



# Article A Computational Approach for Pathway-Based Systemic Drug Influence

Shinuk Kim



Citation: Kim, S. A Computational Approach for Pathway-Based Systemic Drug Influence. *Processes* **2021**, *9*, 1063. https://doi.org/ 10.3390/pr9061063

Academic Editor: Ioannis Nikolakakis

Received: 6 May 2021 Accepted: 15 June 2021 Published: 17 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). College of Engineering, Sangmyung University, Cheonan 31066, Korea; kshinuk@smu.ac.kr; Tel.: +82-41-550-5452

**Abstract:** Drug repositioning is a well-known method used to reduce the time, cost, and development risks involved in bringing a new drug to the market. The rapid expansion of high-throughput datasets has enabled computational research that can suggest new potential uses for existing drugs. Some computational methods allow the prediction of potential drug targets of a given disease from a systematic network. Despite numerous efforts, the path of many drugs' efficacy in the human body remains unclear. Therefore, the present study attempted to understand drug efficacy by systematically focusing on functional gene sets. The purpose of this study was to carry out modeling to identify systemic gene networks (called drug paths) in drug-specific pathways. In our results, we found five different paths for five different drugs.

Keywords: paths of drug efficacy; parameter search; systemic gene networks

## 1. Introduction

The utilization of existing drugs to treat new diseases, which is called drug repositioning, is a process that may help eliminate some of the issues associated with bringing a new drug to market, a 10–15-year process that requires considerable effort and an average cost of 2–3 billion United States dollars [1–3]. Despite overcoming these issues, the average approval rate of new drugs for clinical use is only 10%. Drug repositioning methods have been developed on the basis of the similarity of chemical compounds or ligands, the three-dimensional (3D) structure of proteins [4,5], and network-based algorithms for drug-target interactions and drug-disease interactions [6–9]. Structural approaches have identified potential drug-target interactions using 3D structures. The disadvantages of this method include the time-consuming computational 3D simulation process and the associated limitations of being applied without first obtaining 3D structure information. Ligand-based approaches have identified the target protein by comparing candidate ligands [5,7]. One limitation of this method is that no or few known ligands for the target protein may exist [10]. A network-based model that consists of nodes and edges has indicated that a higher number of connected nodes will produce a greater amount of influence [11,12]. In many cases, a network-based algorithm combined with a mathematical model was applied to develop dynamics of drug studies [13–15]. These computational drug repositioning methods usually follow four steps: data collection from a drug bank or gene-protein interaction, a computational model, evaluation of the method using new datasets, and validation through clinical trials [7]. Although network-based models have revealed functional relationships and drug candidates for repositioning, the drug-disease and drug-target relationships are constructed on the basis of prior knowledge, such as protein-protein interactions [2]. In previous studies, we adopted inverse problems and/or ordinary differential equations for inferring parameters of gene–gene interaction [16,17]. Both methods contributed to identifying interactions between nodes, such as genes, drugs, and targets.

Novelty of this study is that we use a mathematical model of an inverse problem for inferring gene–gene networks in enrichment pathway instead of statistical approaches. In particular, this mathematical approach has an advantage of identifying parameters without

any prior information. Our method integrates pathway information and differentially expressed gene datasets with a mathematical computational algorithm. Therefore, the proposed study aimed to find the most effected pathway of a given drug as well as systemic gene interactions in the pathway.

## 2. Materials and Methods

#### 2.1. Materials

We downloadedMCF7 breast cancer cell line (series GSE5258-GPL3921) updated in 2017 from Gene Expression Omnibus (GEO) [18]. The datasets included 218 total samples divided into 182 drug-induced cancer cell lines and 36 control samples from the Affymetrix-HThuman genome U133A array platform converting 22,268 probes to 13,235 gene names. Datasets were rearranged by eliminating those located under unknown gene names and averaging others listed under duplicate gene names. Our new method (see Section 2.2) requires the statistical analyses in which normal and tumor samples are used previously [19]. In this study, we set the minimum sample size to five arbitrarily. We extracted five drug-induced datasets consisting of five or more samples: LY294002, valproic acid, sirolimus, trichostatin A, and wortmannin-induced datasets. These datasets consisted of eight, seven, seven, seven, and five samples, respectively.

The targets of gene markers by stage, grade, and drug sensitivity are unstable among datasets, and thus drug-path methods are introduced [20,21]. Numerous gene sets are available in open source; here, we used the Kyoto encyclopedia of genes and genomes (KEGG) pathway information [22].

## 2.2. Methods

Even though all the pathways play important roles in drug efficacy, we were concerned with identifying the pathway most affected by a drug. Therefore, we employed gene set enrichment analysis (GSEA) downloaded from [23] and obtained it using the enrichment score. Any detailed information regarding statistical analyses involved were stated previously [20].

An inverse algorithm [16,24,25] consists of four subroutines: the optimization method, a line search for finding the next direction for the parameters, objective function for determining whether the algorithm terminates or repeats by computing the distance between real datasets, and a mathematical model to generate synthetic datasets and parameters in matrix form. The method used in the study was also examined and validated in our previous studies [16,25,26]. In this study, we only used cancer datasets to identify the gene–gene network (parameters in an inverse algorithm) of the enrichment pathways. The parameter predictions for an inverse algorithm, however, require both real datasets and synthetic datasets. Synthetic datasets, which are required to search parameters, will be updated in each iteration. By adding white Gaussian noise [27–29] to a real dataset, we generated initial synthetic data instead of random permutations, since the variation in the real dataset was about 50% of the mean values. In the case of the datasets used to detect signals such as an interaction, noise must be reduced [27]. In this study, however, noise was added to find interactions. The schematic overall flow chart of the methods is shown in Figure 1. In brief, the formula for generating synthetic datasets was  $x_{syn} = x_{real} + \varepsilon$ , where  $\varepsilon \in (-0.1, 0.1)$ ;  $\varepsilon$ is white Gaussian noise;  $x_{syn}$  and  $x_{real}$  are synthetic and real datasets, respectively. In each iteration,  $x_{syn}$  is newly updated. We adapted the  $L_2$  norm for objective function f(x), which can be rewritten as  $||L_2|| = \sqrt{\sum (x_{real} - x_{syn})^2}$ . The Broyden–Fletcher–Goldfarb–Shanno (BFGS) method is one that uses variable metric methods in multidimensions for finding the minimum of objective function, which is called an optimization method. The advantage of this method is that it uses a quasi-Newton method using the quadratic convergence of a Newton method of a nonlinear model, which is faster and more accurate than other optimization methods [24]. The drug modeling is written as  $|X_{ij}| = [p_{ik}] |X_{kj}|$ , and the



Figure 1. The overall flow chart of the inverse problem.

## 3. Results

We identified the top enrichment pathways characterized by metabolism using GSEA on the basis of each drug-induced dataset. These pathways include nitrogen, selenoamino acid and cysteine, porphyrin chlorophyll, naphthalene, and anthracene degradation metabolism pathway from LY294002-, valproic acid-, sirolimus-, trichostatin A-, and wortmannin-induced datasets. The cutoff values depend on size of genes in a pathway.

## 3.1. The Nitrogen Metabolism Pathway from LY294002-Induced Datasets

LY294002 is a known inhibitor of oncogenes [30,31]. We performed GSEA to compare LY294002-induced cancer datasets (with eight samples and 13,235 genes) to normal datasets. The nitrogen metabolism (NitroM) pathway was obtained as the top enrichment pathway. To perform an inverse algorithm using NitroM and including 23 unique genes, we obtained 23 positive relations and 2 negative relations with a 0.46 cutoff. Only two genes, asparagine synthetase (ASNS) and CA14, were donors, while all 23 genes in the NitroM pathway were acceptors (Figure 2).





NitroM is required for cell growth and proliferation and is essential for the synthesis of a variety of biomolecular compounds [32]. A prior study described a current clinical investi-

gation of some drugs, including eryaspase targeting NitroM; the target gene L-asparaginase of eryaspase may be involved in reducing metastasis in breast cancer patients [33]. These studies support our result that the drug's target gene is ASNS in the NitroM pathway and also that LY294002 is one candidate drug for the treatment of breast cancer.

#### 3.2. Selenoamino Acid Metabolism from Valproic Acid-Induced Datasets

We performed GSEA to compare seven valproic acid-induced samples to 36 normal samples with 13,235 genes and determined the selenoamino acid metabolism (SelAM), including only 23 genes (Figure 3). Using these 23 genes of the SelAM pathway, we carried out an inverse problem to reveal the systemic functional relationships between those genes in this pathway. We revealed 21 positive relationships and 5 negative relationships using a cutoff of 0.45. The yellow nodes are acceptor nodes, while the green nodes are donors; the red lines indicate negative relationships, and gray lines show positive relationships.



**Figure 3.** Diagram of gene–gene interactions in the SelAM pathway induced by valproic acid. The red and black lines represent inhibition and activation, respectively.

In this case, adenosylhomocysteinase (AHCY) was negatively related to five genes, and methionine adenosyltransferase 1A (MAT1A) was positively connected with 21 genes, which may imply that these two genes could be valproic acid drug target biomarkers.

#### 3.3. Cysteine Metabolism Pathway from Sirolimus-Induced Datasets

A GSEA was performed to compare seven sirolimus-induced cancer samples to normal samples with 13,235 genes. We obtained the cysteine metabolism (CystM) pathway, including 13 genes with the highest enriched score. The gene networks are shown (cutoff: 0.15, -0.15) with six donors and five acceptors (Figure 4). In this pattern, we found a larger number of donors than acceptors.



**Figure 4.** Gene–gene interactions in the CystM pathway induced by sirolimus. The red and black lines represent inhibition and activation, respectively.

The results of studies that have investigated the associations between cysteine and breast cancer risk are inconsistent. Some cohort studies have reported positive associations [34,35], while another cohort study did not observe an association [36]. A fourth study reported a negative association with breast cancer risk [37]. In addition, a recent study reported that balancing the cysteine pathway was a therapeutic benefit in neurodegenerative disorders [38].

#### 3.4. Porphyrin Chlorophyll Metabolism from Trichostatin A-Induced Datasets

We carried out a GSEA to compare seven trichostatin A-induced cancer samples to normal datasets and obtained the porphyrin chlorophyll metabolism (PorChM) as the top enrichment pathway consisting of 28 genes and 29 gene–gene networks (cutoff > 0.6 or <-0.6). UGT2B4 and UGT1A10 are donor genes, and all their negative interactions are shown in Figure 5.



**Figure 5.** Gene–gene interactions in the PorChM pathway induced by trichostatin A. The red lines represent inhibition.

Trichostatin A, a histone deacetylase inhibitor, is a target therapeutic drug of hematological malignancies, such as breast cancer [39,40]. However, no systemic influences of trichostatin A have been discovered.

## 3.5. Naphthalene and the Anthracene Degradation Pathway from Wortmannin-Induced Datasets

Wortmannin is a known inhibitor of oncogenes [30]. We performed a GSEA to compare five samples of wortmannin-induced datasets with normal datasets and obtained the Naphthalene\_and Anthracene\_Degradation (NaphAD) pathway, including 14 genes in the top differentially expressed pathway. The gene–gene interactions of this pathway are shown in Figure 6 with a cutoff > 6.0 and <-6.0. Only two genes affected 12 genes; PRMT8 (chr 12p 13.32) is an inhibitor, while PRMT5 (chr 14q11.2) is an activator.



**Figure 6.** Gene–gene interactions in the NaphAD pathway induced by wortmannin. The red and black lines represent inhibition and activation, respectively.

A previous study reported that PRMT8 was correlated with survival in patients with breast cancer and gastric cancer [41]. That study demonstrated that a high level of PRMT8 expression was associated with positive survival in breast cancer and negative survival in gastric cancer.

#### 4. Discussion

We analyzed five different drug datasets and obtained five different pathways and systematic gene networks from each pathway (Figure 7). In general, the number of donors was much smaller than that of acceptors, and some relationships between donors were connected either positively or negatively. In the case of trichostatin A, all donors negatively interacted with all acceptors. In contrast, there were more positive interactions than negative ones for LY294002 and valproic acid. Interestingly, in the case of wortmannin, most acceptors were negatively affected by PRMT8 and positively influenced by PRMT5. The only exceptions were DHRS7, PRMT2, and LCMT2, which were negatively affected by PRMT8. In the case of sirolimus, the six donors and five acceptors contributed 4 negative interactions and 19 positive interactions.



**Figure 7.** Potential systemic interactions via pathways, which imply systemic drug influence. The red and black lines represent inhibition and activation, respectively.

These results imply that drugs not only directly influence their target gene but also systemically affect hierarchical gene networks in a disease during treatment. Therefore, exploring systemic patterns of drug influence is a better therapeutic method than investigating only one target gene of a drug. From this perspective, the presented method provides a computational approach to determine the pathway-based systemic drug influence; given that the patterns of relationships between gene networks and drugs are the same, many drugs can therefore be repurposed to treat other diseases.

## 5. Conclusions

In this work of pathway information and inverse problem, we proposed gene–gene interaction-based drug paths in the enriched pathway showing that four out of five drugs have an effect on some genes, which then regulate other genes in the pathway. Although those newly suggested drug paths remain to be validated, the current findings of the study suggest that screening network-based drug efficacy via a specific pathway may offer an unprecedented opportunity of applying a pre-existing drug to a new disease with the same pattern in the selected pathway.

**Funding:** This research was funded by a National Research Foundation of Korea (NRF, MSIT), grant number NRF-2017R1A2B4010684.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** Data can be freely ordered from GEO datasets (https://www.ncbi.nlm. nih.gov/geo/query/acc.cgi?acc=GSE5258, accessed on 17 June 2021).

Conflicts of Interest: The author declares no conflict of interest.

## References

- Pushpakom, S.; Iorio, F.; Eyers, P.A.; Escott, K.J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; McNamee, C.; et al. Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 2019, *18*, 41–58. [CrossRef]
- 2. Badkas, A.; De Landtsheer, S.; Sauter, T. Topological network measures for drug repositioning. Brief. Bioinform. 2020. [CrossRef]
- 3. Lotfi Shahreza, M.; Ghadiri, N.; Mousavi, S.R.; Varshosaz, J.; Green, J.R. A review of network-based approaches to drug repositioning. *Brief. Bioinform.* **2018**, *19*, 878–892. [CrossRef]
- Adasme, M.F.; Bolz, S.N.; Adelmann, L.; Salentin, S.; Haupt, V.J.; Moreno-Rodriguez, A.; Nogueda-Torres, B.; Castillo-Campos, V.; Yepez-Mulia, L.; De Fuentes-Vicente, J.A.; et al. Repositioned drugs for chagas disease unveiled via structure-based drug repositioning. *Int. J. Mol. Sci.* 2020, 21, 8809. [CrossRef] [PubMed]
- 5. Cheng, A.C.; Coleman, R.G.; Smyth, K.T.; Cao, Q.; Soulard, P.; Caffrey, D.R.; Salzberg, A.C.; Huang, E.S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75. [CrossRef] [PubMed]
- 6. Luo, H.; Li, M.; Wang, S.; Liu, Q.; Li, Y.; Wang, J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics* **2018**, *34*, 1904–1912. [CrossRef]
- Luo, H.; Li, M.; Yang, M.; Wu, F.X.; Li, Y.; Wang, J. Biomedical data and computational models for drug repositioning: A comprehensive review. *Brief. Bioinform.* 2020, 22, 1604–1619. [CrossRef]
- Ashburn, T.T.; Thor, K.B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 2004, *3*, 673–683. [CrossRef] [PubMed]
- 9. Wang, W.; Yang, S.; Zhang, X.; Li, J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* **2014**, *30*, 2923–2930. [CrossRef] [PubMed]
- Keiser, M.J.; Roth, B.L.; Armbruster, B.N.; Ernsberger, P.; Irwin, J.J.; Shoichet, B.K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 2007, 25, 197–206. [CrossRef] [PubMed]
- 11. Pastor-Satorras, R.; Vespignani, A. Epidemic spreading in scale-free networks. Phys. Rev. Lett. 2001, 86, 3200–3203. [CrossRef]
- 12. Albert, R.; Jeong, H.; Barabasi, A.L. Error and attack tolerance of complex networks. *Nature* 2000, 406, 378–382. [CrossRef]
- 13. Yin, N.; Ma, W.; Pei, J.; Ouyang, Q.; Tang, C.; Lai, L. Synergistic and antagonistic drug combinations depend on network topology. *PLoS ONE* **2014**, *9*, e93960. [CrossRef]
- 14. Wang, P.; Lu, J.; Yu, X. Identification of important nodes in directed biological networks: A network motif approach. *PLoS ONE* **2014**, *9*, e106132. [CrossRef]
- 15. Wang, P. Statistical identification of important nodes in biological systems. J. Syst. Sci. Complex 2021, 1–17. [CrossRef]
- 16. Kim, S.; Kim, J.; Cho, K.H. Inferring gene regulatory networks from temporal expression profiles under time-delay and noise. *Comput. Biol. Chem.* 2007, *31*, 239–245. [CrossRef]
- 17. Park, S.M.; Hwang, C.Y.; Choi, J.; Joung, C.Y.; Cho, K.H. Feedback analysis identifies a combination target for overcoming adaptive resistance to targeted cancer therapy. *Oncogene* **2020**, *39*, 3803–3820. [CrossRef]
- Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.P.; Subramanian, A.; Ross, K.N.; et al. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006, 313, 1929–1935. [CrossRef] [PubMed]
- Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 2005, 102, 15545–15550. [CrossRef] [PubMed]
- Van't Veer, L.J.; Dai, H.; van de Vijver, M.J.; He, Y.D.; Hart, A.A.; Mao, M.; Peterse, H.L.; van der Kooy, K.; Marton, M.J.; Witteveen, A.T.; et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, *415*, 530–536. [CrossRef] [PubMed]
- 21. Zeng, H.; Qiu, C.; Cui, Q. Drug-path: A database for drug-induced pathways. Database 2015, 2015, bav061. [CrossRef] [PubMed]
- Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000, *28*, 27–30. [CrossRef] [PubMed]
   Subramanian, A.; Kuehn, H.; Gould, J.; Tamayo, P.; Mesirov, J.P. GSEA-P: A desktop application for Gene Set Enrichment Analysis.
- Bioinformatics 2007, 23, 3251–3253. [CrossRef] [PubMed]
  24. Press, W.H. Numerical Recipes in C: The Art of Scientific Computing, 2nd ed.; Cambridge University Press: Cambridge, UK, 1992.
- Kim, S.; Kreider, K. Parameter identification for nonlinear elastic and viscoelastic plates. *Appl. Numer. Math.* 2006, 56, 1538–1554. [CrossRef]
- 26. Kim, S. Identifying dynamic pathway interactions based on clinical information. Comput. Biol. Chem. 2017, 68, 260–265. [CrossRef]
- 27. Ideker, T.; Dutkowski, J.; Hood, L. Boosting signal-to-noise in complex biology: Prior knowledge is power. *Cell* **2011**, 144, 860–863. [CrossRef]
- 28. Aquino, G.; Rocco, A. Bimodality in gene expression without feedback: From Gaussian white noise to log-normal coloured noise. *Math. Biosci. Eng.* **2020**, *17*, 6993–7017. [CrossRef]
- 29. MATLAB. Version (R2020a); The MathWorks Inc.: Portola Valley, CA, USA, 2020.

- Walker, E.H.; Pacold, M.E.; Perisic, O.; Stephens, L.; Hawkins, P.T.; Wymann, M.P.; Williams, R.L. Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. *Mol. Cell* 2000, *6*, 909–919. [CrossRef]
- 31. Gharbi, S.I.; Zvelebil, M.J.; Shuttleworth, S.J.; Hancox, T.; Saghir, N.; Timms, J.F.; Waterfield, M.D. Exploring the specificity of the PI3K family inhibitor LY294002. *Biochem. J.* **2007**, 404, 15–21. [CrossRef]
- 32. Kurmi, K.; Haigis, M.C. Nitrogen metabolism in cancer and immunity. Trends Cell Biol. 2020, 30, 408–424. [CrossRef]
- 33. Knott, S.R.V.; Wagenblast, E.; Khan, S.; Kim, S.Y.; Soto, M.; Wagner, M.; Turgeon, M.O.; Fish, L.; Erard, N.; Gable, A.L.; et al. Asparagine bioavailability governs metastasis in a model of breast cancer. *Nature* **2018**, *554*, 378–381. [CrossRef]
- 34. Chou, Y.C.; Lee, M.S.; Wu, M.H.; Shih, H.L.; Yang, T.; Yu, C.P.; Yu, J.C.; Sun, C.A. Plasma homocysteine as a metabolic risk factor for breast cancer: Findings from a case-control study in Taiwan. *Breast Cancer Res. Treat.* 2007, 101, 199–205. [CrossRef]
- 35. Lin, J.; Lee, I.M.; Song, Y.; Cook, N.R.; Selhub, J.; Manson, J.E.; Buring, J.E.; Zhang, S.M. Plasma homocysteine and cysteine and risk of breast cancer in women. *Cancer Res.* 2010, *70*, 2397–2405. [CrossRef]
- 36. Zhang, S.M.; Willett, W.C.; Selhub, J.; Hunter, D.J.; Giovannucci, E.L.; Holmes, M.D.; Colditz, G.A.; Hankinson, S.E. Plasma folate, vitamin B6, vitamin B12, homocysteine, and risk of breast cancer. *J. Natl. Cancer Inst.* **2003**, *95*, 373–380. [CrossRef]
- Zhang, S.M.; Willett, W.C.; Selhub, J.; Manson, J.E.; Colditz, G.A.; Hankinson, S.E. A prospective study of plasma total cysteine and risk of breast cancer. *Cancer Epidemiol. Biomark. Prev.* 2003, 12, 1188–1193.
- Paul, B.D.; Sbodio, J.I.; Snyder, S.H. Cysteine Metabolism in Neuronal Redox Homeostasis. *Trends Pharmacol. Sci.* 2018, 39, 513–524. [CrossRef] [PubMed]
- 39. Imai, Y.; Maru, Y.; Tanaka, J. Action mechanisms of histone deacetylase inhibitors in the treatment of hematological malignancies. *Cancer Sci.* **2016**, *107*, 1543–1549. [CrossRef]
- 40. Vigushin, D.M.; Ali, S.; Pace, P.E.; Mirsaidi, N.; Ito, K.; Adcock, I.; Coombes, R.C. Trichostatin A is a histone deacetylase inhibitor with potent antitumor activity against breast cancer in vivo. *Clin. Cancer Res.* **2001**, *7*, 971–976.
- 41. Hernandez, S.J.; Dolivo, D.M.; Dominko, T. PRMT8 demonstrates variant-specific expression in cancer cells and correlates with patient survival in breast, ovarian and gastric cancer. *Oncol. Lett.* **2017**, *13*, 1983–1989. [CrossRef] [PubMed]