Identifying drug combinations of individual cancer patients by personalized drug controller method

Wei-Feng Guo[1], Shao-Wu Zhang[1*], Tao Zeng[2*], Luonan Chen[1,2,3*]

[1] *Key Laboratory of Information Fusion Technology of Ministry of Education, School of Automation, Northwestern Polytechnical University, Xi'an 710072, China,*

[2] *Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology , Shanghai Institutes for Biological Sciences, Chinese Academy of sciences, Shanghai 200000, China*

[3] *Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan*

Correspondence: *zhangsw@nwpu.edu.cn,zengtao@sibs.ac.cn,lnchen@sibs.ac.cn*

Abstract

It is a big challenge to identify patient-specific drug combinations based on cancer omics data. However, most conventional methods used for identifying personalized therapies require large sample sizes and focuse on the population as a whole and we still lack a feasible mathematical framework to circumvent this key problem in the clinical application of precision medicine. This work presents a personalized drug controller method (PDC) to identify drug combinations of individual cancer patients, by exploring the transition state information from a disease state to a normal state, thus providing novel insights into tumor heterogeneity in complex patient ecosystems. We validate the effectness of PDC in terms of the accurate prediction of drug combinations on two benchmark cancer datasets. We can also discover personalized key control genes (KCGs) even if the KCGs are hidden in transcription profiles by exploring their network and structural characteristics. Furthermore, we provide computationally derived side effect signatures for drug combinations based on patient-specific KCGs to enhance patient stratification and prognostication. The experimental results strongly support that PDC is effective for personalized drug therapy and treatment, and can fill the gap between network control theory and the personalized drug discovery problem.

Keywords: Tumor heterogeneity; Network control theroy; Patient ecosystems; Drug combination; Drug side effect

Introduction

The clinical efficacy of targeted cancer therapies is hampered by the development of acquired resistance. In an effort to combat targeted therapy drug resistance, drug combinations are increasingly being used as standard of care treatment strategies for several cancers (Jia et al. 2009). A current critical challenge is how to develop efficient *in silico* methods for discovering effective drug combinations from the analysis of large data sets and complex dynamic processes. The number of potential drug combinations generated from such analyses is astronomical, and these combinations cannot all be tested in a rational and rigourous manner. Therefore, scientists are inundated with large amounts of data, and no feasible way of distilling said data into viable drug combinations (Sun et al. 2013). Recent studies have highlighted the importance of targeting key oncogenes for disease control. Designing novel combinatorial strategies to target these fundamental cancer genes may result in synergistic antitumor activity and broader application of current therapies (Phillips 2008; Han et al. 2017; Quan et al. 2018; Yip et al. 2018). A number of systems biology tools have been developed to identify key genes, which can be specifically targeted for drug development (Bush et al. 2006; Hunter et al. 2007; Liu et al. 2010; Wan et al. 2010; Ling et al. 2011; Zhang 2012; Chang et al. 2015; Wright and Ziegler 2017).

Cancer is increasingly viewed as a “patient ecosystem”, a system in which tumor genes cooperate and can also adapt and evolve to changing conditions in their microenvironment(Merlo et al. 2006; Greaves and Maley 2012; van der Wijst et al. 2018). Precision or personalized medicine requires a different type of clinical trial that focuses on individual, not average, responses to therapy. It has been reported that at present, some of the most commonly prescribed drugs are effective in less than 25% of the people for whom they are prescribed (Schork 2015). Therefore, recognition that researchers need to consider individual heterogeneity is driving huge interest for identifying drugs in precision medicine. However, most of the existing detection methods for combinatorial drugs ignore personalized sample information, and fail to

identify patient-specific combinational drugs, which is a necessary and important facet of precision medicine. Therefore, new techniques are urgently needed to better understand tumor heterogeneity and its impact on cancer development, and to generate personalized drug combinations.

To solve this problem, we introduced network control theory to model the control role of drugs (as controllers) on the transition state of gene co-expression networks from tumor state to normal state. Network control theory considers how to choose key network elements as drivers, the activation of which may drive the entire network towards a desired control objective or state based on proper control signals (Liu et al. 2011; Gao et al. 2014; Ruths and Ruths 2014; Wu et al. 2014a; Guo et al. 2017; Guo et al. 2018d). Recent studies on network controllability have offered powerful mathematical frameworks to understand diverse biological systems at a network level (Jgt et al. 2017; Guo et al. 2018c; Li et al. 2018). However, these existing methods mainly focus on the statistical study of tumor heterogeneity at a population level, and cannot be directly applied to the analysis of individual patients. This is primarily due to a gap between the network control theory and the applications in the personalized patient system. More efficient methods are needed to obtain personalized state transition networks that capture the phenotypic transitions between normal and disease states, which is a rate-limiting step of network control methods. However, most of the current network control methods assume unspecified linear dynamics or linearized non-linear dynamics in the complex networks, which may only give an incomplete view of the network control properties of the nonlinear dynamic biological system. To address these limitations, we present a novel personalized drug controller method (PDC), which sheds light on the cancer transition network from a cancer state to a normal state, as well as on the cancer ecosystem of each patient at an individual level by integrating patient data with the network control framework. The impetus behind PDC was to develop efficient models for pushing a complex, nonlinear, individual biological system from disease attractor (i.e., a steady disease state) to normal attractor (i.e., a steady normal state) through potential drug action on specific target genes. Our PDC framework mainly includes two parts: a paired

Single-Sample-Network (SSN) method for constructing the state transition network between normal state and tumor state; and a novel nonlinear control scheme based on undirected networks (NCUA) for detecting both personalized key control genes (KCGs) and personalized combinatorial drugs specific for KCGs.

We have applied PDC to breast cancer (BRCA) and lung cancer (LUSC) datasets obtained from TCGA for the benchmark analysis. We discovered that: i) the nonlinear control method NCUA performed better than other existing network control methods in terms of precision for discovering personalized combinatorial drugs; and ii) the prior gene (or protein) interaction network can improve the precision for predicting combinational drugs. Actually, one significant advantage of our method is to explore network and structural statistics of patient-specific genes, including tests of mutations, differential expression, hub characteristic, and of co-occurrence in gene interaction networks. The following unexpected findings were generated from these analyses: i) the personalized KCGs tended to be genes with insignificant differential expression; ii) the mutational profiles were specific to each patient; and iii) the KCGs were structurally enriched in hub genes and functionally connected in gene interaction networks for most cancer patients.

To elucidate the mechanisms of drug-target action, we quantified the side effects of drug pairs on KCGs by evaluating the *side effect score* for each patient (Torres and Altafini 2016). We estimated the side effect of all of the drug pairs on KCGs for each patient and obtained the number of drug pairs with an *aggravating effect* (side effect score >0), and the number of drug pairs with *an enhancing effect* (side effect score <0). By considering the drug pairs with an *aggravating effect* or an *enhancing effect*, we found that the patient samples in the breast cancer TCGA dataset could be clearly classified into different groups with significant risk assessment. This demonstrates that combinational drug side effect signatures may enable patient stratification and prognostication based on the actionability of drug combinations. In conclusion, our PDC can provide key insights into tumor heterogeneity and its impact on the development and clinical application of personalized cancer therapy.

Results

Overview of personalized drug controller

To identify personalized combinational drugs, we designed PDC to model the role of drugs (as controllers) on the transition from tumor state to normal state. The steps of PDC mainly included: 1) a paired Single-Sample-Network (SSN) method was developed to construct the transition state gene network between a normal state and a tumor state. In that, the co-expression network for each normal and tumor sample was constructed by SSN (Liu et al. 2016) and we integrated the two co-expression networks and the gene interaction networks to obtain the personalized transition state gene network between normal state and tumor state; and 2) a novel nonlinear structural control method (NCUA) based on feedback vertex sets (Fiedler et al. 2013a; Mochizuki et al. 2013) was developed and used for identifying personalized key control genes (KCGs). This was done to ensure that the system state would asymptotically be changed in the personalized transition network, and further prioritize the personalized combinatorial drugs by considering the control impact on the personalized KCGs. These two main methodologies are described in more detail in the methods section. Our PDC consisted of two main parts. One is a paired single sample network method (paired SSN) rather than an original SSN, which we used to construct personalized transition state gene networks. In these networks, edges existed in both the gene interaction network and the differential co-expression network for each patient (**Figure 1 and Figure S1 (a)**) based on integration of the matched genetic information of normal and disease samples from the same patient. The personalized transition state gene networks are undirected and weighted graphs in which nodes represent genes, and a pair of nodes denotes the significant co-expression difference between normal state and tumor state in the gene interaction networks, and the edge weight indicates the differential fold change co-expression between normal sample and tumor sample for a specific patient. For the second part of PDC, we designed an improved nonlinear network control method based on

feedback vertex sets (NCUA), where we considered the optimization of target nodes towards the desired control objective, instead of only evaluating the controllability of the networks. To demonstrate the effectiveness of NCUA as an important factor, we also evaluated the precision of other existing control methods to predict drug combinations, including Liu's linear network control method (Liu method)(Liu et al. 2011), Minimum Dominating Sets based control method (MDS method) (Nacher and Akutsu 2012), Single sample controller method (SCS method) (Guo et al. 2018a), and minimum Directed-networks Feedback vertex set based method (DFVS method) (Chakradhar et al. 1995; Fiedler et al. 2013b). The detailed computational procedures are listed in the Supplementary Methods section of the **Supplementary Manuscript**).

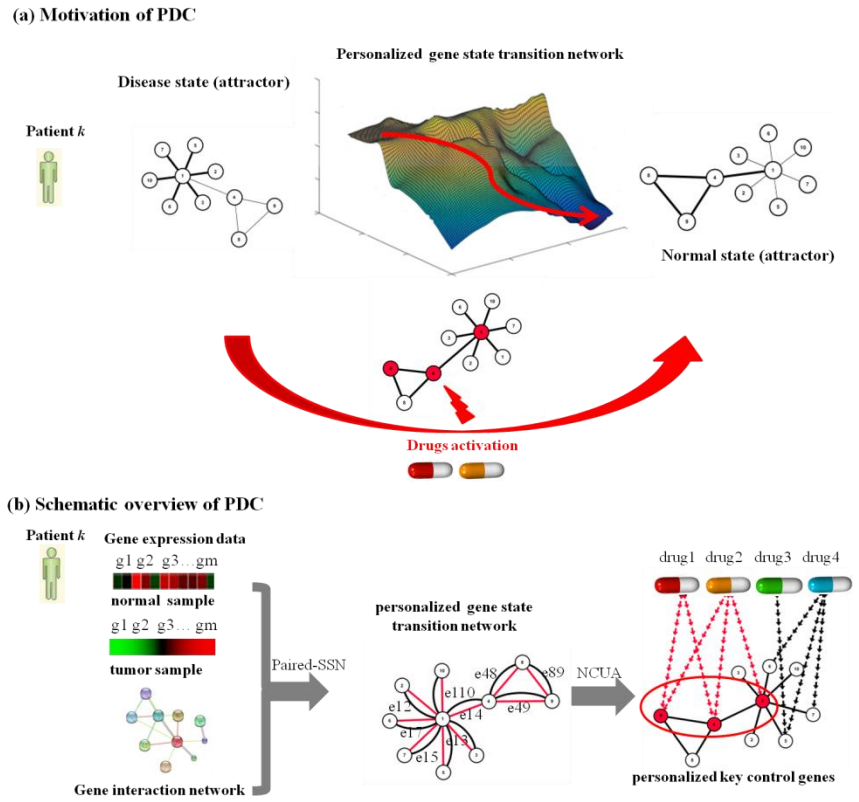

Figure 1 Overview of personalized drug controller strategy for identifying personalized drug combinations. (a) The impetus behind PDC was to design and integrate methods for constructing personalized transition state gene networks to capture the phenotypic transitions between normal and disease attractors, and developing control principles where we push a complex, nonlinear individual system in cancer from disease attractor to normal attractor through drug activation to a feasible subset of network nodes. The assumption of our PDC is that a gene network exists in each attractor (normal attractor or disease attractor) for each patient, where each edge denotes the co-expression of a pair of interacted genes. (b) We developed a paired single sample network method (paired SSN) for constructing personalized transition state gene networks to capture the phenotypic transitions between normal and disease state through the integration of expression data and gene interaction networks. By considering the nonlinear dynamics of transition state gene networks, we also designed a novel nonlinear network control method based on feedback vertex set control (NCUA), where we considered the optimization of input nodes (personalized KCGs) towards the desired control objective, instead of only evaluating the controllability of the networks. Based on their impact on the personalized KCGs (red circles), we extracted the subnetworks connecting the drugs and personalized KCGs (red edges) and prioritized the personalized drug combinations. The details of PDC are depicted in Supplementary Figure S1.

Collection of benchmark datasets and reference networks

The personalized drug controller strategy (PDC) frames the identification of personalized drug combinations as a network control problem in a two-layer multimodal network of two node types: the personalized transition state gene network, and the drug and genes interaction network. On one hand, the personalized transition state gene network describes gene relationships to capture the phenotypic transitions between normal and disease states. On the other hand, the drug and genes links describe the genes targeted by a given drug.

First, to obtain personalized transition state gene networks, we collected gene expression data and gene interaction data. The gene expression data for breast cancer (BRCA, 112 paired samples including normal and tumor sample data) and lung cancer (LUSC, 49 paired samples) were obtained from TCGA. Our PDC used the reference Multidatasets Gene Interaction Network (MGIN), which combines data mining from the literature with theoretical predictions (Hou and Ma 2014). The reference MGIN consisted of 11,648 genes and 211,794 edges, which integrated a variety of sources, including the network used in MEMo (Wu et al. 2010; Ciriello et al. 2012), as well as the most recently curated information from Reactome (Croft et al. 2010), the NCI-Nature Curated PID (Schaefer et al. 2009), and KEGG (Kanehisa et al. 2011). To further demonstrate the importance of prior MGIN network, we also obtained two other types of interaction networks. One was the Synthetic Lethality genes interactions Database (SynLethDB), which is a comprehensive database of a large set of synthetic lethal gene pairs that can lead to cellular or organismal death upon simultaneous mutation. From the Synthetic Lethality Database, 6513 genes and 19,955 synthetic lethal gene pairs for human tumors were collected based on literature review or theoretical prediction. Another interaction network was the human signal PPI network (sPPI) constructed by Vinayagam et al. (Vinayagam et al. 2011). The human sPPI network consists of 6339 proteins and 34813 edges, where the edge corresponds to the hierarchy of signal flow between the interacting proteins.

Second, to obtain the drug-gene interaction data, we searched DCDB (Liu et al. 2014b) and the Drug–Gene Interaction database (DGIdb), DrugBank and the Therapeutic Target Database (TTD)(Law et al. 2014; Hong et al. 2016; Wagner et al. 2016), and 342 documented combinations containing the predicted combinatorial drugs that can target the synthetic lethal gene pairs, were identified (Quan et al. 2018). Among them, 32 and 17 were reported in the literature for efficiently treating breast cancer and lung cancer, respectively. In this work, to analyze the risk assessment of cancer patients, the human drug-target network with positive and negative interactions information was used to add another functional element to the network-based investigation of drug-target interactions, including the activation, inhibition, and pharmacological actions of drug-gene interactions (Torres and Altafini 2016). Analysis of the functional enrichment of KCGs with the Unified Medical Language System (UMLS) and UMLS::similarity identified 2341 breast cancer-related genes (Aronson 2001; Liu et al. 2014a; Quan et al. 2018). All of the data resources used in this study, including the cancer benchmark datasets and the reference network datasets, and the source code of PDC with NCUA and other network control methods, are provided for free at http://sysbio.sibcb.ac.cn/cb/chenlab/software.htm.

PDC accurately prioritizes personalized anti-cancer drug combinations

By calculating the number of targeted KCG for each of the drug combinations, PDC allowed us to rank the anti-cancer drug combinations and select the top k (k=1,2,⋯5) ranked anti-cancer drug combinations for each patient. To verify the effectiveness of our PDC approach, the drug combinations annotated in the Clinical Anti-cancer Combinational drugs (CAC) were applied to assess the precision of the top-ranked anti-cancer drug combinations from PDC using the formula:

$$P = mean\left(p_k \right),$$

where p_k denotes the fraction of the top k predicted drug combinations within CAC for treating cancer.

The key aspect of our PDC model is using NCUA (a nonlinear structure control method) to identify the KCGs (The details are seen in Methods). By comparing the performance of NCUA to that of other network control methods, i.e., Liu method(Liu et al. 2011), MDS method(Nacher and Akutsu 2012), SCS method (Guo et al. 2018a), and DFVS method(Chakradhar et al. 1995; Fiedler et al. 2013b)**,** the highest precision can be achieved in NCUA's predictions as seen in **Figure 2 (a1) and Figure 2 (b1)**. For example, as seen in **Figure 2 (a1),** among the top 5 predicted drug combinations in BRCA and LUSC cancer data sets, PDC's concordance with NCUA in SynLethDB network and multidatasets gene interaction network (MGIN) network was >80% and >50% respectively, while the precision of other control methods (Liu method and MDS and SCS and DFVS) was <60% and <40% respectively. Although both NCUA and MDS focus on undirected network, NCUA is a nonlinear control method while MDS is a linear control method, suggesting that the nonlinear control strategy may be more accurate than linear control methods. As shown in **Figure 2 (a1) and Figure 2 (b1),** in comparison to other directed network control methods, including linear control methods (Liu method and SCS method) and nonlinear Directed-networks FVS-based control method (DFVS), the nonlinear control method performed better for our analysis using NCUA in the SynLethDB network and MGIN network for BRCA dataset and in MGIN network and sPPI network for the LUSC dataset. One reason that NCUA is superior to other directed network control methods on most datasets is the dynamic of the constructed transition state gene network is undirected and nonlinear. In summary, predictions from NCUA are more accurate at identifying effective drug combinations, and is therefore an important factor for our designed PDC.

Next, to demonstrate the effect of network topological structure on our PDC, we used the SynLethDB network, MGIN network, and sPPI network. The results of our PDC with NCUA on the BRCA and LUSC datasets are also shown in **Figure 2 (a1)**

and **Figure 2 (b1).** The results showed that MGIN network, which contains more gene interaction information than the other two networks for BRCA and LUSC datasets, allowed for more accurate drug combination predictions in NCUA.

To further demonstrate the importance of prior gene interaction networks, we recalculated the precision of the predicted drug combinations by utilizing the differential expression network alone, as seen in Supplementary **Figure S2**. We observed that the precision of the differential expression network alone performed worse than the integration of the differential expression network and the gene interaction networks. **Figure S3** shows the results of the precision of balance parameters on different networks, where we can see that the parameters had different effects on different networks. In that, when the parameters were too large, it may have weakened the performance of NCUA. This again demonstrated that proper balence parameter of NCUA is a key factor affecting the precision of predicting drug combinations. Therefore, a map of prior gene interaction networks can indeed improve the discovery of efficient drug combinations and the choice of proper prior network structure is also an important factor for our PDC. From the data illustrated in Figure 2 (a2) and Figure 2(b2), we can see that the number of KCGs for our NCUA is not the minimum among all of the control methods, suggesting that the minimum number of input control nodes was not the unique objective for controlling the dynamic of the network.

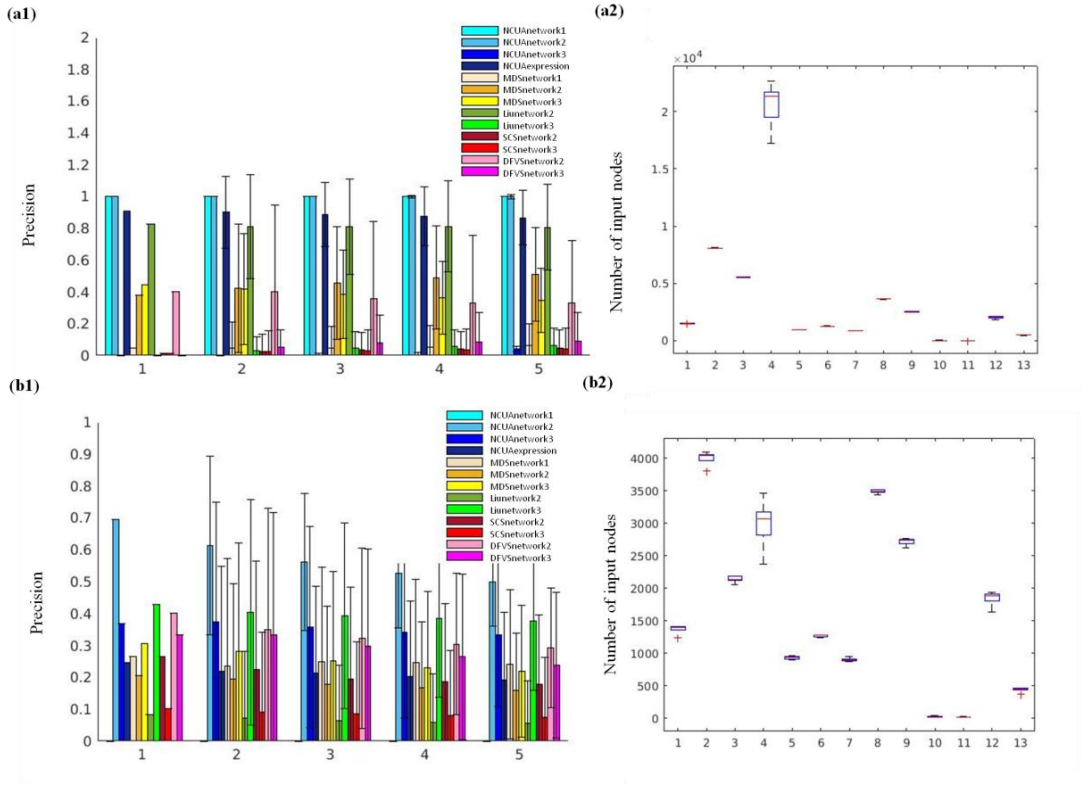

Figure 2. The results of PDC for prioritizing personalized anti-cancer drug combinations. (a1, b1) PDC with NCUA made accurate predictions for patient-specific drug combinations in comparison to other network control methods. The x-axis denotes results of the top k (k=1,2,3,…,5) predicted drug combinations. The colors represent the different methods in different networks, NCUA in SynLethDB network (NCUANetwork1), NCUA in MGIN network (NCUANetwork2), NCUA in sPPI network (NCUAnetwork3), NCUA in expression network without prior gene interaction network information (NCUAexpression), MDS in SynLethDB network and MGIN network and sPPI network (MDSNetwork1, MDSNetwork2, MDSnetwork3), Liu's linear control method in MGIN network and sPPI network (LiuNetwork2, LiuNetwork3), SCS method in MGIN network and sPPI network (SCSNetwork2, SCSNetwork3), and direced minimum FVS control method in MGIN network and sPPI network (DFVSNetwork2, DFVSNetwork3). The y-axis denotes the mean result of the fraction of the top *k* (*k*=1,2,3,…,5) predicted drug combinations within the clinical combinational drugs for treating BRCA and LUSC. (a2, b2) The x-axis represents the different methods (1-13: NCUANetwork1, NCUANetwork2, NCUANetwork3, NCUAexpression, MDSNetwork1, MDSNetwork2, MDSnetwork3, LiuNetwork2, LiuNetwork3, SCSNetwork2, SCSNetwork3, DFVSNetwork2, DFVSNetwork3), while the y-axis denotes the number of identified KCGs by the aforementioned 13 control methods for BRCA and LUSC.

Exploring functional and structural properties of personalized KCG

In this section, we illustrate and discuss the functional and structural property of personalized KCGs determined by PDC on MGIN (BRCA_result1) and SynLethDB network (BRCA_result2) for the BRCA dataset, and on MGIN (LUSC_result1) and sPPI (LUSC_result2) for the LUSC dataset (since these results have better performance in terms of combinational drugs prediction). The mutational level test,

expression level test, epistasis-interaction test, and the hub network index of KCGs were performed and compared with random gene sets.

The mutational level test evaluates whether the identified patient-specific KCGs are significantly enriched in the individual patient-specific mutational profiles (SNP mutation data). In **Figure 3**, we show the statistical significance of the mutational level test based on patient-specific KCGs. In some instances, the P-values were >0.05, particularly for BRCA_result2 and LUSC_result2, demonstrating that there was divergence in the mutational profiles of the patients. The differential expression level test (**Method**) evaluated whether the differential expression level of the personalized KCGs was significantly higher than the random gene sets. In **Figure 3**, we show the P-value of the differential expression level, and observed that in the different expression level tests, the p-values of most patients were <0.05 for BRCA_result1 and BRCA_result2, and LUSC_result1 and LUSC_result2. This suggests that the personalized KCGs tended to be the genes with insignificant differential expression levels. The experimental results of the mutational level test and the differential expression level test **demonstrate that** the properties of the personalizeds KCG are masked at the transcriptional level, as shown by the expression and mutational profiles.

To assess the network characteristics of KCGs in gene interaction networks, we calculated the ratio of the degree of KCGs to the whole network degree in MGIN networks as the hub network level of KCGs (**Method**). In **Figure 4 (a, b),** we show that the KCGs tended to be genes with high degrees. We also performed epistasis-interaction tests (**Method**) to see if the identified KCGs were epistatic. In that, we investigated whether the KCGs were significantly enriched in the MGIN, which contains more synergistic information (as shown in **Figure 2**), in comparison to random gene pairs. Based on the statistical significance of the epistasis-interaction tests shown in **Figure 3**, we observed that most P-values were <0.05 for the BRCA_result1 and BRCA_result2, and LUSC_result1 and LUSC_result2, which provided evidence that KCG interactions were significantly and functionally connected in MGINs for most of cancer patients. The results of hub network level and

epistasis-interaction tests showed that the KCGs were structurally enriched in hub genes and functionally connected in gene interaction networks for most of cancer patients, demonstrating that the properties of KCGs can be discovered by network level data.

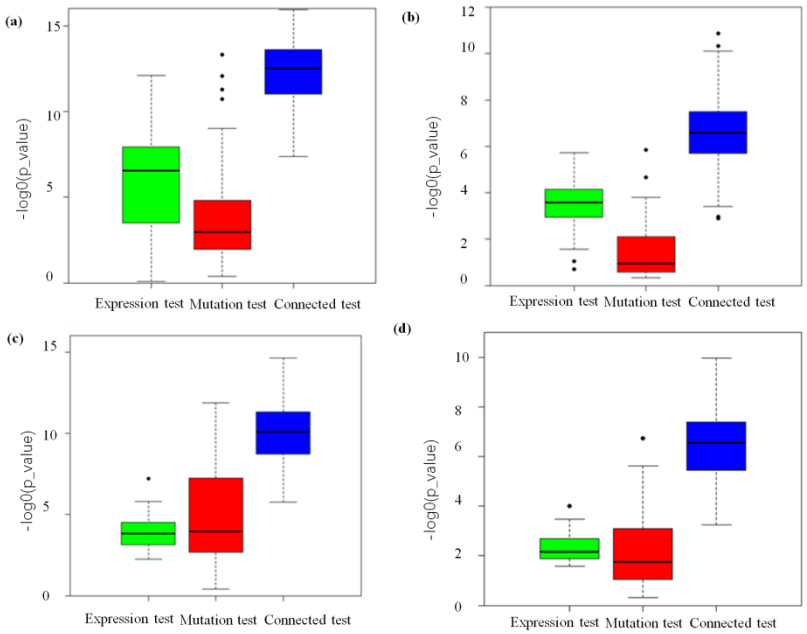

Figure 3 The functional relationships of the patient-specific KCGs. (a) The results of the mutational level, expression level, and epistasis-interaction tests for KCGs identified by PDC on MGIN (BRCA_result1) for BRCA dataset. (b) The results of the mutational level, expression level, and epistasis-interaction tests for KCGs identified by PDC on SynLethDB network (BRCA_result2) for BRCA dataset. (c) The results of the aforementioned tests of the KCGs identified by PDC on MGIN (LUSC_result1) for LUSC dataset. (d) The results of the aforementioned tests of the KCGs identified by PDC on sPPI (LUSC_result2) for LUSC dataset.

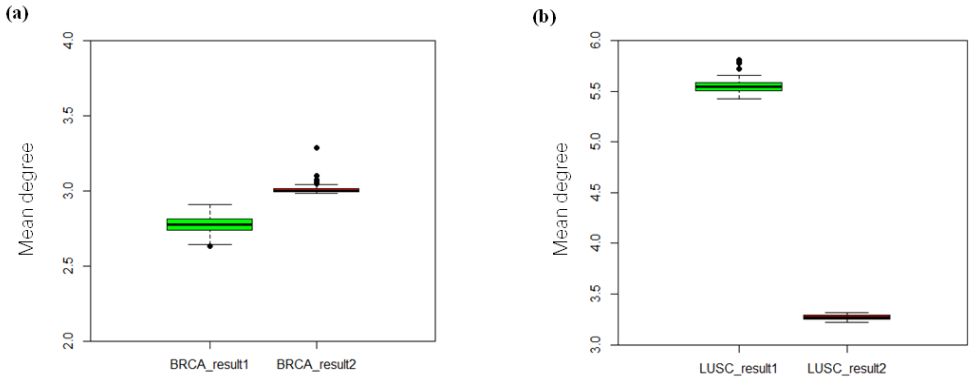

Figure 4 The distribution of the hub level of the patient-specific KCGs. (a) The distribution of the hub level (ratio of network degree of KCG to whole network degree) on SynLethDB network and MGIN network and sPPI network (BRCA_result1, BRCA_result2) for BRCA patients; (b) The results of the hub level test (ratio of network degree of KCG to whole network degree) on MGIN and sPPI network (LUSC_result1, LUSC_result2) for LUSC patients.

Assessing patient risk by side effect quantification on KCGs

Another important task in the context of multi-component therapies is to understand the joint effect of two or more drugs acting on the targets (Campillos et al. 2008). By targeting the KCGs (results on SynLethDB network) within the cancer census gene sets (Hou and Ma 2014; Jia and Zhao 2014; Bertrand et al. 2015), we calculated the number of drug pairs with aggravating effects (side effect score >0) and the number of drug pairs with enhancing effects (side effect score <0). Based on this data, we found that patients in the breast cancer dataset could be classified into three distinct subtypes, as shown in Figure 5 (a). In Figure S5, we also show the results of the comparative analysis of other schemes, including PDC results within cancer census gene sets on MGIN and sPPI network and PDC results alone. The results demonstrate that the SynLethDB network provided better patient classification.

We also considered the fraction of patient deaths as an index of significant risk assessment, as listed in **Figure 5(b)**. **Figure 5(a**) and **Figure 5(b)** demonstrate that the greater the number of side effects associated with drug pairs, the higher the risk to a particular patient subtype. As shown in **Figure 5(b)**, clinical disease stage (I, II, and III) was not associated with significant risk assessment. Furthermore in **Figure 5(c)**, we give the results of patient stratification and prognostication based on our identified subtypes, which support that the side effect signature of drug combinations may enable patient stratification and prognostication with significant risk assessment.

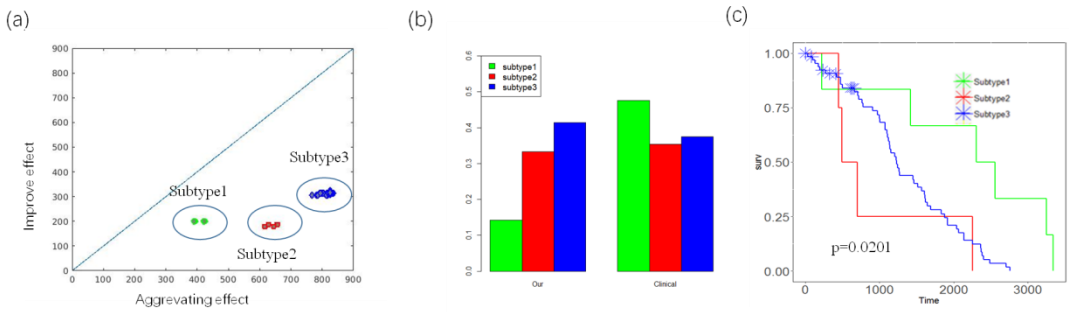

Figure 5 The results of the patient risk assessment by side effect quantification of the personalized KCGs. (a) The classification of patients in the breast cancer dataset by the number of drug pairs with an aggravating effect and an enhancing effect; (b) The comparison results of the fraction of dead patients (used as an index of significant risk assessment) based on our identified subtypes and clinically subtypes; (c) The results of patient stratification and prognostication based on our identified subtypes.

To obtain a sufficiently comprehensive list of breast cancer-related genes, 2341 breast cancer-related genes were identified by using the Unified Medical Language System (UMLS) (Aronson 2001; Liu et al. 2014a; Quan et al. 2018). We identified the KCGs within this gene set as the key control breast cancer genes. Next, we applied the enrichment analysis of the personalized breast cancer KCGs in KEGG pathways to identify potential disease-related pathways. The full list of results is provided in **Table 1**.

For each patient, we performed an enrichment analysis to determine if the breast cancer KCGs were enriched in certain biological pathways. Of the 15 identified biological pathways, 10 of them (66.7%) have been previously reported as breast cancer-related pathways. This suggests that our PDC will be very useful in identifying potential breast cancer-related pathways. Interestingly, we found that although most of the previously reported breast cancer-related pathways were enriched in the pathways with high frequency ($f>0.6$), some pathways with low frequency ($f<0.3$) have still been reported by other publications. **Table 1** also gives the frequency of enriched pathways in different subtypes. We observed that there were some novel enriched pathways in the high risk subtypes (subtype 2 and subtype 3), including the Prolactin signaling pathway and Toll-like receptor signaling pathway. The results demonstrated that the breast cancer-related pathways will be dominant in different patients based on the patient-specific KCGs. Additionally, **Table 1** provides the results of the hub network level, differential expression level, and side effect level tests of the KCGs related to the identified pathways. We found that: i) the p-value of related pathway KCGs in the differential expression level test was significant, demonstrating that the identified enrichment pathways were active in cancer development; ii) the related pathway KCGs were also structurally enriched in hub genes; and iii) the side effect level value of high risk pathways were higher, demonstrating the validity of our predicted high risk pathways.

Table 1. The information of disease-related pathways identified in breast cancer. The first column is the identified disease-related pathways from the personalized genetic-control breast cancer gene interactions. The second column is the frequency for all samples. The third, fourth, and fifth columns are the frequency of subtypes 1, 2, and 3, respectively. The sixth, seventh, eighth columns are the mean values of hub network level (the ratio value in the hub network level), differential expression level (p-value in the differential expressed level test, the alternative hypothesis (H1) is $n_i > n_0$), and the side-effect level tests (the ratio of the number of drug pairs with an enhancing effect to the number of drug pairs with an aggravating effect) of the KCGs related to the identified pathways (intersected with breast cancer genes). Note that all the results in the sixth, seventh, and eighth columns are the mean values of all patient in the BRCA dataset.

Pathway_name	Frequency	Subtype1	Subtype2	Subtype3	Hubs	Expression	Side-effect
Adipocytokine signaling pathway(Rose et al. 2004)	0.4375	0.4285	0.6666	0.4242	10.622	0.0282	40.5714
Choline metabolism in cancer(Xu et al. 2008)	1	1	1	1	15.084	0.0330	83.5357
Endometrial cancer(Fisher et al. 1994)	1	1	1	1	16.113	0.0319	56.4286
Epstein-Barr virus infection(Bonnet et al. 1999)	1	1	1	1	15.167	0.0280	5.2143
ErbB signaling pathway(Kurokawa and Arteaga 2003)	1	1	1	1	16.14	0.0287	76.4286
Fc epsilon RI signaling pathway	0.7678	0.7142	1	0.7575	13.968	0.0326	3.9286
HIF-1 signaling pathway	0.9642	1	1	0.9595	6.8579	0.0326	85.7143
p53 signaling pathway(Sheikh et al. 1994)	1	1	1	1	12.054	0.0305	1.3036
Progesterone-mediated oocyte maturation	1	1	1	1	16.895	0.0295	14.4286
Prolactin signaling pathway(Llovera et al. 2000)	0.4285	0	0.5	0.4545	15.091	0.0313	255.1786
TNF signaling pathway	1	1	1	1	13.193	0.0294	37.4286
Apoptosis(Tanaka et al. 2000)	0.1517	0.2857	0.3333	0.1313	12.631	0.0292	34.6786
Herpes simplex infection(Toda et al. 1998)	0.5892	0.4285	0.8333	0.5858	13.863	0.0260	36.4643
NOD-like receptor signaling pathway	0.1607	0.4285	0.1666	0.1414	10.622	0.02826	7.2143
Toll-like receptor signaling pathway(Cai et al. 2011)	0.0982	0	0.1667	0.1010	15.084	0.0330	37.4286

Discussion

As drug combinations become a mainstay in cancer therapy, integrative analytical approaches are being increasingly incorporated into cancer genomics. It is already recognized that cancer patients are not all the same and unique drug combinations may be relevant for individual patients or a specific subset of patients (Schork 2015). Tumor heterogeneity in cancer patients has been a major obstacle in understanding cancer and in applying precision medicine to the treatment of cancer. Many conventional methods used to study cancer heterogeneity require large sample sizes and focus on statistical analysis at the population level. Therefore, it remains challenging to predict personalized drug combinations based on patient-specific data rather than on patient cohort data. As a result, new analytical tools are needed to address this problem.

An increasing number of studies have demonstrated that targeting key disease genes can provide critical information for drug discovery and drug repurposing (Phillips 2008; Han et al. 2017; Quan et al. 2018; Yip et al. 2018). Coincidentally, network control methods mainly focus on how to identify key nodes to control the whole network state from the initial state to the desired state. In this study, we aimed to develop techniques that bridge the connections between network control theory and the identification of personalized key disease genes (or KCGs) related to the transition between normal state and disease state. In fact, researchers have studied how to control biological network dynamics for more than a decade, which therefore can provide an effective model for drug-induced network changes and prediction of personalized drug combinations (Li et al. 2018). Recently, we began to focus on how to study the control problem of personalized patient systems, and introduced a tool called Single Sample Controller (Guo et al. 2018a), which applies network control to a single patient ecosystem to find driver genes related to the transition between normal and tumor state. Despite intense efforts to develop a scalable and accurate key control gene identification algorithm (Guo et al. 2018a), this SCS algorithm can only use the linear control techniques, which contradict the dynamics of biological

networks, particularly for a patient-specific system. Currently, we still lack a rigorous mathematical controllability model for nonlinear dynamic personalized patient system due to the unavailability of personalized data with small sample numbers but large dimension genes. Here, we considered the issue of identification of personalized drug combinations (a typical control problem on nonlinear biological networks in cancer genomics), and proposed a novel and efficient framework, called PDC (software freely avalable at http://sysbio.sibcb.ac.cn/cb/chenlab/software.htm) to fill the gap between network control theory and biological network control problem. The key consideration of PDC is to construct proper transition network between any two attractors of biological networks. In that, we can push a complex, nonlinear individual system in cancer from attractor (tumor) to attractor (normal) through by perturbing a feasible subset of network nodes (e.g., drug-target genes). In this study, we present a personalized drug controller method (PDC), which revealed the transition mechanism from a cancerous state to a normal state, as well as the cancer ecosystem of each patient at an individual level. Technically, PDC considers how to construct the personalized network capturing the phenotypic transitions between normal and disease states. This is accomplished by utilizing a paired SSN method to obtain the personalized transition network and applying our nonlinear structural control algorithm, called NCUA, based on feedback vertex sets to find the optimal patient-specific KCGs based on phenotype transitions, which can control the individuals from normal state to disease state, or vice versa. We applied PDC to BRCA and LUSC datasets from TCGA, whose validation results suggested that PDC with NCUA can identify known drug combinations for a single patient more efficiently than other existing network control methods. Our PDC approach also enabled patient stratification and prognostication with significant risk assessment. Therefore, this study has provided novel insights into tumor heterogeneity in complex patient ecosystems.

Methods

The framework of PDC

Constructing personalized transition networks by using a paired-SSN method

First, the co-expression network of the tumor sample network or normal sample network for each patient was constructed based on statistical perturbation analysis of one sample against a group of given reference samples (e.g., choosing the normal sample data of all of the patients as the reference data) with SSN method (Liu et al. 2016). If the differential PCC (ΔPCC) of an edge was statistically significantly large based on the evaluation of our SSN method, this edge would be kept for the single sample (normal sample or tumor sample) from each patient (**Figure 1 and Figure S1**). The ΔPCC of an edge between gene *i* and *j* and its significant z-score were calculated by using the following formulae:

$$\begin{aligned} \Delta PCC_{ij} &= \left| PCC_{ij}^{n+1} - PCC_{n}^{n} \right| \\ Z &= \Delta PCC_{ij} / ((1-(PCC_{ij}^{n})^{2}) / (n-1)) \end{aligned}, \quad (1)$$

where PCC_{ij}^{n} is the Pearson correlation coefficient (PCC) of an edge between gene *i* and *j* in the reference network with *n* samples; PCC_{ij}^{n+1} is the Pearson correlation coefficient of the edge between gene *i* and *j* in the perturbed network with one additional sample, given a new sample (normal sample or tumor sample) for each patient being added to the reference group. In addition, the P-value for an edge could also be obtained from the statistical Z-value. All of the edges with significant differential correlations (e.g., p-value<0.05) were used to constitute the SSN for the normal sample or tumor sample.

We then constructed the personalized differential co-expression network between normal sample network and tumor sample network where the edge between gene *i* and gene *j* existed if the p-value of the edge was less than (greater than) 0.05 in the tumor network but greater than (less than) 0.05 in the normal network for their corresponding patient. Additionally, we defined the differential log2 fold-change

between the tumor network edge and the normal network edge as the weight of the reserve edge by using the following formula:

$$e_{ij} = \left| \log_2 \left(\left| \frac{\Delta PCC_{ij}^{Tumor}}{\Delta PCC_{ij}^{Normal}} \right| \right) \right| . \tag{2}$$

Finally, we obtained personalized transition state gene networks where edges existed in both gene interaction network and the differential co-expression network for each patient (**Figure 1 and Figure S1 (a)**). By integrating gene interaction networks and personalized differential co-expression networks, personalized transition state gene networks couljd reveal which gene pairs were involved in the disease development for each patient. A personalized transition state gene network is an undirected graph where nodes represent genes, and a pair of nodes is connected by an edge if there is a significant expression correlation difference for the gene pairs between the tumor network and normal network. In addition, the edge weight was quantified by the differential fold change of gene pairs between normal sample and tumor sample for a patient, and the node weight is a confidence score for each gene denoting the regulatory impact on the state transition as calculated with the following formula:

$$w_i = \sum_{j \in N(i)} e_{ij} , \tag{3}$$

where $N(i)$ represents the neighborhood nodes set of node i in the gene state transition network.

Identifying personalized KCGs based on phenotypic transitions by NCUA method

Structural controllability analysis of network (Lin 1974) has become an important branch of network control. Particularly for biological complex networks, structural control has been widely applied and discovered many interesting properties of biological systems. However, the existing control methods (Liu et al. 2011; Gao et al. 2014; Wu et al. 2014b; Guo et al. 2017; Guo et al. 2018d) cannot be directly applied

to the above constructed personalized transition state gene networks with a nonlinear and undirected dynamic because they are focused on linear dynamic directed networks. Therefore, an effective nonlinear network control strategy is required to characterize personalized transition state gene networks, and support selecting specific KCGs based on phenotypic changes.

Theoretically, the following typical dynamic ODE can be used to represent the dynamic behavior of a patient-specific system:

$$dx_k^i / dt = f(x_k^i, x_{I_k}^i), \tag{4}$$

where x_k^i denotes the expression state of gene k for the patient I, I_k represents the neighborhood gene set of gene k, $f()$ represents the dynamic behavior control law of the patient-specific network for patient i, taking into account continuous differentiability, dissipativity, and decay condition (Fiedler et al. 2013a; Mochizuki et al. 2013). This dynamic equation represents the dynamic behavior of the gene expression level in the patient-specific network for patient i. In this work, we were interested in how to change the gene expression level from the disease state to normal state by using the optimal KCG set under this nonlinear dynamic behavior condition.

In this paper, a novel nonlinear control algorithm of undirected networks (NCUA) was designed based on structural control theory with Feedback vertex set (FVS) to ensure that the state of a system can be determined by the state of genes under nonlinear dynamic laws. The definition of FVS is a subset of nodes in the graph, such that the removal of the set leaves the graph without cycles. Therefore, the impetus behind our NCUA was to define an index to represent the quality of the KCG and apply a linear integer program to identify the optimal FVS with maximum quality as the KCG set. **Figure S1 (b)** illustrates the process of our NCUA for discovering the optimal KCG set. The details of NCUA are provided below.

1) Constructing dominant gene sets based on the FVS structure control theory

In a directed network, if ***I*** is used to represent the Feedback Vertex Set (FVS), then the state of the entire network can be determined by the state of the feedback

node set *I*, called FVS structural control theory (Fiedler et al. 2013a; Mochizuki et al. 2013). For a personalized transition state network, we assume that each edge is bi-directional and converts the personalized transition state gene network into a bipartite network where the nodes and the edges of the original network act as the up and bottom nodes of the bipartite network. For the bipartite graph, if the node v_i in the up side is one of nodes for edge v_j in the bottom side, v_i and v_j are linked in the bipartite network. Based on the structure control theory of FVS, the state of the dominant set *M* in the up side of the bipartite network that can cover the low side of node can determine the state of all nodes in the network.

2) Selecting optimal KCGs from candidate dominant gene sets by using linear integer programming

It is known that different dominant gene sets will have different controlling costs on patient-specific networks. Therefore, we introduced an index (e.g., cost function) to indicate the quality of the selected KCG**:**

$$W(M) = \sum w_i r_i - \lambda \sum r_i \,, \tag{5}$$

where w_i denotes the confidence score of gene *i* by summarizing the scores of its connected edges in a patient-specific network by using formula (4); r_i is an indicative variable, when gene *i* in the up side node set in the bipartite network is within the KCG**,** $r_i = 1$, otherwise $r_i = 0$; $\sum w_i r_i$ denotes the whole score of candidate gene sets; $\sum r_i$ denotes the number of candidate genes; and λ is the balance parameter to regulate the whole score of candidate KCG set and the number of candidate KCGs.

For the control problem modeled by PDC, we expected that the selected set of KCGs would not only contain the minimum gene number, but would also have the maximum differential fold change scores of genes. Therefore, it could be used to

further measure the quality of the candidate KCG set, and we selected the optimal KCGs by solving the following linear integer programming:

$$\begin{aligned} &\max W(M) = \sum w_i r_i - \lambda \sum r_i \\ &s.t. \sum_{i \in N(u)} r_i \geq 1 \left(\text{every } u \in V_L\right), r_i = \{0,1\} \end{aligned} \quad , \tag{6}$$

where V_L denotes the node set of the low side in the (converted) bipartite graph and N_u denotes the neighborhood node set in the bipartite graph; the objective function is to obtain the KCG set with the minimum number but the maximum differential fold change scores. The restriction condition is to ensure that all the edges of the original network in the bipartite network can be covered. Thus, under the nonlinear dynamic behavior, the state of all of the nodes (genes) in a patient-specific network can be regulated by the KCG. The above optimization problems can be effectively solved by the classical LP-based classic branch and bound method (Lenstra 1983; Williams 1990).

Note that the nonlinear control problem of undirected networks (NCUA) is firstly proposed in our recent new paper (Guo et al. 2018b). This new version NCUA used here focus on the optinal key control nodes set (KCGs) while the first version focus on the minimum number of key control nodes set and the multiple control configurations in controlling the undirected nonlinear networks. From this perspective, this version NCUA is more suitable for network control problems in both theory and application.

3) Prioritizing personalized anti-cancer drug combinations

In this work, the impact score of a given drug combination was calculated based on the number of targeting personalized KCGs for each cancer patient using nonlinear structural control method based on feedback vertex sets. We ranked and prioritized potential personalized drug combinations based on their impact on personalized KCGs.

Statistical test of the functional and structural properties of personalized KCGs

(1) **Mutational level test.** This test examined whether personalized KCGs were significantly enriched in patient-specific mutational profiles. We used the following test to evaluate the mutational level (SNP mutation data) of KCGs. The random data sets with the same number of disease genes were first obtained and we then computed z-score as:

$$z_i = (n_i - mean(SD_i)) / (std(SD_i)), \quad (7)$$

where n_i is the mutational level (the intersected gene number of disease genes and genetic mutations) of KCGs for patient i. SD_i is the distribution of the mutational level of generated random genes with the same number of disease genes. Mean and standard deviation of SD_i were computed from 100 simulations of random data sets. Based on the z-score, we obtained the empirical P-value, i.e., P_i (modeled as Gaussian distribution). The significant P-values for the mutational level were obtained for each patient, where the alternative hypothesis (H1) is $n_i > n_0$, and n_0 is mean value of the mutational level of generated random genes with the same number of disease genes.

(2) **Differential expression level test.** This test determined if the differential expression level of personalized KCGs were significantly higher than the random gene sets. We used the test outlined above to evaluate the differential expression level of the personalized genes. In the calculation of z-score, n_i is the differential expression level (the absolute value of log2 fold change between normal expression data and tumor expression data) of KCGs for patient i. SD_i is the distribution of the differential expression gene level of generated random genes with the same number of KCGs. Mean and standard deviation of SD_i was computed from 100 simulations of random data sets. For the differential expression level test, the alternative hypothesis (H1) is $n_i < n_0$, where n_0 is mean

value of the differential expression gene level of generated random genes with the same number of disease genes.

(3) **Hub network level.** To assess the network characteristics of KCGs in gene interaction networks, we calculated the ratio of degree of KCGs to the whole network degree in MGIN as the hub network level of KCGs.

(4) **Epistasis-interaction test.** Recently many studies have proposed that targeting interacting disease-related genes (epistatic genes) may bring synergistic effects for disease control, and drug combinations aimed at epistatic genes may have higher medicinal potential (Phillips 2008; Han et al. 2017; Quan et al. 2018; Yip et al. 2018). For the epistasis-interaction test, we calculated the number of edges among the disease genes. We then generated 100 random networks, each of which maintained the topological characteristics of the original network (degree of each node) and recalculated the number of edges among the disease genes. We then used a one-tailed Fisher exact test to obtain the p-value for evaluating whether the KCGs were significantly more connected than the random genes.

Quantification of side effect scores on targets

In this section, we introduce the method for calculating the *side effect score* of a given drug pair on the drug-target network (Torres and Altafini 2016). The classification of drug actions into positive and negative modes of action allowed us to characterize the effect of multiple drugs acting on the targets. Two drugs sharing the same target tended to reinforce their effect on the target if their modes of action were similar, while they tended to mitigate the overall effect if they had opposing modes of action. The three possible combinations of a drug pair effect on a common target (+,+), (−,−), and (+,−) are shown in **Figure S4**. The two combinations (+,+) and (−,−) will be referred to as coherent, as the action of one drug was reinforced by the presence of a second drug. In the remaining combination (+,−), the presence of a second drug counteracted (and hence in general reduced) the action of the first drug. This configuration will be called incoherent. If all drug-target interactions can be classified

as *pharmacological* (on-target actions) and *non-pharmacological* (off-target actions) by attaching signs to the mechanisms of action(Torres and Altafini 2016), the *side effect score* can be calculated by using the following formula:

$$b_{ij} = \mathrm{sign}(s_{ij}^{+-} - s_{ij}^{++} - s_{ij}^{--}), \tag{8}$$

where s_{ij} is the number of common targets of the drug pair (i, j), which are not pharmacological targets in DrugBank for at least one of the two drugs; $s_{ij}^{++}, s_{ij}^{++}, s_{ij}^{--}$ specify theirs signs.

Availability of data and materials

All of the data resources used in this study, including the cancer benchmark datasets and the reference network datasets , are provided for free at http://sysbio.sibcb.ac.cn/cb/chenlab/software.htm. The source code of PDC with NCUA and other network control methods, are also provided for free at http://sysbio.sibcb.ac.cn/cb/chenlab/software.htm.

Acknowledgements

This paper was supported by the National Natural Science Foundation of China (61873202, 61473232, 91430111, 91439103, 91529303, 31771476, 81471047 and 11871456) and National Key R&D Program (2017YFA0505500), Strategic Priority Re-search Program of the Chinese Academy of Sciences (No. XDB13040700), National Key R&D Program (Special Project on Precision Medicine) (2016YFC0903400), and Natural Science Foundation of Shanghai (17ZR1446100). We thank Professor Jianxi Gao from Rensselaer Polytechnic Institute, USA and Professor Tatsuya Akutsu from Kyoto University, Japan for giving our useful and valuable comments. We also thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Disclosure declaration

The authors declare that they have no competing interests.

References

Aronson AR. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proceedings / AMIA Annual Symposium AMIA Symposium* **2001**(1): 17.

Bertrand D, Chng KR, Sherbaf FG, Kiesel A, Chia BK, Sia YY, Huang SK, Hoon DS, Liu ET, Hillmer A. 2015. Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. *Nucleic acids research* **43**(7): e44-e44.

Bonnet M, Guinebretiere J-M, Kremmer E, Grunewald V, Benhamou E, Contesso G, Joab I. 1999. Detection of Epstein-Barr virus in invasive breast cancers. *Journal of the National Cancer Institute* **91**(16): 1376-1381.

Bush SW, Dudek MS, Ritchie DM. 2006. Parallel multifactor dimensionality reduction: a tool for the large-scale analysis of gene-gene interactions. *Bioinformatics* **22**(17): 2173.

Cai Z, Sanchez A, Shi Z, Zhang T, Liu M, Zhang D. 2011. Activation of Toll-like receptor 5 on breast cancer cells by flagellin suppresses cell proliferation and tumor growth. *Cancer research*.

Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. 2008. Drug target identification using side-effect similarity. *Science* **321**(5886): 263-266.

Chakradhar ST, Balakrishnan A, Agrawal VD. 1995. An exact algorithm for selecting partial scan flip-flops. *Journal of Electronic Testing* **7**(1-2): 83-93.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience,4,1(2015-02-25)* **4**(1): 7.

Ciriello G, Cerami E, Sander C, Schultz N. 2012. Mutual exclusivity analysis identifies oncogenic network modules. *Genome research* **22**(2): 398-406.

Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B. 2010. Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research*: gkq1018.

Fiedler B, Kurosawa G, Saito D. 2013a. Dynamics and Control at Feedback Vertex Sets. I: Informative and Determining Nodes in Regulatory Networks. *Journal of Dynamics & Differential Equations* **25**(3): 563-604.

Fiedler B, Mochizuki A, Kurosawa G, Saito D. 2013b. Dynamics and control at feedback vertex sets. I: Informative and determining nodes in regulatory networks. *Journal of Dynamics and Differential Equations* **25**(3): 563-604.

Fisher B, Costantino JP, Redmond CK, Fisher ER, Wickerham DL, Cronin WM, contributors N. 1994. Endometrial cancer in tamoxifen-treated breast cancer patients: findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *JNCI: Journal of the National Cancer Institute* **86**(7): 527-537.

Gao J, Liu Y-Y, D'Souza RM, Barabási A-L. 2014. Target control of complex networks. *Nature communications* **5**.

Greaves M, Maley CC. 2012. Clonal evolution in cancer. *Nature* **481**(7381): 306.

Guo W-F, Zhang S-W, Liu L-L, Liu F, Shi Q-Q, Zhang L, Tang Y, Zeng T, Chen L. 2018a.

Discovering personalized driver mutation profiles of single samples in cancer by network control strategy. *Bioinformatics* **34**(11): 1893-1903.

Guo W-F, Zhang S-W, Zeng T, Li Y, Gao J, Chen L. 2018b. A novel structure-based control method for analyzing nonlinear dynamics in biological networks. *bioRxiv*: 503565.

Guo WF, Zhang SW, Liu LL, Liu F, Shi QQ, Zhang L, Tang Y, Zeng T, Chen L. 2018c. Discovering personalized driver mutation profiles of single samples in cancer by network control strategy. *Bioinformatics* **34**(11): 1893.

Guo WF, Zhang SW, Shi QQ, Zhang CM, Zeng T, Chen L. 2018d. A novel algorithm for finding optimal driver nodes to target control complex networks and its applications for drug targets identification. *BMC genomics* **19**(Suppl 1): 924.

Guo WF, Zhang SW, Wei ZG, Zeng T, Liu F, Zhang J, Wu FX, Chen L. 2017. Constrained target controllability of complex networks. *Journal of Statistical Mechanics Theory & Experiment* **6**(6): 063402.

Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC. 2017. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nature Biotechnology* **35**(5): 463.

Hong Y, Chu Q, Hong LY, Lin T, Jin Z, Yan YC, Feng X, Zhe C, Feng Z, Zong CY. 2016. Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Research* **44**(Database issue): D1069-D1074.

Hou JP, Ma J. 2014. DawnRank: discovering personalized driver genes in cancer. *Genome medicine* **6**(7): 56.

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A. 2007. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics* **39**(7): 870.

Jgt Z, Yang G, Albert R. 2017. Structure-based control of complex networks with nonlinear dynamics. *Proceedings of the National Academy of Sciences of the United States of America*: 7234-7239.

Jia J, Zhu F, Ma X, Cao ZW, Li YX, Chen YZ. 2009. Mechanisms of drug combinations: interaction and network perspectives. *Nature Reviews Drug Discovery* **8**(2): 111-128.

Jia P, Zhao Z. 2014. VarWalker: personalized mutation network analysis of putative cancer genes from next-generation sequencing data. *PLoS computational biology* **10**(2): e1003460.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2011. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*: gkr988.

Kurokawa H, Arteaga CL. 2003. ErbB (HER) receptors can abrogate antiestrogen action in human breast cancer by multiple signaling mechanisms. *Clinical cancer research* **9**(1): 511s-515s.

Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V. 2014. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research* **42**(Database issue): 1091-1097.

Lenstra HW. 1983. Integer Programming with a Fixed Number of Variables. *Mathematics of Operations Research* **8**(4): 538-548.

Li M, Gao H, Wang J, Wu F-X. 2018. Control principles for complex biological networksLi et al.Control principles for biological networks. *Briefings in Bioinformatics*: bby088-bby088.

Lin CT. 1974. Structural controllability. *Automatic Control, IEEE Transactions on* **19**(3): 201-208.

Ling SY, Yang C, Xiang W, Yu W. 2011. GBOOST: a GPU-based tool for detecting gene–gene interactions in genome–wide case control studies. *Bioinformatics* **27**(9): 1309.

Liu CC, Tseng YT, Li W, Wu CY, Mayzus I, Rzhetsky A, Sun F, Waterman M, Chen JJ, Chaudhary PM. 2014a. DiseaseConnect: a comprehensive web server for mechanism-based disease-disease connections. *Nucleic Acids Research* **42**(Web Server issue): 137-146.

Liu X, Wang Y, Ji H, Aihara K, Chen L. 2016. Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Research* **44**(22): e164-e164.

Liu X, Wang Y, Kelly R, Romdhane R. 2010. AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm. *Bmc Research Notes* **3**(1): 117.

Liu Y-Y, Slotine J-J, Barabási A-L. 2011. Controllability of complex networks. *Nature* **473**(7346): 167.

Liu Y, Wei Q, Yu G, Gai W, Li Y, Chen X. 2014b. DCDB 2.0: a major update of the drug combination database. *Database* **2014**: bau124.

Llovera M, Pichard C, Bernichtein S, Jeay S, Touraine P, Kelly PA, Goffin V. 2000. Human prolactin (hPRL) antagonists inhibit hPRL-activated signaling pathways involved in breast cancer cell proliferation. *Oncogene* **19**(41): 4695.

Merlo LM, Pepper JW, Reid BJ, Maley CC. 2006. Cancer as an evolutionary and ecological process. *Nature Reviews Cancer* **6**(12): 924.

Mochizuki A, Fiedler B, Kurosawa G, Saito D. 2013. Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. *Journal of Theoretical Biology* **335**(4): 130-146.

Nacher JC, Akutsu T. 2012. Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. *New Journal of Physics* **14**(7): 73005-73028(73024).

Phillips PC. 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* **9**(11): 855.

Quan Y, Liu MY, Liu YM, Zhu LD, Wu YS, Luo ZH, Zhang XZ, Xu SZ, Yang QY, Zhang HY. 2018. Facilitating Anti-Cancer Combinatorial Drug Discovery by Targeting Epistatic Disease Genes. *Molecules* **23**(4): 736.

Rose D, Komninou D, Stephenson G. 2004. Obesity, adipocytokines, and insulin resistance in breast cancer. *Obesity reviews* **5**(3): 153-165.

Ruths J, Ruths D. 2014. Control profiles of complex networks. *Science* **343**(6177): 1373-1376.

Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. 2009. PID: the pathway interaction database. *Nucleic acids research* **37**(suppl 1): D674-D679.

Schork NJ. 2015. Personalized medicine: time for one-person trials.

Sheikh MS, Li X-S, Chen J-C, Shao Z-M, Ordonez JV, Fontana JA. 1994. Mechanisms of regulation of WAF1/Cip1 gene expression in human breast carcinoma: role of p53-dependent and independent signal transduction pathways. *Oncogene* **9**(12): 3407-3415.

Sun X, Vilar S, Tatonetti NP. 2013. High-throughput methods for combinatorial drug discovery. *Science Translational Medicine* **5**(205): 205rv201.

Tanaka K, Iwamoto S, Gon G, Nohara T, Iwamoto M, Tanigawa N. 2000. Expression of survivin and its relationship to loss of apoptosis in breast carcinomas. *Clinical Cancer Research* **6**(1): 127-134.

Toda M, Rabkin SD, Martuza RL. 1998. Treatment of human breast cancer in a brain metastatic model by G207, a replication-competent multimutated herpes simplex virus 1. *Human gene therapy*

9(15): 2177-2185.

Torres NB, Altafini C. 2016. Drug combinatorics and side effect estimation on the signed human drug-target network. *BMC systems biology* **10**(1): 74.

van der Wijst MG, de Vries DH, Brugge H, Westra H-J, Franke L. 2018. An integrative approach for building personalized gene regulatory networks for precision medicine. *Genome Medicine* **10**(1): 96.

Vinayagam A, Stelzl U, Foulle R, Plassmann S, Zenkner M, Timm J, Assmus HE, Andradenavarro MA, Wanker EE. 2011. A directed protein interaction network for investigating intracellular signal transduction. *Science Signaling* **4**(189): rs8.

Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, Krysiak K, Deng P, Mcmichael JF, Eldred JM. 2016. DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Research* **44**(Database issue): D1036-D1044.

Wan X, Yang C, Yang Q, Xue H, Tang NLS, Yu W. 2010. Predictive rule inference for epistatic interaction detection in genome-wide association studies. *Bioinformatics* **26**(1): 30-37.

Williams HP. 1990. Integer and Combinatorial Optimization. *Journal of the Operational Research Society* **41**(2): 177-178.

Wright MN, Ziegler A. 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* **077**(1).

Wu F-X, Wu L, Wang J, Liu J, Chen L. 2014a. Transittability of complex networks and its applications to regulatory biomolecular networks. *Scientific reports* **4**.

Wu FX, Wu L, Wang J, Liu J, Chen L. 2014b. Transittability of complex networks and its applications to regulatory biomolecular networks. *Scientific Reports* **4**(4): 4819.

Wu G, Feng X, Stein L. 2010. A human functional protein interaction network and its application to cancer data analysis. *Genome biology* **11**(5): R53.

Xu X, Gammon MD, Zeisel SH, Lee YL, Wetmur JG, Teitelbaum SL, Bradshaw PT, Neugut AI, Santella RM, Chen J. 2008. Choline metabolism and risk of breast cancer in a population-based study. *The FASEB Journal* **22**(6): 2045-2052.

Yip DK, Chan LL, Pang IK, Jiang W, Tang N, Yu W, Yip KY. 2018. A network approach to exploring the functional basis of gene-gene epistatic interactions in disease susceptibility. *Bioinformatics* **34**(10).

Zhang Y. 2012. A novel bayesian graphical model for genome-wide multi-SNP association mapping. *Genetic Epidemiology* **36**(1): 36-47.