

## METHODOLOGY ARTICLE

# PMTDS: a computational method based on genetic interaction networks for Precision Medicine Target-Drug Selection in cancer

Varshini Vasudevaraja<sup>1,2</sup>, Jamie Renbarger<sup>3</sup>, Ridhhi Girish Shah<sup>2</sup>, Garrett Kinnebrew<sup>2</sup>, Murray Korc<sup>4,5</sup>, Limei Wang<sup>1,2</sup>, Yang Huo<sup>1,2</sup>, Enze Liu<sup>1,2</sup>, Lang Li<sup>1,2,3,\*</sup> and Lijun Cheng<sup>1,2,3,\*</sup>

<sup>1</sup> Department of Biomedical Informatics, College of Medicine, Ohio State University, Columbus, OH 43210, USA

<sup>2</sup> Department of Medical and Molecular Genetics, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

<sup>3</sup> Department of Pediatrics, hematology/oncology, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

<sup>4</sup> Department of Medicine, Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

<sup>5</sup> Pancreatic Cancer Signature Center, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

\* Correspondence: lali@iu.edu, lijcheng@iupui.edu

Received April 29, 2017; Revised August 9, 2017; Accepted September 3, 2017

**Background:** Precision medicine attempts to tailor the right therapy for the right patient. Recent progress in large-scale collection of patients' tumor molecular profiles in The Cancer Genome Atlas (TCGA) provides a foundation for systematic discovery of potential drug targets specific to different types of cancer. However, we still lack powerful computational methods to effectively integrate multiple omics data and protein-protein interaction network technology for an optimum target and drug recommendation for an individual patient.

**Methods:** In this study, a computation method, Precision Medicine Target-Drug Selection (PMTDS) based on genetic interaction networks is developed to select the optimum targets and associated drugs for precision medicine style treatment of cancer. The PMTDS system includes three parts: a personalized medicine knowledgebase for each cancer type, a genetic interaction network-based algorithm and a single patient molecular profiles. The knowledgebase integrates cancer drugs, drug-target databases and gene biological pathway networks. The molecular profiles of each tumor consists of DNA copy number alteration, gene mutation, and tumor gene expression variation compared to its adjacent normal tissue.

**Results:** The novel integrated PMTDS system is applied to select candidate target-drug pairs for 178 TCGA pancreatic adenocarcinoma (PDAC) tumors. The experiment results show known drug targets (EGFR, IGF1R, ERBB2, NR112 and AKR1B1) of PDAC treatment are identified, which provides important evidence of the PMTDS algorithm's accuracy. Other potential targets PTK6, ATF, SYK are, also, recommended for PDAC. Further validation is provided by comparison of selected targets with, both, cell line molecular profiles from the Cancer Cell Line Encyclopedia (CCLE) and drug response data from the Cancer Therapeutics Response Portal (CTRP). Results from experimental analysis of forty six individual pancreatic cancer samples show that drugs selected by PMTDS have more sample-specific efficacy than the current clinical PDAC therapies.

**Conclusions:** A novelty target and drug priority algorithm PMTDS is developed to identify optimum target-drug pairs by integrating the knowledgebase base with a single patient's genomics. The PMTDS system provides an accurate and reliable source for target and off-label drug selection for precision cancer medicine.

**Keywords:** precision medicine; drug target; algorithm; pancreatic adenocarcinoma; biological pathway; cancer

## INTRODUCTION

Identification of drug targets and biomarkers are becoming the key factors for treating cancer patients in clinical practice [1]. Classical clinical trials yield a handful of measurements from thousands of people. Factors like tumor heterogeneity, drug biomarker validation and the difficulty in combining huge volumes of molecular and drug data impose challenges in providing personalized cancer medicine [2]. Finding the right drug with the right dose for the right person harboring a particular molecular feature is still an unsolved problem. Several ongoing “one-person trials”, not average, for precision medicine responses to therapy are under way [3]. They include the molecularly targeted therapy based on tumor molecular profiling versus conventional therapy for advanced cancer (SHIVA) trial [4], the rational therapeutics based on the analysis of matched tumor and normal biopsies in subjects with advanced malignancies (WINTHER) trial [5], the personalized medicine for patients with advanced cancer in the (IMPACT) Phase I trial [6] and the molecular analysis for therapy choice (NCI-MATCH) Phase II trial launched by US National Cancer Institute (NCI) [7]. One important rationale behind the recent trials is to allow patients from various cancer types that share the same targets to receive the same drug therapies. This pan-cancer target identification is well supported by the strong pan-cancer mutation [8] and copy number alteration [9] patterns observed from the data in The Cancer Genomics Atlas (TCGA, <http://cancergenome.nih.gov/>) project. In order to produce interpretable and reproducible results, several key aspects of these treatment algorithms must be thoughtfully defined prior to starting the trials. There is an urgent need to develop more efficiently therapeutic agents that selectively target particular genetic profiles or molecular features based on pan-omics data [1,10].

Another critical rationale that drives precision medicine cancer therapy, either in clinical practice or trials, is the popularity of off-label drug use [11] especially in aggressive cancer types which have only limited therapeutic options. However, proper selection of off-label drugs in these cancer types suffers from the lack of knowledge base and computational search strategies [12]. With the availability of clinical tumor samples from TCGA data [13] and other large-scale cancer cell line drug screening data, such as the Cancer Cell Line Encyclopedia (CCLE) [14] and the Cancer Therapeutics Response Portal (CTRP) [15], we anticipate that bioinformatics data integration and analysis will make up this knowledge gap. Advances in high-throughput next-generation sequencing (NGS) technologies now allow the identification of actionable molecular alterations (e.g., clinically relevant driver mutations) in an individual

patient in clinical practice [16]. Precision medicine is gradually becoming a reality with the development of the ability to identify the right drug for the right patient based on their tumor molecular profiles [17]. Reference [18] has proposed a single-gene to single drug method for cancer treatment selection based on a patient’s combined gene expression, mutation and DNA copy number alteration profile. Experimental application results for triple negative breast cancer in cell line drug screening are supported by the CCLE. However, this approach does not consider that genes exert their function by signaling pathways and gene-gene interactions in a systematic biological network. Gene mutations and copy number alterations typically cause loss of gene function and pathway inactivity which result in topology changes of the systematic biological network. Analysis of the systematic network is helpful to detect the continuous gene signature transduction from upstream to downstream and observe a whole gene signal variation systematically. In this way, the global effects of a central gene control deficiency due to gene expression dysregulation, copy number deletion or insertion variation can be observed simultaneously. Pathway network-based target-drug selection is promising for drug therapy selection (i.e., one-person trials) by precision oncology derived from cancer pan-omics data [19].

In this study, a personalized medicine knowledgebase, which integrates cancer drugs, drug-target databases and biological pathway networks, is constructed first. Based on this knowledgebase, a pathway-based target-drug discovery algorithm (Precision Medicine Drug-Target Selection, PMTDS) is developed to identify targets personalized to individual patients and associated drugs by multilevel-omics integration, including DNA copy number alteration, gene mutation, and gene expression variation compared to adjacent normal tissue. The PMTDS system takes advantage of gene control network structure information to seek target genes from top to bottom in the gene control network while avoiding selecting genes as targets when they exhibit functional mutations or are part of inactivated pathways. The optimum druggable targets and associated drugs are then recommended.

The integrated system of knowledgebase and PMTDS algorithm is tested for target and associated drug selection in 178 pancreatic adenocarcinoma (PDAC) tumors from TCGA. A total of 110 targets and 22 drugs for 39 PDAC tumors are recommended. In order to validate the reliability of target-drug selection by the PMTDS algorithm for PDAC tumors, molecular profiles from the CCLE and drug response data from the CTRP datasets are integrated. Experimental results of 40 targets and 12 associated drugs on 46 pancreatic cancer cell lines show PMTDS selected drugs have more efficacy than the current clinical PDAC therapies on the tested PDAC

tumor lines. The PMTDS system provides an accurate and reliable source for off-label drug selection for precision cancer medicine.

## SYSTEM PRINCIPLE

### Precision Medicine Target-Drug Selection (PMTDS) system

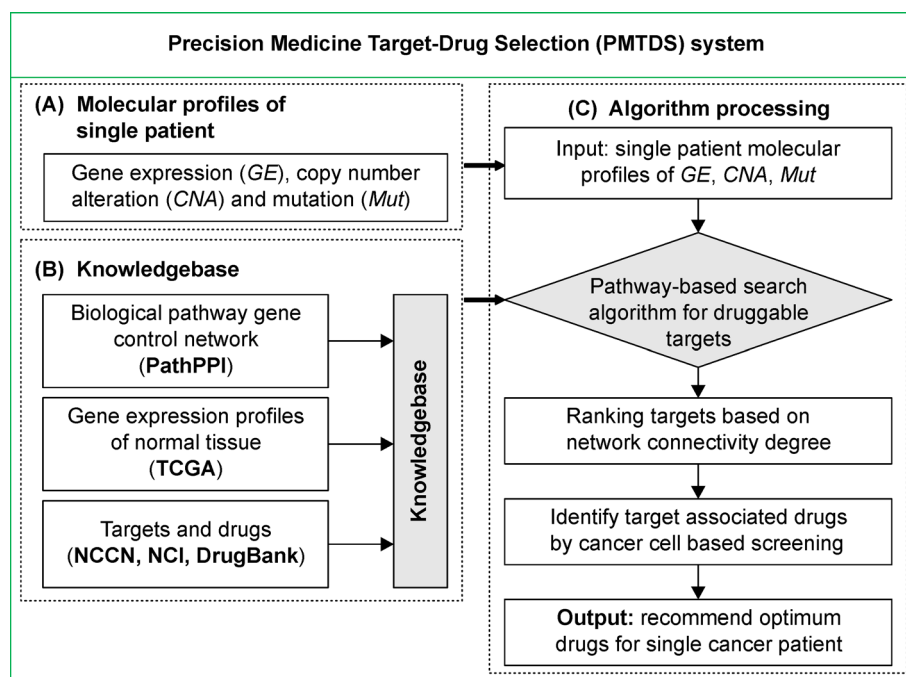
The personalized Precision Medicine Target-Drug Selection (PMTDS) system is based on individual patients' unique molecular profiles integrated with the PMTDS algorithm and knowledgebase to identify optimal targets and recommend associated FDA approved cancer drugs for each patient. Figure 1 describes the schema of the entire PMTDS system including connections between subsystems. The patient molecular profiles include raw counts from RNA sequencing as gene expression (*GE*), copy number alteration (*CNA*) and gene mutation (*Mut*). The knowledgebase includes the gene control network of biological pathways, gene expression profiles of normal tissue, and druggable targets and drugs datasets. It provides important prior knowledge for the PMTDS search algorithm. A pathway-based search algorithm (Figure 1C) is used to select the possible targets and associated cancer drugs for an individual cancer patient by connecting the knowledgebase and patient's *GE*, *CNA* and *Mut* profiles. It is possible that more than a single druggable target is selected per patient. Genes with high

connectivity degree in the biological pathway network, referred to as control hubs, are given priority as potential druggable targets. The potential targets and their drugs are then validated using datasets of large scale drug screening on cancer cell lines from CCLE and CTRP. The most effective drug by the cancer cell line screening is recommended as the optimum therapy for the individual cancer patient.

In this study, biological pathways are collected from PathPPI [20] and gene expression profiles of normal tissue are from Level 3 RNA sequencing Version 2 datasets from TCGA, which are based on the Illumina HiSeq 2000 platform. MapSplice [21] is used for sequence alignment and RSEM [22] is used to perform gene expression quantitation. Cancer drugs are collected from National Cancer Institute (NCI) Drug Dictionary (<https://www.cancer.gov/publications/dictionaries/cancer-drug>) and National Comprehensive Cancer Network (NCCN) Drugs & Biologics Compendium (<https://www.nccn.org/>) before December, 2016. Targets of these cancer drugs are retrieved from DrugBank (<https://www.drugbank.ca/>). Integration of targets and cancer drugs was performed as previously described [18].

### Optimum target and drug selection for single patient based on PMTDS pathway-based search algorithm

The PMTDS pathway-based algorithm is based on five

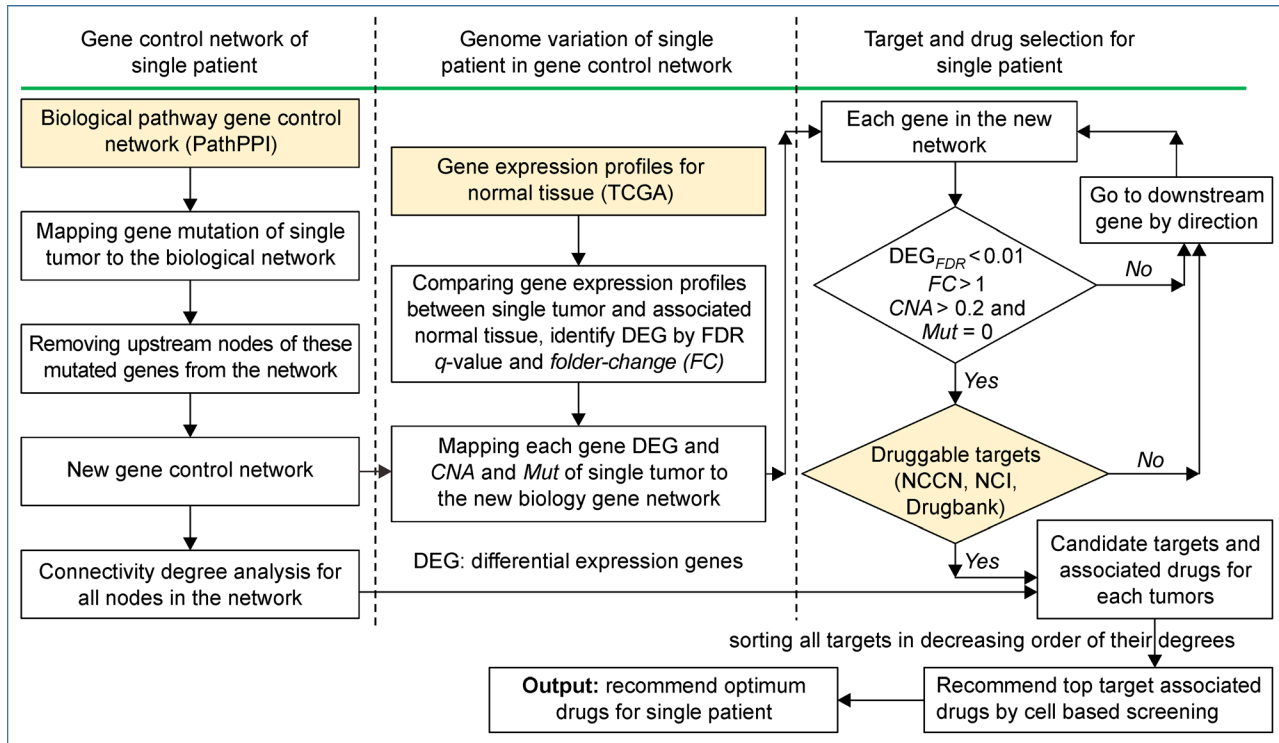


**Figure 1.** The framework of Precision Medicine Target-Drug Selection (PMTDS) system. PMTDS includes three sections: (A) molecular profiles of a single patient; (B) knowledgebase; and (C) algorithm processing.

primary assumptions: (i) The gene interaction network is shared by all types of tumors and the PathPPI biological gene control network is regarded as an inherent prior knowledgebase. (ii) Gene mutation causes functional aberration and effects its biological network topology. (iii) All anti-cancer drugs function as inhibitors, antagonists or low affinity receptors for their targets. The agonists/activators are not considered here. Only genes which are upregulated in tumor versus normal tissue can be usefully targeted by cancer drugs. (iv) In a biological network, genes that are highly interconnected are informally referred to as hub genes. The upregulated hub genes are more effective as druggable targets and are, thus, prioritized for recommendation. (v) There is a many-to-many relationship between targets and drugs, and off-target phenomena can be disregarded.

The components of the PMTDS system implementation are described as follows (Figure 2): (i) Patient-specific gene control network construction: a biological pathway network from PathPPI is considered the standard gene control network in all of cancer patients. The mutated genes of a single patient are mapped to the standard biological network. The immediately adjacent upstream genes of these mutated genes of the patient are removed from the standard network on the assumption

that blockade of signal delivery prevents these genes' functions. This results in a modified gene interaction pathway network specific to a single patient. The connectivity degree of all nodes/genes in the new network is calculated. (ii) Patient-specific genome variation in the gene control network: differential gene expression between the individual patient sample and its corresponding normal tissue group is calculated from raw RNA sequencing counts obtained from TCGA using the *edgeR* software package [23]. The quantile-adjusted conditional maximum likelihood (qCML) method is used to measure the differential expression between tumor and normal tissue of each gene in the patient-specific gene control network, resulting in per-gene *false discovery rate (FDR)* and *fold-change (FC)* values. For each gene in the gene control network from step (i), genome-wide variation for a single patient's tumor is analyzed using a combination of gene expression *FDR* and *FC*, copy number alteration (*CNA*) and binary mutation status. (iii) Patient-specific target and drug selection: based on our assumptions, the genes with upregulated expression in tumor versus normal tissues, high *CNA* amplification and no mutation in the new network are selected as candidates for druggable targets. Default cutoffs for candidate targets are *FDR* less than 0.1, *FC* greater than 1, *CNV* greater



**Figure 2.** The flowchart of target and drug selection for a single patient based on the genetic network search algorithm in the Precision Medicine Target-Drug Selection (PMTDS) system. There are three steps in the algorithm, including patient-specific gene control network construction, patient-specific genome variation of the gene control network and target and drug selection for a single patient. The rectangles colored with a pale pink background each represent a knowledgebase.

than 0.4 and no mutation ( $Mut = 0$ ). All candidate targets are ranked in decreasing order of connectivity degree in the patient-specific gene control network. The efficacy of top target associated drugs is determined using large scale cancer cell line drug screening data from CTRP. The most effective drug is recommended for the patient. Supplementary Materials provide all biological network data, targets and associated cancer drugs list, patient molecular profiles and drug recommendation results for individual patients, these datasets are described in detail in materials and methods. The PMTDS system functions are developed in the Python programming language environment.

**RESULTS**

Pancreatic cancer is the third leading cause of cancer death in the United States. About 85% of pancreatic cancers are pancreatic ductal adenocarcinomas (PDACs). Despite decades of effort, PDAC has the shortest survival time of all major cancers, with a five-year survival rate of only ~8%, in large part because it is treatment-recalcitrant [24,25]. Treatment of PDAC remains challenging as only certain patients benefit from the standard treatment combinations of gemcitabine and the epidermal growth factor receptor (EGFR) inhibitor erlotinib or gemcitabine and nab-paclitaxel [25]. The majority of PDAC tumors

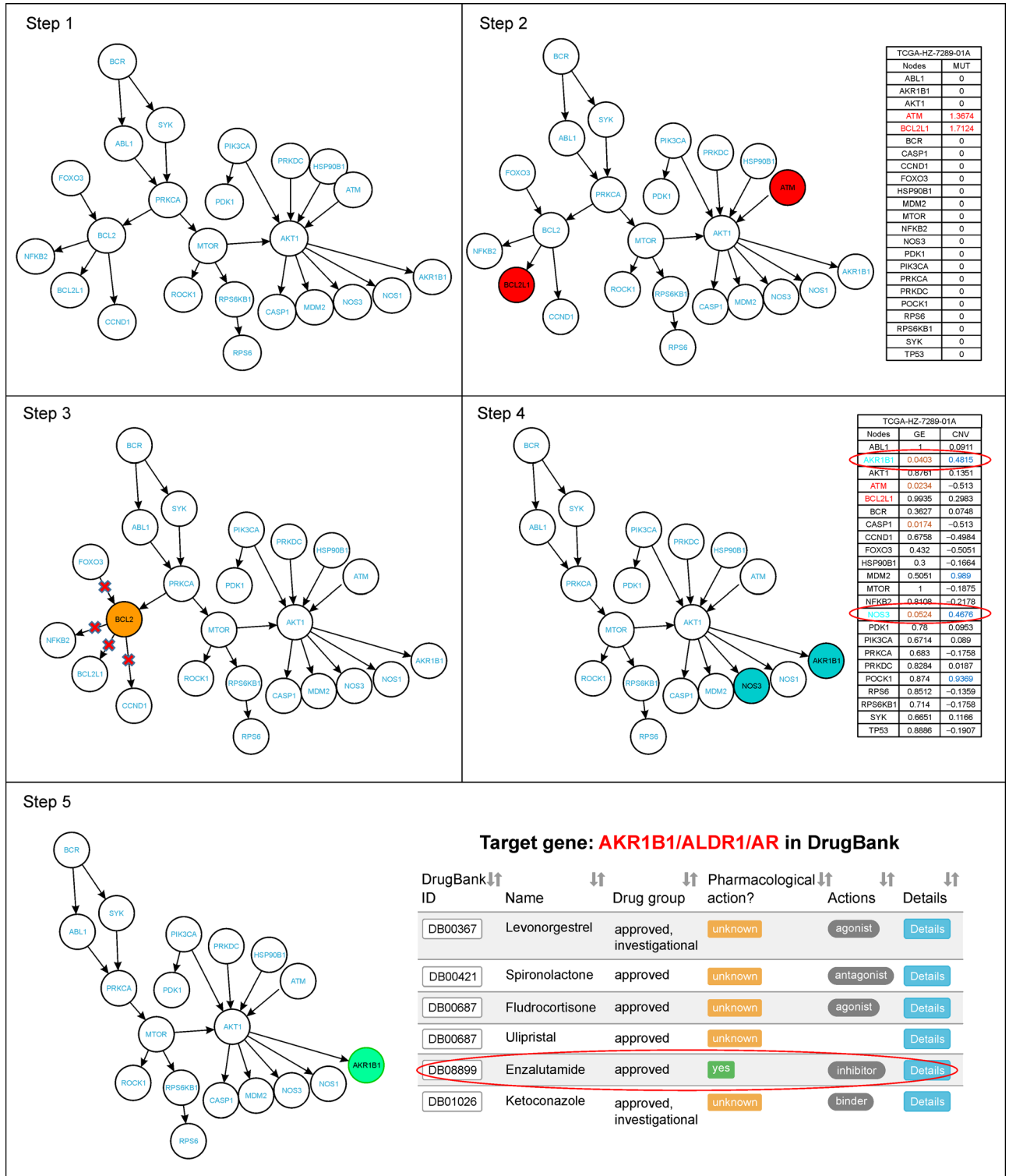
are insensitive to many chemotherapeutic drugs and target-based drugs [24, 25]. Gene network-based methods can help reveal PDAC-associated molecular mechanisms and key targets through a systematic biology strategy.

**Case study 1: Target-drug selection based on the PI3K/AKT/MTOR pathway for pancreatic ductal adenocarcinomas patients in TCGA**

We show how the PMTDS system can, for a single PDAC patient (such as TCGA barcode TCGA-HZ-7289-01A), select an optimum target and associated drug based on the PI3K/AKT/mTOR pathway (Figure 3). In this case the PMTDS systems proceeds in the following 5 steps. Step (i): The PI3K/AKT/MTOR pathway is used as the biological gene control network for running the PMTDS algorithm. Figure 3 shows the signaling pathway with all the nodes/genes and their corresponding edges/interactions. The biological pathway is translated into a network with the help of the NetworkX python module. The node degrees are calculated to determine hub genes in the network which will be used later in the drug-selection process. Step (ii): Using patient TCGA-HZ-7289-01A as an example, patient-specific mutated genes ATM and BCL2L1 are mapped to the PI3K/AKT/MTOR pathway

**Table 1. Selected targets and their frequencies for 39 PDAC patients by the PMTDS system.**

Targeted gene	# of patients	Targeted gene	# of patients	Targeted gene	# of patients	Targeted gene	# of patients	Targeted gene	# of patients
NR1I2/PXR	6	CD55	2	ADCY1	1	CSF3R	1	KLKB1	1
PTK6	6	CPE	2	ADRB2	1	CTSK	1	LY96	1
CEACAM1	5	CYTH3	2	AKR1B1	1	CXCL12	1	MAP3K9	1
F5	5	EGFR	2	ALDH6A1	1	EGLN3	1	MAPT	1
HNF1A	4	FABP4	2	ATF3	1	F12	1	MDM2	1
PCK1	4	FCER1G	2	AURKA	2	FGFR1	1	EPOR	1
PCSK2	4	GHR	2	BAG1	1	FGG	1	MOCOS	1
PLA2G4A	4	GNAI1	2	BDNF	1	FLT1	1	NRP1	1
ASNS	3	GNAS	2	C1R	1	FMO1	1	PDGFRB	1
CD4	3	ITGB4	2	C1S	1	GARS	1	PIK3CG	1
DAPP1	3	KLK1	2	C3	1	GLRX	1	PLAU	1
DRD1	3	NCF2	2	C5	1	GLS2	1	PVR	1
FLT3	3	PDGFRA	2	CAMK2B	1	GPRC5A	1	SLC2A1	1
GUCY2C	3	PRKCG	2	CAMK4	1	HNF4A	1	SLC2A2	1
LYZ	3	PRKCI	2	CCND1	1	HNF4G	1	SLC6A4	1
MMP9	3	PTGS2	2	CD1A	1	HSPA8	1	SLC7A1	1
PRLR	3	SERPINA1	2	CD2	1	ICAM1	1	TGM2	1
SCNN1A	3	SERPINA5	2	CDC25B	1	IGF1R	1	TLR2	1
SNAP25	3	SYK	2	COMP	1	IL6R	1	TNFRSF10B	1
TPI1	3	THBD	2	CRABP2	1	ITK	1	TOP2A	1
ACSS1	2	ACE2	1	CSF1R	1	KIF2C	1	VCP	1
ADA	2	ACSS2	1	CSF2RA	1	KIT	1	VDAC2	1



**Figure 3. A PI3K/AKT/MTOR pathway-based selection of targeted therapy for a pancreatic cancer patient.** In each step, nodes represent genes and edges are directed interactions between genes. Nodes of interest in each step are highlighted in color.

network. Step (iii): Immediately adjacent upstream nodes of the mutated genes ATM and BCL2L1 are removed. BCL2L1 has BCL2 as its predecessor whereas ATM has no upstream nodes. Here BCL2 and its interactions with other genes are removed from the biological network. A new network is produced after elimination, which will be used for identifying drug targets for this patient. Step (iv): All network nodes are inspected to find potential druggable targets with high gene expression, copy number amplification and no mutation according to pre-selected thresholds (default: *fold-change* > 1, *FDR* adjusted *q*-value < 0.1, *CNV* > 0.2 and *Mut* = 0). Figure 3 shows that AKR1B1 and NOS3 have high gene expression and copy number amplification and are not mutated. Thus, AKR1B1 and NOS3 are selected as potential targets for the patient's therapy. Step (v): The target-drug knowledgebase is then used to select drugs targeting either AKR1B1 or NOS3, such as by DrugBank annotation. There are multiple drugs targeting AKR1B1 (aliases ALDR1, AR), but none for the NOS3 protein. The FDA approved cancer drug Enzalutamide, an AR inhibitor, is prioritized in recommendation for patient therapy. Therefore, the targeted therapy for patient TCGA-HZ-7289-01A is the drug Enzalutamide targeting AKR1B1.

### Case study 2: Target-drug selection for PDAC patients in TCGA based on PMTDS system

The PMTDS system is used to identify druggable targets and corresponding FDA approved cancer drugs for 178 individual PDAC patients from TCGA. The experiment result shows 39 patients have at least one druggable target/gene at default thresholds (*FC* > 1, *FDR* < 0.01, *CNA* > 0.2, *mutation* = 0) while the remaining 139 patients do not have any targets selected (results see Tables 1–2, Supplementary File 1 lists all samples annotation, Supplementary File 2 lists the knowledgebase of FDA approved cancer drug and associated target). Selected targets and their frequencies for 39 PDAC patients by the PMTDS system is reported in Table 1. NR1I2/PXR, PTK6, CEACAM1, and F5 are the top 5 most frequently targeted proteins. In recent years, the National Cancer Institute of Canada Clinical Trials Group (NCIC-CTG) has demonstrated a significant survival benefit from the addition of the epidermal growth factor receptor (EGFR) and NR1I2 inhibitor, erlotinib to gemcitabine for patients with pancreatic cancer. The PMTDS system found the target proteins EGFR and NR1I2 for 10 PDAC patients, completely consistent with current clinical treatment.

**Table 2. Recommended targets and corresponding drugs for 39 PDAC patients by PMTDS system.**

Patient_IDs	Targets	Drugs	Patient_IDs	Targets	Drugs
TCGA-2J-AAB6-01A	PTK6	Vandetanib	TCGA-HV-A5A3-01A	PTK6	Vandetanib
TCGA-2J-AAB8-01A	PRKCI	Tamoxifen	TCGA-HV-A7OP-01A	NR1I2/PXR	Erlotinib
TCGA-2J-AABE-01A	F5	Drotrecogin alfa	TCGA-HV-AA8X-01A	PLA2G4A	Aldesleukin, epirubicin
TCGA-2J-AABH-01A	CEACAM1	Arcitumomab	TCGA-HZ-7289-01A	AKR1B1	Sulindac inhibitor
TCGA-2J-AABI-01A	EGFR	Lapatinib	TCGA-HZ-8005-01A	NR1I2/PXR	Erlotinib
TCGA-2J-AABP-01A	PDGFRA	Sunitinib	TCGA-HZ-A8P0-01A	CD55	Chloramphenicol
TCGA-2J-AABU-01A	VCP	Phosphoaminophosphonic acid-adenylate, ester	TCGA-HZ-A9TJ-01A	PRKCI	Tamoxifen
TCGA-2J-AABV-01A	NR1I2/PXR	Erlotinib	TCGA-IB-7646-01A	NR1I2/PXR	Erlotinib
TCGA-2L-AAQA-01A	TOP2A	Doxorubicin	TCGA-IB-7647-01A	PCK1	Phosphoenolpyruvate
TCGA-3A-A9IJ-01A	EGFR	Lapatinib	TCGA-IB-7886-01A	CEACAM1	Arcitumomab
TCGA-3A-A9IL-01A	NR1I2/PXR	Erlotinib	TCGA-IB-7887-01A	NR1I2/PXR	Erlotinib
TCGA-3A-A9IN-01A	MAPT	Paclitaxel, docetaxel	TCGA-IB-7893-01A	NR1I2/PXR	Erlotinib
TCGA-3A-A9IV-01A	FLT3	Sunitinib, ponatinib	TCGA-IB-8127-01A	GUCY2C	Linaclotide
TCGA-3A-A9IZ-01A	PTGS2	Etoposide	TCGA-IB-A7M4-01A	PTK6	Vandetanib
TCGA-FB-A4P6-01A	F5	Drotrecogin alfa	TCGA-L1-A7W4-01A	PTK6	Vandetanib
TCGA-FB-A5VM-01A	NR1I2/PXR	Erlotinib	TCGA-LB-A8F3-01A	PLA2G4A	Aldesleukin, epirubicin
TCGA-FB-AAPP-01A	PCSK2	Insulin regular, insulin porcine	TCGA-S4-A8RO-01A	CCND1	Arsenic trioxide
TCGA-FB-AAPU-01A	PLA2G4A	Aldesleukin, epirubicin	TCGA-US-A776-01A	SCNN1A	Triamterene, amiloride
TCGA-FB-AAQ0-01A	FCER1G	Benzylpenicilloyl, polylysine	TCGA-FB-AAQ2-01A	KIF2C	Hesperidin
TCGA-FB-AAQ1-01A	NR1I2/PXR	Erlotinib			

The PMTDS system selects multiple druggable targets for each pancreatic cancer patient. However, the single optimum target and corresponding FDA approved cancer drugs are our final aim. The CTRP database contains data on cell line models of cancer and provides an important source for drug-targetable susceptibilities that specific genomic alterations impart on human cancers. The area under the plot of drug concentration at a weighted percent viability (for treated vs. untreated) in cell (called “area under the curve” or AUC) gives insight into the extent of exposure to a drug and its ability to kill cancer cells in CTRP experiment. The drug with the lowest average AUC is recommended as the optimum drug for the patient among several drugs against a target. If some drugs are not available in CTRP, all drugs are output together. Table 2 lists the final optimum target and associated FDA approved cancer drugs recommendations for 39 of the PDAC patients from TCGA. However, the remaining 139 patients do not have any targets recommended. For these patients, the PMTDS system will recommend the standard first-line gemcitabine chemotherapy.

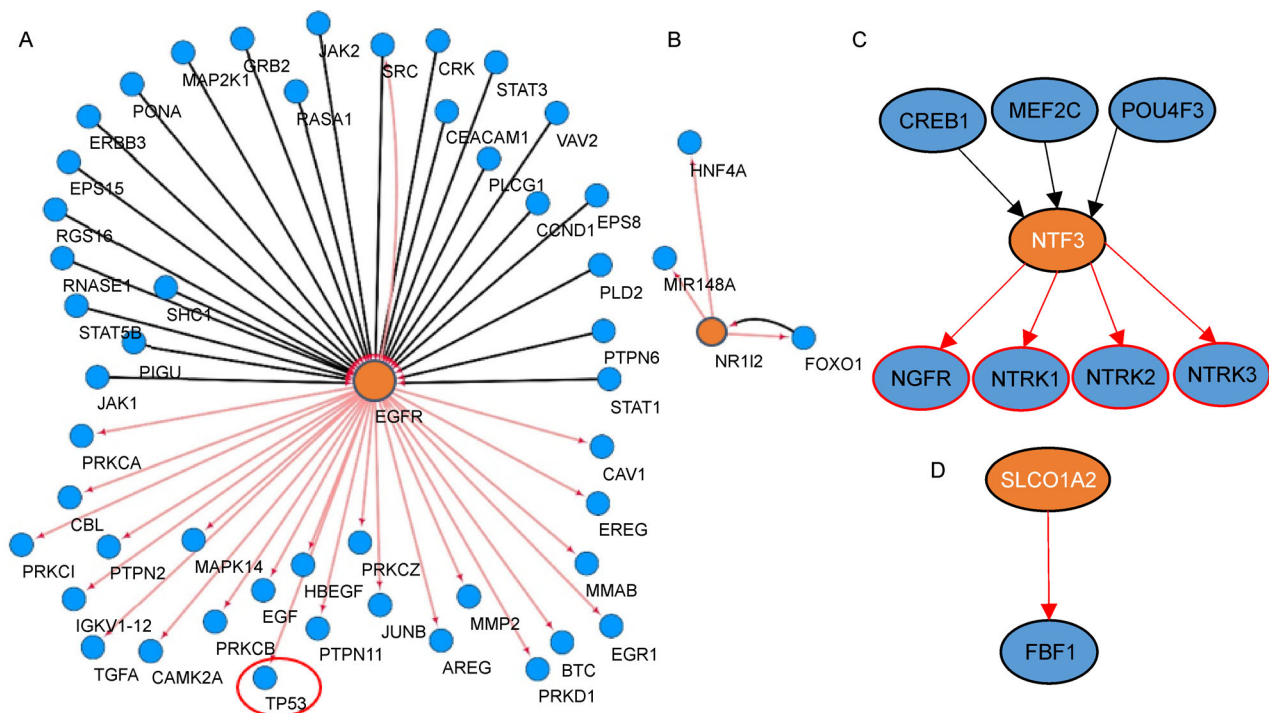
**Drug-target distribution in PDAC patients without systems biology approach**

According to Ref. [18], the single node gene search method for precision medicine is an effective way to select target-drug combinations in the clinic based on patient *GE*, *CNA* and *mutation* profiles. We applied this method for 178 pancreatic cancer patients. The frequency of each target’s selection is presented in Table 3. The proteins KCN10, CHRNA4, SYT2, KCNB2, and NYP are the 5 most frequently occurring targets in this cohort. NR1I2/PXR and EGFR are selected for targeted therapy in 10 patients. Additionally, we compared the results

between the single node approach and the PMTDS pathway-based approach (Tables 1 and 3). The target genes’ distribution by single node search is more concentrated than in the PMTDS pathway-based methods, 37 targets verse 110 targets. Only 12 out of 37 target genes by the single node approach are selected by the PMTDS system. The reason for this appears to be that the missing targets are not included in the biological pathway knowledgebase, with the exception to genes NTF3 and SLCO1A2. The biological pathway knowledgebase cannot cover all human genes. There are 3,649 genes with 7,927 pairs of genes interaction in the biological network (Supplementary File 3). Moreover, the previously studied targets NR1I2/PXR and EGFR are less frequently selected by network-based selection than the single node selection method. This occurs, because the downstream genes of NR1I2 and EGFR are frequently mutated, causing NR1I2 or EGFR to be culled from the potential targets, because they are upstream of mutated genes. Figure 4 shows the genes directly connected to these four genes. TP53 is one of the most frequently mutated genes associated with EGFR in pancreatic cancer. NTF3 and SLCO1A2 do not keep high connectivity degrees in single patient biological networks and lose the chance to be selected as targets by PMTDS algorithm. Hence, a powerful biological pathway knowledgebase is an important issue for target selection and extension of the pathway knowledgebase will improve the accuracy of future work. One possible deficiency of the current method is that nodes with a high degree are likely to be removed, which may cause such nodes and their downstream nodes to be less likely targeted. More work is needed to determine how detrimental this bias is and what possible solutions may be found. In the future we plan to consider methods which penalize nodes upstream to

**Table 3. Target selection for 178 pancreatic patients based on the single node gene search method.**

Targeted gene	# of patients	PMTDS	Targeted gene	# of patients	PMTDS	Targeted gene	# of patients	PMTDS
KCNJ10	25	None	MGAM	10	None	CHRN4	6	None
CHRNA4	18	None	PCK1	10	Yes	FCGR3B	6	None
SYT2	17	None	KCNK1	9	None	KRT12	6	None
KCNB2	16	None	TFR2	9	None	PCSK2	6	Yes
NYP	14	None	KCNH6	8	None	PTK6	6	Yes
ATP4A	14	None	PRKCG	8	Yes	CAMK2B	5	Yes
TUBB1	13	None	ALDH3A1	7	None	CEACAM1	5	Yes
GUCY2C	12	Yes	CHIT1	7	None	DBH	5	None
NR1I2/PXR	12	Yes	CRABP1	7	Yes	DRD1	5	Yes
NPSR1	11	None	F5	7	Yes	GPT	5	None
NTF3	11	None	GRIN2C	7	None	EGFR	2	Yes
SLCO1A2	11	None	CACNA1C	6	None			
TNNC2	11	None	CACNA1G	6	None			



**Figure 4. The upstream and downstream genes of notable targets NTF3, SLCO1A2, NR112 and EGFR of pancreatic cancer in the biological network knowledgebase.** Black arrows represent edges directed towards the gene of interest and red arrow represent edges directed away. The EGFR-SRC and NR112-FOXO1 connections are duplicated due to reciprocal, and thus cyclic, connections between these nodes.

mutated genes without entirely removing them from the network.

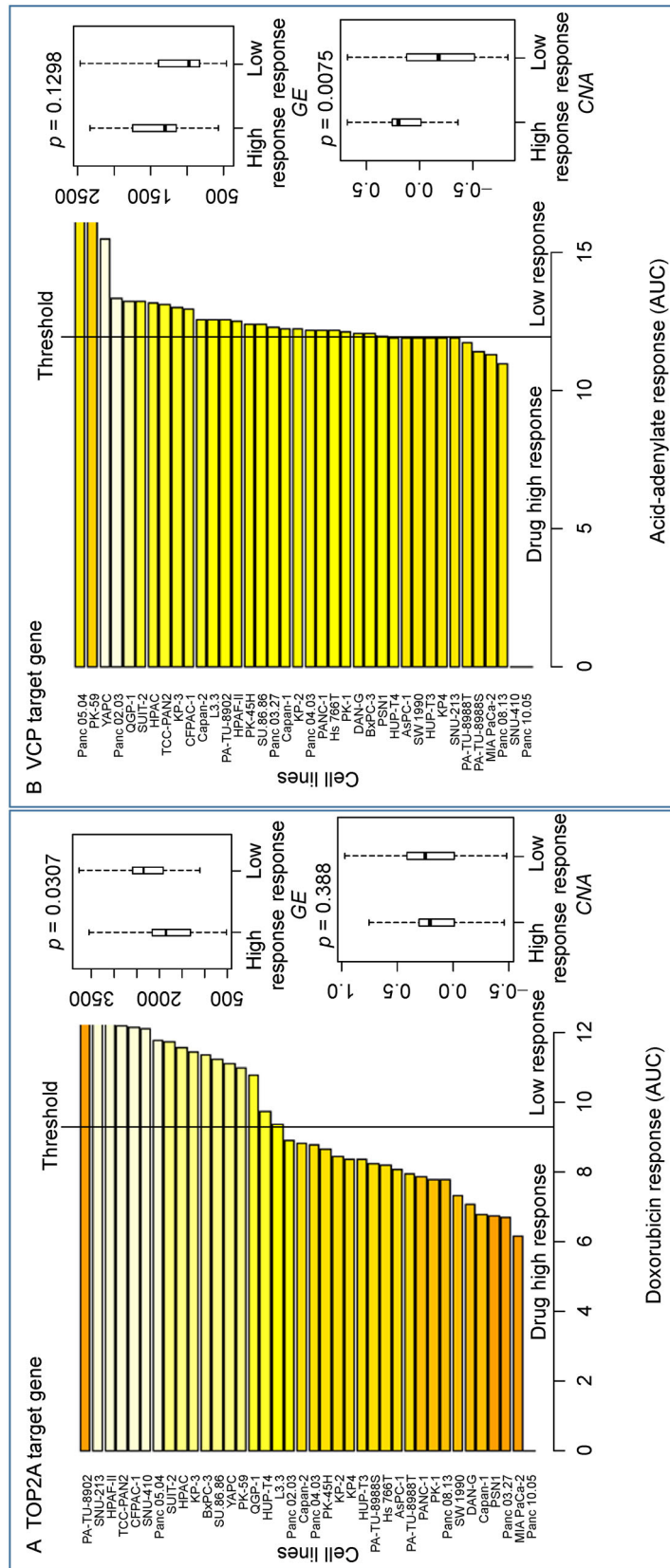
**Pancreatic cell line validation for off-label drug selection**

By linking drug activity with the functional complexity of cancer genomes, systematic pharmacogenomic profiling in cancer cell lines provides a powerful biomarker discovery platform to guide cancer therapeutic strategies. Drug screenings including 46 pancreatic cell lines and 481 drugs from CCLE and CTRP are used to validate potential targets and drugs for pancreatic cancer by the PMTDS system. By comparing the pancreatic tumor samples with cancer cells and connecting the target gene *GE*, *CNV*, *Mut* variation in cancer cells with their drug response (AUC), we are able to estimate drug efficacy for the patient. Cancer cells with high gene expression, copy number amplification and no mutation for a specific target are likely to be affected by its associated drugs. Figure 5 describes two example targets (TOP2A and VCP) and their sensitivity to doxorubicin and Acid-Adenylate in 41

pancreatic cell lines. The *GE*, *CNV* and *Mut* are compared with drug response after splitting into two groups, the strong drug response (low AUC) group and the weak drug response group (high AUC), in 41 cancer cell lines. The pancreatic cancer cells with high *GE*, copy number amplification and no mutation have a significantly increased drug response versus the group with low *GE*, copy number reduction and mutated genes for both TOP2A and VCP. This experiment validates the reliability of target-drug selection by the PMTDS system at the cell line level.

**DISCUSSION**

Selection of the optimum target and corresponding drug for each patient is a critical step toward precision cancer medicine. The PMTDS algorithm is based on the novel idea of integrating signaling pathways and biological networks to observe individual patients’ gene control network variations and assist in selecting patient-optimum pairs of targets and drugs. The current literature shows evidence that the target-drug pairs selected by the



**Figure 5. Drug sensitivity analysis in 38 PDAC cancer cell lines against targets TOP2A and VCP.** Gene expression (GE) and copy number alteration (CNA) associated with drug response (Area Under Drug Response Curve, AUC) in two group's high response and low response are compared by t-test, where the threshold is the mean of AUC. TOP2A has a significant GE correlation with doxorubicin response, and VCP has a significant CNA correlation with acid-adenylate response.

PMTDS algorithm are feasible and appropriate options for the treatment of pancreatic cancer. Representative examples include AKR1B1/AR inhibitors (e.g., Sulindac, Enzalutamide) [26], GARS inhibitors (e.g., Glycine), and IGF1R/EGFR inhibitors (e.g., Insulin, Erlotinib, Bevacizumab [27]. Further clinical trials and mouse models will give better insight into how these new targets can help in treatment decisions [28,29].

One target-drug pair, CEACAM1 and its inhibitor arcitumomab, with external evidence of its importance in pancreatic cancer is selected by PMTDS in 5 patients. Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (CEACAM1) also known as cluster of differentiation 66a (CD66a), encodes a member of the carcinoembryonic antigen (CEA) family. CEACAM1 is absent on normal pancreatic cell, but it is expressed widely in pancreatic cancer [30]. It has been investigated as a novel biomarker for pancreatic cancer [31]. CEACAM1 interacts with TIM-3, a molecule expressed by Tc1 (Cytotoxic T-cells) cells, and participates in immune regulation in pathological conditions, exerting a dual role in the regulation of cell migration. The binding of CEACAM1 expressed by cancer cells with TIM-3 expressed on cytotoxic T-cells induces apoptotic signals in the immune cells. Blockade of CEACAM1 binding restores cytolytic activity by the immune cells against the cancer cells [32]. Our results hint that administration of a CEACAM1 inhibitor will result in increased immune cell activation in pancreatic cancer. Pancreatic cancer has proven difficult to treat with conventional drugs and similarly resistant to initial immunotherapy approaches. The therapeutic approach of combining arcitumomab-based immunotherapy with chemotherapy, such as gemcitabine, may provide a novel strategy to improve the clinical efficacy of pancreatic cancer treatment. Such research is still going on.

Another important gene, activating transcription factor-3 (ATF3), is selected by PMTDS as a druggable target. ATF3 regulates the cell cycle and apoptosis, and has increased expression in pancreatic cancer. Saloman *et al.* [33] have reported using ATF3 immunolabeling to determine if damage to the pancreas of either cerulein-treated or PDAC animals caused an injury response. Further studies may identify potential targets for early detection, prevention, and treatment of pancreatic cancer.

Some important results are found for a target which shows ambiguous reports in recent literatures. The target PTK6 with inhibitor vandetanib is identified for 6 pancreatic cancer patients by the PMTDS system. Studies in Refs. [34,35] have reported that PTK6 is overexpressed in pancreatic cancer and promotes pancreatic cancer cell migration and invasion via ERK pathway signaling activation [35]. PTK6 is a potential target for PDAC

therapy. PTK6 overexpression also increased the S-phase fraction 48 hours after first line gemcitabine treatment in PDAC [34]. These phenomena are in accordance with our findings. However, in recent years, *Lancet Oncology* reported adding vandetanib to gemcitabine failed to significantly prolong overall survival among patients with advanced pancreatic cancer [36]. There are many possible reasons for this result. The PMTDS system selects optimum target-drug pair based on a single therapy without considering drug synergies or drug-drug interactions. This will be addressed in future research.

However, some targets which have already proven to be effective in targeted treatment do not appear among the targets selected by the pathway-based PMTDS algorithm. These targets include ERBB2, CDK1, CDK4, HDAC1 and HDAC4. While the annotations for all drug-targets in DrugBank were present in our knowledge base, these genes frequently failed to meet the criteria for druggable targets. For some patients there was no significant gene expression of ERBB2 in the tumor data and for some it was mutated. Using our algorithm, if a gene is found to be mutated when selecting it as a target, the algorithm goes to downstream genes and inspects their gene expression, copy number alteration and mutation. In this case PTGS2 is downstream gene of ERBB2 and we have selected it as a target instead of ERBB2. AKR1B1 is downstream of both CDK1 and HDAC1 and is present in our selected targeted therapy list. Similarly, TP53's downstream genes MDM2 and SCL6A4 were selected as targets in some patients.

Our treatment algorithm has some limitations that needs to be considered in future works. First is that we did not categorize the patient tumor mutation data as functional or non-functional. We considered only the frequency of mutation in a patient. Differentiating mutations into functional and non-functional groups might improve the accuracy of selecting targets and their corresponding drugs. Also, copy number alteration where we consider only in the context of amplifications of genes. We did not consider deletions, which also plays a key role for molecular abnormalities within cancer patients. Secondly, some patients obtained more than one drug-target based on their genetic profile. We identified the best by finding the targets that act as hubs in entire pathway and drugs that are FDA approved or cancer drugs. A better way of scaling the accurate single drug-target should be designed in future. Additionally, the DrugBank database version we used for our algorithm was version 4.0 which has been superseded by a more recent version. The updated version contains a larger number of FDA approved drug-targets and cancer drugs. The third limitation is that we considered only three molecular alterations: gene expression, copy number

alteration and mutation. If we were to include other abnormalities like methylation, protein expression, or microRNA expression it would give a better perspective of the patients' molecular profile. Finally, the biological pathway data we used contains only the regulatory interactions. Large number of non-regulatory genes are not characterized in our knowledge base. Thus, treatment selection is carried out only at the transcriptomics level. More efforts should be made to run the treatment algorithm at level of protein and molecular interactions.

Our treatment selection algorithm integrates multiple disparate data sources, including patient gene expression, copy number alteration, mutation, drug-target data as well as biological pathway data. We have shown that it selects drug-targets for most of the PDAC patients in TCGA and carried out cell line drug response and sensitivity analysis that support the efficacy of the recommended off-label drug usage. Thus, this algorithm is able to act as an accurate and reliable source for off-label drug selection while providing direction for further experiments on mouse models and cell lines which may help in providing improved treatment strategies in precision medicine. On the other hand, there is still much work to be done in the future. Discovering the potential novel targets, like tumor suppressor (lost-of-function) or oncogenes (gain-of-function) for cancer patient would be an interesting potential extension of this work. Off-label drug selection based on pharmacological function, such as activator or inhibitor, agonist or antagonist, could be performed. The systematic connectivity structure of biological networks is often informative with respect to important function genes, such as, target genes or oncogenes. Finding these patterns could be used to extract more useful information from the biological networks. Finally, replacing upstream node removal with an importance score by integrating node's degree, downstream mutations, downstream network size and DEGs may be helpful for observing targets and drugs selection and would help decrease the current bias towards the removal of high-degree nodes.

## MATERIALS AND METHODS

### Genomics datasets and analysis

Pancreatic adenocarcinoma (PDAC) tumors data from TCGA

Patient molecular profile data, including gene expression (*GE*), copy number alteration (*CNA*) and mutation (*MUT*), were downloaded from The Cancer Genome Atlas (TCGA, <https://cancergenome.nih.gov/>) data portal [37] level three data. To gene expression profiles, there were 183 samples in the PDAC category from TCGA, among which 178 were tumors and 4 were from adjacent-

tumor normal tissue (Supplementary File 1). Sample data along with their clinical annotations were downloaded. The RNA sequencing raw counts which were obtained were based on the Illumina hiseq\_rnaseq version 2 platform. mRNA expression profiling for 17,814 genes was collected. The differential gene expression for each patient versus the four normal tissue samples was analyzed by using EdgeR [23] in the R programming language. An *FDR* *q*-value < 0.01 was used as the threshold for significant differentially expressed genes (DEG). At the same time, the expression *folder-change* was used to capture up- or downregulation of each gene relative to the normal group. The differentially expressed gene sets were then annotated for each patient and were arranged in the form of data matrix. For copy number alteration profiles, TCGA uses the Affymetrix 6.0 single nucleotide polymorphism array platform. Copy number alteration analysis was performed by measuring the strength of probe binding within segments. The number alteration of chromosomes was inferred and normalized based upon specific linear calibration curves relative to the reference sequence hg19 in TCGA. The circular binary segmentation (CBS) algorithm was used to normalize the segmentations (generally,  $\log_2(\text{CN}/2)$ ) for further analysis. Positive segmentation values express *CNA* amplification and arm-level gains while negative values represent deletion. Using bedtools, each segments variations were mapped to the chromosome location to calculate per-gene *CNA*. 28,918 genes were then categorized based on *CNA* thresholds of 0.4 for high copy gain, 0.2 for gain, -0.2 for loss and -0.4 for homozygous loss. For mutation data, DNA sequence mutation annotation format (.maf) files based on the Illumina GAIIX platform were downloaded from TCGA. ANNOVAR (<http://annovar.openbioinformatics.org/>) software was used to identify functional genetic variants detected. ANNOVAR output included detailed mutation types, such as non-sense, missense, somatic and germline mutations using the 1000 Genomes Project and dbSNP database as reference to identify somatic and germline mutations, from which a binary signal matrix was generated where no mutation was denoted as 0 and any somatic mutation as 1.

PDAC cell line data from CCLE and CTRP

Cell lines have become a widespread tool for cancer research and drug screening. The Cancer Cell Line Encyclopedia (CCLE, <http://www.broadinstitute.org/ccle>) project has conducted a detailed genetic characterization of a large panel of human cancer 1,096 cancer cell lines [14]. Gene expression, copy number alterations, and mutations of forty-one pancreatic cancer cell lines were downloaded from CCLE. The gene expression profile

from CCLE is based on the Affymetrix HU133 Plus 2.0 microarray and contains 30,000 genes tested by 54,675 probe sets. Each gene is targeted by several probe sets and the average value of the probe sets targeting a gene was used to represent that genes expression. Copy number alterations in CCLE were interrogated using the same platform, Affymetrix SNP 6.0, and method as with the TCGA samples. DNA mutation data from CCLE was limited to hybrid capture of 1,650 genes. Mutation analysis on CCLE datasets was performed as described above for TCGA samples. ANNOVAR gene-based mutation annotation [38] was performed to identify somatic and germline mutations. Only somatic mutations and functional mutations (not germline mutations) were considered for our data analysis.

Cancer Therapeutics Response Portal (CTRP v2.0, <https://portals.broadinstitute.org/ctrp/>) provided screenings of more than 481 small molecules on 664 cancer cell lines [15]. Drug response (AUC) data for 481 candidate cancer drugs in 41 pancreatic cancer cell lines were collected from CTRP.

### Cancer drugs and their targets

Lists of FDA approved cancer drugs were collected from the National Cancer Institute (NCI, <https://www.cancer.gov/publications/dictionaries/cancer-drug>) and the National Comprehensive Cancer Network (NCCN, <https://www.nccn.org/>). At the time of collection, the 2016 version NCI drug dictionary contained 416 FDA approved drugs related to 39 cancer subtypes. The 2015 version NCCN contained 239 unique FDA approved drugs covering 51 cancer subtypes. All drugs were normalized by brand name and DrugBank ID (<http://www.drugbank.ca/>).

DrugBank version 4.0 includes more than 6,400 drugs. Over 2,000 drugs are FDA approved biotech or small molecules and more than 3,200 drugs are in the experimental stage. Drug-target information is annotated for more than 14,000 proteins or other targets [39,40]. However, currently the database contains a total of 8,250 chemicals/drugs. 229 peptide drugs and around 2,016 small molecule drugs are FDA approved. After integrating NCCN and NCI drugs and targets datasets with DrugBank, 141 cancer drugs and 1,792 FDA approved drugs were composed into 1,086 pairs of cancer drug/drug-targets (Supplementary File 2). Detailed information on this process can be found in previously published work [18].

### Gene-gene interaction data

Gene regulatory interaction data was downloaded from the PathPPI database [20]. PathPPI is an integrated

database of all biological human pathways and protein interactions. There are two kinds of protein-protein interactions (PPI) in the PathPPI database: biological PPIs and technical PPIs. The biological protein-protein interactions are regulatory or directed pathways which includes processes like biochemical reaction regulation (BRR), transport regulation (TR), transport with biochemical reaction regulation (TBRR), complex assembly regulation (CAR) and expression regulation (ER) whereas the technical protein-protein interactions are all undirected pathway interactions which are categorized based on complex assembly interaction (CAI), genetic interaction (GI) and molecular interaction (MI) [20]. Technical PPIs, which include computationally predicted PPIs, were not used in this study since they lack clear *in vivo* relevance. For the PMTDS system, only biological protein-protein interaction data was used to directly find up- and down-stream genes based on drug-targets and transcriptomics data. The 7,888 protein complex interactions in the biological PPIs were filtered out, because we did not consider protein complex based mechanisms in the PMTDS algorithm (Supplementary File 3).

### Gene annotation

The HGNC (HUGO Gene Nomenclature Committee, <http://www.genenames.org/>) database provides researchers with standard gene names for the human genome to avoid the complexity of multiple overlapping and conflicting nomenclature systems. The database currently consists of around 24,000 genes and their corresponding approved gene symbols. Each gene has a unique HGNC ID which makes it easier to identify the gene type. Gene are also annotated with other information including gene synonyms, uniprot ids, refseq ids, previous gene symbols and a functional description about each gene, aids in translating from the NCBI or other database [41]. For this work, all raw data was first translated to HGNC gene names and IDs before analysis. Subsequently, all-omics data analysis and sample profiling was keyed to HGNC IDs. This uniform gene annotation acts as the key for molecular profile data, drug-target data and biological pathway data analysis.

### Programming code

The PMTDS system is developed in Python. The code is shared as a supplementary code file available by download at: <https://github.com/varshivasu7/PMTDS>.

### SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at DOI 10.1007/s40484-017-0126-1.

## ACKNOWLEDGEMENTS

This work was supported by NIH Funding 1U54HD090215-01.

## COMPLIANCE WITH ETHICS GUIDELINES

The authors Varshini Vasudevaraja, Jamie Renbarger, Ridhhi Girish Shah, Garrett Kinnebrew, Murray Korc, Limei Wang, Yang Huo, Enze Liu, Lang Li and Lijun Cheng declare they have no conflict of interests.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## REFERENCES

- Le Tourneau, C., Kamal, M., Tsimberidou, A.-M., Bedard, P., Pierron, G., Callens, C., Rouleau, E., Vincent-Salomon, A., Servant, N., Alt, M., *et al.* (2015) Treatment algorithms based on tumor molecular profiling: the essence of precision medicine trials. *J. Natl. Cancer Inst.*, 108, djv362 <https://doi.org/10.1093/jnci/djv362>
- Ciardiello, F., Arnold, D., Casali, P. G., Cervantes, A., Douillard, J.-Y., Eggermont, A., Eniu, A., McGregor, K., Peters, S., Piccart, M., *et al.*, (2014) Delivering precision medicine in oncology today and in future—the promise and challenges of personalised cancer medicine: a position paper by the European Society for Medical Oncology (ESMO). *Ann. Oncol.*, 25, 1673–1678
- Schork, N. J. (2015) Personalized medicine: time for one-person trials. *Nature*, 520, 609–611
- Le Tourneau, C., Delord, J.-P., Gonçalves, A., Gavaille, C., Dubot, C., Isambert, N., Campone, M., Trédan, O., Massiani, M.-A., Mauborgne, C., *et al.* (2015) Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol.*, 16, 1324–1334
- Rodon, J., Soria, J. C., Berger, R., Batist, G., Tsimberidou, A., Bresson, C., Lee, J. J., Rubin, E., Onn, A., Schilsky, R. L., *et al.* (2015) Challenges in initiating and conducting personalized cancer therapy trials: perspectives from WINTHER, a Worldwide Innovative Network (WIN) Consortium trial. *Ann. Oncol.*, 26, 1791–1798
- Tsimberidou, A.-M., Wen, S., Hong, D. S., Wheler, J. J., Falchook, G. S., Fu, S., Piha-Paul, S., Naing, A., Janku, F., Aldape, K. *et al.* (2014) Personalized medicine for patients with advanced cancer in the phase I program at MD Anderson: validation and landmark Analyses. *Clin. Cancer Res.*, 20, 4827–4836
- Mullard, A. (2015) NCI-MATCH trial pushes cancer umbrella trial paradigm. *Nat. Rev. Drug Discov.*, 14, 513–515
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature*, 502, 333–339
- Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., Lawrence, M. S., Zhang, C. Z., Wala, J., Mermel, C. H., *et al.* (2013) Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.*, 45, 1134–1140
- Chin, L., Andersen, J. N. and Futreal, P. A. (2011) Cancer genomics: from discovery science to personalized medicine. *Nat. Med.*, 17, 297–303
- Conti, R. M., Bernstein, A. C., Villafior, V. M., Schilsky, R. L., Rosenthal, M. B. and Bach, P. B. (2013) Prevalence of off-label use and spending in 2010 among patent-protected chemotherapies in a population-based cohort of medical oncologists. *J. Clin. Oncol.*, 31, 1134–1139
- Gupta, S. K. and Nayak, R. P. (2014) Off-label use of medicine: Perspective of physicians, patients, pharmaceutical companies and regulatory authorities. *J. Pharmacol. Pharmacother.*, 5, 88–92
- Witkiewicz, A. K., McMillan, E. A., Balaji, U., Baek, G., Lin, W.-C., Mansour, J., Mollae, M., Wagner, K.-U., Koduru, P., Yopp, A., *et al.* (2015) Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.*, 6, 6744
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483, 603–607
- Seashore-Ludlow, B., Rees, M. G., Cheah, J. H., Cokol, M., Price, E. V., Coletti, M. E., Jones, V., Bodycombe, N. E., Soule, C. K., Gould, J., *et al.* (2015) Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.*, 5, 1210–1223
- Vargas, A. J. and Harris, C. C. (2016) Biomarker development in the precision medicine era: lung cancer as a case study. *Nat. Rev. Cancer*, 16, 525–537
- Jameson, J. L. and Longo, D. L. (2015) Precision medicine — personalized, problematic, and promising. *N. Engl. J. Med.*, 372, 2229–2234
- Cheng, F., Hong, H., Yang, S. and Wei, Y. (2016) A bioinformatics approach for precision medicine off-label drug drug selection among triple negative breast cancer patients. *J. Am. Med. Inform. Assoc.*, 23, 741–749
- Cheng, F., Hong, H., Yang, S. and Wei, Y. (2017) Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief. Bioinformatics*, 18, 682–697
- Tang, H., Zhong, F., Liu, W., He, F. and Xie, H. (2015) PathPPI: an integrated dataset of human pathways and protein-protein interactions. *Sci. China Life Sci.*, 58, 579–589
- Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., He, X., Mieczkowski, P., Grimm, S. A., Perou, C. M., *et al.* (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.*, 38, e178
- Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. and Dewey, C. N. (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, 26, 493–500
- Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of

- digital gene expression data. *Bioinformatics*, 26, 139–140
24. Hidalgo, M. (2012) New insights into pancreatic cancer biology. *Ann. Oncol.*, 23, x135–x138
  25. Kruger, S., Haas, M., Ormanns, S., Bächmann, S., Siveke, J. T., Kirchner, T., Heinemann, V. and Boeck, S. (2014) Translational research in pancreatic ductal adenocarcinoma: current evidence and future concepts. *World J. Gastroenterol.*, 20, 10769–10777
  26. Chung, Y. T., Matkowskyj, K. A., Li, H., Bai, H., Zhang, W., Tsao, M.-S., Liao, J. and Yang, G.-Y. (2012) Overexpression and oncogenic function of aldo-keto reductase family 1B10 (AKR1B10) in pancreatic carcinoma. *Mod. Pathol.*, 25, 758–766
  27. Philip, P. A. (2008) Targeted therapies for pancreatic cancer. *Gastrointest. Cancer Res.*, 2, S16–S19
  28. Scott, A. J. and Wilkinson, J. C. (2016) HNF1A, KRT81, and CYP3A5: three more straws on the back of pancreatic cancer? *Transl. Cancer Res.*, 5, S253–S256
  29. Thompson, M. R., Xu, D. and Williams, B. R. G. (2009) ATF3 transcription factor and its emerging roles in immunity and cancer. *J. Mol. Med. (Berl.)*, 87, 1053–1060
  30. Simeone, D. M., Ji, B., Banerjee, M., Arumugam, T., Li, D., Anderson, M. A., Bamberger, A. M., Greenson, J., Brand, R. E., Ramachandran, V., *et al.* (2007) CEACAM1, a novel serum biomarker for pancreatic cancer. *Pancreas*, 34, 436–443
  31. Harsha, H. C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadran, S., Gollapudi, S., Balakrishnan, L., Dwivedi, S. B., Telikicherla, D., Selvan, L. D. N., *et al.* (2009) A compendium of potential biomarkers of pancreatic cancer. *PLoS Med.*, 6, e1000046
  32. Huang, Y.-H., Zhu, C., Kondo, Y., Anderson, A. C., Gandhi, A., Russell, A., Dougan, S. K., Petersen, B.-S., Melum, E., Pertel, T., *et al.* (2015) CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. *Nature*, 517, 386–390
  33. Saloman, J. L., Albers, K. M., Li, D., Hartman, D. J., Crawford, H. C., Muha, E. A., Rhim, A. D. and Davis, B. M. (2016) Ablation of sensory neurons in a genetic model of pancreatic ductal adenocarcinoma slows initiation and progression of cancer. *Proc. Natl. Acad. Sci. USA*, 113, 3078–3083
  34. Ono, H., Basson, M. D. and Ito, H. (2015) PTK6 potentiates gemcitabine-induced apoptosis by prolonging s-phase and enhancing DNA damage in pancreatic cancer. *Mol. Cancer Res.*, 13, 1174–1184
  35. Ono, H., Basson, M. D. and Ito, H. (2014) PTK6 promotes cancer migration and invasion in pancreatic cancer cells dependent on ERK signaling. *PLoS One*, 9, e96060
  36. Middleton, G., Palmer, D. H., Greenhalf, W., Ghaneh, P., Jackson, R., Cox, T., Evans, A., Shaw, V. E., Wadsley, J., Valle, J. W., *et al.* (2017) Vandetanib plus gemcitabine versus placebo plus gemcitabine in locally advanced or metastatic pancreatic carcinoma (ViP): a prospective, randomised, double-blind, multicentre phase 2 trial. *Lancet Oncol.*, 18, 486–499
  37. Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. (Pozn.)*, 19, A68–A77
  38. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, 38, e164
  39. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, 34, D668–D672
  40. Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B. and Hassanali, M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, 36, D901–D906
  41. Bruford, E. A., Lush, M. J., Wright, M. W., Sneddon, T. P., Povey, S. and Birney, E. (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.*, 36, D445–D448