bioRxiv preprint doi: https://doi.org/10.1101/112854; this version posted September 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Subject Section

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

Justin Nelson[1], Scott W. Simpkins[1], Hamid Safizadeh[2,3], Sheena C. Li[4], Jeff S. Piotrowski[5], Hiroyuki Hirano[4], Yoko Yashiroda[4], Hiroyuki Osada[4,*], Minoru Yoshida[4,*], Charles Boone[4,6,*] and Chad L. Myers[1,2,*]

[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, [3]Department of Electrical and Computer Engineering, University of Minnesota-Twin Cities, Minneapolis, Minnesota 55455, USA, [4]RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan 3510198, [5]Yumanity Therapeutics, Cambridge, MA, USA 02139 and [6]University of Toronto, Donnelly Centre, Toronto, Ontario, Canada

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: Chemical-genomic approaches that map interactions between small molecules and genetic perturbations offer a promising strategy for functional annotation of uncharacterized bioactive compounds. We recently developed a new high-throughput platform for mapping chemical-genetic (CG) interactions in yeast that can be scaled to screen large compound collections, and we applied this system to generate CG interaction profiles for more than 13,000 compounds. When integrated with the existing global yeast genetic interaction network, CG interaction profiles can enable mode-of-action prediction for previously uncharacterized compounds as well as discover unexpected secondary effects for known 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, bioprocesses (GO terms), and genes for CG information including direct CG interactions, bioprocesses, and gene-level target predictions. MOSAIC also provides access to chemical structure information of screened molecules, chemical-genomic profiles, and the ability to search for compounds sharing structural and functional similarity. This resource will be of interest to chemical biologists for discovering new small molecule probes with specific modes-of-action as well as computational biologists interested in analyzing CG interaction networks.

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

Contact: chadm@umn.edu, charlie.boone@utoronto.ca, yoshidam@riken.jp, or hisyo@riken.jp

1 Introduction

Discovering the mode-of-action of compounds is an important step in developing effective treatments for diseases or controlling the spread of pathogens. With the advent of efficient whole-genome sequencing methods that enable comprehensive genotyping of individuals, there are an increasing number of candidate drug targets implicated in specific diseases, which presents a need for new drug discovery paradigms that maximize these opportunities. The standard methods in high-throughput screening are relatively inefficient at linking small molecules to cellular targets and thus need to be improved in order to develop a rich collection of tool compounds for probing biological pathways and generating drug leads. We have yet to build a collection of small molecules that are able to target a substantial fraction of the human proteome.

Chemical-genomic approaches reveal compound-dependent phenotypes in a systematic, unbiased manner to predict the cellular role of an uncharacterized bioactive compound. For example, studies have tested compounds against the complete gene deletion collection that is available for the genetically tractable yeast *Saccharomyces cerevisiae* (Giaever et al., 1999; Parsons et al., 2004; Parsons et al., 2006; Hillenmeyer et al., 2008; Costanzo et al., 2010;

bioRxiv preprint doi: https://doi.org/10.1101/112854; this version posted September 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hoepfner et al., 2014; Lee et al., 2014; Giaever and Nislow, 2014; Wildenhain et al., 2016). In these approaches, a collection of genetic mutants is assayed for sensitivity or resistance to a compound to generate a compound-specific functional profile. We recently developed a new high-throughput platform for mapping CG interactions in yeast for large-scale compound screening and, using this, we generated a library of CG profiles for various compound collections (Piotrowski *et al*., 2017). This represents one of the most extensive chemical-genomic datasets to date, including informative profiles for more than 1,000 natural products and derivatives, hundreds of clinically relevant compounds, and other small molecules. Here, we present MOSAIC, a web interface that provides access to this rich chemical-genomic dataset.

2 Methods

MOSAIC, the website for accessing the database, is located at http://mosaic.cs.umn.edu. This website displays data obtained by screening the RIKEN Natural Product Depository (NPDepo) as well as diversity libraries from the NCI/NIH, and the Published Kinase Inhibitor Set from Glaxo-Smith-Kline (Piotrowski *et al*., 2017). The data housed in the database were derived from the experiments described in (Piotrowski *et al*., 2017) and were processed and analyzed using BEAN-counter (https://github.com/csbio/BEAN-counter) and CG-TARGET (Simpkins *et al*., 2017) tools.

MOSAIC was designed using a combination of HTML, javascript with jquery, AJAX, javascript object notation (JSON), PHP, and MySQL. This provides a responsive environment for querying the CG interactions.

MOSAIC provides an interface for querying CG profiles, gene- and bioprocess-level target 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 standard name, or a Gene Ontology (GO) biological process term. Multiple terms can be input by separating each term with a semicolon. The results page will return different data and predictions depending on whether a compound, gene, or process has been queried. If a query includes more than one type of input, the terms can be selected by using the tabs (Conditions, Genes, or GO Terms) and tables.

MOSAIC displays four relevant sets of information related to the mode-of-action of compounds. The first is the set of direct CG interactions, which reflect the sensitivity or resistance of each specific deletion mutant to a given compound. The second is the set of gene-level target predictions, which measure the correlation between the CG interaction profile and a genetic interaction profile. A high correlation indicates that the compound phenocopies the effect of a gene deletion, suggesting the compound perturbs the function of the corresponding gene. The third set of information is the bioprocess-level target prediction, which lists GO terms that are over-represented among the top gene-level target predictions as described in (Simpkins *et al*., 2017). Finally, MOSAIC provides access to similarities between pairs of compounds based on either their chemical structure (Safizadeh *et al*., 2017) or their CG interaction profile.

3 Conclusion

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

Funding

This work was partially supported by the National Institutes of Health (R01HG005084, R01GM104975) and the National Science Foundation (DBI 0953881) to CM, the Canadian Institutes of Health Research, FDN-143264, to CB, the JSPS KAKENHI Grant Numbers 15H04483, to CB. MY is supported by JSPS KAKENHI Grant Number 26221204. HO is a research fellow of the Japan Society for the Promotion of Science. SWS was supported by an NSF Graduate Research Fellowship (00039202), an NIH Biotechnology Training Grant (T32GM008347), and a one year BICB fellowship from the University of Minnesota.

Conflict of Interest: none declared.

References

Costanzo, M. et al. (2010). The genetic landscape of a cell. Science, **327(5964)**, 425-431

Giaever, G. and Nislow, C. (2014). The Yeast Deletion Collection: A Decade of Functional Genomics. Genetics, **197(2)**, 451-465

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.

Hoepfner, D. *et al*. (2014). High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. Microbiology Research, **169(2-3),**107-120

Lee, A. Y. et al. (2014). Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science, **344(6180)**, 208–11.

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. Nature Chemical Biology. Advanced online copy.

Safizadeh, H. et al. (2017). Improving prediction of compound function from chemical structure using chemical-genetic networks. bioRxiv. doi:10.1101/112698.

Simpkins, S. W. et al. (2017). Large-scale interpretation of chemical-genetic interaction profiles using a genetic interaction network. bioRxiv. doi:10.1101/111252.

Wildenhain, J. et al. (2016). Systematic chemical-genetic and chemical-chemical interaction datasets for prediction of compound synergism. *Scientific Data*, **3(160095)**