bioRxiv preprint doi: https://doi.org/10.1101/662502; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

MOBN: an interactive database of multi-omics biological networks

Cheng Zhang[1,#,*], Muhammad Arif[1,#], Xiangyu Li[1], Sunjae Lee[1], Abdellah Tebani[1], Wenyu Zhou[2], Brian D. Piening[3], Linn Fagerberg[1], Nathan Price[4], Leroy Hood[4], Michael P. Snyder[2], Jens Nielsen[5], Mathias Uhlen[1], Adil Mardinoglu[1,6,*]

[1]Science for Life Laboratory, KTH – Royal Institute of Technology, Stockholm, SE-171 21, Sweden
[2]Department of Genetics, Stanford University, Stanford, CA 94305, USA
[3]Earle A Chiles Research Institute and Providence Cancer Institute, Portland, OR 97213, USA
[4]Institute of Systems Biology, Seattle, USA
[5]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden
[6]Centre for Host–Microbiome Interactions, Dental Institute, King's College London, London, SE1 9RT, United Kingdom

[#]These authors contributed equally.

*To whom correspondence should be addressed.

Abstract

Summary: The associations among different omics are essential to understand human wellness and disease. However, very few studies have focused on collecting and exhibiting multi-omics associations in a single database. Here, we present an interactive database of multi-omics biological networks (MOBN) and describe associations between clinical chemistry, anthropometrics, plasma proteome, plasma metabolome and gut microbiome obtained from the same individuals. MOBN allows the user to interactively explore the association of a single feature with other omics data and customize its specific context (e.g. male/female specific). MOBN is designed for users who may not have a formal bioinformatics background to facilitate research in human wellness and diseases.

Availability: The database is accessible at http://multiomics.inetmodels.com without any limitation.

Contact: cheng.zhang@scilifelab.se, adilm@scilifelab.se

1 Introduction

During the past decade, the development of high-throughput technologies has dramatically decreased the cost of generating large-scale multi-omics datasets (Goodwin, et al., 2016), and has opened up the possibilities to comprehensively and systematically study human wellness and disease (Karczewski and Snyder, 2018). Although analysis of individual omics methodologies has been proven to be useful in different clinical applications, integration of multiple omics data may offer novel insights and provide a more comprehensive understanding of biological functions in the human body (Hasin, et al., 2017). For instance, a recent study integrated time series phenomics, fluxomics and metabolomics data from subjects with various degrees of liver fat and revealed a novel glycine and serine deficiency phenotype in the patients with non-alcoholic fatty liver disease (Mardinoglu, et al., 2017). In another study, longitudinal phenomics, transcriptomics, metagenomics, and metabolomics data has been obtained from 10 subjects over a two-week period to illustrate the benefit of an isocaloric low-carbon diet for non-alcoholic fatty liver disease patients (Mardinoglu, et al., 2018). Moreover, several other studies have also demonstrated the use of longer-period longitudinal multi-omics data in capturing the dynamics of human diseases systematically (Chen, et al., 2012; David, et al., 2014; Smarr, 2012).

In order to provide a better framework to facilitate these types of investigations, we created an interactive database of multi-omics biological networks (MOBN), a user-friendly database which provides exploratory capabilities and interactive visualization of clinical chemistry, anthropometrics, plasma proteins, metabolites, and gut microbiomes associated with user-queried features (Figure 1). The data in MOBN are obtained from recent studies with large-scale multi-omics data collected on a series of individuals (Supplementary Info). To our knowledge, MOBN is the first database that provide association between multi-omics data obtained from the same subjects in physiological context rather than from text mining.

Fig. 1. Concept of MOBN.

2 Datasources and analysis

The data included in the MOBN are obtained from three independent studies with large-scale personalized longitudinal multi-omics data: (1) the Swedish SciLifeLab SCAPIS Wellness Profiling project which collected multi-omics data from 98 different subjects during 6 visits in 2 years (Bergstrom, et al., 2015), (2) the P100 Wellness study (Price, et al., 2017) where 108 subjects are followed for 9 month and profiled with different omics data (accessible phs001363.v1.p1 from dbGaP) and (3) the iPOP diet perturbation study which employed multi-omics data to describe the systematic dynamics in human with weight gain and weight loss during 3 visits within 1 year (Piening, et al., 2018). All of these studies provided information of the anthropometrics, clinical

bioRxiv preprint doi: https://doi.org/10.1101/662502; this version posted June 8, 2019. 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 4.0 International license.

chemistry, plasma proteome, plasma metabolome and gut microbiome for each individual involved in the study, and thus, the consensus of these multi-omics data is included in MOBN.

Data from the different studies were first standardized to apply the same omics feature names across all datasets, and normalized to remove bias. Different omics d+ata were subsequently corrected based on the age (for all networks) and sex (for non-gender specific networks) using trimmed mean robust regression (Price, et al., 2017). The corrected data were consequently utilized to determine the inter- and intra-omics associations based on Spearman's correlation, and the final association networks were constructed by collecting associations with significant correlation (FDR < 5%). All networks can be downloaded from the database under the 'Downloads' section.

3 Features

MOBN provides a free, flexible and simple-to-use visualization of multi-omics association networks that are freely-accessible. Users can customize their specific network based on their individual interests as shown in Figure 2.

Firstly, MOBN provides a broad selection of networks. Users may select networks constructed based on a specific study with a specific context, such as gender-specific networks or insulin resistance/sensitive networks. In addition, both cross-sectional and delta networks are constructed using the methodology presented in the P100 wellness study (Price, et al., 2017) for each network. Cross-sectional networks present multi-omics correlations in the context of individualized variation, while delta networks allow users to investigate features that co-vary within the same time intervals.

To search within a specific network, users need to input the commonly known name (gene symbol, metabolite name etc.) of (up to 3) analytes of interests to search and/or select from the drop-down list. Once the search is executed, the network will be generated and visualized in the graphic area. Next, users can filter the types of omics data shown in the graphic area based on their interests by checking/unchecking the corresponding box, showing only first neighbor edges and tailoring the network based on the number of edges and/or a specified statistical property, e.g. FDR.

All information about each analyte (network nodes) and related associations (network edges) is shown in the table area: the analyte name, short description, unit, correlation and P value of the association etc. In addition, wherever possible, analytes are linked with external databases such as KEGG (Kanehisa, et al., 2017), Human Protein Atlas (Thul, et al., 2017; Uhlen, et al., 2015; Uhlen, et al., 2017), Uniprot (UniProt, 2019) and HMDB (Wishart, et al., 2018) to facilitate further biological interpretation and investigation. Moreover, users can easily retrieve information provided in their network, export as a tab-separated file and analyze the networks with other tools following the 'Tutorial' section on the website.

4 Conclusion

MOBN is the first database that shows multi-omics associations based on personalized data. It's designed in a simple and user-friendly way, and accessible to a wide range of users including bench scientists with limited to no formal bioinformatics background. In this context, we expect MOBN to help catalyze deeper multi-omic investigations that may reveal novel biological mechanisms underlying human wellness and disease.

Acknowledgements

We appreciate the data sharing from dbGaP. This work was supported in part by the Robert Wood Johnson Foundation, the M.J. Murdock Charitable Trust, NIH grants 2P50GM076547, ES017885, RC2HG005805, and Arivale.

Funding

This work has been supported by the Knut and Alice Wallenberg Foundation.

Conflict of Interest: none declared.

References

Bergstrom, G., *et al.* The Swedish CArdioPulmonary BioImage Study: objectives and design. *J Intern Med* 2015;278(6):645-659.

Chen, R., *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 2012;148(6):1293-1307.

David, L.A., *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol* 2014;15(7):R89.

Goodwin, S., McPherson, J.D. and McCombie, W.R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016;17(6):333-351.

Hasin, Y., Seldin, M. and Lusis, A. Multi-omics approaches to disease. *Genome Biol* 2017;18(1):83.

Kanehisa, M., *et al.* KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* 2017;45(D1):D353-D361.

Karczewski, K.J. and Snyder, M.P. Integrative omics for health and disease. *Nat Rev Genet* 2018;19(5):299-310.

Mardinoglu, A., *et al.* Personal model-assisted identification of NAD(+) and glutathione metabolism as intervention target in NAFLD. *Mol Syst Biol* 2017;13(3):916.

Mardinoglu, A., *et al.* An Integrated Understanding of the Rapid Metabolic Benefits of a Carbohydrate-Restricted Diet on Hepatic Steatosis in Humans. *Cell Metab* 2018;27(3):559-571 e555.

Piening, B.D., *et al.* Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. *Cell Syst* 2018;6(2):157-170 e158.

Price, N.D., *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat Biotechnol* 2017;35(8):747-756.

Smarr, L. Quantifying your body: a how-to guide from a systems biology perspective. *Biotechnol J* 2012;7(8):980-991.

Thul, P.J., *et al.* A subcellular map of the human proteome. *Science* 2017;356(6340).

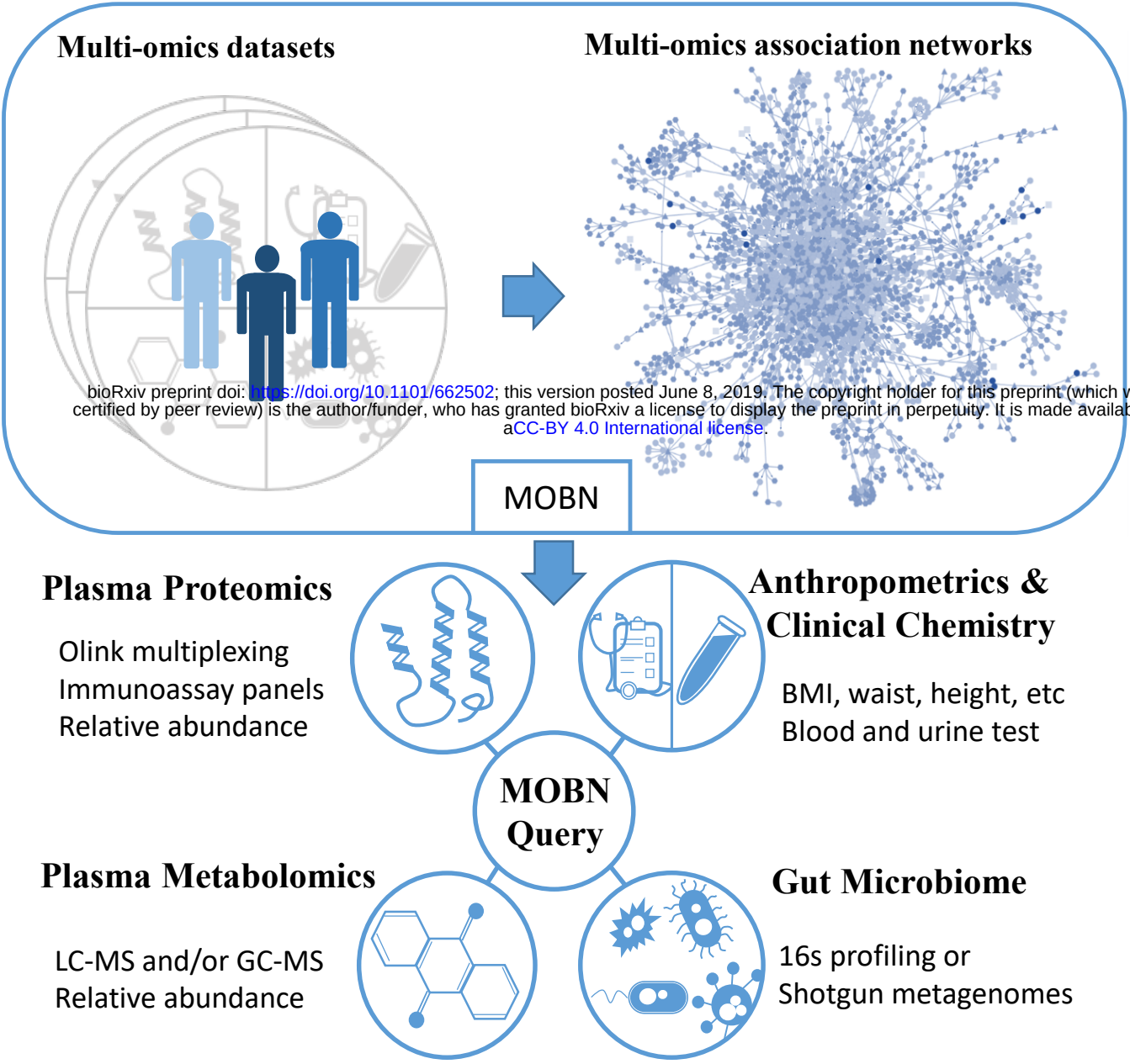
bioRxiv preprint doi: https://doi.org/10.1101/662502; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Uhlen, M., *et al.* Proteomics. Tissue-based map of the human proteome. *Science* 2015;347(6220):1260419.

Uhlen, M., *et al.* A pathology atlas of the human cancer transcriptome. *Science* 2017;357(6352).

UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47(D1):D506-D515.

Wishart, D.S., *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* 2018;46(D1):D608-D617.



bioRxiv preprint doi: https://doi.org/10.1101/662502; this version posted June 8, 2019. The copyright holder for this preprint (which w
certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made availab
aCC-BY 4.0 International license.

MOBN
Multi-Omics Biological Network

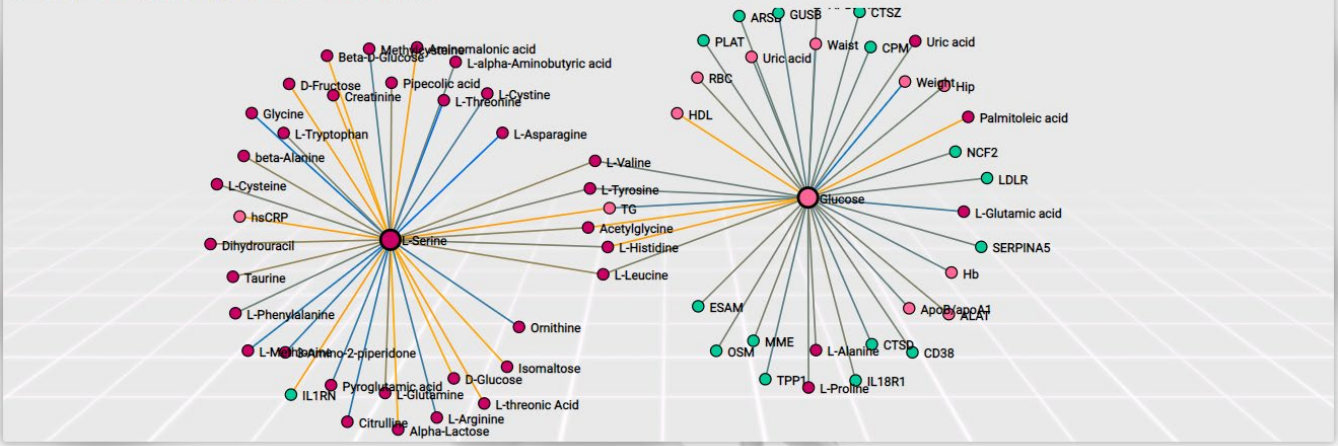
Home About Tutorial Downloads

sysmedicine

Cross-Sectional (Overall) | Glucose (CLINICAL) × L-Serine (METABOLITE) × | Search

Blue edge: postive corr. Orange edge: negative corr. Left click on edge Right click on node

Nodes Legend: Protein Metabolite Microbiome Anthropometric/Clinical Chemistry



Network filters | Save | Citation

Filter node by analyte type:

- [x] Proteins
- [x] Metabolites
- [x] Microbiomes
- [x] Anthropometrics/Clinical Chemistry

First Neighbor:

- [ ] Includes

NOTE:
Increasing the number of nodes will increase the loading time
Less than three genes to search
Less than 50 maximal nodes for multiple analytes

Maximum number of nodes (min-1 max-200): 36

Edge Pruning Parameter (-log10 P-Adjusted) (min-2 max-50): 2

Searched Nodes | All Nodes | Proteins | Metabolites | Microbiomes | Anthropometrics/Clinical Chemistry | Edges

Save Searched Genes as CSV file

#	Symbol	Location
1	Glucose	CLINICAL
2	L-Serine	METABOLITE