

Single-Cell RNA Sequencing Identifies Molecular Biomarkers Predicting Late Progression to CDK4/6 Inhibition in Patients with HR+/HER2- Metastatic Breast Cancer

Linjie Luo^{1,7,**}, Peng Yang^{2,3,7}, Sofia Mastoraki¹, Xiayu Rao⁴, Yan Wang¹, Nicole M. Kettner¹, Akshara Singareeka Raghavendra⁵, Debasish Tripathy⁵, Senthil Damodaran⁵, Kelly K. Hunt⁶, Jing Wang⁴, Ziyi Li³, Khandan Keyomarsi^{1,*}

¹Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

²Department of Statistics, Rice University, Houston, TX, USA

³Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

⁴Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

⁵Department of Breast Medical Oncology, Division of Cancer Medicine, The University of MD Anderson Cancer Center, Houston, TX, USA

⁶Department of Breast Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

⁷These authors contributed equally

*Correspondence: kkeyomar@mdanderson.org

**Co-correspondence: lluo2@mdanderson.org

Figure S1. UMAP of each sample and differentially expressed gene analysis on tumor cells

- (A) Distinguishing tumor versus non-tumor cells via InferCNV across each sample.
- (B) UMAP plot across 18 samples in the embedding space.
- (C) Volcano plot showing differentially expressed genes between the BL and EP samples.
- (D) Volcano plot showing differentially expressed genes between the BL and LP samples.
- (E) Volcano plot showing differentially expressed genes between the EP and LP samples.
- (F) MP analysis heatmap showing the normalized module scores across group comparisons.

Figure S1

A

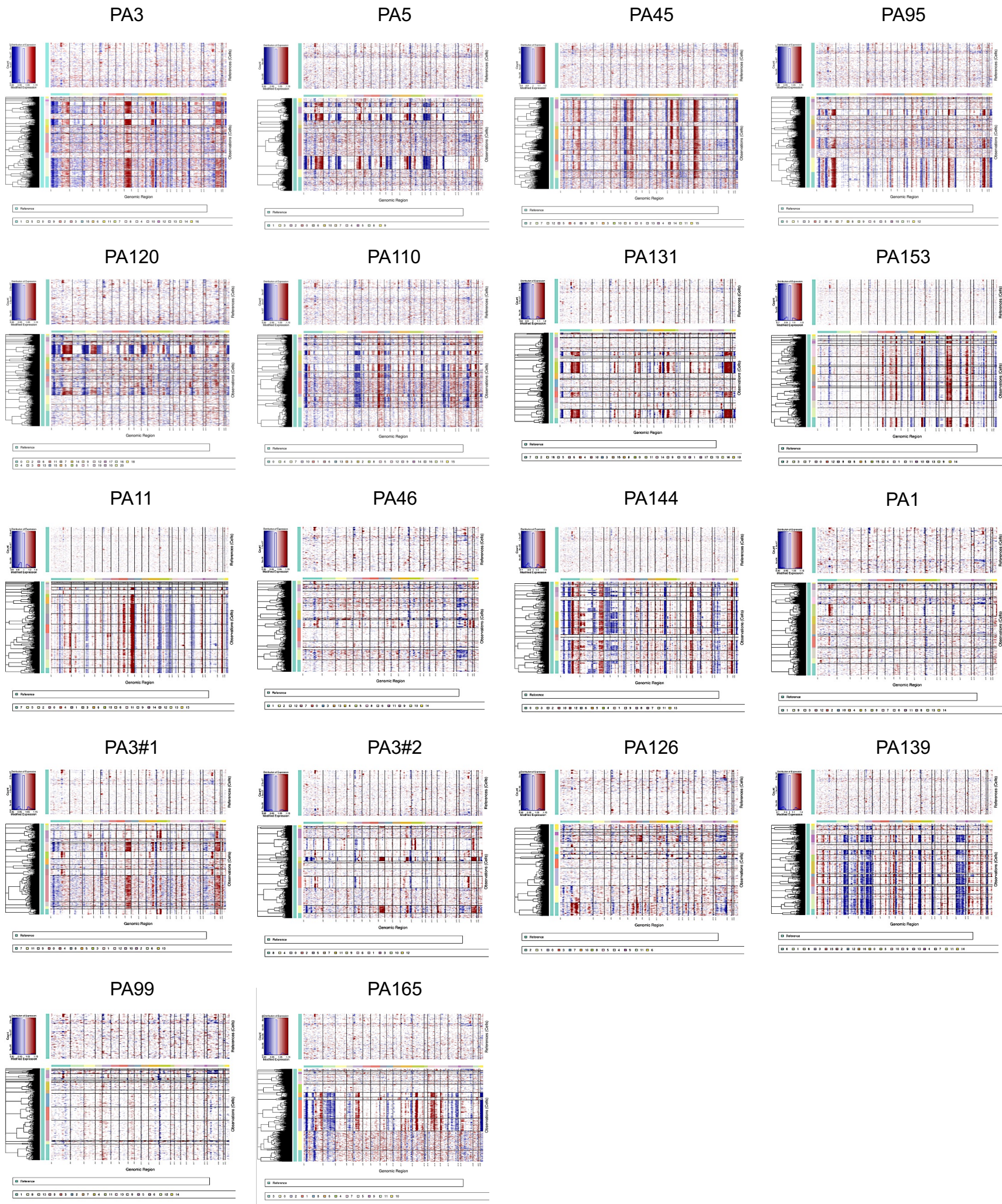


Figure S1

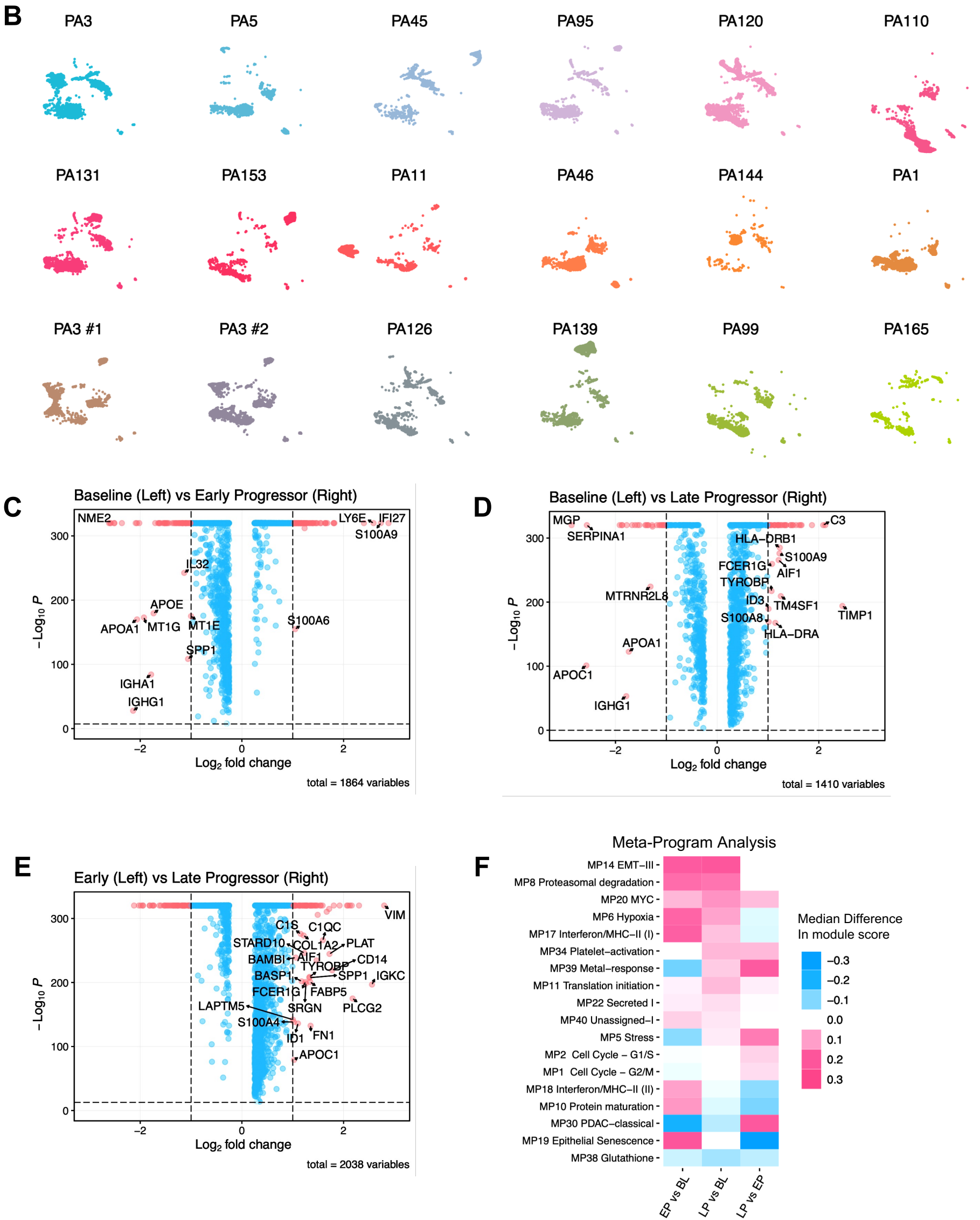


Figure S2. Tumor cell analysis based on metastatic site

(A) Heatmap showing the expression of top 10 differentially expressed genes, stratified by different sample statuses, annotated by groups and metastatic sites.

(B) Volcano plot showing differentially expressed genes between the pleural effusion (left) and liver metastases (right) in BL samples.

(C) Volcano plot showing differentially expressed genes between the pleural effusion (left) and liver metastasis (right) in LP samples.

(D) Dot plot displaying the top 10 differentially expressed genes among BL, EP, and LP samples from pleural effusion.

(E) Dot plot displaying the top 15 differentially expressed genes between BL and LP samples from liver metastases.

Figure S2

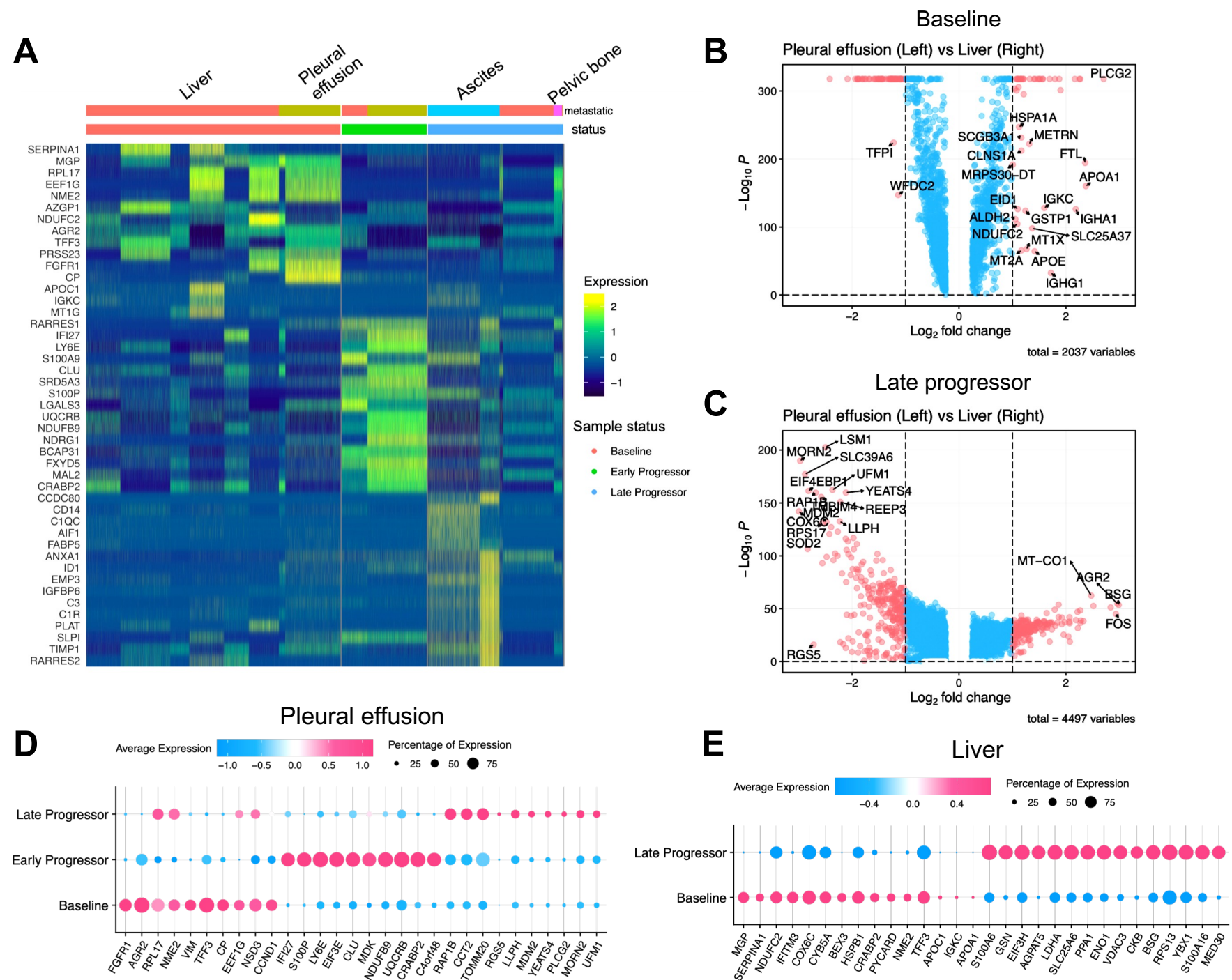


Figure S3. Non-tumor cells of minor-annotation cell subtype analysis

- (A) Chord diagram showing ligand-receptor (L-R) interactions between 2 cell types from BL samples.
- (B) Chord diagram showing L-R interactions between 2 cell types from EP samples.
- (C) Chord diagram showing L-R interactions between non-tumor cells from LP samples.
- (D) UMAP plots of minor non-tumor cell types stratified by different sample statuses (BL, EP, and LP).
- (E) Dot plot showing the marker gene expression for each minor non-tumor cell type in (A).
- (F) Cell numbers for each non-tumor cell type.
- (G) Cell fraction of double-negative T (dnT) cells, gamma delta T (gdT) cells, and mucosal-associated invariant T (MAIT) cell subtypes in BL, EP, and LP samples.

Figure S3

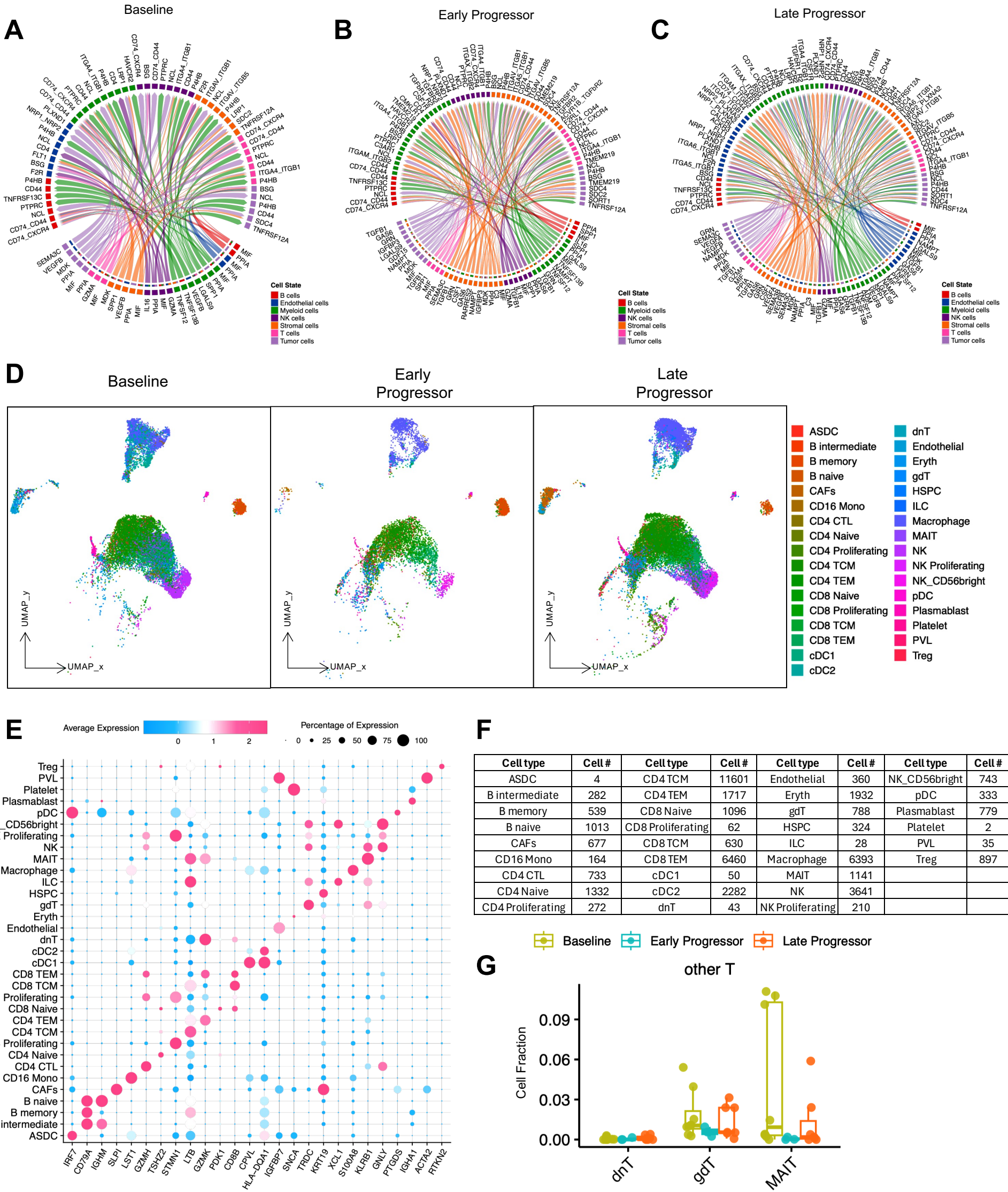


Figure S4. T cell analysis based on metastatic site

- (A) Expression of stress genes in CD4⁺ T cells from BL, EP, and LP samples.
- (B) Module score analysis of CD4⁺ T_{STR} cells among BL, EP, and LP samples.
- (C) Expression of stress genes in CD8⁺ T cells from BL, EP, and LP samples.
- (D) Module score analysis of CD8⁺ T_{STR} cells among BL, EP, and LP samples.
- (E) UMAP plots of 12 different subtypes of T cells stratified by metastatic site, including ascites, liver, pelvic bone, and pleural effusion.
- (F) Bar plot showing the relative fraction of T cell subtypes for each sample, annotated by group and metastatic site.
- (G) Dot plot displaying the top 15 differentially expressed genes between pleural effusion and liver metastasis from BL T cells.
- (H) Dot plot displaying the top 15 differentially expressed genes between pleural effusion and liver metastasis from LP T cells.

Figure S4

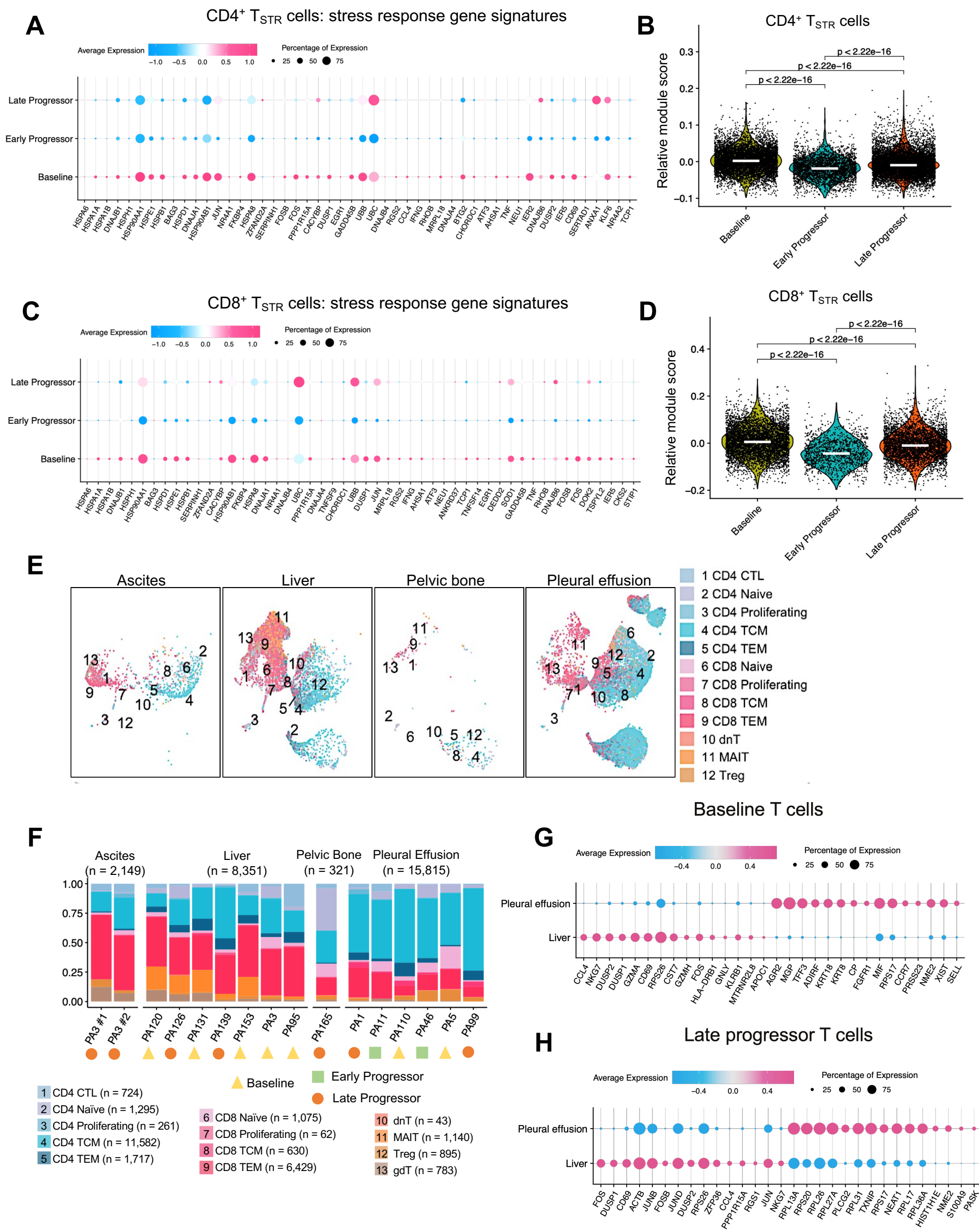


Figure S5. Analysis of other non-tumor cells

(A) t-SNE plot of 13 different subtypes of myeloid cells in the embedding space.

(B) Dot plot showing the expression of M1 and M2 macrophage gene signatures across each cluster from (A).

(C) Dot plots displaying the top 5 differentially expressed genes in each cluster from (A).

(D) UMAP plot of B cell populations stratified by 4 different cell subtypes.

(E) Comparison of fractions of different B cell subtypes among BL, EP, and LP samples.

(F) Violin plots showing significant changes of expression of *CSF1R* and *CCL2* among BL, EP, and LP samples. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

Figure S5

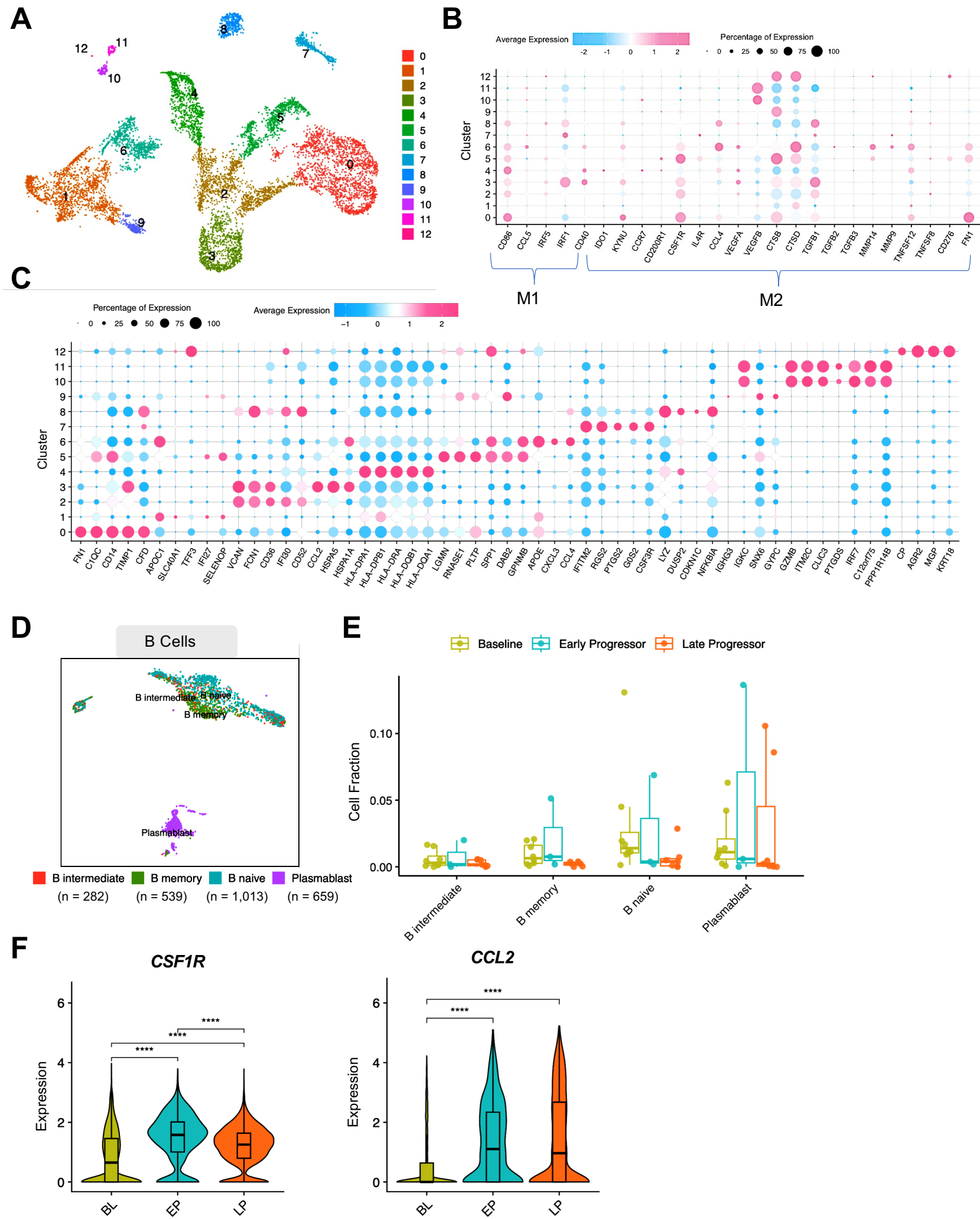


Figure S6. Longitudinal biopsy analysis

(A) Heatmap generated from MP analysis displaying the median module score differences from tumor cells between each sample.

(B) Ligand-receptor (L-R) interaction analysis conducted using CellChat in the BL sample PA3.

(C) Chord diagram illustrating L-R interactions between 2 cell types from the BL sample PA3.

(D) Chord diagram illustrating L-R interactions between 2 cell types from LP samples.

(E) UMAP plot of non-tumor cells with minor non-tumor cell types of annotation and Harmony integration.

(F) Alluvial plot showing the dynamic changes of non-tumor cell fractions across 3 longitudinal samples.

The color legend is the same as in panel (E).

Figure S6

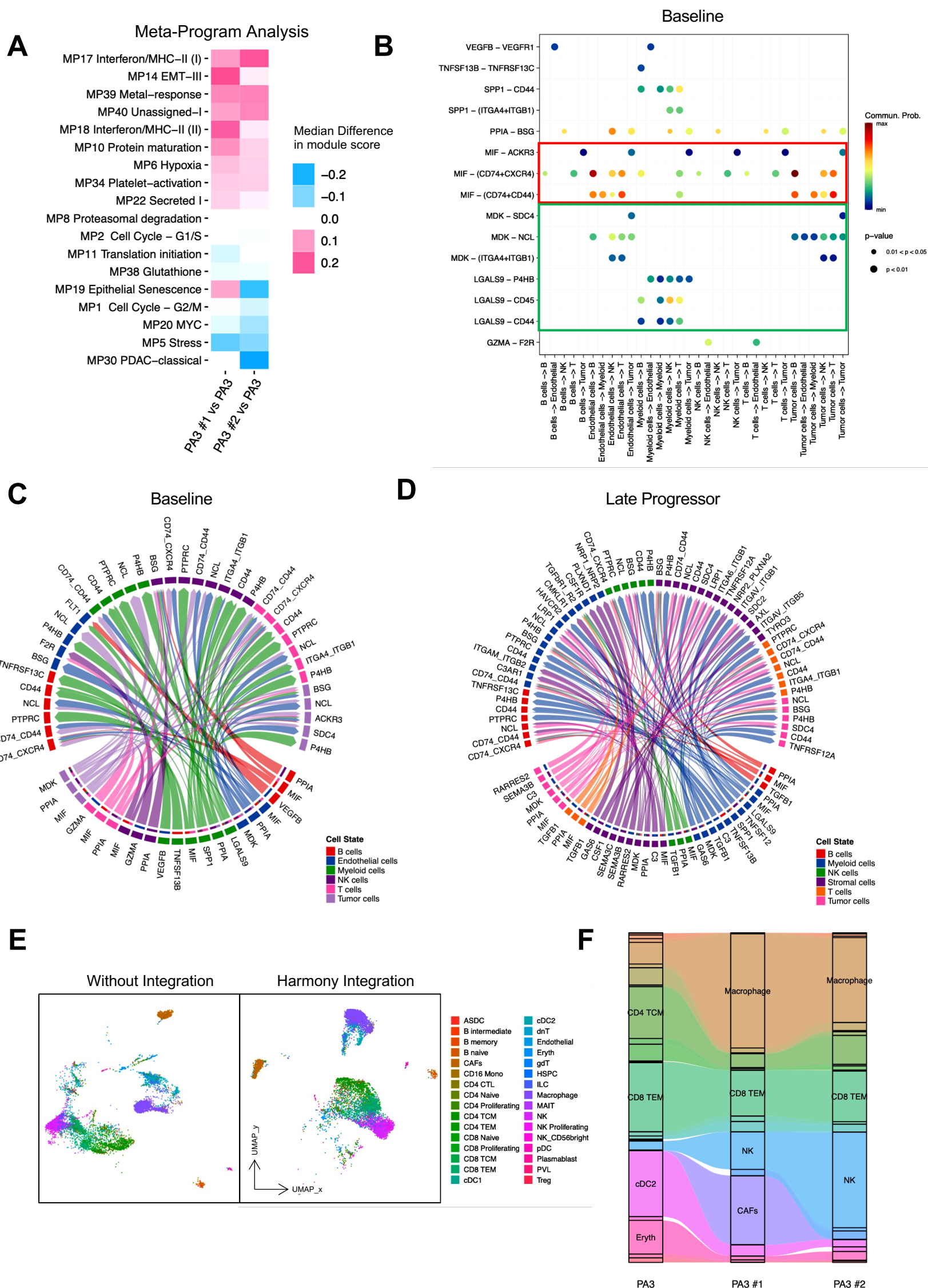
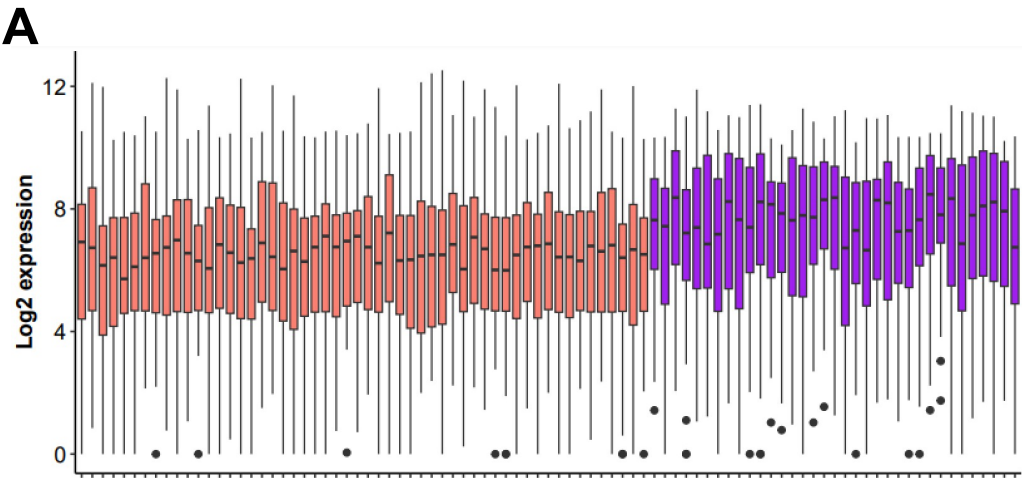


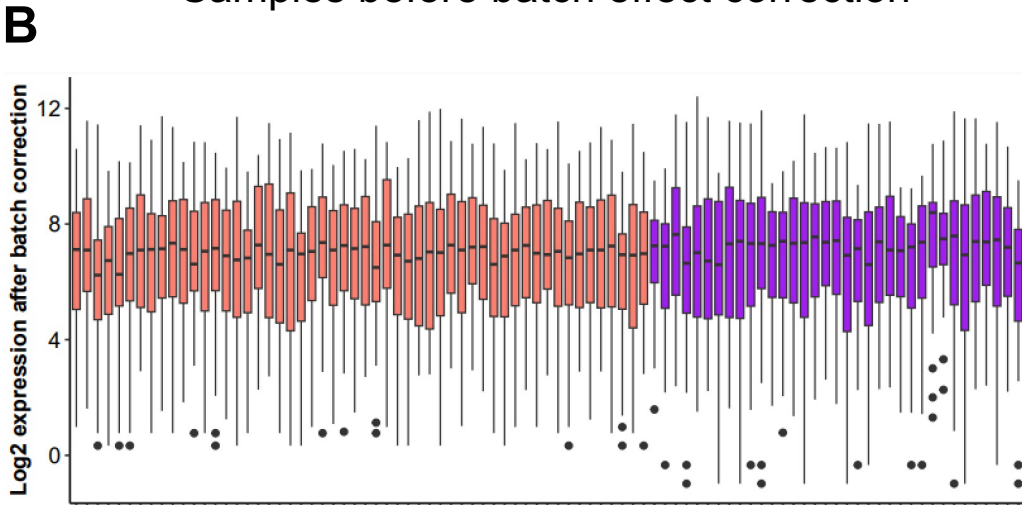
Figure S7. Predictive biomarker signature validation

- (A) Log2 expression of signature genes across all MDACC samples before the batch effect correction.
- (B) Log2 expression of signature genes across all MDACC samples after the batch effect correction.
- (C) KM plot illustrating median PFS in the signature-high group compared to the -low group of 89 combined FFPE and fresh tissue samples from the MD Anderson cohort.
- (D) KM plot indicating median PFS in the signature-high group compared to the -low group of 19 EP samples from the Korean cohort.
- (E) KM plot indicating median PFS in the signature-high group compared to the -low group of 42 LP samples from the Korean cohort.

Figure S7



Samples before batch effect correction



Samples after batch effect correction

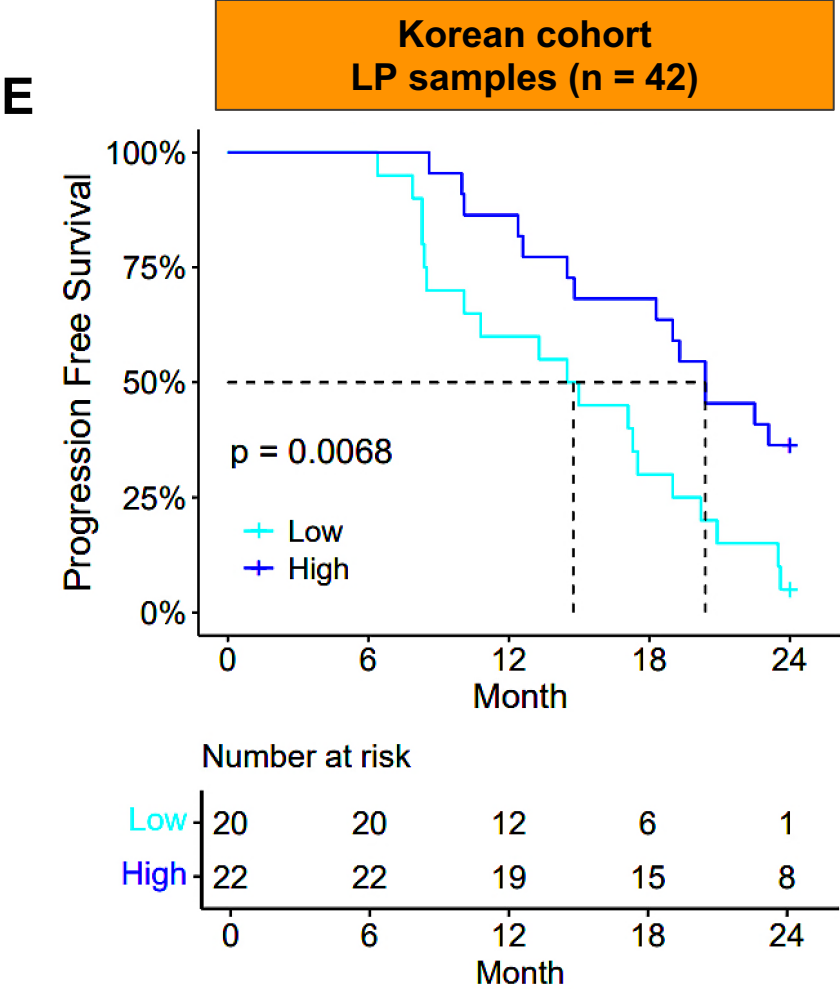
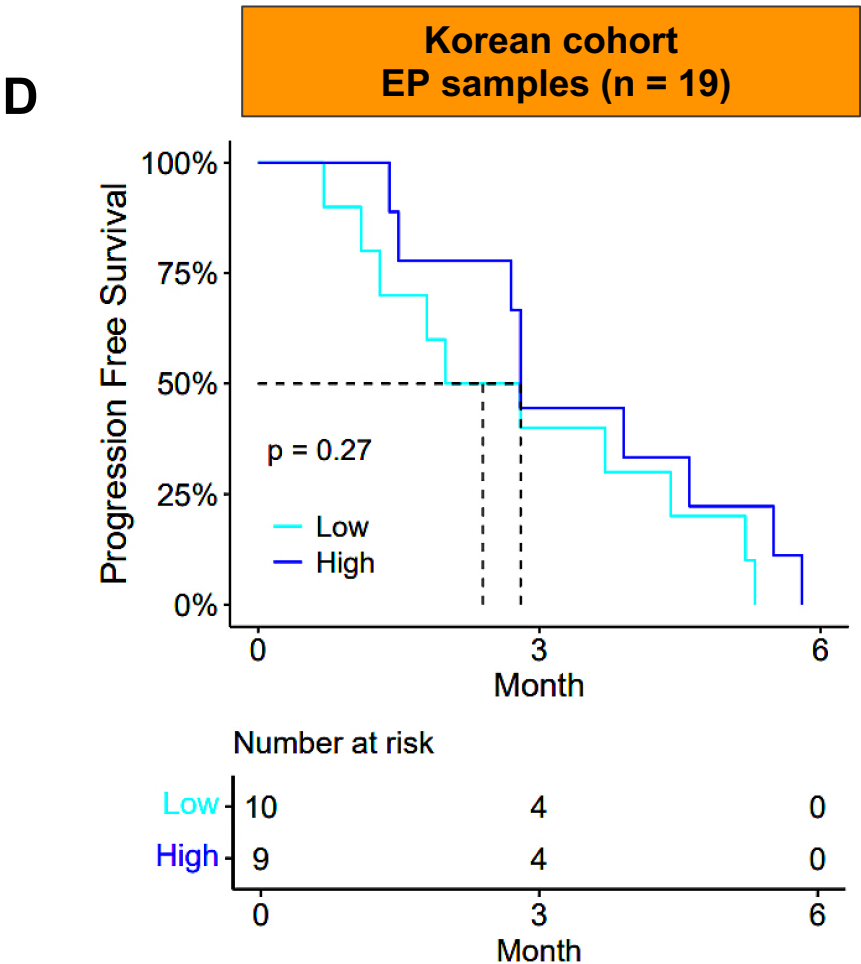
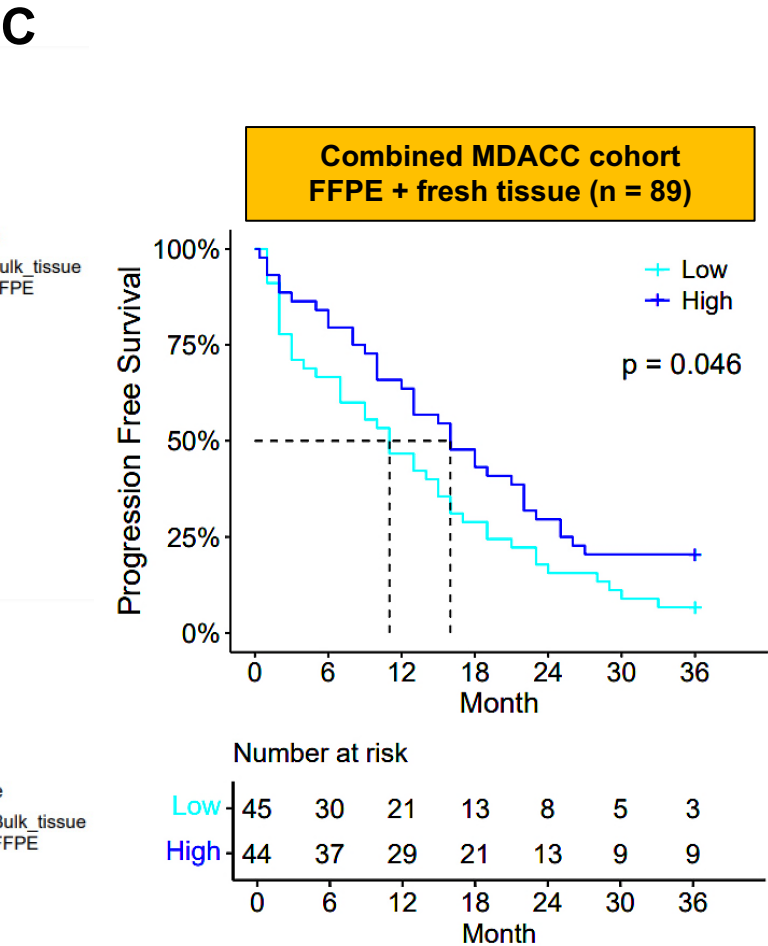
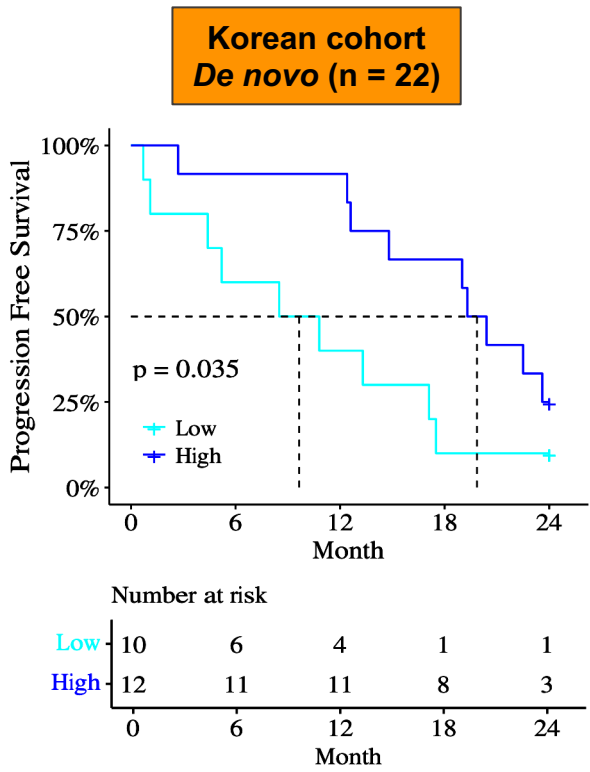


Figure S8. Broad applicability of gene signatures across clinical and biological contexts

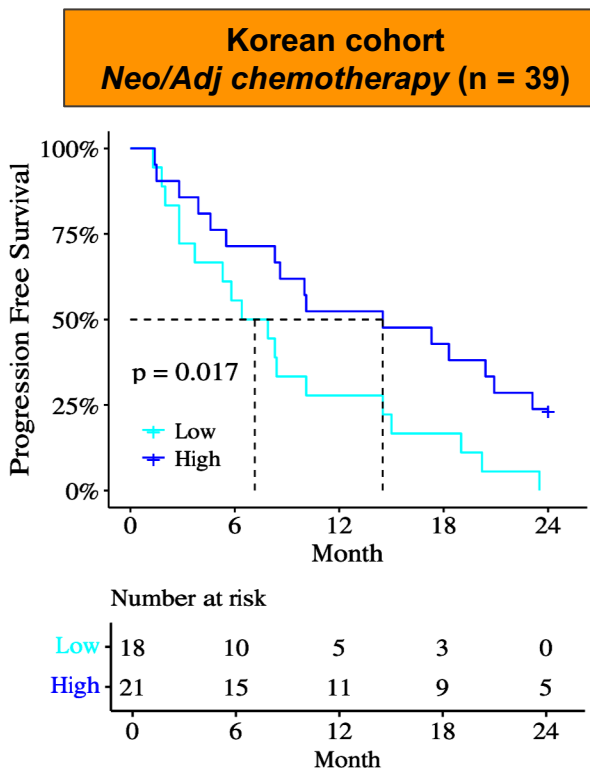
- (A) KM plot illustrating PFS in the signature-high group compared to the -low group in 22 *de novo* (untreated) samples from the Korean cohort from the Korean cohort.
- (B) KM plot illustrating PFS in the signature-high group compared to the -low group in 39 samples from patients who received neoadjuvant or adjuvant chemotherapy in the Korean cohort.
- (C) KM plot illustrating PFS in the signature-high group compared to the -low group in 36 primary breast tumors from the Korean cohort.
- (D) KM plot indicating PFS in the signature-high group compared to the -low group of 33 bone metastasis tissue samples from the MD Anderson cohort.
- (E) KM plot indicating PFS in the signature-high group compared to the -low group of 23 liver metastasis tissue samples from the MD Anderson cohort.
- (F) KM plot indicating PFS in stratified samples from different metastatic sites, including 223 bone, 62 liver, 38 lung, and 18 pleural effusions, from the master database of the MD Anderson cohort. Censored data points indicate patients who are still undergoing treatment and have not yet developed progression.
- (G) KM plot indicating PFS in the signature-high group compared to the -low group in 36 samples treated with CDK4/6i plus aromatase inhibitor (e.g., letrozole) from the MD Anderson cohort.
- (H) KM plot indicating PFS in the signature-high group compared to the -low group in 16 samples treated with CDK4/6i plus fulvestrant from the MD Anderson cohort. Log-rank test p-values are displayed.

Figure S8

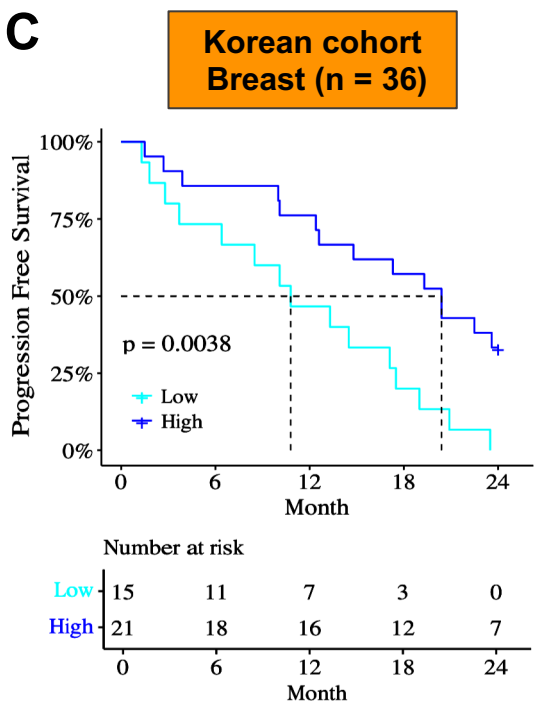
A



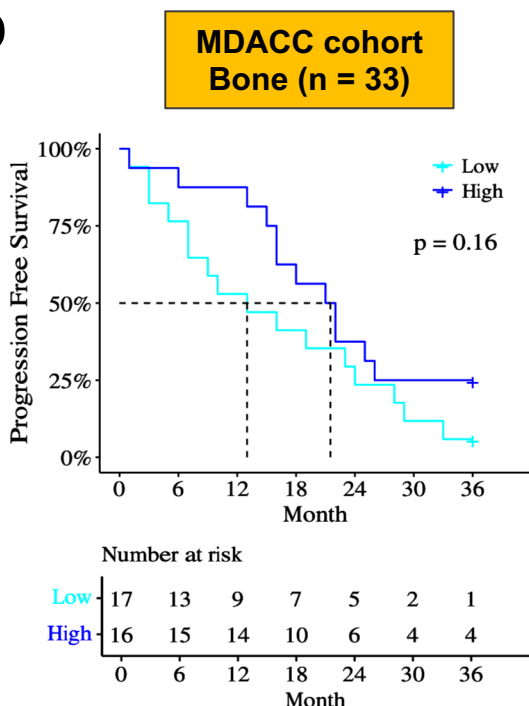
B



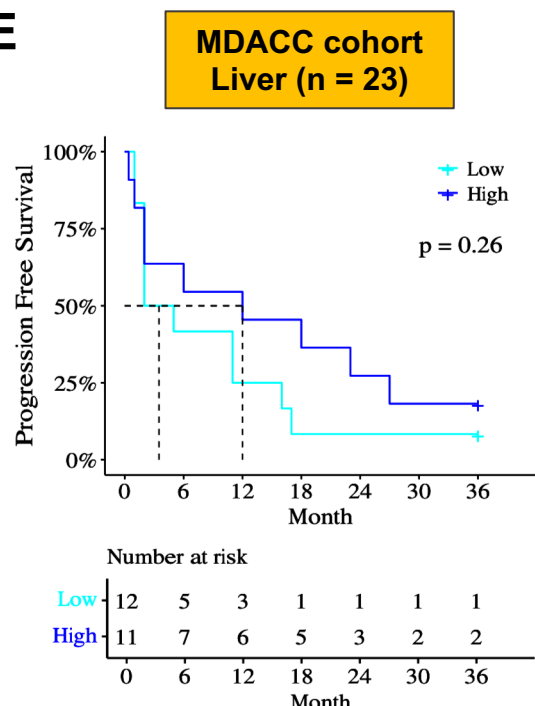
C



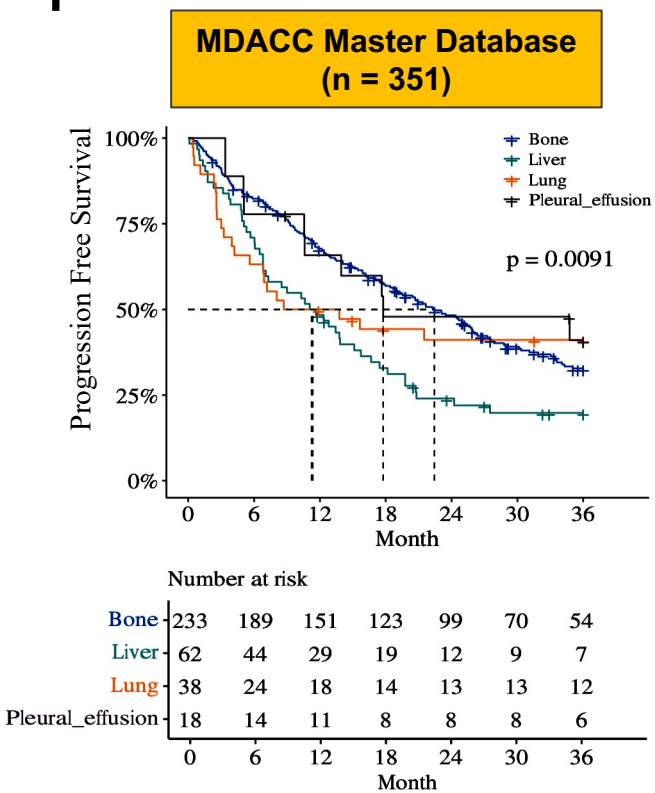
D



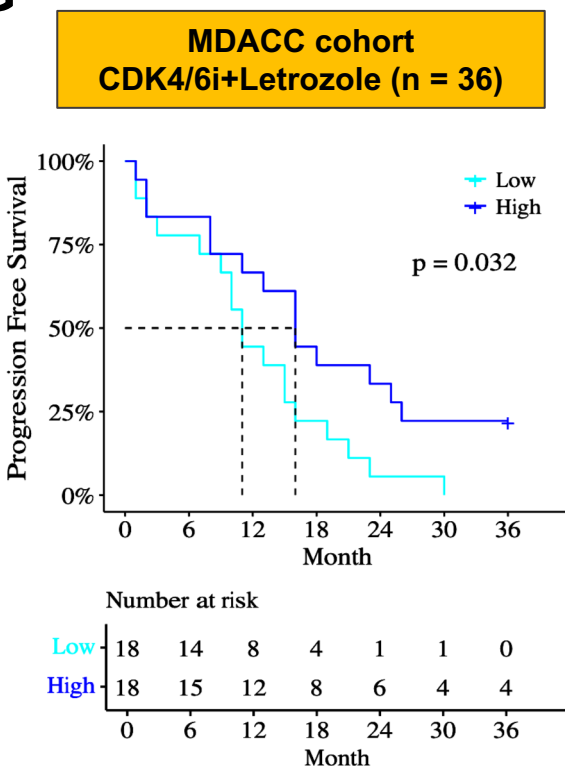
E



F



G



H

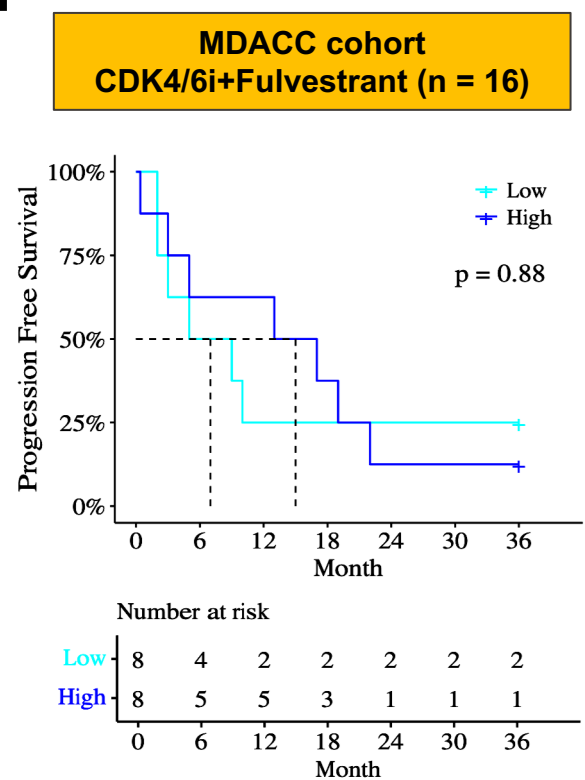


Figure S9. Differential gene expression between CDK4/6i plus letrozole and fulvestrant in EP samples from the FFPE MD Anderson cohort

(A) Heatmap showing the expression of the top 20 differentially expressed genes, stratified by endocrine treatment regimen in EP patients from the FFPE MD Anderson Cohort. CDK4/6i plus letrozole (n = 9) and CDK4/6i plus fulvestrant (n = 7).

Figure S9

A

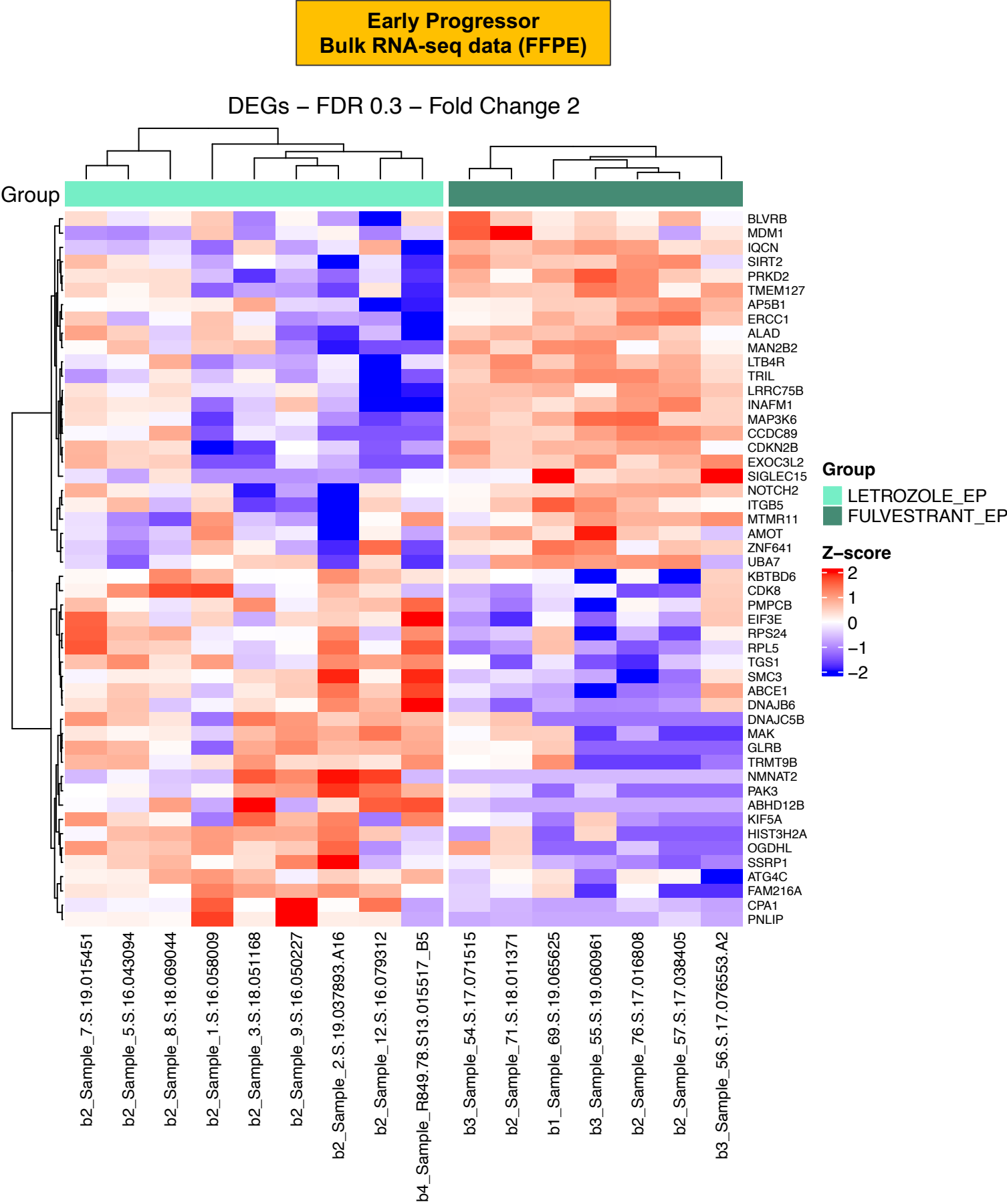
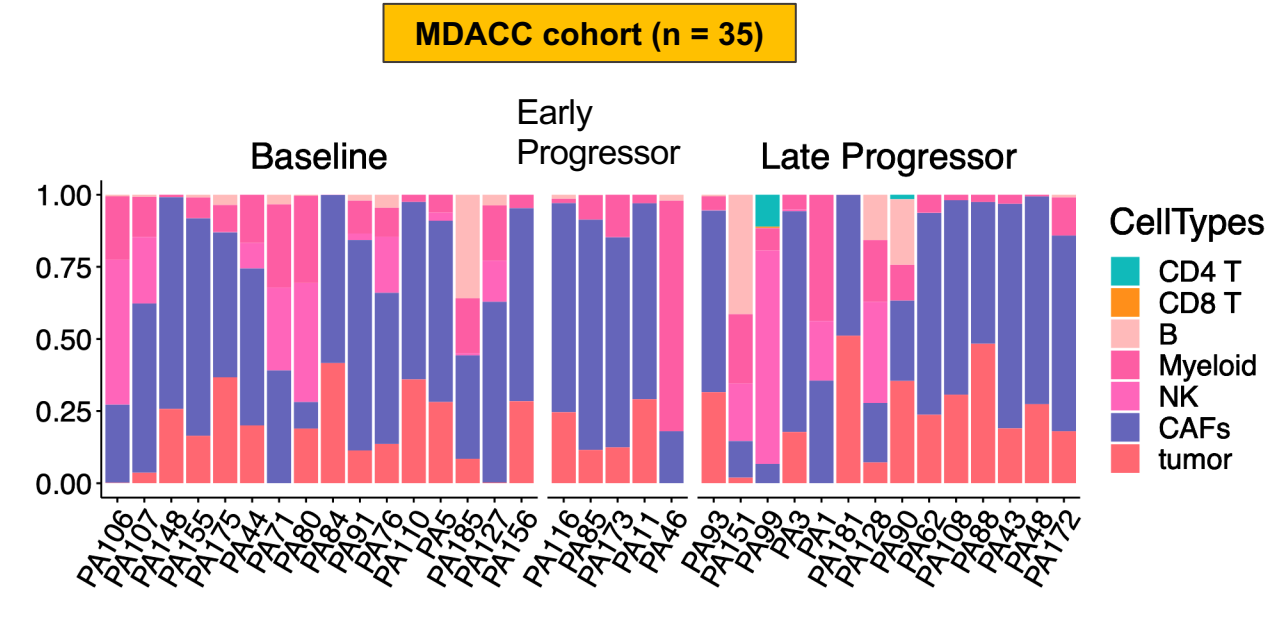


Figure S10. Predictive cell type signature validation

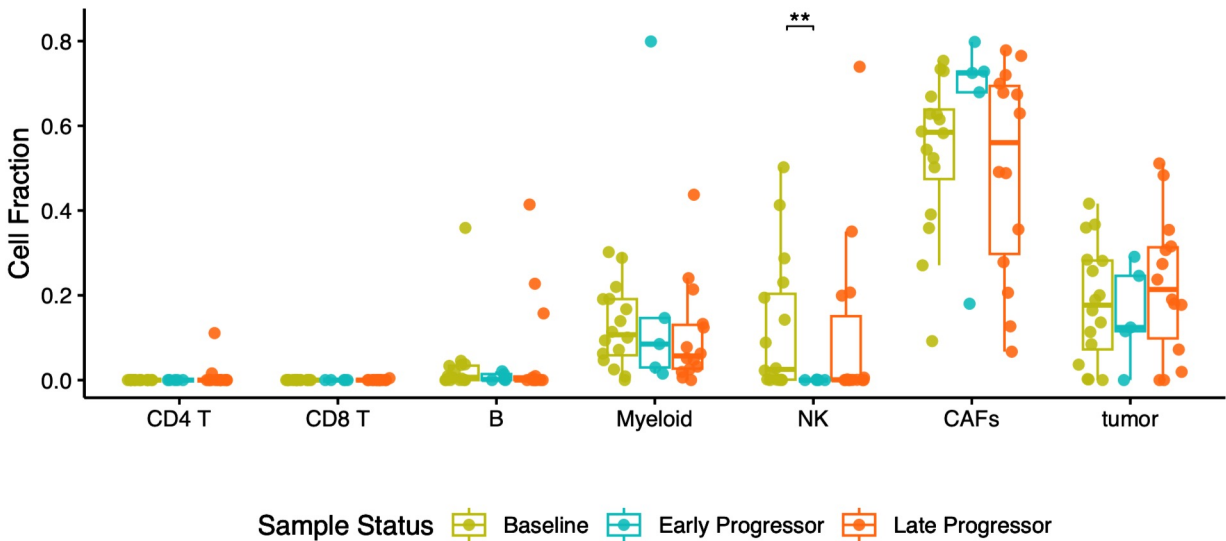
- (A) Bar plot showing the relative fractions of tumor and major non-tumor cell types in samples from the MD Anderson Cohort.
- (B) Cell fraction frequencies of major non-tumor cell types across sample statuses.
- (C) Alluvial plot showing the dynamic changes in major non-tumor cell subtype fractions across longitudinal samples from an EP patient (BRO7F.061) from the Korean cohort.
- (D) Alluvial plot illustrating the dynamic changes in major non-tumor cell subtype fractions across longitudinal samples from an LP patient (BRO7F.021) from the Korean cohort.

Figure S10

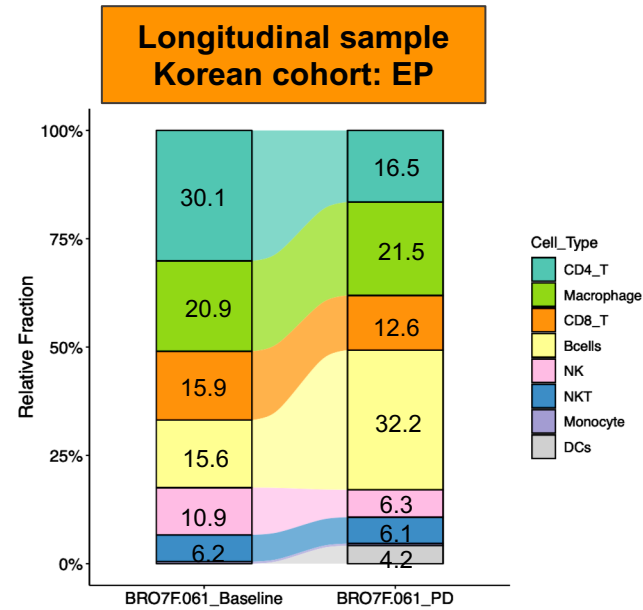
A



B



C



D

