

iNetModels 2.0: an interactive visualization and database of multi-omics data

Muhammad Arif^{1,†}, Cheng Zhang^{1,2,*,†}, Xiangyu Li¹, Cem Güngör³, Buğra Çakmak³, Metin Arslantürk³, Abdellah Tebani^{4,5}, Berkay Özcan³, Oğuzhan Subaş³, Wenyu Zhou⁶, Brian Piening⁷, Hasan Turkez⁸, Linn Fagerberg¹, Nathan Price⁹, Leroy Hood⁹, Michael Snyder⁶, Jens Nielsen¹⁰, Mathias Uhlen¹ and Adil Mardinoglu^{1,11,*}

¹Science for Life Laboratory, KTH – Royal Institute of Technology, Stockholm SE-171 21, Sweden, ²School of Pharmaceutical Sciences & Key Laboratory of Advanced Drug Preparation Technologies, Ministry of Education, Zhengzhou University, Zhengzhou, Henan Province, PR 450001, China, ³Bash Biotech Inc, 600 West Broadway, Suite 700, San Diego, CA, USA, ⁴Department of Metabolic Biochemistry, Rouen University Hospital, 76000 Rouen, France, ⁵Normandie Univ, UNIROUEN, CHU Rouen, INSERM U1245, 76000 Rouen, France, ⁶Department of Genetics, Stanford University, Stanford, CA 94305, USA, ⁷Providence Cancer Center, Oregon Area, Portland, OR, USA, ⁸Department of Medical Biology, Faculty of Medicine, Atatürk University, Erzurum, Turkey, ⁹Institute of Systems Biology, Seattle, USA, ¹⁰Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden and ¹¹Centre for Host–Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London SE1 9RT, UK

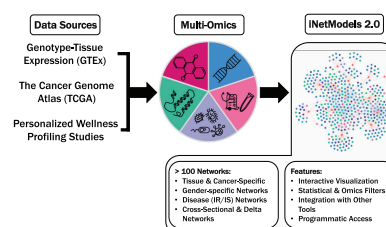
Received January 21, 2021; Revised March 10, 2021; Editorial Decision March 29, 2021; Accepted March 29, 2021

ABSTRACT

It is essential to reveal the associations between various omics data for a comprehensive understanding of the altered biological process in human wellness and disease. To date, very few studies have focused on collecting and exhibiting multi-omics associations in a single database. Here, we present iNetModels, an interactive database and visualization platform of Multi-Omics Biological Networks (MOBNs). This platform describes the associations between the clinical chemistry, anthropometric parameters, plasma proteomics, plasma metabolomics, as well as metagenomics for oral and gut microbiome obtained from the same individuals. Moreover, iNetModels includes tissue- and cancer-specific Gene Co-expression Networks (GCNs) for exploring the connections between the specific genes. This platform allows the user to interactively explore a single feature's association with other omics data and customize its particular context (e.g. male/female specific). The users can also register their data for sharing and visualization of the MOBNs and GCNs. Moreover, iNetModels allows users who do not have a bioinformatics background to facilitate human well-

ness and disease research. iNetModels can be accessed freely at <https://inetmodels.com> without any limitation.

GRAPHICAL ABSTRACT



INTRODUCTION

During the past decade, the development of high-throughput technologies has dramatically decreased the cost of generating large-scale multi-omics datasets (1). This has opened up the possibilities to study human wellness and diseases systematically (2). Although analysis of individual omics methodologies has been proven beneficial in different clinical applications, integrating multi-omics data may offer novel insights and provide a more comprehensive understanding of biological functions in the human body in health and disease (3). For instance, a recent study integrated time series phenomics, metabolomics and

*To whom correspondence should be addressed. Tel: +46 8 790 60 00; Fax: +46 8 790 65 00; Email: adilm@scilifelab.se
Correspondence may also be addressed to Cheng Zhang. Email: cheng.zhang@scilifelab.se

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

fluxomics data from the subjects with various degrees of liver fat and revealed that non-alcoholic fatty liver disease (NAFLD) is associated with glycine and serine deficiency (4). Another longitudinal phenomics, transcriptomics, metagenomics and metabolomics data have been generated for 10 subjects during a two-week follow-up study. This study has illustrated the rapid metabolic benefits of an isocaloric carbohydrate-restricted diet on NAFLD patients and revealed the molecular mechanisms associated with the metabolic changes (5). Moreover, several other studies have also demonstrated the benefit of performing longitudinal multi-omics data analysis in systematically capturing human diseases' dynamics (6–8).

To provide a better framework for facilitating these types of investigations, we created iNetModels. This user-friendly platform provides exploratory capabilities and interactive and intuitive visualization of clinical chemistry, anthropometric parameters, plasma proteins, plasma metabolites, oral microbiome and gut microbiome associated with the user-queried features (Figure 1A). The data in iNetModels are obtained from recent studies, where large-scale Multi-Omics Biological Networks (MOBNs) analyses have been performed for individuals with different metabolic conditions. Moreover, we retrieved data from The Genotype-Tissue Expression (GTEx) Project and The Cancer Genome Atlas (TCGA), created normal tissue- and cancer-specific Gene Co-expression Networks (GCNs) and presented the networks in the iNetModels (Figure 1A). The user can simultaneously query for 1–5 features, visualize the selected features and their neighbouring features, download the associated network in both table and figure format, and analyse them using independent network analysis tools, including Cytoscape (9) and iGraph (10). We also encourage users to upload their networks into iNetModels and make those networks accessible to a broader audience for creating an open platform to share and visualize their networks. To our knowledge, iNetModels is the first database that provides associations between the multi-omics data obtained from the same individuals in a physiological context rather than using a text mining method.

PLATFORM DESCRIPTION AND FEATURES

iNetModels 2.0 is a web-based platform that includes two main features: a database and an interactive visualization of multi-omics network analysis (Figure 1B). It is an updated and improved version of the previous work, TCSBN, released in 2017 (11). First, we reconstructed the GCNs presented in TCSBN based on more recent datasets and expanded it to 85 different tissue- and cancer-specific networks. Second, we broadened the platform by adding 20 MOBNs from multiple independent studies. Finally, we improved the backend and frontend of the platform for a better user experience. To our knowledge, this platform is the only publicly available platform that enables the exploration of MOBNs that were generated based on personalized omics data.

Data sources

The tissue- and cancer-specific GCNs were constructed using GTEx (v8) and TCGA (v27.0-fix) primary tumour data,

respectively. We retrieved the gene transcripts per million (TPM) from GTEx Portal and the fragment per kilobase of transcript per million mapped reads (FPKM) files of the TCGA program from the Genomics Data Commons (GDC) portal.

MOBNs were generated as the consensus of clinical variables, plasma proteomics, plasma metabolomics and metagenomics data based on from three independent longitudinal wellness profiling studies: (i) SCAPIS-SciLifeLab Wellness Profiling study (12): clinical variables, metabolomics, proteomics and gut metagenomics data from 4 visits over a year from 101 individuals (50–65 years old), (ii) P100 study (13): clinical variables, metabolomics and proteomics data collected from three visits of 108 individuals (21–89+ years old) in over 9 months, (iii) Integrative Personal Omics (14): clinical variables, metabolomics and proteomics data from three visits over weight gain and loss period (97 days) from 10 insulin-sensitive (IS) and 13 insulin-resistance (IR) overweight individuals. Moreover, we generated MOBNs for two different clinical trials where combined metabolic activators (CMA) administered to NAFLD and COVID-19 patients: (i) NAFLD CMA administration: clinical variables, metabolomics, proteomics and metagenomics data for the oral and gut microbiome of 31 NAFLD patients from three visits over 70 days, and (ii) COVID-19 CMA administration: Clinical variables, metabolomics and proteomics data collected 93 COVID-19 patients from two visits over 14 days.

Network generations

Each dataset was pre-processed independently before the network generation. The gene expression data were filtered to remove lowly expressed genes (TPM < 1 or FPKM < 1) in each tissue or cancer type to avoid bias. For the multi-omics data, we corrected them by removing the effects of age (in all networks) and sex (in the non-gender-specific networks) using trimmed mean robust regression (13). We combined the data from each omics into a matrix to generate a network and analysed it using the Spearman correlation function from the SciPy package. Each data source was processed independently. After filtering for striking correlations between the pairs, we filtered the pairs with FDR < 0.05. Furthermore, we performed community detection analysis using the Leiden algorithm (15) in the iGraph package to identify sub-network clusters for downstream analysis.

Moreover, in the longitudinal wellness profiling studies, we calculated both cross-sectional and delta networks. These networks are generated based on the methodology presented in the P100 wellness study (13). Cross-sectional networks were calculated by correlating data from all visits to represent correlations in the context of individualized variation. Meanwhile, delta networks were calculated by correlating the analyte changes between visits to allow users to investigate features that co-vary within the same time intervals.

The number of the nodes and edges in the generated networks varying between 5673 (liver cancer)—12 581 (testis) nodes and 63 292 (endocervix)—132 223 286 (testis) edges in GCNs, whereas 279 (Delta IS)—1042 (NAFLD CMAs)

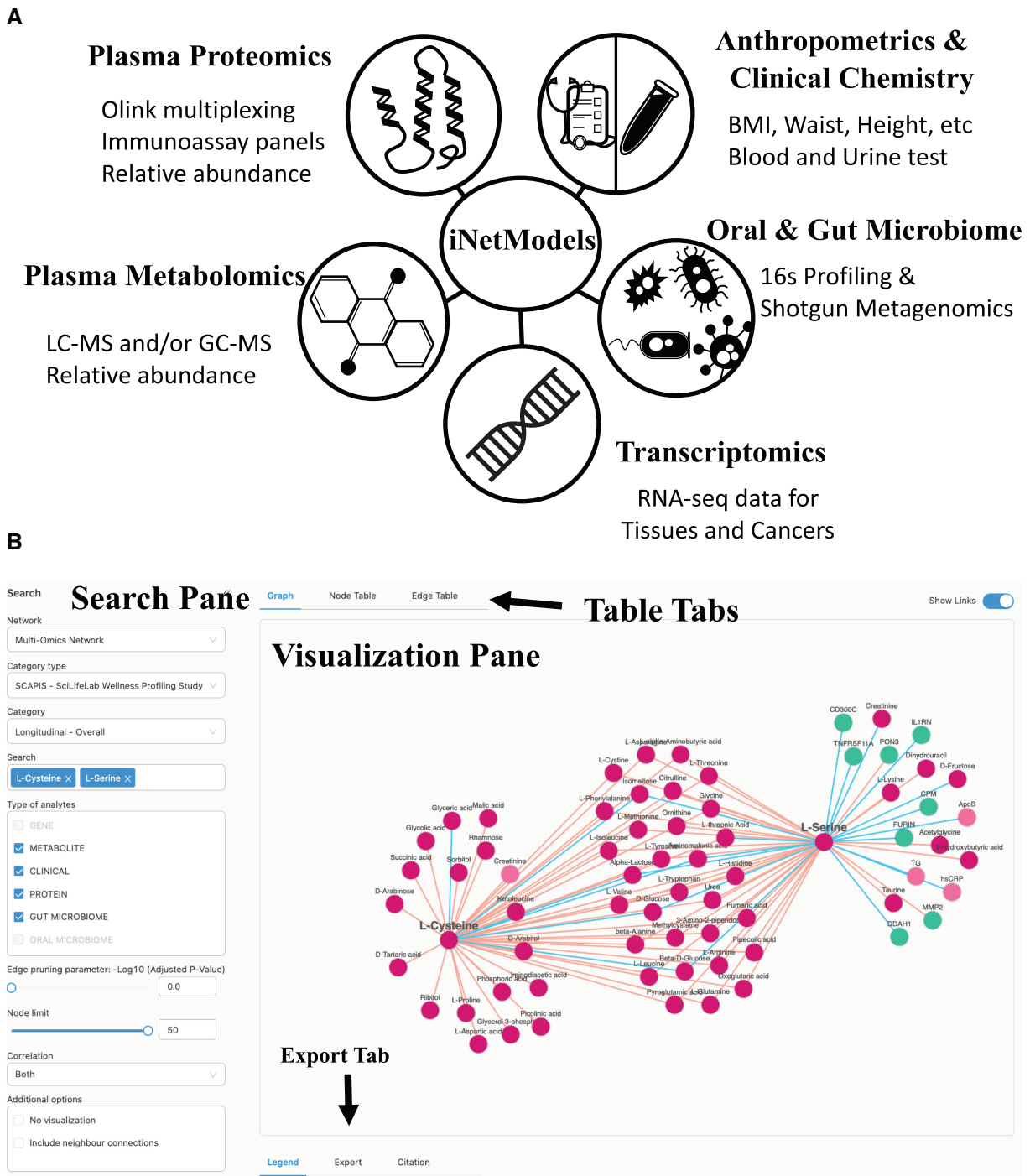


Figure 1. (A) Summary of multi-omics data presented in iNetModels (B) Working page of iNetModels.

nodes and 1034 (Delta IS)—566 390 (SCAPIS—SciLifeLab Longitudinal Male) edges in MOBNs (Supplementary Table S1).

Features

The iNetModels platform provides users with a vast number of pre-computed biological networks. Users may choose the suitable networks from tissue- or cancer-specific GCNs

to gender or IR/IS-specific MOBNs based on their study’s focus. To search within a specific network (Figure 1B, Supplementary Figure S1), firstly, users need to select the specific network category (GCNs or MOBNs), then select the specific network type (normal tissue, cancer, or multi-omics study), and subsequently select the specific network. Following that, users need to input the commonly known names of analytes (gene, protein, metabolite name etc.) of interests using the free text and/or drop-list. Optionally,

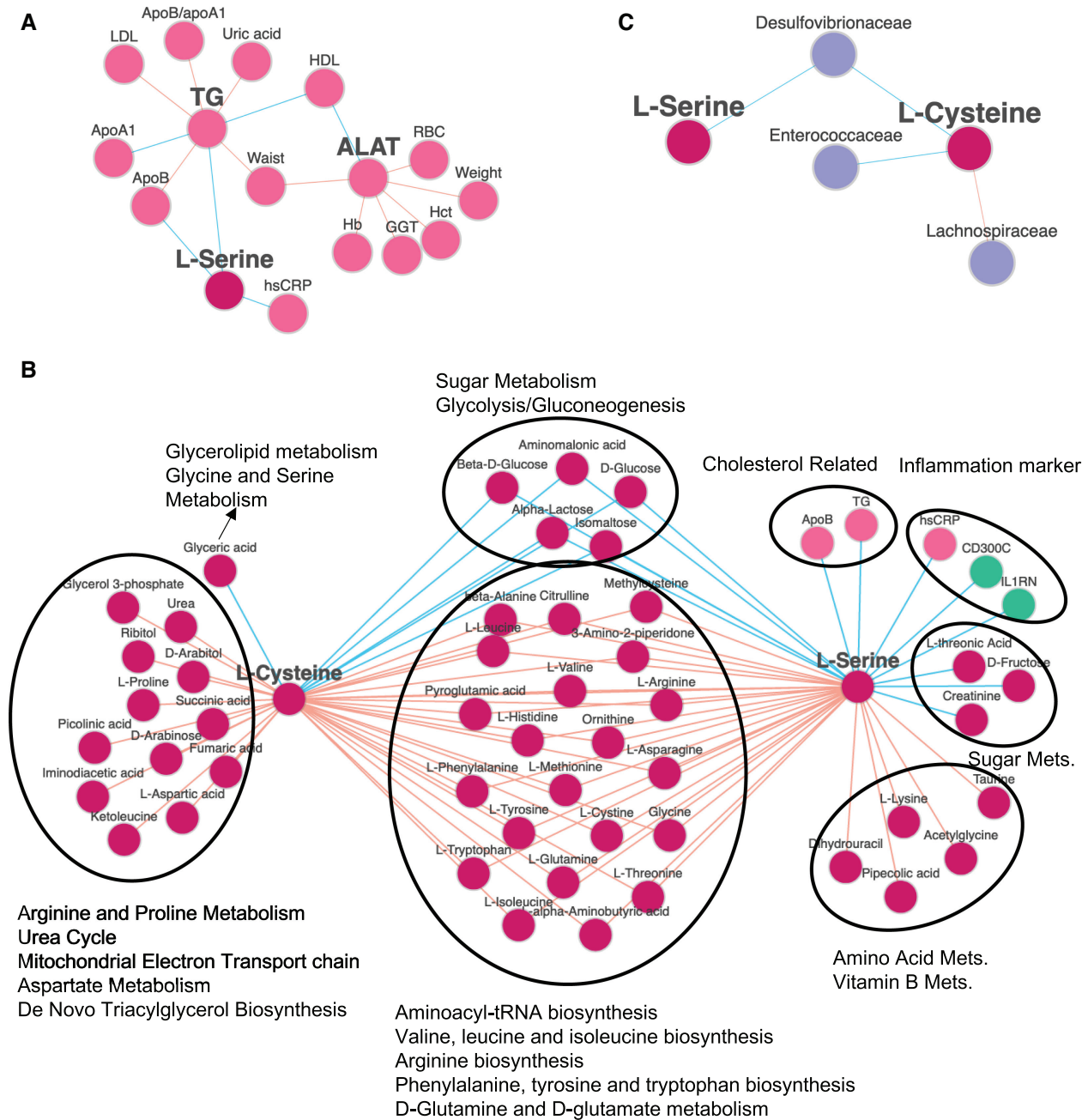


Figure 2. (A) Validation of the hypothesis about the supplementation of L-Serine that was associated with the decrease in the plasma triglycerides levels and liver enzyme (ALAT) in the SCAPIS-SciLifeLab Wellness Profiling study. (LINK). (B) The two main components of the supplementation (L-cysteine and L-serine) and their neighbouring analytes are presented based on multi-omics biological networks analysis (LINK). (C) The two main components of the supplementation (L-cysteine and L-serine) and their associations with the species in the gut microbiome are presented (LINK). All networks in this figure were taken from the cross-sectional overall SCAPIS-SciLifeLab Wellness Profiling study.

users can filter the network based on the analyte types (in multi-omics networks) and statistical properties, e.g. FDR and Spearman correlation ranking (positive, negative, or both correlations). In the web interface, the number of neighbouring nodes is limited to a maximum of 50 neighbours per-queried nodes to avoid browser unresponsiveness. Furthermore, in the ‘additional options’ box, users can choose to include neighbours’ connections in the networks

and disable network visualization when only network tables are needed.

Once the network is generated, the users can interactively explore the network, download the network as a figure and its content in a table format for further exploration and downstream analysis. 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 the *P*-value of the significant associations etc. Besides, wherever possible, analytes are linked with external databases such as KEGG (16), Human Protein Atlas (17–19), Uniprot (20) and HMDB (21) to facilitate further biological interpretation and investigation. All of this information can be downloaded directly and is compatible with other network analysis tools and software, such as Cytoscape or iGraph package in Python and R. By using these tools, users can merge multiple networks and perform additional downstream analysis.

Moreover, in the iNetModels 2.0, we implemented programmatic access to the database using an in-house-built Python package that can be found under the ‘API’ section. With the API, users can retrieve more extensive networks (>50 neighbours per-queried nodes) programmatically. We currently limit the query to one query per second.

Case study related to administration of CMA in NAFLD

One of our platform’s unique features and superior strengths is that iNetModels 2.0 is the first and only platform supporting the exploration of MOBNs based on personalized data. This is a great advantage to avoid bias since all data were analysed in a paired manner.

In our recent study (22), we tested a potential therapeutic strategy for NAFLD patients through CMA administration. We provided CMA to 10 subjects involved in the trial and collected plasma samples during the day to generate proteomics and metabolomics data. The data generated in the clinical trial were analysed using metabolic modelling. The results of the analysis were validated by performing animal experiments, where L-serine was supplemented to mouse and a reduction in the liver triglycerides (TG) and markers of liver tissue functions, e.g. ALAT, ASAT, and ALP was observed. We validated these results in two independent MOBNs in iNetModels 2.0 (Figure 2A, Supplementary Figure S2A). Our analysis revealed that two metabolic activators, including L-serine and L-cysteine are positively associated with metabolites related to branched amino-acid metabolism (i.e. L-valine, L-isoleucine and L-leucine) and negatively associated with plasma glucose level, supporting the main findings of the study (Figure 2B). The MOBNs also showed that L-serine is associated negatively with cholesterol-related clinical variables (ApoB, TG) and several other inflammation markers (hsCRP, IL1RN and CD300C).

Based on the same network, we can filter to show only specific analyte types (Figure 2C, Supplementary Figure S2B–D). For example, we used the same network to show the association of L-cysteine and L-serine with only the gut microbiome (Figure 2C), as dysbiosis in the gut microbiome has been associated with NAFLD (23,24). We observed that L-cysteine had a positive correlation with the abundance of Lachnospiraceae and a negative correlation with the abundance of the Enterococcaceae family. Independent studies have shown that the abundance of Lachnospiraceae was increased in cirrhosis (25,26), whereas the abundance of Enterococcaceae was decreased in cirrhosis (25). We also found that the levels of both L-serine and L-cysteine were negatively correlated with the abundance of the Desulfobibri-

onaceae family, which was increased with the severity of NAFLD (27).

CONCLUSION

iNetModels is a unique platform that gathers a broad spectrum of biological networks, from tissue- and cancer-specific GCNs to MOBNs based on personalized data. This platform may help researchers to perform exploration and validation experiments, identify functional relationships between the analytes, and most importantly, provide new insights into the biological experiments and ultimately identify potential drug targets and biomarkers. The case study about the supplementation of CMA in NAFLD patients has shown the efficient usage of this platform in testing and validating hypotheses as well as confirming results from experiments or clinical trials. This platform is designed in a user-friendly way and it is freely accessible to a wide range of users, including bench scientists with limited or no formal bioinformatics background. We also envisage that this platform will be a key resource for computational biologists working in omics data integration, network science, systems biology and systems medicine. In this context, we expect iNetModels to be an essential resource for more in-depth multi-omics analysis that may reveal novel molecular mechanisms underlying human wellness and disease.

In the future, in addition to the inclusion of more studies, we are planning to expand the functionalities of the platform. First, with the increasing number of personalized wellness profiling study, we plan to develop a method to build a consensus MOBN based on various studies to increase the robustness of the findings generated from this platform. Second, we plan to add a functional analysis feature to show the enriched pathways or biological processes by integrating the selected network information to other databases, such as KEGG (16) or Metabolic Atlas (28). Finally, we will add integration with users’ quantitative omics data or statistical inference results to identify the significantly altered nodes in various perturbations.

DATA AVAILABILITY

iNetModels can be accessed freely by everyone at <https://inetmodels.com> without any limitation. All codes used to generate the network and the network data are available under the ‘API’ and ‘Help’ section of the website.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

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. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the

analyses described in this manuscript were obtained from the GTEx Portal on 10/21/20. The results shown here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

FUNDING

Knut and Alice Wallenberg Foundation; Bash Biotech Inc., San Diego, CA, USA. Funding for open access charge: Knut and Alice Wallenberg Foundation; Bash Biotech Inc, San Diego, CA, USA.

Conflict of interest statement. Cem Güngör, Buğra Çakmak, Metin Arslantürk, Berkay Özcan, Oğuzhan Subaş are the employees of Bash Biotech Inc., San Diego, CA, USA. The other authors declare no conflict of interest.

REFERENCES

- Goodwin,S., McPherson,J.D. and McCombie,W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.*, **17**, 333–351.
- Karczewski,K.J. and Snyder,M.P. (2018) Integrative omics for health and disease. *Nat. Rev. Genet.*, **19**, 299–310.
- Hasin,Y., Seldin,M. and Lusis,A. (2017) Multi-omics approaches to disease. *Genome Biol.*, **18**, 83.
- Mardinoglu,A., Bjornson,E., Zhang,C., Klevstig,M., Soderlund,S., Stahlman,M., Adiels,M., Hakkarainen,A., Lundbom,N., Kilicarslan,M. *et al.* (2017) Personal model-assisted identification of NAD(+) and glutathione metabolism as intervention target in NAFLD. *Mol. Syst. Biol.*, **13**, 916.
- Mardinoglu,A., Wu,H., Bjornson,E., Zhang,C., Hakkarainen,A., Rasanen,S.M., Lee,S., Mancina,R.M., Bergentall,M., Pietilainen,K.H. *et al.* (2018) An integrated understanding of the rapid metabolic benefits of a carbohydrate-restricted diet on hepatic steatosis in humans. *Cell Metab.*, **27**, 559–571.
- Chen,R., Mias,G.I., Li-Pook-Tham,J., Jiang,L., Lam,H.Y., Chen,R., Miriami,E., Karczewski,K.J., Hariharan,M., Dewey,F.E. *et al.* (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, **148**, 1293–1307.
- David,L.A., Materna,A.C., Friedman,J., Campos-Baptista,M.I., Blackburn,M.C., Perrotta,A., Erdman,S.E. and Alm,E.J. (2014) Host lifestyle affects human microbiota on daily timescales. *Genome Biol.*, **15**, R89.
- Smarr,L. (2012) Quantifying your body: a how-to guide from a systems biology perspective. *Biotechnol. J.*, **7**, 980–991.
- Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Csardi,G. and Nepusz,T. (2006) The igraph software package for complex network research. *InterJ. Complex Syst.*, **1695**, 1–9.
- Lee,S., Zhang,C., Arif,M., Liu,Z., Benfeitas,R., Bidkhorji,G., Deshmukh,S., Al Shobky,M., Lovric,A., Boren,J. *et al.* (2018) TCSBN: a database of tissue and cancer specific biological networks. *Nucleic Acids Res.*, **46**, D595–D600.
- Tebani,A., Gummesson,A., Zhong,W., Koistinen,I.S., Lakshmikanth,T., Olsson,L.M., Boulund,F., Neiman,M., Stenlund,H., Hellstrom,C. *et al.* (2020) Integration of molecular profiles in a longitudinal wellness profiling cohort. *Nat. Commun.*, **11**, 4487.
- Price,N.D., Magis,A.T., Earls,J.C., Glusman,G., Levy,R., Lausted,C., McDonald,D.T., Kusebauch,U., Moss,C.L., Zhou,Y. *et al.* (2017) A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.*, **35**, 747–756.
- Piening,B.D., Zhou,W., Contrepois,K., Rost,H., Gu Urban,G.J., Mishra,T., Hanson,B.M., Bautista,E.J., Leopold,S., Yeh,C.Y. *et al.* (2018) Integrative personal omics profiles during periods of weight gain and loss. *Cell Syst.*, **6**, 157–170.
- Traag,V.A., Waltman,L. and van Eck,N.J. (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.*, **9**, 5233.
- Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
- Thul,P.J., Akesson,L., Wiking,M., Mahdessian,D., Geladaki,A., Ait Blal,H., Alm,T., Asplund,A., Bjork,L., Breckels,L.M. *et al.* (2017) A subcellular map of the human proteome. *Science*, **356**, eaal3321.
- Uhlen,M., Fagerberg,L., Hallstrom,B.M., Lindskog,C., Oksvold,P., Mardinoglu,A., Sivertsson,A., Kampf,C., Sjostedt,E., Asplund,A. *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
- Uhlen,M., Zhang,C., Lee,S., Sjostedt,E., Fagerberg,L., Bidkhorji,G., Benfeitas,R., Arif,M., Liu,Z., Edfors,F. *et al.* (2017) A pathology atlas of the human cancer transcriptome. *Science*, **357**, eaan2507.
- The UniProt, C. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
- Wishart,D.S., Feunang,Y.D., Marcu,A., Guo,A.C., Liang,K., Vazquez-Fresno,R., Sajed,T., Johnson,D., Li,C., Karu,N. *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, **46**, D608–D617.
- Zhang,C., Bjornson,E., Arif,M., Tebani,A., Lovric,A., Benfeitas,R., Ozcan,M., Juszcak,K., Kim,W., Kim,J.T. *et al.* (2020) The acute effect of metabolic cofactor supplementation: a potential therapeutic strategy against non-alcoholic fatty liver disease. *Mol. Syst. Biol.*, **16**, e9495.
- Oh,T.G., Kim,S.M., Caussy,C., Fu,T., Guo,J., Bassirian,S., Singh,S., Madamba,E.V., Bettencourt,R., Richards,L. *et al.* (2020) A universal gut-microbiome-derived signature predicts cirrhosis. *Cell Metab.*, **32**, 878–888.
- Aron-Wisniewsky,J., Vigliotti,C., Witjes,J., Le,P., Holleboom,A.G., Verheij,J., Nieuwdorp,M. and Clement,K. (2020) Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders. *Nat. Rev. Gastroenterol. Hepatol.*, **17**, 279–297.
- Bajaj,J.S., Heuman,D.M., Hylemon,P.B., Sanyal,A.J., White,M.B., Monteith,P., Noble,N.A., Unser,A.B., Daita,K., Fisher,A.R. *et al.* (2014) Altered profile of human gut microbiome is associated with cirrhosis and its complications. *J. Hepatol.*, **60**, 940–947.
- Chen,Y., Yang,F., Lu,H., Wang,B., Chen,Y., Lei,D., Wang,Y., Zhu,B. and Li,L. (2011) Characterization of fecal microbial communities in patients with liver cirrhosis. *Hepatology*, **54**, 562–572.
- Zhang,X., Coker,O.O., Chu,E.S., Fu,K., Lau,H.C.H., Wang,Y.X., Chan,A.W.H., Wei,H., Yang,X., Sung,J.J.Y. *et al.* (2020) Dietary cholesterol drives fatty liver-associated liver cancer by modulating gut microbiota and metabolites. *Gut*, **70**, 761–774.
- Robinson,J.L., Kocabas,P., Wang,H., Cholley,P.E., Cook,D., Nilsson,A., Anton,M., Ferreira,R., Domenzain,I., Billa,V. *et al.* (2020) An atlas of human metabolism. *Sci. Signal*, **13**, eaaz1482.