

RESEARCH ARTICLE

# Deciphering Signaling Pathway Networks to Understand the Molecular Mechanisms of Metformin Action

Jingchun Sun<sup>1,2</sup>, Min Zhao<sup>2</sup>, Peilin Jia<sup>2</sup>, Lily Wang<sup>3</sup>, Yonghui Wu<sup>1</sup>, Carissa Iverson<sup>4,5</sup>, Yubo Zhou<sup>6</sup>, Erica Bowton<sup>7</sup>, Dan M. Roden<sup>8,9</sup>, Joshua C. Denny<sup>2</sup>, Melinda C. Aldrich<sup>4,5,10</sup>, Hua Xu<sup>1\*</sup>, Zhongming Zhao<sup>2,11,12\*</sup>

**1** School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America, **2** Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **3** Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **4** Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **5** Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, **6** National Center for Drug Screening, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, People's Republic of China, **7** Institute for Clinical and Translational Research, School of Medicine, Vanderbilt University, Nashville, Tennessee, United States of America, **8** Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **9** Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **10** Division of Epidemiology, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **11** Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **12** Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America

\* [hua.xu@uth.tmc.edu](mailto:hua.xu@uth.tmc.edu) (HX); [zhongming.zhao@vanderbilt.edu](mailto:zhongming.zhao@vanderbilt.edu) (ZZ)



**OPEN ACCESS**

**Citation:** Sun J, Zhao M, Jia P, Wang L, Wu Y, Iverson C, et al. (2015) Deciphering Signaling Pathway Networks to Understand the Molecular Mechanisms of Metformin Action. *PLoS Comput Biol* 11(6): e1004202. doi:10.1371/journal.pcbi.1004202

**Editor:** Xianghong Jasmine Zhou, University of Southern California, UNITED STATES

**Received:** October 1, 2014

**Accepted:** February 13, 2015

**Published:** June 17, 2015

**Copyright:** © 2015 Sun et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This project is partially supported by the National Institutes of Health (grant numbers R01LM011177, R01LM 010685, K07CA172294, P50CA90949, P50CA095103, P50CA098131, P30CA068485, RC2 GM092618, ULTR000445), the Cancer Prevention & Research Institute of Texas Rising Star Award (CPRIT R1307), 2013 NARSAD Young Investigator Award, and American Cancer Society Institutional Research Grant pilot project (#IRG-58-009-55) and Ingram Professorship Funds. The funders had no role in study design, data

## Abstract

A drug exerts its effects typically through a signal transduction cascade, which is non-linear and involves intertwined networks of multiple signaling pathways. Construction of such a signaling pathway network (SPNetwork) can enable identification of novel drug targets and deep understanding of drug action. However, it is challenging to synopsise critical components of these interwoven pathways into one network. To tackle this issue, we developed a novel computational framework, the Drug-specific Signaling Pathway Network (DSPATH-Net). The DSPATHNet amalgamates the prior drug knowledge and drug-induced gene expression via random walk algorithms. Using the drug metformin, we illustrated this framework and obtained one metformin-specific SPNetwork containing 477 nodes and 1,366 edges. To evaluate this network, we performed the gene set enrichment analysis using the disease genes of type 2 diabetes (T2D) and cancer, one T2D genome-wide association study (GWAS) dataset, three cancer GWAS datasets, and one GWAS dataset of cancer patients with T2D on metformin. The results showed that the metformin network was significantly enriched with disease genes for both T2D and cancer, and that the network also included genes that may be associated with metformin-associated cancer survival. Furthermore, from the metformin SPNetwork and common genes to T2D and cancer, we generated a subnetwork to highlight the molecule crosstalk between T2D and cancer. The follow-up network analyses and literature mining revealed that seven genes (*CDKN1A*, *ESR1*, *MAX*,

collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

*MYC*, *PPARGC1A*, *SP1*, and *STK11*) and one novel MYC-centered pathway with *CDKN1A*, *SP1*, and *STK11* might play important roles in metformin's antidiabetic and anticancer effects. Some results are supported by previous studies. In summary, our study 1) develops a novel framework to construct drug-specific signal transduction networks; 2) provides insights into the molecular mode of metformin; 3) serves a model for exploring signaling pathways to facilitate understanding of drug action, disease pathogenesis, and identification of drug targets.

## Author Summary

A deep understanding of a drug's mechanisms of actions is essential not only in the discovery of new treatments but also in minimizing adverse effects. Here, we develop a computational framework, the Drug-specific Signaling Pathway Network (DSPathNet), to reconstruct a comprehensive signaling pathway network (SPNetwork) impacted by a particular drug. To illustrate this computational approach, we used metformin, an anti-diabetic drug, as an example. Starting from collecting the metformin-related upstream genes and inferring the metformin-related downstream genes, we built one metformin-specific SPNetwork via random walk based algorithms. Our evaluation of the metformin-specific SPNetwork by using disease genes and genotyping data from genome-wide association studies showed that our DSPathNet approach was efficient to synopsise drug's key components and their relationship involved in the type 2 diabetes and cancer, even the metformin anticancer activity. This work presents a novel computational framework for constructing individual drug-specific signal transduction networks. Furthermore, its successful application to the drug metformin provides some valuable insights into the mode of metformin action, which will facilitate our understanding of the molecular mechanisms underlying drug treatments, disease pathogenesis, and identification of novel drug targets and repurposed drugs.

## Introduction

Most drugs exert their therapeutic actions through interactions with specific protein targets. These target proteins are dominated by two categories: enzymes that catalyze reactions essential for the functioning of organisms, and receptors that transmit signals by interacting with messenger molecules [1,2]. The interactions of drugs and their targets initiate the signal transduction cascade that is usually propagated by the involved proteins and multiple pathways. These proteins and pathways act in the mode of crosstalk networks [3]. The process of such signaling transduction converts the chemical signals to a specific cellular response such as gene expression, cell division, and inhibition of cell death and apoptosis [4]. The signaling cascade usually ends at the recipients of chemical signals such as transcription factors (TFs), which have specific binding sites on DNA and play critical roles in the gene expression regulation [5]. In complex diseases such as cancer [6,7], neuropsychiatric disorders [8], and diabetes [9], these molecules involved in the signal transduction cascade that are altered and, thus, become attractive targets for disease treatment [10,11]. Therefore, targeting signaling pathways has become an important approach to discovering new drugs through traditional experimental methods [12,13] and to predicting drug repositioning through systematic approaches [14]. However, the primary challenge for utilizing signal transduction pathways for drug discovery is to synopsise

the drug signaling pathways into one comprehensive system, including the major causal genetic factors for pathology of the complex disease and the most elemental components in the drug action.

Recent high-throughput technologies such as array-based mRNA and microRNA expression, genome-wide association studies (GWAS), and next-generation sequencing (NGS) have provided massive amounts of data, enabling investigation of drug effect through pharmacogenomic network approaches. For example, the Connectivity Map (CMap, build 02) studied the effect of 1,309 small chemicals on gene expression in four cultured human cells [15]. Furthermore, multiple reliable drug-centered databases such as DrugBank [16], KEGG (Kyoto Encyclopedia of Genes and Genomes) DRUG [17], PharmGKB (The Pharmacogenomics Knowledge Base) [18], and STITCH (Search Tool for Interactions Chemicals) [19], provide comprehensive and detailed drug information for computational discovery and/or drug design. Therefore, it is possible to integrate known drug targets, genes involved in drug pharmacokinetics (PK) and pharmacodynamics (PD) processes, drug-induced gene expression data, and disease-gene associations. Additionally, network-assisted approaches have become powerful tools to explore disease-gene, gene-gene, as well as drug-target associations in pharmacology and human disease [20–23]. Therefore, we hypothesized that the construction of a signaling pathway network to connect the upstream components and downstream signal recipients for an individual drug would increase power to identify genes that play critical roles in drug action or disease development.

In this study, we develop a computational framework, called DSPathNet, to construct one signaling pathway network (SPNetwork) for a particular drug via amalgamating drug knowledge with drug-induced gene expression data. The main purposes are to capture the principal components in the drug signal transduction process and to provide an alternative approach to identifying critical elements and modules (subnetworks) relevant to drug action. We illustrate the utility of DSPathNet using the metformin, one of the most widely prescribed anti-diabetic drugs in the world which has been recently shown to be useful for cancer treatment and prevention in people at higher risk [24–26]. We started with the collection of known drug-related genes and inference of TFs from metformin-induced gene expression data. Considering that most of the known drug-related genes participate in PK and PD processes and are located in the upstream of the signaling cascade based on their function, we defined them as “metformin upstream genes.” Likewise, we defined the TFs that receive and transmit the chemical signals at the end of the signaling cascade as “metformin downstream genes.” After overlaying the two sets of genes onto human SPNetwork, we employed random walk algorithms to construct a metformin-specific SPNetwork. The random walk-based methodology aims to identify the pathways that are closest to the known disease genes compared to other methods [27] and offers the best predictive performance [28]. The network is expected to enrich with signaling genes involved in metformin signal transduction. We performed the comprehensive gene enrichment analyses of the network using the disease genes of type 2 diabetes (T2D) from GWAS catalog [29], cancer genes from Cancer Gene Census [30], one T2D GWAS [31], three cancer GWAS [32,33], and one novel GWAS of cancer patients with T2D using metformin from BioVU [34]. The enrichment analysis results showed that the network contained a significant number of T2D and cancer disease genes and genes related to metformin action, indicating that the framework is promising as a method to identify critical genes involved in disease pathology and drug action. Additionally, the metformin-specific SPNetwork generated here provides potential metformin targets and molecular insights for further delineating the mechanism of metformin action.

## Results

### DSPathNet, a novel computational framework for exploring drug-specific signaling pathway network

In this study, we develop a novel computational framework to build a Drug-specific Signaling Pathway Network, namely DSPathNet, for constructing a signaling pathway network (SPNetwork) for an individual drug of interest. The drug-specific SPNetwork is expected to contain critical components in the drug's signal transduction cascade. These components are genes that harbor genetic variations contributing to the pathology of the drug indication or drug response. Thus, the drug-specific SPNetwork would facilitate our understanding of the molecular mechanisms of drug action, disease pathogenesis, and identification of novel drug targets. To prove the principle, we utilized the drug metformin as an example to evaluate the framework.

[Fig 1](#) outlines the framework to build the metformin-specific SPNetwork and [S1 Table](#) summarizes the data sources, software and evaluation data used in the study. Briefly, we first collected metformin upstream genes from multiple sources and inferred metformin downstream genes from metformin-induced gene expression data. We compiled a human SPNetwork from two databases, Pathway Commons [35] and TRANSFAC [36], as a background pathway system for all signal transduction processes in humans. To weight the association of each node with metformin action, we assigned a functional similarity score to each node based on their Gene Ontology (GO) annotations and metformin upstream genes. The human SPNetwork included 37,881 edges and 4,367 nodes. Then, we utilized the metformin upstream and downstream genes as seeds to produce the metformin-specific SPNetwork from the human SPNetwork via random walk approaches. In this process, we applied a crossing network strategy to generate the drug-specific SPNetwork from background human SPNetwork by longitudinal and lateral movements. Finally, we computationally evaluated the metformin-specific SPNetwork by examining the enrichment of genes in the network using two types of data. The first includes the disease genes of type 2 diabetes (T2D) and cancer, the two diseases in which metformin has been actively studied. The second contains the individual genotyping data from five GWAS datasets: one T2D GWAS dataset, three cancer GWAS datasets, and one GWAS dataset of cancer patients with T2D treated by metformin. Our evaluation results indicated that the metformin-specific SPNetwork was significantly enriched with genes with mutations that could contribute to the pathology of T2D and cancer, and genes that may be associated with metformin-associated cancer survival ([Table 1](#)). To further investigate the molecular mechanisms underlying metformin action, we built a crosstalk subnetwork based on common genes to T2D and cancer, network topology, and functional analyses. We revealed several critical components, modules, and pathways that might be involved in metformin action.

### Major steps to improve DSPathNet's performance

In order to generate a complete and reliable SPNetwork, we extensively collected the metformin related genes, rigorously selected the expressed genes induced by metformin, and comprehensively compared the performance using T2D GWAS data after the SPNetwork generation. For each step, we provide the detailed information as below.

**Collection of metformin upstream genes.** We first collected the 46 genes related to metformin from two databases DrugBank and PharmGKB. Among them, 21 genes existed in the 4,367 genes in the human SPNetwork. To collect the metformin-related genes to the maximum extent possible, we further performed literature mining on the MEDLINE abstracts to identify the gene entities that have a relation with metformin by calculating the semantic distance

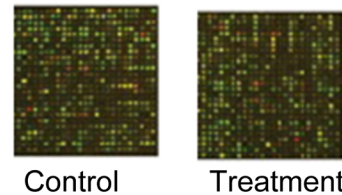
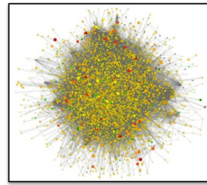
Step 1: Data collection

Metformin upstream genes

Human SPNetwork

Metformin downstream genes

Drug targets (DrugBank)  
Pharmacogenomic genes  
Drug PD/PK pathways  
Literature mining

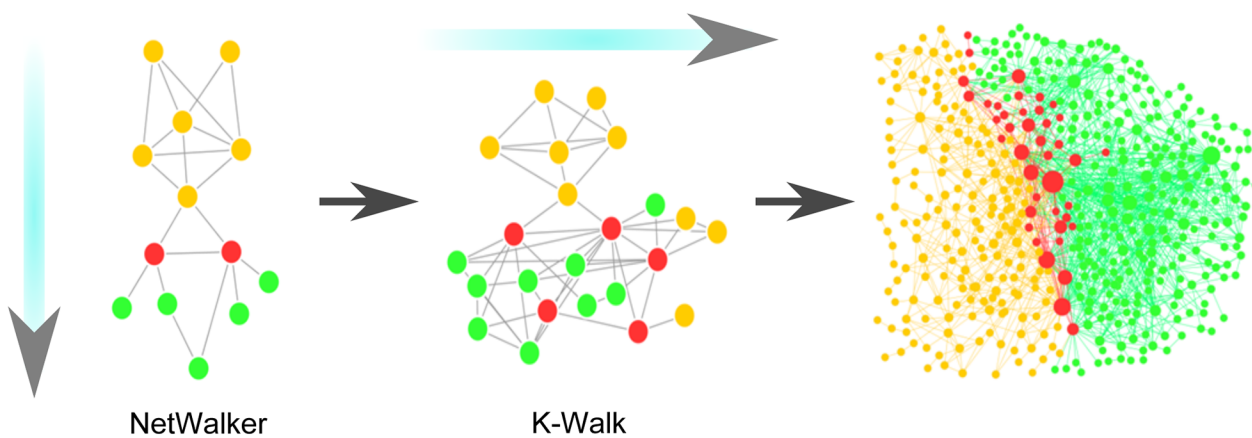


Step 2: Generation of metformin-specific SPNetwork

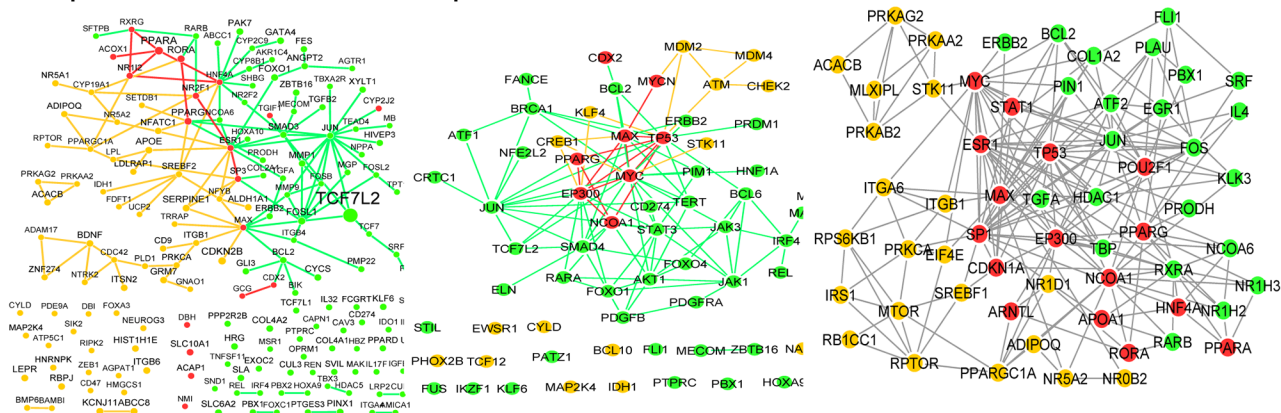
Longitudinal movement

Lateral movement

Metformin-specific SPNetwork



Step 3: Evaluation and interpretation



Subnetwork enriched with T2D-related genes

Subnetwork enriched with cancer genes

Crosstalk subnetwork of metformin in T2D and cancer

**Fig 1. Overview of DSPathNet, a novel computational framework to construct a drug-specific signaling pathway network (SPNetwork): metformin as a case.** Step 1: we collected the metformin upstream genes from multiple sources and inferred metformin downstream genes from metformin-induced gene expression data. We also compiled one human SPNetwork. Step 2: we utilized the metformin upstream and downstream genes as seeds to generate a metformin-specific SPNetwork from the human SPNetwork. The process involved longitudinal and lateral movements. Step 3: we utilized disease genes and genome-wide association studies (GWAS) data to evaluate if the metformin-specific SPNetwork was enriched with disease genes for type 2 diabetes (T2D)

and cancer, genes associated with metformin action. Furthermore, we derived a crosstalk network of metformin action for T2D and cancer in order to identify key components in the metformin signal transduction via network topological and functional analysis. The nodes in orange correspond to the drug-related upstream genes, the nodes in green to the drug-related downstream genes, and the nodes in red to the nodes common to the upstream and downstream gene networks.

doi:10.1371/journal.pcbi.1004202.g001

among the hidden topics uncovered by Latent Dirichlet Allocation (LDA) model [37]. We obtained 29 genes. Among them, ten overlapped with the 46 genes and 19 were uniquely identified by the literature searching method. Of these 19 genes, 15 were found in the human SPNetwork (S2 Table, S1 Fig). Collectively, we obtained a total of 65 genes that were regarded as metformin upstream genes, among which 36 genes could be mapped to human SPNetwork.

**Inference of metformin downstream genes.** We inferred the metformin downstream genes based on gene expression data in cancer cells after metformin treatment from Connectivity Map (CMap) (build 02) [15]. Among the ten gene expression datasets of metformin treatments (S3 Table), four had significantly consistent with each other (absolute value of the enrichment score > 0.5 and FDR  $q$ -value < 0.001) (Fig 2 and S2 Fig) by performing the gene set enrichment analysis (GSEA) [38]. Then, based on the top and bottom 100 probes for the four treatments, we identified 140 up-regulated and 215 down-regulated genes, respectively. From these genes, we identified 29 TFs whose targets were significantly enriched in up-regulated genes and 38 TFs whose targets were significantly enriched in down-regulated genes (Hypergeometric test  $P$ -value < 0.05) compared to the pairs of TFs and their targets (Materials and Methods). There was one TF (TEAD4) shared between the two sets of TFs. Thus, we identified 66 TFs in total (S4 Table). Among these TFs, only one TF (JUN) was observed in the list of the up-regulated genes and two TFs (SMAD3 and NR1I2) in the down-regulated genes. Our observation is in general agreement with previous reports that many TFs are not regulated at the transcriptional level [39,40].

**Generation and evaluation of metformin-specific SPNetwork.** We noticed that only two genes (*PPARG* and *NR1I2*) were common between the metformin upstream gene list and the metformin downstream gene list (Fig 3A). The observation indicated that some of the key components in the metformin signal transduction cascade were missed in the two sets of metformin-related genes. To address this issue, we employed a two-step strategy of random walk-based propagation to recruit more genes via a sequential two-step strategy from the human

**Table 1. Comparison of genes in metformin-specific signaling pathway network with T2D and cancer genes and genes with smallest P-value less than 0.05 in five GWAS data sets.**

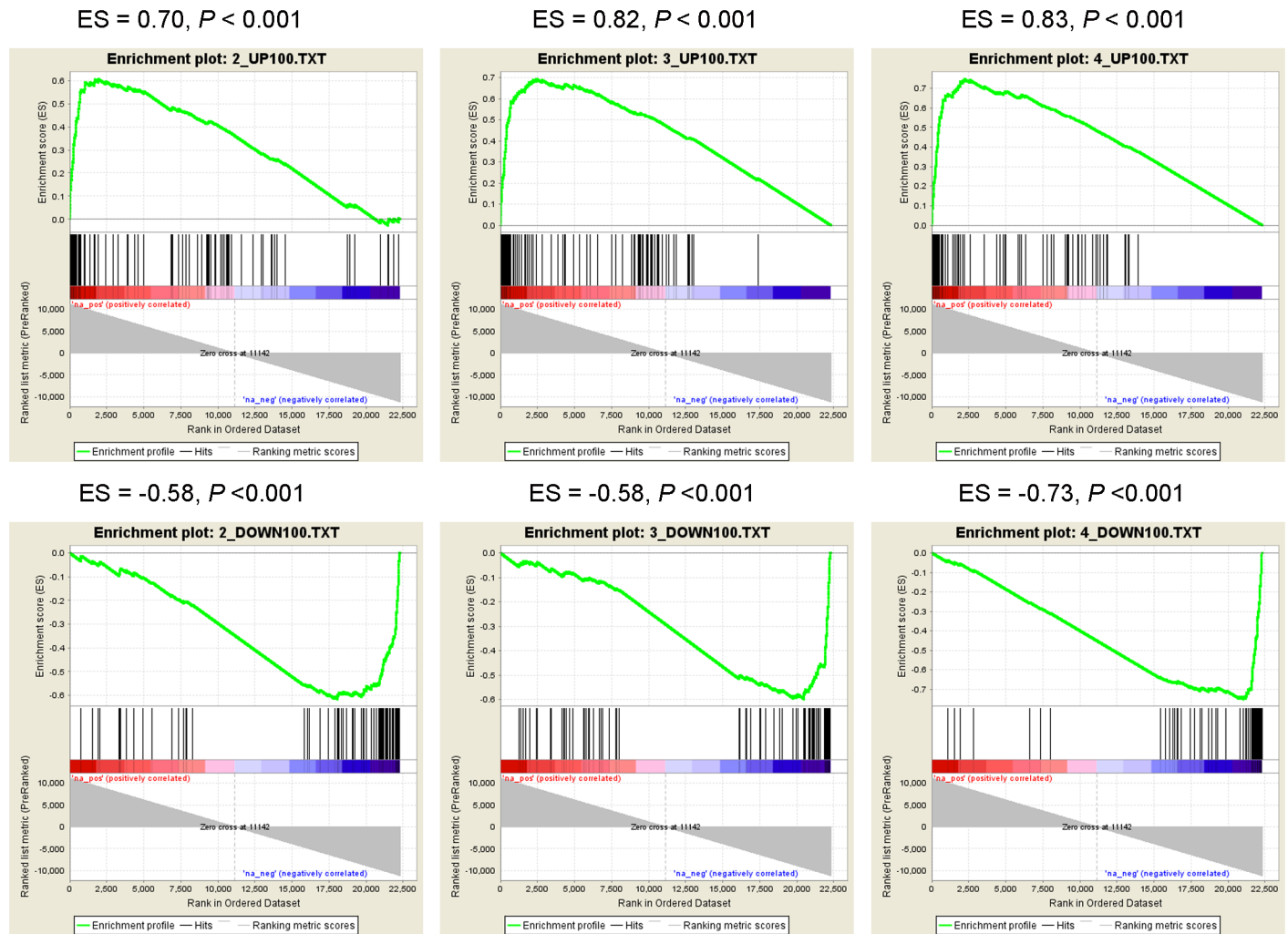
Data	Number of genes <sup>a</sup>	Number of genes with smallest $P < 0.05$	Hypergeometric test $P$ -value <sup>b</sup>
T2D disease genes	131	11	$1.36 \times 10^{-4}$
T2D GWAS	445	169	$3.08 \times 10^{-5}$
Cancer genes	509	64	$1.64 \times 10^{-29}$
Breast cancer GWAS	469	157	0.0144
Pancreatic cancer GWAS	468	170	0.0120
Prostate cancer GWAS	469	172	0.0053
Metformin GWAS <sup>c</sup>	458	177	0.0181

<sup>a</sup> For the five GWAS data sets, each number denotes the number of genes with genotyping data in corresponding GWAS data. The T2D genes were extracted from the GWAS Catalog database and the cancer genes were obtained from Cancer Gene Census.

<sup>b</sup> For disease genes, the hypergeometric test was performed by comparing with all protein-coding genes in the human. For GWAS data, the hypergeometric test was performed by comparing with genotyping data in the corresponding GWAS data set.

<sup>c</sup> Metformin GWAS: the GWAS for identifying genetic variants associated with survival among cancer patients with T2D using metformin.

doi:10.1371/journal.pcbi.1004202.t001

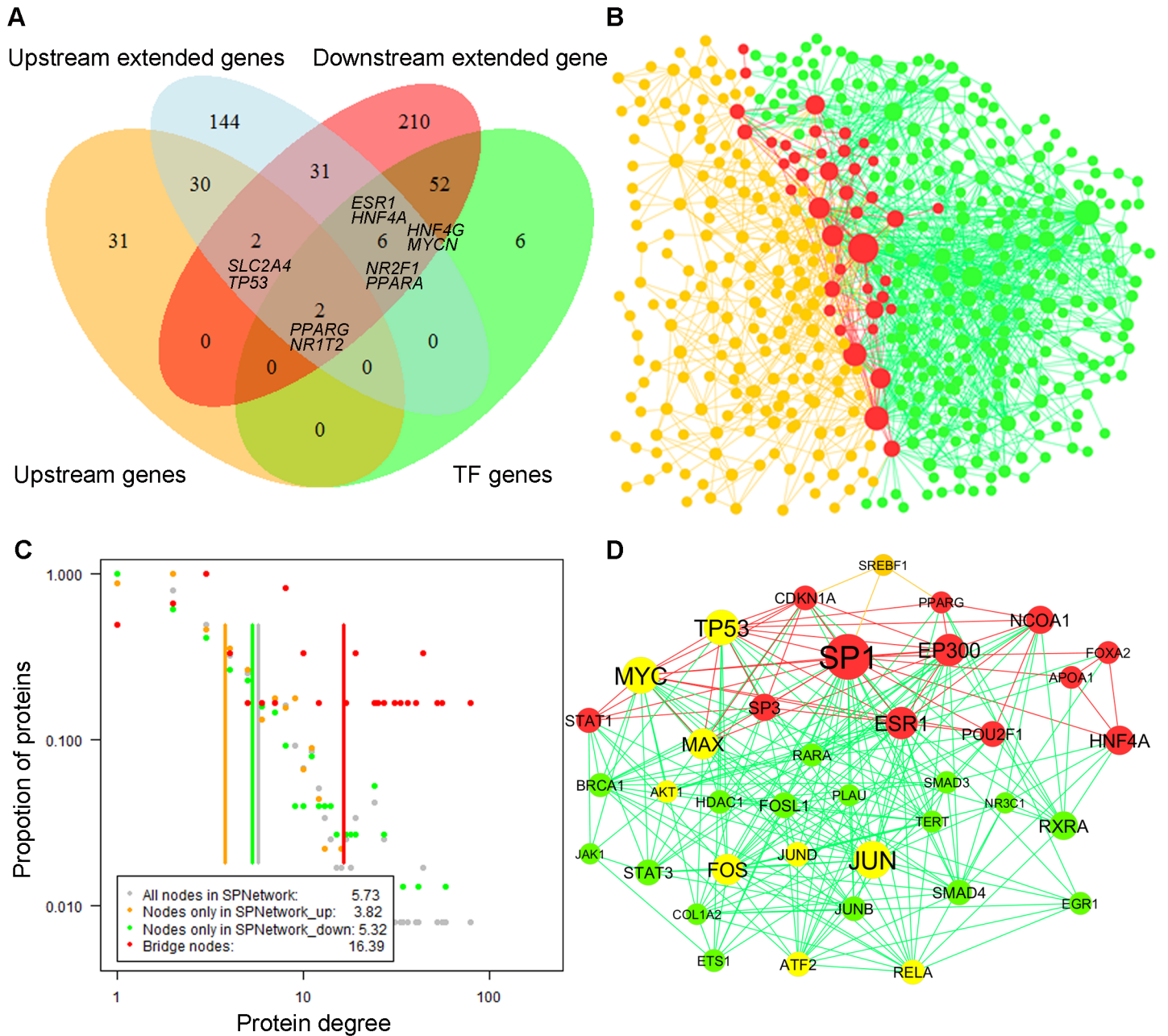


**Fig 2. Gene Set Enrichment Analysis (GSEA) enrichment score curves of metformin-induced probes in three treatments vs. treatment 1.** The four sets of probes of metformin treatments were obtained from the gene expression profiles from Connectivity Map. The three treatment instance IDs are 2, 3, and 4. The graphs on the top panels represent the ranked, non-redundant, and up-regulated probes in the second, third, and fourth treatment groups compared with probes in the first treatment group. The graphs on the bottom panels represent the ranked, non-redundant, and down-regulated probes in second, third, and fourth treatment groups compared with probes in the first treatment group. In each graph, probes on the far left (red) correlated with the most up-regulated probes in the treatment 1 and probes on the far right (blue) correlated with the most down-regulated probes in treatment 1. In each graph, the vertical black lines indicate the position of each of the probes of the studied probe set in the ordered, non-redundant data set. The green curve denotes the ES (enrichment score) curve, the running sum of the weighted enrichment score in GSEA.

doi:10.1371/journal.pcbi.1004202.g002

SPNetwork (Materials and Methods). Table 2 summarizes the number of nodes and edges generated at each step. Through two network movements, we obtained 215 upstream extended genes and 303 downstream extended genes. Then we generated one upstream network by the direct links of metformin upstream extended genes (SPNetwork\_up) and one downstream network by the direct links of metformin downstream extended genes (SPNetwork\_down). They had 41 common nodes and 84 common links. After merging the two networks by their common nodes and common links, we obtained a metformin-specific network with 477 nodes and 1,366 edges.

Compared to the two common genes between the metformin upstream genes and downstream genes, the overlap was increased 20.5 times (Fig 3A). Among the 41 nodes, besides the two common genes (*PPARG* and *NR1T2*), two genes belonged to the metformin upstream



**Fig 3. Metformin-specific signaling pathway network (SPNetwork).** A) A four-way Venn diagram summarizes the number of shared genes among metformin upstream genes represented by 'Upstream genes', metformin downstream genes ('TF genes'), metformin upstream extended genes in the metformin upstream network ('Upstream extended genes'), and metformin downstream extended genes in the metformin downstream network ('Downstream extended genes'). B) Metformin-specific SPNetwork with 477 nodes and 1366 edges. The nodes and edges in orange correspond to nodes and edges only in the metformin upstream network. The nodes and edges in green correspond to the nodes and edges only in the metformin downstream network. And the nodes and edges in red correspond to the nodes and edges common to the metformin upstream network and the metformin downstream network. C) Degree distributions and average degrees (vertical lines) of the four gene sets in the metformin-specific SPNetwork. The four gene sets are 41 common nodes, 174 nodes only in the metformin upstream network (SPNetwork\_up), 262 nodes only in the metformin downstream network, all 477 nodes in the metformin-specific SPNetwork (SPNetwork\_down). The Y-axis represents the proportion of proteins having a specific degree. D) The subnetwork of the 38 hub nodes extracted from metformin-specific SPNetwork. The legends for orange nodes and edges, red nodes and edges, and green nodes and edges are same as those in the subFig B. The nodes in yellow correspond to the genes that exist in the pathway 'MAPK signaling pathway' according to KEGG annotation.

doi:10.1371/journal.pcbi.1004202.g003



**Table 2. Summary of genes and hypergeometric tests at each step in the process of metformin-specific SPNetwork construction.**

	#. genes			#. nodes <sup>a</sup> (All/largest)	Network		<i>P</i> -value <sup>b</sup>
	Upstream	Downstream	Overlap		#. nodes with <i>P</i> < 0.05	#. edges	
Starting seed	65	66	2	98/74	36	179	0.02
Longitudinal movement	103	125	9	219/151	74	384	0.01
Lateral movement	215	303	41	477/473	169	1,366	3.08 × 10 <sup>-5</sup>

<sup>a</sup> The first number denotes the number of the nodes in the whole network while the second number denotes the number of nodes of the largest subnetwork in the network.

<sup>b</sup> The *P*-value was calculated from the hypergeometric test by comparing the number of T2D-related genes that have at least one SNP with *P*-value less than 0.05d with all genotyping genes in the T2D GWAS data based on.

doi:10.1371/journal.pcbi.1004202.t002

genes (*SLC2A4* and *TP53*), and six genes belonged to metformin downstream genes (*ESR1*, *HNF4A*, *HNF4G*, *MYCN*, *NR2F1*, and *PPARA*). The remaining 31 genes (75.6%) were novel linkers, suggesting they might play important roles in metformin action. Considering that these 41 nodes act as bridges to link the SPNetwork\_up and SPNetwork\_down, we defined them as bridge nodes.

To assess the ability of recruiting disease genes of each step, we produced one corresponding network for each step and perform disease gene enrichment analysis based on the T2D GWAS data from the Wellcome Trust Case Control Consortium (WTCCC) T2D study [31]. The concept is based on that the more significant enrichment of disease genes the corresponding network has, the more powerful the network-generating method is. Table 2 summarizes the corresponding evaluation *P*-values. Starting from the unique 129 genes of metformin upstream genes and TF genes, a network with 98 nodes and 179 edges was produced (S3 Fig). The largest module contained 74 nodes and 178 edges, which indicated that metformin upstream genes and downstream genes could regulate each other to a certain degree. Among 98 nodes, 93 had genotyping data in T2D GWAS, of which 36 genes belonging to the T2D-related genes. Compared with all genes with genotyping data in the GWAS, the hypergeometric test *P*-value was 0.02.

In the first step (longitudinal movement), from upstream genes, we obtained 103 genes, of which 36 were upstream genes and 67 were novel genes (S4A Fig). From metformin downstream genes, we obtained 125 genes that contained 62 metformin downstream genes and 63 novel genes (S4B Fig). Between the 103 genes and the 125 genes, there were nine common genes. Then, a subnetwork was created by their direct interactions with 219 nodes (S5 Fig). In the network, the largest module included 151 (68.9%) of the 219 genes, indicating that about one-third of genes (68, 31.1%) could not be recruited in the biggest subnetwork. In addition, among the 219 nodes, 207 had genotyping data, in which 74 were T2D-related genes. Compared with all genes with genotyping data in the T2D GWAS, this network module is statistically enriched with genes having small *P* values (Hypergeometric test, *P*-value: 0.01), but the significance is not very strong.

In the second step (lateral movement), from the metformin upstream longitudinal genes (103), we obtained 215 genes, including 34 metformin upstream genes, 60 metformin upstream longitudinal genes, and 121 novel genes (S4A Fig). From the metformin downstream longitudinal genes (125), we obtained 303 genes comprised of 60 metformin upstream genes, 56 metformin upstream longitudinal genes, and 187 novel genes (S4B Fig). After merging their direct interactions, we obtained a network with 477 nodes and 1,366 edges. Among the 477 nodes, 473 nodes (99.2%) formed a big module. The network had the strongest association with T2D-

related genes ( $P$ -value:  $3.08 \times 10^{-5}$ ). More importantly, the number of common nodes increased to 41. Therefore, the crossing movement strategy is promising to capture the cascade of signal flow and complexity of cross-talking among different pathways involved in signal transduction from the upstream genes and downstream genes.

### Metformin-specific SPNetwork provides a valuable source for understanding metformin action

The final metformin-specific SPNetwork generated above comprised 477 nodes and 1,366 edges (Fig 3B, S5 Table). Among the 477 nodes, 215 belonged to metformin upstream network, while 303 belonged to metformin downstream network. There were 41 bridge nodes between them. Thus, 174 genes were unique to the metformin upstream network, and 262 genes were unique to the metformin downstream network. From here, we refer to the three gene sets as upstream genes (number of genes: 174), downstream genes (262), and bridge genes (41) for follow-up network topological and functional analyses.

To explore the topological properties of this SPNetwork, we calculated node degrees (connectivity) and their distribution [41]. In this network, degree values of nodes ranged from 1 to 79 and the average degree was 5.73. The degree distribution was strongly right-skewed, indicating that most nodes had a low degree and only a small portion of the nodes had a high degree (Fig 3C). The nodes with a high degree act as hubs in the network and hold the whole network together [41]. In biological networks, hubs are more likely to be essential genes [42] and disease genes [43–45]. Using the hub defining method proposed by Yu et al. [46], we determined 38 hubs whose degrees were larger than 14. Among them, one gene (*PPARG*) belonged to both metformin upstream and downstream gene sets, two genes (*TP53* and *SREBF1*) were metformin upstream genes, 13 belonged to the metformin downstream gene set only, and 22 were novel genes. After extracting these hubs from metformin-specific SPNetwork, we generated a hub-centered subnetwork (Fig 3D). Among the 38 hubs, 19 (50.00%) are included in ‘pathway in cancer’ and 9 (23.68%) in ‘MAPK signaling pathway’ according to KEGG pathway annotation. The MAPK signaling pathway plays important roles in the pathology of both cancer [47] and diabetes [48]. Thus, the 477 genes had two genes belonging to metformin upstream and downstream genes, 33 to the metformin upstream genes, 58 to the metformin downstream genes, and 384 novel genes (S6 Table). The novel genes may provide a valuable resource for further investigation of the pathology of T2D and cancer, and the metformin action.

We further examined pathway enrichment in these 477 nodes based on KEGG pathway annotation using the online tool WebGestalt [49]. We identified 69 significant pathways (adjusted  $P$ -value  $< 1.00 \times 10^{-4}$ ) (S7 Table). According to the KEGG pathway first-level category annotation (Materials and Methods), 12 pathways belonged to ‘environmental information processes,’ nine to ‘cellular processes,’ 18 to ‘organismal systems,’ and 29 to ‘human disease.’ Among these 12 environmental information processes pathways, eight were signal transduction pathways, of which the top three pathways were ‘MAPK signaling pathway’ (32 genes, adjusted  $P$ -value:  $3.39 \times 10^{-22}$ ), ‘mTOR signaling pathway’ (13 genes, adjusted  $P$ -value:  $6.39 \times 10^{-14}$ ) and ‘ErbB signaling pathway’ (15 genes, adjusted  $P$ -value:  $1.89 \times 10^{-13}$ ). Among the 18 pathways related to organismal systems, five belonged to the endocrine system, of which the top three pathways were ‘adipocytokine signaling pathway’ (22 genes, adjusted  $P$ -value:  $3.19 \times 10^{-25}$ ), ‘PPAR signaling pathway’ (22 genes, adjusted  $P$ -value:  $5.36 \times 10^{-25}$ ), and ‘insulin signaling pathway’ (23 genes, adjusted  $P$ -value:  $1.91 \times 10^{-19}$ ). Among the 29 pathways related to human disease, 15 were directly related to cancer. Importantly, the pathway ‘type II diabetes’ (10 genes, adjusted  $P$ -value:  $1.12 \times 10^{-10}$ ) and the ‘maturity onset diabetes of the young’ (8 genes, adjusted  $P$ -value:  $1.94 \times 10^{-10}$ ) were among the

enriched pathways. Together, the evidence indicates that the metformin-specific SPNetwork involves both diabetes and cancer at the pathway level.

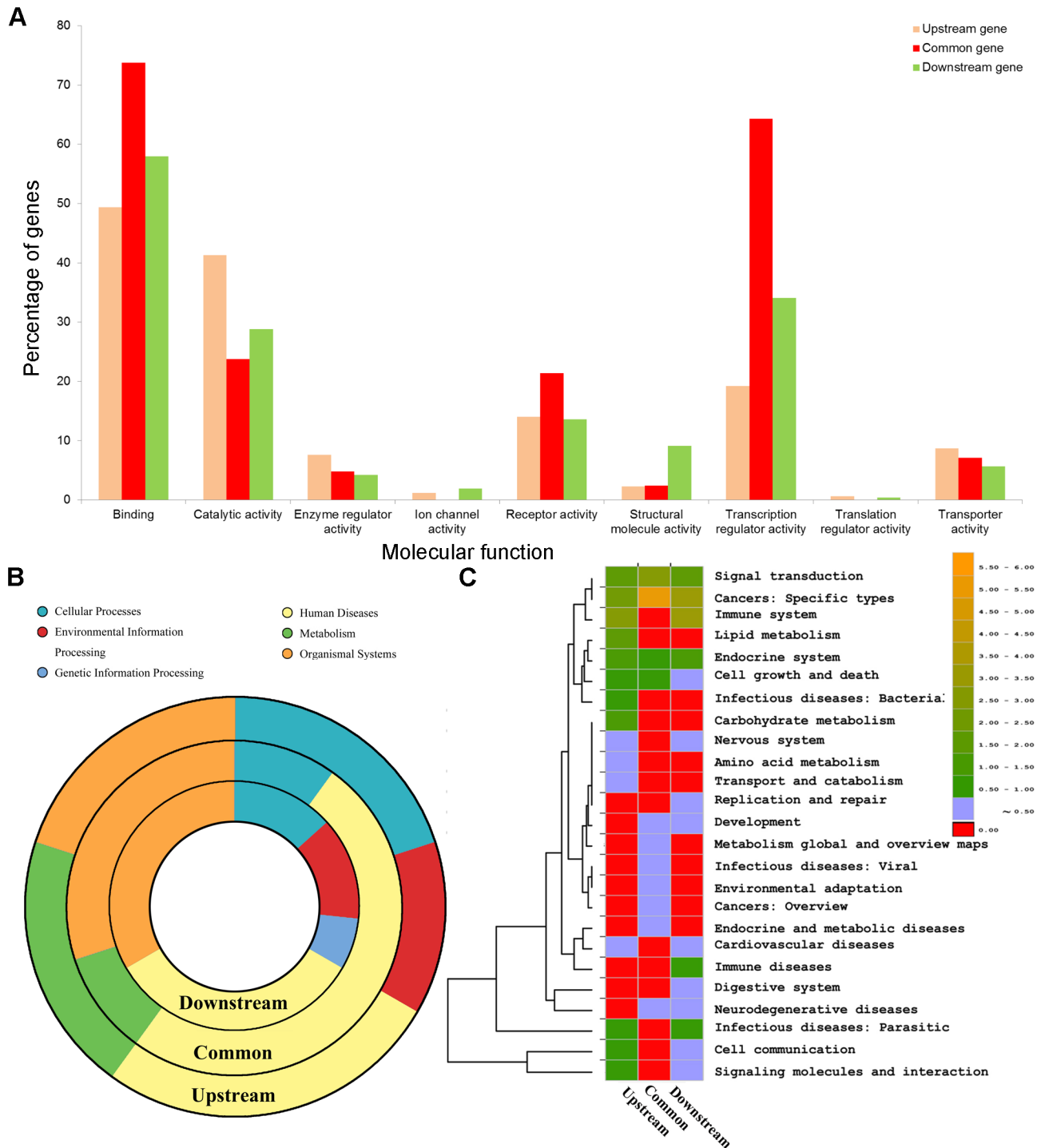
## Network topological and functional properties of bridge genes

In the metformin-specific SPNetwork, there were 41 genes (bridge genes) common to both the metformin upstream and downstream networks. As mentioned above, most of them (31, 75.6%) were novel linkers (S6 Table). To interrogate their roles, we compared them with upstream genes (174) and downstream genes (262) via network topological and functional analyses, as described below.

**Bridge genes tended to have higher degree.** In the metformin-specific SPNetwork, the average degree of these bridge genes was 16.39, significantly higher than that of upstream genes (3.82, Wilcoxon's test,  $P$ -value:  $2.41 \times 10^{-6}$ ) or that of downstream genes (5.32,  $P$ -value:  $1.69 \times 10^{-6}$ ) (Fig 3C). The result indicated that the bridge genes strongly connected in the metformin-specific SPNetwork. In line with this, the bridge nodes are more likely to be the hubs: 15 out of 41 (36.9%) as compared to 1 out of 174 upstream genes (0.57%) or 22 in 262 downstream genes (7.6%). These 15 genes were *SP1* (degree: 79 in the metformin-specific network), *MYC* (55), *TP53* (52), *EP300* (44), *ESR1* (44), *MAX* (40), *HNF4A* (36), *NCOA1* (33), *SP3* (31), *POU2F1* (27), *STAT1* (25), *CDKN1A* (24), *APOA1* (19), *FOXA2* (19), and *PPARG* (17). This observation indicates that these nodes might play important roles to maintain the network topology that is important for biological function.

**Bridge genes had different functional tendencies.** To further explore the functional characteristics of these bridge genes, we first compared them with upstream genes and downstream genes based on the GO Molecular Function domain using the online tool PANTHER Classification System [50] (Fig 4A). The proportion of genes in the following three GO terms were higher in the bridge genes than that in the upstream or downstream genes: binding (GO:0005488), receptor activity (GO:0004872), and transcription regulator activity (GO:0030528). However, for the following three GO terms, the proportion of upstream genes was significantly higher than that in other two gene sets: catalytic activity (GO:0003824), enzyme regulator activity (GO:0030234), and transporter activity (GO:0005215). For the downstream genes, only one GO term, structural molecule activity (GO:0005198), had a higher proportion.

We also examined the enriched pathways in the three sets of genes according to the KEGG enrichment analyses using the tool WebGestalt. By applying an adjusted  $P$ -value of less than 0.05, we found that 92 pathways were significantly enriched in the 174 upstream genes, 105 pathways in the 262 downstream genes, and 28 pathways in 41 common genes (S8 Table). To simplify the comparison, we categorized them into seven categories at the first level and 43 categories at the second level in the KEGG pathway annotation system (Materials and Methods). To represent the relative abundance of the pathways, we further calculated a  $Z$ -score for each category at the second level (Materials and Methods). Accordingly, among the 92 pathways for upstream genes, 73 pathways were grouped into the five first-level categories and 15 second-level categories ( $Z$ -score  $> 0$ ). Among 106 enriched pathways in downstream genes, 86 pathways were grouped in five first-level categories and 15 second-level categories ( $Z$ -score  $> 0$ ). All of the 28 enriched pathways in 41 common genes were categorized into five first-level categories and 11 second-level categories ( $Z$ -score  $> 0$ ) (S9 Table). Fig 4B summarizes the comparison of the three sets of genes at the first-level category. We observed that each of the three sets of genes had their own participating tendency in particular biological processes. For example, among the 73 enriched pathways in the upstream genes, 15 (20.55%) belonged to the metabolism category, which was substantially higher than that in the common genes (1, 3.57%) or that



**Fig 4. Functional comparison of the common genes, upstream network genes, and downstream network genes.** The common genes were those found in both metformin upstream network and downstream network. The upstream network genes were those only belonging to the metformin upstream network. The downstream network genes were those only belonging to the metformin downstream network. A) Proportion of genes of interest in Gene Ontology (GO) molecular function domain. B) Comparison of proportion of enriched pathway in the three gene sets at the first-level category of KEGG annotation. C) The clustering of enriched pathways for the three gene sets at second-level category of KEGG annotation.

doi:10.1371/journal.pcbi.1004202.g004

in the downstream genes (0). Among the 28 enriched pathways in the common genes, 15 (53.57%) belonged to the human diseases, which was higher than that in the upstream genes (19, 26.03%) or that in the downstream genes (32, 37.21%). Fig 4C further shows the pathway comparison of the three sets of genes at the second level. While all genes in the three gene sets were enriched in the pathways related to cancer, compared to upstream and downstream genes, the bridge genes were the likeliest to be involved in the cancer-related pathways.

Among the 41 bridge nodes, 25 were TFs according to TRANSFAC database. Among them, eight (ESR1, HNF4A, HNF4G, MYCN, NR1I2, NR2F1, PPARA, PPARG) were inferred metformin-related TFs based on metformin-induced gene expression data and the remaining 16 TFs were identified as the novel linkers between metformin upstream and downstream network. They were ARNTL, CDX2, EP300, FOXA2, MAX, MYC, NCOA1, PHOX2A, POU2F1, RORA, RXRG, SP1, SP3, STAT1, TGIF1, and USF2. Among them, MYC is encoded by a well-known oncogene that acts as a pluripotent modulator of transcription during normal cell growth and proliferation [51]. Interestingly, several other TFs cooperate with MYC under some particular conditions such as CDX2 [52], MAX [53], SP1 [54–57], SP3 [58], and STAT1 [54]. For example, CDX2, one caudal-related homeobox transcription factor, mediates E-selectin ligand expression in colon cancer cells with MYC together [52].

In summary, our network and functional analyses indicated that these common genes act as bridges between the metformin upstream and downstream networks so that they might act in metformin-specific SPNetwork. Therefore, these bridge genes, especially the novel genes, are warranted for further investigation of their roles in the signal transduction cascade of metformin action.

### Metformin-specific SPNetwork is significantly enriched with T2D associated genes

Since metformin is a well-studied drug for T2D treatment, the metformin-specific SPNetwork was expected to contain genes that have genetic association with T2D. To examine this expectation, we comprehensively performed enrichment analysis using two sets of genes. The first one contained 131 genes collected from 66 T2D GWAS studies curated by the NHGRI GWAS Catalog database (April 1, 2014) [29]. Those genes have been reported to be significantly associated with T2D based on GWA studies. Here, we selected these genes having at least one SNP with  $P$ -value less than  $1.0 \times 10^{-8}$  as T2D associated genes. The second set included the T2D-related genes from the WTCCC T2D study [31] as mentioned above.

Among the 477 nodes in the metformin-specific SPNetwork, 11 genes were found in the first set of 131 genes. Compared to the human protein-coding genes (20,716), the network was significantly enriched for T2D associated genes (Hypergeometric test,  $P$ -value:  $1.36 \times 10^{-4}$ ). Similarly, among the 131 T2D disease genes, 43 existed in the human SPNetwork (4,367). Thus, compared to all nodes in the human SPNetwork, the metformin-specific SPNetwork was significantly enriched for T2D associated genes too ( $P$ -value:  $3.62 \times 10^{-3}$ ). These 11 genes were *CDKN2B*, *HNF1A*, *HNF4A*, *IRS1*, *ITGB6*, *KCNJ11*, *LEP*, *PPARD*, *PPARG*, *SND1*, and *TCF7L2*. Among them, *KCNJ11*, *PPARG*, and *TCF7L2* have the strongest genetic association among genes that appear in the T2D GWAS studies based on a comprehensive review [59].

Among the 477 genes in metformin-specific SPNetwork, 445 had genotyping data from WTCCC T2D GWAS dataset. Among them, 169 genes belonged to T2D-related genes. Compared with all genes with genotyping data in the GWAS, the network was significantly enriched with T2D-related genes (Hypergeometric test  $P$ -value:  $3.08 \times 10^{-5}$ ). We further compared the 169 genes with the genes having genotyping data in the human SPNetwork. Among the 4,367 nodes in the human SPNetwork, 3,446 genes had genotyping data, in which 1,048 genes were

T2D-related genes. Thus, the metformin-specific SPNetwork was significantly enriched for the T2D-related genes as compared to the whole human SPNetwork in this study ( $P$ -value:  $7.47 \times 10^{-3}$ ). Fig 5A shows the comparison of  $P$ -value distributions of genes in whole GWAS data (T2D GWAS), human SPNetwork, and metformin-specific SPNetwork. These comparisons indicate that the network is enriched with genes that might be involved in the pathology of T2D.

We further generated a subnetwork for 169 nominally significant genes with T2D (Fig 5B) by their direct links. Among the 169 genes, 50 genes had SNPs whose  $P$ -values were less than 0.01 in the WTCCC T2D GWAS. In addition, there were six genes observed in both the 131 GWAS Catalog genes and the 169 genes; they are *CDKN2B*, *ITGB6*, *KCNJ11*, *PPARD*, *PPARG*, and *TCF7L2*. Among them, the SNP rs4506565 in gene *TCF7L2* has the strongest significance ( $P = 5.68 \times 10^{-13}$ ). *TCF7L2* encodes a transcription factor that regulates the transcription of several genes. It is a key element in the WNT signaling pathway, which has been reported to contribute to T2D risk significantly [59].

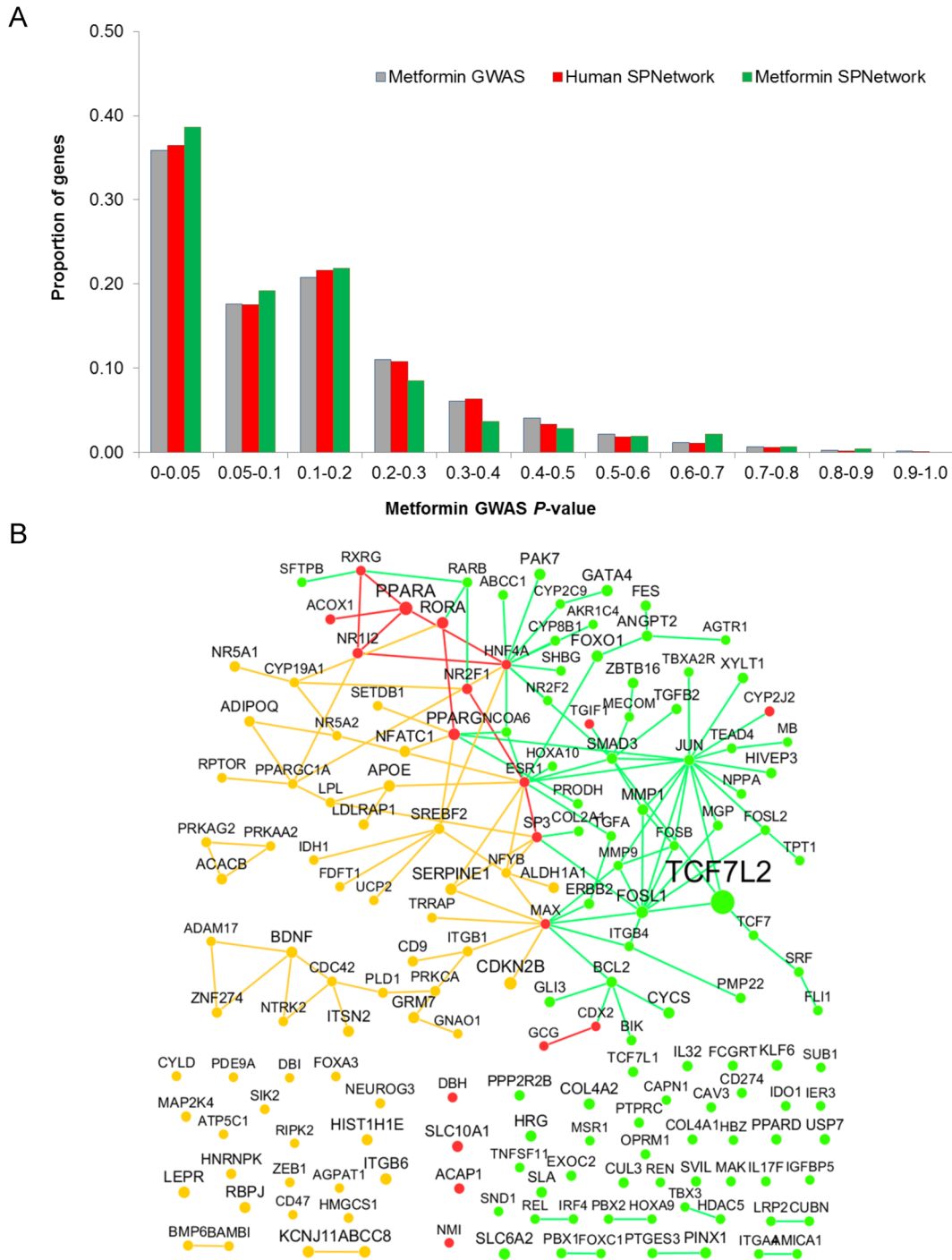
### Metformin-specific SPNetwork is enriched with cancer genes

Above pathway analysis indicated that the metformin-specific SPNetwork was significantly associated with cancer-related pathways. Here, we further examined if the SPNetwork is enriched with cancer genes from four data sets. The first one included 509 cancer genes downloaded from the Cancer Gene Census (December 11, 2013, <http://cancer.sanger.ac.uk/cosmic/census>). Among them, 64 genes were included in the metformin-specific SPNetwork. Compared to all human genes or the protein-coding genes in the human SPNetwork, the network was significantly enriched with cancer genes (Hypergeometric test,  $P$ -value:  $1.64 \times 10^{-29}$  and  $6.48 \times 10^{-8}$ , respectively). Interestingly, 3 of the 64 genes (*HNF1A*, *PPARG*, and *TCF7L2*) were in the T2D GWAS Catalog gene list, and 21 genes belonged to 169 T2D-related genes (see above). This observation strongly indicates that metformin may affect the shared genetic risk factors between T2D and cancer. Such information provides clues for how metformin acts in T2D and cancer treatments. This observation also provides evidence for epidemiological studies of metformin in both T2D and cancer [50].

Additionally, we performed the GSEA of the metformin-specific SPNetwork using three cancer GWAS datasets from the Cancer Genetic Markers of Susceptibility (CGEMS) projects (breast cancer [32], pancreatic cancer [33], and prostate cancer [32]). Table 1 summarizes the corresponding gene numbers in each GWAS dataset. Compared with all genes with genotyping in each GWAS dataset, the metformin SPNetwork was slightly significantly enriched in nominally significantly associated genes (Hypergeometric test  $P$ -values: 0.0144, 0.0120, and 0.0053 for breast, pancreatic, and prostate cancer, respectively). Though the results of these statistical tests are not as robust as that of the genotyping data from the T2D GWAS study, the results confirm that the metformin-specific SPNetwork was enriched with genetic factors associated with cancer development.

### Metformin-specific SPNetwork is enriched with genes associated with overall survival of cancer patients with T2D using metformin

From above analyses, the metformin-specific SPNetwork is enriched with genes associated with T2D and cancer. Several studies over the last few years have demonstrated that patients using metformin have reduced cancer risk and improved cancer survival in T2D patients [24,26,60,61]. Thus, we evaluated whether metformin-specific network enrich genes associated with cancer survival among cancer patients with T2D using metformin. In this study, we took advantage of GWAS data of cancer subjects with T2D treated with metformin from BioVU



**Fig 5.** (A) Comparison of gene-level *P*-value distribution in T2D GWAS among three gene sets from the metformin-specific SPNetwork, human SPNetwork, and genes covered by T2D GWAS. (B) Interactions were extracted from metformin-specific SPNetwork. These interactions occur between T2D-related genes, namely, their smallest *P*-value less than 0.05 in T2D GWAS. The legends for orange nodes, red nodes, and green nodes are same as in Fig 3.

doi:10.1371/journal.pcbi.1004202.g005

[34,62] (Materials and Methods). Hereafter, this dataset is referred as “metformin GWAS.” Among the 477 nodes in the metformin-specific SPNetwork, 458 genes had genotyping and 177 genes were nominally significantly ( $P$ -value  $< 0.05$ ) associated with T2D with better survival. Compared with all genes with genotyping data in the metformin GWAS data, the metformin-specific SPNetwork was enriched with nominally significant genes too (Hypergeometric test,  $P$ -value: 0.0181). We further compared the  $P$ -value distribution of metformin GWAS data for three gene sets: the metformin-specific SPNetwork, human SPNetwork, and all genes in metformin GWAS data set (S6 Fig). The genes in the metformin SPNetwork had the highest proportion of  $P$ -values ( $P$ -value  $< 0.05$ ) in metformin GWAS data at the gene level.

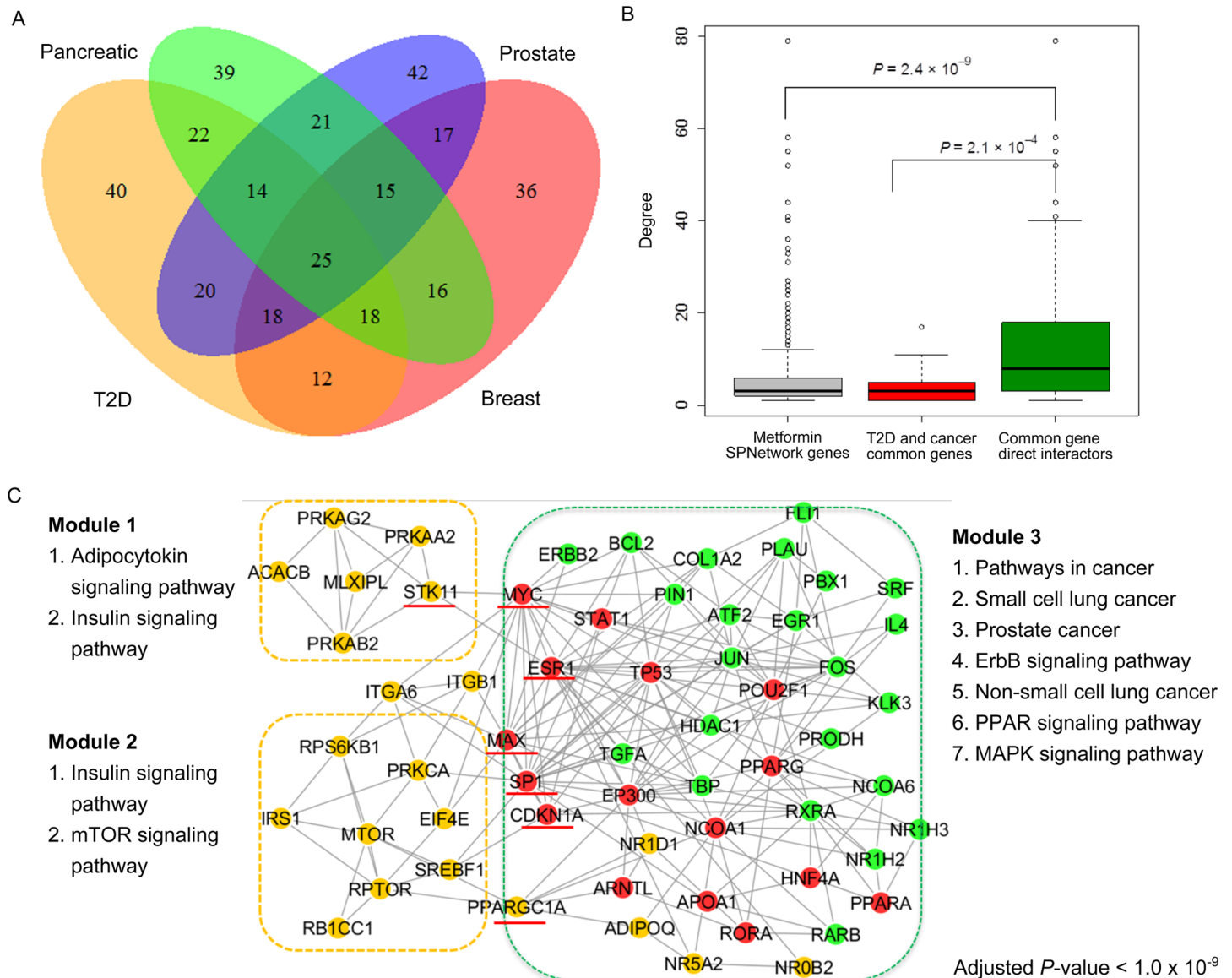
Among the 177 genes, 81 genes were included in the 169 genes whose smallest  $P$ -values were less than 0.05 in T2D GWAS data. While most of them did not link to each other (S7 Fig), these 81 genes directly linked to other 175 genes to form a subnetwork that included 256 nodes and 910 edges. This feature indicated that the 81 genes and their direct interactors dominated the metformin-specific SPNetwork. For example, the 256 nodes accounted for 53.7% of all nodes and the 910 edges accounted for 66.6% of all edges in the metformin-specific SPNetwork. Additionally, among the 81 genes, 17 belonged to ‘pathway in cancer’: *COL4A1*, *COL4A2*, *ERBB2*, *GLI3*, *ITGB1*, *MECOM*, *MMP1*, *PLD1*, *PRKCA*, *RARB*, *RXRG*, *SMAD3*, *TCF7L1*, *TCF7L2*, *TGFA*, *TGFB2*, and *ZBTB16*. Collectively, the above observations indicate that the network was enriched in genes that might contribute to overall survival among cancer patients with metformin therapy.

### Crosstalk subnetwork intertwines the key genes for metformin action in T2D and cancer

From above analyses, we observed that the metformin-specific SPNetwork was enriched with genes associated with T2D and cancer, and genes associated with metformin-associated cancer survival. To gain more insights into how metformin act in T2D and cancer treatment, we generated a subnetwork to synopsis the crosstalk between T2D and cancer based on the common genes with nominal significance ( $P$ -value  $< 0.05$ ) among the four GWAS data sets (T2D, CGEMS breast cancer, pancreatic cancer, and prostate cancer). There were 25 genes common to all the four gene sets (Fig 6A), and there were only five edges in the metformin-specific SPNetwork (S8 Fig). By further examining degree distributions of the common 25 genes and their direct interactors (71 genes), we found that their interactors had significantly more interactions than the 25 genes as well as all the genes in the metformin-specific SPNetwork (Wilcoxon’s test  $P$ -value:  $2.1 \times 10^{-4}$  and  $2.4 \times 10^{-9}$ , respectively) (Fig 6B). The 25 genes included one hub (*PPARG*) while the 71 genes included 21 of the 38 hub nodes in the metformin-specific SPNetwork. Similarly, the 25 genes contained three bridge nodes while the 71 genes contained 15 of the 41 bridge nodes between metformin upstream and downstream network. These observations indicate that the interactors of the 25 common nodes were more likely to play important roles for signal transduction.

Starting with the 25 genes and their 71 interactors, we assembled a subnetwork by their direct links among 96 nodes. The subnetwork comprised 96 nodes and 269 edges (S9 Fig). To further explore the metformin treatment mechanisms in T2D and cancer through the protein modules, we utilized software CFinder to perform network cluster and community analysis [63]. We required each node in one module participate at least one 3-vertex clique. Accordingly, we obtained three modules, which contained 6, 9, and 51 genes, respectively (S10 Fig). We found no gene shared between the first and second modules, but one gene (*STK11*) common to the first and third modules, or five genes (*EIF4E*, *PPARGC1A*, *PRKCA*, *RPS6KB1*, and *SREBF1*) common to the second and third modules. All the genes of the first and second





**Fig 6. Common genes and a crosstalk subnetwork between T2D and cancer.** (A) The four-way Venn diagram summarizes the number of shared genes among the four gene sets with smallest  $P$ -value less than 0.05 in the T2D GWAS and the three types of cancer GWAS data sets (breast, pancreatic, prostate) in metformin-specific SPNetwork. (B) Degree comparison of common genes among the four gene sets in A, common genes' direct interactors, and all genes in metformin-specific SPNetwork. (C) A crosstalk subnetwork of metformin action for T2D and cancer with three modules and enriched pathways. The legends for orange nodes, red nodes, and green nodes are same as in Fig 3. The nodes with underlines are key components in the metformin signal transduction process.

doi:10.1371/journal.pcbi.1004202.g006

modules belonged to metformin upstream network while most of the genes in the third module belonged to metformin downstream network. We merged them to form a network, which included 60 nodes and 210 edges (Fig 6C). Since this subnetwork was generated from common genes to T2D and cancer genotyping data, we defined it as a crosstalk subnetwork of metformin action in T2D and cancer.

We realized that, if we removed the nodes (CDKN1A, ESR1, MAX, MYC, PPARGC1A, STK11, and SP1), the connections among three modules would be lost (S11 Fig). Among them, three (MAX, MYC, and SP1) were both the bridge nodes and hub nodes. Therefore, these

seven nodes might be functionally critical in the metformin signal transduction cascade. To further explore how the three modules and the seven key nodes might be related to metformin treatment in term of biological function meaning, we performed the KEGG pathway enrichment analysis on each module. [Table 3](#) summarizes the enriched pathways for each module (adjusted  $P$ -value  $< 1.0 \times 10^{-4}$ ). We labeled the enriched KEGG pathways (adjusted  $P$ -value  $< 1.0 \times 10^{-9}$ ) for each module in [Fig 6C](#).

In the first and second modules, there were two common pathways: adipocytokine signaling pathway and insulin signaling pathway. Adipocytokine signaling pathway was the top pathway in the first module (adjusted  $P$ -value:  $2.01 \times 10^{-13}$ ). The adipocytokine is a group of cytokines secreted by adipose tissue, which contributes to the development of insulin resistance, T2D, and cardiovascular disease [64,65]. The insulin signaling pathway, the top pathway in the second module, plays important roles in many complex diseases such as diabetes, obesity [66], and neurological disorders [67]. In addition, the mTOR signaling pathway and ErbB signaling pathway were also enriched in the second module. There were 28 pathways enriched in the third community. According to KEGG pathway annotation at the second level, 15 of these 28 pathways belonged to human disease, six to signal transduction, and three belonged to the endocrine system, one to cell communication, one to cell growth and death, one to development, and one to environmental adaptation. Among the 15 human disease related pathways, 11 were for specific types of cancer. Therefore, the three modules reflected different biological processes involved in T2D and cancer. Additionally, the pathway analyses highlighted the seven nodes that are not only topological linkers but also functional linkers in the crosstalk SPNetwork of metformin action in T2D and cancer.

### Literature mining further reveals a novel MYC-centered pathway may play critical roles in metformin action

Starting from above crosstalk subnetwork and the seven key nodes, we manually checked their publications and integrated the experimental evidence for further understanding their roles in the metformin actions. Through careful review, we summarized their function and action together and found that a novel MYC-centered pathway was hidden under the crosstalk subnetwork, which may play important roles in metformin action in T2D and cancer ([Fig 7](#)). The Myc-centered pathway included AMPK, STK11, MYC, SP1, and CDKN1A, which formed two small motifs: AMPK-STK11-MYC and MYC-SP1-CDKN1A.

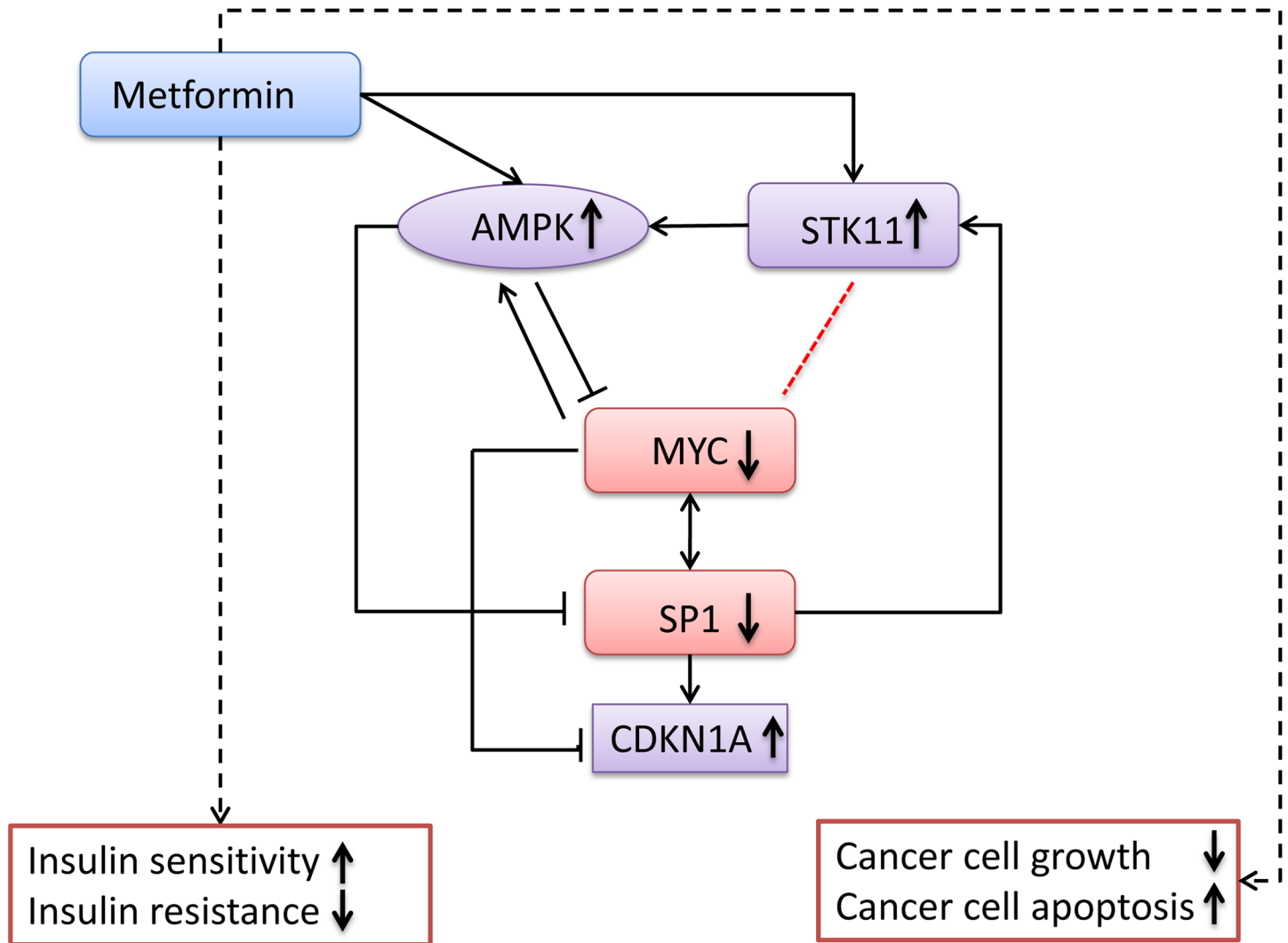
It is well known that metformin exerts anti-diabetes and anti-cancer effects via mitochondrial complex I inhibition [68,69]. Mitochondrial complex I inhibition increases AMP/ATP ratio, which activates AMP-activated protein kinases (AMPKs) [70] to cause human disease [71]. In the crosstalk subnetwork, the first module contained core members of AMPK signaling pathways (PRKAA2, PRKAB2, and PRKAG2), which is linked to the second and third modules through the STK11-MYC interaction. The gene *LKB1* encodes a key upstream activator of AMPK [51] and is known to be inactivated through mutations during lung carcinogenesis [72]. Furthermore, the metformin induces activation of *LKB1* [73]. For the MYC and *LKB1*, several lines of evidence show they are in opposite action in tumor. For example, *LKB1* is over-expressed partly by degradation of MYC protein to inhibit lung carcinoma cell proliferation [74]. Nevertheless, their direct relationship is not clear. Recent studies have shown that metformin has an ability to reduce MYC protein level in vivo and in vitro in several types of cancer, including lung cancer [75] and prostate cancer [76]. Based on the integrative network and function analyses with experimental evidence, we suggested a feed-forward loop (AMPK-STK11-MYC) exists in metformin action. This network motif may act cohesively to strengthen the inhibition of MYC expression.

**Table 3. KEGG pathways overrepresented in genes in three modules of the reduced common network between T2D and cancer for metformin.**

Pathway	#. interest genes	Adjusted <i>P</i> -value <sup>a</sup>
<b>Module # 1</b>		
Adipocytokine signaling pathway	5	$2.01 \times 10^{-13}$
Insulin signaling pathway	4	$3.00 \times 10^{-9}$
Hypertrophic cardiomyopathy (HCM)	3	$1.83 \times 10^{-7}$
<b>Module # 2</b>		
Insulin signaling pathway	7	$1.05 \times 10^{-15}$
mTOR signaling pathway	4	$1.18 \times 10^{-9}$
Adipocytokine signaling pathway	3	$1.04 \times 10^{-6}$
ErbB signaling pathway	3	$1.65 \times 10^{-6}$
Aldosterone-regulated sodium reabsorption	2	$6.64 \times 10^{-5}$
Type II diabetes mellitus	2	$7.25 \times 10^{-5}$
Acute myeloid leukemia	2	$8.77 \times 10^{-5}$
<b>Module # 3</b>		
Pathways in cancer	19	$6.89 \times 10^{-26}$
Small cell lung cancer	8	$2.94 \times 10^{-12}$
Prostate cancer	7	$2.06 \times 10^{-10}$
ErbB signaling pathway	7	$2.06 \times 10^{-10}$
Non-small cell lung cancer	6	$6.09 \times 10^{-10}$
PPAR signaling pathway	6	$2.54 \times 10^{-9}$
MAPK signaling pathway	8	$8.80 \times 10^{-9}$
Huntington's disease	7	$1.66 \times 10^{-8}$
Focal adhesion	7	$2.74 \times 10^{-8}$
Colorectal cancer	5	$7.08 \times 10^{-8}$
Hepatitis C	6	$7.15 \times 10^{-8}$
Adipocytokine signaling pathway	5	$9.45 \times 10^{-8}$
Leishmaniasis	5	$1.17 \times 10^{-7}$
Thyroid cancer	4	$1.76 \times 10^{-7}$
Bladder cancer	4	$7.65 \times 10^{-7}$
Cell cycle	5	$1.46 \times 10^{-6}$
Wnt signaling pathway	5	$3.53 \times 10^{-6}$
Glioma	4	$3.76 \times 10^{-6}$
Pancreatic cancer	4	$4.82 \times 10^{-6}$
Chronic myeloid leukemia	4	$5.43 \times 10^{-6}$
Circadian rhythm—mammal	3	$6.86 \times 10^{-6}$
TGF-beta signaling pathway	4	$8.65 \times 10^{-6}$
Osteoclast differentiation	4	$4.40 \times 10^{-6}$
Toxoplasmosis	4	$4.75 \times 10^{-5}$
Insulin signaling pathway	4	$5.44 \times 10^{-5}$
mTOR signaling pathway	3	$7.46 \times 10^{-5}$
Endometrial cancer	3	$7.46 \times 10^{-5}$
Jak-STAT signaling pathway	4	$7.67 \times 10^{-5}$

<sup>a</sup>Adjusted *P*-value was calculated from hypergeometric test following by Benjamini-Hochberg multiple testing correction.

doi:10.1371/journal.pcbi.1004202.t003



**Fig 7. A novel metformin action pathway.** Solid lines indicate the proposed mechanisms as supported by experimental evidence from literature. The two black dashed lines indicate the drug effects. The red dashed line indicates the relationship is existed but the direction is unknown. The arrows beside the gene names or biological processes indicate the metformin effects. Up-arrows indicate the corresponding genes or processes are up-regulated while the down-arrows indicate the corresponding genes or process are down-regulated.

doi:10.1371/journal.pcbi.1004202.g007

In addition, in the crosstalk subnetwork, three nodes (CDKN1A, MYC, and SP1) formed a 3-node clique. The network small motif bridges the three modules together. The SP1 is a TF that binds to the GC-rich motif of numerous genes' promoters and is involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling. It has been reported that SP1 could cooperate with MYC to activate transcription of the human telomerase reverse transcriptase gene (*TERT*), which is responsible for maintenance of the length of telomeres and its defects may lead to diseases including cancer [77]. During the process of carcinogenesis, expression of *MYC* and *SP1* is known to be up-regulated [78]. It has been reported that metformin has an ability to down-regulate *MYC* [75,76] and *SP1* [61]. Additionally, *MYC* [79,80] and *SP1* [81,82] are also the key transcription factors involved in the regulation of insulin and insulin regulated gene transcription. *MYC* could directly induce both impaired insulin secretion and loss of  $\beta$ -cell mass

[83]. SP1 could regulate the upstream target STK11 expression [84,85]. MYC could activate AMPK in multiple cell lines [86]. AMPK activation could reduce SP1 translocate from cytoplasm to nucleus [87]. The *CDKN1A*, a cyclin-dependent kinase inhibitor p21, inhibits proliferation both in vitro and in vivo. After metformin treatment, the expression of *CDKN1A* is upregulated in hepatocellular carcinoma [88] and bladder cancer cells [89]. Additionally, multiple lines of evidence have demonstrated that MYC can suppress the expression of *CDKN1A* in cancer like colorectal cancer [90]. Therefore, taken all evidence together with the crosstalk network, we propose a new biological pathway for metformin action focused on four key nodes (CKDN1A, MYC, SP1, and STK11) (Fig 7). The pathway highlights several new questions, which may have been missed by previous studies. Specifically, we speculate that MYC and its networks are the key downstream targets of metformin. Further investigations are needed to illustrate this mechanism.

## Discussion

In this study, we developed a computational framework (DSPathNet) to construct a signaling pathway network for a given drug, specifically, metformin. The framework first collected metformin upstream genes from different data sources and inferred chemical signaling receptor TFs based on metformin-induced gene expression data. Then, a metformin-specific SPNetwork was produced using the random walk-based algorithms by applying longitudinal and lateral movements starting from metformin upstream genes and downstream TFs. By examining the enrichment of disease genes in the network, the metformin-specific SPNetwork proved to be enriched with genes that could contribute to the pathology of T2D and cancer, or reducing cancer risk in T2D patients undergoing metformin treatment. Starting from the genes common to T2D and cancer GWAS data, we further produced a crosstalk subnetwork of metformin action in T2D and cancer. Through comprehensive network and functional analyses and literature mining, we identified seven critical genes (*CDKN1A*, *ESR1*, *MAX*, *MYC*, *PPARGC1A*, *STK11*, and *SP1*), some of which have been implicated in previous studies. Furthermore, the MYC and its motifs were suggested to play important roles in metformin action. In summary, this study has the following major results: 1) we developed a computational framework for building drug-specific signaling pathway networks; 2) we generated a metformin-specific signaling pathway network that is significantly enriched with genes associated with T2D, cancer, or metformin-associated cancer survival, and 3) we pinpointed the MYC-centered pathway that may play important roles in metformin action. These results demonstrate that the computational framework effectively integrates various types of data, such as prior drug knowledge and drug-induced gene expression to identify critical genetic factors responsible for drug indications and drug response. This framework is a novel approach that provided a broader and deeper understanding of metformin actions in both T2D and cancer. This computational approach can be applied to other drugs as well.

This framework applies a new network generation strategy that focused on a drug of interest. In our framework, we utilized the gene expression data to infer the drug related gene expression regulators TFs, which is different from the methods that have been developed to infer signaling pathway networks directly from gene expression data [91]. As we know, the gene expression represents the transcriptional changes in the downstream genes of a pathway and provides an indirect view of pathway structure and gene activity after modulation of the system. Thus, the gene expression cannot directly represent the activity state of many signaling components that mediated the cellular response [92]. It is well known that the signal transduction network is not linear; rather it is quite complex [3]. During the development of this framework, we observed only two genes overlapped between metformin upstream genes and

downstream genes. This small overlap presents us with a big challenge: how to fill the gap to rebuild a complete cascade for drug action? To tackle this challenge, we proposed a novel strategy from background human SPNetwork through both longitudinal and lateral movements. For the longitudinal movement, we employed the software NetWalker that implemented the random walk with a starting probability. For the lateral movement, we took advantage of K-Walk algorithm that simulates random walks in the network using a Markov Chain to build the most relevant subnetwork. In this study, we combined them together to achieve our goal. [Table 2](#) summarizes the number of genes in each step and the hypergeometric tests based on the number of genes with smallest *P*-value less than 0.05 in the corresponding network compared to all genotyping data in T2D GWAS data. The evaluation results indicated that the process is promising since it recruited more informative genes; the significance of the association between the network and disease-related mutation signals became stronger.

However, the major concern regarding the framework is to rebuild a complete and reliable human SPNetwork and to control false positives from both public data and prediction results caused by the computational tools. To balance these two factors, we rigorously compiled the information involved in the signaling pathways, extensively collected the drug related data from multiple data sources, applied rigorous parameters during the use of computational approaches, and performed comprehensive evaluations for metformin-specific SPNetwork. To increase the accuracy of results, we only included the protein-protein pairs with experimental evidence and excluded the pairs only involved in the protein complexes. Thus, the coverage of the human SPNetwork was lower than a typical protein-protein interaction network; it contained only 37,881 edges and 4,367 proteins. With the rapid development of human experimental technologies, we believe more data with higher coverage and accuracy will become available, which will enable the construction of a more comprehensive signaling pathway network with high quality. To collect as many metformin-related genes as possible, in addition to the public databases DrugBank and PharmGBK, we further performed literature mining from PubMed abstracts, which provided an additional 19 genes. To ensure the accuracy of TF inference, we only utilized gene expression data from the four treatments of metformin that showed significant consistency with each other. To comprehensively evaluate if the metformin-specific SPNetwork was enriched with mutation signals of T2D and cancer, we not only took advantage of the well-studied disease genes but also individual genotyping data from GWAS data sets. Thus, our framework has the ability to recruit more key components in the drug signal transduction process. It could be potentially applied to other drugs for the purpose of deciphering their signaling pathway networks and identifying critical genes. Another limitation of this framework is the absence of a control network representing the normal state. The signaling network at the normal state may provide additional insights into drug action. However, it is very difficult and challenging to construct a normal-state signaling transduction network for drug action. Though some pathway data sources such as KEGG provide the relevant signaling networks in the normal state, most of them only provide a limited view by focusing on one or two related pathways. Compared to these individual pathway networks, the metformin-specific SPNetwork provides a comprehensive view by including many well-known metformin-related pathways, T2D-related pathways, and cancer-related pathways ([Results](#)).

This computational framework is strongly dependent on the available literature about the investigated drugs. Thus, it is not suitable for these drugs or chemicals that do not have many basic research reports. However, it is known that, during the drug development, most of them cannot be approved by FDA even after entering the clinical trials [93]. Furthermore, as the time and costs for developing novel drugs dramatically increased recently, many drug developers prefer to find new uses for existing drugs including the approved and non-approved drugs. As more large-scale data become publicly available, researchers could utilize the framework to

build a SPNetwork for each drug of interest, and then examine the relationship between the network and disease genes, or calculate network similarities with the known drugs for a certain indication. These relationship or network similarities may provide more clues for drug repurposing at the network level. Therefore, the framework will be promising for identification of drugs that may be used to treat secondary indications by constructing and comparing the drug-specific SPNetworks. Moreover, since the drug-specific SPNetwork contains comprehensive information regarding the drug action of the components, we speculated that some off-targets might be included in the network. Thus, our network approach can be extended to evaluate the association between drugs and their potential side effects. However, it is challenging to identify large-scale side effect data associated with genes or their proteins. So far, several studies have used the available biochemical data to determine candidate targets for specific side effects [94–96]. Such data is limited and likely with a high false positive rate. When more relevant data becomes available in future, our approach will be applied to assess drugs' side effects.

An important output of this study is the metformin-specific SPNetwork consisted of metformin related genes, metformin related TFs, and many novel genes. The network provides a valuable gene pool for further investigation of metformin action. Metformin has been used to treat diabetic disorders for many years because of its ability to lower glucose levels and improve insulin sensitivity [97]. Recently, several findings from epidemiological studies have shown that metformin can reduce cancer risk and improve cancer survival in the T2D patients [60,98,99], including a recent electronic health record (EHR) study we participated in that demonstrated the effect was seen for many cancer types [26]. However, the molecular mechanisms underlying metformin action are complex and remain unclear, especially for its ability of decreased cancer risk [100,101]. In this study, we first constructed a complex metformin-specific SPNetwork and then produced a crosstalk subnetwork from the metformin-specific SPNetwork. This subnetwork contained three modules highlighting different pathways (Fig 6C). The first and second modules were enriched with genes from the insulin signaling pathway and adipocytokine signaling pathway, and the third module was enriched with genes involved in cancer related pathways. The adipocytokine signaling pathway contains the major components of AMPK signaling pathway according to KEGG annotation. Through seven nodes, the first and second modules were linked to the third module. These observations suggest that the metformin possibly affects the AMPK signaling pathway and the insulin signaling pathway directly, which subsequently decrease the chance of cancer development. This outlook is consistent with a previous review [102].

The seven nodes act as bridges linking the first and second modules to the third module. We predicted they might play critical roles in the metformin signaling transduction process (Fig 6C). Among them, two genes (*PPARGC1A* and *STK11*) belonged to metformin upstream genes; one (*ESR1*) to metformin downstream genes; four genes (*CDKN1A*, *MAX*, *MYC*, and *SP1*) were both hubs and bridge nodes. It is well known that gene *STK11*, also known as *LKB1*, encodes a member of the serine/threonine kinase family that regulates cell polarity and functions as a tumor suppressor [103]. Additionally, previous studies have shown that mutations in the *STK11* gene influence insulin sensitivity and metformin efficacy [104,105]. The *MYC* gene encodes a protein that plays a role in cell cycle progression, apoptosis, and cellular transformation [106]. It has been shown that *MYC* gene plays important roles in the anticancer metabolic effects of metformin [75,76]. The *PPARGC1A* gene encodes a transcriptional coactivator that regulates the genes involved in energy metabolism. Its variant rs2970852 has been reported to modify the effects of metformin on triacylglycerol levels [107]. Recent studies have shown that gene regulation induced by metformin involves the transcription factor SP1 in cancers [61,108]. Moreover, the expression of *CDKN1A* (also known as *P21*) is upregulated in hepatocellular carcinoma [88] and bladder cancer cells [89] after metformin treatment. The evidence

from these studies suggests that our approach is effective for identifying the key components in the signaling pathway. To further investigate detailed information for these genes, more experimental validations are needed. To our knowledge, there is no any positive evidence for the association of the genes *ESR1* and *MAX* of the seven critical genes with metformin action. Thus, they are two novel genes for further experimental validation.

In addition to the DSPathNet framework to effectively recruit critical components in the mode of drug action, there are other ways to expand this approach. First, integrating multiple layers of data involving the signal cascade beyond gene expression data into a comprehensive method might improve our ability to identify the association between the genetic changes and their response to drugs. Second, although we have shown the utility of two sources for compiling the human SPNetwork, there are other data worth exploring such as those involved in the metabolism, protein phosphorylation, and protein kinase and phosphatase interactions. While this study focused on one medication, metformin, the computational framework is broadly applicable to any drug for which induced gene expression data is available. Moreover, several experimental data sources are available for further data integration and mining such as the Connectivity Map project [15], Genomics of Drug Sensitivity in Cancer [109], Cancer Cell Line Encyclopedia (CCLE) [110], and anticancer compounds in breast cancer [111]. Finally, analyzing the crosstalk among different types of diseases in the context of networks will offer an intriguing opportunity to explore the underlying molecular mechanisms of drug action, which will provide an alternative approach for drug repurposing.

## Materials and Methods

### Compilation of one human SPNetwork with weighted nodes

Before generating the metformin-specific SPNetwork, we need a global signal transduction network for humans as the background network. We therefore integrated signaling transduction related associations from Pathway Commons with experimental evidence [25], and TF-TF/target pairs from TRANSFAC [26]. The Pathway Commons database collected publicly available pathways from multiple organisms with over 1,400 pathways and 687,000 interactions. We first downloaded the edge data specific for humans from the Pathway Commons (release 2011.10). Since the interactions that occur within the protein complexes do not reveal the flow of signaling information [3], we excluded the edges that came from the same complex. This process resulted in 33,614 pairs among 3,502 proteins. Additionally, we obtained 1,325 pairs among 487 TFs, and 2,723 pairs between 428 TFs and 1,315 targets downloaded from TRANSFAC database (release 2011.4). The TRANSFAC database manually collects eukaryotic TFs, their genomic binding sites, and DNA binding profiles with experimental evidence [112]. After merging the two data sets and removing the redundancies, we obtained a network with 37,881 edges and 4,367 nodes. This network was used to represent global signaling pathways in humans.

To further weight the association of each node in human SPNetwork with metformin action, we assigned a functional similarity score by calculating its functional similarity to the metformin upstream genes using the R package GoSemSim based on GO annotations [113]. GO annotations have three functional domains ( $k$ ): molecular function (MF), biological process (BP), and cellular component (CC). First, for a given node  $i$  in each domain ( $k$ ), we calculated its score as  $Score_i = \sum_{j=1}^n Score_{i,j}/n$ , where  $n$  is the number of existing scores between node  $i$  and metformin upstream gene  $j$ . Second, for the given node  $i$  in all domains, we calculated a final score as  $\hat{S} = \sum_{k=1}^N Score_k/N$ , where  $N$  is the number of the domains having scores for the node.



## Inference of metformin downstream genes

Gene expression profiles of cancer cells following drug treatment are useful for better understanding cellular changes reflective of drug treatment [114]. In this study, we integrated the known TF-target association and drug-induced gene expression data to infer the metformin downstreams. We first comprehensively collected the TF-target associations, then calculated the up- or down-regulated genes from drug-induced gene expression data, and finally performed the hypergeometric test to evaluate the over-representation of the up- or down-regulated genes in multiple TF target gene datasets.

To compile a target gene set for each TF comprehensively, we downloaded data from two sources: TRANSFAC Professional (release 2011.4) and MSigDB database [38]. From the TRANSFAC database, we extracted known TFs and their targets in human. From the MSigDB, we downloaded the gene sets that share one TF binding site. The gene sets were derived from a comparative analysis of human, mouse, rat, and dog genomes and were organized by TF binding motifs. Genes associated with different binding motifs that correspond to a common transcription factor were combined into one gene set. After merging the two data sets, we obtained 666 human TFs and 8,502 human targets.

To calculate the potential differentially expressed genes induced by metformin, we downloaded ten gene expression datasets from Connectivity Map website (version 2.0). The gene expression datasets were generated from metformin treated cell lines. We calculated the ranked probes by using the method described in Lamb et al. [15] and selected the top 100 and bottom 100 probes in each treatment to represent the differentially expressed probes [115]. We examined the expression consistency among them using the software GSEA. We noticed that, among ten metformin treatment data sets, four had the highest consistency based on GSEA analysis [38]. Therefore, we utilized these four treatment gene expression data to perform the GSEA leading edge analysis to detect differentially expressed probes. Then, by mapping the differently expressed probes to genes using Ingenuity Pathway Analysis Tool (<http://www.ingenuity.com/>), we obtained the up-regulated genes and down-regulated genes.

Finally, we performed the hypergeometric test to evaluate the over-representation of the up- or down-regulated genes in the different TF gene sets. The TFs with *P*-value less than 0.05 were identified as significant TFs related to metformin action and their genes as metformin downstream genes.

## Construction of metformin-specific SPNetwork

Considering that the signal transduction cascade is not linear, we adopted a two-step strategy to construct the metformin-specific SPNetwork from the human SPNetwork. More specifically, in the first step, we utilized the software NetWalker to expand metformin upstream genes and downstream genes for longitudinal conduction [116]. The NetWalker implements the random walk with a starting probability. In this study, we gave equal starting probability of 0.5 to each gene in the metformin upstream genes and downstream genes and required those nodes with both local *P*-value < 0.05 and global *P*-value < 0.05 as the expanded genes. In the second step, we expanded the nodes from in the first step by lateral movement by applying the K-Walk method implemented in the Python package GenRev [117]. The K-Walk algorithm simulates random walks in the network using a Markov Chain to build the most relevant subnetwork, connecting seed nodes by walk a fixed length *L* or up to a maximal length *L*<sub>max</sub> in a large network. A subnetwork is obtained by keeping only edges that are above a minimal relevance threshold. The threshold is automatically fixed after the subnetwork has the maximum score. As such, the limited K-Walk algorithm computes edge and node relevance from random walks connecting the seed nodes [118].

## GWAS data sets

We used one T2D GWAS data set, three cancer GWAS data sets, and one GWAS data set for T2D patients with metformin treatment. The T2D GWAS data was individual-level genotype data generated from the WTCCC [31]. The three cancer GWAS datasets were generated by the Cancer Genetic Markers of Susceptibility (CGEMS) project: breast cancer [32], pancreatic cancer [33], and prostate cancer [32]. We downloaded the genotype data from the National Center for Biotechnology Information (NCBI) dbGaP with approved access for the CGEMS project. For these four GWAS datasets, we first removed individuals with genotyping rate < 95% and SNPs with missing rate >5%. A single SNP associated test was conducted using the Armitage trend test for SNPs with a minor allele frequency (MAF) > 0.05. [S10 Table](#) summarizes the data.

T2D cancer patients from Vanderbilt University Medical Center (VUMC) were identified using the Synthetic Derivative (SD), a de-identified copy of the electronic health records from VUMC. Eligible subjects were individuals who 1) had a cancer diagnosis (excluding non-melanoma skin cancers) between January 1, 1995 and December 31, 2010 identified through the Vanderbilt tumor registry, and 2) were older than 18 years at the time of cancer diagnosis. Using a previously developed algorithm [119,120], we identified T2D subjects having at least two pieces of clinical information in their medical record: 1) ICD9 code for type 2 diabetes, 2) medications for type 2 diabetes, or 3) clinical labs suggestive of T2D (random glucose >200 mg/dl or hemoglobin A1c > 6.5%). Individuals without at least two of the above types of information were excluded. At least two mentions of metformin use (mono-therapy or combined therapeutic) and one mention of metformin use within 5 years after cancer diagnosis were required for study inclusion. Individuals on other T2D medications were excluded from analysis. Subjects were followed for overall mortality that was determined through linkage with the Vanderbilt tumor registry. Physician-reported European descent individuals with an available DNA sample in the Vanderbilt biobank (BioVU) [121] were genotyped on either the Illumina HumanOmni1-Quad or the Illumina HumanOmni5-Quad. Only the consensus single nucleotide polymorphisms (SNPs) between the two genotyping platforms were used. Standard quality control (QC) procedures were applied to remove individuals and autosomal SNPs not meeting standard QC criteria (i.e. related individuals, discordant sex, sample efficiency < 98%, genotyping efficiency < 98%, deviations from Hardy-Weinberg equilibrium ( $p < 1 \times 10^{-6}$ ), and MAF < 5%). Palindromic SNPs were also removed. After QC, 461 individuals and 551,745 SNPs remained. Principal components were estimated using EIGENSTRAT [122]. The association between each SNP, assuming an additive genetic model, and overall survival was examined using Cox proportional hazards models, adjusted for age, sex and one principal component, using the GenABEL package of R [123]. The GWAS analysis of this set is ongoing and will be reported in a separate publication.

In this study, we defined the genes having at least one SNP with nominal *P*-value less than 0.05 as disease or drug related genes. The SNP is located in the gene's region or its 20kb up- or down-stream sequence based on the gene annotation and human reference genome build 36 for T2D GWAS study and cancer GWAS studies and build 37 for metformin GWAS study.

## Pathway enrichment, network analysis and visualization

To identify pathways overrepresented in gene sets, we performed KEGG pathway enrichment analyses using WebGestalt [49] (version 1/30/2013). Given a list of genes, a hypergeometric test was performed for the enrichment of these genes, which was implemented in the WebGestalt tool. To control the error rate in the analysis results, WebGestalt also provides a corrected *P*-value based on the Benjamini-Hochberg method [124]. To summarize the enriched

pathways, we took advantage of KEGG pathway category annotation, which included the two-level categories and represent the relative abundance of the pathways [125]. These pathways are grouped into seven categories at the first level of KEGG annotation and 43 categories at the second level of KEGG annotation. At the second-level category, we further calculated a Z-score for each category to represent the KEGG pathway relative abundance:  $Z\text{-score} = \frac{x-u}{\sigma}$ , where  $x$  is the number of pathways in one category in the first or second level,  $u$  is the mean of the pathway number in the first or second category,  $\sigma$  is the standard deviation of the pathway number in the first or second category. The pathway categories were selected for further analysis if their Z-scores were higher than zero.

In this study, we adopted the statistical design for gene set enrichment analysis [126] to compare a gene set (A) in the drug-specific network to a reference gene set (B). The design has been commonly used to conduct the gene annotation enrichment analysis [127]. Suppose that the gene set (A) has  $n$  genes, of which most genes ( $n'$ ) belong to the reference gene set ( $m$ ). Among  $n'$  gene,  $k$  genes belong to a given category (C). And the reference gene set has  $j$  genes belong to the same category (C). Based on the definition of the hypergeometric test, we performed the hypergeometric test to get a  $P$ -value to evaluate the significance of enrichment for category C in the gene set A.

For network property analysis, we calculated degree of each node and degree distribution of all nodes, which are the most basic measures of biological networks [41]. The node degree (connectivity) is the number of links of a node in the network. If degree distribution of one network follows a power law, the network would have only a small portion of nodes with a large number of links (i.e., hubs) [41]. To determine the hubs in metformin-specific SPNetwork, we adopted the method utilized by Yu et al. [46], as we did in a previous study. We first drew a degree distribution for the whole network to define a specific degree value as a cut-off point (S12 Fig). If a node has the degree greater than the cut-off value, then the node is a hub. To identify the modules, we performed the cluster and community analysis using the software CFinder (version 2.0.5) [63]. CFinder is a fast program to locate and visualize overlapping, densely interconnected groups of nodes in undirected network. We required each node in the module being involved in at least one 3-vertex clique. We visualized the networks using Cytoscape (version 3.2) [128].

## Supporting Information

**S1 Fig. The three-way Venn diagram summarizes the number of shared genes among the three gene sets.** The “Human SPNetwork node” represents the genes corresponding to nodes in the human SPNetwork, ‘Gene\_46’ represents the metformin-related genes obtained from DrugBank and PharmGKB, and ‘Gene\_29’ represents the metformin-related gene obtained by literature searching approach.  
(TIFF)

**S2 Fig. GSEA (Gene Set Enrichment Analysis) enrichment score curve for six probe sets of six treatments (Instance IDs: 61, 1694, 1816, 1858, 5068, and 5487) compared to the probes from one treatment (Instance ID: 1).** In each graph, the vertical black lines indicate the position of each of the probes of the studied probe set in the ordered, non-redundant data set. The green curve corresponds to the ES (enrichment score) curve, which is the running sum of the weighted enrichment score in GSEA.  
(PDF)

**S3 Fig. Network of metformin upstream gene and downstream genes.** This network was generated by mapping them129 unique genes of metformin upstream genes and TF genes to

human SPNetwork. The nodes and edges in orange correspond to nodes and edges only in the metformin upstream network. The nodes and edges in green correspond to the nodes and edges only in the metformin downstream network. And the nodes and edges in red correspond to the nodes and edges common to the metformin upstream network and the metformin downstream network.

(PNG)

**S4 Fig. Summary of genes by longitudinal and lateral movements from metformin upstream genes and downstream genes via three-way Venn diagrams.** A) Summary of the number of shared genes among metformin upstream genes represented by 'Upstream gene', the genes obtained by longitudinal movement represented by 'Longitudinal gene' based on 'Upstream gene', and the genes obtained by lateral movement based on 'Longitudinal gene'. B) Summary of the number of shared genes among metformin downstream genes represented by 'Downstream gene', the genes obtained by longitudinal movement represented by 'Longitudinal gene' based on 'Downstream gene', and the genes obtained by lateral movement based on 'Longitudinal gene'.

(PNG)

**S5 Fig. Network of extended genes of metformin upstream genes and downstream genes by longitudinal movement.** The network was generated by mapping the unique 219 genes of extended genes of metformin upstream gene and downstream genes by longitudinal moving to the human SP Network. The legends for orange nodes, red nodes, and green nodes are same as in [S3 Fig](#).

(PNG)

**S6 Fig. P-value distribution of metformin GWAS data of the metformin-specific SPNetwork, human SPNetwork, and metformin GWAS.** The details of the data were provided in Materials and Methods section.

(PNG)

**S7 Fig. The subnetwork for 81 genes.** The genes were common to the 169 genes whose smallest *P*-values were less than 0.05 in T2D GWAS data and the 177 genes had at least one SNP with *P*-value less than 0.05 in metformin GWAS data. The legends for orange nodes, red nodes, and green nodes are same as in [S3 Fig](#).

(PNG)

**S8 Fig. The subnetwork for 25 genes.** These genes were common among the 169 genes whose smallest *P*-values were less than 0.05 in T2D GWAS data, 157 genes whose smallest *P*-values were less than 0.05 in breast cancer WAS data, 170 genes whose smallest *P*-values were less than 0.05 in pancreatic cancer GWAS data, 172 genes whose smallest *P*-values were less than 0.05 in prostate cancer GWAS data. The legends for orange nodes and edges, red nodes and edges, and green nodes and edges are same as in [S3 Fig](#).

(PNG)

**S9 Fig. The subnetwork for 25 common genes and their direct interactors.** The 25 common genes that were among the T2D GWA study and the three cancer GWA studies. The legends for orange nodes and edges, red nodes and edges, and green nodes and edges are same as in [S3 Fig](#).

(PNG)

**S10 Fig. The networks for three modules.** The legends for orange nodes and edges, red nodes and edges, and green nodes and edges are same as in [S3 Fig](#).

(PNG)

**S11 Fig.** Seven highlighted nodes in yellow in the subnetwork for 25 common genes and their direct interactors (A) and three 3-clique communities after removing the highlighted nodes (B). The legends for orange nodes and edges, red nodes and edges, and green nodes and edges are same as in [S3 Fig.](#)

(TIFF)

**S12 Fig. Degree distribution of the 477 nodes in metformin-specific SPNetwork.** This distribution is used for determination of hubs.

(TIFF)

**S1 Table. Summary of data sources, software, and evaluation data used in the study.**

(DOCX)

**S2 Table. Metformin upstream genes and their sources.**

(DOCX)

**S3 Table. List of metformin treatments from Connectivity Map database.**

(DOCX)

**S4 Table. Metformin downstream genes encoding transcription factors inferred from metformin-induced gene expression data from Connectivity Map.**

(DOCX)

**S5 Table. Pairs of metformin-specific signaling pathway network (SPNetwork).**

(XLSX)

**S6 Table. List of genes in the metformin-specific SPNetwork.**

(XLSX)

**S7 Table. KEGG pathways overrepresented in 477 genes in metformin-specific SPNetwork.**

(XLSX)

**S8 Table. KEGG pathways overrepresented in upstream genes (174) only belonging to metformin upstream network, downstream genes (262) only belonging to metformin downstream network, and genes (41) common to metformin upstream network and downstream network.**

(XLSX)

**S9 Table. First-level and secondary level categories of the KEGG pathway overrepresented in upstream genes (174) only belonging to metformin upstream network, downstream genes (262) only belonging to metformin downstream network, and genes (41) common to metformin upstream network and downstream network.**

(XLSX)

**S10 Table. Summary of three cancer GWAS data.**

(DOCX)

## Acknowledgments

We thank Drs. Bing Zhang, Qi Liu, Jing Zhu, and Jing Wang for valuable discussion and Lana Olson for performing quality control of the genotyping data. We thank Dr. Anupama E Gururaj for critically reading and improving an earlier draft of the manuscript.

## Author Contributions

Conceived and designed the experiments: JS HX ZZ. Performed the experiments: JS. Analyzed the data: JS MZ. Contributed reagents/materials/analysis tools: JS PJ LW YW CI EB DMR JCD MCA ZZ. Wrote the paper: JS YZ MCA HX ZZ. Edited and revised the manuscript: JS MZ PJ LW YW CI YZ EB DMR JCD MCA HX ZZ.

## References

- Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1: 727–730. PMID: [12209152](#)
- Rask-Andersen M, Almen MS, Schiøth HB (2011) Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* 10: 579–590. doi: [10.1038/nrd3478](#) PMID: [21804595](#)
- Kholodenko B, Yaffe MB, Kolch W (2012) Computational approaches for analyzing information flow in biological networks. *Sci Signal* 5: re1. doi: [10.1126/scisignal.2002961](#) PMID: [22510471](#)
- Persidis A (1998) Signal transduction as a drug-discovery platform. *Nat Biotechnol* 16: 1082–1083. PMID: [9831041](#)
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 10: 252–263. doi: [10.1038/nrg2538](#) PMID: [19274049](#)
- Shaw RJ, Cantley LC (2006) Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* 441: 424–430. PMID: [16724053](#)
- Pouyssegur J, Dayan F, Mazure NM (2006) Hypoxia signalling in cancer and approaches to enforce tumour regression. *Nature* 441: 437–443. PMID: [16724055](#)
- Karam CS, Ballon JS, Bivens NM, Freyberg Z, Girgis RR, et al. (2010) Signaling pathways in schizophrenia: emerging targets and therapeutic strategies. *Trends Pharmacol Sci* 31: 381–390. doi: [10.1016/j.tips.2010.05.004](#) PMID: [20579747](#)
- Jin T, Liu L (2008) The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus. *Mol Endocrinol* 22: 2383–2392. doi: [10.1210/me.2008-0135](#) PMID: [18599616](#)
- Bianco R, Melisi D, Ciardiello F, Tortora G (2006) Key cancer cell signal transduction pathways as therapeutic targets. *Eur J Cancer* 42: 290–294. PMID: [16376541](#)
- Freyberg Z, Ferrando SJ, Javitch JA (2010) Roles of the Akt/GSK-3 and Wnt signaling pathways in schizophrenia and antipsychotic drug action. *Am J Psychiatry* 167: 388–396. doi: [10.1176/appi.ajp.2009.08121873](#) PMID: [19917593](#)
- Akhurst RJ, Hata A (2012) Targeting the TGFbeta signalling pathway in disease. *Nat Rev Drug Discov* 11: 790–811. doi: [10.1038/nrd3810](#) PMID: [23000686](#)
- Sebolt-Leopold JS, English JM (2006) Mechanisms of drug inhibition of signalling molecules. *Nature* 441: 457–462. PMID: [16724058](#)
- Jin G, Fu C, Zhao H, Cui K, Chang J, et al. (2012) A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. *Cancer Res* 72: 33–44. doi: [10.1158/0008-5472.CAN-11-2333](#) PMID: [22108825](#)
- Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313: 1929–1935. PMID: [17008526](#)
- Knox C, Law V, Jewison T, Liu P, Ly S, et al. (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* 39: D1035–1041. doi: [10.1093/nar/gkq1126](#) PMID: [21059682](#)
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109–114. doi: [10.1093/nar/gkr988](#) PMID: [22080510](#)
- Sangkuhl K, Berlin DS, Altman RB, Klein TE (2008) PharmGKB: understanding the effects of individual genetic variants. *Drug Metab Rev* 40: 539–551. doi: [10.1080/03602530802413338](#) PMID: [18949600](#)
- Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, et al. (2014) STITCH 4: integration of protein-chemical interactions with user data. *Nucleic Acids Res* 42: D401–407. doi: [10.1093/nar/gkt1207](#) PMID: [24293645](#)
- Arrell DK, Terzic A (2010) Network systems biology for drug discovery. *Clin Pharmacol Ther* 88: 120–125. doi: [10.1038/clpt.2010.91](#) PMID: [20520604](#)

21. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 4: 682–690. doi: [10.1038/nchembio.118](https://doi.org/10.1038/nchembio.118) PMID: [18936753](https://pubmed.ncbi.nlm.nih.gov/18936753/)
22. Vidal M, Cusick ME, Barabasi AL (2011) Interactome networks and human disease. *Cell* 144: 986–998. doi: [10.1016/j.cell.2011.02.016](https://doi.org/10.1016/j.cell.2011.02.016) PMID: [21414488](https://pubmed.ncbi.nlm.nih.gov/21414488/)
23. Leung EL, Cao ZW, Jiang ZH, Zhou H, Liu L (2013) Network-based drug discovery by integrating systems biology and computational technologies. *Brief Bioinform* 14: 491–505. doi: [10.1093/bib/bbs043](https://doi.org/10.1093/bib/bbs043) PMID: [22877768](https://pubmed.ncbi.nlm.nih.gov/22877768/)
24. Ben Sahra I, Le Marchand-Brustel Y, Tanti JF, Bost F (2010) Metformin in cancer therapy: a new perspective for an old antidiabetic drug? *Mol Cancer Ther* 9: 1092–1099. doi: [10.1158/1535-7163.MCT-09-1186](https://doi.org/10.1158/1535-7163.MCT-09-1186) PMID: [20442309](https://pubmed.ncbi.nlm.nih.gov/20442309/)
25. Pierotti MA, Berrino F, Gariboldi M, Melani C, Mogavero A, et al. (2013) Targeting metabolism for cancer treatment and prevention: metformin, an old drug with multi-faceted effects. *Oncogene* 32: 1475–1487. doi: [10.1038/onc.2012.181](https://doi.org/10.1038/onc.2012.181) PMID: [22665053](https://pubmed.ncbi.nlm.nih.gov/22665053/)
26. Xu H, Aldrich MC, Chen Q, Liu H, Peterson NB, et al. (2014) Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *J Am Med Inform Assoc* (In press).
27. Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12: 56–68. doi: [10.1038/nrg2918](https://doi.org/10.1038/nrg2918) PMID: [21164525](https://pubmed.ncbi.nlm.nih.gov/21164525/)
28. Navlakha S, Kingsford C (2010) The power of protein interaction networks for associating genes with diseases. *Bioinformatics* 26: 1057–1063. doi: [10.1093/bioinformatics/btq076](https://doi.org/10.1093/bioinformatics/btq076) PMID: [20185403](https://pubmed.ncbi.nlm.nih.gov/20185403/)
29. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106: 9362–9367. doi: [10.1073/pnas.0903103106](https://doi.org/10.1073/pnas.0903103106) PMID: [19474294](https://pubmed.ncbi.nlm.nih.gov/19474294/)
30. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4: 177–183. PMID: [14993899](https://pubmed.ncbi.nlm.nih.gov/14993899/)
31. Wellcome Trust Case Control C (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661–678. PMID: [17554300](https://pubmed.ncbi.nlm.nih.gov/17554300/)
32. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, et al. (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39: 870–874. PMID: [17529973](https://pubmed.ncbi.nlm.nih.gov/17529973/)
33. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, et al. (2009) Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet* 41: 986–990. doi: [10.1038/ng.429](https://doi.org/10.1038/ng.429) PMID: [19648918](https://pubmed.ncbi.nlm.nih.gov/19648918/)
34. Bowton E, Field JR, Wang S, Schildcrout JS, Van Driest SL, et al. (2014) Biobanks and electronic medical records: enabling cost-effective research. *Sci Transl Med* 6: 234cm233. doi: [10.1126/scitranslmed.3008604](https://doi.org/10.1126/scitranslmed.3008604) PMID: [24786321](https://pubmed.ncbi.nlm.nih.gov/24786321/)
35. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, et al. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 39: D685–690. doi: [10.1093/nar/gkq1039](https://doi.org/10.1093/nar/gkq1039) PMID: [21071392](https://pubmed.ncbi.nlm.nih.gov/21071392/)
36. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34: D108–110. PMID: [16381825](https://pubmed.ncbi.nlm.nih.gov/16381825/)
37. Wu Y, Liu M, Zheng WJ, Zhao Z, Xu H (2012) Ranking gene-drug relationships in biomedical literature using Latent Dirichlet Allocation. *Pac Symp Biocomput*: 422–433. PMID: [22174297](https://pubmed.ncbi.nlm.nih.gov/22174297/)
38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550. PMID: [16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
39. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, et al. (2005) Mining for regulatory programs in the cancer transcriptome. *Nat Genet* 37: 579–583. PMID: [15920519](https://pubmed.ncbi.nlm.nih.gov/15920519/)
40. Liu Y, Ringner M (2007) Revealing signaling pathway deregulation by using gene expression signatures and regulatory motif analysis. *Genome Biol* 8: R77. PMID: [17498287](https://pubmed.ncbi.nlm.nih.gov/17498287/)
41. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5: 101–113. PMID: [14735121](https://pubmed.ncbi.nlm.nih.gov/14735121/)
42. Zotenko E, Mestre J, O'Leary DP, Przytycka TM (2008) Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput Biol* 4: e1000140. doi: [10.1371/journal.pcbi.1000140](https://doi.org/10.1371/journal.pcbi.1000140) PMID: [18670624](https://pubmed.ncbi.nlm.nih.gov/18670624/)

43. Ivanic J, Yu X, Wallqvist A, Reifman J (2009) Influence of protein abundance on high-throughput protein-protein interaction detection. *PLoS One* 4: e5815. doi: [10.1371/journal.pone.0005815](https://doi.org/10.1371/journal.pone.0005815) PMID: [19503833](https://pubmed.ncbi.nlm.nih.gov/19503833/)
44. Wachi S, Yoneda K, Wu R (2005) Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* 21: 4205–4208. PMID: [16188928](https://pubmed.ncbi.nlm.nih.gov/16188928/)
45. Sun J, Zhao Z (2010) A comparative study of cancer proteins in the human protein-protein interaction network. *BMC Genomics* 11 Suppl 3: S5. doi: [10.1186/1471-2164-11-S3-S5](https://doi.org/10.1186/1471-2164-11-S3-S5) PMID: [21143787](https://pubmed.ncbi.nlm.nih.gov/21143787/)
46. Yu H, Greenbaum D, Xin Lu H, Zhu X, Gerstein M (2004) Genomic analysis of essentiality within protein networks. *Trends Genet* 20: 227–231. PMID: [15145574](https://pubmed.ncbi.nlm.nih.gov/15145574/)
47. Dhillon AS, Hagan S, Rath O, Kolch W (2007) MAP kinase signalling pathways in cancer. *Oncogene* 26: 3279–3290. PMID: [17496922](https://pubmed.ncbi.nlm.nih.gov/17496922/)
48. Gehart H, Kumpf S, Ittner A, Ricci R (2010) MAPK signalling in cellular metabolism: stress or wellness? *EMBO Rep* 11: 834–840. doi: [10.1038/embor.2010.160](https://doi.org/10.1038/embor.2010.160) PMID: [20930846](https://pubmed.ncbi.nlm.nih.gov/20930846/)
49. Wang J, Duncan D, Shi Z, Zhang B (2013) WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 41: W77–83. doi: [10.1093/nar/gkt439](https://doi.org/10.1093/nar/gkt439) PMID: [23703215](https://pubmed.ncbi.nlm.nih.gov/23703215/)
50. Shikata K, Ninomiya T, Kiyohara Y (2013) Diabetes mellitus and cancer risk: review of the epidemiological evidence. *Cancer Sci* 104: 9–14. doi: [10.1111/cas.12043](https://doi.org/10.1111/cas.12043) PMID: [23066889](https://pubmed.ncbi.nlm.nih.gov/23066889/)
51. Nesbit CE, Tersak JM, Prochownik EV (1999) MYC oncogenes and human neoplastic disease. *Oncogene* 18: 3004–3016. PMID: [10378696](https://pubmed.ncbi.nlm.nih.gov/10378696/)
52. Sakuma K, Aoki M, Kannagi R (2012) Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF/bFGF-induced epithelial-mesenchymal transition. *Proc Natl Acad Sci USA* 109: 7776–7781. doi: [10.1073/pnas.1111135109](https://doi.org/10.1073/pnas.1111135109) PMID: [22547830](https://pubmed.ncbi.nlm.nih.gov/22547830/)
53. Grinberg AV, Hu CD, Kerppola TK (2004) Visualization of Myc/Max/Mad family dimers and the competition for dimerization in living cells. *Mol Cell Biol* 24: 4294–4308. PMID: [15121849](https://pubmed.ncbi.nlm.nih.gov/15121849/)
54. Chavez L, Bais AS, Vingron M, Lehrach H, Adjaye J, et al. (2009) In silico identification of a core regulatory network of OCT4 in human embryonic stem cells using an integrated approach. *BMC Genomics* 10: 314. doi: [10.1186/1471-2164-10-314](https://doi.org/10.1186/1471-2164-10-314) PMID: [19604364](https://pubmed.ncbi.nlm.nih.gov/19604364/)
55. Kyo S, Takakura M, Taira T, Kanaya T, Itoh H, et al. (2000) Sp1 cooperates with c-Myc to activate transcription of the human telomerase reverse transcriptase gene (hTERT). *Nucleic Acids Res* 28: 669–677. PMID: [10637317](https://pubmed.ncbi.nlm.nih.gov/10637317/)
56. Cheng YW, Wu TC, Chen CY, Chou MC, Ko JL, et al. (2008) Human telomerase reverse transcriptase activated by E6 oncoprotein is required for human papillomavirus-16/18-infected lung tumorigenesis. *Clin Cancer Res* 14: 7173–7179. doi: [10.1158/1078-0432.CCR-08-0850](https://doi.org/10.1158/1078-0432.CCR-08-0850) PMID: [19010833](https://pubmed.ncbi.nlm.nih.gov/19010833/)
57. Parisi F, Wirapati P, Naef F (2007) Identifying synergistic regulation involving c-Myc and sp1 in human tissues. *Nucleic Acids Res* 35: 1098–1107. PMID: [17264126](https://pubmed.ncbi.nlm.nih.gov/17264126/)
58. Gartel AL, Ye X, Goufman E, Shianov P, Hay N, et al. (2001) Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. *Proc Natl Acad Sci USA* 98: 4510–4515. PMID: [11274368](https://pubmed.ncbi.nlm.nih.gov/11274368/)
59. Schwenk RW, Vogel H, Schurmann A (2013) Genetic and epigenetic control of metabolic health. *Mol Metab* 2: 337–347. doi: [10.1016/j.molmet.2013.09.002](https://doi.org/10.1016/j.molmet.2013.09.002) PMID: [24327950](https://pubmed.ncbi.nlm.nih.gov/24327950/)
60. Landman GW, Kleefstra N, van Hateren KJ, Groenier KH, Gans RO, et al. (2010) Metformin associated with lower cancer mortality in type 2 diabetes: ZODIAC-16. *Diabetes Care* 33: 322–326. doi: [10.2337/dc09-1380](https://doi.org/10.2337/dc09-1380) PMID: [19918015](https://pubmed.ncbi.nlm.nih.gov/19918015/)
61. Nair V, Pathi S, Jutooru I, Sreevalsan S, Basha R, et al. (2013) Metformin inhibits pancreatic cancer cell and tumor growth and downregulates Sp transcription factors. *Carcinogenesis* 34: 2870–2879. doi: [10.1093/carcin/bgt231](https://doi.org/10.1093/carcin/bgt231) PMID: [23803693](https://pubmed.ncbi.nlm.nih.gov/23803693/)
62. Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR (2010) Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin Transl Sci* 3: 42–48. doi: [10.1111/j.1752-8062.2010.00175.x](https://doi.org/10.1111/j.1752-8062.2010.00175.x) PMID: [20443953](https://pubmed.ncbi.nlm.nih.gov/20443953/)
63. Adamcsek B, Palla G, Farkas IJ, Derenyi I, Vicsek T (2006) CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22: 1021–1023. PMID: [16473872](https://pubmed.ncbi.nlm.nih.gov/16473872/)
64. Esteve E, Ricart W, Fernandez-Real JM (2009) Adipocytokines and insulin resistance: the possible role of lipocalin-2, retinol binding protein-4, and adiponectin. *Diabetes Care* 32 Suppl 2: S362–367. doi: [10.2337/dc09-S340](https://doi.org/10.2337/dc09-S340) PMID: [19875582](https://pubmed.ncbi.nlm.nih.gov/19875582/)
65. Stolar MW (2002) Insulin resistance, diabetes, and the adipocyte. *Am J Health Syst Pharm* 59 Suppl 9: S3–8. PMID: [12489380](https://pubmed.ncbi.nlm.nih.gov/12489380/)
66. White MF (2003) Insulin signaling in health and disease. *Science* 302: 1710–1711. PMID: [14657487](https://pubmed.ncbi.nlm.nih.gov/14657487/)



67. Chiu SL, Cline HT (2010) Insulin receptor signaling in the development of neuronal structure and function. *Neural Dev* 5: 7. doi: [10.1186/1749-8104-5-7](https://doi.org/10.1186/1749-8104-5-7) PMID: [20230616](https://pubmed.ncbi.nlm.nih.gov/20230616/)
68. Owen MR, Doran E, Halestrap AP (2000) Evidence that metformin exerts its anti-diabetic effects through inhibition of complex 1 of the mitochondrial respiratory chain. *Biochem J* 348 Pt 3: 607–614. PMID: [10839993](https://pubmed.ncbi.nlm.nih.gov/10839993/)
69. Wheaton WW, Weinberg SE, Hamanaka RB, Soberanes S, Sullivan LB, et al. (2014) Metformin inhibits mitochondrial complex I of cancer cells to reduce tumorigenesis. *Elife* 3: e02242. doi: [10.7554/eLife.02242](https://doi.org/10.7554/eLife.02242) PMID: [24843020](https://pubmed.ncbi.nlm.nih.gov/24843020/)
70. Hardie DG, Ross FA, Hawley SA (2012) AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat Rev Mol Cell Biol* 13: 251–262. doi: [10.1038/nrm3311](https://doi.org/10.1038/nrm3311) PMID: [22436748](https://pubmed.ncbi.nlm.nih.gov/22436748/)
71. Hardie DG (2014) AMP-activated protein kinase: a key regulator of energy balance with many roles in human disease. *J Intern Med* (In press).
72. Shackelford DB, Shaw RJ (2009) The LKB1-AMPK pathway: metabolism and growth control in tumour suppression. *Nat Rev Cancer* 9: 563–575. doi: [10.1038/nrc2676](https://doi.org/10.1038/nrc2676) PMID: [19629071](https://pubmed.ncbi.nlm.nih.gov/19629071/)
73. Shaw RJ, Lamia KA, Vasquez D, Koo SH, Bardeesy N, et al. (2005) The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin. *Science* 310: 1642–1646. PMID: [16308421](https://pubmed.ncbi.nlm.nih.gov/16308421/)
74. Liang X, Nan KJ, Li ZL, Xu QZ (2009) Overexpression of the LKB1 gene inhibits lung carcinoma cell proliferation partly through degradation of c-myc protein. *Oncol Rep* 21: 925–931. PMID: [19287990](https://pubmed.ncbi.nlm.nih.gov/19287990/)
75. Blandino G, Valerio M, Cioce M, Mori F, Casadei L, et al. (2012) Metformin elicits anticancer effects through the sequential modulation of DICER and c-MYC. *Nat Commun* 3: 865. doi: [10.1038/ncomms1859](https://doi.org/10.1038/ncomms1859) PMID: [22643892](https://pubmed.ncbi.nlm.nih.gov/22643892/)
76. Akinyeke T, Matsumura S, Wang X, Wu Y, Schaffer ED, et al. (2013) Metformin targets c-MYC oncogene to prevent prostate cancer. *Carcinogenesis* 34: 2823–2832. doi: [10.1093/carcin/bgt307](https://doi.org/10.1093/carcin/bgt307) PMID: [24130167](https://pubmed.ncbi.nlm.nih.gov/24130167/)
77. Artandi SE, DePinho RA (2010) Telomeres and telomerase in cancer. *Carcinogenesis* 31: 9–18. doi: [10.1093/carcin/bgp268](https://doi.org/10.1093/carcin/bgp268) PMID: [19887512](https://pubmed.ncbi.nlm.nih.gov/19887512/)
78. Wu KJ, Grandori C, Amacker M, Simon-Vermot N, Polack A, et al. (1999) Direct activation of TERT transcription by c-MYC. *Nat Genet* 21: 220–224. PMID: [9988278](https://pubmed.ncbi.nlm.nih.gov/9988278/)
79. Laybutt DR, Weir GC, Kaneto H, Lebet J, Palmiter RD, et al. (2002) Overexpression of c-Myc in beta-cells of transgenic mice causes proliferation and apoptosis, downregulation of insulin gene expression, and diabetes. *Diabetes* 51: 1793–1804. PMID: [12031967](https://pubmed.ncbi.nlm.nih.gov/12031967/)
80. Kaneto H, Sharma A, Suzuma K, Laybutt DR, Xu G, et al. (2002) Induction of c-Myc expression suppresses insulin gene transcription by inhibiting NeuroD/BETA2-mediated transcriptional activation. *J Biol Chem* 277: 12998–13006. PMID: [11799123](https://pubmed.ncbi.nlm.nih.gov/11799123/)
81. Samson SL, Wong NC (2002) Role of Sp1 in insulin regulation of gene expression. *J Mol Endocrinol* 29: 265–279. PMID: [12459029](https://pubmed.ncbi.nlm.nih.gov/12459029/)
82. Beitner-Johnson D, Werner H, Roberts CT Jr., LeRoith D (1995) Regulation of insulin-like growth factor I receptor gene expression by Sp1: physical and functional interactions of Sp1 at GC boxes and at a CT element. *Mol Endocrinol* 9: 1147–1156. PMID: [7491107](https://pubmed.ncbi.nlm.nih.gov/7491107/)
83. Cheung L, Zervou S, Mattsson G, Abouna S, Zhou L, et al. (2010) c-Myc directly induces both impaired insulin secretion and loss of beta-cell mass, independently of hyperglycemia in vivo. *Islets* 2: 37–45. doi: [10.4161/isl.2.1.10196](https://doi.org/10.4161/isl.2.1.10196) PMID: [21099292](https://pubmed.ncbi.nlm.nih.gov/21099292/)
84. Lutzner N, De-Castro Arce J, Rosl F (2012) Gene expression of the tumour suppressor LKB1 is mediated by Sp1, NF-Y and FOXO transcription factors. *PLoS One* 7: e32590. doi: [10.1371/journal.pone.0032590](https://doi.org/10.1371/journal.pone.0032590) PMID: [22412893](https://pubmed.ncbi.nlm.nih.gov/22412893/)
85. Tsai LH, Chen PM, Cheng YW, Chen CY, Sheu GT, et al. (2014) LKB1 loss by alteration of the NKX2-1/p53 pathway promotes tumor malignancy and predicts poor survival and relapse in lung adenocarcinomas. *Oncogene* 33: 3851–3860. doi: [10.1038/onc.2013.353](https://doi.org/10.1038/onc.2013.353) PMID: [23995788](https://pubmed.ncbi.nlm.nih.gov/23995788/)
86. Nieminen AI, Eskelinen VM, Haikala HM, Tervonen TA, Yan Y, et al. (2013) Myc-induced AMPK-phospho p53 pathway activates Bak to sensitize mitochondrial apoptosis. *Proc Natl Acad Sci USA* 110: E1839–1848. doi: [10.1073/pnas.1208530110](https://doi.org/10.1073/pnas.1208530110) PMID: [23589839](https://pubmed.ncbi.nlm.nih.gov/23589839/)
87. Wen JP, Liu C, Bi WK, Hu YT, Chen Q, et al. (2012) Adiponectin inhibits KISS1 gene transcription through AMPK and specificity protein-1 in the hypothalamic GT1-7 neurons. *J Endocrinol* 214: 177–189. doi: [10.1530/JOE-12-0054](https://doi.org/10.1530/JOE-12-0054) PMID: [22582096](https://pubmed.ncbi.nlm.nih.gov/22582096/)
88. Cai X, Hu X, Cai B, Wang Q, Li Y, et al. (2013) Metformin suppresses hepatocellular carcinoma cell growth through induction of cell cycle G1/G0 phase arrest and p21CIP and p27KIP expression and downregulation of cyclin D1 in vitro and in vivo. *Oncol Rep* 30: 2449–2457. doi: [10.3892/or.2013.2718](https://doi.org/10.3892/or.2013.2718) PMID: [24008375](https://pubmed.ncbi.nlm.nih.gov/24008375/)

89. Zhang T, Guo P, Zhang Y, Xiong H, Yu X, et al. (2013) The antidiabetic drug metformin inhibits the proliferation of bladder cancer cells in vitro and in vivo. *Int J Mol Sci* 14: 24603–24618. doi: [10.3390/ijms141224603](https://doi.org/10.3390/ijms141224603) PMID: [24351837](https://pubmed.ncbi.nlm.nih.gov/24351837/)
90. Gartel AL, Radhakrishnan SK (2005) Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer Res* 65: 3980–3985. PMID: [15899785](https://pubmed.ncbi.nlm.nih.gov/15899785/)
91. Markowitz F (2010) How to understand the cell by breaking it: network analysis of gene perturbation screens. *PLoS Comput Biol* 6: e1000655. doi: [10.1371/journal.pcbi.1000655](https://doi.org/10.1371/journal.pcbi.1000655) PMID: [20195495](https://pubmed.ncbi.nlm.nih.gov/20195495/)
92. Shimoni Y, Fink MY, Choi SG, Sealfon SC (2010) Plato's cave algorithm: inferring functional signaling networks from early gene expression shadows. *PLoS Comput Biol* 6: e1000828. doi: [10.1371/journal.pcbi.1000828](https://doi.org/10.1371/journal.pcbi.1000828) PMID: [20585619](https://pubmed.ncbi.nlm.nih.gov/20585619/)
93. DiMasi JA, Feldman L, Seckler A, Wilson A (2010) Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther* 87: 272–277. doi: [10.1038/clpt.2009.295](https://doi.org/10.1038/clpt.2009.295) PMID: [20130567](https://pubmed.ncbi.nlm.nih.gov/20130567/)
94. Yang L, Chen J, He L (2009) Harvesting candidate genes responsible for serious adverse drug reactions from a chemical-protein interactome. *PLoS Comput Biol* 5: e1000441. doi: [10.1371/journal.pcbi.1000441](https://doi.org/10.1371/journal.pcbi.1000441) PMID: [19629158](https://pubmed.ncbi.nlm.nih.gov/19629158/)
95. Bender A, Scheiber J, Glick M, Davies JW, Azzaoui K, et al. (2007) Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* 2: 861–873. PMID: [17477341](https://pubmed.ncbi.nlm.nih.gov/17477341/)
96. Kuhn M, Al Banchaabouchi M, Campillos M, Jensen LJ, Gross C, et al. (2013) Systematic identification of proteins that elicit drug side effects. *Mol Syst Biol* 9: 663. doi: [10.1038/msb.2013.10](https://doi.org/10.1038/msb.2013.10) PMID: [23632385](https://pubmed.ncbi.nlm.nih.gov/23632385/)
97. Salpeter SR, Buckley NS, Kahn JA, Salpeter EE (2008) Meta-analysis: metformin treatment in persons at risk for diabetes mellitus. *Am J Med* 121: 149–157. doi: [10.1016/j.amjmed.2007.09.016](https://doi.org/10.1016/j.amjmed.2007.09.016) PMID: [18261504](https://pubmed.ncbi.nlm.nih.gov/18261504/)
98. Libby G, Donnelly LA, Donnan PT, Alessi DR, Morris AD, et al. (2009) New users of metformin are at low risk of incident cancer: a cohort study among people with type 2 diabetes. *Diabetes Care* 32: 1620–1625. doi: [10.2337/dc08-2175](https://doi.org/10.2337/dc08-2175) PMID: [19564453](https://pubmed.ncbi.nlm.nih.gov/19564453/)
99. Lee MS, Hsu CC, Wahlqvist ML, Tsai HN, Chang YH, et al. (2011) Type 2 diabetes increases and metformin reduces total, colorectal, liver and pancreatic cancer incidences in Taiwanese: a representative population prospective cohort study of 800,000 individuals. *BMC Cancer* 11: 20. doi: [10.1186/1471-2407-11-20](https://doi.org/10.1186/1471-2407-11-20) PMID: [21241523](https://pubmed.ncbi.nlm.nih.gov/21241523/)
100. Gong L, Goswami S, Giacomini KM, Altman RB, Klein TE (2012) Metformin pathways: pharmacokinetics and pharmacodynamics. *Pharmacogenet Genomics* 22: 820–827. doi: [10.1097/FPC.0b013e3283559b22](https://doi.org/10.1097/FPC.0b013e3283559b22) PMID: [22722338](https://pubmed.ncbi.nlm.nih.gov/22722338/)
101. Quinn BJ, Kitagawa H, Memmott RM, Gills JJ, Dennis PA (2013) Repositioning metformin for cancer prevention and treatment. *Trends Endocrinol Metab* 24: 469–480. doi: [10.1016/j.tem.2013.05.004](https://doi.org/10.1016/j.tem.2013.05.004) PMID: [23773243](https://pubmed.ncbi.nlm.nih.gov/23773243/)
102. Pernicova I, Korbonits M (2014) Metformin—mode of action and clinical implications for diabetes and cancer. *Nat Rev Endocrinol* 10: 143–156. doi: [10.1038/nrendo.2013.256](https://doi.org/10.1038/nrendo.2013.256) PMID: [24393785](https://pubmed.ncbi.nlm.nih.gov/24393785/)
103. Hezel AF, Bardeesy N (2008) LKB1; linking cell structure and tumor suppression. *Oncogene* 27: 6908–6919. doi: [10.1038/onc.2008.342](https://doi.org/10.1038/onc.2008.342) PMID: [19029933](https://pubmed.ncbi.nlm.nih.gov/19029933/)
104. Lopez-Bermejo A, Diaz M, Moran E, de Zegher F, Ibanez L (2010) A single nucleotide polymorphism in STK11 influences insulin sensitivity and metformin efficacy in hyperinsulinemic girls with androgen excess. *Diabetes Care* 33: 1544–1548. doi: [10.2337/dc09-1750](https://doi.org/10.2337/dc09-1750) PMID: [20357370](https://pubmed.ncbi.nlm.nih.gov/20357370/)
105. Goldenberg N, Glueck CJ (2008) Is pharmacogenomics our future? Metformin, ovulation and polymorphism of the STK11 gene in polycystic ovary syndrome. *Pharmacogenomics* 9: 1163–1165. doi: [10.2217/14622416.9.8.1163](https://doi.org/10.2217/14622416.9.8.1163) PMID: [18681789](https://pubmed.ncbi.nlm.nih.gov/18681789/)
106. Collier HA, Grandori C, Tamayo P, Colbert T, Lander ES, et al. (2000) Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. *Proc Natl Acad Sci USA* 97: 3260–3265. PMID: [10737792](https://pubmed.ncbi.nlm.nih.gov/10737792/)
107. Franks PW, Christophi CA, Jablonski KA, Billings LK, Delahanty LM, et al. (2014) Common variation at PPARGC1A/B and change in body composition and metabolic traits following preventive interventions: the Diabetes Prevention Program. *Diabetologia* 57: 485–490. doi: [10.1007/s00125-013-3133-4](https://doi.org/10.1007/s00125-013-3133-4) PMID: [24317794](https://pubmed.ncbi.nlm.nih.gov/24317794/)
108. Hahn SS, Tang Q, Zheng F, Zhao S, Wu J, et al. (2014) Repression of integrin-linked kinase by anti-diabetes drugs through cross-talk of PPARgamma- and AMPKalpha-dependent signaling: role of AP-2alpha and Sp1. *Cell Signal* 26: 639–647. doi: [10.1016/j.cellsig.2013.12.004](https://doi.org/10.1016/j.cellsig.2013.12.004) PMID: [24361375](https://pubmed.ncbi.nlm.nih.gov/24361375/)

109. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483: 570–575. doi: [10.1038/nature11005](https://doi.org/10.1038/nature11005) PMID: [22460902](https://pubmed.ncbi.nlm.nih.gov/22460902/)
110. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483: 603–607. doi: [10.1038/nature11003](https://doi.org/10.1038/nature11003) PMID: [22460905](https://pubmed.ncbi.nlm.nih.gov/22460905/)
111. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, et al. (2012) Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci USA* 109: 2724–2729. doi: [10.1073/pnas.1018854108](https://doi.org/10.1073/pnas.1018854108) PMID: [22003129](https://pubmed.ncbi.nlm.nih.gov/22003129/)
112. Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, et al. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 31: 3576–3579. PMID: [12824369](https://pubmed.ncbi.nlm.nih.gov/12824369/)
113. Yu G, Li F, Qin Y, Bo X, Wu Y, et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26: 976–978. doi: [10.1093/bioinformatics/btq064](https://doi.org/10.1093/bioinformatics/btq064) PMID: [20179076](https://pubmed.ncbi.nlm.nih.gov/20179076/)
114. Cohen AL, Soldi R, Zhang H, Gustafson AM, Wilcox R, et al. (2011) A pharmacogenomic method for individualized prediction of drug sensitivity. *Mol Syst Biol* 7: 513. doi: [10.1038/msb.2011.47](https://doi.org/10.1038/msb.2011.47) PMID: [21772261](https://pubmed.ncbi.nlm.nih.gov/21772261/)
115. Cheng J, Xie Q, Kumar V, Hurle M, Freudenberg JM, et al. (2013) Evaluation of analytical methods for connectivity map data. *Pac Symp Biocomput*: 5–16. PMID: [23424107](https://pubmed.ncbi.nlm.nih.gov/23424107/)
116. Zhang B, Shi Z, Duncan DT, Prodduturi N, Marnett LJ, et al. (2011) Relating protein adduction to gene expression changes: a systems approach. *Mol Biosyst* 7: 2118–2127. doi: [10.1039/c1mb05014a](https://doi.org/10.1039/c1mb05014a) PMID: [21594272](https://pubmed.ncbi.nlm.nih.gov/21594272/)
117. Zheng S, Zhao Z (2012) GenRev: exploring functional relevance of genes in molecular networks. *Genomics* 99: 183–188. doi: [10.1016/j.ygeno.2011.12.005](https://doi.org/10.1016/j.ygeno.2011.12.005) PMID: [22227021](https://pubmed.ncbi.nlm.nih.gov/22227021/)
118. Dupont P, Callut J, Dooms G, Monette J-N, Deville Y (2006) Relevant subgraph extraction from random walks in a graph. Research report UCL/FSA/INGI 2006–07.
119. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, et al. (2012) Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *J Am Med Inform Assoc* 19: 212–218. doi: [10.1136/amiajnl-2011-000439](https://doi.org/10.1136/amiajnl-2011-000439) PMID: [22101970](https://pubmed.ncbi.nlm.nih.gov/22101970/)
120. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, et al. (2010) Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* 86: 560–572. doi: [10.1016/j.ajhg.2010.03.003](https://doi.org/10.1016/j.ajhg.2010.03.003) PMID: [20362271](https://pubmed.ncbi.nlm.nih.gov/20362271/)
121. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, et al. (2008) Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 84: 362–369. doi: [10.1038/clpt.2008.89](https://doi.org/10.1038/clpt.2008.89) PMID: [18500243](https://pubmed.ncbi.nlm.nih.gov/18500243/)
122. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909. PMID: [16862161](https://pubmed.ncbi.nlm.nih.gov/16862161/)
123. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23: 1294–1296. PMID: [17384015](https://pubmed.ncbi.nlm.nih.gov/17384015/)
124. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 57: 289–300.
125. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, et al. (2007) The human microbiome project. *Nature* 449: 804–810. PMID: [17943116](https://pubmed.ncbi.nlm.nih.gov/17943116/)
126. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33: W741–748. PMID: [15980575](https://pubmed.ncbi.nlm.nih.gov/15980575/)
127. Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13. doi: [10.1093/nar/gkn923](https://doi.org/10.1093/nar/gkn923) PMID: [19033363](https://pubmed.ncbi.nlm.nih.gov/19033363/)
128. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431–432. doi: [10.1093/bioinformatics/btq675](https://doi.org/10.1093/bioinformatics/btq675) PMID: [21149340](https://pubmed.ncbi.nlm.nih.gov/21149340/)