

Poster presentation

Quantitative systems-level determinants of drug targets

Lixia Yao*¹ and Andrey Rzhetsky²

Address: ¹Department of Biomedical Informatics, Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA and ²Department of Medicine, Department of Human Genetics, Institute for Genomics and Systems Biology, and Computation Institute, University of Chicago, IL 60637, USA

Email: Lixia Yao* - ly2142@columbia.edu

* Corresponding author

from Fourth International Society for Computational Biology (ISCB) Student Council Symposium
Toronto, Canada. 18 July 2008

Published: 30 October 2008

BMC Bioinformatics 2008, 9(Suppl 10):P6 doi:10.1186/1471-2105-9-S10-P6

This abstract is available from: <http://www.biomedcentral.com/1471-2105/9/S10/P6>

© 2008 Yao and Rzhetsky; licensee BioMed Central Ltd

Background

Modern drug discovery tends to understand disease processes at the molecular level and then determine optimal molecular targets for drug intervention. Inferences have been made from all available drug targets, such as how many drug targets there are, or how many novel drug targets could be potentially found in the human genome, to what functional families these proteins belong, and what structural properties make them bind to small molecules tightly and specifically. But all these are very intuitive and qualitative. The key question of which gene or protein in a disease process could be a successful drug target remains unanswered.

Results

We analyzed specific systems-level properties of human genes and proteins targeted by 919 FDA-approved drugs [1] and identified a number of quantitative measures that distinguish them from other genes and proteins at a highly significant level. Compared to an average gene and its encoded protein(s), successful drug targets are more highly connected in a molecular interaction network, but are far from being the most highly connected; they have higher betweenness values, lower entropies of tissue expression, and lower ratios of non-synonymous to synonymous single-nucleotide polymorphisms (see Figure 1). We also tested the performance of different classification algorithms (see Figure 2). Furthermore, we have identified human tissues significantly over- or under-

targeted relative to the full spectrum of genes active in each tissue.

We also built a machine-learning model to demonstrate the usefulness of these quantitative descriptors for predicting drug targets. With increasing availability of experimental data, we foresee that screening the whole human genome for potential novel drug targets could be feasible in near future.

Conclusion

We found that genes associated with successful FDA-approved drugs have a number of properties at the network, sequence, and tissue-expression levels that significantly distinguish them from other human genes. Although the drug-target-selection guidelines that we suggest cannot replace expensive experiments, they can help pharmaceutical researchers narrow the prospective set of drug targets at the earliest stage of a drug development project. Specifically, when the pharmaceutical company must decide which target to pursue among pathologic pathways that are not fully understood, connectivity, betweenness, C_{ratio} , and entropy might be useful quantitative estimates of each prospective target's expected success rate.

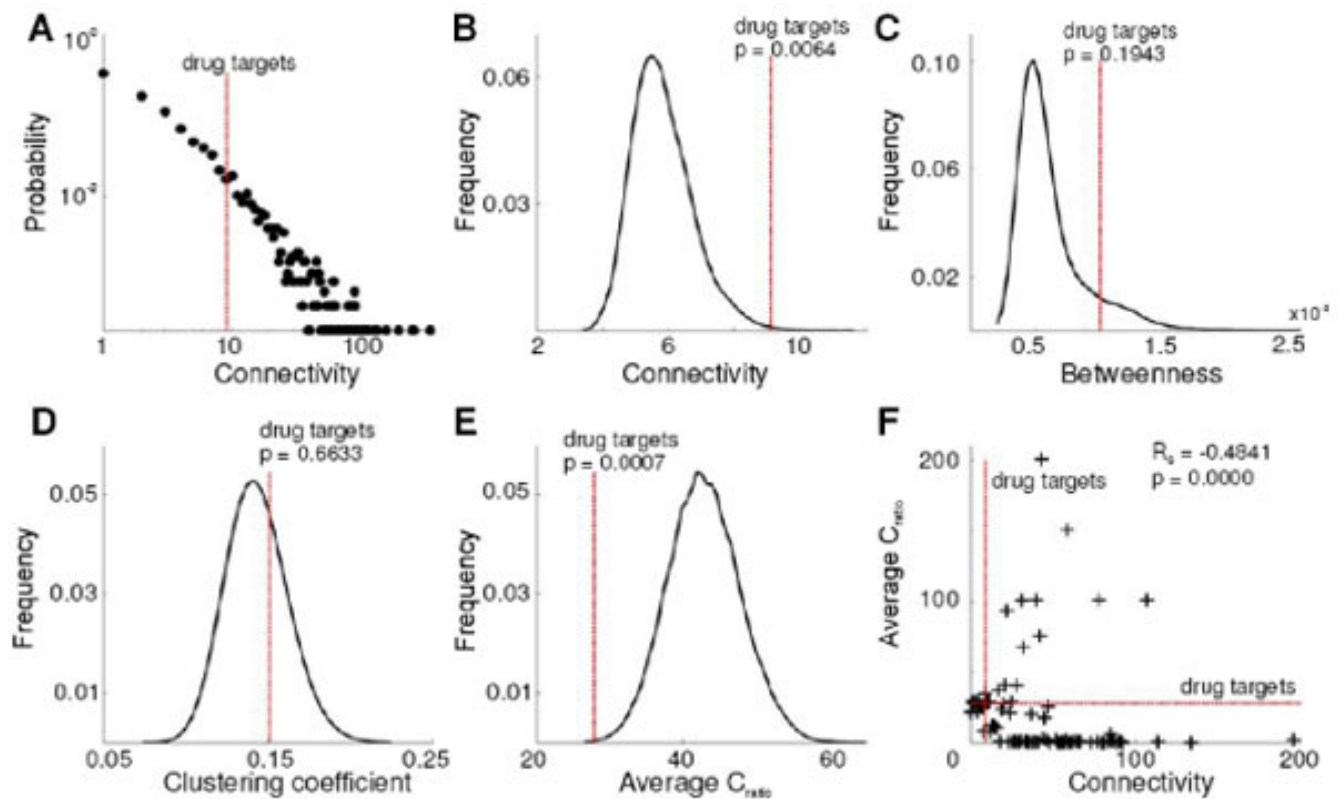


Figure 1

(A) Connectivity distribution for the entire molecular network; (B) Targets of the successful drugs are significantly more connected than an average gene in the network, but are not the most highly connected genes in the network; (C) Drug targets tend to have higher than average betweenness values; (D) The successful drug targets are not statistically different from the rest of the genes in terms of their clustering coefficients; (E-F): Analysis of the ratio (C_{ratio}) of the number of non-synonymous to synonymous single-nucleotide polymorphisms (SNPs): (E) Successful drug targets have significantly smaller C_{ratio} than human genes on average. (F) The value of C_{ratio} tends to correlate negatively with the gene's connectivity in the network.

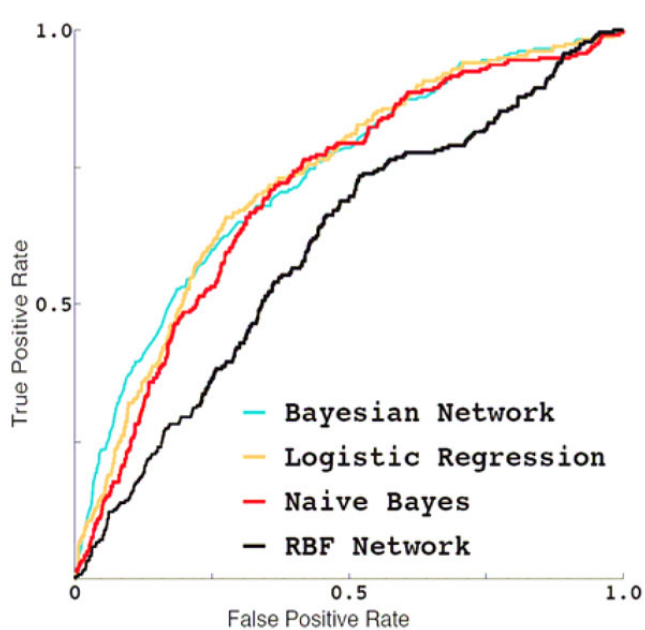


Figure 2
Receiver operating characteristic (ROC) curves for the four classification algorithms that we used in this study. All four methods that we tested performed significantly better than baseline (ROC score of 0.5, corresponding to a random-guess method). The logistic regression performed best.

References

1. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J: **DrugBank: A comprehensive resource for in silico drug discovery and exploration.** *Nucleic Acids Res* 2006, **34**:D668-D672.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."
Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

