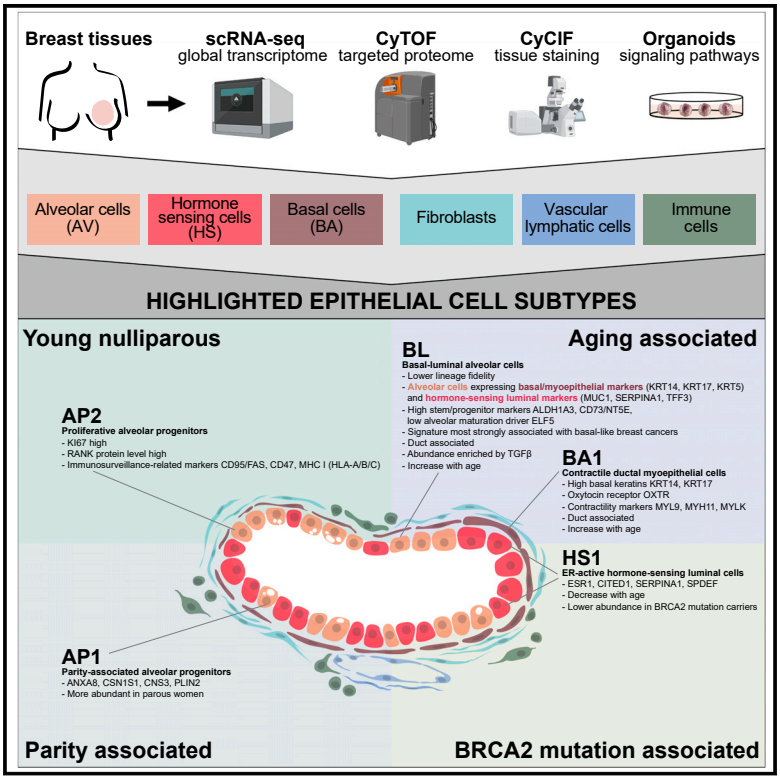
Resource

Developmental Cell

A human breast atlas integrating single-cell proteomics and transcriptomics

Graphical abstract



Highlights

- Multimodal single-cell analyses identify breast epithelial and stromal subtypes
- Spatially distinct epithelial subsets are linked with age, parity, and *BRCA2* status
- Alveolar cells with poor transcriptional lineage fidelity accumulate with age
- Subtypes of the three major epithelial lineages are maintained in organoid cultures

Authors

G. Kenneth Gray, Carman Man-Chung Li, Jennifer M. Rosenbluth, Laura M. Selfors, ..., Judy E. Garber, Deborah A. Dillon, Joan S. Brugge

Correspondence

joan_brugge@hms.harvard.edu

In brief

Gray, Li, Rosenbluth, Selfors et al. generate a multi-dimensional atlas of breast tissues and organoids. Distinct epithelial subtypes are found to be associated with age, parity, and *BRCA2* mutation. An alveolar subset termed basal-luminal cells displays poor transcriptional lineage fidelity and gene expression signatures associated with aggressive basal-like breast cancers.

Gray et al., 2022, Developmental Cell *57*, 1400–1420
June 6, 2022
https://doi.org/10.1016/j.devcel.2022.05.003

CellPress

CellPress

Developmental Cell

Resource

A human breast atlas integrating single-cell proteomics and transcriptomics

G. Kenneth Gray,[1,9] Carman Man-Chung Li,[1,9] Jennifer M. Rosenbluth,[1,2,3,9] Laura M. Selfors,[1,9] Nomeda Girnius,[1,4] Jia-Ren Lin,[4] Ron C.J. Schackmann,[1] Walter L. Goh,[1] Kaitlin Moore,[1] Hana K. Shapiro,[1] Shaolin Mei,[4] Kurt D'Andrea,[5] Katherine L. Nathanson,[5] Peter K. Sorger,[4] Sandro Santagata,[4,6] Aviv Regev,[7,8] Judy E. Garber,[2] Deborah A. Dillon,[6] and Joan S. Brugge[1,10,*]

[1]Department of Cell Biology, Harvard Medical School (HMS), Boston, MA 02115, USA
[2]Department of Medical Oncology, Dana-Farber Cancer Institute (DFCI), Boston, MA 02115, USA
[3]Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, USA
[4]The Laboratory of Systems Pharmacology (LSP), HMS, Boston, MA 02115, USA
[5]Department of Medicine, Division of Translation Medicine and Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[6]Department of Pathology, Brigham and Women's Hospital (BWH), Boston, MA 02115, USA
[7]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[8]Present address: Genentech, 1 DNA Way, South San Francisco, CA 94080, USA
[9]These authors contributed equally
[10]Lead contact
*Correspondence: joan_brugge@hms.harvard.edu
https://doi.org/10.1016/j.devcel.2022.05.003

SUMMARY

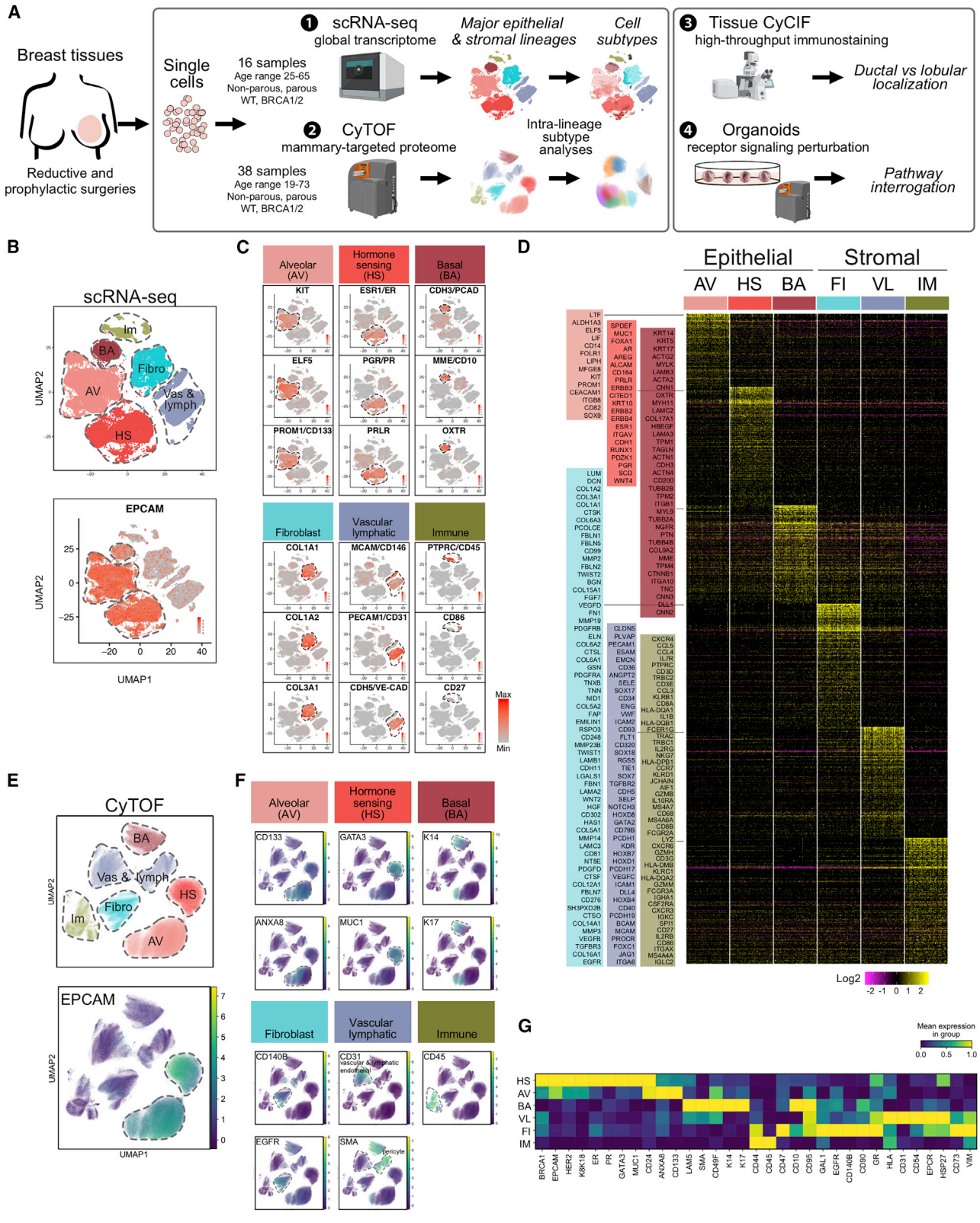
The breast is a dynamic organ whose response to physiological and pathophysiological conditions alters its disease susceptibility, yet the specific effects of these clinical variables on cell state remain poorly annotated. We present a unified, high-resolution breast atlas by integrating single-cell RNA-seq, mass cytometry, and cyclic immunofluorescence, encompassing a myriad of states. We define cell subtypes within the alveolar, hormone-sensing, and basal epithelial lineages, delineating associations of several subtypes with cancer risk factors, including age, parity, and *BRCA2* germline mutation. Of particular interest is a subset of alveolar cells termed basal-luminal (BL) cells, which exhibit poor transcriptional lineage fidelity, accumulate with age, and carry a gene signature associated with basal-like breast cancer. We further utilize a medium-depletion approach to identify molecular factors regulating cell-subtype proportion in organoids. Together, these data are a rich resource to elucidate diverse mammary cell states.

INTRODUCTION

Breast cancer is heterogeneous, comprising distinct subtypes with unique therapeutic vulnerabilities possibly rooted in their distinct cells of origin (Visvader and Stingl, 2014). However, investigation of early tumor formation and progression has been limited by an underdeveloped, occasionally conflicting annotation of mammary lineages. Existing annotations are often defined by a small number of markers that inadequately capture mammary intra-lineage heterogeneity. Many questions remain regarding the diversity of the mammary cellular landscape and how it changes in relation to cancer risk factors, such as age, parity, and germline mutations in breast cancer predisposition genes such as *BRCA1/2*. Thus, there is a pressing need to define breast cell types/states more comprehensively at high resolution from a broad spectrum of samples.

The mammary gland is composed of an epithelial bilayer and a supportive stroma (Inman et al., 2015; Polyak and Kalluri, 2010). Within the epithelium, basal/myoepithelial (BA) cells contract to help express milk during lactation and provide structural support to two luminal populations: alveolar (AV) and hormone-sensing (HS) cells. AV cells give rise to milk-producing cells during lactation, whereas HS cells integrate endocrine signals to control mammary growth and differentiation. AV and HS cells have historically been termed luminal progenitors and mature luminal cells, respectively, based on *in vitro* assays (Fu et al., 2020; Shackleton et al., 2006; Stingl et al., 2006). However, recent post-natal lineage tracing in murine glands show these two cell types are not hierarchically organized but rather are independent (Blaas et al., 2016; Davis et al., 2016; Elias et al., 2017; Lilja et al., 2018; Rios et al., 2014; Rodilla et al., 2015; Van Keymeulen et al., 2017; Wang et al., 2017). These mammary epithelial cell (MEC) lineages undergo dramatic alterations in morphology and function throughout a human life, including in puberty, pregnancy, and menopause (Fu et al., 2020; LaBarge et al., 2016; Slepicka et al., 2021).

Many physiological/genetic variables and cancer risk factors influence MEC composition. Depending on cohorts analyzed and methods employed, previous studies have reported various



(legend on next page)

alterations in MEC proportions and properties associated with aging, parity, and germline *BRCA1/2* mutations (Benz, 2008; Britt et al., 2007; Choudhury et al., 2013; Ding et al., 2019; Garbe et al., 2012; Honeth et al., 2015; Karaayvaz-Yildirim et al., 2020; King et al., 2004; Lim et al., 2009; Mote et al., 2004; Nolan et al., 2016; Pal et al., 2021; Pelissier Vatter et al., 2018; Proia et al., 2011; Russo et al., 1992; Shalabi et al., 2021). These studies have provided somewhat fragmented views of the mammary cell landscape; a systematic, high-resolution taxonomy of the breast can provide a framework for large-scale studies to delineate the effects of risk factors on cancer development.

Recent advances in single-cell technologies like single-cell RNA sequencing (scRNA-seq) and mass cytometry (cytometry by time-of-flight; CyTOF) have enabled a more comprehensive evaluation of mammary cell diversity (Bach et al., 2017; Bhat-Nakshatri et al., 2021; Colacino et al., 2018; Giraddi et al., 2018; Henry et al., 2021; Hu et al., 2021; Kanaya et al., 2019; Knapp et al., 2017; Li et al., 2020; Mahendralingam et al., 2021; Nguyen et al., 2018; Pal et al., 2021, 2017; Pelissier Vatter et al., 2018; Rosenbluth et al., 2020; Scheele et al., 2017; Twigger et al., 2022; Wuidart et al., 2018). However, a comprehensive atlas of breast cell types has not yet been constructed due to high inter-individual variability, highlighting a need for broader studies to capture full intra-lineage cell diversity. In particular, a multi-omic approach may reveal nonsynonymous levels of heterogeneity (Chung et al., 2019; Ding et al., 2020; Hao et al., 2021).

Here, we report a high-resolution breast cell taxonomy generated by integrating scRNA-seq (n = 16), CyTOF (n = 38), and cyclic immunofluorescence (CyCIF; n = 53). We analyzed samples from reductive mammoplasties and prophylactic mastectomies encompassing a broad spectrum of ages, parities, and germline *BRCA1/2* mutation statuses. Organoid modulation experiments uncovered potential molecular regulation of several MEC subtypes, thus demonstrating a feasible approach to illuminate targetable changes in high-risk individuals' breasts. Collectively, this multi-omic, single-cell approach yields a breast atlas of cell lineages and intra-lineage subtypes, uncovers altered MEC subtypes associated with cancer risk factors, and enables signaling pathway interrogation in cell subtypes *in vitro*. This multifaceted approach may serve as a template for studying premalignant changes in the breast and other tissues.

RESULTS

scRNA-seq and CyTOF identify breast epithelial and stromal cell types

To generate a high-resolution portrait of breast cell types, scRNA-seq and CyTOF were utilized to provide complementary views of cell diversity at the mRNA and protein levels (Figure 1A). The scRNA-seq was performed on 52,681 cells from 16 breast tissues, including tissues from noncarriers (n = 3) and carriers of germline mutations in *BRCA1* (n = 6), *BRCA2* (n = 6), or *RAD51C* (n = 1), spanning a range of ages (25–65) and parities (Table S1). Cells were analyzed by unsupervised graph-based clustering and visualized using Uniform Manifold Approximation and Projection (UMAPs) (see STAR Methods) (Figure S1A). MECs and stromal cells were distinguished by *EPCAM* expression (Figure 1B). Clusters were annotated as specific cell types based on canonical marker expression of the three established MEC lineages (AV, HS, and BA) and three stromal populations (fibroblasts, vascular/lymphatic cells, and immune cells) (Figures 1B and 1C). All cell types were detected in each sample, although their proportions varied substantially (Figure S1B). Gene expression signatures were generated for each cell type, which can serve as a benchmark for future cell-type identification and isolation (Figure 1D; Table S2).

Concurrently, CyTOF was performed on 38 breast tissues (Table S1; ages 19–73, including noncarriers [n = 17] or carriers of germline mutations in *BRCA1* [n = 9], *BRCA2* [n = 11], or *RAD51C* [n = 1]), using our previously reported antibody panel recognizing 40 breast development and tumorigenesis markers (Rosenbluth et al., 2020) (Figures S1C and S1D). To identify major cell types, the expression of these markers was quantified in 10,699,281 cells and visualized by UMAP for a subset of these cells (n = 751,970) (see STAR Methods). CyTOF revealed the same cell types as scRNA-seq (Figure 1E). EPCAM marked luminal MECs, but little or no EPCAM protein was detectable in BA cells despite their expression of low levels of *EPCAM* mRNA (Figures 1B and 1E). Nevertheless, BA cells were clearly distinct from stromal cells based on co-expression of BA lineage markers (K14, K17, and CD10) (Figures 1F and 1G). Luminal cells were further subdivided into the AV and HS types based on the expression of key markers CD133 and ANXA8 (AV) or GATA3 and MUC1 (HS) (Figures 1F and 1G). Stromal cells were identified by canonical markers for fibroblasts (CD140B and EGFR), vascular/lymphatic cells (CD31 for endothelial cells and smooth muscle actin [SMA] for pericytes), and immune cells (CD45) (Figures 1F and 1G). Each major cell type was present in each sample, albeit in considerably variable proportions (Figure S1E). Cell-type proportions correlated between modalities for three shared samples analyzed (Figure S1F).

scRNA-seq reveals diverse cell subtypes

The heterogeneity within each lineage was investigated by sub-clustering of the scRNA-seq data of each major lineage (see STAR Methods) (Figure 2A). Among stromal cells, sub-clustering

Figure 1. Major mammary epithelial and stromal cell types identified by scRNA-seq and CyTOF

(A) Schematic of approach using scRNA-seq, CyTOF, CyCIF, and organoid cultures.

(B) UMAPs of scRNA-seq breast cells colored by cell type (top) or *EPCAM* expression (bottom) (n = 16). Cells fell into three MEC (AV, alveolar; HS, hormone-sensing; BA, basal/myoepithelial) and three stromal types (Fibro, fibroblasts; Vas & Lymph, vascular and lymphatic cells; Im, immune cells).

(C) Canonical lineage marker expression in distinct MEC and stromal cell types from scRNA-seq.

(D) Heatmap of cell-type-specific markers for each cell type as identified by multi-pairwise differential gene expression analysis of scRNA-seq data. Select markers are highlighted. See Table S2.

(E) UMAPs of breast cells analyzed by CyTOF (top) and EPCAM expression (bottom) (n = 38). Three samples overlap with the scRNA-seq cohort. See Table S1.

(F) Canonical lineage marker expression in MEC and stromal cell types from CyTOF data.

(G) Matrix plot of lineage markers for each cell type in the CyTOF data.

See also Figure S1.

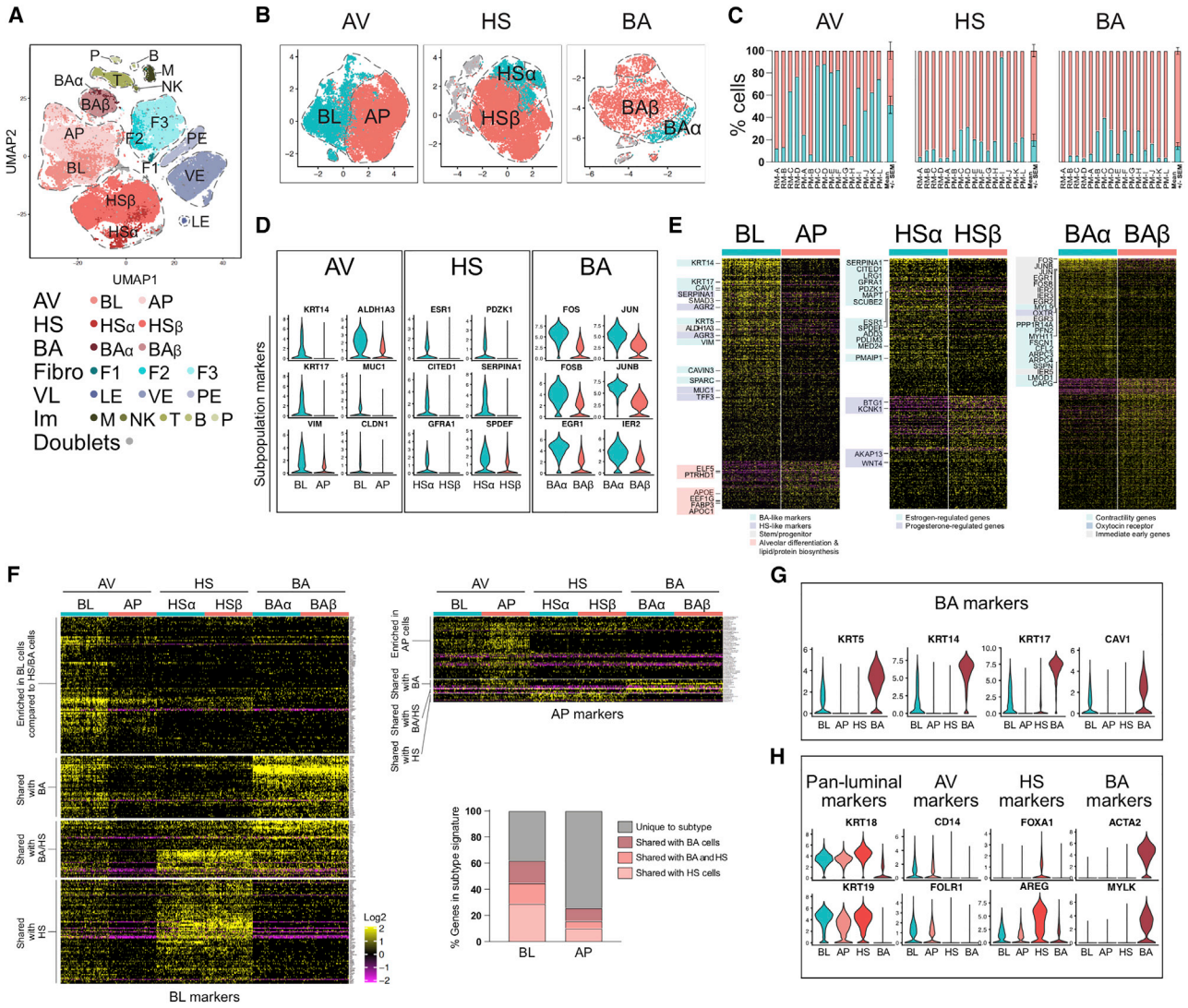


Figure 2. Diverse cell subtypes identified by scRNA-seq

(A) Sub-clustering identifies six MEC and 11 stromal subtypes. Minor subclusters (gray) were discarded as potential doublets due to unusually high gene counts and aberrant marker expression.

(B) UMAPs of the subtypes within the AV (BL, AP), HS (HSα, HSβ), and BA lineages (BAα, BAβ). Gray cells are as in (A).

(C) MEC subtype proportions across 16 samples. Color code is identical to (B).

(D) Violin plots of MEC subtype marker expression.

(E) Heatmaps showing MEC subtype-specific gene expression signatures. See Table S2.

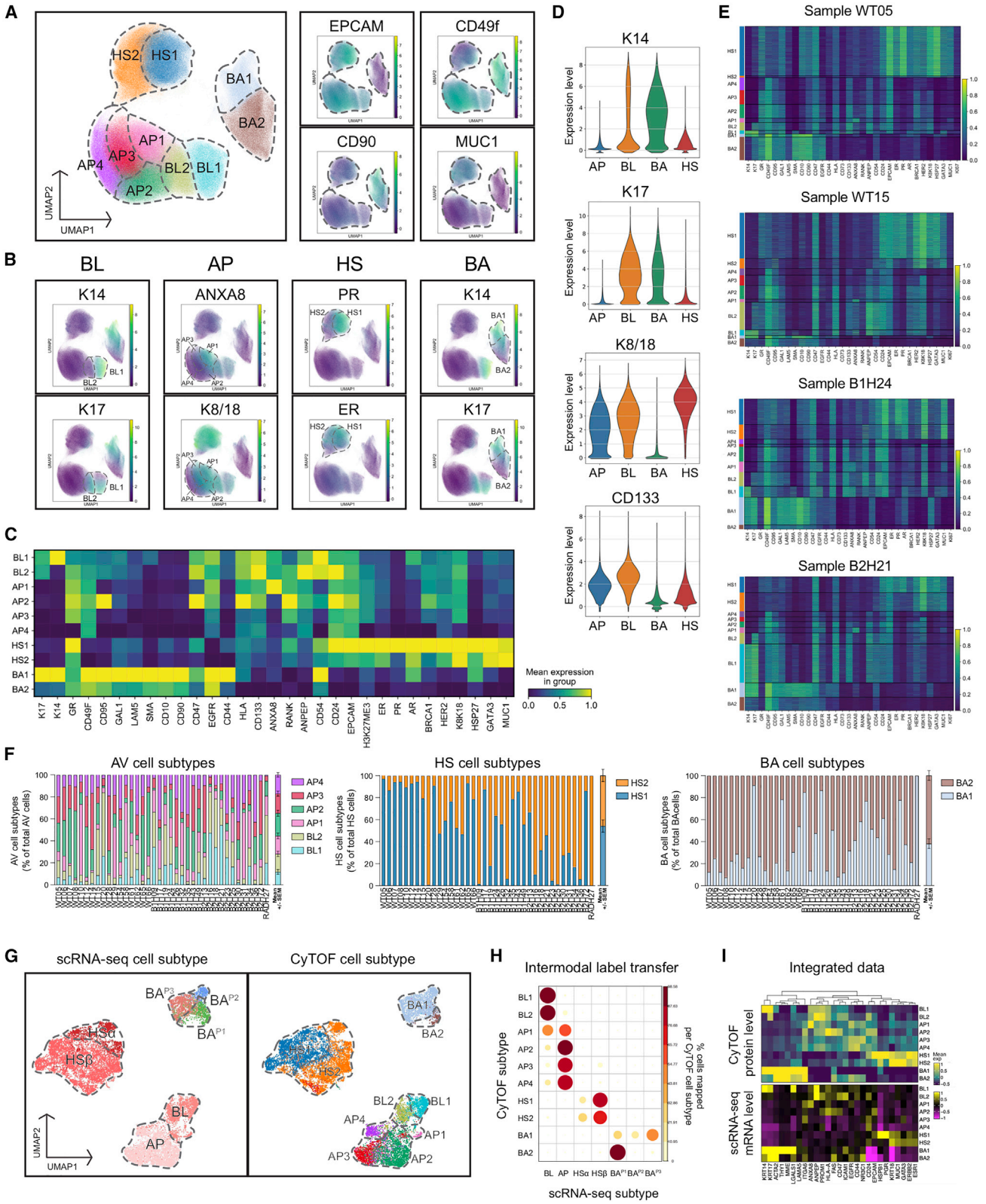
(F) Heatmaps of BL- (left) and AP-specific (right) gene signature expression across MEC subtypes, with BL signature showing substantial overlap with HS or BA genes. Bar graph displays the relative proportion of genes within the BL or AP signatures unique to the subtype or shared with the other lineages.

(G) Violin plots showing example BA markers.

(H) Violin plots showing exemplar marker expression across the indicated cell types, demonstrating BL cells express pan-luminal and AV markers but not some key HS and BA markers.

See also Figure S2.

identified three fibroblast, three vascular/lymphatic, and five immune subtypes whose proportions were highly heterogeneous (Figures S2A–S2D). The fibroblast subtypes (F1–F3) were distinguished by high expression of (1) hormone receptors (HRs; *ESR1*, androgen receptor [*AR*], *PRLR*, and *LEPR*) and cathepsin proteases (*CTSB/D/F*), (2) tubulins (*TUBB2A/B* and *TUBB6*), and (3) fibulins (*FBLN2/5*) and ductal branching regulator *SPRY1* (Koledova et al., 2016) (Figures S2C and S2D). Vascular/lymphatic cells included lymphatic endothelial cells (*LYVE1*, *PDPN*, and *PROX1*), vascular endothelial cells (*SOX17*, *PLVAP*, and *SELE*), and pericytes/smooth muscle cells (*NOTCH3*, *ACTA2*, and *RGS5*) (Figures S2C and S2D) (Kalucka et al., 2020; Li et al., 2020). Immune cells consisted of T cells (CD8$^+$ or CD4$^+$), myeloid cells (*FCGR3A*, *CSF2RA*, and *LYVE1*), plasma cells (*IGKC* or *IGLC2/3/7*), B cells, and natural killer cells (Figures S2C–S2E). These results are consistent with known breast stroma



(legend on next page)

composition (Polyak and Kalluri, 2010). Finally, differential gene expression analyses provided systematically defined cell-subtype-specific signatures (Figure S2D; Table S2).

Although the three MEC types identified via scRNA-seq are consistent with previous work (Henry et al., 2021; Nguyen et al., 2018), sub-clustering performed on each MEC lineage revealed less well-characterized subtypes (Figures 2B and 2C). Interestingly, two AV subtypes, termed basal-luminal (BL) and AV progenitor (AP) cells, were identified. Despite their overall AV phenotype (Figures 2A and S2F), BL cells, unlike AP cells, co-expressed intermediate levels of several BA (*KRT5/14/17* and *CAV1*) and HS markers (*MUC1*, *SERPINA1*, and *AGR2/3*) (Figures 2D–2G and S2F; Table S2). Relative to AP cells, BL cells expressed high levels of the progenitor gene *ALDH1A3* (Colacino et al., 2018; Eirew et al., 2012; Ginestier et al., 2007; Shehata et al., 2012) and low levels of the AV maturity driver *ELF5* (Chakrabarti et al., 2012; Choi et al., 2009; Oakes et al., 2008) (Figures 2D and 2E). Collectively, these patterns suggest the BL state reflects less constrained transcriptional lineage fidelity within the AV lineage. Such features are commonly observed in cancer but rarely in normal tissues (Granit et al., 2018; Seldin and Macara, 2020), raising important questions about the roles of BL cells in development and tumor initiation. Of note, despite expressing several BA and HS markers, BL cells were clearly distinct from these lineages by low/no expression of many canonical BA (*ACTA2* and *MYLK*) and HS genes (*FOXA1* and *AREG*) and by expression of AV markers (*CD14* and *FOLR1*) (Figure 2H). As the BL expression profile was not a simple linear combination of AP, BA, and HS profiles (Figure 2F), BL cells were not multiplets but a bona fide AV subtype. In contrast to BL cells, AP cells were characterized by high lipid and protein biosynthesis regulator expression (Figure 2E; Table S2), suggesting a more committed AV state primed for milk production (Bach et al., 2017).

Two subpopulations (HSα and HSβ) within the HS lineage were identified. Gene set enrichment analysis (GSEA), comparing these subtypes, identified an estrogen-responsive signature as the top distinguishing gene set (Table S2). The HSα subtype expressed higher levels of estrogen receptor (ER, encoded by *ESR1*), ER target genes (*PDZK1*, *SERPINA1*, and *SPDEF*), and selective ER coactivator *CITED1* (McBryan et al., 2007; Yahata et al., 2001) (Figures 2D and 2E; Table S2). HSβ cells highly expressed canonical progesterone receptor (PR) target genes (*KCNK1*, *AKAP13*, and *WNT4*) (Figure 2E; Table S2). Despite differential expression of PR target genes, *PGR* (PR) mRNA levels were not strikingly different between subtypes (Table S2). Overall, these expression patterns suggest HSα and HSβ cells have more active ER and PR signaling, respectively, likely reflecting changes in HS state in response to hormone fluctuations.

Sub-clustering of BA cells identified two subtypes (BAα and BAβ) dominantly distinguished by immediate-early genes (IEGs; *FOS*, *JUN*, *EGR1*, and *IER2*) (Figures 2D and 2E; Table S2). Although the possibility that BA cells express IEGs differentially *in situ* cannot be excluded, a tissue dissociation stress response can produce such differential gene expression among skeletal muscle cell subtypes (van den Brink et al., 2017; van Velthoven et al., 2017), and BA cells are particularly susceptible to alterations during dissociation (Engelbrecht et al., 2021; Gao et al., 2016), potentially due to matrix detachment stress (Debnath and Brugge, 2005). Of note, the BAα cells also exhibited higher expression of contractility genes (*OXTR, MYL9*, *MYH11*, and *FSCN1*) (Figure 2E; Table S2).

To eliminate this possible technical noise from the sub-clustering, regression analysis was performed to remove gene expression variation correlating with a 10-gene IEG signature (see STAR Methods). Re-clustering BA cells post-regression uncovered three clusters (Figure S2G). Differential expression analysis revealed two BA clusters (BA^{P3} and BA^{P2}) displayed higher contractility gene levels (*ACTG2*, *MYL9*, and *OXTR*) (Figure S2H), compared with a third (BA^{P1}). BA^{P3} and BA^{P2} cells also expressed higher levels of cytokine and extracellular matrix genes, respectively (Figure S2H).

The MEC subtypes described above illuminate important aspects of intra-lineage heterogeneity hitherto underappreciated in breast scRNA-seq studies. To verify the robust generalizability of this clustering scheme across datasets with varying sample sizes and tissue processing protocols (Engelbrecht et al., 2021; Gao et al., 2016; Hines et al., 2014), three published scRNA-seq datasets (Hu et al., 2021; Pal et al., 2021; Twigger et al., 2022) were scored with each subtype signature defined in this study, and each cell subtype defined herein was indeed identified in each study (Figures 2I and 2J).

CyTOF delineates complementary and contrasting MEC heterogeneity

Sub-clustering was also performed on MECs within the CyTOF dataset. MECs (n = 716,984) were identified *in silico* and then clustered (see STAR Methods), which partitioned the cells into 10 clusters (Figure 3A). Clusters were annotated based on known marker expression, again identifying HS, BA, and AV cells (Figure 3A) and revealing subtypes within each (Figures 3B–3E). Similarly to the major lineages, the subtypes showed widely varying proportions across individuals (Figures 3E and 3F; Data S1).

Figure 3. Diverse cell subtypes identified by CyTOF and integrated with scRNA-seq

(A) UMAPs of CyTOF clusters of MECs (left) and expression of key AV, HS, and BA markers (right).

(B) Example markers delineate distinct MEC subclusters.

(C) Matrix plot showing marker expression across MEC subclusters.

(D) Violin plots of protein levels of BA (K14, K17), pan-luminal (K8/18), and AV (CD133) markers.

(E) Heatmaps of MEC subtypes in four representative samples.

(F) Within-lineage distribution of cell subtypes across all samples.

(G) UMAPs of the integrated data colored by cell subtypes defined by scRNA-seq (left) or CyTOF (right).

(H) Dot plot of MEC subtype association across modalities after intermodal label transfer from CyTOF to scRNA-seq data. Dot size and color represent the percentage of cells assigned to each integrated cluster.

(I) Matrix plots of CyTOF markers comparing average protein (top) and mRNA (bottom) expression. Plotted values are mean normalized and scaled. See Table S3. See also Figure S3.

CyTOF revealed an AV population expressing K14 and K17 to a level comparable to BA cells, thus mirroring the BL cells identified by scRNA-seq (Figures 3B–3D). These BL cells also specifically expressed CD73, albeit in a small fraction of BL cells from each donor, consistent with scRNA-seq (Figures 3E and S3A). As $CD73^{+}$ luminal MECs display multipotency *in vitro* (Roy et al., 2013), the expression of this marker in BL cells suggests plasticity in this population. In addition, CyTOF delineated further diversity within the AV lineage, including two BL subtypes ($K14^{+}$ $K17^{+/-}$ BL1 and $K14^{-}$ $K17^{+}$ BL2 cells) and four AP subtypes (AP1–AP4; Figures 3B, 3C, and 3E). AP1 cells were distinguished by high ANXA8 levels. AP2 cells highly expressed immunosurveillance-related markers such as HLA-A/B/C (MHC I proteins), CD95 (FAS), and CD47 (Figure 3C). MHC I is upregulated in proliferative cells in various tissues including the mammary gland (Agudo et al., 2018); consistent with this, the AP2 cluster was the most proliferative MEC subtype (Figure S3B). The AP2 cluster also expressed the highest levels of receptor activator of NF-κB (RANK) (Figure 3C), a key mediator of proliferation, differentiation, and transformation in MECs in response to progesterone (Cao et al., 2001; Fata et al., 2000; Mulac-Jericevic et al., 2003; Schramek et al., 2010). The AP3/AP4 clusters were primarily characterized by their low expression of the above markers, with AP4 cells further distinguished by low K8/18 levels.

HS1 cells were differentiated from HS2 cells by higher levels of four HRs, chiefly PR-B but also ER, glucocorticoid receptor (GR), and AR (Figures 3B, 3C, and 3E). This demarcation based on HR levels contrasted with the transcriptomic clusters, which identified diversity centered on HR activity. Two BA subtypes were also identified, with BA1 cells exhibiting high, tightly correlated expression of K14 and K17, compared with BA2 cells (Figures 3B, 3C, and S3C). The degree of correlation between K14 and K17 in BA1 (but not in BL1) cells was unexpected, as these type I acidic keratins do not heterodimerize (Moll et al., 1982). Besides K14/17, BA1 cells exhibited slightly higher levels of other BA markers (CD90, CD10, and SMA) (Figure 3C).

Next, the interrelatedness of the distinct cell subsets defined independently by CyTOF and scRNA-seq was interrogated. Two separate approaches (see STAR Methods) were used to transfer CyTOF subtype labels to scRNA-seq cells and thereby integrate the two modalities. Cells with concordant subtype assignments were selected for further analyses. This stringent approach aligned the major cell types across the two modalities with high accuracy (Figure S3D) and identified MECs in the scRNA-seq closely resembling each CyTOF subtype (Figures 3G and 3H). Examination of the CyTOF marker gene equivalents confirmed that the mRNA and protein levels in each integrated cluster closely mirrored one another (Figures 3I and S3E). These results confirmed that whereas clustering yielded finer or coarser groupings in each modality separately, the axes of variation were concordant such that CyTOF labels separated well onto cell profiles from scRNA-seq (Figure 3G).

Following integration, differential gene expression analyses and GSEA were performed to determine the differences between the integrated cell subtypes (Figure S3F; Table S3). These analyses reinforced many patterns suggested by CyTOF and identified transcriptional profiles/pathways associated with each cluster. For example, BA1 cells were enriched for *KRT14/17* and contractility genes/gene sets, suggesting the $K14^{hi}$ $K17^{hi}$ BA1 cells identified by CyTOF encompass the contractile BA subtypes identified by scRNA-seq. AP2 cells highly expressed several immune-related genes (*CD47*, *TAPBP*, and *CTSS*). AP4 cells displayed a striking enrichment for ribosomal subunit genes and depletion of mitochondrially encoded genes; such a pattern resembles reported poised differentiation or activation states of the hematopoietic system (Athanasiadis et al., 2017; Ricciardi et al., 2018; Wolf et al., 2020). Compared with BL2 cells, BL1 cells expressed higher levels of BA genes (*KRT14/5* and *VIM*), lower levels of AV genes (*ELF5*, *STAT5A*, and *MFGE8*), and higher levels of BL-unique genes (*RARRES2*, *PTN*, and kallikreins), suggesting BL1 cells exhibit particularly high lineage infidelity. *SOX9*, required to induce mixed BL gene expression in normal and precancerous AV cells (Christin et al., 2020; Guo et al., 2012), was also identified as a BL1 marker gene.

HS cells presented an intriguingly more complex portrait post-integration (Figures 3G–3H), likely owing to their dynamic responsiveness to hormone cycles, which may produce elaborate, even discordant, mRNA-protein expression patterns (Cagnet et al., 2018; Métivier et al., 2003; Murrow et al., 2020; Pascual et al., 2020). HS1 cells highly expressed *ESR1*, ER-regulated genes like *PGR*, and ER gene sets (Table S3), as predicted by CyTOF. Cross-comparison of HSα/β and HS1/2 cells demonstrated that cells overlapping between HSα and HS1 (HSα-1) had the highest levels of *ESR1*, *CITED1*, and ER-regulated genes, whereas HSβ-2 had the lowest of these (Figure S3G; Table S3). Interestingly, *GATA3* showed the inverse pattern (Figure S3G), suggesting reciprocal regulation of ER and a key transcriptional partner (Asselin-Labat et al., 2007). GATA3 blocks ER-mediated transcription of HS specifier *SPDEF* (Buchwalter et al., 2013), and accordingly, HSβ-2 cells expressed the lowest *SPDEF* levels (Table S3). HSβ-1 cells expressed high levels of *PTHLH* (Figure S3G), a fibroblast conditioner during hormone-independent murine embryonic mammogenesis (Hiremath and Wysolmerski, 2013). Of note, *PTHLH* expression closely tracked PR target genes *WNT4* and *TNFSF11* (Figure S3G) (Brisken et al., 2000; Lee et al., 2013; Rajaram et al., 2015); future studies may elucidate whether *PTHLH* is hormone-dependent in humans and how this may have contributed to breast evolution, as other *PTHLH* network components were subject to recent evolutionary pressure affecting breast development (Grossman et al., 2010; Kamberov et al., 2013; Sabeti et al., 2007; Southby et al., 1990; Sugimoto et al., 1999; Voutilainen et al., 2012). Although requiring further investigation, these analyses highlight the rich intricacies of this lineage and the power of a multipronged single-cell approach to untangle multilayered phenotypic heterogeneity.

Together, these transcriptomic and mass cytometry data delineated distinct MEC subtypes (Figure S3H) and markedly expanded known cell-subtype heterogeneity of the breast.

Low levels of HS1 cells in older and *BRCA2*^{mut} breast tissue

Previous studies have demonstrated differences in MEC states connected with *BRCA1/2* status, age, and parity (Karaayvaz-Yildirim et al., 2020; LaBarge et al., 2016; Lim et al., 2009; Slepicka et al., 2019), all of which are associated with breast cancer risk. Thus, the correlation of these variables with MEC

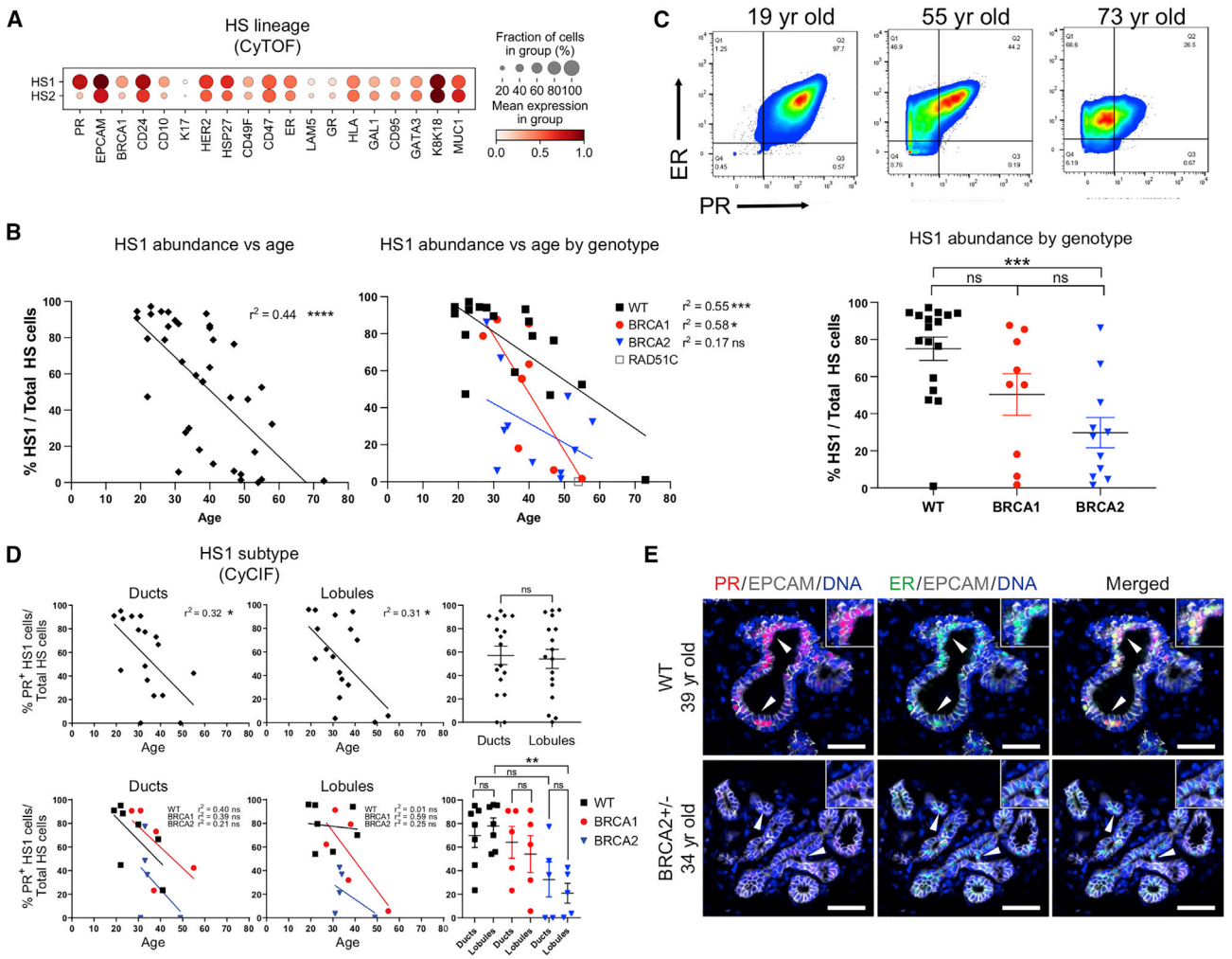


Figure 4. HS subtypes correlate with age and inherited *BRCA2* mutation

(A) Dot plot showing CyTOF marker expression in the HS subtypes.

(B) HS1 abundance by age and genotype from CyTOF (n = 38). Age correlation was performed by simple regression; genotype association analysis, by one-way ANOVA. Error bars represent mean ± SEM.

(C) Biaxial plots of ER/PR in HS cells from three donors of different ages.

(D) CyCIF quantification of HS1 cells in breast sections (n = 17). Top: the percent of ($K19^+$ and/or $ECAD^+$) ER^+ PR^+ HS1 cells out of total ($K19^+$ and/or $ECAD^+$) ER^+ HS cells within ducts or lobules was plotted against donor age (left and middle) and in aggregate (right). Bottom: similar plots showing samples by genotype. Age correlations and error bars are as in (B). Duct/lobule comparisons in aggregate or by genotype were analyzed by paired t test or two-way ANOVA with Bonferroni correction, respectively.

(E) Representative CyCIF of HS1 cells (ER^+ PR^+; arrowheads) in a noncarrier and a $BRCA2^{mut}$ carrier. Scale bars, 50 μm.

See also Figure S4.

types/subtype abundance was explored. These variables were analyzed in the CyTOF dataset to take advantage of its larger sample size and broader representation of ages, parities, and genotypes (Table S1; Figure S3I). Whereas no differences in the abundance of MEC lineages were identified based on these variables (Figure S3J), several MEC subtypes clearly associated with cancer risk factors.

Among HS subtypes (Figure 4A), a decrease in PR^{hi} HS1 cells relative to PR^{lo} HS2 cells was detected with increasing age (Figures 4B and S4A). Biaxial plots of ER/PR for each sample further demonstrated a reduction in PR with age (Figure 4C). This age effect dovetails with a known decrease in PR with age (Yang et al., 2013), consistent with reduced estrogenic signaling at menopause (Endogenous Hormones and Breast Cancer Collaborative Group et al., 2011). Notably, $BRCA2^{mut}$ carriers had fewer HS1 cells independent of age (Figure 4B). Accordingly, PR was the most significant protein expression difference between $BRCA2^{mut}$ noncarrier and carrier HS cells (Figure S4B), even among premenopausal donors with no history of oophorectomy or chemotherapy (Figure S4C; Table S1). Thus, the association with *BRCA2* mutation could not be explained by exogenous hormonal perturbations but instead may be due to intrinsic biological differences in $BRCA2^{mut}$ carriers.

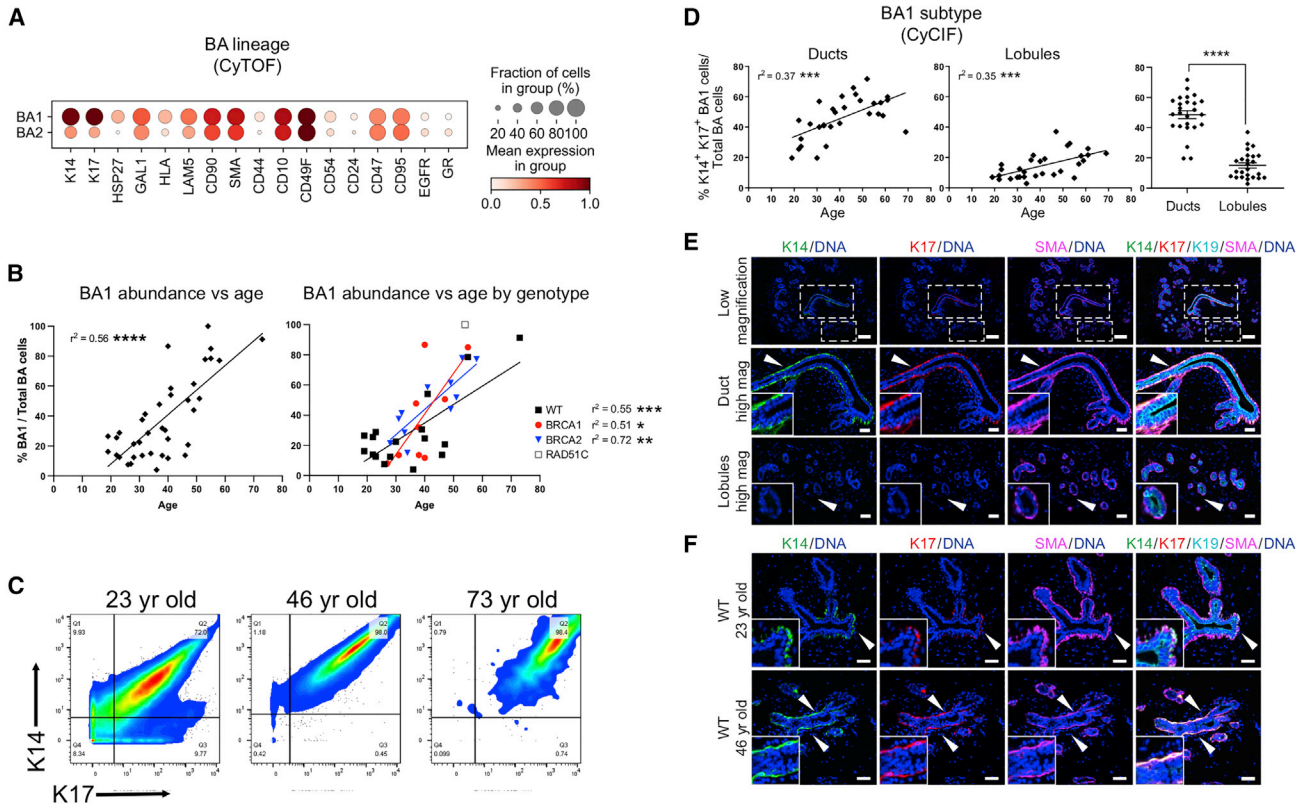


Figure 5. BA subtypes exhibit altered proportion and localization with age

(A) Dot plot showing CyTOF markers expressed in BA subtypes.

(B) BA1 abundance by age and genotype from CyTOF data (n = 38). Age correlation was performed by simple regression.

(C) Biaxial plots of K14/K17 in BA cells of three donors of different ages.

(D) CyCIF quantification of BA1 proportion in breast sections. The percent of BA1 subtype ($K14^+$ $K17^+$ SMA^+ and [$K19^-$ and/or $ECAD^-$]) area out of total BA cell-type (SMA^+ and [$K19^-$ and/or $ECAD^-$])) area within ducts (n = 30) or lobules (n = 30) was plotted against donor age (left and middle) and plotted in aggregate (n = 25) (right). Age correlations are as in (B). Duct versus lobule comparison was analyzed by paired t test. Error bars represent mean ± SEM (related to Figure S4H).

(E) Representative CyCIF of BA1 cells in the ducts and lobules of a young *BRCA1/2*^{mut} noncarrier (age 23), showing preferential localization of BA1 cells ($K14^+$ $K17^+$; arrowheads) within the BA regions (SMA^+) in ducts. Scale bars, 50 μm.

(F) Representative CyCIF of BA1 cells in the ducts of the same young donor shown in (E) compared with an older noncarrier, displaying increased abundance of BA1 cells within the BA region in the older donor. Same markers and scale bars as in (E).

See also Figure S4.

These associations were validated by CyCIF on 17 samples confirmed to be histologically benign. Tissues were iteratively stained for markers (Figure S4D) including ER, PR-B, E-cadherin (ECAD), K19, and EPCAM and were segmented to obtain a total of 11,659 ER^+ HS cells (see STAR Methods). The percentage of ER^+ PR^+ cells out of all ER^+ cells was used to approximate the HS1 cell fraction. These results confirmed that HS1 cell abundance was lower in older persons and *BRCA2*^{mut} carriers across ages (Figures 4D, 4E, and S4E). In addition, HS1 cells were detected in both ducts and lobules, with a significantly lower proportion of HS1 cells in lobules of *BRCA2*^{mut} tissues (Figure 4D). Overall, these results highlight age-dependent changes within the HS lineage and reveal an unexpected difference in the HS cells of *BRCA2*^{mut} carriers.

The BA1/BA2 ratio increases with age

Analysis of the BA clusters (Figure 5A) revealed that the ratio of K14/$K17^{hi}$ BA1 cells to K14/$K17^{lo}$ BA2 cells increased with age but not between genotypes (Figures 5B, S4F, and S4G). Biaxial plots of BA cells also demonstrated a consistent increase in the proportion of BA1 cells in older donors (Figure 5C).

This age association was validated via CyCIF on 30 breast tissues. K14 is more highly expressed in BA cells of ducts versus lobules (Santagata et al., 2014), and CyCIF indeed revealed higher BA K14 and K17 expression in ducts (Figures 5D, 5E, S4H, and S4I). Consistent with CyTOF, CyCIF showed the proportion of $K14^{hi}$ $K17^{hi}$ BA1 cells increased with age (Figures 5D and 5F). Interestingly, this age effect was detected in both structure types, suggesting the age accumulation of BA1 cells cannot be explained entirely by a higher proportion of ductal relative to lobular cells as a result of menopause-associated lobular atrophy. Therefore, these findings reveal a previously unappreciated aspect of age-associated alterations in the BA lineage.

AV subtypes associated with parity, aging, and breast tumor subtypes

The six CyTOF AV subtypes (AP1–AP4, BL1–BL2) (Figure 6A) were assessed with respect to multiple clinical variables

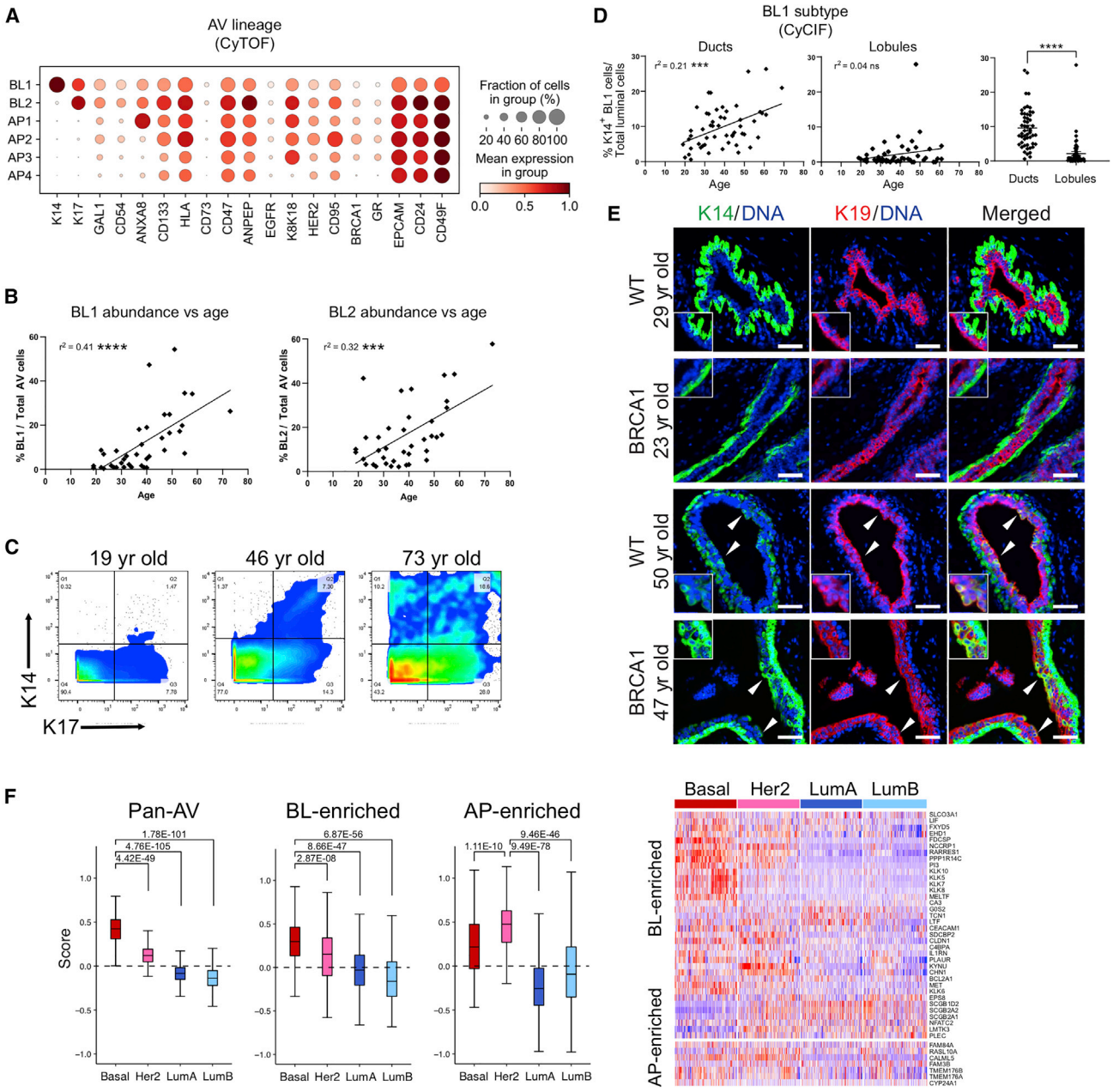


Figure 6. AV subpopulations associated with age, parity, and breast cancer subtypes

(A) Dot plot of CyTOF marker expression in AV subtypes.

(B) BL1/2 abundance by age from CyTOF (n = 38). Age correlation was done by simple regression.

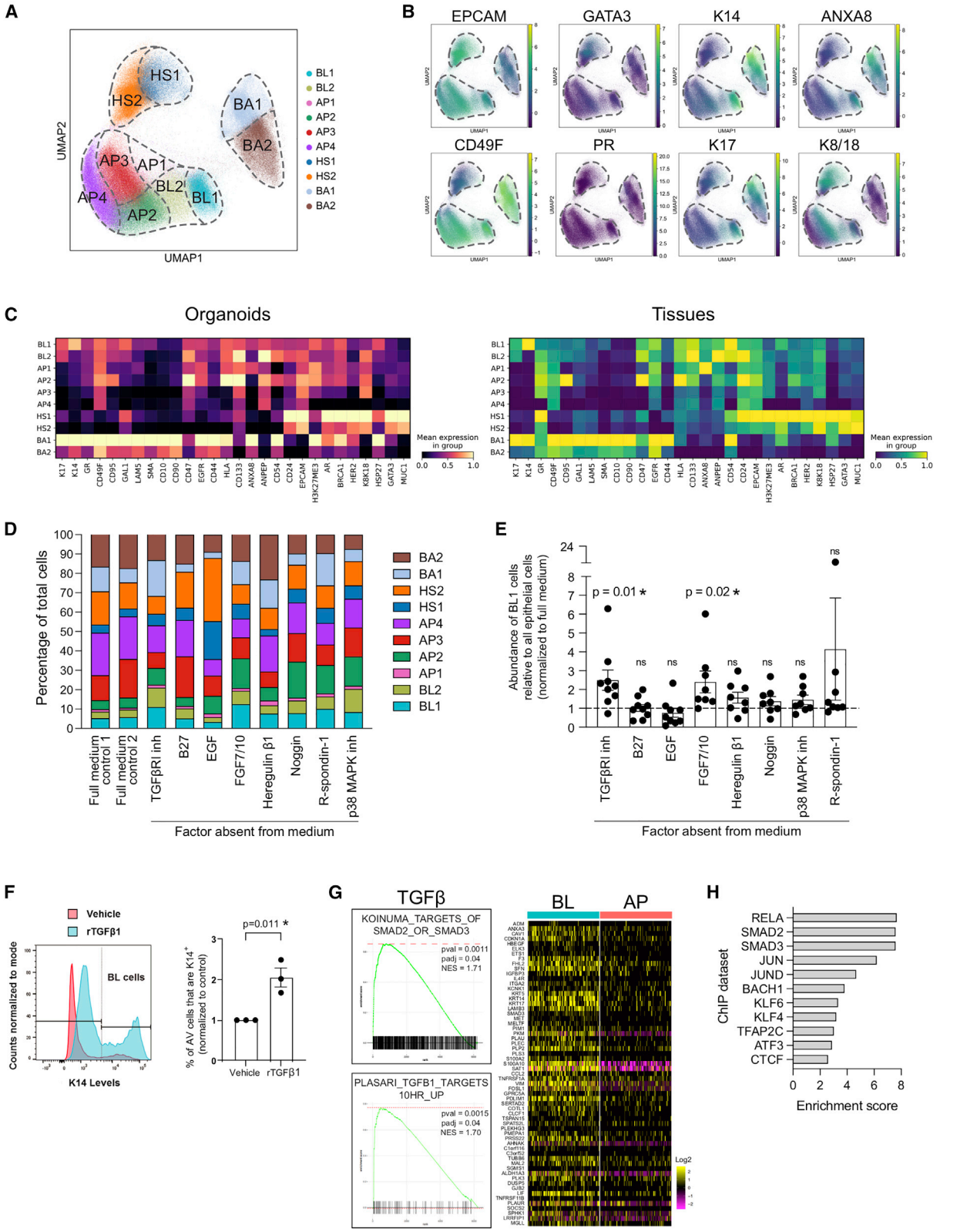
(C) Biaxial plots of K14/17 in AV cells (showing $K14^+$ $K17^{+/-}$ BL1 and $K14^-$ $K17^+$ BL2 cells) of three CyTOF samples from donors of different ages.

(D) CyCIF quantification of BL1 cells in breast sections. The percent of BL1 cells ($K14^+$ and $K19^+$ or $ECAD^+$) out of total luminal cells ($K19^+$ or $ECAD^+$) within ducts (n = 52) or lobules (n = 51) was plotted against donor age (left and middle) and plotted in aggregate for sections with sufficient representation of both structures (n = 51) (right). Age correlations were done by simple regression. Duct/lobule comparison was analyzed by paired t test. Error bars represent mean ± SEM (related to Figure S5E).

(E) Representative examples of BL1 cells ($K14^+$ $K19^+$; arrowheads) in younger and older *BRCA1*^{mut} noncarriers and carriers from CyCIF. Scale bars, 50 μm.

(F) Boxplots (left) showing associations between scRNA-seq signatures of AV cell type/subtypes and transcriptomic profiles of breast cancer subtypes in METABRIC (basal-like, $HER2^+$, and luminal A and B). Boxes represent the first, second (median), and third quartiles; whiskers span 1.5× interquartile range from the first/third quartile. Aggregate and pairwise comparisons among tumor subtypes were analyzed by Kruskal-Wallis and Wilcoxon rank-sum tests, respectively. Heatmap (right) showing gene signature expression across cancer subtypes. Tumors were randomly downsampled to match the subtype with the smallest sample size (related to Figure S6A and Table S4).

See also Figures S5 and S6.



(legend on next page)

(Figures S5A–S5D). For most subtypes, the analysis of parity effects was confounded by strong co-correlation with age. However, the relative proportion of ANXA8$^+$ AP1 cells was significantly higher in parous donors, and this pattern was the opposite of that of age, suggesting independence from this variable in this case (Figures S5B and S5C). ANXA8 marks quiescent AV cells in murine mammary glands after involution (Iglesias et al., 2015), but its parity association in humans has not been characterized. Interestingly, among the top genes specific to AP1 cells (Figure S3F; Table S3) were several milk-related and casein-cluster genes (*CSN1S1*, *FDCSP*, and *PLIN2*), whose expression further highlights this cell subtype's parity association.

CyTOF identified a striking increase in the relative abundance of BL subtypes with age, independent of *BRCA1/2* genotype or parity (Figures 6B, S5A, S5B, and S5D). Biaxial plots further demonstrated enrichment of BL1/2 cells in older donors (Figure 6C). BL1 cells, which displayed a more pronounced increase with age than BL2 cells, were analyzed by CyCIF on two independent breast tissue collections, one local (31 samples yielding 159,393 K19$^+$ and/or ECAD$^+$ luminal cells) and one from the Komen Tissue Bank (21 samples yielding 64,597 K19$^+$ luminal cells). An analysis of the combined set revealed K14$^+$ BL1 cells were localized primarily in ducts and increased in abundance with age (Figures 6D, 6E, S5E, and S5F). Additional immunohistochemistry (IHC) staining of 39 samples also detected K14$^+$ BL1 cells in the luminal layer of ducts in mammary terminal ductal lobular units (TDLUs) and showed that the fraction of TDLUs containing BL cells increased with age (Figure S5G), consistent with the CyTOF and CyCIF analyses.

The age-dependent accumulation of BL cells and their hybrid gene expression evocative of low lineage fidelity suggested changes in the AV differentiation state in older women who face an elevated breast cancer risk. To interrogate the connection between BL cells and breast cancer, tumors from the METABRIC (Curtis et al., 2012) were scored with respect to gene signatures stringently filtered to include only genes most unique to each cell type/subtype (see STAR Methods; Table S4). The BL-unique signature most strongly associated with basal-like breast cancers (Figures 6F and S6A). A similar pattern was observed for a pan-AV signature, whereas an AP signature scored more highly in HER2 tumors. Interestingly, GSEA also identified several basal breast cancer gene sets enriched in BL cells (Figure S6B; Table S2), further underscoring the similarity of BL cells to basal-like breast cancer. By contrast, HS signatures, irrespective of subtype, tightly associated with luminal A/B tumors (Figures S6A and S6C), and BA signatures correlated poorly with every tumor subtype (Figures S6A and S6D). These results support that basal-like breast cancer originates in AV cells (Bach et al., 2021; Gusterson et al., 2005; Molyneux et al., 2010; Smart et al., 2011) and demonstrate that BL-specific genes (*LIF*, *RARRES1*, and *FXYD5*) are more highly expressed in basal-like tumors.

Organoid cultures preserve MEC subtypes for functional interrogation *in vitro*

Mammary organoid cultures preserve the major MEC types *in vitro*, and depleting specific growth factors and signaling inhibitors from the organoid medium alters the relative proportion of lineages in culture (Rosenbluth et al., 2020). Here, the degree to which cell subtypes were maintained in organoid culture was assessed by comparing organoids with primary tissue using CyTOF (see STAR Methods). All MEC subtypes were retained *in vitro*, albeit to varying degrees (Figure 7A), and organoids maintained unique features of the primary tissue MEC subtypes. For example, BA1 cells maintained highly correlated K14 and K17 expression (Figure S7A). Moreover, confocal microscopy detected EPCAM and K14 co-expression in acinar organoids, demonstrating BL cell preservation (Figure S7B). However, organoids had lower PR expression than fresh tissues, with consequently low percentages of HS1 cells (Figures 7A and 7B). Organoid medium lacks estrogens, and its component phenol red bears only mild estrogenic bioactivity (Welshons et al., 1988). Thus, cultures may lack sufficient ER signaling to induce robust PR expression. Additionally, progesterone (present in medium component B27 supplement) downregulates PR expression (Davaadelger et al., 2019), so low PR levels may thus be the result of continuous, rather than cyclic, exposure of MECs to progesterone *in vitro*. Nevertheless, marker expression patterns in organoid subtypes closely mirrored primary tissues overall (Figure 7C).

The effects of medium components on MEC subtypes were investigated by generating a series of organoid media, which each lacked one component, that was used to establish a set of organoids derived from the same tissue for CyTOF analysis (see STAR Methods). Data from six new sets and three previously published ones (Rosenbluth et al., 2020) were used to assess the proportions of cell subtypes under different media conditions (Figures 7D and S7C–S7E). To infer potential direct/indirect effects of these factors, the expression levels of genes

Figure 7. Organoids as a model system to identify signaling pathway dependencies in cell subtypes

(A) UMAP of CyTOF on organoid cultures of normal breast tissues (noncarriers = 4, *BRCA1*^{mut} carriers = 3, *BRCA2*^{mut} carriers = 2). Cells from organoid cultures were projected onto the clusters defined in the primary tissue UMAP (Figure 3A) based on similar protein expression.

(B) UMAPs colored by expression of the indicated markers associated with MEC subtypes.

(C) CyTOF matrix plots showing marker expression of each cell subtype in organoids (left) and primary tissues (right; duplicate of Figure 3C).

(D) Proportion of MEC subtypes grown in full or altered media with removal of each of the indicated components. Mean of 9 cultures is shown.

(E) Relative proportion of BL1 cells in organoids cultured without the indicated media components. Student's t test. Error bars represent mean ± SEM.

(F) Flow cytometry analysis of BL1 cells (K14$^+$ EPCAM^{hi} CD49F^{hi}) as a fraction of total AV cells (EPCAM^{hi} CD49F^{hi}) in organoids cultured with recombinant TGF-β1 (rTGF-β1) or vehicle control (t test). Error bars are as in (E).

(G) TGF-β pathway-related GSEA gene sets enriched in the scRNA-seq BL signature. Heatmap shows BL markers associated with the TGF-β gene sets.

(H) Enrichment score of transcription factors (TFs) with significant overlap (p adj. < 0.05) between their putative target genes curated from published ChIP-seq datasets and the BL signature. Only TFs with detectable expression (>10%) in BL cells are shown. For TFs with multiple enriched ChIP datasets, the highest-scored dataset is shown (related to Figure S7J and Table S2).

See also Figure S7.

targeted by these factors (Figure S7F) and changes in proliferation in cell subtypes under these conditions (Figure S7G) were analyzed.

Removal of individual medium components shifted the distribution of MEC subtypes (Figures 7D, S7C, and S7E). For example, without EGF, both HS subtypes became more prominent (Figures 7D and S7E), likely reflecting an indirect effect of their growth advantage over EGF-dependent AV and BA cells (Ciarloni et al., 2007; Mallepell et al., 2006), which expressed higher EGFR levels than HS cells (Figures 7C and S7F) and proliferated less overall in the absence of EGF (Figure S7G). By contrast, HS cells expressed higher levels of HER2, ERBB3, and ERBB4 (Figures 7C and S7F), and commensurate with this, HS2 (and to a lesser extent HS1) decreased in relative abundance and proliferation in cultures missing heregulin β1 (Figures 7D, S7E, and S7G), a growth factor that binds ERBB3/4 and thereby activates HER2 (Carraway et al., 1994; Plowman et al., 1993). Among BA cells, BA2 cells decreased in the absence of endogenous BMP inhibitor noggin. That BA2 cells displayed a decrease in proliferation (Figure S7G) and expressed BMPR2 (Figure S7F) are consistent with a direct inhibitory effect of BMP signaling on this subtype (Zabala et al., 2020).

Removal of TGF-βRI inhibitor or FGF7/10 significantly increased the percentage of BL cells (Figures 7D, 7E, and S7E), suggesting inverse regulation of BL cells by TGF-β and FGF signaling. BL2 cells also became more abundant upon removal of noggin, reinforcing the notion that TGF-β family signaling regulates BL abundance (Figures 7D and S7E). These increases showed inter-individual heterogeneity (Figure 7E), likely due to variable amounts of TGF-β and FGF family members available in culture (potentially from other cell types or the basement membrane extract lots utilized); however, the regulation of BL cells by TGF-β signaling was confirmed by treating organoids with recombinant TGF-β1, which increased the fraction of $K14^+$ BL1 cells within the AV lineage (Figure 7F). Furthermore, BL1 cells cultured without TGF-βRI inhibitor expressed higher levels of BL markers (K17, GAL1, and ANPEP) (Figure S7H), further supporting a direct role for TGF-βRI in regulating the BL phenotype rather than a proportional increase due to the reduction of other subtypes.

Consistent with these results, GSEA of the scRNA-seq BL signature identified an enrichment of TGF-β pathway activation gene sets (Figure 7G; Table S2). In addition, assessment of TGF-βRI signaling mediators showed increased expression of *SMAD3* (and to a lesser extent *SMAD2*) in BL cells relative to AP cells (Figure S7I), and BL cells expressed the type II receptor *TGFBR2* and low levels of type I receptors *TGFBR1* and *ACVR1B* (Figures S7F and S7I). Considering the pleiotropic role of TGF-β signaling in regulating MEC and breast cancer plasticity (Moses and Barcellos-Hoff, 2011) and the mixed-lineage BL phenotype, the BL signature was evaluated with respect to a ChIP-seq signature database to identify other transcription factors (TFs) possibly active in BL cells (see STAR Methods). This analysis identified nine TFs (e.g., *RELA*, *JUN*, *KLF4*, *ATF3*, and *CTCF*) besides *SMAD2/3* with significant overlap between their target genes and the BL signature (Figure 7H; Table S2) and with detectable expression in BL cells (Figure S7J). Many of these TFs enhance MEC and/or breast cancer stem/progenitor activity (Gu et al., 2021; Guo et al., 2006; Lin et al., 2015; Nishi et al., 2014; Richter et al., 2013; Sau et al., 2016), and *ATF3* and *CTCF* are genetically altered in breast cancer (Curtis et al., 2012; Koboldt et al., 2012). AP-1, NF-κB, and KLF4 also modulate the activity, expression, and pro-survival effects of TGF-β in breast cells (Kwok et al., 2009; Luo, 2017; Tiwari et al., 2013; Yin et al., 2008, 2010). Of note, *ATF3* and inflammatory/NF-κB gene sets are specifically enriched in BL1 versus BL2 cells (Table S3). Although intriguing, further work is needed to verify the role of these TFs in BL cell state maintenance or induction.

Collectively, the studies described herein demonstrate a feasible approach to identify factors regulating cell subtypes by integrating scRNA-seq, mass cytometry, histology, and organoid cultures.

DISCUSSION

Our high-resolution portrait of the breast demonstrates that all three MEC lineages contain multiple subtypes distinguished by specific markers, gene signatures, microanatomical localizations, and sensitivity to growth factor or signaling modulation. Our analyses surveyed tissues encompassing multiple ages, parity and menopause statuses, and cancer-relevant genotypes to comprehensively capture the landscape of distinct cell identities at both the transcriptomic and proteomic levels, providing a generalizable classification of cell subtypes. We also provide evidence that several MEC subtypes are altered in association with breast cancer risk modulators such as age, *BRCA2* mutation, and parity. Finally, we demonstrated organoid cultures can be used to identify factors that differentially regulate MEC subtype proportions. Overall, these findings advance our understanding of breast intra-lineage heterogeneity beyond previous profiles of major cell lineages (Hu et al., 2021; Knapp et al., 2017) or limited cell subtypes (Bhat-Nakshatri et al., 2021; Henry et al., 2021) through our parallel multi-omic approach and broad coverage of clinical variables.

A key contribution of our breast taxonomy is the identification of greater AV diversity than previously appreciated, particularly the robust identification of BL cells across platforms. Although $K14^+$ luminal breast cells have been observed previously (Arendt et al., 2014; Nagle et al., 1986; Wetzels et al., 1991), we provide herein a multifaceted molecular and spatial characterization of these cells. We show BL cells are an AV subset that accumulates predominantly in ducts with age, expresses a subset of BA and HS genes/markers, and thereby displays less restricted lineage fidelity. The properties of BL cells imply increased plasticity, which requires future functional verification. However, $CD73^+$, $ALDH1A3^+$, and TDLU-localized $K14^+K19^+$ MECs (all BL subsets) possess higher *in vitro* plasticity/clonogenicity (Colacino et al., 2018; Roy et al., 2013; Villadsen et al., 2007). Interestingly, genetic manipulation of AV differentiation determinants (*Tnfrsf11a*, *Elf5*, and *Atp6v0a2*, and Notch regulators *Pofut1* and *Rbpsuh*) often results in the accumulation of luminal cells with basal features (Buono et al., 2006; Chakrabarti et al., 2012; Cordero et al., 2016; Pamarthy et al., 2016; Pellegrini et al., 2013), thus suggesting the age accumulation of BL cells in humans may be a consequence of altered AV differentiation cues. Notably though, luminal $K14^+$ MECs are evident in human infants (Anbazhagan et al., 1998), and even our youngest donor tissues (age 19) contained BL cells. Determining the provenance

and physiology of BL cells will be imperative to uncovering the basis for their age accumulation and potential roles in pathology.

Similarly to BL cells, basal-like breast cancers display mixed-lineage features (Badve et al., 2011; Granit et al., 2018; Prat and Perou, 2011) but are luminal in putative origin and overall molecular features (Gusterson, 2009; Santagata and Ince, 2014). Mixed keratin expression is also commonly observed in luminal cells in response to oncogenic insult, even prior to overt transformation (Christin et al., 2020; Koren et al., 2015; Rios et al., 2019; Shakya et al., 2008; Van Keymeulen et al., 2015). As attenuated lineage fidelity is a key property associated with transformation (Ge and Fuchs, 2018) and BL cells share molecular similarities with basal-like breast cancer, these cells represent an important population to study with respect to tumor initiation.

Among AP cells, ANXA8^{+} AP1 cells are of interest due to their parity association. In mice, ANXA8 directly induces quiescent differentiation in post-involution AV cells (Iglesias et al., 2015). That AP1 cells may represent a "memory" of parity is underscored by their expression of casein-cluster genes, which display sustained hypomethylation and expression into murine primiparity (Bach et al., 2017; Dos Santos et al., 2015). This molecular imprint may facilitate the more rapid and high-output lactogenesis of multiparous mammals (Ingram et al., 2001, 1999; Lang et al., 2012; Zuppa et al., 1988). The presence of AP1 cells may be a useful biomarker for studying insufficient human lactation (Farah et al., 2021), and the highly differentiated nature of AP1 cells alongside the largely protective effects of parity against breast cancer (Slepicka et al., 2019) also raises the question of whether AP1 cells resist transformation.

The most proliferative MEC subtype in our CyTOF data is the RANK^{hi} AP2 cells. Their highly proliferative status supports the notion they may be particularly susceptible to malignant transformation (Gonzalez-Suarez et al., 2010; Schramek et al., 2010). Whereas AP2 cells resemble reported proliferative RANK^{+} luminal cells enriched in *BRCA1*^{mut} carriers (Nolan et al., 2016), AP2 cells were not associated with *BRCA1* genotype in our study. Differences in cohorts, methods, and cell-type definitions employed may have influenced genotype associations. A further intriguing aspect of AP2 cells is their high expression of immunosurveillance-related genes/markers. AV subsets may moonlight as surveillants of mastitis-inducing pathogens (Kendrick et al., 2008), and proliferative MECs are more susceptible to immune clearance (Agudo et al., 2018). The possible relationships among RANK, immunosurveillance, and AP proliferation warrant future investigation, especially in the context of tumorigenesis (Gómez-Aleza et al., 2020).

The identification of PR^{hi} HS1 and PR^{lo} HS2 cells via CyTOF is potentially relevant to breast cancer risk. We found that HS1 cells are less common in older donors (likely due to hormone changes) and *BRCA2*^{mut} carriers. The complexity and controversy surrounding the role of progestogens in breast cancer initiation and maintenance (Brisken, 2013, 2014; Carroll et al., 2017; Muti, 2014) extend to the PR expression status of *BRCA1/2*^{mut} breast tissue (King et al., 2004; Mote et al., 2004). However, the pro-differentiation hormonal milieus of gestation and breastfeeding seemingly afford less protection against breast and ovarian cancer for *BRCA2*^{mut} carriers versus noncarriers or *BRCA1*^{mut} carriers (Eoh et al., 2021; Jernström et al., 1999; Kotsopoulos et al., 2012; Milne et al., 2010; Pan et al., 2014; Terry et al., 2018; Tryggvadottir et al., 2003). Progesterone signaling inhibition has been proposed as a prevention strategy for familial breast cancer based on murine studies of *Brca1*-mediated triple-negative tumorigenesis (Gonzalez-Suarez et al., 2010; Lee et al., 2021; Nolan et al., 2017, 2016; Poole et al., 2006; Schramek et al., 2010) and on epidemiological studies combining *BRCA1/2*^{mut} carriers (Hickey, 2013; Widschwendter et al., 2015, 2013), but our data suggest the role of progesterone in breast tumorigenesis may vary by specific genotype. As ER^{+} breast tumors (typically PR$^{lo/-}$ luminal B tumors) predominate over triple-negative tumors in *BRCA2*^{mut} but not *BRCA1*^{mut} carriers (Ha et al., 2017; Larsen et al., 2013; Sorlie et al., 2003), it is critical to clarify the roles of HS2 cells and progesterone in *BRCA2*-associated tumorigenesis.

BA subpopulations exhibit distinct transcriptomic and protein heterogeneity. The scRNA-seq uncovered three BA subtypes distinguished by genes associated with contractility, cytokines, and matrix production. By contrast, CyTOF revealed two BA subtypes consisting of K14^{hi} ductal and K14^{lo} lobular BA cells (Santagata et al., 2014). Integration of the two modalities linked the more contractile BA subtypes identified by scRNA-seq to BA1 cells. Furthermore, we found BA1 cells accumulate within both ducts and lobules with age, suggesting BA cells undergo age-dependent differentiation state changes beyond the effects of lobular atrophy. Throughout human mammogenesis, BA cells in newly formed breast structures display low/no K14/17 expression, reminiscent of BA2 cells (Anbazhagan et al., 1998; Wetzels et al., 1991); thus, the greater prevalence of BA1 cells with age may signify enhanced myoepithelial differentiation in extant structures concurrent with declining production of new structures with age. The decline in organoid BA2 cells in the absence of EGF and noggin supports this hypothesis, as EGF and BMP signaling cascades stimulate and suppress production of new mammary structures, respectively (Ciarloni et al., 2007; Zabala et al., 2020). The contractile signature of BA1 cells also suggests a role of parity (which itself correlates with age) in promoting the BA1 phenotype. Future studies may focus on age-related regulators of BA differentiation and their effects on BA cell tumor suppression (Polyak and Hu, 2005).

Our CyTOF dataset of organoids provides a wealth of information on cell-subtype distribution and marker expression in response to different growth factors and signaling stimuli. The influence of TGF-β on BL cells in organoid culture is particularly noteworthy, as the TGF-β pathway is a vital regulator of mammary development, plasticity, and tumorigenesis (Moses and Barcellos-Hoff, 2011). Various other molecular pathways, such as Wnt and EGF, also affected MEC subtype balance *in vitro* and are crucial mediators of mammary differentiation and disease (Hynes and Watson, 2010; Yu et al., 2016). These results and the hypotheses they prompt demonstrate the power of combining multiple single-cell techniques with small-scale organoid screening, an approach generalizable to studies of other tissues/cancers.

Limitations of the study

Although we discerned some significant correlations herein, confounding variables precluded definitive assessments of others. In many cases, we could not distinguish the effects of age, menopause, surgery type, and genotype. For example, whereas we associate BL cells with aging, others have

connected BL-like phenotypes with *BRCA1/2* status (Proia et al., 2011; Shalabi et al., 2021). Our data suggest a trend toward higher abundance of BL cells in *BRCA1/2*^{mut} carriers (Figure S5A), but the much more profound contribution of age precluded a conclusive evaluation of genotype. Definitively untangling the effects of age and genotype has proven formidable absent multi-institutional efforts to assemble larger datasets. Other relevant covariates, such as surgery type and prior chemotherapy (Wang et al., 2019; Wu et al., 2016), further compound the complexity of the challenge. Such issues have rendered a consensus on the genotype effects on the breast elusive (King et al., 2004; Lim et al., 2009; Mote et al., 2004; Pal et al., 2021). More highly powered observational studies and/or carefully controlled perturbation experiments will be required to parse the independent effects of these intricate, interdependent variables on the breast and the interactions among them.

Finally, although we elucidated common themes across platforms and highlighted the complexity of mRNA/protein differences, our studies were performed in parallel and therefore could not confirm RNA and protein co-expression. Even though *in silico* integration enriched our MEC annotations, technological advances will ideally enable the simultaneous, high-throughput, single-cell detection of RNA and protein in multiple cell compartments (Chung et al., 2021; Stoeckius et al., 2017).

Collectively, our study serves as a benchmark for integrating transcriptomic and mass cytometric analyses to uncover novel, physiologically relevant lineage subtypes in normal and cancer-prone tissues, paving a path for future multi-omic investigations to chart a comprehensive atlas of human tissues.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human mammary tissues
- METHOD DETAILS
  - Tissue dissociation
- SCRNA-SEQ SAMPLE PREPARATION
- CyTOF SAMPLE PREPARATION
- CyTOF DATA ANALYSIS
- INTEGRATION OF CyTOF AND SCRNA-SEQ DATA
- CyCIF TISSUE STAINING
- CyCIF DATA ANALYSIS
- IHC TISSUE STAINING
- ORGANOID CULTURES
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.devcel.2022.05.003.

ACKNOWLEDGMENTS

We gratefully acknowledge the teams who facilitated tissue collection for these studies, including the BWH breast surgery team led by Dr. Tari King; the BWH plastic surgery team, including Drs. Indranil Sinha and Bo Pomahac; the BWH Faulkner pathologists led by Dr. Stephen Pochebit; and the specimen collection group led by Samantha Stokes, Ashka Patel, and the breast biorepository staff. We would like to thank Dr. Michael Steinbaugh (Harvard Chan Bioinformatics Core), Drs. Ayshwarya Subramanian and Matan Hofree (Broad Institute), and members of the AR laboratory (Broad; Danielle Dionne, Lan Nguyen, Julia Waldman, Michael Cuoco, Michal Slyper, Toni Delorey, Christopher Rodman, and Orit Rozenblatt-Rosen) for scRNA-seq assistance. We also thank the DFCI Flow Cytometry Core led by John Daley and Suzan Lazo, the DFCI Mass Cytometry Core led by Nicole Paul and Eric Haas, the HMS Nikon Imaging Center led by Dr. Jennifer Waters, Amgen for providing the RANK antibody (Agreement #2016741159), the HMS LSP (particularly Rumana Rashid), and Dr. Margaret Shipp (DFCI) for providing indium isotopes. We are also grateful for feedback and assistance from all JSB lab members, particularly Francesca Silvestri, Alexis Cook, Griffin Boedicker, and Dr. Yaara Oren. We deeply appreciate invaluable editorial feedback provided by Drs. M. Angelica Martinez-Gakidis (JSB lab) and Marie Bao (HMS). This work was supported in part by an NCI F31 CA228200 (G.K.G.), a Susan G. Komen Postdoctoral Fellowship (C.M.-C.L.), a Croucher Postdoctoral Fellowship (C.M.-C.L.), the American Cancer Society (J.M.R.), the Sidney Farber Scholars Program (J.M.R.), the Terri Brodeur Breast Cancer Foundation (J.M.R.), the Breast Cancer Alliance (J.M.R.), a Clinical Scientist Development Award 2021197 from the Doris Duke Charitable Foundation (J.M.R.), a Komen Scholar Award SAC180002 (J.S.B.), an NCI R35 CA242428 (J.S.B.), an NCI SPORE grant 1P50CA168504 to the Dana-Farber/Harvard Cancer Center (D.A.D.), an NCI U54 CA225088 (P.K.S.), the Ludwig Center at Harvard (P.K.S. and S.S.), the Ludwig Institute for Cancer Research (P.K.S. and S.S.), the Basser Center for BRCA (K.L.N.), the Breast Cancer Research Foundation (K.L.N.), a Gray Foundation Team Sciences Award (J.S.B., S.S., D.A.D., J.-R.L., P.K.S., and K.L.N.), a Goldberg Family Research Fund gift (J.S.B.), a Gray Foundation gift (J.S.B.), and an Anbinder Cancer Research Fund gift (J.S.B.).

AUTHOR CONTRIBUTIONS

C.M.-C.L. harvested tissues with H.K.S. to perform scRNA-seq with support from A.R. L.M.S. and C.M.-C.L. analyzed scRNA-seq data. G.K.G., R.C.J.S., and W.L.G. collected CyTOF samples, and G.K.G. and R.C.J.S. performed CyTOF. G.K.G. analyzed CyTOF data. L.M.S. and G.K.G. integrated CyTOF/scRNA-seq. J.M.R. and L.M.S. performed clinical correlations. J.M.R. and K.M. established organoids and performed organoid experiments. Supported by P.K.S., S.S., and J.S.B., N.G., J.-R.L., W.L.G., H.K.S., and S.M. performed CyCIF. N.G. analyzed CyCIF data. C.M.-C.L. and H.K.S. performed IHC with review by D.A.D. K.D. and K.L.N. performed DNA sequencing and analyses. D.A.D. and J.E.G. orchestrated tissue procurement. D.A.D. performed pathological reviews. G.K.G., C.M.-C.L., J.M.R., L.M.S., N.G., and J.S.B. conceived this study and wrote the manuscript with input from all authors. G.K.G., C.M.-C.L., J.M.R., and L.M.S. contributed equally and are listed alphabetically. They reserve the right to alter the first author order in personal communications.

DECLARATION OF INTERESTS

J.S.B. is a scientific advisory board (SAB) member of Frontier Medicines and eFFECTOR Therapeutics and was an Agios Pharmaceuticals SAB member until January 2022. D.A.D. is on the SAB for Oncology Analytics, Inc., has consulted for Novartis, and receives research support from Canon, Inc. J.E.G. is a paid consultant for Helix and an uncompensated consultant for Konica Minolta and Earli. K.L.N. is a Nest Genomics SAB member. R.C.J.S. and W.L.G. are current employees of Genmab B.V. and Carcell Therapeutics, respectively. There are no conflicts of interest with this work.

P.K.S. is an SAB or Board of Directors member of Applied Biomath, Nanostring, RareCyte Inc., and Glencoe Software, which distributes a commercial version of OMERO. In the last five years, P.K.S. has received research funding from Novartis and Merck. S.S. is a consultant for RareCyte Inc. A.R. is a cofounder and equity holder of Celsius Therapeutics and an equity holder in

Immunitas. She was an SAB member of ThermoFisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics, and Asimov until July 31, 2020. She has been an employee of Genentech since August 1, 2020 and has equity in Roche. P.K.S., S.S., and A.R. declare that these relationships have not influenced this manuscript's content.

G.K.G. is of no relation to the Gray Foundation leadership. J.S.B. is on the Advisory Board of *Developmental Cell*.Inclusion and diversity statement
One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper received support from a program designed to increase minority representation in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: December 2, 2021
Revised: March 23, 2022
Accepted: May 2, 2022
Published: May 25, 2022

REFERENCES

Agudo, J., Park, E.S., Rose, S.A., Alibo, E., Sweeney, R., Dhainaut, M., Kobayashi, K.S., Sachidanandam, R., Baccarini, A., Merad, M., and Brown, B.D. (2018). Quiescent tissue stem cells evade immune surveillance. Immunity *48*, 271–285.e5.

Anbazhagan, R., Osin, P.P., Bartkova, J., Nathan, B., Lane, E.B., and Gusterson, B.A. (1998). The development of epithelial phenotypes in the human fetal and infant breast. J. Pathol. *184*, 197–206.

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

Asselin-Labat, M.L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J., et al. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat. Cell Biol. *9*, 201–209.

Athanasiadis, E.I., Botthof, J.G., Andres, H., Ferreira, L., Lio, P., and Cvejic, A. (2017). Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. Nat. Commun. *8*, 2045.

Bach, K., Pensa, S., Grzelak, M., Hadfield, J., Adams, D.J., Marioni, J.C., and Khaled, W.T. (2017). Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat. Commun. *8*, 2128.

Bach, K., Pensa, S., Zarocsinceva, M., Kania, K., Stockis, J., Pinaud, S., Lazarus, K.A., Shehata, M., Simões, B.M., Greenhalgh, A.R., et al. (2021). Time-resolved single-cell analysis of Brca1 associated mammary tumourigenesis reveals aberrant differentiation of luminal progenitors. Nat. Commun. *12*, 1502.

Badve, S., Dabbs, D.J., Schnitt, S.J., Baehner, F.L., Decker, T., Eusebi, V., Fox, S.B., Ichihara, S., Jacquemier, J., Lakhani, S.R., et al. (2011). Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. Mod. Pathol. *24*, 157–167.

Benz, C.C. (2008). Impact of aging on the biology of breast cancer. Crit. Rev. Oncol. Hematol. *66*, 65–74.

Bhat-Nakshatri, P., Gao, H., Sheng, L., McGuire, P.C., Xuei, X., Wan, J., Liu, Y., Althouse, S.K., Colter, A., Sandusky, G., et al. (2021). A single-cell atlas of the healthy breast tissues reveals clinically relevant clusters of breast epithelial cells. Cell Rep. Med. *2*, 100219.

Blaas, L., Pucci, F., Messal, H.A., Andersson, A.B., Josue Ruiz, E., Gerling, M., Douagi, I., Spencer-Dene, B., Musch, A., Mitter, R., et al. (2016). Lgr6 labels a rare population of mammary gland progenitor cells that are able to originate luminal mammary tumours. Nat. Cell Biol. *18*, 1346–1356.

Brisken, C. (2013). Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. Nat. Rev. Cancer *13*, 385–396.

Brisken, C. (2014). Reply to Is progesterone a neutral or protective factor for breast cancer? Nat. Rev. Cancer *14*, 146.

Brisken, C., Heineman, A., Chavarria, T., Elenbaas, B., Tan, J., Dey, S.K., McMahon, J.A., McMahon, A.P., and Weinberg, R.A. (2000). Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. Genes Dev. *14*, 650–654.

Britt, K., Ashworth, A., and Smalley, M. (2007). Pregnancy and the risk of breast cancer. Endocr. Relat. Cancer *14*, 907–933.

Buchwalter, G., Hickey, M.M., Cromer, A., Selfors, L.M., Gunawardane, R.N., Frishman, J., Jeselsohn, R., Lim, E., Chi, D., Fu, X., et al. (2013). PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. Cancer Cell *23*, 753–767.

Buono, K.D., Robinson, G.W., Martin, C., Shi, S., Stanley, P., Tanigaki, K., Honjo, T., and Hennighausen, L. (2006). The canonical Notch/RBP-J signaling pathway controls the balance of cell lineages in mammary epithelium during pregnancy. Dev. Biol. *293*, 565–580.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cagnet, S., Ataca, D., Sflomos, G., Aouad, P., Schuepbach-Mallepell, S., Hugues, H., Krust, A., Ayyanan, A., Scabia, V., and Brisken, C. (2018). Oestrogen receptor α AF-1 and AF-2 domains have cell population-specific functions in the mammary epithelium. Nat. Commun. *9*, 4723.

Cao, Y., Bonizzi, G., Seagroves, T.N., Greten, F.R., Johnson, R., Schmidt, E.V., and Karin, M. (2001). IKKalpha provides an essential link between RANK signaling and cyclin D1 expression during mammary gland development. Cell *107*, 763–775.

Carraway, K.L., III, Sliwkowski, M.X., Akita, R., Platko, J.V., Guy, P.M., Nuijens, A., Diamonti, A.J., Vandlen, R.L., Cantley, L.C., and Cerione, R.A. (1994). The erbB3 gene product is a receptor for heregulin. J. Biol. Chem. *269*, 14303–14306.

Carroll, J.S., Hickey, T.E., Tarulli, G.A., Williams, M., and Tilley, W.D. (2017). Deciphering the divergent roles of progestogens in breast cancer. Nat. Rev. Cancer *17*, 54–64.

Chakrabarti, R., Wei, Y., Romano, R.A., DeCoste, C., Kang, Y., and Sinha, S. (2012). Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling. Stem Cells *30*, 1496–1508.

Choi, Y.S., Chakrabarti, R., Escamilla-Hernandez, R., and Sinha, S. (2009). Elf5 conditional knockout mice reveal its role as a master regulator in mammary alveolar development: failure of Stat5 activation and functional differentiation in the absence of Elf5. Dev. Biol. *329*, 227–241.

Choudhury, S., Almendro, V., Merino, V.F., Wu, Z., Maruyama, R., Su, Y., Martins, F.C., Fackler, M.J., Bessarabova, M., Kowalczyk, A., et al. (2013). Molecular profiling of human mammary gland links breast cancer risk to a p27(+) cell population with progenitor characteristics. Cell Stem Cell *13*, 117–130.

Christin, J.R., Wang, C., Chung, C.Y., Liu, Y., Dravis, C., Tang, W., Oktay, M.H., Wahl, G.M., and Guo, W. (2020). Stem cell determinant SOX9 promotes lineage plasticity and progression in basal-like breast cancer. Cell Rep. *31*, 107742.

Chung, C.Y., Ma, Z., Dravis, C., Preissl, S., Poirion, O., Luna, G., Hou, X., Giraddi, R.R., Ren, B., and Wahl, G.M. (2019). Single-cell chromatin analysis of mammary gland development reveals cell-state transcriptional regulators and lineage relationships. Cell Rep. *29*, 495–510.e6.

Chung, H., Parkhurst, C.N., Magee, E.M., Phillips, D., Habibi, E., Chen, F., Yeung, B.Z., Waldman, J., Artis, D., and Regev, A. (2021). Joint single-cell measurements of nuclear proteins and RNA in vivo. Nat. Methods *18*, 1204–1212.

Ciarloni, L., Mallepell, S., and Brisken, C. (2007). Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development. Proc. Natl. Acad. Sci. USA *104*, 5455–5460.

Colacino, J.A., Azizi, E., Brooks, M.D., Harouaka, R., Fouladdel, S., McDermott, S.P., Lee, M., Hill, D., Madden, J., Boerner, J., et al. (2018).

Heterogeneity of human breast stem and progenitor cells as revealed by transcriptional profiling. Stem Cell Rep. *10*, 1596–1609.

Cordero, A., Pellegrini, P., Sanz-Moreno, A., Trinidad, E.M., Serra-Musach, J., Deshpande, C., Dougall, W.C., Pujana, M.A., and González-Suárez, E. (2016). Rankl impairs lactogenic differentiation Through inhibition of the prolactin/Stat5 pathway at midgestation. Stem Cells *34*, 1027–1039.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Davaadelger, B., Choi, M.R., Singhal, H., Clare, S.E., Khan, S.A., and Kim, J.J. (2019). BRCA1 mutation influences progesterone response in human benign mammary organoids. Breast Cancer Res. *21*, 124.

Davis, F.M., Lloyd-Lewis, B., Harris, O.B., Kozar, S., Winton, D.J., Muresan, L., and Watson, C.J. (2016). Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. Nat. Commun. *7*, 13053.

Debnath, J., and Brugge, J.S. (2005). Modelling glandular epithelial cancers in three-dimensional cultures. Nat. Rev. Cancer *5*, 675–688.

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. *38*, 737–746.

Ding, L., Su, Y., Fassl, A., Hinohara, K., Qiu, X., Harper, N.W., Huh, S.J., Bloushtain-Qimron, N., Jovanović, B., Ekram, M., et al. (2019). Perturbed myoepithelial cell differentiation in BRCA mutation carriers and in ductal carcinoma in situ. Nat. Commun. *10*, 4182.

Dos Santos, C.O., Dolzhenko, E., Hodges, E., Smith, A.D., and Hannon, G.J. (2015). An epigenetic memory of pregnancy in the mouse mammary gland. Cell Rep. *11*, 1102–1109.

Dou, J., Liang, S., Mohanty, V., Cheng, X., Kim, S., Choi, J., Li, Y., Rezvani, K., Chen, R., and Chen, K. (2020). Unbiased integration of single cell multi-omics data. Preprint at bioRxiv. https://doi.org/10.1101/2020.12.11.422014.

Eirew, P., Kannan, N., Knapp, D.J., Vaillant, F., Emerman, J.T., Lindeman, G.J., Visvader, J.E., and Eaves, C.J. (2012). Aldehyde dehydrogenase activity is a biomarker of primitive normal human mammary luminal cells. Stem Cells *30*, 344–348.

Elias, S., Morgan, M.A., Bikoff, E.K., and Robertson, E.J. (2017). Long-lived unipotent Blimp1-positive luminal stem cells drive mammary gland organogenesis throughout adult life. Nat. Commun. *8*, 1714.

Endogenous Hormones and Breast Cancer Collaborative Group, Key, T.J., Appleby, P.N., Reeves, G.K., Roddam, A.W., Helzlsouer, K.J., Alberg, A.J., Rollison, D.E., Dorgan, J.F., Brinton, L.A., et al. (2011). Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. Br. J. Cancer *105*, 709–722.

Engelbrecht, L.K., Twigger, A.-J., Ganz, H.M., Gabka, C.J., Bausch, A.R., Lickert, H., Sterr, M., Kunze, I., Khaled, W.T., and Scheel, C.H. (2021). A strategy to address dissociation-induced compositional and transcriptional bias for single-cell analysis of the human mammary gland. Preprint at bioRxiv. https://doi.org/10.1101/2021.02.11.430721.

Eoh, K.J., Park, E.Y., Chang, Y.J., Ha, H.I., Hong, J., Huang, D., Nam, E.J., and Lim, M.C. (2021). The preventive effect of breastfeeding against ovarian cancer in BRCA1 and BRCA2 mutation carriers: A systematic review and meta-analysis. Gynecol. Oncol. *163*, 142–147.

Farah, E., Barger, M.K., Klima, C., Rossman, B., and Hershberger, P. (2021). Impaired lactation: review of delayed lactogenesis and insufficient lactation. J. Midwifery Womens. Health *66*, 631–640.

Fata, J.E., Kong, Y.Y., Li, J., Sasaki, T., Irie-Sasaki, J., Moorehead, R.A., Elliott, R., Scully, S., Voura, E.B., Lacey, D.L., et al. (2000). The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. Cell *103*, 41–50.

Fu, N.Y., Nolan, E., Lindeman, G.J., and Visvader, J.E. (2020). Stem cells and the differentiation hierarchy in mammary gland development. Physiol. Rev. *100*, 489–523.

Gao, H., Dong, Q., Chen, Y., Zhang, F., Wu, A., Shi, Y., Bandyopadhyay, A., Daniel, B.J., Huang, C., and Sun, L.Z. (2016). Murine mammary stem/progenitor cell isolation: different method matters? SpringerPlus *5*, 140.

Garbe, J.C., Pepin, F., Pelissier, F.A., Sputova, K., Fridriksdottir, A.J., Guo, D.E., Villadsen, R., Park, M., Petersen, O.W., Borowsky, A.D., et al. (2012). Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia. Cancer Res. *72*, 3687–3701.

Ge, Y., and Fuchs, E. (2018). Stretching the limits: from homeostasis to stem cell plasticity in wound healing and cancer. Nat. Rev. Genet. *19*, 311–325.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555–567.

Giraddi, R.R., Chung, C.Y., Heinz, R.E., Balcioglu, O., Novotny, M., Trejo, C.L., Dravis, C., Hagos, B.M., Mehrabad, E.M., Rodewald, L.W., et al. (2018). Single-cell transcriptomes distinguish stem cell state changes and lineage specification programs in early mammary gland development. Cell Rep. *24*, 1653–1666.e7.

Gómez-Aleza, C., Nguyen, B., Yoldi, G., Ciscar, M., Barranco, A., Hernández-Jiménez, E., Maetens, M., Salgado, R., Zafeiroglou, M., Pellegrini, P., et al. (2020). Inhibition of RANK signaling in breast cancer induces an anti-tumor immune response orchestrated by CD8+ T cells. Nat. Commun. *11*, 6335.

Gonzalez-Suarez, E., Jacob, A.P., Jones, J., Miller, R., Roudier-Meyer, M.P., Erwert, R., Pinkas, J., Branstetter, D., and Dougall, W.C. (2010). RANK ligand mediates progestin-induced mammary epithelial proliferation and carcinogenesis. Nature *468*, 103–107.

Granit, R.Z., Masury, H., Condiotti, R., Fixler, Y., Gabai, Y., Glikman, T., Dalin, S., Winter, E., Nevo, Y., Carmon, E., et al. (2018). Regulation of cellular heterogeneity and rates of symmetric and asymmetric divisions in triple-negative breast cancer. Cell Rep. *24*, 3237–3250.

Grossman, S.R., Shlyakhter, I., Karlsson, E.K., Byrne, E.H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., et al. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. Science *327*, 883–886.

Gu, V.W., Cho, E., Thompson, D.T., Cassady, V.C., Borcherding, N., Koch, K.E., Wu, V.T., Lorenzen, A.W., van der Heide, D.M., White, J.R., et al. (2021). AP-2γ is required for maintenance of multipotent mammary stem cells. Stem Cell Rep. *16*, 106–119.

Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zürrer-Härdi, U., Bell, G., et al. (2012). Slug and Sox9 cooperatively determine the mammary stem cell state. Cell *148*, 1015–1028.

Guo, W., Pylayeva, Y., Pepe, A., Yoshioka, T., Muller, W.J., Inghirami, G., and Giancotti, F.G. (2006). Beta 4 integrin amplifies ErbB2 signaling to promote mammary tumorigenesis. Cell *126*, 489–502.

Gusterson, B. (2009). Do 'basal-like' breast cancers really exist? Nat. Rev. Cancer *9*, 128–134.

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.

Ha, S.M., Chae, E.Y., Cha, J.H., Kim, H.H., Shin, H.J., and Choi, W.J. (2017). Association BRCA Mutation Types Imaging Features Pathologic Findings in patients With breast cancer With BRCA1 and BRCA2 mutations. AJR Am. J. Roentgenol. *209*, 920–928.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.

Henry, S., Trousdell, M.C., Cyrill, S.L., Zhao, Y., Feigman, M.J., Bouhuis, J.M., Aylard, D.A., Siepel, A., and Dos Santos, C.O. (2021). Characterization of gene expression signatures for the identification of cellular heterogeneity in the developing mammary gland. J. Mammary Gland Biol. Neoplasia *26*, 43–66.

Hickey, M. (2013). Untangling BRCA mutations, sex hormones, and cancer risk. Lancet Oncol. *14*, 1151–1152.

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.

Hiremath, M., and Wysolmerski, J. (2013). Parathyroid hormone-related protein specifies the mammary mesenchyme and regulates embryonic mammary development. J. Mammary Gland Biol. Neoplasia *18*, 171–177.

Honeth, G., Schiavinotto, T., Vaggi, F., Marlow, R., Kanno, T., Shinomiya, I., Lombardi, S., Buchupalli, B., Graham, R., Gazinska, P., et al. (2015). Models of breast morphogenesis based on localization of stem cells in the developing mammary lobule. Stem Cell Rep. *4*, 699–711.

Hu, L., Su, L., Cheng, H., Mo, C., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T., Lin, B., et al. (2021). Single-cell RNA sequencing reveals the cellular origin and evolution of breast cancer in BRCA1 mutation carriers. Cancer Res. *81*, 2600–2611.

Hynes, N.E., and Watson, C.J. (2010). Mammary gland growth factors: roles in normal development and in cancer. Cold Spring Harb. Perspect. Biol. *2*, a003186.

Iglesias, J.M., Cairney, C.J., Ferrier, R.K., McDonald, L., Soady, K., Kendrick, H., Pringle, M.A., Morgan, R.O., Martin, F., Smalley, M.J., et al. (2015). Annexin A8 identifies a subpopulation of transiently quiescent c-kit positive luminal progenitor cells of the ductal mammary epithelium. PLoS One *10*, e0119718.

Ingram, J., Woolridge, M., and Greenwood, R. (2001). Breastfeeding: it is worth trying with the second baby. Lancet *358*, 986–987.

Ingram, J.C., Woolridge, M.W., Greenwood, R.J., and McGrath, L. (1999). Maternal predictors of early breast milk output. Acta Paediatr. *88*, 493–499.

Inman, J.L., Robertson, C., Mott, J.D., and Bissell, M.J. (2015). Mammary gland development: cell fate specification, stem cells and the microenvironment. Development *142*, 1028–1042.

Jernström, H., Lerman, C., Ghadirian, P., Lynch, H.T., Weber, B., Garber, J., Daly, M., Olopade, O.I., Foulkes, W.D., Warner, E., et al. (1999). Pregnancy and risk of early breast cancer in carriers of BRCA1 and BRCA2. Lancet *354*, 1846–1850.

Kalucka, J., de Rooij, L.P.M.H., Goveia, J., Rohlenova, K., Dumas, S.J., Meta, E., Conchinha, N.V., Taverna, F., Teuwen, L.A., Veys, K., et al. (2020). Single-cell transcriptome atlas of murine endothelial cells. Cell *180*, 764–779.e20.

Kamberov, Y.G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., Yang, Y., Li, S., Tang, K., Chen, H., et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell *152*, 691–702.

Kanaya, N., Chang, G., Wu, X., Saeki, K., Bernal, L., Shim, H.J., Wang, J., Warden, C., Yamamoto, T., Li, J., et al. (2019). Single-cell RNA-sequencing analysis of estrogen- and endocrine-disrupting chemical-induced reorganization of mouse mammary gland. Commun. Biol. *2*, 406.

Karaayvaz-Yildirim, M., Silberman, R.E., Langenbucher, A., Saladi, S.V., Ross, K.N., Zarcaro, E., Desmond, A., Yildirim, M., Vivekanandan, V., Ravichandran, H., et al. (2020). Aneuploidy and a deregulated DNA damage response suggest haploinsufficiency in breast tissues of BRCA2 mutation carriers. Sci. Adv. *6*, eaay2611.

Kendrick, H., Regan, J.L., Magnay, F.A., Grigoriadis, A., Mitsopoulos, C., Zvelebil, M., and Smalley, M.J. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. BMC Genomics *9*, 591.

King, T.A., Gemignani, M.L., Li, W., Giri, D.D., Panageas, K.S., Bogomolniy, F., Arroyo, C., Olvera, N., Robson, M.E., Offit, K., et al. (2004). Increased progesterone receptor expression in benign epithelium of BRCA1-related breast cancers. Cancer Res. *64*, 5051–5053.

Knapp, D.J.H.F., Kannan, N., Pellacani, D., and Eaves, C.J. (2017). Mass cytometric analysis reveals viable activated caspase-3^{+} luminal progenitors in the normal adult human mammary gland. Cell Rep. *21*, 1116–1126.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., et al. (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Koledova, Z., Zhang, X., Streuli, C., Clarke, R.B., Klein, O.D., Werb, Z., and Lu, P. (2016). SPRY1 regulates mammary epithelial morphogenesis by modulating EGFR-dependent stromal paracrine signaling and ECM remodeling. Proc. Natl. Acad. Sci. USA *113*, E5731–E5740.

Koren, S., Reavie, L., Couto, J.P., De Silva, D., Stadler, M.B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., et al. (2015). PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. Nature *525*, 114–118.

Kotsopoulos, J., Lubinski, J., Salmena, L., Lynch, H.T., Kim-Sing, C., Foulkes, W.D., Ghadirian, P., Neuhausen, S.L., Demsky, R., Tung, N., et al. (2012). Breastfeeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. Breast Cancer Res. *14*, R42.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Kwok, S., Rittling, S.R., Partridge, N.C., Benson, C.S., Thiyagaraj, M., Srinivasan, N., and Selvamurugan, N. (2009). Transforming growth factor-beta1 regulation of ATF-3 and identification of ATF-3 target genes in breast cancer cells. J. Cell. Biochem. *108*, 408–414.

LaBarge, M.A., Mora-Blanco, E.L., Samson, S., and Miyano, M. (2016). Breast cancer beyond the age of mutation. Gerontology *62*, 434–442.

Lang, S.L., Iverson, S.J., and Bowen, W.D. (2012). Primiparous and multiparous females differ in mammary gland alveolar development: implications for milk production. J. Exp. Biol. *215*, 2904–2911.

Larsen, M.J., Kruse, T.A., Tan, Q., Lænkholm, A.V., Bak, M., Lykkesfeldt, A.E., Sørensen, K.P., Hansen, T.V., Ejlertsen, B., Gerdes, A.M., and Thomassen, M. (2013). Classifications within molecular subtypes enables identification of BRCA1/BRCA2 mutation carriers by RNA tumor profiling. PLoS One *8*, e64268.

Lee, H.J., Gallego-Ortega, D., Ledger, A., Schramek, D., Joshi, P., Szwarc, M.M., Cho, C., Lydon, J.P., Khokha, R., Penninger, J.M., and Ormandy, C.J. (2013). Progesterone drives mammary secretory differentiation via RankL-mediated induction of Elf5 in luminal progenitor cells. Development *140*, 1397–1401.

Lee, O., Bosland, M.C., Wang, M., Shidfar, A., Hosseini, O., Xuei, X., Patel, P., Schipma, M.J., Helenowski, I., Kim, J.J., et al. (2021). Selective progesterone receptor blockade prevents BRCA1-associated mouse mammary tumors through modulation of epithelial and stromal genes. Cancer Lett. *520*, 255–266.

Li, C.M., Shapiro, H., Tsiobikas, C., Selfors, L.M., Chen, H., Rosenbluth, J., Moore, K., Gupta, K.P., Gray, G.K., Oren, Y., et al. (2020). Aging-associated alterations in mammary epithelia and stroma revealed by single-cell RNA sequencing. Cell Rep. *33*, 108566.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Lilja, A.M., Rodilla, V., Huyghe, M., Hannezo, E., Landragin, C., Renaud, O., Leroy, O., Rulands, S., Simons, B.D., and Fre, S. (2018). Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. Nat. Cell Biol. *20*, 677–687.

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

Lin, C.C., Sharma, S.B., Farrugia, M.K., McLaughlin, S.L., Ice, R.J., Loskutov, Y.V., Pugacheva, E.N., Brundage, K.M., Chen, D., and Ruppert, J.M. (2015). Kruppel-like factor 4 signals through microRNA-206 to promote tumor initiation and cell survival. Oncogenesis *4*, e155.

Lin, J.R., Izar, B., Wang, S., Yapp, C., Mei, S., Shah, P.M., Santagata, S., and Sorger, P.K. (2018). Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. eLife *7*, e31657.

Luo, K. (2017). Signaling cross talk between TGF-β/Smad and other signaling pathways. Cold Spring Harb. Perspect. Biol. *9*.

Mahendralingam, M.J., Kim, H., McCloskey, C.W., Aliar, K., Casey, A.E., Tharmapalan, P., Pellacani, D., Ignatchenko, V., Garcia-Valero, M.,

Palomero, L., et al. (2021). Mammary epithelial cells have lineage-rooted metabolic identities. Nat. Metab. *3*, 665–681.

Mallepell, S., Krust, A., Chambon, P., and Brisken, C. (2006). Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. Proc. Natl. Acad. Sci. USA *103*, 2196–2201.

McBryan, J., Howlin, J., Kenny, P.A., Shioda, T., and Martin, F. (2007). ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. Oncogene *26*, 6406–6419.

Métivier, R., Penot, G., Hübner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751–763.

Milioli, H.H., Vimieiro, R., Tishchenko, I., Riveros, C., Berretta, R., and Moscato, P. (2016). Iteratively refining breast cancer intrinsic subtypes in the METABRIC dataset. BioData Min. *9*, 2.

Milne, R.L., Osorio, A., Ramón y Cajal, T., Baiget, M., Lasa, A., Diaz-Rubio, E., de la Hoya, M., Caldés, T., Teulé, A., Lázaro, C., et al. (2010). Parity and the risk of breast and ovarian cancer in BRCA1 and BRCA2 mutation carriers. Breast Cancer Res Treat *119*, 221–232.

Moll, R., Franke, W.W., Schiller, D.L., Geiger, B., and Krepler, R. (1982). The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell *31*, 11–24.

Molyneux, G., Geyer, F.C., Magnay, F.A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell *7*, 403–417.

Moses, H., and Barcellos-Hoff, M.H. (2011). TGF-beta biology in mammary development and breast cancer. Cold Spring Harb. Perspect. Biol. *3*, a003277.

Mote, P.A., Leary, J.A., Avery, K.A., Sandelin, K., Chenevix-Trench, G., Kirk, J.A., and Clarke, C.L.; kConFab Investigators (2004). Germ-line mutations in BRCA1 or BRCA2 in the normal breast are associated with altered expression of estrogen-responsive proteins and the predominance of progesterone receptor A. Genes Chromosomes Cancer *39*, 236–248.

Mulac-Jericevic, B., Lydon, J.P., DeMayo, F.J., and Conneely, O.M. (2003). Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. Proc. Natl. Acad. Sci. USA *100*, 9744–9749.

Murrow, L.M., Weber, R.J., Caruso, J.A., McGinnis, C.S., Phong, K., Gascard, P., Borowsky, A.D., Desai, T.A., Thomson, M., Tlsty, T., et al. (2020). Changes in epithelial proportions and transcriptional state underlie major premenopausal breast cancer risks. Preprint at bioRxiv. https://doi.org/10.1101/430611.

Muti, P. (2014). Is progesterone a neutral or protective factor for breast cancer? Nat. Rev. Cancer *14*, 146.

Nagle, R.B., Böcker, W., Davis, J.R., Heid, H.W., Kaufmann, M., Lucas, D.O., and Jarasch, E.D. (1986). Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. J. Histochem. Cytochem. *34*, 869–881.

Nguyen, Q.H., Pervolarakis, N., Blake, K., Ma, D., Davis, R.T., James, N., Phung, A.T., Willey, E., Kumar, R., Jabart, E., et al. (2018). Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nat. Commun. *9*, 2028.

Nishi, M., Sakai, Y., Akutsu, H., Nagashima, Y., Quinn, G., Masui, S., Kimura, H., Perrem, K., Umezawa, A., Yamamoto, N., et al. (2014). Induction of cells with cancer stem cell properties from nontumorigenic human mammary epithelial cells by defined reprogramming factors. Oncogene *33*, 643–652.

Nolan, E., Lindeman, G.J., and Visvader, J.E. (2017). Out-RANKing BRCA1 in mutation carriers. Cancer Res. *77*, 595–600.

Nolan, E., Vaillant, F., Branstetter, D., Pal, B., Giner, G., Whitehead, L., Lok, S.W., Mann, G.B.; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab), and Rohrbach, K., et al. (2016). RANK ligand as a potential target for breast cancer prevention in BRCA1-mutation carriers. Nat. Med. *22*, 933–939.

Oakes, S.R., Naylor, M.J., Asselin-Labat, M.L., Blazek, K.D., Gardiner-Garden, M., Hilton, H.N., Kazlauskas, M., Pritchard, M.A., Chodosh, L.A., Pfeffer, P.L., et al. (2008). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. Genes Dev. *22*, 581–586.

Pal, B., Chen, Y., Vaillant, F., Capaldo, B.D., Joyce, R., Song, X., Bryant, V.L., Penington, J.S., Di Stefano, L., Tubau Ribera, N., et al. (2021). A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J. *40*, e107333.

Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A.C., Wilcox, S., Fu, N., Liu, K.H., Jackling, F.C., et al. (2017). Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat. Commun. *8*, 1627.

Pamarthy, S., Mao, L., Katara, G.K., Fleetwood, S., Kulshreshta, A., Gilman-Sachs, A., and Beaman, K.D. (2016). The V-ATPase a2 isoform controls mammary gland development through Notch and TGF-β signaling. Cell Death Dis. *7*, e2443.

Pan, H., He, Z., Ling, L., Ding, Q., Chen, L., Zha, X., Zhou, W., Liu, X., and Wang, S. (2014). Reproductive factors and breast cancer risk among BRCA1 or BRCA2 mutation carriers: results from ten studies. Cancer Epidemiol. *38*, 1–8.

Pascual, R., Martín, J., Salvador, F., Reina, O., Chanes, V., Millanes-Romero, A., Suñer, C., Fernández-Miranda, G., Bartomeu, A., Huang, Y.S., et al. (2020). The RNA binding protein CPEB2 regulates hormone sensing in mammary gland development and luminal breast cancer. Sci. Adv. *6*, eaax3868.

Pelissier Vatter, F.A., Schapiro, D., Chang, H., Borowsky, A.D., Lee, J.K., Parvin, B., Stampfer, M.R., LaBarge, M.A., Bodenmiller, B., and Lorens, J.B. (2018). High-dimensional phenotyping identifies age-emergent cells in human mammary epithelia. Cell Rep. *23*, 1205–1219.

Pellegrini, P., Cordero, A., Gallego, M.I., Dougall, W.C., Muñoz, P., Pujana, M.A., and Gonzalez-Suarez, E. (2013). Constitutive activation of RANK disrupts mammary cell fate leading to tumorigenesis. Stem Cells *31*, 1954–1965.

Plowman, G.D., Green, J.M., Culouscou, J.M., Carlton, G.W., Rothwell, V.M., and Buckley, S. (1993). Heregulin induces tyrosine phosphorylation of HER4/p180erbB4. Nature *366*, 473–475.

Polyak, K., and Hu, M. (2005). Do myoepithelial cells hold the key for breast tumor progression? J. Mammary Gland Biol. Neoplasia *10*, 231–247.

Polyak, K., and Kalluri, R. (2010). The role of the microenvironment in mammary gland development and cancer. Cold Spring Harb. Perspect. Biol. *2*, a003244.

Poole, A.J., Li, Y., Kim, Y., Lin, S.C., Lee, W.H., and Lee, E.Y. (2006). Prevention of Brca1-mediated mammary tumorigenesis in mice by a progesterone antagonist. Science *314*, 1467–1470.

Prat, A., and Perou, C.M. (2011). Deconstructing the molecular portraits of breast cancer. Mol. Oncol. *5*, 5–23.

Proia, T.A., Keller, P.J., Gupta, P.B., Klebba, I., Jones, A.D., Sedic, M., Gilmore, H., Tung, N., Naber, S.P., Schnitt, S., et al. (2011). Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. Cell Stem Cell *8*, 149–163.

Rajaram, R.D., Buric, D., Caikovski, M., Ayyanan, A., Rougemont, J., Shan, J., Vainio, S.J., Yalcin-Ozuysal, O., and Brisken, C. (2015). Progesterone and Wnt4 control mammary stem cells via myoepithelial crosstalk. EMBO J. *34*, 641–652.

Rashid, R., Gaglia, G., Chen, Y.A., Lin, J.R., Du, Z., Maliga, Z., Schapiro, D., Yapp, C., Muhlich, J., Sokolov, A., et al. (2019). Highly multiplexed immunofluorescence images and single-cell data of immune markers in tonsil and lung cancer. Sci. Data *6*, 323.

Ricciardi, S., Manfrini, N., Alfieri, R., Calamita, P., Crosti, M.C., Gallo, S., Müller, R., Pagani, M., Abrignani, S., and Biffo, S. (2018). The translational machinery of human $CD4^+$ T cells is poised for activation and controls the switch from quiescence to metabolic remodeling. Cell Metab. *28*, 961.

Richter, A., Nissen, N., Mailänder, P., Stang, F., Siemers, F., Kruse, C., and Danner, S. (2013). Mammary gland-derived nestin-positive cell populations can be isolated from human male and female donors. Stem Cell Res. Ther. *4*, 78.

Rios, A.C., Capaldo, B.D., Vaillant, F., Pal, B., van Ineveld, R., Dawson, C.A., Chen, Y., Nolan, E., Fu, N.Y., Jackling, F.C., et al. (2019). Intraclonal plasticity in mammary tumors revealed through large-scale single-cell resolution 3D imaging. Cancer Cell *35*, 618–632.e616.

Rios, A.C., Fu, N.Y., Lindeman, G.J., and Visvader, J.E. (2014). In situ identification of bipotent stem cells in the mammary gland. Nature *506*, 322–327.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rodilla, V., Dasti, A., Huyghe, M., Lafkas, D., Laurent, C., Reyal, F., and Fre, S. (2015). Luminal progenitors restrict their lineage potential during mammary gland development. PLoS Biol. *13*, e1002069.

Rosenbluth, J.M., Schackmann, R.C.J., Gray, G.K., Selfors, L.M., Li, C.M., Boedicker, M., Kuiken, H.J., Richardson, A., Brock, J., Garber, J., et al. (2020). Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. Nat. Commun. *11*, 1711.

Roy, S., Gascard, P., Dumont, N., Zhao, J., Pan, D., Petrie, S., Margeta, M., and Tlsty, T.D. (2013). Rare somatic cells from human breast tissue exhibit extensive lineage plasticity. Proc. Natl. Acad. Sci. USA *110*, 4598–4603.

Russo, J., Rivera, R., and Russo, I.H. (1992). Influence of age and parity on the development of the human breast. Breast Cancer Res. Treat. *23*, 211–218.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913–918.

Santagata, S., and Ince, T.A. (2014). Normal cell phenotypes of breast epithelial cells provide the foundation of a breast cancer taxonomy. Expert Rev. Anticancer Ther. *14*, 1385–1389.

Santagata, S., Thakkar, A., Ergonul, A., Wang, B., Woo, T., Hu, R., Harrell, J.C., McNamara, G., Schwede, M., Culhane, A.C., et al. (2014). Taxonomy of breast cancer based on normal cell phenotype predicts outcome. J. Clin. Invest. *124*, 859–870.

Sau, A., Lau, R., Cabrita, M.A., Nolan, E., Crooks, P.A., Visvader, J.E., and Pratt, M.A. (2016). Persistent activation of NF-κB in BRCA1-deficient mammary progenitors drives aberrant proliferation and accumulation of DNA damage. Cell Stem Cell *19*, 52–65.

Scheele, C.L., Hannezo, E., Muraro, M.J., Zomer, A., Langedijk, N.S., van Oudenaarden, A., Simons, B.D., and van Rheenen, J. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. Nature *542*, 313–317.

Schramek, D., Leibbrandt, A., Sigl, V., Kenner, L., Pospisilik, J.A., Lee, H.J., Hanada, R., Joshi, P.A., Aliprantis, A., Glimcher, L., et al. (2010). Osteoclast differentiation factor RANKL controls development of progestin-driven mammary cancer. Nature *468*, 98–102.

Seldin, L., and Macara, I.G. (2020). DNA damage promotes epithelial hyperplasia and fate mis-specification via fibroblast inflammasome activation. Dev. Cell *55*, 558–573.e6.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

Shakya, R., Szabolcs, M., McCarthy, E., Ospina, E., Basso, K., Nandula, S., Murty, V., Baer, R., and Ludwig, T. (2008). The basal-like mammary carcinomas induced by Brca1 or Bard1 inactivation implicate the BRCA1/BARD1 heterodimer in tumor suppression. Proc. Natl. Acad. Sci. USA *105*, 7040–7045.

Shalabi, S.F., Miyano, M., Sayaman, R.W., Lopez, J.C., Jokela, T.A., Todhunter, M.E., Hinz, S., Garbe, J.C., Stampfer, M.R., Kessenbrock, K., et al. (2021). Evidence for accelerated aging in mammary epithelia of women carrying germline BRCA1 or BRCA2 mutations. Nat. Aging *1*, 838–849.

Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I.A., Avril, S., Prater, M., Eirew, P., Caldas, C., Watson, C.J., and Stingl, J. (2012). Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. Breast Cancer Res. *14*, R134.

Slepicka, P.F., Cyrill, S.L., and Dos Santos, C.O. (2019). Pregnancy and breast cancer: pathways to understand risk and prevention. Trends Mol. Med. *25*, 866–881.

Slepicka, P.F., Somasundara, A.V.H., and Dos Santos, C.O. (2021). The molecular basis of mammary gland development and epithelial differentiation. Semin. Cell Dev. Biol. *114*, 93–112.

Smart, C.E., Wronski, A., French, J.D., Edwards, S.L., Asselin-Labat, M.L., Waddell, N., Peters, K., Brewster, B.L., Brooks, K., Simpson, K., et al. (2011). Analysis of Brca1-deficient mouse mammary glands reveals reciprocal regulation of Brca1 and c-kit. Oncogene *30*, 1597–1607.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. USA *100*, 8418–8423.

Southby, J., Kissin, M.W., Danks, J.A., Hayman, J.A., Moseley, J.M., Henderson, M.A., Bennett, R.C., and Martin, T.J. (1990). Immunohistochemical localization of parathyroid hormone-related protein in human breast cancer. Cancer Res. *50*, 7710–7716.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993–997.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sugimoto, T., Shiba, E., Watanabe, T., and Takai, S. (1999). Suppression of parathyroid hormone-related protein messenger RNA expression by medroxyprogesterone acetate in breast cancer tissues. Breast Cancer Res. Treat. *56*, 11–23.

Terry, M.B., Liao, Y., Kast, K., Antoniou, A.C., McDonald, J.A., Mooij, T.M., Engel, C., Nogues, C., Buecher, B., Mari, V., et al. (2018). The influence of number and timing of pregnancies on breast cancer risk for women With BRCA1 or BRCA2 mutations. JNCI Cancer Spectr. *2*, pky078.

Tiwari, N., Meyer-Schaller, N., Arnold, P., Antoniadis, H., Pachkov, M., van Nimwegen, E., and Christofori, G. (2013). Klf4 is a transcriptional regulator of genes critical for EMT, including Jnk1 (Mapk8). PLoS One *8*, e57329.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.

Tryggvadottir, L., Olafsdottir, E.J., Gudlaugsdottir, S., Thorlacius, S., Jonasson, J.G., Tulinius, H., and Eyfjord, J.E. (2003). BRCA2 mutation carriers, reproductive factors and breast cancer risk. Breast Cancer Res. *5*, R121–R128.

Twigger, A.J., Engelbrecht, L.K., Bach, K., Schultz-Pernice, I., Pensa, S., Stenning, J., Petricca, S., Scheel, C.H., and Khaled, W.T. (2022). Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk. Nat. Commun. *13*, 562.

van den Brink, S.C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., Robin, C., and van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat. Methods *14*, 935–936.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell *174*, 716–729.e27.

Van Keymeulen, A., Fioramonti, M., Centonze, A., Bouvencourt, G., Achouri, Y., and Blanpain, C. (2017). Lineage-restricted mammary stem cells sustain

the development, homeostasis, and regeneration of the estrogen receptor positive lineage. Cell Rep. *20*, 1525–1532.

Van Keymeulen, A., Lee, M.Y., Ousset, M., Brohée, S., Rorive, S., Giraddi, R.R., Wuidart, A., Bouvencourt, G., Dubois, C., Salmon, I., et al. (2015). Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. Nature *525*, 119–123.

van Velthoven, C.T.J., de Morree, A., Egner, I.M., Brett, J.O., and Rando, T.A. (2017). Transcriptional profiling of quiescent muscle stem cells in vivo. Cell Rep. *21*, 1994–2004.

Villadsen, R., Fridriksdottir, A.J., Rønnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M.A., Bissell, M.J., and Petersen, O.W. (2007). Evidence for a stem cell hierarchy in the adult human breast. J. Cell Biol. *177*, 87–101.

Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev. *28*, 1143–1158.

Voutilainen, M., Lindfors, P.H., Lefebvre, S., Ahtiainen, L., Fliniaux, I., Rysti, E., Murtoniemi, M., Schneider, P., Schmidt-Ullrich, R., and Mikkola, M.L. (2012). Ectodysplasin regulates hormone-independent mammary ductal morphogenesis via NF-κB. Proc. Natl. Acad. Sci. USA *109*, 5744–5749.

Wang, C., Christin, J.R., Oktay, M.H., and Guo, W. (2017). Lineage-biased stem cells maintain estrogen-receptor-positive and -negative mouse mammary luminal lineages. Cell Rep. *18*, 2825–2835.

Wang, H., Xiang, D., Liu, B., He, A., Randle, H.J., Zhang, K.X., Dongre, A., Sachs, N., Clark, A.P., Tao, L., et al. (2019). Inadequate DNA damage repair promotes mammary transdifferentiation, leading to BRCA1 breast. Cancer Cell *178*, 135–151.e119.

Welshons, W.V., Wolf, M.F., Murphy, C.S., and Jordan, V.C. (1988). Estrogenic activity of phenol red. Mol. Cell. Endocrinol. *57*, 169–178.

Wetzels, R.H., Kuijpers, H.J., Lane, E.B., Leigh, I.M., Troyanovsky, S.M., Holland, R., van Haelst, U.J., and Ramaekers, F.C. (1991). Basal cell-specific and hyperproliferation-related keratins in human breast cancer. Am. J. Pathol. *138*, 751–763.

Widschwendter, M., Burnell, M., Fraser, L., Rosenthal, A.N., Philpott, S., Reisel, D., Dubeau, L., Cline, M., Pan, Y., Yi, P.C., et al. (2015). Osteoprotegerin (OPG), the endogenous inhibitor of receptor activator of NF-κB ligand (RANKL), is dysregulated in BRCA mutation carriers. EBioMedicine *2*, 1331–1339.

Widschwendter, M., Rosenthal, A.N., Philpott, S., Rizzuto, I., Fraser, L., Hayward, J., Intermaggio, M.P., Edlund, C.K., Ramus, S.J., Gayther, S.A., et al. (2013). The sex hormone system in carriers of BRCA1/2 mutations: a case-control study. Lancet Oncol. *14*, 1226–1232.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

Wolf, T., Jin, W., Zoppi, G., Vogel, I.A., Akhmedov, M., Bleck, C.K.E., Beltraminelli, T., Rieckmann, J.C., Ramirez, N.J., Benevento, M., et al. (2020). Dynamics in protein translation sustaining T cell preparedness. Nat. Immunol. *21*, 927–937.

Wu, A., Dong, Q., Gao, H., Shi, Y., Chen, Y., Zhang, F., Bandyopadhyay, A., Wang, D., Gorena, K.M., Huang, C., et al. (2016). Characterization of mammary epithelial stem/progenitor cells and their changes with aging in common marmosets. Sci. Rep. *6*, 32190.

Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., Centonze, A., Dannau, A., Dubois, C., Van Keymeulen, A., et al. (2018). Early lineage segregation of multipotent embryonic mammary gland progenitors. Nat. Cell Biol. *20*, 666–676.

Yahata, T., Shao, W., Endoh, H., Hur, J., Coser, K.R., Sun, H., Ueda, Y., Kato, S., Isselbacher, K.J., Brown, M., and Shioda, T. (2001). Selective coactivation of estrogen-dependent transcription by CITED1 CBP/p300-binding protein. Genes Dev. *15*, 2598–2612.

Yang, X.R., Figueroa, J.D., Hewitt, S.M., Falk, R.T., Pfeiffer, R.M., Lissowska, J., Peplonska, B., Brinton, L.A., Garcia-Closas, M., and Sherman, M.E. (2013). Estrogen receptor and progesterone receptor expression in normal terminal duct lobular units surrounding invasive breast cancer. Breast Cancer Res. Treat. *137*, 837–847.

Yin, X., Dewille, J.W., and Hai, T. (2008). A potential dichotomous role of ATF3, an adaptive-response gene, in cancer development. Oncogene *27*, 2118–2127.

Yin, X., Wolford, C.C., Chang, Y.S., McConoughey, S.J., Ramsey, S.A., Aderem, A., and Hai, T. (2010). ATF3, an adaptive-response gene, enhances TGF{beta} signaling and cancer-initiating cell features in breast cancer cells. J. Cell Sci. *123*, 3558–3565.

Yu, Q.C., Verheyen, E.M., and Zeng, Y.A. (2016). Mammary development and breast cancer: A Wnt perspective. Cancers (Basel) *8*, 65.

Zabala, M., Lobo, N.A., Antony, J., Heitink, L.S., Gulati, G.S., Lam, J., Parashurama, N., Sanchez, K., Adorno, M., Sikandar, S.S., et al. (2020). LEFTY1 is a dual-SMAD inhibitor that promotes mammary progenitor growth and tumorigenesis. Cell Stem Cell *27*, 284–299.e8.

Zuppa, A.A., Tornesello, A., Papacci, P., Tortorolo, G., Segni, G., Lafuenti, G., Moneta, E., Diodato, A., Sorcini, M., and Carta, S. (1988). Relationship between maternal parity, basal prolactin levels and neonatal breast milk intake. Biol. Neonate. *53*, 144–147.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
ANPEP (151Eu, CyTOF)	Biolegend	Cat#301701; RRID: AB_314177
ANXA8 (161Dy, CyTOF)	R&D Systems	Cat#AF8105; RRID: n/a
CD133 (173Yb, CyTOF)	Miltenyi Biotec	Cat#130-108-062; RRID: AB_2660891
CD73 (170Er, CyTOF)	Biolegend	Cat#344002; RRID: AB_2154067
RANK (175Lu, CyTOF)	Amgen	Cat#N-1H8; RRID: n/a
AR (143Nd, CyTOF)	Cell Signaling Technology	Cat#5153; RRID: AB_10691711
ERα (172Yb, CyTOF)	Cell Signaling Technology	Cat#13258; RRID: AB_2632959
PRβ (164Dy, CyTOF)	Cell Signaling Technology	Cat#3157; RRID: AB_2252606
GATA3 (141Pr, CyTOF)	Miltenyi Biotec	Cat#130-108-061; RRID: AB_2651829
MUC1 (149Sm, CyTOF)	Biolegend	Cat#355602; RRID: AB_2561642
HSP27 (153Eu, CyTOF)	DSHB	Cat#CPTC-HSPB1-3; RRID: AB_2617269
CD10 (163Dy, CyTOF)	Biolegend	Cat#312202; RRID: AB_314913
CD90 (152Sm, CyTOF)	Biolegend	Cat#328101; RRID: AB_940390
K14 (144Nd, CyTOF)	R&D Systems	Cat#MAB3164; RRID: AB_2265623
K17 (166Er, CyTOF)	Cell Signaling Technology	Cat#12509; RRID: AB_2797939
LAM5 (162Dy, CyTOF)	DSHB	Cat#P3H9; RRID: AB_2619590
SMA (160Gd, CyTOF)	eBioscience	Cat#14-9760-80; RRID: AB_2572995
CD24 (158Gd, CyTOF)	Biolegend	Cat#311102; RRID: AB_314851
EPCAM (150Nd, CyTOF)	Biolegend	Cat#324202; RRID: AB_756076
HER2 (176Yb, CyTOF)	Cell Signaling Technology	Cat#2165; RRID: AB_10692490
K8/18 (145Nd, CyTOF)	DSHB	Cat#TROMA-I; RRID: AB_531826
CD49F (155Gd, CyTOF)	Biolegend	Cat#313602; RRID: AB_345296
EGFR (147Sm, CyTOF)	Biolegend	Cat#352901; RRID: AB_10916396
KI67 (146Nd, CyTOF)	Miltenyi Biotec	Cat#130-108-060; RRID: AB_2652564

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
p53 (165Ho, CyTOF)	Cell Signaling Technology	Cat#2524; RRID: AB_331743
BRCA1 (169Tm, CyTOF)	Abcam	Cat# ab16780; RRID: AB_2259338
HLA-A/B/C (167Er, CyTOF)	Biolegend	Cat#311402; RRID: AB_314871
H3K27me3 (175Lu, CyTOF)	Cell Signaling Technology	Cat#9733; RRID: AB_2616029
CD47 (209Bi, CyTOF)	Fluidigm	Cat#3209004B; RRID: n/a
CD54 (142Nd, CyTOF)	Biolegend	Cat#353101; RRID: AB_11204422
CD95 (168Er, CyTOF)	Miltenyi Biotec	Cat#130-108-066; RRID: AB_2660559
GAL1 (171Yb, CyTOF)	R&D Systems	Cat#AF1152; RRID: AB_2136626
GR (156Gd, CyTOF)	Cell Signaling Technology	Cat#3660; RRID: AB_11179215
CD45 (089Y, CyTOF)	Fluidigm	Cat#3089003B; RRID: AB_2661851
EPCR (148Nd, CyTOF)	Biolegend	Cat#351902; RRID: AB_10895923
CD140B (154Sm, CyTOF)	Cell Signaling Technology	Cat#4564; RRID: AB_2236927
VIM (174Yb, CyTOF)	Cell Signaling Technology	Cat#5741; RRID: AB_10695459
CD31 (113In, CyTOF)	Biolegend	Cat#303102; RRID: AB_314328
CD44 (115In, CyTOF)	Biolegend	Cat#103001; RRID: AB_312952
Cleaved PARP1 (159Tb, CyTOF)	eBioscience	Cat#14-6668-80; RRID: AB_10667889
K14 (FITC, CyCIF)	Abcam	Cat#ab77684;RRID: AB_2265437
K19 (AF555, CyCIF)	Abcam	Cat#ab203444; RRID: AB_2857974
K17 (AF647, CyCIF)	Abcam	Cat#ab196199; RRID: AB_2868588
E-cadherin (AF594, CyCIF)	Cell Signaling Technology	Cat#7687; RRID: AB_2797633
SMA (AF555, CyCIF)	Abcam	Cat#ab202509; RRID: AB_2868435
PRβ (AF488, CyCIF)	Abcam	Cat#ab199224; RRID: AB_2728808
ERα (AF647, CyCIF)	Cell Signaling Technology	Cat#57761; RRID: AB_2799533
EPCAM (AF488, CyCIF)	Cell Signaling Technology	Cat#5198; RRID: AB_10692105
K14 (IHC)	BioLegend	Cat#905301; RRID: AB_2565048
Critical Commercial Assays		
Chromium Single Cell 3' Assay	10x Genomics	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
scRNA-seq data	This study	GEO: GSE180878; SRA: SRP329970; Broad Institute Single Cell Portal: Study identifier: SCP1731; Synapse Sage Bionetworks: https://doi.org/10.7303/syn26560310
CyTOF primary tissue raw data	This study	Mendeley Data: https://data.mendeley.com/datasets/pcftzv8w63/1
CyTOF primary tissue processed data	This study	Mendeley Data: https://data.mendeley.com/datasets/vs8m5gkyfn/1
CyTOF organoid data	This study	Mendeley Data: https://data.mendeley.com/datasets/f2v94hj7jm/1
scRNA-seq data	Hu et al., 2021	Data obtained via personal communication with Hu et al.
scRNA-seq data	Pal et al., 2021	GEO: GSE161529
scRNA-seq data	Twigger et al., 2022	Array Express: E-MTAB-9841
Software and Algorithms		
R (v3.5.1, 4.0.1)	R project	RRID: SCR_001905
Cell Ranger (v3)	10x Genomics	RRID:SCR_017344
Seurat (v3, v4)	Butler et al., 2018; Stuart et al., 2019	RRID:SCR_016341
edgeR (v3.24.3)	Robinson et al., 2010	RRID:SCR_012802
fgsea (v 1.8.0)	http://doi.org/10.18129/B9.bioc.fgsea	RRID:SCR_020938
Python (v3.8)	Python Software Foundation	RRID:SCR_008394
FlowJo v10	BD Biosciences	RRID:SCR_008520
MATLAB (v 2020b)	Mathworks	RRID:SCR_001622
Ilastik	Rashid et al., 2019	RRID:SCR_015246
Prism (v9)	Graphpad software	RRID:SCR_002798

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Joan Brugge (joan_brugge@hms.harvard.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw and processed scRNA-seq data are deposited at the Sequence Read Archive (SRA: SRP329970) and Gene Expression Omnibus (GEO: GSE180878), respectively. The scRNA-seq data are also available at the Single Cell Portal with the study identifier SCP1731 and Synapse Sage Bionetworks: https://doi.org/10.7303/syn26560310. CyTOF data are deposited at Mendeley Data (primary tissue raw data, Mendeley Data: https://doi.org/10.17632/pcftzv8w63.1; primary tissue processed data, Mendeley Data: https://doi.org/10.17632/vs8m5gkyfn.1; organoid data, Mendeley Data: https://doi.org/10.17632/f2v94hj7jm.1). Scripts for computational analyses are available as Data S2.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human mammary tissues

All donor samples are listed in Table S1. Specimens were obtained from Brigham & Women's Hospital or Faulkner Hospital on the day of surgery, processed to single cells and analyzed by scRNA-seq on the same day, fixed and frozen as single cells for subsequent

CyTOF analysis, or used to generate organoid cultures or formalin-fixed, paraffin-embedded (FFPE) sections. The scRNA-seq dataset contains 16 samples total, including 4 elective reduction mammoplasties (3 non-mutation carriers and 1 carrying *RAD51C* mutation) and 12 prophylactic mastectomies (6 carrying *BRCA1* mutations and 6 carrying *BRCA2* mutations), spanning a range of 25-65 years of age. The CyTOF dataset contains 38 donors total, including noncarrier (n=17), *BRCA1* (n=9), *BRCA2* (n=11), and *RAD51C* (n=1) mutations, spanning a range of 19-73 years of age. The CyCIF dataset contains 53 samples total, pooling from the Brigham and Faulkner hospitals (n=32) and from the Komen Tissue Bank (n=21), including noncarrier (n=27), *BRCA1* (n=15), and *BRCA2* (n=11), spanning a range of 19-69 years of age. This study was reviewed by the Harvard Medical School Institutional Review Board and deemed not human subjects research. Donors gave their informed consent to have their anonymized tissues used for scientific research purposes. Whole-exome or targeted sequencing was performed on most noncarrier tissues used in scRNA-seq and CyTOF analyses to confirm the absence of germline mutations in common cancer predisposition genes.

METHOD DETAILS

Tissue dissociation

Cells were then pelleted by centrifugation, further dissociated by sequentially pipetting with 10, 5, and 1-ml pipette tips, and recentrifuged. The resulting cell pellet was used directly to establish organoid culture as previously described (Rosenbluth et al., 2020) or further dissociated into single cells for scRNA-seq or CyTOF analysis. For scRNA-seq, cells were treated for 5 minutes with TrypLE (Gibco 12605010), filtered through a 40 μm strainer, resuspended in PBS containing 0.04% BSA, counted manually under the microscope, and loaded onto the 10X Chromium platform immediately. For CyTOF, cells were subjected to red blood cell lysis (Biolegend 420301), a 5-minute treatment with Trypsin (Corning 25053CI), and a 5-minute treatment with 5 U/ml dispase (Stem Cell Technologies 07913) and 0.1 mg/ml DNase (Fisher Scientific NC0208431). Cells were then filtered through a 40 μm strainer, resuspended in PBS, counted manually under the microscope, and used for CyTOF staining/fixation.

scRNA-SEQ SAMPLE PREPARATION

A total of 7,000-8,000 viable cells per sample were loaded for single cell capturing and cDNA library generation using the 10X Chromium 3' library construction kit v2 following the manufacturer's instructions (including 3 washes to minimize ambient RNA). Libraries were sequenced by Illumina HiSeq X Ten. Paired-end reads were processed and mapped to the GRCh38-3.0.0 human genome using Cell Ranger v3.0.

Filtering eliminated genes detected in <20 cells, and discarded cells with (1) UMI counts < 1,000, (2) gene counts < 500 or > 8,000, and (3) mitochondrial gene ratio > 10%, resulting in the detection of 20,437 genes in 52,681 cells (see Table S5) across a total of 16 samples, with a median of approximately 3,000 cells from each sample. A median of 1,756 genes and 6,116 transcripts were captured per cell.

The filtered data were then analyzed using Seurat v3 (Butler et al., 2018; Stuart et al., 2019). Briefly, raw counts were normalized with the SCTransform function with multiple regression variables, including nCount_RNA, percent.mt, S.Score, and G2M.Score. Cells were then clustered using K-nearest neighbor (KNN) graphs and the Louvain algorithm using the first 50 dimensions from principal component analysis. Clustered cells were visualized by UMAP embedding using the default settings in Seurat. Major cell type populations were identified using canonical gene markers. Cell type-specific gene expression signatures for each major population were derived by identifying genes that consistently exhibited expression levels >1.3-fold higher (with FDR < 0.05) than each of the other cell types in multiple pairwise differential expression analyses using edgeR (Robinson et al., 2010). Prior to edgeR, the Seurat FindMarkers function was applied to select features detected in >10% of either cell population in the comparison. To adjust for baseline differences between individuals, all differential analyses were performed with sample donor as a blocking variable. To further identify subpopulations within each major population, $EPCAM^+$ epithelial clusters and $EPCAM^-$ stromal clusters were first subsetted, renormalized, and reclustered. Subsequently each major cell type population was reidentified within the epithelial or stromal subset using canonical markers and was then further individually subsetted, renormalized, and subclustered. In some cases, minor subclusters present in only one or two donors out of 16 were merged with a neighboring subcluster in order to ensure that all subpopulations were generalizable, accounting for ≥10% of cells within the major cell type in at least 3 donors or ≥2% of cells in at least 8 donors. Any minor subclusters with unusually high gene counts (>5,000) and aberrant marker expression were excluded as potential doublets for all downstream analyses. Subpopulation-specific gene expression signatures within each major population were derived using the same pairwise edgeR analysis approach above. All heatmaps for visualizing marker gene expression were median-centered and down-sampled to 100 cells per cell type. Gene set enrichment analysis (GSEA) (Liberzon et al., 2015; Subramanian et al., 2005) was performed on the edgeR-derived cell subtype signatures using the fgsea package and MSigDB v7 collection with FDR cutoff at 0.1. In Figure S2F, gene sets of AV, HS, and BA markers were derived from the respective cell type-specific signatures listed in Table S2.

For reanalysis of published scRNA-seq datasets, raw data were obtained from three published human scRNA-seq studies (Hu et al., 2021; Pal et al., 2021; Twigger et al., 2022). Samples from total normal mammary tissue were reanalyzed using the same methodology described above, except cell subclusters were not merged due to limited sample sizes in some studies. For Hu et al., analysis was limited to a single sample per donor by selecting samples harvested from the contralateral side of cancer occurrence. Analysis was performed using R v4.0.1, Seurat v4.1.0, and ComplexHeatmap v2.4.3.

Each tissue was minced using razor blades and digested in a solution containing DMEM/F12 (Gibco 11330), 2 mM GlutaMax (Gibco 35050), 10 mM HEPES (Gibco 15630), 50 U/mL Penicillin-Streptomycin (Gibco 15070), and 1 mg/ml collagenase XI (Sigma C9407). Tissue digestion was performed at 37°C with constant shaking at 150 rpm, either for 2 hours (for organoids and scRNA-seq) or 2-4 hours (for CyTOF).

For regression analysis of immediate early genes (IEGs), the addModuleScore function in Seurat v3 was applied to the epithelial cells using the top 10 genes differentially expressed between BAα and BAβ cells, all of which are IEGs and are more highly expressed in BAα compared to BAβ (*FOS*, *JUNB*, *ZFP36*, *SOCS3*, *JUN*, *EGR1*, *DUSP1*, *FOSB*, *BTG2*, *IER2*). Raw counts were renormalized as described above except that the IEG module score was added to the list of variables to regress. Subclustering analysis of the post-regression data was performed as described above.

The association of AV, HS, and BA cell type/subtype signatures with breast cancer subtypes was analyzed using tumor microarray data from METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) (Curtis et al., 2012). More stringently refined versions of the signatures for each cell type/subtype described above were derived for each group (see Table S4). Pan-AV signature was defined as genes more highly expressed in AV cells than HS cells and BA cells (FC > 1.5, FDR <0.05), excluding genes differentially expressed between HS and BA cells, and excluding genes differentially expressed between the two AV subtypes. Similarly, pan-HS signature was defined as genes more highly expressed in HS cells than AV and BA cells, not differentially expressed between AV and BA, or between the HS subtypes. Pan-BA signature was defined as genes more highly expressed in BA cells relative to AV and HS but not differentially expressed between AV and HS or between the BA subtypes post-IEG-regression. For cell subtypes, the BL-enriched signature was defined as genes more highly expressed in BL cells relative to AP cells (FC > 1.5, FDR <0.05) and not differentially expressed between AP cells and HS or BA cells. A similar approach was applied to generate the other subtype-specific signatures for the AP, HSα, HSβ subtypes, as well as for the IEG-regressed BA^{P1}, BA^{P2}, and BA^{P3} subtypes. The resulting filtered BA^{P1}-enriched signature contained too few genes and was therefore discarded. Signatures are available in Table S4. Signature scores were defined as the difference between the average expression of genes in the target signature and a randomly sampled control gene set with equivalent expression distribution, as described by (Nguyen et al., 2018). For all other cell type/subtype signatures, correlation score was calculated for each gene signature mapping to the transcriptomic profiles of breast cancer subtypes (basal-like, HER2+, Luminal A, and Luminal B), with p-values calculated using the Kruskal-Wallis test (for aggregate comparison) and Wilcoxon rank sum tests (pairwise comparisons). Tumor molecular subtype designations were from (Milioli et al., 2016).

Enrichr ChEA analysis (Kuleshov et al., 2016) was performed using the Independent Enrichment AnalysisAppyter on the BL signature, the background set of genes detected in >10% of BL or AP cells, and the ChEA_2016 library. The resulting enriched transcription factors were filtered for expression in at least 10% of BL cells.

CyTOF SAMPLE PREPARATION

CyTOF antibody validation and sample preparation was performed as previously described (Rosenbluth et al., 2020). In brief, up to twenty samples at a time were barcoded using the Cell-ID 20-Plex Pd Barcoding Kit (Fluidigm 201060) and then pooled together. These pooled cells were then stained with an extracellular antigen antibody cocktail for 30 min at room temperature and washed twice with Cell-Staining Medium (Fluidigm 201068). They were then permeabilized with 90% ice-cold methanol on ice for 30 minutes, washed three times with CSM, and subsequently stained with an intracellular antigen antibody cocktail for 30 min at room temperature. The cells were subsequently washed twice with CSM and then fixed for 30 min with 4% PFA (EMS 15710) at room temperature. The sample was then stained with the iridium DNA intercalator (Fluidigm 201192A) overnight according to manufacturer's instructions. The following day, the sample was then washed twice with Milli-Q water, resuspended in bead water (1:10 4-Element EQ Beads [Fluidigm 201078] in Milli-Q water), and run on the Helios Mass Cytometer.

CyTOF DATA ANALYSIS

After acquisition of CyTOF data, the data were normalized to the bead signal, converted to an FCS format, and then debarcoded using the CyTOF Software v7 from Fluidigm, thus yielding one file per sample. These samples then underwent quality control hand-gating in Flowjo (v10). Samples were gated on Event Length, Gaussian parameters, 140Ce-Beads, DNA Content (191Ir and 193Ir), and viability stain (195Pt). This gating yielded live single cells, which were then imported into the single-cell analysis platform Scanpy (Wolf et al., 2018). Samples were transformed using the arcsinh transform and downsampled to no more than 20,000 cells per donor.

Analyses of the primary breast tissue were performed in four batches. One batch displayed poor quality staining in all samples and was excluded from downstream analyses. The remaining three batches (n=38 donors; n=50 samples; see Tables S1 and S5) were batch-corrected using the Combat algorithm native to Scanpy and were thereafter analyzed together. The data were embedded as a UMAP using the intersection of all antibodies used in all three experiments. The resulting UMAP was then subjected to Leiden clustering (Traag et al., 2019). Clusters with fewer than 10,000 cells were typically strongly sample-driven and were excluded based on this and their relative rarity. The remaining clusters were used for all downstream analyses. During the analysis of the cluster markers, it was noted that certain antibody channels caused spillover in predicted patterns which created false signal in some circumstances. Notably, K17 created spillover in HLA, and K14 created spillover in K8/18, GATA3, and SMA. Compensation was not feasible at the time the experiments were performed, but these technical artefacts were disregarded during the interpretation of the data.

In order to analyze epithelial cells specifically, each sample was embedded as a t-distributed stochastic neighbor embedding (t-SNE) in FlowJo, and epithelial cells were isolated by hand-gating. Subsequently, no more than 20,000 cells per donor were imported, and the collective data were analyzed in the same manner as the combined epithelial and stromal data described above.

The organoid samples were stained as described above and subjected to the same quality controls. Remnant non-epithelial cells were removed by embedding each batch of organoids as UMAPs and removing non-epithelial clusters from the data frame. All remaining high-quality epithelial cells (see Table S5) were projected onto the primary breast tissue epithelial UMAP using the Ingest algorithm from Scanpy. This approach allowed the organoid samples to be analyzed directly with respect to the primary samples.

INTEGRATION OF CyTOF AND scRNA-SEQ DATA

In order to integrate the CyTOF and scRNA-seq data, the names of CyTOF markers used for clustering were converted to their gene symbol equivalent. Multi-antigen markers K8/18 and HLA-A/B/C were converted to *KRT18* and *HLA-A*, respectively, because those genes most closely mirrored the distribution of the protein data when compared to *KRT8* and *HLA-B/C*. CyTOF data were also converted from arcsinh to log1p space. Subsequently, two methods were utilized to transfer CyTOF cluster labels to the non-regressed scRNA-seq data. In the first, utilizing Scanpy, the MAGIC imputation algorithm (t=4) (van Dijk et al., 2018) was applied to the scRNA-seq data, the Combat algorithm was applied to both datasets, and then Ingest was used to transfer the cluster labels. In the second, the bindSC algorithm (Dou et al., 2020) was applied to the non-imputed sequencing data. Both datasets were normalized and scaled in Seurat v4, with the matrix X set as the protein matrix, Y as the scRNA-seq matrix of 3,000 highly variable genes, Z as the scRNA-seq matrix containing only the protein-homologous genes, α as 0.1, and λ as 0.7. Only those cells with matching cluster designations (n=14,947) were used for the integration results. No doublets were included in the integration analysis. Differential gene expression analyses and GSEA were performed as described above using the original, non-imputed scRNA-seq data without regression of immediate-early genes.

CyCIF TISSUE STAINING

FFPE tissue sections were subjected to multiplex immunofluorescence staining (CyCIF) as previously described (Lin et al., 2018). Briefly, tissue sections were pre-processed using the automated Leica BOND RX instrument before blocking with Intercept (PBS) Blocking Buffer (Li-Cor Cat# 927-70001) for 1 hr at room temperature. Slides were incubated with fluorophore-conjugated primary antibodies overnight, stained with Hoechst 33342 (1 μg/ml) and imaged to determine background fluorescence. Bleaching between cycles was accomplished by incubating slides in a PBS-based solution containing 4.5% H_2O_2 and 30 mM NaOH in the presence of tabletop LED lights for 1 hr at room temperature. After fluorophore inactivation, slides were washed in 1x PBS and subjected to the following cycle of antibody staining. The following antibodies were used: K14 (FITC; Abcam ab77684, RRID:AB_2265437), K19 (AF555; Abcam ab203444, RRID:AB_2857974), K17 (AF647; Abcam ab196199, RRID:AB_2868588), E-cadherin (AF594; Cell Signaling Technology 7687, RRID:AB_2797633), SMA (AF555; Abcam ab202509, RRID:AB_2868435), PRβ (AF488; Abcam ab199224, RRID:AB_2728808), ERα (AF647; Cell Signaling Technology 57761, RRID:AB_2799533), and EPCAM (AF488; Cell Signaling Technology 5198, RRID:AB_10692105). Imaging was performed using the CyteFinder slide scanning fluorescence microscope (RareCyte, Seattle, WA).

CyCIF DATA ANALYSIS

Raw images were stitched, and the multiple cycles were registered to generate an OMETIFF file using ASHLAR (https://github.com/labsyspharm/ashlar). Visual inspection determined which samples had extensive cell loss or tissue damage and had to be excluded from analysis. Epithelial structures were manually extracted using the ROI Manager in ImageJ. Areas of tissue that were damaged, poorly stained, out of focus, or contained debris were not extracted and thus were excluded from further analysis. Subsequently, segmentation was performed using the object classifier Ilastik (ilastik.org) as described previously (Rashid et al., 2019) with some modifications. Approximately three random 250x250 pixel images were sampled from each ROI to construct a training set using MATLAB (version 2020b). These images were inputted into Ilastik, and pixels were assigned as nuclear, cytoplasmic, or background. Once training was complete, Ilastik was used to assign a probability of each pixel in the ROIs to the three object classes defined. Based on these probability masks, a watershed transformation was performed to identify single nuclei using custom MATLAB scripts. Cytoplasmic area was determined by dilating the nuclear mask. Markers restricted to either the nucleus or cytoplasm were used to check segmentation quality by inspecting ks-density plots of those markers in both compartments to confirm enrichment in the correct cellular compartment. Additionally, cytoplasmic and nuclear mask images were overlayed onto the original Hoechst image to qualitatively confirm successful segmentation.

Before quantifying single cells, over- and under-segmented nuclei were removed by applying nuclear perimeter cutoffs based on visual inspection (<10 and >=30 pixels). Exclusion of the lowest 5% of Hoechst signal removed out-of-focus nuclei. Thresholds for markers were set based on intensity distribution across cells in each ROI and a visual comparison of the immunofluorescence image with a centroid graph labeling positive and negative cells in the ROI. Once thresholds were established, cell populations defined by specific marker combinations were counted for each ROI, and the ROIs were aggregated for each sample to yield a percentage for each population in ducts and lobules in each tissue section. The above approach was used for analysis of

AV cells and HS cells. However, because the BA cell population did not segment well using the above approach due to their morphology, an ImageJ script was instead applied to calculate the percentage of $K14^+$ $K17^+$ SMA^+ and ($K19^-$ and/or $ECAD^-$) area out of SMA^+ and ($K19^-$ and/or $ECAD^-$) area for each ROI and then aggregated the ROIs to determine a single percentage for all ducts and lobules per sample.

IHC TISSUE STAINING

FFPE tissue sections (n=39) were stained with antibody against K14 (BioLegend 905301, RRID:AB_2565048) at 1:2000 overnight, then probed with biotinylated anti-rabbit secondary antibody (Vector Labs BA-1000) and avidin/biotin peroxidase (Vector Labs PK-6100), and visualized by 3,3′-Diaminobenzidine (DAB, Sigma D4418). The brown stain was allowed to develop without over-saturation, such that any difference in low versus high signal intensity was discernable. On average, 13 (range 5-44) terminal ducts in TDLUs per patient were scored for $K14^+$ BL1 cells in the luminal layer by a breast pathologist in a blinded fashion.

ORGANOID CULTURES

Organoids were cultured as previously described (Rosenbluth et al., 2020). In growth factor perturbation experiments, one of the following components was removed from the culture medium: the TGFβRI inhibitor A83-01, B27 supplement, EGF, FGF7/10, Heregulin β1, Noggin, R-spondin 1, or the p38 MAPK inhibitor SB202190. Data from three previously published organoid cultures (Rosenbluth et al., 2020) were re-analyzed and results were combined with six additional organoid cultures that were newly generated. One culture (ORG82) was only grown in a subset of the different media formulations (full medium, -EGF, -B27 supplement, and -TGFβRI inhibitor) due to insufficient tissue. Where indicated, 20 ng/mL recombinant TGFβ (Peprotech) were added to the culture medium for 14 days. Flow cytometry analysis and confocal microscopy of immunostained organoid cultures were performed as previously described (Rosenbluth et al., 2020).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details in this study, including the statistical tests and number of samples used, are listed in the figure legends. Unless otherwise specified, group comparisons were performed using t-test when two groups were compared, and one-way or two-way ANOVA when three or more groups were compared. Age correlations were analyzed by simple regression. Statistical analyses were performed by using the GraphPad Prism software or R. All bar graphs represent mean +/- SEM. All *p-value*s were denoted as * for $p \leq 0.05$, ** for $p \leq 0.01$, *** for $p \leq 0.001$, and **** for $p \leq 0.0001$. For all statistical analyses and unless otherwise specified, *p-value*s < 0.05 were deemed significant.

ADDITIONAL RESOURCES

scRNA-seq data: https://singlecell.broadinstitute.org/single_cell (Study identifier: SCP1731); https://doi.org/10.7303/syn26560310CyTOF primary - --- tissue raw data: https://data.mendeley.com/ (https://doi.org/10.17632/pcftzv8w63.1)
CyTOF primary tissue processed data: https://data.mendeley.com/ (https://doi.org/10.17632/vs8m5gkyfn.1)
CyTOF organoid data: https://data.mendeley.com/ (https://doi.org/10.17632/f2v94hj7jm.1).