nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-018-0047-6

In the format provided by the authors and unedited.

Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases

Masahiro Kanai[1,2,3], Masato Akiyama[2], Atsushi Takahashi[2,4], Nana Matoba[2], Yukihide Momozawa[5], Masashi Ikeda[6], Nakao Iwata[6], Shiro Ikegawa[7], Makoto Hirata[8], Koichi Matsuda[9], Michiaki Kubo[10], Yukinori Okada[1,2,11,13]* and Yoichiro Kamatani[2,12,13]*

[1]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. [2]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [4]Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center, Osaka, Japan. [5]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [6]Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan. [7]Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan. [8]Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [9]Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [10]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [11]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan. [12]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. 13These authors jointly supervised this work: Yukinori Okada and Yoichiro Kamatani. *e-mail: yokada@sg.med.osaka-u.ac.jp; yoichiro.kamatani@riken.jp

nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-018-0047-6

In the format provided by the authors and unedited.

Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases

Masahiro Kanai[1,2,3], Masato Akiyama[2], Atsushi Takahashi[2,4], Nana Matoba[2], Yukihide Momozawa[5], Masashi Ikeda[6], Nakao Iwata[6], Shiro Ikegawa[7], Makoto Hirata[8], Koichi Matsuda[9], Michiaki Kubo[10], Yukinori Okada[1,2,11,13]* and Yoichiro Kamatani[2,12,13]*

[1]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. [2]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [4]Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center, Osaka, Japan. [5]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [6]Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan. [7]Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan. [8]Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [9]Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [10]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [11]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan. [12]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. 13These authors jointly supervised this work: Yukinori Okada and Yoichiro Kamatani. *e-mail: yokada@sg.med.osaka-u.ac.jp; yoichiro.kamatani@riken.jp

Supplementary Information

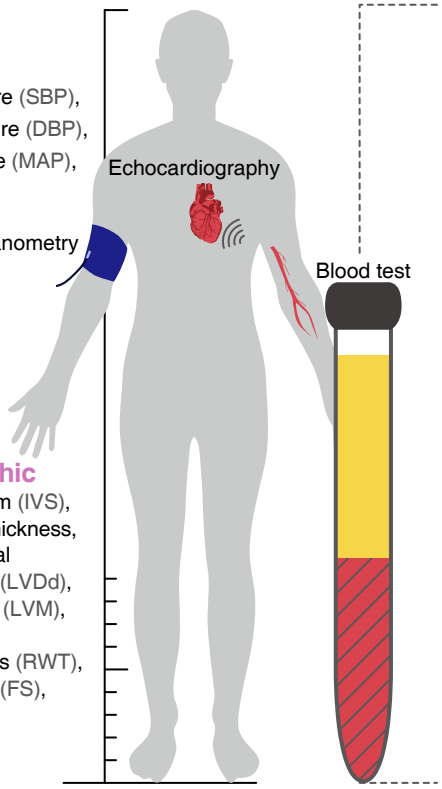
Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases

Kanai, M. *et al.*

Supplementary Figures

Supplementary Figure 1 | Overview of the study.
Summary of the studied quantitative traits, the additional data integration as well as the flow of analyses.

GWAS of the 58 quantitative traits in Japanese individuals (N = 162,255)



Blood pressure
Systolic blood pressure (SBP),
Diastolic blood pressure (DBP),
Mean arterial pressure (MAP),
Pulse pressure (PP)

Echocardiographic
Interventricular septum (IVS),
Posterior wall (PW) thickness,
Left ventricular internal dimension in diastole (LVDd), systole (LVDs), mass (LVM),
LVM index (LVMI),
Relative wall thickness (RWT),
Fractional shortening (FS),
Ejection fraction (EF),
E/A ratio (E/A)

Anthoropometric
Adult height,
Body mass index (BMI)

Additional GWAS results from BBJ

Metabolic
Total cholesterol (TC), HDL-C, LDL-C, Triglyceride (TG),
Blood sugar (BS), Hemoglobin A1c (HbA1c)

Protein
Total protein (TP), Albumin (Alb), Non-albumin protein (NAP),
Albumin/globulin ratio (A/G)

Kidney-related
Blood urea nitrogen (BUN), Serum creatinine (sCr),
Estimated glomerular filtration rate (eGFR), Uric acid (UA)

Electrolyte
Serum sodium (Na), Potassium (K), Chloride (Cl),
Calsium (Ca), Phosphorus (P)

Liver-related
Total bilirubin (TBil), Zinc sulfate turbidity test (ZTT),
Aspartate aminotransferase (AST), Alanine aminotransferase (ALT),
Alkaline phosphatase (ALP), Gamma-glutamyl transferase (GGT)

Other biochemical
Activated partial thromboplastin time (APTT), Prothrombin time (PT),
Fibrinogen (Fbg), Creatine kinase (CK), Lactate dehydrogenase (LDH),
C-reactive protein (CRP)

Hematological
White blood cell count (WBC), Neutrophil (Neutro), Eosinophil (Eosino),
Basophil (Baso), Monocyte (Mono), Lymphocyte (Lym) count,
Red blood cell count (RBC), Hemoglobin (Hb), Hematocrit (Ht),
Mean corpuscular volume (MCV), Mean corpuscular hemoglobin (MCH),
Mean corpuscular hemoglobin concentration (MCHC),
Platelet count (Plt)

1,407 trait-associated loci (763 unique loci)
($P < 5.0 \times 10^{-8}$)

Pleiotropy among the quantitative traits
339 pleiotropic loci (119 inter-categorical)

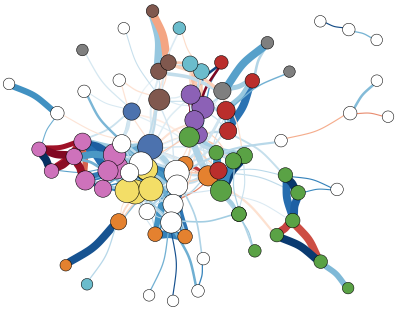
Complex diseases
Cardiometabolic (n = 6),
Immune-related (n = 6),
Hematologic (n = 1),
Psychiatric (n = 2),
Musculoskeletal (n = 2),
Cancer (n = 7),
Others (n = 6)

Additional 30 case-control GWAS results in Japanese

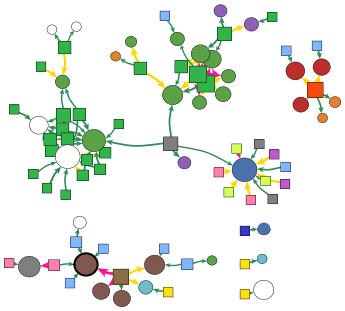
Genetic correlations across the quantitative traits
173 significant correlations (100 inter-categorical)

Genetic correlations
across the quantitative traits and diseases
68 significant genetic correlations

Heritability enrichments
in the cell-type specific annotations
384 significant heritability enrichments



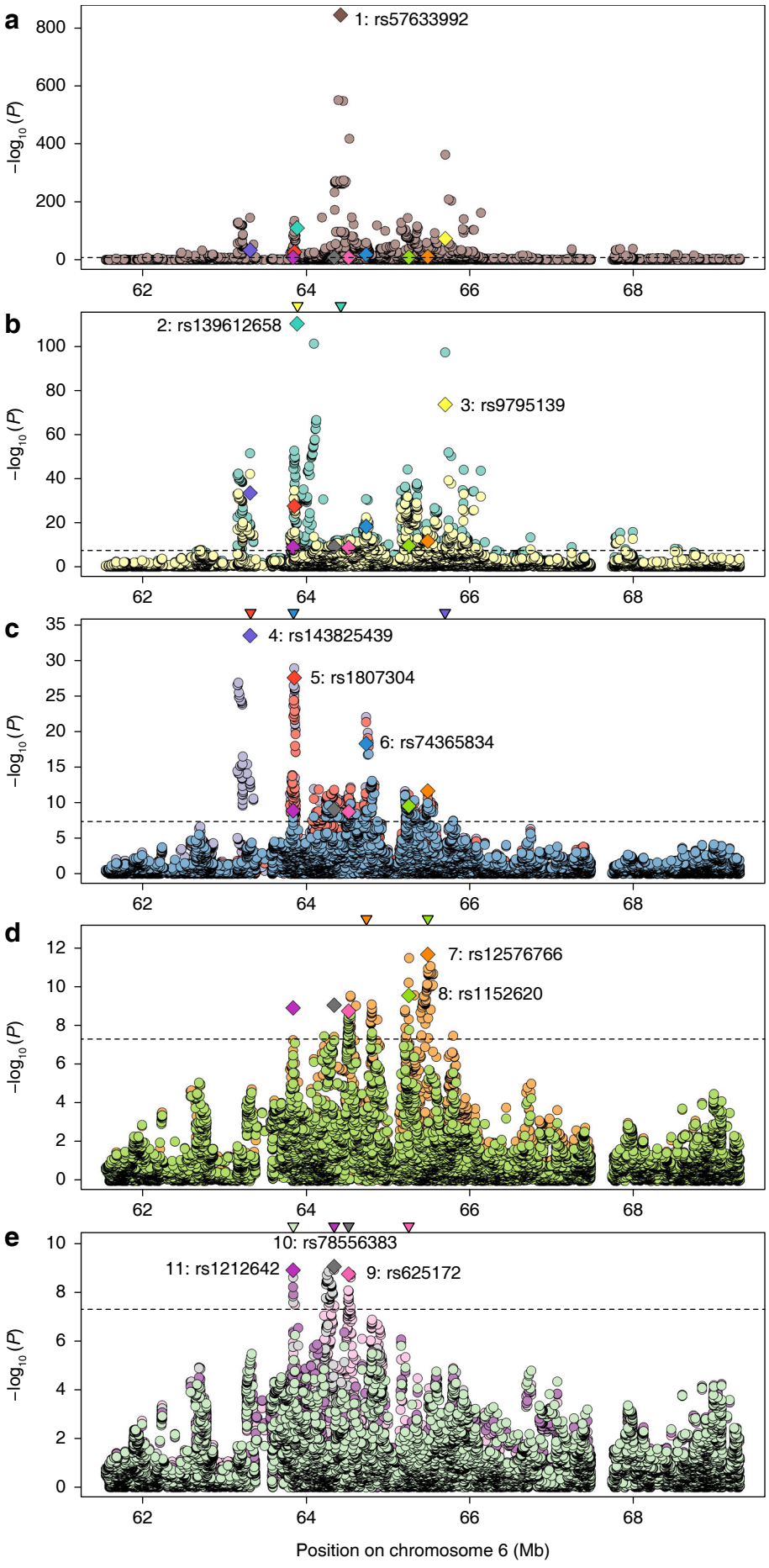
Network visualization



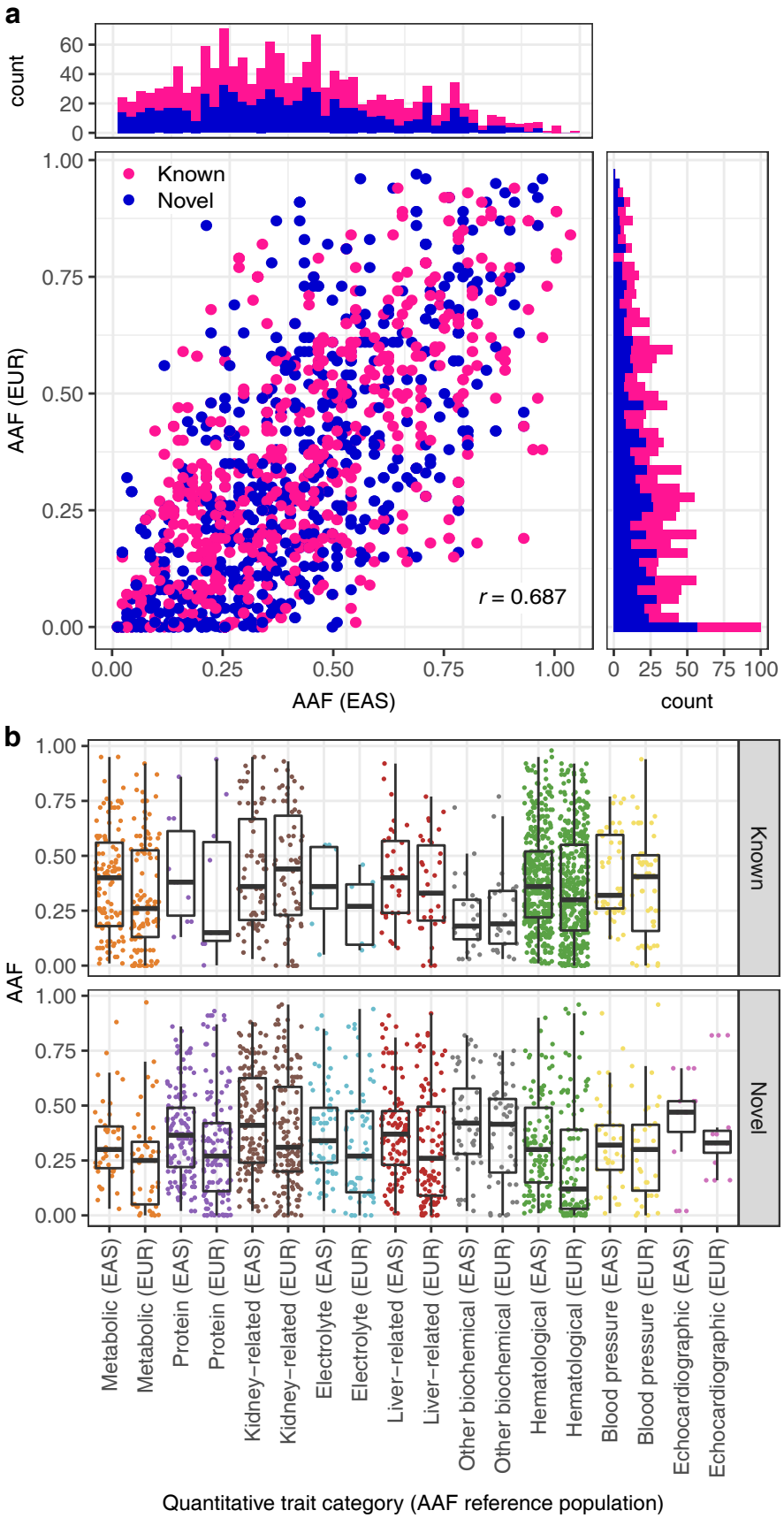
Supplementary Figure 2 | Regional plots of 11q13.1 for UA.

Stepwise conditional analysis identified 11 independent signals on 11q13.1 for UA (n = 109,029). Distinct colors correspond to distinct association results, from the nominal result (no conditioning) to the 11th-conditioned result (conditioned on all of the 11 independent signals). Each diamond represents the most significant variant in each step and the same color triangle represents the position of the additionally conditioned variant in the step. GWAS was conducted using a linear regression model under the assumption of additive allelic effects of the SNP dosages. The dotted horizontal lines represent a genome-wide significance threshold of $P = 5.0 \times 10^{-8}$.

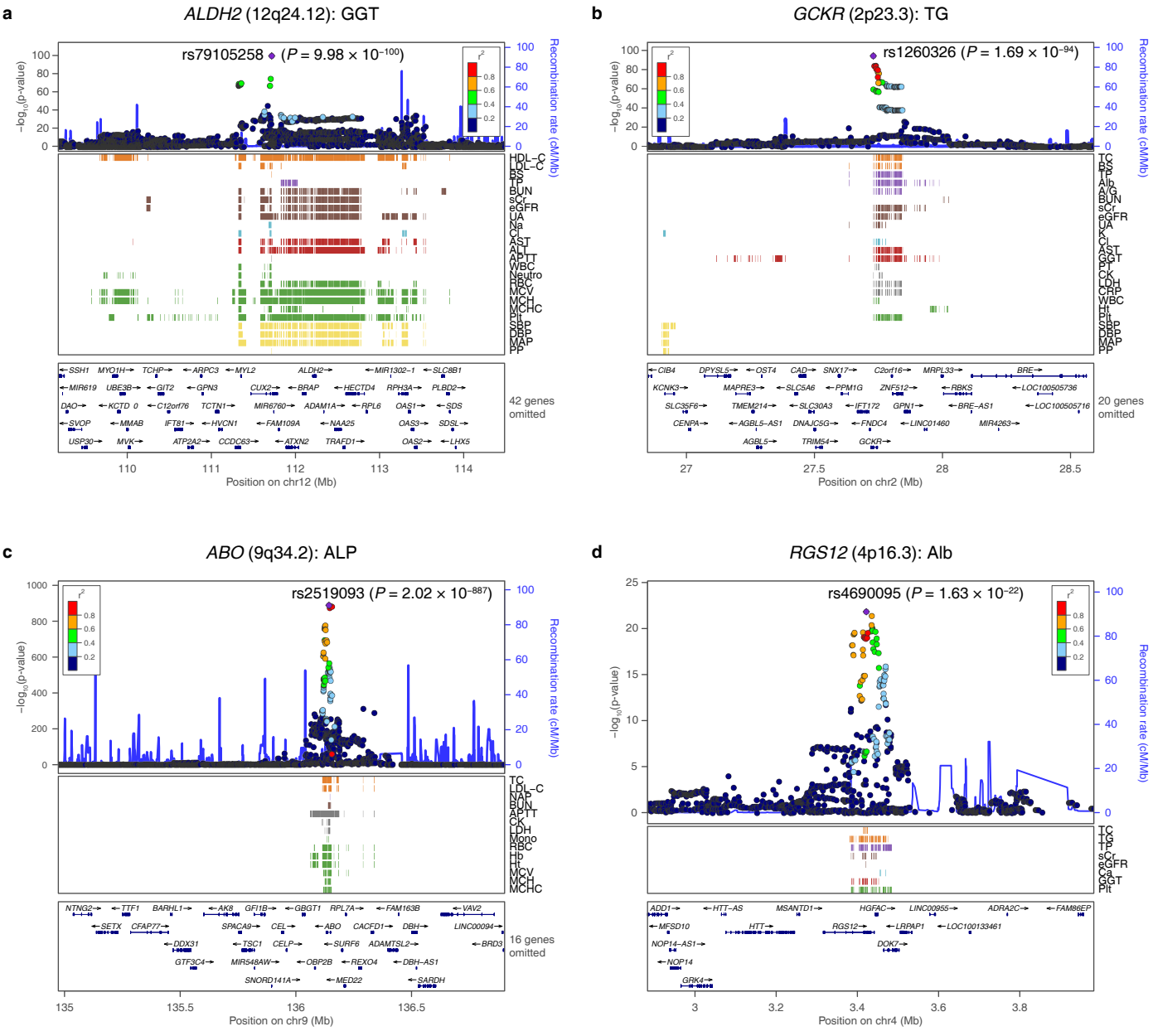
a, the nominal association result. **b–e**, the conditional association results, conditioned on the corresponding independent signals.



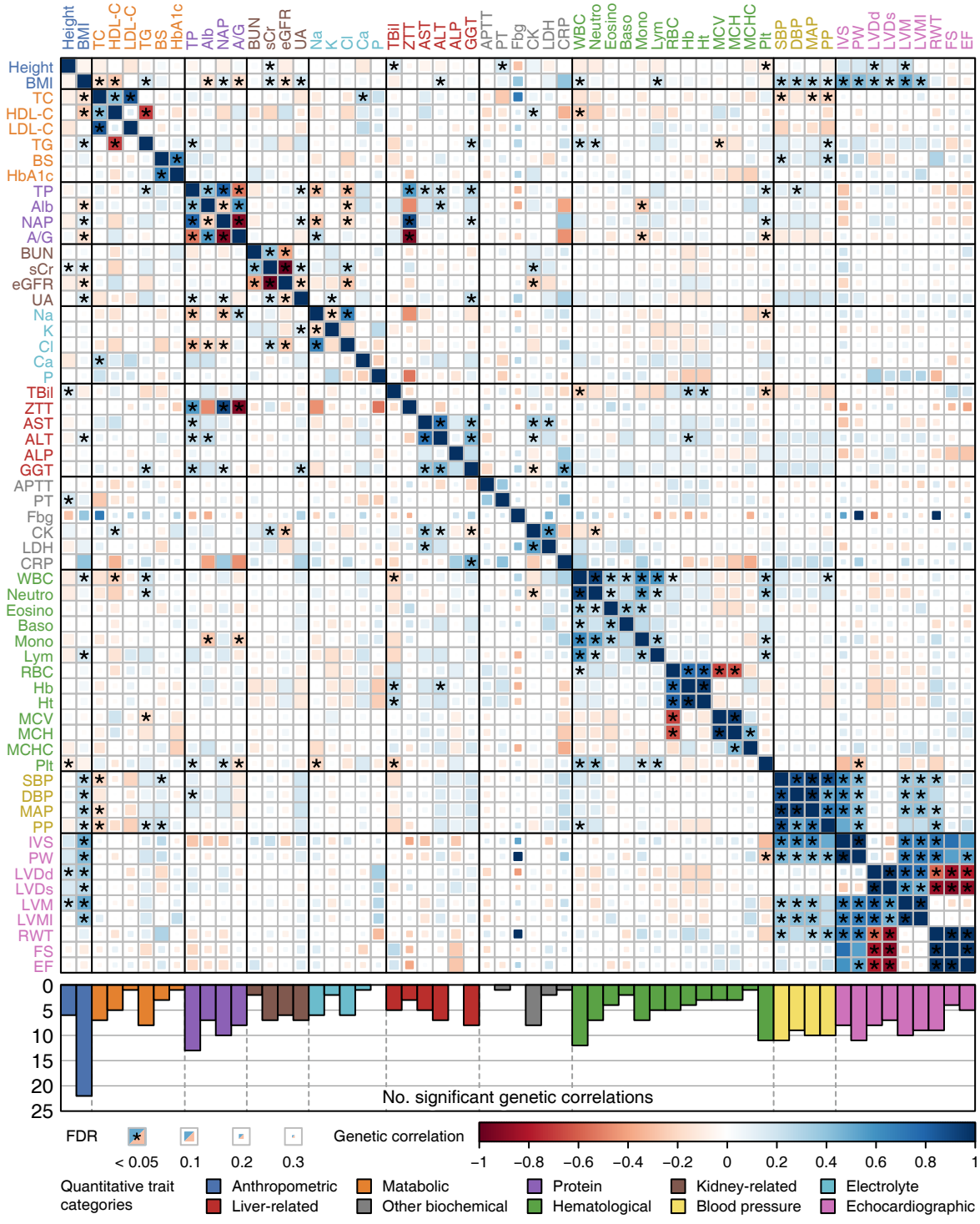
Supplementary Figure 3 | Alternative allele frequencies (AAF) of the 1,407 identified loci in East Asians and Europeans. a, A scatter plot of AAF (main) with histograms showing AAF spectra in East Asians (top) and in Europeans (right). Correlation was measured by Pearson's correlation. Known loci are colored in red, while novel loci are in blue. **b**, Boxplots of AAF with jitter for the 1,407 identified loci, grouped by quantitative trait category, population, and novelty.



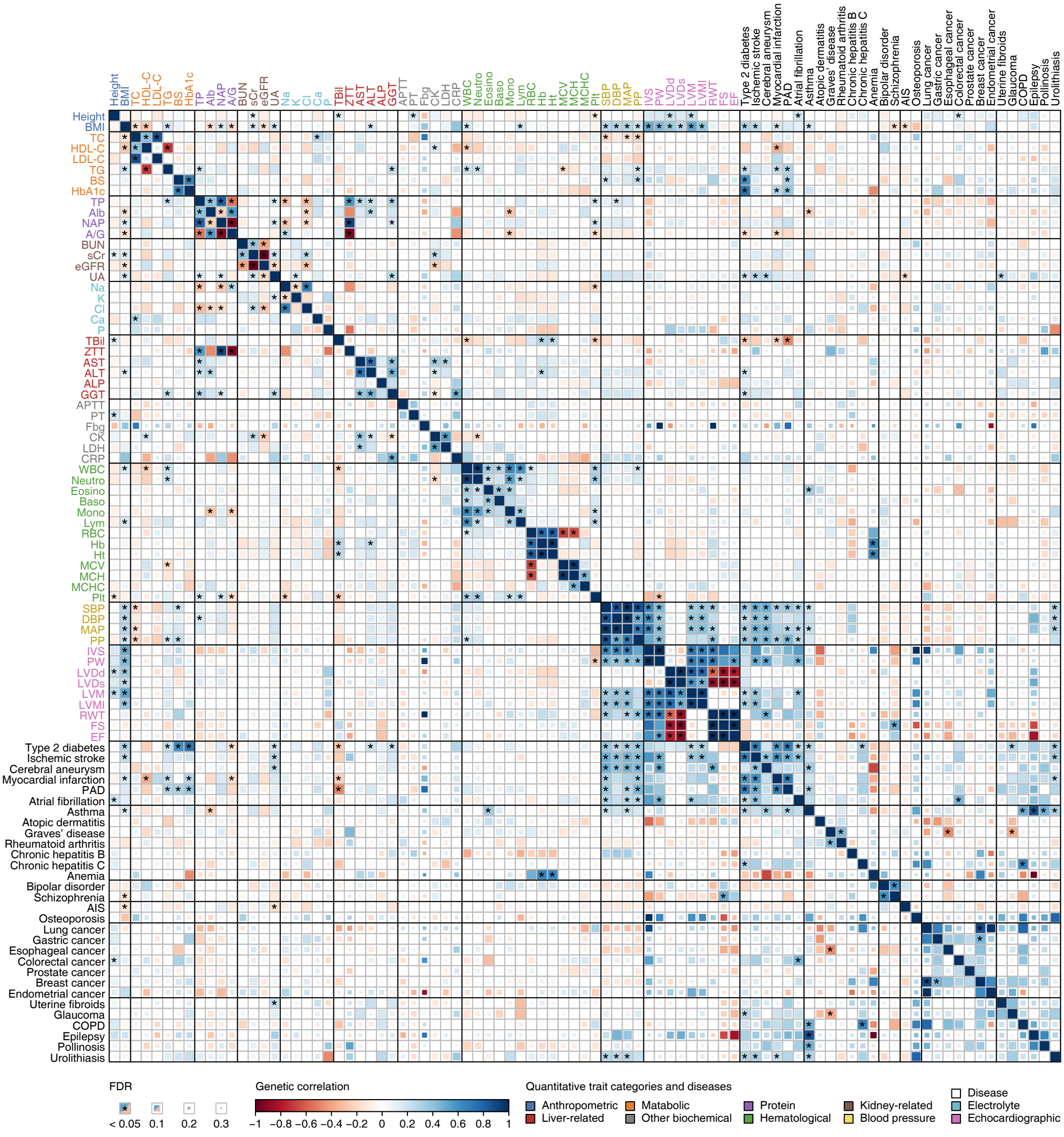
Supplementary Figure 4 | Abundant pleiotropic loci. Each regional plot shows associations of each locus for the most significantly associated trait. Middle rugs represent significant variants ($P < 5.0 \times 10^{-8}$) for the other quantitative traits. **a**, *ALDH2* (12q24.12) for Gamma-glutamyl transferase (GGT, $n = 118{,}309$). **b**, *GCKR* (2p23.3) for triglyceride (TG, $n = 105{,}597$). **c**, *ABO* (9q34.2) for alkaline phosphatase (ALP, $n = 105{,}030$). **d**, *RGS12* (4p16.3) for albumin (Alb, $n = 102{,}223$). GWAS was conducted using a linear regression model under the assumption of additive allelic effects of the SNP dosages.



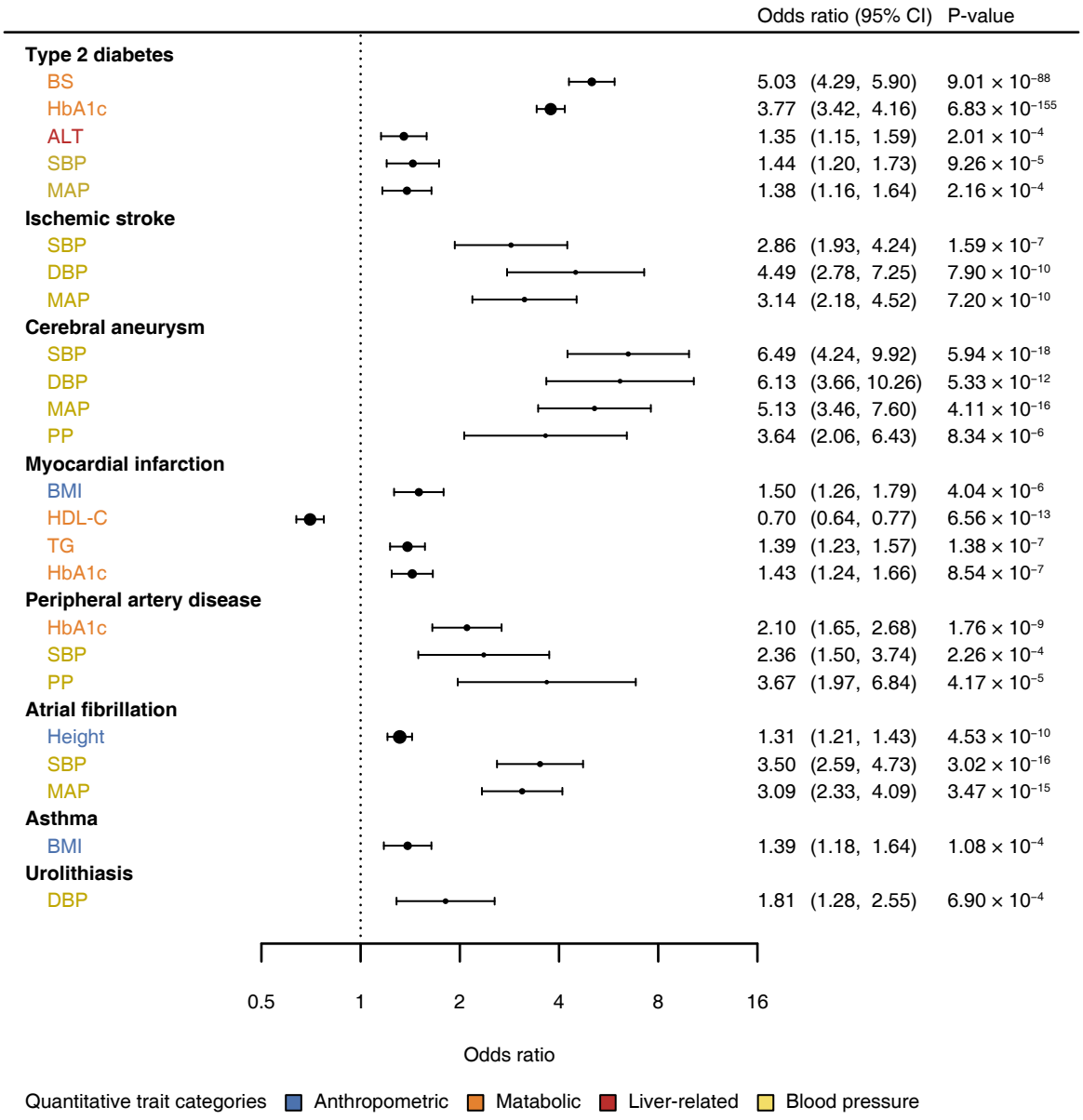
Supplementary Figure 5 | Genetic correlations across the 59 quantitative traits. Pairwise genetic correlations (n = 3,916) were estimated via bivariate LD score regression. Only correlations among the 59 quantitative traits (n = 1,711) are shown (see **Supplementary Fig. 6** for the full results). Positive genetic correlations are colored in blue, while negative correlations are in red. FDR was calculated using the Benjamini–Hochberg method to correct for multiple testing of all 3,916 correlations. Larger squares correspond to more significant FDR. Significant correlations (FDR < 0.05) are shown as full-sized squares with an asterisk. Bottom bars show the number of significant genetic correlations for each quantitative trait.



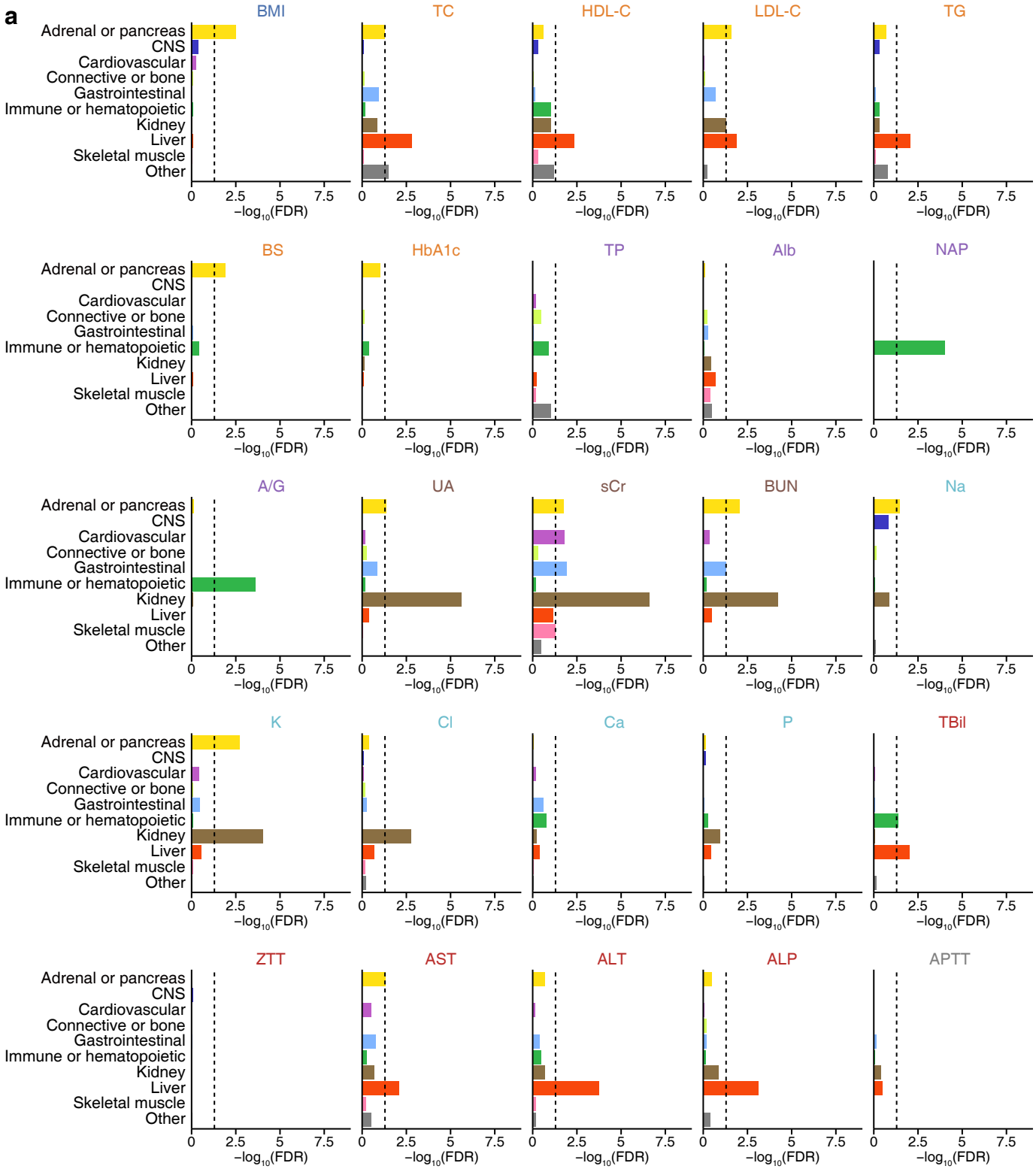
Supplementary Figure 6 | All genetic correlations across the 59 quantitative traits and 30 diseases. This figure integrates **Fig. 2** and **Supplementary Fig. 5**, providing a full genetic correlation matrix of the present study. Pairwise genetic correlations (n = 3,916) were estimated via bivariate LD score regression. FDR was calculated using the Benjamini–Hochberg method to correct for multiple testing of all 3,916 correlations. Positive genetic correlations are colored in blue, while negative correlations are in red. Larger squares correspond to more significant FDR. Significant correlations (FDR < 0.05) are shown as full-sized squares with an asterisk. PAD, peripheral artery disease. AIS, adolescent idiopathic scoliosis. COPD, chronic obstructive pulmonary disease.

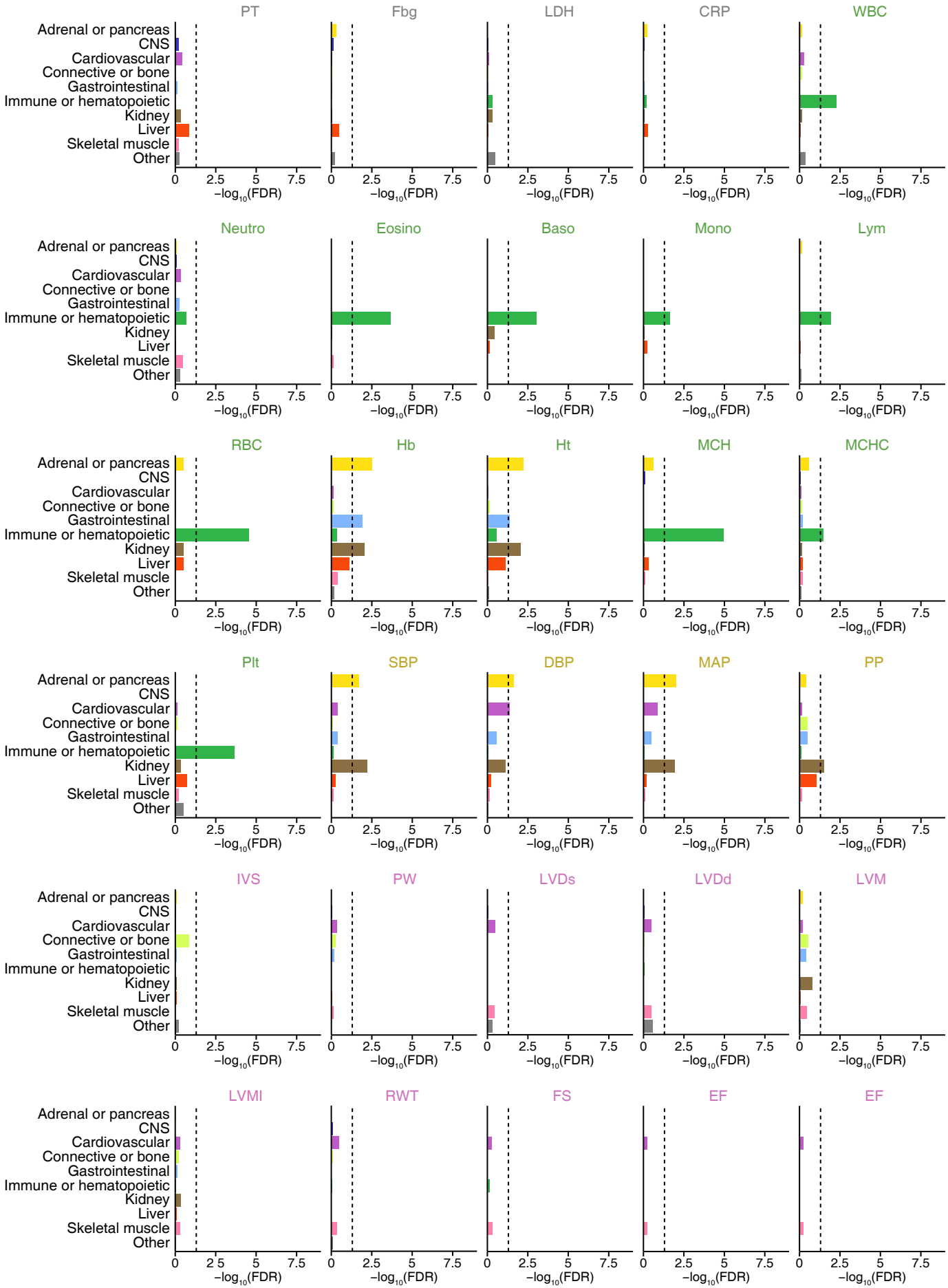


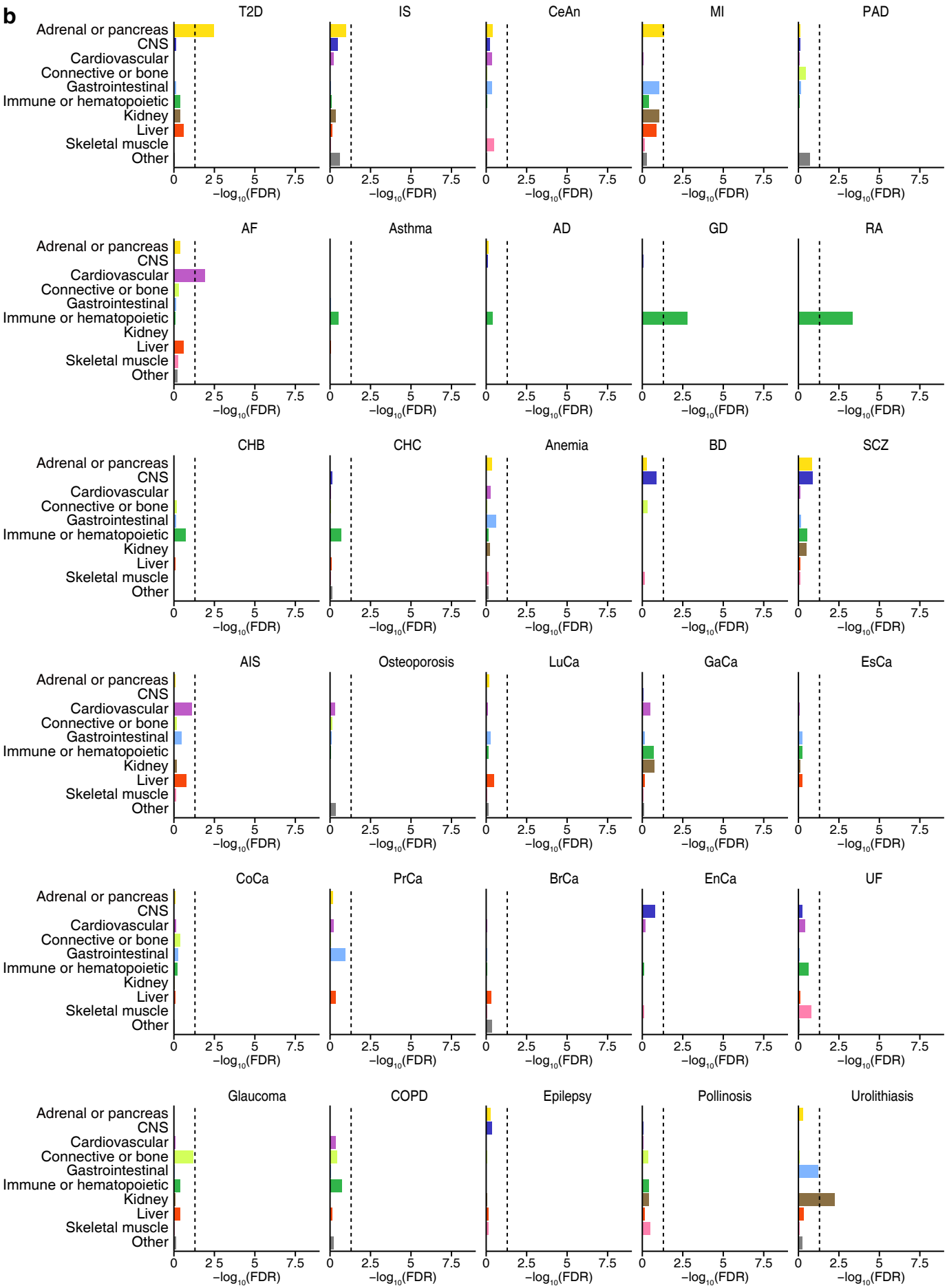
Supplementary Figure 7 | Significant causal associations with complex diseases. Forest plots showing the results of Mendelian Randomization analysis. Each dot represents the estimated odds ratio of disease risk, and its size is inversely proportional to standard error. The errorbar corresponds to the 95% confidence intervals.



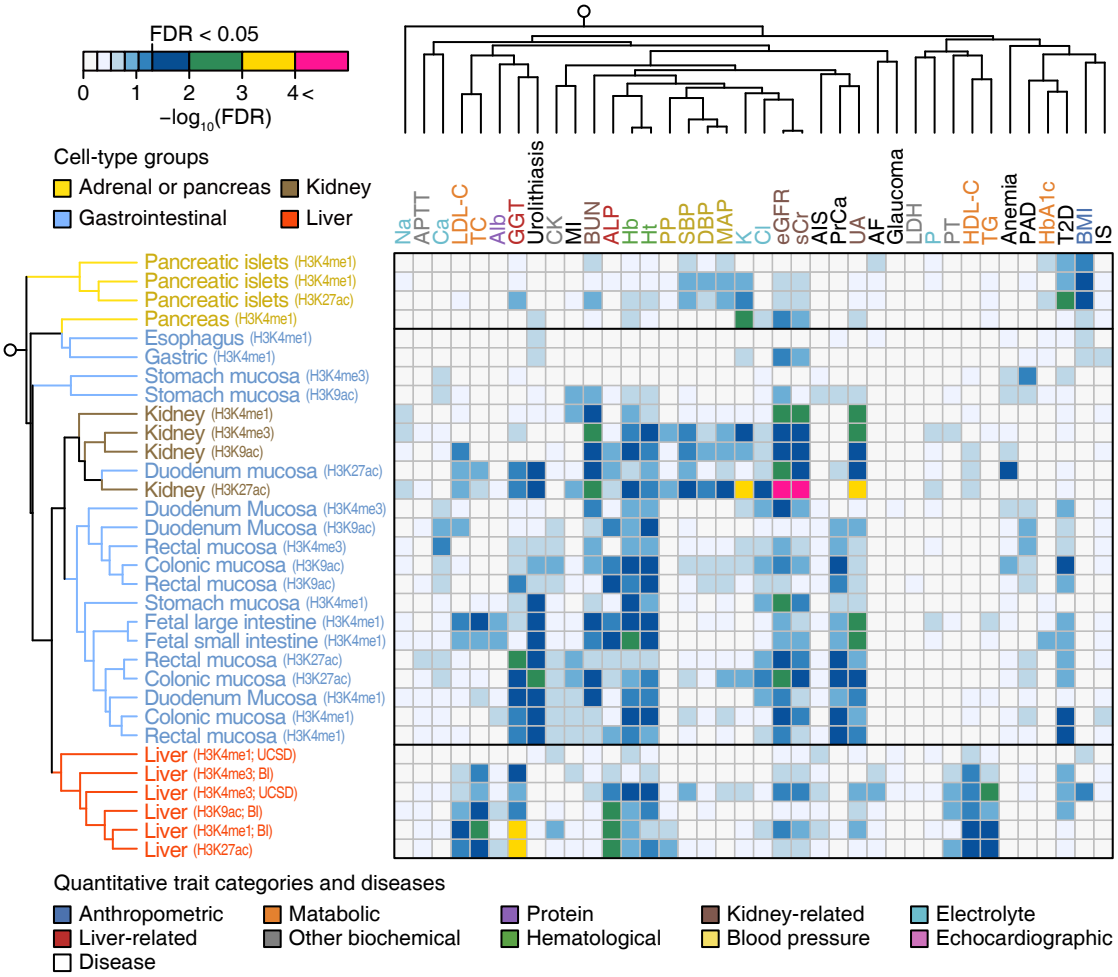
Supplementary Figure 8 | Additional results of heritability enrichment in the 10 cell-type groups. This figure supplements **Fig. 4b**. **a**, Significance of heritability enrichment for the remaining 54 quantitative traits and **b**, the 30 diseases estimated via stratified LD score regression. FDR was calculated using the Benjamini–Hochberg method. Dot line represents FDR = 0.05.




PT
Fbg
LDH
CRP
WBC
Neutro
Eosino
Baso
Mono
Lym
RBC
Hb
Ht
MCH
MCHC
Plt
SBP
DBP
MAP
PP
IVS
PW
LVDs
LVDd
LVM
LVMI
RWT
FS
EF
EF
Adrenal or pancreas
CNS
Cardiovascular
Connective or bone
Gastrointestinal
Immune or hematopoietic
Kidney
Liver
Skeletal muscle
Other
0
2.5
5
7.5
−log10(FDR)


b
T2D
IS
CeAn
MI
PAD
AF
Asthma
AD
GD
RA
CHB
CHC
Anemia
BD
SCZ
AIS
Osteoporosis
LuCa
GaCa
EsCa
CoCa
PrCa
BrCa
EnCa
UF
Glaucoma
COPD
Epilepsy
Pollinosis
Urolithiasis
Adrenal or pancreas
CNS
Cardiovascular
Connective or bone
Gastrointestinal
Immune or hematopoietic
Kidney
Liver
Skeletal muscle
Other
0
2.5
5
7.5
−log10(FDR)

Supplementary Figure 9 | Selected cluster of gastointestinal, kidney, or liver cell-types and various relevant traits and diseases. This is another distinct cluster identified in hierarchical clustering of heritability enrichments in the 220 cell-type specific annotations estimated via stratified LD score regression (**Fig. 5**). FDR was calculated using the Benjamini–Hochberg method. T2D, type 2 diabetes. IS, ischemic stroke. MI, myocardial infarction. PAD, peripheral artery disease. AF, atrial fibrillation. AIS, adolescent idiopathic scoliosis. PrCa, Prostate cancer.



Supplementary Note

1. The BioBank Japan Project (BBJ)

The BioBank Japan Project (http://biobankjp.org/english/index.html) was started in 2003 as a leading project of the Ministry of Education, Culture, Sports, Science and Technology, Japan. This project constructed a large, patient-based biobank, referred to as the Biobank Japan (BBJ). The BBJ collected DNA, serum and clinical information from 199,982 patients with any of 47 target diseases between fiscal years of 2003 and 2007[1,2]. Patients were recruited from 66 hospitals of 12 medical institutes throughout Japan (Osaka Medical Center for Cancer and Cardiovascular Diseases, the Cancer Institute Hospital of Japanese Foundation for Cancer Research, Juntendo University, Tokyo Metropolitan Geriatric Hospital, Nippon Medical School, Nihon University School of Medicine, Iwate Medical University, Tokushukai Hospitals, Shiga University of Medical Science, Fukujuji Hospital, National Hospital Organization Osaka National Hospital, and Iizuka Hospital). All patients were diagnosed by physicians at the cooperating hospitals.

Clinical information was collected through interviews and reviews of medical records using a standard questionnaire. The common items for all patients addressed during interview included height, weight, systolic and diastolic blood pressure. The common items for all patients collected from medical record reviews included information about birth year, sex, drugs used within one month before the study enrollment, and routine laboratory examination data (including complete blood count, urinalysis and biochemical tests such as measurements of liver enzymes, kidney function, lipids, blood sugar and HbA1c). Echocardiogram information was a disease-specific item and collected only from patients with cardiovascular diseases (myocardial infarction, unstable angina, stable angina, arrhythmia, heart failure, and peripheral arterial disease), dyslipidemia or diabetes. Most common diseases registered at BBJ were dyslipidemia (43,812 participants, 21.9%) and diabetes (39,697 participants, 19.9%). Details of study design, sample collection and baseline clinical information were described elsewhere[1,2].

2. Additional GWAS results of two anthropometric traits in Japanese

We obtained two anthropometric GWAS summary statistics of body mass index (BMI)[3] and adult height from the ongoing studies in BBJ, which contain the association results for the imputed variants using the 1000 Genomes Project East Asian samples as reference. The same quality control criteria for preparing both phenotypes and genotypes were applied, as described elsewhere[3].

3. Additional case-control GWAS results of 30 complex diseases in Japanese

We obtained 30 case-control GWAS results of complex diseases from both published[4–9] and

unpublished studies in the Japanese population. We summarized all diseases used in the present study, number of samples, cases/controls description, and publication status in **Table 2**.

Most of GWAS cases were recruited from the BBJ except for bipolar disorder[8], schizophrenia (collected by M.I., N.I., and the advanced Collaborative Study of Mood Disorder [COSMO] team), and adolescent idiopathic scoliosis (AIS, collected by S.I. and collaborating hospitals)[7]. Anemia was defined as hemoglobin level of < 12 g/dL for females and < 13 g/dL for males among the BBJ participants irrespective of registered diseases. As controls for GWAS, we used genome-wide screening data of three population-based controls derived from the Tohoku Medical Megabank organization[10], the Japan Public Health Center–based Prospective study and the Japan Multi-Institutional Collaborative Cohort Study[11]. Details of GWAS data of three population-based cohorts were described in the previous report[9]. For the controls for GWASs of rheumatoid arthritis[5], Graves' disease[6], AIS[7], atopic dermatitis[4], bipolar disorder[8], and schizophrenia, we used combined genome-wide screening data of various diseases (referred to as BBJ disease-mix) obtained from BBJ GWAS database. Diseases used as controls in each case-control GWAS was described in the previous reports[4–8]. In all of the unpublished studies, we obtained GWAS summary statistics of each disease for the imputed variants using the 1000 Genomes Project East Asian samples as reference.

Reference

1. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27,** S2–S8 (2017).
2. Hirata, M. *et al.* Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases. *J. Epidemiol.* **27,** S9–S21 (2017).
3. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49,** 1458–1467 (2017).
4. Hirota, T. *et al.* Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.* **44,** 1222–1226 (2012).
5. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506,** 376–381 (2013).
6. Okada, Y. *et al.* Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nat. Genet.* **47,** 798–802 (2015).
7. Ogura, Y. *et al.* A Functional SNP in BNC2 Is Associated with Adolescent Idiopathic Scoliosis. *Am. J. Hum. Genet.* **97,** 337–342 (2015).
8. Ikeda, M. *et al.* A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol. Psychiatry* 1–9 (2017). doi:10.1038/mp.2016.259
9. Low, S.-K. *et al.* Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat. Genet.* **49,** 953–958 (2017).
10. Kuriyama, S. *et al.* The Tohoku Medical Megabank Project: Design and Mission. *J. Epidemiol.* **26,** 493–511 (2016).
11. Hamajima, N. & J-MICC Study Group. The Japan Multi-Institutional Collaborative Cohort Study (J-MICC Study) to detect gene-environment interactions for cancer. *Asian Pac. J. Cancer Prev.* **8,** 317–23 (2007).