Application of network diffusion approaches to drug screenings: A perspective on multi-layered networks derived from drugs and cell lines

Vigneshwari Subramanian[1], Bence Szalai[1], Luis Tobalina[1], Julio Saez-Rodriguez[1,2#]

[1]RWTH Aachen University, Faculty of Medicine, Joint Research Centre for Computational Biomedicine, MTZ, Pauwelsstrasse 19, 52074-Aachen, Germany
[2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus CB10 1SD Hinxton, United Kingdom

#Corresponding author (saezrodriguez@combine.rwth-aachen.de)

Introduction
In the big data era, network diffusion approaches are frequently used to gain more insights on networks representing interactions between biological macromolecules. Successful applications of network diffusion approaches in identifying relevant disease genes and prioritizing genes for drug sensitivity predictions have been reported in literature[1,2,3,4]. However, the majority of these studies rely on networks representing a single type of information. A recent publication showed that using networks with interconnected layers, each layer representing a different type of information, allowed to highlight genes for diseases that are not well-studied[5]. These multi-layered networks are known as multiplex heterogeneous networks. Motivated by this study, we built a multi-layered network that incorporates information on protein-protein interactions, drug-drug similarities, cell line-cell line similarities and co-expressed genes, and used it to investigate the interactions between drugs, targets and cancer cell lines.

Methods
The data required for constructing co-expression, drug-drug and cell line-cell line similarity networks was retrieved from the Genomics of Drug Sensitivity in Cancer (GDSC)[6] Web portal. The Protein-Protein Interaction (PPI) network was generated based on the data retrieved from Reactome, Biocarta, KEGG, InnateDB and Uniprot databases. Bipartite associations (Drug-Target, Drug-Cell line and Cell line – Mutation / CNV), which define the interactions between the different types of networks, were generated using the information extracted from GDSC and ChEMBL. Random Walk with Restart algorithm was applied to these multiplex heterogeneous networks (RWR-MH) using drug nodes as seeds for the restart step. Different combinations of networks and bipartite associations were explored.

Results
Using the RWR-MH algorithm, we retrieved a prioritized list of genes associated to each drug. The results from ANOVA models testing for the significance of the association between the genomic status (mutation / CNV) of the gene and the drug response show these prioritized genes to be among the most significant ones. Drug response prediction models built using gene expression data of highly ranked genes are as predictive as those built using all the available genes. Taken together, these results show that the multiplex heterogeneous network-based approach is efficient in identifying the most relevant genes related to drug response. Additionally, the method allows to retrieve gene-drug associations even when the drug's mode of action is not well-characterized. Further, extending this approach to patient data by generating patient-cell line networks could aid patient-specific drug response predictions.
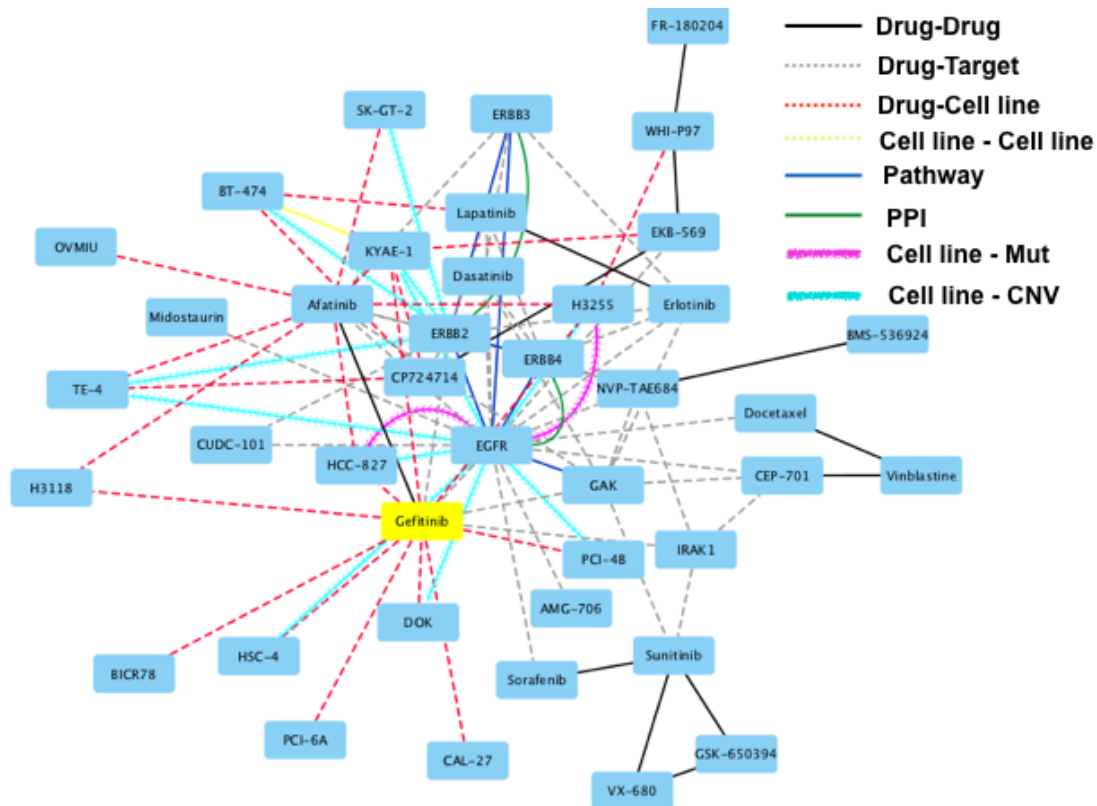
Figure 1. Example of an interaction network involving the top 20 drugs / genes obtained using Gefitinib as the seed node.

References

1. Emad, A., Cairns, J., Kalari, K.R. et al. (2017) Genome Biol. 18, DOI:10.1186/s13059-017-1282-3
2. Franke, L., Bakel, H.V., Fokkens, L. (2006) Am. J. Hum Genet. 78 (6), pp. 1011-1025.
3. Gorge, R.A., Liu, J.Y., Feng, L.L. (2006) Nucleic Acids Res. 34 (19), e130.
4. Li, Y., Patra, J.C. (2010) Bioinformatics 26 (9), pp. 1219-1224.
5. Valdeolivas, A., Tichit, L., Navarro, C. et al. (2017) bioRxiv. DOI:10.1101/134734
6. Iorio, F., Knijnenburg, T.A., Vis, D.J. et al. (2016) Cell, 166(3), pp. 740-754.