iScience, Volume *24*

Supplemental Information

A computational framework of host-based drug repositioning for broad-spectrum antivirals against RNA viruses

Zexu Li, Yingjia Yao, Xiaolong Cheng, Qing Chen, Wenchang Zhao, Shixin Ma, Zihan Li, Hu Zhou, Wei Li, and Teng Fei

Supplemental Figure 1

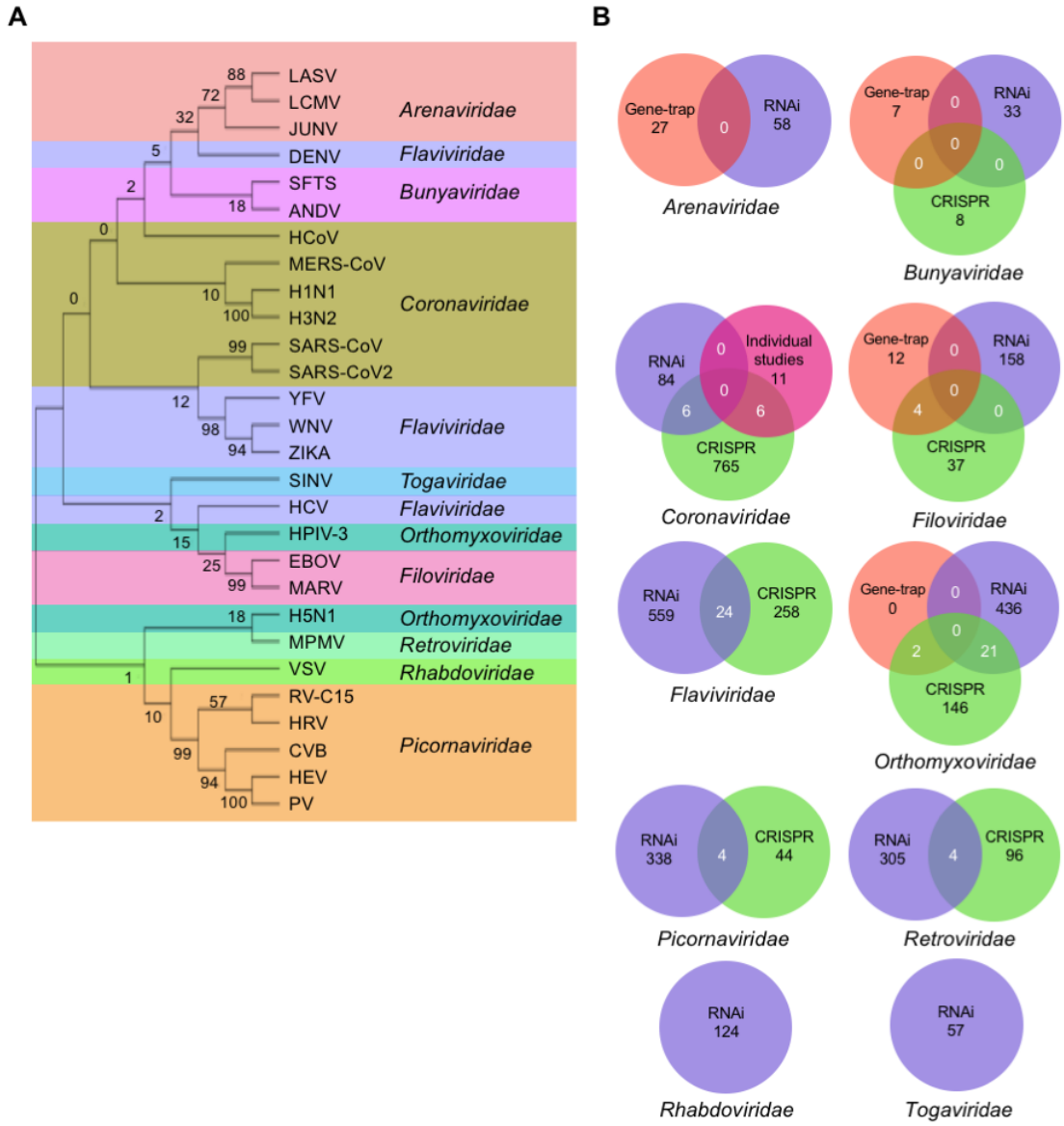

Figure S1. HDG collection for different RNA viruses, Related to Figure 2

(A) The phylogenetic tree for interrogated RNA viruses was constructed with protein sequence of viral RNA polymerase RdRp gene using maximum parsimony method.

(B) The venn diagrams of HDGs for indicated RNA virus families retrieved from different screening platforms.

Supplemental Figure 2

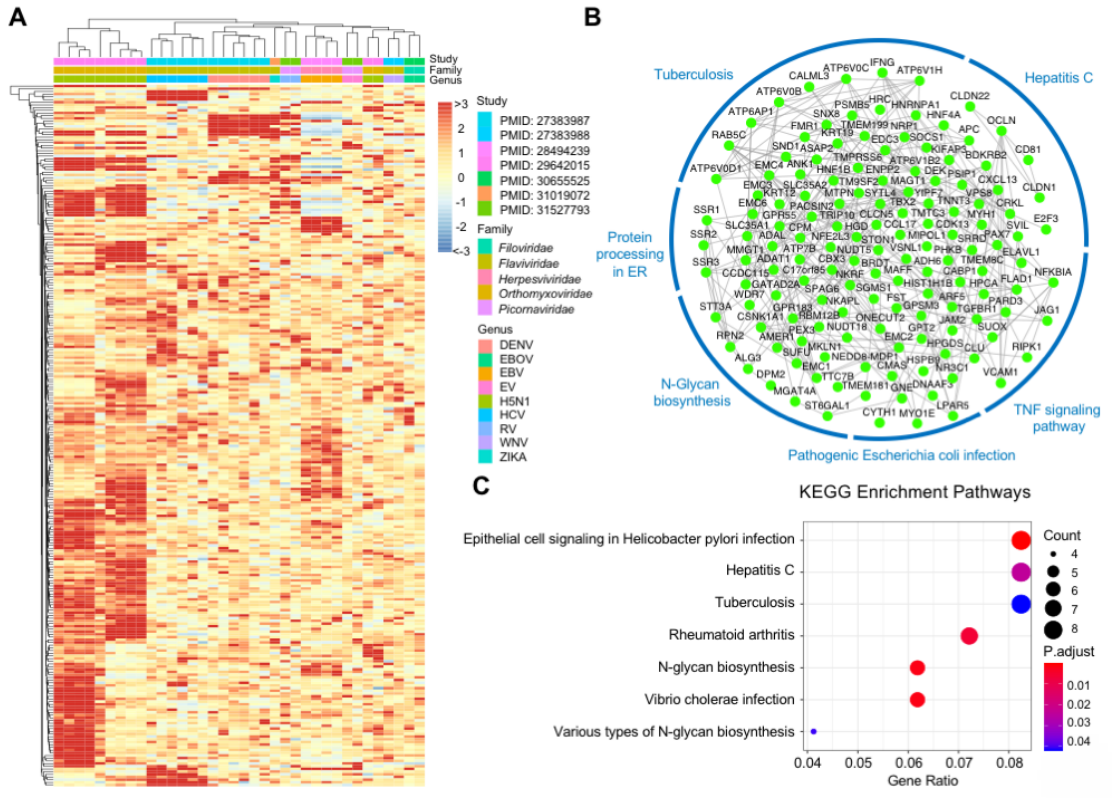

Figure S2. Re-analysis of CRISPR screening data for HDGs, Related to Figure 3

(A) The heatmap clustering of corresponding gene's β score calculated by MAGeCK-VISPR for multiple CRISPR screen studies related to HDG identification. HDG would have a high β score indicating a positive selection against corresponding virus challenge.

(B) The protein-protein interaction network for all the HDGs identified from re-analyzed CRISPR screens.

(C) Functional category enrichment analysis by KEGG for HDGs identified from re-analyzed CRISPR screens.

Supplemental Figure 3

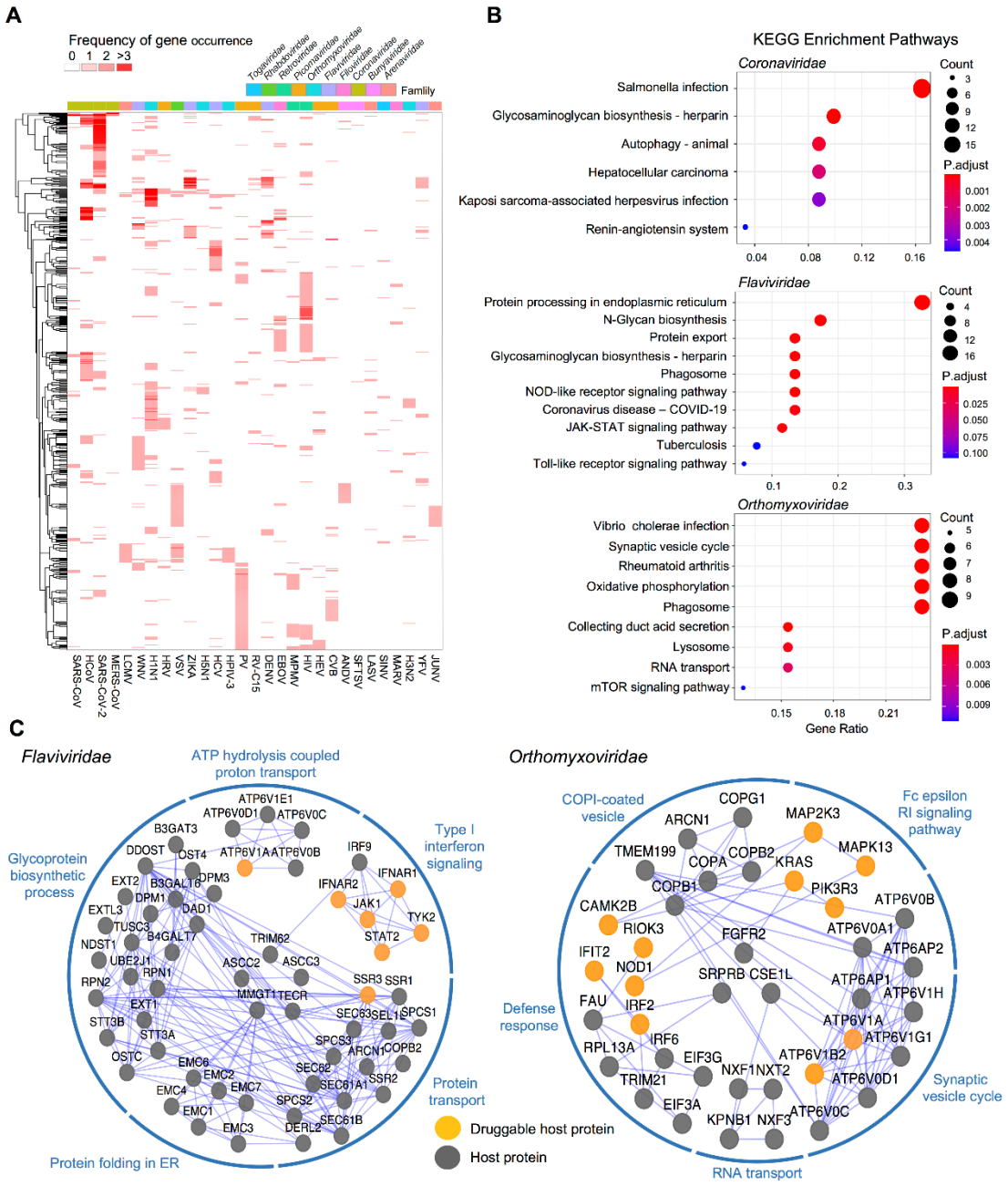

Figure S3. Comparative analysis and characterization of HDGs for indicated RNA virus families, Related to Figure 3

(A) The landscape of all the collected HDGs for indicated RNA viruses. The occurrence frequency of each HDG across studies was indicated by color legend.

(B) KEGG enrichment analysis of HDGs for the three indicated virus families. The size of the dot indicates the number of HDGs in the corresponding terms. The color of the dot represents the value of Benjamini and Hochberg FDR–adjusted p-value.

(C) The protein-protein interaction network of HDGs for *Flaviviridae* and *Orthomyxoviridae* virus families. Each HDG is presented as a node. The edge

between two nodes indicates a protein-protein interaction. The druggable HGDs with targeted drug candidates predicted in this study were highlighted.

Supplemental Figure 4

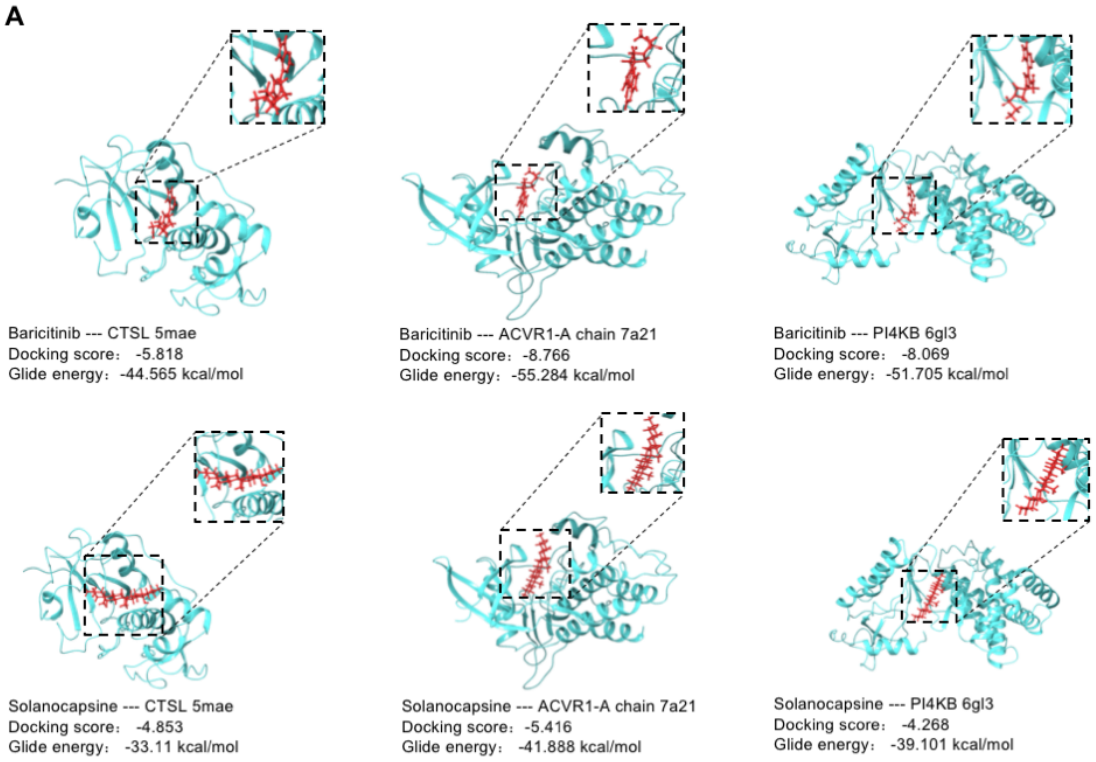

Figure S4. Molecular docking analysis for indicated drug-target pairs, Related to Figure 4

Molecular docking analysis showing the potential binding pockets between the repurposed drug Baricitinib and natural compound Solanocapsine with their corresponding targeted host factors.

Supplemental Tables

Table S1. Compendium of host dependency genes for multiple RNA viruses, Related to Figure 1 (attached dataset)

Table S2. Sequence sources for phylogenetic analysis, Related to Figure 2 (attached dataset)

Table S3. Re-analysis of CRISPR screening data, Related to Figure 3 (attached dataset)

Table S4. Functional gene enrichment analysis of host dependency genes, Related to Figure 3 (attached dataset)

Table S5. List of drug-target interactions and repurposed drug candidates, Related to Figure 4 (attached dataset)

Table S6. Summary of host dependency genes with repurposed drugs or natural compounds, Related to Figure 4

Virus Family	Druggable host dependence gene		High confidence host dependence gene
	Predicted	Known*	
Coronaviridae	29	32	166
Flaviviridae	79	8	81
Orthomyxoviridae	22	13	63

* Interaction Database: BindingDB, DGIdb3.0, DrugCentral, Stitch, ChEMBL

Table S7. Joint P-score ranking: the top ten repurposed FDA-approved drugs against *Flaviviridae* viruses, Related to Table 1

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint P-score
Alatrofloxacin	Bacterial infection	3086677	EMC2 RABGEF1 ARCN1 KRT31 COPB2 TRIM62 UGP2 SPCS3 DPM1 TEAD3	N/A	1.578
Fostamatinib	Chronic immune thrombocytopenia	11671467	NUAK2 TGFBR1 TYK2 JAK1 CHUK	SYK	1.240
Ozenoxacin	Impetigo	9863827	ARCN1 EMC2 RABGEF1 UGP2 COPB2 RPS6KL1 TRIM62 DPM1 KRT31 CKAP5	GyrA	1.191
Grepafloxacin	Bacterial infection	72474	UGP2 CKAP5 IFNAR1 SPCS3 SEC62 NOP56 UBE2J1 RPS6KL1 SSR2 OST4	N/A	1.091
Roxithromycin	Bacterial infection	6915744	SSR1 SEC62 SEC63 SSR2 CD81 IFNAR1 EMC3 DPM3 STT3A B3GALT6	N/A	0.810
Gramicidin D	Bacterial infection	45267103	DERL2 B4GALT7 EMC4 MMGT1 SEC61B OST4 DDOST DAD1 TUSC3 VMP1	N/A	0.732
Triclosan	Bacterial infection	5564	SEC61A1 VMP1 OST4 CD81 TUSC3 DAD1 DERL2 TMEM41B MMGT1 B3GALT6	N/A	0.700
Tofacitinib	Rheumatoid arthritis	9926791	TYK2 JAK1 NUAK2 TGFBR1 CHUK	JAK3 JAK2 JAK1	0.697
Omadacycline	Bacterial infection	54697325	UGP2 NOP56 CKAP5 ARCN1 COPB2 RPN1 CD81 SEC62 SPCS3 RPN2	N/A	0.600
Lasofoxifene	Osteoporosis	216416	ATP6V1E1 ATP6V0C ATP6V0D1 ATP6V0B ATP6V1A CHUK	ESR1、ESR2	0.597

N/A: Not applicable

Table S8. Joint PN-score ranking: the top ten repurposed FDA-approved drugs against *Flaviviridae* viruses, Related to Table 2

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint PN-score
Alatrofloxacin	Bacterial infection	3086677	EMC2 RABGEF1 ARCN1 KRT31 COPB2 TRIM62 UGP2 SPCS3 DPM1 TEAD3	N/A	1.566
Ozenoxacin	Impetigo	9863827	ARCN1 EMC2 RABGEF1 UGP2 COPB2 RPS6KL1 TRIM62 DPM1 KRT31 CKAP5	N/A	1.181
Grepafloxacin	Bacterial infection	72474	UGP2 CKAP5 IFNAR1 SPCS3 SEC62 NOP56 UBE2J1 RPS6KL1 SSR2 OST4	N/A	1.083
Roxithromycin	Bacterial infection	6915744	SSR1 SEC62 SEC63 SSR2 CD81 IFNAR1 EMC3 DPM3 STT3A B3GALT6	N/A	0.804
Gramicidin D	Bacterial infection	45267103	DERL2 B4GALT7 EMC4 MMGT1 SEC61B OST4 DDOST DAD1 TUSC3 VMP1	N/A	0.727
Triclosan	Bacterial infection	5564	SEC61A1 VMP1 OST4 CD81 TUSC3 DAD1 DERL2 TMEM41B MMGT1 B3GALT6a	N/A	0.695
Omadacycline	Bacterial infection	54697325	UGP2 NOP56 CKAP5 ARCN1 COPB2 RPN1 CD81 SEC62 SPCS3 RPN2	N/A	0.596
Lasofoxifene	Osteoporosis	216416	ATP6V1E1 ATP6V0C ATP6V0D1 ATP6V0B ATP6V1A CHUK	ESR1、ESR2	0.550
Candicidin	Bacterial infection	10079874	KAT8 WDR7 STAT2 UBE2J1 USP11 ASCC2 SEC63 RABGEF1 SEL1L TRIM62	N/A	0.536
Troleandomycin	Bacterial infection	202225	CKAP5 USP11 SEC62 SEC63 EMC7 B3GALT6 SSR1 SSR2 OSTC SSR1	N/A	0.392

N/A: Not applicable

Table S9. DeepCPI P-score ranking: the top ten repurposed natural compounds against *Flaviviridae* viruses, Related to Table 3

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI P-score
Lysergol	MOL005261	14987	Semen Pharbitidis	EMC2 ARCN1 RPN1 ATP6V1E1 TEAD3 COPB2 AAR2 RABGEF1 ATP6V0D1 IRF9	3.918
Atropine	MOL002219	174174	Lycii Cortex, Hyoscyami Semen	TEAD3 AAR2 PAPSS1 ARCN1 RPN1 COPB2 ATP6V0D1 IRF9 ASCC2 KAT8	3.591
Costaclavine	MOL008145	160462	Ricini Semen	ATP6V1E1 EMC2 ATP6V0D1 SEC61A1 RPN1 ARCN1 RPN2 HSPA13 DPM3 AAR2	3.545
Solanocapsine	MOL007356	73419	Solanum Nigrum	EMC2 RABGEF1 ATP6V1E1 TEAD3 KRT31 STAT2 ATP6V0D1 TRIM62 DPM1 KAT8	3.389
Chanoclavine	MOL005260	5281381	Semen Pharbitidis	RPN2 ATP6V1E1 ARCN1 RPN1 AAR2 EMC2 TEAD3 ATP6V0D1 KAT8 HSPA13	3.321
Triptofordin B1	MOL003232	122391803	Tripterygii Radix	IRF9 SEC61A1 RPN2 TEAD3 ASCC2 TMEM41B B4GALT7 IL4I1 STAT2 NDST1	2.931
Penniclavin	MOL005257	115247	Semen Pharbitidis	EMC2 ATP6V0D1 ARCN1 HSPA13 ATP6V0C SPCS3 TPT1 TMEM41B RPN2 UGP2	2.620
Picrasidine D	MOL012140	5316876	Asari Radix et Rhizoma	KAT8 RABGEF1 AAR2 TEAD3 STAT2 ARCN1 ASCC2 IRF9 TRIM62 RPS6KL1	2.806
elymoclavine	MOL005267	16758153	Semen Pharbitidis	EMC2 ATP6V0D1 ARCN1 SEC61A1 HSPA13 RPN1 DPM3 SPCS3 TPT1 TMEM41B	2.519
13,14-dehydrosophoridine	MOL006573		Sophorae Flavescentis Radix	IRF9 ARCN1 AAR2RABGEF1 STAT2 COPB2 TRIM62 PAPSS1 KAT8 RABEPK	2.614

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S10. DeepCPI PN-score ranking: the top ten repurposed natural compounds against *Flaviviridae* viruses, Related to Table 4

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI PN-score
N-[6-(9-acridinylamino)hexyl]benzamide	MOL005935	146515	Sophora Japonica L.	NDST1 TMEM41B DERL2 VMP1 MMGT1 EMC3 ATP6V0D1 RPN2 TECR EMC1	0.535
14b-pregnane	MOL009604	176992	Lycii Fructus	ATP6V0D1 SPCS2 SPCS3 UGP2 ASCC3 MMGT1 CHUK VMP1 EMC2 DPM1	0.465
(S)-Canadine	MOL001455	21171	Chelidonii Herba	SEC61A1 ATP6V0C RPN2 TMEM41B ATP6V0B DERL2 HSPA13 VMP1 MMGT1 SPCS3	0.460
(R)-Canadine	MOL002903	443422	Coptidis Rhizoma	SEC61A1 ATP6V0C RPN2 TMEM41B ATP6V0B DERL2 HSPA13 VMP1 MMGT1 SPCS3	0.460
Androstane	MOL003790	6857536	Styrax	MMGT1 SPCS3 SPCS2 VMP1 CHUK ASCC3 HSPA13 UGP2 SEC61A1 TMEM41B	0.407
Narcein	MOL009329	8564	Papaveris Pericarpium	DERL2 OST4 NDST1 EXT2 TECR MMGT1 ATP6V0B SSR2 CD81 STT3A	0.404
malkangunin	MOL005360	90473155	Stemonae Radix	EMC1 PAPSS1 COPB2 EXTL3 ATP6V1A AAR2 ASCC3 SEL1L NDST1 USP11	0.390
alpha-berbine	MOL008182	164543	Dichroae Radix	SEC61A1 TMEM41B ATP6V0C ATP6V0B DERL2 SEC61B OST4 MMGT1 CD81 SPCS3	0.363
yohimbine	MOL008488	8969	Uncariae Ramulus Cumuncis	SEC61A1 DERL2 TMEM41B ATP6V0C VMP1 ATP6V0B TECR MMGT1 SPCS3 SEC61B	0.359
coryincine	MOL008635	3058605	Uncariae Ramulus Cumuncis	SEC61A1 DERL2 TMEM41B ATP6V0C VMP1 ATP6V0B TECR MMGT1 SPCS3 SEC61B	0.359

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S11. Joint P-score ranking: the top ten repurposed FDA-approved drugs against *Orthomyxoviridae* viruses, Related to Table 1

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint P-score
Fostamatinib	Chronic immune thrombocytopenia	11671467	CAMK2B CLK3 FGFR2 GSK3A MAP2K3 MAPK13 RIOK3	SYK	2.724
Lasofoxifene	Osteoporosis	216416	ATP6AP1 ATP6V0A1 ATP6V0B ATP6V0C ATP6V0D1 ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1 ESR2	1.259
Baricitinib	Rheumatoid arthritis	44205240	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	JAK1 JAK2 JAK3 TYK2	0.869
Tofacitinib	Rheumatoid arthritis	9926791	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13 FGFR2	JAK3 JAK2 JAK1	0.859
Lusutrombopag	Thrombocytopenia	49843517	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	MPL	0.814
Bosutinib	Chronic Myelogenous Leukemia	5328940	MAPK13 ATP6V1H ATP6V1G1 ATP6AP1 ATP6V0D1 ATP6V1A	SRC ABL1	0.799
Raloxifene	Breast cancer	5035	ATP6AP1 ATP6V0B ATP6V0C ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1、ESR2	0.787
Etoricoxib	Rheumatoid arthritis	123619	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	COX2	0.770
Hydroxychloroquine	Rheumatoid arthritis	3652	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	N/A	0.667
Eltrombopag	Thrombocytopenia	135449332	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 ATP6V1A FGFR2	CD110	0.637

N/A: Not applicable

Table S12. Joint PN-score ranking: the top ten repurposed FDA-approved drugs against *Orthomyxoviridae* viruses, Related to Table 2

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint PN-score
Fostamatinib	Chronic immune thrombocytopenia	11671467	CAMK2B CLK3 FGFR2 GSK3A MAP2K3 MAPK13 RIOK3	SYK	1.751
Lasofoxifene	Osteoporosis	216416	ATP6AP1 ATP6V0A1 ATP6V0B ATP6V0C ATP6V0D1 ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1、ESR2	1.212
Raloxifene	Breast cancer	5035	ATP6AP1 ATP6V0B ATP6V0C ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1、ESR2	0.724
Lusutrombopag	Thrombocytopenia	49843517	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	MPL	0.632
Baricitinib	Rheumatoid arthritis	44205240	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	JAK1 JAK2 JAK3 TYK2	0.617
Etoricoxib	Rheumatoid arthritis	123619	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	COX2	0.553
Vancomycin	Bacterial infection	14969	ATP6V1A ATP6AP1 ATP6V0B ATP6V0C ATP6V0D1 ATP6V1G1 ATP6V1H ATP6V1B2 PGD MAPK13	N/A	0.526
Eltrombopag	Thrombocytopenia	135449332	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 ATP6V1A FGFR2	CD110	0.495
Hydroxychloroquine	Rheumatoid arthritis	3652	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	N/A	0.495
Ibandronate	Osteoporosis	60852	ATP6V0A1 ATP6V1H	FPPS	0.459

N/A: Not applicable

Table S13. DeepCPI P-score ranking: the top ten repurposed natural compounds against *Orthomyxoviridae* viruses, Related to Table 3

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI P-score
Lysergol	MOL005261	14987	Semen Pharbitidis	ARCN1 ATP6V0D1 COPB2 COPG1 FAU IFIT5 IVNS1ABP KPNB1 NXT2 PIK3R3	4.410
Atropine	MOL002219	174174	Lycii Cortex, Hyoscyami Semen	ARCN1 ATP6V0D1 COPB2 COPG1 EIF3G IFIT5 IRF6 IVNS1ABP PGD PIK3R3	3.988
Costaclavine	MOL008145	160462	Ricini Semen	ARCN1 ATP6V0D1 COPG1 EIF3A FAU IFIT5 KPNB1 NXT2 PIK3R3 SLC1A3	3.935
Chanoclavine	MOL005260	5281381	Semen Pharbitidis	ARCN1 ATP6V1B2 COPG1 EIF3A EIF3G FAU IFIT5 KPNB1 NXT2 PIK3R3	3.868
Solanocapsine	MOL007356	73419	Solanum Nigrum	ATP6V0D1 COPB2 COPG1 EIF3A IFIT5 KPNB1 NXT2 PIK3R3 RIOK3 TRIM21	3.770
Penniclavin	MOL005257	115247	Semen Pharbitidis	EIF3A KPNB1 ATP6V0D1 SLC1A3 ARCN1 IFIT5 STARD5 FAU IFIT2 ATP6V0C	3.151
Picrasidine D	MOL012140	5316876	Asari Radix et Rhizoma	IRF6 TRIM21 ARCN1 COPG1 PHF3 AKAP13 EIF3A PIK3R3 SF3A1 NXF1	3.149
elymoclavine	MOL005267	440904	Semen Pharbitidis	ARCN1 ATP6V0D1 EIF3A FAU IFIT2 IFIT5 KPNB1 PIK3R3 SLC1A3 STARD5	3.099
(-)-9alpha-hydroxysophoramine	MOL006570	50695119	Sophorae Flavescentis Radix	IRF6 IVNS1ABP EIF3G PIK3R3 COPG1 BUB3 NXF3 SF3A1 EIF3A ARCN1	3.006
Triptofordin B1	MOL003232	122391803	Tripterygii Radix	IVNS1ABP IRF6 COPG1 ATP6V0A1 TRIM21 NXT2 ATP6AP2 EIF3A NXF1 SLC1A3	2.984

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S14. DeepCPI PN-score ranking: the top ten repurposed natural compounds against *Orthomyxoviridae* viruses, Related to Table 4

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI PN-score
Deoxycamptothecine	MOL008209	169724	Andrographis Herba	ARCN1 ATP6V0D1 ATP6V1G1 ATP6V1H EIF3A EIF3G FAU IFIT2 IFIT5 NXT2	0.908
Lysergol	MOL005261	14987	Semen Pharbitidis	ARCN1 ATP6V0D1 COPB2 COPG1 FAU IFIT5 IVNS1ABP KPNB1 NXT2 PIK3R3	0.690
Chanoclavine	MOL005260	5281381	Semen Pharbitidis	ARCN1 ATP6V1B2 COPG1 EIF3A EIF3G FAU IFIT5 KPNB1 NXT2 PIK3R3	0.685
Costaclavine	MOL008145	160462	Ricini Semen	ARCN1 ATP6V0D1 COPG1 EIF3A FAU IFIT5 KPNB1 NXT2 PIK3R3 SLC1A3	0.608
elymoclavine	MOL005267	440904	Semen Pharbitidis	ARCN1 ATP6V0D1 EIF3A FAU IFIT2 IFIT5 KPNB1 PIK3R3 SLC1A3 STARD5	0.580
Solanocapsine	MOL007356	73419	Solanum Nigrum	ATP6V0D1 COPB2 COPG1 EIF3A IFIT5 KPNB1 NXT2 PIK3R3 RIOK3 TRIM21	0.562
Penniclavin	MOL005257	115247	Semen Pharbitidis	EIF3A KPNB1 ATP6V0D1 SLC1A3 ARCN1 IFIT5 STARD5 FAU IFIT2 ATP6V0C	0.531
Vitexifolin C	MOL011912	11033408	Viticis Fructus	KPNB1 FAU ATP6V0D1 MAP2K3 FGFR2 EIF3A IFIT2 COPB2 ATP6V0B CMAS	0.428
Atropine	MOL002219	174174	Lycii Cortex, Hyoscyami Semen	ARCN1 ATP6V0D1 COPB2 COPG1 EIF3G IFIT5 IRF6 IVNS1ABP PGD PIK3R3	0.426
Isolimonic acid	MOL013443	131752314	Aurantii Fructus Immaturus	IFIT5 ARCN1 NXF1 RIOK3 KPNB1 FAU USP46 EIF3A IFIT2 ATP6V0D1	0.420

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S15. Key parameters of molecular docking analysis, Related to Figure 4

PubChem CID	Drug or natural compound	PDB ID	Target	Docking score	Glide energy (kcal/mol)
44205240	Baricitinib	5MAE	CTSL	-5.818	-44.565
44205240	Baricitinib	6EIS	DYRK1A-A chain	-9.431	-48.038
44205240	Baricitinib	6EIS	DYRK1A-B chain	-8.288	-48.185
44205240	Baricitinib	6EIS	DYRK1A-C chain	-8.289	-46.526
44205240	Baricitinib	6EIS	DYRK1A-D chain	-6.793	-44.422
44205240	Baricitinib	6GL3	PI4KB	-8.069	-51.705
44205240	Baricitinib	7A21	ACVR1-A chain	-8.766	-55.284
44205240	Baricitinib	7A21	ACVR1-B chain	-8.413	-50.806
73419	Solanocapsine	5MAE	CTSL	-4.853	-33.11
73419	Solanocapsine	6EIS	DYRK1A-A chain	-4.128	-34.175
73419	Solanocapsine	6EIS	DYRK1A-B chain	-4.297	-37.025
73419	Solanocapsine	6EIS	DYRK1A-C chain	-4.961	-29.42
73419	Solanocapsine	6EIS	DYRK1A-D chain	-5.011	-34.605
73419	Solanocapsine	6GL3	PI4KB	-4.268	-39.101
73419	Solanocapsine	7A21	ACVR1-A chain	-5.416	-41.888
73419	Solanocapsine	7A21	ACVR1-B chain	-5.413	-42.105
14987	Lysergol	6FYV	CLK4	-7.132	-23.175
5316876	Picrasidine D	6FYV	CLK4	-6.939	-31.901
154417	Hyoscyamine	6FYV	CLK4	-6.279	-18.605
14987	Lysergol	5T4E	DPP4	-6.196	-31.612
154417	Hyoscyamine	5T4E	DPP4	-6.064	-29.032

Table S16. 2D structures of the top drug candidates, Related to Tables 1-4 (attached dataset)

Transparent Methods

Host dependency gene collection and literature mining

By systematically searching the literature to date, studies performing genetic screening for human-specific HDGs corresponding to RNA viruses were collected. Screens for DNA viruses or in non-human cells were not included with an exception for SARS-CoV-2 virus-related screens. We collected all the recently published viral resistance CRISPR screens against SARS-CoV-2 virus, with 5 studies in human cells and 1 study in Vero-E6 cells (Table S1). Under this criteria, data from 63 studies with different genetic perturbation techniques (CRISPR knockout, RNAi and haploid gene-trap mutagenesis) were collected. These studies identified virus-specific HDGs for 29 RNA viruses spanning 10 RNA virus families. Due to the high interest for *Coronaviridae* virus family, we collected additional 34 individual gene-focused non-screening studies to include as many *Coronaviridae* HDGs as possible. A gene is defined as a HDG when it meets any of the following criteria: 1) Its loss-of-function impedes or reduces viral infection or activity by experimental evidence in non-screen studies; 2) It has been clearly classified into HDG group in screen studies; 3) When HDG group is not specified in screen studies, we took the top ~5% of all the interrogated genes in the positive selection list as HDGs with a custom log fold change cutoff in CRISPR knockout or RNAi screens challenged by the corresponding virus. The detailed information concerning to these literatures and HDGs was summarized in Table S1. For *Coronaviridae*, *Flaviridae* and *Orthomyxoviridae* viruses, we only took a subset of HDGs that occurred more than once within its corresponding family as high confidence HDGs for further analysis. In general, around one hundred HDGs for each group of the above three virus families were used for molecular characterization and drug repurposing analysis (Table S6).

Phylogenetic tree construction

The sequences of nucleic acid and protein corresponding to viral RNA-dependent RNA polymerase (RdRp) gene for indicated RNA viruses were downloaded from online sources (https://www.ncbi.nlm.nih.gov) and were used for phylogenetic tree analysis (Table S2). The nucleic acid and protein sequences were analyzed by Multiple Sequence Alignment in Muscle calculation using MEGA X software. The phylogenetic tree was subsequently constructed based on neighbor-joining (NJ) method or maximum parsimony (MP) method using pairwise phylogenetic distance with 1000 bootstrap replicates.

Re-analysis of CRISPR screening data

Among the 25 CRISPR screening studies, we downloaded the raw sequencing or read count data from 7 studies wherever these raw data were available. We re-analyzed these CRISPR screening data to re-call the HDGs using the same

MAGeCK-VISPR pipeline (Li et al., 2015). In total, 36 samples across the 9 viruses are included in the analysis. The beta scores of each screening, generated by MAGeCK-VISPR, were combined together and normalized using quantile normalization. Next, we filtered the data using the following two thresholds: First, the maximum of the beta score of a gene across all the samples must be greater than 3. Second, the average beta score of a gene across all the samples must be greater than 1. After filtering, 261 genes were retained as positively selected HDG hits. Then hierarchical clustering and protein-protein interaction network was performed using StringDB.

KEGG and GO enrichment analysis

The high confidence HDGs for *Coronaviridae*, *Flaviridae* and *Orthomyxoviridae* viruses (166, 81 and 63, respectively) were used for this analysis (Table S6). KEGG and GO enrichment analysis were performed using clusterProfiler R package with a strict cutoff of p-value < 0.001 and false discovery rate (FDR) < 0.05 (Yu et al., 2012). Enrichment analyses were visualized using the R package clusterProfiler with default settings.

Network analysis

The input HDGs were uploaded to the STRING database (version 11.0, https://string-db.org) and high confidence protein-protein interactions (PPIs) were extracted with a minimum required interaction score ≥ 0.7. Next, the interactions were imported into Cytoscape 3.2.1 software to visualize PPI Network. The druggable HDG-encoding proteins with predicted drug candidates in this study and proteins classified into certain functional protein complexes or biological processes are highlighted.

Drug candidate selection for repurposing

FDA-approved drug information was extracted from DrugBank database (version 5.1.7, released 2020-07-02; https://www.drugbank.ca) corresponding to 2352 marketed drugs with InChI (the IUPAC International Chemical Identifier) key information. Natural compound information is downloaded from Traditional Chinese Medicine Systems Pharmacology (TCMSP) online database (version 2.3, released 2014-05-31; https://tcmspw.com/tcmsp.php) which is a unique systems pharmacology platform of Chinese herbal medicines (Ru et al., 2014). To select the most favorable compound candidates, we filtered the pool of 1455 natural compounds by requiring each candidate passing the criteria of oral bioavailability (OB) ≥ 30.0 %, drug-likeness (DL) ≥ 0.18 and blood-brain barrier (BBB) ≥ -0.30, and finally ended up with 1062 selected natural compounds for the downstream DTI analysis.

DTI retrieval from related databases

Known drug-target interactions were extracted according to annotated information associated with related drugs, compounds or target genes from multiple databases including BindingDB (updated 2020-03-01), DGIdb3.0 (version 3.0.2), DrugCentral (version 10.12) and Stitch (version 5.0) (Cotto et al., 2018; Gilson et al., 2016; Kuhn et al., 2010; Ursu et al., 2019). The high confidence HDGs for *Coronaviridae*, *Flaviridae* and *Orthomyxoviridae* viruses were used for the DTI analysis (Table S6). One HDG may be associated with multiple drugs or compounds. Only FDA-approved drugs and selected natural compounds were considered for compiling these known DTI information for drug repurposing.

DTI prediction by DeepCPI

The source code of DeepCPI can be downloaded from https://github.com/FangpingWan/DeepCPI. The binding activity score for each drug-target pair was predicted by providing the InChl key information of a drug or compound and the amino acid sequence of a protein target from UniProt database. We applied DeepCPI on 4,563 high confidence DTIs out of 7,444,710 putative pairs (3,030 druggable proteins and 2,457 FDA-approved drugs) extracted from DGIdb3.0 database (version 3.0.2) as a benchmark analysis and determined an optimal threshold with a normalized z-score ≥ 0.641 (sensitivity: 73%; specificity: 51.9%) by receiver operating characteristics (ROC) analysis. We then used this cutoff to filter confident DTI in our analysis for virus-related HDGs and FDA-approved drugs as well as selected natural compounds.

DTI prediction by DTINet

The source code of DTINet can be downloaded from https://github.com/luoyunan/DTINet. The drug-protein interactions and protein-protein interactions were extracted from UniProt database. The drug-disease associations and protein-disease associations were extracted from the Therapeutic Target Database (Wang et al., 2020). The drug-drug interactions were extracted from the BioSNAP Network database (http://snap.stanford.edu/biodata/). Then the Jaccard similarity for these interactions/associations was calculated to further augment the heterogeneity. A heterogeneous network (including three types of nodes and five types of edges) are constructed using these diverse drug-related and protein-related information for the prediction task. The informative, but low-dimensional feature vector was obtained by integrating the diverse information from the heterogeneous network by combining the network diffusion algorithm (random walk with restart, RWR) with a dimensionality reduction scheme (diffusion component analysis, DCA). The restart probability is set to 0.50 and the maximum number of iterations is set to 20. Intuitively, the low-dimensional feature vector is used to encode the relational properties (e.g., similarity), association information and topological context of each drug (or protein) node

in the heterogeneous network. Finally, the score for each drug-protein pair was calculated based on the feature vectors by DTINet default parameters. Similar to DeepCPI analysis, we also applied DTINet on the benchmark datasets and determined an optimal threshold with a normalized z-score ≥ 0.973 (sensitivity: 88.9%; specificity: 63.8%) by ROC analysis. We then use this cutoff to filter confident DTI in our analysis for virus-related HDGs and FDA-approved drugs. Due to the insufficient prior data for proper modeling, DTINet was not applied for natural compound DTI analysis.

Prioritizing repurposed drug candidates

The repurposed FDA-approved drugs were prioritized by both known DTI and predicted DTI with high confidence. The candidate drugs were ranked by predicted DTI scores with known DTI annotation accompanied to the drug if any. We adopted two ranking methods to prioritize these candidates. The first ranking method only considers the HDG target-associated DTIs. For FDA-approved drugs with both DeepCPI and DTINet DTI prediction, we extracted mutual confident DTIs by both prediction algorithms and the mean of normalized z-score by each prediction tool was calculated as a positive score (P-score). A joint P-score by the sum of DeepCPI and DTINet P-score was employed to rank the drug candidates. The second ranking method not only considers HDG targets, but also incorporates non-HDG targets and common essential gene targets to evaluate drug promiscuousness and cytotoxicity effects. In addition to P-score, we introduced a negative score for DTIs between a given drug and non-HDG (among 3,030 druggable proteins in DGIdb3.0 database) or essential gene targets (676 core essential gene-encoded proteins) (Wang et al., 2019). An arbitrary weight was set for positive score (1) and negative score (-0.333) for multiplexing to generate a PN-score. For FDA-approved drugs, a joint PN-score was reported by adding the DeepCPI and DTINet PN-score together, and used for ranking the drugs. For natural compounds, we also employed these two ranking methods using either DeepCPI P-score or DeepCPI PN-score.

The detailed formula was as follows:

For a given drug-target pair, we calculated the DTI score t_{CPI} and t_{Net} by DeepCPI and DTINet, respectively. By collecting all the DTI scores, two score matrices T_{CPI} and T_{Net} were defined to quantify the confidence of predicted DTIs:

$$\begin{cases} T_{CPI} \in \mathbb{R}^{l \times k} \\ T_{Net} \in \mathbb{R}^{l \times k} \end{cases} \quad (1)$$

Where, l refers to the length of drug list and k refers to the length of target list.

To ensure them comparable, the score matrices T_{CPI} and T_{Net} were normalized by Z-Score measurement:

$$\begin{cases} Z_{CPI} = \frac{x_{CPI} - \mu_{CPI}}{\sigma_{CPI}}, x_{CPI} \in T_{CPI} \\ Z_{Net} = \frac{x_{Net} - \mu_{Net}}{\sigma_{Net}}, x_{Net} \in T_{Net} \end{cases} \quad (2)$$

Where, μ is mean value of the scores and σ is standard deviation of the scores.

We further applied an optimal threshold (as discussed above, 0.641 and 0.973 were used for Z_{CPI} and Z_{Net}, respectively) to filter the non-significant scores and only keep the confident DTI scores:

$$Z_{CPI\_sig} = \begin{cases} z, if\ z \geq 0.641 \\ 0, if\ z < 0.641 \end{cases} \quad z \in Z_{CPI} \quad (3)$$

$$Z_{Net\_sig} = \begin{cases} z, if\ z \geq 0.973 \\ 0, if\ z < 0.973 \end{cases} \quad z \in Z_{Net} \quad (4)$$

For each FDA-approved drug, the mean value of the normalized z-scores was defined as its positive score:

$$\begin{cases} P\_score_{CPI} = \sum_{i=1}^{k} z_i^{CPI} / k \\ P\_score_{Net} = \sum_{i=1}^{k} z_i^{Net} / k \end{cases} \quad (5)$$

Similar as above, we defined negative scores $N\_score_{druggable}$ and $N\_score_{essentialome}$ for non-HDG and essential gene targets, respectively. The final negative was the sum of $N\_score_{druggable}$ and $N\_score_{essentialome}$:

$$\begin{cases} N\_score_{CPI} = N\_score_{Druggable_{CPI}} + N\_score_{Essentialome\_CPI} \\ N\_score_{Net} = N\_score_{Druggable_{Net}} + N\_score_{Essentialome\_Net} \end{cases} \quad (6)$$

The PN-score was the sum of weighted positive score and negative score:

$$\begin{cases} PN\_score_{CPI} = 1 * P\_score_{CPI} + (-0.333) * N\_score_{CPI} \\ PN\_score_{Net} = 1 * P\_score_{Net} + (-0.333) * N\_score_{Net} \end{cases} \quad (7)$$

Here, we defined a joint P-score by the sum of P_score_{CPI} and P_score_{Net} for each drug:

$$Joint\_P\_score = P\_score_{CPI} + P\_score_{Net} \quad (8)$$

The joint PN-score was the sum of PN_score_{CPI} and PN_score_{Net} for each drug:

$$Joint\_PN\_score = PN\_score_{CPI} + PN\_score_{Net} \quad (9)$$

Molecular Docking

The structures of target protein were downloaded from PDB database (http://www.rcsb.org). The drug or compound structures were downloaded from TCMSP and PubChem database (https://pubchem.ncbi.nlm.nih.gov). The structures of proteins and compounds were imported into prime tool of Maestro

(version 11.8.012) suite of Schrödinger software (released 2018-4). Next the preprocessing step was performed by adding hydrogens and missing atoms as well as removing water molecules for the proteins using the Protein Preparation tool. Ligand preprocessing was performed using default settings with Ligprep tool of Maestro software. Then, the top-ranked potential binding site was defined using Receptor Grid Generation tool. Glide tool was used to detect the interactions between ligands and proteins. The docking score ≤ -6 was considered as a high confidence binding event between tested ligand and protein. The Glide energy for each docking pair was also shown in Table S15. The 2D structures of the top candidate drugs were presented in Table S16.

Supplemental References

Cotto, K.C., Wagner, A.H., Feng, Y.Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic acids research *46*, D1068-D1073.

Gilson, M.K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic acids research *44*, D1045-1053.

Kuhn, M., Szklarczyk, D., Franceschini, A., Campillos, M., von Mering, C., Jensen, L.J., Beyer, A., and Bork, P. (2010). STITCH 2: an interaction network database for small molecules and proteins. Nucleic acids research *38*, D552-556.

Li, W., Koster, J., Xu, H., Chen, C.H., Xiao, T., Liu, J.S., Brown, M., and Liu, X.S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome biology *16*, 281.

Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., Li, P., Guo, Z., Tao, W., Yang, Y.*, et al.* (2014). TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. Journal of cheminformatics *6*, 13.

Ursu, O., Holmes, J., Bologa, C.G., Yang, J.J., Mathias, S.L., Stathias, V., Nguyen, D.T., Schurer, S., and Oprea, T. (2019). DrugCentral 2018: an update. Nucleic acids research *47*, D963-D970.

Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z.*, et al.* (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nature protocols *14*, 756-780.

Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., Zhang, R., Zhu, J., Ren, Y., Tan, Y.*, et al.* (2020). Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. Nucleic acids research *48*, D1031-D1041.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology *16*, 284-287.