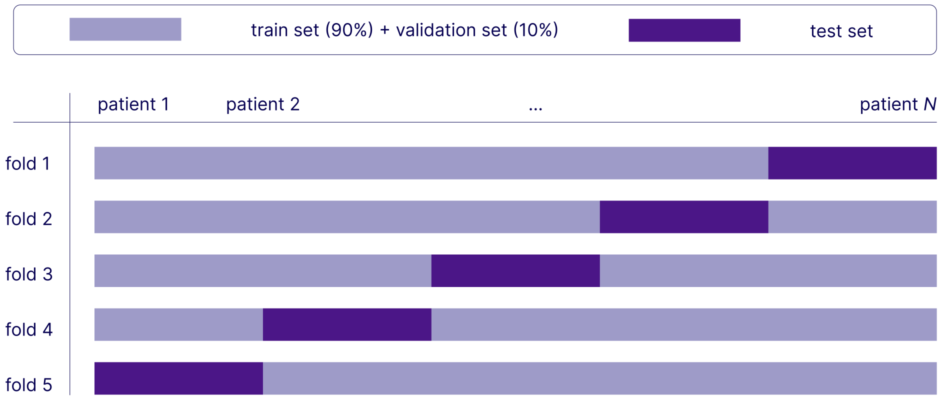
Supplementary tables and figures

Supplementary Table 1: Number of Whole Slide Images and unique number of patients used from each cancer type in TCGA. Source data are provided in the Source Data File.

TCGA project	number WSIs	number patients
TCGA-BRCA	1,130	1,059
TCGA-UCEC	563	502
TCGA-LUAD	536	473
TCGA-THCA	517	504
TCGA-KIRC	514	508
TCGA-LUSC	510	476
TCGA-SKCM	474	432
TCGA-HNSC	464	442
TCGA-COAD	455	447
TCGA-BLCA	451	380
TCGA-PRAD	448	401
TCGA-STAD	411	386
TCGA-LIHC	373	359
TCGA-KIRP	299	275
TCGA-GBM	237	102
TCGA-PAAD	202	176
total	7,584	6,922

Supplementary Table 2: Number of Whole Slide Images and unique number of patients used from each cancer type in CPTAC. Source data are provided in the Source Data File.

CPTAC project	number WSIs	number patients
CPTAC-BRCA	106	106
CPTAC-CCRCC	302	211
CPTAC-COAD	103	103
CPTAC-GBM	94	94
CPTAC-LUAD	222	222
CPTAC-LSCC	109	108
CPTAC-PDA	146	146
total	1,081	989



Supplementary Figure 1: **Data splitting.** First, five folds were made each of which consist of a 'global' train (80%) and test set (20%). The 'global' train set was further split into a train (90%) and validation (10%) set. In each fold i, validation set i was used to determine the optimal point to stop training model i, which was then evaluated on test set i. Afterwards, predictions on patients from test sets i ($i = 1..5$) were concatenated for performance evaluation.

Supplementary Table 3: Number of genes significantly well predicted in the TCGA test sets. Source data are provided in the Source Data File.

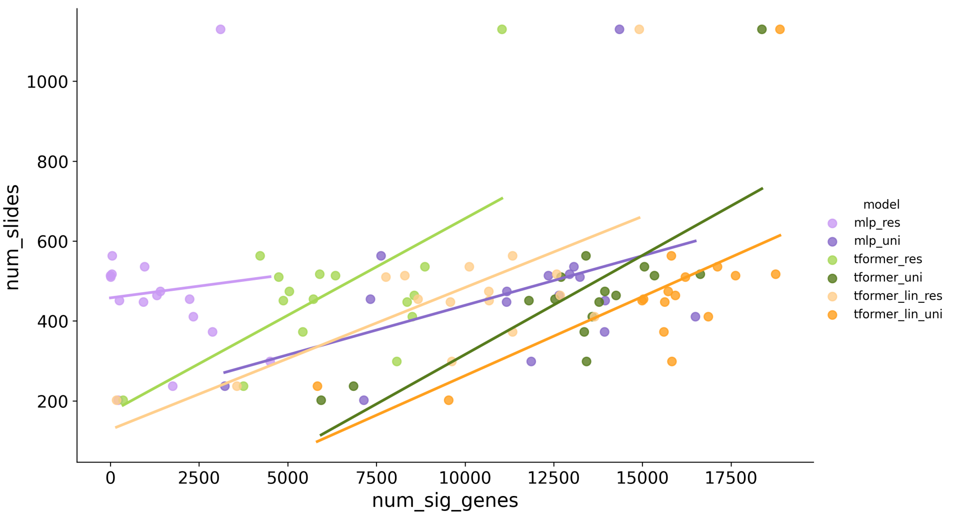
	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
BRCA	3,106	14,355	11,036	18,368	14,906	18,878
THCA	38	12,943	5,902	16,633	12,584	18,758
KIRC	0	12,344	6,342	15,337	8,300	17,623
LUAD	957	13,066	8,860	15,056	10,115	17,116
STAD	2,334	16,490	8,509	13,587	13,651	16,855
LUSC	11	13,238	4,740	12,706	7,769	16,210
HNSC	1,312	12,647	8,563	14,253	12,675	15,928
KIRP	4,500	11,861	8,071	13,424	9,620	15,828
UCEC	46	7,624	4,218	13,407	11,335	15,810
SKCM	1,403	11,179	5,041	13,937	10,670	15,728
PRAD	927	11,170	8,367	13,775	9,582	15,627
LIHC	2,874	13,927	5,418	13,355	11,334	15,604
COAD	2,231	7,331	5,723	12,530	8,668	15,028
BLCA	253	13,944	4,871	11,797	10,675	14,989
PAAD	219	7,144	355	5,942	170	9,535
GBM	1,750	3,223	3,749	6,854	3,559	5,831
average	1,372	11,405	6,235	13,185	9,725	15,334

Supplementary Table 4: Median correlation coefficient of top 1000 genes in TCGA test sets. (top 1000 genes defined independently within each model and cancer type). Source data are provided in the Source Data File.

	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
THCA	0.116	0.553	0.336	0.563	0.528	0.652
LIHC	0.227	0.506	0.283	0.480	0.418	0.560
KIRP	0.377	0.485	0.365	0.471	0.390	0.548
KIRC	0.080	0.418	0.225	0.462	0.332	0.537
STAD	0.265	0.500	0.346	0.405	0.407	0.536
BRCA	0.200	0.411	0.325	0.458	0.396	0.534
PAAD	0.149	0.494	0.168	0.371	0.140	0.523
BLCA	0.094	0.458	0.250	0.425	0.378	0.512
PRAD	0.126	0.464	0.321	0.417	0.345	0.503
LUAD	0.130	0.415	0.284	0.404	0.306	0.495
SKCM	0.166	0.420	0.225	0.449	0.366	0.483
COAD	0.215	0.328	0.263	0.385	0.345	0.480
HNSC	0.166	0.420	0.277	0.393	0.377	0.479
LUSC	0.094	0.360	0.197	0.337	0.266	0.443
UCEC	0.081	0.309	0.180	0.329	0.293	0.413
GBM	0.230	0.300	0.285	0.354	0.316	0.367
average	0.170	0.428	0.271	0.419	0.350	0.504

Supplementary Table 5: Normalized RMSE of top 1000 genes in TCGA test sets. (top 1000 genes defined independently within each model and cancer type). Source data are provided in the Source Data File.

	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
COAD	0.240	0.251	0.225	0.206	0.202	0.191
UCEC	0.234	0.180	0.211	0.181	0.190	0.166
PAAD	0.272	0.216	0.137	0.171	0.124	0.161
KIRP	0.242	0.234	0.175	0.167	0.162	0.153
THCA	0.309	0.268	0.223	0.181	0.182	0.146
HNSC	0.178	0.198	0.164	0.153	0.154	0.140
LUSC	0.224	0.197	0.160	0.168	0.159	0.139
PRAD	0.290	0.196	0.163	0.155	0.142	0.138
KIRC	0.189	0.205	0.148	0.159	0.148	0.138
BLCA	0.228	0.206	0.170	0.154	0.160	0.137
LIHC	0.220	0.202	0.164	0.138	0.156	0.123
GBM	0.221	0.232	0.098	0.115	0.093	0.122
STAD	0.180	0.198	0.149	0.142	0.142	0.119
BRCA	0.181	0.146	0.150	0.128	0.135	0.118
SKCM	0.131	0.169	0.136	0.113	0.135	0.112
LUAD	0.171	0.173	0.151	0.127	0.143	0.107
average	0.219	0.204	0.164	0.154	0.152	0.138



Supplementary Figure 2: **Relation num_slides and num_sig_genes.** Each dot represents for a particular model and in particular cancer type, how many WSIs were available and how many genes were well-predicted. The dots are colored according to the model type, and a regression line is plotted across dots within each model type. Source data are provided in the Source Data File.

Supplementary Table 6: Effect of downsampling training dataset size of TCGA-BRCA on performance metrics for SEQUOIA, with num_genes: number of significantly predicted genes, med_r and std_r: median and standard deviation of correlation coefficient of top 1000 genes, med_rmse and std_rmse: median and standard deviation on normalized RMSE of top 1000 genes. The percentage indicates the fraction of the slides that were retained. Source data are provided in the Source Data File.

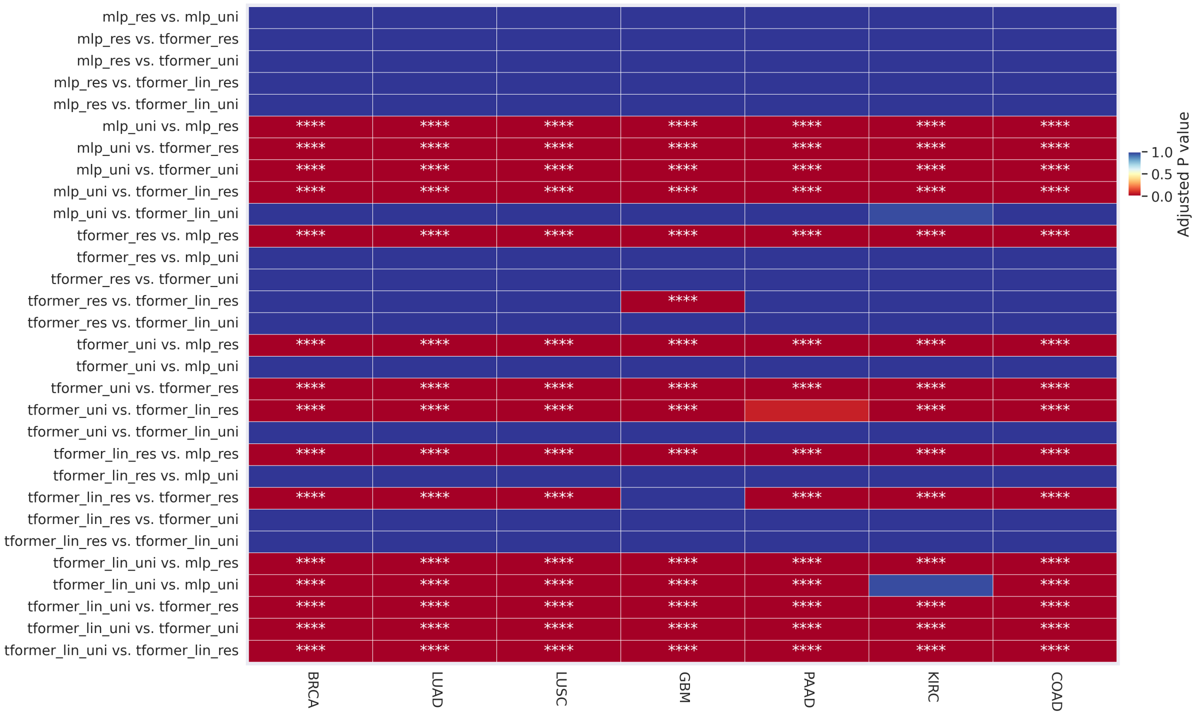
	cancer	num_genes	med_r	std_r	med_rmse	std_rmse
0	BRCA-40%	15726	0.513	0.048	0.754	0.325
1	BRCA-30%	13684	0.509	0.054	0.776	0.345
2	BRCA-20%	9706	0.479	0.054	0.781	0.357

Supplementary Table 7: Median correlation coefficient of top 1000 genes in CPTAC. (top 1000 genes defined independently within each model and cancer type). Source data are provided in the Source Data File.

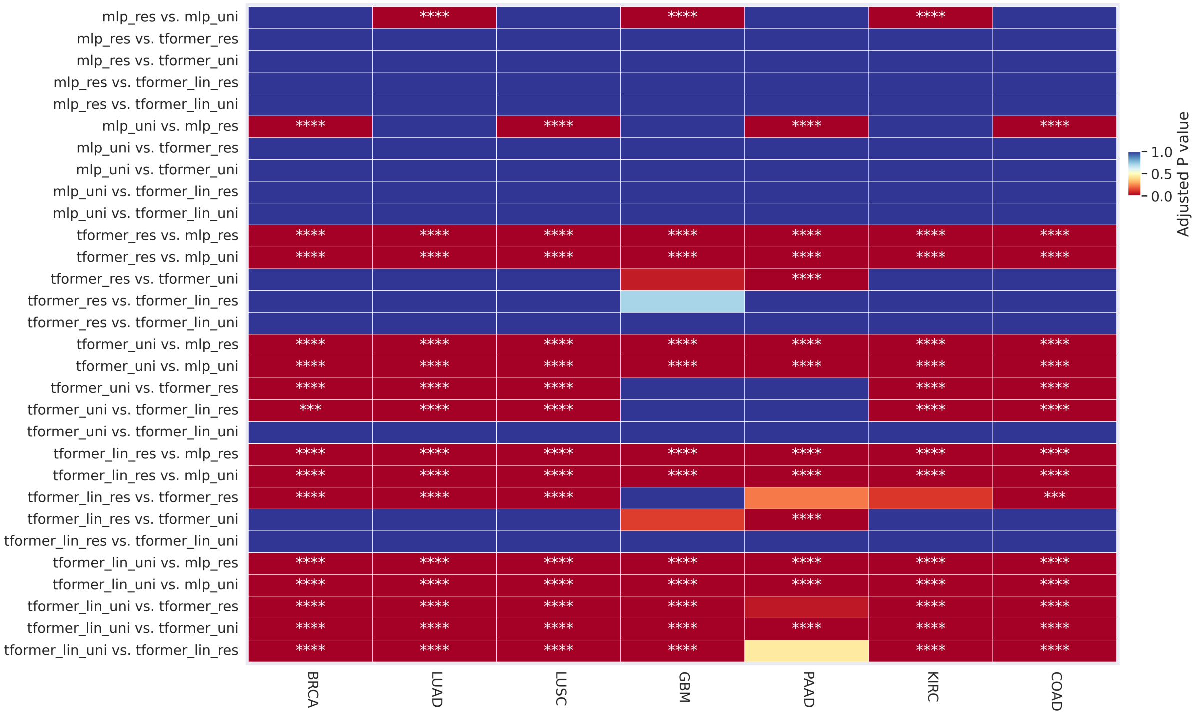
	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
BRCA	0.466	0.596	0.544	0.582	0.562	0.636
LUAD	0.371	0.534	0.414	0.489	0.468	0.578
LUSC (LSCC)	0.235	0.509	0.316	0.490	0.343	0.543
KIRC (CCRCC)	0.233	0.528	0.337	0.471	0.370	0.526
COAD	0.262	0.426	0.272	0.379	0.326	0.499
GBM	0.210	0.403	0.254	0.294	0.237	0.422
PAAD (PDA)	0.194	0.249	0.203	0.221	0.221	0.321
average	0.282	0.464	0.334	0.418	0.361	0.504

Supplementary Table 8: Normalized RMSE of top 1000 genes in CPTAC. (top 1000 genes defined independently within each model and cancer type). Source data are provided in the Source Data File.

	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
PAAD (PDA)	0.291	0.222	0.163	0.166	0.161	0.160
LUAD	0.190	0.215	0.178	0.167	0.171	0.150
LUSC (LSCC)	0.192	0.185	0.161	0.144	0.158	0.137
KIRC (CCRCC)	0.183	0.194	0.155	0.143	0.154	0.135
BRCA	0.179	0.163	0.147	0.138	0.141	0.131
GBM	0.215	0.242	0.124	0.126	0.125	0.116
COAD	0.152	0.147	0.132	0.122	0.130	0.114
average	0.200	0.195	0.151	0.144	0.149	0.135



Supplementary Figure 3: **CPTAC evaluation (correlation coefficients)**. The table shows the P value calculated for each pairwise comparison of models using a one-sided Mann-Whitney U test for the hypothesis that values from model A are larger than values from model B. Legend: * for $p < 0.05$, ** for $p <= 0.01$, *** for $p <= 0.001$ and **** for $p <= 0.0001$. Source data are provided in the Source Data File.



Supplementary Figure 4: **CPTAC evaluation (normalized RMSE)**. The table shows the P value calculated for each pairwise comparison of models using a one-sided Mann-Whitney U test for the hypothesis that values from model A are smaller than values from model B. Legend: * for $p < 0.05$, ** for $p <= 0.01$, *** for $p <= 0.001$ and **** for $p <= 0.0001$. Source data are provided in the Source Data File.

Supplementary Table 9: Number of genes that validate to both TCGA test set and CPTAC. Source data are provided in the Source Data File.

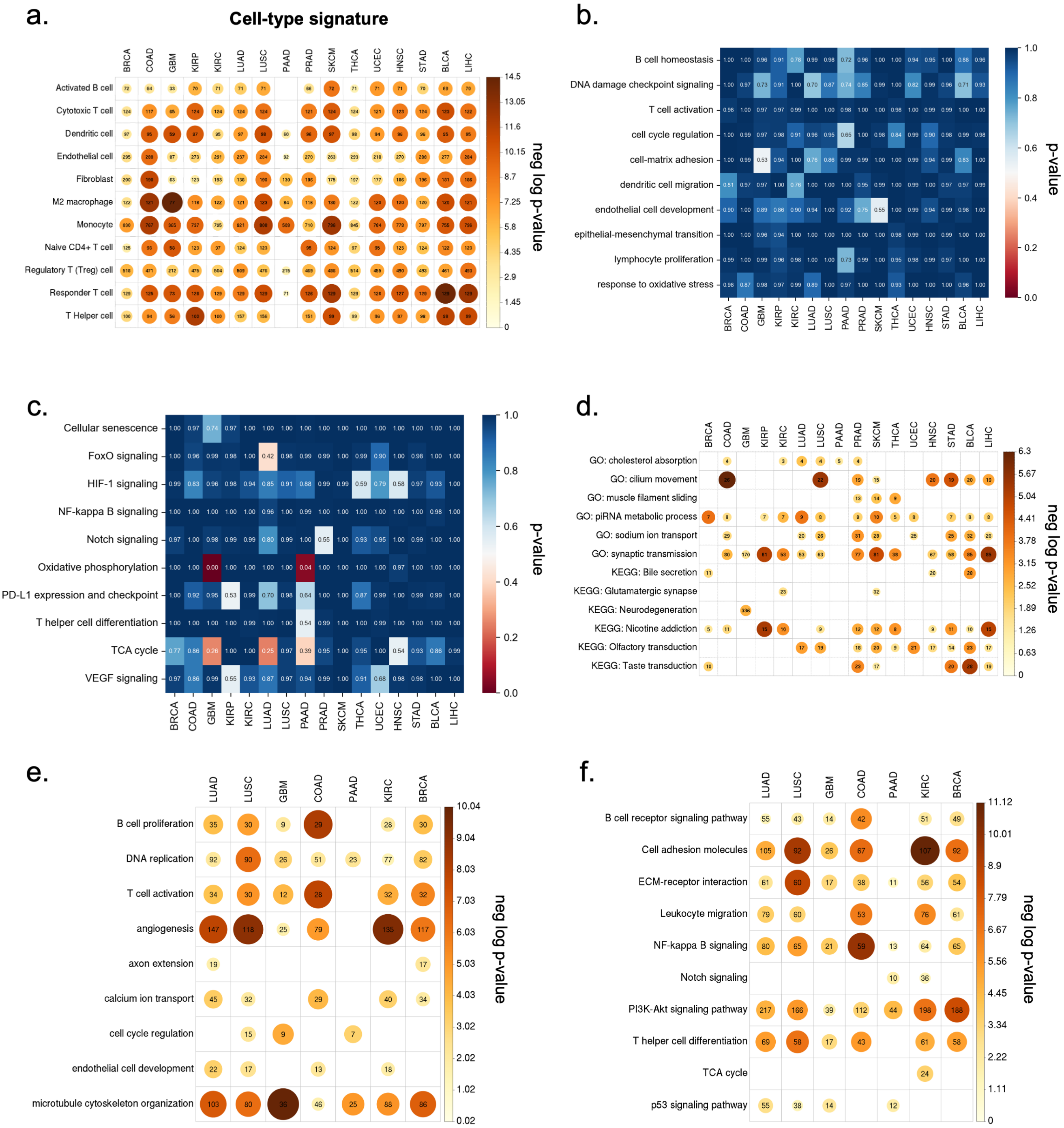
	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
LUAD	498	8,294	3,726	8,166	5,116	12,422
KIRC	0	6,899	2,532	8,698	3,937	10,477
BRCA	1,343	7,569	5,530	8,243	6,858	9,418
LUSC	0	6,128	780	5,914	2,029	8,610
COAD	140	1,944	242	2,798	1,159	5,784
GBM	9	1,039	94	495	90	1,816
PAAD	0	439	6	227	4	1,589
average	284	4,616	1,844	4,934	2,741	7,159

Supplementary Table 10: Median correlation coefficient of top 1000 genes for genes that validate to both TCGA test set and CPTAC (correlation coefficient from evaluation on CPTAC). Source data are provided in the Source Data File.

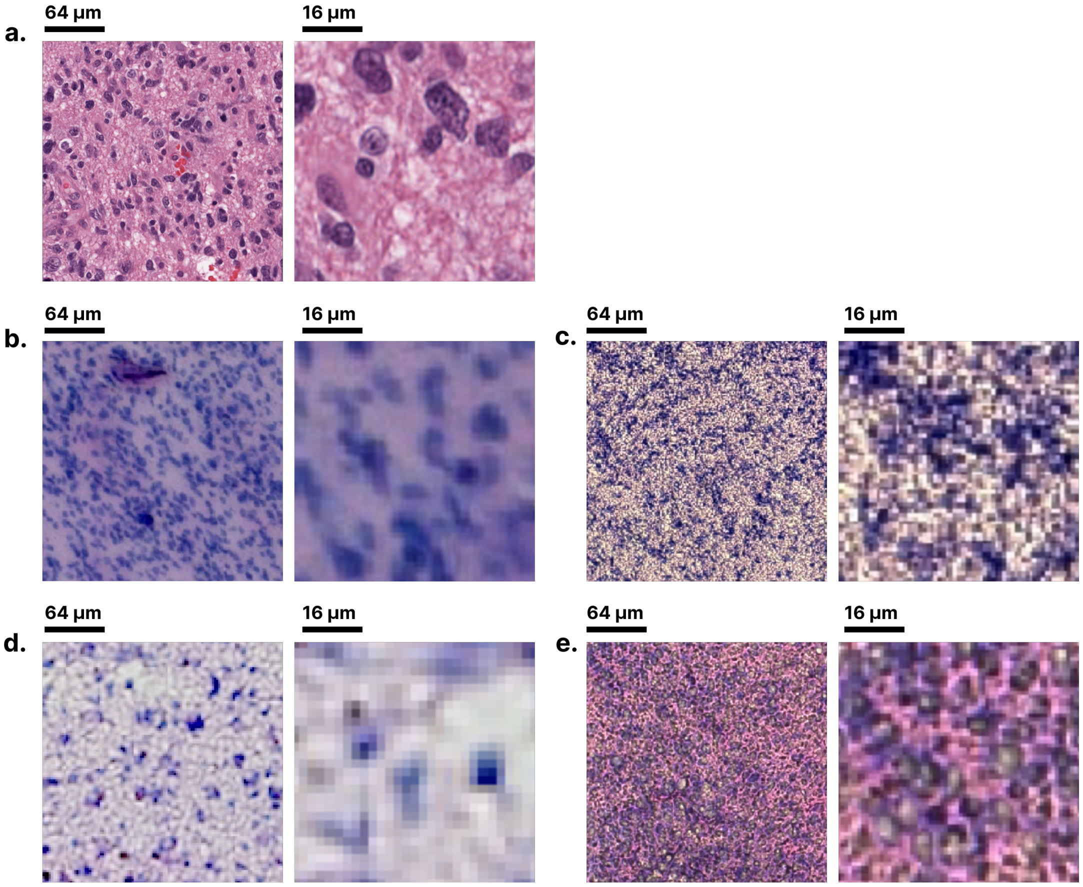
	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
BRCA	0.393	0.595	0.543	0.582	0.562	0.636
LUAD	0.333	0.531	0.406	0.488	0.463	0.578
LUSC	0.000	0.503	0.268	0.488	0.320	0.543
KIRC	0.000	0.525	0.308	0.470	0.356	0.525
COAD	0.245	0.391	0.260	0.365	0.298	0.498
GBM	0.245	0.333	0.254	0.271	0.258	0.391
PAAD	0.000	0.225	0.214	0.223	0.216	0.274
average	0.174	0.443	0.322	0.412	0.353	0.492

Supplementary Table 11: Normalized RMSE of top 1000 genes for genes that validate to both TCGA test set and CPTAC. (RMSE from evaluation on CPTAC). Source data are provided in the Source Data File.

	mlp_res	mlp_uni	tformer_res	tformer_uni	tformer_lin_res	tformer_lin_uni
PAAD	0.000	0.859	0.555	0.608	0.414	0.296
GBM	1.000	0.985	0.573	0.430	0.485	0.195
LUAD	0.395	0.236	0.206	0.175	0.191	0.152
LUSC	0.000	0.231	0.405	0.160	0.244	0.144
KIRC	0.000	0.224	0.214	0.152	0.190	0.139
BRCA	0.322	0.174	0.160	0.141	0.147	0.133
COAD	0.401	0.288	0.420	0.170	0.264	0.132
average	0.303	0.428	0.362	0.262	0.277	0.170



Supplementary Figure 5: **Characterization of the well-predicted genes.** a) Heatmap showing the significant P values for cell-type signatures enriched with the well-predicted genes. Color and size of the circles represent the negative log-transformed P values. Integers represent the absolute gene count in each category, and non-significant categories are left in blank. b-c) P values from the gene-set enrichment analysis of the inaccurately predicted genes using the gene sets from b) gene ontology and c) KEGG. d) Heatmap showing the P values for gene sets enriched with the inaccurately predicted genes across cancer types. Color and size of the circles represent the negative log-transformed P values. Integers represent the absolute gene count in each category, and non-significant categories are left in blank. e-f) Heatmaps showing the significant P values for pathways enriched with well-predicted genes in both the TCGA and CPTAC datasets using the e) gene ontology and f) KEGG pathways. Source data for all panels are provided in the Source Data File.

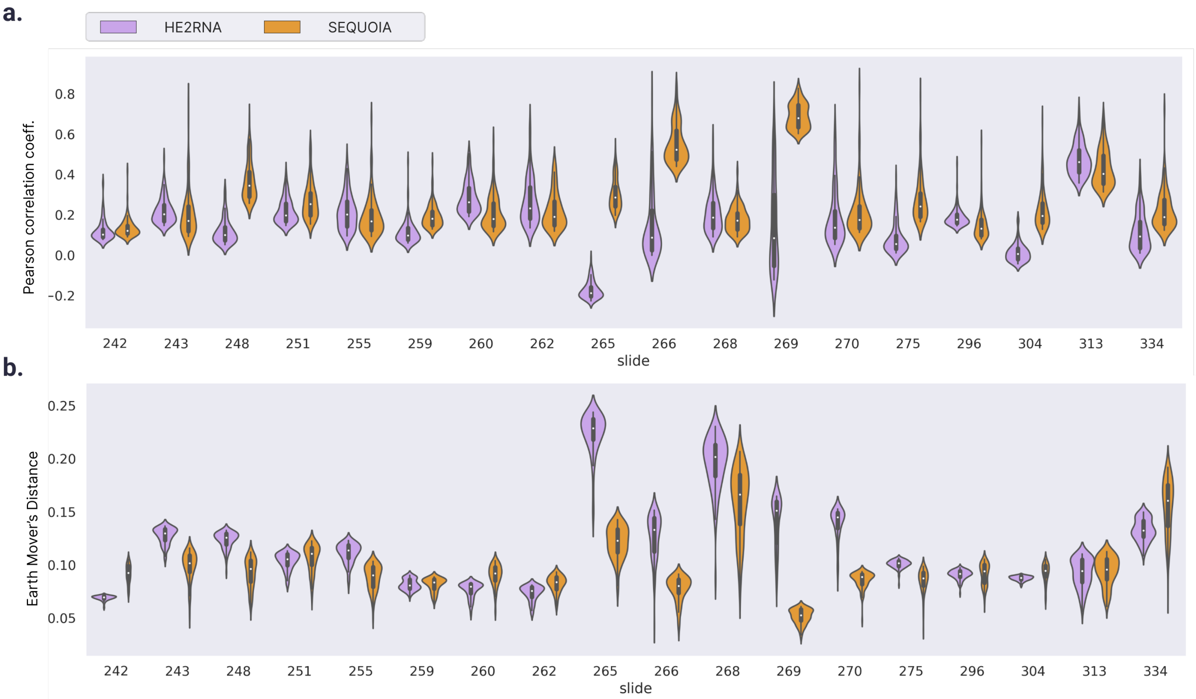


Supplementary Figure 6: **Comparison of H&E quality of TCGA-GBM slides to slides from spatial GBM dataset.** a) Example of quality of TCGA-GBM slide, with more zoomed-in version on the right. b-e) Four examples of quality from spatial GBM slides.

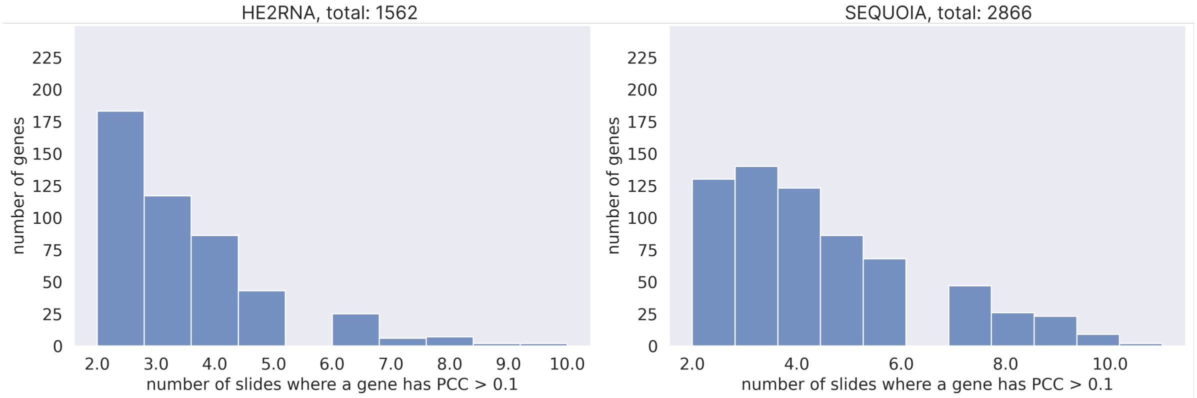
Supplementary Table 12: median EMD and Pearson correlation coefficient for top 100 genes for each model on the spatial GBM dataset. Source data are provided in the Source Data File.

	emd_HE2RNA	emd_SEQUOIA
268	0.202	0.166
334	0.132	0.160
265	0.229	0.123
251	0.105	0.110
243	0.130	0.101
248	0.125	0.096
313	0.094	0.095
304	0.087	0.094
296	0.091	0.094
242	0.069	0.092
260	0.079	0.092
255	0.113	0.090
270	0.145	0.088
275	0.101	0.087
262	0.075	0.083
259	0.080	0.083
266	0.133	0.080
269	0.151	0.052
average	0.119	0.099

	corr_HE2RNA	corr_SEQUOIA
269	0.084	0.678
266	0.086	0.522
313	0.461	0.402
248	0.100	0.344
265	-0.189	0.285
251	0.196	0.253
275	0.055	0.241
304	0.006	0.194
334	0.092	0.189
262	0.232	0.189
259	0.097	0.180
260	0.261	0.178
270	0.135	0.174
268	0.186	0.171
243	0.202	0.170
255	0.201	0.167
296	0.175	0.131
242	0.101	0.121
average	0.138	0.255



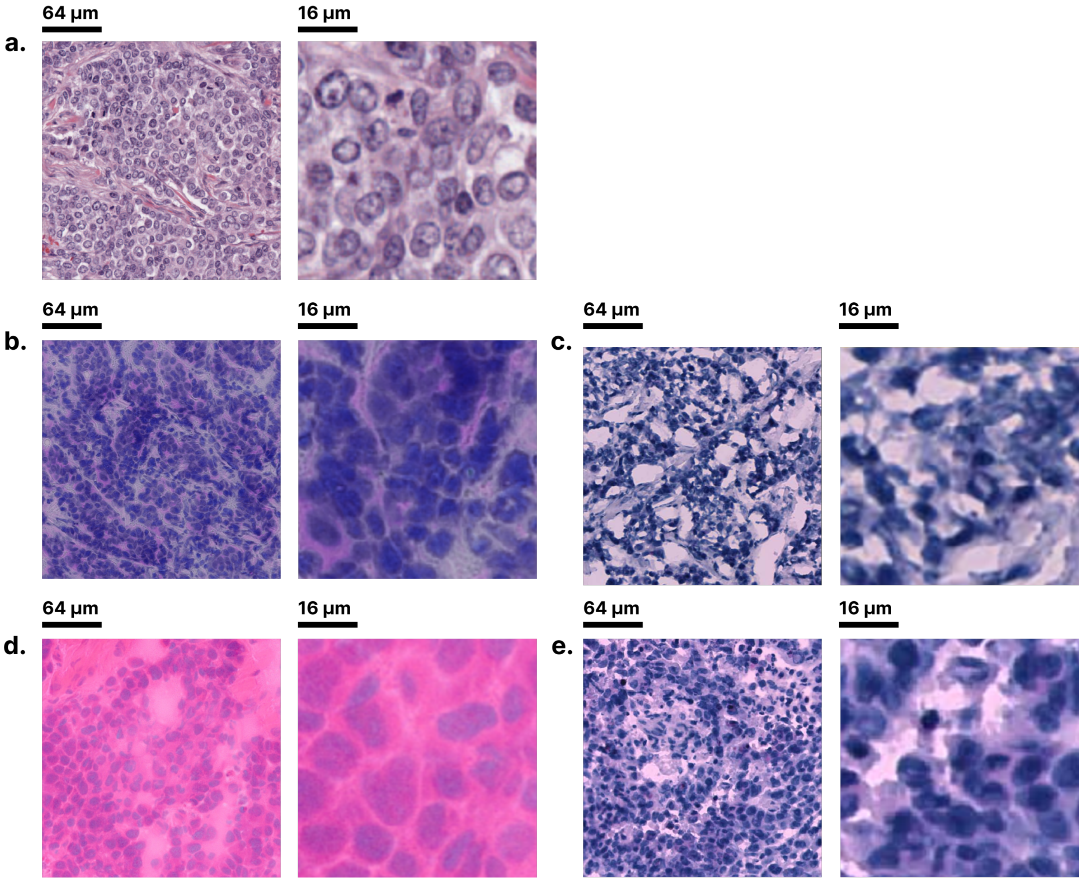
Supplementary Figure 7: **Performance for gene expression prediction of the spatial GBM dataset.** a) Distribution of Pearson correlation coefficients for the top 100 most accurately predicted genes with highest correlation coefficients obtained from each model. b) Distribution of EMD values for the top 100 most accurately predicted genes with lowest EMD values obtained from each model. Source data are provided in the Source Data File.



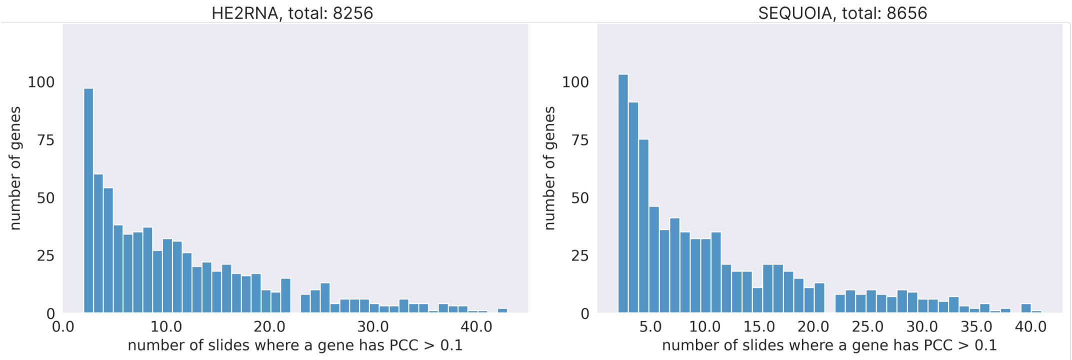
Supplementary Figure 8: **Predictability of genes across >2 slides in the spatial GBM cohort.** Source data are provided in the Source Data File.

gene	N
PDIA4	11
GALNT2	11
CTSH	10
EXOSC10	10
S100A13	10
PLXDC2	10
LMNA	10
PDLIM7	10
PTBP1	10
DVL2	10
FCGBP	10
ALDH9A1	9
COLGALT1	9
LPCAT2	9
ALDOC	9
TUBB	9
FAM3C	9
TRIM24	9
TCF12	9
PLOD3	9
FGFRL1	9
COL6A1	9
SERPINH1	9
TGFBI	9
TENT5A	9
TFDP2	9
LAMC1	9
FXYD5	9
C1R	9
SLC4A7	9
RAB3GAP2	9
VSIR	9
CREBL2	9
GJA1	9
AKAP8	8
CBX4	8
PCOLCE	8
FAM171B	8
PPP1R1B	8
SMPD4	8
LOXL2	8
DAZAP1	8
MMP14	8
ELAVL1	8
SFPQ	8
PLS3	8
KPNA2	8
UGDH	8
MYD88	8
TWSG1	8
FURIN	8
NONO	8
CSRP1	8
INTS3	8
BHLHE41	8
PLOD1	8
CAVIN1	8
ACIN1	8
HNRNPH1	8
STOM	8
HLA-DRB1	7
PTPRC	7
HLA-DPB1	7
MESD	7
FKBP10	7
HLA-DQB1	7
ZNF274	7
LAIR1	7
PIK3R1	7
ARHGEF2	7
PHF19	7
MTA1	7
RIN2	7
EFNB2	7
FAM20C	7
MLLT11	7
ACSL4	7
P4HB	7
ATL3	7
CAPRIN2	7
BIN1	7
VEZF1	7
P3H1	7
CPQ	7
CAMK2D	7
ELK3	7
LAMB2	7
THRA	7
TLR4	7
CTSZ	7
PROSER1	7
SLC16A3	7
PCGF3	7
NPEPL1	7
ATP6V1A	7
OSBPL3	7
SLC12A7	7
ZYX	7
RICTOR	7
RMND5A	7
DPP10	7
COL6A2	7
B3GNT2	7
TBXAS1	7
HNMT	7
TRIP6	7
CYBRD1	7
IL13RA1	6
FNDC3B	6
ZNF711	6
PCGF2	6
LYN	6
PHACTR4	6
ATL1	6
DKC1	6
TMEM248	6
NBPF15	6
GMPS	6
S100A16	6
BCL3	6
CHSY1	6
NEIL2	6
RIN3	6
PTGDS	6
B3GNT9	6
IQGAP1	6
FGF1	6
ADD3	6
ANXA1	6
CDCA5	6
SPOCD1	6
NCALD	6
DNAJC9	6
FUBP1	6
PLCXD1	6
FZD1	6
AP3M1	6
SPRY2	6
RPS6KA5	6
LILRB4	6
RNF144A	6
SYTL2	6
TRIM28	6
RNPS1	6
ANXA4	6
INTS11	6
SPRED2	6
MCM7	6
SRPK1	6
LRRC8D	6
SDC4	6
SUCLG2	6
TJAP1	6
ALCAM	6
SHOC2	6
MICALL2	6
EPC1	6
SOX4	6
PABPN1	6
MEX3A	6
CTSB	6
PHACTR3	6
EHD2	6
LITAF	6
HES6	6
FAM102B	6
CASP7	6
EZH2	6
IL4R	6
CYSTM1	6
MYO1C	6
TRIP10	6
DPYSL4	6
GALNT15	6
TRO	6
SOX11	5
RBM14	5
TNR	5
PDPN	5
DPYD	5
HNRNPD	5
ARHGAP18	5
SAT1	5
F13A1	5
CLU	5
SLC25A37	5
ELAVL2	5
RELB	5
MFSD1	5
TNFAIP2	5
MARCKSL1	5
PTPN9	5
TAGLN2	5
ACAP3	5
CSNK2A1	5
SNRPD3	5
GLIPR1	5
PLEKHB1	5
PYGL	5
TRADD	5
SLC44A1	5
AC018647.1	5
SNX32	5
ABHD15	5
SLA	5
KIF21B	5
AQP1	5
RAB40B	5
KIF2C	5
CADM2	5
FZD7	5
PRR3	5
SERPING1	5
AMOTL2	5
EMP1	5
FOXK2	5
NKIRAS2	5
SMARCB1	5
ELOVL2	5
AIFM2	5
ATAD3B	5
DHPS	5
HEBP1	5
PIPOX	5
RNF135	5
RIN1	5
ENPP4	5
KANK2	5
SOX6	5
FMNL1	5
PTK7	5
ACSL3	5
HLA-DQA1	5
PTPN18	5
STRIP1	5
GPR173	5
FCHSD2	5
LIG3	5
TUT4	5
XPO5	5
POLD4	5
RETREG1	5
ATP6AP2	5
RSBN1L	5
NRP1	5
SHROOM2	5
CYB5B	5
PPP1R3B	5
CDC5L	5
CEBPD	5
ANXA2	5
SIX5	5
ETV5	5
CHURC1	5
CD163	5
CHI3L1	5
BHLHE40	5
C1RL	5
SETDB1	5
CHI3L2	5
S100A4	5

Supplementary Table 13: Genes that have PCC> 0.1 in N slides in the spatial GBM cohort. Only shown for $N >= 5$. Source data are provided in the Source Data File.



Supplementary Figure 9: **Comparison H&E quality of TCGA-BRCA slides to slides from spatial BRCA dataset.** a) Example of quality of TCGA-BRCA slide, with more zoomed-in version on the right. b-e) Four examples of quality from spatial BRCA slides.



Supplementary Figure 10: **Predictability of genes across > 2 slides in the spatial BRCA cohort.** Source data are provided in the Source Data File.

Supplementary Table 14: median PCC for top 100 genes for HE2RNA ('HE2') versus SEQUOIA ('SEQ') on the spatial BRCA dataset. Source data are provided in the Source Data File.

study	corr_HE2	corr_SEQ
SPA145	0.507	0.512
SPA143	0.507	0.496
SPA146	0.517	0.485
SPA148	0.508	0.456
SPA134	0.368	0.373
TENX53	0.301	0.362
SPA142	0.399	0.346
SPA140	0.440	0.341
TENX13	0.182	0.317
TENX39	0.254	0.301
TENX14	0.273	0.289
SPA138	0.403	0.286
TENX68	0.313	0.285
SPA117	-0.041	0.281
SPA133	0.318	0.280
SPA135	0.191	0.272
SPA60	0.292	0.270
SPA139	0.406	0.257
SPA149	-0.056	0.248
SPA141	0.389	0.221
SPA111	0.032	0.206
SPA115	0.091	0.198
SPA61	0.180	0.187
SPA116	0.021	0.187
SPA102	0.198	0.183
SPA53	0.121	0.182
SPA128	0.160	0.181
SPA131	0.137	0.176
SPA151	0.221	0.167
SPA107	0.074	0.162
SPA152	0.138	0.161
SPA153	0.248	0.160
SPA114	0.146	0.156
SPA75	0.268	0.150
SPA95	0.100	0.149
SPA76	0.189	0.148
SPA92	0.148	0.147
SPA58	0.048	0.146
SPA70	0.110	0.141
SPA94	0.114	0.140
SPA67	0.156	0.139
SPA62	0.063	0.136
SPA68	0.135	0.135
SPA100	0.109	0.134
SPA121	0.088	0.134
SPA91	0.114	0.133
SPA55	0.109	0.131
SPA56	0.097	0.131
SPA126	0.107	0.131
SPA90	0.118	0.129
SPA127	0.111	0.128
SPA101	0.113	0.127
SPA87	0.058	0.127
SPA69	0.087	0.124
SPA73	0.108	0.124
SPA98	0.168	0.122
SPA79	0.066	0.121
SPA51	0.069	0.120
SPA129	0.118	0.119
SPA77	0.174	0.117
SPA123	0.046	0.116
SPA154	0.103	0.116
SPA124	0.137	0.116
SPA125	0.158	0.116
SPA54	0.124	0.114
SPA88	0.047	0.112
SPA106	0.101	0.112
SPA85	0.150	0.110
SPA65	0.108	0.109
SPA109	0.040	0.109
SPA113	0.152	0.108
SPA105	0.063	0.105
SPA84	0.069	0.102
SPA72	0.081	0.099
SPA108	0.112	0.099
SPA89	0.095	0.099
SPA118	0.137	0.098
SPA86	0.054	0.095
SPA64	0.058	0.093
SPA96	0.139	0.089
SPA57	0.072	0.086
SPA122	0.130	0.085
SPA119	0.133	0.084
SPA130	0.075	0.082
SPA93	0.144	0.081
SPA110	0.119	0.075
SPA63	0.086	0.074
SPA104	0.063	0.071
SPA59	0.075	0.052
SPA83	-0.003	0.046
SPA82	-0.008	0.045
SPA81	-0.043	0.027
average	0.154	0.168

Supplementary Table 15: median EMD for top 100 genes for HE2RNA ('HE2') versus SEQUOIA ('SEQ') on the spatial BRCA dataset. Source data are provided in the Source Data File.

	emd_HE2	emd_SEQ
SPA129	0.098	0.140
SPA117	0.224	0.139
SPA61	0.169	0.135
SPA128	0.111	0.133
SPA89	0.097	0.132
SPA102	0.111	0.132
SPA109	0.119	0.129
SPA131	0.105	0.128
SPA154	0.162	0.127
SPA118	0.168	0.126
SPA116	0.203	0.123
SPA87	0.100	0.123
SPA113	0.110	0.123
SPA86	0.128	0.122
SPA130	0.105	0.122
SPA94	0.170	0.122
SPA111	0.105	0.120
SPA88	0.115	0.119
SPA104	0.106	0.119
SPA62	0.150	0.118
SPA72	0.131	0.117
SPA114	0.112	0.117
SPA92	0.125	0.116
SPA141	0.124	0.115
SPA152	0.147	0.115
SPA65	0.089	0.114
SPA107	0.139	0.113
SPA106	0.153	0.113
SPA115	0.108	0.112
SPA83	0.124	0.112
SPA82	0.138	0.112
SPA59	0.119	0.111
SPA119	0.161	0.111
SPA84	0.133	0.110
SPA64	0.104	0.110
SPA93	0.159	0.110
SPA75	0.096	0.109
SPA151	0.122	0.109
SPA108	0.087	0.109
SPA60	0.158	0.108
SPA121	0.180	0.107
SPA110	0.094	0.105
SPA95	0.180	0.105
SPA105	0.153	0.105
SPA91	0.140	0.105
SPA142	0.129	0.103
SPA81	0.141	0.102
SPA139	0.092	0.100
SPA148	0.113	0.100
SPA79	0.148	0.099
SPA146	0.100	0.098
SPA76	0.114	0.097
SPA63	0.093	0.097
SPA58	0.092	0.097
SPA90	0.146	0.097
SPA77	0.116	0.096
SPA153	0.136	0.095
SPA149	0.198	0.095
SPA67	0.118	0.094
SPA73	0.129	0.093
SPA85	0.129	0.093
SPA143	0.106	0.092
SPA134	0.102	0.092
SPA140	0.104	0.091
SPA135	0.105	0.091
SPA51	0.085	0.090
TENX68	0.125	0.089
SPA55	0.102	0.089
SPA138	0.107	0.089
SPA133	0.088	0.089
SPA122	0.125	0.089
SPA145	0.107	0.088
SPA96	0.124	0.088
SPA57	0.106	0.087
SPA54	0.095	0.087
SPA53	0.115	0.086
SPA56	0.085	0.085
SPA127	0.117	0.084
SPA68	0.109	0.084
TENX53	0.100	0.084
SPA69	0.075	0.084
SPA124	0.129	0.083
SPA70	0.083	0.083
SPA125	0.085	0.083
SPA101	0.118	0.083
SPA98	0.127	0.082
SPA123	0.137	0.079
SPA126	0.088	0.077
TENX13	0.107	0.077
TENX39	0.085	0.077
SPA100	0.087	0.077
TENX14	0.100	0.075
average	0.121	0.103

gene	N	gene	N	gene	N	gene	N	gene	N
YWHAZ	41	COL1A2	32	CD6	28	ZNF703	25	TSPAN13	22
DCN	40	ILF2	32	CD79A	28	GATA3	25	ENO1	22
HLA-DPB1	40	ERBB3	31	CST3	28	CSTB	25	YBX1	22
CYB5R1	40	TUBB	31	LTB	28	RRM2	25	TIMP2	21
TMSB10	40	LIMD2	31	SLC44A4	28	SNRPG	24	CD7	21
AEBP1	38	SDC1	31	SF3B3	28	ZG16B	24	RBM17	21
NASP	38	STMN1	31	IL32	27	AGR2	24	MSN	21
HDGF	37	HTRA1	31	HLA-DMB	27	AR	24	TAP2	21
PCNA	36	IGFBP4	30	GAPDH	27	HMGB2	24	TUBA1C	21
LRP1	36	SOD2	30	CD37	27	C1S	24	MCM3	21
TPI1	36	LAMB2	30	SERBP1	27	IVD	24	SFT2D2	21
COL1A1	36	CCL5	30	TLE3	27	REEP5	24	LRRC32	21
HMGA1	35	HSPD1	30	COTL1	27	MMP2	23	RPS7	21
LUM	35	C1R	30	TMPO	26	TOP2A	23	TMEM119	21
SF3B6	34	JAK3	29	TSC22D3	26	ITGB5	23	CXXC5	21
MCM7	34	EVL	29	RHOB	26	LDHB	23	PIM2	21
CCT5	34	CORO1A	29	POSTN	26	HLA-DQA1	23	CDT1	20
UQCRH	33	MRPS5	29	ANTXR1	26	CDC123	23	RANBP1	20
COL3A1	33	KDM4B	29	TFF3	26	RNF166	23	PFKP	20
MAGED2	33	CEBPB	29	HID1	26	ARSG	23	PSMB2	20
FAM174B	33	GSTP1	29	DEK	26	CD27	23	PAICS	20
ALG3	33	PLOD1	29	LOXL1	25	SLC39A6	23	BTG2	20
XBP1	33	EMILIN1	29	SNRPD1	25	KLF2	22	TPX2	20
CIRBP	33	WFS1	28	MCM6	25	TMEM123	22	CDC25B	20
SPARC	32	EHD2	28	CTSC	25	DCAF13	22	RRP1	20
SERPINF1	32	COL5A1	28	BIRC5	25	MRPL37	22	FOXM1	20
MXRA8	32	COL6A3	28	KPNA2	25	MLPH	22	MYBL2	20

Supplementary Table 16: Genes that have PCC$>$ 0.1 in N slides in the spatial BRCA cohort. Only shown for $N >= 20$. Source data are provided in the Source Data File.