Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category

OXFORD

Subject Section

Drug Target Interaction prediction using network information

Radhika Saraf [1,*]

[1] Electrical and Computer Engineering, Texas A&M University, College Station, 77840, US .

[*] To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The overall goal is to identify the mechanisms of action of drugs used in osteosarcoma treatment. Natural compounds and herbal medicine are being used in studies with success. In order to develop a drug derived from natural compounds to treat this cancer, we must know their targets and possible downstream effects. In this paper, we use drug target interaction(DTI) prediction techniques to predict targets of the drug Cryptotanshinone, a Chinese herbal medicine. We also look to estimate the efficacies of different drug combinations on osteosarcoma cell lines to come up with the most effective treatment strategy.

Results: We use Boolean representation techniques to model osteosarcoma. The Boolean network is developed with the help of drug-target information from the DTI prediction algorithm. The output of the Boolean network is a weighted apoptotic fraction. This output is cross-validated by data from actual biological experiments.

Contact: saraf.radhika@tamu.edu

Supplementary information: Supplementary data are available at *https://github.tamu.edu/sarafradhika/DTICryptotanshinone* online.

1 Introduction

The drug development process requires complete information about the effects of any potential drug. At the preclinical trial stage, it is important to model the action of the drug and predict its behaviour in animals and humans. The targets of the drug must be known in order to define a functional model. Predicting the efficacy of different drug combinations will help development. The focus of this work is to study the mechanism of action of the drug Cryptotanshinone in osteosarcoma, a type of bone cancer. We will also estimate the theoretical efficacies of various drug combinations of Cryptotanshinone and other drugs.

Osteosarcoma affects children, adolescents and young adults predominantly Isakoff *et al.* (2015). A drug can have different effects in adults as compared to children. In order to develop Cryptotanshinone as a drug for osteosarcoma, it is important to take all these factors into consideration during its modeling. The way to choose the relevant links and genes in this huge network of candidates is to look at the drugs that have been used with positive results. Natural compounds seem to be more effective in treating the cancer. There is not enough data about these compounds to completely predict their downstream effects. The challenge is twofold; firstly to build a network that represents the cell signaling pathways in osteosarcoma, and then secondly, to isolate the points in this network where Cryptotanshinone acts. This paper attempts to use drug target interaction prediction as outlined in Luo *et al.* (2017) to elucidate the mechanism of action of such natural compounds.

Biological experiments are performed that involve techniques such as gene silencing to identify possible targets of drugs. These experiments usually investigate only one or two pathways. The unique and complex biology of osteosarcoma makes it difficult to pinpoint drug targets in this manner, and to effectively predict accurate drug efficacy in the late phases of clinical trials Isakoff *et al.* (2015). The key lies in integrating the pathways and examining and modeling cross talk between them. This means that finding the target of a drug will require a search among a large number of candidate genes. The motivation behind employing bioinformatics tool is clear.Given that there are a lot of nodes and not enough data points, methods such as Bayesian networks might not yield accurate results. Without an accurate model, drug target prediction cannot be performed effectively in such methods. Additionally, exact inference on Bayesian networks with loops is not straightforward. Drug target prediction is done using machine learning and deep learning, since these

methods allow for drug-target interaction to be modeled without complete knowledge of the system. We look at the method called DTINet Luo *et al.* (2017), in detail in this work. DTINet uses information from a heterogeneous network to learn feature representations of drugs and targets to aid the discovery of drug-target interaction. The heterogeneous network comprises edges representing different relationships between drugs, diseases and proteins. DTINet learns the features based on similarity scores of the various nodes.

2 Approach

First step is to curate information from individual biological experiments and construct a Boolean network representation of osteosarcoma. Following which, the effects of drugs whose targets are known in this network can be quantified using a metric. We use apoptotic fraction as our metric as shown in Equation (1), where we give the drug as input to the network and measure a weighted average of apoptotic genes as output.

$$\text{Apoptosis Ratio} = \frac{\sum \text{Pro-Apoptotic factors}}{\sum \text{Anti-Apoptotic factors}} \tag{1}$$

Next, we will choose the relevant information from the datasets in Luo *et al.* (2017). The features for Cryptotanshinone can be calculated from the databases and techniques described in Luo *et al.* (2017). Once we construct a heterogeneous network, we will survey the various matrix factorization techniques to project the network into a common feature space. This paper attempts to employ the matrix factorization techniques to learn the targets of Cryptotanshinone.

It is known that STAT3 is a target of Cryptotanshinone, and this will be used as a control. We also know the set of possible targets based on pharmacokinetic and biological experiments Chen *et al.* (2011) and the predicted target is most likely from this set. Incorporating the predicted drug targets of Cryptotanshinone into the Boolean model, we will evaluate the model against the results of the actual biological experiments.

3 Methods

Osteosarcoma is a cancer of the bones that are undergoing growth. It has a complex biology and traditional models do not satisfactorily represent it. In normal bone cells, we must consider a few common pathways such as cell survival (PI3K/Akt), cell proliferation (MAPK/ERK) as seen in Figure 1 and mitochondrial apoptotic pathways. Since bone cells interact with red blood cells, we can look at angiogenesis (JAK/STAT) as in Figures **??** and RBC ceramide synthesis pathways. We usually also consider immune system (KEAP1/NRF2), where we can look at the extrinsic apoptosis pathways and their interaction with the inflammation(NKκB) and hypoxia (H1F) pathways. Lastly, we look at metabolic pathways (Glutamate to Glutathione), particularly those that interact with the immune system. Next, we will look to add stemness pathways (Notch, Hedgehog, Wnt), since osteosarcoma involves mesenchymal stem cells that mature into osteoblasts.

The various gene interactions in osteosarcoma can be represented by biological pathways, some of which are all well documented Kanehisa and Goto (2000); Kanehisa *et al.* (2016a,b). Some of the interconnections derived during modeling these pathways are based on the interpretation of different research papers Thomas *et al.* (2010); Fleury *et al.* (2002); Tait and Green (2010); Mendoza *et al.* (2012); Saleiro and Platanias (2015); Corazza *et al.* (2009); Flashner-Abramson *et al.* (2016); Liu *et al.* (2003); Chen *et al.* (2011); Wei *et al.* (2015); Chiarini *et al.* (2011); Yue *et al.* (2016); Jamil *et al.* (2008) by the author of the present paper. We consider only a subset of all possible interconnections and signaling pathways in the cell, since the cancer of interest to us here is osteosarcoma.

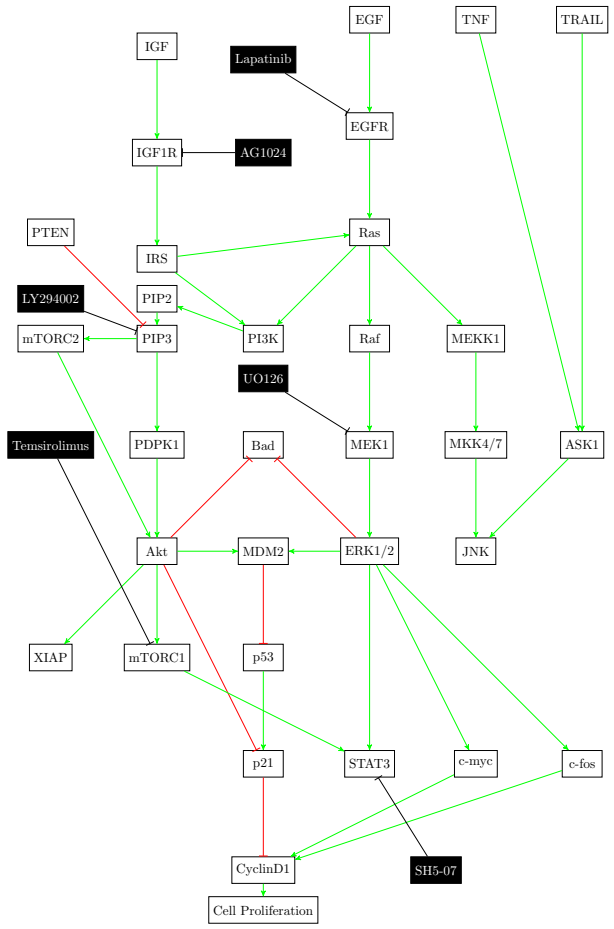

Fig. 1. JNK, p53, PI3K/AKT/mTOR and MAPK/ERK pathways.

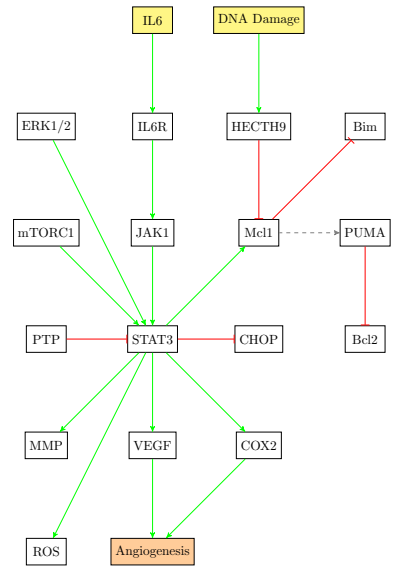

Fig. 2. STAT3 pathway.

We model the biological signaling pathways that we have discussed in the background section as a Boolean network. Each gene is a node and its direct interaction with another gene is represented as an edge. Gene expression is binarily quantized: a gene, if expressed is considered to be

ON (State 1) and if not expressed, is considered to be OFF (State 0). If two or more genes interact to activate or inhibit a third gene, such relationships are modelled with the use of logic gates. The genetic regulatory network can then be thought of as a multi-input multi-output (MIMO) digital logic circuit.

For the drug target interaction prediction, DTINet uses the information from the heterogeneous network comprises edges representing different relationships between drugs, diseases and proteins. In this paper, we look at protein-disease and drug-disease association in particular, since this information is readily available for natural compounds. First, the similarity matrices for these association are calculated using jaccard similarity measure. Next, the algorithm performs a random walk across this heterogenous network to obtain a compact feature. The random walk with restart (RWR) involves the minimization of Kullback-Leibler (KL) divergence between observed diffusion states and those estimated from the network. This feature representation is obtained by taking a projection from drug space to target space using drug and protein information and their interaction matrix. Finally, the inductive matrix completion is performed using the method in Natarajan and Dhillon (2014). Inductive matrix completion is the method to find low rank decomposition of a sparse matrix whose entries are missing. This is similar to the algorithm that recommender systems use in sites such as NetFlix.

Assume that $M = WH^T$ is low-rank where $W \in \mathcal{R}^{N_g \times k}$ and $H \in \mathcal{R}^{N_d \times k}$ are of rank $k << m, n$.
Apply the model on the interaction matrix $P \approx WH^T$, we could solve the following optimization problem:

$$\min_{W,H} \sum_{i,j} (P_{ij} - x_i W^T H y_j^T) + \frac{\lambda}{2}(||W||_F^2 + ||H||_F^2)$$

For the purpose of our paper, the drug-protein interaction matrix is similar to M and the drug features form a vector X, protein features Y such that $P = XMY^T$. W and H are the obtained components and the predicted interaction can be calculated by using them to calculate P.

We add to the datasets used in DTINet Luo *et al.* (2017) the information relevant to Cryptotanshinone. The important proteins under investigation currently that make these pathways are all present in the database, and the ones that are not are those that can be eliminated from the model or their relevant information has been added wherever required.

4 Discussion

The results obtained from DTINet were not satisfactory. The positive control failed and it did not predict strong association of Cryptotanshinone with STAT3. Finally, the data was changed to reflect STAT3 as a target of Cryptotanshinone, and then the Boolean modeling was performed.

The next simulation was run to test which single drug is the most effective in combination with Cryptotanshinone. The results are shown in Figure 5. The theoretical efficacies for these drug combinations have thus, been estimated in terms of their capability to induce apoptosis.

The apoptosis ratio is a measure of the relative change in apoptosis upon a change in conditions. The apoptosis ratio will change depending on different factors such as the values of the inputs, the presence of certain faults or the application of a drug. Changing the input combination to the Boolean network will change the value of the apoptosis ratio Saraf *et al.* (2018).

The results obtained from the Boolean network were then compared with the results of the biological experiments outlined in Saraf *et al.* (2018).

The cellular apoptosis occurring in osteosarcoma cells with respect to time is displayed in Figure 1. The Y-axis shows the apoptotic fraction, which corresponds to the percentage of apoptosis occurring in the cell line in the given time. Table 1explains the legend in Figure 1 in greater detail.

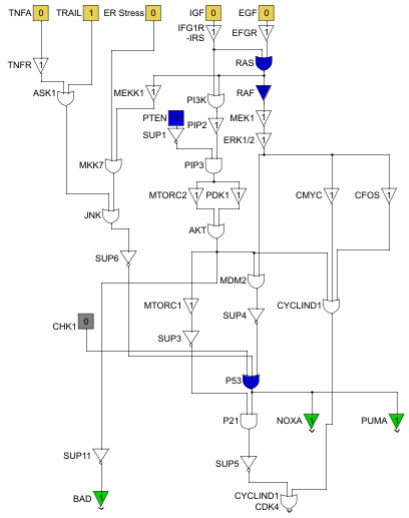

Fig. 3. Boolean Network for the cell survival pathway.

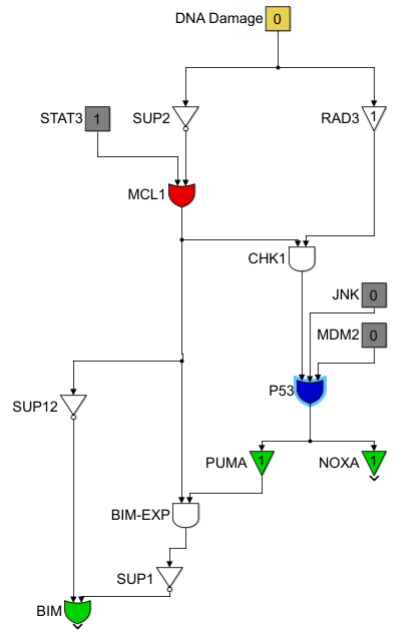

Fig. 4. Boolean Network for the p53 pathway.

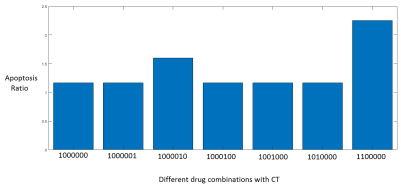

Fig. 5. Apoptosis by CT in combination with a single drug.

5 Conclusion

The comparison of the results from the Boolean network and the biological experiment tells us that the assumption that STAT3 is the target of

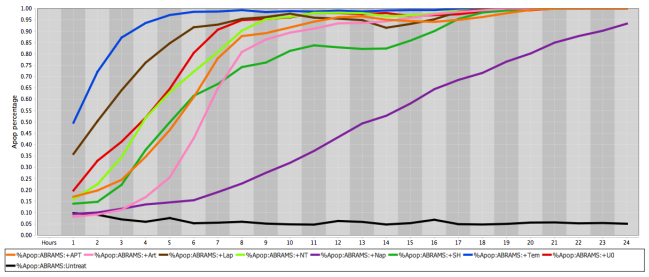

Fig. 6. Experimental results for each single drug in combination with CT.

Table 1. Legend for Figure 1

Abbreviation	Drug Combination
Cry	Cryptotanshinone $50\mu M$
+Ly	LY294002 $10\mu M+$ Cryptotanshinone $50\mu M$
+Tem	Temsirolimus $10\mu M+$ Cryptotanshinone $50\mu M$
+U0	U0126 $10\mu M+$ Cryptotanshinone $50\mu M$
+Lap	Lapatinib $10\mu M+$ Cryptotanshinone $50\mu M$
+SH	SH5-07 $10\mu M+$ Cryptotanshinone $50\mu M$
+AG	AG1024 $10\mu M+$ Cryptotanshinone $50\mu M$
Untreat	Untreated

Cryptotanshinone is valid. The discrepancies are not large. Better results could be obtained by finding other possible targets of Cryptotanshinone. There is a need to collect more information about the drug, such as its binding affinities with kinases.

The results of the DTINet prediction need improvement to achieve better target prediction. The dataset used needs to contain more information about natural drugs. Additionally, there exist newer methods to learn the features that could be employed, such as Wan *et al.* (2018).

Acknowledgements

I would like to thank Dr Shen for his constant guidance, Dr Datta for his advise on the osteosarcoma modeling and Dr Bittner for performing the necessary biological experiments.

Funding

This work has been supported by the Texas A&M University.

References

Chen, W., Lu, Yinand Chen, G., and Huang, S. (2011). Cryptotanshinone has diverse effects on cell cycle events in melanoma cell lines with different metastatic capacity. *Cancer Chemotherapy and Pharmacology*, **68**, 17–27.

Chiarini, F., Grimaldi, C., Ricci, F., Tazzari, P., Iacobucci, I., Martinelli, G., Pagliaro, P., McCubrey, J., Amadori, S., and Martelli, A. M. (2011). Temsirolimus, an allosteric mtorc1 inhibitor, is synergistic with clofarabine in aml and aml leukemia initiating cells. *Blood*, **118**, 2596.

Corazza, N., Kassahn, D., Jakob, S., Badmann, A., and Brunner, T. (2009). Trail-induced apoptosis: Between tumor therapy and immunopathology. *Annals of the New York Academy of Science*, **1171**, 50–âŁ“58.

Flashner-Abramson, E., Klein, S., Mullin, G., Shoshan, E., Song, R., Shir, A., Langut, Y., Bar-Eli, M., Reuveni, H., and Levitzki, A. (2016). Targeting melanoma with nt157 by blocking stat3 and igf1r signaling. *Oncogene*, **35**, 2675âŁ“–2680.

Fleury, C., Mignotte, B., and VayssiÃ¨re, J.-L. (2002). Mitochondrial reactive oxygen species in cell death signaling. *Biochimie*, **84**(2-3), 131–141.

Isakoff, M. S., Bielack, S. S., Meltzer, P., and Gorlick, R. (2015). Osteosarcoma : Treatment and a collaborative pathway to success. *Journal Of Clinical Oncology*, **33**, 3029–3035.

Jamil, S., Mojtabavi, S., Hojabrpour, P., Cheah, S., and Duronio, V. (2008). An essential role for mcl-1 in atr-mediated chk1 phosphorylation. *Molecular Biology of the Cell*, **19**(8), 3212–âŁ“3220.

Kanehisa, M. and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**(1), 27âŁ“30.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2016a). Kegg: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, **45**(1), D353–âŁ“D361.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016b). Kegg: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, **44**(1), D457–âŁ“D462.

Liu, H., Ma, Y., Cole, S. M., Zander, C., Chen, K.-H., Karras, J., and Pope, R. M. (2003). Serine phosphorylation of stat3 is essential for mcl-1 expression and macrophage survival. *Blood*, **102**(1), 344–352.

Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, Chen, L., and Zeng, J. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications*, **8**(573).

Mendoza, M. C., Er, E. E., and Blenis, J. (2012). The ras-erk and pi3k-mtor pathways:cross-talk and compensation. *Trends in Biochemical Sciences*, **36**(6), 320–328.

Natarajan, N. and Dhillon, I. S. (2014). Inductive matrix completion for predicting geneâŁ“disease associations. *Bioinformatics*, **30**(12), i60–i68.

Saleiro, D. and Platanias, L. C. (2015). Intersection of mtor and stat signaling in immunity. *Trends in Immunology*, **36**(1), 21–29.

Saraf, R. S., Datta, A., Sima, C., Hua, J., Lopes, R., and Bittner, M. (2018). An in-silico study examining the induction of apoptosis by cryptotanshinone in metastatic melanoma cell lines. *BMC Cancer*, **18**(855).

Tait, S. W. G. and Green, D. R. (2010). Mitochondria and cell death: outer membrane permeabilization and beyond. *Molecular Cell Biology*, **11**(9), 621–632.

Thomas, L. W., Lam, C., and Edwards, S. W. (2010). Mcl-1; the molecular regulation of protein function. *FEBS Letters*, **584**(14), 2981–âŁ“2989.

Wan, F., Hong, L., Xiao, A., Jiang, T., and Zeng, J. (2018). Neodti: Neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. *Bioinformatics*.

Wei, Y.-h., Tang, H.-x., Liao, Y.-d., Fu, S.-l., Xu, L.-q., Chen, G., Zhang, C., Ju, S., Liu, Z.-g., You, L.-k., Yu, L., and Zhou, S. (2015). Effects of insulin-like growth factor 1 receptor and its inhibitor ag1024 on the progress of lung cancer. *Journal of Huazhong University of Science and Technology [Medical Sciences]*, **35**(6), 834–841.

Yue, P., Lopez-Tapia, F., Paladino, D., Li, Y., Chen, C.-H., Namanja, A. T., Hilliard, T., Chen, Y., Tius, M. A., and Turkson, J. (2016). Hydroxamic acid and benzoic acid–based stat3 inhibitors suppress human glioma and breast cancer phenotypes in vitro and in vivo. *Cancer Research*, **76**(3), 652–663.