Cell Reports Medicine, Volume *2*

Supplemental information

A single-cell atlas of the healthy breast tissues reveals clinically relevant clusters of breast epithelial cells

Poornima Bhat-Nakshatri, Hongyu Gao, Liu Sheng, Patrick C. McGuire, Xiaoling Xuei, Jun Wan, Yunlong Liu, Sandra K. Althouse, Austyn Colter, George Sandusky, Anna Maria Storniolo, and Harikrishna Nakshatri

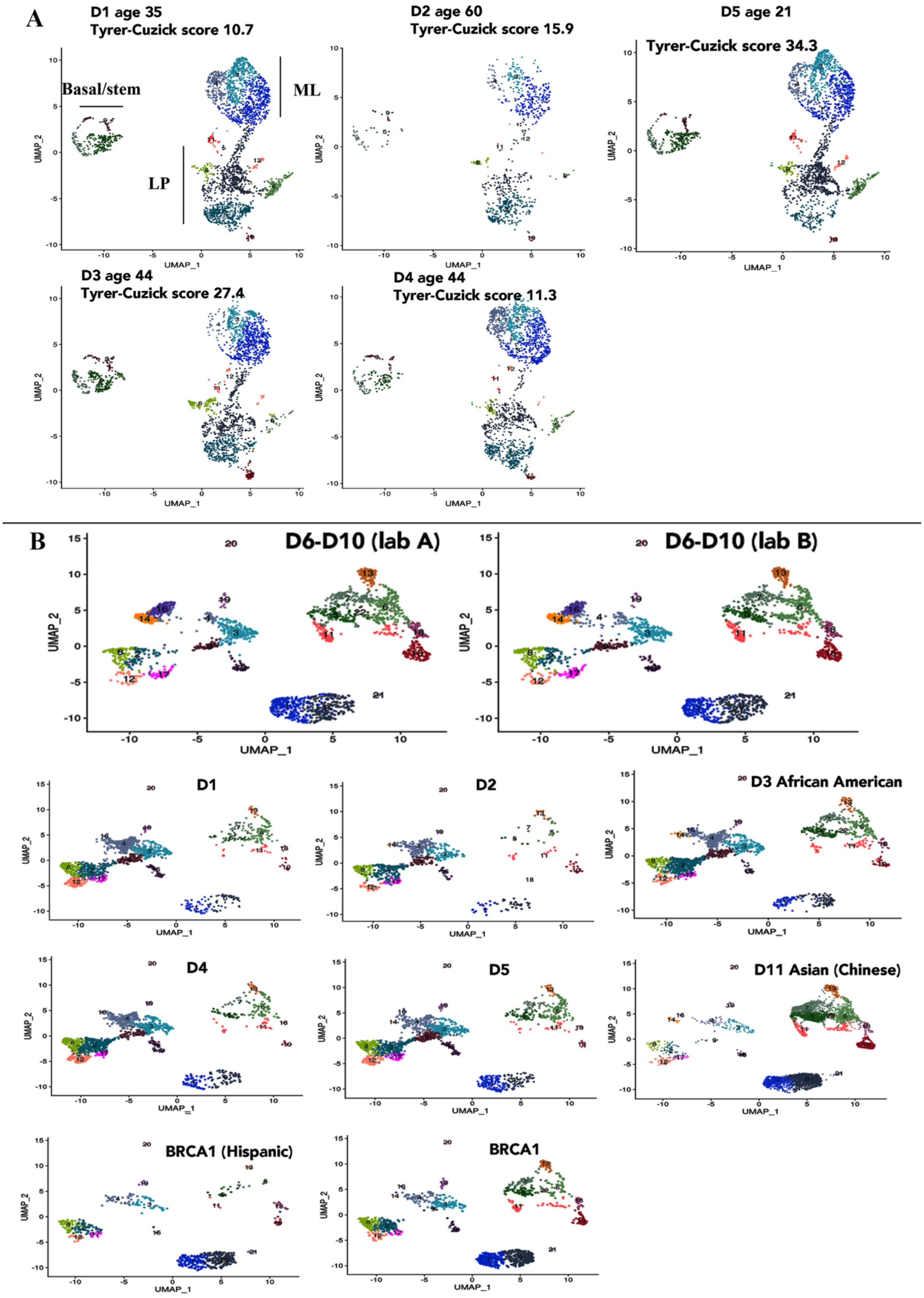
A
D1 age 35
Tyrer-Cuzick score 10.7
Basal/stem
ML
LP
D2 age 60
Tyrer-Cuzick score 15.9
D5 age 21
Tyrer-Cuzick score 34.3
D3 age 44
Tyrer-Cuzick score 27.4
D4 age 44
Tyrer-Cuzick score 11.3
UMAP_1
UMAP_2
B
D6-D10 (lab A)
D6-D10 (lab B)
D1
D2
D3 African American
D4
D5
D11 Asian (Chinese)
BRCA1 (Hispanic)
BRCA1

Figure S1: Epithelial clusters in individual samples. A) Epithelial cell clusters in breast tissues of D1 to D5. B) Epithelial cell clusters in first five samples sequenced individually, additional samples sequenced as a pool in two labs (D6-D10), individual sample from an Asian (Chinese), and a BRCA1 mutation carrier. Clustering was done with Seurat and Loupe browser was used to explore various gene expression. Related to Figures 1 and 4.

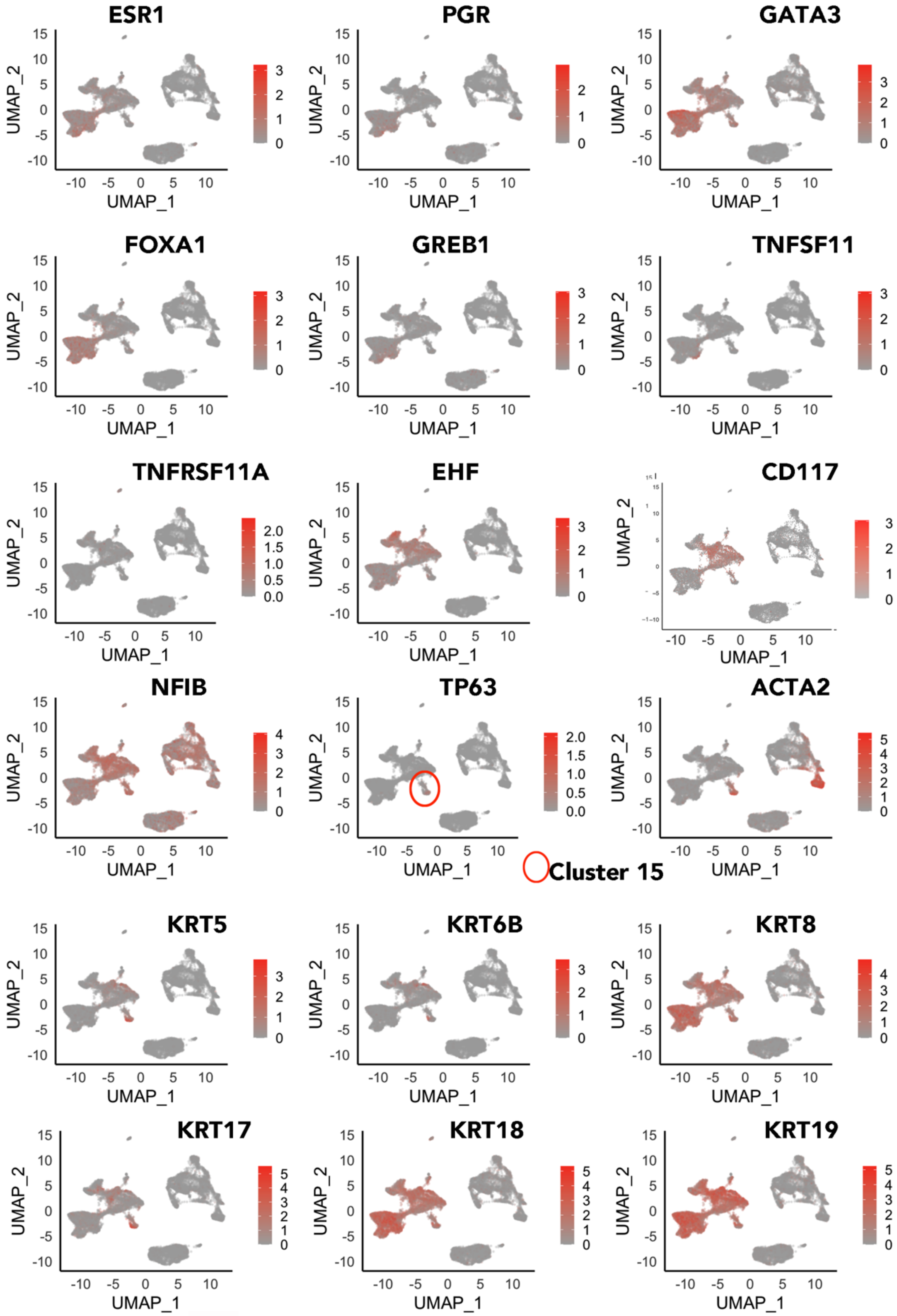
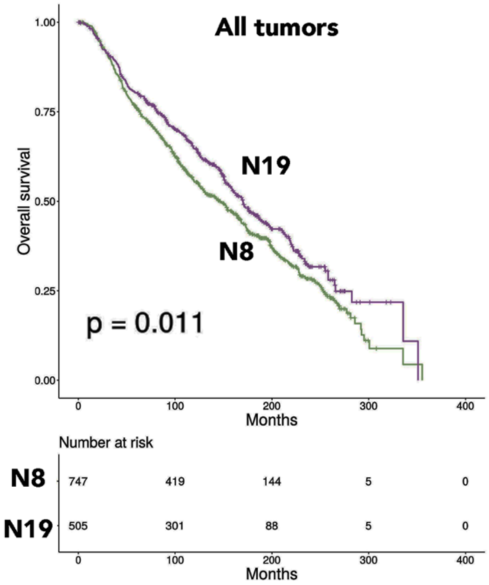
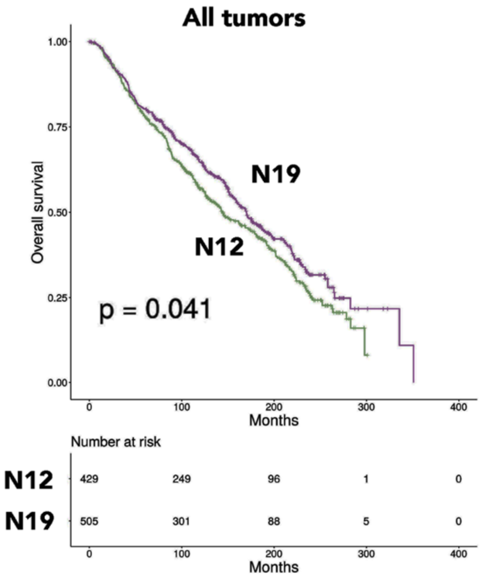
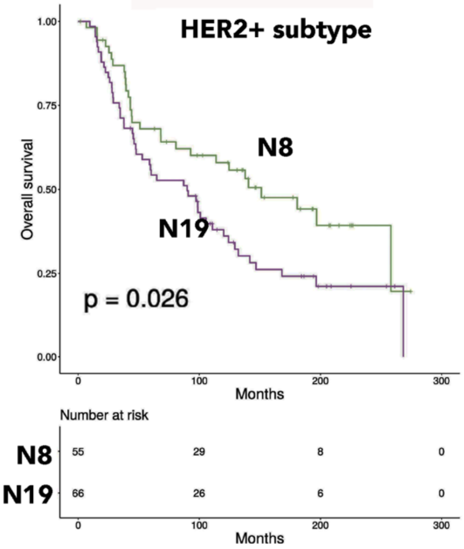
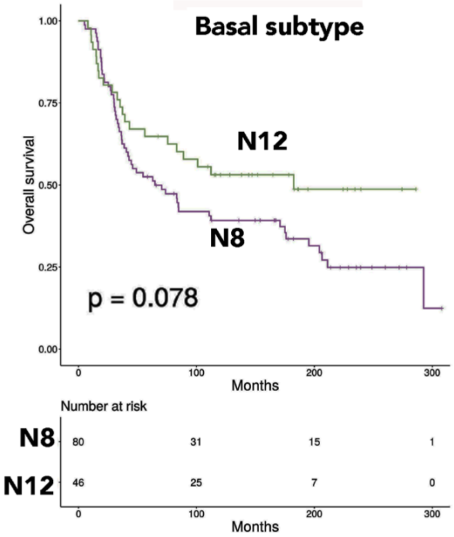

ESR1
PGR
GATA3
FOXA1
GREB1
TNFSF11
TNFRSF11A
EHF
CD117
NFIB
TP63
ACTA2
Cluster 15
KRT5
KRT6B
KRT8
KRT17
KRT18
KRT19
UMAP_1
UMAP_2
15
10
5
0
-5
-10

Figure S2: Expression patterns of known mature luminal (ML), luminal progenitor (LP) and basal/stem/myoepithelial cell-enriched marker genes and different keratins in various clusters. Related to Figure 4.


All tumors
Overall survival
1.00
0.75
0.50
0.25
0.00
N19
N8
p = 0.011
0
100
200
300
400
Months
Number at risk
N8
747
419
144
5
0
N19
505
301
88
5
0
0
100
200
300
400
Months


All tumors
Overall survival
1.00
0.75
0.50
0.25
0.00
N19
N12
p = 0.041
0
100
200
300
400
Months
Number at risk
N12
429
249
96
1
0
N19
505
301
88
5
0
0
100
200
300
400
Months


HER2+ subtype
Overall survival
1.00
0.75
0.50
0.25
0.00
N8
N19
p = 0.026
0
100
200
300
Months
Number at risk
N8
55
29
8
0
N19
66
26
6
0
0
100
200
300
Months


Basal subtype
Overall survival
1.00
0.75
0.50
0.25
0.00
N12
N8
p = 0.078
0
100
200
300
Months
Number at risk
N8
80
31
15
1
N12
46
25
7
0
0
100
200
300
Months

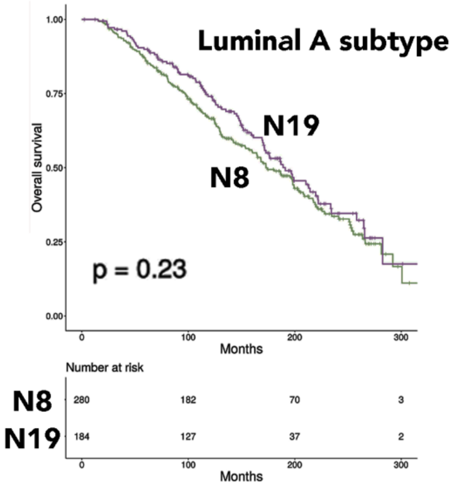

Luminal A subtype
Overall survival
1.00
0.75
0.50
0.25
0.00
N19
N8
p = 0.23
0
100
200
300
Months
Number at risk
N8
280
182
70
3
N19
184
127
37
2
0
100
200
300
Months

Figure S3: Prognostic value of cluster-enriched genes in various intrinsic subtypes of breast cancer. Data were generated using METABRIC datasets. Related to Figure 4.

Table S1: Information on breast tissue donors. Related to data presented in Figures 1 and 4

Donor number	**Race/ethnicity**	**Age**	**BMI**	**Gender**	**Times pregnant**	**Family history of BC**	**Tyrer-Cuzick-lifetime risk score**
D1	White- non-Hispanic or Latina	35	24.5	Female	1	NA	10.7
D2	White- non-Hispanic or Latina	60	24.9	Female	4	Sisters	15.9
D3	African American	44	29	Female	4	Mother, Grandmother and Aunt	27.4
D4	White- non-Hispanic or Latina	42	27.8	Female	1	No	11.3
D5	White- non-Hispanic or Latina	21	35.4	Female	0	Paternal Grandmother	34.3
D6	White- non-Hispanic or Latina	33	23.4	Female	2	Paternal Grandmother	20.1
D7	White- non-Hispanic or Latina	27	22.7	Female	0	No	14.8
D8	White- non-Hispanic or Latina	56	23.5	Female	3	No	8.8
D9	White- non-Hispanic or Latina	24	30.5	Female	3	NA	12.0
D10	White- non-Hispanic or Latina	34	26.8	Female	2	No	14.3
D11	Asian	43	25.6	Female	3	Mother	13

NA=Not available.

Tyrer-Cuzick score >20 indicates higher risk.

Table S4: Overlap in gene expression between clusters of two analyses. Related to Figures 1, 4 and 6A.

Analysis-2 clusters	Peak similarity to analysis 1	P value of overlap	Differentiation status in both sets of analyses	Number of samples in TCGA with cluster-specific gene expression enrichment	Number of samples in METABRIC with cluster-specific gene expression enrichment
N0	C12	1.63E-58	Luminal progenitor	34	58
N1	C12	2.28E-60	Luminal progenitor	9	38
N2	C1,3,4	2.40E-148	Mature luminal		
N3	C0	6.30E-182	Luminal progenitor	8	36
N4	C2	8.38E-95	Luminal Progenitor	47	
N5	C5,7,9	1.90E-235	Basal		
N6	C5,7,9	1.40E-106	Basal		
N7	C5,7,9	8.80E-138	Basal		
N8	C1,3,4	6.80E-168	Mature luminal	44	748
N9	C5,7,9	4.5E-294	Luminal progenitor/Basal		
N10	C5,7,9	6.12E-40	Basal		
N11	C5,7,9	2.60E-183	Basal		
N12	C1,3,4	1.46E-89	Mature luminal	185	429
N13	C5,7,9	1.3E-160	Basal	8	
N14	C6	1.68E-63	Luminal progenitor		
N15	C5,7,9	4.16E-41	Luminal progenitor/basal	43	88
N16	C5,7,9	1.7E-299	Luminal progenitor/basal		
N17	C1,3,4	1.20E-107	Mature luminal	615	
N18	C5,7,9	<5.00e-324	Basal		
N19	C11	2.23E-75	Luminal progenitor	90	505
N20	C5,7,9	1.23E-81	Basal		
N21	C5,7,9	3.00E-256	Basal		
N22	C5,7,9	<5.00e-324	Basal		

Table S5: Overall survival- PROC PHREG with PDK4 and TBX3 H-scores as dichotomous variable in multivariable models. Related to Figure 7.

			PDK4				TBX3			
Obs	**Group**	**H-Score Category Parameter***	**p-value**	**Point Estimate**	**Lower 95% Wald Confidence Limit**	**Upper 95% Wald Confidence Limit**	**p-value**	**Point Estimate**	**Lower 95% Wald Confidence Limit**	**Upper 95% Wald Confidence Limit**
1	OS	High vs Low	**0.0431**	1.382	1.010	1.890	**0.0333**	0.721	0.533	0.974
2	OS for ER+	High vs Low	**0.0131**	1.600	1.104	2.319	0.1884	0.789	0.554	1.123
3	OS for ER-	High vs Low	0.5535	1.239	0.610	2.514	0.0709	0.503	0.238	1.060
4	OS for Endocrine Therapy=Yes	High vs Low	0.0847	1.431	0.952	2.152	0.3786	0.837	0.564	1.243
5	OS for Endocrine Therapy=No	High vs Low	0.0976	1.608	0.917	2.822	0.0505	0.589	0.346	1.001
6	OS for ER+ and ET=Yes	High vs Low	**0.0300**	1.614	1.048	2.488	0.9396	0.983	0.637	1.519
7	OS for ER+ and ET=No	High vs Low	0.1473	1.956	0.789	4.847	**0.0323**	0.422	0.191	0.930
8	OS for ER- and ET=Yes	High vs Low	0.9707	0.961	0.117	7.887	0.3166	0.324	0.036	2.939
9	OS for ER- and ET=No	High vs Low	0.2016	1.710	0.751	3.893	0.1874	0.559	0.235	1.327
10	OS for ER+/PR+/HER2-=Yes	High vs Low	0.1131	1.510	0.907	2.513	0.4029	0.792	0.458	1.368
11	OS for ER+/PR+/HER2-=No	High vs Low	0.7579	1.146	0.483	2.717	**0.0036**	0.250	0.099	0.635

*referent group listed second
ET= Endocrine therapy
OS=Overall survival
Obs= Observations