Expanded View Figures

Figure EV1. scRNA-seq coverage and transcriptional profiling of normal breast epithelium.

A Scatterplot showing the number of cells and the number of genes expressed after quality checking and filtering for each of the 79 samples and 10x runs listed in Table EV4.

B Flow cytometric analysis of pre- and post-menopausal tissue using CD49f and EpCAM to fractionate lineage-negative breast cells into epithelial and stromal cells. Cell clusters corresponding to stromal, basal, LP, and ML cell populations were visible for each tissue specimen, but the percentage of cells in each population varied between samples (pre- vs post-menopausal).

C t-SNE plot of the integrated scRNA-seq profiles of epithelial cells from 11 reduction mammoplasties colored by tissue specimen (left) or by cell cluster (right, with Seurat cluster resolution set to 0.015). Cluster 4 (red) corresponds to stromal cells.

D Box plots of lineage-specific expression signatures by epithelial cell cluster. Each epithelial cell was interrogated with the expression signatures for the human basal, LP, ML, and stromal cell types. Each panel shows the same cells and cell colors as in Fig 1E. Labels 1–4 correspond to clusters in Fig 1E. The four panels show basal, LP, ML, and stromal expression signatures, respectively. Vertical axis shows average expression of cell population signature genes as log2 counts per million. Box plots show quartiles, minimum, and maximum.

E Heat map of pseudo-bulk samples showing marker genes for each epithelial cluster. The top 20 marker genes were identified for each cluster by differential expression analysis of the pseudo-bulk read counts. Color bars at the top of the plot indicate the cluster and patient sample.

F Same t-SNE plot as Fig 2A and B showing combined profiles of total tissue cells from reduction mammoplasties but colored by epithelial lineage expression scores. Expression levels confirm the green, blue, and red clusters as basal, LP, and ML, respectively.

G As for (F) but colored by expression of EMT marker genes (*ZEB1/2, SNAI1*).

H Heat map of non-epithelial pseudo-bulk samples showing expression of immune and stromal lineage genes from Novershtern *et al* (2011) and Jeffrey *et al* (2006). Sample colors correspond to cell clusters in Fig 2D.

© 2021 The Authors

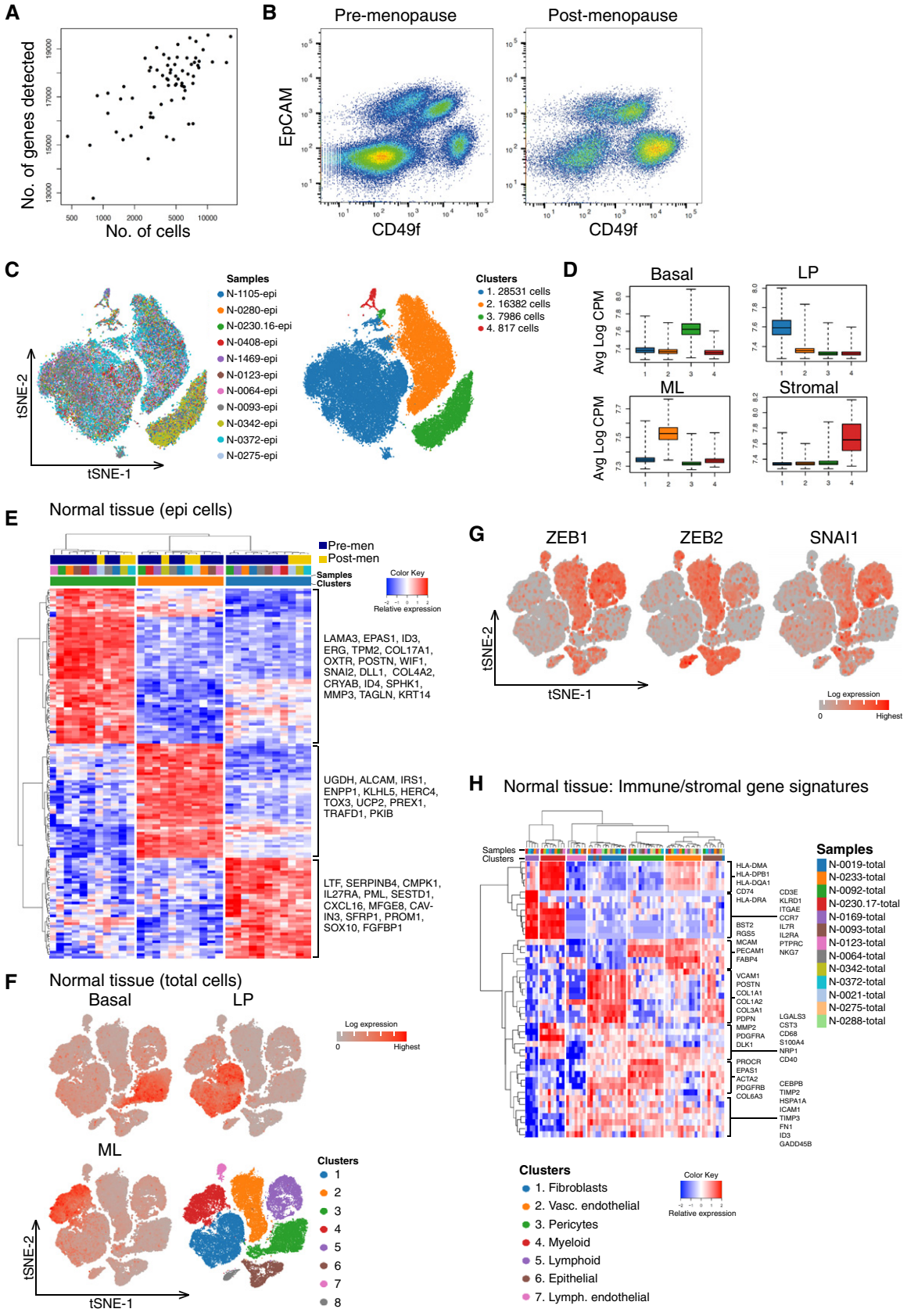
A
No. of genes detected
No. of cells
B
Pre-menopause
Post-menopause
EpCAM
CD49f
CD49f
C
Samples
N-1105-epi
N-0280-epi
N-0230.16-epi
N-0408-epi
N-1469-epi
N-0123-epi
N-0064-epi
N-0093-epi
N-0342-epi
N-0372-epi
N-0275-epi
Clusters
1. 28531 cells
2. 16382 cells
3. 7986 cells
4. 817 cells
tSNE-2
tSNE-1
D
Basal
LP
ML
Stromal
Avg Log CPM
E
Normal tissue (epi cells)
Pre-men
Post-men
Samples
Clusters
Color Key
Relative expression
LAMA3, EPAS1, ID3, ERG, TPM2, COL17A1, OXTR, POSTN, WIF1, SNAI2, DLL1, COL4A2, CRYAB, ID4, SPHK1, MMP3, TAGLN, KRT14
UGDH, ALCAM, IRS1, ENPP1, KLHL5, HERC4, TOX3, UCP2, PREX1, TRAFD1, PKIB
LTF, SERPINB4, CMPK1, IL27RA, PML, SESTD1, CXCL16, MFGE8, CAVIN3, SFRP1, PROM1, SOX10, FGFBP1
F
Normal tissue (total cells)
Basal
LP
ML
Log expression
0
Highest
Clusters
1
2
3
4
5
6
7
8
tSNE-2
tSNE-1
G
ZEB1
ZEB2
SNAI1
tSNE-2
tSNE-1
Log expression
0
Highest
H
Normal tissue: Immune/stromal gene signatures
Samples
Clusters
HLA-DMA
HLA-DPB1
HLA-DQA1
CD74
HLA-DRA
CD3E
KLRD1
ITGAE
CCR7
IL7R
IL2RA
PTPRC
NKG7
BST2
RGS5
MCAM
PECAM1
FABP4
VCAM1
POSTN
COL1A1
COL1A2
COL3A1
PDPN
MMP2
PDGFRA
DLK1
LGALS3
CST3
CD68
S100A4
NRP1
CD40
PROCR
EPAS1
ACTA2
PDGFRB
COL6A3
CEBPB
TIMP2
HSPA1A
ICAM1
TIMP3
FN1
ID3
GADD45B
Samples
N-0019-total
N-0233-total
N-0092-total
N-0230.17-total
N-0169-total
N-0093-total
N-0123-total
N-0064-total
N-0342-total
N-0372-total
N-0021-total
N-0275-total
N-0288-total
Clusters
1. Fibroblasts
2. Vasc. endothelial
3. Pericytes
4. Myeloid
5. Lymphoid
6. Epithelial
7. Lymph. endothelial
Color Key
Relative expression

Figure EV1.

Figure EV2. Immunohistochemistry and inter-patient variability among breast tumors.

A Immunostaining of representative ER^+, TNBC, and $HER2^+$ primary tumors for expression of ER, PR, HER2, and pan-cytokeratin. H&E sections are shown in the left panels. Scale bar, 200 μm.

B–D Same t-SNE plots as in Fig 6A–C showing the integrated profiles of total cells from 8 TNBC (4 non-*BRCA1*, 4 *BRCA1*-mutated), 6 $HER2^+$, and 13 ER^+ tumors, respectively. Cells are colored by tissue sample (top panels) or cell cluster (bottom panels). Cluster resolutions: 0.1, 0.07, and 0.1 for TNBC, $HER2^+$, and ER^+, respectively.

© 2021 The Authors

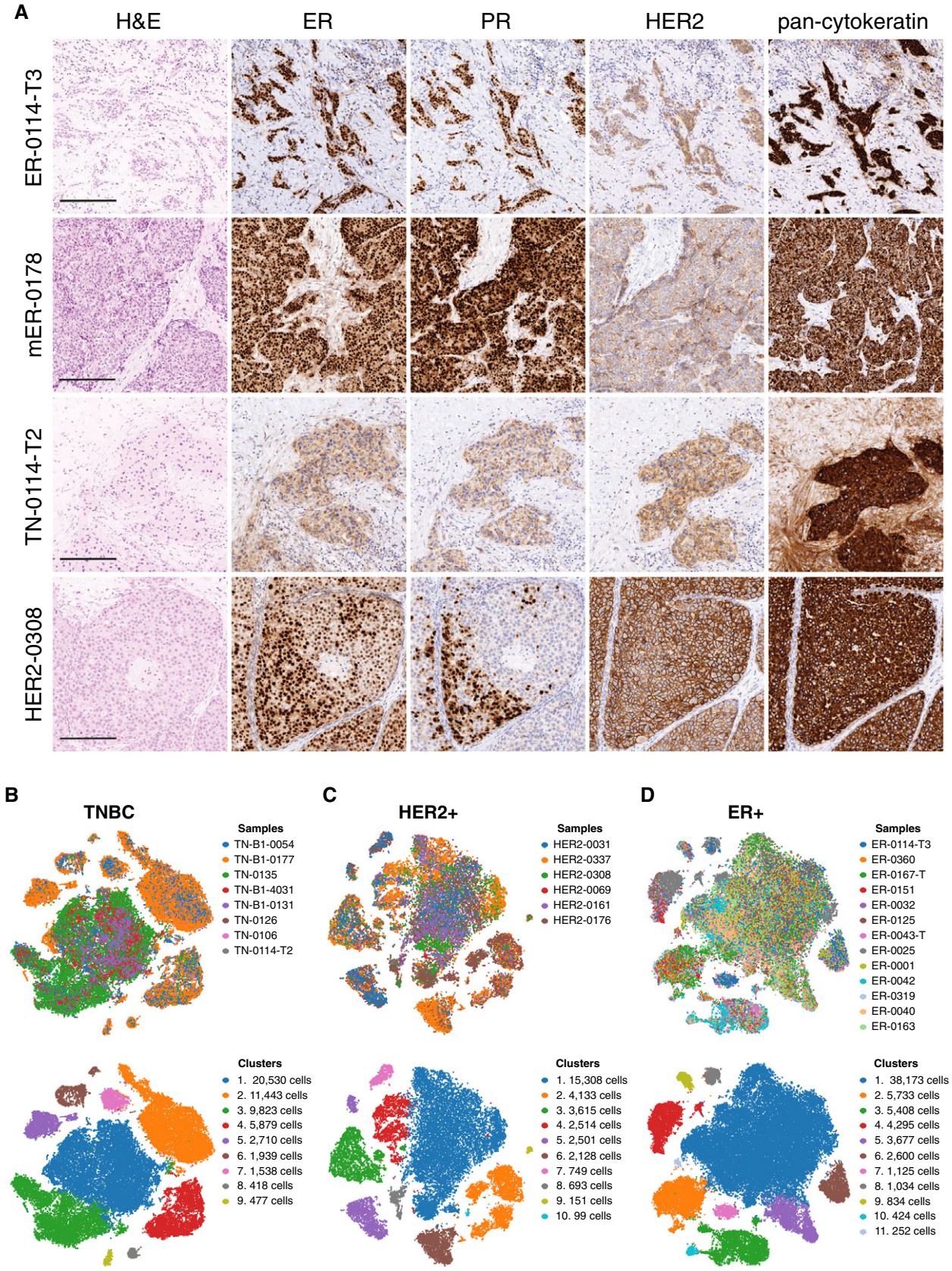

Figure EV2.

© 2021 The Authors

Figure EV3. Characterization of the expression profiles of breast cancer cells in different subtypes.

A Box plots showing transcriptional correlations between epithelial subsets (basal, LP, and ML) and cell clusters present in TNBC, $HER2^+$, and ER^+ tumors. LP = Luminal progenitor, ML= Mature luminal cells. Cell numbers are given in Fig EV2B–D. Box plots show quartiles, minimum, and maximum.

B t-SNE plots of epithelial cells from TNBC and $HER2^+$ tumors, respectively (*EPCAM*$^+$ cells in Fig 6B and C). Left panels show reclustering (resolutions 0.1 and 0.05). Right panel shows *MKI67* expression.

C Same t-SNE plots as in (B) and Fig 6E, depicting distribution of cells expressing epithelial markers (*KRT5* and *KRT8*), hormone receptors (*ESR1, PGR*), *ERBB2* receptor, EMT signature genes (*ZEB1*, *ZEB2*, *SNAI1*, *SNAI2*, *VIM*, and *TWIST1*) and *BCL2*.

D Enrichment of KEGG pathways in clusters identified from (B) (Fisher's exact test). For TBNC, cluster 2 vs 1; $HER2^+$, cluster 3 vs the rest.

E Box plots of breast cancer subtype signatures from the TCGA by epithelial cell cluster for the combined ER^+ tumor dataset. Each panel shows the same cells and cell colors as for the combined dataset in Fig 6E. Vertical axis shows average expression of cell population signature genes as log2 counts per million. Box plots show quartiles, minimum, and maximum.

© 2021 The Authors

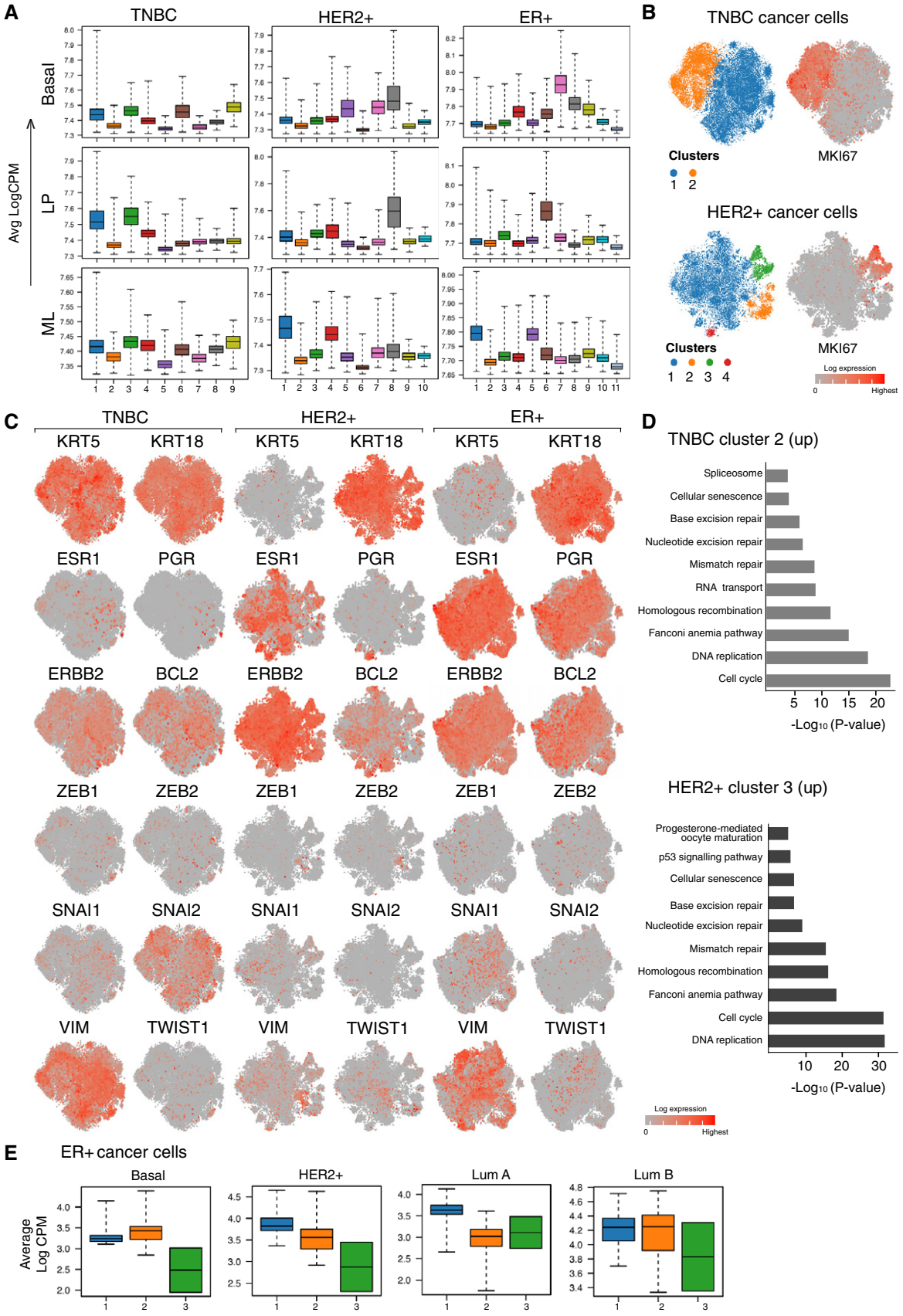
A
TNBC
HER2+
ER+
Basal
LP
ML
Avg LogCPM
B
TNBC cancer cells
Clusters
MKI67
1 2
HER2+ cancer cells
Clusters
MKI67
1 2 3 4
Log expression
0
Highest
C
TNBC
HER2+
ER+
KRT5
KRT18
ESR1
PGR
ERBB2
BCL2
ZEB1
ZEB2
SNAI1
SNAI2
VIM
TWIST1
D
TNBC cluster 2 (up)
Spliceosome
Cellular senescence
Base excision repair
Nucleotide excision repair
Mismatch repair
RNA transport
Homologous recombination
Fanconi anemia pathway
DNA replication
Cell cycle
5 10 15 20
-Log10 (P-value)
HER2+ cluster 3 (up)
Progesterone-mediated oocyte maturation
p53 signalling pathway
Cellular senescence
Base excision repair
Nucleotide excision repair
Mismatch repair
Homologous recombination
Fanconi anemia pathway
Cell cycle
DNA replication
10 20 30
-Log10 (P-value)
Log expression
0
Highest
E
ER+ cancer cells
Basal
HER2+
Lum A
Lum B
Average Log CPM

Figure EV3.

Figure EV4. Unique T-cell subsets mark the tumor microenvironments of the three breast cancer subtypes.

A t-SNE plots showing T cells (clusters 1 and 5 for TNBC; clusters 2 and 7 for $HER2^+$; cluster 1 for ER^+ tumors) identified in Fig 7A–C. Reclustering revealed at least five T-cell subsets in TNBC and $HER2^+$ tumors, and four T-cell populations in ER^+ tumors (cluster resolutions: 0.2, 0.7, and 0.2, respectively).

B Same t-SNE plots as in (A) but colored by expression of T-cell markers and of MKI67 to identify proliferating T-cell clusters.

C Heat maps showing pseudo-bulk expression of top DE genes within the T-cell subsets identified in the t-SNE plots shown in (A). T_{EM} = Effector memory T cells; Treg = Regulatory T cells; T_{RM} = Tissue-resident memory T cells; NK = Natural Killer cells.

© 2021 The Authors

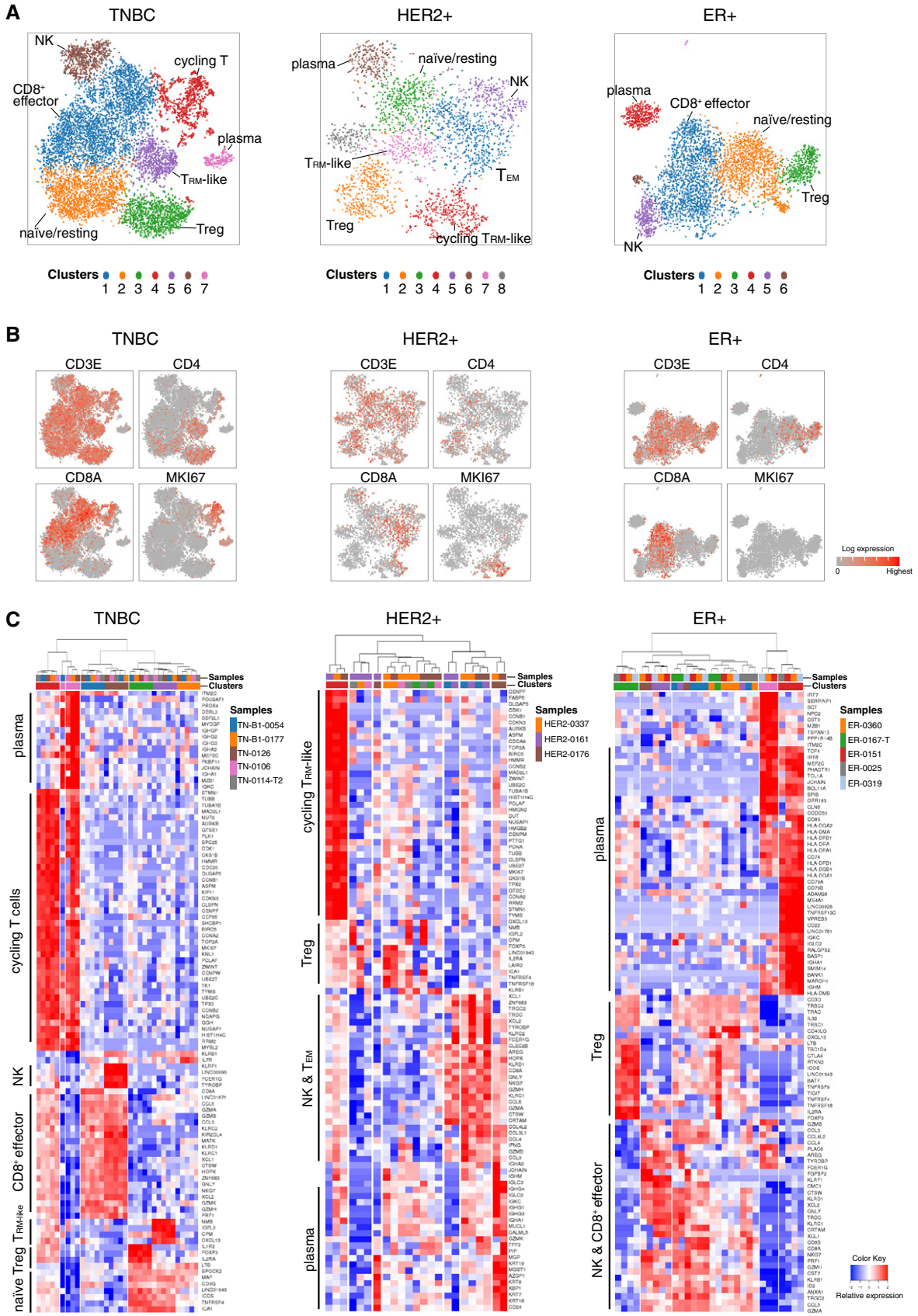
A
TNBC
HER2+
ER+
NK
cycling T
CD8+ effector
plasma
TRM-like
naïve/resting
Treg
naïve/resting
NK
TEM
cycling TRM-like
plasma
CD8+ effector
naïve/resting
Treg
NK
Clusters
1 2 3 4 5 6 7
Clusters
1 2 3 4 5 6 7 8
Clusters
1 2 3 4 5 6
B
TNBC
HER2+
ER+
CD3E
CD4
CD8A
MKI67
Log expression
0
Highest
C
TNBC
HER2+
ER+
Samples
Clusters
Samples
TN-B1-0054
TN-B1-0177
TN-0126
TN-0106
TN-0114-T2
HER2-0337
HER2-0161
HER2-0176
ER-0360
ER-0167-T
ER-0151
ER-0025
ER-0319
plasma
cycling T cells
NK
CD8+ effector
TRM-like
naïve Treg
cycling TRM-like
Treg
NK & TEM
plasma
plasma
Treg
NK & CD8+ effector
Color Key
Relative expression

Figure EV4.

Figure EV5. scRNA-seq profiles of male breast tumors and matching tumor-LN pairs.

A Combined t-SNE transcriptome profiles of two ER^+ male tumors; mER-0068-T (blue) and mER-0178T (yellow). Bottom panel shows the corresponding t-SNE cell clusters, as shown in the top panel, but colored by cell clusters.

B t-SNE plots showing gene expression of epithelial markers and steroid hormone receptors.

C Box plots of lineage-specific expression signatures by epithelial cell cluster. Each epithelial cell was interrogated with the expression signatures for the human basal, LP, ML, and stromal cell types. Each panel shows the same cells and cell colors as in (A). Vertical axis shows average expression of cell population signature genes as log2 counts per million. Box plots show quartiles, minimum, and maximum.

D Tumor-LN pairs: interrogation of cell clusters identified in Fig 9A for expression of canonical immune (*CD68*, *CD19*, *CD4*) and stromal (*COL1A2*, *S100A4*, *PDGFRA*) genes.

E Maps of inferred copy number variation (CNV) for the combined tumor-LN scRNA-seq expression data for the clusters indicated in Fig 9A. Tumor cells can be distinguished from normal (N) by abundance of CNVs.

© 2021 The Authors

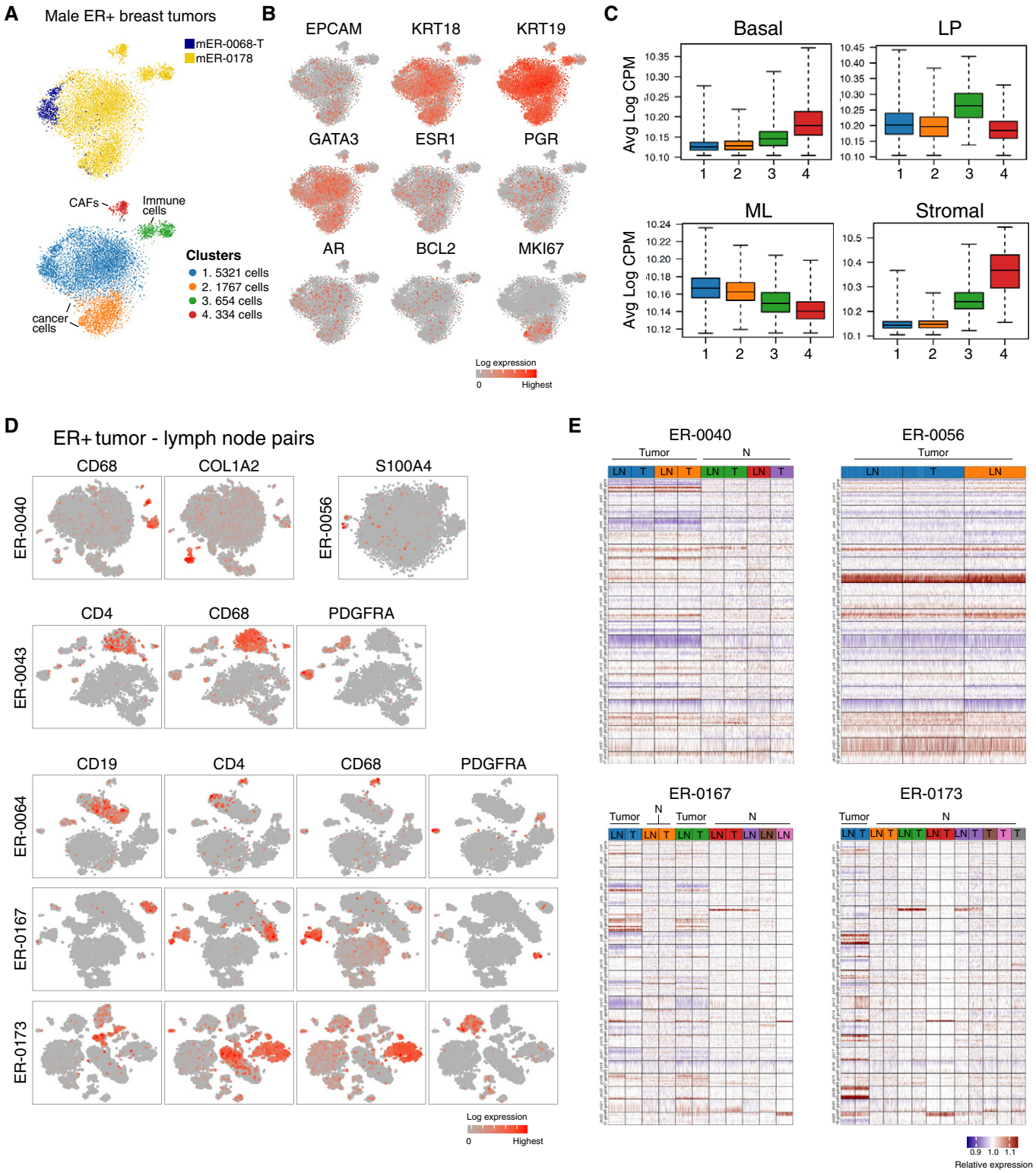
A
Male ER+ breast tumors
mER-0068-T
mER-0178
CAFs
Immune cells
cancer cells
Clusters
1. 5321 cells
2. 1767 cells
3. 654 cells
4. 334 cells
B
EPCAM
KRT18
KRT19
GATA3
ESR1
PGR
AR
BCL2
MKI67
Log expression
0
Highest
C
Basal
LP
ML
Stromal
Avg Log CPM
D
ER+ tumor - lymph node pairs
CD68
COL1A2
S100A4
CD4
PDGFRA
CD19
ER-0040
ER-0056
ER-0043
ER-0064
ER-0167
ER-0173
E
ER-0040
ER-0056
ER-0167
ER-0173
Tumor
N
LN
T
Relative expression

Figure EV5.