Supplementary Materials for

Molecular classification of geriatric breast cancer displays distinct senescent subgroups of prognostic significance

Xia Wu *et al*.

* Corresponding authors: Hao Ke at kehao@ncu.edu.cn;
Limin Zhao at zhaolimin@ncu.edu.cn

This PDF file includes:

Figures. S1 to S11

Other Supplementary Materials for this manuscript include the following:

Table S1 to S5

Figure S1. The analysis of 25 genes related to senescence in breast cancer.

(A) The network analysis was performed to reveal the interactions among these 25 genes in elderly breast cancer. Genes were grouped into four clusters (hcluster) based on their expression, with red lines indicate positive correlations and blue lines indicate negative correlations. The size of the circles reflects the *P* value representing the impact of the genes on patient survival. Two genes, HSPA8 and UBC, clustered together, while the remaining genes were distributed among three other clusters. (B) Displaying the expression levels of the 25 genes in both elderly breast cancer and normal samples. *P* value were determined using Student's *t*-test, with blue representing normal sample expression and red representing tumor sample expression. (C) Analysis of single nucleotide variations (SNVs) in the 25 senescence genes among 445 elderly breast cancer patients. (D) Co-occurrence analysis of genetic alterations in the 25 senescence genes among 445 elderly breast cancer patients. There are significant co-mutations between MDM2 and MAPKAPK5, as well as HMGA2 and CCNA2. (E) Univariate Cox analysis of the expression of the 25 senescence genes. HSPA8 and E2F2 were negatively correlated with patient survival, with a hazard ratio greater than 1 and statistical significance. *$P < 0.01$, ** $P < 0.001$, *** $P < 0.0001$, **** $P < 0.00001$.

Figure S2. Unsupervised clustering analysis displays two distinct senescent clusters in elderly breast cancer.

(A) The ConsensusClusterPlus R package was used to calculate the optimal cutoff values for patients. Consistency Matrix (CM) plots values from white to blue on a color

scale for different K values. (B-H) Consensus matrix heatmaps represent sample matrices where values range from 0 to 1, indicating the likelihood of belonging to the same cluster, with colors from white to dark blue. (I) Consensus cumulative distribution function (CDF) plots use histograms with 100 bins to calculate cumulative distributions for each K corresponding to the consensus matrix. CDF plots help determine the optimal K value. (J) Delta area plot compares the relative changes in the area under the curve of the CDF between K and K-1. Points with minimal increase were selected as the optimal K value; in this case, K=2 was chosen. (K) Tracking plots show samples (rows) across different K values (columns) using heatmaps, allowing qualitative assessment of unstable clustering and sample stability. (L) PCA reduction of 25 senescence genes in 513 elderly breast cancer patients. Red represents Cluster1, and blue represents Cluster2.

Figure S3. Two senescent clusters show different gene enrichment pathways.

(A) Univariate Cox analysis of patient information (Gender, Stage, ER, PR, HER-2, and Cluster) (B) Multivariate Cox analysis of patient information (Gender, Stage, ER, PR, HER-2, and Cluster). (D-E) Illustrate the differential Gene Set Enrichment Analysis (GSEA) between Cluster1 and Cluster2.

Figure S4. Construction of the Sene_Signature model.

(A) Model Comparison: We conducted a comprehensive evaluation of machine learning models using accuracy parameters such as AUC, MCC, F1 score, recall, and

Positive predictive value (PPV). (B) Discrepancies in survival outcomes between the senescence score model and single genes, risk scores, and biological pathway enrichment scores.

Figure S5. Sene_Signature score associated with clinical features.

(A-B) Lollipop chart showing the correlation between 25 senescence genes and senescence scores in the TCGA (A) and METRBIC (B) data. The smaller the *P* value, the redder the color, and the size of the circle maps to the size of the correlation. (C) The Recurrence-Free Survival differences between the high and low evaluation groups in the GSE2990 dataset. (D) The sample information table is represented by the Sankey plot, which corresponds to cluster information, high and low scores, PAM50 typing information, and TMN tumor staging information. (E) The box chart of Sene_Signature difference between triple negative breast cancer and non-triple-negative breast cancer shows that there is significant difference between the two groups through Wilcoxon-test.

Figure S6. The correlation between Sene_Signature and genomic mutations

(A) The box plot of TMB scores in the high and low scoring groups in METABRIC shows that red represents the high scoring group and blue represents the low scoring group (Student's *t*-test). (B-C) Mutation information is provided for the top 20 genes with the highest number of mutations in high and low scoring patients in TCGA (B) and METABRIC (C).

Figure S7. Identification of characteristic genes for the single-cell transcriptomic data.

(A) Using the "FindAllMarkers" function to identify characteristic genes within 6 cell types from the single-cell transcriptomic data of 70 breast cancer patients. Columns represent cell types, rows represent genes, and the heatmap displays gene expression within each cell type.

Figure S8. The relationship between Sene_Signature and tumor immune infiltration from scRNA-seq data.

(A) Integration of 70 single-cell transcriptomic datasets, where different colors represent distinct patients. From the visual results, batch effects appear to have been removed. (B) UMAP dimensionality reduction results for the subdivision of T-cell subtypes, grouped into 12 clusters. (C) Mapping patient information onto the UMAP results for T-cell subtypes, with each color representing a different patient. No noticeable batch effects are observed in the data.

Figure S9. Drug sensitivity analysis for 25 senescence-related genes.

(A) The correlation between 25 senescence genes and drug IC50 was analyzed in the GDSC database, and the corresponding drug names were listed for the 25 genes. Red indicates that higher gene expression is associated with lower drug sensitivity, while blue indicates that higher gene expression is associated with greater drug sensitivity.

Addition, larger circles represent more significant *P* value.

Figure S10. Sene_Signature predicts drug sensitivity of elderly breast cancer.

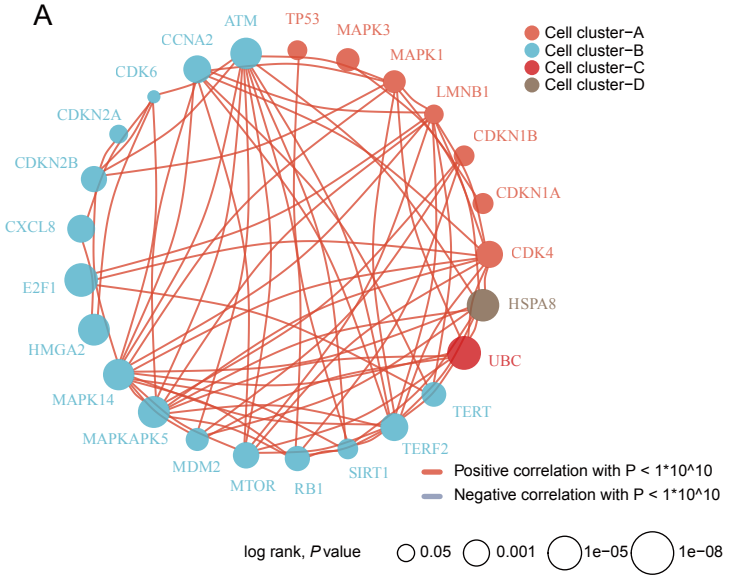
(A-B) Differential analysis of IC50 values between high and low Sene_Signature groups in elderly breast cancer patients from the TCGA (A) and METABRIC (B) datasets. (C-D) Correlation analysis was conducted between drug sensitivity predictions and Sene_Signature scores in the TCGA (C) and METABRIC (D) datasets.

FigureS11. EGFR inhibitor sensitivity analysis in high and low Sene_Signature groups.

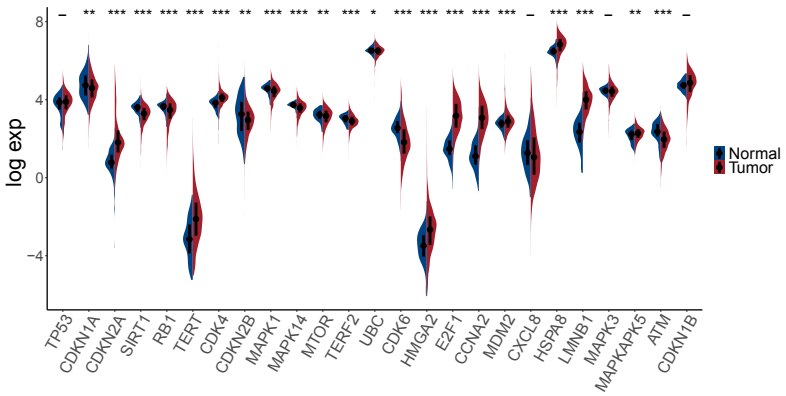
(A) Eight breast cancer patients in GSE33658 predicted the high and low scores through the senescence model, and checked the patients' drug response to EGFR inhibitor Gefitinib.

Figure S1

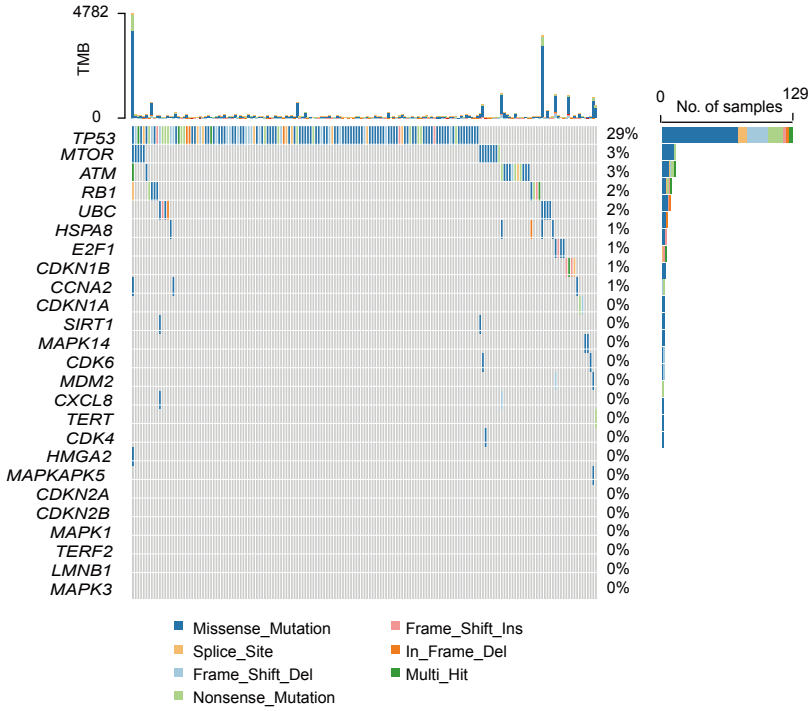
A



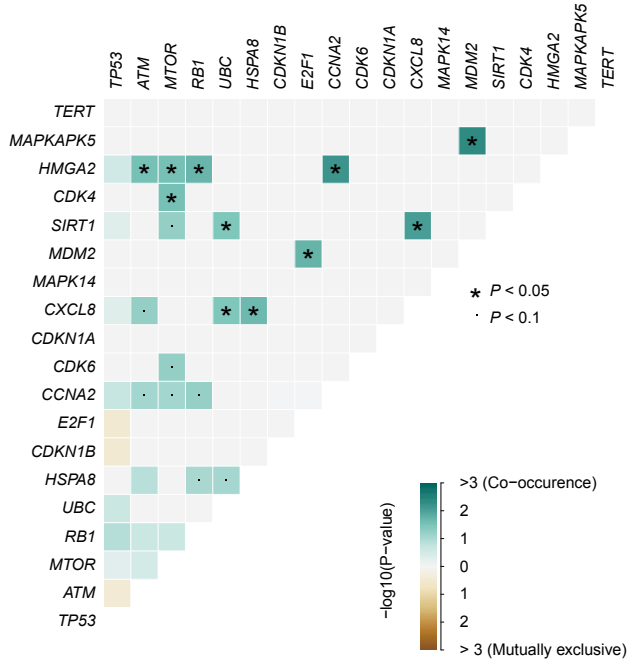
B



C



D



E

	P value	HR
TERT	0.967	0.987
CDKN2B	0.168	0.987
CDK6	0.658	0.994
CDKN2A	0.486	0.996
SIRT1	0.728	0.996
MAPK3	0.290	0.997
CDKN1B	0.270	0.998
TP53	0.987	1.000
CDKN1A	0.333	1.001
UBC	0.168	1.001
HSPA8	0.003	1.001
LMNB1	0.262	1.003
CDK4	0.125	1.004
MAPK1	0.204	1.004
CCNA2	0.367	1.004
CXCL8	0.654	1.004
RB1	0.276	1.007
MAPK14	0.351	1.008
MTOR	0.462	1.009
E2F1	0.028	1.010
MDM2	0.227	1.011
ATM	0.273	1.026
MAPKAPK5	0.374	1.029
TERF2	0.082	1.030
HMGA2	0.321	1.058

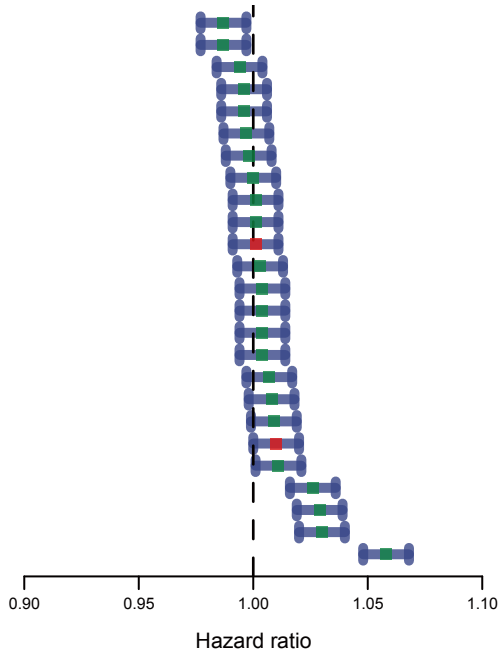
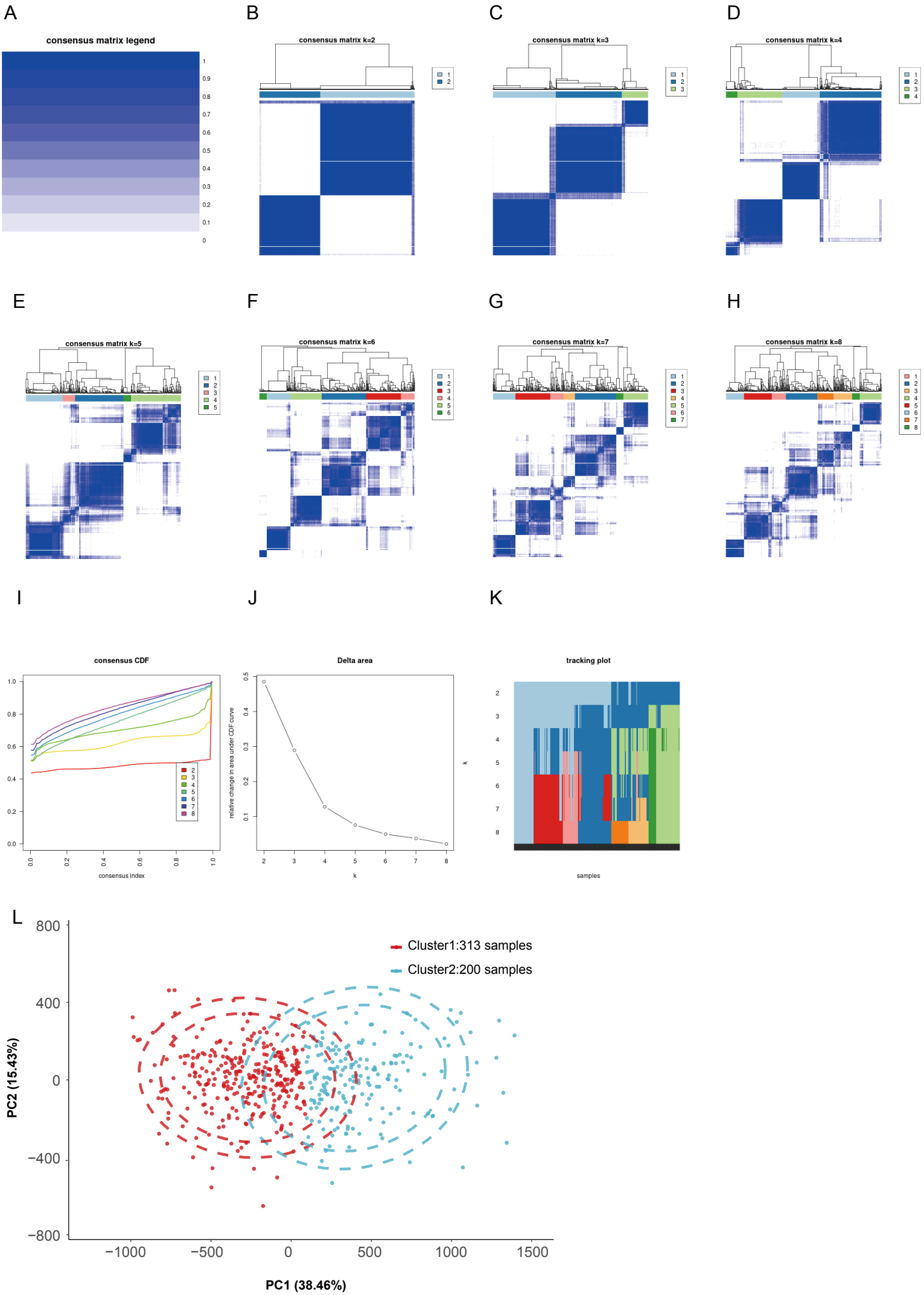
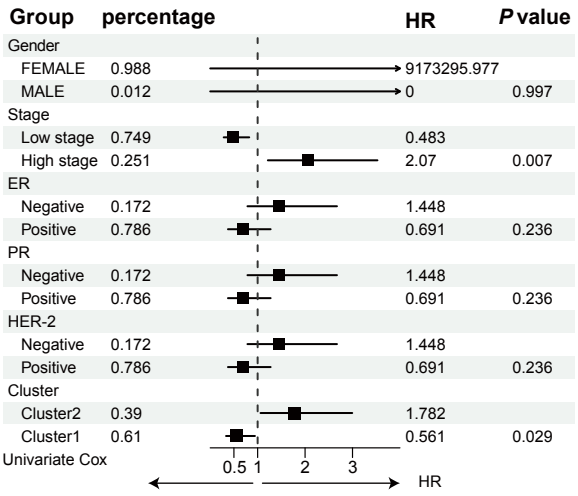


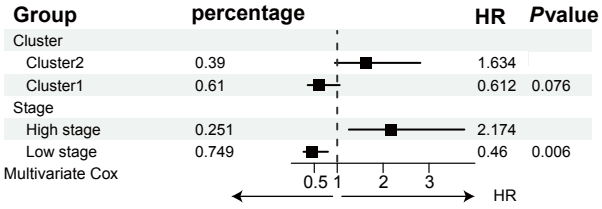
Figure S2


A
consensus matrix legend
B
consensus matrix k=2
C
consensus matrix k=3
D
consensus matrix k=4
E
consensus matrix k=5
F
consensus matrix k=6
G
consensus matrix k=7
H
consensus matrix k=8
I
consensus CDF
CDF
consensus index
J
Delta area
relative change in area under CDF curve
k
K
tracking plot
samples
L
Cluster1:313 samples
Cluster2:200 samples
PC2 (15.43%)
PC1 (38.46%)

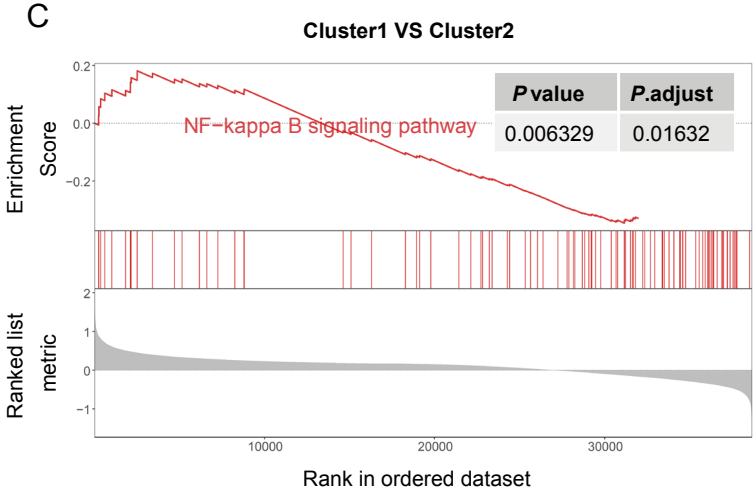
A


Group
percentage
HR
P value
Gender
FEMALE 0.988 9173295.977
MALE 0.012 0 0.997
Stage
Low stage 0.749 0.483
High stage 0.251 2.07 0.007
ER
Negative 0.172 1.448
Positive 0.786 0.691 0.236
PR
Negative 0.172 1.448
Positive 0.786 0.691 0.236
HER-2
Negative 0.172 1.448
Positive 0.786 0.691 0.236
Cluster
Cluster2 0.39 1.782
Cluster1 0.61 0.561 0.029
Univariate Cox
0.5 1 2 3
HR

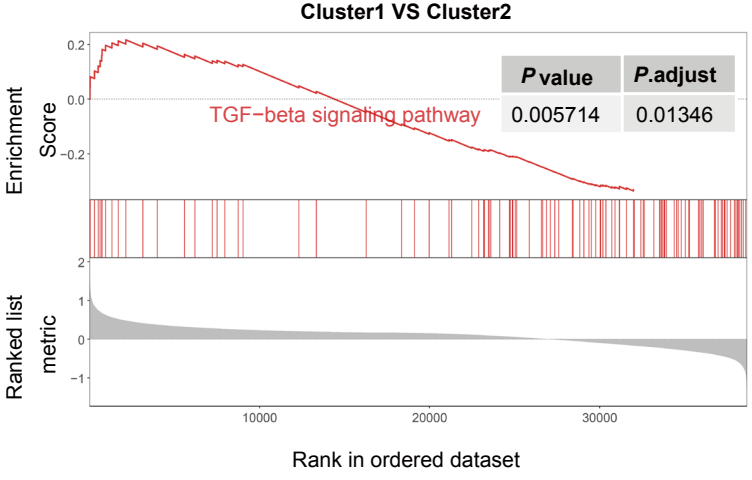
B


Group
percentage
HR
Pvalue
Cluster
Cluster2 0.39 1.634
Cluster1 0.61 0.612 0.076
Stage
High stage 0.251 2.174
Low stage 0.749 0.46 0.006
Multivariate Cox
0.5 1 2 3
HR

C


Cluster1 VS Cluster2
P value
P.adjust
NF−kappa B signaling pathway
0.006329
0.01632
Enrichment Score
Ranked list metric
Rank in ordered dataset

D


Cluster1 VS Cluster2
P value
P.adjust
TGF−beta signaling pathway
0.005714
0.01346
Enrichment Score
Ranked list metric
Rank in ordered dataset

E

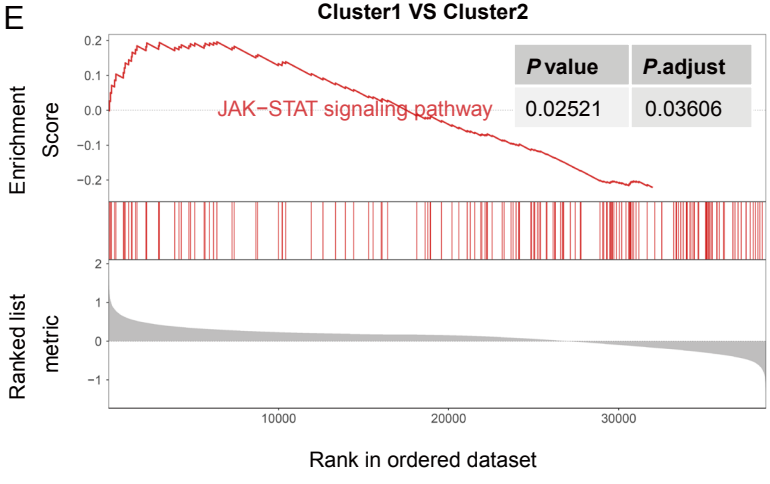

Cluster1 VS Cluster2
P value
P.adjust
JAK−STAT signaling pathway
0.02521
0.03606
Enrichment Score
Ranked list metric
Rank in ordered dataset

Figure S4

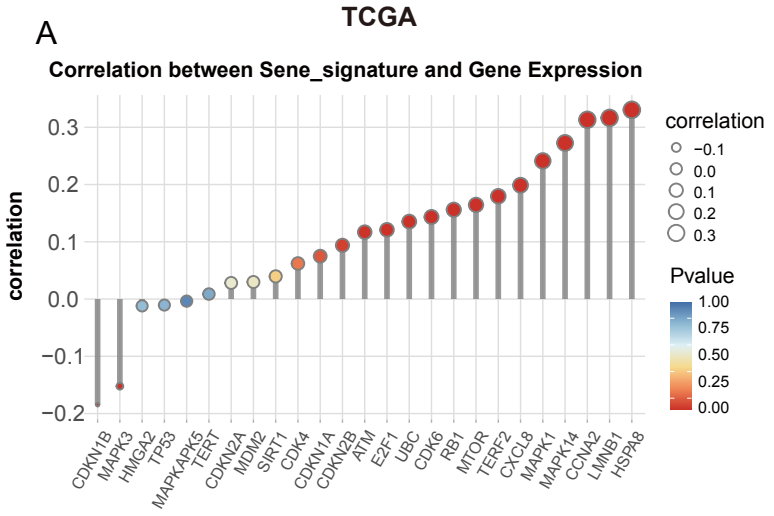
A

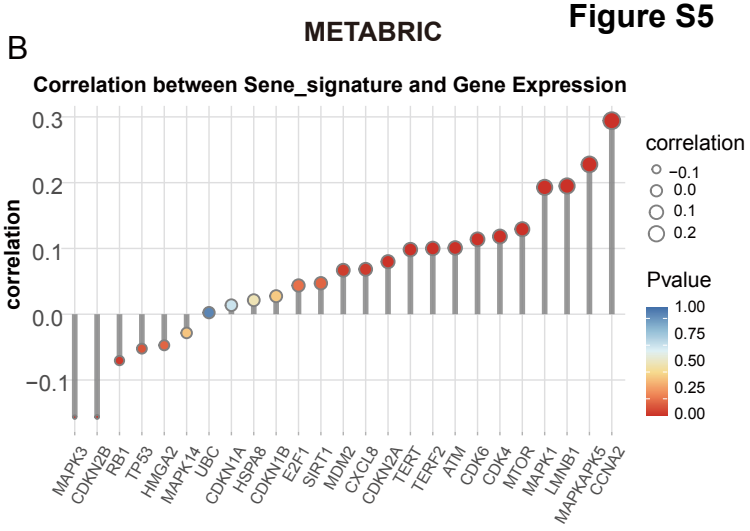
Method	AUC	MCC	F1Score	Recall	PPV
OneClassDetection	0.65330	0.2170231	0.5784314	0.7375	0.4758065
GradientBoostingRegressor	0.63555	0.2161001	0.5573770	0.6375	0.4951456
DecisionTreeRegressor	0.60975	0.2150721	0.5433526	0.5875	0.5053763
KNeighborsRegressor	0.68885	0.2100731	0.5512821	0.5375	0.5657895
LinearRegressor	0.61940	0.1896500	0.5502646	0.6500	0.4770642
SupportVectorRegression	0.73080	0.1705952	0.3589744	0.2625	0.5675676

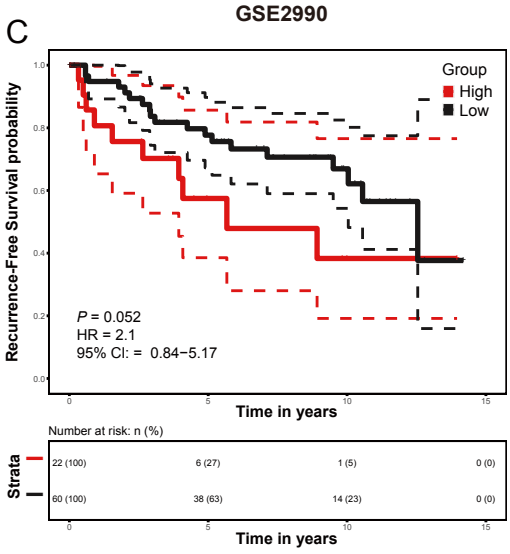
B

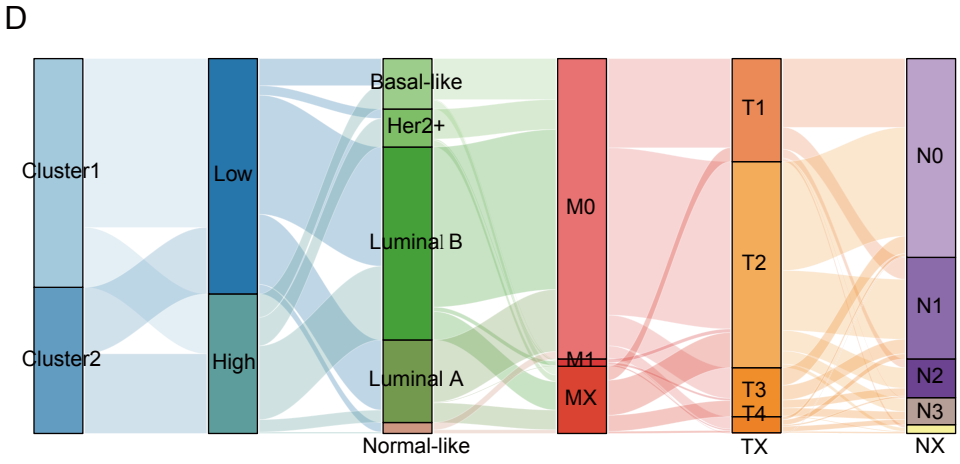
Models	HR	*P* value	Low	Upper
Sene_Signature	5.783	0.017	1.375	24.317
SERPINE1.Sig	1.001	0.382	0.999	1.002
ASF1B.Sig	1.004	0.033	1.000	1.008
ORC6.Sig	1.060	0.004	1.019	1.103
SKP2.Sig	1.034	0.250	0.977	1.095
8_DNA_Repair.Sig	1.273	0.000	1.144	1.415
13_Epigenetic.Sig	1.003	0.283	0.997	1.009
Immune_Score	1.342	0.565	0.493	3.654
CD8 T effector	0.891	0.393	0.684	1.161
APM	1.082	0.676	0.749	1.563
Immune checkpoint	1.040	0.756	0.814	1.327
EMT	1.238	0.158	0.920	1.665
Pan-F-TBRS	0.861	0.092	0.723	1.025
Angiogenesis	0.858	0.218	0.672	1.095
Cell cycle	0.939	0.636	0.724	1.218
DNA replication	1.153	0.222	0.917	1.449
NER	1.151	0.353	0.855	1.549
DNA damage repair	0.923	0.558	0.706	1.207
Homologous recombination	1.281	0.059	0.990	1.656
Mismatch repair	1.202	0.133	0.945	1.528
WNT target	0.871	0.291	0.674	1.125
TMEscore	1.208	0.086	0.974	1.498
Courtois-Cox et al.	1.765	0.045	1.012	3.077
de Magalhães et al.	1.220	0.667	0.494	3.013
Fridman et al.	0.339	0.052	0.114	1.007
GOBP_AGING	0.529	0.409	0.116	2.404
Regulski et al.	1.099	0.878	0.331	3.653
CellAge	1.200	0.793	0.308	4.678
HumanAge	1.334	0.657	0.374	4.755
Kyng et al.	0.921	0.862	0.365	2.325
Lee et al.	0.662	0.546	0.173	2.529
Lu et al.	0.938	0.905	0.329	2.677
Rodwell et al.	0.554	0.272	0.194	1.588
Monet et al.	2.309	0.009	1.231	4.332
Visala et al.	1.085	0.865	0.427	2.757

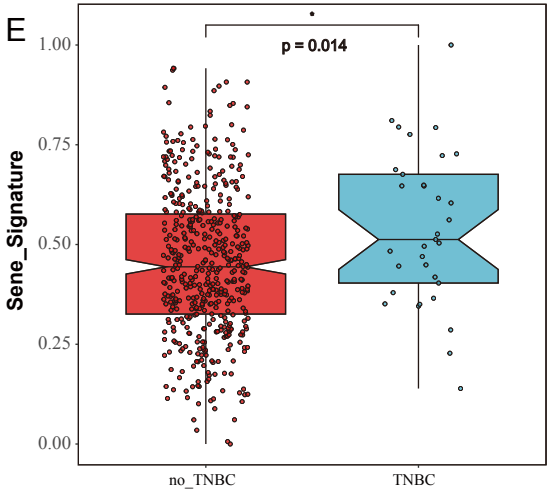
Figure S5


A
TCGA
Correlation between Sene_signature and Gene Expression
correlation
0.3
0.2
0.1
0.0
−0.1
−0.2
CDKN1B MAPK3 HMGA2 TP53 MAPKAPK5 TERT CDKN2A MDM2 SIRT1 CDK4 CDKN1A CDKN2B ATM E2F1 UBC CDK6 RB1 MTOR TERF2 CXCL8 MAPK1 MAPK14 CCNA2 LMNB1 HSPA8
correlation
−0.1
0.0
0.1
0.2
0.3
Pvalue
1.00
0.75
0.50
0.25
0.00

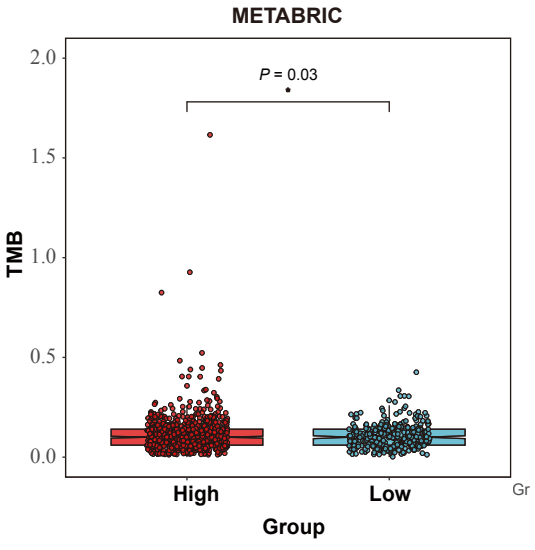

B
METABRIC
Correlation between Sene_signature and Gene Expression
correlation
0.3
0.2
0.1
0.0
−0.1
MAPK3 CDKN2B RB1 TP53 HMGA2 MAPK14 UBC CDKN1A HSPA8 CDKN1B E2F1 SIRT1 MDM2 CXCL8 CDKN2A TERT TERF2 ATM CDK6 CDK4 MTOR MAPK1 LMNB1 MAPKAPK5 CCNA2
correlation
−0.1
0.0
0.1
0.2
Pvalue
1.00
0.75
0.50
0.25
0.00


C
GSE2990
Recurrence-Free Survival probability
Group
High
Low
P = 0.052
HR = 2.1
95% CI: = 0.84−5.17
Time in years
Number at risk: n (%)
Strata
22 (100)
6 (27)
1 (5)
0 (0)
60 (100)
38 (63)
14 (23)
0 (0)
Time in years

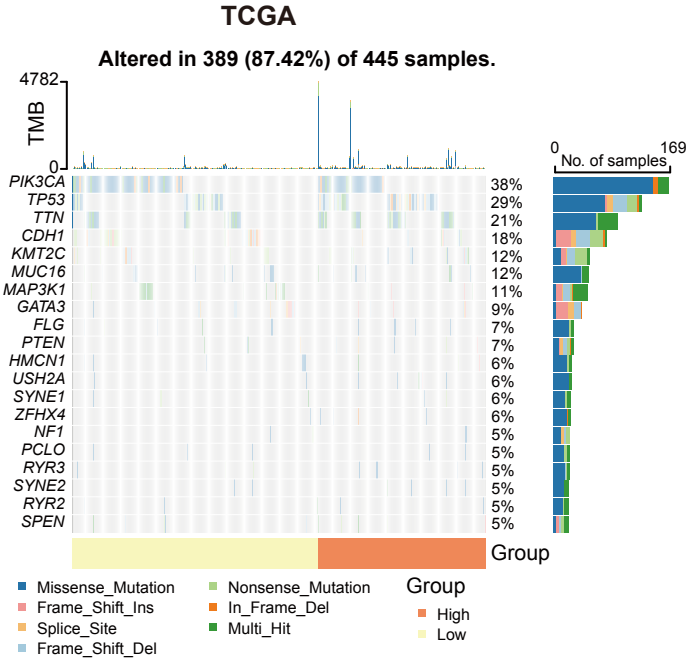

D
Cluster1
Cluster2
Low
High
Basal-like
Her2+
Luminal B
Luminal A
Normal-like
M0
M1
MX
T1
T2
T3
T4
TX
N0
N1
N2
N3
NX


E
p = 0.014
Sene_Signature
1.00
0.75
0.50
0.25
0.00
no_TNBC
TNBC

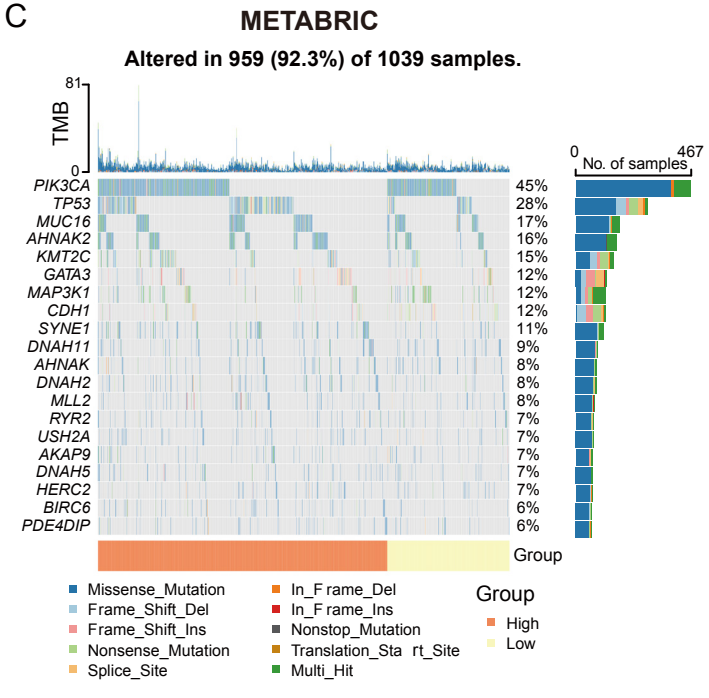
A


METABRIC
P = 0.03
TMB
2.0
1.5
1.0
0.5
0.0
High
Low
Group

B


TCGA
Altered in 389 (87.42%) of 445 samples.
TMB
4782
0
0
No. of samples
169
PIK3CA 38%
TP53 29%
TTN 21%
CDH1 18%
KMT2C 12%
MUC16 12%
MAP3K1 11%
GATA3 9%
FLG 7%
PTEN 7%
HMCN1 6%
USH2A 6%
SYNE1 6%
ZFHX4 6%
NF1 5%
PCLO 5%
RYR3 5%
SYNE2 5%
RYR2 5%
SPEN 5%
Group
Missense_Mutation
Frame_Shift_Ins
Splice_Site
Frame_Shift_Del
Nonsense_Mutation
In_Frame_Del
Multi_Hit
Group
High
Low

C


METABRIC
Altered in 959 (92.3%) of 1039 samples.
TMB
81
0
0
No. of samples
467
PIK3CA 45%
TP53 28%
MUC16 17%
AHNAK2 16%
KMT2C 15%
GATA3 12%
MAP3K1 12%
CDH1 12%
SYNE1 11%
DNAH11 9%
AHNAK 8%
DNAH2 8%
MLL2 8%
RYR2 7%
USH2A 7%
AKAP9 7%
DNAH5 7%
HERC2 7%
BIRC6 6%
PDE4DIP 6%
Group
Missense_Mutation
Frame_Shift_Del
Frame_Shift_Ins
Nonsense_Mutation
Splice_Site
In_Frame_Del
In_Frame_Ins
Nonstop_Mutation
Translation_Start_Site
Multi_Hit
Group
High
Low

A

Figure S7

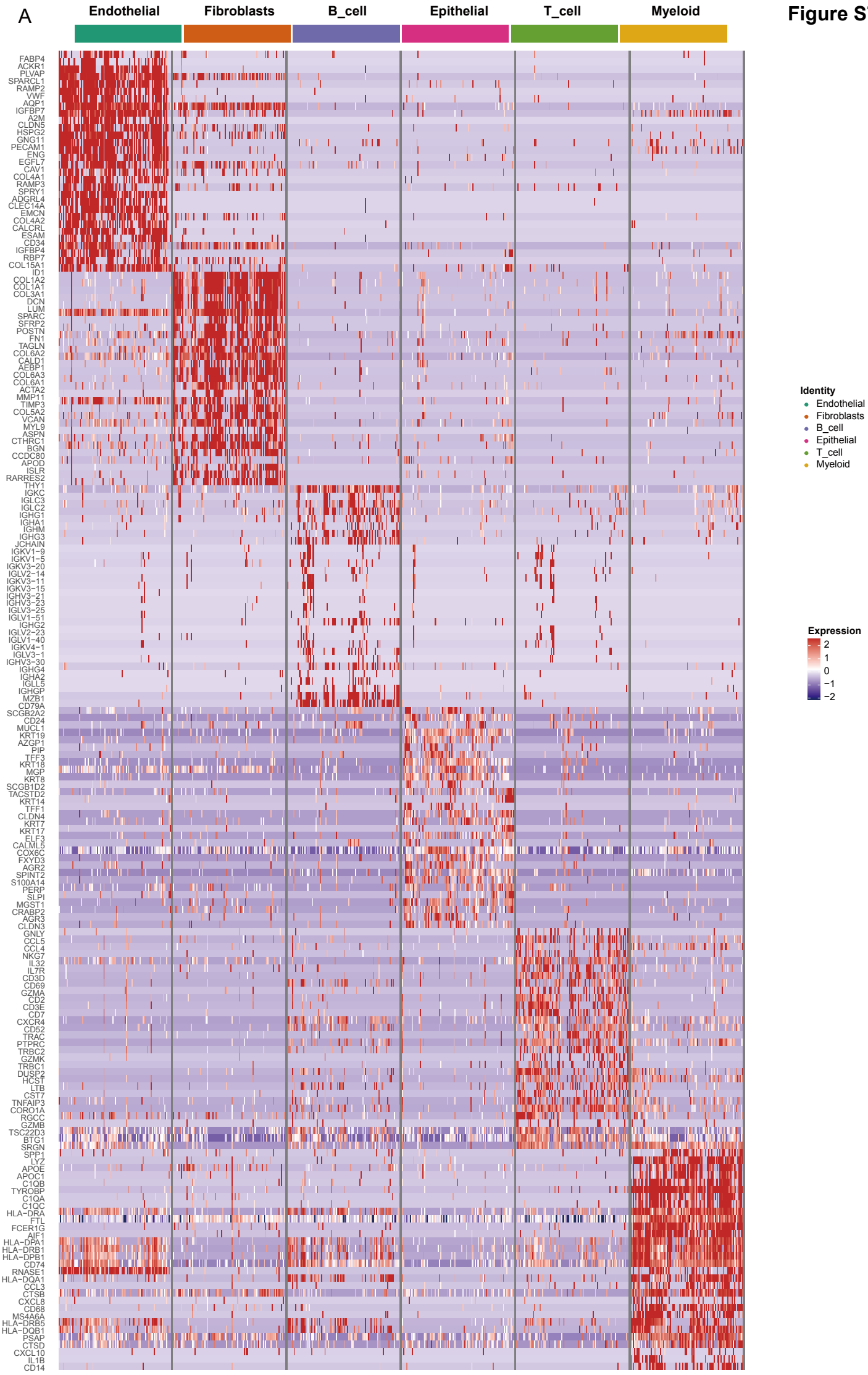
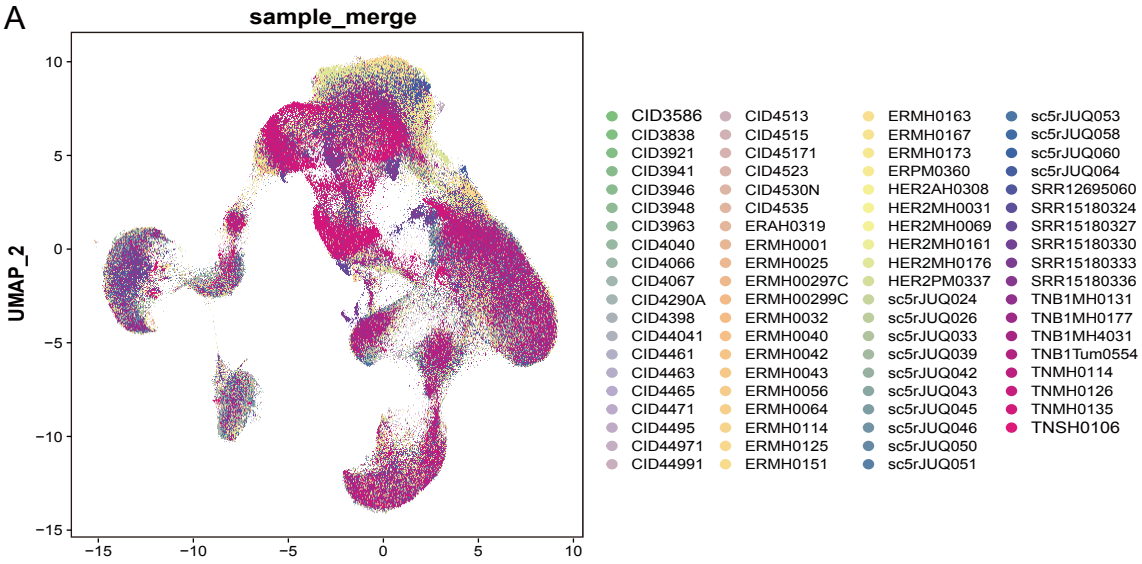
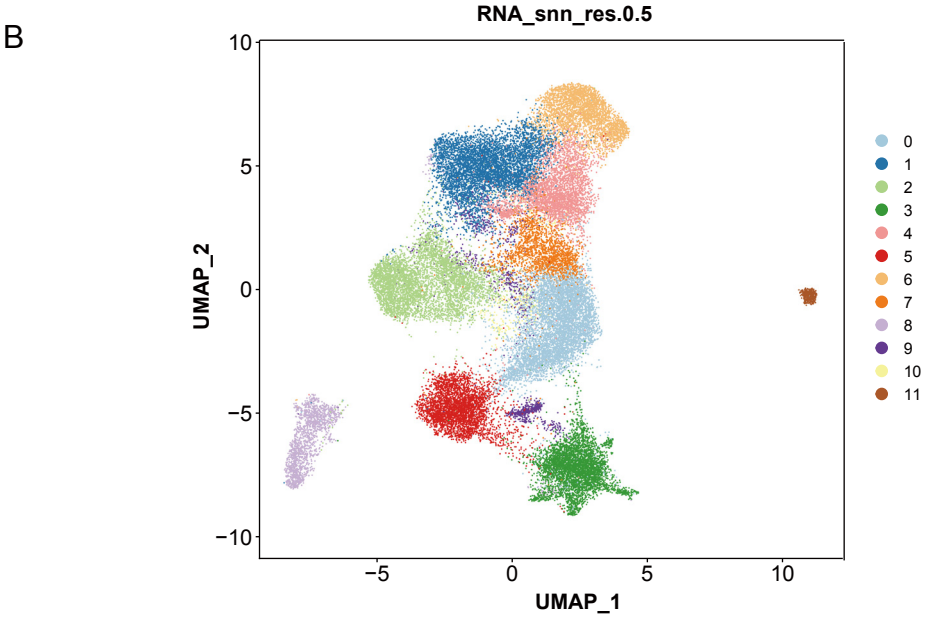

Endothelial
Fibroblasts
B_cell
Epithelial
T_cell
Myeloid
FABP4
ACKR1
PLVAP
SPARCL1
RAMP2
VWF
AQP1
IGFBP7
A2M
CLDN5
HSPG2
GNG11
PECAM1
ENG
EGFL7
CAV1
COL4A1
RAMP3
SPRY1
ADGRL4
CLEC14A
EMCN
COL4A2
CALCRL
ESAM
CD34
IGFBP4
RBP7
COL15A1
ID1
COL1A2
COL1A1
COL3A1
DCN
LUM
SPARC
SFRP2
POSTN
FN1
TAGLN
COL6A2
CALD1
AEBP1
COL6A3
COL6A1
ACTA2
MMP11
TIMP3
COL5A2
VCAN
MYL9
ASPN
CTHRC1
BGN
CCDC80
APOD
ISLR
RARRES2
THY1
IGKC
IGLC3
IGLC2
IGHG1
IGHA1
IGHM
IGHG3
JCHAIN
IGKV1-9
IGKV1-5
IGKV3-20
IGLV2-14
IGKV3-11
IGKV3-15
IGHV3-21
IGHV3-23
IGLV3-25
IGLV1-51
IGHG2
IGLV2-23
IGLV1-40
IGKV4-1
IGLV3-1
IGHV3-30
IGHG4
IGHA2
IGLL5
IGHGP
MZB1
CD79A
SCGB2A2
CD24
MUCL1
KRT19
AZGP1
PIP
TFF3
KRT18
MGP
KRT8
SCGB1D2
TACSTD2
KRT14
TFF1
CLDN4
KRT7
KRT17
ELF3
CALML5
COX6C
FXYD3
AGR2
SPINT2
S100A14
PERP
SLPI
MGST1
CRABP2
AGR3
CLDN3
GNLY
CCL5
CCL4
NKG7
IL32
IL7R
CD3D
CD69
GZMA
CD2
CD3E
CD7
CXCR4
CD52
TRAC
PTPRC
TRBC2
GZMK
TRBC1
DUSP2
HCST
LTB
CST7
TNFAIP3
CORO1A
RGCC
GZMB
TSC22D3
BTG1
SRGN
SPP1
LYZ
APOE
APOC1
C1QB
TYROBP
C1QA
C1QC
HLA-DRA
FTL
FCER1G
AIF1
HLA-DPA1
HLA-DRB1
HLA-DPB1
CD74
RNASE1
HLA-DQA1
CCL3
CTSB
CXCL8
CD68
MS4A6A
HLA-DRB5
HLA-DQB1
PSAP
CSD
CXCL10
IL1B
CD14
Identity
Endothelial
Fibroblasts
B_cell
Epithelial
T_cell
Myeloid
Expression
2
1
0
−1
−2

Figure S8


A
sample_merge
UMAP_2
10
5
0
−5
−10
−15
−15
−10
−5
0
5
10
CID3586
CID3838
CID3921
CID3941
CID3946
CID3948
CID3963
CID4040
CID4066
CID4067
CID4290A
CID4398
CID44041
CID4461
CID4463
CID4465
CID4471
CID4495
CID44971
CID44991
CID4513
CID4515
CID45171
CID4523
CID4530N
CID4535
ERAH0319
ERMH0001
ERMH0025
ERMH00297C
ERMH00299C
ERMH0032
ERMH0040
ERMH0042
ERMH0043
ERMH0056
ERMH0064
ERMH0114
ERMH0125
ERMH0151
ERMH0163
ERMH0167
ERMH0173
ERPM0360
HER2AH0308
HER2MH0031
HER2MH0069
HER2MH0161
HER2MH0176
HER2PM0337
sc5rJUQ024
sc5rJUQ026
sc5rJUQ033
sc5rJUQ039
sc5rJUQ042
sc5rJUQ043
sc5rJUQ045
sc5rJUQ046
sc5rJUQ050
sc5rJUQ051
sc5rJUQ053
sc5rJUQ058
sc5rJUQ060
sc5rJUQ064
SRR12695060
SRR15180324
SRR15180327
SRR15180330
SRR15180333
SRR15180336
TNB1MH0131
TNB1MH0177
TNB1MH4031
TNB1Tum0554
TNMH0114
TNMH0126
TNMH0135
TNSH0106


B
RNA_snn_res.0.5
UMAP_2
UMAP_1
10
5
0
−5
−10
−5
0
5
10
0
1
2
3
4
5
6
7
8
9
10
11

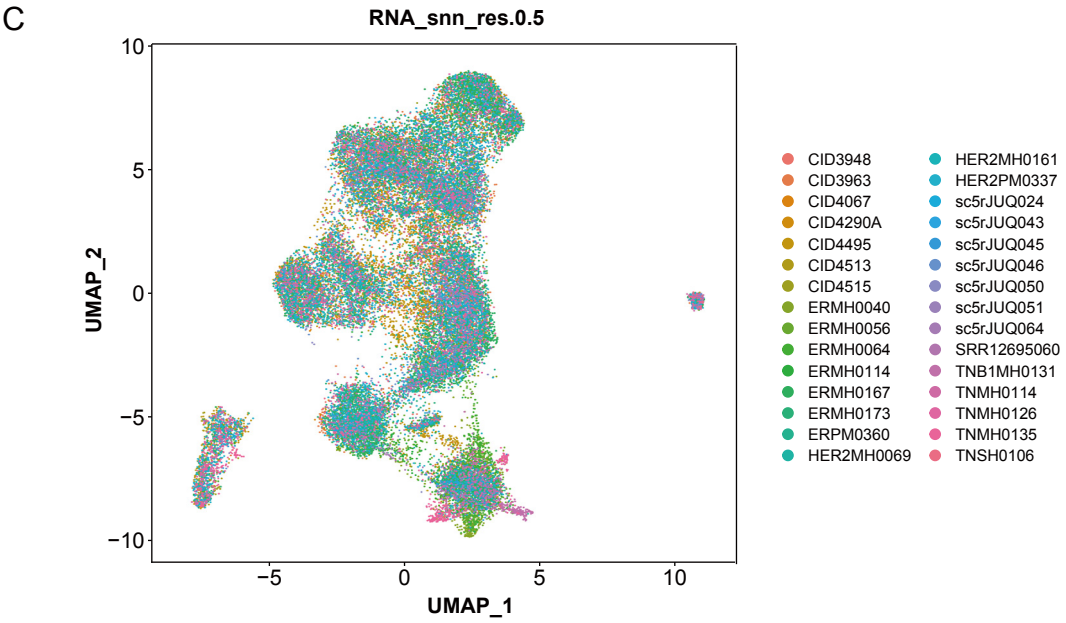
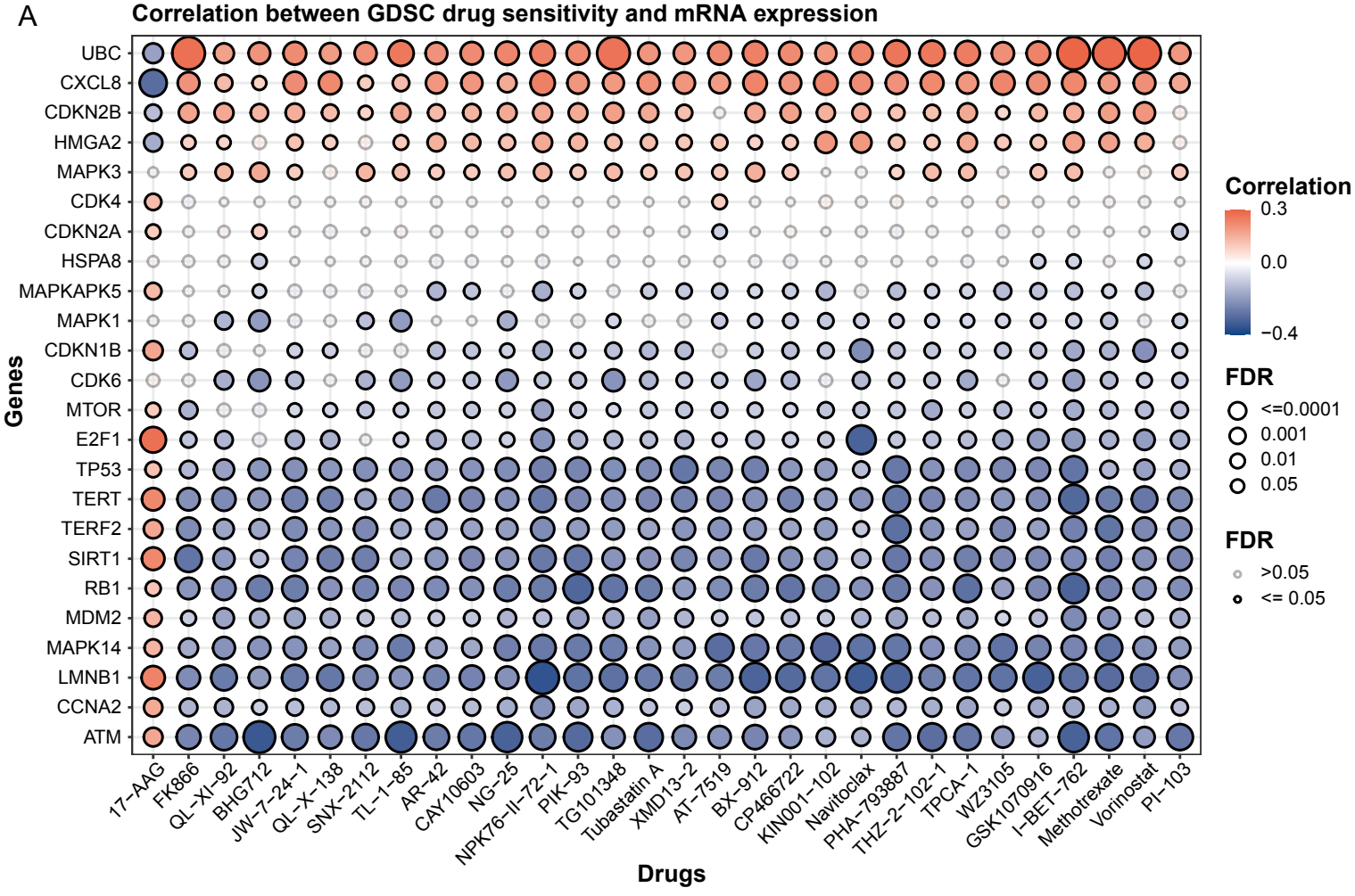

C
RNA_snn_res.0.5
UMAP_2
UMAP_1
10
5
0
−5
−10
−5
0
5
10
CID3948
CID3963
CID4067
CID4290A
CID4495
CID4513
CID4515
ERMH0040
ERMH0056
ERMH0064
ERMH0114
ERMH0167
ERMH0173
ERPM0360
HER2MH0069
HER2MH0161
HER2PM0337
sc5rJUQ024
sc5rJUQ043
sc5rJUQ045
sc5rJUQ046
sc5rJUQ050
sc5rJUQ051
sc5rJUQ064
SRR12695060
TNB1MH0131
TNMH0114
TNMH0126
TNMH0135
TNSH0106

Figure S9


A
Correlation between GDSC drug sensitivity and mRNA expression
Genes
UBC
CXCL8
CDKN2B
HMGA2
MAPK3
CDK4
CDKN2A
HSPA8
MAPKAPK5
MAPK1
CDKN1B
CDK6
MTOR
E2F1
TP53
TERT
TERF2
SIRT1
RB1
MDM2
MAPK14
LMNB1
CCNA2
ATM
17−AAG
FK866
QL−XI−92
BHG712
JW−7−24−1
QL−X−138
SNX−2112
TL−1−85
AR−42
CAY10603
NG−25
NPK76−II−72−1
PIK−93
TG101348
Tubastatin A
XMD13−2
AT−7519
BX−912
CP466722
KIN001−102
Navitoclax
PHA−793887
THZ−2−102−1
TPCA−1
WZ3105
GSK1070916
I−BET−762
Methotrexate
Vorinostat
PI−103
Drugs
Correlation
0.3
0.0
−0.4
FDR
<=0.0001
0.001
0.01
0.05
FDR
>0.05
<= 0.05

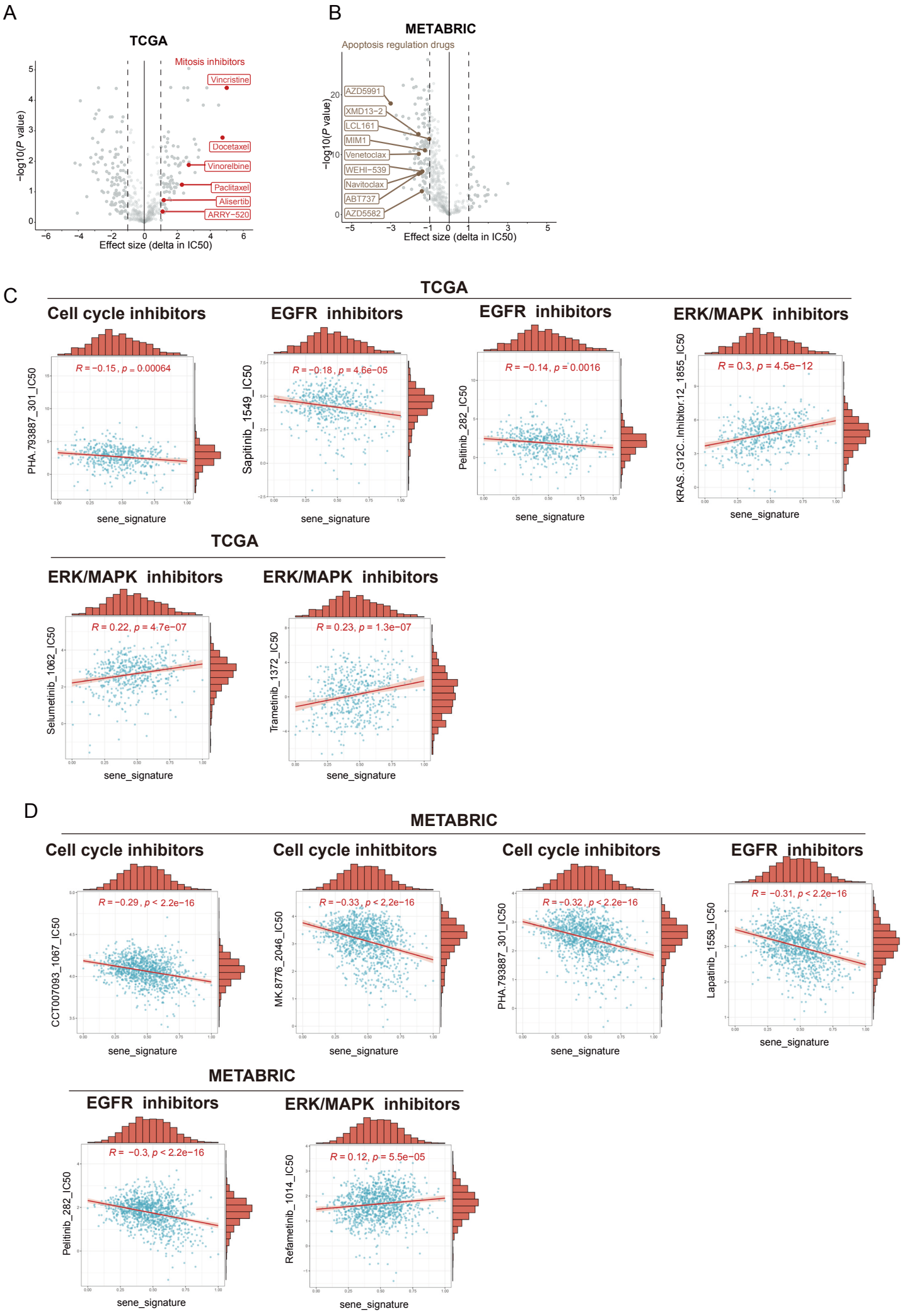

A
TCGA
Mitosis inhibitors
Vincristine
Docetaxel
Vinorelbine
Paclitaxel
Alisertib
ARRY−520
−log10(P value)
Effect size (delta in IC50)
B
METABRIC
Apoptosis regulation drugs
AZD5991
XMD13−2
LCL161
MIM1
Venetoclax
WEHI−539
Navitoclax
ABT737
AZD5582
−log10(P value)
Effect size (delta in IC50)
C
TCGA
Cell cycle inhibitors
EGFR inhibitors
EGFR inhibitors
ERK/MAPK inhibitors
PHA.793887_301_IC50
R = −0.15, p = 0.00064
Sapitinib_1549_IC50
R = −0.18, p = 4.6e−05
Pelitinib_282_IC50
R = −0.14, p = 0.0016
KRAS..G12C..Inhibitor.12_1855_IC50
R = 0.3, p = 4.5e−12
sene_signature
TCGA
ERK/MAPK inhibitors
ERK/MAPK inhibitors
Selumetinib_1062_IC50
R = 0.22, p = 4.7e−07
Trametinib_1372_IC50
R = 0.23, p = 1.3e−07
sene_signature
D
METABRIC
Cell cycle inhibitors
Cell cycle inhitbitors
Cell cycle inhibitors
EGFR inhibitors
CCT007093_1067_IC50
R = −0.29, p < 2.2e−16
MK.8776_2046_IC50
R = −0.33, p < 2.2e−16
PHA.793887_301_IC50
R = −0.32, p < 2.2e−16
Lapatinib_1558_IC50
R = −0.31, p < 2.2e−16
sene_signature
METABRIC
EGFR inhibitors
ERK/MAPK inhibitors
Pelitinib_282_IC50
R = −0.3, p < 2.2e−16
Refametinib_1014_IC50
R = 0.12, p = 5.5e−05
sene_signature

Figure S11