

Supplementary Tables

Table S1 The detailed information of patients analyzed in manuscript.

Cancer type	Abrr	Gene expression from TCGA Sizes	Copy number variation from TCGA Sizes	LncRNA expression from TANRIC Sizes	Clinical data from cBioPortal Sizes
Stomach adenocarcinoma	STAD	415	441	285	443
Prostate adenocarcinoma	PRAD	297	331	374	499
Skin cutaneous melanoma	SKCM	254	469	226	478
Brain lower grade glioma	LGG	469	463	486	530
Breast invasive carcinoma	BRCA	1102	1080	837	1105
Ovarian serous cystadenocarcinoma	OV	266	572	412	607
Lung squamous cell carcinoma	LUAD	490	494	488	522
Lung adenocarcinoma	LUSC	490	490	220	504
Head and neck squamous cell carcinoma	HNSC	451	452	426	530
Glioblastoma multiforme	GBM	169	535	154	607
Colon adenocarcinoma	COAD	192	429	157	453
Rectal adenocarcinoma	READ	72	162	71	162

(As described in previously TCGA studies, COAD and READ were merged and formed one cancer sample set COAD/READ (CR) in subsequent analysis.)

Table S2. The number of copy number altered PCGs and lncRNAs across cancers.

Cancer Type	PCG(amp)	PCG(del)	lncRNA(amp)	lncRNA(del)
BRCA	1914	211	293	20
GBM	99	15	25	9
HNSC	945	168	145	19
LUAD	1760	428	267	51
LUSC	1613	170	283	33
PRAD	104	969	35	210
LGG	507	421	134	125
STAD	1236	156	147	13

CR	545	141	34	8
SKCM	775	190	135	29
OV	3075	228	496	29

amp, amplification; del, deletion

Table S3 The number of mutually exclusive modules and driver lncRNAs across cancers.

Abbreviation	Sample sizes	Candidate lncRNA number	Driver lncRNA number	Module number
STAD	441	160	7	15
PRAD	331	245	45	60
SKCM	469	164	2	6
LGG	463	259	41	35
BRCA	1080	313	88	282
OV	572	525	133	1547
LUAD	494	318	69	379
LUSC	490	316	48	97
HNSC	452	164	27	55
GBM	535	34	14	54
CR	591	42	9	25

Table S4 The overlap of candidate driver lncRNAs identified by MiTranscriptome and our method.

Cancer types	MiTranscriptome	Our results	Overlap	P value
HNSC	382	27	3	3.72e-02
LUSC	497	48	9	4.37e-05
LUAD	481	69	7	9.76e-03
PRAD	412	45	6	2.02e-03
STAD	64	7	1	3.19e-02
BRCA	521	88	5	2.36e-01
Total	1577	238	71	6.28e-15

Table S5 The overlap of candidate driver lncRNAs identified by CRISPRi and our method.

Cancer types	CRISPRi	Our results	Overlap	P value
BRCA	55	88	3	5.10e-03
GBM	37	14	0	1.0e-00
Total	286	378	27	1.97e-08

Table S6 Cox proportional hazards models of driver lncRNAs across cancers. # indicated not significant.

Overall Survival Analysis				
Univariate Analysis			Multivariate Analysis	
Ensembl ID	P value	Hazard ratio	P value	Hazard ratio
head and neck squamous cell carcinoma				
ENSG00000246889	2.50E-02	1.11	3.50E-02	1.11
ENSG00000254604	2.90E-02	1.11	3.80E-02	1.10
ENSG00000254605	1.90E-02	1.12	3.00E-02	1.11
lung adenocarcinoma				
ENSG00000246695	2.60E-02	1.36	#	#
ENSG00000256894	1.90E-02	1.37	4.40E-02	1.33
ovarian serous cystadenocarcinoma				
ENSG00000254024	1.07E-02	1.18	7.18E-03	1.19
ENSG00000226686	1.89E-04	1.26	1.70E-03	1.22
ENSG00000267254	4.66E-03	1.20	2.08E-02	1.16
ENSG00000267470	2.30E-03	1.19	9.57E-03	1.16
ENSG00000246263	7.42E-03	1.20	6.70E-03	1.20
ENSG00000236144	1.38E-03	1.23	2.45E-03	1.22
ENSG00000247081	1.76E-03	1.23	9.50E-04	1.24
ENSG00000260708	2.21E-02	1.31	3.77E-02	1.28
ENSG00000267014	5.05E-03	1.16	1.82E-02	1.13
ENSG00000225975	2.87E-03	1.21	1.18E-02	1.18
ENSG00000261824	8.56E-04	1.24	5.96E-03	1.19
ENSG00000267213	1.44E-02	1.15	1.33E-02	1.15
ENSG00000266935	1.47E-03	1.22	6.63E-03	1.19
ENSG00000267053	2.95E-03	1.22	1.00E-02	1.20
ENSG00000267439	3.95E-05	1.31	1.15E-04	1.29

ENSG00000267575	3.59E-04	1.24	3.49E-03	1.19
ENSG00000232677	2.95E-03	1.22	1.00E-02	1.20
ENSG00000180458	2.99E-03	1.19	1.23E-02	1.16
ENSG00000233527	3.66E-03	1.20	1.48E-02	1.17
ENSG00000251136	3.51E-04	1.32	7.56E-05	1.36
ENSG00000260032	4.77E-03	1.26	#	#
ENSG00000270127	4.47E-03	0.76	3.41E-02	0.81
ENSG00000257084	3.69E-02	1.16	#	#
ENSG00000261437	1.10E-03	1.27	7.17E-04	1.29
ENSG00000205885	4.63E-02	1.15	#	#
ENSG00000253217	2.13E-02	1.16	1.17E-02	1.18
ENSG00000256967	5.00E-02	1.15	#	#
ENSG00000253854	1.52E-03	1.28	1.19E-03	1.28
ENSG00000259424	7.55E-03	0.78	#	#
ENSG00000259583	7.63E-03	0.78	4.77E-02	0.83
ENSG00000232386	7.63E-03	0.78	4.77E-02	0.83
ENSG00000267381	3.46E-03	1.18	1.43E-02	1.15
ENSG00000266916	3.13E-03	1.19	1.33E-02	1.16
ENSG00000250899	3.09E-02	1.16	#	#
ENSG00000177406	4.75E-02	1.14	#	#
ENSG00000266930	1.01E-02	1.16	9.88E-03	1.16
ENSG00000253878	4.16E-03	1.24	3.46E-03	1.25

breast invasive carcinoma

ENSG00000228801	3.38E-02	1.27	#	#
ENSG00000251136	1.15E-02	1.29	#	#

brain lower grade glioma

ENSG00000241764	9.74E-03	1.74	#	#
ENSG00000231856	3.22E-03	0.39	4.59E-02	0.46
ENSG00000217576	3.19E-03	0.39	4.54E-02	0.46
ENSG00000234418	3.51E-02	1.55	#	#
ENSG00000228031	3.18E-02	1.52	#	#
ENSG00000238099	6.57E-03	0.39	#	#
ENSG00000260917	3.33E-16	0.12	1.55E-09	0.16
ENSG00000203497	1.11E-16	0.11	1.02E-07	0.18
ENSG00000267289	4.37E-02	2.66	#	#
ENSG00000224138	1.75E-02	1.64	#	#
ENSG00000269386	4.24E-02	2.67	#	#
ENSG00000225559	4.42E-02	1.48	#	#
ENSG00000227869	9.61E-03	1.75	#	#

lung squamous cell carcinoma

ENSG00000224046	2.38E-02	0.66	2.28E-02	0.64
ENSG00000182165	2.64E-02	0.67	2.57E-02	0.65

ENSG00000253414	9.07E-03	0.77	5.29E-03	0.76
ENSG00000264868	1.40E-02	0.65	1.58E-02	0.64
ENSG00000232019	1.88E-02	0.66	2.17E-02	0.65
ENSG00000242687	2.82E-02	0.68	2.94E-02	0.67
glioblastoma multiforme				
ENSG00000234707	1.15E-02	1.08	#	#
Disease Free Survival Analysis				
Univariate Analysis			Multivariate Analysis	
Ensembl ID	P value	Hazard ratio	P value	Hazard ratio
head and neck squamous cell carcinoma				
ENSG00000255310	3.80E-02	0.57	2.50E-02	0.53
ENSG00000269918	3.80E-02	0.57	2.50E-02	0.53
ENSG00000246089	2.70E-02	0.54	1.70E-02	0.51
ovarian serous cystadenocarcinoma				
ENSG00000247081	3.20E-03	1.21	9.31E-03	1.18
ENSG00000254353	8.50E-03	1.47	1.45E-02	1.44
ENSG00000267213	4.46E-02	1.13	1.45E-02	1.16
ENSG00000269983	8.50E-03	1.47	1.45E-02	1.44
ENSG00000251136	2.70E-03	1.26	3.73E-03	1.24
ENSG00000260032	2.54E-04	1.36	7.63E-03	1.27
ENSG00000270127	1.88E-02	0.81	#	#
ENSG00000270017	3.55E-02	0.82	#	#
ENSG00000261437	5.26E-03	1.23	1.60E-02	1.19
ENSG00000253854	1.06E-02	1.22	2.65E-02	1.18
ENSG00000259424	1.91E-02	0.81	#	#
ENSG00000259583	3.51E-02	0.83	#	#
ENSG00000232386	3.51E-02	0.83	#	#
ENSG00000242622	4.04E-02	1.24	#	#
ENSG00000232453	2.40E-02	1.33	1.86E-02	1.35
ENSG00000253878	3.02E-02	1.17	#	#
brain lower grade glioma				
ENSG00000241764	8.38E-06	2.15	3.37E-04	1.95
ENSG00000234418	4.81E-06	2.20	1.96E-04	2.01
ENSG00000228031	1.14E-05	2.11	7.00E-04	1.89
ENSG00000238099	6.43E-03	0.41	#	#
ENSG00000204934	9.09E-06	2.17	1.33E-04	2.05
ENSG00000260917	2.21E-11	0.20	7.94E-08	0.24
ENSG00000261342	3.04E-02	2.45	#	#
ENSG00000203497	0.00E+00	0.13	5.51E-11	0.17
ENSG00000267299	3.99E-02	2.54	#	#

ENSG00000267289	3.42E-02	2.61	#	#
ENSG00000270157	1.35E-05	2.10	4.50E-04	1.93
ENSG00000224138	2.94E-06	2.24	2.64E-04	1.99
ENSG00000269386	3.78E-02	2.57	#	#
ENSG00000261455	4.74E-06	2.21	1.45E-04	2.04
ENSG00000182648	2.60E-06	2.23	#	#
ENSG00000228775	1.35E-05	2.10	4.50E-04	1.93
ENSG00000225559	6.75E-06	2.14	4.37E-04	1.93
ENSG00000240449	9.09E-06	2.17	1.33E-04	2.05
ENSG00000227869	8.18E-06	2.18	3.78E-04	1.96
ENSG00000260231	1.38E-05	2.10	4.72E-04	1.93
ENSG00000239377	9.09E-06	2.17	1.33E-04	2.05
lung squamous cell carcinoma				
ENSG00000253414	8.37E-03	0.71	3.52E-03	0.68
ENSG00000265943	2.10E-04	1.89	5.26E-04	1.85
ENSG00000266729	6.75E-03	1.64	6.10E-03	1.66
ENSG00000246889	2.32E-02	1.18	#	#
ENSG00000244124	1.78E-02	0.71	2.57E-02	0.72
ENSG00000248932	5.72E-03	0.64	8.98E-03	0.65

Table S7 Univariate and multivariate Cox proportional hazards analysis of disease free survival for 429 brain lower grade glioma patients according to AC000123.4 copy number status.

Subgroup	Univariate Analysis		Multivariate Analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
All patients (N=429)				
Age, modeled as a continuous variable	1.02(1.01-1.04)	1.52e-03	1.02(1.01-1.03)	7.43e-03
Gender				
Male vs. Female	0.85(0.61-1.20)	0.36	0.84(0.60-1.19)	0.33
Histological diagnosis				
Oligoastrocytoma vs. Astrocytoma	0.56(0.35-0.89)	1.35e-02	0.56(0.35-0.88)	1.32e-02
Oligodendroglioma vs. Astrocytoma	0.64(0.44-0.94)	2.15e-02	0.66(0.45-0.99)	4.39e-02

Grade				
G3 vs. G2	1.55(1.10-2.18)	1.31e-02	1.40(0.98-2.00)	6.76e-02
Continuous copy number				
AC000123.4	2.12(1.49-3.02)	3.06e-05	1.99(1.38-2.88)	2.44e-04

(Abbreviations: CI, confidence interval.)

Table S8 Univariate and multivariate Cox proportional hazards analysis of disease free survival for 429 brain lower grade glioma patients according to AC000123.4 expression levels.

Subgroup	Univariate Analysis		Multivariate Analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
All patients (N=429)				
Age, modeled as a continuous variable	1.02(1.01-1.04)	1.42e-03	1.02(1.01-1.04)	9.76e-04
Gender				
Male vs. Female	0.84(0.60-1.18)	0.32	0.83(0.59-1.17)	0.29
Histological diagnosis				
Oligoastrocytoma vs. Astrocytoma	0.57(0.36-0.90)	1.58e-02	0.71(0.44-1.14)	0.15
Oligodendroglioma vs. Astrocytoma	0.65(0.45-0.95)	2.73e-02	0.73(0.48-1.10)	0.13
Grade				
G3 vs. G2	1.54(1.09-2.17)	1.47e-02	1.30(0.90-1.87)	0.16
Continuous expression				
AC000123.4	4.41(2.46-7.80)	5.92e-07	3.46(1.92-6.22)	3.44e-05

(Abbreviations: CI, confidence interval.)

Table S9 Univariate and multivariate Cox proportional hazards analysis of disease free survival for 399 brain lower grade glioma patients without EGFR amplification according to AC000123.4 copy number status.

Subgroup	Univariate Analysis	Multivariate Analysis
----------	---------------------	-----------------------

	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
All patients (N=399)				
Age, modeled as a continuous variable	1.01(1-1.02)	0.04	1.01(0.99-1.02)	0.06
Gender				
Male vs. Female	0.84(0.59-1.19)	0.34	0.83(0.59-1.18)	0.30
Histological diagnosis				
Oligoastrocytoma vs. Astrocytoma	0.51 (0.31-0.83)	0.007	0.47(0.29-0.77)	0.002
Oligodendroglioma vs. Astrocytoma	0.70(0.48-1.03)	0.071	0.65(0.44-0.96)	0.03
Continuous copy number				
AC000123.4	1.86(1.26-2.76)	0.0018	1.91 (1.27-2.86)	0.0017

(Abbreviations: CI, confidence interval.)

Table S10 The miRNAs from miRBase v22 that located in the genomic regions of candidate driver lncRNAs.

Pri-miRNA	miRNA	Driver miRNA evidence	LncRNA name	Driver lncRNA evidence	Cancer type
hsa-mir-31	hsa-miR-31-3p,	PMID: 22264793; 19524507;	MIR31HG	PMID:25908244; 28639619;	LUSC
	hsa-miR-31-5p	29051494; 24582497		27334046; 26549028	
hsa-mir-570	hsa-miR-570-3p,	PMID:29874097;26084609;	LINC00969		HNSC
	hsa-miR-570-5p	29795113			
hsa-mir-1204	hsa-mir-1204	PMID:29555976;25756509;	PVT1	PMID:26481356;23175614;	HNSC, LUAD, OV
		22110125; 18314482		25043044;29752292;29731168	
hsa-mir-1205	hsa-mir-1205	PMID:29852168;	PVT1	PMID:26481356;23175614;	HNSC, LUAD, OV
		doi:10.1158/1538-7445.AM2015-187		25043044;29752292;29731168	
hsa-mir-1206	hsa-miR-1206	PMID:28628559;23077621	PVT1	PMID:26481356;23175614;	HNSC, LUAD, OV

				25043044;29752292;29731168	
hsa-mir-1207	hsa-miR-1207-3p,	PMID:28235236;26337084;	PVT1	PMID:26481356;23175614;	HNSC, LUAD, OV
	hsa-miR-1207-5p	29387230; 27852979		25043044;29752292;29731168	
hsa-mir-200c	hsa-miR-200c-3p,	PMID:20579395;17804704;	U47924.29		OV,BRCA
	hsa-miR-200c-5p	28436968; 19665978			
hsa-mir-141	hsa-miR-141-3p,	PMID:24285464;22479552;	U47924.29		OV,BRCA
	hsa-miR-141-5p	24000293; 28112170			
hsa-mir-924	hsa-miR-924	PMID:29753758;27525335	LINC00669		CR
hsa-mir-5583-2	hsa-miR-5583-5p,		LINC00669		CR
	hsa-miR-5583-3p				
hsa-mir-24-2	hsa-miR-24-3p,	PMID:25943634;27625395;	MIR24-2		OV
	hsa-miR-24-2-5p	22911661; 21463514			
hsa-mir-4999	Has-miR-4999-5p,		RAB11B-AS1	PMID: 29928484	LGG
	hsa-miR-4999-3p				

Supplementary Figure legends

Figure S1 The distribution of semantic similarity score between candidate GO terms and hallmark-associated GO terms. Red arrows indicate Rest-similarity score that was calculated between each curated GO term of a certain hallmark and the rest of other curated GO terms of this hallmark.

Figure S2 Heatmaps show enrichment of hallmark-associated PCGs in additional pathways that were downloaded from Synapse (syn1741407). The P-value were calculated using the hypergeometric test.

Figure S3 Construction of mutually exclusive networks from binary copy number alteration profiles. For each pair of genes, a 2x2 contingency table was built to investigate their relationship of CAN. Gene pair g1 and g2 were significantly mutually exclusive by a hypergeometric test ($p < 0.05$). The co-occurrence of gene pair g3 and g4 was never observed in any samples.

Figure S4 The lncRNAs and protein-coding genes in the top 10 most significant wide peak regions in BRCA, HNSC, LUAD, LUSC, PRAD, LGG, STAD, CR, SKCM and OV. The numbers of protein-coding genes (left) and lncRNAs (right) in each peak are indicated in parentheses. Pie chart of each peak shows the proportion of PCG (blue) and lncRNA (green).

Figure S5 The basic statistics of mutually exclusive module across 11 major cancer types. (A) Stacked bars indicate the distribution of module size. (B) Box plot shows frequencies of CNA for PCGs from hallmark-associated mutually exclusive modules (left) and other PCGs (right) (*: $p < 0.05$, permutation test).

Figure S6 The number of predicted lncRNA from real data (red arrow) and random CNAs profiles (left) and expression profiles (right) across 11 cancer types (curve) (*: $p < 0.05$, permutation test).

Figure S7 QQ-plot of p-values that evaluate the hallmark associated function of mutually exclusive modules. The proportion of p-values are shown in the bottom.

Figure S8 The overlap of candidate driver genes with Cancer Gene Census (CGC). Statistical significance is assessed by the hypergeometric test with $p < 0.05$.

Figure S9 Proportion of lncRNAs from modules lie within early-replicating regions. (A-C) Bar plots show the difference of the fraction of lncRNAs from modules and other lncRNAs that located in early replicating regions derived from all of the 8 cell lines, IMR90 and MCF-7, respectively. The ME lncRNAs in figures indicate lncRNAs from mutually exclusive modules.

Figure S10 Heatmaps show cancer-associated hallmarks affected by cancer-specific driver lncRNAs.

Figure S11 Heatmap shows cancer-associated hallmarks affected by PVT1 in HNSC, LUAD and OV.

Figure S12 Kaplan–Meier plots of DFS of LGG patients grouped by copy number status of lncRNA AC000123.4. The newly added clinical data was downloaded on 1 July, 2016 from cBioPortal website that includes 35 disease-free patients and 11 disease-progressed patients.

Supplementary Figures

Figure S1

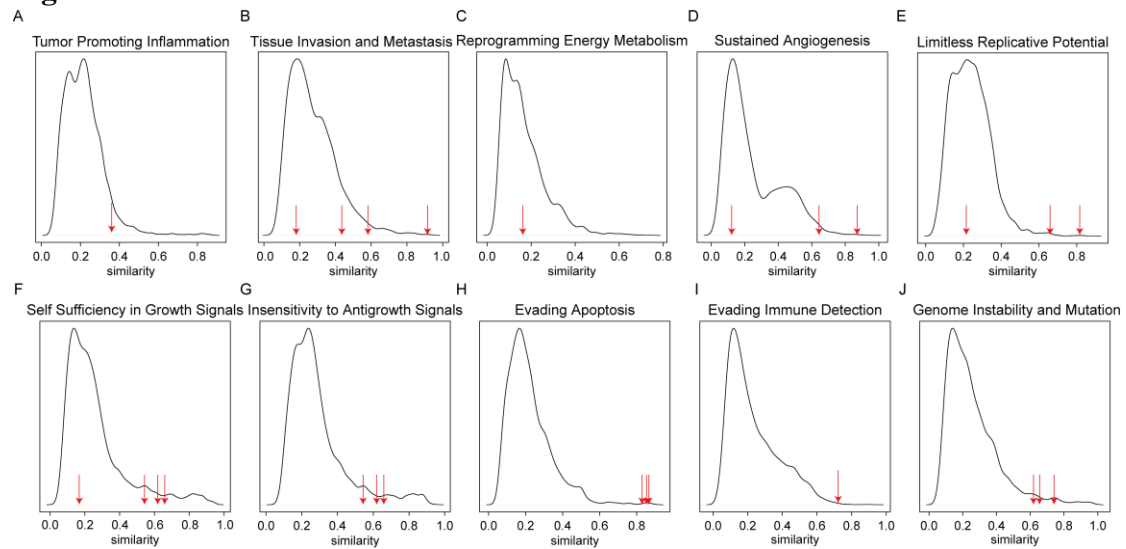


Figure S2

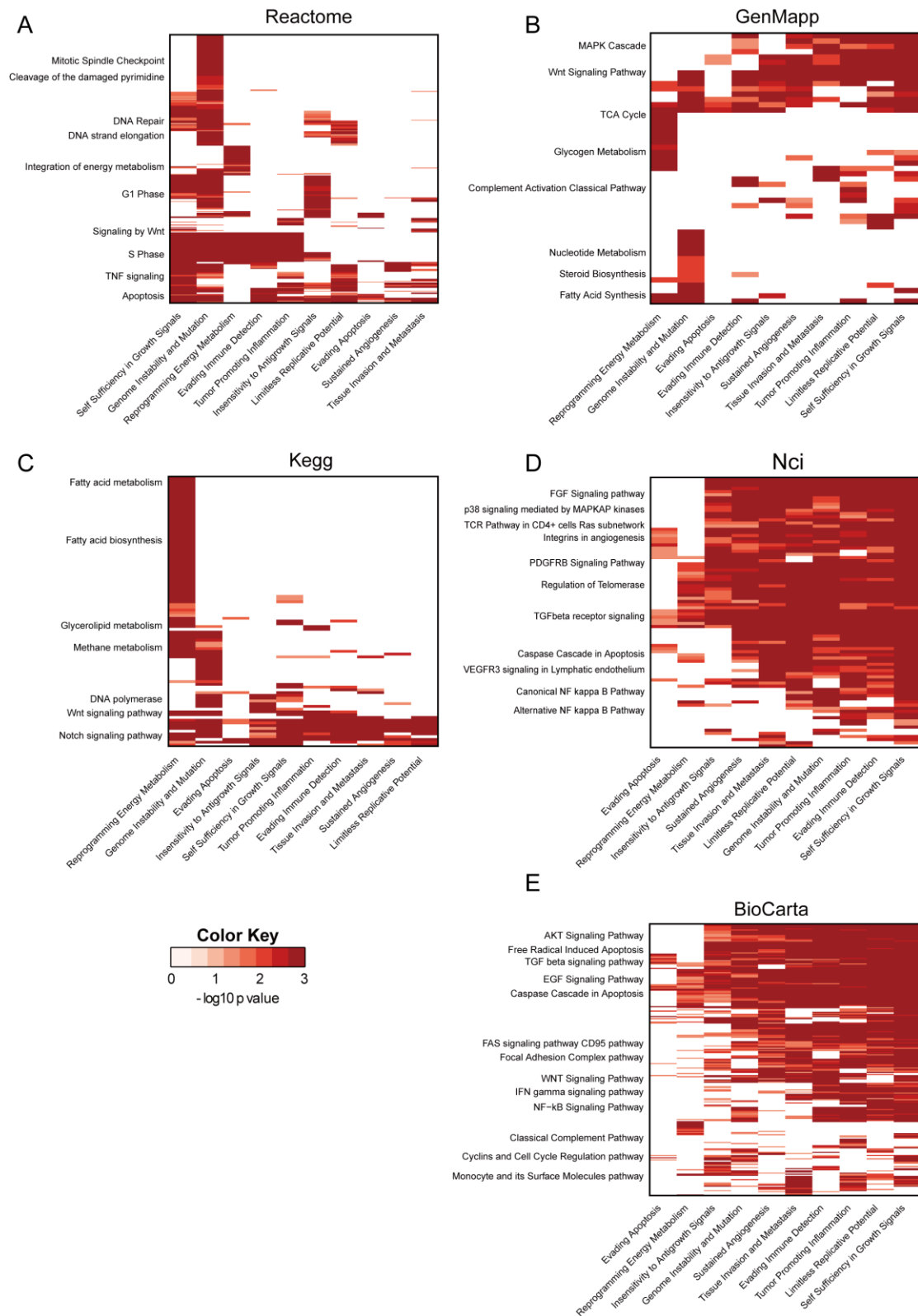


Figure S3

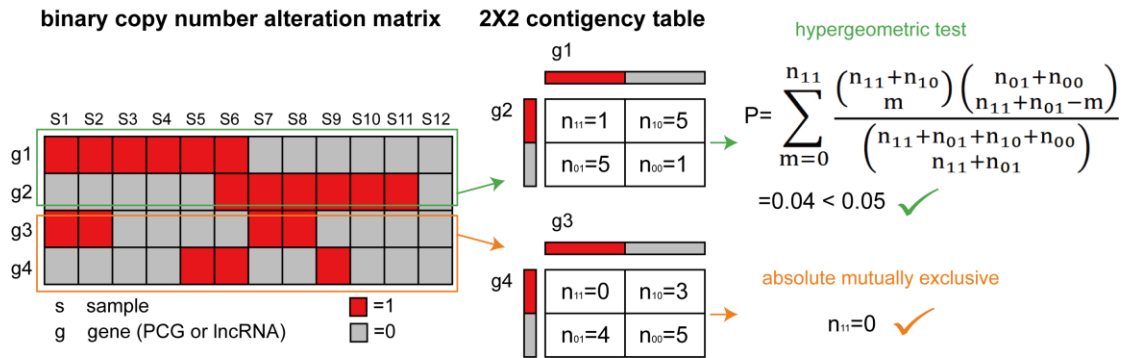


Figure S4

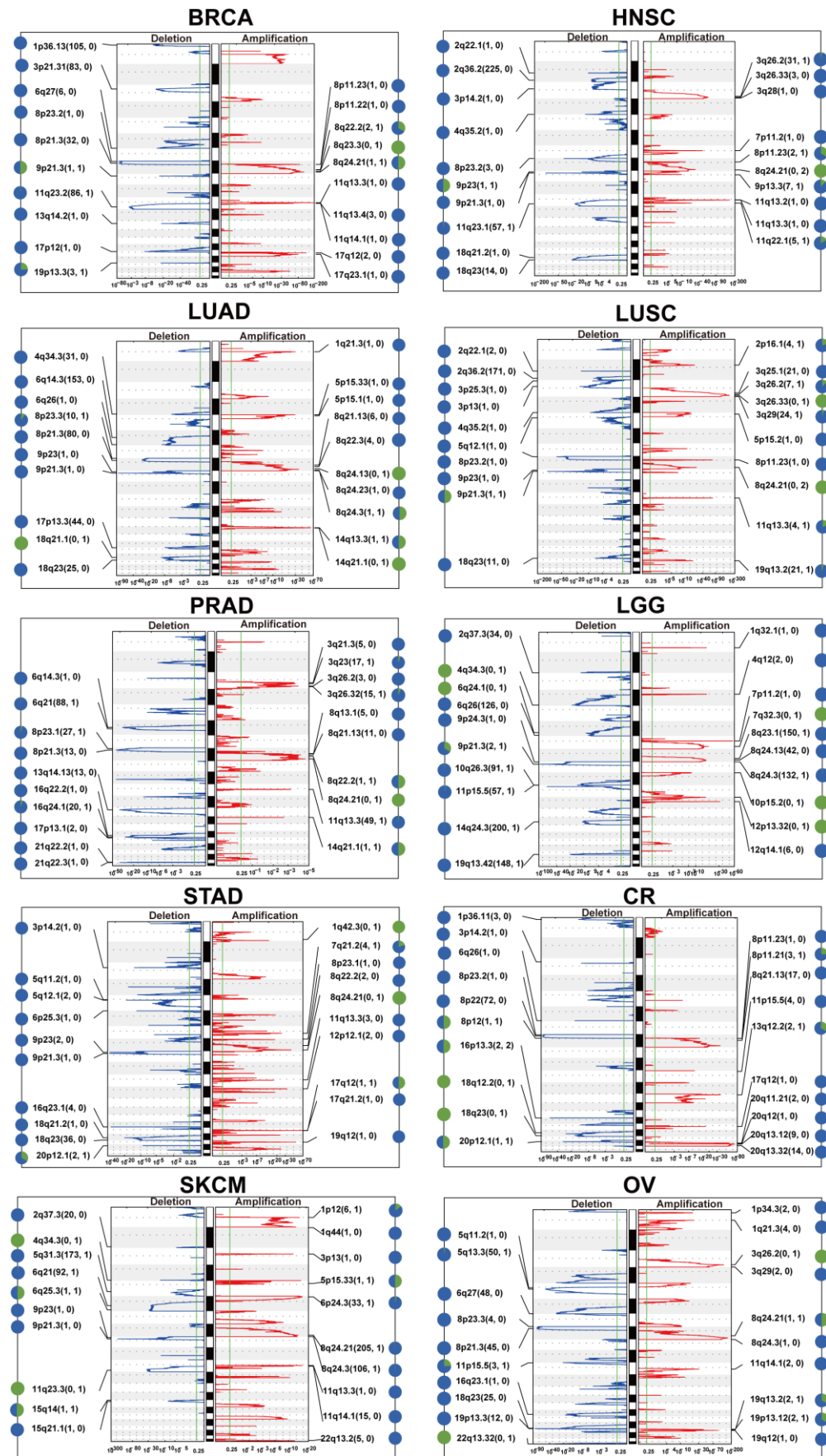


Figure S5

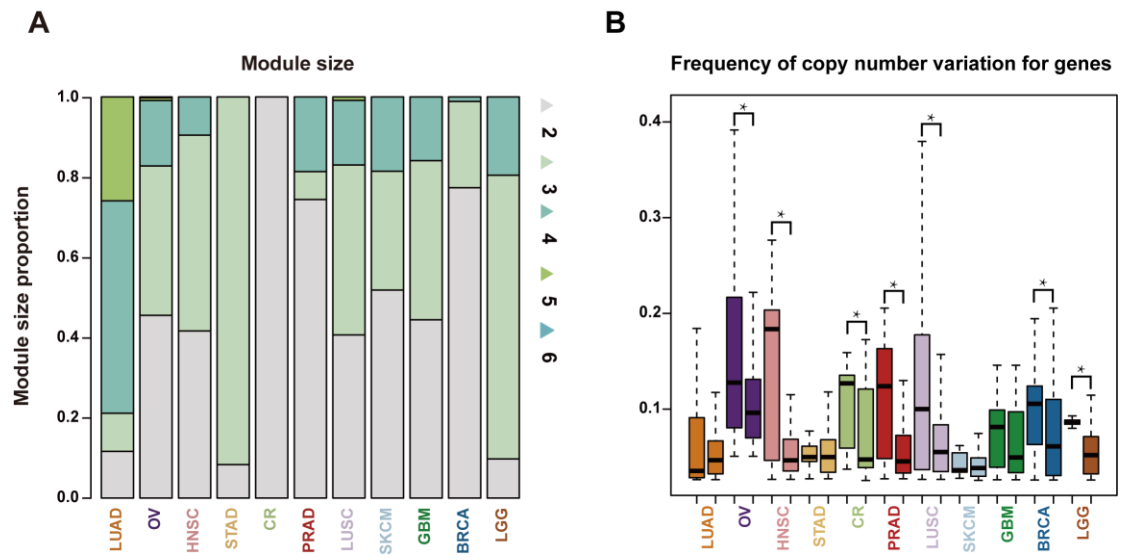
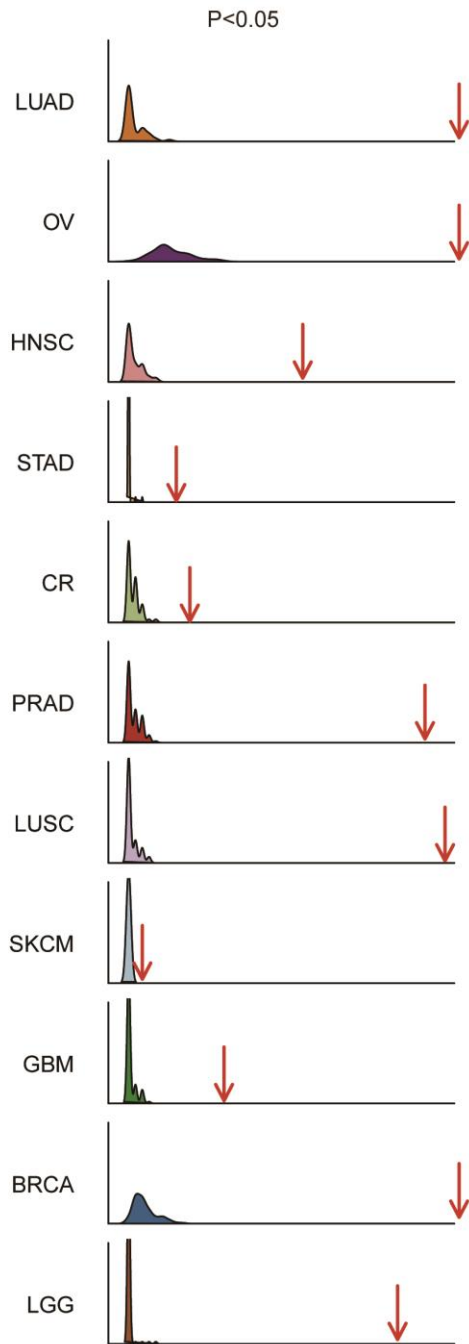


Figure S6

A

Number of predicted lncRNAs
from random CNV profiles



B

Number of predicted lncRNAs
from random expression profiles

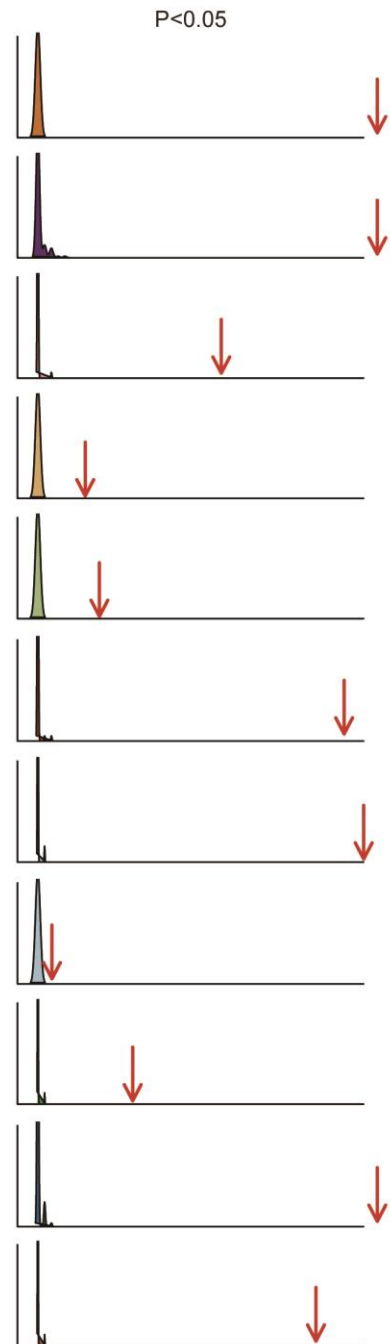


Figure S7

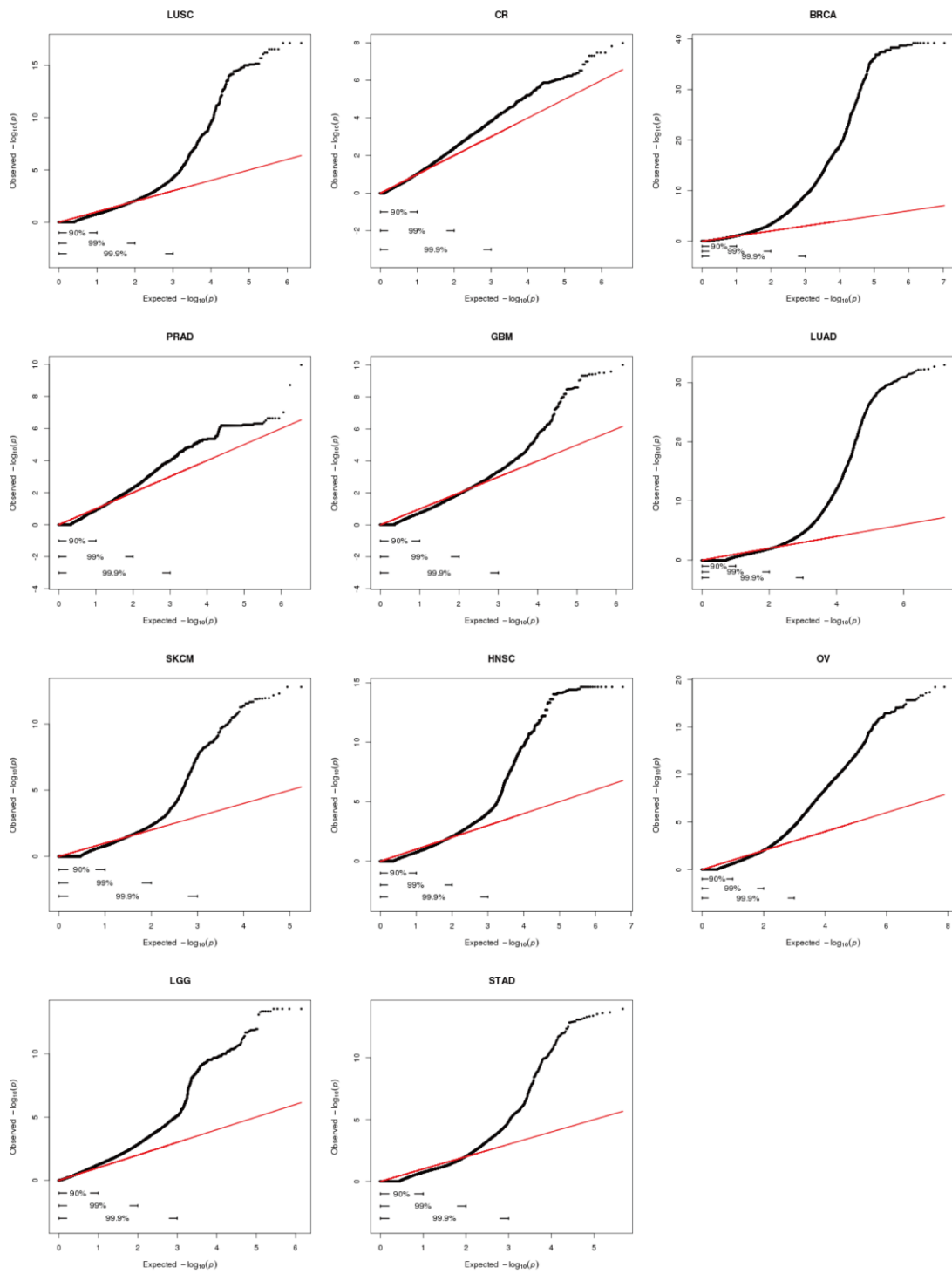


Figure S8



LUAD: lung adenocarcinoma HNSC: head and neck squamous cell carcinoma PRAD: prostate adenocarcinoma LGG: brain lower grade glioma
 STAD: stomach adenocarcinoma GBM: glioblastoma multiforme LUSC: lung squamous cell carcinoma BRCA: breast invasive carcinoma
 SKCM: skin cutaneous melanoma OV: ovarian serous cystadenocarcinoma CR: Colon adenocarcinoma&Rectum adenocarcinoma

Figure S9

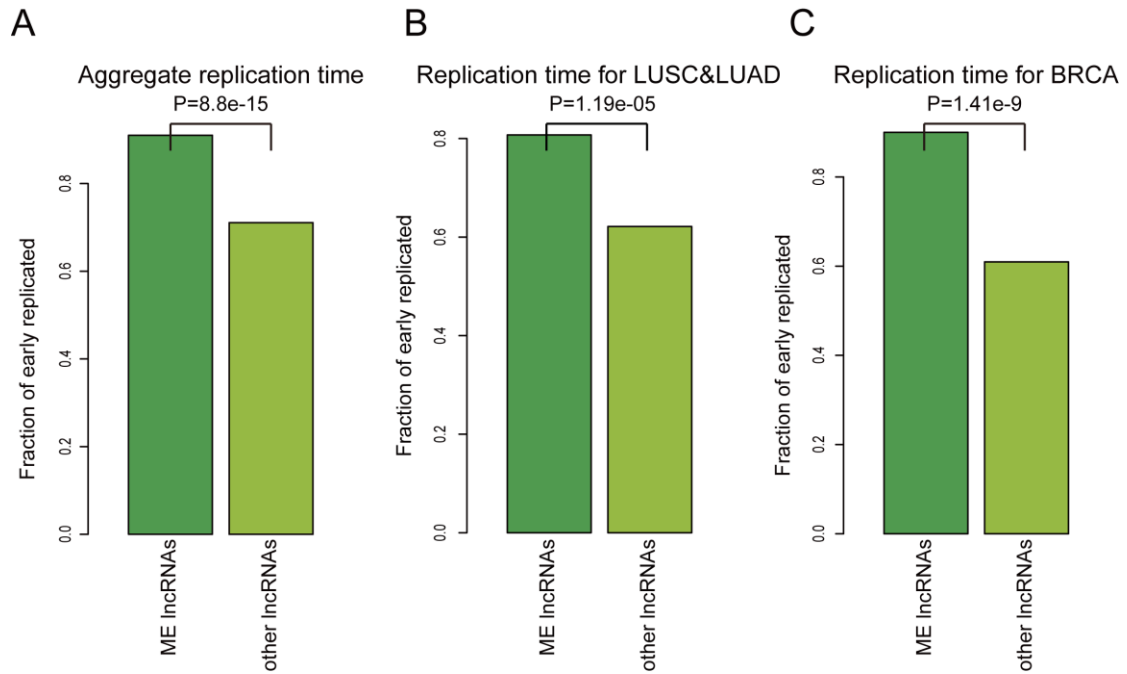


Figure S11

hallmarks affected by PVT1 across cancers

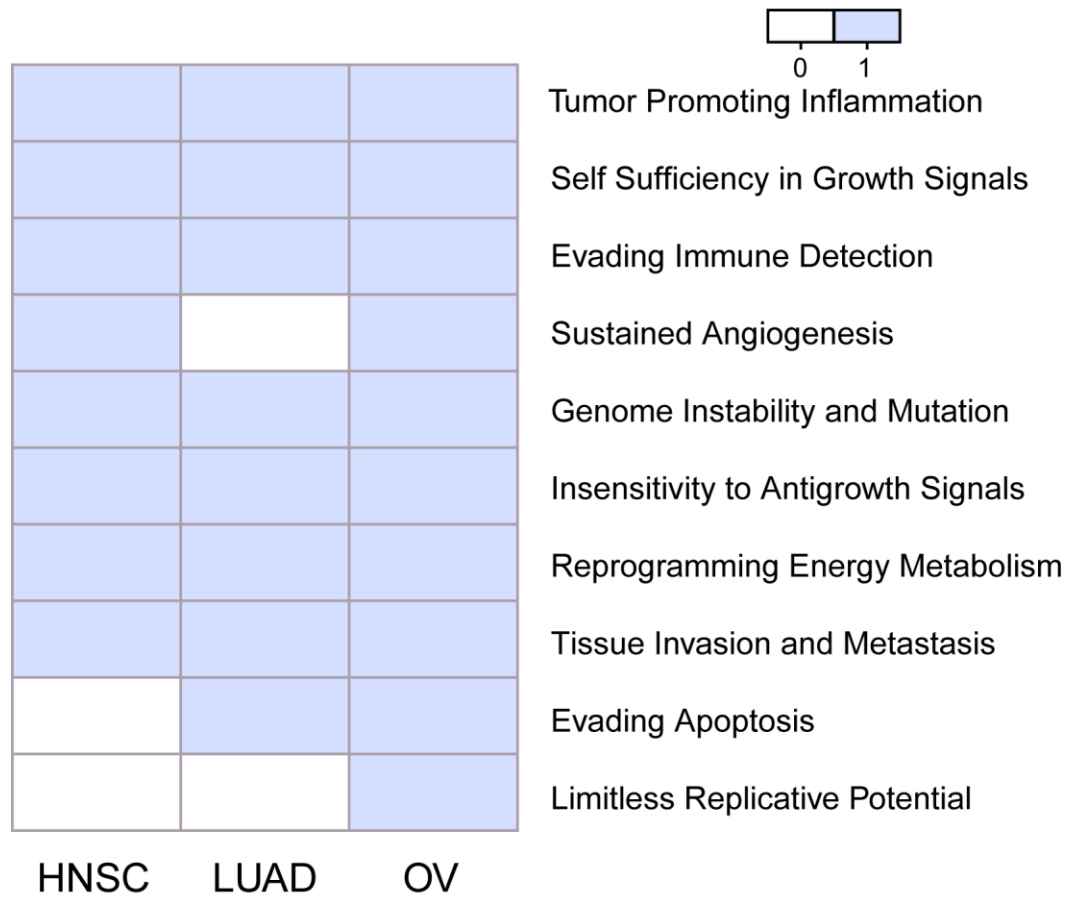


Figure S12

