Virtanen et al. Breast Cancer Research (2021) 23:66
https://doi.org/10.1186/s13058-021-01444-5

Breast Cancer Research

RESEARCH ARTICLE Open Access

High-throughput surface marker screen on primary human breast tissues reveals further cellular heterogeneity

Siru Virtanen[1], Reiner Schulte[2], John Stingl[1], Carlos Caldas[1,3] and Mona Shehata[1,4*]

Abstract

Background: Normal human breast tissues are a heterogeneous mix of epithelial and stromal subtypes in different cell states. Delineating the spectrum of cellular heterogeneity will provide new insights into normal cellular properties within the breast tissue that might become dysregulated in the initial stages of cancer. Investigation of surface marker expression provides a valuable approach to resolve complex cell populations. However, the majority of cell surface maker expression of primary breast cells have not been investigated.

Methods: To determine the differences in expression of a range of uninvestigated cell surface markers between the normal breast cell subpopulations, primary human breast cells were analysed using high-throughput flow cytometry for the expression of 242 cell surface proteins in conjunction with EpCAM/CD49f staining.

Results: We identified 35 surface marker proteins expressed on normal breast epithelial and/or stromal subpopulations that were previously unreported. We also show multiple markers were equally expressed in all cell populations (e.g. CD9, CD59, CD164) while other surface markers were confirmed to be enriched in different cell lineages: CD24, CD227 and CD340 in the luminal compartment, CD10 and CD90 in the basal population, and CD34 and CD140b on stromal cells.

Conclusions: Our dataset of CD marker expression in the normal breast provides better definition for breast cellular heterogeneity.

Keywords: Normal breast, Surface markers, Breast epithelial cells, Stromal, Luminal progenitor, Antibody screen

Background

The human breast is a complex steroid-responsive organ which undergoes morphological and structural changes depending on the reproductive stage. The breast epithelium is composed of two known cell types, an outer layer of myoepithelial/basal cells and an inner luminal layer composed of separate secretory and hormone receptor-positive populations. These populations are organised into a series of ductal networks, surrounded by stromal cells and adipocytes [1–3]. This breast network is structured via a main stem or primary duct ending in a cluster of sac-like lobules termed terminal ductal lobular units (TDLUs). The origins and development of breast cancer revealed that most breast cancers originate from a single TDLUs [4]. Historically mammographic and histology analyses were limited in defining the exact cell compartment responsible for neoplastic transformation. Reliance on immunostaining for specific keratin (K) markers classifying breast cell types has led to discrepancy. K5 and K14 are often referred to as basal keratins based on their expression in the mouse mammary gland, specifically within the basal layer of the ducts, yet they

* Correspondence: ms2140@cam.ac.uk
[1]CRUK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK
[4]Medical Research Council Cancer Unit, University of Cambridge, Hills Road, Cambridge CB2 0XZ, UK
Full list of author information is available at the end of the article

were also expressed within the luminal layer of TDLUs of human breast tissues, therefore making cell identity difficult to interpret using these markers [5, 6]. A better understanding of the cellular heterogeneity existing in the breast epithelium and different cell states provides useful clues to how these cell types transform into the distinct breast cancer subtypes.

Many studies have relied on in vitro and in vivo assays to understand the hierarchical organisation and the progenitor/stem capacities of breast epithelial cells. One of the earliest studies used a combination of cell surface markers including EpCAM (ESA), CD10, CD49f (Integrin α6) and MUC1 (CD227) to identify the basal and luminal populations via flow cytometry [7, 8]. Subsequently, different cell isolation protocols and cell surface marker combinations were utilised to identify dissimilar subpopulations adding to the complexity with minimal overlap between studies [9–11]. Currently, the combination of two key cell surface markers, EpCAM and CD49f, are widely used as differentiation markers to identify the basal, luminal progenitor (LP), mature luminal (ML) and stromal compartments of the normal breast [12–14]. Investigating breast cellular heterogeneity has taken a leap forward with the enhancement of single cell omic studies. Single cell transcriptome analysis of primary human breast tissue confirmed the three main epithelial cell types and has highlighted that there are additional cell states within each cell population [15, 16]. The different cell states are essential to predicting a cellular trajectory hierarchy. Validating these novel cell states is problematic due to technical difficulties in isolating viable live cells based on their transcriptomic profile. The cell surface proteome is central to many biological functions which reflect cell fate, yet expression patterns of many cell surface markers in the human breast cell subpopulations are poorly defined.

Here, we identify specific CD marker expression patterns within the breast epithelium and stromal cell populations to generate a searchable dataset. We developed an analysis platform using standard flow cytometry and multiplexing for the simultaneous examination of epithelial and stromal cell populations. This protocol allowed us to identify and quantify the abundance of hundreds of CD markers on single cell suspensions of reduction mammoplasty specimens. Our data presents opportunities for new antibody panels that focus on stricter definitions of the cellular states of the human breast. Characterisation of CD proteins expressed by each breast subpopulation is informative as it will not only improve cell state classifications but may also provide insights into biological function.

Methods

Dissociation of human mammary tissue

All primary human materials were derived from reduction mammoplasties at Addenbrookes Hospital, Cambridge, UK, under full informed consent and in accordance with the National Research Ethics Service, Cambridgeshire 2 Research Ethics Committee approval (08/H0308/178) as part of the Adult Breast Stem Cell Study. All tissue donors had no previous history of cancer and were premenopausal (37–43 years old). Reduction mammoplasty specimens were transferred from the operating room on ice in sterile DMEM/F12 1:1 media (Invitrogen) supplemented with 5% FBS (Gibco/Invitrogen). Tissues were dissociated into single cell suspension as described previously [17]. Briefly, tissue was manually minced and incubated in DMEM/F12 1:1 medium with 10 mM Hepes plus 2% BSA, 5 μg/ml insulin (Invitrogen), 50 μg/ml gentamycin, 300 U/ml collagenase (Sigma) and 100 U/ml hyaluronidase (Sigma) with gentle shaking at 37 °C, overnight or for 16 h. Tissue fragments were harvested by washing with DMEM/F12 and spinning at 450*g* for 5 min at 4 °C. Fragments were triturated in trypsin-EDTA (0.25%; Stem Cell Technologies) for 2–3 min following a red blood cell lysis using ammonium chloride solution (Stem Cell Technologies). Cells were then washed in HBSS without calcium or magnesium, supplemented with 2% FBS, and centrifuged. Cells were then triturated in dispase 5 U/ml and 50 μg/ml DNase I for 1 min, followed by a final wash in HBSS plus 2% FBS and centrifuged.

Surface protein screening using lyoplate technology

Single cell suspensions from two human mammary reductions were pooled together and analysed using a commercial antibody screen, the BD Lyoplate™ Human Cell Surface Marker Screening Panel (BD Biosciences), containing AlexaFluor®647-conjugated antibodies with specificity for 242 cell surface markers and 9 isotype controls, arrayed across three 96-well plates. The cell surface marker antibody screen was performed twice using a total of 4 individual mammary reduction samples. 3–4 × 10^5 breast cells were used for each antibody to ensure sufficient cells analysed to obtain a reliable positive signal. A detailed list of the antibodies can be found in Supplementary Table 1. Staining was performed as described by the manufacturer's protocol with minor modifications. Briefly, the lyophilized antibodies were reconstituted with 110 μl of deionised water. One hundred microliters of breast cell suspension was aliquoted into three new 96-well plates at a density of 3–4 × 10^5 cells/well. Twnety microliters of the reconstituted antibody was added to cells and incubated on ice for 20 min. The cells were then washed twice with HBSS plus 2% FBS and centrifuged at 300×*g* for 5 min to remove

any unlabelled antibody. The cell pellet was incubated with the following primary antibodies: CD31-APC/Cy7, CD45-APC/Cy7, epithelial cell adhesion molecule (EpCAM)-PE, CD49f-PE/Cy7 (BioLegend). CD45 and CD31 were used to deplete contaminating haematopoietic and endothelial cells (collectively termed Lin+ cells). Cells were incubated with 4′,6-diamidino-2-phenylindole (DAPI, Invitrogen) before a final wash and data was acquired by flow cytometry using an LSR II flow cytometer (BD Biosciences) with a high-throughput sample attachment on the instrument, and 250,000–350,000 events per well were collected. The lyoplate workflow is shown in Fig. 1.

Surface protein screening data analysis

Data analysis was accomplished using FlowJo v10 software (FlowJo LLC, Treestar, USA). The gating strategy (Figure S1) was designed to remove debris, dead and Lin+ cells. EpCAM and CD49f markers were used to discriminate between the basal, luminal progenitor, mature luminal and stromal cell types. To analyse each population for its AlexaFluor®647 positivity, a 1% positive events in the AlexaFluor®647 gate was the minimum criteria positive selection for each cell surface marker. Less than 1% event detections were deemed as negative cell surface markers and recorded as zero. Analytical data of percentage of AlexaFluor®647 positive events were exported to Excel and associated to sample ID, plate number row and column. To determine signal intensity, histograms were generated, and the control isotype median fluorescence intensity (MFI) was calculated using FlowJo. Bisector gating on the histogram was used to discriminate between positive and negative populations. Positivity was calculated as being 3 robust standard deviations of the control MFI. Selecting the positive population, the median, minimum and maximum fluorescence intensities were exported to Excel. Using the minimum and maximum values, variation in positive marker signal was categorised into 4 groups: 0 − < 1 log fluorescence intensity; 1 − > 1 and < 1.5 log fluorescence intensity; 1.5 − > 1.5 and < 2 log fluorescence intensity and 2 − > than 2 log fluorescence intensity.

Aldehyde dehydrogenase (ALDH) flow cytometry assay

Human breast single cell suspensions were treated to detect the enzyme activity of aldehyde dehydrogenase (ALDH) using the Aldefluor Kit (StemCell Technologies) as per the manufacturer's instructions. The cells were then preblocked with 10% normal rat serum (Sigma) and stained with the following antibodies: CD31-APC/Cy7 (Clone WM-59), CD45-APC/Cy7 (Clone HI30), EpCAM-PE, CD49f-PE/Cy7 (Clone GoH3) (all from BioLegend) in combination with one of the following antibodies CD140b-AF647 (Clone 28D4), CD142-AF647 (Clone HTF-1), CD26-AF647 (Clone M-A261), CD34-AF647 (Clone 581), CD340 (Her2)-AF647 (Clone Neu24.7), CD39-AF647 (Clone TU66), CD44-AF647 (Clone G44-26), CD49c-AF647 (Clone C3 II.1), CD66 (a,c,d,e)-AF647 (Clone B1.1/CD66), CD54-AF647 (Clone LB-2), CD55-AF647 (Clone IA10), CD13-AF647 (Clone WM15), CD73-AF647 (Clone AD2), CD15s-AF647 (Clone CSLEX1), CD151-AF647 (Clone 14A2.H1), CD166-AF647 (Clone 3A6), CD282-AF647 (Clone 11G7), CD63-AF647 (Clone H5C6), CD75-AF647 (Clone LN1), SSEA-4-AF647 (Clone MC813-70), TRA-1-81-AF647 (Clone TRA-1-81), CLA-Biotin-AF647 (Clone HECA-452), CD15-AF647 (Clone HI98) (all from BD Biosciences). Cells were then filtered through a 30-μm cell strainer and incubated with DAPI. Human cells were separated using an Influx cell sorter (Becton Dickinson). Single-stained control cells were used to perform compensation manually. Gates were set in reference to fluorescence-minus-one controls. The ALDH+ gate was set in reference to control populations incubated with the ALDH inhibitor DEAB in addition to Aldefluor. Flow cytometry data were analysed using FlowJo™ software.

In vitro colony-forming assays

Flow-sorted human luminal progenitor cells were seeded into 60 mm plates with 2.5 × 10^5 irradiated NIH-3 T3 feeder cells. The cultures were maintained in Human EpiCult-B (StemCell Technologies) supplemented with 5% FBS (StemCell Technologies) and 50 μg/ml gentamicin for 48 h and then the media changed to serum-free

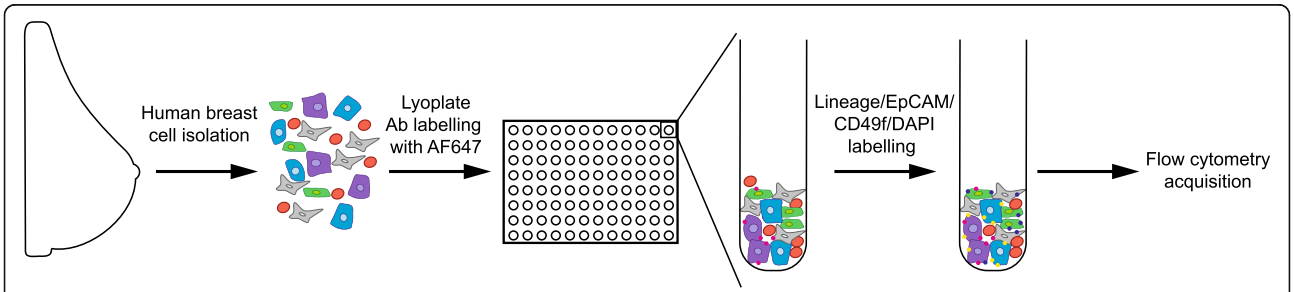


Fig. 1 Schematic diagram depicting the experimental overview of the human breast surface protein marker antibody screen

conditions and maintained for an additional 12 days. Colonies were fixed with acetone to methanol (1:1), stained with Giemsa (Fisher Scientific) and enumerated under a microscope.

Statistical analysis

Data presented are the mean of multiple independent experiments and the standard error of the mean. One-way analysis of variance was used to test multiple groups followed by Tukey's post-test to test significant differences between pairs of results. Comparisons between just two groups were analysed by t-test. Significance was set at $^{*}P < 0.05$ and $^{**}P < 0.01$.

Results

To explore the heterogeneity of normal breast epithelial and stromal cells and to generate a dataset of surface protein expression, we subjected human reduction mammoplasty specimens to a panel of monoclonal antibodies specific for 242 human cell surface proteins using the BD Lyoplate system. Primary human breast tissue from two healthy donors per antibody screen was dissociated to single cells and pooled. Single cell suspensions were arrayed on the 96 well plates containing the Alexa-Fluor®647-conjugated lyoplate antibodies and controls. Subsequently, tagged cells were then subjected to the widely used flow cytometry staining protocol (Fig. 1). Flow cytometry (FC) analysis gating allowed the elimination of doublets, debris and endothelial/haematopoietic cells. The breast epithelial subpopulations and stromal compartments were then gated to identify negative and positive antibody markers (Figure S1a-b). The inclusion of the breast epithelial flow antibody strategy was imperative to eliminate the number of false positive surface markers irrelevant to the stromal/epithelial content of the normal human breast.

Analysis of the screen revealed 78 out of the 242 lyoplate cell surface proteins were positive in the breast epithelial/stromal compartments (Fig. 2, Figure S2a). Without the inclusion of lineage or live/dead markers, the number of positive antibodies increased to 144 and 168, respectively (Figure S2b-c). The mean percentage of positive cells for each cell surface marker (greater than 1% positive) of the different epithelial/stromal populations was calculated (Fig. 2). As expected, our screen positively identified a number of well-known breast basal and luminal epithelial cell surface proteins including CD10, CD24, CD44, CD227, CD340 and EGFR (Fig. 2) [7, 10–12, 18, 19]. Furthermore, we identified positive expression of CD49a, CD49b, CD49c, CD47, CD54, CD73, CD90, CD95, CD151, CD271, HLA-ABC, HLA-DR, SSEA-4 and CD201 markers which were reported in primary human breast cells and tissue [20–25] and on breast organoids [26]. Surprisingly, CD117 (C-Kit), a well-known surface marker expressed on breast epithelial cells, was not detected as being positive in this screen. C-Kit [27], along with CD105 [28] were detected in breast epithelial or stromal cells, respectively, via FC. However, these studies used different clones for CD117/C-Kit and CD105 compared to the antibodies in this screen, highlighting that different antibody clones may yield contrasting results. Although the complete list of CD markers was not included in this screen, the screen contained a number of surface markers not previously examined in breast tissues. The screen identified 35 surface markers that were novel and a further 8 less characterised markers in the normal breast epithelial/stromal compartments (Fig. 2). The less characterised markers are of interest, as these markers were previously reported as having expression in normal breast tissues; however, no distinction between luminal or basal cell types was documented [29–35]. Quantification revealed several of the novel and less characterised markers were widely expressed in breast epithelial cells. For instance, CD9, CD59 and CD164 expression was detected in greater than 80% of all epithelial subpopulations (Fig. 2). Other novel markers including CD40 and CD120b were expressed in 5% or less of each epithelial subpopulation, demonstrating the heterogeneity of marker expression in the normal breast.

Unsupervised hierarchical clustering of the 78 positive surface markers showed several expression clusters between the different subpopulations (Fig. 3a). We observed distinct clusters exhibiting high expression in both epithelial and stromal populations (CD44, CD54, CD59, CD164, HLA A,B,C) or enriched in epithelial populations (CD9, CD49c, CD49e, CD55, and CD66(a,c,d,e)). These data indicate that these markers may contribute towards a general biological function. Other clusters of CD markers were restricted to epithelial subcompartments including the luminal cluster encompassing of CD24, CD227, CD46, CD321, CD166 and CD340 cell surface markers, the luminal progenitor cluster (EGFR, CD282 and CLA), and the basal cell cluster (CD10, CD200, CD271, CD142, CD201 and CD104), suggesting more specialised function in these cell types. Of note, 62 of the 78 positive surface markers were expressed on stromal cells (Fig. 2), yet only a few of these markers were restricted to the stromal compartment (Fig. 3a). It is also notable that several markers were expressed in both stromal and luminal populations including CD13, CD75, CD95, CD107a, Hem. Prog. Cell and GD2 (Figs. 2 and 3a).

Marker positivity gives indication of the proportion of cells expressing these markers; it does not indicate the signal intensity or heterogenous marker expression. We generated histograms of the positive identified surface markers to determine whether heterogenous expression

Antibody	Basal	LP	NC	Stromal
CD1a	0	0	0	0
CD1b	0	0	0	0
CD1d	0	0	0	0
CD2	0	0	0	0
CD3	0	0	0	0
CD4	0	0	0	0
CD4v4	0	0	0	0
CD5	0	0	0	0
CD6	0	0	0	0
CD7	0	0	0	0
CD8a	0	0	0	0
CD8b	0	0	0	0
CD9*	92.466	87.133	87.366	25.933
CD10	43.2	4.166	0.956	11.246
CD11a	0	0	0	0
CD11b	0	0	0	0
CD11c	0	0	0	0
CD13	4.415	76.9	17.35	79.2
CD14	0	0	0	0
CD15*	2.135	21	5	0
CD15s	2.095	20.7	2.16	1.1
CD16	0	0	0	0
CD18	3.145	0	0	13.4
CD19	0	0	0	0
CD20	0	0	0	0
CD21	0	0	0	0
CD22	0	0	0	0
CD23	0	0	0	0
CD24	1.82	82.5	82.933	0.673
CD25	0	0	0	0
CD26	8.935	6.805	1.68	30.5
CD27	0	0	0	0
CD28	0	0	0	0
CD29	42.35	8.505	1.8	50.35
CD30	0	0	0	0
CD32	0	0	0	0
CD33	0	0	0	0
CD34	4.95	3.803	3.71	74.2
CD35	0	0	0	0
CD36*	2.186	0.446	0	5.346
CD37	0	0	0	0
CD38	0	0	0	0
CD39	2.375	1.47	0.5	40.75
CD40	2.855	5.42	0.54	4.42
CD41a	0	0	0	0
CD41b	0	0	0	0
CD42a	0	0	0	0
CD42b	0	0	0	0
CD43	0	0	0	0.546
CD44	92.733	59.833	40.633	86.833
CD45	2.826	3.143	1.493	29.466
CD45RA	0	0	0	0
CD45RB	0	0	0	0
CD45RO	0	0	0	0
CD46*	8.52	72.7	68.6	10.98
CD47	29.95	54.25	48.3	79.7
CD48	0	0	0	0
CD49a	16.95	2.45	0	37.35
CD49b	92.166	90.7	75.7	3.916
CD49c	93.1	53.166	58.866	13
CD49d	1.98	0.98	1.03	7.266
CD49e	8.44	12.176	12.106	70.766
CD50	0	0	0	0
CD51/61	0	0	0	0
CD53	0	0	0	0
CD54	76.05	73.85	68.95	57.65
CD55	47.85	55.85	22.95	11.625
CD56	0	0	0	0
CD57	14.45	17.4	3.015	1.37
CD58	10.035	7.62	0.895	24.555
CD59*	98.366	98.866	94.3	80.1
CD61	0.743	1.17	0.413	3.533
CD62E	0	0	0	0
CD62L	0	0	0	0
CD62P	0	0	0	0
CD63	37.533	22.366	34.166	49.566
CD64	0	0	0	0
CD66(a,c,d,e)	51.3	81.75	55.15	1.135
CD66b	0	0	0	0
CD66f	0	0	0	0

Antibody	Basal	LP	NC	Stromal
CD69	5.745	2.335	0	0
CD70	0.955	0	0	4.205
CD71	1.406	1.3	0	7.366
CD72	0	0	0	0
CD73	4.463	5.83	5.903	73.266
CD74	2.236	3.223	1.01	24.8
CD75	0.796	5.066	6.046	4.013
CD77	0	0	0	0
CD79b	0	0	0	0
CD80	0	0	0	0
CD81	0	0	0	0
CD83	0	0	0	0
CD84	1.89	0	0	0
CD85	2.17	0.97	0.5	7.105
CD86	0	0	0	0
CD87	0	0	0	0
CD88	0	0	0	0
CD89	0	0	0	0
CD90	39.7	5.38	1.68	58.5
CD91	0	0	0	0
CDw93	0	0	0	0
CD94	0	0	0	0
CD95	0.835	4.395	0.5	3.845
CD97	0	0	0	0
CD98	7.245	25.7	11	5.215
CD99	0	0	0	0
CD99R	0	0	0	0
CD100	0	0	0	0
CD102	0	0	0	0
CD103	0	0	0	0
CD105	0	0	0	0
CD106	0	0	0	0
CD107a*	0	8.865	12.3	2.095
CD107b	0	0	0	0
CD108	0	0	0	0
CD109	0	0	0	0
CD112	0	0	0	0
CD114	0	0	0	0
CD116	0	0	0	0
CD117	0	0	0	0
CD118 (LIFR cptr)	0	0	0	0
CD119	0	1.63	2.445	0
CD120a	6.93	3.5	2.09	3.905
CD121a	0	0	0	0
CD121b	0	0	0	0
CD122	0	0	0	0
CD123	0	0	0	0
CD124	0	0	0	0
CD126	0	0	0	0
CD127	0	0	0	0
CD128b	0	0	0	0
CD130	0	0	0	0
CD134	0	0	0	0
CD135	0	0	0	0
CD137	0	0	0	0
CD137 Ligand	0	0	0	0
CD138	0	0	0	0
CD140a	0	0	0	0
CD140b	2.05	1.315	2.625	65.65
CD141	0.935	0	0	1.985
CD142	38.9	15.4	17.05	5.085
CD144	0	0	0	0
CD146	0	0	0	1.875
CD147*	6.445	19.4	14.4	10.905
CD150	0	0	0	0
CD151	42.25	18.2	32.85	29.75
CD152	0	0	0	0
CD153	0	0	0	0
CD154	0	0	0	0
CD158a	0	0	0	0
CD158b	0	0	0	0
CD161	0	0	0	0
CD162	0	0	0	0
CD163	0	0	0	0
CD164	81.7	90.85	80.3	71.1
CD165	0	0	0	0
CD166	0	25.75	73.45	1.14
CD171	0	0	0	0
CD172b	0	0	0	0
CD177	0	0	0	0

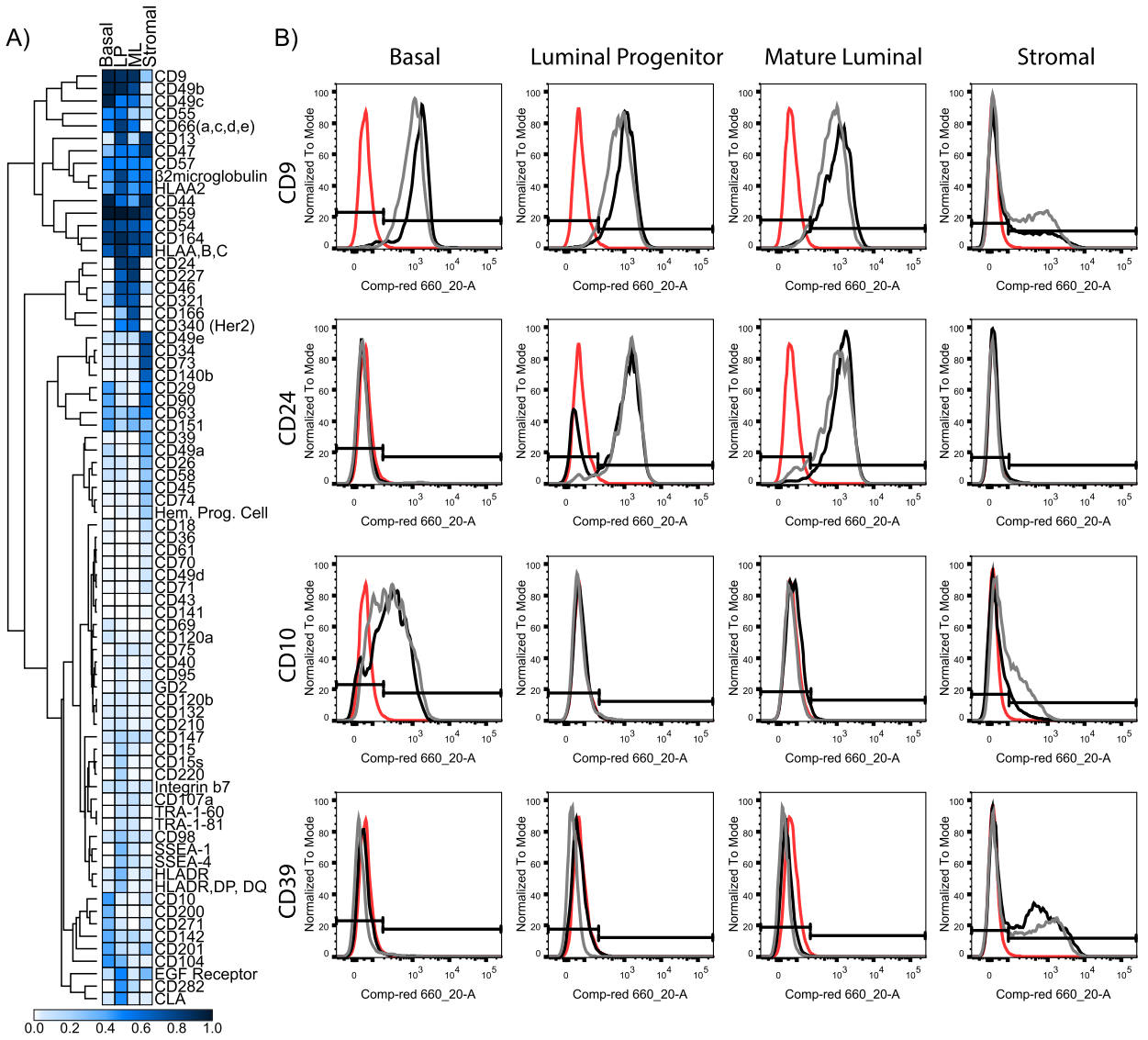
Antibody	Basal	LP	NC	Stromal
CD178	0	0	0	0
CD180	0	0	0	0
CD181	0	0	0	0
CD183	0	0	0	0
CD184	0	0	0	0
CD193	0	0	0	0
CD195	0	0	0	0
CD196	0	0	0	0
CD197	0	0	0	0
CD200	34.7	2.37	0.55	1.635
CD205	0	0	0	0
CD206	0	0	0	0
CD209	0	0	0	0
CD220	0	20.55	1.695	0
CD221	0	0	0	0
CD226	0	0	0	0
CD227	0	64.45	82.95	0.58
CD229	0	0	0	0
CD231	0	0	0	0
CD235a	0	0	0	0
CD243	0	0	0	0
CD244	0	0	0	0
CD255 (Tweak)	0	0	0	0
CD268	0	0	0	0
CD271	31.7	5.28	0	9.51
CD273	0	0	0	0
CD274	0	0	0	0
CD275	0	0	0	0
CD278	0	0	0	0
CD279	0	0	0	0
CD282	0.535	45.2	16.475	0
CD305	0	0	0	0
CD309	0	0	0	0
CD314	0	0	0	0
CD321 (F11 Rcptr)*	15.53	68.7	68	1.28
CDw327	0	0	0	0
CDw328	0	0	0	0
CDw329	0	0	0	0
CD335	0	0	0	0
CD336	0	0	0	0
CD337	0	0	0	0
CD338	0	0	0	0
CD340 (Her2)	0.98	49.45	59.85	0.38
abTCR	0	0	0	0
β2microglobulin	44.05	79.35	48.2	58.15
BLTR-1	0	0	0	0
CLIP	0	0	0	0
CMRF-44	0	0	0	0
CMRF-56	0	0	0	0
EGF Receptor	13.4	49.05	6.075	32.5
fMLP Receptor	0	0	0	0
γδTCR	0	0	0	0
Hem. Prog. Cell	0.5	3.165	2.065	22.65
HLAA,B,C	66.9	87.05	68.15	72.4
HLAA2	40.1	71.05	43.6	56.2
HLADQ	0	0	0	0
HLADR	5.89	32.2	3.52	5.205
HLADR,DP, DQ	3.76	27.1	2.325	3.345
Invariant NKT	0	0	0	0
GD2	1.39	5.575	1.49	7.27
MIC A/B	0	0	0	0
NKB1	0	0	0	0
SSEA-1	0	32.1	9.11	0
SSEA-4	0.655	26.65	10.46	2.815
TRA-1-60	0	13.7	6.71	0
TRA-1-81	0	12.85	6.695	0
Vb 23	0	0	0	0
Vb 8	0	0	0	0
CD104	45.05	33.25	3.77	2.63
CD120b	4.13	5.85	3.34	2.77
CD132	4.31	6.3	3.005	2.175
CD201	41.3	15.25	4.165	28.85
CD210	4.61	9.595	2.385	3.3
CD212	0	0	0	0
CD267	0	0	0	0
CD294	0	0	0	0
SSEA-3	0	0	0	0
CLA	6.98	47.05	3.795	4.25
Integrin b7	8.32	17.25	4.54	6.48

Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 Lyoplate analysis of surface marker expression patterns in different mammary epithelial and stromal subpopulations. FACS based expression analysis of 242 surface markers using the BD Lyoplate™ Human Surface Marker Screening Panel on primary human breast cell populations (blue: basal, green: luminal progenitor (LP), purple: mature luminal (ML), orange: stromal). The values represent the mean percentage of positive cells for each surface marker antibody within two pooled donor samples from the two antibody screen replicates. Zero indicated that the percentage of cells ranged from 0 to 1% within the positive gating. CD markers underlined indicate unreported expression in breast epithelial/stromal cells. CD markers with an asterisk indicate less characterised expression in breast epithelial and stromal cells

patterns exist in the epithelial or stromal subpopulations (Fig. 3b, Figure S3). We observed that several markers had strong signal intensities including CD9 and CD59 in all populations. Luminal cells expressing CD24, CD49b, and CD13 in the LP populations all showed strong signal intensities. Basal cells expressing CD44, CD49b and CD49c also showed strong signal intensity. However, the vast majority of markers displayed diverse fluorescence intensities, suggestive of heterogeneity marker expression. Examining the minimum and maximum signal intensities (Table 1), a small number of markers displayed a spread of signal intensities greater than 1.5 logs. Many of the markers with a broad signal intensity were only detected in less than 5% of the subpopulation. This is evident in CD36, CD39, CD73 for the basal population; CD34 for the LP population; CD29, CD34, CD39 and CD73 for the ML population, indicating that whilst these populations may have some heterogenous expression, the overall proportion of cells expressing these markers are low. The stromal population contained markers that had the most heterogenous expression, especially for cells expressing CD9, CD13, CD26, CD34, CD39, CD44, CD49a, CD54 and CD73 (Table 1, Figure S3). To validate the specificity of the screening panel, we selected well known positive markers in breast epithelial (CD44, CD340) and stromal cells (CD140b [14], CD34 [25], CD26 [28]), as well as novel/less characterised epithelial (CD142, CD49c, CD66, CD54, CD55) and novel stromal (CD39) CD markers identified from the screen (Fig. 3b) for expression analysis in an additional two independent donor samples. The resulting FC analyses indicated that all positive surface markers selected from the screen for validation were also detected in subsequent donor samples (Fig. 3c, Figure S4), however, at times the proportion of cell positivity differed. We observed antibodies such as CD140b containing 4.5-fold higher proportion in the basal compartment and a 2-fold reduction in the stromal compartment. CD142 contained a 2-3-fold reduction in luminal and stromal compartments, but a small increase in the basal compartment (Fig. 3c). Other antibodies including CD54 and CD55 showed comparable proportions between the screens and the subsequent donor samples (Fig. 3c). Whilst proportions differed in some cases between the screen and validation assays, the trend of positivity was the same, i.e. CD140b expression was most frequently detected in the stromal compartment (Fig. 2), and this trend was observed in the subsequent donor samples (Fig. 3c). This demonstrates that we have generated a robust dataset as a resource for identifying a selection of CD marker expression on normal human breast cells.

The luminal compartment is considered to be the cell of origin for most breast cancers and understanding the heterogeneity of surface marker expression in normal cells may illuminate differences in cell state with relevance to cancer initiation and progression. Focusing on the luminal compartment we investigated a selection of novel and less characterised surface markers identified from the lyoplate screen and confirmed expression in a further 3–5 donor samples (Fig. 4a). The markers were selected based on the following criteria: (i) dominant expression in the LP population (CLA, CD15s and CD15), (ii) high expression in the luminal population (CD13, CD282, TRA-1-81 and SSEA-4), (iii) moderate expression in the luminal population (CD63 and CD151) or (iv) dominant expression in the ML population (CD166). Although CD73 and CD75 markers were strongly expressed in the stromal compartment, positive cells were detected in the luminal compartments and were included for further analysis. CD166 and CD151 were also selected for further investigation. CD166 and CD151 have previously been detected in the ML and basal compartments, respectively; however, detection in the LP compartment is not well documented and warranted confirmation in a further 3–5 donor samples (Fig. 4a). FC analysis confirmed expression patterns reported in the lyoplate screen. However, we observed a range of positive cells in the luminal compartments between the different donor samples (Fig. 4a). CD13 and CD73 surface markers exhibited at least a twofold range of positive cells (Fig. 4a). CD15s, CD15, CD282 and CLA displayed a wide range of positive cells, where some donor samples exhibited a lower proportion of positive cells, between 2.5% and 33%, for these markers whilst other donors contained 50% to 100% of LP cells expressing these markers (Fig. 4a), showing the disparate variability of marker expression on human breast cells. The proportion of cells that were positive for surface markers in the ML population also varied between donor samples (Fig. 4a). Again, whilst proportions differed between the screens and the further validation assays, the trend of positivity was the same, i.e., CD13 expression was most



C)

Surface marker	Basal	Luminal Progenitor	Mature Luminal	Stromal
CD140b[a]	9.725 (7.08 - 12.37)	1.556 (0.142 - 2.97)	0.224 (0.124 - 0.323)	33.49 (28.17 - 38.8)
CD142[a]	46.29 (39.34 - 53.24)	4.88 (3.61 - 6.15)	9.566 (6.7 - 12.432)	2.8 (2.24 - 3.36)
CD26[a]	2.637 (1.834 - 3.44)	3.265 (1.951 - 4.58)	0.778 (0.515 - 1.04)	9.273 (9.135 - 9.41)
CD34[a]	2.968 (2.89 - 3.045)	0.265 (0.153 - 0.377)	0.062 (0.029 - 0.095)	30.55 (21.57 - 39.53)
CD340 (Her2)[a]	1.213 (1.02 - 1.406)	56.96 (54.91 - 59)	70.53 (59.02 - 82.03)	0.38 (0.121 - 0.638)
CD39[a]	4.354 (2.307 - 6.4)	1.753 (1.675 - 1.83)	0.22 (0.21 - 0.23)	37.29 (22.385 - 52.2)
CD44[a]	92.76 (90.465 - 95.045)	68.3 (58.7 - 77.9)	53.62 (51.475 - 55.77)	77.5 (65.6 - 89.4)
CD49c[a]	97.8 (96.835 - 98.76)	73.76 (60.81 - 86.7)	84.15 (80.76 - 87.54)	65.91 (53.02 - 78.8)
CD66 (a,c,d,e)	48.63 (4.29 - 96.4)	78.52 (44.2 - 100)	49.4 (13.66 - 99.12)	0.906 (0.075 - 2.46)
CD54	90.02 (81.71 - 99.17)	85.22 (75.5 - 96.3)	72.47 (43.175 - 98.13)	56.69 (43.67 - 74.7)
CD55	45.15 (25.305 - 72.35)	62.64 (40.7 - 83.8)	35.2 (17.185 - 57.15)	14.86 (2.97 - 35.3)

Fig. 3 Validation of positive surface markers. **A** Heatmap showing expression of the positive detected surface markers antibodies of basal, LP, ML and stromal populations from the Lyoplate screen analysis. **B** Example histograms of show intensity staining for CD9, CD10, CD24 and CD39 compared with isotype controls (red) and the two replicate antibody screens (black and grey) in basal, LP, ML and stromal cells on a log scale. **C** 11 positive surface markers that were selected to validate the Lyoplate screen in human mammary epithelial and stromal subpopulations. Values represent the mean percentage of expression followed by the range of expression in bracket). [a] represent $n = 2$ independent human breast donor samples, otherwise $n = 3$

Table 1 Minimum and maximum signal intensities. Cell surface markers with 1.5 log fluorescence intensities or greater are highlighted in the basal (blue), luminal progenitor (green), mature luminal (purple) and stromal (orange) compartments

Basal

	Lyoplate screen 1					Lyoplate screen 2				
Antibody	**Median**	**rSD**	**Min**	**Max**	**Log spread**	**Median**	**rSD**	**Min**	**Max**	**Log spread**
CD9	1890	885	339	5138	1	1263	684	186	4843	1
CD10	489	373	141	2302	1	508	471	144	3825	1
CD13	204	166	119	3825	1	707	1466	105	12052	2
CD15	767	1044	103	625	0	243	639	141	7050	1.5
CD15s	236	115	144	592	0	485	474	134	2965	1
CD18	813	1	158	1177	1	267	141	128	2587	1
CD24	238	37.5	141	765	0	1248	1072	136	3400	1
CD26	204	133	101	4940	1	214	145	107	1627	1
CD29	308	128	144	2488	1	269	119	147	4940	1
CD34	333	1370	134	9493	1.5	1314	804	147	4303	1
CD36	301	845	136	8768	1.5	938	2086	155	12793	2
CD39	360	491	134	6776	1.5	373	481	107	5345	1.5
CD40	173	55.6	98	581	0	232	244	105	2050	1
CD44	973	566	141	3400	1	1599	864	152	5561	1.5
CD45	465	835	107	3144	1	826	462	107	1972	1
CD46	207	66.3	101	592	0	207	131	105	1450	1
CD47	455	171	105	2258	1	202	84.9	105	1396	1
CD49a	230	120	103	1861	1	296	196	112	4656	1
CD49b	1409	723	161	4476	1	1148	546	186	3536	1
CD49c	949	565	131	4137	1	1345	728	147	4749	1
CD49d	199	159	103	2258	1	408	447	98	1861	1
CD49e	208	89.6	103	1319	1	207	81.8	107	1691	1
CD54	437	284	94	2965	1	1386	939	103	7481	1.5
CD55	334	158	107	1031	1	463	234	105	1566	1
CD57	328	423	107	3606	1	223	228	116	1112	1
CD58	199	63.5	110	855	0	250	151	103	3083	1
CD59	1127	521	176	3825	1	1977	700	459	5138	1
CD61	206	134	103	1536	1	174	1	96	179	0
CD63	292	122	136	1450	1	347	167	134	1507	1
CD66(a,c,d,e)	355	214	139	6261	1.5	595	425	134	6385	1.5
CD69	271	177	136	672	0	357	176	139	1627	1
CD70	231	131	134	1031	0	240	126	144	1246	1
CD71	248	69.5	136	711	0	271	93.1	136	698	0
CD73	274	497	144	6643	1.5	505	1281	144	8944	1.5
CD74	390	253	139	1566	1	381	239	136	1396	1
CD75	258	82.8	136	1031	0	594	724	179	2537	1
CD84	209	1	131	1269	1	307	134	134	1222	1
CD85	258	150	136	3206	1	252	109	141	1897	1
CD90	155	108	49	1482	1	274	269	53	3175	1
CD95	105	31	54	322	0	107	29.8	54	203	0
CD98	127	48.1	58	373	0	150	73.1	77	1057	0
CD107a	94.8	26	64	481	0	100	39.6	70	461	0
CD119	87.6	13.6	52	226	0	92.1	17.8	54	161	0
CD120a	117	56.5	68	524	0	121	71.2	70	988	0
CD140b	282	389	64	1182	1	161	180	68	1992	1
CD141	108	102	50	925	0	121	63.5	73	1057	0
CD142	158	99	70	1033	0	238	206	77	1903	1
CD147	107	30.3	70	350	0	158	71.2	79	1105	1
CD151	176	86.9	75	1081	1	249	133	73	1659	1
CD164	265	146	70	1323	1	599	423	77	3015	1
CD166	90.6	17.9	68	261	0	90.6	17	70	251	0
CD200	120	53.5	73	524	0	205	122	73	946	0
CD220	89.4	14	27	164	0	97	23.9	64	278	0
CD227	104	91.5	68	1033	0	90	15.3	48	241	0
CD271	182	97.1	70	905	0	141	67.8	73	865	0
CD282	88.8	30.8	59	675	0	78.7	18.6	51	296	0
CD321 (F11 Rcpt	149	83.6	68	1240	1	87.3	20.7	66	370	0
CD340 (Her2)	91.8	28.8	59	388	0	82.5	18.1	47	230	0
b2micro globulin	227	179	68	2076	1	147	88.6	68	1920	1
EGF Receptor	135	62.6	63	870	0	92.4	25.7	61	388	0
Hem. Prog. Cell	93.6	41.5	61	615	0	82.8	33.6	47	407	0
HLAA,B,C	466	374	71	5210	1.5	187	122	54	2304	1
HLAA2	283	240	71	2915	1	124	69.2	66	1120	1
HLADR	109	224	61	3598	1	122	124	68	1823	1
HLADR,DP, DQ	121	335	66	2915	1	110	88.7	63	1178	1
Disialoganglioside G	94.2	129	63	2304	1	87.9	222	40	1305	1
SSEA-1	81.6	18.1	56	278	0	82.2	37.9	52	448	0
SSEA-4	87	32.5	49	470	0	92.1	54.8	66	938	0
TRA-1-60	84.9	18	49	299	0	85.2	34.4	47	177	0
TRA-1-81	80.8	14.6	54	292	0	78.7	11.5	54	370	0
CD104	158	76.1	71	696	0	398	221	66	1446	1
CD120b	93.6	84.2	45	482	0	118	61.9	59	892	0
CD132	85.5	24.8	38	299	0	122	61.5	54	615	0
CD201	123	51.6	52	407	0	254	137	66	2187	1
CD210	95.5	30.4	52	388	0	112	55.5	54	987	0
utaneous Lymph. An	93.9	41.9	47	849	0	119	72.9	63	938	0
Integrin b7	97.3	34.5	45	370	0	106	44.8	54	731	0

Luminal Progenitor

	Lyoplate screen 1					Lyoplate screen 2				
Antibody	**Median**	**rSD**	**Min**	**Max**	**Log spread**	**Median**	**rSD**	**Min**	**Max**	**Log spread**
CD9	1308	737	178	4996	1	921	610	124	3811	1
CD10	275	84.8	161	719	0	325	175	164	1993	1
CD13	2210	2226	173	11853	2	1245	1704	164	11323	2
CD15	624	683	168	8812	1.5	458	658	178	7181	1.5
CD15s	498	658	171	3726	1	480	386	168	3987	1
CD18	638	318	186	3944	1	358	275	186	1746	1
CD24	1554	983	182	6125	1.5	1695	980	161	5989	1.5
CD26	239	84.5	130	4885	1	263	338	130	2662	1
CD29	288	95.7	171	1564	1	339	190	182	5469	1.5
CD34	295	162	171	12407	2	1617	748	164	5854	1.5
CD36	275	93.6	178	1797	1	412	791	161	3987	1
CD39	294	117	133	1203	1	1074	1818	135	7181	1.5
CD40	242	67	133	689	0	251	90.2	127	852	0
CD44	538	503	168	5595	1.5	463	373	161	6862	1.5
CD45	318	735	133	4996	1	220	84.6	130	1907	1
CD46	648	368	127	2722	1	725	414	141	2976	1
CD47	652	305	121	2178	1	251	78.9	130	1012	1
CD49a	246	96.8	133	1371	1	222	54.8	127	871	0
CD49b	802	582	148	4669	1	595	326	135	2381	1
CD49c	441	286	168	2722	1	378	266	168	3184	1
CD49d	229	51	138	634	0	287	214	133	1993	1
CD49e	222	69.2	106	931	0	269	2400	138	2329	1
CD54	644	480	138	3256	1	1074	963	135	6267	1.5
CD55	456	277	135	1785	1	587	492	133	4077	1
CD57	460	453	138	3184	1	215	53.6	130	1993	1
CD58	284	98.6	135	909	0	244	69.7	116	719	0
CD59	1409	798	198	4775	1	2526	1128	536	7347	1.5
CD61	229	62.6	111	608	0	220	43.6	124	341	0
CD63	329	147	154	1313	1	387	197	157	1950	1
CD66(a,c,d,e)	797	1018	161	9653	1.5	1379	1295	154	9222	1.5
CD69	263	73.3	164	2381	1	313	122	164	1203	1
CD70	259	57.9	164	1152	0	256	48.7	164	571	0
CD71	278	80.8	154	852	0	293	94.7	154	800	0
CD73	290	187	164	5854	1.5	373	278	157	5854	1.5
CD74	255	66.9	154	1127	0	258	79.6	154	704	0
CD75	357	671	161	8420	1.5	518	558	161	4462	1
CD84	254	59.8	161	2083	1	287	89.2	157	835	0
CD85	266	74.9	161	1993	1	246	45.7	161	1257	1
CD90	127	110	63	2752	1	123	182	63	2477	1
CD95	150	38.1	81	352	0	162	52.5	83	358	0
CD98	209	103	81	641	0	195	86.7	96	694	0
CD107a	160	75.2	96	1015	0	154	589.5	97	751	0
CD119	133	20.7	69	207	0	139	28.8	79	291	0
CD120a	134	46.2	101	667	0	153	52.4	94	1015	0
CD140b	326	191	83	559	0	253	351	101	1692	1
CD141	135	44.6	71	260	0	138	26.7	99	380	0
CD142	189	124	101	955	0	181	122	101	1405	1
CD147	162	68.1	94	1123	1	214	96.7	108	1623	1
CD151	170	67.2	99	1123	1	194	104	101	864	0
CD164	372	269	99	2039	1	826	647	113	4827	1
CD166	225	160	99	1464	1	173	97.2	104	936	0
CD200	142	64.8	90	365	0	165	107	99	1078	0
CD220	155	58.3	96	527	0	165	60.8	92	559	0
CD227	405	476	94	6501	1.5	409	418	101	7388	1.5
CD271	245	158	99	1349	1	149	65.8	104	955	0
CD282	258	157	82	1244	1	151	67.9	83	670	0
CD321 (F11 Rcpt	388	179	81	1214	1	171	84.3	86	762	0
CD340 (Her2)	218	112	81	1047	0	157	86.4	78	762	0
b2micro globulin	416	241	83	2006	1	248	161	86	1483	1
EGF Receptor	254	133	83	840	0	134	51.1	83	420	0
Hem. Prog. Cell	119	61.2	81	1073	0	102	22.3	61	544	0
HLAA,B,C	921	584	136	5177	1.5	317	202	76	1768	1
HLAA2	603	430	83	2793	1	219	156	78	1559	1
HLADR	249	336	76	3608	1	253	378	64	3341	1
HLADR,DP, DQ	231	266	81	2164	1	290	399	88	4791	1
Disialoganglioside G	117	85.3	71	2220	1	143	117	76	949	0
SSEA-1	307	457	68	3256	1	180	281	76	3341	1
SSEA-4	418	587	71	3798	1	201	429	76	3428	1
TRA-1-60	137	232	64	3093	1	213	251	64	2396	1
TRA-1-81	141	278	68	3608	1	164	245	68	2164	1
CD104	119	37.8	73	290	0	205	130	73	1127	1
CD120b	114	32.4	64	277	0	141	74.8	68	1047	0
CD132	116	41.5	64	391	0	143	73.8	61	1021	0
CD201	130	56.9	64	451	0	154	105	59	1100	1
CD210	125	49.3	59	303	0	133	63.9	64	926	0
utaneous Lymph. An	287	195	68	1185	1	404	338	66	2220	1
Integrin b7	136	67.6	66	643	0	131	59.9	66	675	0

Mature Luminal

	Lyoplate screen 1					Lyoplate screen 2				
Antibody	**Median**	**rSD**	**Min**	**Max**	**Log spread**	**Median**	**rSD**	**Min**	**Max**	**Log spread**
CD9	1370	984	115	6104	1.5	991	683	157	4192	1
CD10	270	52.8	161	921	0	257	57.6	159	1923	0
CD13	822	1484	169	11949	2	1099	1629	173	13435	2
CD15	474	724	169	10974	2	616	1081	167	5108	1
CD15s	346	314	166	1960	1	420	385	157	5160	1.5
CD18	371	497	157	11143	2	334	224	172	3270	1
CD24	1703	972	166	5688	1.5	1314	956	176	5705	1.5
CD26	205	61.5	114	2287	1	203	51.9	117	1430	1
CD29	336	1212	169	6194	1.5	364	428	166	6506	1.5
CD34	259	91.9	160	18646	2	1972	1366	171	18629	2
CD36	265	93.7	162	2917	1	323	204	116	987	0
CD39	318	713	124	7534	1.5	1289	1674	115	6736	1.5
CD40	201	47.8	115	419	0	214	74.7	108	936	0
CD44	440	299	154	3073	1	487	465	141	8540	1.5
CD45	205	421	119	3015	1	228	88.2	114	1638	1
CD46	534	338	119	2060	1	828	476	110	3323	1
CD47	555	230	124	1955	1	238	74.5	106	931	0
CD49a	207	129	121	2950	1	197	53.6	124	1389	1
CD49b	540	316	124	2178	1	443	263	118	1913	1
CD49c	393	198	169	2085	1	487	235	173	1449	1
CD49d	200	43.7	121	474	0	301	236	116	1276	1
CD49e	218	61	111	711	0	235	177	116	1913	1
CD54	413	288	121	2480	1	1292	1007	106	6792	1.5
CD55	271	137	113	1148	1	333	237	116	2646	1
CD57	250	218	124	1998	1	193	35.6	116	991	0
CD58	209	55.4	111	493	0	209	53.2	113	568	0
CD59	928	597	126	4277	1	1744	994	219	6943	1.5
CD61	201	59.4	106	874	0	186	31.5	98	338	0
CD63	320	142	150	1331	1	380	188	147	1719	1
CD66(a,c,d,e)	519	390	147	5447	1.5	669	653	159	5569	1.5
CD69	252	89.2	144	1578	1	305	120	153	1545	1
CD70	258	62.4	153	931	0	245	44	156	893	0
CD71	259	60.5	153	697	0	268	71.4	153	697	0
CD73	378	505	162	10341	2	431	415	159	5447	1.5
CD74	274	145	147	1359	1	328	193	159	1998	1
CD75	486	649	166	7585	1.5	607	726	159	4277	1
CD84	246	45.3	150	616	0	280	87.2	150	726	0
CD85	279	95.6	153	1418	1	245	95.2	159	1682	1
CD90	92.4	38.4	49	1302	1	102	32.8	55	924	0
CD95	126	35.9	70	345	0	126	36.7	67	305	0
CD98	157	62.5	74	586	0	165	70.8	85	586	0
CD107a	147	56.7	81	444	0	141	57.9	88	574	0
CD119	117	24.1	70	263	0	133	40.3	70	368	0
CD120a	114	19.9	74	311	0	166	79.4	83	906	0
CD140b	282	205	76	777	0	135	187	85	1735	1
CD141	109	10.1	74	233	0	124	27.1	74	345	0
CD142	172	152	85	1416	1	164	117	95	2179	1
CD147	179	120	92	1384	1	165	203	83	2334	1
CD151	154	59.2	92	562	0	328	167	92	1323	1
CD164	336	242	100	2443	1	445	347	92	3220	1
CD166	288	168	78	1384	1	311	154	85	1106	1
CD200	105	11.6	67	185	0	118	24.6	78	318	0
CD220	111	20.5	63	263	0	118	22.3	78	286	0
CD227	1159	1013	88	9188	1.5	899	1187	90	7270	1.5
CD271	120	56.1	88	728	0	115	19.9	90	562	0
CD282	175	129	84	1030	0	117	38	86	538	0
CD321 (F11 Rcpt	427	205	77	1432	1	176	73.6	79	739	0
CD340 (Her2)	258	159	84	1252	1	184	86.4	91	584	0
b2micro globulin	259	165	91	1369	1	168	78.4	77	622	0
EGF Receptor	159	106	79	806	0	126	47.2	77	622	0
Hem. Prog. Cell	129	119	82	982	0	106	30.4	71	492	0
HLAA,B,C	539	424	84	3240	1	216	128	79	1198	1
HLAA2	356	312	77	2198	1	159	83.7	84	940	1
HLADR	123	64.2	68	1309	1	136	126	79	1713	1
HLADR,DP, DQ	120	64.9	77	1281	1	137	136	79	1309	1
Disialoganglioside G	116	50.4	75	1793	1	105	76.7	62	664	0
SSEA-1	129	158	73	3552	1	175	427	73	4174	1
SSEA-4	255	555	79	2957	1	143	197	82	3025	1
TRA-1-60	130	276	66	3806	1	140	147	66	1339	1
TRA-1-81	137	313	71	2760	1	121	109	71	1339	1
CD104	110	30.6	60	90	0	127	55.7	77	525	0
CD120b	107	20.6	58	201	0	137	75.4	71	1146	1
CD132	115	40.6	71	264	0	133	69	64	861	0
CD201	115	46	66	360	0	149	88.7	68	1049	0
CD210	120	49.3	66	324	0	126	62	64	584	0
utaneous Lymph. An	133	71.8	71	622	0	141	79.8	73	880	0
Integrin b7	113	42.3	64	408	0	130	61.7	68	664	0

Stromal

	Lyoplate screen 1					Lyoplate screen 2				
Antibody	**Median**	**rSD**	**Min**	**Max**	**Log spread**	**Median**	**rSD**	**Min**	**Max**	**Log spread**
CD9	546	955	73	9744	1.5	616	746	73	5356	1.5
CD10	180	129	77	2080	1	242	205	76	2315	1
CD13	969	1002	79	9173	1.5	1802	1757	80	14366	2
CD15	224	1324	62	5993	1.5	105	41	78	4812	1
CD15s	157	70.8	81	675	0	143	142	91	2681	1
CD18	180	231	75	1388	1	235	303	88	3702	1
CD24	160	104	82	1950	1	171	238	83	2852	1
CD26	696	907	79	9202	1.5	427	691	71	9323	1.5
CD29	288	449	78	4244	1	248	217	75	3393	1
CD34	2127	2027	83	14445	2	1360	1168	78	16873	2
CD36	388	1150	73	8571	1.5	774	1721	78	9409	1.5
CD39	741	1157	77	13502	2	897	1165	77	8656	1.5
CD40	125	68	53	678	0	144	95.7	51	948	0
CD44	1017	1003	76	6830	1.5	2063	1854	76	11686	2
CD45	1536	1024	53	3350	1	845	317	56	1646	1
CD46	178	129	51	1041	1	134	87.4	52	986	0
CD47	905	715	57	5296	1.5	487	324	53	2728	1
CD49a	357	537	58	7245	1.5	386	544	60	10135	2
CD49b	131	76.2	58	880	0	121	71.8	62	846	0
CD49c	704	687	76	4301	1	702	899	77	4772	1
CD49d	144	76.6	64	738	0	239	186	58	1739	1
CD49e	366	256	51	2176	1	528	343	55	2222	1
CD54	519	632	57	7716	1.5	1529	1926	60	13319	2
CD55	163	107	57	1285	1	217	205	55	1887	1
CD57	119	114	64	972	0	100	27.8	60	798	0
CD58	221	109	53	916	0	325	335	49	6663	1.5
CD59	841	648	84	7245	1.5	1278	1033	145	11739	2
CD61	153	98.8	44	798	0	106	30.5	49	280	0
CD63	364	286	70	2785	1	357	236	82	2315	1
CD66(a,c,d,e)	146	65.1	76	1450	1	174	266	82	4975	1
CD69	154	86.1	84	1186	1	188	133	80	1311	1
CD70	140	54.1	84	2089	1	210	152	84	2047	1
CD71	166	78.6	80	670	0	174	80.9	80	670	0
CD73	693	762	82	8391	1.5	561	694	82	11497	2
CD74	193	140	80	1509	1	203	167	80	1670	1
CD75	188	239	84	3087	1	316	372	78	3024	1
CD84	128	70.4	72	400	0	177	111	82	1139	1
CD85	177	288	78	2565	1	229	222	84	3642	1
CD90	378	459	53	4112	1	991	1020	45	8443	1.5
CD95	141	63.1	80	507	0	132	49.5	68	536	0
CD98	136	57	68	556	0	151	126	98	2166	1
CD107a	128	45.6	86	507	0	126	44.9	86	412	0
CD119	108	18.2	67	213	0	116	28.8	72	463	0
CD120a	173	162	84	1096	1	185	198	88	2166	1
CD140b	243	177	80	1888	1	271	197	80	2209	1
CD141	147	65.9	80	695	0	143	80.5	90	824	0
CD142	144	93.2	84	1279	1	147	108	82	1951	1
CD147	139	57.1	92	599	0	171	84.5	84	1015	0
CD151	254	211	84	1925	1	225	220	88	2002	1
CD164	330	262	88	2801	1	625	548	84	4009	1
CD166	114	28.3	82	415	0	123	35.6	82	454	0
CD200	134	51.5	80	423	0	138	94.8	92	1254	1
CD220	107	17.4	56	175	0	116	29.3	74	265	0
CD227	144	172	80	1552	1	136	59.2	80	735	0
CD271	208	198	86	2746	1	156	113	86	1963	1
CD282	82.8	23.7	72	674	0	77.6	9.22	46	292	0
CD321 (F11 Rcpt	105	77.2	68	804	0	85.2	32.4	58	428	0
CD340 (Her2)	94.2	31.6	61	262	0	85.2	74.4	51	186	0
b2micro globulin	290	285	73	2781	1	236	218	73	2497	1
EGF Receptor	183	121	58	869	0	116	50.2	66	461	0
Hem. Prog. Cell	201	172	63	2014	1	121	66	63	804	0
HLAA,B,C	500	512	73	4530	1	305	328	73	3359	1
HLAA2	348	342	66	3359	1	202	201	68	2366	1
HLADR	110	87.1	68	1353	1	117	87.9	63	963	0
HLADR,DP, DQ	101	63.2	56	1156	1	109	78.5	61	1353	1
Disialoganglioside G	183	400	68	4175	1	197	336	63	4290	1
SSEA-1	79.9	15.9	56	418	0	77.8	11.6	54	428	0
SSEA-4	109	85	56	963	0	119	84.2	63	1097	1
TRA-1-60	82.2	16.8	49	69	0	76.7	10.3	51	123	0
TRA-1-81	80.8	18.2	54	221	0	77.3	17.6	44	191	0
CD104	84	20	47	199	0	101	51.5	56	469	0
CD120b	83.7	22.2	44	211	0	108	66.9	51	846	0
CD132	89.7	30.2	54	269	0	105	58.8	54	623	0
CD201	136	142	56	2183	1	196	256	49	3954	1
CD210	91.2	33.9	49	343	0	107	64.5	49	707	0
utaneous Lymph. An	96.4	37.9	54	428	0	106	57.8	58	623	0
Integrin b7	99.2	44.6	51	388	0	99.5	50	54	578	0

frequently detected in the LP and stromal compartments (Figs. 2 and 4a). These data highlighted the complexities of surface marker expression and that inter-individual variation did not deviate expression patterns in the different epithelial and stromal populations.

To interrogate whether expression of surface makers enriches proliferative capacity in the LP population, purified LP cells were seeded into colony-forming assays. Surprisingly, only cells that were positive for CD13, CD15s or TRA-1-81 surface markers had increased colony-forming capacity (CFC), with the rest of the antibodies tested showing no differences in CFC between cells positive and negatively expressing these markers (Fig. 4b). Previous reports show that ALDH enriches for detection of progenitor/stem activity [3, 10, 36]. Assessing the differential expression of ALDH and the individual surface marker, we resolved the LP population into four subtypes: ALDH+surface marker Ab+, ALDH+ Ab−, ALDH−Ab+, ALDH−Ab− groups (Figure S5a). Twelve surface markers were assessed for co-expression and only CD73, SSEA-4 or TRA-1-81 surface markers overlapped with ALDH expression. For instance, 6.8% of LP cells co-expressed ALDH and CD73 compared to 1.6% of LP cells that expressed CD73 only. Similar patterns were observed for the other two markers, with 7.0% of LP cells being ALDH$^+$SSEA-4$^+$ vs 3.6% LP cells SSEA-4$^+$ and 13.2% of LP cells expressed ALDH$^+$TRA-1-81$^+$ vs 5.0% TRA-1-81$^+$ LP cells (Figure S5a). The remaining markers were either equally distributed between ALDH positive and negative expressing cells (i.e. CD13, CD282, CD63, CD75 or CLA) or had inverse expression with ALDH expressing cells (i.e. CD15s, CD151, CD16 or CD15) (Figure S5a). Interestingly, majority of ALDH positive cells were also CD13 and CD282 positive, whilst the remaining markers only contained a smaller proportion of co-expression with ALDH positivity (Figure S5a). This result demonstrates even greater heterogeneity of the LP progenitor population beyond that can be further refined by CD markers and ALDH expression. Furthermore, colony-forming assays show that co-expression of CD73, CD282 or TRA-1-81 cell surface markers together with ALDH enriches progenitor capacity. Co-expression of ALDH and CD151 or CD15 markers increased progenitor capacity compared to cells positive for CD151 or CD15 alone (Figure S5b). An exception to this finding was CLA. CLA labelled cells had the lowest progenitor activity, suggesting CLA expression may indicate a committed LP cell subtype (Figure S5b). Showing that some of the novel/less characterised markers identified may determine different LP cell states.

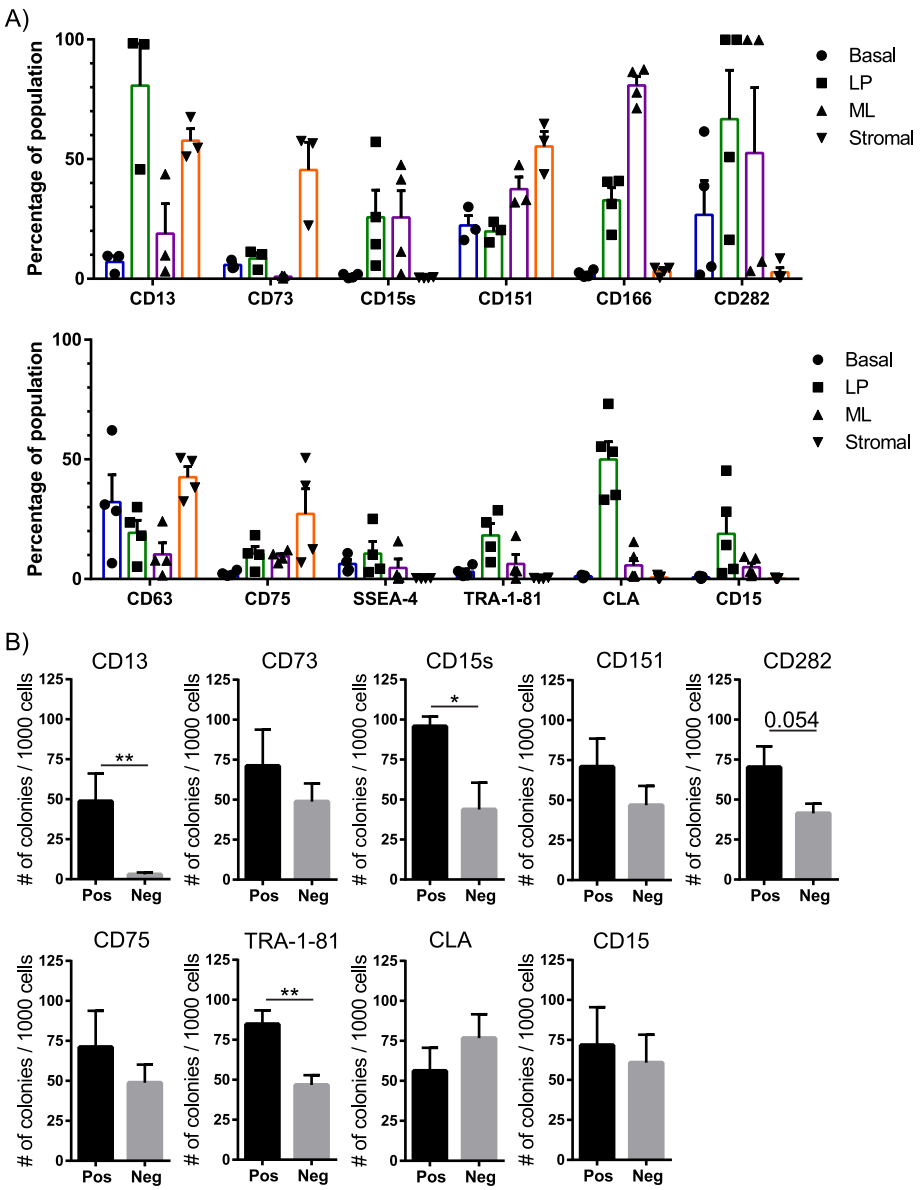


Fig. 4 Lyoplate screen identifies novel luminal progenitor markers. **A** Analysis of variability in expression of 12 surface markers enriched in the luminal progenitor population. Bar charts show percentage of positive marker cells in each epithelial/stromal population, all error bars represent SEM. n = 3–5 independent human breast donor samples. **B** Bar chart showing the colony-forming ability of the luminal progenitor population from positive and negative surface marker expressing cells. n = 3–5 independent human breast donor samples, all error bars represent SEM. Statistical significance was calculated using two-tailed *t* test. Statistically significant differences are indicated by asterisks. * $P < 0.05$, ** $P < 0.001$

Discussion

This screen uncovered greater diversity of surface marker expression among epithelial and stromal cell lineages in normal breast tissues than what is currently reported and, whilst not a complete study of all possible cell surface markers, is a starting point for generating an overview of all surface marker expression patterns on breast epithelial and stromal cells. We identified pan-breast tissue markers such as CD9, CD54, CD59, CD164 or HLA-A,B,C that were strongly expressed in majority of epithelial and stromal breast cells, luminal lineage enriched markers including CD13, CD15, CD24, CD75, CD166, CD227, CD282 and markers that were enriched within the basal compartments, such as CD49a, CD90, CD200, CD271. This screen confirms CD expression of several well-characterised breast epithelial markers (CD10, CD24, CD44, CD227), and identified several novel surface markers including CD15s, CD75, CD164, CD282, TRA-1-81, among others.

Here we compiled a searchable dataset of surface marker expression for human breast epithelial and stromal populations that allows greater ability to refine the CDs that are functionally important for human breast development. For instance, we found that CD13 a proteolytic enzyme also known as Anpep, was strongly expressed in the LP population. *Anpep*$^{-/-}$ knockout mice have delayed mammary gland development during pregnancy attributed to reduced branching morphogenesis within the duct [37]. Furthermore, transgenic mice that overexpress human ANPEP display a reciprocal phenotype including mammary glands that are hyper-branched during pregnancy [37]. The LP population is known as the secretory luminal cell type involved in alveologenesis and milk production during pregnancy. The reporting of CD13 expression in the luminal compartment, especially the LP population supports the hypothesis that CD13 positive cells may contribute to breast morphogenesis during pregnancy. This finding demonstrates the potential for our screen in identifying different cell states within the epithelial compartment, data from which can be used to explore the role of these cells within the breast tissue development.

Breast cellular heterogeneity remains a key obstacle in understanding the transition of normal cells towards cancer and how different breast cancer subtypes develop. Our screen provides a starting point for identifying novel as well as other less characterised cell surface markers that could be useful for diagnostic as well as predictive of disease progression or defining invasive tumours. The ability to identify a cell type based on marker expression/s that enables cancer development can then be used as a therapy target. CD44 has been the subject of intense breast cancer research for several decades and is considered one such example of a surface marker that is used diagnostically and for therapy. The COSMIC database reported 3.7% of breast cancers overexpress CD44 whilst 2.8% of breast cancer samples contain mutations in CD44 [38]. However, the data surrounding the role of CD44 in cancer stem cells (CSCs) or its prognostic ability can be conflicting [39]. This screen reveals that CD44 is highly expressed in all normal breast epithelial populations and corroborates previous immunostaining [5]. Therefore, it is evident that CD44 marks several cell states in normal breast and breast cancer tissues, including cells that have CSC and non-CSC roles. Our screen assessed several known CD markers and identified several novel (i.e. CD63, CD98 and CD164) and less characterised (i.e. CD46, CD107a and CD321) breast epithelial markers, of which are overexpressed in at least 5% of breast cancer samples in the COSMIC database [38]. Many studies reporting overexpression of particular markers are not always substantiated when considering the proportion of expression detected in the normal tissue. For example, CD9 is overexpressed in 10% of breast cancers [38] and has been implicated in breast tumour invasion [40], yet CD9 is expressed in approximately 90% of all normal breast epithelial cells, highlighting that CD9 may mark diverse cell functions in the different epithelial cell lineages. Use of this dataset can determine the cell types containing expression of the surface marker in the normal breast tissue and whether these CD markers are then overexpressed in cancer.

Focused investigations on a single surface marker can assist in understanding biological function of that particular marker. However, combinatorial analysis of markers will enhance our understanding of cell states in normal breast biology and tumour heterogeneity. Multiplatform single cell technologies have rapidly identified the proteomic landscape of normal breast tissue and breast cancer. However, the use of surface markers without a clear understanding of the expression pattern in breast epithelial/stromal populations may lead to interpretation difficulties of the omic data generated. Mass cytometry/imaging mass cytometry techniques have enhanced the single cell phenotypic capacity by simultaneously detecting up to triple the number of markers achieved by conventional flow cytometry. Greater proteomic and spatial architecture atlas of the breast tumour ecosystem [22, 41, 42] and normal tissue across aged breasts [43] has yielded better connections between different cell lineages. However, these datasets are limited by the availability of known surface markers for breast tissue including CD44, Her2/CD340, EGFR, CD24 markers [22, 42, 44]. A recent publication utilised several less described surface markers in relation to normal breast biology including CD47, CD54, CD73 and CD95 [26]. Using our resource, CD47 and CD54 were detected in all breast cell subpopulations at a frequency of 30-80% for each cell population, whilst CD73 and CD95 positive cells were predominately located in the LP and stromal compartments.

Conclusions

Our resource can enhance multiplatform system such as complex surface marker staining, mass cytometry, single cell omic studies for cell lineage clarity. Using this surface marker dataset, we have identified cell lineage antibodies in addition to the standard panel of Lineage/EpCAM/CD49f which can be used to investigate the variation in epithelial and stromal compartments. These panels include (but not limited to) CD15s/CD73/CLA for further investigation into the LP compartment, whilst targeting the ML population can be carried out with the addition of CD166/CD227/CD340. Investigating heterogeneity within the basal compartment can be performed using CD29/CD142/CD271 antibodies, and antibodies

targeting the stromal compartment include CD34/CD39/CD140b. These panels can be used in conventional cytometry for recoverable cellular material and further functional studies into normal breast and cancer development. Currently, mass cytometry/imaging mass cytometry datasets are limited to using known and available antibodies and many cell surface markers have not been previously reported. Multiplexing many of the surface markers that were identified in this study allows further investigation into spatial locations and relationships between different cell types in order to understand normal/disease development and functions in the breast.

Abbreviations

ALDH: Aldehyde dehydrogenase; CD: Cluster of differentiation; CFC: Colony-forming capacity; FC: Flow cytometry; K: Keratin; LP: Luminal progenitor; ML: Mature luminal; TDLU: Terminal ductal lobular unit

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13058-021-01444-5.

Additional file 1: Supplemental Figure S1. Gating strategy of antibody screen. Gating strategy to eliminate debris, doublets, dead and endothelial cells and to select the epithelial and stromal subpopulations. Percentage of positive cells for each antibody was determined based on gates drawn from the isotype control for each of the subpopulations. Gating strategy illustrated represents A) a negative and B) a positive surface marker.

Additional file 2: Supplemental Figure S2. Multiplexing reduces the number of positive surface marker antibodies. Pie charts depicting the proportion of positive surface marker detected in human breast single cell suspension containing A) all live cells depleted of endothelial cells, B) all live cell types and C) all cell types.

Additional file 3: Supplemental Figure S3. Positive surface marker expression patterns in different mammary epithelial and stromal subpopulations. Histograms show intensity staining for all positive identified antibody surface markers compared with isotype controls (red) and the duplicates of the screen in the basal, LP, ML and stromal (Black and Grey) cells on a log scale.

Additional file 4: Supplemental Figure S4. Validation of lyoplate screen. Representative FACS analysis depicting surface marker expression in the different epithelial/stromal subpopulations (blue: basal, green: LP, purple: ML, orange: stromal positive surface marker cells, grey illustrates negative cells).

Additional file 5: Supplemental Figure S5. Luminal progenitor activity in surface marker and ALDH expression. A) Analysis of variability in expression of ALDH and the 12 surface markers in the luminal progenitor populations. Bar charts show percentage of positive marker cells in each of the LP subpopulations, all error bars represent SEM. n=3-5 independent human breast donor samples. B) Stacked bar chart showing the colony forming ability of the luminal progenitor ALDH-Ab+/ALDH+Ab+ or ALDH+Ab-/ALDH-Ab- subpopulations. n=3-5 independent human breast donor samples, error bars represent SEM. Statistical significance was calculated using an ANOVA and Tukey's multiple comparison test. Statistical significance differences are indicated by asterisks * $p< 0.05$ and ** $p< 0.01$.

Additional file 6.

Acknowledgements

The authors thank G. Tarulli and E. Watson for critical reading of the manuscript. The authors thank L. Jones for facilitating and coordinating access to human tissue.

Authors' contributions

SV, RS and MS designed the study. SV, RS and MS performed the Lyoplate screen. MS performed the validation experiments, data analysis and interpreted the data. SV and RS participated in the paper editing. MS supervised the project and wrote the manuscript. CC and JS secured funding for the project. All authors read and approved the final manuscript.

Funding

This work was funded by Cancer Research UK, the University of Cambridge, and NIHR Cambridge Biomedical Research Centre. MS is currently funded by Tzo Tze Ang and Eashwar Krishnan.

Availability of data and materials

The dataset generated, used and analysed in this study available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

We received approval for all primary human breast material use, under full informed consent and in accordance with the National Research Ethics Service, Cambridgeshire 2 Research Ethics Committee approval (08/H0308/178) as part of the Adult Breast Stem Cell Study.

Consent for publication

Not applicable

Competing interests

JS is a paid consultant for StemCell Technologies Inc. CC is a member of the External Science Panel of AstraZeneca, and his laboratory has received research grants (administered by the University of Cambridge) from Genentech, Roche, AstraZeneca and Servier. All remaining authors declare that they have no competing interests.

Author details

[1]CRUK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK. [2]Cambridge Institute for Medical Research, Cambridge University, Cambridge CB2 0XY, UK. [3]Cambridge Breast Unit, Addenbrookes Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge, UK. [4]Medical Research Council Cancer Unit, University of Cambridge, Hills Road, Cambridge CB2 0XZ, UK.

Received: 9 June 2020 Accepted: 31 May 2021
Published online: 13 June 2021

References

1. Petersen OW, Høyer PE, van Deurs B. Frequency and distribution of estrogen receptor-positive cells in normal, nonlactating human breast tissue. Cancer Res. 1987;47(21):5748–51.
2. Sleeman KE, Kendrick H, Robertson D, Isacke CM, Ashworth A, Smalley MJ. Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. J Cell Biol. 2007;176(1):19–26. https://doi.org/10.1083/jcb.200604065.
3. Shehata M, Teschendorff A, Sharp G, Novcic N, Russell IA, Avril S, et al. Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. Breast Cancer Res. 2012;14(5):R134. https://doi.org/10.1186/bcr3334.
4. Wellings SR, Jensen HM, Marcum RG. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. J Natl Cancer Inst. 1975;55(2):231–73.
5. Santagata S, Thakkar A, Ergonul A, Wang B, Woo T, Hu R, et al. Taxonomy of breast cancer based on normal cell phenotype predicts outcome. J Clin Invest. 2014;124(2):859–70. https://doi.org/10.1172/JCI70941.
6. Gusterson BA, Ross DT, Heath VJ, Stein T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. Breast Cancer Res. 2005;7(4):143–8. https://doi.org/10.1186/bcr1041.
7. Stingl J, Eaves CJ, Kuusk U, Emerman JT. Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. Differentiation. 1998;63(4):201–13. https://doi.org/10.1111/j.1432-0436.1998.00201.x.

8. Stingl J, Eaves CJ, Zandieh I, Emerman JT. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. Breast Cancer Res Treat. 2001;67(2):93–109. https://doi.org/10.1023/A:1010615124301.
9. Gudjonsson T, Villadsen R, Nielsen HL, Rønnov-Jessen L, Bissell MJ, Petersen OW. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. Genes Dev. 2002;16(6):693–706. https://doi.org/10.1101/gad.952602.
10. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome. Cell Stem Cell. 2007;1(5):555–67. https://doi.org/10.1016/j.stem.2007.08.014.
11. Garbe JC, Pepin F, Pelissier FA, Sputova K, Fridriksdottir AJ, Guo DE, et al. Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia. Cancer Res. 2012;72(14):3687–701. https://doi.org/10.1158/0008-5472.CAN-12-0157.
12. Keller PJ, Lin AF, Arendt LM, Klebba I, Jones AD, Rudnick JA, et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. Breast Cancer Res. 2010;12(5):R87. https://doi.org/10.1186/bcr2755.
13. Villadsen R, Fridriksdottir AJ, Rønnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA, et al. Evidence for a stem cell hierarchy in the adult human breast. J Cell Biol. 2007;177(1):87–101. https://doi.org/10.1083/jcb.200611114.
14. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009;15(8):907–13. https://doi.org/10.1038/nm.2000.
15. Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N, et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nat Commun. 2018;9:1–12.
16. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature. 2015;526(7571):131–5. https://doi.org/10.1038/nature15260.
17. Shehata M, Stingl J. Purification of distinct subsets of epithelial cells from normal human breast tissue. Methods Mol Biol. 2017;1501:261–76. https://doi.org/10.1007/978-1-4939-6475-8_13.
18. Stingl J, Raouf A, Emerman JT, Eaves CJ. Epithelial Progenitors in the Normal Human Mammary Gland. J Mammary Gland Biol Neoplasia. 2005;10(1):49–59. https://doi.org/10.1007/s10911-005-2540-7.
19. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, et al. Defining the cellular precursors to human breast cancer. PNAS. 2012;109(8):2772–7. https://doi.org/10.1073/pnas.1017626108.
20. Suzuki RN, Entwistle A, Atherton AJ, Clarke C, Lakhani SR, O'Hare MJ. The expression patterns of integrin subunits on human breast tissues obtained during pregnancy. Cell Biol Int. 2002;26(7):593–8. https://doi.org/10.1006/cbir.2002.0880.
21. Yang XH, Richardson AL, Torres-Arzayus MI, Zhou P, Sharma C, Kazarov AR, et al. CD151 accelerates breast cancer by regulating α6 integrin function, signaling, and molecular organization. Cancer Res. 2008;68(9):3204–13. https://doi.org/10.1158/0008-5472.CAN-07-2949.
22. Wagner J, Rapsomaniki MA, Chevrier S, Anzeneder T, Langwieder C, Dykgers A, et al. A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. Cell. 2019;177:1330–45 e18.
23. Kumar B, Prasad M, Bhat-Nakshatri P, Anjanappa M, Kalra M, Marino N, et al. Normal breast-derived epithelial cells with luminal and intrinsic subtype-enriched gene expression document interindividual differences in their differentiation Cascade. Cancer Res. 2018;78(17):5107–23. https://doi.org/10.1158/0008-5472.CAN-18-0509.
24. Nakshatri H, Anjanappa M, Bhat-Nakshatri P. Ethnicity-dependent and -independent heterogeneity in healthy normal breast hierarchy impacts tumor characterization. Sci Rep. 2015;5:1–14.
25. Fridriksdottir AJ, Villadsen R, Morsing M, Klitgaard MC, Kim J, Petersen OW, et al. Proof of region-specific multipotent progenitors in human breast epithelia. PNAS. 2017;114(47):E10102–11. https://doi.org/10.1073/pnas.1714063114.
26. Rosenbluth JM, Schackmann RCJ, Gray GK, Selfors LM, Li CM-C, Boedicker M, et al. Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. Nat Commun. 2020;11(1):1711. https://doi.org/10.1038/s41467-020-15548-7.
27. Fridriksdottir AJ, Kim J, Villadsen R, Klitgaard MC, Hopkinson BM, Petersen OW, et al. Propagation of oestrogen receptor-positive and oestrogen-responsive normal human breast cells in culture. Nat Commun. 2015;6(1):8786. https://doi.org/10.1038/ncomms9786.
28. Morsing M, Klitgaard MC, Jafari A, Villadsen R, Kassem M, Petersen OW, et al. Evidence of two distinct functionally specialized fibroblast lineages in breast stroma. Breast Cancer Res. 2016;18(1):108. https://doi.org/10.1186/s13058-016-0769-2.
29. Jamil F, Peston D, Shousha S. CD9 immunohistochemical staining of breast carcinoma: Unlikely to provide useful prognostic information for routine use. Histopathology. 2001;39(6):572–7. https://doi.org/10.1046/j.1365-2559.2001.01296.x.
30. Croce MV. Lewis x is Highly Expressed in Normal Tissues: a Comparative Immunohistochemical Study and Literature Revision. Pathol. Oncol. Res. 2007;13(2):130–8. https://doi.org/10.1007/BF02893488.
31. Defilippis RA, Chang H, Dumont N, Rabban JT, Chen Y-Y, Fontenay G V, et al. CD36 Repression Activates a Multicellular Stromal Program Shared by High Mammographic Density and Tumor Tissues. 2012.
32. Thorsteinsson L, O'Dowd GM, Harrington PM, Johnson PM. The complement regulatory proteins CD46 and CD59, but not CD55, are highly expressed by glandular epithelium of human breast and colorectal tumour tissues. APMIS. 1998;106(7-12):869–78. https://doi.org/10.1111/j.1699-0463.1998.tb00233.x.
33. Wang Q, Yao J, Jin Q, Wang X, Zhu H, Huang F, et al. LAMP1 expression is associated withpoor prognosis in breast cancer. Oncol Lett. 2017;14(4):4729–35. https://doi.org/10.3892/ol.2017.6757.
34. Pinheiro C, Albergaria A, Paredes J, Sousa B, Dufloth R, Vieira D, et al. Monocarboxylate transporter 1 is up-regulated in basal-like breast carcinoma. Histopathology. 2010;56(7):860–7. https://doi.org/10.1111/j.1365-2559.2010.03560.x.
35. Naik MU, Naik TU, Suckow AT, Duncan MK, Naik UP. Attenuation of junctional adhesion molecule-A is a contributing factor for breast cancer cell invasion. Cancer Res. 2008;68(7):2194–203. https://doi.org/10.1158/0008-5472.CAN-07-3057.
36. Eirew P, Kannan N, Knapp DJHF, Vaillant F, Emerman JT, Lindeman GJ, et al. Aldehyde Dehydrogenase Activity Is a Biomarker of Primitive Normal Human Mammary Luminal Cells. Stem Cells. 2012;30(2):344–8. https://doi.org/10.1002/stem.1001.
37. Kolb AF, Sorrell D, Lassnig C, Lillico S, Carlisle A, Neil C, et al. Mammary gland development is delayed in mice deficient for aminopeptidase N. Transgenic Res. 2013;22(2):425–34. https://doi.org/10.1007/s11248-012-9654-7.
38. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941–7. https://doi.org/10.1093/nar/gky1015.
39. Louderbough JMV, Schroeder JA. Understanding the dual nature of CD44 in breast cancer progression. Mol. Cancer Res. 2011;9(12):1573–86. https://doi.org/10.1158/1541-7786.MCR-11-0156.
40. Rappa G, Green TM, Karbanová J, Corbeil D, Lorico A. Tetraspanin CD9 determines invasiveness and tumorigenicity of human breast cancer cells. Oncotarget. 2015;6(10):7970–91. https://doi.org/10.18632/oncotarget.3419.
41. Ali HR, Jackson HW, Zanotelli VRT, Danenberg E, Fischer JR, Bardwell H, et al. Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. Nat Cancer. 2020;1(2):163–75. https://doi.org/10.1038/s43018-020-0026-6.
42. Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, et al. The single-cell pathology landscape of breast cancer. Nature. 2020;578(7796):615–20. https://doi.org/10.1038/s41586-019-1876-x.
43. Pelissier Vatter FA, Schapiro D, Chang H, Borowsky AD, Lee JK, Parvin B, et al. High-Dimensional Phenotyping Identifies Age-Emergent Cells in Human Mammary Epithelia. Cell Rep. 2018;23(4):1205–19. https://doi.org/10.1016/j.celrep.2018.03.114.
44. Giesen C, Wang HAO, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods. 2014;11(4):417–22. https://doi.org/10.1038/nmeth.2869.

Publisher's Note