MOSAIC: a chemical-genetic interaction data repository and web resource for exploring chemical modes of action

Justin Nelson[1], Scott Simpkins[1] and Chad L. Myers[2,*]

[1]Bioinformatics and Computational Biology Program, University of Minnesota-Twin Cities, Minneapolis, Minnesota 55455, USA [2]Department of Computer Science and Engineering, University of Minnesota-Twin Cities, Minneapolis, Minnesota 55455, USA

*To whom correspondence should be addressed.

Abstract:

Summary: Chemical genomic approaches that map interactions between compounds and genetic perturbations offer a promising new strategy for annotation of uncharacterized bioactive compounds. We recently developed a new high-throughput platform for mapping chemical-genetic interactions in yeast that can be scaled to screen large compound collections, which has been applied to generate chemical genetic (CG) profiles for more than 13,000 compounds. When integrated with existing yeast genetic interaction network data, these CG interaction profiles can enable the prediction of modes of action for previously uncharacterized compounds as well as the discovery of new effects for approved drugs. To facilitate future analysis of these valuable data, we developed a public database and web interface named MOSAIC. The website provides a convenient interface for querying compounds, genes, or biological processes (GO terms) for direct chemical genetic interactions or target predictions. MOSAIC also provides direct access to chemical structural information of screened molecules and the ability to search for compounds sharing structural similarity. This resource will be of interest both to chemical biologists interested in discovering new probes with specific modes of action as well as computational biologists interested in analyzing chemical genetic interaction networks.

Availability: MOSAIC is available at http://mosaic.cs.umn.edu.

Contact: chadm@umn.edu

1 Introduction

Discovering the mode-of-action of compounds is an important step in developing effective treatments for diseases or controlling the spread of dangerous pathogens. Unfortunately, recent trends show that the cost of healthcare continues to increase, partially driven by increasing costs related to drug development and research (Kleinke, 2001). However, there is an increasing number of candidate drug targets available to treat diseases. This fundamental tension between the breadth of candidate drug targets we are able to discover and an increasingly sluggish and expensive therapeutic development process that is not able to capitalize on these targets suggests we need to develop new paradigms for drug discovery. A survey of the standard methods in high-throughput screening of small-molecules reveals one potential source of the problem: almost all current strategies are based on a "bottom-up" approach. For example, the standard paradigm for screening chemical libraries typically involves developing a specific biochemical assay (traditionally in vitro) to detect activity against a desired target and then applying it on a large scale to a large library of candidate small molecules (Bibette, 2012; Macarron *et al*., 2011). Consequently, such an approach requires a significant commitment to the target protein or pathway before one develops a specific assay. Thus, it is not surprising that we have yet to build a comprehensive collection of small molecules that are able to target a significant fraction of the human genome.

Chemical genomic screens offer a compelling "top-down" alternative to the traditional approach of high-throughput screening of small molecules. Unlike traditional high-throughput screens designed around detecting compounds that hit a pre-defined target, chemical genomic approaches do the opposite by revealing compound-dependent phenotype(s) in a systematic, unbiased manner and can be used to predict the target of an uncharacterized bioactive compound. A series of these methods utilize a complete barcoded gene deletion collection, which is available for the genetically tractable yeast Saccharomyces cerevisiae (Giaever *et al*., 1999; Lee *et al*., 2014; Parsons *et al*., 2004; Parsons *et al*., 2006). The deletion mutants that comprise the collection are typically assayed for sensitivity or resistance to a particular compound of interest to generate a functional profile. Methods have been developed for both homozygous deletion and heterozygous deletion collections and have incorporated large-scale biological networks to make mode-of-action predictions (Giaever *et al*., 1999; Hillenmeyer *et al*., 2008; Lee *et al*., 2014). We recently developed a new high-throughput platform for mapping chemical-genetic interactions in yeast that can be scaled to screen large compound collections, which has been applied to generate chemical genetic profiles (Piotrowski *et al*., 2017). This represents the largest chemical genomic dataset to date, including profiles for more than 13,000 natural products, small molecules or clinically relevant compounds. Here, we present MOSAIC, a database website which allows access to this rich chemical genomic dataset.

2 Methods

The website for accessing the database is located at http://mosaic.cs.umn.edu. MOSAIC displays data obtained by screening the RIKEN NPDepo Library as well as compound collections from the NIH, NCI and a GSK Kinase Inhibitor collection (Piotrowski *et al.*, 2017). The data housed in the database was derived from the approach described in (Piotrowski *et al*., 2017) and has been analyzed using the BEAN-counter (Simpkins, 2017) and CG-TARGET (Simpkins *et al.*, 2017) tools.

The website was designed using a combination of hypertext markup language (HTML), javascript with jquery, asynchronous javascript, and XML (AJAX), javascript object notation (JSON), PHP and MySQL. The website uses PHP to deliver an HTML page to the client, with javascript and jquery for client side scripting. In order to retrieve data from the MySQL database, the website utilizes AJAX to run a PHP script on the server that queries the MySQL database and returns a JSON object back to the client to be displayed. This provides a quick environment for querying the chemical genetic interaction data.

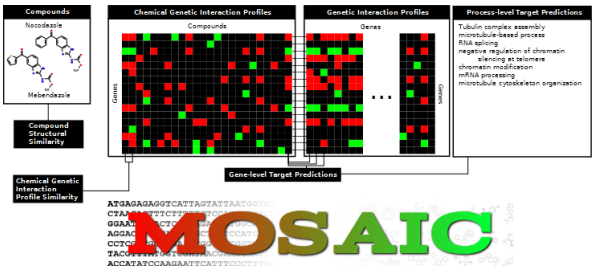

Fig. 1. Overview of data sources and mode-of-action predictions for the MOSAIC website. These data sources and mode-of-action predictions include chemical genetic interaction profiles, gene-level target predictions, process-level target predictions, compound structural similarity and chemical genetic interaction profile similarity. Chemical genetic interaction profiles are generated using BEAN counter. Gene-level target predictions come from similarities between the chemical genetic interaction profiles and genetic interaction profiles. Process-level target predictions are Gene Ontology enrichments within Gene-level Target Predictions. Compound structural similarity and profile similarity compare compound structures or chemical genetic interaction profiles with other compounds structures and profiles respectively.

3 Usage

MOSAIC provides an interface for querying chemical genomic data, target predictions and process predictions. The MOSAIC interface has a search bar available in the middle or top of the browser that will accept a compound name, a gene common name or ORF, or a gene ontology biological process term. Multiple terms can be input by separating each term by a semicolon. Depending on whether a compound, gene, or process has been queried the results page will return different data and predictions. If a query included more than one term, the terms can be selected by using the tabs (Conditions, Genes, GO Terms) and tables.

3.1 Chemical Genetic Profiles

The sensitivity or resistance of deletion mutants in the presence of a compound is provided as part of the chemical genetic interaction data. MOSAIC provides a table of significant chemical genetic interaction scores for compound queries or genes queries present in the yeast diagnostic pool. A negative chemical genetic interaction corresponds to a deletion mutant that is sensitive to a compound. Conversely, a positive chemical genetic interaction corresponds to a deletion mutant that is resistant to a compound. This data is provided as part of MOSAIC, either in the form of a chemical genetic interaction profile for compound queries or as a list of conditions that a deletion mutant is sensitive or resistant to for gene queries.

3.2 Gene-level Target Prediction

A gene-level target prediction is made based on the similarity between the chemical genetic interaction profile and a genetic interaction profile (Parsons *et al*., 2004; Parsons *et al*., 2006). Compound and gene queries both return the table of relevant predicted molecular target predictions. These molecular target predictions, called Gene-level Target Predictions, are calculated using a dot product between a chemical genetic profile and an $\ell 2$ normalized genetic interaction profile. This score highlights genetic interaction profiles that a chemical genetic interaction profile phenocopies while penalizing chemical genomic profiles with low signal (i.e. few strong chemical-genetic interactions). Chemical genetic interaction profiles that phenocopy a genetic interaction profile are possibly caused by that compound having a targeted effect on the corresponding protein or closely related proteins. Using the MOSAIC website, researchers can quickly browse and form hypotheses based on these potential molecular target predictions.

3.3 Process-level Target Prediction

A Gene Ontology biological process prediction is made based on an enrichment analysis of gene-level target predictions as described in (Simpkins *et al.*, 2017). The table of Gene Ontology biological process predictions is displayed for compound queries and the table of compounds predicted to a Gene Ontology biological process term are displayed for Gene Ontology queries. For example, Nocodazole is a potent tubulin destabilizer known to target the tubulin complex. Aquery of "Nocodazole" will show a process prediction to "tubulin complex assembly" (GO:0007021). These Gene Ontology predictions, called process-level target predictions, are calculated by comparing the observed similarity of the compound's chemical genetic profile with the GO term's genes' genetic interaction profile to an empirical null distribution generated by sampling random sets of genes of the same size (Simpkins *et al.*, 2017). MOSAIC allows researchers to quickly generate a list of compounds predicted to target a specific cellular process of interest, enabling efficient identification of candidate compounds with the desired mode of action.

3.4 Chemical and Profile Similarity

The discovery of similar compounds can be useful for expanding lists of candidate compounds. MOSAIC provides access to both compound structural similarity and chemical genetic interaction profile similarity for compound quries. These two similarity approaches can be compared to understand how compound substructure modulates the effect of a compound on a cell.

4 Conclusion

The MOSAIC database provides convenient access to compound chemical genetic interaction data, mode of action predictions, and structural information for a large collection of both characterized and uncharacterized compounds. This resource can serve as a starting point for discovery of new probes with specific modes of action or further characterization of novel compounds.

Acknowledgements

We thank Jeff Piotrowski, Sheena Li, and Charlie Boone for valuable feedback in developing the features of the web interface for MOSAIC.

Funding

This work was partially supported by the National Institutes of Health (R01HG005084, R01GM104975) and the National Science Foundation (DBI 0953881). CM is a fellow in the Canadian Institute for Advanced Research (CIFAR) Genetic Networks Program. Computing resources and data storage services were partially provided by the Minnesota Supercomputing Institute and the UMN Office of Information Technology, respectively. SS was supported by an NSF Graduate Research Fellowship.

Conflict of Interest: none declared.

References

Balkwill, F. (2004). Cancer and the chemokine network. Nature Reviews Cancer, 4(7), 540–550.
Bibette, J. (2012). Gaining confidence in high-throughput screening. Proceedings of the National Academy of Sciences, 109, 649-650.
Giaever, G. *et al*. (1999). Genomic profiling of drug sensitivities via induced haploinsufficiency. Nature Genetics, 21(3), 278–83.
Hillenmeyer, M. E. *et al*. (2008). The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science, 320(5874), 362–365.
Kleinke, J. D. (2001). The Price Of Progress: Prescription Drugs In The Health Care Market. Health Affairs, 20(5), 43–60.
Lee, A. Y. *et al*. (2014). Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science (New York, N.Y.), 344(6180), 208–11.
Macarron, R. *et al*. (2011). Impact of high-throughput screening in biomedical research. Nature Reviews Drug Discovery, 10, 188-195.
Parsons, A. B. *et al*. (2004). Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nature Biotechnology, 22(1), 62–9.
Parsons, A. B. *et al*. (2006). Exploring the Mode-of-Action of Bioactive Compounds by Chemical-Genetic Profiling in Yeast. Cell, 126(3), 611–625.
Piotrowski, J. S. *et al*. (2017). Functional Annotation of Chemical Libraries across Diverse Biological Processes. Submitted.
Simpkins, S. BEAN-counter, (2017), GitHub repository, https://github.com/RussianImperialScott/BEAN-counter
Simpkins, S. *et al*. (2017). In preparation
van 't Veer, L. J. *et al*. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature, 415(6871), 530–536.
Viatour, P. *et al*. (2005). Phosphorylation of NF-κB and IκB proteins: implications in cancer and inflammation. Trends in Biochemical Sciences, 30(1), 43–52.