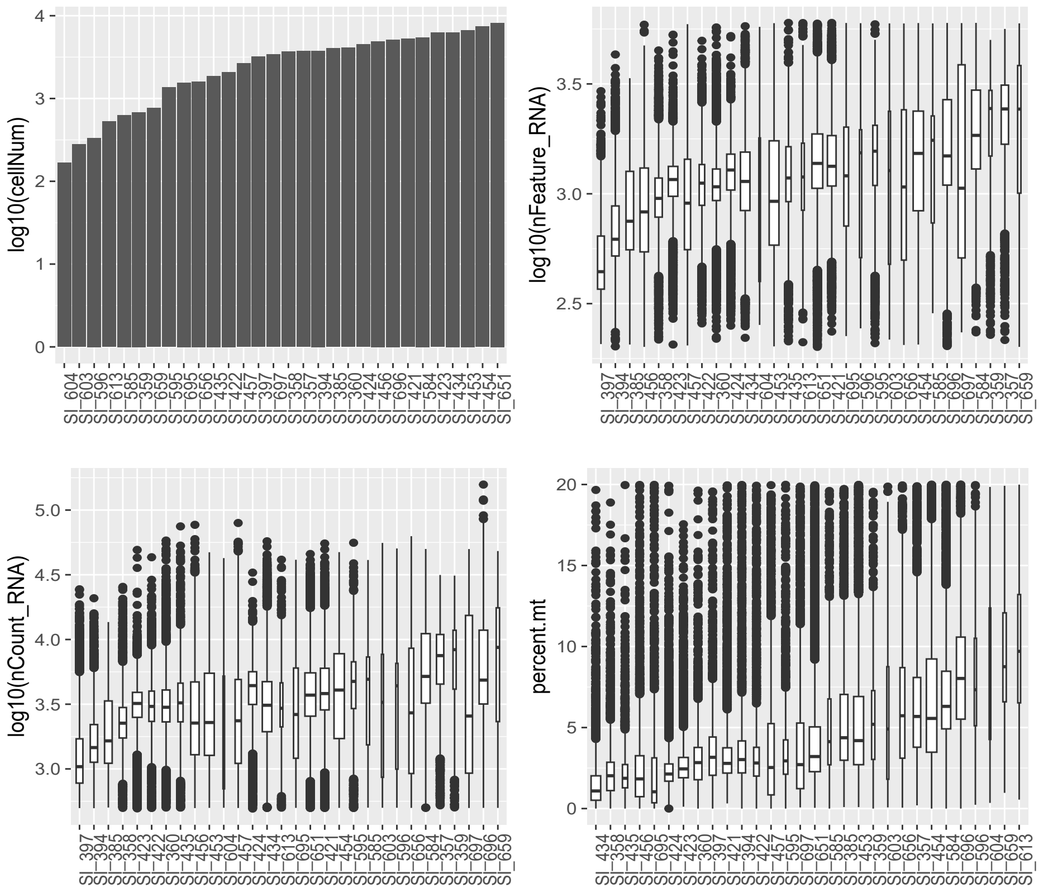
**iScience, Volume *27***

## Supplemental information

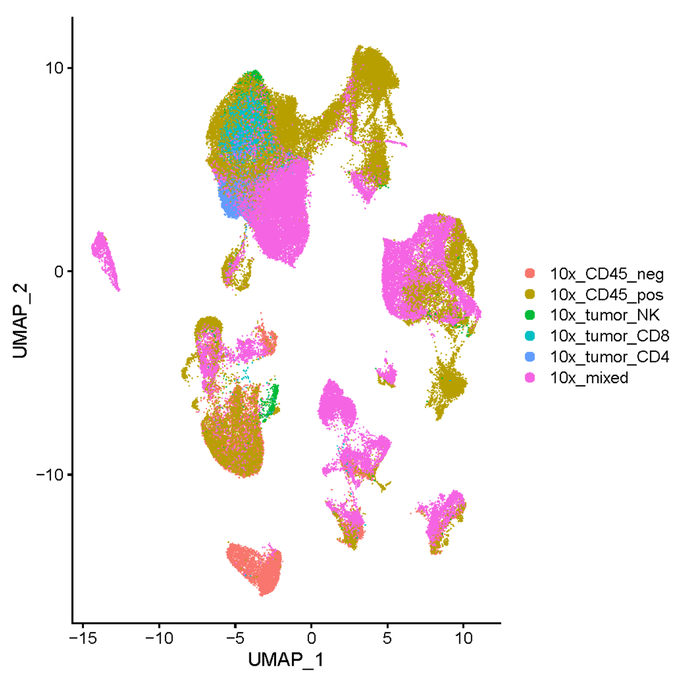
# Single-cell transcriptomic-informed deconvolution of bulk data identifies immune checkpoint blockade resistance in urothelial cancer

**Li Wang, Sudeh Izadmehr, John P. Sfakianos, Michelle Tran, Kristin G. Beaumont, Rachel Brody, Carlos Cordon-Cardo, Amir Horowitz, Robert Sebra, William K. Oh, Nina Bhardwaj, Matthew D. Galsky, and Jun Zhu**

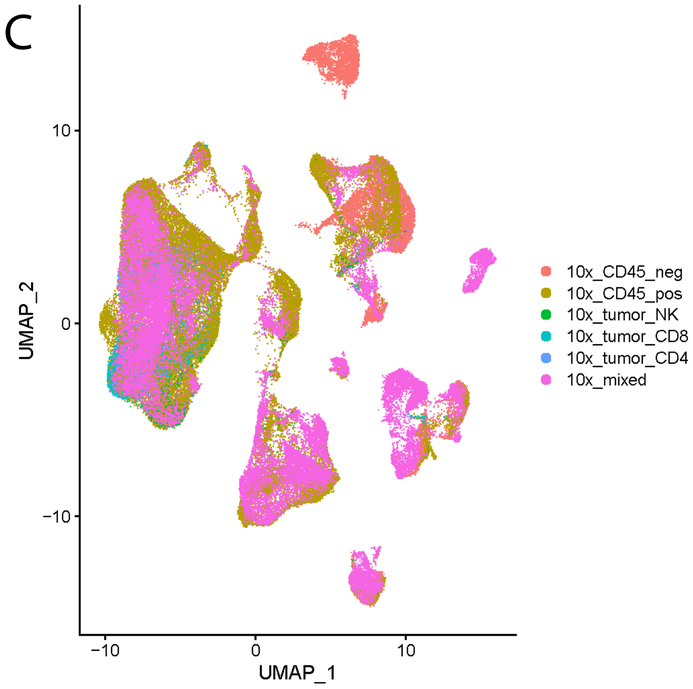
A




B




C




D

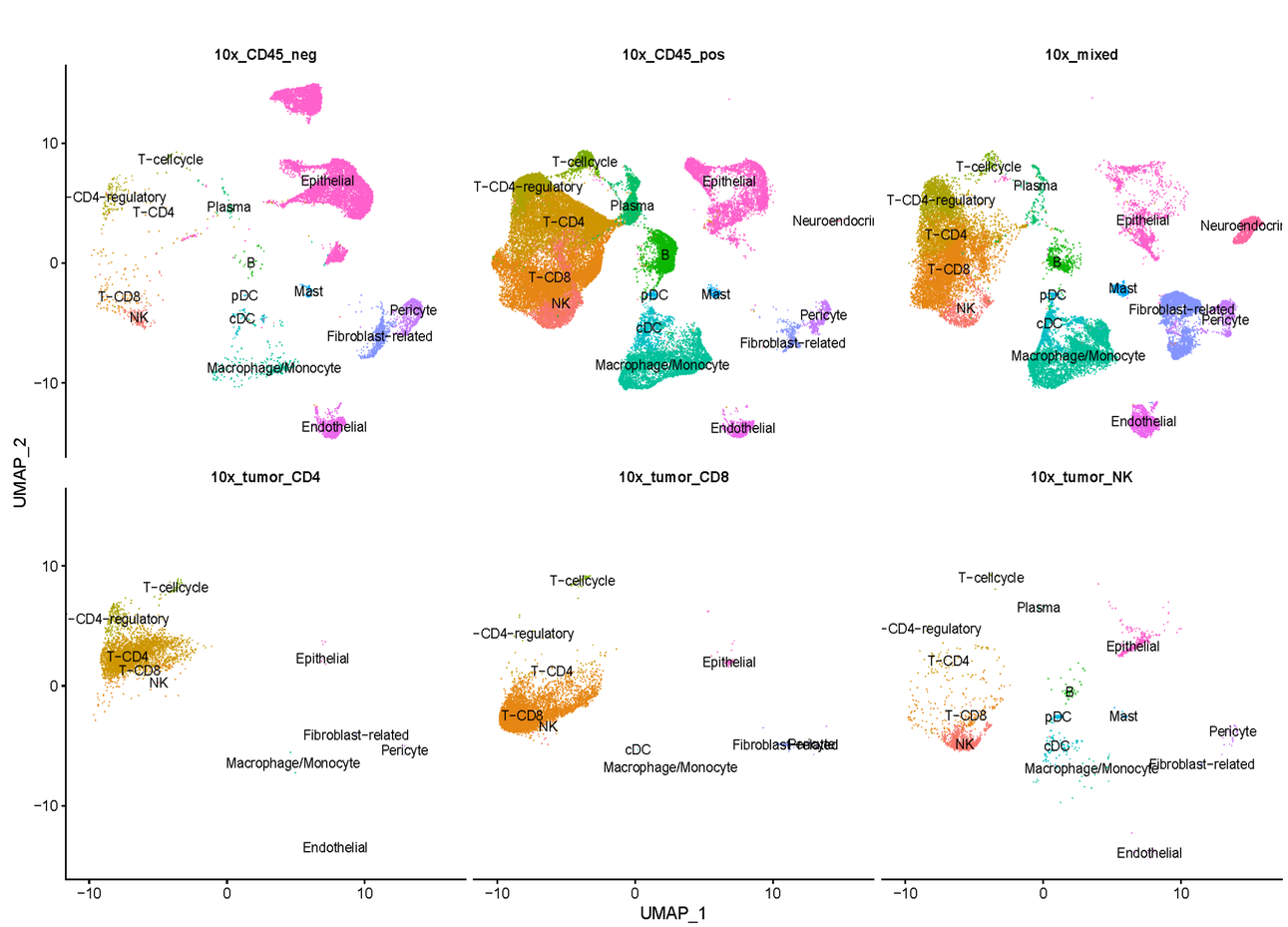
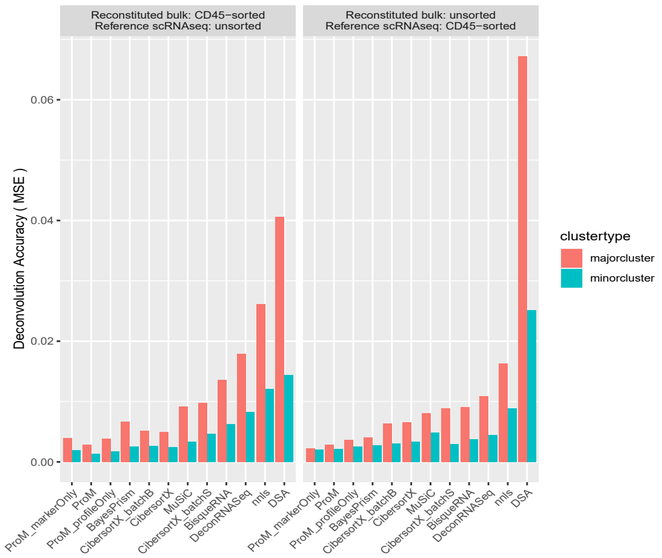
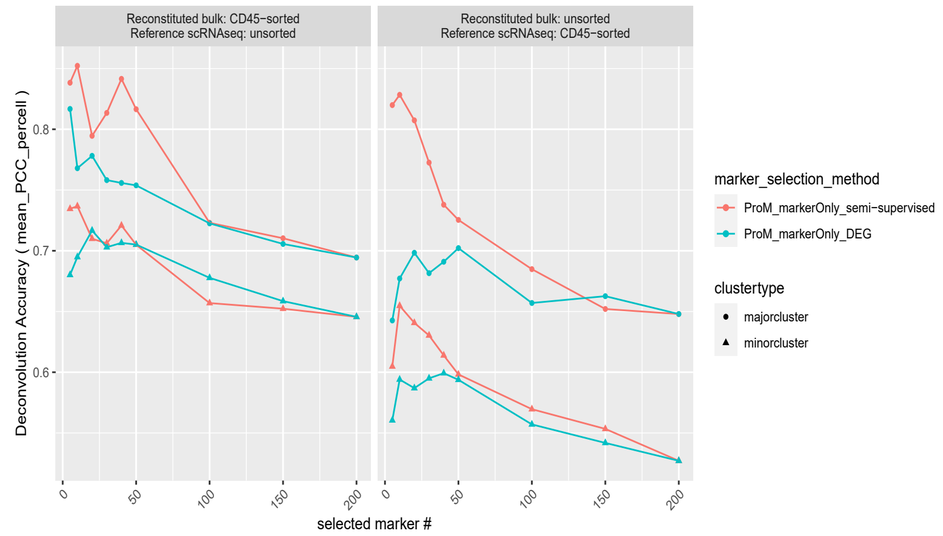



Figure S1

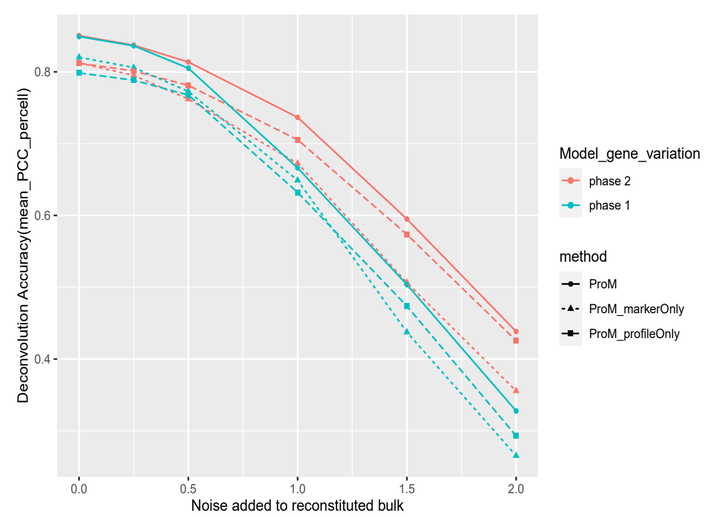
A



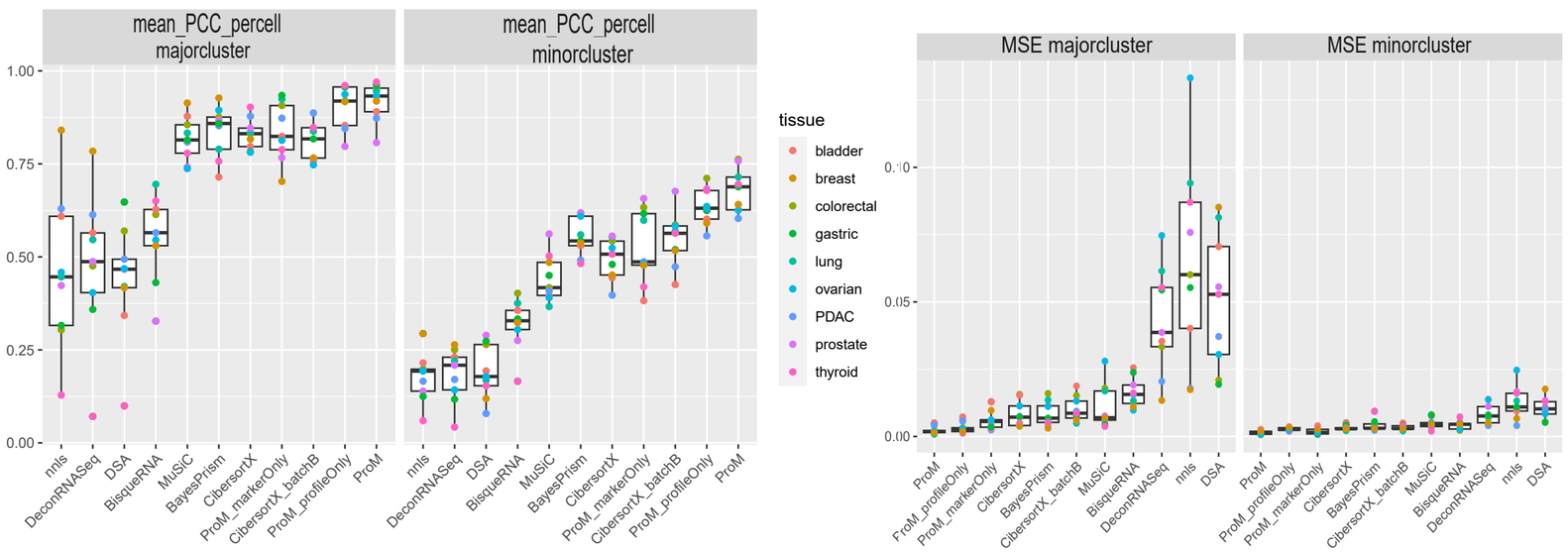

B



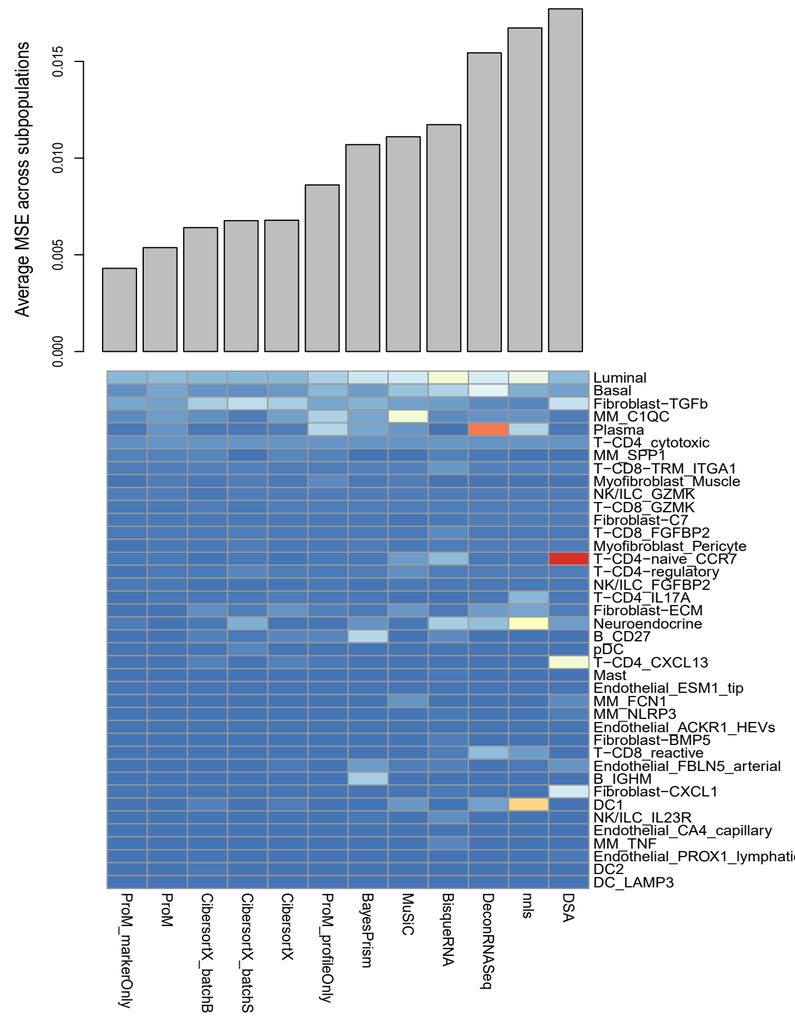

C




D




E




F

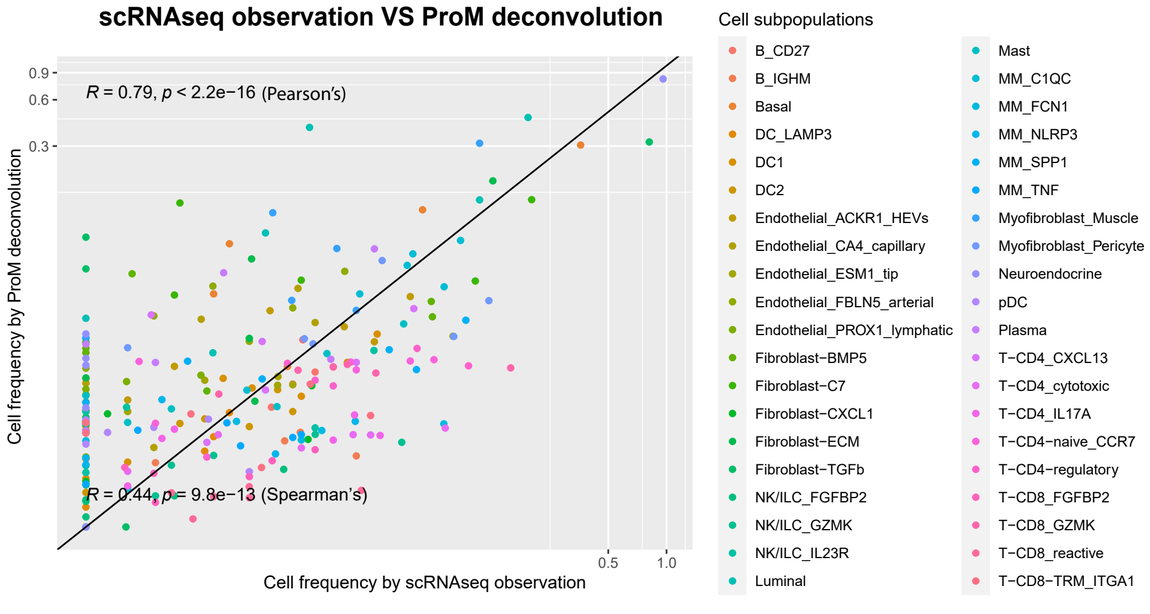
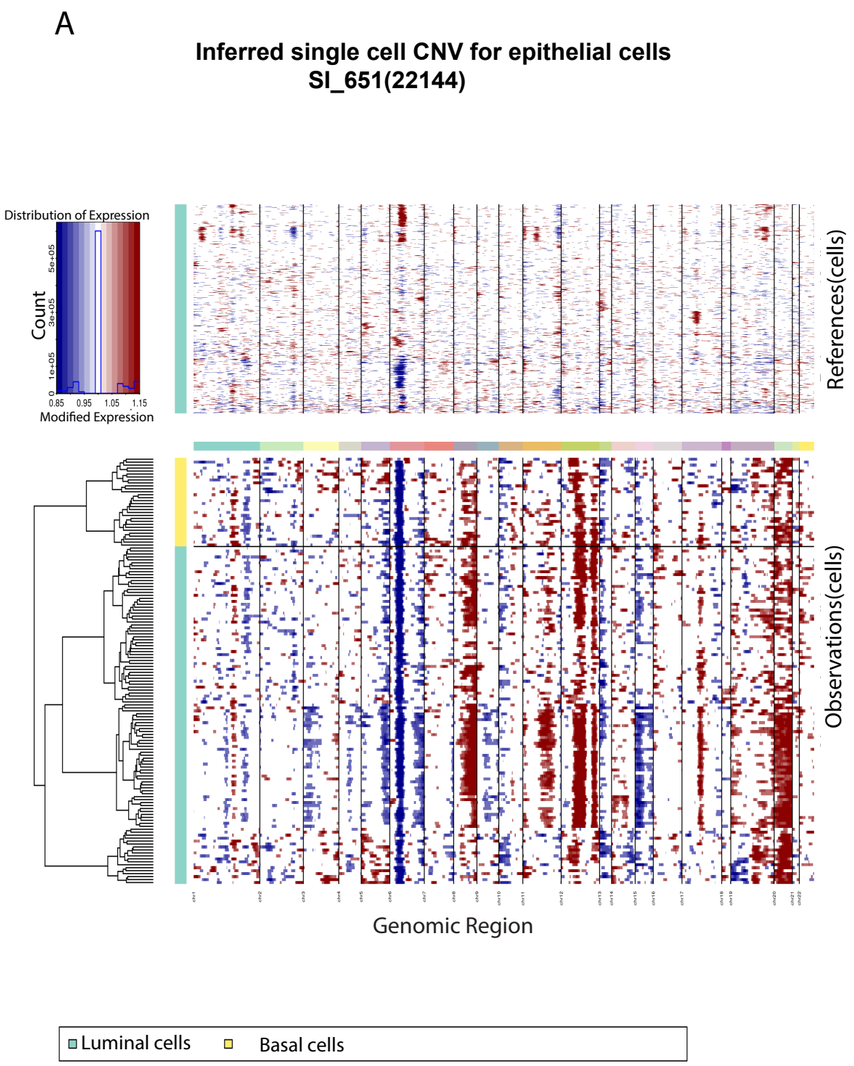
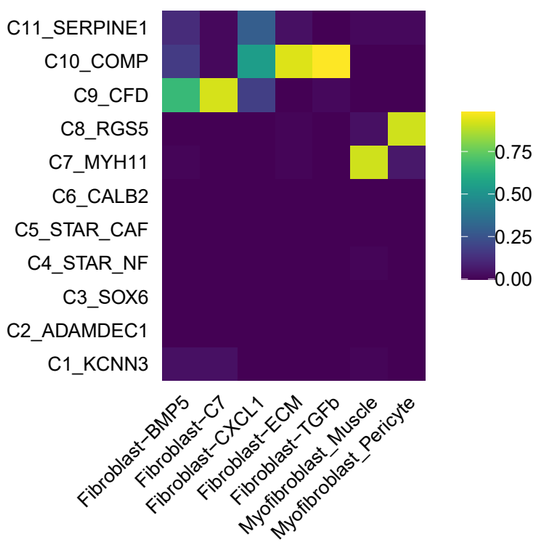



**Figure S2**

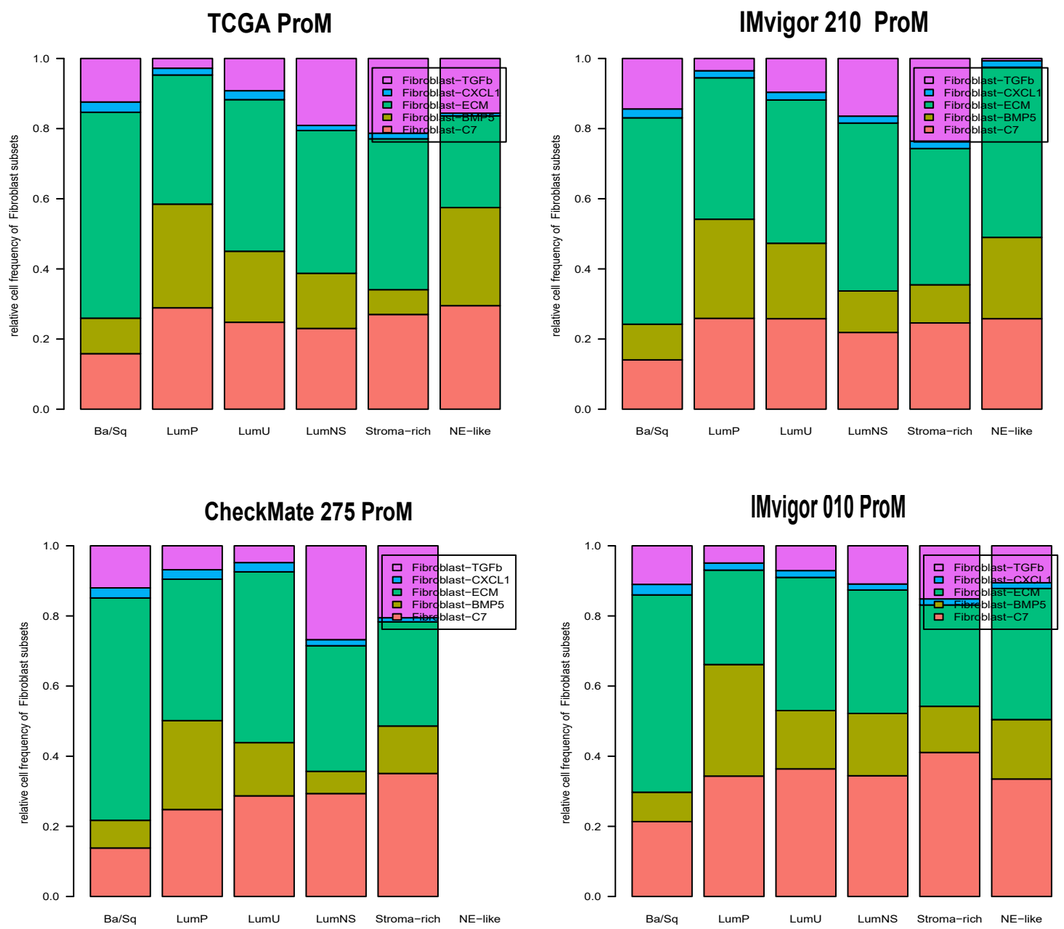
A



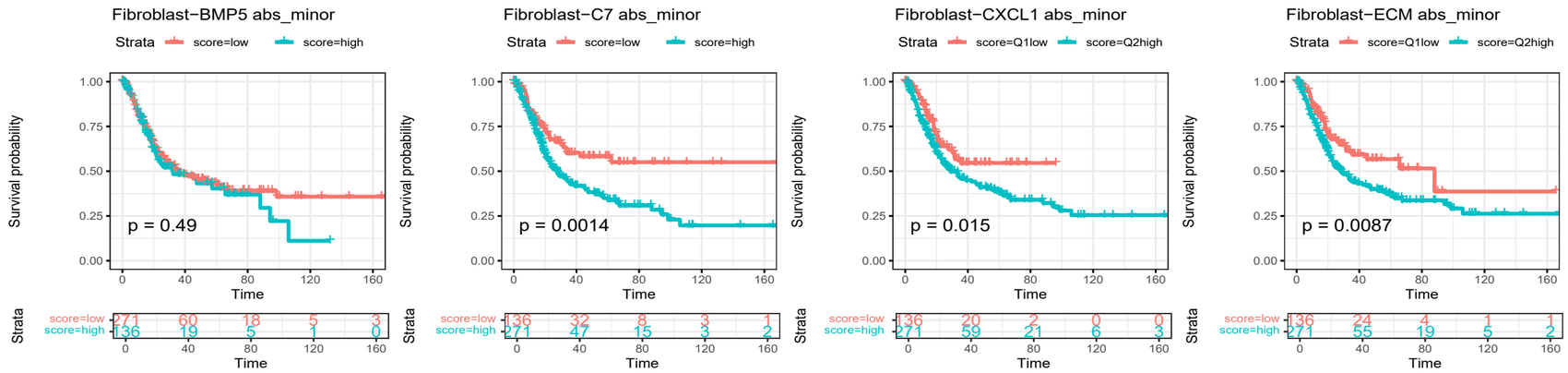

B




C




D




E

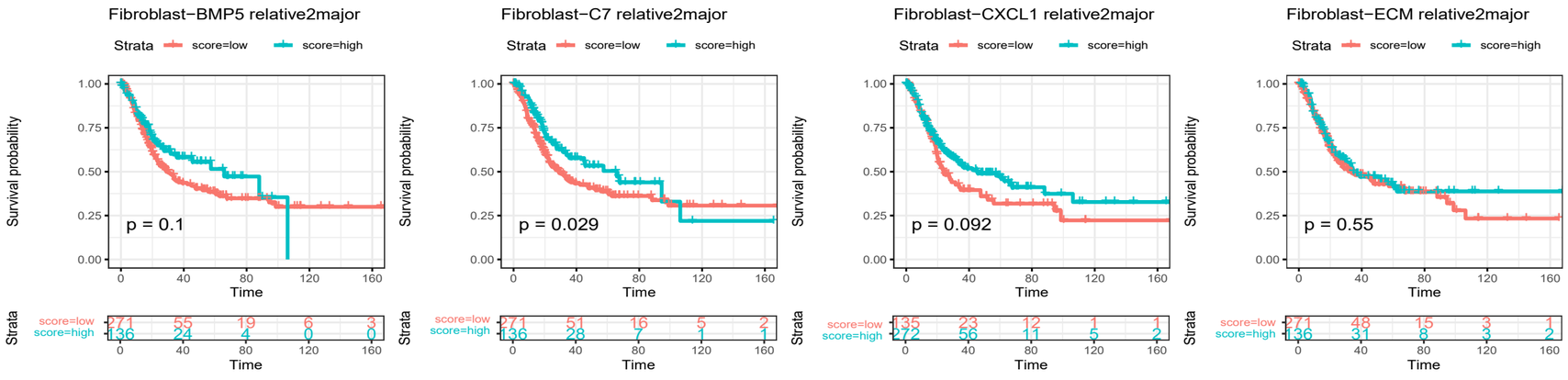



**Figure S3**

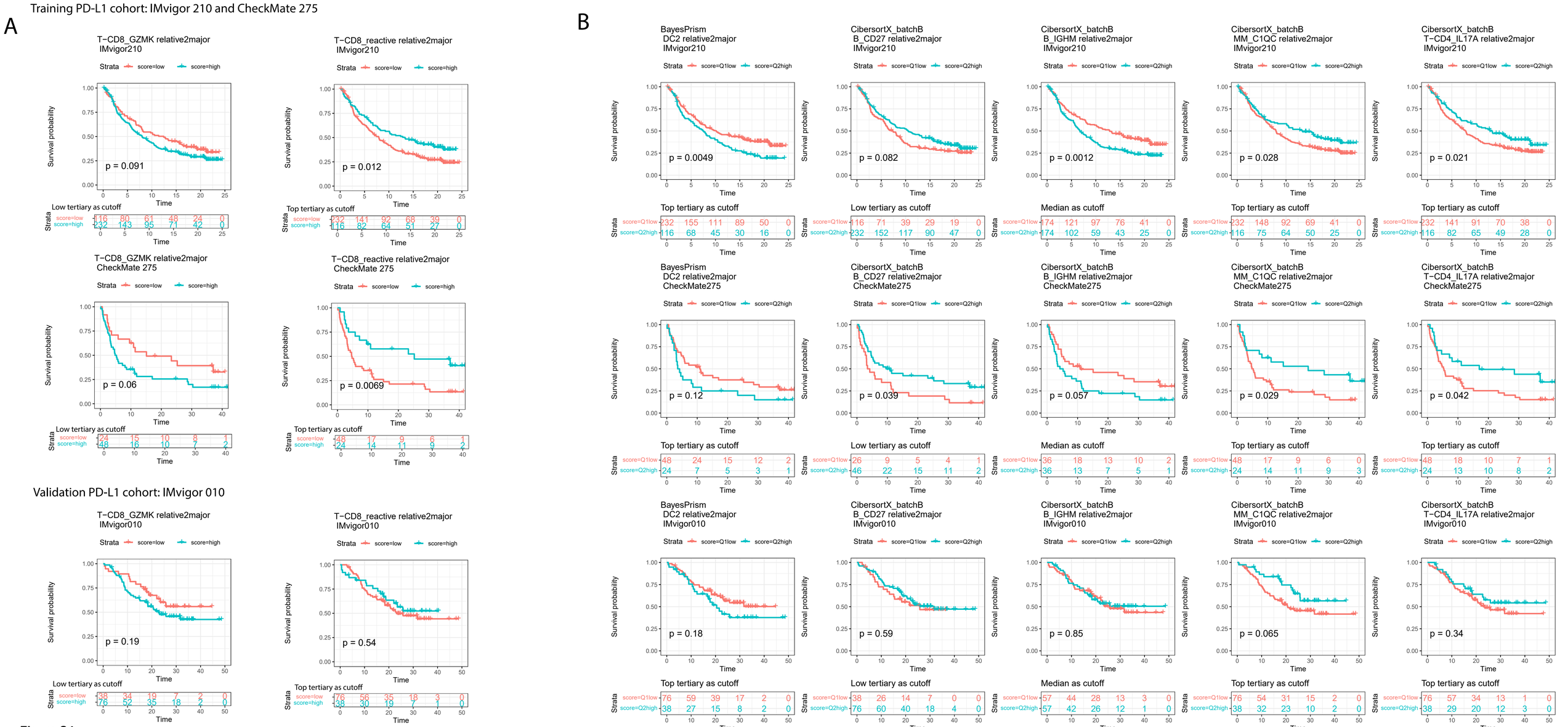
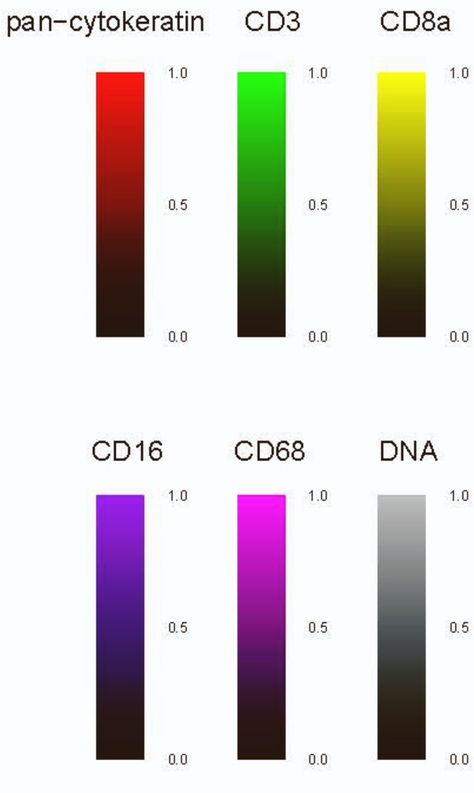
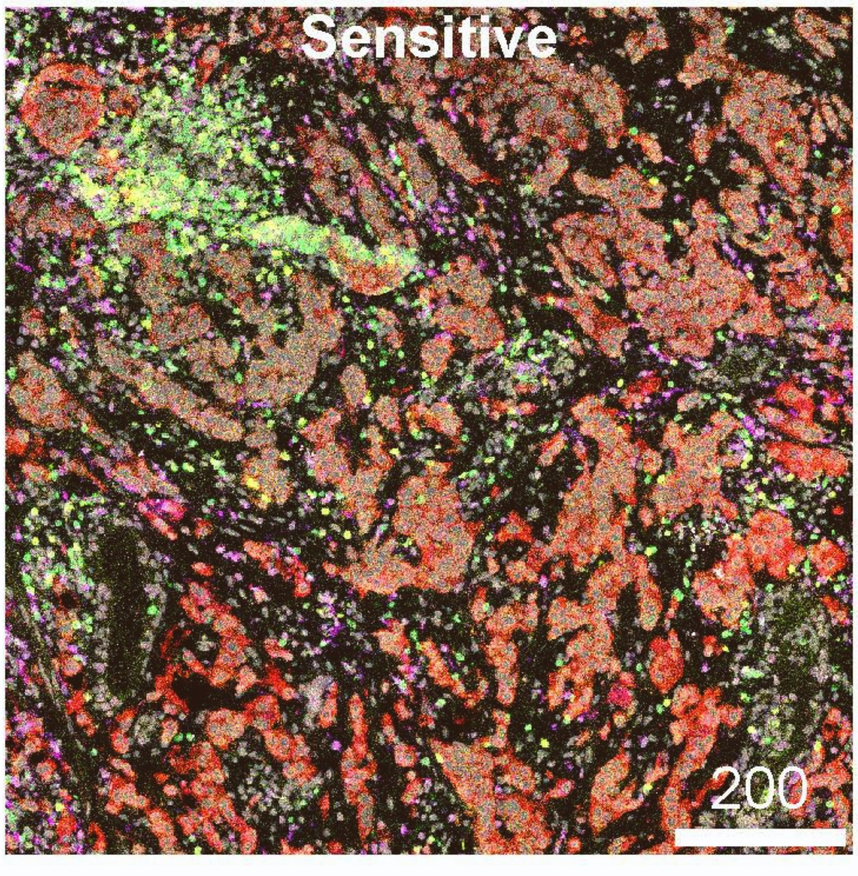
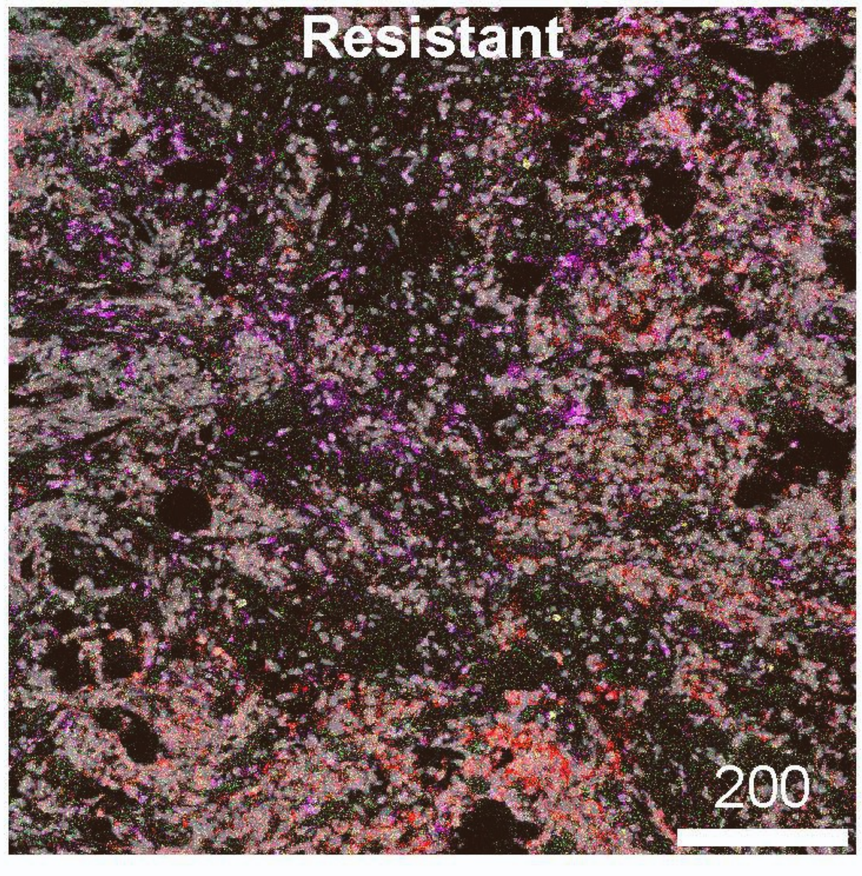

A
Training PD-L1 cohort: IMvigor 210 and CheckMate 275
T−CD8_GZMK relative2major
IMvigor210
p = 0.091
Low tertiary as cutoff
T−CD8_reactive relative2major
IMvigor210
p = 0.012
Top tertiary as cutoff
T−CD8_GZMK relative2major
CheckMate 275
p = 0.06
Low tertiary as cutoff
T−CD8_reactive relative2major
CheckMate 275
p = 0.0069
Top tertiary as cutoff
Validation PD-L1 cohort: IMvigor 010
T−CD8_GZMK relative2major
IMvigor010
p = 0.19
Low tertiary as cutoff
T−CD8_reactive relative2major
IMvigor010
p = 0.54
Top tertiary as cutoff
B
BayesPrism
DC2 relative2major
IMvigor210
p = 0.0049
Top tertiary as cutoff
CibersortX_batchB
B_CD27 relative2major
IMvigor210
p = 0.082
Low tertiary as cutoff
CibersortX_batchB
B_IGHM relative2major
IMvigor210
p = 0.0012
Median as cutoff
CibersortX_batchB
MM_C1QC relative2major
IMvigor210
p = 0.028
Top tertiary as cutoff
CibersortX_batchB
T−CD4_IL17A relative2major
IMvigor210
p = 0.021
Top tertiary as cutoff
BayesPrism
DC2 relative2major
CheckMate275
p = 0.12
Top tertiary as cutoff
CibersortX_batchB
B_CD27 relative2major
CheckMate275
p = 0.039
Low tertiary as cutoff
CibersortX_batchB
B_IGHM relative2major
CheckMate275
p = 0.057
Median as cutoff
CibersortX_batchB
MM_C1QC relative2major
CheckMate275
p = 0.029
Top tertiary as cutoff
CibersortX_batchB
T−CD4_IL17A relative2major
CheckMate275
p = 0.042
Top tertiary as cutoff
BayesPrism
DC2 relative2major
IMvigor010
p = 0.18
Top tertiary as cutoff
CibersortX_batchB
B_CD27 relative2major
IMvigor010
p = 0.59
Low tertiary as cutoff
CibersortX_batchB
B_IGHM relative2major
IMvigor010
p = 0.85
Median as cutoff
CibersortX_batchB
MM_C1QC relative2major
IMvigor010
p = 0.065
Top tertiary as cutoff
CibersortX_batchB
T−CD4_IL17A relative2major
IMvigor010
p = 0.34
Top tertiary as cutoff
Survival probability
Time
Strata
score=low
score=high
score=Q1low
score=Q2high

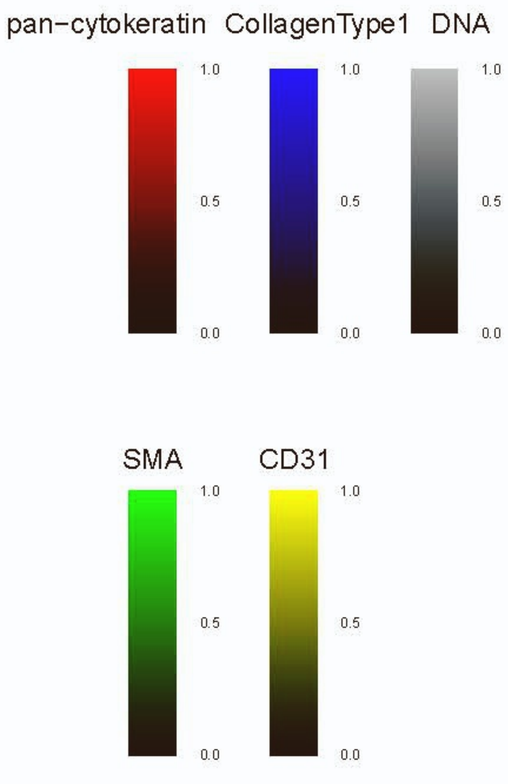
**Figure S4**

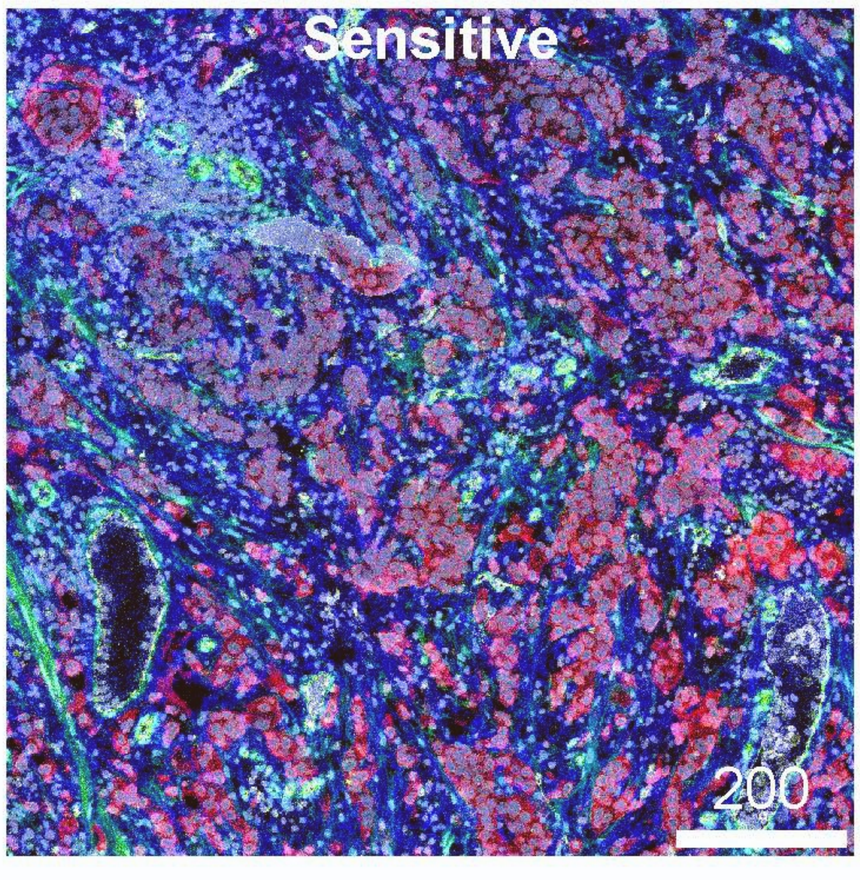
## Pixel-level marker expression by IMC

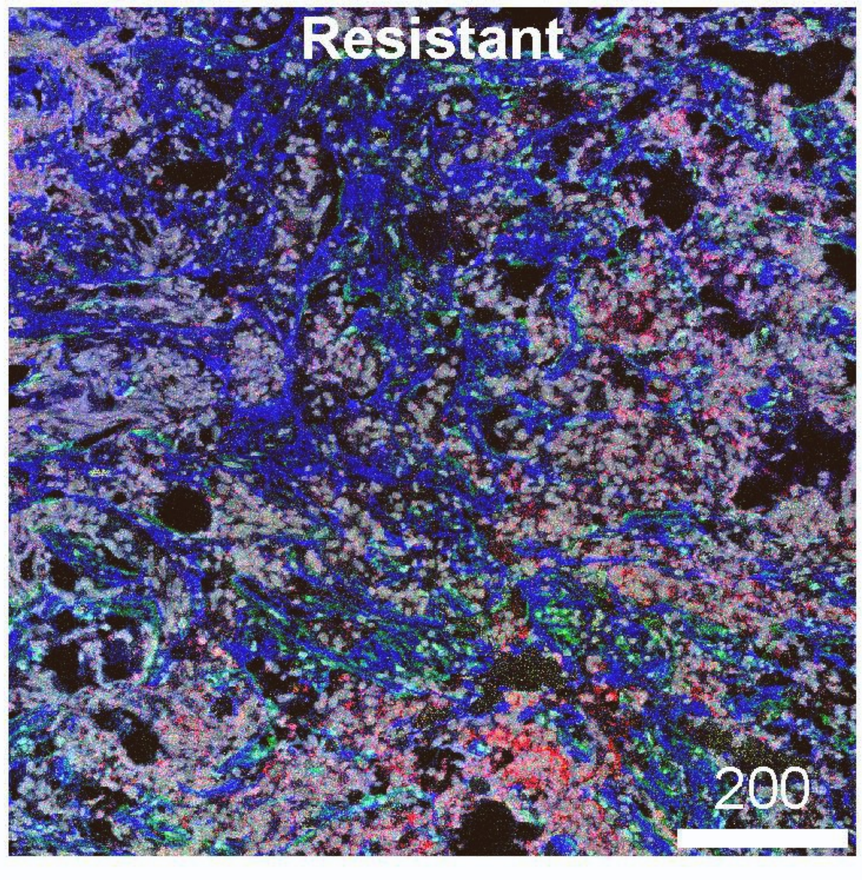


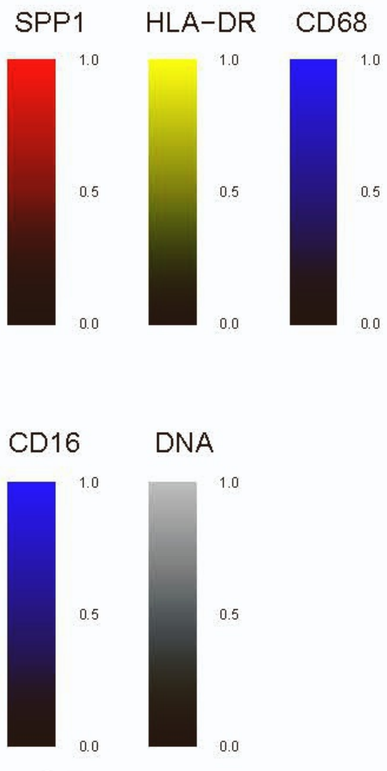



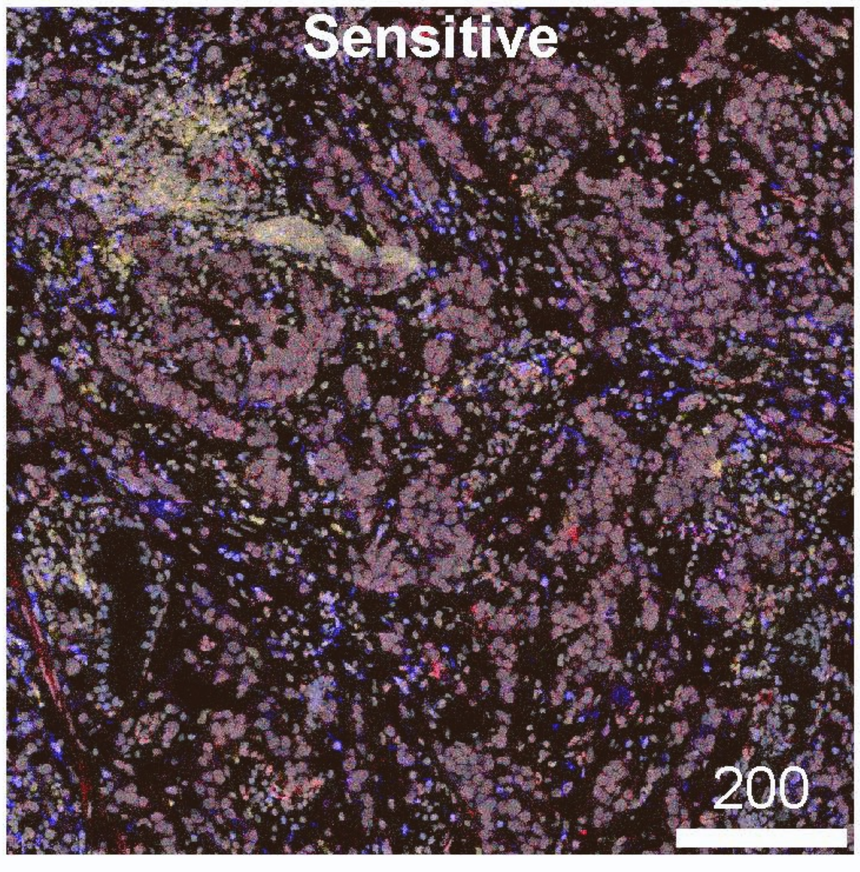



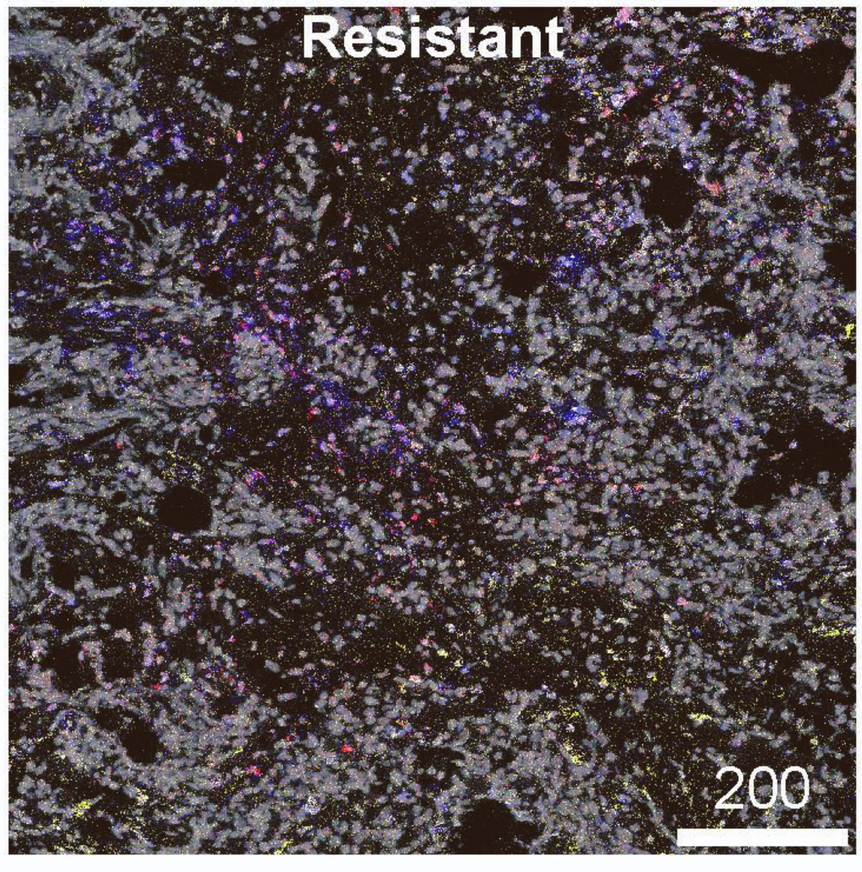




**Figure S5**

**Supplementary Figure Legends**

**Figure S1 Integrative analysis of UC single cell cohort**; **related to Figure 1 (A) QC plots for each sample in UC single cell cohort** (B) UMAP of all cells before removing "batch effects" colored by different sorting techniques used. (C) UMAP of all cells after removing "batch effects" colored by different sorting techniques used. (D) UMAP of all cells after removing "batch effects" colored by cell types and split by the various sorting techniques.

**Figure S2 Evaluation of ProM's deconvolution performance; related to Figure 2** (A) Deconvolution accuracies of different algorithms using reconstituted bulk RNAseq profiles in UC. Y-axis indicates Mean Square Error (MSE) between deconvoluted and reconstituted cell frequency averaged across allthe minor or major cell clusters. (B) Deconvolution accuracies when different numbers of cell markers were selected and different marker ranking strategies were employed in marker-based ProM deconvolution. (C) Deconvolution accuracies of ProM with and without cross-platform gene variation modeling. The x-axis represents different levels of cross-platform noise added to the reconstituted bulk profiles. (D) Deconvolution accuracies of different algorithms using reconstituted bulk RNAseq from pan cancer dataset. (E) Deconvolution accuracies of different deconvolution algorithms using paired single cell and bulk RNAseq data. The color in the heatmap represents the MSE between deconvoluted and observed frequency for particular cell subpopulations. The barplot above the heatmap indicates the average MSE across all cell subpopulations. (F) Scatter plot of ProM deconvoluted cell frequency form bulk RNAseq sample versus the observed cell frequency via matched scRNAseq sample for all 40 subpopulations.

**Figure S3 Heterogeneity of fibroblast subpopulations; related to Figure 3 and Figure 4** (A) The copy number variation (CNV) for each cell inferred from its scRNAseq profile using R package inferCNV (http://github/broadinstitute/inferCNV) for scRNAseq sample SI_651 or the paired bulk sample 22144. The upper-panel is the inferred CNV for non-epithelial cells (immune and stromal cells), and the lower-panel is that for epithelial cells split into luminal and basal groups. (B) Mapping of Fibroblast-related cells in our scRNAseq cohort to fibroblast cell annotations in a pan-cancer scRNAseq study[1]. Colors in the heatmap represent the percentage of a certain cell subpopulation (x-axis) predicted as the other cell subpopulations (y-axis) by the SciBet model. (C) The relative frequency of the five fibroblast-subpopulations by ProM deconvolution averaged within each tumor subtype in four UC bulk RNAseq datasets. (D)(E) Kaplan-Meier (KM) curves of overall survival for TCGA BLCA samples grouped by the frequency of each fibroblast subpopulation among all cells (D) or by the relative frequency of each fibroblast subpopulation among all five fibroblast subpopulations (E). Three different cutoffs, i.e., the lower tertiary vs. the median and higher tertiary, were used to split patients into low and high cell frequency group. The cutoff resulting in the most significant p-value (log rank test) was shown here. The p-value of likelihood ratio test is 0.20, 0.05, 0.8, 0.019 (D) and 0.11, 0.21, 0.37, 0.79 (E), respectively, when cell frequency was treated as continuous variable in Cox-regression model.

**Figure S4 Survival Analysis of UC treated with PD-1/PD-L1 blockade; related to Figure 5.** (A)KM curves of overall survival for patients grouped by T-CD8_GZMK and T-CD8_reactive as estimated by ProM deconvolution in three PD-L1 clinical cohorts. .**(B)** Using the same training datasets (IMvigor210 and CheckMate275) and the same criteria as in ProM analysis, the other nine deconvolution methods detected a total of five subpopulations significantly associated with overall survival (four by CibersortX_batchB and one by MuSiC). None of the findings by CibersotX_batchB and MuSiC can be replicated in the validation dataset (IMvigor010).

**Figure S5 Exploratory study of MM_SPP1 using Imaging Mass Cytometry; related to Figure 5.** Visualization of pixel-level intensity of immune cell markers(upper), stromal cell markers(middle) and MM subpopulation markers(lower) within region of interest (ROI) from one patient sensitive (left) or resistant (right) to CPI.

**Supplementary Reference**

1. Qian, J., Olbrecht, S., Boeckx, B., Vos, H., Laoui, D., Etlioglu, E., Wauters, E., Pomella, V., Verbandt, S., and Busschaert, P. (2020). A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell research *30*, 745-762.

**Supplementary Tables**

**Table S1 samples in the paired single cell and bulk RNAseq cohort: related to Figure 1**

| Patient ID | scRNAseq | | | | Bulk RNAseq |
|---|---|---|---|---|---|
| | Sample ID | Cell sorting | Specimen | Published | Sample ID |
| PI_21 | SI_357 | 10x_CD45_neg | TURBT | PMID: 32111252 PMID: 36129800 | 17760 |
| PI_22 | SI_359 | 10x_CD45_neg | TURBT | PMID: 32111252 PMID: 36129800 | 17766 |
| PI_7 | SI_454 | 10x_CD45_neg | TURBT | PMID: 36099881 PMID: 36129800 | 19639 |
| PI_8 | SI_457 | 10x_CD45_neg | Cystectomy | PMID: 36099881 PMID: 36129800 | 18418 |
| PI_21 | SI_358 | 10x_CD45_pos | TURBT | PMID: 36129800 | 17760 |
| PI_22 | SI_360 | 10x_CD45_pos | TURBT | PMID: 36129800 | 17766 |
| PI_1 | SI_385 | 10x_CD45_pos | TURBT | PMID: 36099881 PMID: 36129800 | 18763 |
| PI_2 | SI_394 | 10x_CD45_pos | TURBT | PMID: 36099881 PMID: 36129800 | NA |
| PI_3 | SI_397 | 10x_CD45_pos | TURBT | PMID: 36099881 PMID: 36129800 | 18859 |
| PI_4 | SI_421 | 10x_CD45_pos | TURBT | PMID: 36099881 PMID: 36129800 | 19425 |
| PI_4 | SI_422 | 10x_tumor_NK | TURBT | NA | NA |
| PI_4 | SI_423 | 10x_tumor_CD8 | TURBT | NA | NA |
| PI_4 | SI_424 | 10x_tumor_CD4 | TURBT | NA | NA |
| PI_5 | SI_434 | 10x_CD45_pos | TURBT | PMID: 36099881 PMID: 36129800 | 16877 |
| PI_6 | SI_435 | 10x_CD45_pos | Cystectomy | PMID: 36099881 PMID: 36129800 | NA |
| PI_7 | SI_453 | 10x_CD45_pos | TURBT | PMID: 36099881 PMID: 36129800 | 19639 |
| PI_8 | SI_456 | 10x_CD45_pos | Cystectomy | PMID: 36099881 PMID: 36129800 | 18418 |
| PI_15 | SI_584 | 10x_unsorted | Cystectomy | PMID: 33837006 PMID: 36099881 | NA |
| PI_15 | SI_585 | 10x_unsorted | Adjacent Normal | PMID: 33837006 | NA |
| PI_9 | SI_595 | 10x_unsorted | Cystectomy | PMID: 33837006 PMID: 36099881 | NA |
| PI_9 | SI_596 | 10x_unsorted | Adjacent Normal | PMID: 33837006 | NA |
| PI_14 | SI_603 | 10x_unsorted | TURBT | PMID: 33837006 | 21753 |
| PI_17 | SI_604 | 10x_unsorted | TURBT | PMID: 33837006 | 21783 |

| | | | | | |
|---|---|---|---|---|---|
| PI_18 | SI_613 | 10x_unsorted | TURBT | PMID: 33837006 PMID: 36099881 | NA |
| PI_14 | SI_651 | 10x_unsorted | Cystectomy | PMID: 33837006 PMID: 36099881 | 22144 |
| PI_23 | SI_656 | 10x_unsorted | Cystectomy | PMID: 33837006 PMID: 36099881 | 22171 |
| PI_16 | SI_659 | 10x_unsorted | Cystectomy | NA | NA |
| PI_24 | SI_695 | 10x_unsorted | TURBT | PMID: 33837006 PMID: 36099881 | 22605 |
| PI_25 | SI_696 | 10x_mixed | TURBT | NA | NA |
| PI_26 | SI_697 | 10x_mixed | Lymph Node | NA | 22612 |

**Table S2 Patient clinical information: related to Figure 1**

| Patient ID | Gender | Age | Stage | Race |
|---|---|---|---|---|
| PI_1 | M | 50-60 | cT2N0M0 | White |
| PI_2 | M | 70-80 | cT0N0M0 | White |
| PI_3 | M | 60-70 | cT3N0M0 | White |
| PI_4 | F | 80-90 | cTaN0M0 | White |
| PI_5 | M | 60-70 | pT4bN3M0 | White |
| PI_6 | M | 80-90 | pT3bN0M0 | White |
| PI_7 | M | 80-90 | cT2N0M0 | White |
| PI_8 | M | 60-70 | yT4N2M0 | Asian.Indian |
| PI_9 | M | 70-80 | pT2N0M0 | White |
| PI_14 | F | 40-50 | cT2N0M0 | Asian |
| PI_15 | M | 60-70 | pT4N0M0 | White |
| PI_16 | M | 60-70 | pT0N0M0 | White |
| PI_17 | F | 70-80 | cT2N0M0 | White |
| PI_18 | F | 80-90 | cT3N0M0 | White |
| PI_21 | F | 80-90 | cT2N0M0 | White |
| PI_22 | M | 50-60 | cT2N0M0 | White |
| PI_23 | M | 70-80 | pT3bN2M0 | White |
| PI_24 | M | 60-70 | cT2N0M0 | White |
| PI_25 | F | 70-80 | pT0N0M0 | Unknown |
| PI_26 | M | 60-70 | pT3aN2M0 | White |

**Table S3 Comparison between ProM and other deconvolution methods; related to Figure 2**

| | Reconstituted bulk: CD45-sorted Reference scRNAseq: unsorted | | | | Reconstituted bulk: unsorted Reference scRNAseq: CD45-sorted | | | | Paired scRNAseq and bulk RNAseq | | TCGA (by Methylation) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Minor cluster | | Major cluster | | Minor cluster | | Major cluster | | Minor cluster | | Major cluster | |
| Comparison with ProM | t-stat* | p-value* | t-stat | p-value | t-stat | p-value | t-stat | p-value | t-stat | p-value | t-stat | p-value |
| ProM_profileOnly | 2.41 | 1.0E-02 | 0.87 | 2.0E-01 | 1.48 | 7.3E-02 | 1.08 | 1.5E-01 | 4.25 | 6.4E-05 | 2.36 | 2.8E-02 |
| ProM_markerOnly | 2.72 | 4.8E-03 | 2.08 | 2.9E-02 | 2.50 | 8.4E-03 | 2.93 | 5.8E-03 | 0.26 | 4.0E-01 | 2.87 | 1.4E-02 |
| CibersortX_batchB | 4.58 | 2.3E-05 | 2.48 | 1.4E-02 | 3.34 | 9.2E-04 | 3.95 | 8.3E-04 | 2.41 | 1.0E-02 | 2.59 | 2.1E-02 |
| CibersortX | 4.85 | 1.0E-05 | 2.90 | 6.2E-03 | 4.01 | 1.3E-04 | 3.01 | 5.0E-03 | 2.36 | 1.2E-02 | 2.91 | 1.3E-02 |
| BayesPrism | 3.89 | 1.9E-04 | 3.17 | 3.7E-03 | 3.61 | 4.3E-04 | 2.75 | 8.2E-03 | 4.54 | 2.7E-05 | 3.18 | 9.6E-03 |
| MuSiC | 5.59 | 9.7E-07 | 2.98 | 5.3E-03 | 5.67 | 7.5E-07 | 3.10 | 4.2E-03 | 4.91 | 8.4E-06 | 2.90 | 1.4E-02 |
| CibersortX_batchS | 7.09 | 8.1E-09 | 3.00 | 5.1E-03 | 7.31 | 4.0E-09 | 4.30 | 4.3E-04 | 5.54 | 1.1E-06 | 4.61 | 1.8E-03 |
| BisqueRNA | 8.92 | 3.0E-11 | 3.91 | 9.0E-04 | 6.85 | 1.7E-08 | 4.99 | 1.2E-04 | 5.25 | 2.9E-06 | 3.37 | 7.5E-03 |
| DeconRNASeq | 9.11 | 1.7E-11 | 3.93 | 8.7E-04 | 8.82 | 3.9E-11 | 7.46 | 2.4E-06 | 6.36 | 8.3E-08 | 4.32 | 2.5E-03 |
| nnls | 10.65 | 2.1E-13 | 4.45 | 3.3E-04 | 8.82 | 4.0E-11 | 8.05 | 1.0E-06 | 5.96 | 3.0E-07 | 4.27 | 2.6E-03 |
| DSA | 12.27 | 2.9E-15 | 4.90 | 1.4E-04 | 8.36 | 1.6E-10 | 12.50 | 6.4E-09 | 8.31 | 1.8E-10 | 4.82 | 1.5E-03 |

*Pearson's correlation coefficient (PCC) between deconvolution results and "ground truth" was calculated for each cell subpopulation (40 minor cell clusters and 15 major clusters).Paired t-test was conduct to compare the difference of PCC between ProM and the other method across all cell subpopulations (sample size = 40 for minor cluster and 15 for major cluster). The t-statistic and one-sided p-value was shown in the table.