

**Supplementary Table 1. Origin of tissue collected from reduction mammoplasty surgeries.**

<b>Patient</b>	<b>Age Range (yrs)</b>	<b>Year Collected</b>	<b>Use of Hormonal Contraceptives</b>	<b>Number of Pregnancies</b>
1	>20	2007	Yes	0
2	>20	2009	Yes	Unknown
3	20-24	2010	Yes	0
4	20-24	2009	Yes	0
5	20-24	2009	No	Unknown
6	20-24	2009	Yes	0
7	25-29	2008	Unknown	Unknown
8	25-29	2009	Yes	0
9	25-29	2009	Unknown	Unknown
10	35-39	2007	Unknown	3
11	35-39	2008	Unknown	Unknown
12	35-39	2008	Mirena IUD	3
13	35-39	2009	Yes	3
14	40-44	2007	No, IVF treatment	0
15	40-44	2009	No, fertility treatment	2
16	50-54	2009	Unknown	Unknown
17	50-54	2007	Unknown	1
18	50-54	2007	Unknown	Unknown
19	55-59	2008	Unknown	Unknown

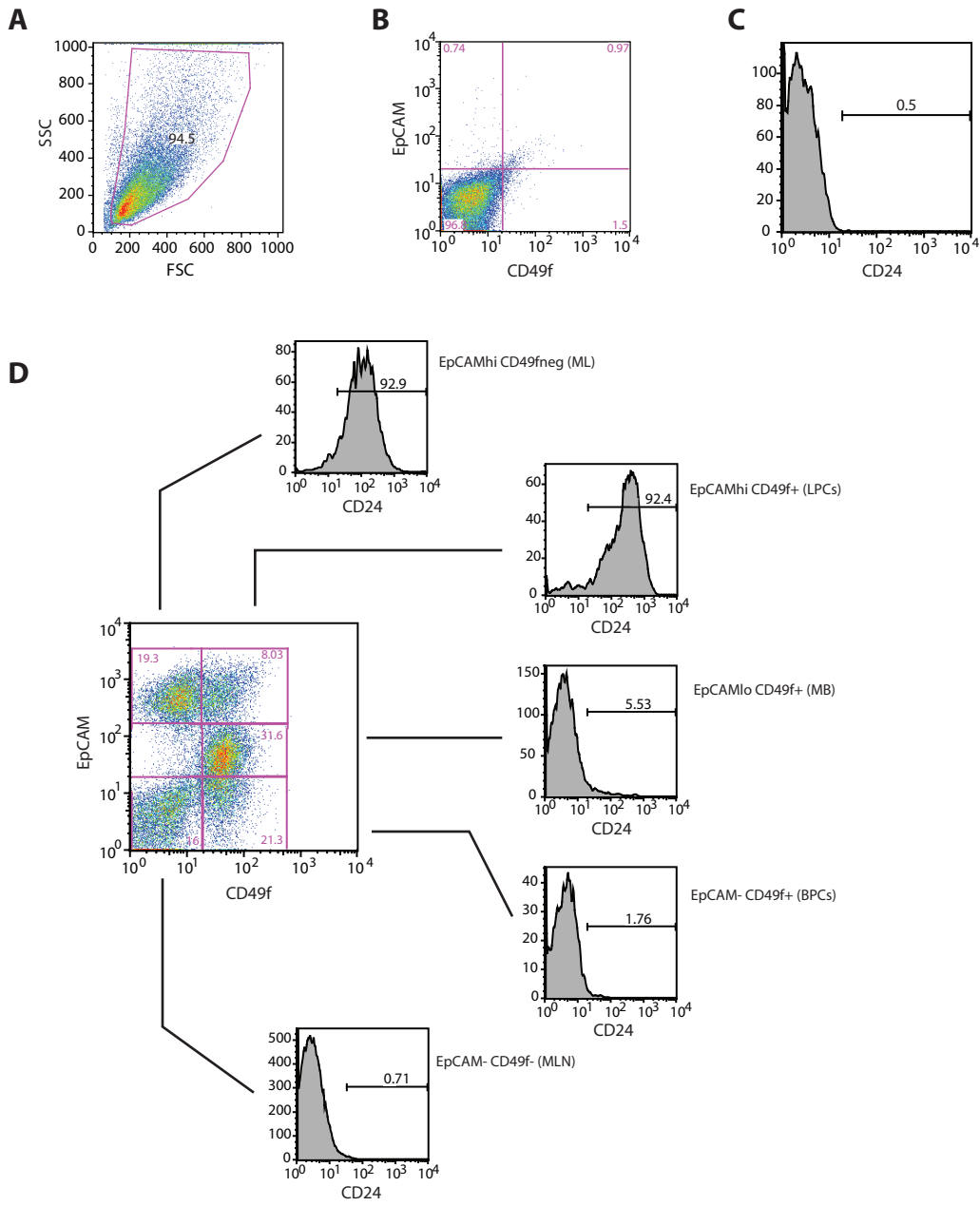
**Supplementary Table 2. Percentage of Mature Luminal, Luminal Progenitor, Mature Basal, and Basal Progenitor Cell Populations.**

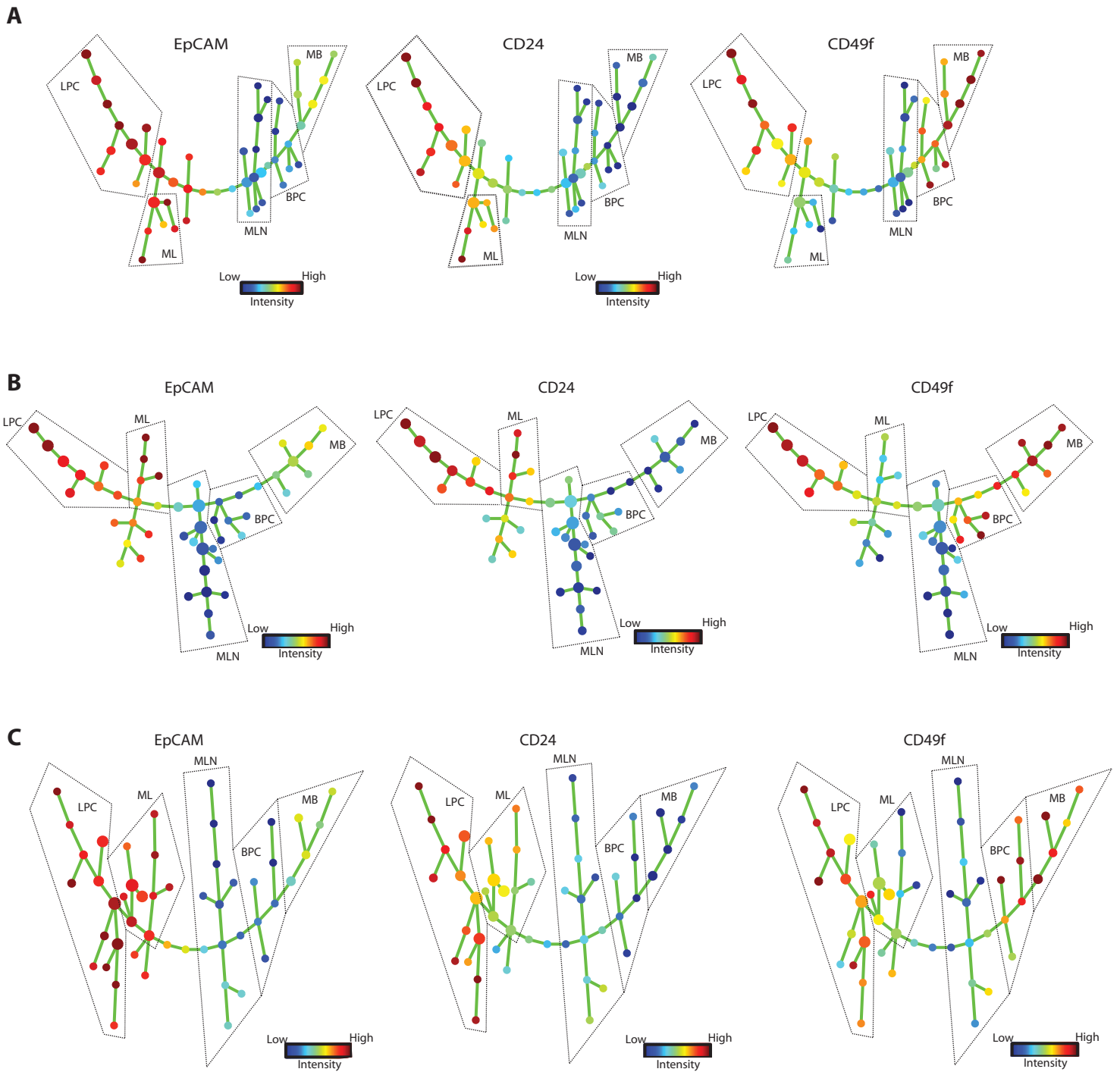
<b>Patient</b>	<b>Age Range (yrs)</b>	<b>EpCAM<sup>hi</sup> CD49f<sup>-</sup> (ML)</b>	<b>EpCAM<sup>hi</sup> CD49f<sup>+</sup> (LPCs)</b>	<b>EpCAM<sup>lo</sup> CD49f<sup>+</sup> (MB)</b>	<b>EpCAM<sup>-</sup> CD49f<sup>+</sup> (BPCs)</b>
1	>20	57.0	13.1	13.6	16.3
2	>20	9.8	38.4	33.4	18.4
4	20-24	27.3	45.7	12.9	14.1
6	25-29	19.1	48.0	13.4	19.6
7	25-29	16.5	58.2	17.9	7.4
8	25-29	79.4	1.6	1.9	17.1
10	35-39	11.0	44.0	15.0	30.0
11	35-39	86.4	1.2	5.1	7.2
12	35-39	32.6	47.8	10.1	9.6
13	35-39	21.0	35.0	11.0	33.0
14	40-44	31.2	13.6	43.2	12.0
15	40-44	41.0	18.0	25.0	16.0
16	45-49	52.6	14.7	21.4	11.4
17	50-54	96.4	0.4	1.0	2.1
19	55-59	63.7	20.2	7.7	8.5

**Supplementary Table 3. Frequency of progenitor cells measured by *in vitro* assays of patient derived mammary epithelial cells.**

<b>Patient</b>	<b>Age Range (yrs)</b>	<b>Freq Adherent Colonies</b>	<b>Enrichment of Colonies</b>	<b>Freq CK8/14 Colonies</b>	<b>Freq Mammo-spheres</b>	<b>Freq Floating Colonies</b>	<b>Freq 3D Colonies</b>
1	>20	0.0074	Basal	0.00047	0.004	NT	NT
2	>20	0.0049	Basal	0.0017	0.0038	0.0023	0.0073
3	20-24	0.0066	Luminal	0.001	0.0039	0.0026	0.0006
4	20-24	0.00084	Basal	0.00028	0.004	0.0035	0.0063
5	20-24	0.015	Luminal	0.003	0.018	0.015	0.0043
6	25-29	0.0093	Basal	0.027	0.0069	0.0022	0.0024
7	25-29	0.0043	Luminal	0.00035	0.0028	0.0019	NT
8	25-29	0.042	Luminal	0.00045	0.0089	0.0082	0.0016
9	25-29	0.014	Basal	0.0016	0.011	0.0007	0.0016
10	35-39	0.0058	Luminal	0.00095	0.014	0.0081	NT
11	35-39	0.0094	Basal	0.0071	0.0051	0.0021	NT
12	35-39	0.0063	Luminal	0.00062	0.0078	0.0086	0.0026
13	35-39	0.003	Luminal	0.00008	0.0091	0.0073	0.0018
14	40-44	0.029	Basal	0.0037	0.0056	0.0033	NT
15	40-44	0.012	Basal	0.00043	0.0056	0.00035	0.0044
16	45-49	0.0072	Luminal	0.0003	0.0045	0.0025	0.00083
17	50-54	0.0018	Luminal	0.00028	0.0036	0.0044	0.001
18	50-54	0.0033	Luminal	0.0014	0.0067	0.0042	0.003
19	55-59	0.0041	Luminal	0.00048	NT	NT	NT

NT=not tested





**Figure S1. Flow cytometry gating strategy for EpCAM, CD49f, and CD24.** (A) Cell population gated to remove cellular debris and doublets. (B) Gates for EpCAM<sup>+</sup> and CD49f<sup>+</sup> cells based on isotype controls for each antibody. (C) Gate for CD24<sup>+</sup> cells based on isotype control for CD24 antibody. (D) Gates for each cellular population: Mature Luminal (ML) EpCAM<sup>hi</sup>CD49f<sup>-</sup>, Luminal Progenitor Cells (LPCs) EpCAM<sup>hi</sup>CD49f<sup>+</sup>, Mature Basal (MB) EpCAM<sup>lo</sup>CD49f<sup>+</sup>, Basal Progenitor Cells (BPC) EpCAM<sup>-</sup>CD49f<sup>+</sup>, Mammary Lineage Negative (MLN) EpCAM<sup>-</sup>CD49f<sup>-</sup>. CD24 expression denoted for each population.

**Figure S2. SPADE applied to mammary epithelial cell flow cytometry data.** (A) SPADE tree derived from the flow cytometry data of cell surface markers EpCAM, CD24, CD49f from patient samples with known lobule composition (n=8 patient samples). Trees are colored by the median intensity of each individual marker. Dashed lines denote cell populations based on cell surface markers: Mature Luminal (ML) EpCAM<sup>hi</sup>CD24<sup>+</sup>CD49f<sup>-</sup>, Luminal Progenitor Cells (LPCs) EpCAM<sup>hi</sup>CD24<sup>+</sup>CD49f<sup>+</sup>, Mature Basal (MB) EpCAM<sup>lo</sup>CD24<sup>-</sup>CD49f<sup>+</sup>, Basal Progenitor Cells (BPC) EpCAM<sup>-</sup>CD24<sup>-</sup>CD49f<sup>+</sup>, Mammary Lineage Negative (MLN) EpCAM<sup>-</sup>CD24<sup>-</sup>CD49f<sup>-</sup>. (B, C) SPADE tree derived from the flow cytometry data of cell surface markers EpCAM, CD24, CD49f from patient samples enriched for Type I/Type II (B) and Type III (C) lobules types (n=4 patient samples/group). Trees are colored by the median intensity of each individual marker.