

Supplementary Information for “Genome-wide discovery of hidden genes mediating known drug-disease association using KDDANet”

Hua Yu ^{1, 4#*}, Lu Lu ^{2, #}, Ming Chen ³, Chen Li ^{2, *}, Jin Zhang ^{1, *}

¹ Center for Stem Cell and Regenerative Medicine, Department of Basic Medical Sciences, and The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China. Institute of Hematology, Zhejiang University, Hangzhou, Zhejiang, China.

² Department of Human Genetics, and Women’s Hospital, Zhejiang University School of Medicine, Hangzhou, China

³ College of Life Sciences, Zhejiang University, Hangzhou, China.

⁴ Lead contact

These authors contributed equally.

* Correspondence: huayu@zju.edu.cn, chenli2012@zju.edu.cn, zhgene@zju.edu.cn

Supplementary Figure Legends

Supplementary Figure 1: Performance evaluation of KDDANet method.

Supplementary Figure 2: General applicability of KDDANet method.

Supplementary Figure 3: Mechanistic relevance of KDDANet prediction results.

Supplementary Figure 4: KDDANet provided novel molecular insights on KDDAs related to cancer.

Supplementary Figure 5: KDDANet uncovered the shared genes mediating multiple KDDAs.

Supplementary Note Legends

Supplementary Note 1: A mini review of state-of-the-art computational tools for facilitating drug repurposing.

Supplementary Note 2: Enrichment analysis of hidden genes mediating KDDA in the KDDANet resulting subnetworks.

Supplementary Note 3: Method for enrichment analysis.

Supplementary Note 4: Method for selecting a suitable γ value.

Supplementary Note 5: KEGG pathway enrichment of KDDANet resulting subnetworks.

Supplementary Note 6: Analysis of KDDANet resulting subnetworks mediating the associations between drugs and cancer using cancer-omics’ datasets.

Supplementary Note 7: An online web server for KDDA decoding.

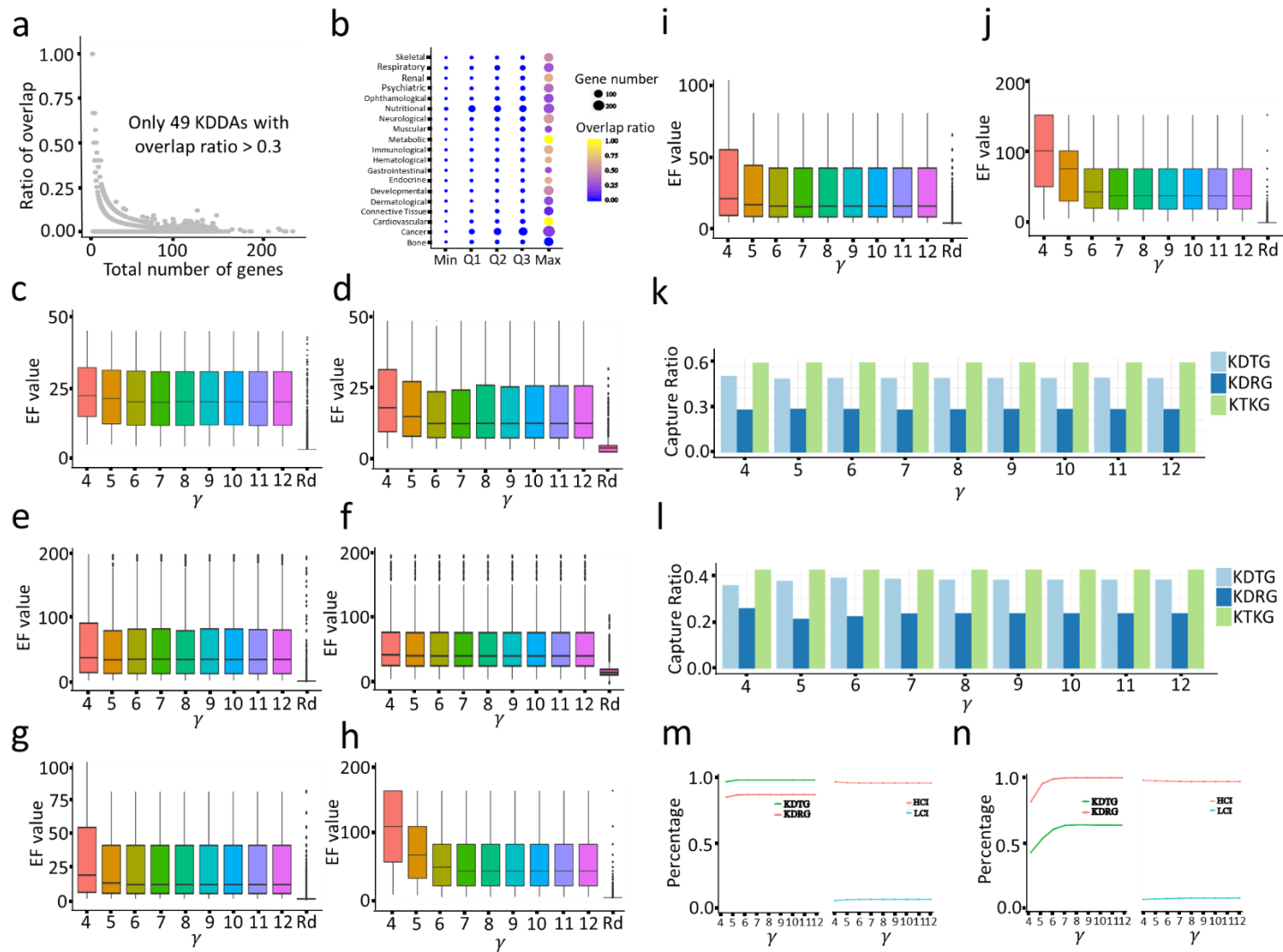
Supplementary Data Legends

Supplementary Data 1: Five different types of gene interaction networks, including HumanNet, HINT+HI2012, iRefIndex, MultiNet and STRINGv10 used in this study.

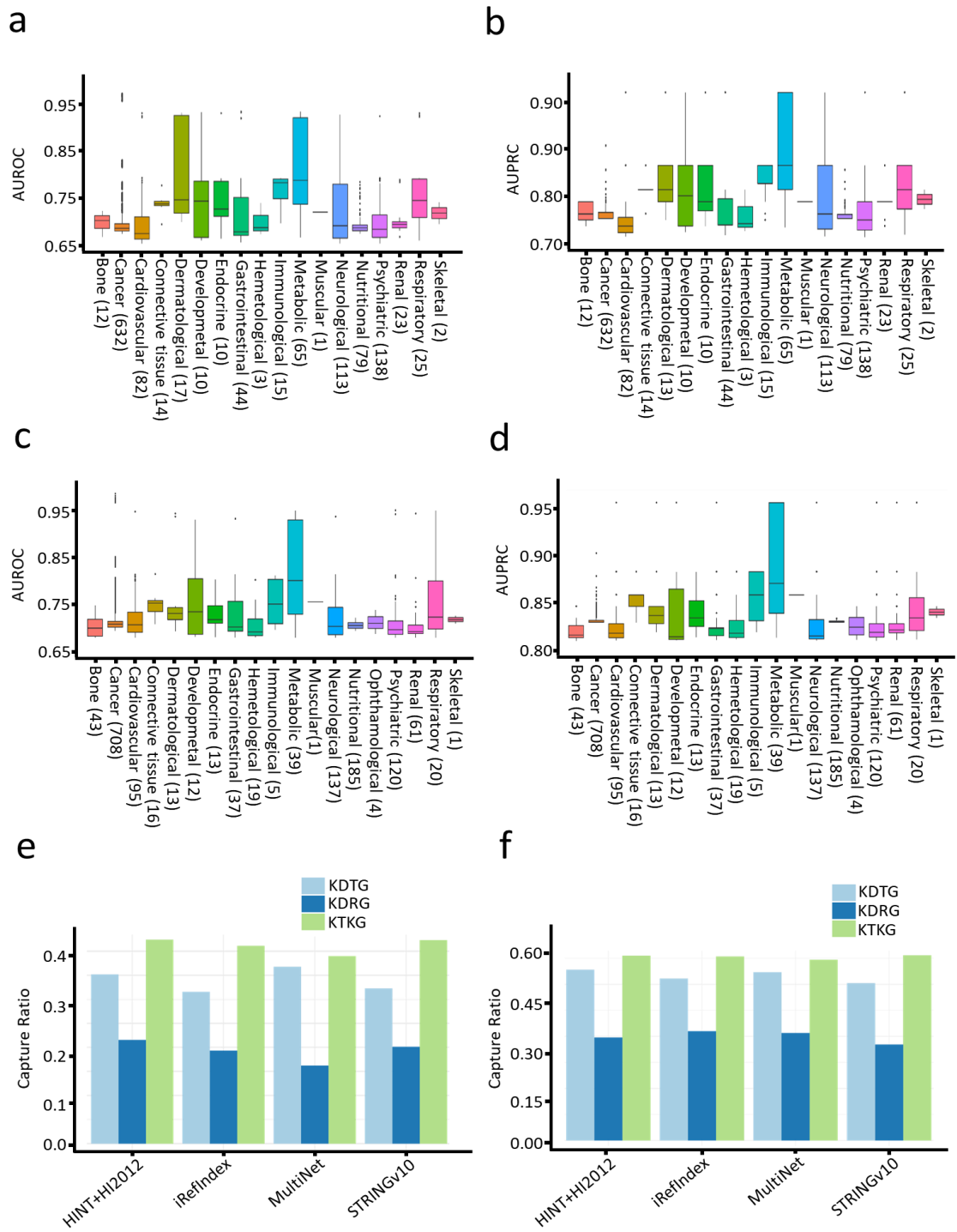
Supplementary Data 2: Drugs and their target genes used in this study.

Supplementary Data 3: Known diseases and their related genes used in this study.

Supplementary Data 4: 53124 KDDAs used in this study.

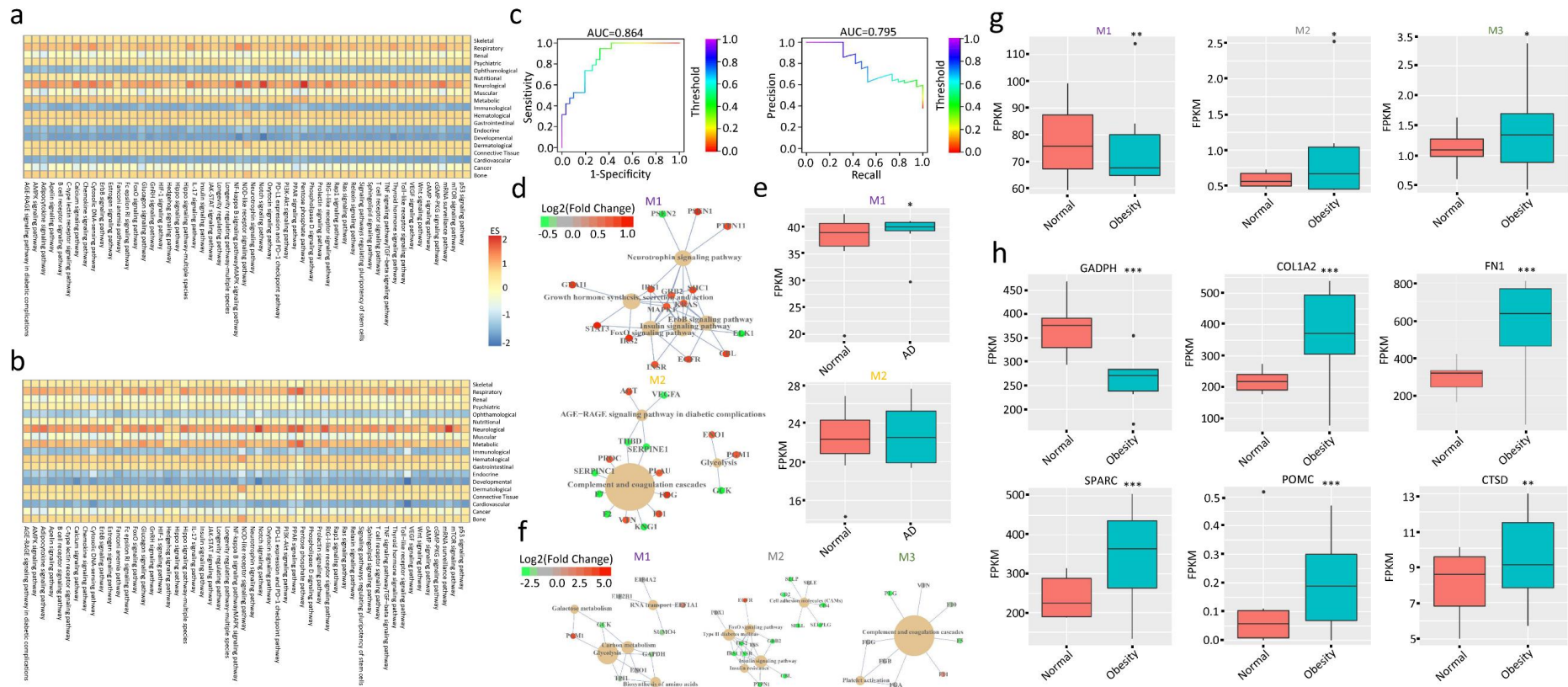


Supplementary Figure 1: Performance evaluation of KDDANet method. **a)** Scatter plot demonstrating the distribution of overlap ratio between KDTGs and KDRGs. The x-axis denoted the total number of KDTGs and KDRGs mediating a KDDA. **b)** Scatter plot demonstrating the Minimum (Min), 25th Quantile (Q1), 50th Quantile (Q2), 75th Quantile (Q3) and Maximum (Max) of overlap ratio and gene number distribution of 53124 KDDAs in different types of diseases; the size and color of point denoted gene number and overlap ratio, respectively. **c)** Boxplot demonstrating the enrichments of KDTGs in KDDANet resulting subnetwork with different γ settings in SDrTDi context, using random permutation as control (Rd). **d)** Similar to as **c)**, demonstrating the enrichments of KDRGs. **e)** Similar to **c)**, demonstrating the enrichments of KDTGs in SDiTDr context. **f)** Similar to **d)**, demonstrating the enrichments of KDRGs in SDiTDr context. **g)** Boxplot demonstrating the enrichment of KTKGs in SDrTDi context, using random permutation as control (Rd). **h)** Similar to **g)**, demonstrating the enrichment of KTKGs in SDiTDr context. **i)** Similar to **g)**, demonstrating the enrichment of NTKGs. **j)** Similar to **h)**, demonstrating the enrichment of NTKGs. **k)** The KDDANet capture ratio of KDTGs, KDRG and KTKGs with different γ setting in SDrTDi context. **l)** The KDDANet capture ratio of KDTGs, KDRGs and KTKGs with different γ setting in SDiTDr context. **m)** Line chart demonstrating the percentage of KDTGs and KDRGs that were incorporated into the KDDANet resulting subnetwork, as well as the percentage of low probability edges with weight smaller than 0.3 in SDrTDi context (HCI, high confidence interaction; LCI, low confidence interaction). **n)** Similar to **m)**, demonstrating the percentages in SDiTDr context. In panels **c-j)**, the centre of the box plots represents the median value and the lower and upper lines represent the 25% and 75% quantile, respectively. The whiskers stand for $1.5 \times$ interquartile range below the lower quartile and above the upper quartile. The points located outside of whiskers denote outliers.



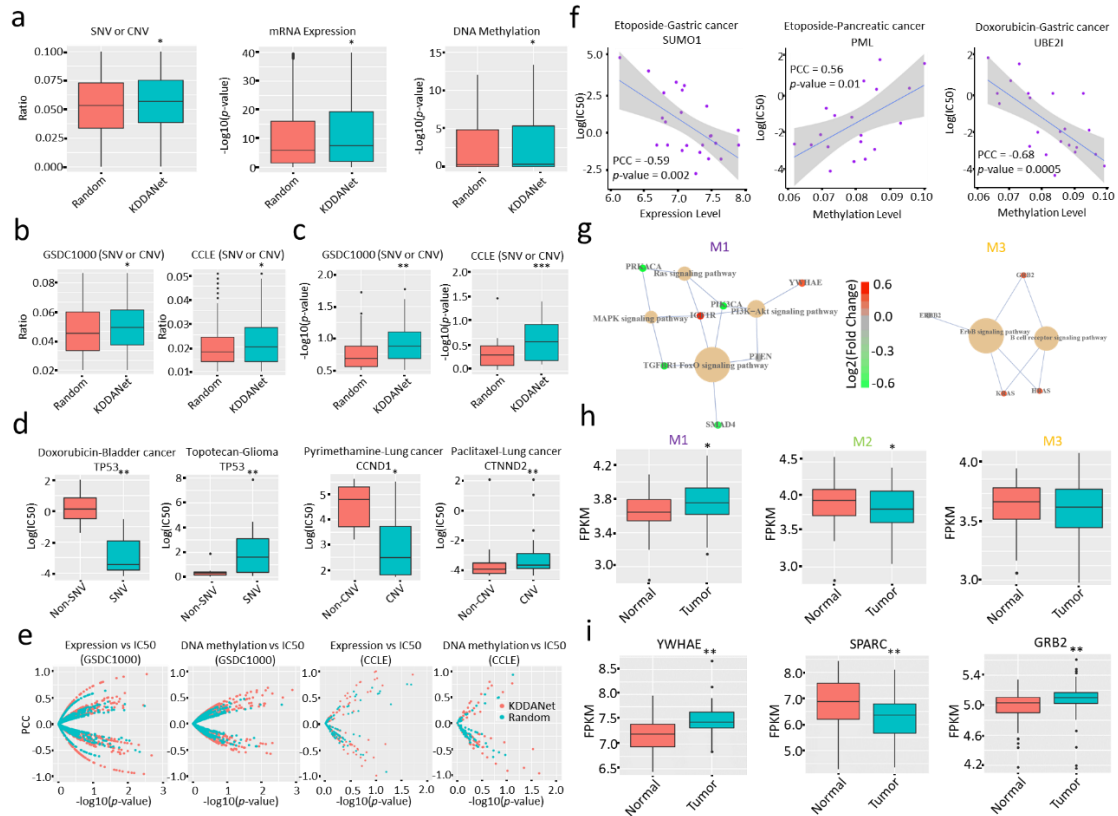
Supplementary Figure 2: General applicability of KDDANet method. **a)** and **b)** AUROC and AUPRC of KDDANet across different types of diseases in SDrTDi context. **c)** and **d)** Similar to **a)** and **b)** in SDiTDi context; Values in parentheses denotes the number of KDDAs. **e)** The KDDANet capture ratio of KDTGs, KDRGs and KTKGs using different types of networks in SDiTDi context. **f)** Similar to **e)** in SDrTDi context. In panels **a-d)**, the

centre of the box plots represents the median value and the lower and upper lines represent the 25% and 75% quantile, respectively. The whiskers stand for $1.5 \times$ interquartile range below the lower quartile and above the upper quartile. The points located outside of whiskers denote outliers.



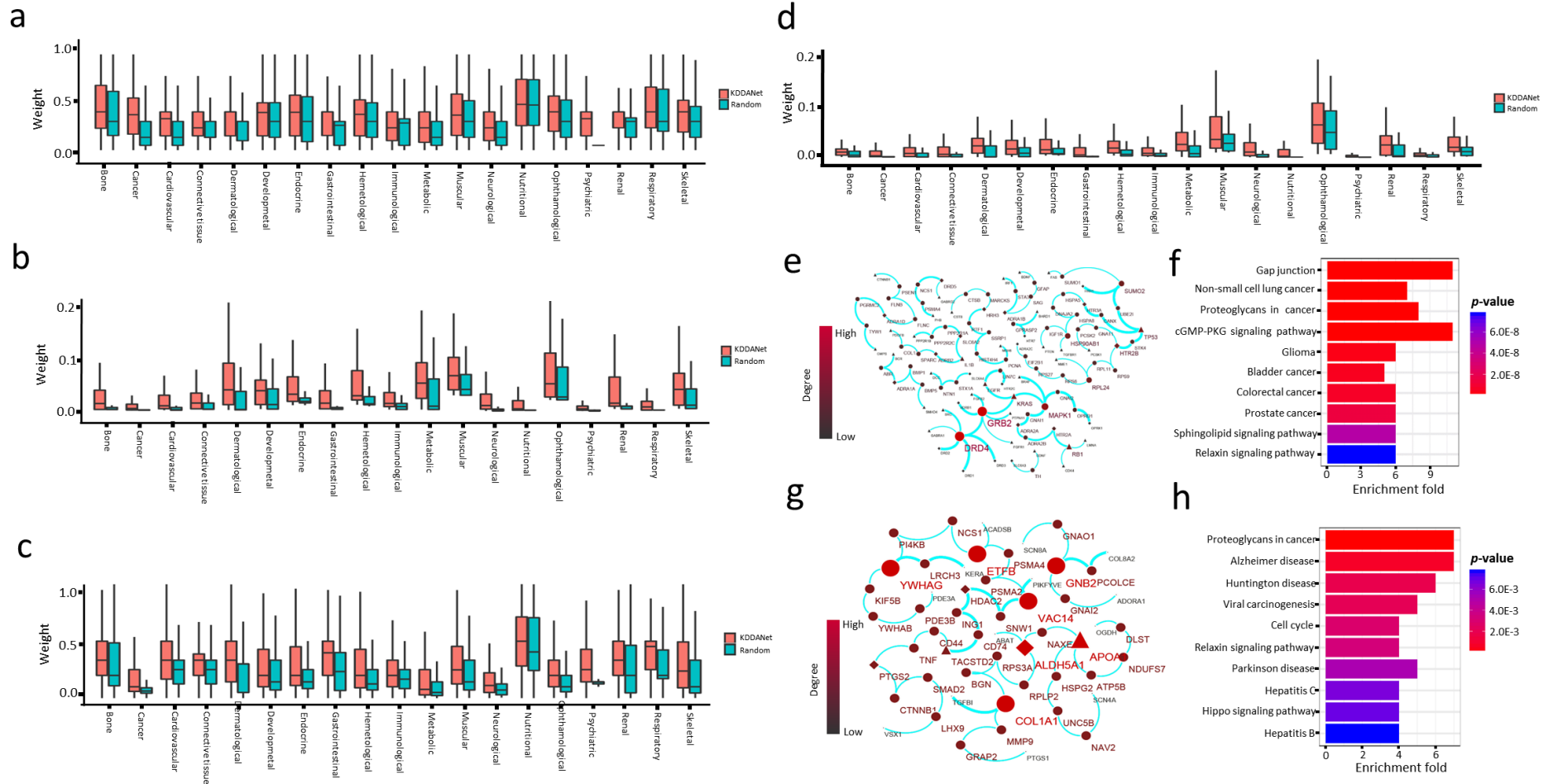
Supplementary Figure 3: Mechanistic relevance of KDDANet prediction results. a) Enrichment Score (ES) heatmap of 53 KEGG pathways on 19 different types of diseases in SDRTDi context. **b)** Similar to **a)**, in SDITDr context. A two-step procedure was used to calculate ES. First, for each disease type, we calculated the frequency of KDDA resulting subnetworks enriched on 53 KEGG pathways and obtained a frequency matrix. Then,

5 for each pathway, we normalized the frequency matrix to a z-score matrix. The value in z-score matrix was defined as ES. **c)** ROC and PR curves
6 of KDDANet resulting subnetwork mediating phyloquinone-Alzheimer's disease (AD) association. **d)** Pathway-gene relationship network
7 demonstrating the enriched KEGG terms and their related genes for module M1 and M2 mediating phyloquinone-AD association. **e)** Averaged
8 expression level of module M1 and M2 genes in normal individuals and AD patients, * p -value < 0.05, calculated by Mann-Whitney U test. **f)**
9 Similar to **d)** pathway-gene relationship network demonstrating the enriched KEGG terms and their related genes for module M1, M2 and M3
10 mediating heparin-obesity association. **g)** Averaged expression level of module M1, M2 and M3 genes in normal individuals and obesity patients,
11 ** p -value < 0.01, * p -value < 0.05, calculated by Mann-Whitney U test. **h)** Expression level of GADPH, COL1A2, FN1, SPARC, POMC and CTSD in
12 normal individuals and obesity patients; *** p -value < 0.001, ** p -value < 0.01, calculated by Mann-Whitney U test. In the pathway-gene
13 relationship networks, the size of a KEGG term node was proportional to its p -value in enrichment analysis; The color of a gene node denoted its
14 fold change of expression level between normal individuals and patients. Fragments Per Kilobase Of Exon Per Million Fragments Mapped, FPKM.
15 In panels **e**, **g** and **h**, the centre of the box plots represents the median value and the lower and upper lines represent the 25% and 75% quantile,
16 respectively. The whiskers stand for $1.5 \times$ interquartile range below the lower quartile and above the upper quartile. The points located outside
17 of whiskers denote outliers.



Supplementary Figure 4: KDDANet provided novel molecular insights on KDDAs related to cancer. **a)** The ratio of TCGA tumor samples harboring SNVs or CNVs of KDDANet resulting subnetwork genes mediating the associations between drugs and cancer and randomly selected genes, and the overall statistic difference of mRNA expression and DNA methylation of KDDANet resulting subnetwork genes mediating the associations between drugs and cancer (randomly selected genes) between tumor samples and normal samples, respectively; * p -value < 0.05, calculated by Mann-Whitney U test. **b)** The ratio of cancer cell lines harboring SNVs or CNVs of KDDANet resulting subnetwork genes mediating the associations between drugs and cancer and randomly selected genes, * p -value < 0.05, calculated by Mann-Whitney U test. **c)** Boxplots demonstrating the overall statistic difference between IC50 values of anti-cancer drugs on cell lines harboring SNVs or CNVs of KDDANet resulting subnetwork genes mediating the associations between drugs and cancer and cell line without SNVs and CNVs of these genes, *** p -value < 0.001, ** p -value < 0.01, calculated by Mann-Whitney U test. **d)** Representative examples demonstrating the difference between IC50 values of anti-cancer drugs on cell lines harboring SNVs or CNVs of KDDANet

35 resulting subnetwork genes mediating the associations between drugs and cancer and
 36 cell lines without SNVs and CNVs of these genes, ** p -value < 0.01, * p -value < 0.05,
 37 calculated by Mann-Whitney U test. **e)** Scatter plot demonstrating the correlation
 38 between expression abundances/methylation levels of KDDANet resulting subnetwork
 39 genes mediating the associations between drugs and cancer (randomly selected genes)
 40 and half maximal inhibitory concentration (IC50) values of anti-cancer drugs in both
 41 GSDC1000 and CCLE cancer cell lines, respectively; PPC, Pearson Correlation
 42 Coefficients; p -value was calculated by cor.test function in R software. **f)**
 43 Representative examples demonstrating the correlation between the expression
 44 abundances/methylation levels of KDDANet resulting subnetwork genes mediating
 45 the associations between drugs and cancer and IC50 values of anti-cancer drugs in
 46 both GSDC1000 and CCLE cancer cell lines, respectively. **g)** Pathway-gene relationship
 47 network demonstrating the enriched KEGG terms and their related genes for M1 and
 48 M3 mediating sotalol (DB00489)-prostate cancer (176807) association. **h)** Averaged
 49 expression level of module M1, M2 and M3 genes in prostate cancer tumor samples
 50 and adjacent normal tissue samples, * p -value < 0.05, calculated by Wilcox signed rank
 51 test. **i)** Expression level of YWHAE, SPARC and GRB2 in prostate cancer tumor samples
 52 and adjacent normal tissue samples, ** p -value < 0.01, calculated by Wilcox signed
 53 rank test. In panels **a-d**, **h** and **i**, the centre of the box plots represents the median
 54 value and the lower and upper lines represent the 25% and 75% quantile, respectively.
 55 The whiskers stand for $1.5 \times$ interquartile range below the lower quartile and above
 56 the upper quartile. The points located outside of whiskers denotes outliers.



Supplementary Figure 5: KDDANet uncovered the shared genes mediating multiple KDDAs. a) Boxplots demonstrating the distributions of weight values in KDDANet meta-subnetwork and random meta-subnetwork for MDiODr in different types of diseases in SDrTDi context. **b)**

60 Boxplots demonstrating the distributions of weight values in KDDANet meta-subnetwork and random meta-subnetwork for MDrODi in different
61 types of diseases in SDrTDi context. **c)** Similar to **a)** in SDiTDi context; **d)** Similar to **b)** in SDiTDi context. **e)** Shared meta-subnetwork mediating
62 the associations between mirtazapine (DB00370) and 9 types of cancers. **f)** Top 10 enriched KEGG terms of shared genes of meta-subnetwork
63 mediating the associations between mirtazapine (DB00370) and 9 types of cancers. **g)** Shared meta-subnetwork mediating the associations
64 between Keratoconus (148300) and three drugs. **h)** Top 10 enriched KEGG terms of shared genes of meta-subnetwork mediating the associations
65 between Keratoconus (148300) and three drugs. In panels **a-d**, the centre of the box plots represents the median value and the lower and upper
66 lines represent the 25% and 75% quantile, respectively. The whiskers stand for $1.5 \times$ interquartile range below the lower quartile and above the
67 upper quartile. The points located outside of whiskers denotes outliers.

Supplementary Note 1: A mini review of state-of-the-art computational tools for facilitating drug repurposing.

With the rapid accumulation of biomedical data, various computational methods, including machine learning, similarity computation and network-based models, have been proposed for facilitating drug repurposing. We here provided a mini summary of state-of-the-art computational approaches designed for predicting drug-drug interactions, drug side-effects, drug combinations, drug indications, drug-target interactions, disease-related genes, disease modules and constructing disease-disease relationship network. For forecasting pharmacokinetic and pharmacodynamic drug-drug interactions and their associated recommendations, Gottlieb et al. designed “INDI”, a similarity measure-based logistic regression classifier¹. In addition, Huang et al. proposed a metric “S-score” that measures the strength of connection between drug targets in the protein-protein interaction network and integrates drug clinical side effects to predict pharmacodynamic drug-drug interactions². For identifying drug side-effects, Atias et al. and Shaked et al. designed the network-based approaches^{3,4}. Wang et al. developed a multilabel classifier by combining chemical structure and gene expression features to predict drug side-effects⁵. Tatonetti et al. presented an adaptive data-driven approach for predicting drug side-effects and drug interaction side-effects⁶. Duffy et al. used the tissue-specific genetic features for drug side-effects prediction in clinical trials⁷. Kuhn et al. integrated phenotypic data obtained during clinical trials with known drug-target relations to identify overrepresented protein-side effect combinations⁸. For screening novel drug combinations, Zhao et al. explored various molecular and pharmacological features of drugs and showed that combinations of such features were an effective predictor⁹. In addition, the network-based methods have been proposed as a promising tool to identify novel drug combination^{10,11}. Interestingly, Kuenzi et al. developed a deep-learning model of visible neural network, called DrugCell, to predict drug response and synergy in human cancer cells¹². To predict novel drug-disease associations and identify drug indications, Pacini et al. developed a R/Cytoscape package, which utilized the correlation of gene expression profiles to infer whether a query drug can be repurposed for a disease¹³. Gottlieb et

al. developed PREDICT, which scored a possible drug-disease association by combining multiple drug-drug and disease-disease similarity measures¹⁴. Guney et al. proposed an unsupervised and unbiased network-based proximity measure between drug targets and disease proteins for scoring drug-disease associations¹⁵. Cheng et al. further showed that the integration of network proximity-based approach and large-scale patient-level longitudinal data complemented by mechanistic in vitro studies offered an effective platform by which to identify and validate novel drug-disease associations and drug indications¹⁶. Besides, Gilvary and Elkhader developed a machine learning and network framework to discover new indications for small molecules¹⁷. For identifying novel drug targets, Paolini et al presented the global mapping of pharmacological space that can identify confidently the human drug targets¹⁸. Yamanishi et al. proposed a new statistical method that formalized the drug-target interaction inference as a supervised learning problem for a bipartite graph and integrated chemical and genomic spaces into a unified pharmacological space¹⁹. Campillos et al. used the side-effect similarity to identify novel drug targets²⁰. Perlman et al. developed "SITAR", which incorporated multiple drug-drug and gene-gene similarity measures for scoring drug and gene associations²¹. Yu et al employed both support vector machine and random forest for identifying multiple drug and target interactions²². Luo et al. developed a computational pipeline, called DTINet, to predict novel drug–target interactions from a constructed heterogeneous network²³. The network propagation algorithm has been also proposed for drug target identification²⁴. Iskar et al. identified and characterized drug-induced transcriptional modules using microarray data²⁵. Silberberg et al. further proposed a novel method to identify drug response pathways by combining drug targets, drug response expression profiles, and the human physical interaction network²⁶. For discovering the disease-related genes and disease modules, multiple computational approaches have been designed²⁷⁻³⁶. For example, Wu et al. designed CIPHER for identifying human disease genes that integrated human protein-protein interactions, disease phenotype similarities, and known gene-phenotype associations to capture the complex relationships between

phenotypes and genotypes²⁷. Gottlieb et al. developed “PRINCIPLE”, which employed the classical network propagation algorithm for associating genes with diseases³⁰. Ghiassian et al. designed “DIAMOnD” for detecting disease modules based on systematic analysis of connectivity patterns of disease proteins in the human interactome³¹. Particularly, Vinayagam et al. identified both disease genes and drug targets based on controllability analysis of the directed human protein interaction network³⁶. Menche et al. used the biomedical literature database and incomplete interactome to construct disease-symptom and disease-disease relationship networks^{37,38}. Moreover, Hofree et al. introduced network-based stratification (NBS) to uncover tumor subtypes by integrating somatic tumor genomes with gene networks³⁹. Collectively, these computational methods helped our understanding of the pathology of diseases and mechanisms of drug actions, and thus promoted the process of drug repurposing.

Supplementary Note 2: Enrichment analysis of hidden genes mediating KDDA in the KDDANet resulting subnetworks.

We first tested whether the KDTGs and KDRGs are contained in KDDANet resulting subnetworks by performing enrichment analysis with a range of γ setting from 4 to 12 with a step 1. For comparing with background, we carried out a permutation test by producing random subnetworks with the same number of genes. We found that the KDTGs and KDRGs have significant enrichment in the resulting subnetworks against random subnetworks (Supplementary Figure 1c-1f). This indicated that KDDANet can effectively capture KDTGs and KDRGs. To examine whether KDDANet can capture true genes mediating KDDA, we introduced two concepts: “known true KDDA genes (KTKGs)” and “novel true KDDA genes (NTKGs)” (see Main Text for definition). Based on these definitions, we then performed enrichment analysis to check whether the KTKGs have captured in our prediction results. As shown in Supplementary Figure 1g and Supplementary Figure 1h, the subnetworks outputted by KDDANet have obviously higher enrichment of KTKGs when compared with the permutation subnetworks. We

further performed enrichment analysis to test whether KDDANet can uncover NTKGs. We found that the resulting subnetworks have still higher enrichment of NTKGs than random permutation subnetworks (Supplementary Figure 1i and Supplementary Figure 1j). We removed the link of each KTKG with drug and disease from the flow network to test the recoverability of KDDANet. Interestingly, we observed that the capture ratio of KTKG in KDDANet resulting subnetwork is higher than KDTG and KDRG (Supplementary Figure 1k and Supplementary Figure 1l). This suggested that KDDANet favor to capture more NTKGs than KDTGs and KDRGs.

Supplementary Note 3: Method for enrichment analysis.

We carried out fold enrichment analysis against full gene network for each predicted KDDA subnetwork to evaluate whether KDDANet can capture true genes mediating KDDA. The fold of enrichment was calculated by the following formula:

$$\frac{p}{q} \times \frac{M}{N} \quad (1)$$

where, p was the number of “KDTGs”/“KDRGs” or the number of true positive genes (“KTKGs”/“NTKGs”) mediating a given KDDA; q was the number of inferred genes in the resulting KDDA subnetwork; M denoted the number of all “KDTGs”/“KDRGs” or the number of all true positive genes (“KTKGs”/“NTKGs”) in the gold standards and N represented the number of all genes in full network.

Supplementary Note 4: Method for selecting a suitable γ value.

As the size and quality of KDDANet output subnetwork depended on a parameter γ . Higher γ values will identify more links between the source node and the sink node but with lower confidence. To obtain suitable values for γ , we run KDDANet with γ values ranging between 4 and 12 with a step of 1. For each of the output subnetworks, we computed the percentage of input, namely KDTGs and KDRGs, that were incorporated into the network, as well as the percentage of low probability edges with weight smaller than 0.3. As shown in Supplementary Figure 1m and Supplementary Figure 1n, we observed that the percentages for KDTGs, KDRGs and

low probability edges contained in resulting subnetworks became stable when $\gamma = 6$ and 8 for SDrTDi and SDiTDr context, respectively. We therefore selected $\gamma = 6$ for SDrTDi and $\gamma = 8$ for SDiTDr as they were the minimal values with which a significant fraction of the input was incorporated while the percentage of low probability edges remained small in respective contexts.

Supplementary Note 5: KEGG pathway enrichment of KDDANet resulting subnetworks.

We asked whether the enriched functions of KDDANet subnetworks were mechanistically related to KDDA by carrying out a global enrichment of all predicted KDDA subnetworks against 53 classical KEGG pathways. We observed that p53 signaling pathway had relatively higher enrichment in KDDA subnetwork linking drug with cancer when comparing with other pathways in both SDrTDi and SDiTDr contexts (Supplementary Figure 3a and Supplementary Figure 3b). For neurological disease, we found that the notch-signaling pathway had significant enrichment in SDrTDi and SDiTDr context (Supplementary Figure 3a and Supplementary Figure 3b). This was consistent with the previous discovery that showed the association between notch-related pathways and neurological disorders⁴⁰. Interestingly, we found that ophthalmological disease, immunological disease, endocrine disease, developmental disease, and cardiovascular diseases have consistent less enrichment on all KEGG pathways than other types of diseases (Supplementary Figure 3a and Supplementary Figure 3b). This indicated that the drugs developed to cure these types of diseases cannot be well explained by the known pathways and the uncovered molecular interactions mediating KDDAs for these diseases was of great value for further experimental studies.

Supplementary Note 6: Analysis of KDDANet resulting subnetworks mediating the associations between drugs and cancer using cancer-omics' datasets.

We tested whether KDDANet provides novel molecular insights on KDDAs mediating

the associations between drugs and cancers using cancer-omics dataset. For this, we required that the type of cancer of KDDA was contained in TCGA, GSDC1000 or CCLE. KDDANet predicted 1847 novel unknown genes mediating KDDAs mediating the associations between drugs and cancer. We found that 149 of these genes were contained in COSMIC Cancer Gene Census with ~96-fold enrichment and adjusted p -value of $5.709\text{e-}10$ (Hypergeometric test and Bonferroni correction). Using different types of TCGA datasets, including mRNA expression, Single Nucleotide Variant (SNV), Copy Number Variant (CNV) and DNA methylation, we checked whether these genes undergo oncogenic alterations in tumor samples. As expected, we found that these KDDA candidate genes have significantly higher SNV and CNV ratio in tumor samples than randomly selected genes (Supplementary Figure 4a). In consistent with this, the expression abundances and methylation levels of these candidates are significantly dysregulated in tumor samples than randomly selected genes, respectively (Supplementary Figure 4a). We further checked whether these genes undergo oncogenic alterations in cancer cell lines. In line with TCGA primary tumor data, we observed that they harbor more SNVs and CNVs in cancer cell lines than randomly selected genes (Supplementary Figure 4b).

We further asked whether the SNVs and CNVs of KDDA genes mediating the associations between drugs and cancers were correlated with the responses of cancer cell lines under anti-cancer drug treatment. For each gene of KDDA subnetwork mediating the association between drug and cancer, we evaluated the difference of drug's half maximal inhibitory concentration (IC50) between cell lines harboring SNVs or CNVs on this gene (HSCs) and cell lines without SNVs and CNVs (WSCs) on this gene using Mann-Whitney U test. We required that the KDDA was contained in GSDC1000 or CCLE. We observed that the difference of IC50 values of anti-cancer drugs between HSCs and WSCs derived from KDDA genes mediating the associations between drugs and cancers was more significant than IC50 values of anti-cancer drugs between HSCs and WSCs derived from randomly selected genes (Supplementary Figure 4c). Such as,

KDDANet predicted, TP53, a well-known tumor suppression gene, mediating doxorubicin-bladder cancer association and topotecan-glioma association. The SNVs of bladder cancer cell line on TP53 was correlated with decreased IC50 values of doxorubicin (Supplementary Figure 4d). And the SNVs of glioma cancer cell line on TP53 was correlated with increased IC50 values of topotecan (Supplementary Figure 4d). A key lung cancer driver gene CCND1 was predicted mediating pyrimethamine-lung cancer association ⁴¹. The CNVs of lung cancer cell lines on CCND1 was correlated with decreased IC50 values of pyrimethamine (Supplementary Figure 4d). CTNND2 encoded δ -Catenin which promoted tumorigenesis and metastasis of lung adenocarcinoma ⁴². CTNND2 was inferred mediating paclitaxel and lung cancer association. The CNVs of lung cancer cell lines on CTNND2 was correlated with decreased IC50 values of paclitaxel (Supplementary Figure 4d). Based on these, we further tested whether the expression abundances and methylation levels of KDDA genes mediating the associations between drugs and cancerd were also associated with the responses of cancer cell lines under anti-cancer drug treatment. When compared with randomly selected genes, the expression abundances and methylation levels of KDDA genes mediating the associations between drugs and cancers were more obviously correlated with IC50 values of anti-cancer drugs in both GSDC1000 and CCLE cancer cell lines, respectively (Supplementary Figure 4e). For example, SUMO1 has been linked to tumorigenesis and invasion in gastric cancer ⁴³. KDDANet inferred SUMO1 mediating etoposide-gastric cancer association. The expression level of SUMO1 was negatively correlated to IC50 value of etoposide in gastric cancer cell lines (Supplementary Figure 4f). It has been reported that PML as potential treatment for pancreatic cancer ⁴⁴. KDDANet discovered PML mediating etoposide-pancreatic cancer association. The methylation level of PML gene was positively correlated to IC50 value of etoposide in pancreatic cancer cell line (Supplementary Figure 4f). E2 ubiquitin-conjugating enzymes have been reported involving in various tumor-promoting processes ⁴⁵. KDDANets predicted UBE2I mediating doxorubicin-gastric cancer. The methylation level of UBE2I gene was negatively correlated with IC50 value of

doxorubicin (Supplementary Figure 4f). We have not found that these correlations are reported in the GSDC1000 and CCLE studies. Collectively, these outcomes indicated that KDDANet provided novel molecular insights into the molecular mechanism of anti-cancer drug response of cancer cell lines.

Supplementary Note 7: An online web server for KDDA decoding.

To help biomedical researchers to understand the pathogenesis of disease and promote drug development, we develop an online web server, <http://www.kddanet.cn>, for facilitating the researchers to explore the subnetwork of genes mediating KDDAs. Our website is divided into three parts: I) querying and browsing subnetwork of genes mediating individual KDDA (iKDDA); II) querying and browsing meta-subnetwork of shared genes mediating multiple KDDAs connecting multiple diseases to one drug (MDiODr) and III) querying and browsing meta-subnetwork of shared genes mediating multiple KDDAs connecting multiple drugs to one disease (MDrODi). For Part I, subnetworks of genes mediating 52878 KDDAs are provided for querying, visualizing, and downloading, respectively. For Part II, 12386 shared meta-subnetworks are contained in our web server. For Part III, 773 shared meta-subnetworks are contained in our web server. We design a bulk download interface and tutorials for facilitating user to explorer KDDANet resulting subnetworks. Overall, we provide a simply and easy-to-use web server for user to decode the hidden genes mediating KDDAs and thus help researchers to understand the pathogenesis of disease, accelerate drug repurposing and disease treatment. We next provided a more in-depth description with some figures generated from our web server and one example as step-by-step schematic illustration to explain the utility of KDDANet. For Part I, the iKDDA page aims to search gene subnetwork mediating the association between a given drug and a given disease. We provided an example of searching gene subnetwork mediating the association between heparin and obesity. For Part II, the MDiODr page aims to search the shared gene meta-subnetwork mediating multiple KDDAs between a given drug and a given disease type. We provided an example of searching gene subnetwork mediating multiple associations between bivalirudin and cancer. For Part III, the

MDiODr page aims to search the shared gene meta-subnetwork mediating multiple KDDAs between a given disease and all its related drugs. We provided an example of searching gene subnetwork mediating multiple associations between 13 drugs and GRACILE syndrome (Finnish lethal neonatal metabolic syndrome). Please see the detailed tutorials as follow and the Help section of our website for detailed description of the utility of an online web version of KDDANet tool.

Tutorials of KDDANet

iKDDA page

This page aims to search gene subnetwork mediating the association between a given drug and a given disease.

iKDDA Search Box

Home iKDDA MDiODr MDrODi Bulk Download

Search

DrugBank ID or drug name OMIM disease ID or disease name Context

DB00407 Heparin 601665 Obesity SDrTDi

SEARCH

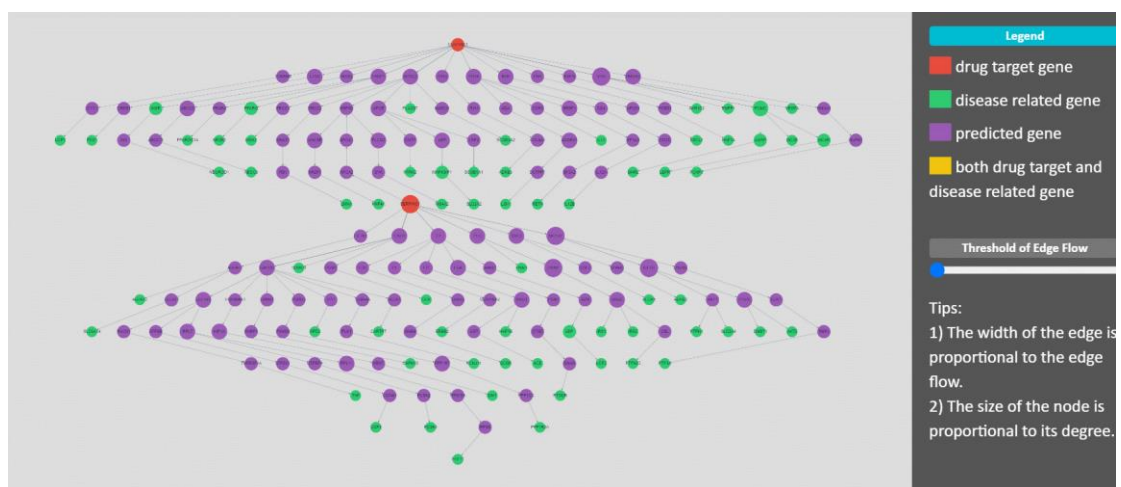
① Select drug from the drop-down list by DrugBank ID or drug name. A prompt will appear when you type characters.

② Select disease from the drop-down list by OMIM disease ID or disease name.

③ Select the context. SDrTDi: Network flow model constructed by using query drug as source node. SDiTDr: Network flow model constructed by using query disease as source node.

iKDDA Search Result

iKDDA Subnetwork



Tips:

1. Move the threshold of Threshold of Edge Flow slider in the right panel, it will dynamically display the updated gene subnetwork by selecting the edge with flow amount larger than the threshold. Bar left: minimal edge flow. Bar right: maximum edge flow.

2. Box selection via modifier key (shift, command, control, alt)

3. Mouse wheel to zoom

4. Grab and drag background to pan

5. Click a gene node to see the NCBI gene information

SCGB1A1 secretoglobin family 1A member 1 [*Homo sapiens* (human)]

Gene ID: 7356, updated on 22-Aug-2020

Summary

Official Symbol

SCGB1A1 provided by HGNC

Official Full Name

secretoglobin family 1A member 1 provided by HGNC

Primary source

[HGNC:HGNC:12523](#)

See related

[Ensembl:ENSG00000149021](#) [MIM:192020](#)

Gene type

protein coding

RefSeq status

REVIEWED

Organism

[Homo sapiens](#)

Lineage

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo

Also known as

UGB; UP1; CC10; CC16; CCSP; UP-1; CCPBP

Summary

This gene encodes a member of the secretoglobin family of small secreted proteins. The encoded protein has been implicated in numerous functions including anti-inflammation, inhibition of phospholipase A2 and the sequestering of hydrophobic ligands. Defects in this gene are associated with a susceptibility to asthma. [provided by RefSeq, May 2010]

Expression

Restricted expression toward lung (RPKM 1367.3) [See more](#)

Orthologs

[mouse](#) [all](#)

Genomic context

Location:

11q12.3

Exon count:

3

See SCGB1A1 in [Genome Data Viewer](#)

Annotation release	Status	Assembly	Chr	Location
109.20200815	current	GRCh38.p13 (GCF_000001405.39)	11	NC_000011.10 (62419033..62423195)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	11	NC_000011.9 (62186507..62190678)

Subnetwork Download

DOWNLOAD

[DB00407_601665_download.txt](#)

File Content

3339	HSPG2	344	APOC2	0.00746269
3339	HSPG2	335	APOA1	0.0223881
6514	SLC2A2	9479	MAPK8IP1	0.0149254
2206	MS4A2	6850	SYK	0.0149254
3596	IL13	1232	CCR3	0.0149254
5506	PPP1R3A	5501	PPP1CC	0.0149254
10554	NA	1967	EIF2B1	0.0298507
3827	KNG1	2160	F11	0.0298507
7356	SCGB1A1	4036	LRP2	0.0149254
56729	RETN	8175	SF3A2	0.0149254

Column 1: Gene 1 EntrezID

Column 2: Gene 1 Name

Column 3: Gene 2 EntrezID

Column 4: Gene 2 Name

Column 5: Edge Flow

MDiODr page

This page aims to search the shared gene meta-subnetwork mediating multiple KDDAs between a given drug and a given disease type.

MDiODr Search Box

Home iKDDA MDiODr MDrODi Bulk Download

Search

DrugBank ID or drug name Disease type Context

DB00006 bivalirudin Cancer SDrTDi

SEARCH

① Select drug from the drop-down list by DrugBank ID or drug name. A prompt will appear when you type characters.

② Select disease type from the drop-down.

③ Select the Context. SDrTDi: Network flow model constructed by using query drug as source node. SDiTDi: Network flow model constructed by using query disease as source node.

MDiODr Search Result

Related Diseases

Related Diseases:

[Lung Neoplasms](#),[Prostatic Neoplasms](#),[Stomach Neoplasms](#),[Urinary Bladder Neoplasms](#),

Click a disease to see the OMIM disease information

211980

LUNG CANCER

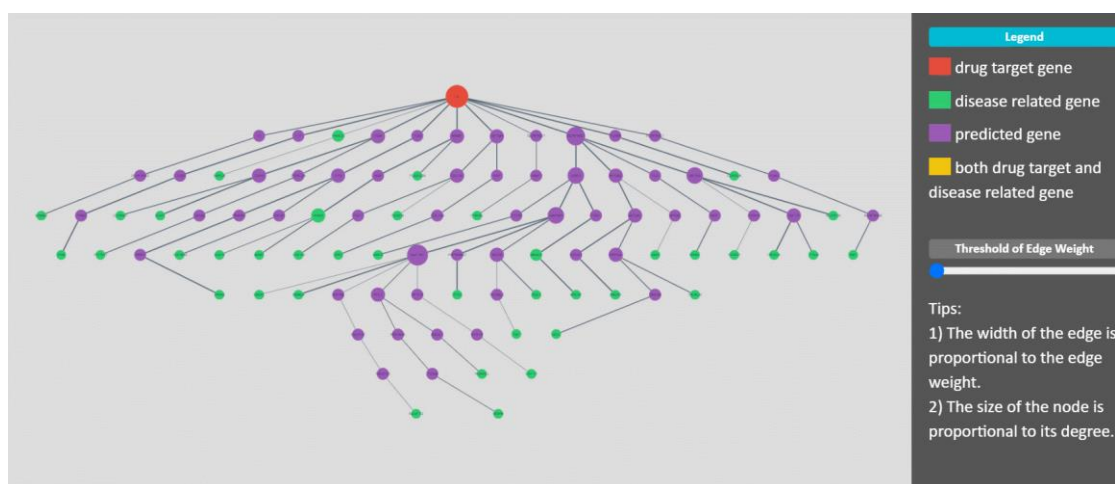
Other entities represented in this entry:

ALVEOLAR CELL CARCINOMA, INCLUDED
 ADENOCARCINOMA OF LUNG, INCLUDED
 NONSMALL CELL LUNG CANCER, INCLUDED
 LUNG CANCER, PROTECTION AGAINST, INCLUDED

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
1q24.3	{Lung cancer, susceptibility to}	211980	AD, SMu	3	FASLG	134638
2q33.1	{Lung cancer, protection against}	211980	AD, SMu	3	CASP8	601763

MDiODr Meta-Subnetwork



Tips:

1. Move the threshold of Threshold of Edge Weight slider in the right panel, it will dynamically display the updated gene subnetwork by selecting the edge with weight larger than the threshold. Bar left: minimal edge weight. Bar right: maximum edge weight.
2. Box selection via modifier key (shift, command, control, alt)

3693. Mouse wheel to zoom
3704. Grab and drag background to pan
3715. Click a gene node to see the NCBI gene information

Full Report

Showing Current Items.

Send to

F2 coagulation factor II, thrombin [*Homo sapiens* (human)]
Gene ID: 2147, updated on 19-Sep-2020

Summary

Official Symbol

Official Full Name

Primary source

See related

Gene type

RefSeq status

Organism

Lineage

Also known as

Summary

F2 provided by HGNC

coagulation factor II, thrombin provided by HGNC

HGNC:HGNC:3535

Ensembl:ENSG00000180210 MIM:176930

protein coding

REVIEWED

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo

PT, THPH1; RPRGL2

This gene encodes the prothrombin protein (also known as coagulation factor II). This protein is proteolytically cleaved in multiple steps to form the activated serine protease thrombin. The activated thrombin enzyme plays an important role in thrombosis and hemostasis by converting fibrinogen to fibrin during blood clot formation, by stimulating platelet aggregation, and by activating additional coagulation factors. Thrombin also plays a role in cell proliferation, tissue repair, and angiogenesis as well as maintaining vascular integrity during development and postnatal life. Peptides derived from the C-terminus of this protein have antimicrobial activity against *E. coli* and *P. aeruginosa*. Mutations in this gene lead to various forms of thrombosis and dysprothrombinemia. Rapid increases in cytokine levels following coronavirus infections can dysregulate the coagulation cascade and produce thrombosis, compromised blood supply, and organ failure. [provided by RefSeq, May 2020]

Annotation information

Expression

Orthologs

Note: This gene has been reviewed for its involvement in coronavirus biology, and is relevant for disease process.

Restricted expression toward liver (RPKM 385.8) [See more](#)

[mouse](#) [all](#)

Genomic context

Location: 11p11.2

Exon count: 14

See F2 in [Genome Data Viewer](#)

Annotation release	Status	Assembly	Chr	Location
109.20200815	current	GRCh38.p13 (GCF_000001405.39)	11	NC_000011.10 (46719213..46739506)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	11	NC_000011.9 (46740743..46761056)

Download

DOWNLOAD

DB00006_Cancer_download.txt

File Content

6041	RNASEL	6059	ABCE1	0.5
6129	RPL7	1915	EEF1A1	0.5
60	ACTB	2597	GAPDH	0.75
335	APOA1	336	APOA2	1
5781	PTPN11	2885	GRB2	0.75
1621	DBH	338	APOB	0.25
5111	PCNA	121504	HIST4H4	0.5
2885	GRB2	7916	PRRC2A	0.75

Column 1: Gene 1 EntrezID

Column 2: Gene 1 name

Column 3: Gene 2 EntrezID

Column 4: Gene 2 Name

Column 5: Edge Weight

MDrODi page

This page aims to search the shared gene meta-subnetwork mediating multiple KDDAs between a given disease and all its related drugs.

MDrODi Search Box

Search

OMIM disease ID or disease name

603358 Finnish lethal neonatal metabolic syndrome ①

Context

SDrTDi ②

SEARCH

① Select disease from the drop-down list by OMIM disease ID or disease name. A prompt will appear when you type characters.

② Select the Context. SDrTDi: Network flow model constructed by using query drug as source node. SDiTDi: Network flow model constructed by using query disease as source node.

MDrODi Search Result

Related Drugs

Related Drugs:

rosiglitazone, Tamoxifen, Ketamine, Tretinoin, Finasteride, nefazodone, Methotrexate, Estradiol, nimesulide, Flutamide, resveratrol, Ethinyl Estradiol, Acetaminophen,

Click a drug to see the DrugBank drug information

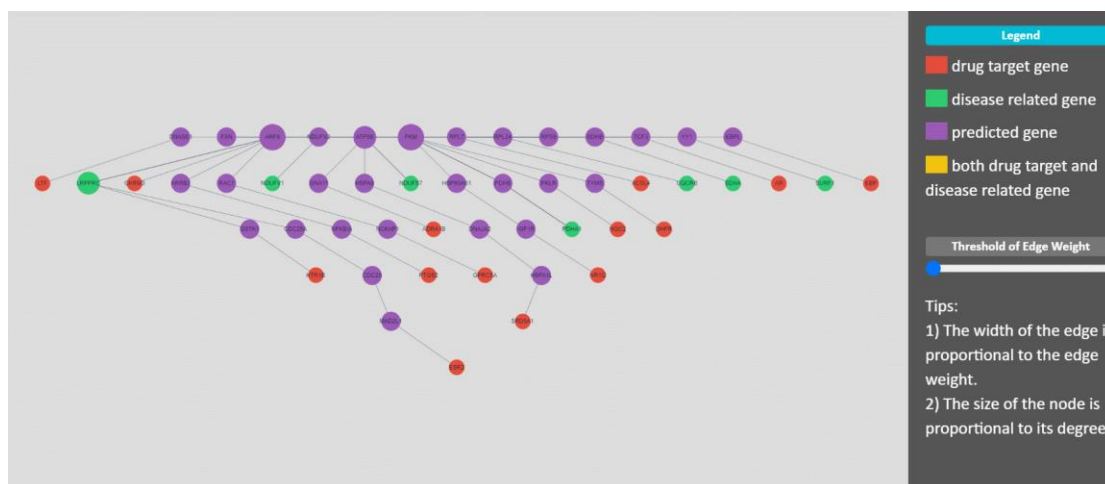
Rosiglitazone

Targets (7)Enzymes (7)Carriers (1)Transporters (2)

IDENTIFICATION

Name	Rosiglitazone	Accession Number	DB00412
Description	Rosiglitazone is an anti-diabetic drug in the thiazolidinedione class of drugs. It is marketed by the pharmaceutical company GlaxoSmithKline as a stand-alone drug (Avandia) and in combination with metformin (Avandamet) or with glimepiride (Avandaryl). Like other thiazolidinediones, the mechanism of action of rosiglitazone is by activation of the intracellular receptor class of the peroxisome proliferator-activated receptors (PPARs), specifically PPAR γ . Rosiglitazone is a selective ligand of PPAR γ , and has no PPAR α -binding action. Apart from its effect on insulin resistance, it appears to have an anti-inflammatory effect: nuclear factor kappa-B (NF κ B) levels fall and inhibitor (IKB) levels increase in patients on rosiglitazone. Recent research has suggested that rosiglitazone may also be of benefit to a subset of patients with Alzheimer's disease not expressing the ApoE4 allele. This is the subject of a clinical trial currently underway.		

MDrODi Meta-Subnetwork



Tips:

1. Move the threshold of Threshold of Edge Weight slider in the right panel, it will dynamically display the updated gene subnetwork by selecting the edge with weight larger than the threshold. Bar left: minimal edge weight. Bar right: maximum edge weight.

2. Box selection via modifier key (shift, command, control, alt)

3. Mouse wheel to zoom

4. Grab and drag background to pan

5. Click a gene node to see the NCBI gene information

Full Report ▾ Send to ▾

Showing Current items.

ATP5F1B ATP synthase F1 subunit beta [*Homo sapiens* (human)]

Gene ID: 506, updated on 6-Sep-2020

Summary

Official Symbol ATP5F1B provided by [HGNC](#)
Official Full Name ATP synthase F1 subunit beta provided by [HGNC](#)
Primary source [HGNC:HGNC:830](#)
See related [Ensembl:ENSG00000110955 MIM:102910](#)
Gene type protein coding
RefSeq status REVIEWED
Organism [Homo.sapiens](#)
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as ATP5B; ATPMB; ATPSB; HEL-S-271
Summary This gene encodes a subunit of mitochondrial ATP synthase. Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation. ATP synthase is composed of two linked multi-subunit complexes: the soluble catalytic core, F1, and the membrane-spanning component, F0, comprising the proton channel. The catalytic portion of mitochondrial ATP synthase consists of 5 different subunits (alpha, beta, gamma, delta, and epsilon) assembled with a stoichiometry of 3 alpha, 3 beta, and a single representative of the other 3. The proton channel consists of three main subunits (a, b, c). This gene encodes the beta subunit of the catalytic core. [provided by RefSeq, Jul 2008]
Expression Ubiquitous expression in heart (RPKM 713.8), kidney (RPKM 376.0) and 25 other tissues [See more](#)
Orthologs [mouse](#) [all](#)

Genomic context

Location: 12q13.3 See ATP5F1B in [Genome Data Viewer](#)

Exon count: 10

Annotation release	Status	Assembly	Chr	Location
109.20200815	current	GRCh38.p13 (GCF_000001405.39)	12	NC_000012.12 (56638175..56645984, complement)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	12	NC_000012.11 (57031959..57039852, complement)

Download

File Content

374291	NDUFS7	506	ATP5B	0.153846153846154
2767	GNA11	147	ADRA1B	0.0769230769230769
3305	HSPA1L	6715	SRD5A1	0.0769230769230769
382	ARF6	409	ARRB2	0.0769230769230769
6152	RPL24	6202	RPS8	0.0769230769230769
5315	PKM	3326	HSP90AB1	0.0769230769230769
5162	PDHB	5315	PKM	0.230769230769231
7528	NA	1773	DNASE1	0.0769230769230769

Column 1: Gene 1 EntrezID

Column 2: Gene 1 name

Column 3: Gene 2 EntrezID

Column 4: Gene 2 Name

Column 5: Edge Weight

Bulk Download page

This page aims to download all KDDA subnetworks for downstream analysis.

Bulk Download

1. iKDDA.zip
2. MDiODr.zip
3. MDrODi.zip

Supplementary References

1. Gottlieb, A., Stein, G. Y., Oron, Y., Ruppín, E. & Sharan, R. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Molecular Systems Biology* **8**, 592, doi:10.1038/msb.2012.26 (2012).
2. Huang, J. *et al.* Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS Computational Biology* **9**, e1002998, doi:10.1371/journal.pcbi.1002998 (2013).
3. Atias, N. & Sharan, R. An algorithmic framework for predicting side effects of drugs. *Journal of Computational Biology* **18**, 207-218, doi:10.1089/cmb.2010.0255 (2011).
4. Shaked, I., Oberhardt, M. A., Atias, N., Sharan, R. & Ruppín, E. Metabolic Network Prediction of Drug Side Effects. *Cell Systems* **2**, 209-213, doi:10.1016/j.cels.2016.03.001 (2016).
5. Wang, Z., Clark, N. R. & Ma'ayan, A. Drug-induced adverse events prediction with the LINCS L1000 data. *Bioinformatics* **32**, 2338-2345, doi:10.1093/bioinformatics/btw168 (2016).
6. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Science Translational Medicine* **4**, 125ra31. doi: 10.1126/scitranslmed.3003377 (2012).
7. Duffy Á, Verbanck M, Dobbyn A, Won HH, Rein JL, Forrest IS, Nadkarni G, Rocheleau G, Do R. Tissue-specific genetic features inform prediction of drug side effects in clinical trials. *Science Advance* **6**, eabb6242. doi: 10.1126/sciadv.abb6242 (2020).
8. Kuhn, M. *et al.* Systematic identification of proteins that elicit drug side effects. *Molecular Systems Biology* **9**, 663, doi:10.1038/msb.2013.10 (2013).

- 444 9. Zhao, X. M. *et al.* Prediction of drug combinations by integrating molecular and
445 pharmacological data. *PLoS Computational Biology* **7**, e1002323, doi:
446 10.1371/journal.pcbi.1002323 (2011).
- 447 10. Cheng, F., Kovács, I. A. & Barabási, A. L. Network-based prediction of drug
448 combinations. *Nature Communications* **10**, 1197, doi:10.1038/s41467-019-09186-x
449 (2019).
- 450 11. Huang, L. *et al.* DrugComboRanker: drug combination discovery based on target
451 network analysis. *Bioinformatics* **30**, i228-236, doi:10.1093/bioinformatics/btu278
452 (2014).
- 453 12. Kuenzi, B. M. *et al.* Predicting Drug Response and Synergy Using a Deep Learning
454 Model of Human Cancer Cells. *Cancer Cell* **38**, 672-684.e676,
455 doi:10.1016/j.ccell.2020.09.014 (2020).
- 456 13. Pacini, C. *et al.* DvD: An R/Cytoscape pipeline for drug repurposing using public
457 repositories of gene expression data. *Bioinformatics* **29**, 132-134,
458 doi:10.1093/bioinformatics/bts656 (2013).
- 459 14. Gottlieb, A., Stein, G. Y., Ruppín, E. & Sharan, R. PREDICT: a method for inferring
460 novel drug indications with application to personalized medicine. *Molecular Systems*
461 *Biology* **7**, 496, doi:10.1038/msb.2011.26 (2011).
- 462 15. Guney, E., Menche, J., Vidal, M. & Barabási, A. L. Network-based in silico drug
463 efficacy screening. *Nature Communications* **7**, 10331, doi:10.1038/ncomms10331
464 (2016).
- 465 16. Cheng, F. & Desai, R. J. Network-based approach to prediction and population-
466 based validation of in silico drug repurposing. *Nature Communications* **9**, 2691,
467 doi:10.1038/s41467-018-05116-5 (2018).
- 468 17. Gilvary, C. & Elkhader, J. A machine learning and network framework to discover
469 new indications for small molecules. *PloS Computational Biology* **16**, e1008098,
470 doi:10.1371/journal.pcbi.1008098 (2020).
- 471 18. Paolini, G. V., Shapland, R. H., van Hoorn, W. P., Mason, J. S. & Hopkins, A. L. Global
472 mapping of pharmacological space. *Nature Biotechnology* **24**, 805-815,
473 doi:10.1038/nbt1228 (2006).
- 474 19. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W. & Kanehisa, M. Prediction of
475 drug-target interaction networks from the integration of chemical and genomic spaces.
476 *Bioinformatics* **24**, i232-240, doi:10.1093/bioinformatics/btn162 (2008).
- 477 20. Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J. & Bork, P. Drug target
478 identification using side-effect similarity. *Science* **321**, 263-266,
479 doi:10.1126/science.1158140 (2008).
- 480 21. Perlman, L., Gottlieb, A., Atias, N., Ruppín, E. & Sharan, R. Combining drug and gene
481 similarity measures for drug-target elucidation. *Journal of Computational Biology* **18**,

133-145, doi:10.1089/cmb.2010.0213 (2011).

22. Yu H, Chen J, Xu X, Li Y, Zhao H, Fang Y, et al. A Systematic Prediction of Multiple Drug-Target Interactions from Chemical, Genomic, and Pharmacological Data. *PLoS ONE* **7**(5): e37608, <https://doi.org/10.1371/journal.pone.0037608> (2012).

23. Luo, Y., Zhao, X., Zhou, J. et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communication* **8**, 573 (2017). <https://doi.org/10.1038/s41467-017-00680-8>.

24. Shnaps, O., Perry, E., Silverbush, D. & Sharan, R. Inference of personalized drug targets via network propagation. *Pacific Symposium on Biocomputing* **21**, 156-167 (2016).

25. Iskar, M. et al. Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. *Molecular Systems Biology* **9**, 662, doi:10.1038/msb.2013.20 (2013).

26. Silberberg, Y., Gottlieb, A., Kupiec, M., Ruppín, E. & Sharan, R. Large-scale elucidation of drug response pathways in humans. *Journal of Computational Biology* **19**, 163-174, doi:10.1089/cmb.2011.0264 (2012).

27. Wu, X., Jiang, R., Zhang, M. Q. & Li, S. Network-based global inference of human disease genes. *Molecular Systems Biology* **4**, 189, doi:10.1038/msb.2008.27 (2008).

28. Karni, S., Soreq, H. & Sharan, R. A network-based method for predicting disease-causing genes. *Journal of Computational Biology* **16**, 181-189, doi:10.1089/cmb.2008.05TT (2009).

29. Vanunu, O., Magger, O., Ruppín, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. *PLoS Computational Biology* **6**, e1000641, doi:10.1371/journal.pcbi.1000641 (2010).

30. Gottlieb, A., Magger, O., Berman, I., Ruppín, E. & Sharan, R. PRINCIPLE: a tool for associating genes with diseases via network propagation. *Bioinformatics* **27**, 3325-3326, doi:10.1093/bioinformatics/btr584 (2011).

31. Ghiassian, S. D., Menche, J. & Barabási, A. L. A DISeAse MOdule Detection (DIAMOND) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Computational Biology* **11**, e1004120, doi:10.1371/journal.pcbi.1004120 (2015).

32. Mazza, A., Klockmeier, K., Wanker, E. & Sharan, R. An integer programming framework for inferring disease complexes from network data. *Bioinformatics* **32**, i271-i277, doi:10.1093/bioinformatics/btw263 (2016).

33. Silberberg, Y., Kupiec, M. & Sharan, R. GLADIATOR: a global approach for elucidating disease modules. *Genome Medicine* **9**, 48, doi:10.1186/s13073-017-0435-z (2017).

34. Carlin, D. E. *et al.* A Fast and Flexible Framework for Network-Assisted Genomic Association. *iScience* **16**, 155-161, doi:10.1016/j.isci.2019.05.025 (2019).
35. Lee, I., Blom, U. M., Wang, P. I., Shim, J. E. & Marcotte, E. M. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Research* **21**, 1109-1121, doi:10.1101/gr.118992.110 (2011).
36. Vinayagam, A. *et al.* Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 4976-4981, doi:10.1073/pnas.1603992113 (2016).
37. Zhou, X., Menche, J., Barabási, A. L. & Sharma, A. Human symptoms-disease network. *Nature Communications* **5**, 4212, doi:10.1038/ncomms5212 (2014).
38. Menche, J. *et al.* Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601, doi:10.1126/science.1257601 (2015).
39. Hofree, M., Shen, J. P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nature Methods* **10**, 1108-1115, doi:10.1038/nmeth.2651 (2013).
40. Lathia, J. D., Mattson, M. P. & Cheng, A. Notch: from neural development to neurological disorders. *Journal of Neurochemistry* **107**, 1471-1481, doi:10.1111/j.1471-4159.2008.05715.x (2008).
41. Gautschi, O., Ratschiller, D., Gugger, M., Betticher, D. C. & Heighway, J. Cyclin D1 in non-small cell lung cancer: a key driver of malignant transformation. *Lung Cancer* **55**, 1-14, doi:10.1016/j.lungcan.2006.09.024 (2007).
42. Huang, F. *et al.* δ -Catenin promotes tumorigenesis and metastasis of lung adenocarcinoma. *Oncology Reports* **39**, 809-817, doi:10.3892/or.2017.6140 (2018).
43. Jin, L., Shen, K., Chen, T., Yu, W. & Zhang, H. SUMO-1 Gene Silencing Inhibits Proliferation and Promotes Apoptosis of Human Gastric Cancer SGC-7901 Cells. *Cellular Physiology and Biochemistry* **41**, 987-998, doi:10.1159/000460836 (2017).
44. Swayden, M. *et al.* PML hyposumoylation is responsible for the resistance of pancreatic cancer. *FASEB Journal* **33**, 12447-12463, doi:10.1096/fj.201901091R (2019).
45. Hosseini, S. M. *et al.* E2 ubiquitin-conjugating enzymes in cancer: Implications for immunotherapeutic interventions. *Clinica Chimica Acta* **498**, 126-134, doi:10.1016/j.cca.2019.08.020 (2019).