The Third International Symposium on Optimization and Systems Biology (OSB'09)
Zhangjiajie, China, September 20–22, 2009

A Network-Based Approach for Identifying Effective Drug Combination

Zikai Wu[1,2] Luonan Chen[1,3,*]

[1]Institute of Systems Biology,
Shanghai University, Shanghai 200444, China
[2]School of Communication and Information Engineering
Shanghai University, Shanghai 200444, China
[3]Department of Electrical Engineering and Electronics
Osaka Sangyo University, Osaka 574-8530, Japan

Abstract To cope with complex diseases, combination regimens or combination drugs are increasingly attracting the attention from the area of drug design. However, most of the combination drugs are developed based on clinical experience or by test-and-trial strategy. In this paper, we present a novel network-based systems biology approach to identify effective drug combinations by exploiting high throughput data. As a pilot study, we apply the proposed method to identify combination drug used to treat Type 2 Diabetes. The results on real biological data demonstrate the effectiveness and efficiency of the proposed method.

Keywords Drug combination; molecular interaction network; active subnetwork; gene expression

1 Introduction

It is a long history to use combination drugs for treating diseases and reducing suffering. Now, combination drug is becoming the standard of care for many complex diseases. As a result, some methods have been proposed to identify effective drug combinations. These methods can be grouped into two classes, i.e. computation (or *in silico*) based methods and experiment (or *in vivo*) based methods.

Dose-response curve is the common basis for computation based methods to identify effective drug combinations. With dose-response curve in hand, one can define the null model to describe the relationship between dose and response of the combination drug whose members have no interaction. Finally, based on the comparison between the predicted dose-response curve of the null model and a real dose-response curve of combination drugs, synergism, additive and antagonism between drugs can be defined accordingly. Under the assumption that two inhibitors acting on a target through similar mechanisms, Loewe proposed an additivity model to predict the combined effect of two inhibitors[5]. By assuming two inhibitors acting through independent mechanisms, Bliss proposed another null model to define combined effect of two inhibitors[6]. Based on

*Corresponding author. E-mail: chen@eic.osaka-sandai.ac.jp, xmzhao@shu.edu.cn

mass action law, Chou and Talalay unified all existing models and proposed a general model to define combined effect of multiple drugs [4].

On the other hand, some important combination drugs have been discovered by experiments. For example, Agrawal et al. found a combination drug for treating Huntington disease based on experiments in Drosophila[8]. There are also many papers describing clinical rules about how to combine 2-3 drugs[9]. Recently, a high-throughput screening method was recently proposed to identify effective combinations of therapeutic compounds[11]. In [10, 12], two stochastic search algorithms were developed to identify effective combinations of drugs. In these two methods, biological response information such as the percentage of cancerous cells being killed was utilized to detect an appropriate solution, i.e. proper dose of each drug.

Although above mentioned methods identified many effective drug combinations, there still exists much room to be improved. For example, the methods mentioned above except clinical experience based methods did not consider side effect of drugs explicitly or sufficiently while searching or evaluating drug combinations. Another limitation of existing methods is that they are blackbox-like methods to some extent, and thereby makes it difficult to explain why the drugs work.

In this paper, we propose a novel network-based method to identify effective drug combinations based on gene expression data of individual drugs. In particular, we construct a background molecular interaction network, and predict gene expression profile of one combination drug based on microarray data of individual drugs with a new computational scheme. We present a new integer programming model to identify active pathways or subnetworks affected by single drugs or combination drug from the background network by exploiting gene expression data. Moreover, we design a score by taking into account efficacy and side-effect based on the identified active subnetworks affected by drugs, and quantitatively evaluate all possible combinations of drugs and identify the best candidates of combination drugs. As a pilot study, we apply the method to identify effective combinations of drugs to treat Type 2 Diabetes. The result on real biological data verifies the effectiveness and efficiency of the proposed method.

2 Materials and Methods

The idea behind the proposed method is that a subnetwork or pathway will be activated and affected in the cellular system after a drug is administrated. Therefore, the active subnetwork can be used to assess the drug's overall effect, and thereby help to identify effective drug combinations by comparing the active subnetworks affected by individual drugs with that by combination drug.

2.1 Identifying subnetworks affected by drug

In our work, a molecular interaction network is represented as a graph $G = (V, E, W)$, where V represents the set of molecules or nodes, E represents the set of interactions between nodes and W is the set of weights defined by the differential expression change of genes. In detail, the weight of gene i is defined as $w_i = |\log T_i/C_i|$, where T_i is the expression value of gene i in a case sample and C_i is the one in a control sample. With the weighted graph generated above, the task therefore becomes into looking for a maximum-score subnetwork. In literature, there are some approaches on identifying active subnet-

works under specific conditions [13, 14, 15]. In this work, we introduce a new network flow model to find the maximum-score subnetwork. The network flow model was originally proposed by Lee and Dooly to model a constrained maximum-weighted connected graph that finds maximum-score subnetworks of size R with a fixed root node[16]. In this case, drug combinations have multiple targets, which are all needed to be included in the subnetwork to be found. Therefore, a dummy node, namely a drug node, and additional constraints are introduced into the original model. The model is formulated as follows:

$$\max_{\{x_i, Z_{ij}\}} \quad Q = \frac{1}{\sqrt{R}} \sum_{i=1}^{|V|} w_i x_i \tag{2.1}$$

$$subject\ to \quad \sum_{j \in T} Z_{Sj} = R \tag{2.2}$$

$$\sum_{j \in V} Z_{ji} - \sum_{k \in \acute{V}} Z_{ik} = 1, \quad i \in H \tag{2.3}$$

$$\sum_{j \in \acute{V}} Z_{ji} - \sum_{k \in \acute{V}} Z_{ik} = x_i, \quad i \in \acute{\acute{V}} \tag{2.4}$$

$$\sum_{j \in \acute{V}} Z_{ji} \leq (R-K) x_i, \quad i \in \acute{\acute{V}} \tag{2.5}$$

$$Z_{Sj} \in \{1, \cdots, R-K+1\}, \quad j \in H \tag{2.6}$$

$$Z_{ij} \in \{0, 1, \cdots, R-K\}, \quad i, j \in \acute{V} \tag{2.7}$$

$$x_i = 1, \quad i \in \overline{H} \tag{2.8}$$

$$x_i \in \{0, 1\}, \quad i \in \acute{\acute{V}} \tag{2.9}$$

where S denotes dummy node, $H = \{H_1, H_2, \cdots, H_K\}$ is the set of targets, $\acute{V} = V - \{S\}$, $\acute{\acute{V}} = \acute{V} - H$, and $\overline{H} = H \bigcup \{S\}$. $\bigcup$ is the operator of union for two sets, and $|V|$ is the number of elements of V. Binary variable x_i ($x_i \in \{0,1\}$) denotes whether the i-th gene is included in the subnetwork. The number of units of flow from the i-th gene to the j-th gene is signified by Z_{ij}. Furthermore, all the constraints introduced aim to ensure a connected subnetwork with R nodes.

Due to the NP-hard nature of this integer programming model, we relax it to a linear programming model in practice. The solution of the linear programming is obtained by an open source software glpsol.exe. Due to the relaxation, non-zero entries in variable x in solution will define the final identified subnetwork. We assume that R is at most as large as 10% of the number of nodes in the network. The lower bound of R is set to the number of nodes in the shortest paths between different drug targets. Finally, the subnetworks with the maximum score will be identified as the subnetwork affected by the drug.

3 Results

We applied our method to Type 2 Diabetes mellitus, which is one of leading complex diseases that threat the health of human being worldwide[17].

The data set used include the microarray data for Metformin and Rosiglitazone. They are all FDA approved drugs to treat Type 2 Diabetes. The combination of Metformin and Rosiglitazone, i.e. Avandamet, was also approved to treat Type 2 Diabetes. Compared with Metformin or Rosiglitazone clinically, Avandamet can improve glycemic control, insulin sensitivity without new tolerability issues[18, 19, 20, 21, 22]. In this paper, we intend to elucidate the mechanism underlying Avandamet and investigate why it works better than Rosiglitazone or Metformin, based on the proposed computational method.

To assess the efficacy of drugs, we used the genes predicted by at least five methods for Type 2 Diabetes from the supporting information of [23]. Subsequently, the list of Type 2 diabetes related genes obtained from OMIM database [24, 25] and the list from [23] were merged. Of the genes in the merged list, 54 genes exist in our background network. These 54 genes constitute the reference set of Type 2 Diabetes related genes. On the other hand, to assess the side-effect of drugs, we used the essential genes which are defined to be the orthologs of essential genes found in mouse, and the list of essential genes in the mouse were obtained from MGI (http://www.informatics.jax.org/). Besides, the background network was constructed by integrating PPI data obtained from HPRD database, Protein-DNA interaction data from TRED database [26] and Signaling pathway data from the supporting information of [27].

In the subnetwork affected by Avandamet (combination drug), 16 Type 2 Diabetes related genes, that is, TCF4, MAPK8IP1, NEUROD1, HNF4A, IRS1, AXL, ERBB2, PCSK2, RBP4, SLC8A1, IKBKAP, 8MARCA4, PMP22, CSF1R, RAG1 and PLCE1 are included. Seven Type 2 Diabetes related genes, that is, IRS2, AKT2, SMARCA4, PMP22, CSF1R, RAG1 and PLCE1 are affected by Metformin. Rosiglitazone affect 12 such genes, that is, TCF4, MAPK8IP1, NEUROD1, IRS1, AXL, ERBB2, PCSK2, RBP4, SLC8A1, CSF1R, RAG1 and SHC1. It can be seen that most of the disease related genes in the subnetworks affected by Metformin or Rosiglitazone are also in the subnetwork affected by Avandamet, which explains why Avandamet (combination drug) outperforms Rosiglitazone or Metformin to some extent. On the other hand, the number of essential innocent genes in the three identified subnetworks are 271, 266 and 242 respectively, which explain why Avandamet will not introduce new tolerability issues to some extent. To quantitatively measure the advantage of Avandamet over Rosiglitazone or Metformin based on the identified subnetworks, the three subnetworks were evaluated by scheme (2.10). Table. 1 gives scores of three subnetworks corresponding to three drugs when increasing the parameter λ from zero to one.

It can be seen from the table that scores of Avandamet are always higher than that of Rosiglitazone or Metformin regardless of the parameter λ, which agrees with the clinical conclusion very well. The results show that our method can successfully identify effective combination drug, i.e. Avandamet, which demonstrates the efficiency of the proposed method and also proves the necessity to understand working mechanism of drugs from perspectives of systems biology.

4 Discussion

In this paper, we present a new method to identify effective drug combinations. Different from existing methods, the proposed method aims to identify effective drug combinations from the perspective of network or systems biology. The main idea is that subnetworks affected by the drug administrated can be used as surrogates of overall impact brought by drug. Especially, the problem of identifying subnetworks affected by one drug including a combination drug was formulated into an integer programming model and solved by relaxing it to a linear programming model. Furthermore, we defined efficacy or side effect respectively by using the differential expression of disease genes and essential genes under study. A new score scheme that considers efficacy and side effect simultaneously was defined and used to evaluate candidate subnetworks and identify effective drug

Table 1: Scores of subnetworks affected by Avandamet, Metformin and Rosiglitazone, in which Avan, Met and Ros denote Avandamet, Metformin and Rosiglitazone respectively.

λ	Score of Avan	Score of Met	Score of Ros
0	-0.149013	-0.152550	-.155560
0.05	-0.109251	-0.125941	-0.118688
0.10	-0.069689	-0.099331	-0.081816
0.15	-0.030026	-0.072722	-0.044944
0.20	0.009636	-0.046112	-0.008072
0.25	0.049299	-0.019502	0.028799
0.30	0.088961	0.007107	0.065671
0.35	0.128623	0.033717	0.102543
0.40	0.168286	0.060326	0.139415
0.45	0.207948	0.086936	0.176287
0.50	0.247611	0.113546	0.213158
0.55	0.287273	0.140155	0.250030
0.60	0.326935	0.166765	0.286902
0.65	0.366598	0.193374	0.323774
0.70	0.406260	0.219984	0.360646
0.75	0.445923	0.246594	0.397517
0.80	0.485585	0.273203	0.434389
0.85	0.525247	0.299813	0.471261
0.90	0.564910	0.326422	0.508133
0.95	0.604572	0.353032	0.545005
1.00	0.644235	0.379641	0.581876

combinations. The pilot study on identifying combination of drugs to treat Type 2 Diabetes shows that our method can successfully identify the approved combination drug, i.e. Metformin&Rosiglitazone. In the future, to verify the reasonableness and demonstrate its power extensively, we will apply our proposed method to identify possible effective drug combinations for treating other complex diseases, such as cancer.

Despite the success of the proposed method, we noticed that there are still some issues that affect its performance and hamper its further application. Firstly, there are no expression data treated with combination drug available in public right now. We believe that the performance of our method will be boosted accordingly if the expression data treated with combination drug are available. Finally, when the method of predicting gene expression profile was used to predict expression profile caused by 3th-order or higher order combinations, expression ratio of some genes may take negative value, which needs to be modified during computation. In near future work, we will modify our method to predict gene expression profiles caused by higher order drug combinations.

Acknowledgements

Dr. Zikai Wu was partly supported by Natural Science Foundation of China under grant No.60873129, No.10802043. Dr. Xingming Zhao was partly supported by Innovation Funding of Shanghai University, Open Funding of National Key Laboratory of

Plant Molecular Genetics, SRF for ROCS, SEM. Dr. Luonan Chen was supported by Key Project of Shanghai Education Committee and JSPS-NSFC collaboration project(10711140116). The authors are indebted to Dr. Yong Wang for enthusiastic comments. The authors would like to thank Yuqing Qiu and Jiguang Wang for helpful discussion.

References

[1] Chen LN, Wang RS and Zhang XS. (2009) Biomolecular networks: methods and applications in systems biology. New Jersey, USA:Wiley Interscience.

[2] Lindsay MA. (2003) Target discovery. Nat Rev Drug Discov. 2, 831-838.

[3] Yang K, Bai HJ, Ouyang Q, Lai LH and Tang C. (2008) Finding multiple traget optimal intervention in disease-related molecular network. Molecular Systems Biology. 4:228.

[4] Chou TC. (2006) Theoretical Basis, Experimental Design, and Computerized Simulation of Synergism and Antagonism in Drug Combination Studies. Pharmacol Rev. 58(3): 621-680.

[5] Loewe S. (1953) The problem of synergism and antagonism of combined drugs. Arzneimittelforschung. 3:285-290.

[6] Bliss CI. (1956) The calculation of microbial assays. Bacteriol Rev. 20:243-258.

[7] Fitzgerald JB, Schoeberl B, Nielsen UB and orger PK. (2006) Systems biology and combination therapy in the quest for clinical efficacy. Nat Chem Bioc. 2(9): 458-466.

[8] Agrawal N, Pallos J, Slepko N, Apostol BL, Bodai L, Chang LW, Chiang AS, Thompson LM and Marsh JL. (2005) Identification of combinatorial drug combination regimens for treatment of Huntington's disease using Drosophila. P Natl Acad Ssi(USA). 102(10): 3777-3781.

[9] Cottarel G and Wierzbowski J. (2007) Combination drugs, an emerging option for antibacterial therapy. TRENDS Biotechnol. 25(12): 548-554.

[10] Calzolari D, Bruschi S, Coquin L, Schofield J, Feala J, Reed JC, McCulloch AD and Paternostro G. (2008) Search Algorithms as a framework for the Optimization of Drug combinations. Plos Comput Biol. 4(12): e1000249. doi:10.137/journal.pcbi.1000249.

[11] Borisy AA, Elliott PJ, Hurst NW, Lee MS, Lehál J, Price ER, Serbedzija G, Zimmermann GR, Foley MA, Stockwell BR and Keith Curtis T. (2003) Systematic discovery of multicomponent therapeutics. P Natl Acad Ssi(USA. 100(13): 7977-7982.

[12] Wong PK, Yu FQ, Shahangian A, Cheng GH, Sun R and Ho CM. (2008) Closed-loop control of cellular functions using comninatory drugs guided by a stochastic search algorithm. P Natl Acad Ssi(USA). 105(13): 5105-5110.

[13] Qiu YQ, Zhang SH. (2008) Uncovering Differentially expressed Pathways with protein Interation and gene expression data. Lecture Notes in Operations Research. 9, 74-82.

[14] Qiu YQ, Zhang SH, Zhang XS and Chen LN. (2009) Identifying Differentially expressed Pathways by high throughput data. IET Systems Biology. 7(2), 309-322.

[15] Zhao XM, Wang RS, Chen LN and Aihara K. (2009) Automatic modeling of signaling pathways based on network flow model. Journal of Bioinformatics and Computational Biology. (In press).

[16] Lee HF and Dooly DR. (1996) Algorithms for the constrained maximum-Weight connected graph problem. Nav Res Log. 43:985-1008.

[17] María T and Tusié L. (2005) Genes and Type 2 diabetes Mellitus. Arch Med res. 36:210-222.

[18] Fonseca V, Rosenstock J, Patwardhan and Salzman A. (2000) Effect of Metformin and Rosiglitazone Combination Therapy in Patients with Type 2 Diabetes Mellitus(A Randomized Controlled Trial). The Journal of the American Medical Association. 283:1695-1702.

[19] Gómez-Perez F, Fanghänel-Salmón G, Barbosa JA, Montes-Villareal J, Berry RA, Warsi G and Gould EM. (2002) Efficacy and safety of Rosiglitazone plus Metformin in Mexicans with type 2 diabetes. Diabetes Metab rev. 18: 127-134.

[20] Bailey CJ and Day C. (2004) Avandamet: combined Metformin-Rosiglitazone treatment for insulin resistence in type 2 diabetes. Int J Clin Pract. 58(9): 867-876.

[21] Rosak C, Petzoldt R, Wolf R, Reblin T, Dehmel B, Seidel D. (2005) Rosliglitazone plus Metformin is effective and well telerated in clinical practice: results from large observational studies in people with type 2 diabetes. Int J Clin Prat. 59(10):1131-1136.

[22] Rosenstock J, Rood J, Cobitz A, Biswas N, Chou H and Garber A. (2006) Initial treatment with Rosiglitazone/Metformin fixed-dose combination therapy compared with monotherapy with either Rosiglitazone or Metformin in patients with uncontrolled type 2 diabetes. Diabetes Obes Metab. 8:650-660.

[23] Tiffin N, Adie E, Turner F, Brunner HG, Van Driel MA, Oti M, Lopez-Bigas N, Ouzounis C, Perez-Iratxeta C and Andrade-Navarro MA. (2006) Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. Nucleic Acids Res. 34(10):3067-3081.

[24] Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 33:D514-D517.

[25] Amberger J, Bocchini CA, Scott AF and Hamosh A.(2009) McKusick's Online Mendelian Inheritance in Man (OMIM^R). Nucleic Acids Res. 37:D793-D796.

[26] Zhao F, Xuan Z, Liu L and Zhang MQ. (2005) TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. Nucl Acids Res. 33(Database Issue):D103-D107.

[27] Cui Q, Yu Z, Purisima EO and Wang E. (2006) Principles of microRNA regulation of a human cellular signaling network. Molecular Systems Biology. 2:46.