Bioinformatics, 35(7), 2019, 1247–1248
doi: 10.1093/bioinformatics/bty763
Advance Access Publication Date: 31 August 2018
Applications Note

Systems biology

Drug Gene Budger (DGB): an application for ranking drugs to modulate a specific gene based on transcriptomic signatures

Zichen Wang, Edward He, Kevin Sani, Kathleen M. Jagodnik, Moshe C. Silverstein and Avi Ma'ayan *

Department of Pharmacological Sciences, BD2K-LINCS Data Coordination and Integration Center, Knowledge Management Center for Illuminating the Druggable Genome, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on May 10, 2018; revised on July 27, 2018; editorial decision on August 24, 2018; accepted on August 29, 2018

Abstract

Summary: Mechanistic molecular studies in biomedical research often discover important genes that are aberrantly over- or under-expressed in disease. However, manipulating these genes in an attempt to improve the disease state is challenging. Herein, we reveal Drug Gene Budger (DGB), a web-based and mobile application developed to assist investigators in order to prioritize small molecules that are predicted to maximally influence the expression of their target gene of interest. With DGB, users can enter a gene symbol along with the wish to up-regulate or down-regulate its expression. The output of the application is a ranked list of small molecules that have been experimentally determined to produce the desired expression effect. The table includes log-transformed fold change, *P*-value and *q*-value for each small molecule, reporting the significance of differential expression as determined by the limma method. Relevant links are provided to further explore knowledge about the target gene, the small molecule and the source of evidence from which the relationship between the small molecule and the target gene was derived. The experimental data contained within DGB is compiled from signatures extracted from the LINCS L1000 dataset, the original Connectivity Map (CMap) dataset and the Gene Expression Omnibus (GEO). DGB also presents a specificity measure for a drug–gene connection based on the number of genes a drug modulates. DGB provides a useful preliminary technique for identifying small molecules that can target the expression of a single gene in human cells and tissues.

Availability and implementation: The application is freely available on the web at http://DGB.cloud and as a mobile phone application on iTunes https://itunes.apple.com/us/app/drug-gene-budger/id1243580241?mt=8 and Google Play https://play.google.com/store/apps/details? id=com.drgenebudger.

Contact: avi.maayan@mssm.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Recent high-throughput genome-wide expression-based drug screening have generated large collections of drug-induced transcriptomic signatures. The LINCS L1000 (Subramanian *et al.*, 2017) dataset, and its precursor the original Connectivity Map (CMap) (Lamb *et al.*, 2006), have systematically profiled the transcriptomic response of several human cell lines to treatment with over 20 000 small molecules that include almost all FDA-approved drugs and many preclinical compounds. Crowdsourcing efforts have also been made to curate hundreds of drug-induced gene expression signatures from the Gene

Expression Omnibus (GEO) (Wang *et al.*, 2016). By integrating and analyzing these datasets, we can prioritize small molecules that are found to significantly modulate single genes. Herein, we present Drug Gene Budger (DGB), a web-based and mobile application utilizing the large collections of aforementioned drug-induced transcriptomic signatures for prioritizing drugs and small molecule compounds to maximally influence the expression of a target gene of interest. DGB provides users with straightforward user interface to select a target gene, and interact with the ranked list of small molecules returned as the query results. Small molecules are ranked by different methods that quantify the previously observed change in mRNA expression of the gene by the application of the drug. We also benchmarked different methods to provide recommendations for selecting the most appropriate methods.

2 Materials and methods

2.1 Processing of transcriptomic datasets into signatures

To identify differentially expressed genes (DEGs) before and after drug treatment, batch effects were removed. ComBat (Johnson *et al.*, 2007) was used for processing the original CMap dataset (Lamb *et al.*, 2006) as well as the LINCS L1000 dataset (Subramanian *et al.*, 2017). SVA (Leek *et al.*, 2012) was applied to remove batch effects for the datasets curated from GEO (Wang *et al.*, 2016). The reason for using two different methods is because for the original CMap and LINCS L1000, the batches are known, whereas for the signatures from GEO the batches are unknown. Next, Limma (Ritchie *et al.*, 2015) was applied to evaluate the statistical significance (*P*-value) of differential expression for each gene. *P*-values were corrected using the Benjamini–Hochberg procedure to yield *q*-values adjusted for multiple hypothesis testing. Genes with *q*-values less than 0.05 were kept as DEGs. The specificity of a drug–gene association in each experiment is defined as the inverse of the number of identified DEGs for the drug.

2.2 Benchmarking small-molecule prioritization methods

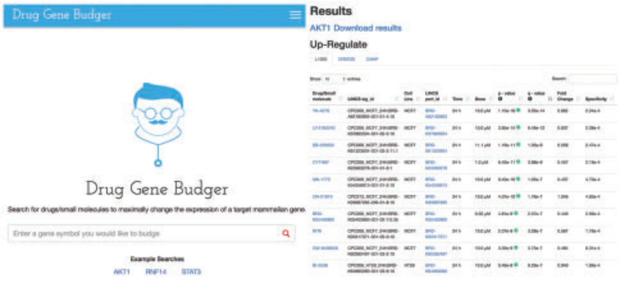
The various prioritization methods were benchmarked using drug-target and protein–protein interactions (PPIs) background knowledge. Briefly, the known protein targets of the drugs were retrieved from DrugBank (Law *et al.*, 2014). PPIs of the protein targets are identified from an updated version of the low-content PPI network aggregated from multiple high-quality resources (Clarke, 2018). To set up the benchmark, for each gene queried, we sorted the drug–gene associations by the different methods and recorded the ranks for drugs that are known to target the gene product protein of the queried gene, or its direct PPIs. The ranks of the expected drugs were scaled between 0 and 1, and then aggregated across different query genes. The cumulative distribution of such scaled ranks were used to evaluate how well each method performs in prioritize drugs for up- or down-regulating their known targets, or the PPI neighborhood around the target.

2.3 Development of DGB web and mobile applications

The DGB web application and mobile applications use Python Flask web framework as the backend. Custom object-relation mappings (ORMs) were written to query the drug–gene associations stored in a MySQL database. The frontend of DGB uses the Bootstrap HTML framework. The mobile version of the application is developed with the React Native framework.

3 Results

After the processing of drug-induced transcriptomic datasets from the three resources: Original CMap, LINCS L1000 and GEO, we obtained 4810 drugs and small molecule compounds and 36 523 017 significant drug–gene associations. To quantify the strength of the up-/down-regulation relationships between a drug and a gene, we implemented the following methods: fold change, *P*-value and *q*-value computed by limma, and the specificity. Benchmarking these methods with prior knowledge about drug targets and PPIs, we observed that *q*-values best prioritize the expected drugs for a given query gene, followed by *P*-value, specificity, and FC for both up- or down-regulated query genes (Supplementary Figs S1 and S2). The DGB web application landing page enable users to enter a gene symbol (Fig. 1, left). The result page outputs six tables corresponding to the results from the three processed datasets in both up- and down-regulations. Each table contains a list of small molecules that produced the respective expression effect sorted by the descending order of their *q*-values (Fig. 1, right). Metadata about the compounds and signatures are also provided with external links to acquire more information about the compounds. To demonstrate DGB, we query it with tumour suppressor genes to find drugs and compounds that up-regulate those genes. Many cancer drugs show up on top, for instance, doxorubicin is predicted to up-regulate BRCA1; dabrafenib and GDC-0879 are predicted to up-regulate VHL; and homoharringtonine and triamterene are predicted to up-regulate TP53.



Search for drugs/small molecules to maximally change the expression of a target mammalian gene.

Fig. 1. (Left) DGB search engine. (Right) Example result page for user-submitted query for AKT1

Funding

This work was partially supported by NIH grant U54HL127624, OT3OD025467 and U24CA224260.

Conflict of Interest: none declared.

References

Clarke,D.K. *et al.* (2018) eXpression2Kinases (X2K) web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Res*, **46**, W171–W179.

Johnson,W.E. *et al.* (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.

Lamb,J. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929–1935.

Law,V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research*, **42**, D1091–D1097.

Leek,J.T. *et al.* (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, **28**, 882–883.

Ritchie,M.E. *et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.

Subramanian,A. *et al.* (2017) A next generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell*, **171**, 1437–1452.e17.

Wang,Z. *et al.* (2016) Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat. Commun.*, **7**, 12846.