A Cancer Gene-Drug Connectivity Map for DrBioRight

Varshini Vakulabharanam, Hu Chen, Yitao Tang, Jun Li, and Han Liang

Department of Bioinformatics and Computational Biology,

The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center

Making Cancer History®

Background

High-throughput molecular profiling technologies have revolutionized biomedical research and generated various omics data for thousands of tumor and cell-line samples in cancer research. To better utilize such data and maximize its utility, in our previous study, we developed a natural language-oriented, artificial intelligence-based analytics platform “DrBioRight” for exploring and analyzing cancer omics data.

Using DrBioRight, one can perform bioinformatics analysis simply by asking biological questions in day-to-day language. Since initial release, DrBioRight has been widely used by >10,000 users from >100 countries. Although integrated with popularly used analytic and visualization modules (e.g., correlation, differential and survival analysis), DrBioRight covers only a limited number of common bioinformatics tools. To address this challenge and increase the utility of DrBioRight, we plan to enlarge its library of analytic and visualization modules and compile more public omics datasets.

In this study, we specifically built a gene-drug connectivity map based on their interactions derived from the public Drug Gene Interaction Database (DGIdb). DGIdb provides gene-drug interactions that are curated by experts’ manual curation and text-mining from public resources, including Drug Target Commons and DrugBank. Such interaction information can help with easily identifying potential drug targets or check whether a cancer-related gene is targetable by a drug.

By incorporating such information, our users can visualize the associated drugs for each queried gene and readily find the corresponding publication information for each gene-drug interaction.

Results

Figure 1 shows the number of interacted drugs for each gene in which there is one gene associated with 770 drugs. Figure 2 shows the number of genes for each interacted drug. For example, there are >2000 drugs only targeting a single gene. Figures 3 and 4 summarize the top genes and drugs ranked by their degrees in the interaction network; the gene NFE2L2 and the drug cetuximab have the highest number of neighbors. Figure 5 shows a sub-connectivity map constructed by the top genes and drugs identified from Figures 3 and 4. Each node represents a drug or gene; each edge represents an interaction claim source, such as Clinical Interpretation of Variants in Cancer, The Druggable Genome Clinical Trial, Targeted Agents in Lung Cancer, “My Cancer Genome Clinical Trial,” “Clearity Foundation Biomarkers,” and Database of Curated Mutations.

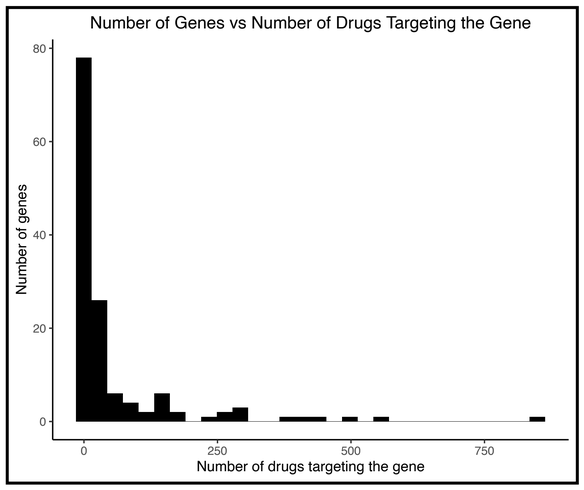

Fig. 1 Drug-gene targets histogram generated by Rstudio.

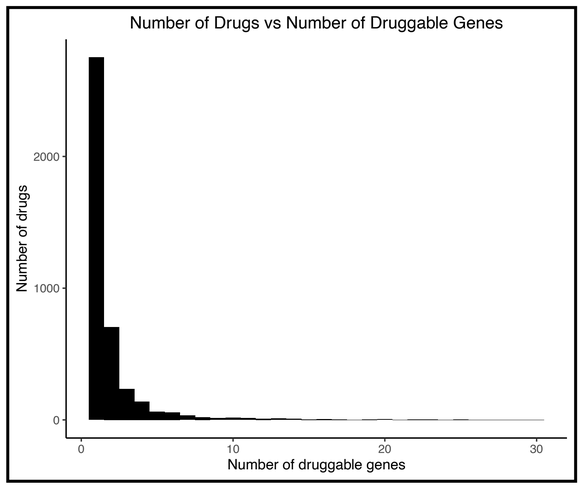

Fig. 2 Druggable genes histogram generated by Rstudio.

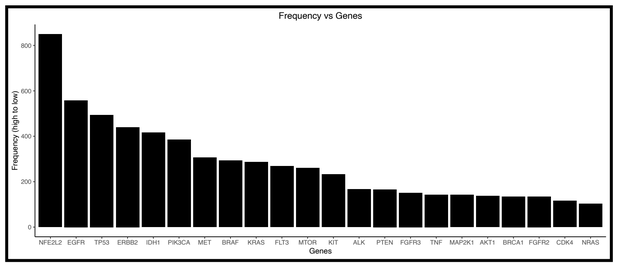

Fig. 3 Gene frequency bar graph generated by Rstudio.

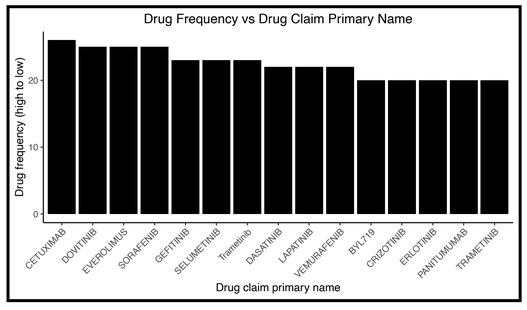

Fig. 4 Drug frequency bar graph generated by Rstudio.

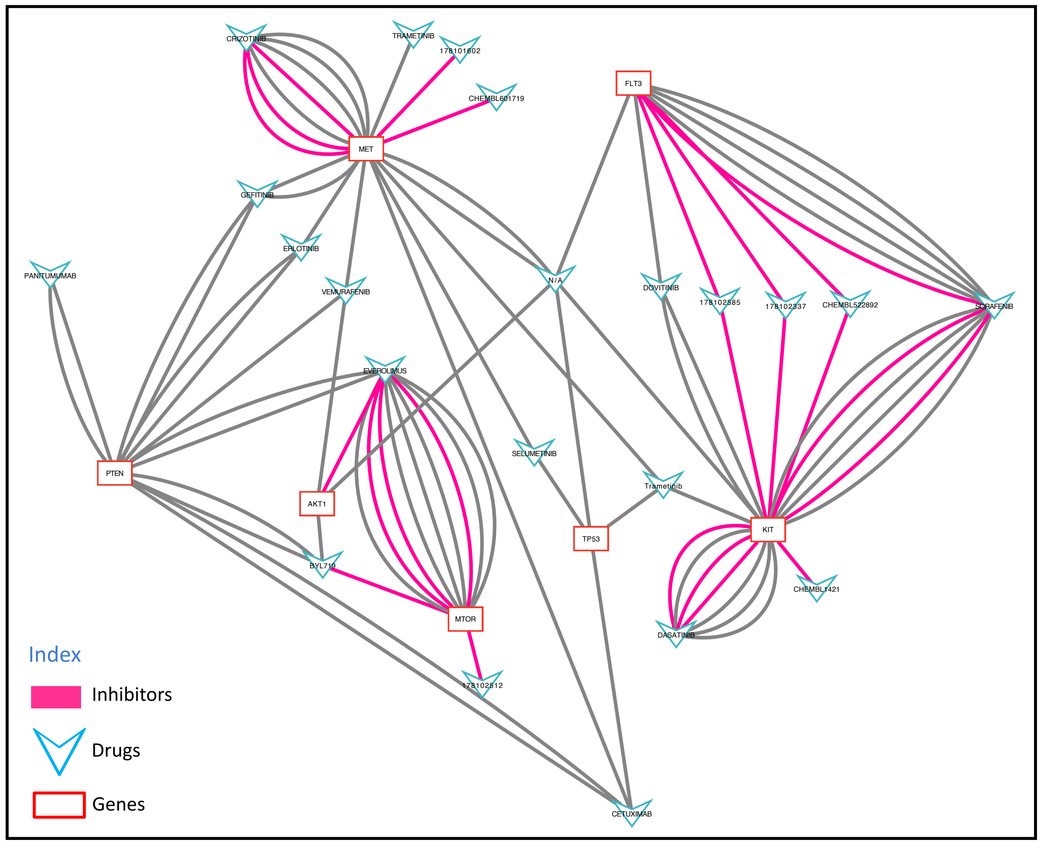

Fig. 5 Connectivity map generated by Cytoscape.

Methods

1. 261 cancer genes were obtained from TCGA highly mutated cancer genes.
2. Interactions (~80,0000) for the 261 cancer genes were filtered from DGIdb.
3. Two histograms were generated to show the gene and drug association.
4. High frequency genes and drugs were sorted to generate the two lists by using cutoff of greater than or equal to 100 for genes and greater than or equal to 20 for drugs.
5. Bar plots were created to show two lists from output of step four.
6. 82 interactions were extracted based on high frequency genes and drugs from step four.
7. Data from step six was used to generate connectivity map using Cytoscape.

Conclusions

Our results serve as the foundation in building a new module for DrBioRight. This new module will help users obtain information on drug-gene interactions conveniently in DrBioRight, rather than by referencing other databases. The new module for DrBioRight will consolidate the data generated in this study to further increase utility and improve the user experience.

Acknowledgements

I would like to thank Dr. Liang and Dr. Li for their support and guidance. I would like to thank my graduate mentors Hu and Yitao for all of their assistance. I would also like to thank the MD Anderson Careers in Cancer Science and Medicine Program for providing me such a great opportunity.

References

1) Li et al. Cancer Cell 2021;39:3-6
2) Cotto et al. Nucleic Acids Research 2018;46:D1068-D1073