Science Advances
AAAS

Supplementary Materials for

CancerVar: An artificial intelligence–empowered platform for clinical interpretation of somatic mutations in cancer

Quan Li, Zilin Ren, Kajia Cao, Marilyn M. Li, Kai Wang*, Yunyun Zhou*

*Corresponding author. Email: wangk@chop.edu (K.W.); zhouy6@chop.edu (Y.Z.)

Published 6 May 2022, *Sci. Adv.* **8**, eabj1624 (2022)
DOI: 10.1126/sciadv.abj1624

The PDF file includes:

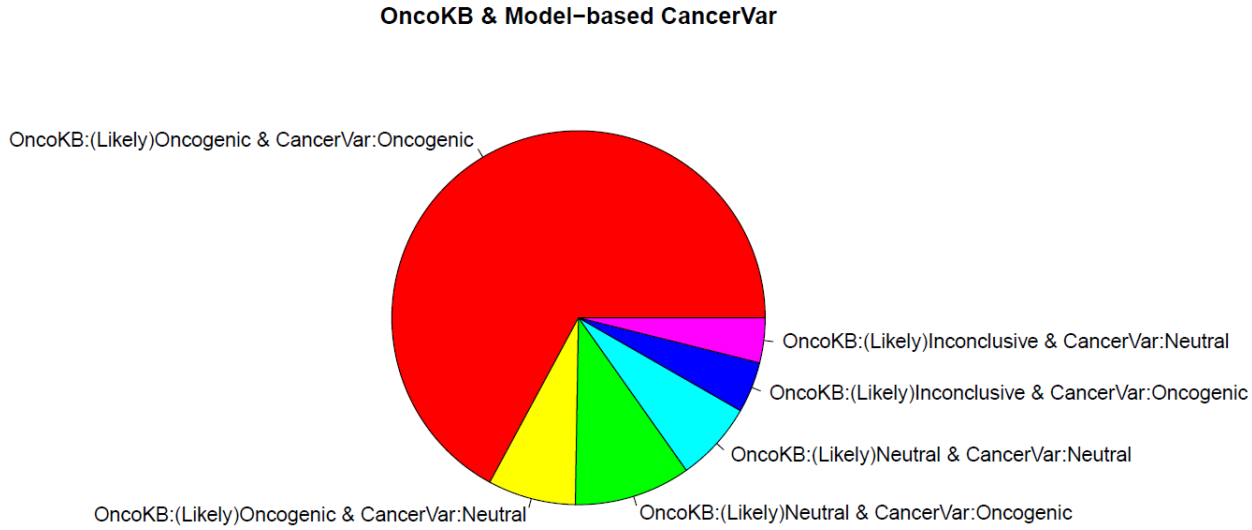
Figs. S1 to S4
Tables S1 and S2, S4 to S6
Legend for table S3

Other Supplementary Materials for this manuscript include the following:

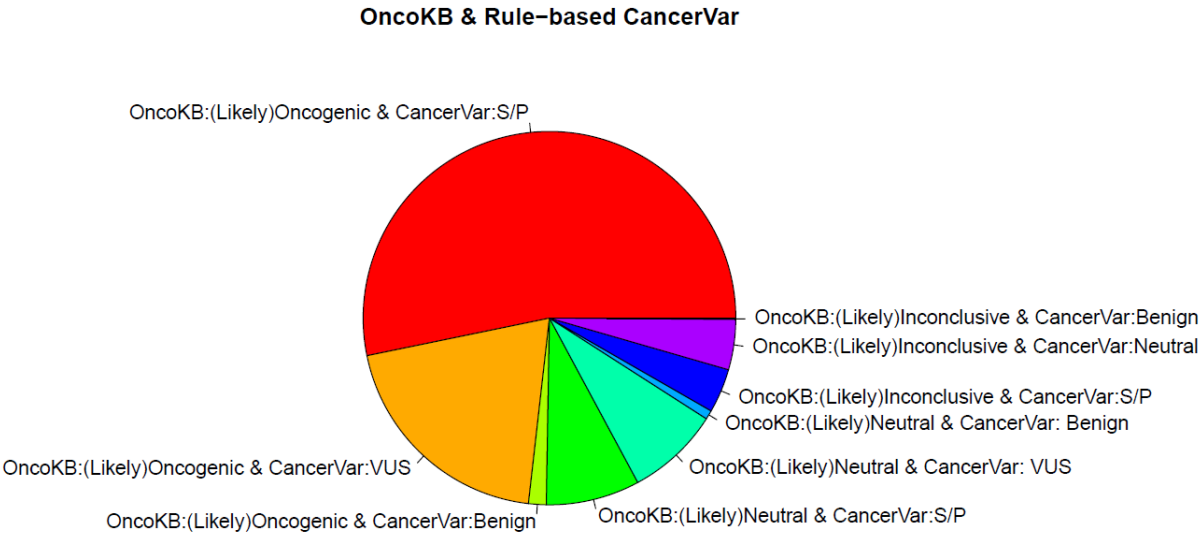
Table S3

Supplementary figures

Figure-S1 Pie chart of the comparison between CancerVar with database of OncoKB

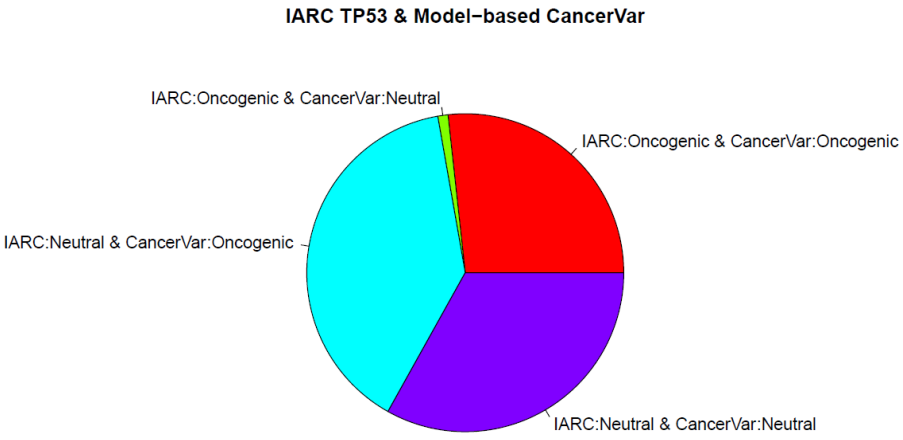


(a)

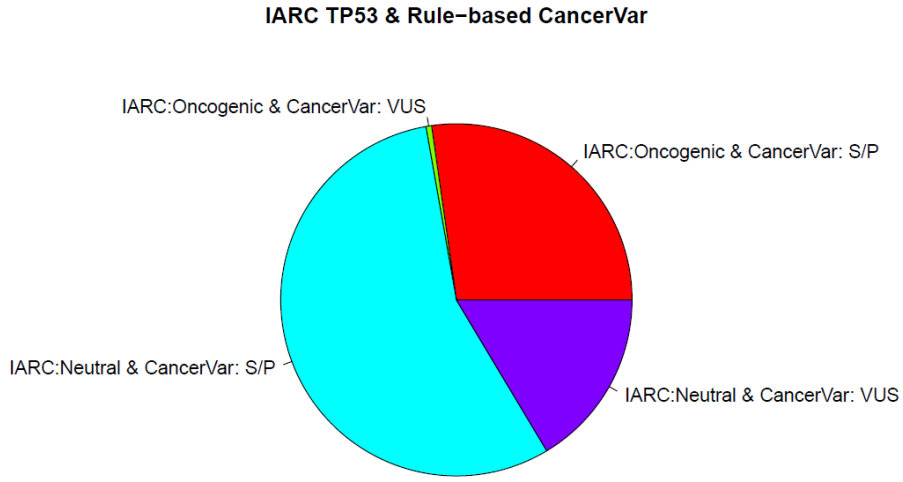


(b)

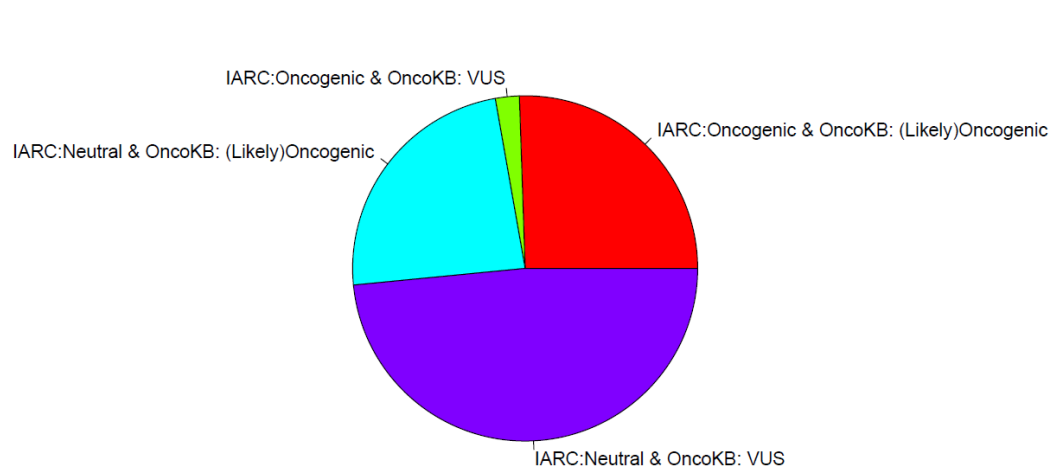
Figure-S2 Pie chart of the comparison between CancerVar, OncoKB with database of IARC TP53



(a)

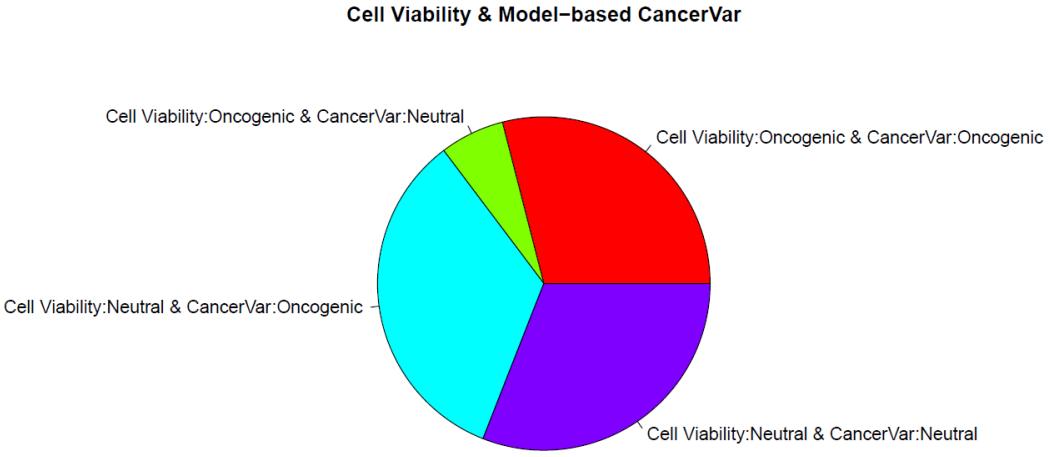


(b)

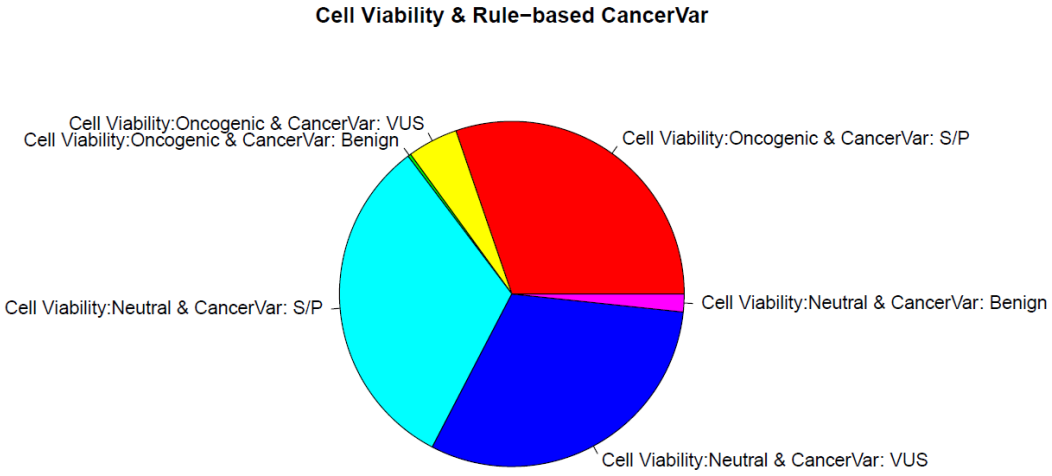


(c)

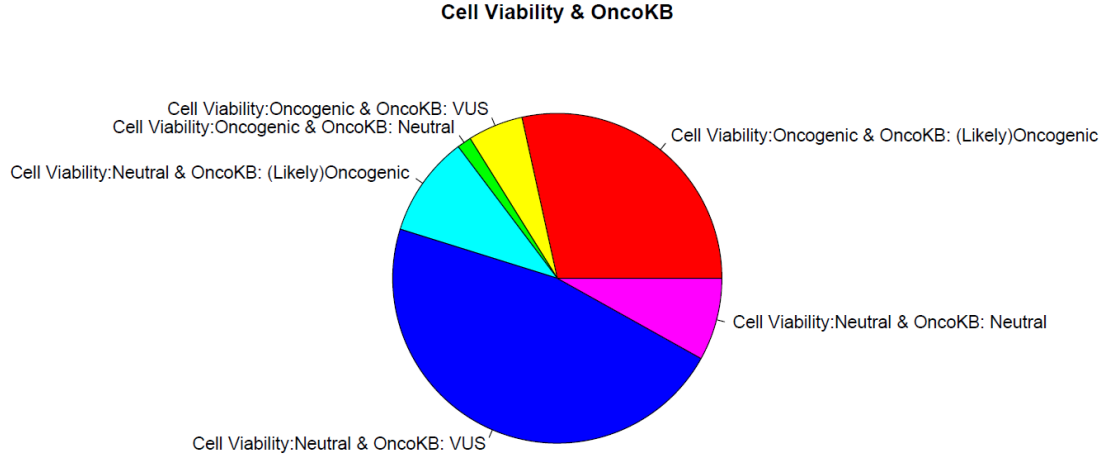
Figure-S3 Pie chart of the comparison between CancerVar,OncoKB with database of *in vitro* cell viability



(a)

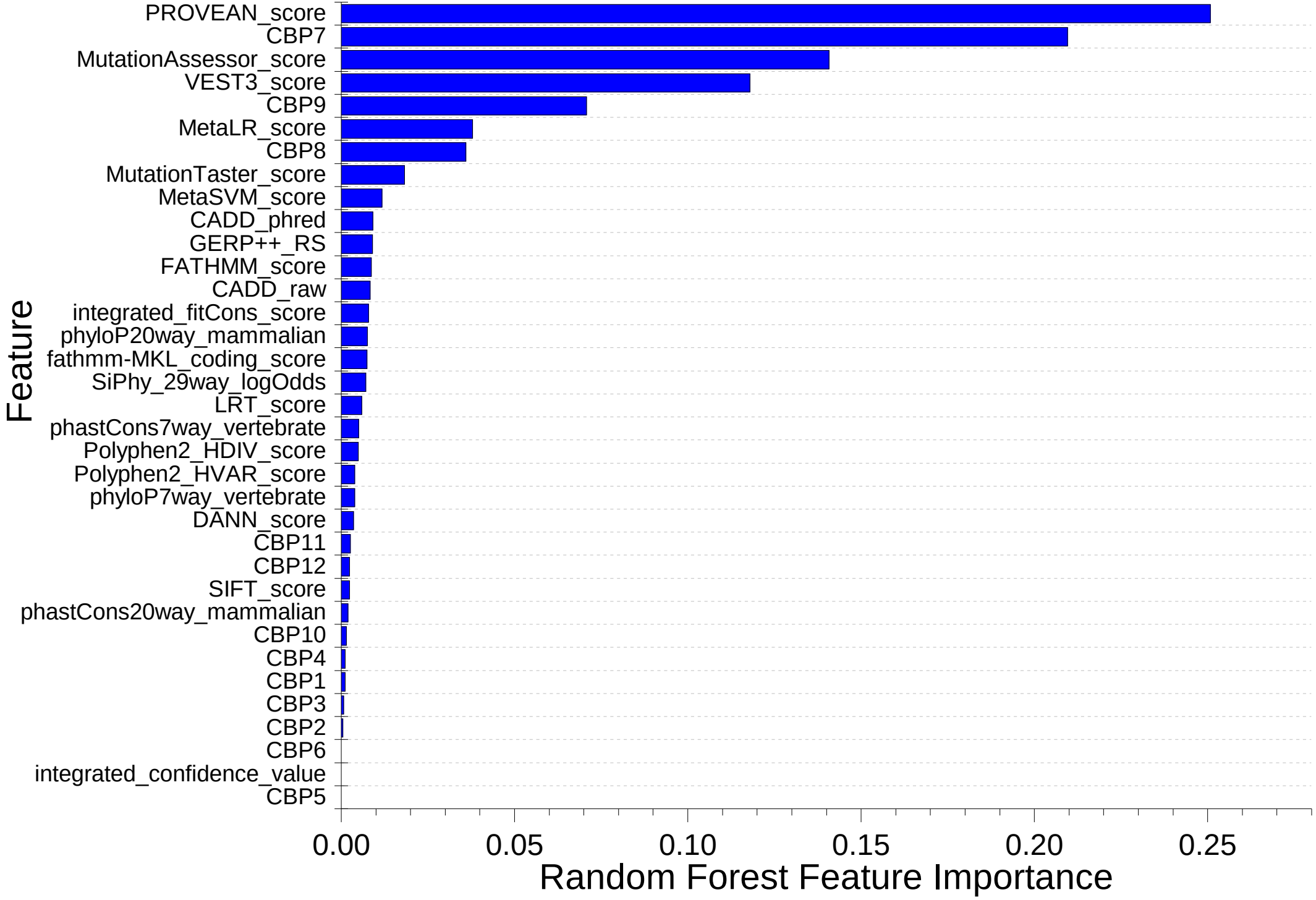


(b)



(c)

Figure-S4 Bar plot of Feature importance from Random Forest



Supplementary tables

Table-S1 Genes complied in CancerVar from 7 databases and 2 driver gene predictions

Source	Version/Date	Sum of unique genes	Notes
OncoKB	2020/05/11	1059	https://www.oncokb.org/cancerGenes
Cosmic	V91	723	https://cancer.sanger.ac.uk/cosmic/census,
CGI/ Cancer Genome Interpreter	2017/05/01	765	https://www.cancergenomeinterpreter.org/genes
IntoGen	2020/02/01	568	https://www.intogen.org/download
CIViC	2020/05/01	428	https://civicdb.org/releases
JAX CKB	CKB core	85	https://ckb.jax.org/gene/grid
PMKB	V1.4.6	610	https://pmkb.weill.cornell.edu/genes
Bailey et al. (2)	2018	299	https://www.cell.com/cell/references/S0092-8674(18)30237-X
MutPanning	2020	460	http://www.cancer-genes.org
ALL(CancerVar)	-	1911	http://cancervar.wglab.org/

Table-S2 CancerVar clinical evidence features and score system

Evidence type	Evidence code	Scores			
		2	1	0	-1
Therapeutic	CBP1	FDA approved or investigational with strong evidence	FDA approved or investigational with strong evidence in different tumour type	None	NA
Diagnostic	CBP2	In PG or reported evidence with consensus	not in PG but with convincing published data	None	NA
Prognostic	CBP3	In PG or reported evidence with consensus	not in PG but with convincing published data	None	NA
Mutation type	CBP4	NA	Activating, LOF (missense, nonsense, indel, splicing), CNAs, fusions	Functionally unknown; mostly missense, in-frame indels; less commonly, other types	NA
Variant frequencies	CBP5	NA	NA	NA	NA
Potential germline	CBP6	NA	NA	NA	NA
Population data	CBP7	NA	Absent or extremely low MAF in ESP, dbSNP, 1000Genome, ExAC, gnomAD	MAF>0.01% in the general population; or high MAF(>5%) in some ethnic populations	NA
Germline data	CBP8	NA	Clinvar/HGMD showed as (likely)Pathogenic		Clinvar/HGMD showed as (likely)benign

Somatic data	CBP9	Showed in >=2 Somatic database: : COSMIC, ICGC,TCGA and My Caner Genome	Showed in one of Somatic database: COSMIC, ICGC,TCGA and My Caner Genome	Do not show in any of Somatic database: COSMIC, ICGC,TCGA and My Caner Genome	NA
Predictive software	CBP10	>=6 predictions showed as pathogenic in total 7 Predictive softwares: SIFT, PolyPhen2, MutationAssessor, MetaLR, GERP++, MetaSVM, FATHMM	>=4 predictions showed as pathogenic in total 7 Predictive softwares: SIFT, PolyPhen2, MutationAssessor, MetaLR, GERP++, MetaSVM, FATHMM	The prediction is Variable or conflict	>=4 predictions showed as (likely)benign in 7 Predictive softwares
Pathway	CBP11	NA	Involve disease-associated pathways or pathogenic pathways	Do not Involve any disease-associated pathways nor pathogenic pathways	NA
Publication	CBP12	Publication reported pathogenic evidence with consensus	Publication reported pathogenic evidence for different tumour types or without consensus	None publication or no convincing evidence for pathogenic	Publications Reported evidence supportive of benign/likely benign

*PG: professional guidelines

Table-S3 48 variants classification comparisons between 20 reporters and CancerVar (showed as excel table)

Table-S4. CancerVar and OncoKB predictions on mutations from the IARC TP53 transactivation dataset.

IARC TP53 Transactivation	Model-based CancerVar		Rule-based CancerVar			OncoKB		
	Oncogenic	Neutral	S/P	VUS	benign	O/LO	VUS	Neutral
Oncogenic	512	20	522	10	0	489	43	0
Neutral	749	634	1069	314	0	455	928	0
Total	1261	654	1591	324	0	944	971	0

Table-S5. CancerVar and OncoKB predictions on mutations from the *in vitro* Cell Viability dataset by Ng. et al. (44)

Cell Viability (*in vitro*)	Model-based CancerVar		Rule-based CancerVar			OncoKB		
	Oncogenic	Neutral	S/P	VUS	benign	O/LO	VUS	Neutral
Oncogenic	208	45	217	34	2	204	39	10
Neutral	242	222	230	222	12	71	335	58
Total	450	267	447	256	14	275	374	68

Table-S6 CancerVar's predictions of whole exome mutations in 10 Oncogenes (ONC) and 10 Tumor suppressor genes (TSG).

Gene	TSG/ ONC	Total exome mutations	Oncogenic variants predicted by CancerVar	
			Number	Frequency
CDKN1A	TSG	1785	22	0.012
EP300	TSG	21735	226	0.010
CREBBP	TSG	21987	322	0.015
ARID1A	TSG	20568	97	0.0047
KDM6A	TSG	13086	55	0.0042
TP53	TSG	3789	467	0.12
NF1	TSG	25737	226	0.0088
NF2	TSG	5472	47	0.0086
APC	TSG	26091	109	0.0042
PIK3R1	TSG	6891	70	0.010
GNA11	ONC	3240	60	0.018
GNAQ	ONC	3240	46	0.014
GNAS	ONC	11997	132	0.011
EGFR	ONC	11661	153	0.013
ERBB2	ONC	11394	88	0.0077
ERBB4	ONC	11781	192	0.016
KRAS	ONC	2061	58	0.028
PIK3CA	ONC	9621	65	0.0067
AKT1	ONC	4329	32	0.0074
CTNNB1	ONC	7038	49	0.007