Supplementary Information

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
[6]University of Toronto, Donnelly Centre, Toronto, Ontario, Canada

Extended Summary of MOSAIC Components

1 Chemical-genetic Interaction Profiles

The sensitivity or resistance of deletion mutants in the presence of a compound is provided as part of the CG interaction data. MOSAIC provides a table of CG interaction scores for compound queries (~13,000 compounds) or gene queries (~300 gene mutants in the yeast diagnostic pool). A negative CG interaction indicates that a deletion mutant is sensitive to a compound; conversely, a positive CG interaction indicates that a deletion mutant is resistant to a compound. These data are provided as part of MOSAIC, either in the form of a CG interaction profile (for a compound query) or as a list of conditions that a deletion mutant is sensitive or resistant to (for a gene query).

2 Gene-level Target Prediction

A gene-level target prediction is made based on the similarity between the CG interaction profile and a genetic interaction profile (Parsons *et al*., 2004; Parsons *et al*., 2006; Costanzo *et al*., 2010). A genetic interaction is when growth of a double mutant differs from expectation based on the single gene deletion effects. Compound and gene queries both return the table of relevant gene-level target prediction scores, which are computed using the inner product between a CG profile and all $\ell 2$-normalized genetic interaction profiles. A high score between a CG profile and a genetic interaction profile indicates that the compound's effect phenocopies those of the gene deletion, suggesting that the compound perturbs the function and/or the proteins encoded by the gene. Using the MOSAIC website, researchers can quickly browse and form hypotheses based on these gene-level target predictions.

3 Bioprocess-level Target Prediction

A Gene Ontology biological process prediction is computed by mapping the gene-level target prediction scores onto GO biological process terms 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 is displayed for Gene Ontology queries. For example, Nocodazole binds tubulin and blocks microtubule formation. A query of "Nocodazole" will return a process prediction to "tubulin complex assembly" (GO:0007021). These Gene Ontology predictions, called bioprocess-level target predictions, are calculated by comparing the observed gene-level target prediction scores within a GO biological process to empirical null distributions of the same scores generated from experimental negative controls, resampled CG profiles, and shuffled gene-level target predictions (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 a desired mode-of-action.

4 Chemical Structure and CG Profile Similarity

Compound similarity searching can be useful for expanding lists of candidate compounds. MOSAIC provides access to both structural similarity and functional similarity (CG interaction profile similarity) for compound queries. The compound structural similarity is calculated using the All-Shortest Path (ASP) structure descriptor (depth 8) combined with the Braun-Blanquet similarity coefficient, whose performance

was superior to several other alternatives in connecting chemical structures to compound functions (Hinselmann *et al*., 2011; Safizadeh *et al*., 2017). Functional similarity measures are calculated using Pearson correlation applied to the CG profile. These two similarity approaches can be compared to understand how structure modulates the effect of a compound on a cell.

References

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.

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

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

Hinselmann G, et al. (2011) jCompoundMapper: An open source Java library and command-line tool for chemical fingerprints. J. Cheminformatics. **3,** 1-14

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