

Electronic Supplementary Information (ESI)

Discovering potential driver genes through an integrated model of somatic mutation profiles and gene functional information

Jianing Xi,^{‡a} Minghui Wang^{‡ab} and Ao Li^{*ab}

^a School of Information Science and Technology, University of Science and Technology of China, Hefei AH230027, China.

^b Centers for Biomedical Engineering, University of Science and Technology of China, Hefei AH230027, China.

*E-mail: aoli@ustc.edu.cn

‡ Joint first authors.

List of Figures

- Fig. S1** The F1 scores of the results of IntDriver under the parameters λ_N and λ_S set to different values. As illustrated, the performances of IntDriver for the datasets of the three types of cancers are robust to the two parameter λ_N and λ_S generally. The F1 scores under λ_N and λ_S set to 0.3 and 0.7 are relative better than those under the other settings among the three cancer datasets BRCA, KIRC and LUSC. 2
- Fig. S2** The pie diagram illustrating the proportions of known benchmarking driver genes in the top ranked 100 genes identified by IntDriver, HotNet2, MUFFINN-DNmax, MUFFINN-DNsum, DriverNet, NBS(k=3), NBS(k=4) and ReMIC, when these approaches are applied on (a) BRCA dataset, (b) KIRC dataset and (c) LUSC dataset. The benchmarking driver genes used in this study are from NCG [1], including cancer-specific experimentally supported CGC genes [2], experimentally supported genes from CGC [2] and statistically inferred candidate genes. Black numbers near the pie diagrams represent the numbers of the detected cancer-specific CGC genes, while the gray numbers represent the numbers of detected CGC genes, NCG genes and other genes. 2
- Fig. S3** The precisions, recalls and F1 scores for NCG benchmarking driver genes of the top 200 genes selected by IntDriver, HotNet2, MUFFINN-DNmax, MUFFINN-DNsum, DriverNet, NBS(k=3), NBS(k=4) and ReMIC on (a) BRCA dataset, (b) KIRC dataset and (c) LUSC dataset respectively. 3

List of Tables

- Table S1** The driver gene candidates identified in the top 200 genes prioritized by IntDriver on BRCA dataset. Their ranks, gene symbols and mutation frequencies are demonstrated in the table, along with whether the genes are experimentally (CGC & NCG) supported or statistically inferred (NCG). The asterisk * represents that the gene is reported to related with breast cancer. 4
- Table S2** The driver gene candidates identified in the top 200 genes prioritized by IntDriver on KIRC dataset. Their ranks, gene symbols and mutation frequencies are demonstrated in the table, along with whether the genes are experimentally (CGC & NCG) supported or statistically inferred (NCG). The asterisk * represents that the gene is reported to related with kidney renal clear cell carcinoma. 5
- Table S3** The driver gene candidates identified in the top 200 genes prioritized by IntDriver on LUSC dataset. Their ranks, gene symbols and mutation frequencies are demonstrated in the table, along with whether the genes are experimentally (CGC & NCG) supported or statistically inferred (NCG). The asterisk * represents that the gene is reported to related with lung squamous cell carcinoma. 6
- Table S4** The p-values of Fisher's exact test accessing the enrichment significance for NCG [1] benchmarking driver genes of the top 50, 100, 150 and 200 genes selected by IntDriver, NotNet2, MUFFINN-DNmax, MUFFINN-DNsum, DriverNet, NBS(k=3), NBS(k=4), ReMIC. 7
- Table S5** Results of functional enrichment analysis of the top 200 genes identified by IntDriver on BRCA datasets. 8
- Table S6** Results of functional enrichment analysis of the top 200 genes identified by IntDriver on KIRC datasets. 9
- Table S7** Results of functional enrichment analysis of the top 200 genes identified by IntDriver on LUSC datasets. 10

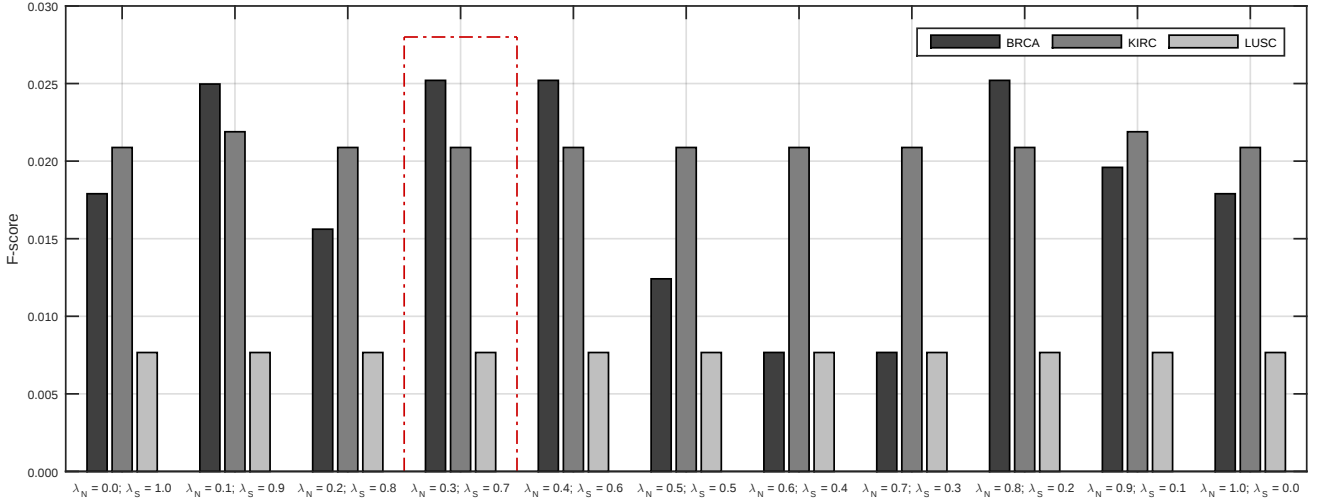


Fig. S1. The F1 scores of the results of IntDriver under the parameters λ_N and λ_S set to different values. As illustrated, the performances of IntDriver for the datasets of the three types of cancers are robust to the two parameter λ_N and λ_S generally. The F1 scores under λ_N and λ_S set to 0.3 and 0.7 are relative better than those under the other settings among the three cancer datasets BRCA, KIRC and LUSC.

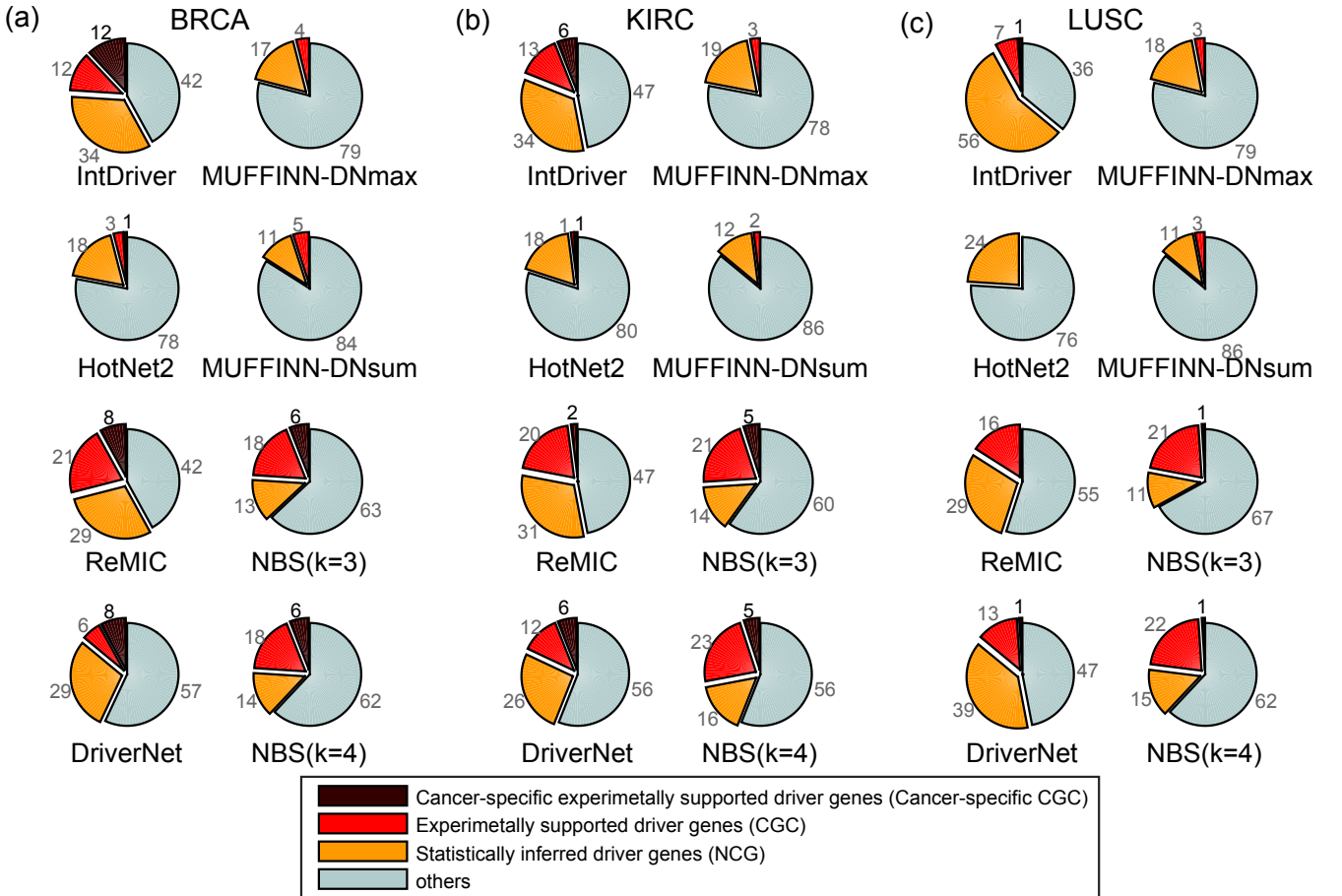


Fig. S2. The pie diagram illustrating the proportions of known benchmarking driver genes in the top ranked 100 genes identified by IntDriver, HotNet2, MUFFINN-DNmax, MUFFINN-DNsum, DriverNet, NBS(k=3), NBS(k=4) and ReMIC, when these approaches are applied on (a) BRCA dataset, (b) KIRC dataset and (c) LUSC dataset. The benchmarking driver genes used in this study are from NCG [1], including cancer-specific experimentally supported CGC genes [2], experimentally supported genes from CGC [2] and statistically inferred candidate genes. Black numbers near the pie diagrams represent the numbers of the detected cancer-specific CGC genes, while the gray numbers represent the numbers of detected CGC genes, NCG genes and other genes.

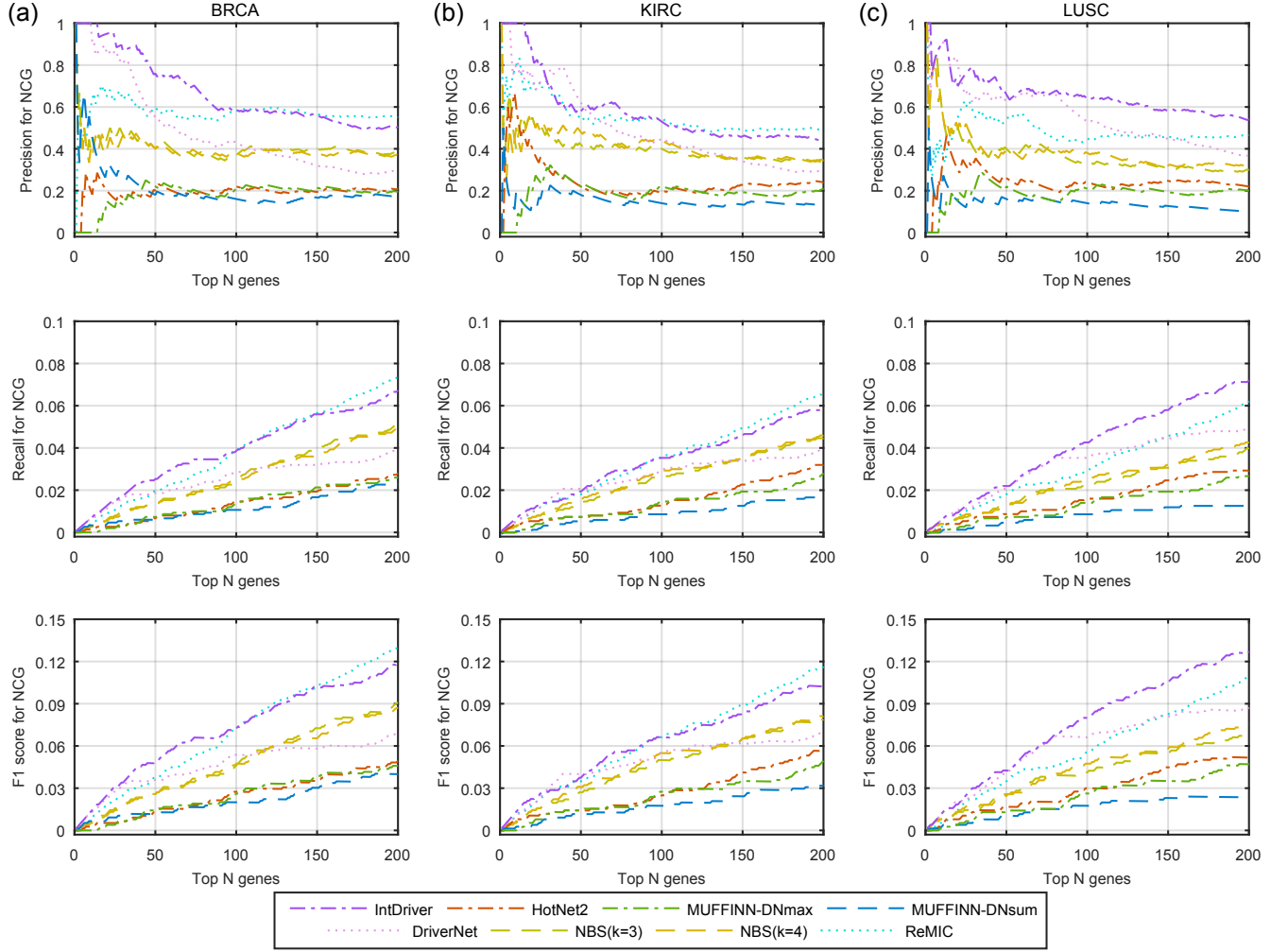


Fig. S3. The precisions, recalls and F1 scores for NCG benchmarking driver genes of the top 200 genes selected by IntDriver, HotNet2, MUFINN-DNmax, MUFINN-DNsum, DirverNet, NBS(k=3), NBS(k=4) and ReMIC on (a) BRCA dataset, (b) KIRC dataset and (c) LUSC dataset respectively.

Table S1. The driver gene candidates identified in the top 200 genes prioritized by IntDriver on BRCA dataset. Their ranks, gene symbols and mutation frequencies are demonstrated in the table, along with whether the genes are experimentally (CGC & NCG) supported or statistically inferred (NCG). The asterisk * represents that the gene is reported to related with breast cancer.

Rank	Symbol	MutFreq	Benchmarking	Rank	Symbol	MutFreq	Benchmarking	Rank	Symbol	MutFreq	Benchmarking
1	PIK3CA	180	CGC* & NCG*	68	COL12A1	11		135	MLLT4	12	CGC & NCG
2	TP53	190	CGC* & NCG*	69	CACNA1B	13		136	UBR5	10	CGC & NCG*
3	GATA3	55	CGC* & NCG*	70	RELN	17		137	CHD4	11	CGC & NCG
4	TTN	86	NCG	71	WDFY3	13		138	KIF4A	9	
5	MAP3K1	40	CGC* & NCG*	72	TG	15	NCG*	139	SRRM2	9	
6	MUC16	55	NCG*	73	LAMA1	12	NCG*	140	ITPKB	8	
7	CDH1	34	CGC* & NCG*	74	FLNC	14	NCG*	141	ASPM	10	
8	USH2A	32	NCG*	75	CACNA1G	14		142	DCAF17	4	
9	OBSCN	30	NCG	76	LRP1	12		143	LASP1	4	CGC & NCG
10	RYR2	33	NCG*	77	BRAT1	8		144	CHD5	8	NCG*
11	BRCA2	24	CGC* & NCG*	78	WNK3	8		145	CCT6B	6	NCG
12	SYNE1	22	NCG	79	ADAMTS5	7		146	DNAH1	12	
13	MAP2K4	21	CGC* & NCG*	80	CADPS	8		147	MGAM	12	NCG
14	BRCA1	19	CGC & NCG*	81	CACNA1C	12		148	DOCK2	6	NCG
15	FLG	28		82	ANKRD24	5		149	HECW2	9	
16	CSMD1	28	NCG*	83	PRX	5		150	MYO9B	9	
17	SPTA1	18	NCG	84	PTPRH	5		151	NUAK2	9	
18	MUC4	26	NCG	85	MYO10	13		152	BCORL1	8	
19	RYR3	24	NCG	86	COL6A3	10		153	BCMO1	4	
20	RUNX1	19	CGC & NCG*	87	CACNA1D	8		154	MYH11	8	CGC & NCG
21	AKT1	13	CGC* & NCG*	88	RAB1F	3		155	RAB3GAP2	9	
22	DMD	17	NCG	89	SP11	16		156	KCNB2	8	
23	APOB	19	NCG	90	HRNR	12	NCG	157	DNM3	3	
24	LRP2	20	NCG	91	PXDN	6	NCG	158	KIRREL	6	
25	DNAH5	14		92	UTRN	11		159	NNT	4	
26	SPEN	19		93	PREX2	13	NCG	160	PLCE1	9	
27	DST	14	NCG	94	VWF	13	NCG	161	HERC2	10	NCG
28	ZFXH3	14	CGC & NCG	95	KAT6B	9	CGC & NCG	162	IL17RC	5	
29	CTCF	15	CGC* & NCG*	96	SVEP1	14		163	SP100	4	
30	ANK1	14		97	LAMA5	13		164	CDC42BPA	8	
31	TBX3	14	CGC* & NCG*	98	PCDHA5	5		165	CENPE	10	
32	RYR1	17	NCG	99	PCDHA1	8		166	ATAD5	7	
33	TBL1XR1	10	CGC & NCG*	100	MED12	11	CGC & NCG	167	TANC2	7	
34	ATM	20	CGC & NCG	101	DNAH7	10	NCG	168	EPHB1	7	NCG*
35	CBFB	9	CGC & NCG*	102	FBN1	13		169	CBX8	3	
36	NCOR1	18	CGC* & NCG*	103	AKAP9	11	CGC & NCG*	170	LAMA3	8	
37	FOXA1	8	CGC & NCG	104	ZFPM2	11		171	SMEK2	5	
38	MUC2	18	NCG	105	ARID2	5	CGC & NCG*	172	POLE	7	
39	MED23	9		106	PTEN	18	CGC & NCG*	173	DSCAML1	9	NCG
40	AFF2	14	NCG*	107	MYH7	11		174	EEA1	5	
41	DNAH3	15		108	ERBB2	8	CGC* & NCG*	175	HCFC1	7	
42	NEB	19	NCG	109	PIK3R1	15	CGC & NCG*	176	PCDHA8	8	
43	SRCAP	12		110	TSC22D1	6	NCG	177	CAPN1	11	
44	UBR4	18		111	SDK1	13		178	MYH9	8	CGC & NCG*
45	XIRP2	15	NCG	112	RIMBP2	10	NCG	179	VPS18	6	
46	HECW1	13		113	COL14A1	11		180	SCN10A	9	
47	MDN1	15		114	ARHGAP35	8		181	NCOA3	7	NCG
48	PLXNA4	12		115	MYO5B	9	NCG	182	FLNB	9	NCG*
49	SSPO	20		116	PKD1	10		183	ERBB3	9	CGC & NCG*
50	SYNE2	19	NCG	117	CACNA1F	8	NCG*	184	LATS2	4	
51	BZRAP1	11		118	MYH7B	9		185	HNFA4	3	
52	BCOR	7	CGC & NCG	119	ZDBF2	10	NCG	186	GRIK2	8	
53	CIT	11	NCG	120	WDR7	8		187	PKHD1	13	NCG*
54	ARID1A	11	CGC* & NCG*	121	PDE3A	11		188	ATN1	9	NCG*
55	CCDC40	4		122	PTPRD	12	NCG*	189	GRIN2C	7	
56	PDE4DIP	7	CGC & NCG	123	CROCC	12	NCG	190	IKBKE	6	
57	PLEC	16	NCG	124	ARID1B	10	CGC* & NCG*	191	RIMS2	7	NCG*
58	BRIP1	10	CGC & NCG	125	F5	11		192	TRPM6	5	NCG
59	CELSR1	5		126	CHD6	9	NCG	193	ADCY8	7	NCG
60	EPG5	9	NCG*	127	DCC	10		194	RNF213	11	CGC & NCG
61	PCNT	11		128	MACF1	11	NCG	195	KIAA2022	10	NCG
62	LRBA	11	NCG*	129	MYT1L	10		196	DYNC2H1	9	NCG*
63	ANK3	11	NCG	130	MYLK	10		197	AMPD1	7	
64	LETM1	5	NCG	131	PREX1	9		198	MYH14	10	
65	MTOR	13		132	CBLB	10	CGC & NCG	199	CACNA1A	10	
66	MUC12	11		133	HSPG2	11		200	MGA	10	NCG
67	CACNA1E	14	NCG	134	CSMD2	13	NCG				

Table S2. The driver gene candidates identified in the top 200 genes prioritized by IntDriver on KIRC dataset. Their ranks, gene symbols and mutation frequencies are demonstrated in the table, along with whether the genes are experimentally (CGC & NCG) supported or statistically inferred (NCG). The asterisk * represents that the gene is reported to related with kidney renal clear cell carcinoma.

Rank	Symbol	MutFreq	Benchmarking	Rank	Symbol	MutFreq	Benchmarking	Rank	Symbol	MutFreq	Benchmarking
1	VHL	248	CGC* & NCG*	68	MUC6	19		135	HLA-DQB1	11	
2	PBRM1	167	CGC* & NCG*	69	MAGEC1	18	NCG	136	CUL9	11	
3	MUC4	103	NCG	70	UHRF1BP1	5		137	NEB	18	NCG
4	TTN	89	NCG	71	ATM	16	CGC & NCG	138	AKAP9	14	CGC & NCG
5	SETD2	62	CGC* & NCG*	72	MKNK2	5		139	EIF4G3	7	
6	BAP1	49	CGC* & NCG*	73	ANPEP	11		140	SLC36A1	5	
7	PABPC1	22	NCG	74	PCSK5	15		141	LRBA	5	NCG
8	MUC16	50	NCG	75	SVEP1	11		142	MLLT3	9	CGC & NCG
9	ZFXH3	12	CGC & NCG	76	RBM26	5		143	MEGF10	8	
10	PTEN	20	CGC & NCG	77	FLG	13		144	PKHD1	14	NCG
11	KDM5C	32	CGC* & NCG*	78	NDUFS1	8		145	PRKCD	5	
12	PDE4DIP	26	CGC & NCG	79	CR1	7	NCG	146	EPRS	5	
13	VCAN	9	NCG	80	PODXL	6		147	PABPC3	12	NCG
14	TP53	16	CGC & NCG*	81	TOM1L2	5		148	LILRB4	8	
15	FBN2	17	NCG	82	MTMR4	6		149	OGDH	6	
16	SPEN	12		83	HLA-DRB5	11		150	LRP2	16	NCG
17	ANK2	12	NCG	84	PLB1	7	NCG	151	OBSCN	10	NCG
18	PRKAG2	15		85	AKR1C1	5		152	TTI1	5	
19	DST	29	NCG	86	MICAL3	8	NCG	153	PCSK2	5	
20	RBX1	7		87	FMN2	16	NCG*	154	POLR2B	5	
21	FAM179B	10		88	HERC1	11		155	PCNT	10	
22	USH2A	21	NCG	89	MYH11	10	CGC & NCG	156	RRP12	7	
23	PCLO	20	NCG	90	SDHA	12		157	KDM5B	7	NCG
24	CASP8AP2	4		91	ADCY8	6	NCG	158	XPO1	6	CGC & NCG
25	SETBP1	9	CGC & NCG	92	PPP1R3A	4	NCG	159	SMARCA4	15	CGC & NCG
26	FREM1	7	NCG	93	GLS	6		160	FN1	8	NCG
27	SYNRG	6		94	NBR1	6		161	UBR4	14	
28	ZBTB38	10		95	LRRK1	11	NCG	162	COL6A3	15	
29	CEP290	9		96	SEC14L4	4		163	MACF1	4	NCG
30	GPRIN1	10		97	CROCC	8	NCG	164	ITIH4	5	
31	ADCY2	6	NCG	98	TATDN1	4		165	UIMC1	6	
32	ABCC3	7		99	WDFY3	12		166	TRPM6	5	NCG
33	HLA-DRB1	16		100	ZNF714	4		167	BAZZ2B	13	NCG
34	CSPG4	6		101	IL17RC	5		168	SSPO	15	
35	MTOR	25		102	STARD9	9		169	INPP5E	3	
36	PAH	7		103	EBF3	9		170	LAMA1	14	NCG
37	ANK3	15	NCG	104	YLP1M1	12		171	CNOT1	13	NCG
38	TCEB1	5		105	ZNFX1	11		172	MUC2	15	NCG
39	TRIOBP	6	NCG	106	DPP10	7	NCG	173	DNAH3	14	
40	PREX1	7		107	DSC2	4		174	MYO5B	12	NCG
41	HUWE1	12		108	AHNAK	10	NCG*	175	RYR1	15	NCG*
42	AFF3	8	CGC & NCG	109	NFASC	6		176	LRP1	15	
43	RYR3	14	NCG	110	SUCLG1	6		177	AGL	5	
44	RYR2	15	NCG	111	EP400	10		178	TSHZ3	8	
45	RIF1	7		112	HLA-B	10	NCG	179	JMJD1C	9	
46	LYST	14		113	ANKRD17	9		180	KCNMA1	4	
47	CUBN	16		114	BCORL1	9		181	CARD10	3	NCG
48	ARID1A	20	CGC* & NCG*	115	PLEC	11	NCG	182	PAK6	5	
49	PTCH1	14	CGC & NCG	116	STAG2	8		183	USP18	6	NCG
50	TNS1	9		117	CABIN1	9		184	SSFA2	6	
51	SYNE2	13	NCG*	118	NBEA	7		185	NSD1	8	CGC & NCG
52	SLIT2	12	NCG	119	LAMA3	6		186	SERPINB4	4	
53	SYNE1	22	NCG	120	DNAH7	13	NCG	187	BAZZ2A	3	
54	GRM8	9	NCG	121	BABAM1	7	NCG	188	CHD5	7	NCG
55	ASMTL	6		122	RCHY1	5		189	BIN2	3	
56	PRSS16	7		123	AQR	8		190	KIAA0232	2	
57	APOB	12	NCG	124	CSMD1	12	NCG	191	COL5A1	8	NCG
58	KRT4	7		125	EML4	3	CGC & NCG	192	AGK	4	
59	COL7A1	7	NCG	126	ASPM	7		193	ASCC3	7	
60	EZH2	5	CGC & NCG	127	NOM1	5		194	CCDC88B	7	
61	TNR	14	NCG	128	CDH8	10		195	TTC17	3	
62	MAGED1	4	NCG	129	PDE8A	3		196	C1orf173	5	
63	VPS13A	8		130	DNAJC14	7		197	RNF40	4	
64	NOTCH2	20	CGC & NCG	131	PDS5A	4		198	TACC3	3	
65	EGFR	8	CGC & NCG	132	NF1	10	CGC & NCG	199	SPZ1	4	
66	LRP1B	19	CGC & NCG*	133	COL4A5	12		200	LRRK2	7	NCG*
67	KCNJ12	18	NCG	134	SIPA1L3	6					

Table S3. The driver gene candidates identified in the top 200 genes prioritized by IntDriver on LUSC dataset. Their ranks, gene symbols and mutation frequencies are demonstrated in the table, along with whether the genes are experimentally (CGC & NCG) supported or statistically inferred (NCG). The asterisk * represents that the gene is reported to related with lung squamous cell carcinoma.

Rank	Symbol	MutFreq	Benchmarking	Rank	Symbol	MutFreq	Benchmarking	Rank	Symbol	MutFreq	Benchmarking
1	TP53	172	CGC* & NCG*	68	NBAS	20		135	LRRK2	27	NCG*
2	TTN	145	NCG*	69	MYCBP2	27	NCG	136	ZNF208	29	
3	MUC16	98	NCG*	70	CTNNA2	34	NCG	137	KALRN	21	NCG
4	FLG	48		71	PEG3	30	NCG	138	COL12A1	25	
5	PKHD1	46	NCG*	72	STAB2	34		139	DYSF	23	
6	RYR2	82	NCG	73	ROS1	26	CGC & NCG*	140	EPHB3	16	
7	LRP1B	74	CGC & NCG*	74	LTBP1	23		141	TLN2	20	
8	LRP2	47	NCG	75	ASH1L	13		142	CCDC141	17	
9	RYR3	57	NCG*	76	SLIT2	19	NCG*	143	HERC2	30	NCG*
10	APOB	41	NCG	77	PAPPA2	40	NCG*	144	SLC8A1	23	
11	MACF1	44	NCG	78	MYH2	31	NCG	145	CHD6	19	NCG
12	USH2A	72	NCG	79	FAT1	32	CGC & NCG	146	FLT1	14	
13	SYNE1	63	NCG	80	BAI3	28	NCG*	147	MUC4	29	NCG
14	DNAH5	43		81	DIDO1	21	NCG	148	ARID1B	13	CGC & NCG
15	DYNC1H1	30		82	USP34	23		149	LRP1	26	
16	BIRC6	34		83	TNC	20		150	COL5A2	18	
17	MDN1	31		84	DNAH7	35	NCG	151	TPR	20	CGC & NCG
18	XIRP2	47	NCG	85	DSCAML1	23	NCG	152	MYH9	15	CGC & NCG
19	NAV3	49	NCG	86	CEP89	4		153	PIK3CG	22	NCG*
20	VPS13D	30		87	NLRP3	20	NCG*	154	UTRN	17	
21	GLI3	13	NCG*	88	SH3GL3	1		155	RANBP2	18	
22	SI	37	NCG	89	LRRC7	35	NCG	156	TEP1	15	
23	AHNAK	36	NCG	90	SHANK2	16		157	KIAA1549	20	CGC & NCG
24	FAM135B	48	NCG	91	LYST	19		158	KAT6A	18	CGC & NCG
25	CSMD1	42	NCG	92	SPHKAP	31	NCG*	159	VCAN	23	NCG
26	SPTA1	49	NCG	93	MYH1	23	NCG	160	COL7A1	21	NCG
27	SLITRK3	32	NCG	94	GRM3	23	NCG*	161	FBN1	22	
28	HCN1	42	NCG*	95	COL6A3	20		162	ABCA1	16	NCG
29	CSMD2	38	NCG	96	MYH8	25	NCG	163	UBE2Q2	1	
30	ARHGAP5	2		97	DNAH1	14		164	COL4A1	18	
31	NRXN1	33		98	RP1	36	NCG	165	PLEC	23	NCG
32	TNR	35	NCG	99	DMBT1	19		166	MKI67	15	
33	PRPF31	0		100	BDP1	13		167	TRPM6	14	NCG
34	AKAP12	8	NCG	101	ADAMTS9	18		168	PXDN	20	NCG
35	SYNE2	26	NCG	102	MED12L	22	NCG	169	GPRASP1	14	
36	KIAA1109	27		103	ANK3	19	NCG	170	PTPRB	20	
37	MYO18B	26	NCG	104	GABRB3	18	NCG	171	CNTNAP2	34	NCG*
38	CACNA1E	24	NCG	105	ADCY2	19	NCG*	172	ZSCAN4	9	
39	PLXNA4	26		106	NRAP	17		173	NOS1	18	
40	RYR1	34	NCG	107	FLNC	23	NCG	174	CDH18	31	
41	COL11A1	41	NCG*	108	CUL9	19		175	SRCAP	17	
42	ALMS1	29	NCG	109	ABCB5	17	NCG	176	SPEN	14	
43	FAT4	34	CGC & NCG	110	PCNX	18		177	PTPRH	15	
44	VPS28	1		111	PCLO	38	NCG	178	MYLK	14	
45	SCN1A	34		112	EPHA5	18	NCG*	179	UBAP2L	8	
46	CTNND2	28		113	PIK3CA	29	CGC & NCG*	180	ADCY8	22	NCG*
47	SH3PXD2A	6		114	SMG1	16		181	PLB1	14	NCG
48	TRRAP	20	CGC & NCG	115	GRIN2B	26		182	PRG4	18	NCG
49	SH3PXD2B	5		116	DST	18	NCG*	183	MXRA5	17	
50	CUBN	38		117	MYT1L	24		184	FMN2	35	NCG
51	SH3KBP1	6		118	NOTCH2	16	CGC & NCG	185	UNC79	21	NCG
52	SH3GLB2	0		119	UNC5A	5		186	MYH4	27	NCG
53	EPHB1	25	NCG*	120	TAF1L	23		187	SLIT3	15	
54	PTFRT	24	NCG	121	ITPR2	16		188	LILRB2	16	
55	PDE4DIP	34	CGC & NCG	122	TRPM2	14		189	CMYA5	18	NCG*
56	DCHS1	19	NCG	123	RNF213	15	CGC & NCG	190	SACS	20	
57	FBN2	29	NCG	124	SETD5	8	NCG	191	NRK	20	NCG
58	LAMA4	21	NCG	125	C1orf173	41		192	NUP205	11	
59	CDH10	38	NCG*	126	HERC1	17		193	PFFIA1	13	
60	CDH12	33	NCG	127	ZNF804A	24	NCG	194	IFT172	13	
61	ANK2	40	NCG	128	LARGE	16		195	MYH13	18	
62	NEB	37	NCG	129	OTOF	18	NCG	196	ZNF33A	15	
63	SMEK1	1		130	GRIN2A	20	CGC & NCG	197	DMXL2	13	
64	OBSCN	33	NCG	131	PRUNE2	17	NCG	198	DNAJC13	16	
65	SRRM2	20		132	OCA2	17		199	FCGBP	19	
66	UBR5	16	CGC & NCG*	133	POLQ	22		200	ZFH3X	20	CGC & NCG
67	CNTNAP5	38		134	NIPBL	15					

Table S4. The p-values of Fisher’s exact test accessing the enrichment significance for NCG [1] benchmarking driver genes of the top 50, 100, 150 and 200 genes selected by IntDriver, NotNet2, MUFFINN-DNmax, MUFFINN-DNsum, DriverNet, NBS(k=3), NBS(k=4), ReMIC.

BRCA				
Rank	50	100	150	200
IntDriver	6.00e-25	1.34e-27	6.99e-38	5.36e-40
HotNet2	1.27e-01	5.81e-03	1.25e-02	4.82e-04
MUFFINN-DNmax	5.01e-02	1.38e-02	1.69e-03	3.36e-03
MUFFINN-DNsum	2.77e-01	2.84e-01	1.33e-01	5.10e-02
DriverNet	1.81e-13	1.72e-14	1.93e-10	7.12e-11
NBS(k=3)	7.47e-07	2.33e-10	7.66e-18	2.45e-21
NBS(k=4)	7.47e-07	5.30e-11	6.50e-14	2.24e-19
ReMIC	1.82e-15	1.34e-27	7.27e-39	2.48e-48
KIRC				
Rank	50	100	150	200
IntDriver	1.89e-14	8.59e-23	1.48e-24	3.06e-29
HotNet2	5.01e-02	3.10e-02	3.98e-04	6.09e-06
MUFFINN-DNmax	5.01e-02	5.81e-03	1.25e-02	4.82e-04
MUFFINN-DNsum	2.77e-01	6.46e-01	7.08e-01	5.89e-01
DriverNet	1.62e-16	3.04e-15	4.05e-12	2.45e-11
NBS(k=3)	1.47e-07	2.41e-12	6.50e-14	3.87e-15
NBS(k=4)	6.88e-10	3.04e-15	6.50e-14	2.67e-16
ReMIC	1.60e-12	8.59e-23	1.06e-28	3.04e-38
LUSC				
Rank	50	100	150	200
IntDriver	9.92e-19	5.83e-34	7.22e-41	2.04e-46
HotNet2	7.89e-03	1.15e-03	2.86e-05	1.25e-04
MUFFINN-DNmax	5.01e-02	1.38e-02	1.25e-02	2.22e-03
MUFFINN-DNsum	3.92e-01	6.46e-01	9.01e-01	5.15e-01
DriverNet	1.32e-17	8.59e-23	5.57e-23	2.24e-19
NBS(k=3)	7.47e-07	5.71e-08	2.18e-09	2.45e-11
NBS(k=4)	3.52e-06	5.30e-11	1.93e-10	1.81e-13
ReMIC	1.60e-12	5.18e-16	9.20e-24	3.38e-33

Table S5. Results of functional enrichment analysis of the top 200 genes identified by IntDriver on BRCA datasets.

Term	Count	Percentage	PValue
hsa04020:Calcium signaling pathway	17	8.67	2.68E-09
hsa04930:Type II diabetes mellitus	9	4.59	2.68E-07
hsa04510:Focal adhesion	15	7.65	8.88E-07
hsa05205:Proteoglycans in cancer	14	7.14	3.72E-06
hsa05213:Endometrial cancer	7	3.57	7.92E-05
hsa05200:Pathways in cancer	17	8.67	9.97E-05
hsa04725:Cholinergic synapse	9	4.59	1.52E-04
hsa05222:Small cell lung cancer	8	4.08	1.75E-04
hsa04919:Thyroid hormone signaling pathway	9	4.59	1.83E-04
hsa04012:ErbB signaling pathway	8	4.08	2.02E-04
hsa04010:MAPK signaling pathway	13	6.63	2.15E-04
hsa05230:Central carbon metabolism in cancer	7	3.57	2.55E-04
hsa04921:Oxytocin signaling pathway	10	5.10	3.48E-04
hsa04713:Circadian entrainment	8	4.08	3.49E-04
hsa04024:cAMP signaling pathway	11	5.61	4.27E-04
hsa05161:Hepatitis B	9	4.59	9.27E-04
hsa04261:Adrenergic signaling in cardiomyocytes	9	4.59	9.69E-04
hsa05221:Acute myeloid leukemia	6	3.06	1.07E-03
hsa05414:Dilated cardiomyopathy	7	3.57	1.10E-03
hsa04512:ECM-receptor interaction	7	3.57	1.32E-03
hsa05215:Prostate cancer	7	3.57	1.41E-03
hsa04210:Apoptosis	6	3.06	1.69E-03
hsa05214:Glioma	6	3.06	2.09E-03
hsa05212:Pancreatic cancer	6	3.06	2.09E-03
hsa04022:cGMP-PKG signaling pathway	9	4.59	2.21E-03
hsa04973:Carbohydrate digestion and absorption	5	2.55	2.71E-03
hsa04151:PI3K-Akt signaling pathway	13	6.63	3.05E-03
hsa05218:Melanoma	6	3.06	3.08E-03
hsa05220:Chronic myeloid leukemia	6	3.06	3.27E-03
hsa05410:Hypertrophic cardiomyopathy (HCM)	6	3.06	4.62E-03
hsa04727:GABAergic synapse	6	3.06	6.65E-03
hsa04911:Insulin secretion	6	3.06	6.65E-03
hsa05223:Non-small cell lung cancer	5	2.55	7.67E-03
hsa04150:mTOR signaling pathway	5	2.55	8.68E-03
hsa04912:GnRH signaling pathway	6	3.06	8.83E-03
hsa05166:HTLV-I infection	10	5.10	9.50E-03
hsa04611:Platelet activation	7	3.57	9.66E-03
hsa05210:Colorectal cancer	5	2.55	1.09E-02
hsa04530:Tight junction	7	3.57	1.23E-02
hsa04723:Retrograde endocannabinoid signaling	6	3.06	1.35E-02
hsa05146:Amoebiasis	6	3.06	1.63E-02
hsa05412:Arrhythmogenic right ventricular cardiomyopathy (ARVC)	5	2.55	1.73E-02
hsa04260:Cardiac muscle contraction	5	2.55	2.08E-02
hsa04724:Glutamatergic synapse	6	3.06	2.17E-02
hsa05100:Bacterial invasion of epithelial cells	5	2.55	2.36E-02
hsa05145:Toxoplasmosis	6	3.06	2.48E-02
hsa04925:Aldosterone synthesis and secretion	5	2.55	2.67E-02
hsa04666:Fc gamma R-mediated phagocytosis	5	2.55	3.00E-02
hsa05010:Alzheimer's disease	7	3.57	3.03E-02
hsa04923:Regulation of lipolysis in adipocytes	4	2.04	4.35E-02
hsa04066:HIF-1 signaling pathway	5	2.55	4.85E-02
hsa04070:Phosphatidylinositol signaling system	5	2.55	4.85E-02
hsa04924:Renin secretion	4	2.04	6.05E-02
hsa04620:Toll-like receptor signaling pathway	5	2.55	6.14E-02
hsa04931:Insulin resistance	5	2.55	6.48E-02
hsa04664:Fc epsilon RI signaling pathway	4	2.04	6.99E-02
hsa04726:Serotonergic synapse	5	2.55	7.02E-02
hsa00562:Inositol phosphate metabolism	4	2.04	7.74E-02
hsa04270:Vascular smooth muscle contraction	5	2.55	8.57E-02
hsa04071:Sphingolipid signaling pathway	5	2.55	8.78E-02
hsa04722:Neurotrophin signaling pathway	5	2.55	8.78E-02
hsa04152:AMPK signaling pathway	5	2.55	9.20E-02

Table S6. Results of functional enrichment analysis of the top 200 genes identified by IntDriver on KIRC datasets.

Term	Count	Percentage	PValue
hsa04512:ECM-receptor interaction	7	3.52	1.32E-03
hsa04510:Focal adhesion	10	5.03	2.31E-03
hsa05200:Pathways in cancer	14	7.04	3.17E-03
hsa05146:Amoebiasis	7	3.52	3.61E-03
hsa04921:Oxytocin signaling pathway	8	4.02	6.58E-03
hsa05222:Small cell lung cancer	6	3.02	6.65E-03
hsa05416:Viral myocarditis	5	2.51	8.17E-03
hsa05332:Graft-versus-host disease	4	2.01	1.08E-02
hsa04066:HIF-1 signaling pathway	6	3.02	1.19E-02
hsa05230:Central carbon metabolism in cancer	5	2.51	1.22E-02
hsa05330:Allograft rejection	4	2.01	1.48E-02
hsa04940:Type I diabetes mellitus	4	2.01	2.07E-02
hsa04151:PI3K-Akt signaling pathway	11	5.53	2.23E-02
hsa05166:HTLV-I infection	9	4.52	2.69E-02
hsa04911:Insulin secretion	5	2.51	3.11E-02
hsa05320:Autoimmune thyroid disease	4	2.01	3.61E-02
hsa04977:Vitamin digestion and absorption	3	1.51	3.75E-02
hsa04120:Ubiquitin mediated proteolysis	6	3.02	4.31E-02
hsa04713:Circadian entrainment	5	2.51	4.41E-02
hsa05206:MicroRNAs in cancer	9	4.52	4.62E-02
hsa04514:Cell adhesion molecules (CAMs)	6	3.02	4.89E-02
hsa04320:Dorso-ventral axis formation	3	1.51	5.45E-02
hsa05205:Proteoglycans in cancer	7	3.52	6.16E-02
hsa05214:Glioma	4	2.01	6.28E-02
hsa05211:Renal cell carcinoma	4	2.01	6.28E-02
hsa04931:Insulin resistance	5	2.51	6.48E-02
hsa05310:Asthma	3	1.51	6.58E-02
hsa00020:Citrate cycle (TCA cycle)	3	1.51	6.58E-02
hsa04115:p53 signaling pathway	4	2.01	6.75E-02
hsa05140:Leishmaniasis	4	2.01	7.74E-02
hsa05145:Toxoplasmosis	5	2.51	8.37E-02
hsa04612:Antigen processing and presentation	4	2.01	9.07E-02

Table S7. Results of functional enrichment analysis of the top 200 genes identified by IntDriver on LUSC datasets.

Term	Count	Percentage	PValue
hsa04510:Focal adhesion	14	7.18	4.07E-06
hsa04020:Calcium signaling pathway	11	5.64	1.59E-04
hsa04921:Oxytocin signaling pathway	10	5.13	2.98E-04
hsa04713:Circadian entrainment	8	4.10	3.07E-04
hsa04611:Platelet activation	9	4.62	3.90E-04
hsa04512:ECM-receptor interaction	7	3.59	1.19E-03
hsa04974:Protein digestion and absorption	7	3.59	1.26E-03
hsa04530:Tight junction	8	4.10	2.68E-03
hsa04724:Glutamatergic synapse	7	3.59	4.67E-03
hsa04970:Salivary secretion	6	3.08	6.40E-03
hsa05146:Amoebiasis	6	3.08	1.50E-02
hsa04024:cAMP signaling pathway	8	4.10	1.88E-02
hsa04151:PI3K-Akt signaling pathway	11	5.64	1.96E-02
hsa04015:Rap1 signaling pathway	8	4.10	2.51E-02
hsa04022:cGMP-PKG signaling pathway	7	3.59	2.64E-02
hsa05222:Small cell lung cancer	5	2.56	2.91E-02
hsa04360:Axon guidance	6	3.08	3.02E-02
hsa05014:Amyotrophic lateral sclerosis (ALS)	4	2.05	3.10E-02
hsa05213:Endometrial cancer	4	2.05	3.43E-02
hsa04977:Vitamin digestion and absorption	3	1.54	3.61E-02
hsa04923:Regulation of lipolysis in adipocytes	4	2.05	4.13E-02
hsa04750:Inflammatory mediator regulation of TRP channels	5	2.56	4.55E-02
hsa04915:Estrogen signaling pathway	5	2.56	4.69E-02
hsa05205:Proteoglycans in cancer	7	3.59	5.68E-02
hsa04720:Long-term potentiation	4	2.05	6.19E-02
hsa04725:Cholinergic synapse	5	2.56	6.61E-02
hsa04918:Thyroid hormone synthesis	4	2.05	7.13E-02
hsa04919:Thyroid hormone signaling pathway	5	2.56	7.14E-02
hsa04971:Gastric acid secretion	4	2.05	7.87E-02
hsa05010:Alzheimer's disease	6	3.08	8.10E-02
hsa05200:Pathways in cancer	10	5.13	9.00E-02

References

- [1] Omer An, Vera Pendino, Matteo D'Antonio, Emanuele Ratti, Marco Gentilini, and Francesca D' Ciccarelli. Ncg 4.0: the network of cancer genes in the era of massive mutational screenings of cancer genomes. *Database*, 2014:bau015, 2014.
- [2] P Andrew Futreal, Lachlan Coin, Mhairi Marshall, Thomas Down, Timothy Hubbard, Richard Wooster, Nazneen Rahman, and Michael R Stratton. A census of human cancer genes. *Nature Reviews Cancer*, 4(3):177–183, 2004.