

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

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

**Clinical measurements can be viewed as useful intermediate phenotypes to promote understanding of complex human diseases. To acquire comprehensive insights into the underlying genetics, here we conducted a genome-wide association study (GWAS) of 58 quantitative traits in 162,255 Japanese individuals. Overall, we identified 1,407 trait-associated loci ( $P < 5.0 \times 10^{-8}$ ), 679 of which were novel. By incorporating 32 additional GWAS results for complex diseases and traits in Japanese individuals, we further highlighted pleiotropy, genetic correlations, and cell-type specificity across quantitative traits and diseases, which substantially expands the current understanding of the associated genetics and biology. This study identified both shared polygenic effects and cell-type specificity, represented by the genetic links among clinical measurements, complex diseases, and relevant cell types. Our findings demonstrate that even without prior biological knowledge of cross-phenotype relationships, genetics corresponding to clinical measurements successfully recapture those measurements' relevance to diseases, and thus can contribute to the elucidation of unknown etiology and pathogenesis.**

Clinical laboratory measurements (e.g., blood test results) are powerful intermediate phenotypes that can be used to diagnose and monitor human diseases. Elucidation of the underlying genetics, as well as inference of genetic relationships to diseases and implicated cell types, can provide clues about disease biology. To this end, GWASs have been conducted to investigate various quantitative traits, including anthropometric<sup>1–3</sup>, metabolic<sup>4,5</sup>, kidney-related<sup>6,7</sup>, hematological<sup>8,9</sup>, and blood pressure traits<sup>10–12</sup>. The interplay between the genetics of quantitative traits and diseases has been assessed by several approaches, such as pleiotropy<sup>13,14</sup>, genetic correlation<sup>15,16</sup>, and Mendelian randomization<sup>17</sup>. For example, recent large-scale studies of body mass index (BMI), a key measure for assessing obesity, revealed shared genetic effects on metabolic traits and the involvement of the central nervous system<sup>2</sup> and immune cells<sup>3</sup> in obesity susceptibility. However, previous studies primarily examined subjects of European ancestry, and each study separately focused on few quantitative traits. For the creation of a comprehensive landscape, additional studies of non-European populations are warranted that simultaneously investigate a wide range of clinical measurements and extensively interrogate their relevance to complex diseases.

Here we report a GWAS of 58 quantitative traits in 162,255 Japanese individuals from the BioBank Japan Project (BBJ)<sup>18,19</sup>, one of the largest non-European single-descent biobanks with detailed phenotypes, to broaden the current knowledge and understanding

of the genetics and biology of these traits. Moreover, we incorporated additional GWASs of complex diseases and traits in Japanese subjects, and evaluated pleiotropy, genetic correlation, and cell-type specificity with respect to the quantitative traits. Our study provides many insights into the genetic basis of various quantitative traits and illuminates the complex genetic links among clinical measurements, complex diseases, and relevant cell types.

## Results

**Genome-wide association analysis of 58 quantitative traits.** We tested 5,961,600 autosomal variants and 147,353 X-chromosome variants (imputed with 1000 Genomes Project Phase 1<sup>20</sup>; Methods) for association with 58 quantitative traits in 162,255 Japanese individuals. The studied traits covered a wide range of clinical measurements, grouped into nine distinct categories (Table 1): metabolic ( $n=6$ ), serum protein ( $n=4$ ), kidney-related ( $n=4$ ), electrolyte ( $n=5$ ), liver-related ( $n=6$ ), other biochemical ( $n=6$ ), hematological ( $n=13$ ), blood pressure ( $n=4$ ), and echocardiographic ( $n=9$ ). The study design is illustrated in Supplementary Fig. 1, and the detailed characteristics of the subjects, phenotype source, and exclusion criteria are described in Supplementary Tables 1 and 2.

Overall, we identified 1,407 trait-associated loci for 53 quantitative traits that satisfied a genome-wide significance threshold of  $P=5.0 \times 10^{-8}$  (Methods). Of these, 679 loci were novel (Table 1 and Supplementary Table 3). When we applied multiple-testing

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

**Table 1 | Overview of the studied quantitative traits**

Category	Trait	Abbreviation	<i>n</i>	Number of loci	(New loci)
Anthropometric <sup>a</sup>	Adult height	Height	159,148	-	-
	Body mass index	BMI	158,284	-	-
Metabolic	Total cholesterol	TC	128,305	32	(5)
	High-density-lipoprotein cholesterol	HDL-C	70,657	31	(9)
	Low-density-lipoprotein cholesterol	LDL-C	72,866	22	(6)
	Triglyceride	TG	105,597	26	(4)
	Blood sugar	BS	93,146	17	(3)
	Hemoglobin A1c	HbA1c	42,790	26	(16)
Protein	Total protein	TP	113,509	33	(31)
	Albumin	Alb	102,223	17	(13)
	Non-albumin protein	NAP	98,538	48	(46)
	Albumin/globulin ratio	A/G	98,626	42	(40)
Kidney-related	Blood urea nitrogen	BUN	139,818	43	(30)
	Serum creatinine	sCr	142,097	66	(51)
	Estimated glomerular filtration rate	eGFR	143,658	69	(48)
	Uric acid	UA	109,029	27	(10)
Electrolyte	Sodium	Na	127,304	14	(14)
	Potassium	K	132,938	14	(14)
	Chloride	Cl	126,402	19	(19)
	Calcium	Ca	71,701	15	(11)
	Phosphorus	P	42,793	8	(5)
	Liver-related	Total bilirubin	TBil	110,207	11
Zinc sulfate turbidity test		ZTT	12,303	4	(4)
Aspartate aminotransferase		AST	134,154	25	(22)
Alanine aminotransferase		ALT	134,182	27	(23)
Alkaline phosphatase		ALP	105,030	28	(19)
γ-glutamyl transferase		GGT	118,309	42	(24)
Other biochemical	Activated partial thromboplastin time	APTT	37,767	10	(5)
	Prothrombin time	PT	58,110	9	(6)
	Fibrinogen	Fbg	18,348	1	(0)
	Creatine kinase	CK	106,080	36	(31)
	Lactate dehydrogenase	LDH	126,319	10	(5)
	C-reactive protein	CRP	75,391	7	(1)
Hematological	White blood cell count	WBC	107,964	36	(7)
	Neutrophil count	Neutro	62,076	21	(3)
	Eosinophil count	Eosino	62,076	18	(1)
	Basophil count	Baso	62,076	26	(10)
	Monocyte count	Mono	62,076	31	(2)
	Lymphocyte count	Lym	62,076	13	(1)
	Red blood cell count	RBC	108,794	54	(11)
	Hemoglobin	Hb	108,769	19	(5)
	Hematocrit	Ht	108,757	22	(8)
	Mean corpuscular volume	MCV	108,256	91	(17)
	Mean corpuscular hemoglobin	MCH	108,054	77	(10)
	Mean corpuscular hemoglobin concentration	MCHC	108,728	39	(10)
	Platelet count	Plt	108,208	78	(19)
Blood pressure	Systolic blood pressure	SBP	136,597	23	(4)
	Diastolic blood pressure	DBP	136,615	18	(3)
	Mean arterial pressure	MAP	136,482	30	(21)
	Pulse pressure	PP	136,249	17	(8)

Continued

**Table 1 | Overview of the studied quantitative traits (continued).**

Category	Trait	Abbreviation	<i>n</i>	Number of loci	(New loci)
Echocardiographic	Interventricular septum thickness	IVS	19,318	0	(0)
	Posterior wall thickness	PW	19,373	0	(0)
	Left ventricular internal dimension in diastole	LVDd	19,676	3	(3)
	Left ventricular internal dimension in systole	LVDs	19,586	4	(4)
	Left ventricular mass	LVM	19,076	0	(0)
	Left ventricular mass index	LVMi	17,837	0	(0)
	Relative wall thickness	RWT	19,311	1	(1)
	Fractional shortening	FS	19,580	4	(4)
	Ejection fraction	EF	19,516	3	(3)
	E/A ratio	E/A	8,600	0	(0)

<sup>a</sup>We obtained the GWAS results for these two anthropometric traits from the ongoing studies in the BioBank Japan Project.

correction to the number of the studied traits, 943 trait-associated loci for 51 traits showed significant associations ( $P < 5.0 \times 10^{-8}/58 = 8.6 \times 10^{-10}$ ), of which 372 loci were novel. Stepwise conditional analysis for each trait-associated locus further identified 267 additional independent signals at 158 trait-associated loci for 39 traits (Supplementary Table 4). We observed multiple additional independent signals at 49 trait-associated loci, with the maximum number of 11 independent signals at 11q13.1 for uric acid (the top associated signal was rs57633992 on *NRXN2*;  $P = 7.30 \times 10^{-845}$ ) (Supplementary Fig. 2). Although the genomic inflation factors ( $\lambda_{GC}$ ) of several traits showed considerable inflation (mean  $\lambda_{GC} = 1.11$ ), linkage disequilibrium (LD) score regression<sup>21</sup> analysis confirmed no existence of substantial confounding biases for all traits (mean intercept: 1.04), as shown in Supplementary Table 5. Given the substantial sample sizes in our GWASs, these statistics suggest that a majority of the inflation was due to polygenic effects, and population stratification and other potential biases were strictly controlled<sup>3,21</sup>. Manhattan, quantile–quantile, and LD score plots are provided in Supplementary Dataset 1. Detailed regional plots for each locus are provided in Supplementary Dataset 2.

Trans-ethnic comparison of the allele frequencies of the identified loci between East Asians and Europeans showed an overall shared allelic spectrum across populations ( $r = 0.687$ ; Supplementary Fig. 3). The novel loci tended to have higher allele frequencies in East Asians than in Europeans, as 60 novel loci (8.8%) were common (minor allele frequency  $\geq 5\%$ ) in East Asians but rare ( $\leq 1\%$ ) in Europeans. Of note, the associated single-nucleotide polymorphisms (SNPs) in 15 unique loci (for example, *ALDH2*, *EGF*, and *SUFU*) were monomorphic in Europeans but had higher frequencies in East Asians ( $\geq 10\%$ ). These observations show the contribution of population-specific factors, such as evolutionary selection pressure, to the identified loci. The percentage of mean heritability of the traits explained by the significant loci was 2.84% (Supplementary Table 6). On average, the known loci in Europeans explained 1.92%, the overall known loci explained 2.03%, and the newly identified loci explained 0.84%. The percentage of global heritability explained by the genome-wide common SNPs was on average 8.60%, which is comparable to previous findings in Europeans (Supplementary Table 5).

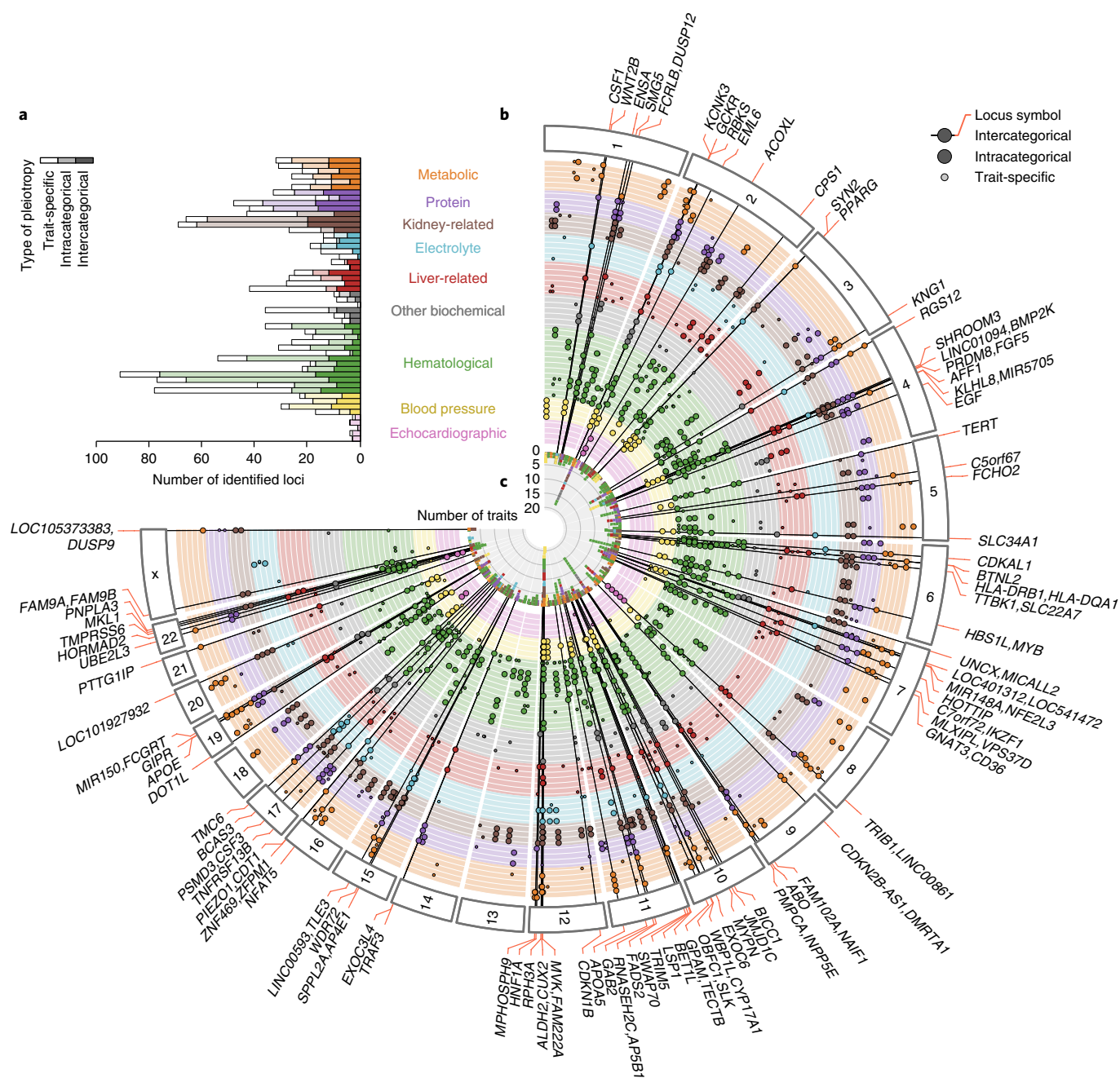
**Pleiotropy of top associated quantitative trait loci.** Pleiotropy, defined here as the sharing of risk alleles across multiple traits, is a key concept in investigations of cross-phenotype relationships across human traits, leading us to decipher a shared genetic etiology underlying a complex genetic architecture<sup>13,14</sup>. To identify major pleiotropic loci, we assessed pleiotropy at the single-locus level across 763 unique loci (derived from the 1,407 trait-associated loci for 53 quantitative traits mentioned above; Methods). We

identified numerous pleiotropic loci among the quantitative traits ( $n = 313$ ), representing approximately 41% of the unique loci (Fig. 1 and Supplementary Table 7). Of these, 88 loci showed pleiotropy across traits in multiple trait categories (intercategorical pleiotropy), whereas the other 225 loci showed pleiotropy across traits in a single category (intracategorical pleiotropy).

We observed the most abundant intercategorical pleiotropy at *ALDH2* (12q24.12), associated with 21 traits in seven categories (Supplementary Fig. 4). The most significant associations were at rs79105258 (the top associated signal was  $\gamma$ -glutamyl transferase (GGT);  $P = 9.98 \times 10^{-100}$ ), which shows high minor allele frequency in East Asians (0.24) but is monomorphic in other ancestral populations<sup>20</sup>. Other pleiotropic loci that showed intercategorical pleiotropy included *GCKR* (2p23.3), associated with 18 traits in seven categories (rs1260326 for triglyceride;  $P = 1.69 \times 10^{-94}$ ); *ABO* (9q34.2), associated with 15 traits in six categories (rs2519093 for alkaline phosphatase;  $P = 2.02 \times 10^{-887}$ ); and *RGS12* (4p16.3), associated with nine traits in six categories (rs4690095 for albumin;  $P = 1.63 \times 10^{-22}$ ). Although *RGS12* (4p16.3) has received little attention as a pleiotropic locus compared with the other loci mentioned<sup>13</sup>, this locus has shown associations with several traits and diseases, including serum lipids<sup>4</sup> and inflammatory bowel disease<sup>22</sup>. Our results expand its associations with additional traits, including kidney function, serum calcium, GGT, and platelet count (Plt).

**Polygenic correlations across quantitative traits.** Another approach to infer genetic overlap across traits is to estimate a genetic correlation, that is, a correlation of causal effect sizes at a genome-wide level<sup>15,16</sup>. Rather than using a single-locus-level analysis, we evaluated genetic correlations under a polygenic model that could take into account the consistency of effect directions, unlike pleiotropy analysis, to disentangle the polygenic architecture of the studied traits. We incorporated additional GWAS results for the anthropometric traits BMI<sup>3</sup> and adult height, obtained from ongoing studies under the BBJ (Supplementary Note 1), to gain a broader perspective on quantitative traits. We carried out bivariate LD score regression<sup>15</sup> to estimate pairwise genetic correlations across the 59 quantitative traits (we excluded the E/A ratio, a marker of heart function, owing to small sample size; Methods). We found 173 significant genetic correlations (false discovery rate (FDR)  $< 0.05$ ), 100 (58%) of which were intercategorical (Supplementary Fig. 5 and Supplementary Table 8).

We observed the greatest number of significant intercategorical genetic correlations with BMI, which showed significant correlations with 22 quantitative traits in seven trait categories (the most significant correlation ( $P = 9.83 \times 10^{-17}$ ) was with mean arterial



**Fig. 1 | Overview of the identified loci and their pleiotropy.** **a**, The number of identified trait-associated loci for each trait, grouped by type of pleiotropy. **b**, Identified loci and their pleiotropy. Each dot corresponds to a trait-associated locus. Larger dots highlight a pleiotropic association. Each radial line connects all dots for an intercategory pleiotropic locus with a locus symbol. **c**, The number of associated traits for each intercategory pleiotropic locus. The orders of the quantitative traits in each trait category are the same as those presented in Table 1.

pressure). Total protein and height had the second highest numbers of correlated categories ( $n = 6$ ), followed by triglycerides, non-albumin protein, and Plt ( $n = 5$ ). Although some of the significant intercategory genetic correlations had been suggested previously (for example, BMI and serum lipids in Europeans<sup>15</sup>), most were newly identified. Notably, most of these links were consistent with observations in epidemiological studies, thus demonstrating the robustness and potential of the genetics-based studies to elucidate novel biological and medical architectures of human traits without prior knowledge (Supplementary Table 8). For example, the observed negative correlation between white blood cell (WBC) count and total bilirubin was suggested in an epidemiological study<sup>23</sup>, but our

study corroborated this correlation on the basis of genetics, thus providing empirical support for the hypothesis of the anti-inflammatory activities of bilirubin<sup>24</sup>.

**Genetic correlations among quantitative traits and diseases.**

Given that clinical measurements are informative as intermediate phenotypes for the assessment of complex human diseases, we reasoned that additional exploration of genetic correlations between quantitative traits and diseases would provide more empirical corroboration of shared genetic architecture, which could illuminate the underlying etiology and pathogenesis. To this end, we additionally incorporated 30 case-control GWAS results for complex

**Table 2 | Summary of the additional case-control GWASs of the 30 complex diseases**

Disease			Number of samples		Case recruitment	Control <sup>a</sup>	PMID
Category	Trait	Abbreviation	Case	Control			
Metabolic disease	Type 2 diabetes	T2D	36,832	28,870	BBJ	Cohort	-
Cardiovascular disease	Ischemic stroke	IS	16,256	27,294	BBJ	Cohort	-
	Cerebral aneurysm	CeAn	2,597	28,870	BBJ	Cohort	-
	Myocardial infarction	MI	12,494	28,870	BBJ	Cohort	-
	Peripheral artery disease	PAD	3,382	28,870	BBJ	Cohort	-
	Atrial fibrillation	AF	8,180	28,612	BBJ	Cohort	28416822
	Allergic disease	Asthma	-	7,570	28,870	BBJ	Cohort
Autoimmune disease	Atopic dermatitis	AD	1,472	7,966	BBJ	Disease-mix	23042114
	Graves' disease	GD	1,961	7,968	BBJ	Disease-mix	26029868
Infectious disease	Rheumatoid arthritis	RA	4,873	17,642	Meta-analysis (Asian)		24390342
	Chronic hepatitis B	CHB	1,092	28,870	BBJ	Cohort	-
	Chronic hepatitis C	CHC	4,988	28,870	BBJ	Cohort	-
Hematologic disease	Anemia	-	7,570	28,870	BBJ	Cohort	-
Psychiatric disorder	Bipolar disorder	BD	2,964	61,887	Collaborators	Disease-mix	28115744
	Schizophrenia	SCZ	1,987	9,788	Collaborators	Disease-mix	-
Musculoskeletal disease	Adolescent idiopathic scoliosis	AIS	2,109	11,140	Collaborators	Case-mix	26211971
	Osteoporosis	-	7,099	28,870	BBJ	Cohort	-
Cancer	Lung cancer	LuCa	3,874	27,178	BBJ	Cohort	-
	Gastric cancer	GaCa	6,171	27,178	BBJ	Cohort	-
	Esophageal cancer	EsCa	1,225	27,178	BBJ	Cohort	-
	Colorectal cancer	CoCa	6,692	27,178	BBJ	Cohort	-
	Prostate cancer	PrCa	5,088	10,682	BBJ	Cohort (male)	-
	Breast cancer	BrCa	5,272	16,496	BBJ	Cohort (female)	-
	Endometrial cancer	EnCa	1,931	17,492	BBJ	Cohort (female)	-
Other	Uterine fibroids	UF	5,720	17,492	BBJ	Cohort (female)	-
	Glaucoma	-	3,980	18,815	BBJ	Cohort	-
	Chronic obstructive pulmonary disease	COPD	3,050	28,870	BBJ	Cohort	-
	Epilepsy	-	2,109	28,870	BBJ	Cohort	-
	Pollinosis	-	6,246	28,870	BBJ	Cohort	-
	Urolithiasis	-	5,536	28,870	BBJ	Cohort	-

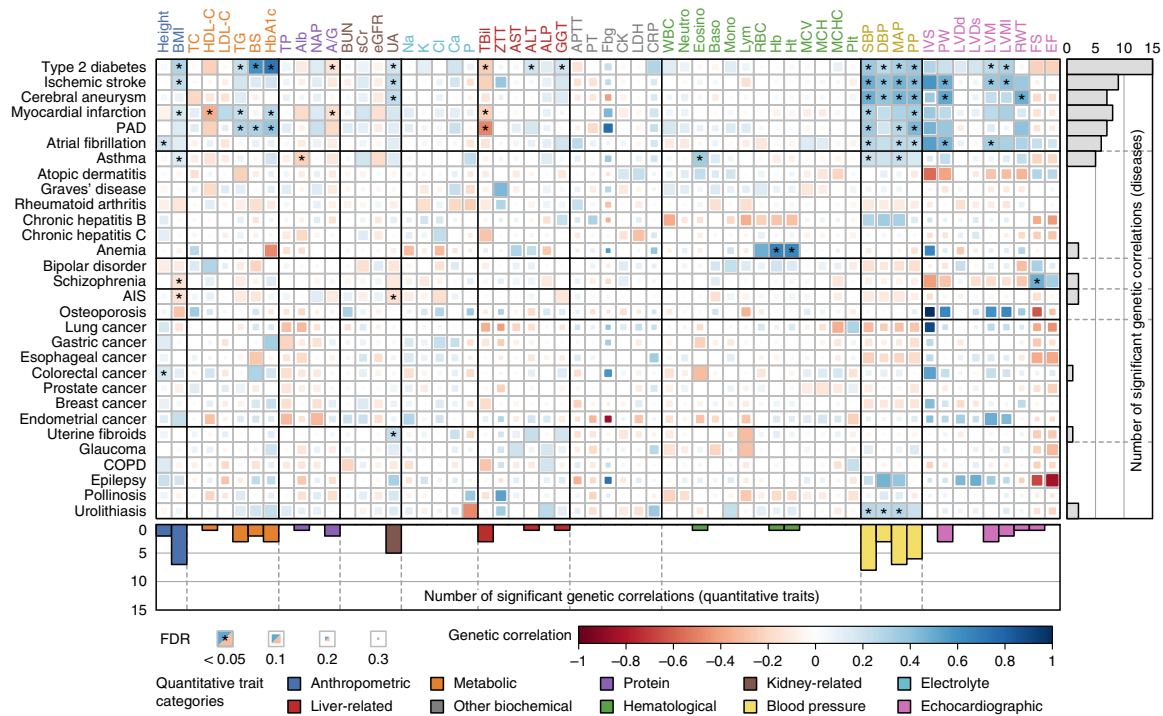
<sup>a</sup>"Cohort" refers to the population-based cohorts in Japan. "Disease-mix" refers to the mixture of the affected subjects obtained from BBJ.

diseases in Japanese individuals (Table 2 and Supplementary Note 1)<sup>25–30</sup>, including cardiometabolic ( $n=6$ ), immune-related ( $n=6$ ), hematologic ( $n=1$ ), psychiatric ( $n=2$ ), and musculoskeletal diseases ( $n=2$ ); cancer ( $n=7$ ); and other diseases ( $n=6$ ).

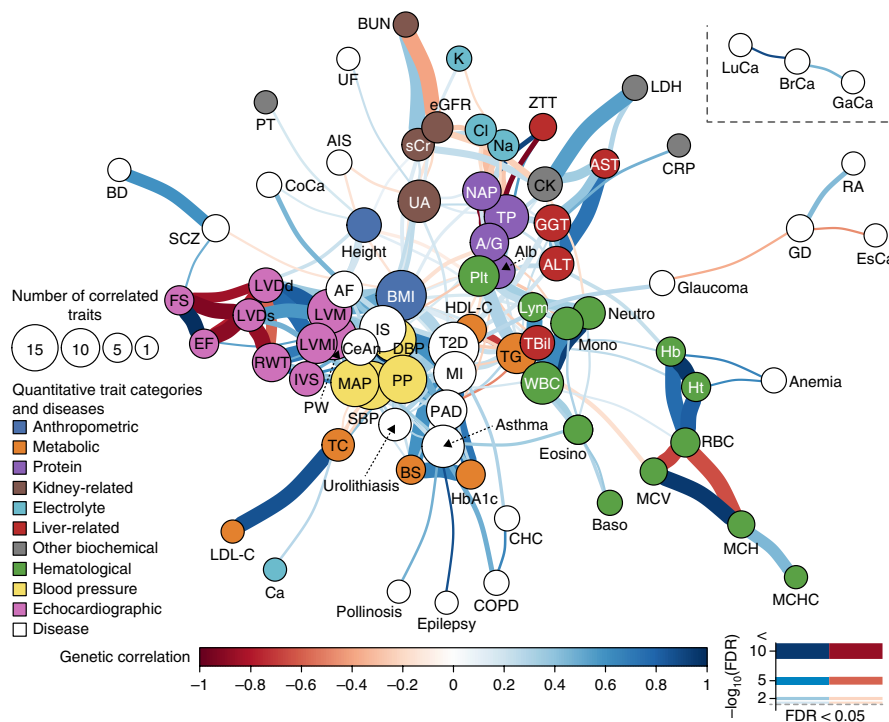
We then estimated pairwise genetic correlations across the 59 quantitative traits and 30 diseases. We identified 68 significant genetic correlations (FDR < 0.05), which supported the biological relevance of associations between clinical measurements and complex diseases (Fig. 2 and Supplementary Table 8; the full results are presented in Supplementary Fig. 6 and Supplementary Table 9). Among the 68 significant correlations, 52 (76.5%) involved cardiometabolic diseases, correlating with quantitative traits in seven categories. Indeed, type 2 diabetes showed the greatest number of significant correlations with quantitative traits ( $n=15$ ), and demonstrated the most significant genetic correlation with hemoglobin A1c ( $r_g=0.724$ ;  $P=2.54 \times 10^{-22}$ ). We also observed other significant correlations, such as those between ischemic stroke and uric acid ( $r_g=0.254$ ;  $P=5.74 \times 10^{-5}$ ), and between myocardial infarction and

albumin/globulin ratio ( $r_g=-0.174$ ;  $P=1.06 \times 10^{-3}$ ). Among the remaining 16 significant genetic correlations (other than for cardiometabolic diseases), the most significant correlation was between asthma and eosinophil count ( $r_g=0.348$ ;  $P=3.76 \times 10^{-4}$ ). Other significant correlations included those between urolithiasis and systolic blood pressure ( $r_g=0.272$ ;  $P=7.22 \times 10^{-4}$ ), asthma and systolic blood pressure ( $r_g=0.214$ ;  $P=8.84 \times 10^{-4}$ ), and colorectal cancer and height ( $r_g=0.164$ ;  $P=2.92 \times 10^{-3}$ ).

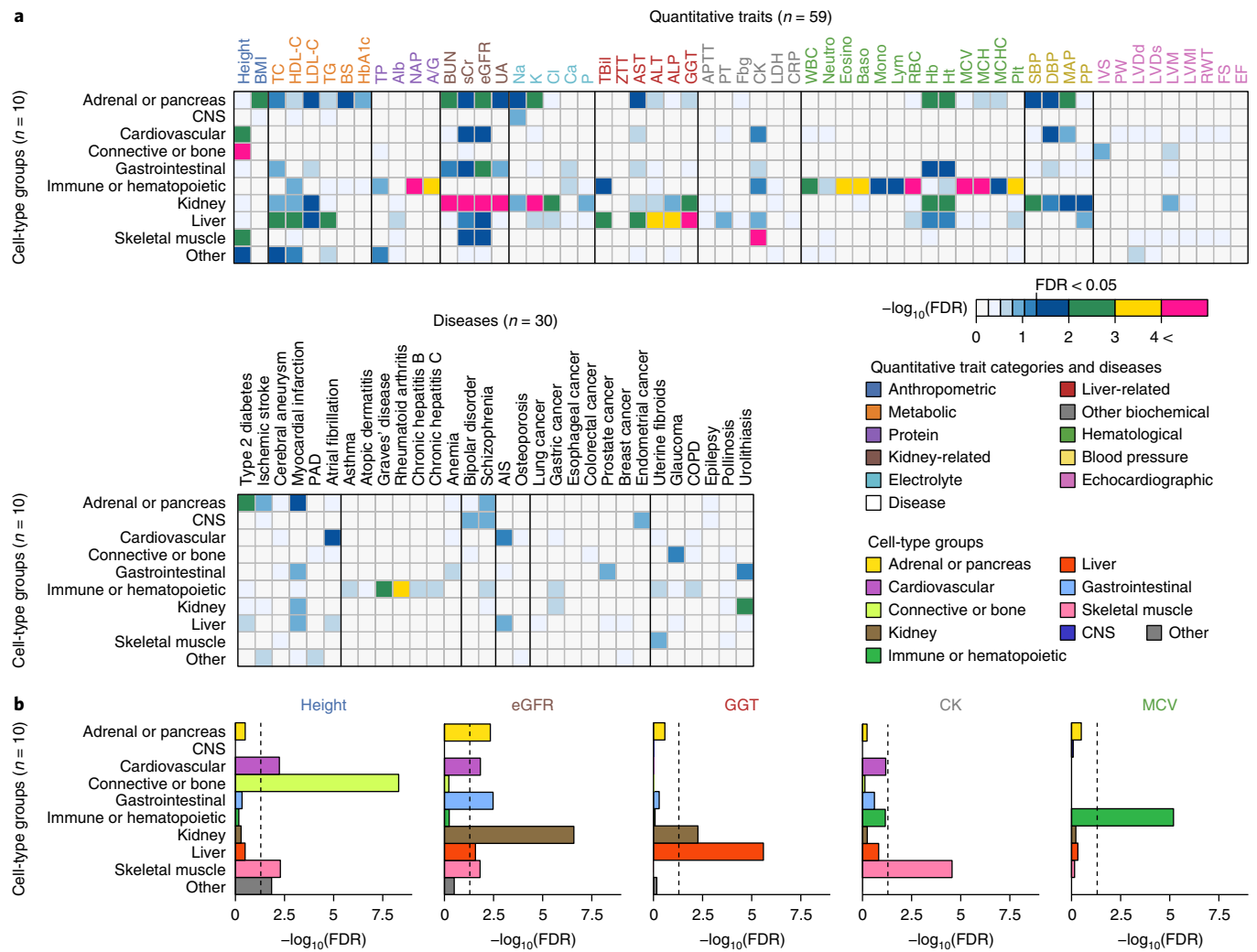
In addition to the suggested genetic correlations in Europeans (type 2 diabetes and BMI; triglycerides, blood sugar, and hemoglobin A1c; coronary artery disease and BMI; and high-density-lipoprotein cholesterol and triglycerides)<sup>15</sup>, we empirically corroborated novel genetic correlations that have been implicated in Mendelian randomization analyses (e.g., type 2 diabetes and alanine aminotransferase<sup>31</sup>; atrial fibrillation and height<sup>32</sup>; asthma and eosinophil count<sup>9</sup>; and colorectal cancer and height<sup>33</sup>) and epidemiological studies (e.g., ischemic stroke and uric acid<sup>34</sup>; myocardial infarction and albumin/globulin ratio<sup>35</sup>; peripheral artery disease and total



**Fig. 2 | Genetic correlations between the 59 quantitative traits and 30 diseases.** Pairwise genetic correlations ( $n = 3,916$ ) were estimated via bivariate LD score regression. Only correlations between the 59 quantitative traits and 30 diseases ( $n = 1,770$ ) are shown (full results are shown in Supplementary Fig. 6). Positive genetic correlations are shown in blue, and negative correlations are in red. FDR was calculated via the Benjamini-Hochberg method to correct for multiple testing of all 3,916 correlations. Larger squares correspond to more significant FDRs. Significant correlations ( $FDR < 0.05$ ) are indicated by asterisks. The bars at the right and bottom of the figure show the number of significant genetic correlations per disease and quantitative trait, respectively. Abbreviations are defined as in Tables 1 and 2.



**Fig. 3 | Genetic correlation network across the 59 quantitative traits and 30 diseases.** Genetically correlated traits and diseases are clustered close together. Each circle represents a trait, and each edge represents a significant genetic correlation ( $FDR < 0.05$ ). Pairwise genetic correlations ( $n = 3,916$ ) were estimated via bivariate LD score regression, and their  $P$  values were corrected via the Benjamini-Hochberg method for multiple testing of all 3,916 correlations. Positive and negative genetic correlations are indicated by color according to the key. Thicker edges correspond to more significant FDRs. The dashed line separates unconnected plots. Definitions of trait abbreviations can be found in Tables 1 and 2.



**Fig. 4 | Heritability enrichment in the ten cell-type groups.** **a**, Significance of heritability enrichment in the ten cell-type groups for the 59 quantitative traits and 30 diseases estimated via stratified LD score regression. **b**, Selected quantitative traits in five quantitative trait categories. FDR was calculated via the Benjamini–Hochberg method. Vertical dashed lines indicate FDR = 0.05. Definitions for trait and disease abbreviations can be found in Tables 1 and 2.

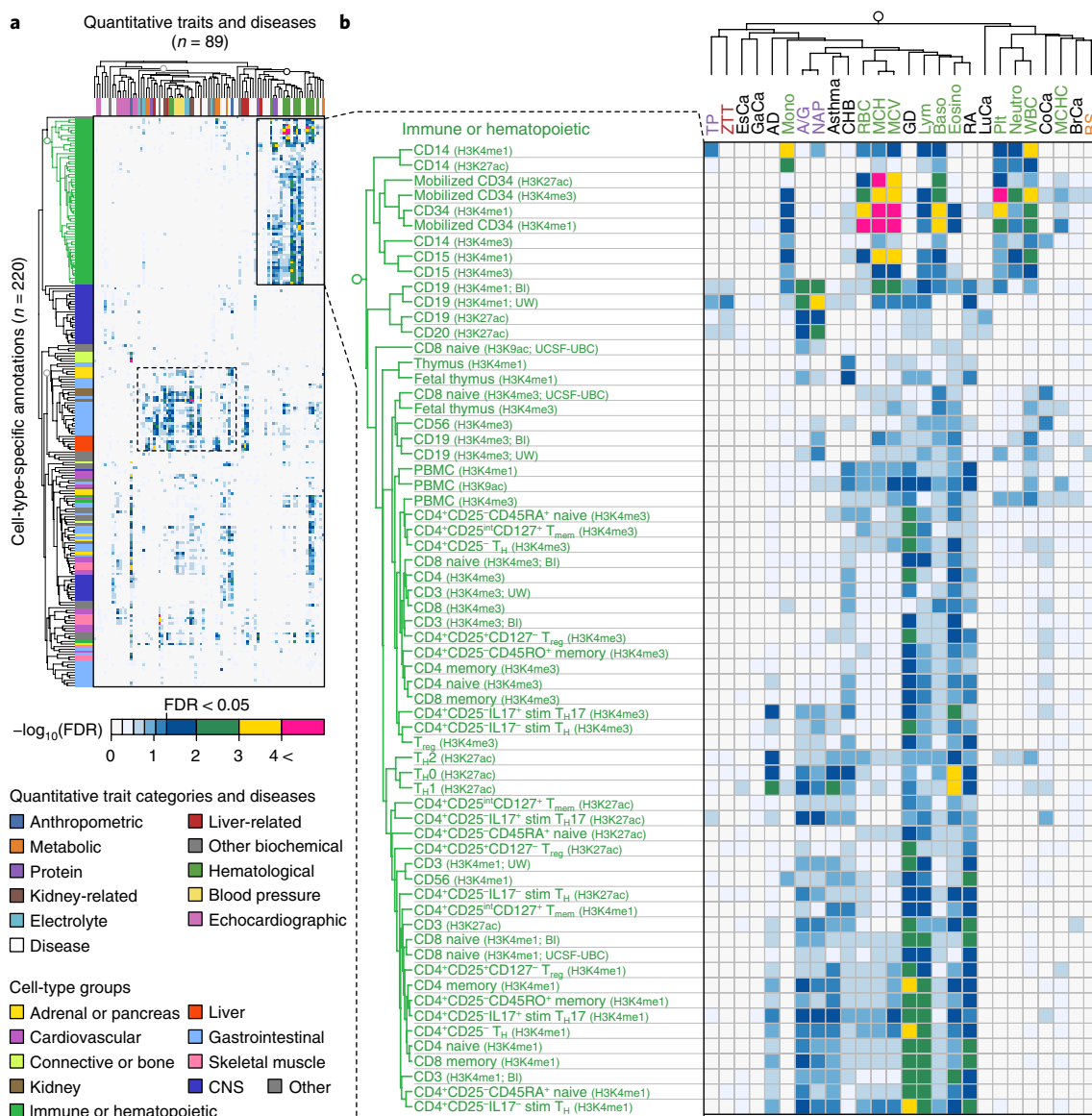
bilirubin<sup>36</sup>; and urolithiasis and blood pressure<sup>37</sup>) (Supplementary Table 8). Thus, we further investigated causal relationships between the significant pairs of quantitative traits and diseases by using a Mendelian randomization approach (Methods). We identified 24 significant causal associations ( $P < 9.43 \times 10^{-4}$  ( $= 0.05/53$ )), 15 of which had not been previously suggested by genetic causal relationships (Supplementary Fig. 7 and Supplementary Table 10). To distinguish bias due to pleiotropy, we further applied MR-Egger regression<sup>38</sup> as a sensitivity test, and confirmed the robustness of the identified causal relationships ( $P > 0.05$  for intercept after Bonferroni correction).

To facilitate understanding of the complex inter-relationships between clinical measurements and diseases, we constructed a network from the genetic correlation matrix (Fig. 3). In the network, the distance between correlated phenotypes is determined by weighting of the magnitudes of their correlations (Methods). Although we constructed our genetic correlation network without prior biological knowledge of cross-phenotype relationships, we observed distinctive clusters of biologically related phenotypes. The largest cluster was composed of cardiometabolic diseases and their biomarkers, interconnected with various clinical measurements, such as kidney-related, liver-related, and hematological traits. The network also depicted cross-disease interplay, including the positive

correlation of autoimmune diseases (rheumatoid arthritis and Graves' disease) and chronic inflammatory diseases (asthma and chronic obstructive pulmonary disease), as well as the negative correlation of glaucoma and Graves' disease. These results suggest that the polygenic landscapes of traits reflect their biological backgrounds, and thus could be used to elucidate the unknown etiology of diseases.

**Shared cell-type specificity among human complex traits.** The identification of trait-relevant cell types is essential for fine-mapping of candidate causal variants, the identification of potent therapeutic targets, and, ultimately, full understanding of disease biology<sup>39–41</sup>. To assess the cell-type specificity of human traits and diseases on the basis of heritability enrichment, we applied stratified LD score regression<sup>39</sup> to the GWAS results for the 59 quantitative traits and 30 diseases using 220 cell-type-specific annotations for histone modifications (H3K4me1, H3K4me3, H3K9ac, and H3K27ac) constructed from the Roadmap Epigenomics Project dataset<sup>39,42</sup>.

To create a broad picture of trait-relevant cell types, we first assessed heritability enrichment in ten major cell-type groups, defined as unions of 220 cell-type-specific annotations, representing their system- or organ-level structure<sup>39</sup>. We observed 72 significant heritability enrichments (FDR < 0.05) in the cell-type groups



**Fig. 5 | Hierarchical clustering of heritability enrichments in the 220 cell-type-specific annotations.** **a**, The significance of heritability enrichment among the 59 quantitative traits and 30 diseases, estimated via stratified LD score regression. Horizontal color bars correspond to the quantitative trait category for each column. Vertical color bars correspond to the cell-type groups for each row. **b**, A selected cluster of immune or hematopoietic cell types and various relevant traits and diseases. Another cluster highlighted in **a** is shown in Supplementary Fig. 9. FDR was calculated via the Benjamini-Hochberg method. T<sub>mem</sub>, memory T cell; T<sub>reg</sub>, regulatory T cell; T<sub>H</sub>, helper T cell; T<sub>H</sub>17, IL-17-producing helper T cell; T<sub>H</sub>1, type 1 helper T cell; T<sub>H</sub>2, type 2 helper T cell; T<sub>H</sub>0, naïve T cell; stim, stimulatory. Full trait and diseases abbreviations can be found in Tables 1 and 2.

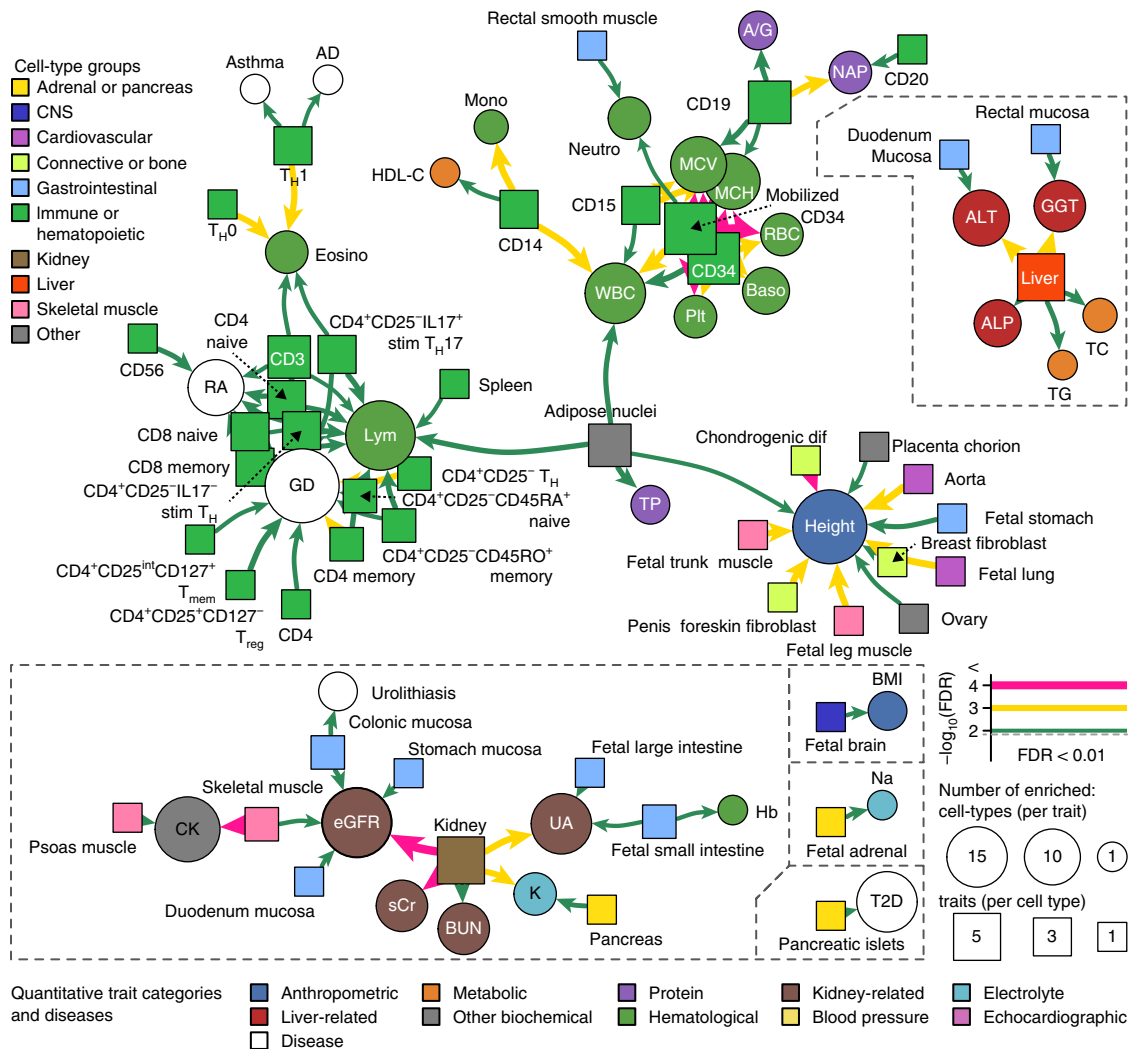
for 44 quantitative traits and diseases (Fig. 4a and Supplementary Table 11). The top significant enrichments in each quantitative trait category included connective or bone for height ( $P=4.89 \times 10^{-9}$ ), kidney for estimated glomerular filtration rate ( $P=2.59 \times 10^{-7}$ ), liver for GGT ( $P=2.54 \times 10^{-6}$ ), immune or hematopoietic for mean corpuscular volume ( $P=6.46 \times 10^{-6}$ ), and skeletal muscle for creatine kinase ( $P=2.77 \times 10^{-5}$ ), consistent with known biology (Fig. 4b and Supplementary Fig. 8). The same held true for the diseases—for example, significant enrichments in immune or hematopoietic for rheumatoid arthritis ( $P=9.19 \times 10^{-6}$ ) and Graves' disease ( $P=3.81 \times 10^{-5}$ ).

Although the cell-type-group-level analysis successfully identified a trait-relevant group for most of the quantitative traits and diseases, we hypothesized that more detailed assessment at the cell-type level would differentiate a trait-relevant cell type within the group. We thus

assessed heritability enrichment in each of the 220 cell-type-specific annotations. We identified 384 significant heritability enrichments (FDR < 0.05) for 50 quantitative traits and diseases (Supplementary Table 12). To further explore the complex systems of trait-relevant cell types, we carried out hierarchical clustering based on the earned profile of heritability enrichment for the 59 quantitative traits and 30 diseases in the 220 cell-type annotations (Fig. 5a).

We observed several distinct clusters that specifically comprised related traits and cell types. The most distinct cluster involved a great majority of immune or hematopoietic cell types enriched in hematological traits and in autoimmune, allergic, and infectious diseases, representing a wide range of immune-related diseases and traits (Fig. 5b). The most significant heritability enrichment was for mean corpuscular hemoglobin in mobilized CD34 ( $P=2.01 \times 10^{-9}$ ; H3K4me1). All CD34-related epigenetic annotations also showed





**Fig. 6 | Cell-type-specificity network across the 59 quantitative traits and 30 diseases in the 220 cell-type-specific annotations.** Significant heritability enrichments connect cell types to phenotypes. Each circle represents a trait, and each square represents a cell type. Arrows denote significant heritability enrichments for the indicated traits. The significance of heritability enrichment was estimated via stratified LD score regression. FDR was calculated via the Benjamini-Hochberg method. Thicker edges/arrows correspond to more significant FDRs. The dashed line separates unconnected plots. We used highly significant enrichments (FDR < 0.01). Definitions for abbreviations are as in Fig. 5 and Tables 1 and 2.

significant enrichments for red blood cell, WBC, and Plt-related hematological traits. Because CD34 is recognized as a marker of hematopoietic progenitor cells<sup>43</sup>, our findings suggest that variants in the regulatory region of CD34<sup>+</sup> primary cells affect hematopoietic cell differentiation and the number of hematopoietic cells.

Finally, to highlight shared cell types involved in human diseases and traits, we constructed a directed network matrix of cell-type-specific heritability annotations (Fig. 6; details are also presented in the Methods section). We identified several independent networks of cell-type specificity. The largest network was composed of three major clusters connected via the significant enrichment of adipose nuclei for (i) WBC count, (ii) lymphocyte count, and (iii) height. In addition to the contribution of CD34, we observed heritability enrichments in regulatory regions of CD14<sup>+</sup> and CD15<sup>+</sup> primary cells for WBC counts and WBC subtypes (i.e., monocytes and neutrophils), representing their specificity in myeloid lineages (CD14 for monocytes and macrophages<sup>44</sup>, and CD15 for granulocyte series cells<sup>45</sup>). Primary cells expressing CD19 and CD20, surface markers of B cells<sup>46</sup>, also showed enrichment for non-albumin protein and albumin/globulin ratio, potentially reflecting immunoglobulin-synthesis functions of B cells. Moreover, various CD4<sup>+</sup> and CD8<sup>+</sup> T cells

showed enrichment for autoimmune diseases such as Graves' disease and rheumatoid arthritis. We note that the enrichment of regulatory T cells ( $T_{reg}$  cells) in Graves' disease, a human autoimmune thyroiditis, is concordant with the biological finding that  $T_{reg}$ -cell-depleted mice develop thyroiditis<sup>47</sup>. Other observed links between allergic diseases (atopic dermatitis and asthma) and helper T cells, or about the contribution of fetal or chondrogenic tissues to height, also supported biological and medical findings.

These results demonstrate that 'individual cell-type level' analysis can successfully recapture the biology of human traits, without prior knowledge of 'consolidated cell-type group-level' analysis. The cell-type-specificity networks pinpoint potent causal cell types that cooperatively affect human phenotypes, providing promising resources for novel therapeutic targets. Nevertheless, integration of cell-type specificity in addition to polygenic genetic correlations clearly expanded the current knowledge of cross-phenotype relationships and underlying genetic mechanisms of diseases.

## Discussion

We have presented one of the largest non-European GWASs of quantitative traits so far, identifying 1,407 trait-associated loci for 53

traits in 162,255 Japanese individuals. By incorporating additional GWAS results for 32 complex diseases and traits in Japanese individuals, we further identified numerous pleiotropic loci, wide-ranging genetic correlations, and distinct cell-type specificity among the quantitative traits and diseases that confirmed or expanded our current understanding of biology.

Our findings suggest that there are complex inter-relationships between clinical measurements and diseases, demonstrating the value of GWASs for a variety of traits in a single large-scale cohort with detailed clinical information. We report novel genetic correlations, some of which are consistent with the results of epidemiological studies. These findings substantially expand the knowledge of genetic relationships across clinical measurements and diseases. We also highlight shared cell-type specificity by linking cell types to diseases. These results shed light on the underlying genetic mechanisms, revealing shared etiology and pathogenesis of complex diseases by using clinical measurements as an intermediate phenotype.

Although our work provided various insights into the genetics corresponding to clinical measurements in Japanese subjects, we should address several limitations of this study. First, we did not have a replication cohort for validation of the identified loci, but the majority of the trait-associated loci were previously reported ( $n=728$ ; 51.7%). This issue partly reflects a dilemma in the present study, namely, that extensive phenotypes were covered simultaneously, which makes replication more challenging. Second, our subjects for each trait mostly overlapped. Although bivariate LD score regression has elegantly modeled overlapping samples and their phenotypic correlation<sup>15</sup>, such sample overlap might exert an upward bias in interpretation of the genetic overlaps. Third, although we adopted a linear regression model for unrelated subjects, the application of a linear mixed model for both related and unrelated subjects could potentially have increased the statistical power of the study<sup>48</sup>. Fourth, the causal inference of clinical measurements for complex diseases in the present study could be limited because of the handling of the single cohort. Further application of Mendelian randomization<sup>17</sup> in independent validation cohorts is warranted. Finally, our cell-type analysis was inevitably limited by the availability of the cell-type-specific annotations regarding the variety of cell types and epigenetic markers. More acquisition and integration of cell-type-specific resources would further facilitate the exploration of causal cell types in human diseases.

In conclusion, we conducted a large-scale GWAS of 58 quantitative traits in Japanese individuals and demonstrated complex inter-relationships with human diseases via pleiotropy, genetic correlation, and cell-type-specificity analyses. We further visualized the results as networks, depicting the genetic links among clinical measurements, human diseases, and relevant cell types. Our findings will contribute to future studies and serve as a fundamental resource for understanding the genetics and biology underlying clinical measurements and human diseases.

**URLs.** BBJ, <https://biobankjp.org/english/index.html>; JENGER, <http://jenger.riken.jp/en/>; 1000 Genomes Project, <http://www.1000genomes.org/>; GWAS catalog, <https://www.ebi.ac.uk/gwas/>; PLINK 1.9, <https://www.cog-genomics.org/plink2/>; ldsc, <https://github.com/bulik/ldsc/>; LD score, <http://data.broadinstitute.org/alkesgroup/LDSCORE/>; MACH, <http://csg.sph.umich.edu/abecasis/MaCH/>; Minimac, <https://genome.sph.umich.edu/wiki/Minimac>; ANNOVAR, <http://annovar.openbioinformatics.org/en/latest/>; R, <https://www.r-project.org/>; Locuszoom, <http://locuszoom.sph.umich.edu/locuszoom/>; Circos, <http://circos.ca/>; NBDC Human Database, <https://humandbs.biosciencedbc.jp/en/>.

## Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41588-018-0047-6>.

Received: 5 June 2017; Accepted: 21 December 2017;  
Published online: 5 February 2018

## References

- Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
- Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
- 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).
- Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
- Surakka, I. et al. The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589–597 (2015).
- Okada, Y. et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat. Genet.* **44**, 904–909 (2012).
- Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7**, 10023 (2016).
- Kamatani, Y. et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **42**, 210–215 (2010).
- Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
- Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
- Liu, C. et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* **48**, 1162–1170 (2016).
- Ehret, G. B. et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* **48**, 1171–1184 (2016).
- Sivakumaran, S. et al. Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* **89**, 607–618 (2011).
- Han, B. et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nat. Genet.* **48**, 803–810 (2016).
- Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).
- Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–R98 (2014).
- Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
- 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).
- 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
- Tsai, W.-N. et al. Serum total bilirubin concentrations are inversely associated with total white blood cell counts in an adult population. *Ann. Clin. Biochem.* **52**, 251–258 (2015).
- Liu, Y. et al. Bilirubin possesses powerful immunomodulatory activity and suppresses experimental autoimmune encephalomyelitis. *J. Immunol.* **181**, 1887–1897 (2008).
- 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).
- Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
- 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).

28. Ogura, Y. et al. A functional SNP in *BNC2* is associated with adolescent idiopathic scoliosis. *Am. J. Hum. Genet.* **97**, 337–342 (2015).
29. Ikeda, M. et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol. Psychiatry* <https://doi.org/10.1038/mp.2016.259> (2017).
30. 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).
31. Liu, J., Au Yeung, S. L., Lin, S. L., Leung, G. M. & Schooling, C. M. Liver enzymes and risk of ischemic heart disease and type 2 diabetes mellitus: a Mendelian randomization study. *Sci. Rep.* **6**, 38813 (2016).
32. Rosenberg, M. A. et al. Genetic variants related to height and risk of atrial fibrillation: the cardiovascular health study. *Am. J. Epidemiol.* **180**, 215–222 (2014).
33. Khankari, N. K. et al. Association between adult height and risk of colorectal, lung, and prostate cancer: results from meta-analyses of prospective studies and Mendelian randomization analyses. *PLoS Med.* **13**, e1002118 (2016).
34. Wu, A. H., Gladden, J. D., Ahmed, M., Ahmed, A. & Filippatos, G. Relation of serum uric acid to cardiovascular disease. *Int. J. Cardiol.* **213**, 4–7 (2016).
35. Azab, B. et al. Value of albumin-globulin ratio as a predictor of all-cause mortality after non-ST elevation myocardial infarction. *Angiology* **64**, 137–145 (2013).
36. Perlstein, T. S., Pande, R. L., Beckman, J. A. & Creager, M. A. Serum total bilirubin level and prevalent lower-extremity peripheral arterial disease: National Health and Nutrition Examination Survey (NHANES) 1999 to 2004. *Arterioscler. Thromb. Vasc. Biol.* **28**, 166–172 (2008).
37. Timio, F., Kerry, S. M., Anson, K. M., Eastwood, J. B. & Cappuccio, F. P. Calcium urolithiasis, blood pressure and salt intake. *Blood Press.* **12**, 122–127 (2003).
38. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
39. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
40. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
41. Trynka, G. et al. Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. *Am. J. Hum. Genet.* **97**, 139–152 (2015).
42. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
43. Sidney, L. E., Branch, M. J., Dunphy, S. E., Dua, H. S. & Hopkinson, A. Concise review: evidence for CD34 as a common marker for diverse progenitors. *Stem Cells* **32**, 1380–1389 (2014).
44. Ziegler-Heitbrock, H. W. L. & Ulevitch, R. J. CD14: cell surface receptor and differentiation marker. *Immunol. Today* **14**, 121–125 (1993).
45. Gadhoun, S. Z. & Sackstein, R. CD15 expression in human myeloid cell differentiation is regulated by sialidase activity. *Nat. Chem. Biol.* **4**, 751–757 (2008).
46. Clark, E. A. & Lane, P. J. L. Regulation of human B-cell activation and adhesion. *Annu. Rev. Immunol.* **9**, 97–127 (1991).
47. Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.* **155**, 1151–1164 (1995).
48. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

### Acknowledgements

We acknowledge the staff of BBJ for their outstanding assistance in collecting samples and clinical information. We also thank the Tohoku Medical Megabank Project, the Japan Public Health Center-based Prospective (JPHC) Study, and the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study for their invaluable contributions to the case-control studies used in this study. We thank the staff of the Japan Scoliosis Clinical Research Group (JSCRG) for their support in recruiting patients to the AIS GWAS used in this study. We are grateful to H. Finucane for helpful discussions and assistance with LD score regression analysis. This research was supported by the Tailor-Made Medical Treatment Program (the BioBank Japan Project) of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) and the Japan Agency for Medical Research and Development (AMED). The study of psychiatric disorders was supported by the Strategic Research Program for Brain Sciences (SRPBS) of AMED. M. Kanai was supported by a Nakajima Foundation Fellowship. Y.O. was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (grants 15H05670, 15H05907, 15H05911, 15K14429, 16H03269, and 16K15738), AMED (grants 16km0405206h0001, 16gm6010001h0001, and 17ek0410041h0001), Takeda Science Foundation, the Uehara Memorial Foundation, the Naito Foundation, Daiichi Sankyo Foundation of Life Science, and Senri Life Science Foundation.

### Author contributions

M. Kanai, M.A., M. Kubo, Y.O., and Y.K. designed the study and wrote the manuscript. K.M., M.H., and M. Kubo collected and managed the BBJ samples. Y.M. and M. Kubo performed genotyping. M. Kanai, M.A., A.T., and N.M. performed statistical analysis. S.I., M.I., and N.I. contributed to data acquisition. Y.O. and Y.K. supervised the study. All authors contributed to and approved the final version of the manuscript.

### Competing interests

The authors declare no competing financial interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41588-018-0047-6>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to Y.O. or Y.K.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Methods

**Subjects.** All the subjects enrolled in this study were collected under the BioBank Japan Project (BBJ). The BBJ is a multi-institutional hospital-based registry that collected DNA, serum, and clinical information of approximately 200,000 patients from 66 hospitals affiliated with 12 medical institutes between fiscal years 2003 and 2007. All study participants had been diagnosed with one or more of 47 target diseases by physicians at the cooperating hospitals as described in the previous reports<sup>18,19</sup>. Written informed consent was obtained from all participants, as approved by the ethics committees of RIKEN Center for Integrative Medical Sciences and the Institute of Medical Sciences, the University of Tokyo. Detailed characteristics of the subjects for each trait are shown in Supplementary Table 1.

**Phenotype.** BBJ collected baseline clinical information through interviews and reviews of medical records using a standardized questionnaire. Among the quantitative traits included in this study, age, height, and weight were retrieved from the self-reported questionnaire for all participants. Laboratory measurements were retrieved from medical records of routine laboratory examination for all participants. Because dyslipidemia and diabetes were the most common diseases registered in the BBJ, around half of the study participants (41.8%) had these two diseases. Echocardiographic traits were retrieved from medical records only for the subjects with cardiovascular diseases, dyslipidemia, and diabetes. The measured values of each quantitative trait (or common log-transformed values if required, to achieve normality) were adjusted for age, sex, top ten principal components of genetic ancestry, disease status (affected versus non-affected) for the 47 target diseases in the BBJ, and any necessary trait-specific covariates in a linear regression model. We then normalized the resulting residuals by applying an appropriate trait-specific transformation ( $Z$ -score or rank-based inverse normal transformation) as detailed in Supplementary Table 2.

**Genotyping and imputation.** We genotyped samples with the Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips. We excluded samples with (i) sample call rate  $< 0.98$ , (ii) closely related individuals identified by identity-by-descent analysis, and (iii) non-East Asian outliers identified by principal component analysis of the studied samples and the three major reference populations (Africans, Europeans, and East Asians) in the International HapMap Project<sup>49</sup>. We then applied standard quality-control criteria for variants, excluding those with (i) SNP call rate  $< 0.99$ , (ii) minor allele frequency  $< 1\%$ , and (iii) Hardy-Weinberg equilibrium  $P$  value  $\leq 1.0 \times 10^{-6}$ . We prephased the genotypes with MACH<sup>50</sup> and imputed dosages with minimac and the 1000 Genomes Project Phase 1 (version 3) East Asian reference haplotypes<sup>50</sup>. For the X chromosome, we performed prephasing and imputation separately for females and males. Imputed SNPs with an imputation quality  $Rsq < 0.7$  were excluded from the subsequent association analysis.

**Genome-wide association analysis.** For each quantitative trait, we conducted a GWAS using a linear regression model under the assumption of additive allelic effects of the SNP dosages via mach2qt<sup>50</sup>. We set a genome-wide significance threshold at the level of  $P = 5.0 \times 10^{-8}$  (ref. 51) and a study-wide significance threshold at the level of  $P = 8.6 \times 10^{-10}$  ( $= 5.0 \times 10^{-8}/58$ ) by applying Bonferroni correction based on the number of studied traits. We defined independent associated loci on the basis of genomic positions at least 1 Mb apart from each other. We call such independent associated loci for each trait ‘trait-associated loci’, and these could overlap other trait-associated loci (i.e., multiple trait-associated loci could be mapped to one unique locus). We considered a trait-associated locus as novel when it was (i) located at a distance of  $> 500$  kb from the nearest locus and (ii) not in LD ( $r^2 < 0.1$ ) in both East Asians and Europeans with the previously reported loci of the same quantitative trait. For the X chromosome, we conducted GWASs separately for females and males, and meta-analyzed association results. We performed stepwise conditional analysis to identify additional independent signals around associated loci (each region  $\pm 500$  kb) by adjusting the most significant variant of the region in each step until none met the genome-wide significance threshold. For extremely significant variants showing  $P < 1.0 \times 10^{-300}$ , we calculated  $P$  values in R (ver. 3.3.1) with the Rmpfr package. We calculated the genomic inflation factor  $\lambda_{GC}$  in R. The variance explained by the significantly associated SNPs was estimated with the formula  $2f(1-f)\beta^2$ , where  $f$  represents the allele frequency and  $\beta$  represents the additive effect. We then summed the resulting values to calculate the total variance explained by the significant SNPs for each of the 53 quantitative traits that showed at least one genome-wide significant locus. We carried out LD score regression<sup>21</sup> with ldsc (v. 1.0.0; commit 23a94fc) to estimate confounding bias and heritability explained by the genome-wide high-quality common SNPs present in the HapMap 3 reference panel. We generated regional plots with LocusZoom<sup>52</sup> (v. 1.3) and R.

**Pleiotropy analysis.** We assessed pleiotropy at a unique locus using the following criteria: top-associated variants of different quantitative traits were (i) in LD ( $r^2 \geq 0.5$ ) or (ii) closely located (physical distance within 25 kb). We calculated  $r^2$  of two variants using PLINK 1.90<sup>53</sup> and the 1000 Genomes Project Phase 3 (version 5) East Asian dataset<sup>50</sup>. We used Circos<sup>54</sup> to visualize the results.

**Additional GWAS results for anthropometric traits and diseases in Japanese subjects.** We additionally obtained two quantitative trait GWAS results for anthropometric traits (BMI<sup>5</sup> and height), and 30 case-control GWAS results for complex diseases in the Japanese population from both published<sup>25–30</sup> and unpublished studies in the BioBank Japan Project (Table 2). For the two anthropometric traits, results for most of the subjects overlapped with those from the present study ( $n = 152,667$  (94.1%) and 153,456 (94.6%) for BMI and height, respectively). For the 30 complex diseases, the 26 disease cases were recruited through BBJ, whereas subjects with rheumatoid arthritis, bipolar disorder, schizophrenia, and adolescent idiopathic scoliosis were recruited by collaborators as described elsewhere<sup>26,28,29</sup>. The controls were constructed from three population-based cohorts (the Tohoku Medical Megabank organization, the Japan Public Health Center-based Prospective study, and the Japan Multi-Institutional Collaborative Cohort Study) or a mixture of the cases in BBJ as detailed in Supplementary Note 1. We incorporated these additional GWAS results into the original GWAS results for the 58 quantitative traits in the subsequent analyses.

**Genetic correlation.** We conducted bivariate LD score regression<sup>15</sup> to quantify genetic correlations across the 59 traits and 30 complex diseases in the Japanese population. To maintain sufficient statistical power<sup>15</sup>, we excluded one GWAS result (E/A ratio of echocardiographic trait) for which the sample size was far less than 10,000. For the regression, we used the East Asian LD score and summary statistics of high-quality common SNPs present in the HapMap 3 reference panel for each available trait or disease. We excluded SNPs found in the major histocompatibility complex (MHC) region (chromosome 6: 25–34 Mb) from the analysis because of its complex LD structure<sup>27,39,55,56</sup>. We defined significant genetic correlations as those with  $FDR < 0.05$ , calculated via the Benjamini-Hochberg method to correct multiple testing of all 3,916 pairwise correlations among the 59 quantitative traits and 30 diseases.

For network visualization, we constructed a network from the genetic correlation matrix of the 59 traits and 30 diseases. Specifically, each phenotype was represented as a node, and the nodes were connected by edges if they were genetically correlated. We assigned a weight to each edge based on the magnitude of the corresponding genetic correlation. To highlight biological patterns in the network and to prevent it from becoming too dense, we used only significant genetic correlations ( $FDR < 0.05$ ). Node layout was determined by the Fruchterman-Reingold algorithm given edge weights, with strongly correlated phenotypes placed closer together. We used R (ver. 3.3.1) with the igraph package for this network analysis.

**Mendelian randomization.** Given the 68 significant genetic correlations between clinical measurements and complex diseases, we carried out a Mendelian randomization analysis for each pair of them to evaluate potential causal effects of clinical measurements on complex diseases. Because most of the samples overlapped in the present study and the disease GWAS, we excluded overlapping samples from disease cases with clinical measurements available for each pair, to avoid potential bias. We selected 53 pairs on the basis of the following criteria: (i) raw genotypes of disease cases were available (i.e., the cases were recruited through BBJ (Table 2 and Supplementary Note 1)), (ii) more than three loci were identified in clinical measurement GWASs, and (iii) unique samples remained after the removal of overlapping samples. We note that this sample exclusion might have led to decreased statistical power compared with that of the original disease GWAS. For each pair, we calculated a weighted genetic risk score by summing the product of risk allele dosage and the effect sizes of the identified alleles influencing each clinical measurement. Associations between the genetic risk score and disease were quantified via a logistic regression model. To further test pleiotropy, we applied MR-Egger regression<sup>38</sup> as sensitivity analysis. We used R (ver. 3.3.1) with the MendelianRandomization package<sup>57</sup>.

**Partitioning heritability.** We carried out stratified LD score regression<sup>39</sup> to partition heritability into multiple functional categories. We used the 220 cell-type-specific and the 10 cell-type-group-specific annotations constructed based on the Roadmap Epigenomics Project<sup>42</sup> available at the authors’ website (see ‘URLs’). Because only European references are provided for partitioning heritability analysis, we generated the East Asian LD Score reference for each annotation using the 1000 Genomes Project Phase 3 (version 5) East Asian reference haplotypes<sup>50</sup> according to standard procedures. For each annotation, we calculated the  $P$  value of the regression coefficient  $\tau_c$  of the annotation. We defined significant heritability enrichments as those with  $FDR < 0.05$ , calculated via the Benjamini-Hochberg method.

We performed hierarchical clustering on the matrix of enrichment significance for the 59 quantitative traits and 30 diseases in the 220 cell-type-specific annotations, using Spearman’s correlation distance and the group average method. We also constructed a network from the matrix to represent the heritability enrichment of cell types to phenotypes. We assigned each phenotype and cell type to a node, and linked a pair of them with an arrow if a cell type was enriched for a phenotype. We assigned a weight to each arrow on the basis of the corresponding enrichment significance. For the sake of clarity, we used only highly significant enrichments ( $FDR < 0.01$ ). Node layout was determined with the Fruchterman-

Reingold algorithm given edge weights, with significantly enriched pairs of phenotypes and cell types placed closer together.

**Life Sciences Reporting Summary.** Further information on experimental design is available in the Life Sciences Reporting Summary.

**Data availability.** GWAS summary statistics of the 58 quantitative traits are publically available at our website (JENGER; see “URLs”) and the National Bioscience Database Center (NBDC) Human Database (Research ID: [hum0014](#)) as open data without any access restrictions. GWAS genotype data from the subjects was deposited at the NBDC Human Database (Research ID: [hum0014](#)).

## References

49. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
50. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
51. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).
52. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
53. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
54. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
55. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
56. Hirata, J. et al. Variants at *HLA-A*, *HLA-C*, and *HLA-DQB1* confer risk of psoriasis vulgaris in Japanese. *J. Invest. Dermatol.* <https://doi.org/10.1016/j.jid.2017.10.001> (2017).
57. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).

## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### ▶ Experimental design

#### 1. Sample size

Describe how sample size was determined.

The BioBank Japan Project was started in 2003 and has recruited approximately 200,000 participants. Clinical information were retrieved from medical records. DNA/serum samples were also collected. Most of them were genotyped by genome-wide SNP genotyping array. We selected as many subjects as possible if their genotype and phenotype (at least, one of the studied 58 quantitative traits) were both available.

#### 2. Data exclusions

Describe any data exclusions.

We excluded samples and variants based on the standard quality control procedure in GWAS. Detailed information on quality controls were described in our manuscript.

#### 3. Replication

Describe whether the experimental findings were reliably reproduced.

We searched for all of the identified loci in GWAS Catalog as well as previous literatures, and confirmed that around half of them robustly replicated the previous findings (51.7%). Because the BioBank Japan is the largest cohort in Japan and we simultaneously investigated more than 50 traits to illustrate a comprehensive landscape, to the best of our knowledge, there are no Japanese or east Asian replication cohort which covers all of the studied traits with enough sample sizes.

#### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

We applied quantitative genetics approach and treated the whole sample set as a single population, not allocated into groups.

#### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

- n/a Confirmed
- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
  - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
  - A statement indicating how many times each experiment was replicated
  - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
  - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
  - The test results (e.g.  $p$  values) given as exact values whenever possible and with confidence intervals noted
  - A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
  - Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

### 7. Software

Describe the software used to analyze the data in this study.

We used publicly available softwares for the analyses. The used softwares were listed and described in the Method section in our manuscript.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* [guidance for providing algorithms and software for publication](#) may be useful for any submission.

## ► Materials and reagents

Policy information about [availability of materials](#)

### 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used.

### 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Not applicable.

### 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Not applicable.

b. Describe the method of cell line authentication used.

Not applicable.

c. Report whether the cell lines were tested for mycoplasma contamination.

Not applicable.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

Not applicable.

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

### 11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Not applicable.

## 12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

We used 162,255 Japanese individuals from the BioBank Japan Project. For each of the studied 58 quantitative traits, GWAS was separately conducted using the subjects whose phenotype of the trait was available. We summarized detailed characteristics of the subjects for each trait in Supplementary Table 1.