ibc Interdisciplinary Bio Central

IBC 2013;5:3, 1-7 • DOI: 10.4051/ibc.2013.5.2.0003

Open Access

FULL REPORT

Systematic Approach for Analyzing Drug Combination by Using Target-Enzyme Distance

Jaesub Park, Sunjae Lee, Kiseong Kim and Doheon Lee*

Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea

Subject areas; Bioinformatics/Computational biology/Molecular modeling

Author contribution; J.S.P.: Writing and design of the whole paper; S.J.I.: Design method, writing part of introduction about metabolic network; K.S.K.: Drug-enzyme distance data acquisition and analysis, interpretation of correlation score distribution; D.H. L.: Management, supervising and advising the whole paper

***Correspondence** and requests for materials should be addressed to D.H.L. (dhlee@kaist.ac.kr).

Editor; Keun Woo Lee, Gyeongsang National University, Korea

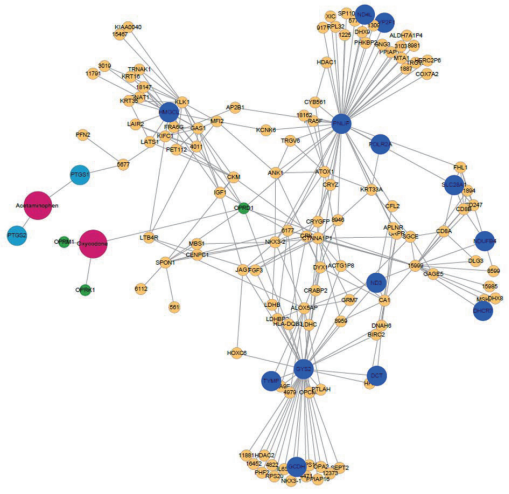
Received April 22, 2013
Accepted April 23, 2013
Published April 25, 2013

Citation; Park, J.S., et al. Systematic Approach for Analyzing Drug Combination by Using Target-Enzyme Distance. IBC 2013, 5:3, 1-7. doi: 10.4051/ibc.2013.5.2.0003

Competing interest; All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

SYNOPSIS

Recently, the productivity of drug discovery has gradually decreased as the limitations of single-target-based drugs for various and complex diseases become exposed. To overcome these limitations, drug combinations have been proposed, and great efforts have been made to predict efficacious drug combinations by statistical methods using drug databases. However, previous methods which did not take into account biological networks are insufficient for elaborate predictions. Also, increased evidences to support the fact that drug effects are closely related to metabolic enzymes suggested the possibility for a new approach to the study drug combinations. Therefore, in this paper we suggest a novel approach for analyzing drug combinations using a metabolic network in a systematic manner. The influence of a drug on the metabolic network is described using the distance between the drug target and an enzyme. Target-enzyme distances are converted into influence scores, and from these scores we calculated the correlations between drugs. The result shows that the influence score derived from the target-enzyme distance reflects the mechanism of drug action onto the metabolic network properly. In an analysis of the correlation score distribution, efficacious drug combinations tended to have low correlation scores, and this tendency corresponded to the known properties of the drug combinations. These facts suggest that our approach is useful for prediction drug combinations with an advanced understanding of drug mechanisms.



Key Words: drug interaction; metabolic network; drug selectivity; synergistic drug combination; protein-protein network

INTRODUCTION

For the past several decades, drug discovery efforts have been fairly successful, leading to effective and efficient treatments of menacing diseases. However, despite the advances in technology and the enormous increase in research budgets, R&D productivity in the area of drug discovery has gradually decreased[1]. Most of the reasons for this phenomenon are related to external factors such as the pharmaceutical environment or difficulties in discovering natural product drug candidates[2]. However, a more fundamental problem is related to the mechanism of drug action itself. Drugs which target only single molecules show limited efficacies and many side effects due to the redundancy and robustness of the complex biological networks in which they act.

To overcome the limitations of drug therapies based on a single target, drug combinations have been proposed as a solution. When using multi-target agents or more than two agents at the same time, various drug targets can be affected simultaneously. In such a case, biological networks related to disease are affected by drugs in various ways. For this reason, drug combinations are more efficacious and adaptable than compared to single-target agents[3]. Moreover, some synergistic drug combinations show improvements in terms of selectivity[4].
The most painstaking way to discover new drug combinations is to verify their effects by experience sequentially. However, given the restrictions related to time and money, it is impossible to undertake clinical trials for all possible combinations. Therefore, predictions of promising drug combinations that show synergistic effects are very important. For this reason, many computational methods have been suggested to predict drug combinations using drug databases. For example, one recent approach for finding drug combinations which integrated molecular and pharmacological data was introduced[5]. Although successful predictions of combinations were made in previous studies, the fundamental principle of a drug combination has yet to be revealed due to the lack of an in-depth consideration of the molecular mechanisms of drugs.

As part of the effort to understand drug actions, previous studies investigated drug mechanisms in several cellular processes. The important cellular processes of the metabolic pathway have also received attention as they are related to the pathogenesis of diseases such as diabetes, cardiovascular disease, neurodegeneration, and bipolar disorder[6-9]. Despite the increasing amount of evidence of multifaceted roles of metabolic pathways in the process of disease pathogenesis, systematic investigations of drug actions on metabolic pathways have been limited thus far. The accumulated information on human metabolic pathways and computational techniques for modeling related perturbation effects has enabled us to delve into issues pertaining to the identification of drug effects. Some studies have also underpinned the substantial utility of investigations into metabolic pathways to identify drug targets and the off-target effects[10,11]. Hence, a systematic investigation can be utilized to unveil the therapeutic effects of drug combinations and expand our understanding of drug combinations at the metabolic level.

Here, we suggest a novel approach for analyzing drug combinations based on the interrelationship between drug targets and enzymes, i.e. devised influential score of drug targets to enzymes based on distances of proteins in a PPI network, in a systematic manner. Evidences to support the fact that drug effects are related to metabolic enzymes have been increased. Some previous studies attempted to predict undesired drug effects via metabolic enzymes in the context of a metabolic network[10]. Other studies considered drug effects via metabolic enzymes in predicting drug-drug interactions[12]. These studies are based on previous report about various drug-drug interactions of which drugs targeted similar sets of metabolic enzymes but impeded enzyme activities to other drugs[13].

Hence we represent degree of drug effects to metabolic pathways in terms of the distance between drug targets and enzymes by a whole metabolic network. Based on the measured distance, we define what are known as influence scores of drugs for all enzymes. In addition, we calculate the correlation between two drugs using the influence score. The results show that there is a distinguishing correlation distribution pertaining to efficacious drug combinations that differs from that of non-efficacious drug combinations. Moreover, this difference in influence scores indicates the usefulness of our approach as a drug combination method.

RESULTS AND DISCUSSION

Subhead 1: Distance between drug targets and enzymes in protein-protein network

To evaluate the influence of drugs on the metabolic network, we used the distance between the drug target and the pertinent enzyme. This distance is defined as the shortest path length in the protein-protein interaction (PPI) network as established by connections between proteins. Because a drug target and an enzyme form a subset of an entire protein, both are included in the PPI network. In this network, a short distance between the drug target and a certain enzyme means that a drug has a considerable amount of influence on the metabolic pathway regulated by that enzyme.

We investigated the target-enzyme distance to determine the properties of the target-enzyme network (Figure 1). The average distance between the drug target and an enzyme was approximately 4.09, when a distance between directly connected nodes

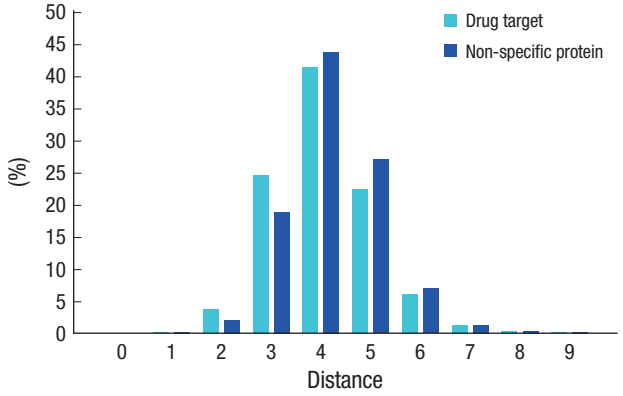


Figure 1. Protein-protein distance histogram.

is regarded as 1. In other words, there are three protein nodes between a certain drug target and an enzyme on average. The distance showing the largest percentage was 4, accounting for approximately 41%, and more than 85% of the distances were covered when including 3 and 5. Despite the large number of drug targets and enzymes, distances between them were very short. This property corresponds to that of a biological network as a scale-free network[14].

When comparing the distance distribution of a drug target and a non-specific protein, which indicates all proteins in PPI network, to an enzyme, we found some interesting results. The average distance between a drug target and an enzyme was 4.09, shorter than that of a non-specific protein between enzymes (4.19). We identified statistical significance of difference in distance distribution of two groups by Kolmogorov–Smirnov test (P-value < $2.2e^{-16}$). In particular, the ratio of short distances (d < 3) of drug targets (3.79%) was one and a half times that of non-specific proteins (2.21%). These results indicate that the drug target is closer to an enzyme than a non-specific protein.

As the drug target becomes closer to the enzyme, it could increase the chances to affect metabolic network than the drug targets of longer distance to the enzyme[15]. Therefore, drugs, of which targets are close to metabolic enzymes, have increasing tendency to treat a disease by bringing a change in the flux of metabolism. In other words, the protein-enzyme distance concisely describes the properties of the drug.

Subhead 2: The influence score and pattern represent the drug action on the metabolic pathway.

Although the target-enzyme distance feasibly represents the characteristic of the drug, that is, increased chances to affect metabolic enzymes, it is inappropriate to use that distance directly to analyze drug combinations. Above all, the distance was not linearly proportional to the biological influence due to the complex biological network. Moreover, infinite distances are difficult to handle when calculating the correlation. Finally,

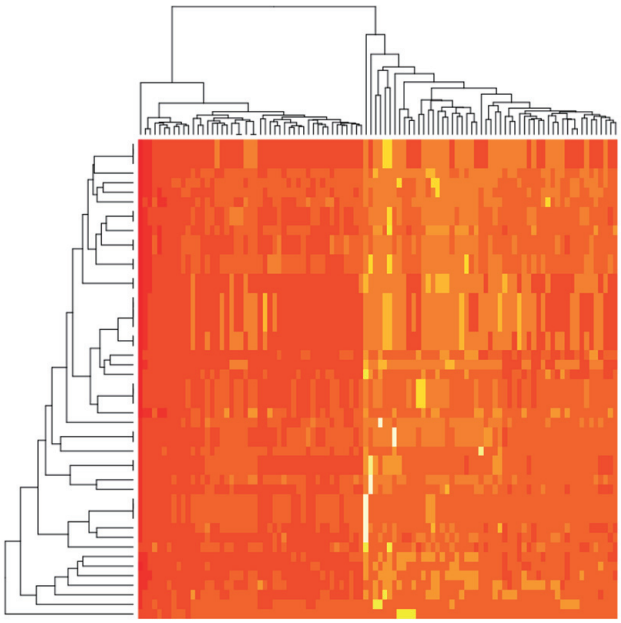


Figure 2. Heat map of a drug influence pattern. Two hundred drugs were selected randomly from all drugs as the sample. The x axis denotes the enzymes and the Y axis represents the drug. A high influence score is represented by the red squares.

there could be many drugs with several drug targets. In such a case, one drug has several distances to certain enzymes, and we had to integrate those distances. Therefore, applying our scoring method, the target-enzyme distances were converted into drug-enzyme influence scores with values between 0 and 1. This score comprehensively denotes the influence of the drug action to the enzyme.

One drug has one influence score for each enzyme. Thus, the set of influence scores for each enzyme can be represented as a type of vector. We termed this vector the influence pattern of the drug in each case.

The influence patterns have possibility to discriminate types of drug combinations, such as synergistic or addictive types, on the metabolic pathway by influence patterns of drug pairs. To remove biased patterns due to drug targets that have no connections to any metabolic enzymes, we exclude drugs whose targets have no paths to metabolic enzymes in a PPI network. A heat map of drug influence patterns shows that drugs can be categorized by the influence pattern which describes the drug action (Figure 2).

Subhead 3: The correlation score distribution depends on the type of effect of the drug combination.

To investigate the properties of the actions of drug combinations on the metabolic network, we calculated the Pearson correlation between influence patterns of two drug combinations. Before analyzing all of the drug combination, we choose two

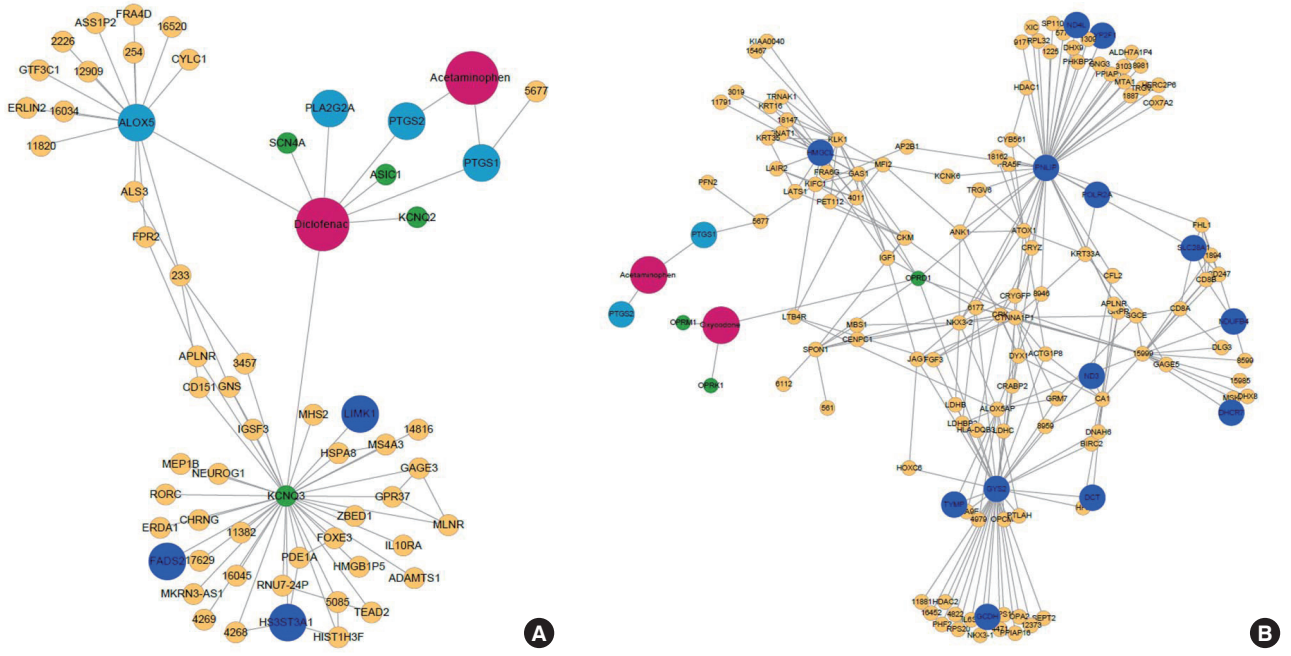


Figure 3. Target-enzyme interaction networks between drug combinations: (A) network of the acetaminophen and diclofenac combination, and (B) network of the acetaminophen and oxycodone combination. The red nodes are drugs, the blue nodes are enzymes, the green nodes are the drug target, and the blue-green nodes are the drug targets and enzymes at the same time.

drug combination cases with which to determine the meaning of the correlation score. One drug combination (acetaminophen and diclofenac) showed a high correlation score (0.664) while also showing a simple PPI network between two drugs (Figure 3A). Two drugs share two target proteins, and they have a similar path to the enzyme. Another drug combination (acetaminophen and oxycodone) had a low correlation score (0.115) and a more complex PPI network than the former combination (Figure 3B). They do not share drug targets, and their path to each enzyme varies. These two cases of protein-enzyme networks indicate that the correlation score expresses the differences in the action mechanisms between the two drug combinations. A high correlation score means that the influence patterns of the two drugs are quite similar while a low correlation score means the opposite.

Finally, we compared the distribution of the correlation scores for each group, categorized according to the effect of the combination. To obtain information about the effect of the drug combination, we used The Drug Combination Database (DCDB), which provides drug combination lists with their efficacy levels and effect types. Here, the term 'efficacious' means that a drug combination was able to produce the expected improvements over other treatments in clinical trials or pre-clinical studies[16].

The result revealed a noticeable difference in the distributions between efficacious and non-efficacious drug combinations. Overall, the density graph of efficacious combinations was weighted with a score of 0.2 and that of non-efficacious combinations ranged widely from -0.2 to 1. On average, the correlation score of efficacious combinations was 0.24, and that of non-efficacious combinations was 0.35. Thus, in general, efficacious combinations had lower correlation scores (Figure 4A). The difference between the shapes of graph of two groups was remarkable. While the non-efficacious combinations resulted in a wide graph, meaning that they had nearly the same number for each range of correlation score, most efficacious combination scored ranged from 0 to 0.3 Almost 68% of the correlation scores of efficacious combinations were in a range between 0 and 0.3, but only 43% of the scores of non-efficacious combinations were in that range. Another feature of efficacious drug combinations was that only 7% of efficacious combinations scored above 0.6, whereas 36% of non-efficacious combinations scored higher than 0.6.

To obtain more insight into drug combinations, we classified efficacious combinations as additive, synergistic, potentiative, or antagonistic combinations according to their effect type[17]. Because there were few potentiative and antagonistic drug combinations in the DCDB, we only looked into additive and synergistic combinations which were statistically significant. Additive drug combinations showed an almost parallel distribution to efficacious drug combinations. However, synergistic combinations interestingly showed a narrower distribution graph than efficacious combinations (Figure 4B). The average score of synergistic combinations (0.23) was lower than that of efficacious combinations (0.24), and the ratio of the score which

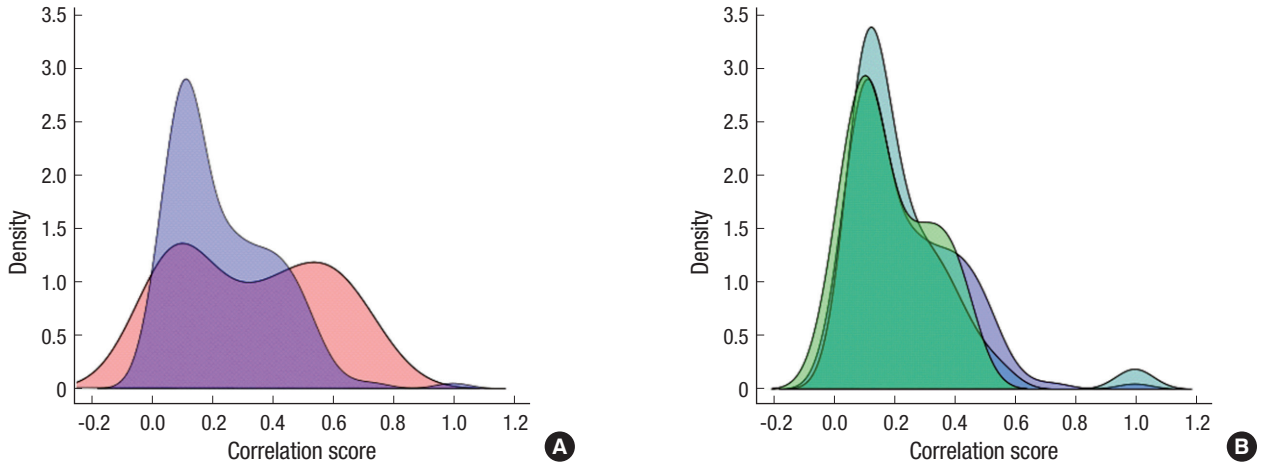


Figure 4. Density of the Pearson correlation scores of the drug combinations: (A) the blue graph is the correlation scores of efficacious drug combination and the red graph is that of non-efficacious drug combinations. (B) The blue, green, and blue-green graphs express the correlation scores of efficacious, additive and synergistic drug combinations, respectively.

ranged between 0 and 0.3 was 74%, which was higher than that of efficacious combinations (65%).

In summary, efficacious drug combinations tended to have low correlation scores, indicating that the influence patterns of the two drugs are not in good agreement with each other. Moreover, this tendency was more remarkable for the synergistic drug combinations. Most synergistic drug combinations had a correlation score between 0 and 0.2. Further work can explain why the influence patterns did not show a clear correlation. One possible explanation comes from recent research about the selectivity of drug combinations. According to the literature, synergies are induced from multi-target interactions which require coordinated actions in a narrower range of cellular phenotypes[4]. Thus, the synergistic effect most likely appears in drug combinations that have a correlation score between 0 and 0.2, which is not too far to be coordinated and not too close to restrict unnecessary interference.

CONCLUSION AND PROSPECTS

In previous studies, drug combinations were analyzed and predicted with a method that did not deal with biological networks considerably. Here, we suggest a novel approach for analyzing the drug combinations using metabolic networks in a systematic manner. The proposed method considers the biological mechanisms of drugs toward metabolism based on the distance between the drug target and an enzyme. Therefore, this approach has the advantage of being able to determine the actual relationships between drugs in cells.

The result pertaining to the different features of drug combinations between the different action types shows that phenotypes of drug combinations can be represented by correlations between influence patterns. In addition, if we improve the accuracy of the correlation scores of influence patterns, it will be quite useful for predicting efficacious drug combinations. Not only judging the synergy effect of drugs but also predicting the detailed action mechanisms of particular drug combinations will be possible.

For an improvement of our method, more drug target and enzyme databases are necessary. Numerous drug combinations cannot have their scores calculated due to the absence of drug target data. Moreover, the deficiency of the PPI data, which does not include half of the enzymes, led to the incompleteness of constructing a metabolic pathway. Nevertheless, our approach provides a new means of understanding drug combinations with biological implications, shedding new light on the selectivity of drug combinations.

MATERIALS AND METHODS

Subhead 1: Drug Combination & Drug Target data

All drug combination datasets were obtained from the DCDB, with 499 drug combinations in total (178 approved and 321 investigational, including 40 unsuccessful cases)[16]. These were categorized into 461 efficacious and 38 non-efficacious drug combinations. For all drug combinations, only pairwise drug combinations were selectively used. For drug target annotations, we used the DrugBank database. DrugBank provided information about 6,714 drugs, and among them, 3,004 contained the target protein data. We selectively used the drugs containing the target protein data and considered it as a complete drug pool, which allowed us to calculate the influence score. Consequently, only drug combinations that consisted of two drugs and drug targets of both drugs as provided from DrugBank were

utilized in our research. Among 499 drug combinations, 138 efficacious and 14 non-efficacious drug combinations were available for calculating the correlations between influence patterns.

Subhead 2: Data sources for network construction

Advances in high-throughput technologies and the accumulation of information from numerous studies have made it possible to use large-scale protein-protein interaction data. We selected a protein-protein interaction database, Human Protein Reference Database (HPRD), which collected interaction clues from curated information in the literature as well as large-scale experiment data[18]. From a public human metabolic network database, the Edinburgh Human Metabolic Network (EHMN) database, which contains human metabolic reactions and corresponding enzymes, an enzyme list was extracted for our calculation. The enzyme set for calculating the distances was parsed from the SBML data base[19]. There were 1,496 enzymes in the SBML data, but only 738 enzymes which existed in the PPI database constituted the enzyme set.

Subhead 3: Calculation of the target-enzyme distance from the protein-protein interaction database

The distance between a drug target and an enzyme calculated from the PPI database. If one protein interacts (specifically at the molecular level, this refers to binding) with another protein, then the distance between the two proteins is 1 and they are connected in the PPI network. When there is no direct interaction between proteins, the distance is the shortest protein-to-protein path length. For example, in Figure 5, the distance between node (a) and node (g) is 2, while that between node (c) and node (f) is 3. If the drug target is the enzyme itself, its distance is set to 0.

In conclusion, the distances between drug target j of drug i and enzyme k were defined using the following equation:

$$\text{Distance}\ (Drugtarget_j^i, Enzyme_k) = min\ (Drugtarget_j^i, Enzyme_k)$$

In this formula, $min\ (Drugtarget_j^i, Enzyme_k)$ denotes the shortest path length between drug target j of drug I and enzyme k. If there is no path to connect the target and enzyme, the distance is set to infinite. The shortest paths are calculated by MathMatica[20].

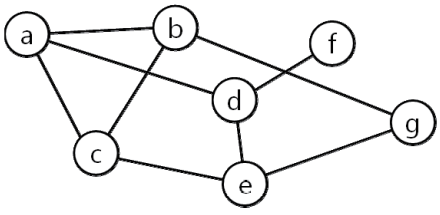


Figure 5. Example of a protein-protein network.

Subhead 4: Influence score

We defined the influence score based on the distance between drug targets and enzymes. When a drug had one target protein, the score was calculated only from that protein. However, if a drug had several targets, the score was calculated from the target which had the shortest distance to each enzyme. Finally, the influence score IS of drug i between enzyme k was defined as follows:

$$IS\ (Drug^i, Enzyme_k) = [\min\ \{\text{Distance}\ (Drugtarget_j^i, Enzyme_k)\}+1]^{-2}$$

To make a score adequate for calculating a correlation, we formulate the influence score to have a value between 0 and 1 intentionally.

ACKNOWLEDGEMENTS

This work was supported by the World Class University program (R32-2008-000-10218-0) and the Basic Research Laboratory grant (2009-0086964) of the Ministry of Science, ICT and Future Planning through the National Research Foundation of Korea.

REFERENCES

1. Ashburn, T. T., and Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 3, 673-683.
2. Li, J. W., and Vederas, J. C. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science* 325, 161-165.
3. Zimmermann, G. R., Lehár, J., and Keith, C. T. (2007). Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov Today* 12, 34-42.
4. Lehar, J., Krueger, A. S., Avery, W., Heilbut, A. M., Johansen, L. M., Price, E. R., Rickles, R. J., Short, G. F., 3rd, Staunton, J. E., Jin, X., et al. (2009). Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nat Biotechnol* 27, 659-666.
5. Zhao, X. M., Iskar, M., Zeller, G., Kuhn, M., van Noort, V., and Bork, P. (2011). Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol* 7, e1002323.
6. Zhang, B. B., Zhou, G., and Li, C. (2009). AMPK: an emerging drug target for diabetes and the metabolic syndrome. *Cell Metab* 9, 407-416.
7. Robert, C., Green, M. (2005). Diagnosis and Management of Alzheimer's Disease and Other Dementias. 2 Edition.: Professional Communications.
8. Vander Heiden, M. G. (2011). Targeting cancer metabolism: a therapeutic window opens. *Nat Rev Drug Discov* 10, 671-684.
9. Chong, Z. Z., Wang, S., Shang, Y. C., and Maiese, K. (2012). Targeting cardiovascular disease with novel SIRT1 pathways. *Future Cardiol* 8, 89-100.
10. Chang, R. L., Xie, L., Bourne, P. E., and Palsson, B. O. (2010). Drug off-target effects predicted using structural analysis in the context of a

metabolic network model. *PLoS Comput Biol* 6, e1000938.
11. Folger, O., Jerby, L., Frezza, C., Gottlieb, E., Ruppin, E., and Shlomi, T. (2011). Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol* 7, 501.
12. Gottlieb, A., Stein, G. Y., Oron, Y., Ruppin, E., and Sharan, R. (2012). INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol* 8, 592.
13. Bibi, Z. (2008). Role of cytochrome P450 in drug interactions. *Nutr Metab (Lond)* 5, 27.
14. Barabasi, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5, 101-113.
15. Huang, J., Niu, C., Green, C. D., Yang, L., Mei, H., and Han, J. D. (2013). Systematic Prediction of Pharmacodynamic Drug-Drug Interactions through Protein-Protein-Interaction Network. *PLoS Comput Biol* 9, e1002998.
16. Liu, Y., Hu, B., Fu, C., and Chen, X. (2010). DCDB: drug combination database. *Bioinformatics* 26, 587-588.
17. Jia, J., Zhu, F., Ma, X., Cao, Z., Li, Y., and Chen, Y. Z. (2009). Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 8, 111-128.
18. Mathivanan, S., Ahmed, M., Ahn, N. G., Alexandre, H., Amanchy, R., Andrews, P. C., Bader, J. S., Balgley, B. M., Bantscheff, M., Bennett, K. L., et al. (2008). Human Proteinpedia enables sharing of human protein data. *Nat Biotechnol* 26, 164-167.
19. Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R., and Palsson, B. O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104, 1777-1782.
20. Wolfram Research, I. (2010). Mathematica Edition: Version 8.0. Champaign, Illinois: Wolfram Research, Inc.