Cell Reports Methods, Volume *1*

Supplemental information

Toward reproducible, scalable, and robust data analysis across multiplex tissue imaging platforms

Erik A. Burlingame, Jennifer Eng, Guillaume Thibault, Koei Chin, Joe W. Gray, and Young Hwan Chang

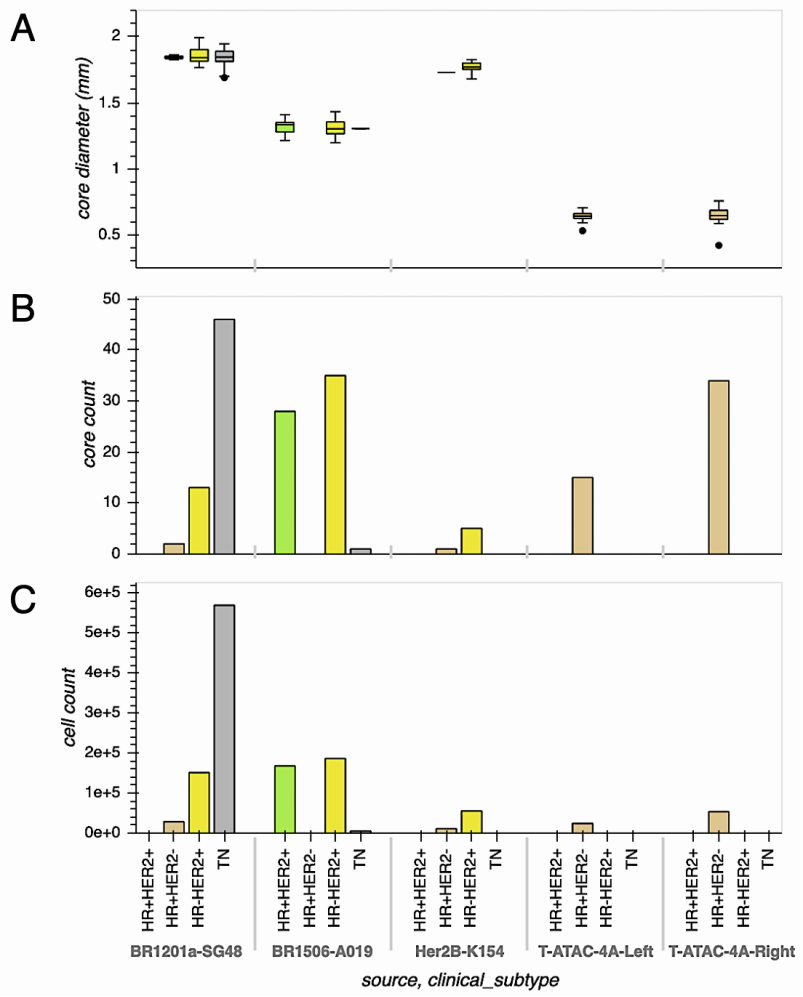
A
core diameter (mm)
2
1.5
1
0.5
B
core count
50
40
30
20
10
0
C
cell count
6e+5
5e+5
4e+5
3e+5
2e+5
1e+5
0e+0
HR+HER2+
HR+HER2-
HR-HER2+
TN
BR1201a-SG48
BR1506-A019
Her2B-K154
T-ATAC-4A-Left
T-ATAC-4A-Right
source, clinical_subtype

Figure S1. Overview of TMA composition and CyCIF panel. Related to Figure 1.
(A) Core diameters split by TMA source and BC subtype. Lines in boxes indicate the medians and whiskers indicate data within 1.5 × interquartile range of the upper and lower quartiles. Outliers are shown as distinct points. (B) Core count split by TMA source and BC subtype. (C) Cell count split by TMA source and BC subtype.

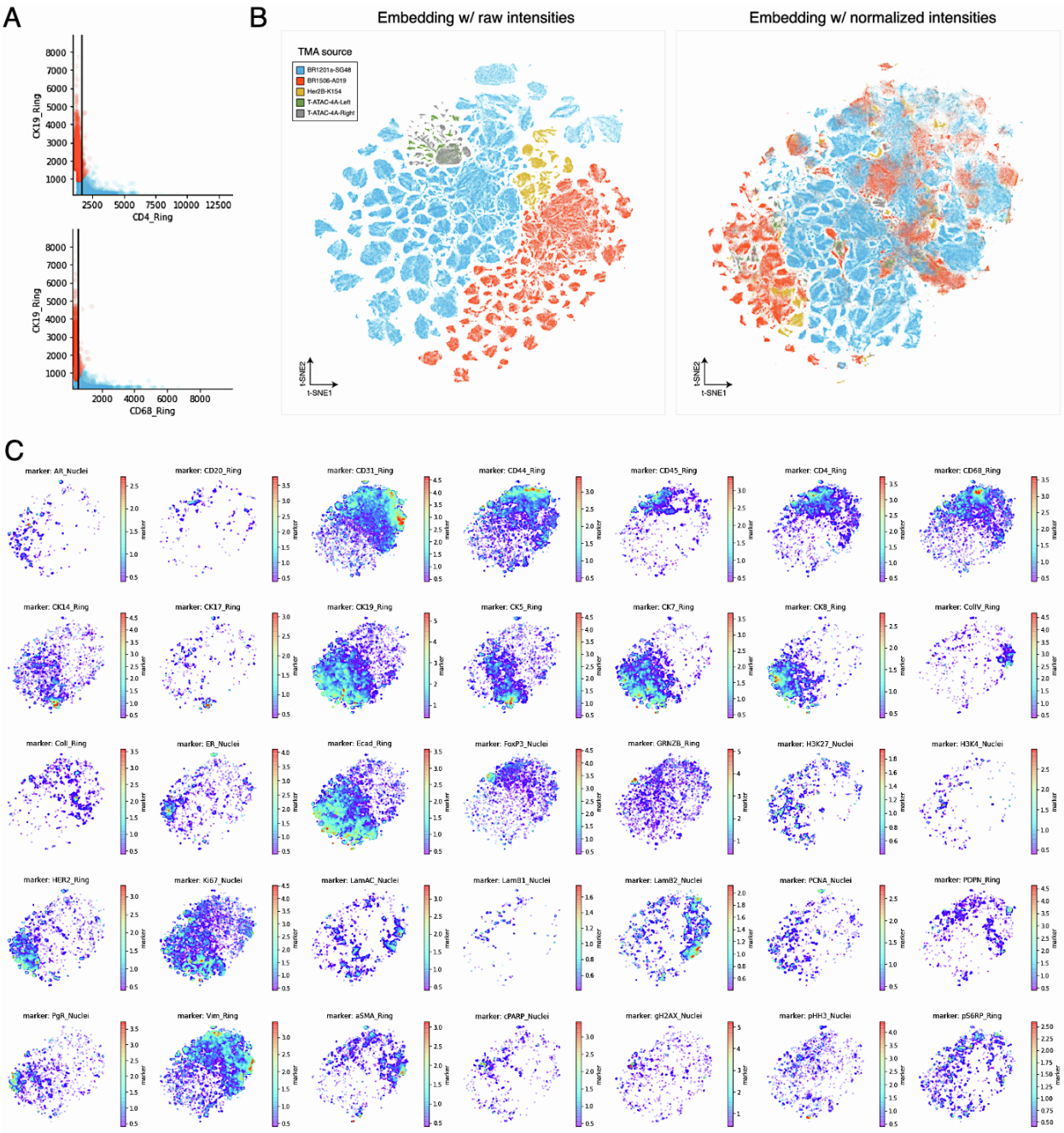
A
CK19_Ring
CD4_Ring
CD68_Ring
B
Embedding w/ raw intensities
Embedding w/ normalized intensities
TMA source
BR1201a-SG48
BR1506-A019
Her2B-K154
T-ATAC-4A-Left
T-ATAC-4A-Right
t-SNE1
t-SNE2
C
marker: AR_Nuclei
marker: CD20_Ring
marker: CD31_Ring
marker: CD44_Ring
marker: CD45_Ring
marker: CD4_Ring
marker: CD68_Ring
marker: CK14_Ring
marker: CK17_Ring
marker: CK19_Ring
marker: CK5_Ring
marker: CK7_Ring
marker: CK8_Ring
marker: ColIV_Ring
marker: Coll_Ring
marker: ER_Nuclei
marker: Ecad_Ring
marker: FoxP3_Nuclei
marker: GRNZB_Ring
marker: H3K27_Nuclei
marker: H3K4_Nuclei
marker: HER2_Ring
marker: Ki67_Nuclei
marker: LamAC_Nuclei
marker: LamB1_Nuclei
marker: LamB2_Nuclei
marker: PCNA_Nuclei
marker: PDPN_Ring
marker: PgR_Nuclei
marker: Vim_Ring
marker: aSMA_Ring
marker: cPARP_Nuclei
marker: gH2AX_Nuclei
marker: pHH3_Nuclei
marker: pS6RP_Ring

Figure S2. Cell mean intensity normalization enables batch compilation of TMA data. Related to Figure 1 and STAR Methods.

(A) Example of RESTORE normalization (Chang et al., 2020) of CD4 and CD68 cell mean intensities for a single TMA core. Cell mean intensities of these immune markers are plotted against the cell mean intensity of epithelial CK19, a mutually exclusive marker. Cells are partitioned into positive (blue) and negative (red) populations. Black lines represent the computed normalization factors. Cells to the right of each line are above the background intensity level for that immune marker. (B) *t*-SNE embeddings of all cells using either raw (left) or normalized (right) cell mean intensities for all markers. Cells are colored according to the TMA from which they originate. A strong batch effect is observed before normalization, leading to partitioning according to TMA of origin. Following normalization, cell phenotypes shared between TMAs are co-localized. However, some TMA-specific partitioning remains due to subtype-specific marker bias within TMAs. The coordinates used in the *t*-SNE plots at right are the same as those used in Figure 1B, where cells are colored by metacluster. In the plot in Figure 1B, it is clear that the HR+ and HER2+ tumor cells aggregated in the lower left correspond to the HR+ and HER2+ TMA cores from the BR1506, Her2B, and T-ATAC cohorts from Figure S2B. (C) *t*-SNE normalized embedding, faceted by marker, showing only cells that have a mean intensity above the normalization factor for that marker. Color scale is log-transformed normalized intensity.

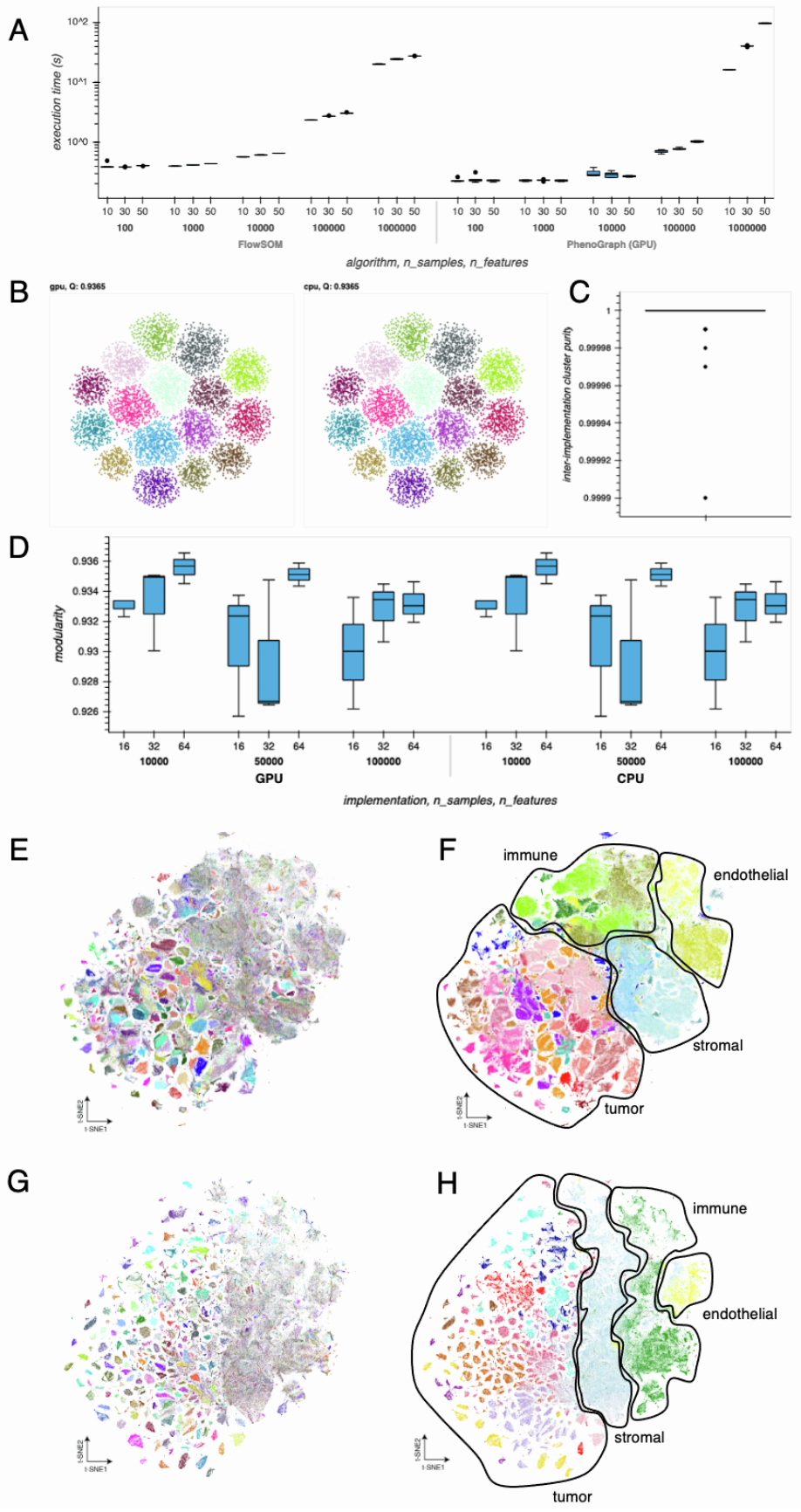
A
execution time (s)
10^2
10^1
10^0
10 30 50
100
1000
10000
100000
1000000
FlowSOM
PhenoGraph (GPU)
algorithm, n_samples, n_features
B
gpu, Q: 0.9365
cpu, Q: 0.9365
C
inter-implementation cluster purity
1
0.99998
0.99996
0.99994
0.99992
0.9999
D
modularity
0.936
0.934
0.932
0.93
0.928
0.926
16 32 64
10000
50000
100000
GPU
CPU
implementation, n_samples, n_features
E
F
immune
endothelial
stromal
tumor
t-SNE2
t-SNE1
G
H
immune
endothelial
stromal
tumor

Figure S3. Benchmarking single-cell phenotyping algorithms and implementations. Related to Figure 1, Figure 3, and STAR Methods.

(A) Comparison of execution time between FlowSOM (Van Gassen et al., 2015) and our GPU implementation of PhenoGraph. Each boxplot represents the results from 5 replicate experiments using the indicated parameters. Lines in boxes indicate the medians and whiskers indicate data within 1.5 × interquartile range of the upper and lower quartiles. Outliers are shown as distinct points. (B) Benchmarking CPU and GPU implementations of PhenoGraph reveals no appreciable difference in quantitative endpoints between implementations. Scatter plots showing *t*-SNE embeddings of synthetic data, colored based on cluster label defined by either implementation. The modularity Q for each clustering result is shown above the plot. (C) Boxplot showing the distribution of GPU vs. CPU cluster purity for synthetic data. The line in the box indicates the median and whiskers indicate data within 1.5 × interquartile range of the upper and lower quartiles. The boxplot is compressed since most of the measurements were exactly or very near 1. Outliers are shown as distinct points. (D) Boxplots showing the distribution of clustering modularities for either implementation for the same synthetic data generated using the indicated number of samples and number of features. Lines in boxes indicate the medians and whiskers indicate data within 1.5 × interquartile range of the upper and lower quartiles. Outliers are shown as distinct points. Each boxplot represents the results from 5 replicate experiments using the indicated parameters. No significant difference was observed in modularity between implementations (ANOVA, P=0.999). (E) For both OHSU and Basel cohorts, tumor cells differ more between samples than immune, stromal, or endothelial cells. The same *t*-SNE embedding of the OHSU dataset from Figure 1B, but with cells colored based on the unique tissue core from which they are derived. (F) The same *t*-SNE embedding of the OHSU dataset from Figure 1B, but with annotations indicating immune, stromal, endothelial, and tumor phenotypic regions. (G) A *t*-SNE embedding of the Basel dataset (Jackson et al., 2020) derived using the same parameters as were used for the OHSU *t*-SNE embedding, with cells colored based on the unique tissue core from which they are derived. (H) The same *t*-SNE embedding as in (G), but with annotations indicating immune, stromal, endothelial, and tumor phenotypic regions.

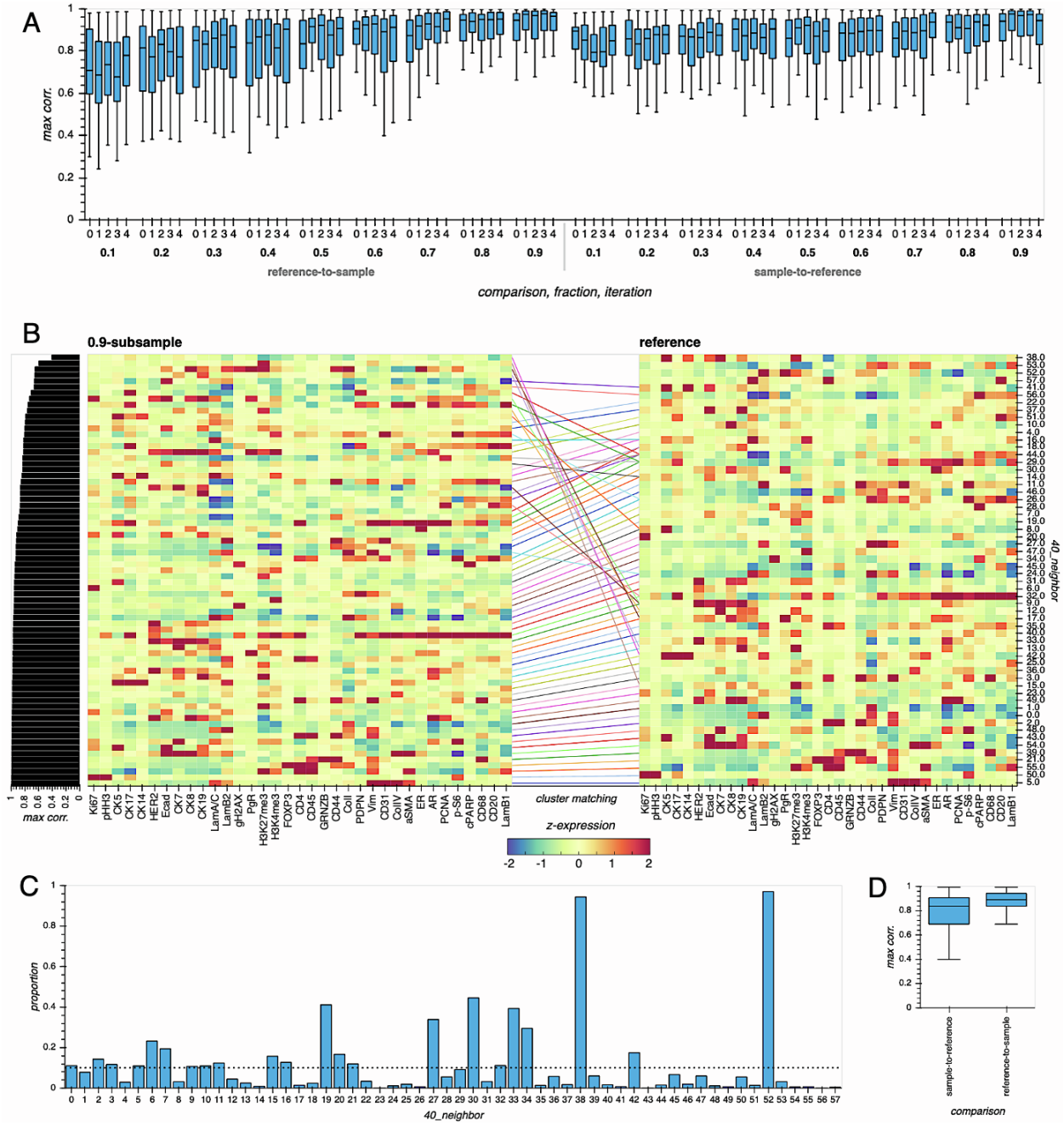
A
max corr.
reference-to-sample
sample-to-reference
comparison, fraction, iteration
B
0.9-subsample
reference
max corr.
cluster matching
z-expression
40 neighbor
C
proportion
40_neighbor
D
max corr.
sample-to-reference
reference-to-sample
comparison

Figure S4. BC cell phenotypes are robust against subsampling and noise. Related to Figure 1 and STAR Methods.

(A) Maximum Pearson's correlations between full reference cell phenotypes and those derived using PhenoGraph on cells from a random fractional sample of TMA cores, iterated five times at each fraction level. Lines in boxes indicate the medians and whiskers indicate data within 1.5 × interquartile range of the upper and lower quartiles. (B) Phenotype matching between the full reference phenotypes and those derived from a 90%-subsampled fraction of TMA cores. Phenotypes are ordered by increasing matching correlation. Matching phenotypes are linked by a line, and lines are colored to discriminate between adjacent or overlapping links. 40_neighbor represents the PhenoGraph cluster labels since we set k=40 when defining the k-nearest neighbor graph in the PhenoGraph routine. The colorbar indicates z-scored marker expression. (C) The proportion of cells from each reference phenotype that correspond to the 10% of TMA cores held out from the 90%-subsampled fraction from (B). Unmatched phenotypes in the full reference correspond in part to cells from held-out cores. The dotted line marks the 10% threshold. (D) Maximum Pearson's correlation between full reference phenotypes and those derived using PhenoGraph on normalized mean intensities from all TMA cores, but with ±20% random noise added to each marker in each core. Lines in boxes indicate the medians and whiskers indicate data within 1.5 × interquartile range of the upper and lower quartiles.

Reference	ME markers					
AR_Nuclei	CK5_Ring	FOXP3_Nuclei	ColIV_Ring			
aSMA_Ring	CK14_Ring	CD45_Ring	CK7_Ring	CK5_Ring	CK19_Ring	
CD20_Ring	CK14_Ring	CK7_Ring	CK5_Ring	CK19_Ring		
CD31_Ring	CK5_Ring	CK19_Ring	CK14_Ring	CK7_Ring	Ecad_Ring	
CD4_Ring	CK19_Ring	CK7_Ring	CK14_Ring	CK5_Ring	Ecad_Ring	
CD44_Ring	CK14_Ring	CK7_Ring	CK5_Ring	CK19_Ring	CD31_Ring	
CD45_Ring	CK19_Ring	CK7_Ring	CK14_Ring	CK5_Ring	CK8_Ring	CD31_Ring
CD68_Ring	CK19_Ring	CK7_Ring	CD31_Ring	CK14_Ring		
CK14_Ring	CD31_Ring	CD68_Ring	Vim_Ring	aSMA_Ring	CD20_Ring	CD45_Ring
CK17_Ring	CD31_Ring	CD68_Ring	Vim_Ring	ColI_Ring	CD45_Ring	
CK5_Ring	CD31_Ring	CD68_Ring	Vim_Ring	CD4_Ring	CD45_Ring	
CK19_Ring	CD68_Ring	CD4_Ring	CD31_Ring	CD45_Ring		
CK7_Ring	CD68_Ring	CD4_Ring	CD31_Ring	CD45_Ring	FOXP3_Nuclei	
CK8_Ring	CD68_Ring	CD4_Ring	CD31_Ring	CD45_Ring		
ColI_Ring	CD45_Ring	CK19_Ring	CK7_Ring	CK14_Ring	CK5_Ring	
ColIV_Ring	CK19_Ring	CK7_Ring	CK14_Ring	CK5_Ring	CD68_Ring	FOXP3_Nuclei
cPARP_Nuclei	Ki67_Nuclei	CK5_Ring	CD31_Ring	CD68_Ring	CK14_Ring	
Ecad_Ring	CD68_Ring	CD4_Ring	CD31_Ring			
ER_Nuclei	CD68_Ring	CD4_Ring	CD31_Ring	FOXP3_Nuclei		
FOXP3_Nuclei	CK19_Ring	CK7_Ring	CK5_Ring	CK14_Ring	CD31_Ring	CK8_Ring
gH2AX_Nuclei	CK8_Ring	CK14_Ring	CK5_Ring	CK7_Ring		
GRNZB_Ring	CK19_Ring	CK7_Ring	CK5_Ring	CD31_Ring	CK14_Ring	aSMA_Ring
H3K27me3_Nuclei	CD31_Ring	CD68_Ring	CD44_Ring			
H3K4me3_Nuclei	CD31_Ring	CD68_Ring	CD44_Ring	CK19_Ring		
HER2_Ring	CD68_Ring	CD44_Ring	CD31_Ring	Vim_Ring	CD4_Ring	
Ki67_Nuclei	cPARP_Nuclei					
LamA/C_Nuclei	CD68_Ring	CD44_Ring	CK19_Ring	CD45_Ring		
LamB1_Nuclei	CD68_Ring	CD44_Ring	CK19_Ring	CD45_Ring	CK14_Ring	CK7_Ring
LamB2_Nuclei	CD68_Ring	CD44_Ring	CK19_Ring	CD31_Ring	CK7_Ring	CK14_Ring
PCNA_Nuclei	CK7_Ring	CD45_Ring	CD31_Ring	CD68_Ring	CK14_Ring	LamB2_Nuclei
PDPN_Ring	CK19_Ring	CK7_Ring	CK14_Ring	CK5_Ring	CD31_Ring	CD68_Ring
PgR_Nuclei	CD68_Ring	CD4_Ring	CD31_Ring	CD20_Ring	aSMA_Ring	Vim_Ring
pHH3_Nuclei	CD31_Ring	CK5_Ring	CK19_Ring	CK14_Ring	GRNZB_Ring	
p-S6_Ring	CK19_Ring	CK5_Ring	CK7_Ring	CK14_Ring		
Vim_Ring	CK19_Ring	CK7_Ring	CD68_Ring	CD45_Ring	Ecad_Ring	

Table S1. Related to Figure 1 and Figure S2. Putative reference and mutually-exclusive (ME) marker pairs used for RESTORE normalization of cell mean intensities. Each marker name indicates from which compartment its mean intensity was extracted. "Ring" indicates that a marker's intensity was extracted from the ring-shaped cytoplasmic segmentation masks derived by subtracting the "Nuclei" segmentation masks from "Cell" segmentation masks.