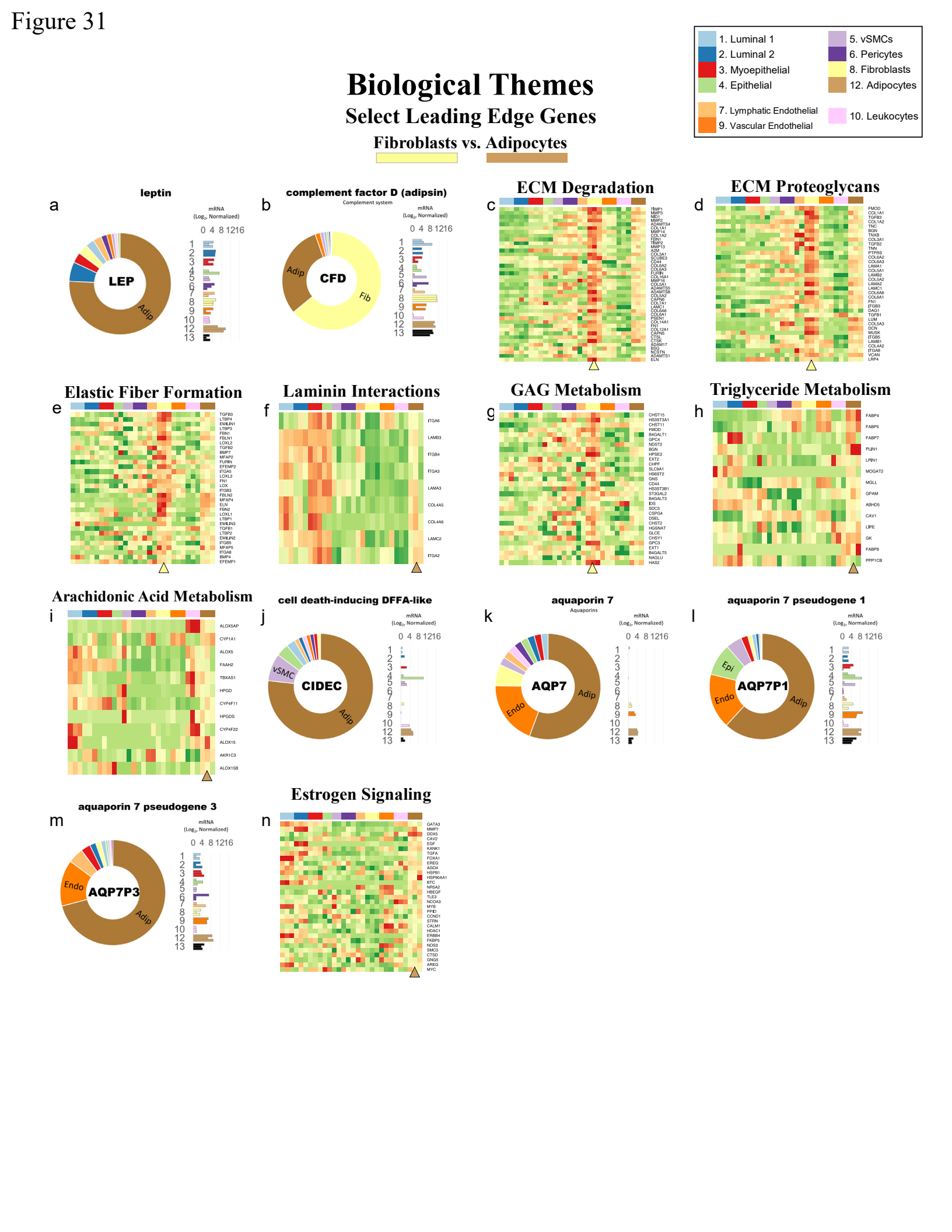

1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
8. Fibroblasts
12. Adipocytes
7. Lymphatic Endothelial
9. Vascular Endothelial
10. Leukocytes
Biological Themes
Select Leading Edge Genes
Fibroblasts vs. Adipocytes
a
leptin
LEP
Adip
b
complement factor D (adipsin)
Complement system
CFD
Adip
Fib
c
ECM Degradation
d
ECM Proteoglycans
e
Elastic Fiber Formation
f
Laminin Interactions
g
GAG Metabolism
h
Triglyceride Metabolism
i
Arachidonic Acid Metabolism
j
cell death-inducing DFFA-like
CIDEC
vSMC
Adip
k
aquaporin 7
Aquaporins
AQP7
Endo
Adip
l
aquaporin 7 pseudogene 1
AQP7P1
Epi
Endo
Adip
m
aquaporin 7 pseudogene 3
AQP7P3
Endo
Adip
n
Estrogen Signaling
mRNA
(Log2, Normalized)
0 4 8 1216

Figure 32

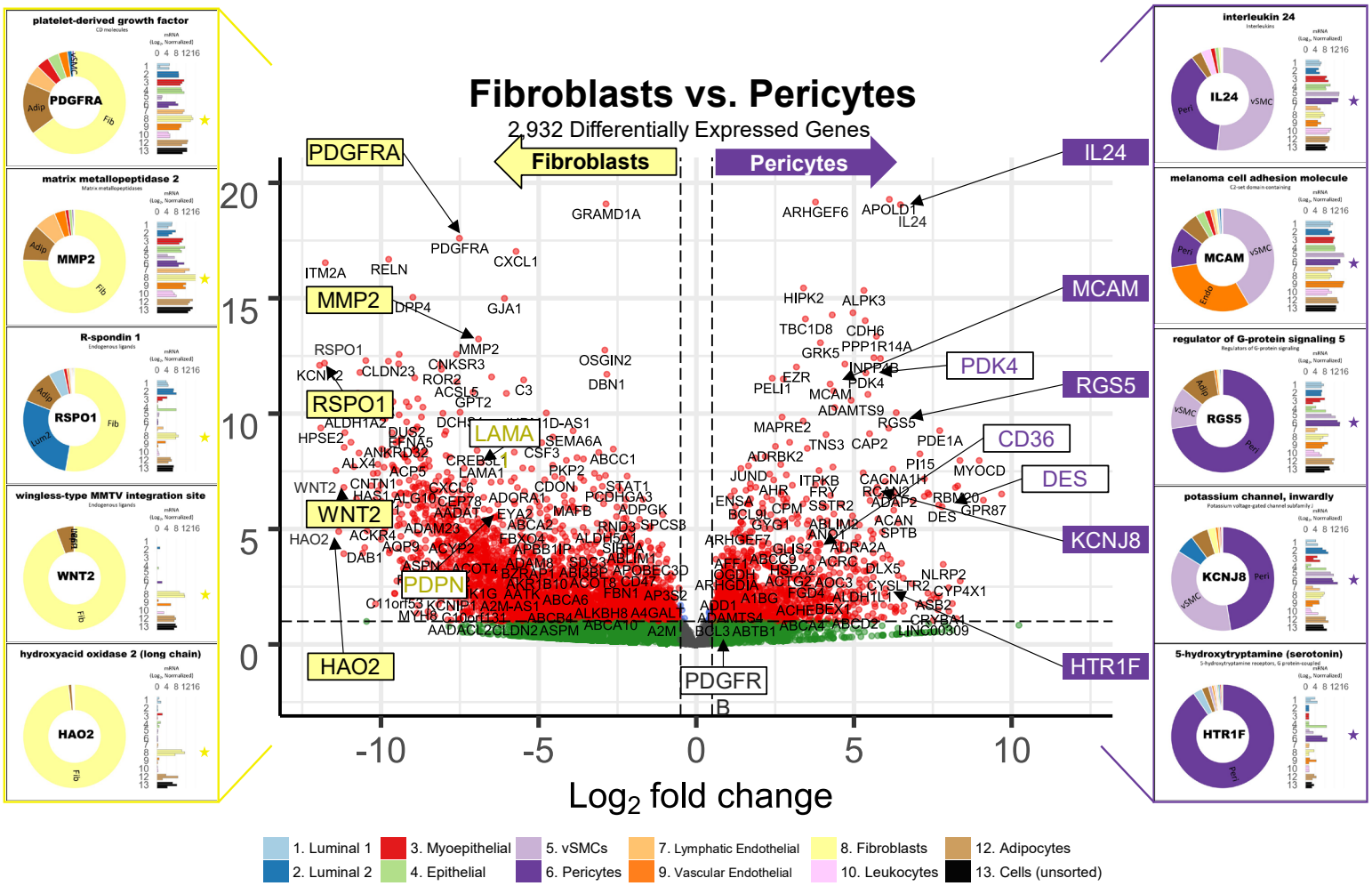

Fibroblasts vs. Pericytes
2,932 Differentially Expressed Genes
Fibroblasts
Pericytes
PDGFRA
MMP2
RSPO1
LAMA
WNT2
PDPN
HAO2
PDGFRB
IL24
MCAM
PDK4
RGS5
CD36
DES
KCNJ8
HTR1F
Log2 fold change
platelet-derived growth factor
matrix metallopeptidase 2
R-spondin 1
wingless-type MMTV integration site
hydroxyacid oxidase 2 (long chain)
interleukin 24
melanoma cell adhesion molecule
regulator of G-protein signaling 5
potassium channel, inwardly
5-hydroxytryptamine (serotonin)
1. Luminal 1
2. Luminal 2
3. Myoepithelial
4. Epithelial
5. vSMCs
6. Pericytes
7. Lymphatic Endothelial
8. Fibroblasts
9. Vascular Endothelial
10. Leukocytes
12. Adipocytes
13. Cells (unsorted)

Figure 33

Enriched Gene Sets & Biological Themes
Mesenchymal Cells:
Pericytes (Pop6,●) vs. Fibroblasts (Pop8,●)

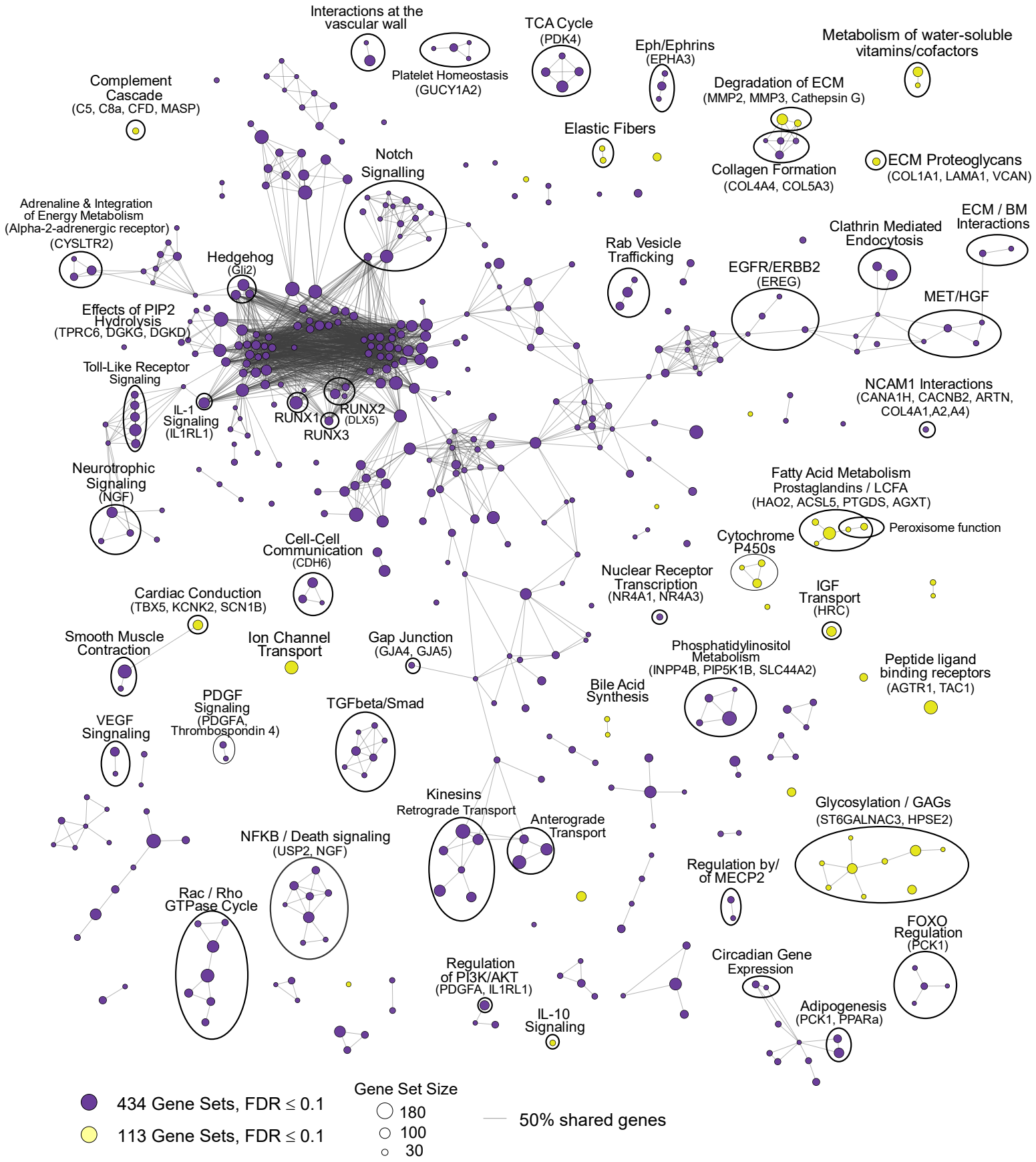


Figure 33—figure supplement 1