Supplementary Materials

Supplementary Tables 2

Table S1 Survival analysis results of eleven genes in public databases 2

Table S2. Validation results of 11-gene model in public database and in-house cohorts 4

Supplementary Figures and legends 5

Figure S1. Detailed cluster information and related analysis. 5

Figure S2. Functional and survival details of EPCAM+ cells related genes. 6

Figure S3. The T cell subsets and relevant analysis. 7

Figure S4. Functional details of five different Ma/Mo subsets. 8

Figure S5. CIBERSORT and XCELL results of tumor microenvironment in TCGA and GSE datasets. 9

Figure S6. Volcano and bubble plots of functional analysis in GSE and bulk datasets. 10

Figure S7. Results of drug sensitivity analysis. 11

Supplementary Tables

Table S1 Survival analysis results of eleven genes in public databases

Genes	Datasets	Group	Median OS (months)	95%CI of median survival time	HR	95%CI of HR	P
HLA-DPB1	TCGA	Higher	59.1	50.0-87.3	0.65	0.48-0.87	0.0039
		Lower	41.5	37.2-49.9			
	K-M	Higher	22.6		0.36	0.17-0.77	0.0057
		Lower	11.6				
FAM83A	TCGA	Higher	41.0	33.2-53.3	1.71	1.27-2.31	0.00034
		Lower	54.1	49.7-112.0			
	K-M	Higher	60.0		1.33	1.13-1.57	0.00071
		Lower	91.0				
ITGB4	TCGA	Higher	36.0	31.7-48.5	1.72	1.28-2.31	0.00026
		Lower	55.1	50.0-87.3			
	K-M	Higher	53.0		1.49	1.31-1.69	＜0.0001
		Lower	91.0				
OAS1	TCGA	Higher	39.9	34.2-51.0	1.50	1.12-2.01	0.0068
		Lower	54.1	46.0-106.0			
	K-M	Higher	52.2		1.36	1.20-1.55	＜0.0001
		Lower	85.0				
FHL2	TCGA	Higher	39.9	33.2-49.7	1.49	1.11-2.00	0.0072
		Lower	54.4	50.5-89.4			
	K-M	Higher	65.1		1.14	1.00-1.29	0.046
		Lower	74.0				
S100P	TCGA	Higher	40.5	34.8-59.3	1.34	1.00-1.79	0.050
		Lower	50.0	45.2-72.5			
	K-M	Higher	62.0		1.17	1.03-1.32	0.016
		Lower	76.0				
FSCN1	TCGA	Higher	38.9	33.2-46.0	1.63	1.21-2.19	0.0011
		Lower	54.4	50.0-87.3			
	K-M	Higher	52.0		1.51	1.33-1.71	＜0.0001
		Lower	90.0				
SFTPD	TCGA	Higher	54.1	42.2-105.6	0.68	0.51-0.92	0.010
		Lower	41.0	37.2-53.3			
	K-M	Higher	88.7		0.66	0.58-0.75	＜0.0001
		Lower	50.0				
DBH-AS1	TCGA	Higher	59.3	42.9-87.3	0.73	0.54-0.98	0.040
		Lower	41.9	38.2-53.3			
	K-M	Higher	101.5		0.74	0.63-0.88	0.00045

		Lower	63.4				
CST3	TCGA	Higher	51.0	47.4-77.3	0.71	0.53-0.95	0.020
		Lower	41.5	35.8-59.9			
	K-M	Higher	93.1		0.69	0.58-0.81	＜0.0001
		Lower	57.0				
SPP1	TCGA	Higher	42.9	34.9-55.1	1.38	1.03-1.86	0.030
		Lower	53.9	46.0-89.4			
	K-M	Higher	57.0		1.34	1.18-1.52	＜0.0001
		Lower	85.0				

Abbreviations: OS, overall survival; HR, hazard ratio; TCGA, The Cancer Genome Atlas; GSE, Gene Expression Omnibus Series ID; K-M, Kaplan-Meier; 95%CI, 95% confidence interval. P values were determined by log-rank test.

Table S2. Validation results of 11-gene model in public database and in-house cohorts

		Median OS (months)	95%CI of Median OS	HR	95%CI of HR	P
TCGA	Higher	59.3	50.5-89.4	0.61	0.46-0.80	0.00028
	Lower	39.9	33.2-49.7			
GSE8894	Higher	NA	NA	0.27	0.13-0.54	0.00025
	Lower	18.5	7.7-32.6			
GSE31210	Higher	NA	NA	0.25	0.13-0.49	0.00041
	Lower	NA	NA			
GSE68571	Higher	NA	NA	0.24	0.11-0.53	0.0018
	Lower	79.5	28.2-NA			
GSE42127	Higher	NA	81.6-NA	0.45	0.28-0.74	0.0019
	Lower	62.4	48.0-NA			
GSE4573	Higher	80.9	54.3-NA	0.61	0.38-0.98	0.039
	Lower	62.3	26.9-94.5			
GSE37745	Higher	64.6	43.7-108.7	0.63	0.45-0.88	0.0048
	Lower	34.0	25.5-43.1			
GSE30219	Higher	85.0	52.0-118.0	0.67	0.51-0.89	0.0049
	Lower	45.0	28.0-71.0			
GSE50081	Higher	NA	71.5-NA	0.62	0.39-0.98	0.040
	Lower	67.2	49.3-NA			
NCC-Bulk	Higher	NA	NA	0.41	0.22-0.76	0.0074
	Lower	NA	NA			
NCC-ICI (n=40 with OS data)	Higher	NA (OS)	24.3-NA	0.37	0.16-0.88	0.025
	Lower	18.3 (OS)	8.8-NA			
NCC-ICI (n=29 with PFS data)	Higher	5.6 (PFS)	1.8-25.2	0.62	0.29-1.34	0.18
	Lower	3.4 (PFS)	1.9-15.3			
NCC-ICI-LUAD (n=22 with OS data)	Higher	NA (OS)	21.2-NA	0.36	0.11-1.12	0.073
	Lower	28.9 (OS)	7.5-NA			
NCC-ICI-LUAD (n=16 with PFS data)	Higher	17.9(PFS)	5.9-NA	0.35	0.11-1.08	0.021
	Lower	2.0(PFS)	1.8-NA			

Abbreviations: OS, overall survival; PFS, progression-free survival; HR,hazard ratio; NA, not applicable; TCGA, The Cancer Genome Atlas; GSE, Gene Expresion Omnibus Series ID; CI, confidence interval; NCC, National Cancer Center; ICI, Immune checkpoint inhibitor. P values were determined by log-rank test.

Supplementary Figures and legends

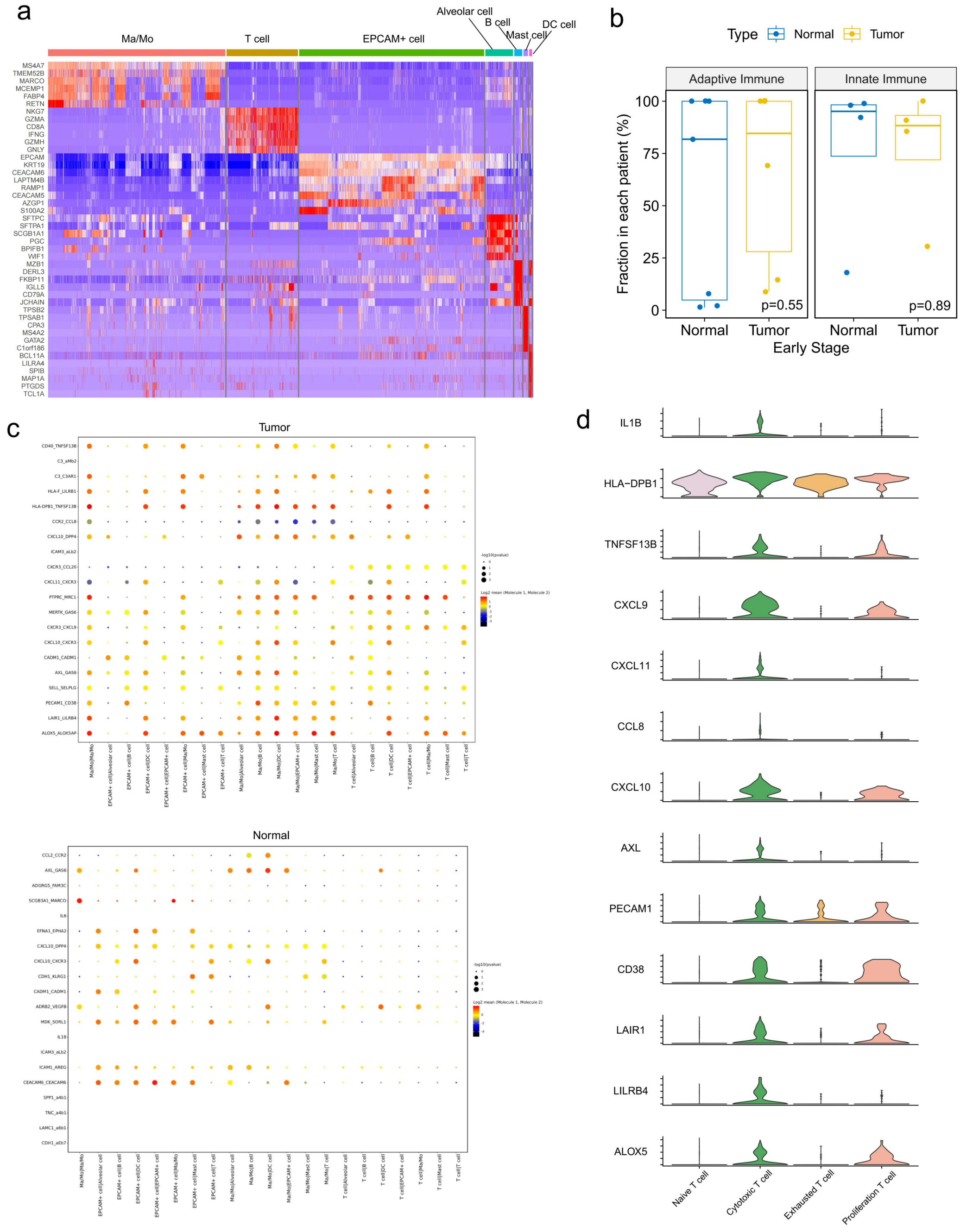

Figure S1. Detailed cluster information and related analysis.

(a) The heatmap of gene expression level grouped by different cell types.

(b) Different fraction of adaptive/innate immune cells between adenocarcinoma and normal samples.

(c) Differentially expressed ligands/receptors in tumor and normal tissues.

(d) Important receptor - ligand molecules and their expression levels between different T cells.

Abbreviations: EPCAM, epithelial cell adhesion molecule; Ma/Mo, macrophage and monocyte.

We used Wilcoxon rank-sum test in different fraction analysis.

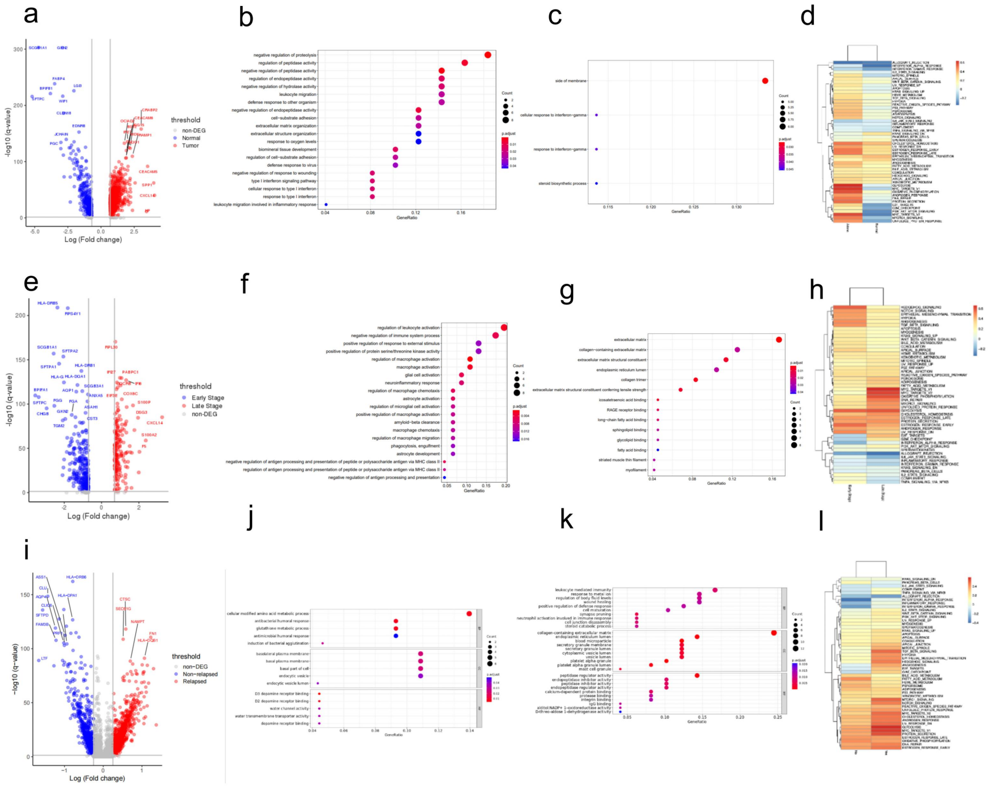

Figure S2. Functional and survival details of EPCAM+ cells related genes.

(a) Volcano map of differential genes between EPCAM+ cells in tumor and normal tissues.

(b) Functional analysis of tumor-related DEGs.

(c) Functional analysis of normal tissue-related DEGs.

(d) Enrichment analysis of DEGs between EPCAM+ cells in tumor vs normal tissues.

(e) Volcano map of differential genes between EPCAM+ cells in patients with early and late stage.

(f) Functional analysis of early stage-related DEGs.

(g) Functional analysis of late stage-related DEGs.

(h) Enrichment analysis of DEGs between EPCAM+ cells in patients with early vs late stage.

(i) Volcano map of differential genes between EPCAM+ cells in relapsed and non-relapsed patients.

(j) Functional analysis of relapse-related DEGs.

(k) Functional analysis of non-relapse-related DEGs.

(l) Enrichment analysis of DEGs between EPCAM+ cells in relapsed vs. non-relapsed patients.

Abbreviations: EPCAM, epithelial cell adhesion molecule; DEG, differentially-expressed gene.

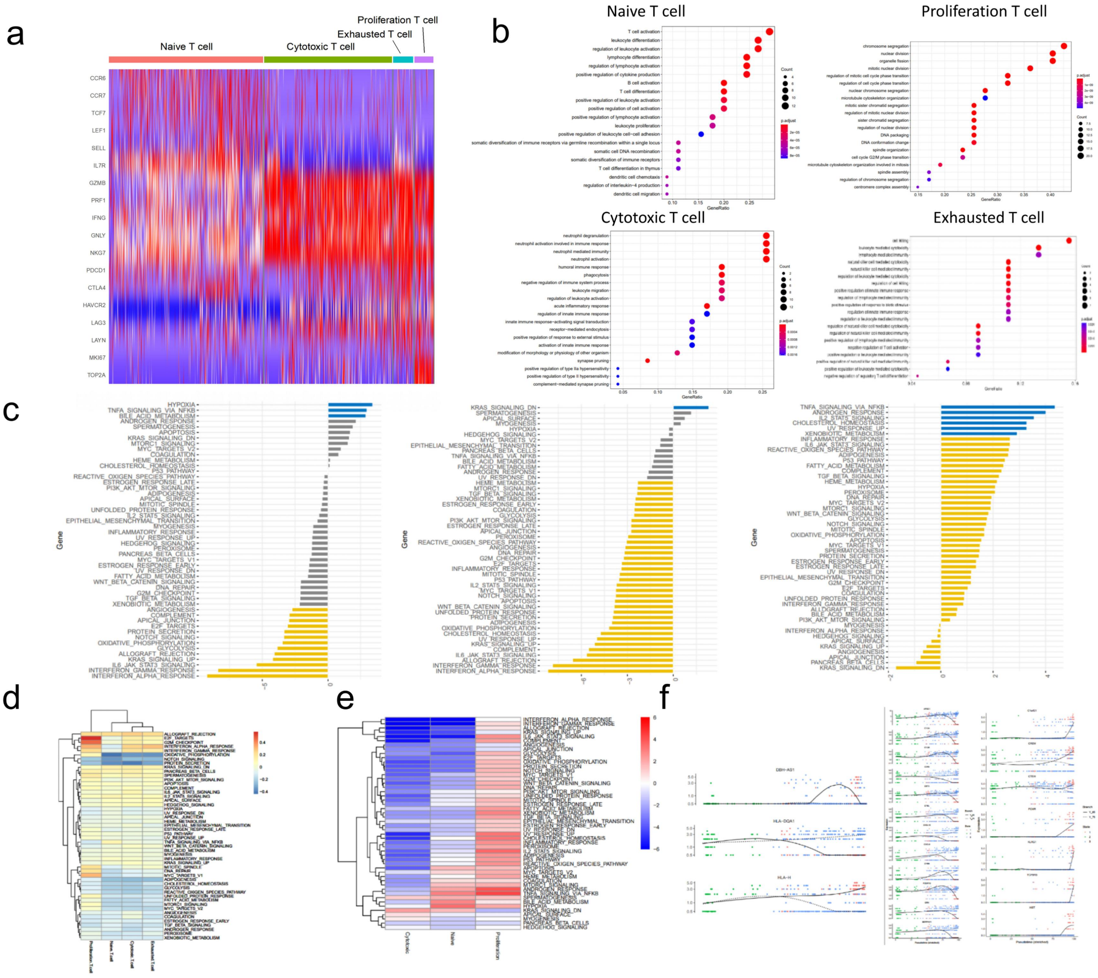

Figure S3. The T cell subsets and relevant analysis.

(a) Heatmap of key genes grouped by four T cell subsets.

(b) Functional analysis of four T cell subsets.

(c) The functions of naïve T cells, toxic T cells and proliferative T cells between tumor tissues and normal tissues.

(d) Hallmark function enrichment analysis of four T cell subsets.

(e) Functional analysis of naïve T cells, toxic T cells and proliferative T cells between tumor tissues and normal tissues.

(f) Identification of gene and T cell differentiation trajectory.

Abbreviations: DEG, differentially expressed genes; HR, hazard ratio.

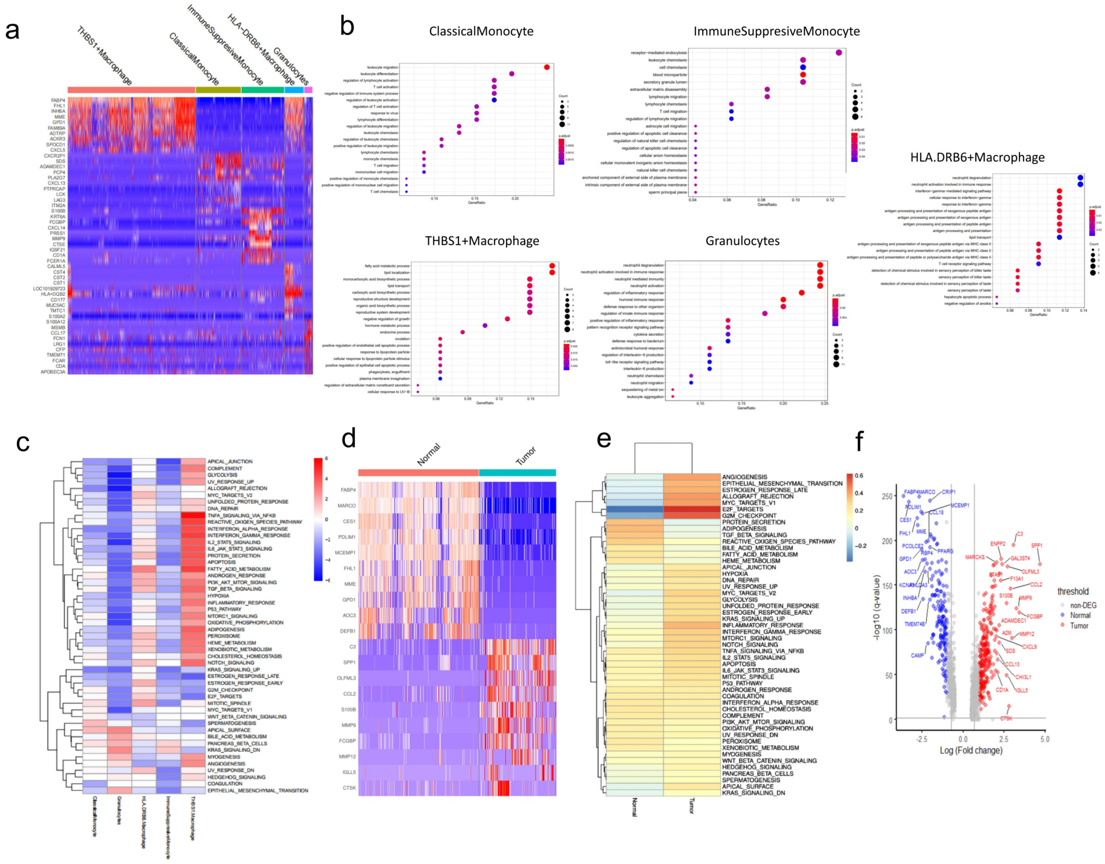

Figure S4. Functional details of five different Ma/Mo subsets.

(a) Heatmap of key genes grouped by different Ma/Mo cell subsets

(b) Bubble plots of enrichment analysis of DEGs between tumor and normal cells in five different Ma/Mo groups.

(c) Hallmark pathways analysis of different Ma/Mo subsets.

(d) Heatmap shows differences in the expression of Ma/Mo-related genes between tumor and normal cells.

(e) Hallmark functional analysis of differentially-expressed genes on Ma/Mo cells between tumor tissues and normal tissues.

(f) Volcano plot of Ma/Mo related DEGs in tumor and normal cells.

Abbreviations: DEG, differentially expressed genes; Ma/Mo, macrophage and monocyte.

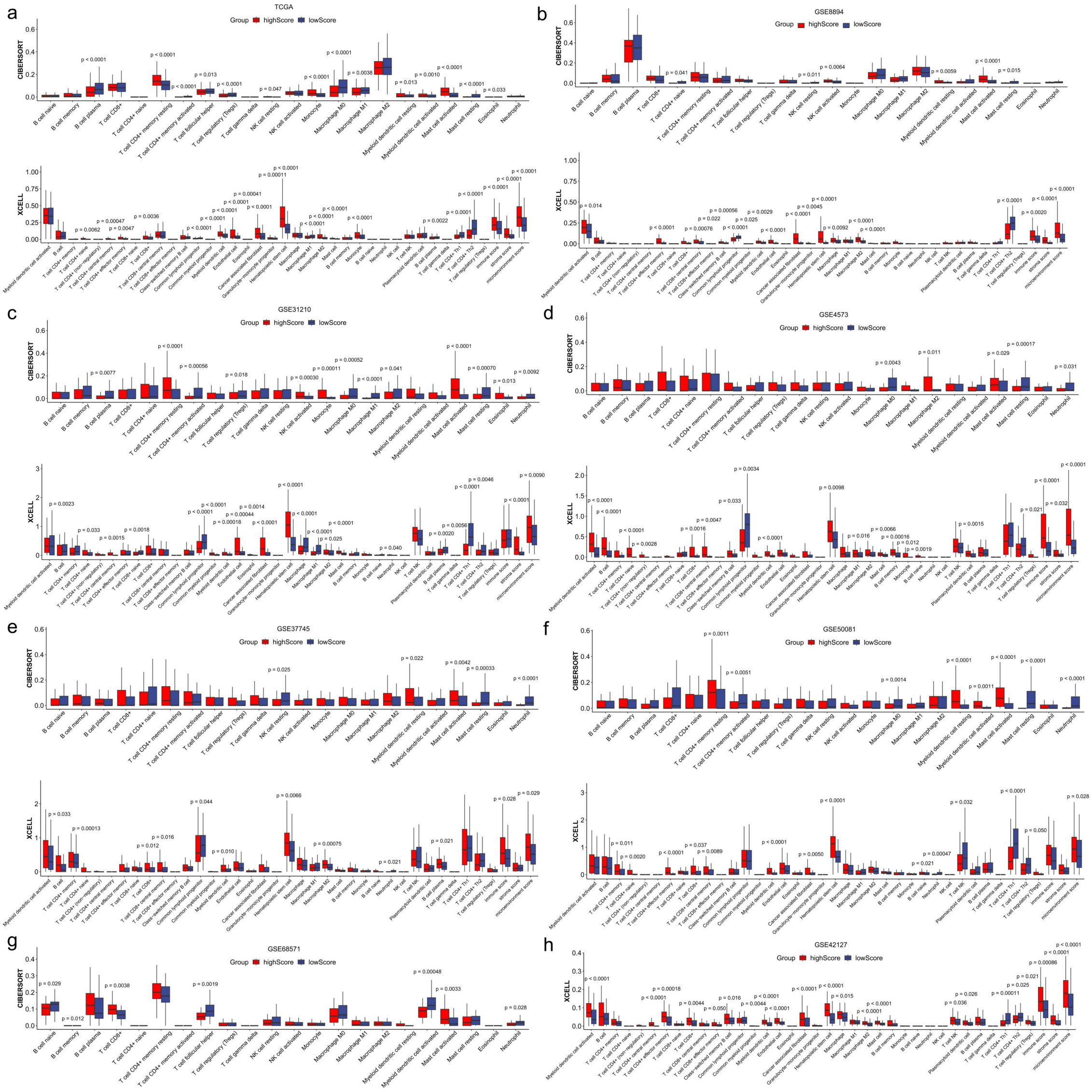

Figure S5. CIBERSORT and XCELL results of tumor microenvironment in TCGA and GSE datasets.

(a) Validation of tumor microenvironment in TCGA.
(b) Validation of tumor microenvironment in GSE8894.
(c) Validation of tumor microenvironment in GSE31210.
(d) Validation of tumor microenvironment in GSE4573.
(e) Validation of tumor microenvironment in GSE37745.
(f) Validation of tumor microenvironment in GSE50081.
(g) Validation of tumor microenvironment in GSE68571.
(h) Validation of tumor microenvironment in GSE42127.

p values were determined by Wilcoxon rank-sum test.

Abbreviations: GSE, Gene Expression Omnibus Series ID; TCGA, The Cancer Genome Atlas.

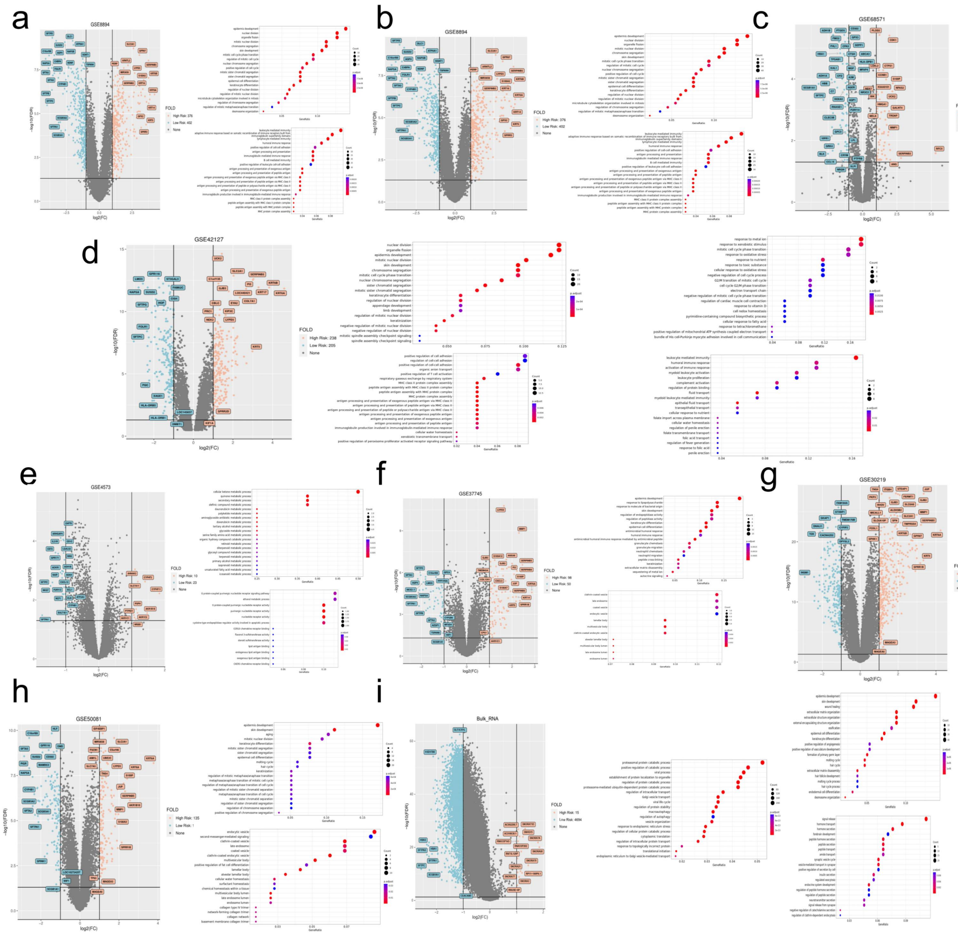

Figure S6. Volcano and bubble plots of functional analysis in GSE and bulk datasets.

(a) Functional analysis in GSE8894.

(b) Functional analysis in GSE31210.

(c) Functional analysis in GSE68571.

(d) Functional analysis in GSE42127.

(e) Functional analysis in GSE4573.

(f) Functional analysis in GSE37745.

(g) Functional analysis in GSE30219.

(h) Functional analysis in GSE50081.

(i) Functional analysis in NCC-bulk dataset.

Abbreviations: FC, fold change; FDR, False Discovery Rate; GSE, Gene Expression Omnibus Series ID, TCGA, The Cancer Genome Atlas; NCC, National Cancer Center.

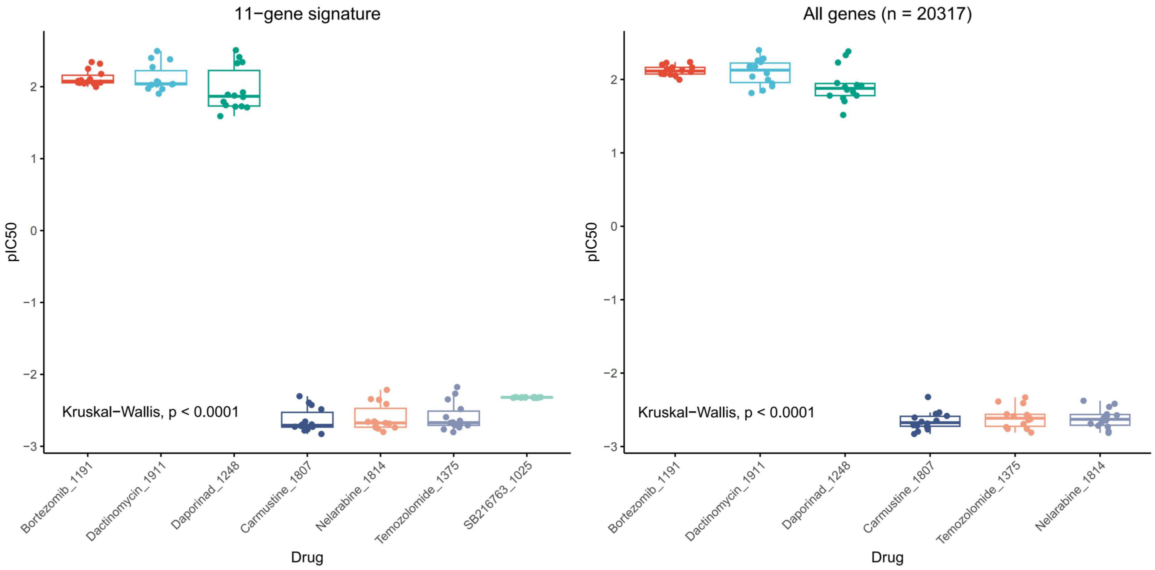

Figure S7. Results of drug sensitivity analysis.
The drug sensitivity analysis test of patients showed that there were significant statistical differences among different drug sensitivities, similar between 11-gene model (left) and all RNA profiles (right). P values were determined by Kruskal-Wallis test.