UCSF

Title

Comprehensive identification, isolation, and culture of human breast cell types.

Permalink

Journal

Authors

Thi, Kate
Del Toro, Katelyn
Licon-Munoz, Yamhilette
et al.

Publication Date

DOI

Comprehensive identification, isolation, and culture of human breast cell types

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

From the [1]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; [2]Department of Biochemistry and Molecular Biology, University of New Mexico School of Medicine, Albuquerque, New Mexico, USA

Reviewed by members of the JBC Editorial Board. Edited by Paul Shapiro

Tissues are formed and shaped by cells of many different types and are orchestrated through countless interactions. Deciphering a tissue's biological complexity thus requires studying it at cell-level resolution, where molecular and biochemical features of different cell types can be explored and thoroughly dissected. Unfortunately, the lack of comprehensive methods to identify, isolate, and culture each cell type from many tissues has impeded progress. Here, we present a method for the breadth of cell types composing the human breast. Our goal has long been to understand the essence of each of these different breast cell types, to reveal the underlying biology explaining their intrinsic features, the consequences of interactions, and their contributions to the tissue. This biological exploration has required cell purification, deep-RNA sequencing, and a thorough dissection of the genes and pathways defining each cell type. While the molecular analysis is presented in an adjoining article, we present here an exhaustive cellular dissection of the human breast and explore its cellular composition and histological organization. Moreover, we introduce a novel FACS antibody panel and rigorous gating strategy capable of isolating each of the 12 major breast cell types to purity. Finally, we describe the creation of primary cell models from nearly every breast cell type—some the first of their kind—and submit these as critical tools for studying the dynamic cellular interactions within breast tissues and tumors. Together, this body of work delivers a unique perspective of the breast, revealing insights into its cellular, molecular, and biochemical composition.

Modern advances in single-cell technologies and cytometry permit high dimensional analyses of numerous simultaneous markers and are capable and well-suited for interrogating complex cell populations (1, 2). These kinds of investigations have led to a more considerable appreciation of the types and properties of cells in complex tissue systems (*e.g.*, the hematopoietic system), have given us insight into normal cell development and physiology, and have been crucial to developing a broader understanding of the malignant state (3, 4). While emerging technologies like single-cell RNA sequencing and digital spatial profiling are helping unravel the cellular heterogeneity in tissues, deep RNA sequencing of purified populations provides a broader assessment of expressed genes and proteins. Unfortunately, the full range of capabilities offered by coupling complex multicolor FACS strategies and deep RNA-sequencing has yet to be applied *fully* to many solid tumor and tissue types, including the breast. Obstacles include the relatively limited availability and amounts of freshly isolated specimens, technicalities associated with producing single-cell suspensions from solid tissues (5), the sparsity of specific cell types causing them to be easily overlooked, as well as the immense amount of sort-time and labor required to collect enough rare cells for deep analysis. Progress in the field is further restricted by the expertise and technical knowledge needed to design and optimize intricate antibody panels (6, 7), properly operate a sophisticated flow sorter, and interpret the resulting complex multi-parameter datasets—information that can be dangerously misleading if not handled carefully (6–9). These are indeed time-consuming and challenging endeavors (10).

A major impediment to analyzing breast tissues by flow cytometry is the absence of an accepted set of markers and strategy for purifying the entire spectrum of cell types constituting this tissue. This is true for adult stem cells and most other cells in the breast, and many technical caveats and perplexing and contradictory claims persist in the literature (5, 11–13). These challenges and enduring questions have long contributed to controversies regarding the nature, relationship, and function of different cell types in breast tissues (5, 14, 15), collectively placing the field—and our understanding of these different cell types—into a state of uncertainty (16).

Resolving the types and properties of all breast cells is vital to understanding tissue biology and the emergence of diseases such as cancer. Achieving such a comprehensive understanding ultimately relies on analyzing each cell type's molecular profile and functionally characterizing these cells. Thus, inclusive and reliable methods are needed to identify, purify, and culture each breast cell lineage. To establish such protocols and cell models, we have revisited our understanding of the breast's cellular composition and organization through exhaustive tissue immunostaining and have matched results with cytometric analyses performed in parallel. We have used this knowledge to develop a comprehensive FACS antibody

Present Address for Rosalyn W. Sayaman: Department of Laboratory Medicine, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA.
* For correspondence: William C. Hines, wchines@salud.unm.edu.

panel and gating strategy that permits objective identification and isolation of every major cell type in the breast. These twelve cell populations include two distinct luminal epithelial fractions, myoepithelial cells, adipocytes, leukocytes (counted here as a single population), pericytes, vascular smooth muscle cells, erythrocytes, adipose-derived mesenchymal stem cells (fibroblasts), lymphatic and vascular endothelial cells, and a newly identified epithelial cell type. This method considerably expands the number and types of cells isolated and cultured from the breast from the two to three commonly reported. Development and validation of this FACS method occurred over many years; Here, we present critical discoveries that guided its evolution—including how newly isolated cell fractions were identified and validated—and how critical pitfalls were encountered and overcome. Furthermore, we demonstrate how nearly every cell type can be seeded into culture and expanded for future analyses.

Most of the breast literature to date has dealt with either the epithelium or stroma, with a focus primarily confined to a few cell types within each of these compartments, such as luminal and myoepithelial cells—or fibroblasts and adipocytes (Fig. 1, *A* and *B*). Widespread adoption of primary cell models has also been limited—essentially to myoepithelial cells, fibroblasts, or mixed unsorted human mammary epithelial cultures (HMECs), which are commercially available and ultimately self-select for faster-growing myoepithelial cells (Fig. S1). Of course, it is understood that the tissue contains additional cell types, for example, cells that compose blood vessels, immune cells, and possibly others (Figs. 2, and 3). Knowing where these different cell types reside in the breast is not always abundantly clear, and markers and methods to objectively classify and prospectively isolate each cell type from the breast have proved difficult to establish. This has left the field without precise methods to purify most cell types from breast tissues—and even less from tumors. Epithelial cells have received the bulk of past attention, with the most commonly reported marker-combination for discriminating these being "epithelial cell adhesion molecule" (EPCAM) and alpha-6 integrin (CD49f) (17–25). In our hands, this marker pair resolved four different cell populations from normal reduction mammoplasty breast tissues (Fig. 4). However, reports using these markers have varied on the number, types, and functional capabilities of the cell populations identified (Ibid). Whether these disparities between studies reflect technical differences or underlying biological diversity was unclear at the outset of this study, but the discord has generated uncertainty. To expand upon these methods and make deductive inferences about each of the different cell types in the breast, we set out to define a comprehensive FACS strategy capable of objectively and unambiguously resolving each cell type from the tissue.

Results and discussion

FACS method development: key examples and guiding principles

To explore the breast's architecture and decipher its cellular composition—and also test different marker combinations
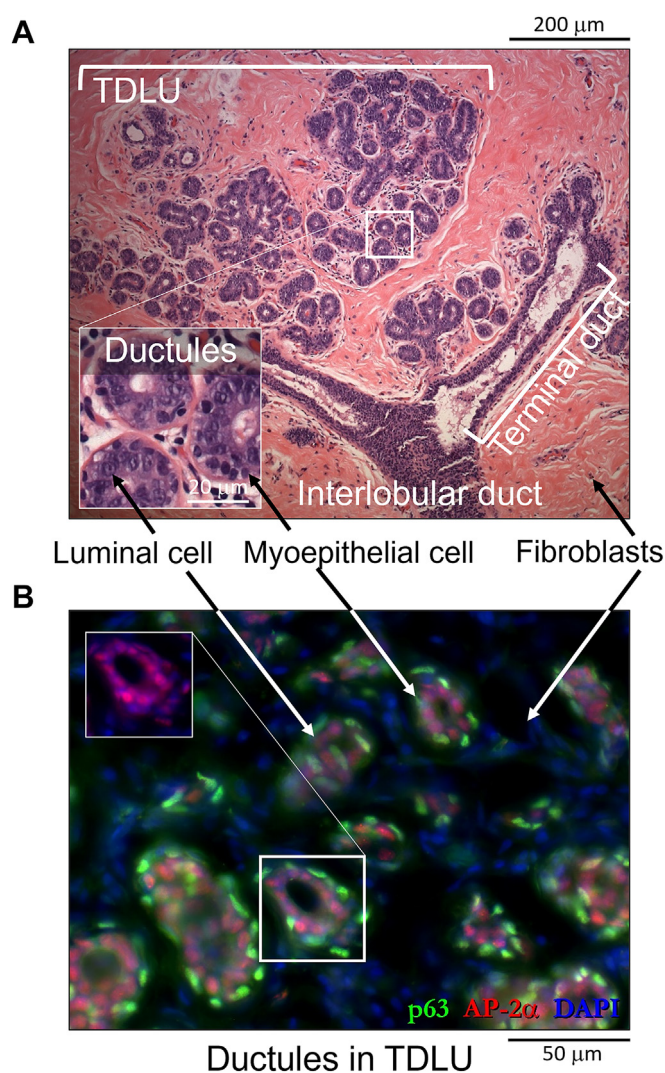

Figure 1. Normal Breast Histology. *A*, H&E staining of a normal breast reveals the breast's compound alveolar glandular architecture (10× objective; 34 years/o female). The terminal ductal lobular units (TDLUs)—also referred to as lobules— along with their associated terminal ducts and an interlobular duct are embedded in an eosin-staining collagenous stroma. The two major epithelial cell types (myoepithelial and luminal epithelial cells) are discernible within the magnified tube-shaped ductules that compose the TDLUs (magnified inset). *B*, immunostaining for the transcription factor AP-2α (*red staining*) reveals nuclei of all epithelial cell types, whereas the superimposed p63 staining (*green*) is limited to the basally located myoepithelial cells (40× objective, 35 years/o female). (*B*, inset) Removal of the p63 overlay reveals myoepithelial AP-2α expression.

used for FACS-isolating cells—we prospectively collected and analyzed normal tissues (reduction mammoplasty specimens) from 65 women, whose ages ranged from 16 to 61 years ($\bar{x}$ = 29.4 ± 9.8 years). We froze, sectioned, and immunostained these tissues with over 60 different primary antibodies (Figs. 5, *A–F*, S2, *A–F*, S3, *A–C*, and Table S1). We further analyzed antibodies displaying cell-type (and cell-surface) specificity by flow cytometry to determine if they could resolve viable cell populations. Our goal was to find markers that maximized resolution between cell types that would, in turn, increase the objectivity of the isolation strategy. Unfortunately, many antibodies failed in this regard despite their marked ability to expose different cell types when used on tissue. The luminal-
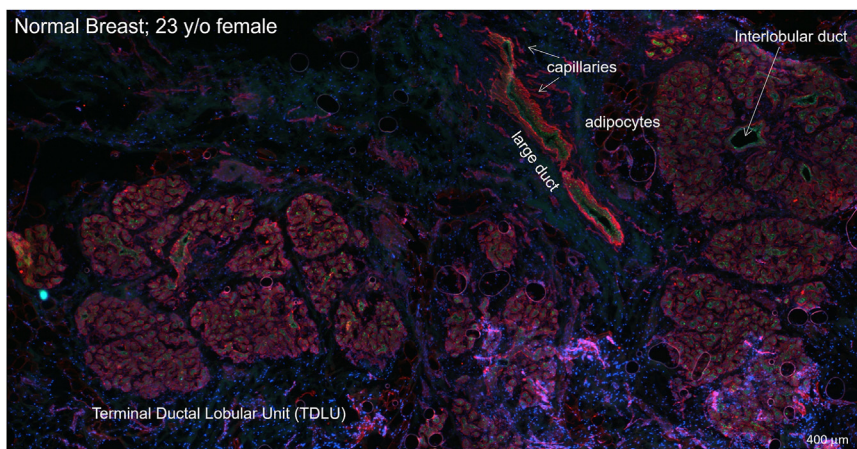
Figure 2. Breast Tissue Landscape. A low-power (4×) microscopic tiled scan of normal breast tissue (reduction mammoplasty, 23-year-old female) immunostained with keratin 18 (*green*) and pan-laminin (*red*) antibodies exposes the arrangement of cell types and tissue elements within the breast. Several large TDLUs, a large lactiferous duct, adipocytes, interlobular duct, and a dense epithelium-adjacent vasculature, are present. Nerve fibers were not observed, consistent with previous anatomical reports describing the breast's sensory innervation as being limited to superficial fascia and an areolar subplexus (55, 56). EVOS FL Auto imaging scan; 4× Plan LWD Fluor 0.13NA objective.

cell-specific antibodies we tested are a prime example: That is, out of the eight we screened by tissue immunostaining, only three adequately resolved luminal cells when measured by FACS, with both CD24 and Muc1 staining more intensely and providing better resolution than the frequently used EpCAM antibody (Figs. 6, *A–H* and S4–S11).

Why so few luminal-cell-specific antibodies could not adequately resolve luminal cells by FACS is puzzling given that most performed exceptionally well when used on tissue sections. We attributed this discrepancy to the distinct detection strategies employed by these two methods, *i.e.*, indirect staining of tissues vs. the direct staining used in FACS. The differential sensitivity of protein markers to trypsin—currently an unavoidable step in preparing cell suspensions for FACS analysis—may also contribute. Nevertheless, the antibody
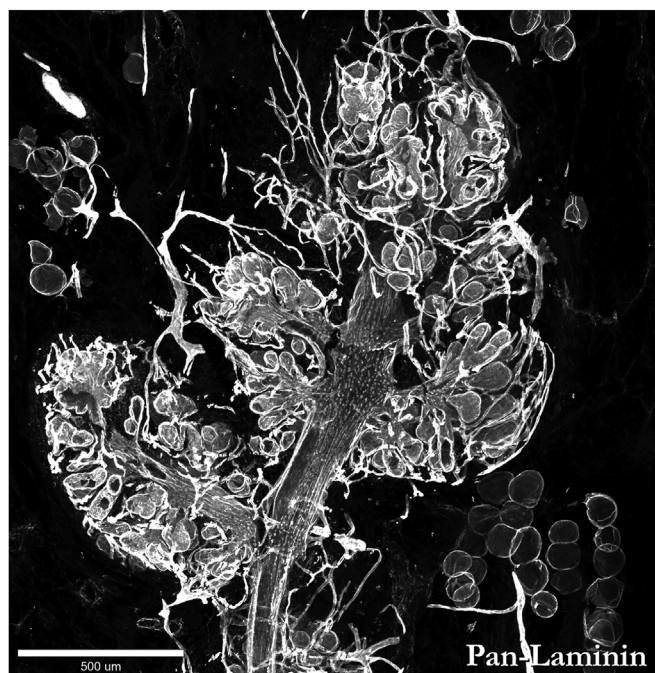

Figure 3. 3-dimensional organization of the breast's stromal and epithelial components. A maximum intensity projection of a confocal Z-stack of a 72.25 μm thick breast tissue section immunostained for pan-laminin. Staining reveals the basement membranes of the epithelium, blood vasculature, and adipocytes. This reduction of mammoplasty tissue was from a 33-year-old female. Imaged on a Zeiss LSM 710 confocal microscope; Plan-Apochromat 10×/0.3 DIC M27 objective; 500 μm scale.
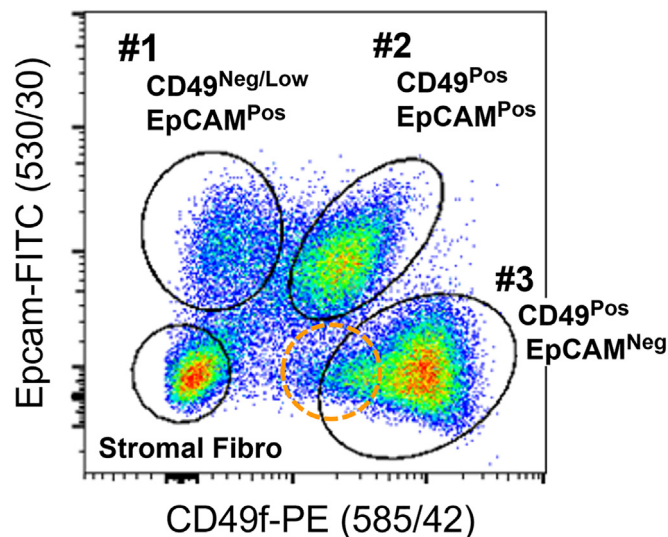

Figure 4. Flow cytometric analysis of breast cells. EpCAM and CD49f staining of normal breast cells resolves three epithelial cell populations by flow cytometry: #1 CD49f[Neg] EpCAM[Pos], #2 CD49f[Pos]EpCAM[Pos], and #3 CD49f[High] EpCAM[Neg]. The dual negative CD49[Neg]EpCAM[Neg] population is often described as "stroma" or "stromal fibroblasts." During the development of our FACS panel, we discovered that if other markers and gates are used to categorize the cells, several other cell populations are revealed, which occupy the general area indicated by the *orange dotted line*. These cell types include two endothelial populations, myoepithelial cells (MEPs) and luminal epithelial cells (LEPs), and a rare epithelial cell population. We have found also that the dual negative 'stromal' area contains fibroblasts, vascular smooth muscle cells, pericytes, luminal epithelial cells, red blood cells, and leukocytes—the relative abundance of which depends on the lineage-depletion and detection strategy employed. This reduction mammoplasty tissue was from a 22-year-old female (sample #N141).
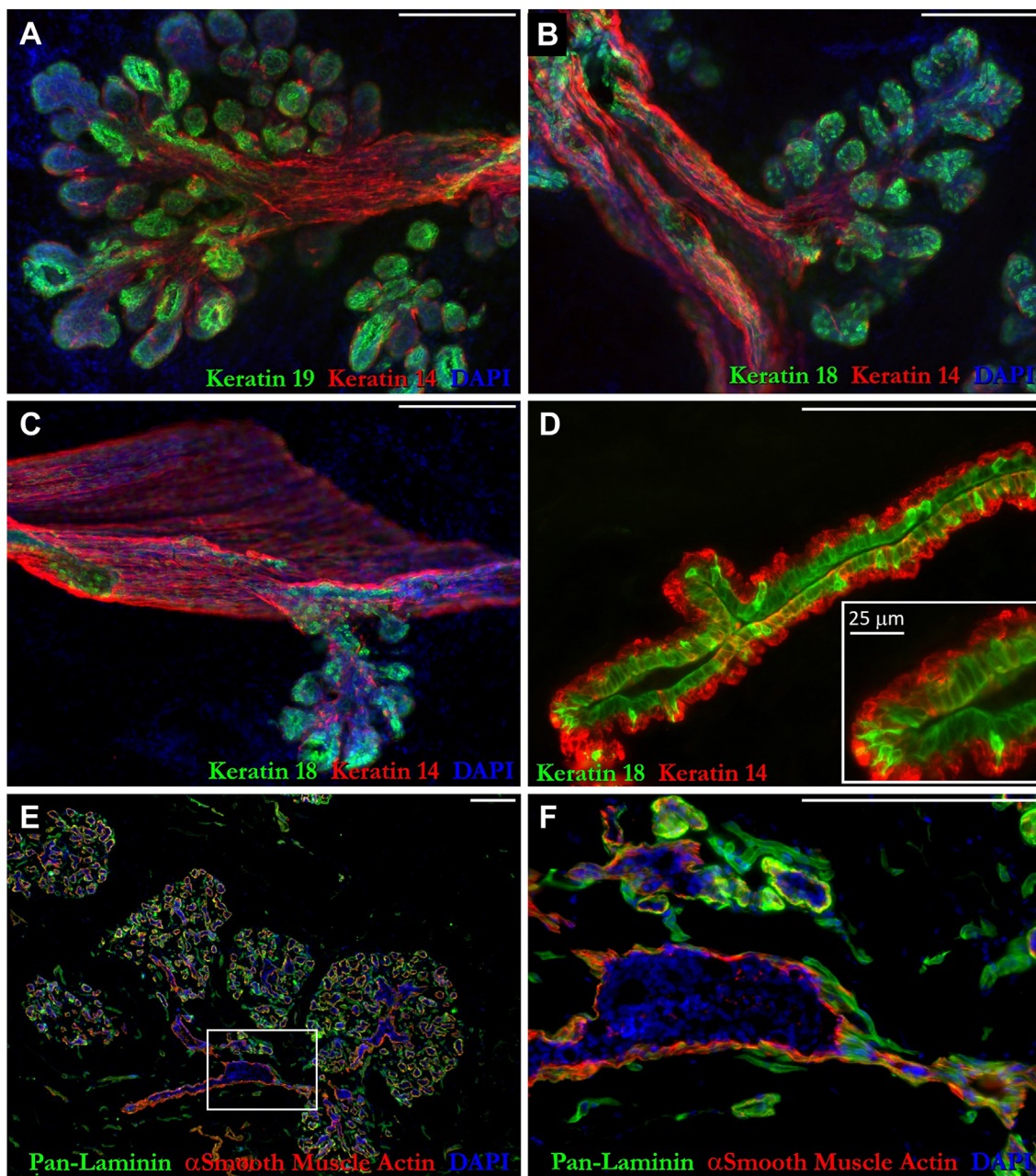
Figure 5. Cellular composition and architecture of breast tissues. *A–C*, representative immunostains of 80 μm thick breast tissue sections illustrating the complex three-dimensional organization of cell types forming the lactiferous ducts and terminal ductal lobular unit structures in the breast. Myoepithelial cells, identified by intense keratin 14 staining, are arranged longitudinally along large, pleated ducts (*A–C*) and often appear as compressed tubes (*C*). Luminal epithelial cells are distinguished by their expression of keratin 18 and 19 (*A–C*). *D*,10 μm breast tissue section stained for keratins 18 and 14. Some luminal cell projections are seen extending basally to the basement membrane (*D*, magnified inset). *E*, A low-power image of 10 μm breast section illustrates the branching glandular architecture and tortuous system of tubules within the TDLUs. *F*, at higher magnification (*F*, *boxed area* in *E*), the entangled pan-laminin staining capillary network running parallel to the glandular epithelium becomes evident. 1 mm scale bar (all images).

screen was successful, and through the exercise, we found that both CD24 and Muc1 antibodies enhanced the resolution of the luminal cell populations. We now needed to choose which of these two antibodies to use in our final staining panel—and testing them together led to a pivotal discovery.

When we co-stained breast cell suspensions with CD24 and Muc1 antibodies (each labeled with distinct fluorophores) and analyzed them by flow cytometry, most luminal cells bound both. However, we discovered that there was always a fraction of luminal cells that repeatedly and reproducibly escaped detection if we used either antibody in isolation. Some cells stained positively for one marker but not the other—and *vice versa* (Fig. 7, *A* and *B*). This observation was critical. If we had used a single marker as is commonly practiced, most of these "missed" luminal epithelial cells would have been incorrectly sorted as myoepithelial cells and gone unnoticed (Fig. S12). This incorrect assignment would have confounded results and possibly produced erroneous conclusions had it gone
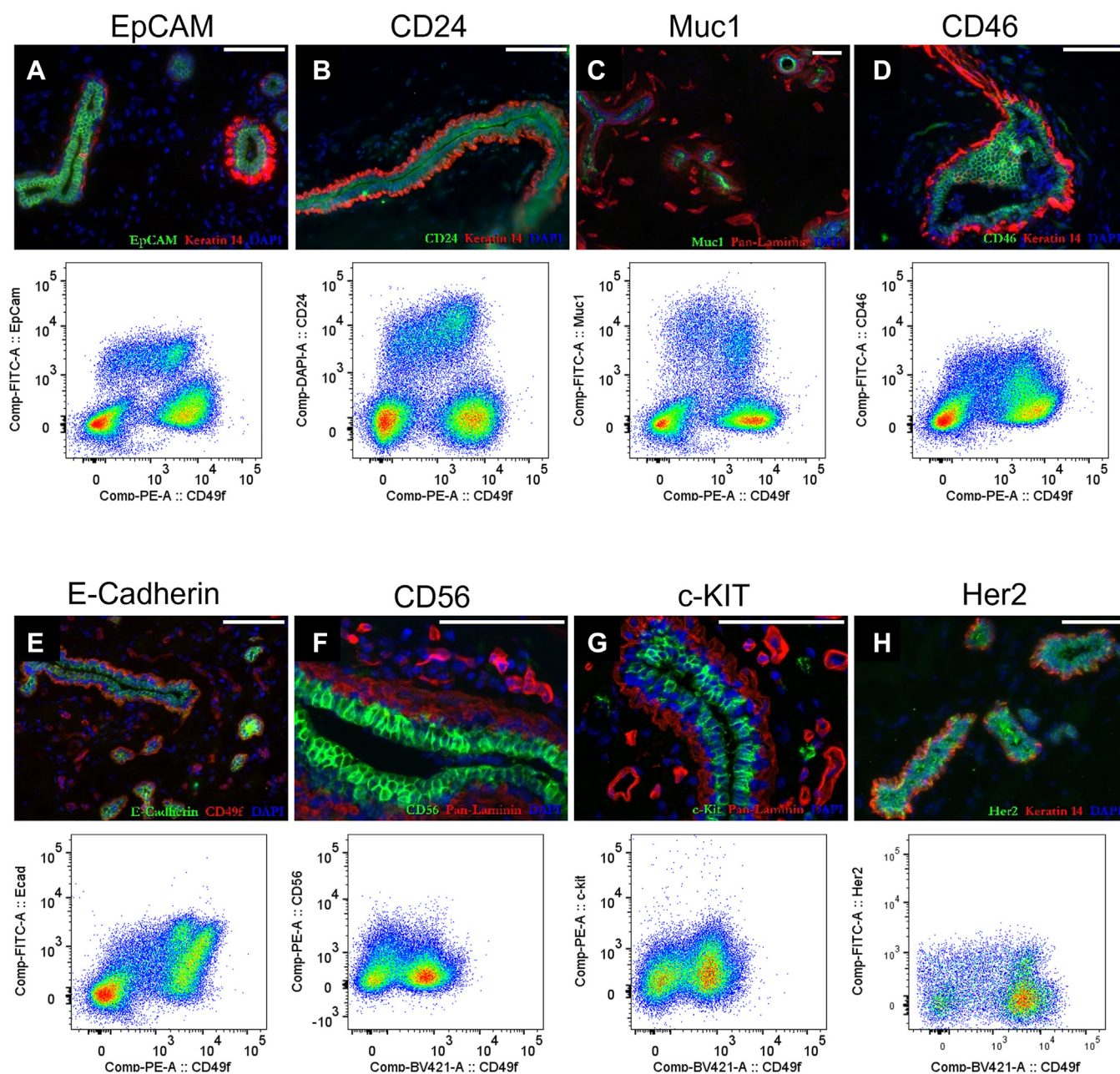
Figure 6. FACS antibody screen identifies CD24 and Muc1. *A–H*, tissue immunostains); Normal breast tissue, immunostained with a panel of antibodies predicted to label luminal epithelial cells; Scale = 0.1 mm. *A–H*, FACS plots, flow cytometry scatter plots of breast cell suspensions (from reduction mammoplasty tissues) stained with each indicated antibody.

unnoticed. To avoid this problem and ensure proper luminal-cell discrimination, we discovered we needed to use both markers and take advantage of the redundancy. We thus used CD24 and Muc1 antibodies in all subsequent experiments. This unexpected finding was an object lesson and set a precedent as we moved forward with panel design. To ensure proper classification of the other cell types, we also needed to build redundancies into their sorting strategies. However, finding markers that would adequately separate each cell type by FACS was a persistent challenge.

A noteworthy example of mistaken identity and low resolution involved CD31—a lineage marker frequently used to identify endothelial cells. Like many other antibodies we tested, CD31 performed exceptionally well on frozen breast sections. The antibody marked capillaries and other vessels in tissues (Figs. 8*A* and S13*A*). However, even after titrating and optimizing antibody staining, we found the resolution in our FACS experiments was often subpar. Identifying this problem was somewhat tortuous. It became apparent only after we had sorted epithelial fractions by FACS, isolated their RNA, and measured more than 100 transcripts using a custom quantitative RT-PCR array. Among the genes tested was the hematopoietic stem cell marker *CD34*, which we curiously found highly expressed in the sorted myoepithelial fraction

A

Viable, single cells
Gated: CD45^{Neg} CD34^{Neg} Thy1^{Neg}
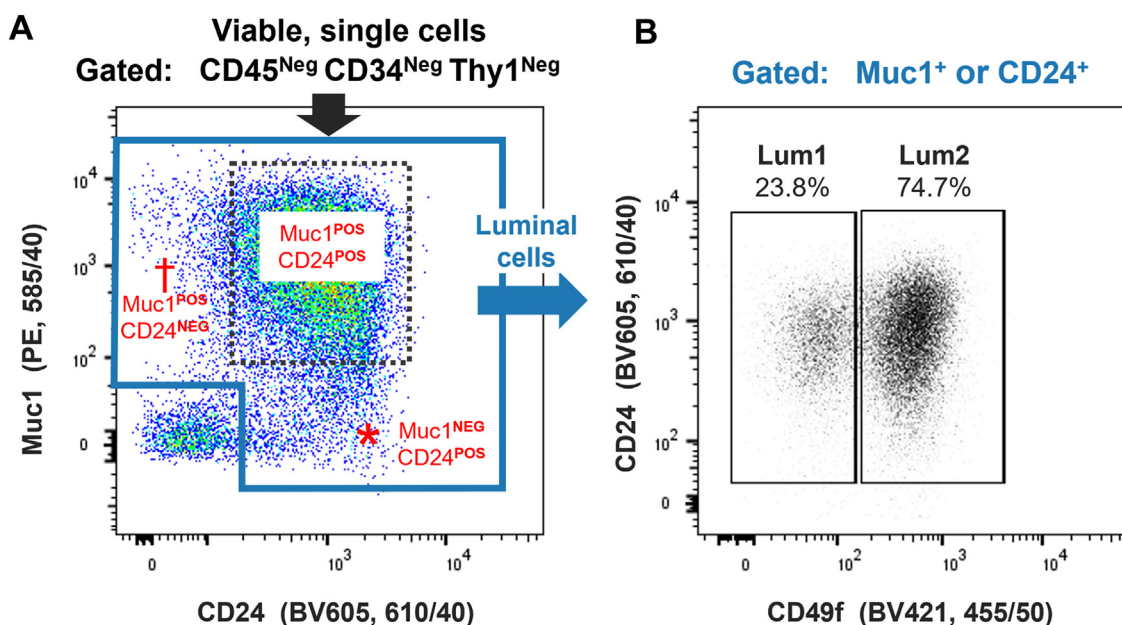
B

Gated: Muc1$^+$ or CD24$^+$

Figure 7. Luminal cells escape detection if single FACS marker is used. *A,* representative FACS analysis of normal breast cells co-stained with luminal epithelial cell markers Muc1 and CD24 (using independent markers: PE and BV605). Cells were backgated for the viability marker To-Pro-3^{Neg} (viable cells), SSCWidth^{Low} (single cells), CD45^{Neg}, CD34^{Neg}, and Thy1^{Neg} (which removed dead cells, cell clusters, leukocytes, fibroblasts, endothelial cells, pericytes, and myoepithelial cells from the analysis). Most remaining cells were indeed stained for both Muc1 and CD24, identifying them as luminal cells (*dotted box*). However, a fraction of luminal cells stained only for one luminal marker, but not the other; *that is,* they stained for Muc1^{POS}CD24^{NEG} (†) or Muc1^{NEG}CD24^{POS} (*). The "luminal cell" gate (*blue outline*) indicates cells that expressed Muc1, CD24, or both. *B,* FACS plot (CD24 vs. CD49f) of the gated luminal cells from the previous plot. Cells were classified as being either "Luminal1" (Lum1) or "Luminal2" (Lum2) cells, based on their respective staining for CD49f (α6-integrin). The machine's bandpass filter used for each fluorophore is indicated on its corresponding axis (*e.g.*, a 610/40 filter was used for detecting the CD24-BV605 antibody).

(Fig. S14*A*). This myoepithelial expression was at odds with our tissue immunostaining results, where we found no evidence of *CD34* in myoepithelial cells (Figs. 8*B*, S13*B*, S14, *B* and *C*, and S15, *A* and *B*). Based on this peculiar result, we questioned the purity of our gated myoepithelial fraction and investigated further. We added a CD34-specific antibody to our FACS panel and repeated the analysis. Upon examining the FACS data, we immediately noticed that cells resolved by CD31 were also stained for CD34 (Fig. 8, *C* and *D*). By comparing the inverse, *that is,* exploring CD31 expression in the CD34^{Pos} gated populations, it became evident that the CD31 antibody was not resolving the entire endothelial cell population (Fig. 8*E*). Because endothelial cells express CD49f at a level similar to myoepithelial cells, those that escaped detection by CD31 were being gated into—and were contaminating—the myoepithelial cell fraction (these were the unknown cells detected in Fig. S12). This endothelial contamination explained why we had initially detected *CD34* mRNA in our FACS-isolated myoepithelial fraction. We rectified this problem by replacing the CD31 antibody with one specific to CD34.

A silver lining to the above exercise is that we discovered that the resolution provided by the CD34 antibody was unmatched by any other marker we have tested to date. In normal tissues, CD34 excelled at identifying endothelial cells when combined with CD49f. Indeed, sorting and culture of the CD34^{Pos} CD49f^{Pos} population produced homogeneous cell lines with prominent endothelial characteristics (Fig. 8*F*). These cells formed endothelial cords and vessel-like structures when seeded on Matrigel or fibroblast feeder layers, respectively (Fig. 8, *G* and *H*). Because CD34 provided excellent resolution, we made it a central marker in our final FACS strategy that we outline in Figure 9.

Comprehensive identification and isolation of cell types

The central nodes of our sorting strategy, including markers used and cell types they each contain, are detailed in the following sections—along with evidence supporting their validation. For some cell types, we needed primary cultures to establish or confirm their identity. In other cases, molecular data from RNA-sequencing experiments provided the necessary clues. In the end, we used a panel of nine different antibodies and the viability dye To-Pro-3 to isolate twelve viable cell types. These antibodies are specific to CD36, CD34, CD49f, Podoplanin, Thy1, CD10, CD24, Muc1, and CD45. The development of this panel and the design of our final gating strategy were guided by tissue immunostaining using these and many other antibodies (Table S1). Select images of tissues stained with several of these markers—from the thousands collected—are provided in Figure 10, along with high-resolution images and RNA-seq-derived transcript levels in each isolated cell fraction (Figs. S18–S29). Additional images are available on our website, www.breastcancerlab.com. The mRNA sequencing and thorough analysis of transcripts from each cell type described here are provided in a separate article (26).

Our final FACS gating strategy, illustrated in Figure 9, comprises five primary nodes: (I) The adipocyte-containing
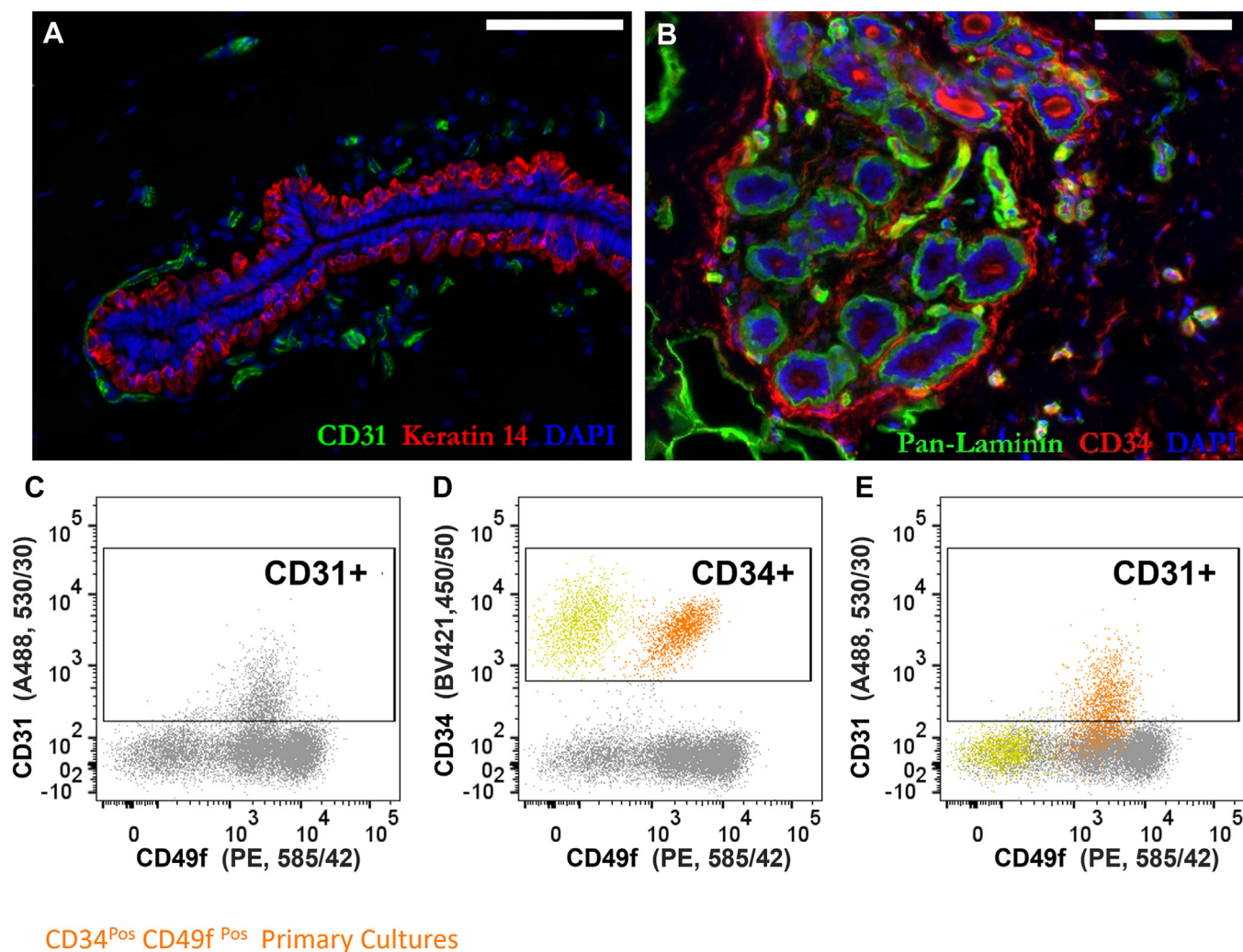
Figure 8. Isolation and validation of vascular endothelial cells. *A*, Breast tissue immunostained for CD31 reveals capillaries adjacent to a mammary lactiferous duct. *B*, CD34-staining labels endothelial and other stromal cell types. *C–E*, FACS analysis of normal breast cells co-stained with CD31 (A488), CD34 (BV421), and CD49f (PE). *C*, CD31 vs. CD49f scatterplot illustrates the incomplete resolution of endothelial cells by the CD31 antibody. *D*, same data as in *C*, now contrasting CD34 vs. CD49f. CD34 provides superior resolution and separates the entire endothelial cell fraction (*orange*) and another unknown cell population (*yellow*) that lacks CD49f. *E*, same data and plot as *C*, retaining the *yellow/orange color* scheme defined in *D*, which illustrates the incomplete separation of the orange endothelial cell fraction. *F–H*, primary CD34^{Pos}CD49f^{Pos} endothelial cell cultures seeded (*F*) on 2D collagen I, (*G*) on top (O/T) of Matrigel with 5% overlay, or (*H*) with a fibroblast feeder layer, stained with CD93-PE to label endothelial cells and reveal vessel-like structures. Cells were cultured in endothelial growth medium (EGM-2, Lonza) supplemented with 10% FBS and incubated in a 5% CO2 humidified incubator at 37 °C. Scale = 100 μm (*A* and *B*); 400 μm (*F*); 200 μm (*G* and *H*).

lipid layer; (II) The CD34-positive population containing fibroblasts and vascular and lymphatic endothelial cells; (III) The Thy1-expressing fraction containing myoepithelial cells and pericytes; (IV) The CD24 OR Muc1 positive (inclusive disjunction) gate that defines the two luminal cell types; and

the final combined CD24^{Neg} AND Muc1^{Neg} exclusion gate that contains CD45^{Pos} leukocytes, erythrocytes, vascular smooth muscle cells and a newly-identified epithelial cell population. This gating strategy has been successfully implemented over a hundred times across numerous samples.
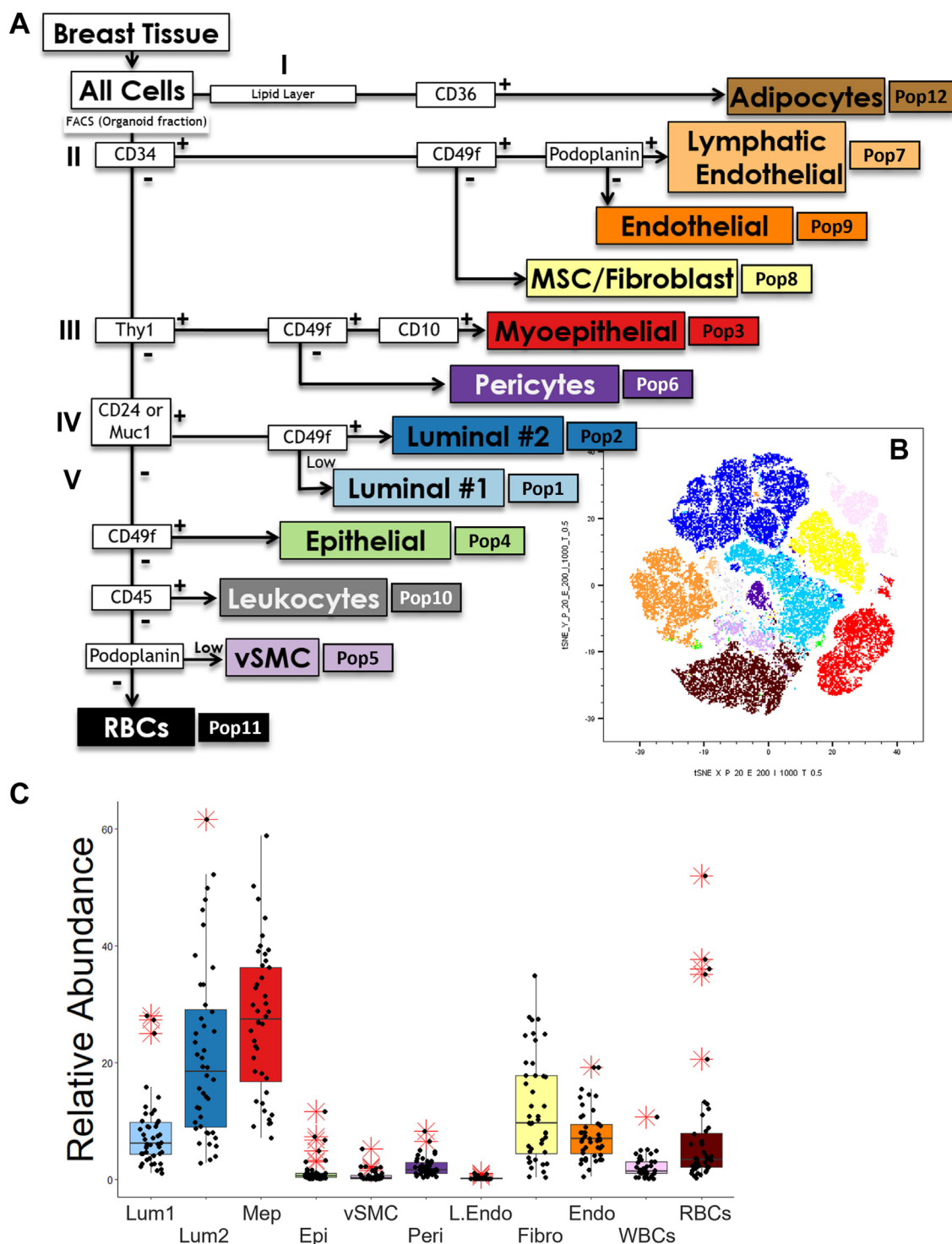
Isolation and culture of human breast cell types

Figure 9. FACS strategy for comprehensive cell type isolation. *A*, purifying the breast cell types begins with enzymatic (collagenase) dissociation of normal breast tissues. This was followed by low-speed centrifugation, yielding an organoid pellet, supernatant, and an adipocyte-enriched lipid layer. Node I) We used the lipid layer for RNA isolation/RNA-sequencing and primary adipocyte culture. Cultured adipocytes were subsequently FACS purified using CD36, a marker identified by RNA-sequencing. All remaining breast cell types were FACS-purified from the organoid cell pellet, after first dissociating the organoids to a cell suspension and staining it with CD34, CD49f, Podoplanin, Thy1, CD10, CD24, Muc1, and CD45 antibodies, along with the viability marker To-Pro-3. As most FACS machines can sort into only four tubes simultaneously, purifying every population required multiple rounds of sorting. The first round of sorting isolated cells into groups (or nodes) based on their shared expression of key markers. These nodes were: II) The CD34^{Pos} fraction; III) The Thy1^{Pos} (CD34^{Neg}) fraction; IV) The CD24^{Pos} and/or Muc^{Pos} (CD34^{Neg} Thy1^{Neg}) fraction; and V) CD34^{Neg} Thy1^{Neg} CD24^{Neg} Muc^{Neg} fraction. After the first round of sorting, two additional rounds of sorting (required to ensure purity) were used to isolate each cell type from their respective nodes, using the indicated markers and Boolean logic scheme. For example, the CD34^{Pos} cells (node II) were FACS interrogated for CD49f and podoplanin. Lymphatic endothelial cells in this fraction were identified by their co-expression of both CD49f and podoplanin; Vascular endothelial cells were identified by their expression of CD49f and lack of podoplanin; and Fibroblasts, by their lack of CD49f. The plus (+) and minus (−) signs indicate visual differences and separation of stained cell fractions—absolute 'negative' expression was not explicitly evaluated. Each FACS-purified population was assigned a number, *i.e.*, pop1-12. The identity of

Node (I) CD34 Pos
Fibroblasts
    Pop8: **CD34^{Pos} CD49f^{Neg}**
Lymphatic Endothelial Cells
    Pop7: **CD34^{Pos} CD49f^{Pos} Podoplanin^{Pos}**
Vascular Endothelial
    Pop9: **CD34^{Pos} CD49f^{Pos} Podoplanin^{Neg}**
LEC = lymphatic endothelial cell
VEC = vascular endothelial cell
F = Fibroblasts; A=adipocytes; M = myoepithelial

Node (II) Thy1 Positive
Gates: CD34^{Neg} CD45^{Neg}
Pericytes:
    Pop6; **Thy1^{Pos} CD49f^{Neg}**
Myoepithelial Cells:
    Pop3; **Thy1^{Pos} CD49f^{Pos} Podoplanin^{Pos} CD10^{Pos}**
C = capillary
P = pericyte;
M = myoepithelial cells

Node (III) CD24/Muc1 Positive
Gates: CD34^{Neg} Thy1^{Neg} CD45^{Neg}
Luminal #1 Cells:
    Pop1; **CD24^{Pos} OR Muc1^{Pos} CD49f$^{Low/Neg}$**
Luminal #2 Cells:
    Pop2; **CD24^{Pos} OR Muc1^{Pos} CD49f^{Pos}**
D = Lactiferous Duct
C = capillary; M = myoepithelial cells
L1 = Lum1
L2 = Lum2

Node (IV) CD24/Muc1 Negative
Gates: CD34^{Neg} Thy1^{Neg} CD24^{Neg} Muc1^{Neg} CD45^{Neg}
Epithelial Cells
    Pop4; CD49f^{Pos}
Leukocytes
    Pop10; **CD45^{Pos}** CD49f^{Neg}
Vascular Smooth Muscle Cells
    Pop5; **Thy1^{Neg} CD49f^{Neg} Podoplanin^{Low}**)
Red Blood Cells/Debris
    Pop11; **CD49f^{Neg} CD45^{Neg} Podoplanin^{Neg}**)
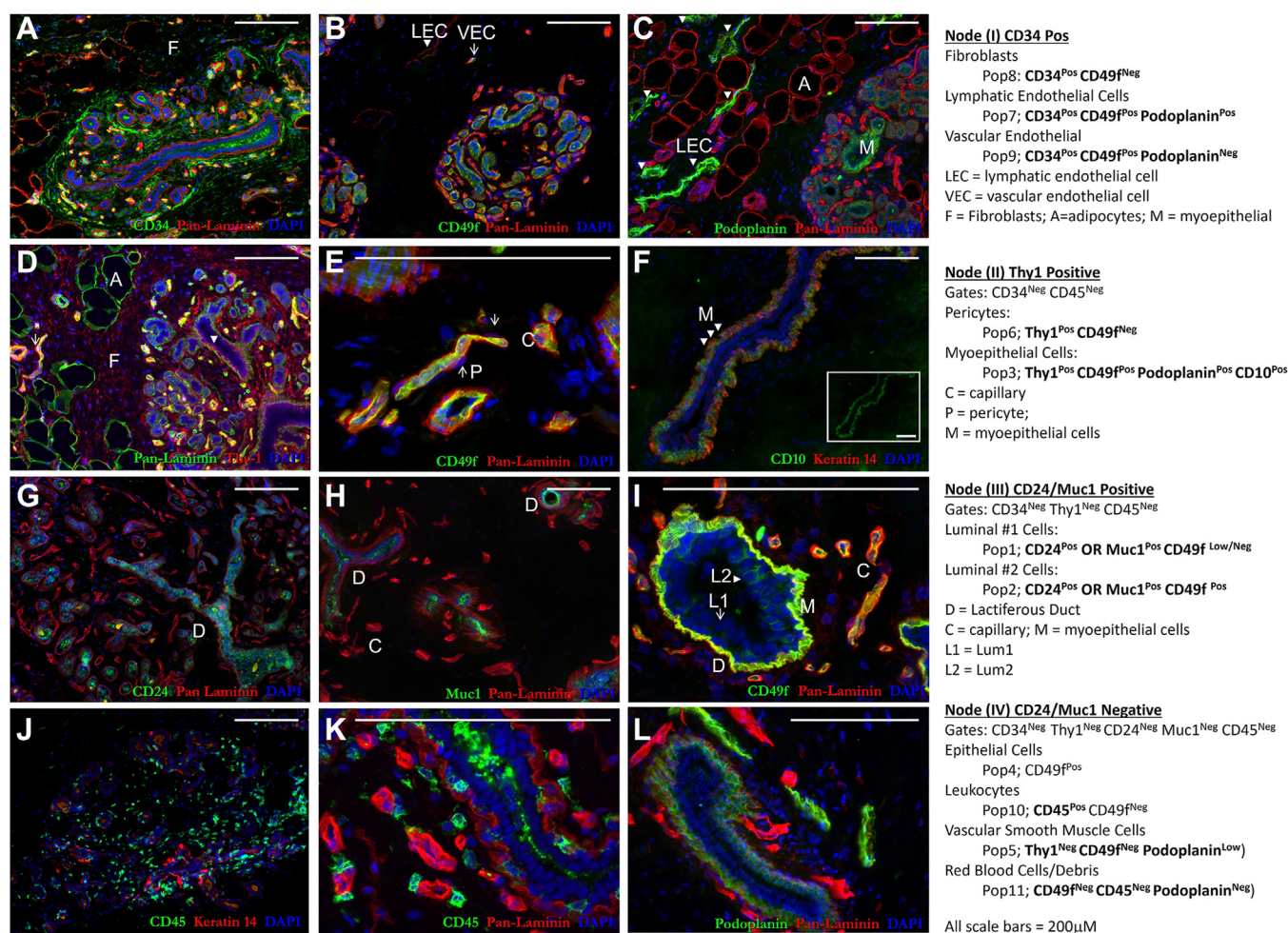
All scale bars = 200µM

Figure 10. FACS marker expression in breast tissues. Representative immunofluorescence images of tissues stained for critical markers used in the FACS isolation strategy (*A*) CD34 is expressed by breast fibroblasts (Pop8) and both endothelial cell types (Pop7&9). *B*, among the CD34 expressing cells, CD49f is expressed by vascular (Pop9, ↑) and lymphatic (Pop7, ▼) endothelial cell types, which can be distinguished by their differentially high expression of podoplanin by the lymphatic endothelial cells (*C*, ▼). Podoplanin is also expressed by myoepithelial cells and fibroblasts—expression that does not interfere with cell sorting because other distinguishing markers identify these different cell types. *D*, Thy1 is widely expressed. Among the CD34 negative cell populations, Thy1 is found in (*E*) pericytes (Pop6, ↑) and (*F*) myoepithelial cells (Pop3,▼); two cell types that are FACS separated by their differential expression of CD49f. *G*, CD24 and (*H*) Muc1 are expressed exclusively by both luminal epithelial cell types. The two luminal cell types are themselves distinguished by their differential expression of (*I*) CD49f; as luminal cells are either CD49f$^{Neg/Low}$ (Pop1, ↑) or CD49f^{Pos} (Pop2, ▼). *J* and *K*, CD45 staining reveals leukocytes (Pop10). *L*, analysis of Podoplanin (and CD49f) is used in the final FACS node to separate the remaining cell types. All staining was performed on normal breast tissues (cryosectioned reduction mammoplasty tissues).

Furthermore, we have reproduced these results on multiple FACS platforms using different fluorophore arrangements and strategies tailored to each machine's capabilities (Figs. S16 and S17).

The isolation process begins by processing tissues to generate cell suspensions suitable for staining and FACS analysis. Each cell type was isolated, identified, and validated through the following steps.

Node1: lipid layer (adipocytes)

To prepare cell suspensions for FACS analysis, we processed and collagenase-digested normal breast tissues from reduction mammoplasties, as described previously (5, 27). Centrifugation of these digests separated an oily adipocyte-containing surface layer from the remaining organoid fraction. We collected these lipid layers, evaluated their cellular content by fluorescent microscopy (staining a small aliquot with Hoechst 33,342), isolated total RNA from each oil fraction, and archived the RNA for later sequencing. Furthermore, we created primary adipocyte cultures by seeding small portions of the oil (100–300 µl) into collagen-coated T-25 flasks, using a ceiling culture method (28). The adherent colonies that emerged contained intracellular lipid-containing vesicles that stained positively for BODIPY (Fig. 11, *A* and *B*). Furthermore,

each of these populations was affirmed or later discovered by using the literature and prior knowledge or validation experiments, *e.g.*, primary cell cultures, immunofluorescence staining, and RNA-sequencing. *B*, an unsupervised tSNE projection of normal breast cells (stained and analyzed by FACS) segregated the cells into discrete populations. Each cell was overlaid with the color associated with its respective FACS gate, collectively providing a visual representation of each cell type's relative proportion in the population. *C*, relative abundance of each cell type from 30 different individuals (44 independent FACS experiments); Outliers (*) are defined as points falling (1.5 × IQR) above/below the 75th and 25th quartile.
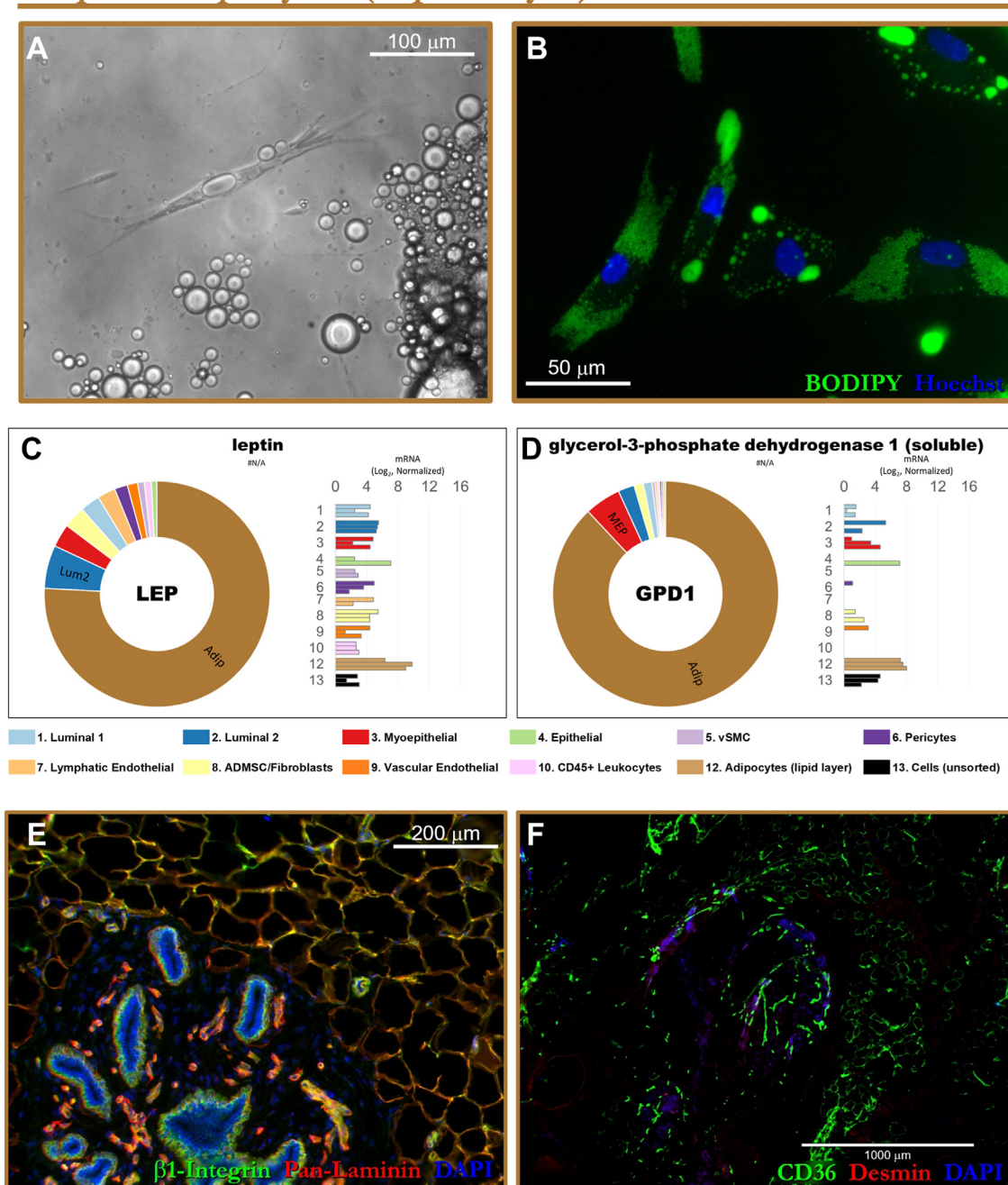
Pop12: Adipocytes (Lipid Layer)

Figure 11. Isolation and validation of adipocytes. *A*, adipocytes (Pop12) were isolated from the oil layer derived from collagenase digested tissues. Roughly 250 μl of oil was injected into tissue culture flasks (filled to the brim with medium) and cultured upside-down for 1 to 2 weeks—known as a 'ceiling' culture. Adipocytes floated to the top and attached to the collagen-I coated surface; After 1 to 2 weeks, most of the medium was removed, the dish flipped over, and the cells propagated as a typical adherent cell line. *A*, phase-contrast image of lipid vesicles within of the primary adipocytes. *B*, nonpolar BODIPY (493/503) staining of lipid droplets within the primary adipocyte cultures. *C* and *D*, RNA from uncultured adipocytes (oil layer from tissue digests) was extracted and subjected to RNA-sequencing. Among the genes uniquely or differentially expressed by these adipocyte fractions included: (*C*) the adipocyte hormone leptin (*LEP*) and (*D*) glycerol-3-phosphate dehydrogenase 1 (*GPD1*)— an enzyme critical to triglyceride synthesis. Normalized mRNA values (rlog, DESeq2) are provided on a log2 scale (horizontal bar graph showing each biological replicate) and linear scale (donut graph of median values), each color-coded by cell type. mRNA from unsorted cells served as a comparator control (Pop13, black bars). *E* and *F*, Adipocytes within breast tissues are revealed by staining tissues for (*E*) β1-integrin (CD29), (*E* and *F*) Pan-laminin, or (*F*) CD36.

transcriptome analyses identified CD36 as a marker differentially expressed by adipocytes that we found could be applied to purify cultured adipocytes from potentially contaminating fibroblasts (Figs. S30, *A*−*C* and 11, *E* and *F*). Genes we found uniquely expressed in the lipid layer included leptin (*LEP*) and glycerol-3-phosphate-dehydrogenase 1 (*GPD1*)—a key enzyme required for re-esterification of fatty acids to form tri-acylglycerol—that, along with other genes, demonstrated the

oil layer was indeed highly enriched with adipocytes (Fig. 11, *C* and *D*).

To isolate the remaining breast cell types by FACS, we dissociated the organoid fractions obtained from the collagenase digestions using trypsin. Antibodies specific to cell surface markers were added to these cell suspensions, which were then analyzed using an advanced flow sorter. Any viable, unused organoids were cryopreserved in liquid nitrogen to allow for repeated FACS analyses. Before collagenase digestion, we thoroughly rinsed the tissues with phosphate-buffered saline (PBS) to remove as much visible blood as possible. We did not use hypotonic RBC-lysis reagents on cell preparations because we were unsure if the osmotic stress would negatively impact other cell types or trigger significant changes in gene expression. Furthermore, we did not apply any method that would alter the proportion of cell types in these cell suspensions, *for example*, by removing cell lineages with magnetic beads or using any other method, as our goal was to identify every cell type in the tissue. Finally, we kept the cells chilled throughout the entire process, except during collagenase digestion, and we used the viability marker (To-Pro-3) during the sorting to exclude compromised cells. Essentially, every cell in these samples was analyzed and binned into 11 FACS-defined cell populations (the adipocytes from the lipid layer bring the total to 12 identified cell types). The first marker we used in our FACS gating scheme was CD34 (Fig. 9).

Node2: CD34 (vascular and lymphatic endothelial cells and fibroblasts)

CD34 was chosen as a central marker in our sorting strategy because of its pronounced ability, when combined with CD49f, to resolve endothelial cells, as introduced in Figure 8. However, when we stained and analyzed breast cells by FACS, we noticed CD34 resolved yet another cell fraction on 2D scatterplots (Figs. 8D and 12A, yellow box). This unknown cloud of cells lacked CD49f. Because its identity was a mystery, we named it "Population 8," or "Pop8" for short.

During the development of our FACS strategy, we assigned numerical designations to each FACS gate to facilitate sample labeling and tracking at the benchside. This practice, exemplified by Pop8, often preceded our full understanding of the gated population's identity. The numbering system originated with the epithelial types (Pops1-3) and expanded from there. We have retained these shorthand notations for brevity and clarity, specifically referring to the cell fractions defined by the FACS gating scheme outlined in Figure 9.

To establish the identity of "Pop8" (CD34^{Pos} CD49f^{Neg}) cells, we co-stained them in subsequent flow cytometry experiments using a collection of antibodies we had assembled while developing our FACS panel. Through this screen, we found Pop8 cells uniformly expressed CD73, Thy1 (CD90), and moderate levels of CD10 (Fig. 12, B–D). Staining by CD10 and Thy1 was surprising because we anticipated the expression of these markers would be limited to myoepithelial cells, as they are frequently used for this purpose (29–31). Yet, Pop8 cells were clearly not myoepithelial cells. When seeded into culture, we found these cells were elongated and had a mesenchymal spindle shape (Fig. 12, E and F). Immunostaining of these cultures demonstrated the cells indeed stained positively for CD10 and Thy1, along with Vimentin, Twist, CD44, CD105, and alpha-1 type I collagen. However, they did not express CD31 or CD36 (Figs. 12, G–N, and S32–S34). These data collectively indicated that Pop8 cells were fibroblasts. When we examined marker expression in breast tissues, we found fibroblasts indeed expressed the pattern of markers defining the pop8 FACS gate (*i.e.*, they were CD34^{Pos} and CD49f^{Neg}; Figs. S35, A and B and S36A). We also observed stromal fibroblasts in the tissues stained faintly for CD10 (Fig. S35B). We were initially surprised by this finding, as we had overlooked this weak fibroblast staining in the past—likely because intense staining of myoepithelial cells overshadows it. Previous reports have indeed described CD10 expression in fibroblasts (within breast tumors) (32), and we should have expected this, as CD10 is an established marker of bone-marrow-derived mesenchymal stem cells (MSC) (33). Interestingly, we discovered Pop8 shares more than just CD10 expression with MSCs.

While developing our FACS panel, and attempting to identify Pop8 cells, we noticed that this cell fraction expressed other prominent markers frequently used to define bone marrow and adipose-derived MSCs (34–36). Along with CD10, these included CD34, CD73, Thy1, CD44, and CD105 (Fig. 12, A–D, K, and L). This unexpected connection prompted us to explore whether Pop8 fibroblasts exhibited lineage-differentiation potentials commonly ascribed to MSCs (37–39). We found they did. Primary cultures of early-passage Pop8 cells, placed under classic adipogenic or osteogenic conditions, exhibited tell-tale signs of directed differentiation, *i.e.*, deposition of calcium and lipid accumulation (Fig. 12, O–Q). Because Pop8 cells exhibited these abilities—and to be consistent with current literature—we defined Pop8 as "adipose-derived mesenchymal stem cells" (ADMSC); however, we most often refer to them as "fibroblasts."

After identifying the CD34-expressing CD49f^{Neg} fraction as fibroblasts and the CD49f^{Pos} fraction as endothelial cells (Fig. 12A), it appeared that we had revealed all CD34-expressing cell types in the breast. We later discovered that this was not the case. We found that lymphatic endothelial cells also express CD34 and that we needed another marker to resolve them. Lymphatic endothelial cells have received little attention in the breast literature, especially in the context of normal breast tissues. As such, we could not locate literature on how to identify, isolate, or culture these cells from the breast. A search of the vascular literature led us to a few potential markers (40, 41), and after testing several antibodies, we discovered that anti-podoplanin worked well. Other antibodies, such as those specific to lyve1 and vegfr3, indeed stained lymphatic endothelium when used to stain tissue sections, but the staining was either not specific or did not produce sufficient resolution in FACS experiments (Figs. 13, A–E, S37, and S38). Podoplanin staining, in contrast, resolved the entire lymphatic endothelial cell fraction (Fig. 13E). We labeled this CD34^{Pos} CD49f^{Pos} podoplanin^{Pos} cell fraction

SASBMB

J. Biol. Chem. (2024) 300(10) 107637 **11**
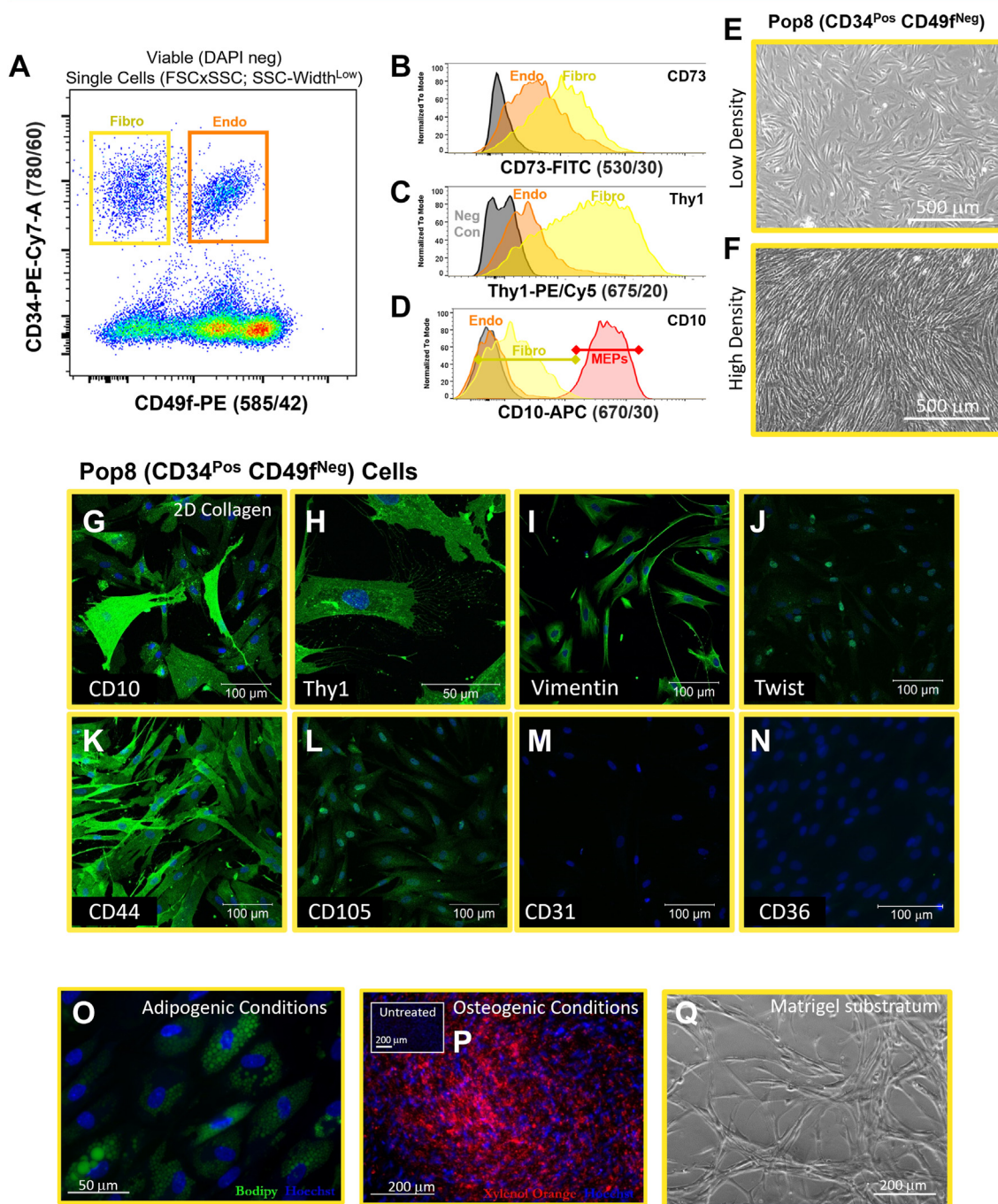
Pop8: Fibroblasts (CD34+ Gate)

Figure 12. Identification and validation of fibroblasts. *A*, flow cytometry analysis of normal breast cells, back gated for viable single cells. A discreet cell population (Pop8, later discovered to be fibroblasts), was discovered in the CD34[Pos] FACS fraction, due to its lack of CD49f staining (*yellow box*), which distinguished it from CD49f[Pos] endothelial cells (Pop9, *orange box*). *B–D*, discovering the identity of Pop8 cells was aided by recognizing this population also stained positive for (*B*) CD73, (*c*) Thy1, and (*d*) CD10, markers frequently used to isolate mesenchymal stem cells. Each histogram overlay displays gated Pop8 fibroblasts ('Fibro', *yellow*), Pop9 endothelial cells ('Endo', *orange*), unstained cells ('neg con', *gray*), and in (*D*) Pop3 myoepithelial cells ('MEPs', *red*). *E and F*, primary Pop8 cultures, imaged at (*E*) low and (*F*) high densities, display a characteristic mesenchymal phenotype. *G–N*, Primary Pop8 (CD34[Pos]CD49f[Neg]) cultures immunostained for CD10, Thy1, vimentin, twist, CD44, CD105, CD31, and CD36. O-Q, Pop8 fibroblasts cultured under (*O*) adipogenic or (*P*) osteogenic conditions; or (*Q*) on top of Matrigel substratum.

"Pop7" and called the remaining podoplanin [Neg] vascular endothelial population "Pop9". When we triple-sorted and seeded Pop7 cells into culture, these rare cells produced uniform colonies with a classical endothelial morphology that retained their differential expression of podoplanin and cultures also differentially stained for NG2 (Figs. 13, *F–H* and S39, *A–F*). We confirmed the two FACS-isolated endothelial fractions (Pop7 and Pop9) by sequencing RNAs extracted from

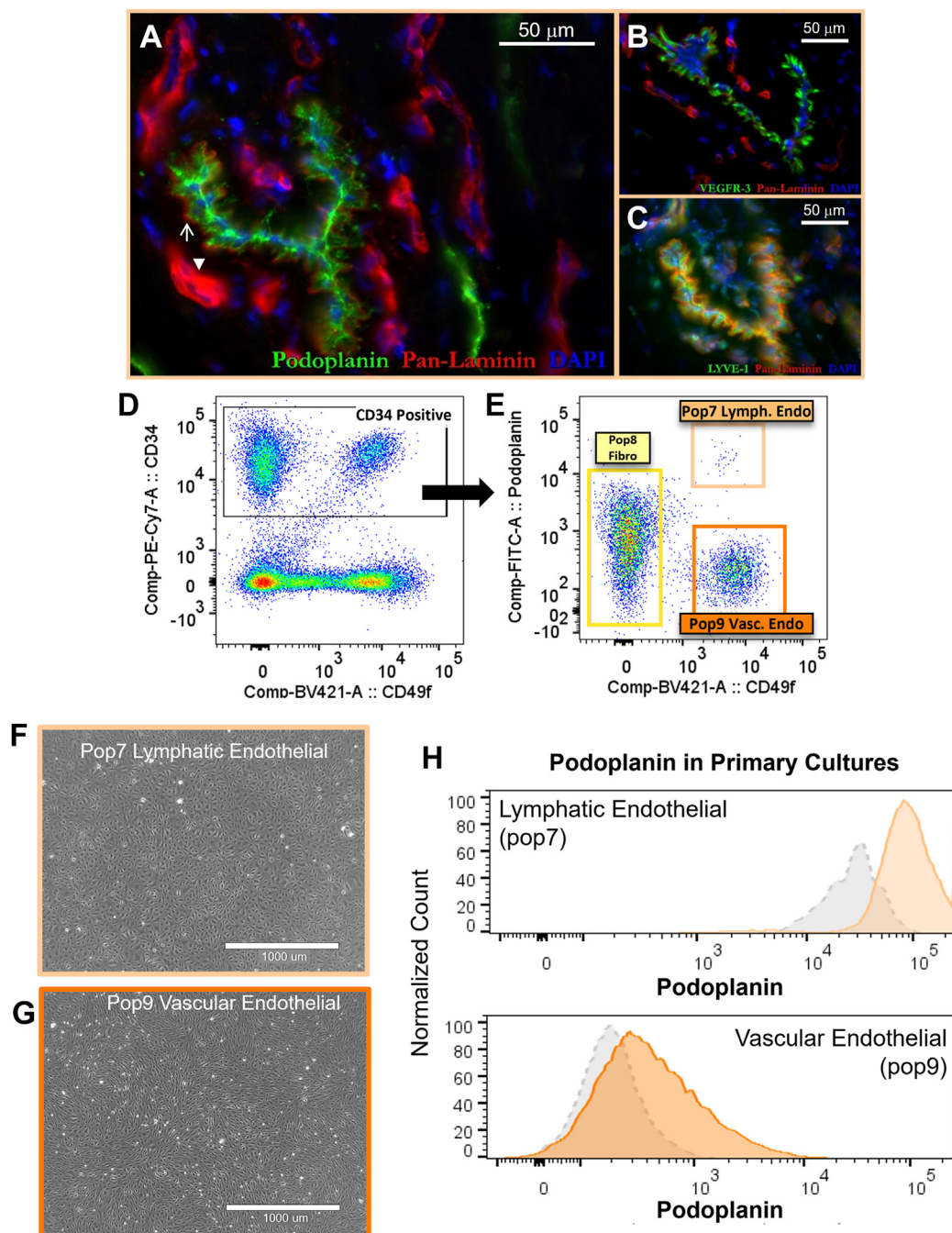## Pops 7&9: Lymphatic and Vascular Endothelial Cells (CD34+ Gate)

Figure 13. Identification and validation of lymphatic endothelial cells. *A–C*, normal breast tissue sections (adjacent, serial sections) immunostained for established lymphatic markers: (*A*) podoplanin, (*B*) vascular endothelial growth factor receptor 3 (VEGFR-3), and (*C*) lymphatic vessel endothelial hyaluronan receptor 1 (LYVE-1). Lymphatic vessels expressed minimal levels of laminin, as indicated by faint pan-laminin staining (↑), which contrasted with intensely stained vascular capillaries (▼). *D*, flow cytometry analysis (scatterplot) of normal breast cells, backgated for viable single cells. CD34 resolved two groups of cells: pop8 (CD49f^{Neg}) fibroblasts and CD49f^{Pos} endothelial cells. *E*, high podoplanin expression in the CD34-gated fraction distinguishes Pop7 lymphatic endothelial cells (*light orange box*) from Pop9 vascular endothelial cells (*dark orange box*). Pop8 fibroblasts are resolved by their lack of CD49f expression, but express low to moderate levels of podoplanin which does not interfere with FACS purification (*yellow box*). F and G, Primary cultures of (*F*) Pop7 lymphatic and (*G*) Pop9 vascular cells both display a classic endothelial phenotype and are themselves indistinguishable by phase microscopy (4× objective; scale = 1000 μm). *H*, FACS histograms of passage 3 Pop7 (lymphatic, *top*) and Pop9 (vascular, *bottom*) endothelial cells, stained with podoplanin. Each trace is contrasted with the podoplanin levels (*gray trace*) measured when the cells were originally sorted 2 months prior (55 days). These data are of isogenic cells sorted from normal breast tissue of a 20-year-old female (sample #N218).

freshly sorted (uncultured) cells. We found that both populations expressed the gene encoding the endothelial marker CD31 (*PECAM1*) and that the two cell types differentially

expressed podoplanin (*PDPN*), *LYVE1*, endothelial-associated Selectin-E (*SELE*), among others (Fig. S40). These data confirmed the different vascular origins of Pop7 (lymphatic)

and Pop9 (vascular) and validated our FACS isolation strategy. Based on the above tissue and cell staining results and transcriptome data collected from sorted cells, we concluded that three CD34-expressing cell types exist in the breast: fibroblasts and vascular and lymphatic endothelial cells.

We have sorted breast cells from over thirty individuals using our sorting scheme–some many times over (using liquid nitrogen archived specimens for multiple repeat analyses). The CD34^{Pos} cell fraction accounts for roughly 7% of all cells in these breast samples, but this proportion varied widely between individuals, ranging between <1 to 19% (Fig. 9*C* and Table S3). The different amount of red blood cells (RBCs) that remain after rinsing the cells likely contributes to some of this variability, as RBCs are included in the entire FACS cell count. We attempted to remove as much external blood as possible by rinsing samples with PBS, but the RBC fraction still ranged between 0 to 52% of all cells (which also includes debris particles; 3 ± 11%, median±sdev). Due to the indefinite nature of negative gates, it is generally more useful to consider cell proportions within defined gates—instead of calculating proportions from all "detected events", a.k.a. "total cells." For example, fibroblasts compose most of the CD34^{Pos} gate (51 ± 21%), followed closely by vascular endothelial cells (43 21%). Lymphatic endothelial cells were in the minority and composed less than 1% of all CD34^{Pos} cells (0.46 ± 1.2%), or between 0.01% and 1.02% of all cells in the tissue, which establishes it as the least abundant cell type in the breast. The isolation and culture of lymphatic endothelial cells from breast tissues have not previously been reported. Here, we describe how to identify, isolate, and culture lymphatic endothelial cells from breast tissues.

To resolve the remaining breast cell types, *that is,* those that do not express CD34, we focused our attention on myoepithelial cells and capitalized on their differential expression of Thy1, which we established as the next marker in our gating scheme (Fig. 9).

Node3: Thy1 (myoepithelial cells and pericytes)

Myoepithelial cells form the epithelial compartment's basal layer and reside between a laminin-rich basement membrane and a sheet of luminal epithelial cells (Figs. 3, 5*D*, and 6). Intracellular markers frequently used to identify myoepithelial cells (MEPs) in breast tissues include alpha-smooth muscle actin, keratin 14, and the transcription factor p63 (Figs. 14, *A* and *B*, S2, S3, and S41, *A* and *B*). Surface markers include Thy1 and CD10 (19, 27, 29, 31), which we incorporated into our staining panel to provide redundancy and ensure proper myoepithelial cell identification (Figs. 9, S21, and S23). However, when we FACS-analyzed Thy1 staining within the CD34-null fraction (Fig. 14, *C* and *D*), we surprisingly discovered that Thy1 resolved two populations, which were distinguished by their different levels of CD49f expression in FACS scatterplots (Fig. 14, *D*–*F*). While the CD49f-null population's identity was a complete mystery, we were confident the CD49f^{Pos} cells, which we designated Pop3, were myoepithelial cells—as high CD49f expression is one of their characteristic features (*e.g.,*

Figs. 4, 10, *B* and *I*, and S14). Furthermore, gated Pop3 cells also expressed significant levels of podoplanin and CD10, which corresponded to myoepithelial staining observed in tissue immunostains (Figs. 14, *F* and *G*, 10, *F* and *L*, S23, and S29). Notably, the CD49f-null fraction, which we named Pop6, did not express either of these two markers, indicating Pop6 was a discreet cell population (Figs. 14, *F*, *G*, and *I*, S43*A*, S20, and S23). Although Thy1 resolved both Pop3 and Pop6 in our FACS strategy, it is important to emphasize that Thy1 was also expressed by Pop8 fibroblasts and, to a lesser degree, by Pop9 endothelial cells (Fig. 14*H*). However, these cell types are removed by the preceding CD34 gate, which is why we placed CD34 analysis before Thy1 in our gating strategy.

To validate the myoepithelial identity and purity of Pop3 gated cells, we seeded these cells into the culture and found, unsurprisingly, that they indeed exhibited the tell-tale cobblestone morphology of myoepithelial cells; Fig. S43, *A* and *B*). These cultures stained strongly and uniformly for myoepithelial markers p63 and keratin 14 (Figs. S43, *C* and *D* and S44) consistent with RNA-seq measurements performed on freshly-sorted uncultured Pop3 cells (Fig. S43, *E* and *F*). Myoepithelial cells are thus defined in our sorting strategy by eight markers, with their full designation being: CD34^{Neg} Thy1^{Pos} CD24^{Neg} Muc1^{Neg} CD45^{Neg} CD49f ^{Pos} Podoplanin^{Pos} CD10^{Pos}. After solidifying the myoepithelial cells' isolation strategy, we turned our attention to the mysterious CD49f-null (Pop6) fraction that we had previously uncovered (the CD49f^{Neg} population in the Thy1^{Pos} gate, Fig. 14*F*).

The identity of Pop6 cells proved to be much more challenging to establish, more so than any other population we had so far encountered. We discovered this population somewhat serendipitously, so we were curious and determined to learn more about these cells. Representing less than 2% of all cells in the breast (1.6 ± 2%, median±sdev), the scarcity of Pop6 cells contributed to our difficulties in identifying them. When we initially seeded Pop6 cells into the culture, they did not attach or grow well, if at all. This was frustrating, yet itself revealing, as it contrasted with the ease and success we had in culturing fibroblasts, myoepithelial cells, and the other cell types to date. It did, however, demonstrate that Pop6 cells were different from the others. After testing different substrata and media, we eventually managed to get Pop6 cells to grow—something we now do routinely. As with every other cell type we cultured, we triple-sorted the cell populations to avoid potential contamination. Sorting multiple cycles was time-consuming yet essential to ensure we had pure populations, and it was especially crucial here as any contaminating cell type could easily overtake this slow-growing culture.

Once Pop6 cells had attached to the culture dishes and begun to grow, we discovered they exhibited a spindle-like mesenchymal phenotype that looked very much like fibroblasts. (Fig. 15*A*). However, there were noticeable differences between pop6 cells and pop8 fibroblasts. First, Pop6 cells did not express podoplanin or CD10 like fibroblasts, and, in culture, they required more aggressive trypsinization and significant tapping of the culture dish to get them to detach. Pop6 cells also possessed a bi- and tri-polar phenotype and had a

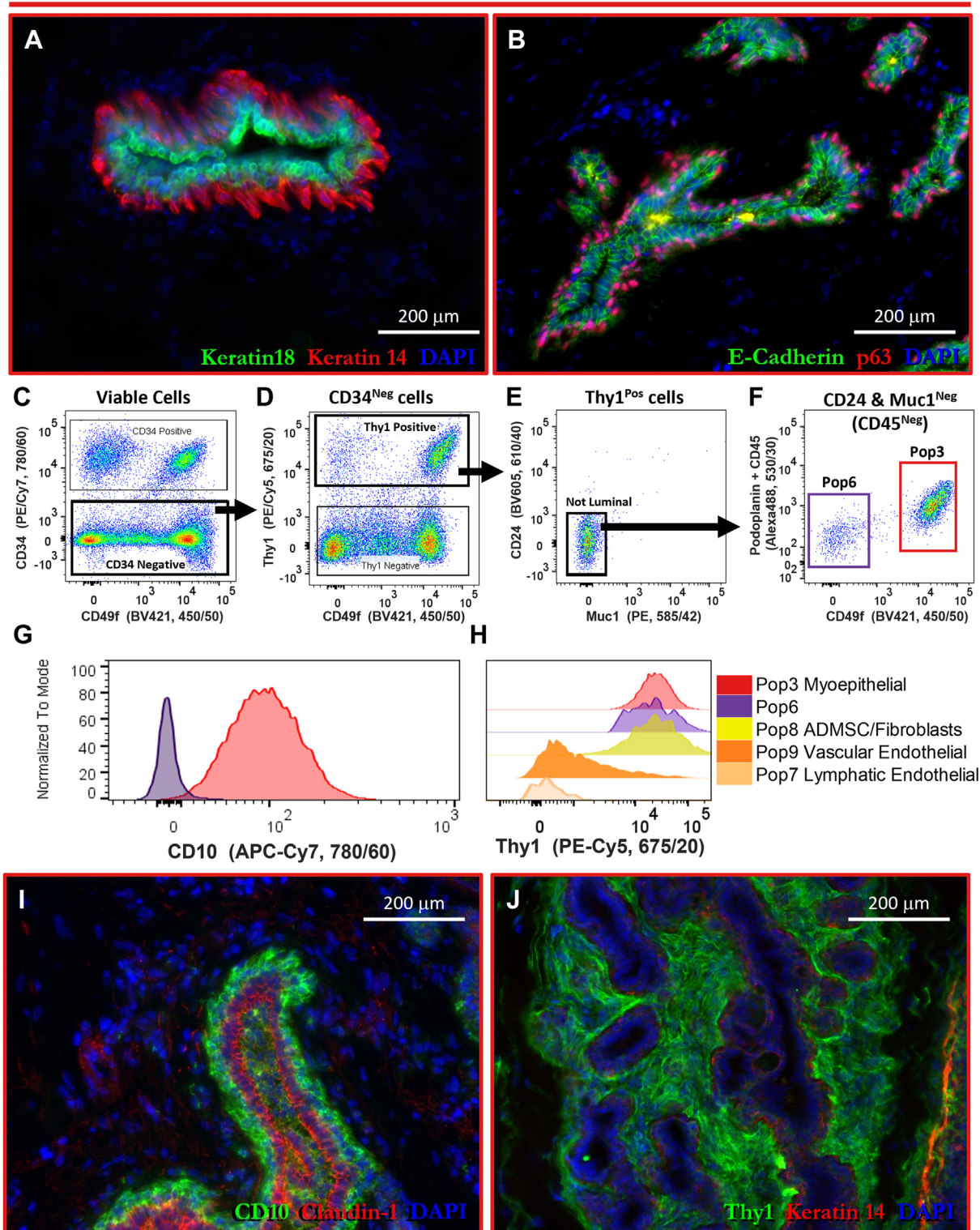# Pop3: Myoepithelial Cells (Thy1+ Gate)

Figure 14. Identification and validation of myoepithelial cells. *A* and *B*, normal breast tissues immunostained for myoepithelial-associated proteins (*A*) keratin 14 and (*B*) p63, counterstained with keratin 18 and E-cadherin, respectively. *C–F*, FACS gating strategy for isolating myoepithelial cells: *C*, CD34 Negative cells are gated and analyzed for (*D*) Thy1 expression. *E*, the CD34^{Neg}Thy1^{Pos} cell fraction was refined by excluding any cell staining positively for CD24 or Muc1 luminal markers. *F*, the resulting CD34^{Neg} Thy1^{Pos} CD24^{Neg} Muc1^{Neg} cell fraction was then interrogated for CD49f and podoplanin (along with CD45 when using the 8-channel FACS strategy depicted in Fig. S16). Intense CD49f staining distinguishes Pop3 myoepithelial cells (*red box*, *F*) from a CD49f^{Neg} cell fraction (Pop6; purple box, f—later identified to be pericytes). *G*, CD10, included as a redundant and confirmatory marker for myoepithelial cell isolation, was indeed differentially expressed between Pop3 myoepithelial and Pop6 cell fractions. *H*, although Thy1 expression is a defining feature of Pop6 and Pop3 myoepithelial cells, it is expressed also by Pop8 fibroblasts and, to a lesser extent, by Pop9 vascular endothelial cells (both of which are removed by the preceding CD34 gate). *I* and *J*, normal breast tissue immunostained with FACS panel markers (*I*) CD10 and (*J*) Thy1, counterstained with claudin-1 and keratin 14 staining.

Pop6: Pericytes (Thy1+ Gate)
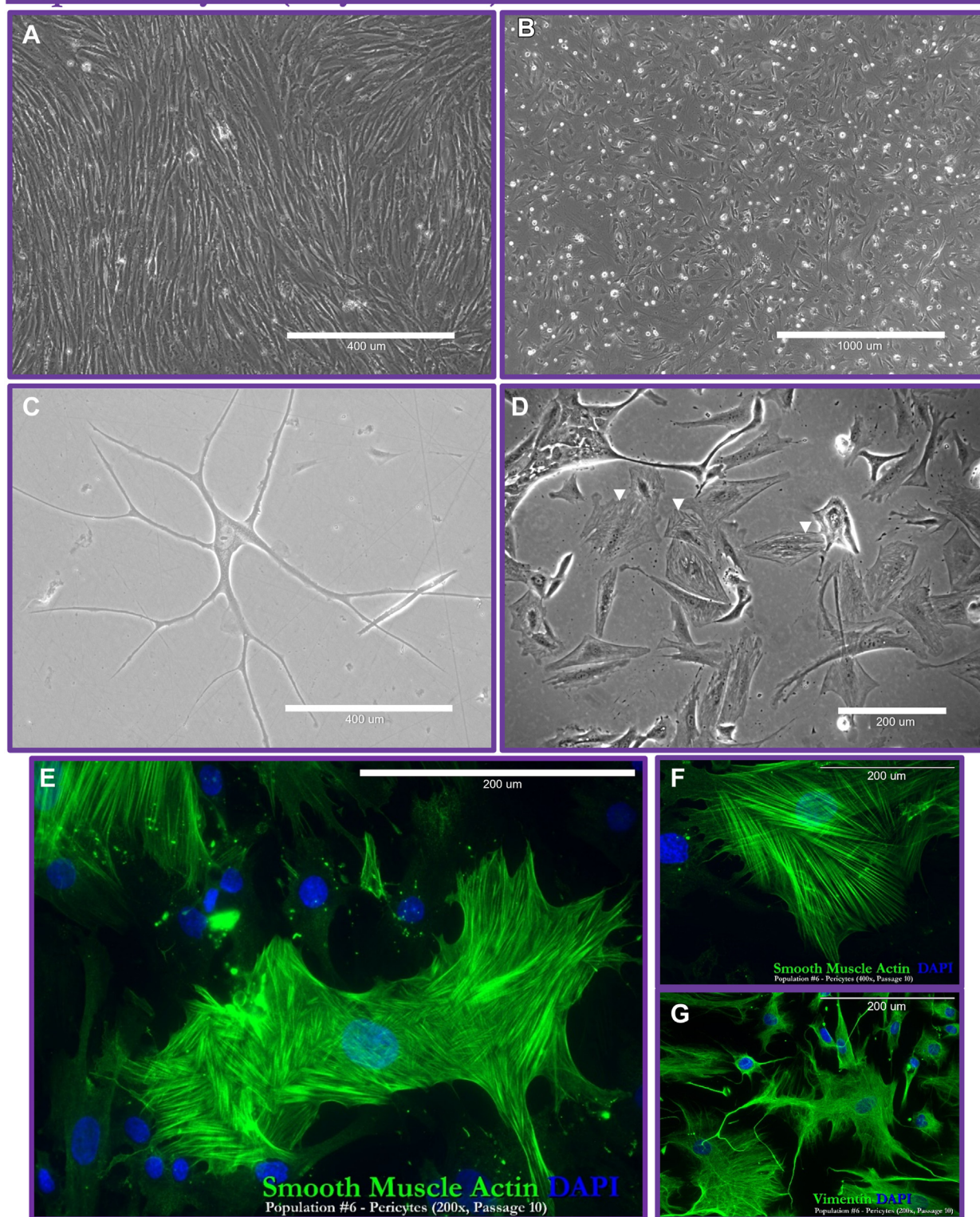
Figure 15. Identification and validation of pericytes I. *A–D*, phase contrast images of primary Pop6 cultures (later identified as pericytes). Prominent stress fibers were often observed, *e.g.*, (d, '▼'). Pop6 cells were FACS-purified from normal breast tissues surgically excised from (*A* and *B*) a 23-year-old, sample #N239; (*C*) a 37-year-old, sample #N274; a (*D*) 22-year-old, sample# N141 female, and a (*E–G*) 22-year-old female, sample# N255 (all reduction mammoplasties). *E–G*, primary pop6 pericyte cultures immunostained for (*E* and *F*) smooth muscle actin and (*G*) vimentin. Scale bars = 200 μm (*D–G*), 400 μm (*A* and *C*), and 1 mm (*B*).

broader and more pronounced cell body than fibroblasts, especially at low cell densities (Figs. 15B and S45, A and B). Pop6 cells also occasionally exhibited a unique astrocyte-like stellate morphology (Figs. 15C and S45, C–E) and would present prominent stress-actin fibers (Fig. 15, D–G), all of which were entirely distinct from Pop8 fibroblast cultures grown in parallel. We were perplexed.

Searching the literature for clues, we identified a 1985 article from Herman and D'Amore (42) that described bovine retinal pericytes, and these cells shared a striking resemblance to our own primary Pop6 cultures (Fig. 15, E and F). This connection guided us to the solution. We returned to our breast tissue staining to determine if periendothelial cells embedded within the basement membrane of blood microvessels expressed the same pattern of markers used to FACS-isolate Pop6 cells. That is, did pericytes in the tissue lack CD34 and CD49f expression? Moreover, did they express Thy1? Remarkably we found they did, indicating Pop6 cells were likely pericytes (Figs. 16, A–D, S46, and S47).

Absolute confirmation of Pop6's pericyte-identity would not come for a couple of years until we finalized our overall gating strategy for the other cell types and sequenced and analyzed RNA isolated from each purified population. We discovered Pop6 cells indeed differentially expressed known pericyte markers that included: desmin (*DES*), *CSPG4*—the gene encoding neuron glial antigen 2 (NG2), basement membrane-associated collagen (*COL4A4*), *RGS5*, *KCNJ8*, *PDGRFB*, and others (Figs. 16E and S48, A–E). Pop6 cells also uniquely expressed serotonin receptor 1F (*HTR1F*, Fig. S48F). Serotonin is a vasoactive peptide and a key modulator of vasoconstriction. The expression of *HTR1F* is notable because its unique expression by Pop6 pericytes helped us later identify Pop5, another mysterious population that uniquely expressed a related isoform.

The lack of defining markers for sorting pure pericytes has been an enduring problem that has raised doubts about the origins and identities of many primary cell models claimed to be pericytes (43). Here, we reveal a technique that overcomes this persistent and nagging problem for breast pericytes, using combinations of markers and by accounting for each cell type in the tissue. Like lymphatic endothelial cells presented above, this is to our knowledge, the first report describing how to identify, purify, and culture breast pericytes.

Although Pop6 pericytes were initially difficult to culture, we later found they would readily survive on collagen I substrate when seeded at higher densities (>5000 cells per cm^2). This requirement stood in stark contrast to fibroblasts, which, under identical conditions, could achieve similar success rates at roughly 1/30th the seeding density required by pericytes (Fig. S45F). Like adipocytes, primary pericytes in culture expressed CD36; They also retained pericyte-specific markers NG2 and did not express CD10, which distinguished them from cultures of isogenic fibroblasts grown in parallel (Figs. S45E, S49, and S50).

The Thy1 FACS node in our gating strategy thus contains two cell types: myoepithelial cells and pericytes. Cells in this node account for roughly 32% (32 ± 13%) of all cells in the tissue, ranging between 10 to 61% (Table S3). Myoepithelial cells compose the bulk of this fraction (90 ± 8%), and it is often, but not always, the most abundant cell type in the breast (27 ± 13%). By contrast, pericytes are among the least abundant cell types in the breast (1.6 ± 2%), composing only 8% ± 7% of cells in the Thy1 node on average.

Within the first three nodes in our isolation strategy (*i.e.*, the lipid layer and the CD34 and Thy1 FACS nodes), we have identified and verified six cell types that can be isolated to purity and cultured successfully. These are adipocytes, fibroblasts, vascular and lymphatic endothelial cells, myoepithelial cells, and pericytes. Continuing our quest, we set our sights on luminal epithelial cells— cells that possess a phenotype associated with the vast majority of breast cancers.

Node4: CD24 or Muc1 (ER^{Pos} and ER^{Neg} Luminal cells)

Enveloping and forming the mammary ductal system's interior luminal space are columnar and cuboidal epithelial cells that perform the gland's main secretory role (Figs. 17, A and B and S51, A and B). To accurately identify luminal epithelial cells by FACS, we discovered at an early stage in our panel design that co-staining with both CD24 and Muc1 antibodies was essential to provide the necessary redundancy and assurance that luminal cells would be properly gated and not contaminate other cell populations (Figs. 6 and 7). A common FACS strategy has been to segregate luminal cells into two populations based on their differential staining of CD49f (17–25) (Fig. 4). Although our upstream markers and gating strategy are distinct, we adhered to this approach (Fig. 17, C–F). We numbered the CD49f$^{Low/Neg}$ and CD49f^{Pos} populations "Pop1" and "Pop2", respectively, with their full antibody designations being: CD34^{Neg}Thy1^{Neg}CD45^{Neg} (CD24 or Muc1)^{Pos} and CD49f$^{Low/Neg}$ (for Pop1); or CD49f^{Pos} (for Pop2, Fig. 9). Others have designated similarly-isolated fractions as being respectively enriched for "mature luminal" and "luminal progenitor' cells— predominantly because the CD49f-null fraction has failed to produce actively growing primary cultures or outgrowths (19–21). These designations have been supported by cell staining, which showed that the CD49f-null cell population has roughly twice as many estrogen receptor alpha expressing (ER^{Pos}) cells than does the CD49f^{Pos} fraction (55 *vs.* 28%) (21). Yet, we found that both would establish growing cultures when we triple-sorted and seeded Pop1 and Pop2 luminal cells into the culture. Both primary cultures exhibited a squamous morphology distinct from the cobblestone appearance of Pop3 MEPs (Figs. 17, G and H and S43A). We attributed the successful culture of these two luminal cell types to the unique methods, stringent gates, medium, and conditions we used, which are indeed different from other reports (17, 19, 21, 23, 25, 44, 45).

Although we found we could culture both luminal populations, there was a notable difference: Compared to their Pop2 counterparts, Pop1 luminal cells required a higher seeding density to establish a growing culture. When we tested colony-forming efficiencies by FACS sorting and seeding limiting numbers of luminal cells into collagen-coated 96 well
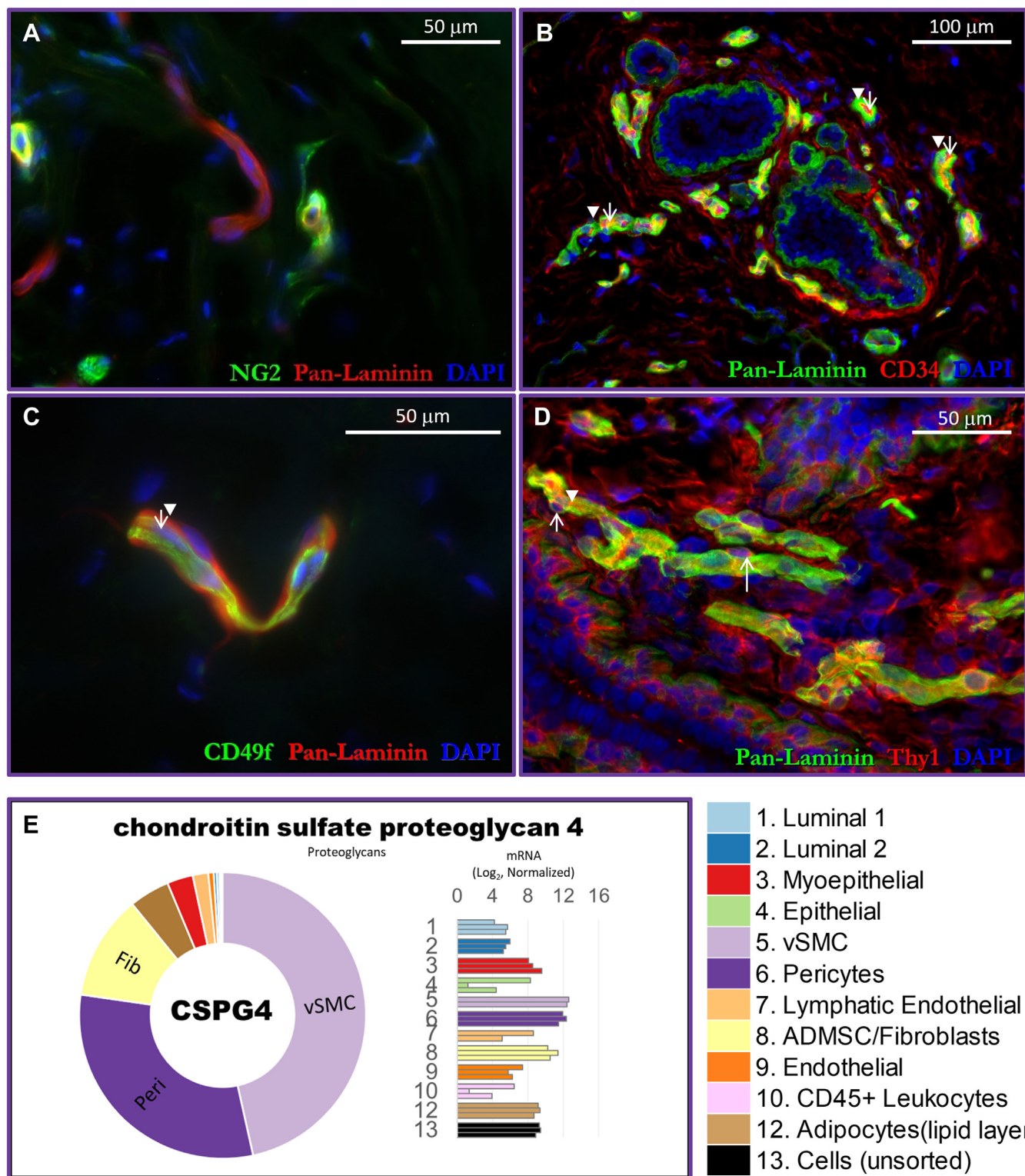
Figure 16. Identification and validation of pericytes II. *A,* Normal breast tissue immunostained for pericyte marker NG2 proteoglycan, which labels a fraction of breast pericytes. *B–D,* breast tissue immunostained for several key FACS markers defining the Pop6 gate (CD34[Neg] Thy1[Pos] CD24[Neg] Muc1[Neg] CD45[Neg] CD49f[Neg] Podoplanin[Neg]): (*B*) CD34, (*C*) CD49f, and (*D*) Thy1. Tissue staining indeed confirmed pericytes (embedded within the laminin-staining basement membrane of vessels) exhibited the pattern of staining observed by FACS. With respect to: (*B*) CD34, pericytes were CD34 negative (▼), whereas the endothelial cells were positive (↑); (*C*) CD49f, pericytes were CD49f negative (▼), whereas the endothelial cells stained positive (↑); (*D*) Thy1, pericytes stained positively for Thy1 (▼), along with the majority of endothelial cells (↑, as shown in Fig. 12*C*). *E,* transcript levels of CSPG4 (the gene encoding NG2) within uncultured FACS-purified cell types, as determined by RNA-sequencing. Pop6 pericytes and Pop5 cells (later identified as vascular smooth muscle cells) indeed expressed the highest levels of this perivascular marker (CSPG4).
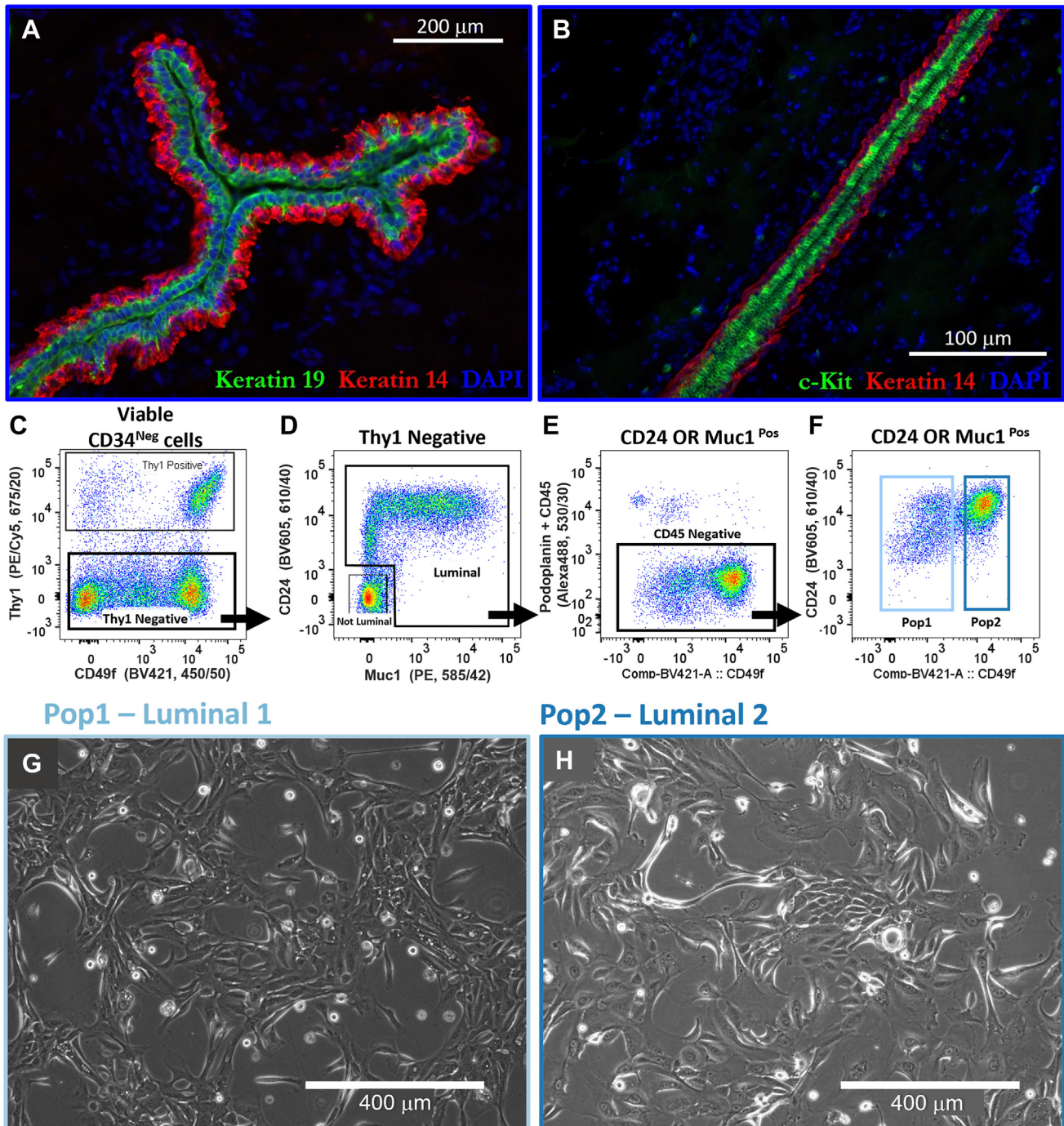
Figure 17. Identification and validation of the luminal cell types. *A* and *B,* normal breast tissues immunostained for luminal epithelial-associated proteins (*A*) keratin 19 and (*B*) c-kit, counterstained with keratin 14. *C–F,* FACS gating strategy for isolating luminal epithelial cells: (*C*) CD34 negative cells (backgated for viability and single cells) were analyzed for Thy1 expression. *D,* Luminal epithelial cells are present within the gated 'Thy1 Negative' fraction, identified by their expression of one, or both, luminal markers: CD24 and Muc1 (denoted by the inclusive disjunction operator, '∨'). *E,* the resulting CD34^{Neg} Thy1^{Neg} CD24^{Pos} ∨ Muc1^{Pos} cell fraction (on the 8-color BD FACS machine, shown here) was further refined by gating out CD45^{Pos} leukocytes and focusing on the 'CD45 Negative' cells (Note: CD45Pos cells are gated earlier when using a FACS machine with additional channels). *F,* differential CD49f expression demarcates two distinguishable luminal cell populations within the CD45^{Neg} fraction: Pop1 (CD49f$^{Neg/Low}$, *light blue box*) and Pop2 (CD49f^{Pos}, *dark blue box*). *G* and *H,* phase contrast images of (*G*) Primary Pop1 (Luminal 1) and (*H*) Pop2 (Luminal 2) cultures (both Passage 2; 21 days post-seeding). Both cultures were initially seeded at 4700 cells/cm2 and sub-cultured at 32,500 cells/cm^2 on bovine collagen I substrata, both of which (density and substrata) proved critical to the successful culture of the luminal cells. To ensure purity, the cell populations were triple sorted (using conservative gating that avoided the transitional space between pop1 and pop2, *gray dashed lines*).

Isolation and culture of human breast cell types

dishes (Fig. S52A), we found 83% of wells seeded with Pop2 cells (at the highest density) became confluent after 35 days (Fig. S52, A and B). In contrast, none of the wells seeded with Pop1 cells—at the same or lower densities (from three independent experiments)—ever produced large colonies. Although some individual Pop1 cells attached in these wells, a few produced small colonies of less than about 10 cells each, they did not exhibit the vigorous growth observed by Pop2 cells. These results seemingly confirmed prior claims that CD49f-null luminal cells lacked progenitor activity. However, we discovered that this was not the case. We also seeded every other well in these experiments with irradiated fibroblasts—to serve as a non-confluent feeder layer. When we examined these neighboring wells, we found Pop1 cells grew dramatically, with half the wells reaching confluency (*i.e.*, those seeded at the highest density; Fig. S52C). The support provided by the feeder layer extended to the lower seeding densities as well, for both Pop1 and Pop2 cells. Remarkably, 4% of wells seeded with just a single Pop2 cell produced clonal cultures. Still, the growth of Pop1 cells was striking and stood in stark contrast to the stagnant growth of those cultured without fibroblasts—and to prior claims that the CD49f-null "mature" luminal cells do not have progenitor activity (Ibid). Our results demonstrate that they do, but it was not clear if stromal support was essential.

The progressive improvement in Pop2's colony-forming efficiency, which occurred when we seeded them at higher densities (Fig. S52B), made us wonder if Pop1 cells just needed an even higher seeding density than 1500 cells/well (4700 cells/cm^2) to survive and thrive. We discovered this was indeed the case. When we seeded Pop1—and Pop2—cells at 15,000 cells/cm^2 or greater (roughly 3× higher than the highest density in prior experiments), they all produced propagating cell lines (n = 3 each). Along with the above experiments, these data provided evidence that Pop1 and Pop2 cells were functionally distinct and that both could establish *ex vivo* cell lines. Knowing where these different FACS-defined luminal cell types existed within the breast epithelium remained a vexing question, however.

The proportions of the two luminal types in our FACS-analyzed samples incidentally helped pinpoint their histological location. While we detected the two luminal populations in every breast sample (n = 32 individuals), their relative abundance within the luminal cell compartment varied widely between individuals, ranging between 3% to 91% for Pop1 and 6% to 96% for Pop2. This unpredictable imbalance (*i.e.*, the ratio of the two luminal cell populations) was notably the most variable feature among individuals—for any cell type (Fig. 18, A and B). A crucial clue to Pop1's identity was uncovered when we stained and analyzed tissues for estrogen receptors. We noticed that one sample, N277, was teeming with ER^{Pos} luminal cells. Nearly every luminal cell in this sample stained positive for ERα (Fig. 18C). This patient was notably taking TriNessa (norgestimate and ethinyl estradiol tablets) for contraceptive purposes; however, we are uncertain if this contributed to the overwhelming abundance of ER^{Pos} luminal cells in this patient's breast tissue. The staining was

nevertheless quite striking. When we examined the tissue by FACS, we noticed a similar luminal-cell imbalance: We found that it had, by far, the most CD49f$^{Low/Neg}$ (Lum1) cells than any other tissue we have ever tested, and it was indeed the sole outlier (Fig. 18B). The correlation between the abundance of ER^{Pos} and Pop1 cells in this tissue, measured respectively by tissue immunofluorescence and FACS, was glaring. In this sample, nearly every luminal cell expressed ER (determined by tissue staining) and consisted, almost entirely, of Pop1 luminal epithelial cells (determined by FACS). This connection hinted that the CD49f$^{Low/Neg}$ luminal cell population (Pop1) were, in essence, ER^{Pos} luminal cells.

To explore the strong parallel relationship between the abundance of ER^{Pos} cells and FACS-defined CD49f^{Neg}Pop1 cells, we co-stained breast tissues with both ER and CD49f antibodies. The results were clear: CD49f (alpha-6-integrin) was strongly expressed along the basal side of myoepithelial cells and nicely traced the outline of the basement membrane (Fig. 19A). The luminal cells were indeed split by those that stained for CD49f and those that did not. For those that did, expression was most intense along the cells' basolateral membrane—and these cells did not express ER. Of the luminal cells that did express ER, we found they did not express CD49f—which we verified by confocal optical sectioning (Fig. 19B and Video S1). These results showed that, between the two luminal epithelial cell types, CD49f and ER expression was mutually exclusive. Therefore, we concluded that the FACS-defined CD49f$^{Low/Neg}$ Pop1 cells are, in fact, ER^{Pos} luminal cells. Immunostaining tissues for markers differentially expressed between Pop1 and Pop2 cells (identified by RNA-seq analysis) solidified this conclusion. Staining results included the mutually coherent expression of estrogen receptor α and progesterone receptor in Pop1 cells (Fig. S54); and mutually exclusive expression between Pop1 and Pop2 of estrogen receptor α and either: keratin 15 or c-kit proto-oncogene (Figs. S55 and S56)—staining that was consistent with our RNA-sequencing results and with prior reports (46, 47).

Both luminal cell types compose, on average, about 26% of all cells in the breast, ranging between 5 and 64% (26 ± 16%), making them collectively the second most abundant cell type in the tissue—just behind myoepithelial cells (27%, Table S3). ER^{Neg} luminal cells (Pop2) were almost always the most prevalent of the two (38/44 cases), composing about 67% of all the luminal cells on average. However, as previously noted, the relative proportion of the luminal cell fractions varied widely between individuals. Why this is so is currently unknown. All FACS-analyzed samples were from assumedly premenopausal women (ages 16–40). Although we did not have the necessary data to stage each individual's position in the estrous cycle, we predict that cycling estrogen and progesterone hormones influence the cell phenotype and the balance of the two luminal cell types.

We have demonstrated in this section that CD24 and Muc1, along with CD49f, resolve two luminal cell types within the Thy1^{Neg} FACS gate. The addition of these two cell types brings the running tally to eight cell types identified so far—all of which we have successfully cultured. The remaining Thy1^{Neg}
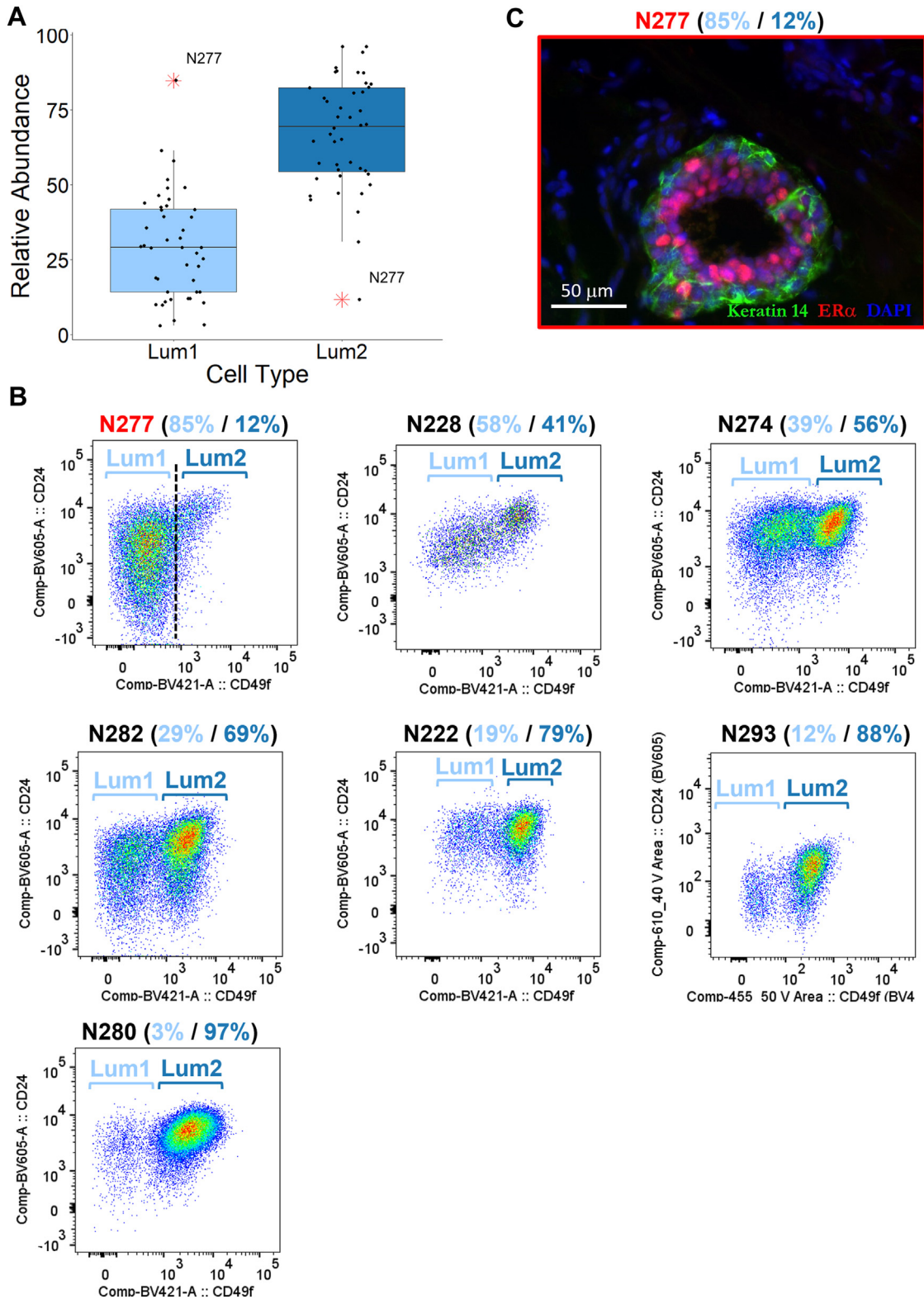
Figure 18. The relative abundance of Pop1 and Pop2 cells varies widely among individuals. *A*, relative abundance of CD49f $^{Low/Neg}$ Luminal 1 (Lum1) and CD49f^{Pos} Luminal 2 (Lum2) cells, measured by FACS, within the luminal cell compartment (all CD24/Muc1 staining breast cells). Displayed are 44 separate analyses of randomly selected normal breast tissues (reduction mammoplasty tissues) from 30 different individuals. The average age of the women was 27 ± 6 years (sdev), with ages ranging between 16 and 40 years *B*, FACS scatterplots of the luminal cell fractions, showing select samples with an uncommonly high proportion of Lum1 cells (N277, N228); nearly equivalent proportions of Lum1 and Lum2 cells (N274, N282, N222); and uncommonly high proportion of Lum2 cells (N293 and N280). *C*, immunofluorescent immunostaining of N277 breast tissue for estrogen receptor alpha indicated that the high proportion of Lum1 cells in this sample coincided with an uncommonly high proportion of ER-staining luminal cells (N277 outlined in *red*).
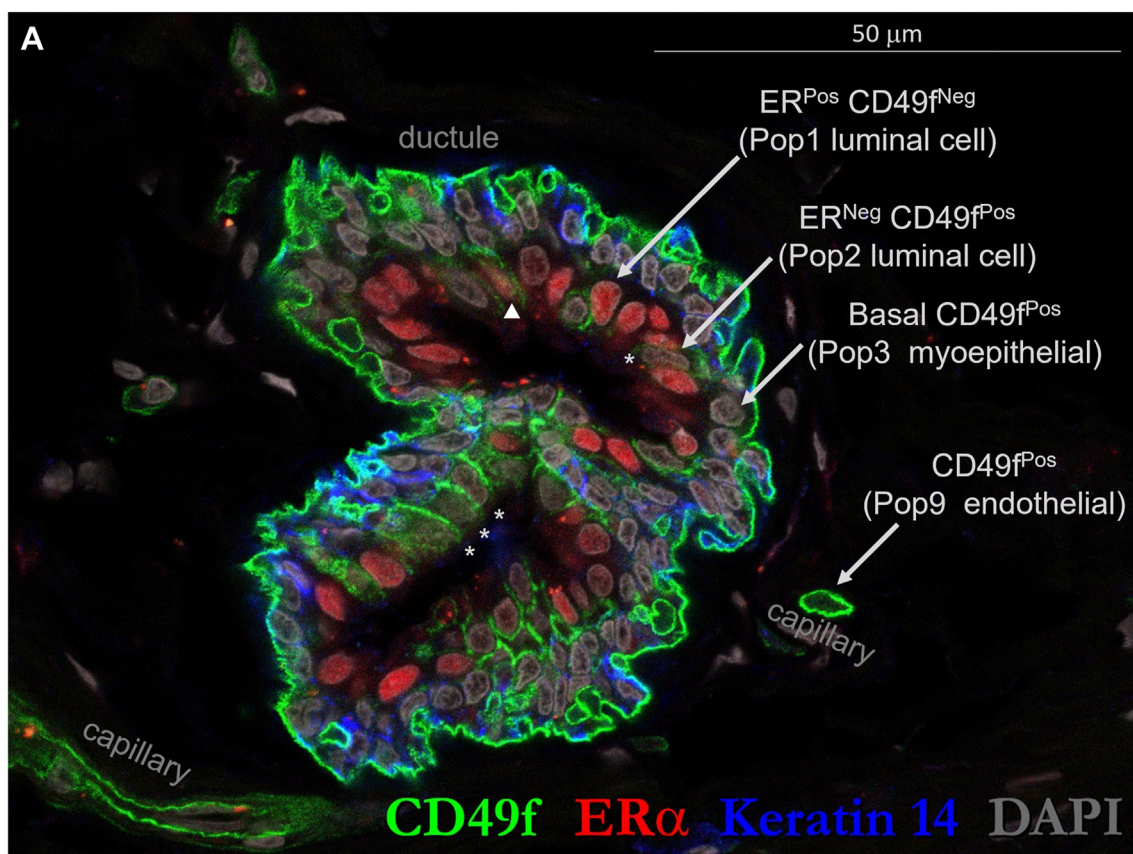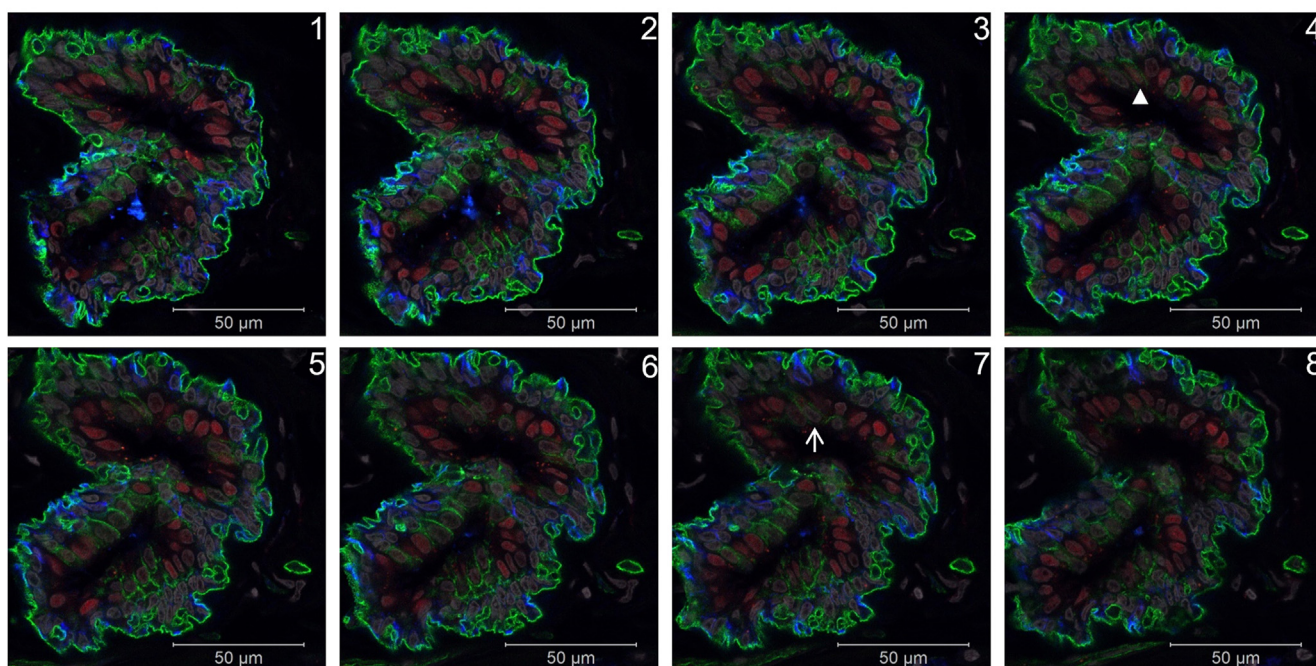
Figure 19. Pinpointing the histological location of Pop1 and Pop2 luminal cells. *A*, confocal image of normal breast tissue immunostained with ERα (*red*) and CD49f (*green*). CD49f (α6 integrin) was present within endothelial cells of small capillaries and along the basal side of myoepithelial cells, which is consistent with CD49f[Pos] staining of these cell types in FACS experiments. Tissue staining also revealed CD49f expression along the basolateral membranes within a fraction of luminal epithelial cells, which is also consistent with FACS results, as this marker defines the Pop2 luminal cell gate (CD49f[Pos]). The tissue staining also showed that these same CD49f-expressing luminal cells did not express ERα. Four such cells are labeled with an *asterisk*, '*'. Estrogen receptor was instead confined to CD49f[Neg] luminal cells that, by FACS, defines Pop1 luminal cells (CD49f[low/neg]). *B*, adjacent confocal planes (b1-8) illustrate how the cellular origin of ERα and CD49f staining is sometimes obscured by the proximity of cell types in tissues. For example, one ERα-stained luminal cell initially appeared to express CD49f (marked by '▼' in a and b4), however, upon examination of adjacent Z-stacked planes, the CD49f staining was instead traced to an adjacent ER-negative luminal epithelial cell (marked by '↑' in b6&7).

cell types, which do not express CD24 or Muc1, compose the final node that we called the "Not-Luminal" or "Null" Gate.

Node5: null gate (leukocytes, erythrocytes, vascular smooth muscle cells, and Pop4 epithelial)

Leukocytes

Of the remaining cell populations to be isolated, those we had anticipated from the outset—and had still not sorted—were the leukocytes and erythrocytes (white and red blood cells). Although the leukocytes comprise a broad range of related cell types with distinct functions, resolving each subpopulation was not our priority. To isolate the bulk leukocyte population, we relied on the frequently used pan-leukocyte marker, CD45. However, we first needed to overcome a technical issue: The other markers in our panel were already occupying each of the fluorescent detectors on our 8-color FACS machine. These were: TO-PRO-3, CD34, CD49f, Podoplanin, Thy1, CD10, Muc1, and CD24. To add CD45, we needed to double up and detect two markers in a single fluorescent channel. This strategy is possible if the two antibodies do not bind to the same cells and if the different cell types that they do bind to can be resolved in another dimension by other antibodies in the panel. After a series of tests, we elected to use the same fluorescent tag (Alexa Fluor 488) for both CD45 and podoplanin. This strategy worked well, as adding CD45 to the panel cleanly separated the leukocyte population, which we numbered Pop10 (Fig. 20A). This CD34^{Neg} Thy1^{Neg} CD24^{Neg}Muc1^{Neg} CD45^{Pos} cell fraction exhibited a unique low side scatter FACS profile typical of leukocytes (Fig. S57). After RNA-sequencing, we found Pop10 indeed differentially expressed the gene encoding the CD45 antigen, protein tyrosine phosphatase receptor (*PTPRC*, Fig. 20B), as well as many other leukocyte-specific markers (Fig. S58, *A–F*). Collectively, these data validated our sorting strategy. Of course, if a cytometer has enough fluorescent detectors to accommodate each marker, doubling up antibodies would not be required, and one could instead add CD45 to an open channel. Using this alternative approach on a different FACS machine—one with 14 available channels (SONY SY3200)—we have successfully and repeatedly replicated the above results, demonstrating the rigor of the FACS strategy when tailored to the machine (Figs. S16 and S17).

After applying the FACS strategy to our collection of breast tissues, we found that the proportion of leukocytes ranged between 0 and 11%, composing 2.1 ± 2% of all cells in the tissue on average. Leukocytes were among the least abundant cell types in the breast (Table S3). Although we did not design our sorting strategy to identify each leukocyte type, we did observe Kit^{Pos} mast cells, CD56^{Pos} natural killer cells, and others when we immunostained tissues with epithelial markers that cross-reacted with these leukocyte types, *that is*, CD117 (Kit), CD56, and CD10 (Figs. S9, S10, 14I, and 17B). When we stained tissue sections, we found leukocytes throughout the tissue—inside and outside blood vessel lumina (Fig. 20C and Video S2). Leukocytes were concentrated near epithelial structures and were especially prevalent within TDLUs

(Figs. 10J, 20, D and E, and S60A). They were most frequently found in the stroma but were also conspicuously present on the epithelial side of the basement membrane, among the luminal and myoepithelial cells, and within the lumen of ducts and ductules (Figs. 20F and S60B). Co-staining tissues for CD45 and CD68, CD3e, CD20, or Kit demonstrated the vast majority of extravascular leukocytes were CD68^{Pos} macrophages, although a small fraction of Kit^{Pos} mast cells, CD3e^{Pos} T-cells, and CD20^{Pos} B-cells were also present (Figs. S60, *C–I* and S59–S66).

To ensure each of the other FACS-sorted cell types were free of leukocytes, we added a CD45^{Neg} gate to each of their respective isolation strategies, and these gates were included in the RNA-sequenced samples (Figs. S16 and S17). We then returned to the final FACS plot in Figure 20A to investigate the remaining cell populations. As expected, red blood cells (erythrocytes) were one of these.

Erythrocytes

Erythrocytes do not express any of the markers within our staining panel. As such, they occupy the left lower quadrant of the final FACS scatter plot; *that is,* they are CD34^{Neg} Thy1^{Neg} CD24^{Neg} Muc1^{Neg} Podoplanin^{Neg} CD45^{Neg} cells (Fig. 20A). We named this null-staining cell fraction "Pop11," and found it was the only population that produced a red cell pellet after being centrifuged (Fig. 21A). Microscopic examination of Pop11 confirmed the presence of biconcave red blood cells (Fig. 21B). Small debris was often present as well, which is expected in a marker-negative gate such as this. Because debris particles, precipitates, and large salt crystals can sometimes trigger the FACS machine into falsely recognizing them as cells (FACS events), we filtered all buffers and collection media through 0.2um polyethersulfone (PES) filters prior to using them for cytometric analyses. We thoroughly rinsed the cells and strained them through 100 and 40um cell strainers—multiple times—prior to analysis, which reduced debris and false-positive events. When we seeded Pop11 into the culture, nothing grew as expected—and RNA isolations did not yield any appreciable RNA. We, therefore, concluded that Pop11 contained red blood cells and small debris particles. Additional purification using an erythrocyte-specific marker, *for example*, CD235a, was not needed as this fraction was not further used or analyzed.

Population 5: vascular smooth muscle cells

When we viewed the final FACS plot of the "Not Luminal" node (Fig. 20A) after identifying leukocytes and erythrocytes, we were surprised that two cell populations remained. The first of these, we named "Pop5." This fraction was not well-defined but was located adjacent to the marker-negative RBCs on the final FACS plot (Fig. 20A). Attempts to culture Pop5 cells were mainly unsuccessful, as single cells would sometimes attach to the dish and only a few cultures were established. Pop5's lack of growth distinguished it from the other non-blood cell types, all of which we could successfully and routinely culture. Therefore, revealing Pop5's identity
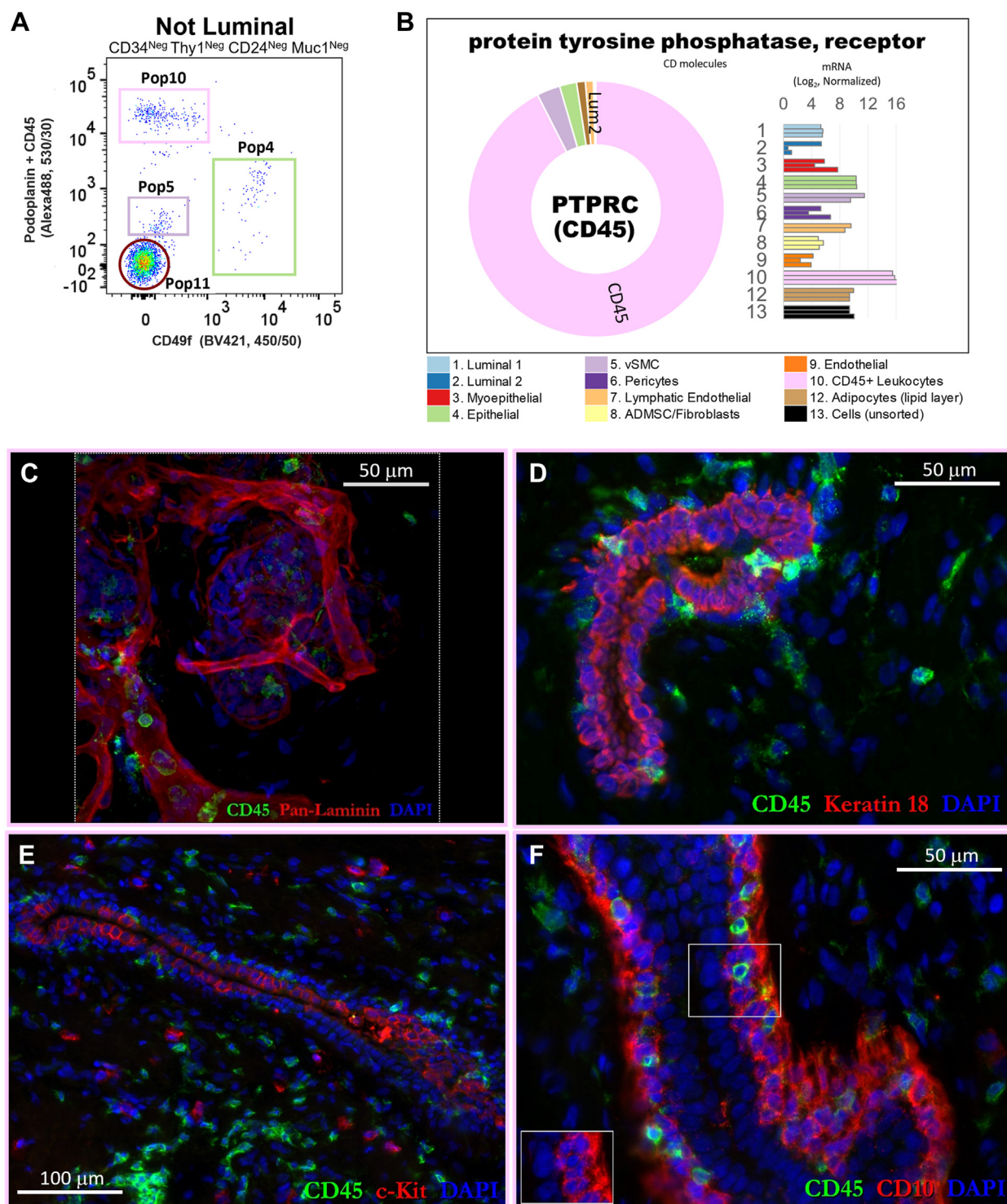
Leukocytes (CD24^Neg & Muc1^Neg Gate)

Figure 20. Identification and validation of Leukocytes. *A*, FACS gating strategy for isolating leukocytes and remaining breast cell types. CD34, Thy1, CD24, and Muc1 negative cells (also backgated for viability and single cells) were analyzed for CD45/Podoplanin and CD49f expression. CD45 and podoplanin antibodies were labeled with the same fluorophore (Alexa488) because all eight fluorescent channels on this FACS machine were in use. This analysis revealed four cell populations: Pop10 (leukocytes), Pop11 (red blood cells –and debris), and two additional, yet unknown, populations. We designated these two as Pop4 and Pop5 (which we later identified to be a rare epithelial cell type and vascular smooth muscle cells, respectively). *B*, transcript levels of *PTPRC* (the gene encoding CD45) within uncultured FACS-purified cell types, as determined by RNA-sequencing. As expected, this pan-leukocyte marker was clearly and differentially expressed by Pop10 leukocytes. *C–F*, leukocytes within breast tissues, revealed by staining for CD45 and (*C*) pan-laminin, to reveal white blood cells (WBCs) within- and outside the vasculature; (*D*) keratin 18, to reveal WBCs proximity to the epithelium; (*E*) c-kit, to reveal mast cells (and Pop2 luminal epithelial cells); and (*F*) CD10, which reveals WBCs that have transmigrated into the epithelial compartment and are

Red Blood Cells (CD24^Neg & Muc1^Neg Gate)
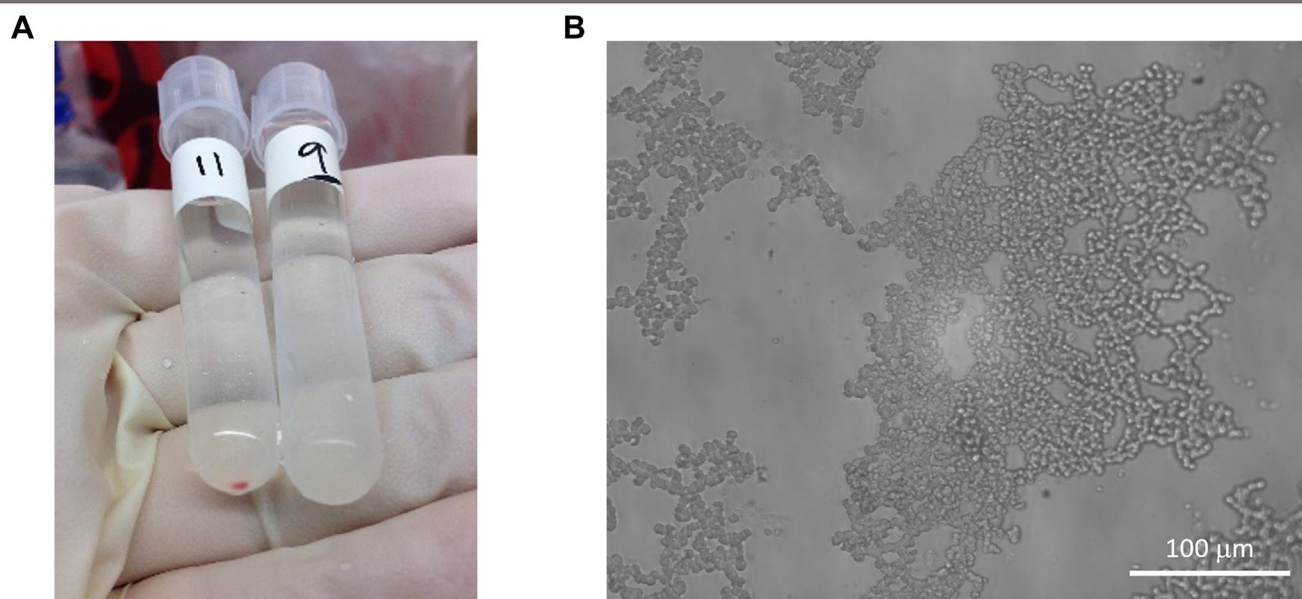
Figure 21. Identification and validation of Red Blood Cells. *A*, as illustrated in the preceding figure (Fig. 20), there was a single cell fraction that did not bind any of the eight antibodies used in our FACS panel. It was designated Pop11. This population was conspicuously the only cell fraction that produced a red cell pellet upon centrifugation, which indicated the presence of erythrocytes. *B*, phase-contrast microscopy confirmed Pop11 indeed consisted of biconcave erythrocytes and unidentifiable debris.

required RNA sequencing, and one of the earliest clues was its unique expression of 5-hydroxytryptamine (serotonin) receptor 1B (*HTR1B*). This gene was one of 13 uniquely expressed by Pop5 (at a 16× stringency level, Fig. 22*A*). Pop5's expression of *HTR1B* was notable because we recalled that Pop6 pericytes uniquely expressed two genes at this stringency, and one was *HTR1F*—an isoform of *HTR1B* (Fig. S48). Although circumstantial at the time, this association was our first clue that Pop5 and Pop6 cells were related, causing us to suspect Pop5 cells were vascular smooth muscle cells (vSMCs). Pericytes and vSMCs are both periendothelial cells and coexist on a related developmental spectrum (43). A key feature of vSMCs is their strong expression of alpha-smooth muscle actin (αSMA, Fig. 22*B*), which we discovered was expressed at high levels by Pop5 cells (Figs. 22*C* and S67) and was significantly higher than pericytes (adj. *p*-value = 1.03×10^{-4}). We also discovered Pop5 differentially expressed additional vSMC-associated genes, including those encoding chymase and calponin (Fig. S68). Other genes differentially expressed between Pop5 and Pop6 cells included melanoma cell adhesion molecule (*MCAM*, a.k.a. CD146) and *CD36*, the latter expressed more highly by Pop6 pericytes; and *MCAM* being highest in Pop5 (Fig. 22, *D* and *E*). When we immunostained tissues for these two markers, we found vSMCs and pericytes indeed displayed this expression pattern, further strengthening our conclusion that Pop5 cells were vSMCs (Figs. 22, *F* and *G* and S69). Nevertheless, there was one more piece of supporting data: The primary marker separating Pop5 from Pop6 in the FACS strategy was thy1—which we used to identify and isolate Pop6 pericytes (and myoepithelial cells). Naturally, we were curious to know whether vSMCs in the tissue lacked thy1. We found they did—*via* thy1 immunostaining of breast tissues (Fig. S71).

After analyzing each breast tissue sample by FACS, we found Pop5 vSMCs were relatively scarce in our preparations, constituting only 0.2% ± 1% of all cells in the tissue. Thus, vascular smooth muscle cells are the second least abundant cell type in the breast, slightly more prevalent than lymphatic endothelial cells. After our successful identification of Pop5, only one population in the final FACS plot remained—and it proved to be the most puzzling of them all.

Population 4: "epithelial" cells

A single population stood apart from all other cell types in the "Not Luminal" node, predominantly due to its expression of alpha-6 integrin (CD49f). We designated this fraction, Pop4 (Fig. 20*A*). Its expression of CD49f was peculiar because we believed, based on our tissue immunostaining, that we had accounted for all CD49f-expressing cell types in the tissue (Figs. S19 and S22). The previously characterized CD49f-expressing cell types include: (a) ER^Neg luminal cells (Pop2—gated by CD24/Muc1), (b) MEPs (Pop3—gated by Thy1), and (c) both endothelial cell types (Pops 7&9—gated by CD34, Fig. 9). Nevertheless, despite removing these cell types in the preceding gates, a small number of CD49f^Pos cells remained, which formed the Pop4 fraction. This population persisted

found situated among CD10^Pos myoepithelial cells. The confocal image used in Figure 20*C* has a different aspect ratio than neighboring images, so the image (surrounded by *dashed lines*) was placed on a *black* background for aesthetic purposes.

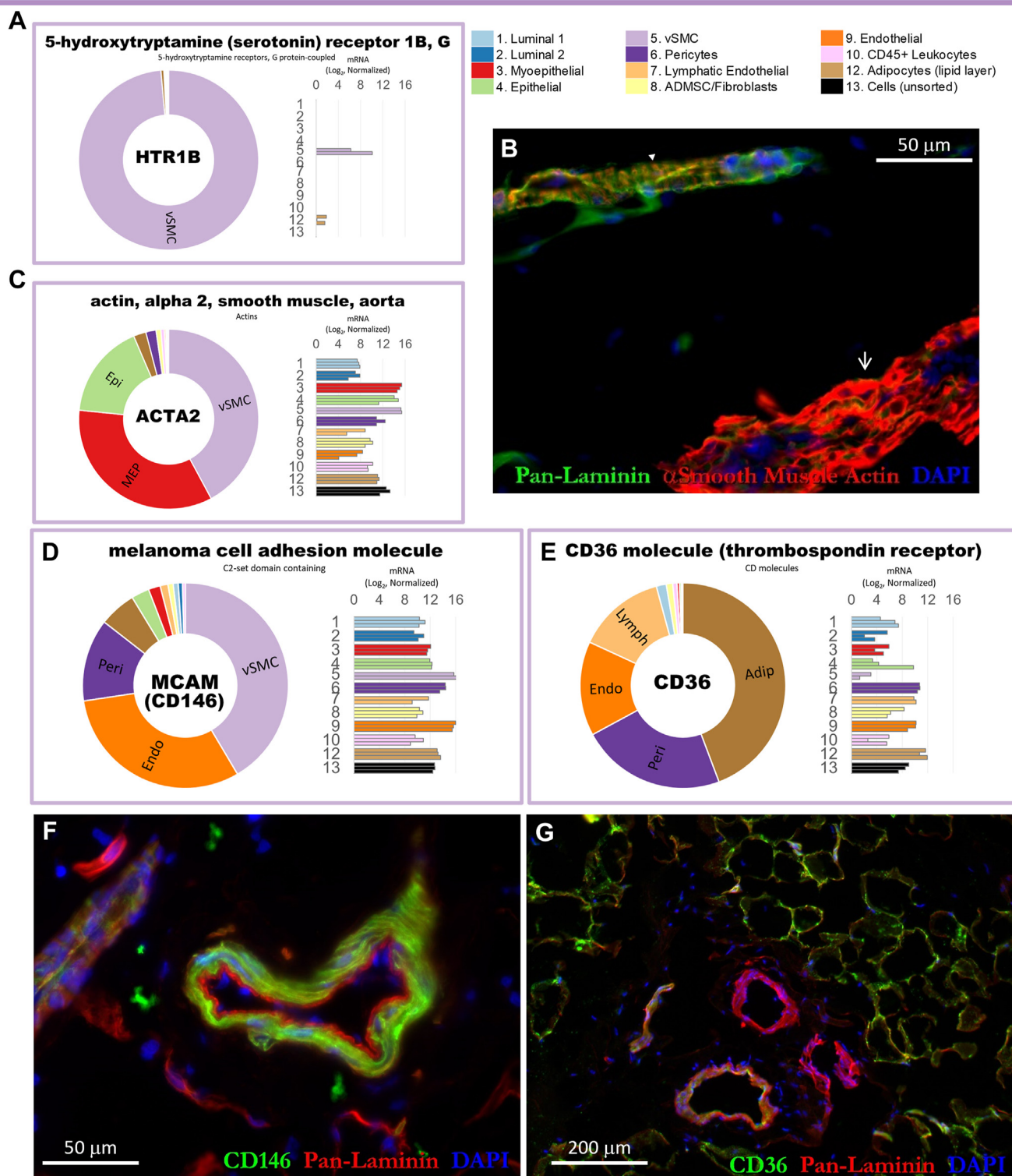# Vascular Smooth Muscle Cells (CD24/Muc1^Neg Gate)

Figure 22. Identification and validation of Vascular Smooth Muscle Cells. *A*, transcript levels of *HTR1B* (the gene encoding 5-hydroxytryptamine receptor 1B) within uncultured FACS-purified cell types, as determined by RNA-sequencing. HTR1B is an isoform of serotonin receptor, similar to the HTR1F isoform, which was uniquely-expressed by breast pericytes. *B*, Breast tissue immunostaining demonstrates alpha smooth muscle actin expression within arteriolar vascular smooth muscle cells ('▼', *top left*); and within ductal myoepithelial cells ('↑', *bottom right*). *C*, *ACTA2* (the gene encoding alpha smooth muscle actin) is differentially expressed between Pop5 vSMCs (*light purple*) and Pop6 Pericytes (*dark purple*). *D* and *E*, transcript levels of (*D*) *MCAM* (CD146) and (*E*) *CD36* in uncultured FACS-purified cell types. *F* and *G*, normal breast tissue immunostained with (*F*) CD146 and (*G*) CD36.

even when we used extreme gating of the above markers to ensure that weakly stained cells at each gate's border were not slipping through, which would inadvertently divert them to the Pop4 gate.

After analyzing many samples, we found Pop4 cells to be quite rare, as they comprised only 0.6% of all cells in the breast (0.6 ± 2%, median±sdev, Table S3). However, in contrast to Pop5 cells found in the same gate, Pop4 cells established growing cultures when seeded into culture. The phenotype of Pop4 cells was most similar to MEPs, but cultures often contained morphologically diverse phase-contrast phenotypes that were unlike homogeneous MEP cultures (Fig. 23, *A* and *B*; compared to S43). Based on this phenotype, we determined Pop4 cells were epithelial. There were also several lines of evidence indicating Pop4 cells were distinct from luminal and myoepithelial cells. For example, in low-density 3D Matrigel cultures, Pop4 cells produced structures consisting of keratin 19 positive cells surrounded by a layer of cells co-expressing both keratin 14 and 19 (Fig. 23, *C* and *D*). In contrast, pure luminal and myoepithelial cells homogenously expressed keratin 19 or 14, respectively (Fig. 23, *E* and *F*). The mixed cell phenotypes persisted when we seeded Pop4 cells onto a 2D collagen substratum—at clonal density—that produced cultures that heterogeneously expressed Muc1, suggesting they possess bipotent progenitor activity (Fig. S72, *A–C*). Lastly, Pop4 cells demonstrated a unique beta-1 integrin (CD29) expression pattern distinct from all other CD49f-expressing cell types (Figs. 23*G* and S72, *D–G*). How Pop4 cells related to the other epithelial cell types was a central question we aimed to clarify through RNA-sequencing.

Unsupervised hierarchical clustering of transcripts from each cell population confirmed Pop4's unique epithelial nature, and it clustered most closely with myoepithelial cells (MEPs). However, a subset of genes showed divergence, aligning Pop4 more closely with luminal cells, underscoring its distinctive gene expression pattern (Fig. S72, *D–H*). The transcriptional analysis identified protocadherin 20 (*PCDH20*) as uniquely expressed in Pop4 cells (Fig. 23*H*). Immunostaining of breast tissues with a PCDH20 antibody revealed rare PCDH20-positive cells, occurring at a frequency of 1 to 3 cells per section. These PCDH20-positive cells were located within the stroma of terminal ductal lobular units (TDLUs), adjacent to the basement membrane but outside the epithelial compartment (Fig. 23*I*). Importantly, these PCDH20-positive cells did not co-express CD45, confirming they were not leukocytes (Fig. 23*J*). Instead, they co-expressed keratin 14, establishing their epithelial identity, consistent with the primary culture phenotype of Pop4 cells (Fig. 23, *I* and *J*).

Due to the investigative nature of identifying Pop5 vascular smooth muscle cells (vSMCs) and Pop4 epithelial cells, these populations did not rely on positive selection using a lineage-specific marker. Future studies of these cells would benefit from single-cell RNA sequencing to evaluate potential heterogeneity in these populations, as well as the addition of redundant markers for their FACS purification.

Our characterization of the four cell types identified within this node (null gate) brings the overall tally to twelve identified and purified breast cell types, marking the conclusion of our isolation scheme.

Summary

The number and different types of cells composing the breast have been clouded with uncertainty, as have their origins and relationships to each other. We are not always sure how best to identify these different cell types, know where in the tissue they are located, how they communicate with each other and are maintained, or what all their respective functions may be. To these questions and more, we have some answers; but not all are sufficient or complete.

Since the emergence of methods permitting the outgrowth of cells from normal breast tissues in the late 1970s (48, 49), the field's most widely adopted primary cell models have essentially represented only two kinds of cells (myoepithelial cells and fibroblasts). The breast, however, is an intricate structure whose function relies on a symphony of many different specialized cell types. Understanding how these cells interact—and knowing the consequences of their interactions—is essential to developing necessary insights into the tissue and the processes that go awry in malignancy. There is thus an urgent need for more context-specific models to study reciprocal interchanges between the cell and the ECM—and among other cell types (50). Despite this widespread acknowledgment, a method to comprehensively identify, isolate, and culture the spectrum of cell types composing this elaborate tissue has not emerged, even in the new era of single-cell transcriptomics.

To establish a comprehensive understanding of the cellular landscape of breast tissues and lay a robust foundation for future studies, we meticulously examined a broad collection of normal breast tissues using both microscopic and flow cytometric approaches. This article constitutes the first of a pair focused on the cellular composition of breast tissues. Here, we present detailed immunostaining results, alongside insights into breast architecture and cellular composition, crucial for interpreting our forthcoming transcriptomic analyses (26).

While our transcriptomic analysis will provide deeper insights into the molecular features defining each identified breast cell type, understanding the breast architecture, cellular composition, and considerations that ensure the purity and identity of each cell population would be advantageous. These topics, details, and an atlas of immunostaining results are presented here, along with primary cell models and preliminary RNA-sequencing data of markers that collectively helped identify and confirm cell identities.

To develop a comprehensive sorting strategy, we sought high-resolution markers and used redundancy rigorously to identify cell types (Figs. 9, S16, and S17). We demonstrate the importance of this approach and show that many luminal epithelial cells would have otherwise been inappropriately sorted into the myoepithelial fraction had this not been discovered and alternative strategies employed (Figs. 7 and S12). Analysis of FACS plots and immunostaining identified CD49f[Neg] and CD49f[Pos] luminal cells as ER[Pos] and ER[Neg] cells, respectively, and we show the proportion of these different

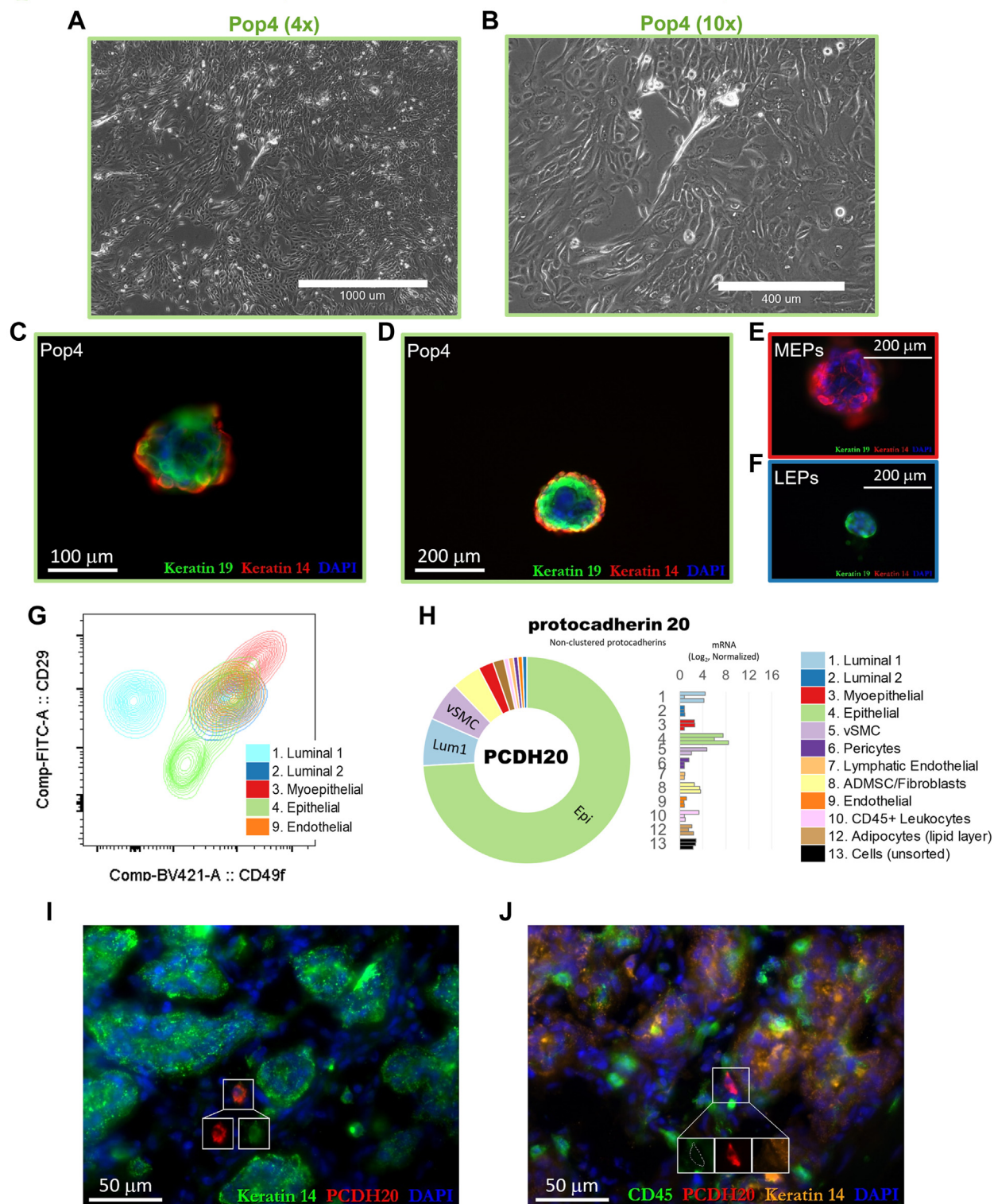# 'Epithelial Cells' (CD24^{Neg} & Muc1^{Neg} Gate)

Figure 23. Identification and validation of 'Pop4' epithelial cells. *A* and *B*, phase-contrast images of primary Pop4 epithelial cultures imaged with (*A*) 4× and (*B*) 10× objectives (from the same well). *C–F*, 3D matrigel cultures of (*C* and *D*) pop4 epithelial cells, (*E*) Pop3 MEPs, (*F*) Pop2 Luminal epithelial cells immunostained for keratin 14 and 19. *G*, Flow cytometry scatter plot of β1-integrin (CD29) and α6-integrin (CD49f) in the five different CD49f-expressing breast cell types. *H*, Transcript levels of protocadherin20 (*PCDH20*), a gene uniquely expressed by Pop4 cells. Transcripts were measured in freshly sorted (uncultured) cell types by RNA sequencing. Normalized mRNA values (rlog) are provided on log2 scale (bar graph of each biological replicate), as well as on a linear scale (donut graph of median value), each color-coded by cell type (Pop4 epithelial cells are *green*). *I* and *J*, Pop4 epithelial cells within breast tissues, revealed by PCDH20 and keratin 14 co-staining, and absence of pan-leukocyte maker CD45 staining.

luminal cell types varies considerably between individuals (Figs. 17–19 and S51–S56). Furthermore, we demonstrate that primary cultures of these different luminal cell fractions, though exhibiting different seeding requirements, can both be propagated in culture without the need for genetic manipulations or extraordinary effort (Figs. 17, G and H and S52, B and C). We have also identified a new and unusual epithelial cell with unknown functions (Pop4), which does not express CD24, Muc1, or Thy1—respective markers of mature luminal and myoepithelial cells (Figs. 9, 23, S16, S17, and S72–S74). Finally, we have identified and described methods for sorting and culturing perivascular pericytes and lymphatic and endothelial cells, that along with the other described models, significantly expand the repertoire of available breast cell models that may be employed.

Our FACS isolation strategy has successfully resolved twelve distinct breast cell types, with potential for more refinement to include a broader array of leukocyte populations residing in the breast (Figs. 9 and 20, S16, S17, and S57–S66). We routinely sort and culture nine FACS-purified cell types from breast tissues using these procedures and methods. These are ER^{Pos} and ER^{Neg} luminal epithelial cells, myoepithelial cells, the novel epithelial cell fraction, pericytes, fibroblasts, lymphatic and vascular endothelial cells, and adipocytes (Pops 1,2,3,4,6,7,8,9,&12; Fig. 24). Several of these breast cell models are the first of their kind.

The ability to purify and culture these breast cell types allows us to now explore the heterogeneity within each population, such as using single-cell RNA-sequencing, and to investigate their functional properties when cultured alone or in different combinations and contexts. We anticipate that the models, procedures, and concepts presented here will enhance clarity and deepen our understanding of the cellular composition and characteristics of breast tissues. These tools are poised to advance our exploration of normal cell and tissue functions and may offer insights into disease processes not only within the breast but also across other tissues and organ systems.
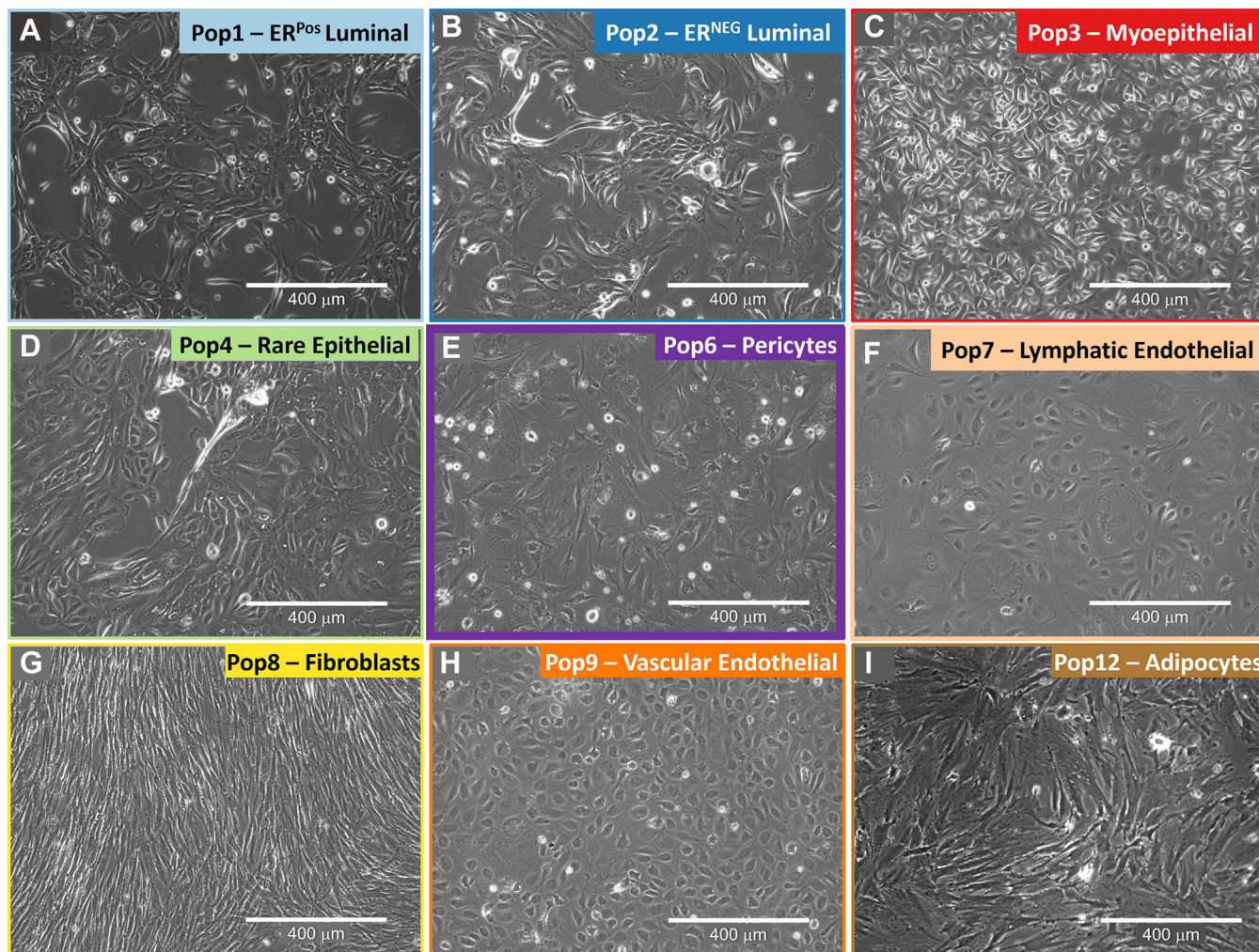
Figure 24. Primary cell models were generated using the FACS strategy. Using the FACS isolation strategy detailed herein, cells were sorted and used to create primary cultures of (*A*) Pop1 ER^{Pos} Luminal epithelial cells, (*B*) Pop2 ER^{Neg} Luminal epithelial cells, (*C*) Pop3 Myoepithelial cells, (*D*) Pop4 Epithelial cells, (*E*) Pop6 Pericytes, (*F*) Pop7 Lymphatic Endothelial cells, (*G*) Pop8 Adipocyte-derived Mesenchymal Stem Cells (Fibroblasts), (*H*) Pop9 Vascular Endothelial cells, and (*I*) Pop12 Adipocytes. Scale = 400 μm. Images reused from Figure 17, *G* and *H* (luminal epithelial cells), 8f (vascular endothelial cells) and 23b (Pop4 epithelial cells).

Isolation and culture of human breast cell types

Experimental procedures

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 the use of anonymous samples through exemption status, according to the Code of Federal Regulations 45 CFR 46.101. Upon receipt, several fragments (roughly 2 cm^2) were embedded in optimal cutting temperature compound (OCT) in tissue cassettes, flash-frozen in nitrogen, and archived at −80 °C 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) [5]. Briefly, this included manually mincing and scoring tissues with a scalpel and incubating them overnight (12–18 h, 37 °C) with 0.1% collagenase I (Gibco/Invitrogen) in Dulbecco's Modified Eagle Medium containing 100 U/ml penicillin, 100 μg/ml streptomycin, and 100 μg/ml Normocin (Invivogen). The resulting divested tissue fragments (organoids) were collected by centrifugation ($100g \times 2$ min) and either archived in liquid nitrogen (90% FBS+10% DMSO) or immediately processed to cell suspensions for FACS analysis. Cells used for RNA-sequencing were derived from fresh unarchived tissues (not previously archived in nitrogen).

Antibodies

A list of antibodies and reagents used in this study is provided in Tb 1. The table includes antibody clone designations, conjugations, isotype, supplier product numbers, dilution factors used, and citations to figures. Antibody specificity was evaluated by immunostaining tissues and cells using redundant antibody combinations and by comparing results to transcriptional data sets (RT-PCR and RNA sequencing).

Cryosectioning and immunostaining

Immunofluorescence was performed on 10 to 80 μm cryosectioned tissue sections (and cell cultures) by fixing samples with 4% paraformaldehyde for 5 min at 23 °C, followed by 4% formaldehyde/0.1% saponin for 5 min at 23 °C. Tissues were subsequently incubated for 20 min in wash buffer (0.1% saponin/10% goat serum in PBS), and incubated with primary antibodies diluted in wash buffer at the indicated antibody dilution ratios (Table S1). Samples were incubated at empirically determined times, typically overnight at 4 °C. Following the primary antibody incubations, samples were washed and incubated with anti-mouse, anti-rabbit, anti-chicken or anti-sheep secondary antibodies, respectively conjugated with Alexafluor 488, 568, 594, or 647 (Thermo), diluted 1:400 in wash buffer. After 1 h incubation at 23 °C, samples were rinsed in PBS and their nuclei were counterstained with 300 nM DAPI (4′,6-Diamidino-2-Phenylindole, Dihydrochloride, Thermo). Coverslips were mounted with Fluoromount G (Southern Biotech). Images were captured using a Zeiss LSM710 confocal microscope, Zeiss Axioscope, or EVOS FL Auto imaging station. Image contrast was applied to the entire image using Photoshop and was annotated with the antibodies used to stain the tissue. Thick tissue sections (30–80 μm) were similarly treated, but primary antibody incubation times were extended to 3 days, and rinse times were extended to 2 to 3 h each.

Cell preparation for flow cytometry and FACS analysis

To prepare cell suspensions for FACS and flow cytometry analysis, organoids were rinsed twice in PBS and pelleted by centrifuging them at $100g \times 2$ min. After removing PBS by aspiration, the organoids were suspended in 1 ml of "Cell Dissociation Reagent" (Sigma# C5914; or Thermo# 13150016), incubated for 2 minutes at room temperature, upon which 3 ml trypsin (or TrypLE) was added (0.25%, Thermo cat#25200072, or TrypLE, Thermo cat#12605010; Note that we have found TrypLE better preserves CD49f staining but worsens CD34 resolution). Samples were incubated by hand/body temperature, thus permitting visual inspection and brief (1-3s) pulse vortexing every 30 to 60 s. After the mixture became cloudy (about 8–10 min), the cells were gently and repeatedly pipetted through a 16-gauge needle until the clumps of cells dissipated. Afterward, we filtered cell suspensions through a 100 μm cell strainer and added 3 ml 0.1% w/v soybean trypsin inhibitor to stop the digest (Sigma# T9128). The suspensions were then filtered through a 40 μm cell strainer, rinsed in 10-20 ml PBS, and pelleted by centrifugation at high centrifugal force ($400g \times 5$ min.—or longer, until all cells had pelleted, which we confirmed by microscopic examination of the supernatant). The cells were then rinsed in 10 ml Hanks balanced salt solution/1% BSA (w/v), counted, and again pelleted by centrifugation. After centrifugation, nearly all the Hanks/BSA was aspirated, leaving the cell pellet with roughly 60 μl of Hanks/BSA. The pellet was resuspended in this small volume, and we added FACS antibodies to the samples and incubated the samples on ice, covered, for 30 min (See Table S1 for antibodies). Following the incubation, the cells were rinsed in Hanks/1% BSA, centrifuged ($400g \times 5$ min), and resuspended in Hanks/BSA with To-Pro-3 viability marker (diluted 1:4000, Thermo T3605). We did not use DNase enzyme or hypotonic RBC lysis solutions on the cell preparations because we found they were unnecessary, and we wanted to avoid any potential deleterious effect these treatments may have on the cells, such as reducing cell viability or altering gene expression levels. We also conscientiously chilled samples during cell preparation and the sorting procedure for similar concerns.

Flow cytometry

The FACS panel and gating strategy outlined herein was developed using a BD FACS Vantage 8-channel cytometer

(FACSDIVA software) and has since been adapted and reproduced using a 14-channel SONY SY3200. PMT voltages were optimized on both instruments by analyzing individually stained compensation beads (AbC Anti-Mouse Bead Kit, Thermo # A10344) at increasing voltage intervals for each detector. Due to the higher autofluorescence of breast cells (compared to polystyrene beads), we found that defining the compensation spillover matrix was optimally set if individually stained breast cells were used as compensation controls. Breast cells were also used to titrate each antibody to identify the optimal staining dilutions. Because the FACS Vantage has only eight available channels, it was necessary to use two markers (podoplanin and CD45) in a single channel (A488), which we experimentally validated. To-Pro-3 was used as the viability marker, so the corresponding 405 nm channel (455/50) could be used for BV421. This channel was dedicated to CD49f to provide a maximum resolution of cell populations, which is essential to this sorting strategy. The considerable compensation typically required between the heavily overlapping dyes PE/Cy5 and To-Pro-3 was not needed, as the To-Pro-3 negative (viable) cells were used for downstream analyses, allowing us to use the PE/Cy5 channel for Thy1. Negative controls consisted of unlabeled beads and cells incubated with isotype control antibodies conjugated to each fluorophore. The additional channels available on the SONY SY3200 allowed us to move CD45 to an independent channel (PE/Cy5), which was available after we moved Thy1 to the PE-Dazzle-594 channel. This adjustment permitted sorting of the CD45[Pos] leukocytes earlier in the gating scheme (Fig. S17). Sort times typically ranged between 4 to 6 h (∼10million cells stained), and 14 to 18 h (50–120 million cells stained— needed to yield enough cells for RNA-sequencing). Cells were chilled at the sample intake and collection chambers, and all samples were chilled on ice during the entire experiment. All media and PBS were filtered through a 0.2 μm filters to prevent the FACS machine from triggering events from crystals and debris. Staining profiles of cells collected at the beginning and end of the sorts (sometimes separated by as many as 18 h) were identical and did not show any signal loss or appreciable decrease in viability. To purify cells for cell culture, we found it necessary to triple-sort the cells. Because our machines could sort cells into a maximum of four tubes, we naturally needed to sort the cells into four groups, containing 2 to 4 populations each. These mixtures were: Tube1 (CD34Pos) and contained Pops 7,8,9; Tube 2 (Thy1[Pos]) contained Pops 3,6; Tube 3 (Luminal) contained Pops 1&2; and Tube 4 (Null gate) contained Pops 4,5,10,11. During the first round of sorting, cells were typically sorted at approximately 8000 cell/sec. We followed this initial enrichment with purity sorts for each cell type, using stringent sort masks, sorting each population into individual tubes at a rate of approximately 2000 cells/sec. For cells earmarked for cell cultures, the cells were sorted a third time to ensure purity of the isolated fractions and were sorted directly into 96-well dishes containing the appropriate medium (at 1–1500 cells/well). In some experiments, irradiated fibroblast feeder layers (1500 cells/well, 30Gy X-ray) were used to support growth and were applied into every other column of wells. FACS data were analyzed using Flowjo software (version v7.6.3–v10.8), Tree Star Inc./Becton, Dickinson & Company).

Cell culture

We generated primary cultures by seeding each cell type onto collagen I (PureCol, Advanced Biomatrix) coated 96-well cell culture dishes (Falcon 353,075) containing M87 (M87+CT+X (51)) for Pops 1,2,3,4; M87 supplemented with an additional 10% fetal bovine serum (FBS) for Pops 5,6,8; and EGM2 (Lonza CC-3162) supplemented with an additional 10% FBS for Pops 7&9. We supplemented media with 1× Penicillin (100 U/ml), 1×/Streptomycin (100 μg/ml), and 0.5× Normocin (50 μg/ml); EGM2 contains gentamicin and amphotericin and was therefore only supplemented with 0.5× normocin for the immediate culture post-FACS. A range of seeding densities was used in experiments, but 1500 cells per (96) well (∼4700 cells/cm^2) was generally sufficient to produce colonies in most wells for each cell type. When the cells were at low cell densities in these initial cultures, the media was initially replenished after 7 days post-seeding, then replenished every 2 to 3 days (Monday/Wednesday/Friday). We did not allow the epithelial cells (Pops1-4) to become confluent, but other cell types (Pops 6,7,8,9, & 12) all fared better when maintained at confluent densities and continued to divide in these conditions. Special attention was given when subculturing the cells, with the goal to minimize time exposed to trypsin. To subculture, we thoroughly rinsed each well 1× PBS, then applied Sigma Cell Dissociation Reagent (Sigma C5914 or Thermo 13,150,016—enough to cover the cells). Once the cells displayed early signs of detachment from the plate, which we monitored by phase microscopy, we added several drops (∼100ul) of 0.25% trypsin (or TrypLE). Typically, within 15 to 30 s, most cell types detach, at which point we gently and repeatedly pipetted the cells to detach them from the surface of the culture dish. We then inactivated the trypsin by adding an equal volume of 0.1% w/v soybean trypsin inhibitor. The cells were subsequently rinsed in excess PBS (10–12 ml), centrifuged, and either archived in cryopreservative medium (75% base medium, 15% FBS, 15% v/v DMSO) or re-seeded in their respective media onto PureCol coated dishes. All cell types were cultured under normoxic conditions at 37 °C with 5%CO_2, in a humidified cell culture incubator (Thermo/Forma).

Adipocyte cultures were created using a modified ceiling culture method (28). Briefly, T-25 flasks were coated with bovine collagen I (PureCol), turned upright, and filled with DMEM/F12 supplemented with 15% FBS and 1× Penicillin (100 U/ml), 1×/Streptomycin (100 μg/ml), and 0.5× Normocin (50 μg/ml). 100-300 μl media was removed and replaced with an equal volume of the lipid layer from the tissue/collagenase digests that had been prefiltered through a 40 μm cell strainer. The flask was capped, inverted, inserted into a secondary container, and placed into a humidified incubator. Having the flask upside-down permits adipocytes to float to the surface and attach to the collagen-coated surface. After 1 week in culture, the medium is removed, the flask is returned to an

upright position, and replenished with 3 ml fresh medium. The adipocyte cultures are subsequently maintained as a typical adherent culture—as described above for the other cell types. At first passage, the cells are harvested, stained, and FACs sorted for CD36 to remove potentially contaminating fibroblasts. Lipids were stained using Oil Red O (Sigma cat#O0625) or BODIPY (4,4-difluoro-1,3,5,7,8-pentamethyl-4-bora-3a,4a-diaza-s-indacene (BODIPY 493/503, Thermo cat#D3922).

Adipogenesis/osteogenesis lineage differentiation of Pop8 ADMSC/fibroblasts

To test the lineage differentiation capabilities of Pop8 fibroblasts, we placed early passage Pop8 fibroblasts into modified adipogenic and osteogenic conditions (37–39, 52, 53). Briefly, early passage (3-5p) Pop8 fibroblasts were seeded at a density of $\sim$53,000 cells/cm^2 in PureCol coated 4-well tissue culture dishes (100,000 cells per well) in M87/10% FBS + P/S + N (pen-strep/normocin-described above). After cells were attached to the dish, the medium was replaced with adipogenic or osteogenic differentiation medium, which was made fresh weekly and replenished bi-weekly (every 3–4 days). Adipogenic Differentiation Medium consisted of DMEM/F12, 10%FBS, 1× Penicillin (100 U/ml), 1×/Streptomycin (100 μg/ml), and 0.5× Normocin (50 μg/ml); Supplemented with: Dexamethasone (1 μM final, Sigma D4902), IBMX (0.5 mM final, Sigma I5879), Insulin (10 μg/ml final, Sigma I6634), Indomethacin (0.2 mM final, Sigma I7378), and Rosiglitizone (1 μM, Cayman 71740) – added fresh. Adipocyte induction medium (AIM) is reportedly stable for 3 weeks without rosiglitazone and 1 to 2 days with rosiglitazone. Differentiation was monitored by microscopy, and cells were maintained until they formed large lipid droplets, which usually took 10 to 21 days. Cells were stained by supplementing the complete growth medium with 2 μg/ml Hoechst 33342 and 10 μg/ml BODIPY, then incubated at 37 °C for 1 h. After the incubation, the cells were rinsed thrice with 1× PBS, and observed by fluorescent microscopy. Osteogenic Differentiation Medium consisted of Alpha-MEM supplemented with 10% FBS, 1× Penicillin (100 U/ml), 1×/Streptomycin (100 μg/ml), 100 nM dexamethasone, 50 μM L-Ascorbic acid 2-phosphate sesquimagnesium salt hydrate (Sigma A8960), 4 mM beta-glycerophosphate (10 mM), 0.5 μM 1α,25-Dihydroxyvitamin D3 (Sigma D1530). After 2 weeks, 20 μM Xylenol Orange (Fluka/Fisher 33825) was added to the medium and incubated for an additional 24 h. The cells were then incubated for an additional hour with Hoechst 33342, rinsed with PBS, and imaged by fluorescent microscopy.

RNA isolation and sequencing

Total RNA was isolated from FACS-sorted cells using Qiagen RNA Mini (>100,000 cells, Qiagen 74104), RNA Micro (<100,000 cells, Qiagen 74004); and from the collagenase lipid layer using Qiagen RNA Lipid Tissue Mini Kit (Qiagen 74,804) according to the to manufacturer's instructions. Total RNA was treated with DNAse I for 30 min at 37 °C (Ambion DNA-*free* DNA Removal kit, Thermo AM1906), then treated with DNAse Inactivation beads. RNA was then concentrated with RNA Clean & Concentrator-5 columns (Zymo R1015), quantified by Nanodrop and Agilent Bioanalyzer assays, and stored at −80 °C. Complete details of RNA sequencing and links to the dataset are provided in a separate article (26). Briefly, mRNA was captured using poly-T coated dishes (Clontech) and SMARTer technology was used for library construction (SMARTer Stranded RNA-seq Kit). Sequencing was performed on Illumina Trueseq v.2 technology, targeting approximately 30 million paired-end reads (on average) for each cell population. We processed raw count data using R programming language (RStudio) and associated Bio-Conductor packages. The count matrix was transformed for analysis, using two methods used in the DESeq2 package; *i.e.*, (i) regularized log transformation (rlog), and (ii) variance-stabilizing transformation (vst) (54). Quality control assessment of expressed transcripts across sorted cell samples was conducted and one sample was removed from statistical analyses, as it was an extreme outlier (Epithelial Population 4 Sample N239). This was determined by: (i) distance matrix, (ii) Principal Component Analysis, and (iii) Kolmogorov-Smirnov test that compares each sample to the reference probability distribution of the pooled data. This sample is however included in downstream visualizations, for example, the bar and donut graphs used throughout the manuscript, which display rlog transformed values on both log2 (bar graphs) and linear (donut graph) scales.

Quantitative RT-PCR array

A custom qRT-PCR array was generated by designing and validating PCR reactions for 148 genes related to breast cancer, cell adhesion and signaling, ECM and ligand pathways, and stem and developmentally related pathways, such as wnt/beta-catenin, notch, TGFbeta, and hedgehog. Primer sets are provided in Table S2 and contain the official gene symbols, pathway/category information, ENSEMBL accession IDs, number of exons per gene, intron spanning information, and sequences for FOR and REV primers. Although not used, we have included also sequences of designed qRT-PCR Taqman probes. Total RNA was isolated from FACS-sorted cells (Qiagen RNA Mini Kit), treated with DNAse I (Ambion DNA-*free*), and reverse transcribed using the SuperScript VILO cDNA Synthesis Kit (Thermo 11,754–050) in 20 ul reactions according to the manufacturer's protocol. After cDNA synthesis, the cDNAs were diluted 8×, by adding 140 ul water to the 20 ul reactions. Transcripts were amplified, by qPCR, using the LightCycler 480 SYBR Green I Master Mix in 10 μl reactions, using 1 μl diluted cDNA in each reaction. Primers, supplied as 300 nM final concentration in the reactions were designed to be 80-200 bp in length and overlapped exons, if possible. All PCR primers are provided in Table S2. PCR reaction parameters: 5 min activation at 95 °C for 5 min; 40 cycles each of melting at 95 °C for 10 s, one-step annealing and extension at 60 °C for 60 s; and was followed by melting curve analysis. Relative levels of transcripts were calculated using the delta Ct method and normalized to those of the *TBP*

reference transcript using the formula: $\%TBP = 2^{-(Ct_{Gene}-Ct_{TBP})}$.

References

1. Krutzik, P. O., Crane, J. M., Clutter, M. R., and Nolan, G. P. (2008) High-content single-cell drug screening with phosphospecific flow cytometry. *Nat. Chem. Biol.* **4**, 132–142
2. Bendall, S. C., and Nolan, G. P. (2012) From single cells to deep phenotypes in cancer. *Nat. Biotechnol.* **30**, 639–647
3. Bendall, S. C., Simonds, E. F., Qiu, P., Amir el, A. D., Krutzik, P. O., Finck, R., *et al.* (2011) Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696
4. Bonnet, D., and Dick, J. E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3**, 730–737
5. Hines, W. C., Su, Y., Kuhn, I., Polyak, K., and Bissell, M. J. (2014) Sorting out the FACS: a devil in the details. *Cell Rep.* **6**, 779–781
6. Mahnke, Y. D., and Roederer, M. (2007) Optimizing a multicolor immunophenotyping assay. *Clin. Lab. Med.* **27**, 469–485
7. Cossarizza, A., Chang, H. D., Radbruch, A., Akdis, M., Andrä, I., Annunziato, F., *et al.* (2017) Guidelines for the use of flow cytometry and cell sorting in immunological studies. *Eur. J. Immunol.* **47**, 1584–1797
8. Herzenberg, L. A., Tung, J., Moore, W. A., and Parks, D. R. (2006) Interpreting flow cytometry data: a guide for the perplexed. *Nat. Immunol.* **7**, 681–685
9. Ashhurst, T. M., Smith, A. L., and King, N. J. C. (2017) High-dimensional fluorescence cytometry. *Curr. Protoc. Immunol.* **119**. https://doi.org/10.1002/cpim.37
10. Mahnke, Y., Chattopadhyay, P., and Roederer, M. (2010) Publication of optimized multicolor immunofluorescence panels. *Cytometry. Part A.* **77**, 814–818
11. Gusterson, B. A., Ross, D. T., Heath, V. J., and Stein, T. (2005) Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res.* **7**, 143–148
12. Yamada, K. M., and Hall, A. (2015) Reproducibility and cell biology. *J. Cell Biol.* **209**, 191–193
13. Bradbury, A., and Pluckthun, A. (2015) Reproducibility: standardize antibodies used in research. *Nature* **518**, 27–29
14. Visvader, J. E. (2009) Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev.* **23**, 2563–2577
15. Petersen, O. W., and Polyak, K. (2010) Stem cells in the human breast. *Cold Spring Harb. Perspect. Biol.* **2**, a003160
16. Smalley, M. J., Kendrick, H., Sheridan, J. M., Regan, J. L., Prater, M. D., Lindeman, G. J., *et al.* (2012) Isolation of mouse mammary epithelial subpopulations: a comparison of leading methods. *J. Mammary Gland Biol. Neoplasia* **17**, 91–97
17. Stingl, J., Eaves, C. J., Zandieh, I., and Emerman, J. T. (2001) Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res. Treat.* **67**, 93–109
18. Stingl, J., Raouf, A., Eirew, P., and Eaves, C. J. (2006) Deciphering the mammary epithelial cell hierarchy. *Cell Cycle* **5**, 1519–1522
19. Villadsen, R., Fridriksdottir, A. J., Ronnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M. A., *et al.* (2007) Evidence for a stem cell hierarchy in the adult human breast. *J. Cell Biol.* **177**, 87–101

20. Eirew, P., Stingl, J., Raouf, A., Turashvili, G., Aparicio, S., Emerman, J. T., et al. (2008) A method for quantifying normal human mammary epithelial stem cells with *in vivo* regenerative ability. *Nat. Med.* **14**, 1384–1389

21. Lim, E., Vaillant, F., Wu, D., Forrest, N. C., Pal, B., Hart, A. H., et al. (2009) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913

22. Bachelard-Cascales, E., Chapellier, M., Delay, E., Pochon, G., Voeltzel, T., Puisieux, A., et al. (2010) The CD10 enzyme is a key player to identify and regulate human mammary stem cells. *Stem Cells* **28**, 1081–1088

23. Keller, P. J., Arendt, L. M., Skibinski, A., Logvinenko, T., Klebba, I., Dong, S., et al. (2011) Defining the cellular precursors to human breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2772–2777

24. Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I. A., Avril, S., et al. (2012) Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res.* **14**, R134

25. Arendt, L. M., Keller, P. J., Skibinski, A., Goncalves, K., Naber, S. P., Buchsbaum, R. J., et al. (2014) Anatomical localization of progenitor cells in human breast tissue reveals enrichment of uncommitted cells within immature lobules. *Breast Cancer Res.* **16**, 453

26. [preprint] Del Toro, K., Sayaman, R. W., Thi, K., Licon-Munoz, Y., and Hines, W. C. (2022) A cellular and transcriptomic dissection of the human breast for studying mechanisms of cell and tissue function. *bioRxiv*. https://doi.org/10.1371/journal.pbio.3002820

27. Hines, W. C., Yaswen, P., and Bissell, M. J. (2015) Modelling breast cancer requires identification and correction of a critical cell lineage-dependent transduction bias. *Nat. Commun.* **6**, 6927

28. Zhang, H. H., Kumar, S., Barnett, A. H., and Eggo, M. C. (2000) Ceiling culture of mature human adipocytes: use in studies of adipocyte functions. *J. Endocrinol.* **164**, 119–128

29. Gusterson, B. A., Monaghan, P., Mahendran, R., Ellis, J., and O'Hare, M. J. (1986) Identification of myoepithelial cells in human and rat breasts by anti-common acute lymphoblastic leukemia antigen antibody A12. *J. Natl. Cancer Inst.* **77**, 343–349

30. Lennon, V. A., Unger, M., and Dulbecco, R. (1978) Thy-1: a differentiation marker of potential mammary myoepithelial cells *in vitro*. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 6093–6097

31. Sapino, A., Macri, L., Gugliotta, P., and Bussolati, G. (1990) Evaluation of proliferating cell types in human and mouse mammary gland by a double immunostaining procedure. *Acta Histochem. Suppl.* **40**, 81–84

32. Kalof, A., Tam, D., Beatty, B., and Cooper, K. (2004) Immunostaining patterns of myoepithelial cells in breast lesions: a comparison of CD10 and smooth muscle myosin heavy chain. *J. Clin. Pathol.* **57**, 625–629

33. Jones, E. A., Kinsey, S. E., English, A., Jones, R. A., Straszynski, L., Meredith, D. M., et al. (2002) Isolation and characterization of bone marrow multipotential mesenchymal progenitor cells. *Arthritis Rheum.* **46**, 3349–3360

34. Schaffler, A., and Buchler, C. (2007) Concise review: adipose tissue-derived stromal cells–basic and clinical implications for novel cell-based therapies. *Stem Cells* **25**, 818–827

35. Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., et al. (2006) Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **8**, 315–317

36. Lin, C. S., Ning, H., Lin, G., and Lue, T. F. (2012) Is CD34 truly a negative marker for mesenchymal stromal cells? *Cytotherapy* **14**, 1159–1163

37. Pittenger, M. F., Mackay, A. M., Beck, S. C., Jaiswal, R. K., Douglas, R., Mosca, J. D., et al. (1999) Multilineage potential of adult human mesenchymal stem cells. *Science* **284**, 143–147

38. Fink, T., and Zachar, V. (2011) Adipogenic differentiation of human mesenchymal stem cells. *Methods Mol. Biol.* **698**, 243–251

39. Liu, Y., Goldberg, A. J., Dennis, J. E., Gronowicz, G. A., and Kuhn, L. T. (2012) One-step derivation of mesenchymal stem cell (MSC)-like cells from human pluripotent stem cells on a fibrillar collagen coating. *PLoS One* **7**, e33225

40. Breiteneder-Geleff, S., Soleiman, A., Kowalski, H., Horvat, R., Amann, G., Kriehuber, E., et al. (1999) Angiosarcomas express mixed endothelial phenotypes of blood and lymphatic capillaries: podoplanin as a specific marker for lymphatic endothelium. *Am. J. Pathol.* **154**, 385–394

41. Podgrabinska, S., Braun, P., Velasco, P., Kloos, B., Pepper, M. S., and Skobe, M. (2002) Molecular characterization of lymphatic endothelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16069–16074

42. Herman, I. M., and D'Amore, P. A. (1985) Microvascular pericytes contain muscle and nonmuscle actins. *J. Cell Biol.* **101**, 43–52

43. Armulik, A., Genove, G., and Betsholtz, C. (2011) Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Dev. Cell* **21**, 193–215

44. Stingl, J., Emerman, J. T., and Eaves, C. J. (2005) Enzymatic dissociation and culture of normal human mammary tissue to detect progenitor activity. *Methods Mol. Biol.* **290**, 249–263

45. Fridriksdottir, A. J., Kim, J., Villadsen, R., Klitgaard, M. C., Hopkinson, B. M., Petersen, O. W., et al. (2015) Propagation of oestrogen receptor-positive and oestrogen-responsive normal human breast cells in culture. *Nat. Commun.* **6**, 8786

46. Westbury, C. B., Reis-Filho, J. S., Dexter, T., Mahler-Araujo, B., Fenwick, K., Iravani, M., et al. (2009) Genome-wide transcriptomic profiling of microdissected human breast tissue reveals differential expression of KIT (c-Kit, CD117) and oestrogen receptor-alpha (ERalpha) in response to therapeutic radiation. *J. Pathol.* **219**, 131–140

47. Moreira, J. M., Cabezon, T., Gromova, I., Gromov, P., Timmermans-Wielenga, V., Machado, I., et al. (2010) Tissue proteomics of the human mammary gland: towards an abridged definition of the molecular phenotypes underlying epithelial normalcy. *Mol. Oncol.* **4**, 539–561

48. Bissell, M. J., Bartholomew, J. C., Folkman, J., Smith, H., and Stampfer, M. (1979) Culture systems for studying malignancy. *Cancer Res.* **39**, 4293–4295

49. Stampfer, M., Hallowes, R. C., and Hackett, A. J. (1980) Growth of normal human mammary cells in culture. *In Vitro* **16**, 415–425

50. Bissell, M. J., and Hines, W. C. (2011) Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat. Med.* **17**, 320–329

51. Garbe, J. C., Bhattacharya, S., Merchant, B., Bassett, E., Swisshelm, K., Feiler, H. S., et al. (2009) Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res.* **69**, 7557–7568

52. Strutt, B., Khalil, W., and Killinger, D. (1996) Growth and differentiation of human adipose stromal cells in culture. *Methods Mol. Med.* **2**, 41–51

53. Lund, P., Pilgaard, L., Duroux, M., Fink, T., and Zachar, V. (2009) Effect of growth media and serum replacements on the proliferation and differentiation of adipose-derived stem cells. *Cytotherapy* **11**, 189–197

54. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550

55. Cooper, A. P. (1840) *On the Anatomy of the Breast-of the Internal Parts of the Breast, or Mammary Gland.* Longman, London, United Kingdom

56. Sarhadi, N. S., Shaw Dunn, J., Lee, F. D., and Soutar, D. S. (1996) An anatomical study of the nerve supply of the breast, including the nipple and areola. *Br. J. Plast. Surg.* **49**, 156–164

ASBMB