

jMorp: Japanese Multi Omics Reference Panel

Shu Tadaka^{1,2}, Daisuke Saigusa^{1,2}, Ikuko N. Motoike^{1,2,3}, Jin Inoue^{1,2}, Yuichi Aoki^{1,2,3}, Matsuyuki Shirota^{1,2,3}, Seizo Koshiba^{1,2}, Masayuki Yamamoto^{1,2} and Kengo Kinoshita^{1,3,4,*}

¹Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryō-machi, Aoba-ku, Miyagi 980-8575, Japan,

²Graduate School of Medicine, Tohoku University, 2-1 Seiryō-machi, Aoba-ku, Miyagi 980-8573, Japan, ³Graduate

School of Information Sciences, Tohoku University, 6-3-09, Aramaki aza Aoba, Aoba-ku, Miyagi 980-8579, Japan and

⁴Institute of Development, Aging and Cancer, Tohoku University, 4-1, Seiryō-machi, Aoba-ku, Miyagi 980-8575, Japan

Received August 10, 2017; Revised October 06, 2017; Editorial Decision October 06, 2017; Accepted October 10, 2017

ABSTRACT

We developed jMorp, a new database containing metabolome and proteome data for plasma obtained from >5000 healthy Japanese volunteers from the Tohoku Medical Megabank Cohort Study, which is available at <https://jmorp.megabank.tohoku.ac.jp>. Metabolome data were measured by proton nuclear magnetic resonance (NMR) and liquid chromatography–mass spectrometry (LC–MS), while proteome data were obtained by nanoLC–MS. We released the concentration distributions of 37 metabolites identified by NMR, distributions of peak intensities of 257 characterized metabolites by LC–MS, and observed frequencies of 256 abundant proteins. Additionally, correlation networks for the metabolites can be observed using an interactive network viewer. Compared with some existing databases, jMorp has some unique features: (i) Metabolome data were obtained using a single protocol in a single institute, ensuring that measurement biases were significantly minimized; (ii) The database contains large-scale data for healthy volunteers with various health records and genome data and (iii) Correlations between metabolites can be easily observed using the graphical viewer. Metabolites data are becoming important intermediate markers for evaluating the health states of humans, and thus jMorp is an outstanding resource for a wide range of researchers, particularly those in the fields of medical science, applied molecular biology, and biochemistry.

INTRODUCTION

Clarifying interactions between the genome and environmental factors is important for personalized healthcare and medication development to overcome problems of missing heritability (1). For this purpose, genome cohort stud-

ies have been carried out in many countries, such as UK biobank in the United Kingdom and deCODE in Iceland, where great effort has been exerted to collect precise health and medical information to define the phenotypes of each participant. Typically, health status information is determined through biochemical tests and questionnaires, which is effective for some genome cohorts, but it also exhibits some limitations. For example, biochemical tests typically assess a limited number of items, and questionnaires contain various kinds of noise such as recall bias and/or self-selection bias, because they rely on participant responses.

To overcome these limitations, a genome cohort recently began using multi-omics analyses to define the molecular phenotypes of each person. The most illustrative example is LifeLines-deep study (2), a subcohort of LifeLines in the Northern Netherlands, where 1500 participants were selected from 167 000 participants (3) and subjected to multi-omics analyses such as DNA methylation analyses, gene expression analyses, plasma metabolome analyses, and gut metagenome analyses, in addition to whole genome analyses. The results of the study indicated that multi-omics analyses with cohort studies are highly effective for analyzing gene-environment interactions (see (4) for gut metagenomics and (5) for disease and methylation/transcription).

In Japan, Tohoku University Tohoku Medical Megabank Organization (ToMMO) (6) conducted multi-omics (metabolome and proteome) analyses of 5093 plasma samples collected from Japanese (male: 2077, female: 3016) residents who participated in the Tohoku Medical Megabank Project Cohort Study. By using some of these data, Koshiba et al detected five interesting associations between the genome and plasma metabolome (TCN000004 formate, TCN000017 asparagine, TCN000019 phenylalanine, TCN000031 proline and TCN000033 glycine) (7). To share the data and promote the development of personalized healthcare, we created a new database known as jMorp for metabolome and proteome data in plasma obtained from 5,093 healthy volunteers in a Japanese population from the Tohoku Medical Megabank Cohort Study.

*To whom correspondence should be addressed. Tel: +81 22 795 7179; Fax: +81 22 795 7179; Email: kengo@ecei.tohoku.ac.jp

Several databases contain metabolome data, such as HMDB (Human Metabolome Database, (8)), MMCD (Madison Metabolomics Consortium Database; a resource for metabolomics research based on nuclear magnetic resonance (NMR) and mass spectrometry (MS), (9)), MetaboLights (a database for metabolomics experiments and the associated metadata, (10)). These databases are important as repositories of raw data, but jMorp has some advantages compared to existing databases: (a) metabolome data were obtained using a single protocol in a single institute, ensuring that measurement biases were significantly minimized; (b) jMorp is built using large-scale cohort data for healthy volunteers with various health records and genome data, and it provides significant GWAS results and (c) correlations between metabolites could be easily observed with the graphical viewer.

Overview of available data and functionalities

Metabolome data were measured by proton NMR and liquid chromatography (LC)–MS, and proteome data were obtained by nanoLC–MS as described in the Methods section. In the current version of jMorp, we have released the concentration distributions of 37 metabolites identified by NMR, peak intensity distributions of 257 metabolites by LC–MS, and observed frequencies of 256 abundant proteins. All distributions of metabolites were prepared for male, female, and all samples and divided by age categories; those of the protein detection rate are shown separately for male, female, and all samples. In addition to the distributions, correlation networks of the metabolites are provided using an interactive network viewer. Concentration data, genotypes, and biological specimens for each participant are provided by Tohoku Medical Megabank under controlled access after carefully checking the data and obtaining approval from the sample access committee because of ethical reasons (Table 1).

Overview of jMorp database (Figure 1)

jMorp mainly consists of three types of pages: search page, metabolite page, and protein page. The search page is the central page that has a search window with toggle buttons for platform categories, ‘metabolites [NMR]’, ‘metabolites [MS]’, and ‘proteins’ (Figures 1A and 2A), and a compound table of compound IDs, compound names, platform categories and basic statistics. Initially, the first 20 entries sorted by compound ID are shown; this section is replaced with the search results after users carry out a search (Figures 1B and 2B). Text search can be entered in the text box at the top of the table, where we have implemented incremental search functionality for a better user experience. In Figure 1B, the search results for ‘Phenyl’ are shown, while the results for ‘Apolipoprotein’ are presented in Figure 2B. Compound IDs and names in the search results are linked to each compound page. Compound pages are prepared independently for metabolites and proteins. Notably, MS data can be searched for a range of m/z values. For example, a search with ‘ $100 < m/z < 200$ ’ will show a compound list with m/z values ranging from 100 to 200. All search results can be downloaded from the download icon at the right top corner of the table.

On a compound page for a metabolite (Figure 1C), users will find four or five sections. First, we provide the distributions of the compound concentration (μM unit) for NMR data or abundance (or corrected peak intensities) for MS data. The distributions are shown for all (grey bars in Figure 1C), male (blue bars), and female (red bars) samples. To the left of the distributions, we show changes in metabolite abundance across age groups, where data are separated by sex. It should be noted that distribution changes with age are important when analyzing omics data, as described below (Figure 3). Third, links to other public compound databases (DB) are provided below the distributions. Links to HMDB (8), KEGG Compound (11) and LIPID MAPS (12) are shown if the focused compound is available in these DBs. In addition, some notes regarding measurements such as m/z , retention time, measurement ion mode, and column mode are shown if the metabolite is observed by ‘metabolites [MS]’. In addition to this basic information, five metabolite pages containing reported genome-wide association study (GWAS) results (7) by using a Manhattan plot with variations in the compound across genotypes at the most significant variant. GWAS results will be added to the future release, when we have obtained GWAS results with new jMorp data and corresponding genomics data. Finally, a table of correlation between the focused metabolite and other metabolites among the population is shown. In the table, the Spearman’s rank correlation coefficient ($|r_s| > 0.2$) and P -value between two metabolites are shown. An interactive correlation network view is also available through the link ‘view as network’ above the correlation table (Figure 1D). The correlation information is one of the main contents of jMorp. As far as we know, this is the first database of metabolite correlations among a healthy population. Some interesting biological and biomedical findings from the correlation data will be described elsewhere.

On a compound page for a protein, there are three or four sections: (i) detection rate of the protein for population, all/male/female (Figure 2C), (ii) links to UniProt and (iii) peptide sequences containing reference or alternative alleles resulting from non-synonymous genomic variants shown in a table format. Measurement information, m/z , charge, and modifications for each peptide are also shown. If there is a peptide sequence for the genomic variant, the variant ID and change in amino acid is annotated in the table. Users can obtain more information about the genomic variant using the link from the mutation to the Integrative Japanese Genome Variation Database (13).

Note that each compound page has its own URL, in the form of [https://jmorp.megabank.tohoku.ac.jp/\[year\]/compounds/\[compound ID\]](https://jmorp.megabank.tohoku.ac.jp/[year]/compounds/[compound ID]); therefore, links from the external server can be easily implemented. The naming rule of the compound ID in jMorp is described in the Methods section.

Age and BMI distributions

In omics analyses, age and body mass index (BMI) are important factors affecting the results, and thus it is important to achieve flat distributions for age and normal distributions for BMI. Figure 3 shows a change in the age and BMI distributions of the analyzed participants for each year. For

Table 1. Basic statistics of the data in jMorp

Data Types	# of Items		# of Samples		
	2015	2016–2017	2015	2016	2017
Basic Info.	3 (sex, age, BMI)		501	1008	5093
Metabolome [NMR]	37		501	1008	5093
Metabolome [MS]	201	257	501	1008	1312
Proteome [MS]	256			501	

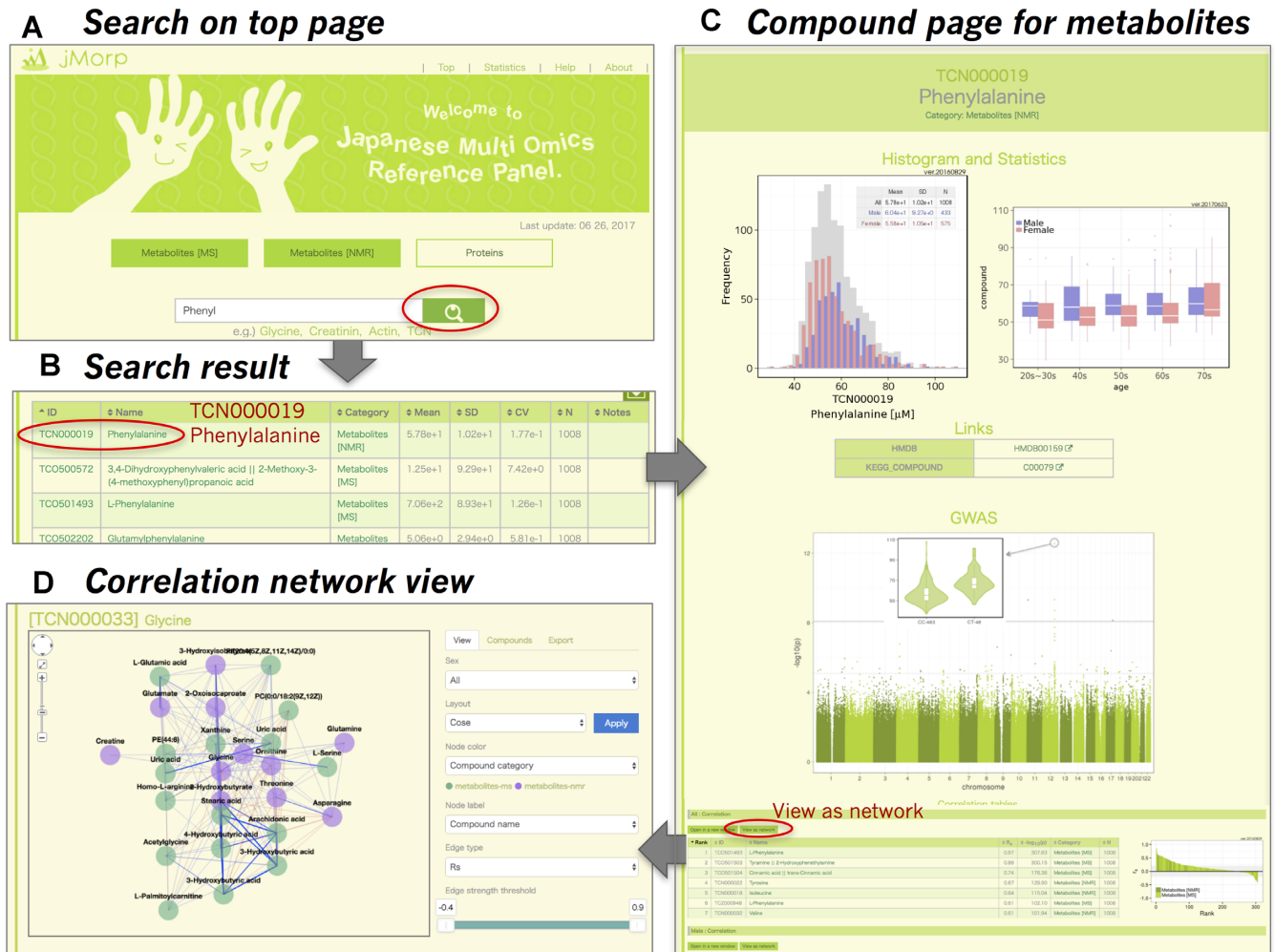


Figure 1. An example of metabolites search in jMorp. Thin red circles in (A), (B) and (C) indicate the link to the next step specified by gray arrows. (A) Search page. (B) Keyword search result. (C) Metabolite page for phenylalanine. The distribution plot of phenylalanine concentration and variation in concentration across age groups are shown. For this metabolite, variations among sex or age groups can be observed. Users can obtain information regarding this metabolite from HMDB or KEGG Compound. GWAS Manhattan plot and variation of concentration among genotypes at the most significant genomic variant relationship to the metabolites are shown. At the bottom of the compound page, correlations with other metabolites are listed. (D) A network viewer of correlation network among metabolites.

example, for phenylalanine, the concentration distribution monotonically increases across age groups, particularly in females (Figure 1C), while glucose concentrations tend to increase according to age (<https://jmorp.megabank.tohoku.ac.jp/2017/compounds/TCN000037>).

Versioning policy

jMorp was first released in July 2015 and gradually updated by increasing the number of samples and implement-

ing various functionalities. All historical data are available at [https://jmorp.megabank.tohoku.ac.jp/\[year\]](https://jmorp.megabank.tohoku.ac.jp/[year]), where [year] can be 2015, 2016, or 2017. The default version is now 2017, as metabolite data obtained by NMR were manually quantified by an expert. We are currently attempting to improve the automated quantification method described in the Methods section. After improvement, we will replace the default version to 2017, but will retain all versions for backward compatibility.

A Search on top page
B Search result

13 items exist

ID	Name	Category	Mean	SD	CV	N	Notes
TCP000002	Apolipoprotein B-100	Proteins					90-100%
TCP000010	Apolipoprotein A-I	Proteins					90-100%
TCP000020	Apolipoprotein A-IV	Proteins					90-100%
TCP000021	Apolipoprotein D	Proteins					90-100%
TCP000028	Apolipoprotein C-III	Proteins					90-100%
TCP000034	Apolipoprotein C-I	Proteins					90-100%
TCP000043	Apolipoprotein A-II	Proteins					90-100%
TCP000055	Apolipoprotein E	Proteins					90-100%
TCP000056	Apolipoprotein M	Proteins					90-100%
TCP000084	Apolipoprotein C-II	Proteins					70-80%
TCP000087	Apolipoprotein L1	Proteins					70-80%
TCP000100	Apolipoprotein F	Proteins					60-70%
TCP000152	Apolipoprotein(a)	Proteins					30-40%

Thin red circles in the original image highlight the 'Proteins' button in (A), the 'Apo' search bar in (B), and the 'rs429358' variant in (C).

C Compound page for proteins

TCP000055
Apolipoprotein E
Category: Proteins

Detection rates

All	Male	Female
90-100%	90-100%	90-100%

Links

UNIPROT P02649 E2

Peptides

Start	End	Sequence	rsID	Modification	m/z	Charge	Modifications
122	130	LGADMEDVYR	rs429358	OX	611.23467	2	M5(Oxidation)
122	132	LGADMEDVCGR	rs429358	C13OR	611.76359	2	C9(Carbamidomethyl)
122	137	LGADMEDVCGRLVQYR	rs429358	C13OR	627.9696	3	C9(Carbamidomethyl)
122	130	LGADMEDVYR	rs429358	C13OR	503.23694	2	
94	108	SELEEQITPVAETR			866.4256	2	O6(Deamidated)
94	108	SELEEQITPVAETR			865.92615	2	
R/I	R/I	AI/MNCTLRK		DEF	460.7166	2	

D genetic variant information

Peptides

INTEGRATIVE Japanese Genome Variation Database
The genome cohort study of Tohoku Medical Megabank Organization

rsID (ex. rs671, rs6721961):
rs429358

Result: 1 row(s)

Position	rsID	Reference allele	Reference allele frequency (1KJPN)	Reference allele count (1KJPN)	Reference allele frequency (2KJPN)	Reference allele count (2KJPN)	Alternative allele	Alternative allele frequency (1KJPN)
chr19:45411941	rs429358	T	0.8947	1759	0.8966	3348	C	0.1053

Tohoku Medical Megabank Organization, Tohoku University

Figure 2. An example of protein search in jMorp. The thin red circles in (A), (B) and (C) indicate the link to the next step indicated by gray arrows. (A) Search page. (B) Keyword search result. (C) Protein page for apolipoprotein E. The detection rate of apolipoprotein E is shown. User can obtain information regarding this metabolite from the UniProt database. Detected peptide information for this protein is shown at the bottom of this page.

METHODS**NMR measurements**

Blood samples from cohort participants were collected using vacutainer tubes containing EDTA-2Na (Venoject II, Terumo Corporation, Tokyo, Japan). Plasma was prepared and stored at -80°C using a MATRIX[®] 2D screw tube (Thermo Scientific, Waltham, MA, USA). Metabolites were extracted using a standard methanol extraction procedure with 200 μl of plasma. Extracted metabolites were suspended in 200 μl of 100 mM sodium phosphate buffer (pH 7.4) in 100% D₂O containing 200 μM d6-DSS. All NMR experiments were performed at 298 K on a Bruker 600 MHz spectrometer (Bruker BioSpin, Germany). Standard 1D nuclear Overhauser effect spectroscopy (NOESY) and Carr-Purcell-Meiboom-Gill (CPMG) spectra were obtained for each plasma sample. All data were processed using the Chenomx NMR Suite (Chenomx, Edmonton, Canada). Metabolites were identified and quantified using the target profiling approach implemented in the Chenomx Profiler module.

The quality control of the cohort sample is important to secure the validity of the clinical result and to create an omics reference. In our analyses, the influences of common pre-analytical variation on the human plasma metabolites were evaluated by the abundance of specific compounds by GC-MS assay as reported in Kamlage *et al.* (14). In addition, glucose and lactate concentration changes were most pronounced in plasma metabolites after the exposure of EDTA blood to 25°C , and thus we also checked the concentrations of glucose and lactate to exclude the low-quality samples.

Metabolite concentrations were manually estimated until the 2016 release, and then automatically estimated from NMR spectra by using several regression models beginning in 2017. More than 1,000 concentration data manually calculated by experts by 2016 were used for training data for later automated quantification. Both linear regression and neural network models were used. A suitable model for each metabolite was selected from the best R-squared (R^2) values as an evaluation index. We provide a reliability score of the estimated concentration on a four-tiered scale: 'Triple Stars

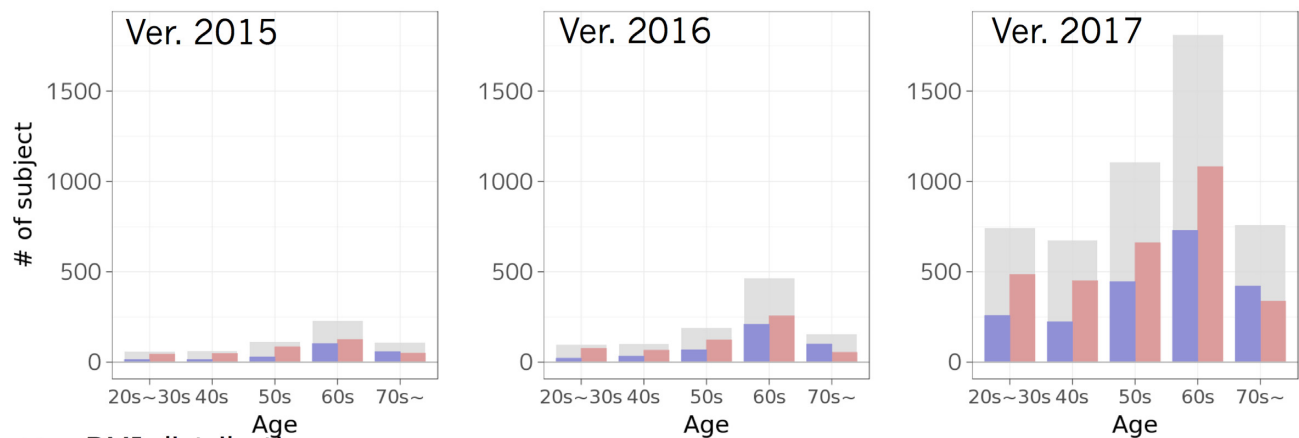
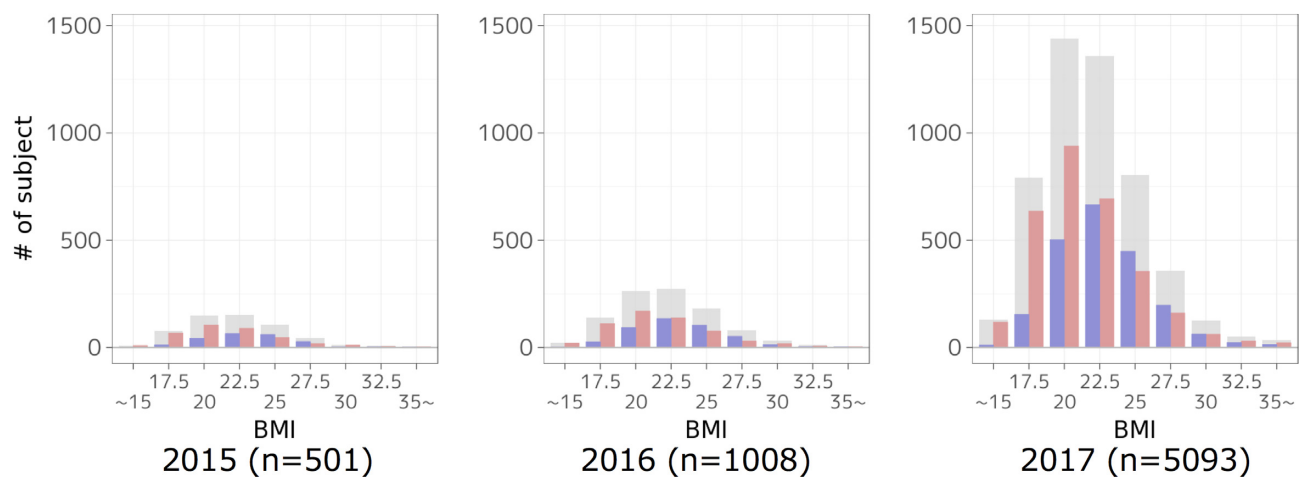
A Age distribution**B** BMI distribution

Figure 3. Change in the distributions of age and BMI of analyzed participants for each year. Distributions are shown for all (grey bars), male (blue bars) and female (red bars) samples, respectively.

(★★), ‘Double Stars (★★★)’, ‘Single Star (★☆☆)’ and ‘Zero Star (☆☆☆)’. Each category corresponds to an R^2 value of ≥ 0.9 , ≥ 0.7 , ≥ 0.6 and < 0.6 , respectively.

MS measurements

Fifty microliters of each plasma sample were transferred into a single well of a 96-well sample collection plate, and total of 1312 plasma samples were added to 19-well plates with a reference quality control used for normalization between batches. ultra-high-performance LC-quadrupole time-of-flight (QTOF)/MS analysis was performed on an Acquity Ultra Performance LC I-class system equipped with a binary solvent manager, sample manager, and column heater (Waters Corp., Milford, MA, USA). This system interfaced with a Waters Synapt G2-Si QTOF MS with electrospray ionization (ESI) system operated in positive ion mode. LC separation was performed using a C18 column (Acquity HSS T3; 150×2.1 mm i.d., $1.8 \mu\text{m}$ particle size; Waters) with a gradient elution of solvent A (water containing 0.01% formic acid) and solvent B (acetonitrile containing 0.01% formic acid) at $400 \mu\text{l min}^{-1}$. The data were collected using MassLynx, v4.1 soft-

ware (Waters Corp.). The LC-FTMS system consisted of a NANOSPACe SI-II HPLC equipped with a dual pump system, auto sampler, and column oven (Shiseido, Tokyo, Japan) as well as a Q Exactive Orbitrap MS (Thermo Fisher Scientific) equipped with a heated-ESI-II source for negative ion mode. LC separation was performed using an HILIC column (ZIC-pHILIC; 100×2.1 mm i.d., $5 \mu\text{m}$ particle size; Sequant, Darmstadt, Germany) with a gradient elution of solvent A (10 mM ammonium bicarbonate in water, pH 9.2) and solvent B (acetonitrile) at $300 \mu\text{l min}^{-1}$. The data were collected using Xcalibur v4.1 software (Thermo Fisher Scientific). The ultra-high-performance LC-QTOF/MS and LC-FTMS operating conditions have been described previously (15).

Proteome analysis

Plasma samples were heat-treated and then reacted with lysyl-endopeptidase. Plasma proteins were subsequently digested by trypsin and desalinated. These plasma samples were analyzed in triplicate by LC-tandem MS. Three mass spectrometers were used to process 501 plasma samples from 501 individuals; 233 samples by Thermo Scientific

Orbitrap Fusion, 156 samples by Thermo Scientific Elite, and 112 samples by Thermo Scientific Q Exactive. Peptide identification from mass spectra was obtained using the SequestHT and Mascot search engines with the UniProt human proteome data set from April 2014 as reference protein sequences. These peptide identification results were integrated using Proteome Discoverer1.4. The abundance of a specific protein in the 501 samples was calculated as the fraction of samples in which the protein was successfully identified. To identify peptides resulting from non-synonymous genomic variations, we created a data set of protein sequences containing alternative alleles found in at least 5% of the ToMMo 1KJPN cohort. This database was also searched for peptide sequences harboring alternative alleles (amino acids). All reference and alternative peptide sequences observed in our plasma proteome analysis are listed on the peptide table page.

Network viewer

The web-based network viewer on each metabolite page for correlation networks was implemented using Cytoscape.js (16). The network representation of a correlation table of a metabolite gives users an overview of correlation relationships among metabolites. When a metabolite is selected as a seed for network construction, the viewer searches metabolites with strong correlations with the seed and draws a network in which a node corresponds to a metabolite and edge corresponds to the correlation relationship between two metabolites. Users can interactively navigate the generated networks and perform basic analysis, such as filtering edges by setting a cutoff threshold for correlation strength. Networks generated by the network viewer can be saved as PNG images or GraphML format files for further analysis.

GWAS view

GWAS with metabolites results are shown as a Manhattan plot and violin plots of the metabolite across genotypes. The Manhattan plot shows the $-\log_{10}P$ values for each single-nucleotide polymorphism along the genomic coordinates from the previous GWAS study (7), while the violin plot displays differences in metabolite concentrations among genotypes for the reported single-nucleotide polymorphism with the highest association P-value.

ToMMo compound ID

In jMorp, compound IDs are shown in the form TCx123456; the first two letters, TC, indicate that the ID is ToMMo compound ID (TC-ID). The following one letter indicates the data source. In the current version, we provide four types of data, each of which has its own one-letter code for the data source: (P) proteome by MS, (N) NMR, (Z) MS metabolome in HILIC mode and (O) MS metabolome in C18 mode. The last six digits are unique numbers for each compound in each data source. In the MS metabolome, data sources can be further divided into positive mode and negative mode. If the six digits are less than 500 000, the compound was identified by negative mode. Otherwise, the compound was found by positive mode.

AVAILABILITY

jMorp is freely available at <https://jmorp.megabank.tohoku.ac.jp>.

ACKNOWLEDGEMENTS

We are indebted to all volunteers who participated in this ToMMo project. We would like to acknowledge all members associated with this project; the member list is available at the following web site: <http://www.megabank.tohoku.ac.jp/english/a170601/>. All computational resources were provided by the ToMMo supercomputer system. We thank Mr. Kota Jin for the web design.

FUNDING

Platform Program for Promotion of Genome Medicine [16815713] from the Japan Agency for Medical Research and Development (AMED); Tohoku Medical Megabank Project of Ministry from Education, Culture, Sports, Science and Technology (MEXT); Center of Innovation Program from Japan Science and Technology Agency (JST). Funding for open access charge: Platform Program for Promotion of Genome Medicine [16815713] from the Japan Agency for Medical Research and Development (AMED). *Conflict of interest statement.* None declared.

REFERENCES

- Manolio,T.A., Collins,F.S., Cox,N.J., Goldstein,D.B., Hindorf,L.A., Hunter,D.J., McCarthy,M.I., Ramos,E.M., Cardon,L.R., Chakravarti,A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.
- Tigchelaar,E.F., Zhernakova,A., Dekens,J.A.M., Hermes,G., Baranska,A., Mujagic,Z., Swertz,M.A., Muñoz,A.M., Deelen,P., C nit,M.C. *et al.* (2015) Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open*, **5**, e006772.
- Scholtens,S., Smidt,N., Swertz,M.A., Bakker,S.J.L., Dotinga,A., Vonk,J.M., van Dijk,F., van Zon,S.K.R., Wijmenga,C., Wollfienbuttel,B.H.R. *et al.* (2015) Cohort profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.*, **44**, 1172–1180.
- Zhernakova,A., Kurilshikov,A., Bonder,M.J., Tigchelaar,E.F., Schirmer,M., Vatanen,T., Mujagic,Z., Vila,A.V., Falony,G., Vieira-Silva,S. *et al.* (2016) Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, **352**, 565–569.
- Bonder,M.J., Luijk,R., Zhernakova,D. V., Moed,M., Deelen,P., Vermaat,M., van Itersson,M., van Dijk,F., van Galen,M., Bot,J. *et al.* (2017) Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.*, **49**, 131–138.
- Kuriyama,S., Yaegashi,N., Nagami,F., Arai,T., Kawaguchi,Y., Osumi,N., Sakaida,M., Suzuki,Y., Nakayama,K., Hashizume,H. *et al.* (2016) The Tohoku medical megabank project: design and mission. *J. Epidemiol.*, **26**, 493–511.
- Koshihara,S., Motoike,I., Kojima,K., Hasegawa,T., Shirota,M., Saito,T., Saigusa,D., Danjoh,I., Katsuoka,F., Ogishima,S. *et al.* (2016) The structural origin of metabolic quantitative diversity. *Sci. Rep.*, **6**, 31463.
- Wishart,D.S., Jewison,T., Guo,A.C., Wilson,M., Knox,C., Liu,Y., Djoumbou,Y., Mandal,R., Aziat,F., Dong,E. *et al.* (2013) HMDB 3.0-The human metabolome database in 2013. *Nucleic Acids Res.*, **41**, D801–D807.
- Cui,Q., Lewis,I.A., Hegeman,A.D., Anderson,M.E., Li,J., Schulte,C.F., Westler,W.M., Eghbalian,H.R., Sussman,M.R. and

- Markley, J.L. (2008) Metabolite identification via the Madison Metabolomics Consortium Database. *Nat. Biotechnol.*, **26**, 162–164.
10. Haug, K., Salek, R.M., Conesa, P., Hastings, J., de Matos, P., Rijnbeek, M., Mahendraker, T., Williams, M., Neumann, S., Rocca-Serra, P. *et al.* (2013) MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.*, **41**, D781–D786.
11. 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.
12. Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Seyama, Y., Shaw, W. *et al.* (2005) A comprehensive classification system for lipids. *J. Lipid Res.*, **46**, 839–862.
13. Yamaguchi-Kabata, Y., Nariyai, N., Kawai, Y., Sato, Y., Kojima, K., Tateno, M., Katsuoka, F., Yasuda, J., Yamamoto, M. and Nagasaki, M. (2015) iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. *Hum. Genome Var.*, **2**, 15050.
14. Kamlage, B., Maldonado, S.G., Bethan, B., Peter, E., Schmitz, O., Liebenberg, V. and Schatz, P. (2014) Quality markers addressing preanalytical variations of blood and plasma processing identified by broad and targeted metabolite profiling. *Clin. Chem.*, **60**, 399–412.
15. Saigusa, D., Okamura, Y., Motoike, I.N., Katoh, Y., Kurosawa, Y., Saijyo, R., Koshiba, S., Yasuda, J., Motohashi, H., Sugawara, J. *et al.* (2016) Establishment of protocols for global metabolomics by LC–MS for biomarker discovery. *PLoS One*, **11**, e0160555.
16. Franz, M., Lopes, C.T., Huck, G., Dong, Y., Sumer, O. and Bader, G.D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.