

Supplementary Information

High Classification Accuracy Does Not Guarantee a Stable Gene Signature: A Selection-Stability Analysis of a Compact RNA-Seq Cancer Classifier

Table S1. Selection frequency of all 381 genes that entered the top-75 panel across the 60 resamples (10 outer cross-validation folds plus 50 repeated 80/20 stratified splits with different seeds). Genes are ordered by panel membership, then selection frequency, The final column indicates membership of the 75-gene panel reported in Table S2 below. The 15 genes selected in at least 80% of resamples ($\geq 48/60$) constitute the stable core and are shown in bold.

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
TCF21	60	1.000	yes
TRPS1	59	0.983	yes
HNF1B	58	0.967	yes
SFTPB	57	0.950	yes
SFTA3	56	0.933	yes
LMX1B	54	0.900	yes
C20ORF56	52	0.867	yes
FOXF1	52	0.867	yes
GATA3	51	0.850	yes
SFTPA1	51	0.850	yes
TFAP2A	49	0.817	yes
HKDC1	49	0.817	yes
SFTPA2	49	0.817	yes
PRLR	48	0.800	yes
MCTP2	47	0.783	yes
NKX2-1	46	0.767	yes
SLC45A3	46	0.767	yes
SCGB2A2	45	0.750	yes
EFHD1	45	0.750	yes
FZD5	45	0.750	yes
KLK3	45	0.750	yes
IHH	43	0.717	yes
C8ORF85	43	0.717	yes
C9ORF167	43	0.717	yes
EN1	42	0.700	yes
SHH	42	0.700	yes
SCTR	42	0.700	yes
KLK2	42	0.700	yes
SFTA2	40	0.667	yes

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
PAX8	39	0.650	yes
NKX3-1	38	0.633	yes
HOXB13	38	0.633	yes
ACPP	35	0.583	yes
OVOL2	33	0.550	yes
SCGB3A2	33	0.550	yes
TMEFF2	33	0.550	yes
HSD17B11	32	0.533	yes
CDH16	32	0.533	yes
EMX2OS	31	0.517	yes
ACSM3	31	0.517	yes
BEND4	30	0.500	yes
GALNT12	30	0.500	yes
SLC6A13	28	0.467	yes
STEAP2	28	0.467	yes
RDH11	24	0.400	yes
SLC3A1	24	0.400	yes
SLC28A1	24	0.400	yes
HHLA2	22	0.367	yes
C17ORF93	22	0.367	yes
PRAC	21	0.350	yes
C15ORF21	19	0.317	yes
ACSM2B	19	0.317	yes
CAPN5	17	0.283	yes
SLC17A3	16	0.267	yes
ADIPOQ	14	0.233	yes
PAGE4	14	0.233	yes
SLC22A2	14	0.233	yes
SFTPD	14	0.233	yes
ALDH1A3	13	0.217	yes
FOLH1	13	0.217	yes
DMGDH	13	0.217	yes
IGSF9	13	0.217	yes
TMEM140	10	0.167	yes
MET	9	0.150	yes
AGXT2	8	0.133	yes
PAX2	8	0.133	yes

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
KLK15	7	0.117	yes
DHDPSL	5	0.083	yes
SEL1L3	4	0.067	yes
LMO7	3	0.050	yes
TESC	3	0.050	yes
C9ORF152	3	0.050	yes
GPA33	3	0.050	yes
HOXA10	1	0.017	yes
ARSE	1	0.017	yes
NAPSA	52	0.867	no
SFTPC	45	0.750	no
CHRNA2	41	0.683	no
CPNE4	40	0.667	no
FXVD2	40	0.667	no
ST6GALNAC1	35	0.583	no
PCA3	35	0.583	no
KLK4	32	0.533	no
ENPP1	32	0.533	no
POU3F3	32	0.533	no
FLJ45983	32	0.533	no
C20ORF151	30	0.500	no
SLC4A4	29	0.483	no
C14ORF105	29	0.483	no
GCNT3	27	0.450	no
TARP	26	0.433	no
OR51E2	26	0.433	no
EMX2	26	0.433	no
PBX1	25	0.417	no
LOC100130933	25	0.417	no
THRSP	25	0.417	no
GAL3ST1	25	0.417	no
KLKP1	25	0.417	no
SALL1	23	0.383	no
IRX5	23	0.383	no
KCNJ16	23	0.383	no
ESRP1	22	0.367	no
ACSM2A	22	0.367	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
TMPRSS2	22	0.367	no
NR1H4	21	0.350	no
ARHGAP6	21	0.350	no
ENPEP	20	0.333	no
MREG	20	0.333	no
PPAP2A	20	0.333	no
GDEP	20	0.333	no
CALML5	19	0.317	no
DPP4	19	0.317	no
FOXA2	19	0.317	no
ROS1	18	0.300	no
GRHL2	18	0.300	no
KCNN2	18	0.300	no
TBC1D9	17	0.283	no
PPP3CA	17	0.283	no
AMN	17	0.283	no
UGT2B7	17	0.283	no
CDH6	16	0.267	no
LMAN1L	16	0.267	no
BHMT	16	0.267	no
TRPM8	15	0.250	no
HOXA13	15	0.250	no
GALNT14	15	0.250	no
HNF4A	15	0.250	no
CUX2	14	0.233	no
SLC30A4	14	0.233	no
MUC21	14	0.233	no
NAT8	14	0.233	no
CUBN	13	0.217	no
GALNT6	13	0.217	no
SLC34A2	13	0.217	no
GATA2	13	0.217	no
ESR1	13	0.217	no
CNKSR1	12	0.200	no
HAND2	12	0.200	no
CAMKK2	12	0.200	no
CRYL1	11	0.183	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
RAB25	11	0.183	no
FEV	11	0.183	no
GLYAT	11	0.183	no
PRODH2	11	0.183	no
FBXO17	10	0.167	no
USH1C	10	0.167	no
TMEM174	10	0.167	no
ACSM5	10	0.167	no
HOXA11AS	10	0.167	no
ALPK2	10	0.167	no
CLRN3	10	0.167	no
SHOX2	9	0.150	no
SEMA5B	9	0.150	no
? 553137	9	0.150	no
C1ORF172	9	0.150	no
PRKCA	9	0.150	no
SLC5A10	8	0.133	no
PABPC1L2A	8	0.133	no
C11ORF41	8	0.133	no
GPX3	8	0.133	no
COL23A1	8	0.133	no
POTEG	8	0.133	no
DHRS7	8	0.133	no
TBL1Y	8	0.133	no
HOMER2	7	0.117	no
BIN1	7	0.117	no
SLC22A3	7	0.117	no
SLC13A1	7	0.117	no
METTL7B	7	0.117	no
ASPA	7	0.117	no
CRABP2	7	0.117	no
TFAP2B	7	0.117	no
STK32A	7	0.117	no
C6ORF223	7	0.117	no
DPCR1	6	0.100	no
STEAP1	6	0.100	no
C4BPA	6	0.100	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
EPB41L4B	6	0.100	no
SLC37A4	6	0.100	no
CREB3L4	6	0.100	no
APOH	6	0.100	no
SLC17A2	6	0.100	no
CHRM1	6	0.100	no
SLC17A1	6	0.100	no
ARHGAP24	6	0.100	no
EPDR1	5	0.083	no
PRSS22	5	0.083	no
GLI3	5	0.083	no
FXYP3	5	0.083	no
SCGB1A1	5	0.083	no
HOXA11	5	0.083	no
C19ORF77	5	0.083	no
C18ORF45	5	0.083	no
PKP3	5	0.083	no
MFAP5	5	0.083	no
SLC2A2	5	0.083	no
TMEM84	5	0.083	no
LOC100131551	4	0.067	no
AMACR	4	0.067	no
TSPAN8	4	0.067	no
CDHR5	4	0.067	no
KCNE3	4	0.067	no
PVRL4	4	0.067	no
SLC6A12	4	0.067	no
HOXC8	4	0.067	no
SI	4	0.067	no
SPDEF	4	0.067	no
CYORF15B	4	0.067	no
ADRB2	4	0.067	no
CLEC18C	4	0.067	no
MOGAT3	4	0.067	no
ABCC4	4	0.067	no
CDH17	4	0.067	no
MUC4	4	0.067	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
MYB	4	0.067	no
PYCR1	4	0.067	no
CDK18	4	0.067	no
SYT7	4	0.067	no
CEACAM6	4	0.067	no
ATP11A	4	0.067	no
NWD1	4	0.067	no
C12ORF27	4	0.067	no
NDUFA4L2	4	0.067	no
NPY	3	0.050	no
PDZD3	3	0.050	no
CPLX3	3	0.050	no
RFPL2	3	0.050	no
ANKRD30A	3	0.050	no
LRRC19	3	0.050	no
CHIA	3	0.050	no
MBIP	3	0.050	no
PDE9A	3	0.050	no
RCBTB2	3	0.050	no
FAM150A	3	0.050	no
GUCY2C	3	0.050	no
ESM1	3	0.050	no
LRRC15	3	0.050	no
PKD2	3	0.050	no
B3GNT6	3	0.050	no
RBP5	3	0.050	no
PRKD1	3	0.050	no
N4BP3	3	0.050	no
RPS4Y1	3	0.050	no
MST1R	3	0.050	no
CDX1	3	0.050	no
TRPV6	3	0.050	no
RAB3B	3	0.050	no
KCNIP4	3	0.050	no
TMEM30B	2	0.033	no
ACAD11	2	0.033	no
SORD	2	0.033	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
GPR35	2	0.033	no
DES	2	0.033	no
CTSE	2	0.033	no
SLC39A5	2	0.033	no
BHMT2	2	0.033	no
TINAG	2	0.033	no
FAM198A	2	0.033	no
TJP3	2	0.033	no
ANKS6	2	0.033	no
ASB5	2	0.033	no
ABCC3	2	0.033	no
SLC22A11	2	0.033	no
NAAA	2	0.033	no
SIX4	2	0.033	no
BCL2L15	2	0.033	no
PLA2G12B	2	0.033	no
USP9Y	2	0.033	no
HNF1A	2	0.033	no
TMEM92	2	0.033	no
SLC15A4	2	0.033	no
TFDP3	2	0.033	no
ATCAY	2	0.033	no
UGT2A3	2	0.033	no
WDR69	2	0.033	no
GBA3	2	0.033	no
LDLRAD3	2	0.033	no
LGALS4	2	0.033	no
FLT1	2	0.033	no
MYO7B	2	0.033	no
SLCO4C1	2	0.033	no
CILP	2	0.033	no
SPOCK3	2	0.033	no
PLUNC	2	0.033	no
DDX3Y	2	0.033	no
SLC39A6	2	0.033	no
SFTA1P	2	0.033	no
ANO7	2	0.033	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
MOSC1	2	0.033	no
TPM3	2	0.033	no
ZNF350	2	0.033	no
ISX	2	0.033	no
SYCP2	2	0.033	no
SLC17A4	2	0.033	no
SCEL	2	0.033	no
PDZK1	2	0.033	no
A1CF	2	0.033	no
MEP1A	2	0.033	no
S100A14	1	0.017	no
C2ORF89	1	0.017	no
ZNF761	1	0.017	no
SLC25A42	1	0.017	no
CYORF15A	1	0.017	no
CABYR	1	0.017	no
HOPX	1	0.017	no
STX19	1	0.017	no
ANGPTL4	1	0.017	no
BMP2	1	0.017	no
DAB2	1	0.017	no
CDX2	1	0.017	no
NR0B2	1	0.017	no
ENAM	1	0.017	no
MMP26	1	0.017	no
MASP1	1	0.017	no
SERPINB1	1	0.017	no
TMEM163	1	0.017	no
FAM149A	1	0.017	no
MYOM3	1	0.017	no
LOC388387	1	0.017	no
FAAH	1	0.017	no
TFAP2C	1	0.017	no
DPYS	1	0.017	no
FERMT1	1	0.017	no
NBLA00301	1	0.017	no
NOX1	1	0.017	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
AGR2	1	0.017	no
ACMSD	1	0.017	no
CAMK1	1	0.017	no
ERBB4	1	0.017	no
SMARCD2	1	0.017	no
ARFGAP3	1	0.017	no
GNMT	1	0.017	no
SNRPN	1	0.017	no
PRUNE2	1	0.017	no
XIST	1	0.017	no
ANKS4B	1	0.017	no
EHF	1	0.017	no
CECR6	1	0.017	no
TBX3	1	0.017	no
CLEC18A	1	0.017	no
FOXA1	1	0.017	no
BBOX1	1	0.017	no
ALDH3B2	1	0.017	no
VAV3	1	0.017	no
TTY15	1	0.017	no
C1ORF116	1	0.017	no
NRIP2	1	0.017	no
ACADL	1	0.017	no
TMSB4Y	1	0.017	no
PIK3R3	1	0.017	no
EIF1AY	1	0.017	no
TRIM15	1	0.017	no
ADAM12	1	0.017	no
TRIM10	1	0.017	no
GMPR	1	0.017	no
UPB1	1	0.017	no
AP1M2	1	0.017	no
DNAJB13	1	0.017	no
AQP4	1	0.017	no
KIF5C	1	0.017	no
TRPM4	1	0.017	no
KDM5D	1	0.017	no

Gene	Selection frequency (/60)	Frequency	In reported 75-gene panel
PLEKHF2	1	0.017	no
CYP2S1	1	0.017	no
ACSL5	1	0.017	no
TRIM55	1	0.017	no
LPCAT1	1	0.017	no
RABEP1	1	0.017	no
EPN3	1	0.017	no
CELSR2	1	0.017	no
TMC5	1	0.017	no
MSMB	1	0.017	no
ERN2	1	0.017	no
EPHB3	1	0.017	no
SLC2A12	1	0.017	no
SLC5A6	1	0.017	no
ATP10B	1	0.017	no
PDE4C	1	0.017	no
HOXA9	1	0.017	no
SLC44A4	1	0.017	no
ANXA4	1	0.017	no

Total distinct genes entering the top 75 at least once: 381. Reported panel genes falling below the 80% threshold: 61 of 75.

Table S2. The reported 75-gene panel, ranked by random-forest importance, with selection frequency across the 60 resamples. Genes selected in at least 80% of resamples, which constitute the stable core, are shown in bold.

Rank	Gene	Selected (/60)	Frequency	Stable core
1	RDH11	24	0.40	no
2	TRPS1	59	0.98	yes
3	TCF21	60	1.00	yes
4	ACSM2B	19	0.32	no
5	HNF1B	58	0.97	yes
6	C8ORF85	43	0.72	no
7	SFTA3	56	0.93	yes
8	SCGB2A2	45	0.75	no
9	MCTP2	47	0.78	no
10	IHH	43	0.72	no
11	FOXF1	52	0.87	yes
12	GALNT12	30	0.50	no

Rank	Gene	Selected (/60)	Frequency	Stable core
13	SLC45A3	46	0.77	no
14	NKX2-1	46	0.77	no
15	OVOL2	33	0.55	no
16	MET	9	0.15	no
17	TFAP2A	49	0.82	yes
18	KLK2	42	0.70	no
19	HOXB13	38	0.63	no
20	PAX8	39	0.65	no
21	SCGB3A2	33	0.55	no
22	C9ORF152	3	0.05	no
23	SCTR	42	0.70	no
24	BEND4	30	0.50	no
25	SLC6A13	28	0.47	no
26	FZD5	45	0.75	no
27	KLK3	45	0.75	no
28	EFHD1	45	0.75	no
29	SLC22A2	14	0.23	no
30	C20ORF56	52	0.87	yes
31	C9ORF167	43	0.72	no
32	PAX2	8	0.13	no
33	SFTPA2	49	0.82	yes
34	SLC17A3	16	0.27	no
35	SFTPB	57	0.95	yes
36	HHLA2	22	0.37	no
37	GATA3	51	0.85	yes
38	KLK15	7	0.12	no
39	EMX2OS	31	0.52	no
40	HKDC1	49	0.82	yes
41	SLC3A1	24	0.40	no
42	SFTPA1	51	0.85	yes
43	SFTPD	14	0.23	no
44	CDH16	32	0.53	no
45	SHH	42	0.70	no
46	LMO7	3	0.05	no
47	PRAC	21	0.35	no
48	SLC28A1	24	0.40	no
49	ACPP	35	0.58	no

Rank	Gene	Selected (/60)	Frequency	Stable core
50	C15ORF21	19	0.32	no
51	TMEFF2	33	0.55	no
52	LMX1B	54	0.90	yes
53	DMGDH	13	0.22	no
54	EN1	42	0.70	no
55	CAPN5	17	0.28	no
56	IGSF9	13	0.22	no
57	TMEM140	10	0.17	no
58	NKX3-1	38	0.63	no
59	DHDPSL	5	0.08	no
60	AGXT2	8	0.13	no
61	ADIPOQ	14	0.23	no
62	FOLH1	13	0.22	no
63	ACSM3	31	0.52	no
64	HSD17B11	32	0.53	no
65	PAGE4	14	0.23	no
66	ARSE	1	0.02	no
67	ALDH1A3	13	0.22	no
68	SFTA2	40	0.67	no
69	STEAP2	28	0.47	no
70	C17ORF93	22	0.37	no
71	GPA33	3	0.05	no
72	TESC	3	0.05	no
73	SEL1L3	4	0.07	no
74	HOXA10	1	0.02	no
75	PRLR	48	0.80	yes

Artificial Intelligence (AI) Disclosure

All aspects of this study, including its conceptualization, research questions, study design, methodology, data acquisition, feature-selection strategy, statistical analyses, interpretation of results, and conclusions, were developed independently by the author. The manuscript was written and revised entirely by the author without the use of generative artificial intelligence for drafting, editing, or scientific interpretation.

The author accepts full responsibility for the integrity, originality, accuracy, and scientific content of this work.