GENETICS

A whole-genome reference panel of 14,393 individuals for East Asian populations accelerates discovery of rare functional variants

Jaeyong Choi[1]†, Sungjae Kim[2]†, Juhyun Kim[1]†, Ho-Young Son[3]†, Seong-Keun Yoo[4]†, Chang-Uk Kim[5], Young Jun Park[6], Sungji Moon[7,8], Bukyoung Cha[3], Min Chul Jeon[1], Kyunghyuk Park[3], Jae Moon Yun[9], Belong Cho[9,10], Namcheol Kim[2], Changhoon Kim[2], Nak-Jung Kwon[2], Young Joo Park[3,11,12], Fumihiko Matsuda[13], Yukihide Momozawa[14], Michiaki Kubo[14], Biobank Japan Project[15], Hyun-Jin Kim[16]*, Jin-Ho Park[9,10]*, Jeong-Sun Seo[2,17]*, Jong-Il Kim[1,3,8,18]*, Sun-Wha Im[19]*

Underrepresentation of non-European (EUR) populations hinders growth of global precision medicine. Resources such as imputation reference panels that match the study population are necessary to find low-frequency variants with substantial effects. We created a reference panel consisting of 14,393 whole-genome sequences including more than 11,000 Asian individuals. Genome-wide association studies were conducted using the reference panel and a population-specific genotype array of 72,298 subjects for eight phenotypes. This panel yields improved imputation accuracy of rare and low-frequency variants within East Asian populations compared with the largest reference panel. Thirty-nine previously unidentified associations were found, and more than half of the variants were East Asian specific. We discovered genes with rare protein-altering variants, including *LTBP1* for height and *GPR75* for body mass index, as well as putative regulatory mechanisms for rare noncoding variants with cell type–specific effects. We suggest that this dataset will add to the potential value of Asian precision medicine.

INTRODUCTION

Predicting nonassayed genotypes, called imputation, is an essential step for large-scale genetic research, especially genome-wide association studies (GWAS) (*1*, *2*). This process usually requires a reference panel that is constructed from large-scale whole-genome sequencing (WGS) data (*3*, *4*). Conventional imputation panels from global consortia such as the 1000 Genomes Project Phase 3 [1KGP3; (*5*)] and the Haplotype Reference Consortium (*3*) are widely used for imputation. The quality of imputed genotypes depends not only on the size (*2*, *6*, *7*) but also the population specificity of the panel (*8*). As the majority of genetic studies and reference panels are biased in favor of Europeans, an imbalance of studied populations within a variety of genetic research is frequently found (*9*). Therefore, increasing genetic diversity and specificity of reference panels will help explain diseases and complex traits of non-European populations (*10*). Substantial efforts have been devoted to increase genetic diversity, including the Uganda Genome Resource (*11*), Singapore 10K (SG10K) (*12*), Northeast Asian Reference Database (NARD) (*8*), and GenomeAsia 100K (*13*). Recently, the Trans-Omics for Precision Medicine (TOPMed) constructed and released a reference panel of 97,256 WGS samples which is, hitherto, the largest panel for genotype imputation (*14*). Their efforts have improved imputation accuracy in African, European, and even in admixed African and Hispanic/Latino populations (*15*). However, this large-scale panel has shown limited improvement for Asian populations.

GWAS have found numerous variants associated with phenotypes. However, the majority of variants are common with small effect sizes and cannot fully explain heritability (*16*). Population allele frequencies (AF) of genetic variants and their effect size are generally inversely correlated (*17*). These low-frequency variants are population specific because they are evolutionarily recent (*18*). Therefore, it is necessary to make efforts to find low-frequency variants that have a substantial effect on phenotype. Population-specific reference panels created from WGS data are essential for genotyping these powerful low-frequency variants.

[1]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea. [2]Macrogen Inc., Seoul, Republic of Korea. [3]Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Republic of Korea. [4]The Marc and Jennifer Lipschultz Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Psomagen Inc., Rockville, MD, USA. [6]Department of Translational Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. [7]Interdisciplinary Program in Cancer Biology, Seoul National University College of Medicine, Seoul, Republic of Korea. [8]Cancer Research Institute, Seoul National University, Seoul, Republic of Korea. [9]Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea. [10]Department of Family Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. [11]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. [12]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea. [13]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [14]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [15]Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [16]National Cancer Control Institute, National Cancer Center, Goyang, Republic of Korea. [17]Asian Genome Center, Seoul National University Bundang Hospital, Gyeonggi, Republic of Korea. [18]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Republic of Korea. [19]Department of Biochemistry and Molecular Biology, Kangwon National University School of Medicine, Gangwon, Republic of Korea.
*Corresponding author. Email: hyunjin@ncc.re.kr (H.-J.K.); pjhn@snu.ac.kr (J.-H.P.); jeongsun@snu.ac.kr (J.-S.S.); jongil@snu.ac.kr (J.-I.K.); imhere@kangwon.ac.kr (S.-W. I.)
†These authors contributed equally to this work.

Here, we present a large-scale WGS reference panel, NARD2, which generates accurately imputed genotypes for the East Asian (EAS) population, particularly those at extremely rare frequencies. We applied the NARD2 reference panel to 72,298 Koreans genotyped with a population-specific array which led to highly accurate imputation quality even at rare frequencies. GWAS was performed for eight phenotypes along with statistical fine-mapping and epigenetic annotation to predict the regulatory mechanism of putative causal variants located in the noncoding regions. In our opinion, our efforts to provide rich diversity in genetics will advance precision medicine.

RESULTS

A total of 14,393 individuals were included in the updated reference panel, NARD2

A basic graphical overview of our study regarding the imputation panel, GWAS, and functional annotation of noncoding variants is depicted in Fig. 1. We previously presented the WGS of Northeastern Asian individuals and built a reference panel to generate accurate genotype imputation dosages (*8*). To achieve improved accuracy for ultrarare variants with nonreference AF below 0.1%, we expanded the NARD to 9583 individuals using raw sequencing data that were obtained from 53 studies and archives. Details regarding the sources of added data are listed in table S1. Moreover, we merged the SG10K panel with our constructed dataset to create a large-scale reference panel with increased genetic diversity of the EAS population.

The 1KGP3 is one of the gold standard panels for a variety of genetic studies and is known to evenly cover global populations including 504 EASs who comprise 20.1% of the total panel size. Many efforts have been devoted to construct reference panels for EASs. The GenomeAsia 100K panel (*13*) contains diverse EAS populations but the size is still limited. Now, reference panels are scaling up to levels of more than 10,000 WGS samples (*14*). Detailed information regarding the number of samples is not available for the TOPMed reference panel, but 9.0% (n = 13,860) of total samples were identified as coming from the Asian population including South Asians, so the number of EASs could be far less than 9.0% of total samples in the TOPMed. In NARD2, 58.3% (n = 8386) of

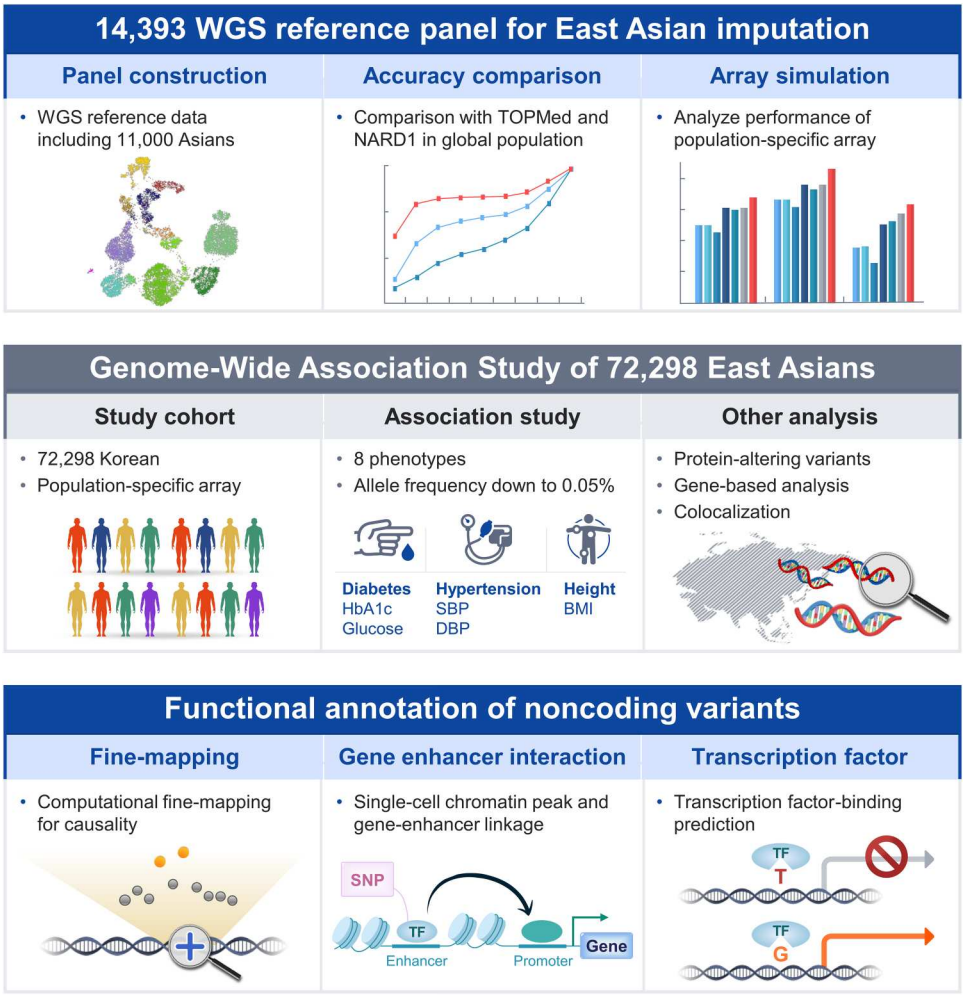

Fig. 1. Study schema. A large WGS reference panel was constructed with 14,393 individuals. GWAS was conducted with 72,298 Koreans imputed with the newly developed reference panel. Functional annotation was performed by applying fine-mapping and epigenetic annotations to GWAS summary statistics. SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index.

the total sample is from EAS populations (fig. S1A). Among these individuals, Korean (KOR), Japanese (JPN), Chinese (CHN), Mongolian (MNG), and Other Asians, including Southeast Asian (SEA) and various tribes in China (Others), account for 16.5, 37.8, 37.0, 5.4, and 3.4%, respectively (fig. S1B).

We performed population analysis to identify the genetic structure of individuals in the panel. The uniform manifold approximation and projection plot clearly showed distinctive clusters of samples by continent, importantly reflecting diversity within the EAS population including KOR, JPN, and CHN (fig. S2).

NARD2 yielded better imputation accuracy for EAS populations than the largest WGS reference panel

To evaluate the imputation performance of our reference panel, we compared the imputation accuracy of global populations including African (AFR), European (EUR), American (AMR), Middle Eastern (ME), SEA, Oceanian (OCE), South Asian (SAS), and MNG/Siberian (SIB) between NARD2 and TOPMed. We randomly selected 100 unrelated samples of each population from the NARD2 WGS dataset and selected genotypes included in the array to create a simulated array data (*3*). Then, we compared genotype imputation results from simulated array data and original genotypes from their WGS (Pearson coefficient of determination, R^2_{PCD}). At AF below 0.2%, we discovered clear differences between NARD2 and TOPMed. NARD2 provided improved R^2_{PCD} for KOR, JPN, MNG, CHN, SAS, OCE, and SEA samples compared with TOPMed, while TOPMed showed better R^2_{PCD} for AFR, EUR, AMR, and ME samples. This comparative pattern was also observed at AF between 0.2 and 0.5% and between 0.5 and 5% bins. As we expected for variants with AF more than 5%, the differences between NARD2 and TOPMed were negligible (Fig. 2A). We also compared the number of variants with an estimated imputation accuracy from Minimac4 (R^2_{Est}) greater than 0.9 for the global population. The number of single-nucleotide polymorphisms (SNP) with a high R^2_{Est} was 6.72, 6.36, and 5.94 million using NARD1 (rephased version of NARD1 merged with 1KGP3); 6.64, 6.43, and 6.96 million using TOPMed; and 7.58, 7.62, and 7.09 million using NARD2 for KOR, JPN, and CHN, respectively (Fig. 2B). The numbers of variants in MNG/SIB, SAS, SEA, and OCE were higher when NARD2 was used, while those in AFR, EUR, AMR, and ME were higher when TOPMed was used (Fig. 2B). The result regarding the number of high-quality variants was also consistent with the aforementioned results supporting improved imputation quality using NARD2 compared with NARD1 and TOPMed.

To precisely investigate the accuracy of reference panels at rare frequency for KOR, we compared average R^2_{PCD} at each rare frequency bin. NARD2 generated better R^2_{PCD} than NARD1 and TOPMed, specifically at AF below 0.1% (R^2_{PCD} = 0.467, 0.502, and 0.693 for TOPMed, NARD1, and NARD2, respectively; Fig. 2C) and from AF of 0.1 to 0.2% (R^2_{PCD} = 0.516, 0.662, and 0.833 for TOPMed, NARD1, and NARD2, respectively; Fig. 2C). Even the imputation performance of NARD1 from very rare to <5% was more accurate than TOPMed. In addition, NARD2 outperformed NARD1 and TOPMed for JPN and CHN at AF below 0.1% and AF of 0.1 to 0.2% (fig. S3). Imputation performance of NARD1 from very rare to <5% was more accurate than TOPMed, but TOPMed had better R^2_{PCD} than NARD1 for CHN (fig. S3B). The percentage of variants with R^2_{Est} greater than 0.9 was higher in NARD2 at every AF bin compared to NARD1 and TOPMed across KOR, JPN, and CHN. The difference in percentage between NARD2 and others increased as the AF decreased (table S2).

As we created simulated genotype array data, we evaluated the importance of population specificity of the genotype array by generating different types of array using KOR simulated array data. From WGS, we prepared simulated array data as follows: Thermo Fisher Scientific′s Affymetrix 6.0; Illumina′s Infinium Asian Screening Array v1.0; Illumina′s Infinium Global Screening Array; Illumina′s HumanOmni1-Quad; Illumina′s Omni2.5 BeadChip; Illumina′s Omni5 BeadChip; and a population-specific microarray (Fig. 2D). The population-specific genotype array was based on the Korean Biobank Array (KCHIP). To efficiently assess imputation performance on different types of array, we imputed chromosome 22 only. The average R^2_{PCD} was higher when population-specific arrays were used compared with the other types of arrays for KOR. We also compared population-specific arrays for JPN (Japonica Array NEO) (*19*, *20*) and CHN (Infinium OmniZhongHua-8 BeadChip). JPN showed similar results with KOR, but in a CHN dataset, Omni5 had the highest R^2_{PCD} compared with the other arrays (fig. S4). The percentages of variants with R^2_{Est} greater than 0.9 were comparable between population-specific arrays and Omni arrays but slightly higher when population-specific arrays were used for KOR and JPN (table S3).

GWAS identifies rare and novel variants after NARD2 imputation

To evaluate a population-specific reference panel for discovery of rare and novel variants that are associated with common traits, we imputed 72,298 Korean genotypes created with a population-specific array, KCHIP. Cohort characteristics are summarized in table S4.

After quality control and filtering, we performed GWAS with 16 million variants with minor allele frequency (MAF) over 0.05% for eight phenotypes: diabetes mellitus (DM), glucose, hemoglobin A1c (HbA1c), hypertension (HTN), systolic blood pressure (SBP), diastolic blood pressure (DBP), height, and body mass index (BMI). We found genomic inflation in some phenotypes. Genomic inflation factor (λ_{GC}) was the highest for height at 1.16 (λ_{GC} = 1.05 to 1.16 for all phenotypes), suggesting polygenicity or possible confounding biases such as population stratification. We quantified the contribution of confounding bias by performing linkage disequilibrium score (LDSC) regression (*21*). For height, LDSC intercept and ratio were 1.01 (SE = 0.00) and 0.04 (SE = 0.01), indicating that polygenicity was the cause of genomic inflation. In total, we observed 347 independent loci, including 39 novel loci with genome-wide significance level (Fig. 3A and table S5). Among novel variants, cohort-level AF for 13 variants was less than 1%.

Noticeably, rs902310682 is a rare population-specific variant associated with height [$P = 6.3 \times 10^{-11}$, beta (SE) = −0.324 (0.050), MAF = 0.0028]. It is located in the intron of *GRM4*, which is a major excitatory neurotransmitter in the central nervous system and is known to be related to height in both Europeans and EASs (*22*). Another novel variant, rs191684511, associated with HbA1c [$P = 7.3 \times 10^{-12}$, beta (SE) = 0.151 (0.022), MAF = 0.0299], is located in the intron of *HHEX*, which is a transcription factor that plays an important role in maintaining delta-cell differentiation and islet function (*23*). This variant has previously been reported in

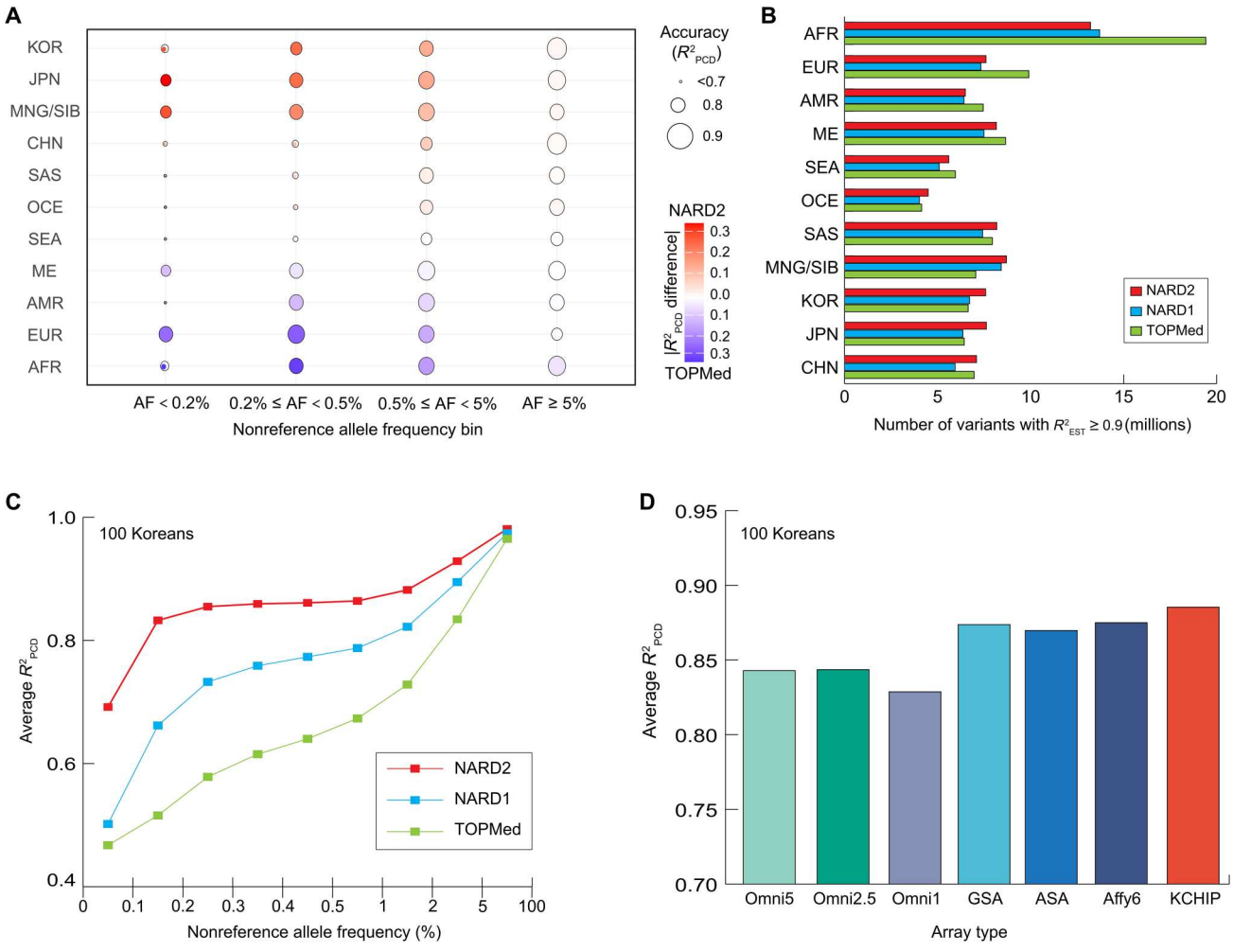

Fig. 2. Imputation accuracy comparison across the reference panels. (**A**) Global population comparison. Nonreference AF bins are determined as follows: AF < 0.2%, 0.2% ≤ AF < 0.5%, 0.5% ≤ AF < 5%, and AF ≥ 5%. Dot size is based on the Pearson coefficient of determination between true genotypes and imputed dosages (R^2_{PCD}) of the more accurate panel. Color represents the intensity of average R^2_{PCD} differences. (**B**) Number of variants with estimated imputation accuracy from Minimac4 (R^2_{Est}) above 0.9 across global populations. Red, blue, and green bars represent the number of imputed variants derived from NARD2, NARD1, and TOPMed, respectively. (**C**) The *x* axis represents the nonreference AF of KOR in the NARD2, and the *y* axis represents the aggregate R^2_{PCD} of variants. Imputation accuracy at each nonreference AF using simulated array data of 100 unrelated KOR individuals. (**D**) Imputation performances using different types of simulated arrays. The *y* axis represents the average R^2_{PCD} of variants in chromosome 22 using different types of microarrays. Each bar represents types of microarrays. Simulated array data were generated on the basis of types of genotype arrays: Affymetrix 6 (Affy6; Thermo Fisher Scientific), Infinium Asian Screening Array v1.0 (ASA; Illumina), Infinium Global Screening Assay (GSA; Illumina), HumanOmni1-Quad (Omni1; Illumina), Omni2.5 BeadChip (Omni2.5; Illumina), Omni5 BeadChip (Omni5; Illumina), and population specific array (Korea Biobank array, KCHIP).

diabetes GWAS performed by Biobank Japan (BBJ) (*24*) but was not significant for HbA1c levels in other EAS GWAS (*25*).

We sought to find overall similarity and colocalized variants between phenotypes. As expected, similarity values and colocalization counts were higher within groups of similar phenotypes, such as blood pressure-related traits (SBP, DBP and HTN) and glucose-related traits (glucose, HbA1c and DM).(Fig. 3B). Similarity between HTN and glucose-related traits, BMI and blood pressure–related traits were slightly higher than other nonsimilar traits, which may relate to complex relationships for general metabolic syndrome. Colocalization also revealed similar results to similarity indices (Fig. 3C).

Because rare variants might not reach genome-wide significance due to the small size effect of the allele carrier, we applied a moderate *P* value threshold to find high-impact protein-altering variants with low frequency. We observed 46 protein coding variants with nominal significance of 1×10^{-5}, and 10 of 17 variants that reached genome-wide significance were novel (table S6). Between all protein-altering variants, 20 (43%) had AF less than 1%.

GPR75 is a member of the guanine nucleotide-binding (G) protein–coupled receptor family that is highly expressed in the brain and involved in regulation of energy metabolism. A study using whole-exome sequencing of 640,000 subjects in the United Kingdom, United States, and Mexico found that protein-truncating *GPR75* variants have a large protective effect against obesity (*26*). Knockout mice $Gpr75^{-/+}$ and $Gpr75^{-/-}$ mice showed resistance to high-fat diet-induced weight gain, impaired glucose tolerance, and insulin sensitivity in an allele dose–dependent manner. In

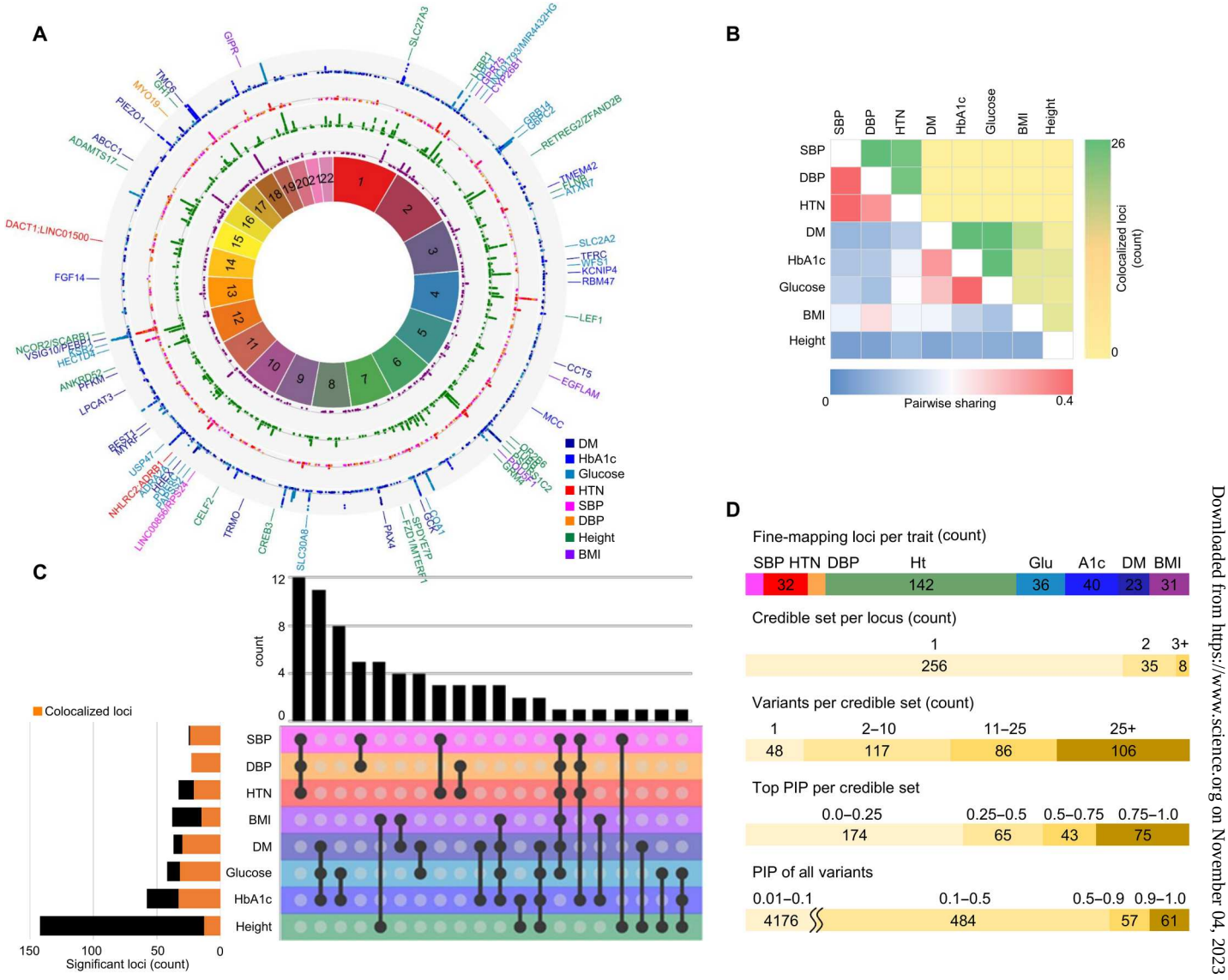

Fig. 3. GWAS and fine-mapping summary. (**A**) Circos Manhattan plot of eight phenotypes, novel independent signals, and protein coding variants are marked in the outermost rim. Maximum *P* value set to 1×10^{-30}, inner gray line denotes genome-wide significance level (5×10^{-8}). (**B**) Bottom left heatmap (blue to red) is the pairwise sharing between GWAS traits, and the upper right heatmap (yellow to green) is the colocalized loci for intersecting traits. (**C**) Upset plot of colocalization among GWAS traits. (**D**) Fine-mapping summary for all 234 loci.

our results, rs80328470 is a missense variant (p.T27A) of *GPR75* and has shown similar direction of effect as protein-truncating variants associated with BMI.

LTBP1 is a member of the family of latent transforming growth factor–β (TGF-β) binding proteins that regulate TGF-β activation. TGF-β signaling has been known to be associated with human height. In a recent study of consanguineous families with homozygous truncating *LTBP1* variants, subjects with LTBP1 deficiency showed connective tissue and skeletal disorders, including short stature (*27*). An in vivo study with zebrafish lines found that *ltbp1* variants affected skin and bone. Variant rs528249193 (p.G1258W) is a missense variant that has rarely been found in populations other than Koreans.

To measure the performance of our updated panel, we compared the R^2_{Est} between NARD2 and TOPMed for significant associations with MAF lower than 1% (table S7). Of the 19 variants, four were not included in the TOPMed panel, and a further two variants had lower R^2_{Est} than our threshold of 0.3. All six variants with low R^2_{Est} in TOPMed had novel associations, one of which was rs528249193, a rare coding variant in *LTBP1*. The other 13 variants had comparable accuracy. TOPMed had slightly higher R^2_{Est} for four variants, but overall, the R^2_{Est} of NARD2 was on average 0.11 higher.

Gene-based analysis reveals genes enriched with rare protein-altering variants

To find phenotype-associated genes concentrated with low-frequency variants, we selected protein-altering variants with MAF

lower than 5% and performed gene-based analysis. Many of the previously reported genes associated with the phenotype ranked highly in the gene-based analysis (table S8).

A total of seven genes were significant for HbA1c, among which well-known genes were *G6PC2* ($P = 2.22 \times 10^{-35}$), *LPCAT 3* ($P = 1.04 \times 10^{-7}$), and *PFKM* ($P = 1.65 \times 10^{-6}$). The association between HbA1c and *TFRC* ($P = 1.38 \times 10^{-6}$) has not been reported previously, but many variants in this gene have been reported to be associated with diabetes (*28*). TFRC encodes a high-affinity transferrin receptor, and it acts in iron transport, which affects glucose metabolism (*29*).

We identified 13 genes associated with height. We found strong associations in the two genes, *CYP26B1* ($P = 2.19 \times 10^{-13}$) and *SLC27A3* ($P = 1.60 \times 10^{-11}$), that were the most significantly identified in the Japanese study (*30*). *LTBP1* has a total of 28 low-frequency protein-altering variants. In addition to rs528249193 mentioned in the previous section, *LTBP1* has other meaningful variants, including rs770326287 [p.N1262S, $P = 6.96 \times 10^{-4}$, beta (SE) = −0.381 (0.113), MAF = 0.0005]. *LTBP1* was not statistically significant in gene-based analysis of the Japan study ($P = 6.96 \times 10^{-2}$).

For HTN and SBP, *RNF213* ($P = 2.29 \times 10^{-14}$ and 1.05×10^{-15}, HTN and SBP, respectively) was significantly associated. Defects in RNF213 are the cause of Moyamoya disease (*31*). *RNF213* has 161 low-frequency protein-altering variants. The strongest coding variant was rs112735431 (HTN: $P = 5.01 \times 10^{-18}$, SBP: $P = 5.77 \times 10^{-6}$, and DBP: $P = 3.07 \times 10^{-6}$), which is specific to EASs.

For BMI, *GIPR* ($P = 7.46 \times 10^{-13}$), *GPR75* ($P = 8.80 \times 10^{-8}$), and *MC4R* ($P = 7.36 \times 10^{-7}$) were found to be associated. They had 14, 7, and 6 low-frequency protein-altering variants, respectively. All three genes encode G protein–coupled receptors expressed in the brain. GIPR is a well-known obesity-promoting hormone in mouse models and human genetic studies (*32*). MC4R is a member of the melanocortin receptor family and is involved in energy balance by interacting with melanocyte stimulating hormone (*33*). Genetic defects in *MC4R* cause obesity (*34*).

We sought to replicate our results using GWAS from BBJ which has similar EAS ancestries. Of our 387 reported associations, 334 could be found in BBJ (table S9). Two-hundred forty-nine signals were replicated with Bonferroni corrected significance level [$P < 1.497 \times 10^{-4}$ (= 0.05/334)], and 49 signals were replicated with nominal significance ($1.497 \times 10^{-4} < P < 0.05$).

Epigenetic annotation reveals putative regulatory mechanism for GWAS causal variants

One of the limitations of GWAS is that the true causal variant may not have the highest significance, and the closest gene may not be the gene that affects the phenotype. To overcome this challenge, we used a two-step approach to find putative causal variants. First, sum of single effects (SuSiE) (*35*) was used to computationally fine-map causal variants for each GWAS locus (Fig. 3D). We found 55 variants with posterior inclusion probability (PIP) over 0.9 (table S10). Second, we reviewed multiple epigenetic databases for putatively causal variants with PIP ≥ 0.1 to search for evidence of relevant epigenetic noncoding variants (table S11).

In 234 GWAS loci of eight phenotypes, we found loci that contained at least one variant that altered protein with high or moderate effect (34, 14.5%), located in a regulatory region predicted to interact with a nearby gene (118, 50.4%) where the causal single-cell type could be specified (121, 51.7%), which modified any transcription factor–binding motif (151, 64.5%). Summing up, 197 (84.2%) loci contained at least one putative causal variant that we could relate to any kind of biological function. In addition, assay for transposase-accessible chromatin (ATAC) peak enrichment for 220 single-cell types was used to prioritize phenotype-related single-cell types (fig. S5).

We provide some examples to illustrate the utility of epigenetic annotation in discovering putative regulatory mechanisms.

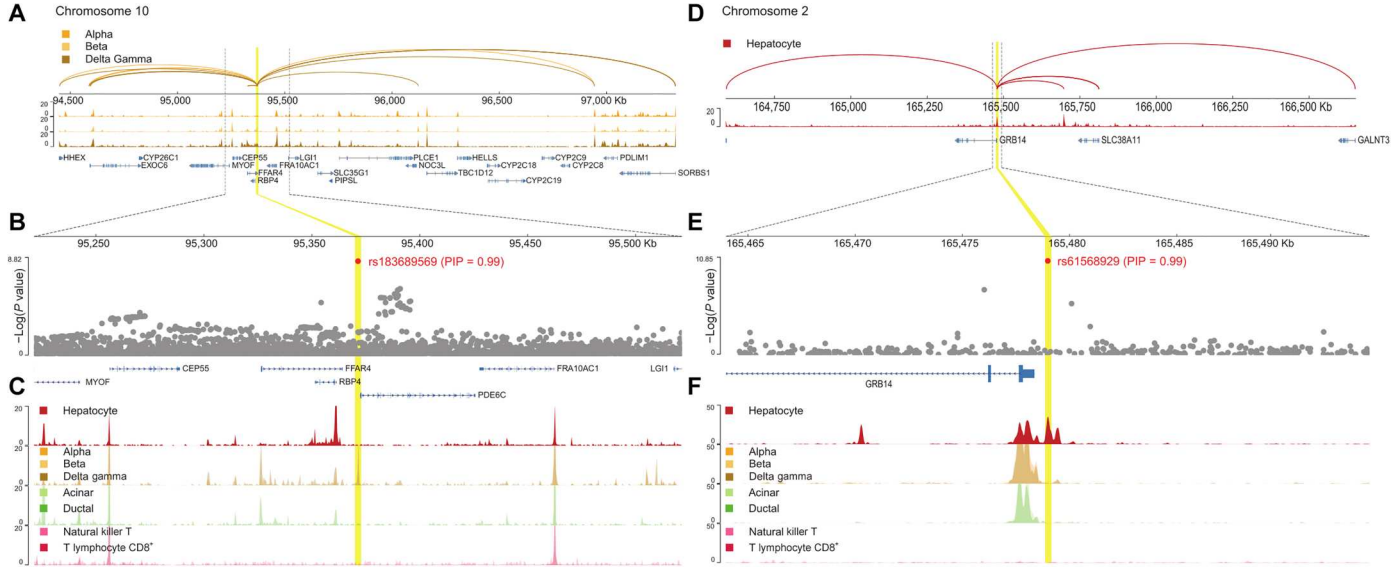

Fig. 4. Epigenetic annotation of putative causal variants. Examples of epigenetic annotation for variant (**A** to **C**) rs183689569 and (**D** to **F**) rs61568929, both associated with blood glucose levels. (A and D) Line in the top represents single-cell–level gene-enhancer interactions predicted by the activity-by-contact (ABC) model, and the color specifies single-cell type. (B and E) Zoomed in Manhattan plot for the region of interest. Variants with *P* value less than the genome-wide significance level (5×10^{-8}) are colored in red. (C and F) Variant of interest is located in single-cell type–specific open chromatin peaks.

Additional examples are illustrated in figs. S6 to S8. Novel glucose-associated variant rs183689569 is found in rare frequencies (MAF = 0.003) among EASs and is not reported in Europeans. Single-cell gene-enhancer link predicted interactions specific to these cell types, interacting with *EXOC6*, *HHEX*, and other nearby genes (Fig. 4, A and B). Chromatin accessibility of the region where rs183689569 resides is specific to pancreatic islets such as alpha, beta, delta, and gamma cells (Fig. 4C). EXOC6 is mostly expressed in pancreatic islet cells and is known to regulate insulin secretion (*36*). Previously mentioned, HHEX is also known to regulate pancreatic islets (*23*). It was predicted that this variant changes the binding affinity for transcription factor DBP which regulates the circadian rhythm of beta cells. Summing the evidence, rs183689569 increases the binding affinity of the transcriptional factor DBP, leading to altered expression of nearby genes in pancreatic islet cells specifically, and may affect blood glucose levels.

Another glucose-associated variant, rs61568929, is located in the promoter of *GRB14*. This variant is found in low frequencies among EASs and is very rare in Europeans. Rs61568929 is predicted to interact with several nearby genes, including *GRB14* (Fig. 4, D and E), and the chromatin conformation is only open in hepatocytes (Fig. 4F). Rs61568929 changes the binding affinity of BHLHA15, which is a transcription activator predicted to be involved in glucose homeostasis (*37*). GRB14 is known to play a role as a negative regulator of the insulin receptor. Knockout of Grb14 in liver improved glucose homeostasis in diet-induced obese mice (*38*). In summary, rs61568929 may alter the binding affinity of transcription factors such as BHLHA15 and modify the expression of *GRB14* in hepatocytes leading to lower blood glucose.

Height-associated causal variant rs11107120 is located in an open chromatin region common to 172 cell types. Epigenetic marks of this region show enhancer-like signatures (fig. S9). Alternative alleles of the variant enhance binding of transcription factor DBP and HLF. From the gene-enhancer activity-by-contact (ABC) model, this locus was predicted to interact with the *SOCS2* gene in four cell types. SOCS2 proteins inhibit growth-promoting cytokine receptor signaling (*39*). The expression of *SOCS2* can be induced by various cytokines including growth hormone and insulin-like growth factor (*40*). SOCS2 protein is involved in insulin-like growth factor 1 (IGF-1) receptor signaling and in the TGF-β pathway (*41*). $Socs2^{-/-}$ mice show increased long bone length and body weight, and most organs are enlarged (*42*). In $Socs2^{-/-}$ mice, growth hormone and IGF-1 signaling are deregulated. These data suggest that SOCS2 may be an essential negative regulator of height and growth.

DISCUSSION

Extremely low genetic diversity results in poor imputation qualities for non-European populations (*9*, *10*). These imputed genotypes might affect the misinterpretation of GWAS results for non-European populations, caused by the presence of false positives or false negatives (*43*, *44*). Although many research groups have constructed reference panels for those underrepresented populations, these panels and databases have not as yet been extensively used in genetic studies. Previously, we constructed a panel called NARD to improve deficient imputation quality for Northeast Asians, and it demonstrated the potential to become an Asian representative reference panel.

In this study, we constructed a large-scale reference panel of 14,393 individuals to provide high-quality imputed genotypes. We included not only EASs but also individuals from diverse populations including AFR, EUR, and SAS in our reference panel, to increase the size of the panel which contributes to the imputation results considerably (*2*, *6*, *7*) as our goal is to build a reference panel that will yield improved imputation accuracies even for rare-frequency variants. Also, diverse populations in NARD2 would be beneficial for genetic research on admixed individuals, and this would broaden the use range of NARD2 in genetic studies. Before assessing imputation performance, we first illustrated the genetic diversity of this recently established panel. Consistent with previous findings (*8*), distinctive clusters among Asian populations, including KOR, JPN, CHN, and MNG/SIB were observed. We then compared imputed dosages and true genotypes by creating simulated array data for 11 representative populations and found that average R^2_{PCD} values were higher when NARD2 was used compared with TOPMed. At ultrarare AF (below 0.2%), we discovered substantial difference in R^2_{PCD} between these two panels of EAS populations. In addition, the number of variants with R^2_{Est} of > 0.9 using NARD2 was also higher than those using TOPMed. This evaluation analysis demonstrates that the improved panel generates more accurate imputed genotypes for Asian populations, despite a smaller number of WGS compared with the TOPMed panel which was recently constructed using more than 90,000 WGS. Increased accuracy of NARD2 suggests that more extensive genetic candidates can be selected for investigating Asian populations.

Using a population-specific genotyping array and a population-matching imputation panel, we were able to obtain highly accurate genotypes which led to the discovery of several novel GWAS variants, especially at low AF. Finding low-frequency coding variants with large effects has recently been a concern of GWAS, and some successful cases have been reported (*45*–*48*). Protein-altering variants affecting certain phenotypes are crucial in understanding pathophysiology and can be directly linked to treatment (*26*, *49*). However, low-frequency variants have inherently population-specific properties, so it is essential to develop a reference panel using WGS for each population to discover them. Here, we report 17 low-frequency protein-altering variants reaching genome-wide significance level, including 10 novel variants, of which many were EAS specific. Gene-based analysis using these low-frequency coding variants found many gene-level associations, including a previously unidentified one between HbA1c and *TFRC*.

Many GWAS variants are positioned in the noncoding region, and the biological mechanism beyond the association remains unknown. We performed statistical fine-mapping, as well as epigenetic annotations to discover the causality and regulatory functions for these variants. Single-cell–level epigenetic resources and gene-enhancer interactions made it possible to pinpoint the target gene that a GWAS variant affects, as well as the cell of origin where the regulation occurs. Rs11107120, which is a putative causal variant for height, was predicted to interact with genes from various cell types, which could be the case as multiple tissue and cell types are known to influence height. By contrast, rs183689569, which is a putative causal variant for blood glucose level, was found to interact only within cell types found in pancreatic islets. Variants in the enhancer region may change the binding affinity of transcription factors, leading to differences in gene regulation. We found putative

causal variants positioned in transcription factor–binding motifs, with some predicted to affect the binding affinity. We also found putative causal variants lying within CCCTC-binding factor (CTCF) binding regions, which could affect topologically associating domains.

Unfortunately, our NARD2 imputation reference panel was only created for the hg19 version of the human genome. More up-to-date reference genomes, such as hg38, CHM13 (*50*), or future references such as the human pan-genome reference (*51*) may improve accuracy. Compared with recent GWAS publications (*52*), our cohort size is relatively small. A larger sample size would enable us to find more population-specific rare and functional variants. The majority of epigenetic annotations publicly available comes from Caucasian sources (*53*), and our approach would only be able to detect genetic regulators common to both ancestries, which is another limitation. Although imputation quality depends on sample size of a reference panel, our results emphasize the importance of population specificity. We anticipate that our ongoing efforts will facilitate precise and accurate genetic research, especially for Asian populations. To this end, we provide this panel in a user-friendly web server (gmi.snu.ac.kr/imputation).

MATERIALS AND METHODS

Variant calling and quality control

Sequencing reads were aligned to the human reference genome (hg19) and genomic variant call format files (gVCFs) were generated using Dynamic Read Analysis for GENomics platform (version 01.003.024.02.00.01.23004). We used hg19 because most array data including the one used in our further analyses are based on hg19, and our purpose was to construct a whole-genome reference panel for GWAS using SNP array data (*54*, *55*). We then divided the gVCFs into 100-kilobase pair (kbp) segments and performed joint genotyping with a batch size of 50 bp using Genome Analysis ToolKit 4′s GenomicsDBImport and GenotypeGVCFs (*56*). Variants based on truth sensitivity of 99.7% were filtered by variant quality score recalibration using resources including Haplotype Map 3.3 (*57*), 1KGP Omni2.5 (*58*), Genome Aggregation Database (gnomAD) r2.2.1 (*59*), 1KGP phase 1, and Mills & 1KGP gold standard (*58*). Variants were additionally filtered for variant quality controls and avoiding potential batch effects across different resources (*60*, *61*) using the following parameters: genotype quality < 20, read depth < 5, and genotype rate per SNP < 85% (*8*, *62*, *63*). In our dataset, 473 WGS were CompleteGenomics (CG) data which has a different format compared with the conventional VCF. Therefore, these CG var files needed to be processed using CG-specific computation tools so we converted these 473 CG var files into gVCFs using cgivar2gvcf 0.1.7 and split blocks using gVCFtools-0.17.0. Then, we filtered low-quality and non-PASS variants in converted VCFs.

After collecting data for constructing the reference panel, we excluded samples with less than 2.5 million SNPs (*64*) and an abnormal ratio of heterozygous to homozygous genotypes. Since our dataset was derived from multiple studies, we evaluated the batch effect of our jointly called set. We conducted principal components analysis (PCA) on our joint-called data before and after variant filtration and compared their genetic distributions of their PC. We separated samples in different batches into large studies with more than 100 global samples including 1KGP3, Human Genome Diversity Project, and Simons Genome Diversity Project, and we classified data from other resources as "Others." For EAS, we further classified data from NARD1 (*8*) and JPN individuals from National Bioscience Database Center, which has the largest sample size among the batches in NARD2. We observed some distribution differences between the batches within EAS and SAS before variant filtration, but batches were more clustered for the data after poor-quality variants had been removed (fig. S10, A and S10B). We also found batch-correct effects of non-Asian populations, for example, EUR after variant filtration (fig. S10C). The number of novel variants was classified on the basis of gnomAD v3.1 and v2 (*59*), Known Variants (*65*), the Exome Aggregation Consortium database (*66*), and the Single Nucleotide Polymorphism Database build 150 (*67*). As samples were from various resources, we performed relationship analysis to identify and remove potential duplicate samples within the dataset using Kinship-based Inference for Genome-wide association studies (KING) (*68*).

Panel construction

We extended the WGS set of 9583 individuals by merging with the SG10K to construct a large-scale reference panel with a total of 14,393 individuals. We removed singletons and variants with more than 49 bp (*69*, *70*) to retain a total of 82.7 million SNPs and 3.92 million indels. Then, the panel was phased in parallel using BEAGLE v5.0 (*71*) by splitting chromosomes into chunks using splitVCF.jar in BEAGLE utilities. We evaluated the effect of choosing the size of the chunk to the number of phasing blocks that reflect the error of phasing. The sizes were prepared from 10,000, 30,000, 50,000, 70,000, and 80,000 variants with a 10% overlap each between the chunks and whole chromosomes. Using chromosome 22 only, phasing error reflected by the number of phasing blocks gradually decreased when the size of the chunk was increased (fig. S11). On the basis of this result, we chose the most efficient chunk size of 80,000 variants with an overlap of 8000 variants between the chunks. The chunks were then merged using mergeVCF.jar in BEAGLE utilities (*71*).

After the panel of 9583 samples was constructed, we extended our reference panel by merging the SG10K panel consisting of 4810 individuals (*12*). Typically, reference panels are merged via reciprocal imputation (*6*). However, we did not merge two panels based on reciprocal imputation because imputed SG10K panel–specific variants in the 9583 panel can be less informative. Approximately 89.1% of SG10K variants with MAF > 0.5% were overlapped with the 9583 set. Therefore, our alternative strategy was to impute NARD2-specific variants in the SG10K panel to rescue relatively more accurate imputed genotypes.

Population genetic structure analysis

For population analysis, we extracted biallelic autosomal SNPs from the dataset and converted them into PLINK binary format (*72*). First, we pruned SNPs with LD squared correlation (R^2_{LD}) > 0.1 within the 50-base sliding window, and then we filtered variants with MAF ≤ 1%. The PCA was carried out using genome-wide complex trait analysis (GCTA) version 1.91.3beta (*73*). We performed additional dimension reduction analysis with 13 PC data using a uniform manifold approximation and projection algorithm (*74*). The analysis was executed with the following parameters: n_neighbors = 200, min_dist = 1.0, n_components = 2, and metric = canberra. The number of inferred ancestral populations

was optimized using the cross-validation error rates of each kbp. Population classification was based on self-described ethnicity information of individual samples obtained from each source of study.

Imputation accuracy measurement

Before constructing simulated array data for imputation accuracy measurement, we identified a total of 12,803 unrelated samples in the panel by performing kinship estimation using KING (*68*). We selected 100 of these unrelated samples from 11 populations based on their self-reported ancestries to avoid potential batch effects across these datasets.

After sample selection for the datasets, we masked genotypes of their WGS to preserve genotypes within the Illumina Omni 2.5 array. As 100 selected samples of the simulated array dataset were included in the reference panel, we imputed these data to the modified NARD1 and NARD2 panels without those selected 100 samples. To measure imputation accuracy of each reference panel, we uploaded these masked genotypes to the TOPMed imputation server to obtain phased and imputed dosages using the TOPMed reference panel. For NARD panels, we conducted phasing and imputation under consistent settings provided by the TOPMed imputation server to avert unwanted effects on imputation accuracy by using different algorithms: Eagle v2.4 (*75*) and Minimac4 (*76*) for phasing and imputation, respectively. Then, we calculated the Pearson coefficient of determination between true genotypes and imputed dosages (R^2_{PCD}). Average R^2_{PCD} at each nonreference AF bin was calculated. For AFR, AMR, and EUR populations, gnomAD v3.1(*59*) was used to define AF bins because it has a larger sample size for these populations than NARD2. Because TOPMed supports only hg38 coordinates, imputed variants were liftovered using Picard v2.18.25 (*56*, *77*). Other populations except KOR, JPN, and CHN used NARD2 total frequency bins. We used KOR-, JPN-, and CHN-specific frequencies because NARD2 has an adequate number of WGS to reflect population-specific frequencies.

Data collection for GWAS

This research project was approved by the institutional review board of Seoul National University Hospital Clinical Research Institute (C-2004-080-1117). De-identified Korean Genome and Epidemiology Study data, which includes three cohorts (city cohort, N = 58,700; rural cohort, N = 8105; and Ansung Ansan Community cohort, N = 5493), were received from the Korea National Institute of Health, Korea Disease Control and Prevention Agency. All data were generated with the Korean Biobank Array and preprocessed according to the Korea Biobank Array Project analysis protocol (*54*). We imputed genotypes with the NARD2 reference panel. Haplotypes of the input genotypes were phased by BEAGLE v5.0 (*71*) using impute = false, ap = true, and gp = true options. We then performed imputation of the phased genotypes by Minimac4 (*76*) using the NARD2 reference panel with "allTypedSites" and "ignoreDuplicates" options. We selected variants with $R^2_{Est} \geq 0.3$, Hardy-Weinberg equilibrium $P \geq 1 \times 10^{-6}$, and variant missing rate < 0.1.

Phenotype data were provided by the Korea National Institute of Health, Korea Disease Control and Prevention Agency. For all quantitative phenotypes, we performed normalization for males and females separately and merged the normalized values afterward. We first inversely normalized each phenotype and then used linear regression with age, age^2, and the first five genotype principal components. The residuals were standardized to a normal distribution. We excluded samples from individuals taking antihypertensive medication and from individuals taking diabetes medication. DM was defined if at least one of the following criteria was met: (i) record of diabetic medication; (ii) HbA1c $\geq$ 6.5%; (iii) fasting blood glucose level $\geq$ 126 mg/dl. HTN was defined if at least one of the following criteria was met: (i) record of antihypertensive medication; (ii) SBP $\geq$ 140 mmHg; (iii) DBP $\geq$ 90 mmHg.

Replication with BBJ

Full summary statistics for seven matching phenotypes and one comparable phenotype (HTN versus use of antihypertensives) were downloaded from https://pheweb.jp (*78*). We compared each variant-trait association after matching the effect allele.

Fine-mapping and epigenetic annotation

Fine-mapping was performed with polygenic functionally informed fine-mapping (*79*) adaptation of SuSiE (*35*) for each locus defined previously with the maximum number of causal variants set to 10. Causal variants were defined as variants within a credible set with PIP higher than 0.9.

Assay for ATAC peak enrichment

To generate enrichment scores for cell type–specific open chromatin peaks, we obtained ATAC peak calls for 220 cell types from Cis-element Atlas (CAT-las) (*80*). We defined cell type–specific peaks as open chromatin regions that were found in less than five cell types among 220 cell types. GWAS summary statistics were used to calculate enrichment scores using FGWAS (*81*) with default parameters.

Functional fine-mapping for noncoding variants

To examine the regulatory mechanism of noncoding variants, we selected putative causal variants, defined as variants within a credible set with PIP higher than 0.1. We curated multiple databases to select relevant functional noncoding variants, including ATAC peak calls and gene-enhancer ABC model (*82*) from CAT-las (*80*), single-cell expression and ATAC data from DESCARTES (*80*, *83*), single-cell expression data from Tabula Sapiens (*84*), DNase hypersensitivity site and gene-enhancer link from EpiMap (*85*, *86*), candidate cis regulatory regions and transcription factor–binding sites from Encyclopedia of DNA Elements (*87*), transcription binding motif from JASPER (*88*), and prediction of transcription binding affinity change from deltaSVM (*89*). Liftover was used to change the coordinates to hg19 when hg19-based data were not provided.

Statistical analysis

Association analysis

Plink2 (*90*) was used to generate association statistics for all variants with MAF larger than 0.05% at the computing server of the Genomic Medicine Institute Research Service Center. We defined GWAS loci as a variant group with a P value less than the genome-wide significance level (5×10^{-8}) when the maximum distance between the variants was less than 1 Mbp. If the distance was larger than 1 Mbp, we split the loci evenly to 1-Mbp sizes considering the P value distribution. GCTA-conditional and joint multiple-SNP analysis (COJO) (*91*) was used to identify independent signals. Novel variants were defined if the variant was not previously

reported, and the variant did not have R^2_{LD} higher than 0.7 with any other reported variants in our study.

LD score

LDSC v1.0.1 (*21*) was used to estimate confounding bias due to population stratification and cryptic relatedness. Univariate LD score was estimated with NARD2 imputed genotypes with default parameters. Full GWAS summary statistics were applied with info-min = 0.3, maf-min = 0.0005 following the same filter criteria used for association.

Protein-altering variants and gene-based analysis

To screen for rare protein-altering variants, we selected coding variants with a *P* value lower than 1×10^{-5} and with high or moderate impact in Ensembl Variant Effect Predictor (*92*). Magma (*93*) multimodel was used for gene-based analysis. To find phenotype-associated genes enriched with low-frequency variants, protein-altering variants with MAF lower than 5% were selected as input. Very rare variants previously not included in the association analysis due to MAF being lower than 0.05% were included in the gene-based analysis as a count in the burden score calculation. Variants within the major histocompatibility complex region were not included.

Colocalization and pairwise sharing

Hypothesis Prioritisation in multi-trait Colocalization (*94*) was used to conduct multi-trait colocalization for each genome-wide significant GWAS locus with others in default parameters considering whether each phenotype was a quantitative or binary trait. To calculate pairwise sharing between phenotypes, we compared the effect size and SE for all independent loci with MashR (*95*) data-driven covariance model. Upset (*96*) was used to visualize data.

Supplementary Materials

This PDF file includes:

Figs. S1 to S11
Tables S1 to S4, S7, S8, and S10
Legends for tables S5, S6, S9, and S11
Legend for list of members of BioBank Japan Cooperative Hospital Group
References

Other Supplementary Material for this manuscript includes the following:

Tables S5, S6, S9, and S11
List of members of BioBank Japan Cooperative Hospital Group

REFERENCES AND NOTES

1. J. Marchini, B. Howie, Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
2. Y. Li, C. Willer, S. Sanna, G. Abecasis, Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).
3. S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A. R. Wood, A. Teumer, H. M. Kang, C. Fuchsberger, P. Danecek, K. Sharp, Y. Luo, C. Sidore, A. Kwong, N. Timpson, S. Koskinen, S. Vrieze, L. J. Scott, H. Zhang, A. Mahajan, J. Veldink, U. Peters, C. Pato, C. M. van Duijn, C. E. Gillies, I. Gandin, M. Mezzavilla, A. Gilly, M. Cocca, M. Traglia, A. Angius, J. C. Barrett, D. Boomsma, K. Branham, G. Breen, C. M. Brummett, F. Busonero, H. Campbell, A. Chan, S. Chen, E. Chew, F. S. Collins, L. J. Corbin, G. D. Smith, G. Dedoussis, M. Dorr, A. E. Farmaki, L. Ferrucci, L. Forer, R. M. Fraser, S. Gabriel, S. Levy, L. Groop, T. Harrison, A. Hattersley, O. L. Holmen, K. Hveem, M. Kretzler, J. C. Lee, M. McGue, T. Meitinger, D. Melzer, J. L. Min, K. L. Mohlke, J. B. Vincent, M. Nauck, D. Nickerson, A. Palotie, M. Pato, N. Pirastu, M. McInnis, J. B. Richards, C. Sala, V. Salomaa, D. Schlessinger, S. Schoenherr, P. E. Slagboom, K. Small, T. Spector, D. Stambolian, M. Tuke, J. Tuomilehto, L. H. Van den Berg, W. Van Rheenen, U. Volker, C. Wijmenga, D. Toniolo, E. Zeggini, P. Gasparini, M. G. Sampson, J. F. Wilson, T. Frayling, P. I. de Bakker, M. A. Swertz, S. McCarroll, C. Kooperberg, A. Dekker, D. Altshuler, C. Willer, W. Iacono, S. Ripatti, N. Soranzo, K. Walter, A. Swaroop, F. Cucca, C. A. Anderson, R. M. Myers, M. Boehnke, M. I. McCarthy, R. Durbin; Haplotype Reference Consortium, A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
4. S. Das, G. R. Abecasis, B. L. Browning, Genotype imputation from large reference panels. *Annu. Rev. Genomics Hum. Genet.* **19**, 73–96 (2018).
5. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
6. J. Huang, B. Howie, S. McCarthy, Y. Memari, K. Walter, J. L. Min, P. Danecek, G. Malerba, E. Trabetti, H. F. Zheng; UK10K Consortium, G. Gambaro, J. B. Richards, R. Durbin, N. J. Timpson, J. Marchini, N. Soranzo, Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015).
7. O. Delaneau, J. F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
8. S. K. Yoo, C. U. Kim, H. L. Kim, S. Kim, J. Y. Shin, N. Kim, J. S. W. Yang, K. W. Lo, B. Cho, F. Matsuda, S. C. Schuster, C. Kim, J. I. Kim, J. S. Seo, NARD: Whole-genome reference panel of 1779 Northeast Asians improves imputation accuracy of rare and low-frequency variants. *Genome Med.* **11**, 64 (2019).
9. Genetics for all. *Nat. Genet.* **51**, 579 (2019).
10. Navigating 2020 and beyond. *Nat. Genet.* **52**, 1 (2020).
11. D. Gurdasani, T. Carstensen, S. Fatumo, G. Chen, C. S. Franklin, J. Prado-Martinez, H. Bouman, F. Abascal, M. Haber, I. Tachmazidou, I. Mathieson, K. Ekoru, M. K. DeGorter, R. N. Nsubuga, C. Finan, E. Wheeler, L. Chen, D. N. Cooper, S. Schiffels, Y. Chen, G. R. S. Ritchie, M. O. Pollard, M. D. Fortune, A. J. Mentzer, E. Garrison, A. Bergström, K. Hatzikotoulas, A. Adeyemo, A. Doumatey, H. Elding, L. V. Wain, G. Ehret, P. L. Auer, C. L. Kooperberg, A. P. Reiner, N. Franceschini, D. Maher, S. B. Montgomery, C. Kadie, C. Widmer, Y. Xue, J. Seeley, G. Asiki, A. Kamali, E. H. Young, C. Pomilla, N. Soranzo, E. Zeggini, F. Pirie, A. P. Morris, D. Heckerman, C. Tyler-Smith, A. A. Motala, C. Rotimi, P. Kaleebu, I. Barroso, M. S. Sandhu, Uganda genome resource enables insights into population history and genomic discovery in Africa. *Cell* **179**, 984–1002.e36 (2019).
12. D. Wu, J. Dou, X. Chai, C. Bellis, A. Wilm, C. C. Shih, W. W. J. Soon, N. Bertin, C. B. Lin, C. C. Khor, M. DeGiorgio, S. Cheng, L. Bao, N. Karnani, W. Y. K. Hwang, S. Davila, P. Tan, A. Shabbir, A. Moh, E. K. Tan, J. N. Foo, L. L. Goh, K. P. Leong, R. S. Y. Foo, C. S. P. Lam, A. M. Richards, C. Y. Cheng, T. Aung, T. Y. Wong, H. H. Ng; SG10K Consortium, J. Liu, C. Wang, Large-scale whole-genome sequencing of three diverse asian populations in Singapore. *Cell* **179**, 736–749.e15 (2019).
13. GenomeAsia100K Consortium, The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106–111 (2019).
14. D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S. B. Lee, X. Tian, B. L. Browning, S. Das, A. K. Emde, W. E. Clarke, D. P. Loesch, A. C. Shetty, T. W. Blackwell, A. V. Smith, Q. Wong, X. Liu, M. P. Conomos, D. M. Bobo, F. Aguet, C. Albert, A. Alonso, K. G. Ardlie, D. E. Arking, S. Aslibekyan, P. L. Auer, J. Barnard, R. G. Barr, L. Barwick, L. C. Becker, R. L. Beer, E. J. Benjamin, L. F. Bielak, J. Blangero, M. Boehnke, D. W. Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F. Casella, B. Chalazan, D. I. Chasman, Y. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A. Correa, J. E. Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K. Dutcher, P. T. Ellinor, L. S. Emery, C. Eng, D. Fatkin, T. Fingerlin, L. Forer, M. Fornage, N. Franceschini, C. Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E. Hall, J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson, R. Kaplan, S. L. R. Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A. Konkle, C. Kooperberg, A. Kottgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K. H. Lin, C. Liu, R. J. F. Loos, L. Garman, R. Gerszten, S. A. Lubitz, K. L. Lunetta, A. C. Y. Mak, A. Manichaikul, A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs, D. A. Meyers, J. L. Mikulla, M. A. Minear, B. D. Mitchell, S. Mohanty, M. E. Montasser, C. Montgomery, A. C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E. North, J. R. O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, J. Pleiness, W. S. Post, B. M. Psaty, D. C. Rao, S. Redline, A. P. Reiner, D. Roden, J. I. Rotter, I. Ruczinski, C. Sarnowski, S. Schoenherr, D. A. Schwartz, J. S. Seo, S. Seshadri, V. A. Sheehan, W. H. Sheu, M. B. Shoemaker, N. L. Smith, J. A. Smith, N. Sotoodehnia, A. M. Stilp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, R. P. Tracy, D. J. Van Den Berg, R. S. Vasan, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S. Weir, S. T. Weiss, L. C. Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-Koch, K. C. Barnes, E. Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E. K. Silverman, P. Qasba, W. Gan; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, G. J. Papanicolaou, D. A. Nickerson, S. R. Browning, M. C. Zody, S. Zollner, J. G. Wilson, L. A. Cupples, C. C. Laurie, C. E. Jaquish, R. D. Hernandez, T. D. O'Connor, G. R. Abecasis, Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
15. M. H. Kowalski, H. Qian, Z. Hou, J. D. Rosen, A. L. Tapia, Y. Shan, D. Jain, M. Argos, D. K. Arnett, C. Avery, K. C. Barnes, L. C. Becker, S. A. Bien, J. C. Bis, J. Blangero, E. Boerwinkle, D. W. Bowden, S. Buyske, J. Cai, M. H. Cho, S. H. Choi, H. Choquet, L. A. Cupples,

M. Cushman, M. Daya, P. S. de Vries, P. T. Ellinor, N. Faraday, M. Fornage, S. Gabriel, S. K. Ganesh, M. Graff, N. Gupta, J. He, S. R. Heckbert, B. Hidalgo, C. J. Hodonsky, M. R. Irvin, A. D. Johnson, E. Jorgenson, R. Kaplan, S. L. R. Kardia, T. N. Kelly, C. Kooperberg, J. A. Lasky-Su, R. J. F. Loos, S. A. Lubitz, R. A. Mathias, C. P. McHugh, C. Montgomery, J. Y. Moon, A. C. Morrison, N. D. Palmer, N. Pankratz, G. J. Papanicolaou, J. M. Peralta, P. A. Peyser, S. S. Rich, J. I. Rotter, E. K. Silverman, J. A. Smith, N. L. Smith, K. D. Taylor, T. A. Thornton, H. K. Tiwari, R. P. Tracy, T. Wang, S. T. Weiss, L. C. Weng, K. L. Wiggins, J. G. Wilson, L. R. Yanek, S. Zollner, K. E. North, P. L. Auer, N. T.-O. F. P. M. Consortium, T. O. Hematology, G. H. Working, L. M. Raffield, A. P. Reiner, Y. Li, Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLOS Genet.* **15**, e1008500 (2019).

16. T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, P. M. Visscher, Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
17. J. H. Park, M. H. Gail, C. R. Weinberg, R. J. Carroll, C. C. Chung, Z. Wang, S. J. Chanock, J. F. Fraumeni Jr., N. Chatterjee, Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18026–18031 (2011).
18. I. Surakka, K. Kristiansson, V. Anttila, M. Inouye, C. Barnes, L. Moutsianas, V. Salomaa, M. Daly, A. Palotie, L. Peltonen, S. Ripatti, Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging. *Genome Res.* **20**, 1344–1351 (2010).
19. M. Sakurai-Yageta, K. Kumada, C. Gocho, S. Makino, A. Uruno, S. Tadaka, I. N. Motoike, M. Kimura, S. Ito, A. Otsuki, A. Narita, H. Kudo, Y. Aoki, I. Danjoh, J. Yasuda, H. Kawame, N. Minegishi, S. Koshiba, N. Fuse, G. Tamiya, M. Yamamoto, K. Kinoshita, Japonica Array NEO with increased genome-wide coverage and abundant disease risk SNPs. *J. Biochem.* **170**, 399–410 (2021).
20. S. Tadaka, E. Hishinuma, S. Komaki, I. N. Motoike, J. Kawashima, D. Saigusa, J. Inoue, J. Takayama, Y. Okamura, Y. Aoki, M. Shirota, A. Otsuki, F. Katsuoka, A. Shimizu, G. Tamiya, S. Koshiba, M. Sasaki, M. Yamamoto, K. Kinoshita, jMorp updates in 2020: Large enhancement of multi-omics data resources on the general Japanese population. *Nucleic Acids Res.* **49**, D536–D544 (2021).
21. B. K. Bulik-Sullivan, P. R. Loh, H. K. Finucane, S. Ripke, J. Yang; Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
22. H. W. Cho, H. S. Jin, Y. B. Eom, A genome-wide association study of novel genetic variants associated with anthropometric traits in koreans. *Front Genet* **12**, 669215 (2021).
23. J. Zhang, L. B. McKenna, C. W. Bogue, K. H. Kaestner, The diabetes gene Hhex maintains delta-cell differentiation and islet function. *Genes Dev.* **28**, 829–834 (2014).
24. M. Kanai, M. Akiyama, A. Takahashi, N. Matoba, Y. Momozawa, M. Ikeda, N. Iwata, S. Ikegawa, M. Hirata, K. Matsuda, M. Kubo, Y. Okada, Y. Kamatani, Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
25. J. Chen, C. N. Spracklen, G. Marenne, A. Varshney, L. J. Corbin, J. Luan, S. M. Willems, Y. Wu, X. Zhang, M. Horikoshi, T. S. Boutin, R. Magi, J. Waage, R. Li-Gao, K. H. K. Chan, J. Yao, M. D. Anasanti, A. Y. Chu, A. Claringbould, J. Heikkinen, J. Hong, J. J. Hottenga, S. Huo, M. A. Kaakinen, T. Louie, W. Marz, H. Moreno-Macias, A. Ndungu, S. C. Nelson, I. M. Nolte, K. E. North, C. K. Raulerson, D. Ray, R. Rohde, D. Rybin, C. Schurmann, X. Sim, L. Southam, I. D. Stewart, C. A. Wang, Y. Wang, P. Wu, W. Zhang, T. S. Ahluwalia, E. V. R. Appel, L. F. Bielak, J. A. Brody, N. P. Burtt, C. P. Cabrera, B. E. Cade, J. F. Chai, X. Chai, L. C. Chang, C. H. Chen, B. H. Chen, K. N. Chitrala, Y. F. Chiu, H. G. de Haan, G. E. Delgado, A. Demirkan, Q. Duan, J. Engmann, S. A. Fatumo, J. Gayan, F. Giulianini, J. H. Gong, S. Gustafsson, Y. Hai, F. P. Hartwig, J. He, Y. Heianza, T. Huang, A. Huerta-Chagoya, M. Y. Hwang, R. A. Jensen, T. Kawaguchi, K. A. Kentistou, Y. J. Kim, M. E. Kleber, I. K. Kooner, S. Lai, L. A. Lange, C. D. Langefeld, M. Lauzon, M. Li, S. Ligthart, J. Liu, M. Loh, J. Long, V. Lyssenko, M. Mangino, C. Marzi, M. E. Montasser, A. Nag, M. Nakatochi, D. Noce, R. Noordam, G. Pistis, M. Preuss, L. Raffield, L. J. Rasmussen-Torvik, S. S. Rich, N. R. Robertson, R. Rueedi, K. Ryan, S. Sanna, R. Saxena, K. E. Schraut, B. Sennblad, K. Setoh, A. V. Smith, T. Sparso, R. J. Strawbridge, F. Takeuchi, J. Tan, S. Trompet, E. van den Akker, P. J. van der Most, N. Verweij, M. Vogel, H. Wang, C. Wang, N. Wang, H. R. Warren, W. Wen, T. Wilsgaard, A. Wong, A. R. Wood, T. Xie, M. H. Zafarmand, J. H. Zhao, W. Zhao, N. Amin, Z. Arzumanyan, A. Astrup, S. J. L. Bakker, D. Baldassarre, M. Beekman, R. N. Bergman, A. Bertoni, M. Bluher, L. L. Bonnycastle, S. R. Bornstein, D. W. Bowden, Q. Cai, A. Campbell, H. Campbell, Y. C. Chang, E. J. C. de Geus, A. Dehghan, S. Du, G. Eiriksdottir, A. E. Farmaki, M. Franberg, C. Fuchsberger, Y. Gao, A. P. Gjesing, A. Goel, S. Han, C. A. Hartman, C. Herder, A. A. Hicks, C. H. Hsieh, W. A. Hsueh, S. Ichihara, M. Igase, M. A. Ikram, W. C. Johnson, M. E. Jorgensen, P. K. Joshi, R. R. Kalyani, F. R. Kandeel, T. Katsuya, C. C. Khor, W. Kiess, I. Kolcic, T. Kuulasmaa, J. Kuusisto, K. Lall, K. Lam, D. A. Lawlor, N. R. Lee, R. N. Lemaitre, H. Li, S. L. Cohort, S. Y. Lin, J. Lindstrom, A. Linneberg, J. Liu, C. Lorenzo, T. Matsubara, F. Matsuda, G. Mingrone, S. Mooijaart, S. Moon, T. Nabika, G. N. Nadkarni, J. L. Nadler, M. Nelis, M. J. Neville, J. M. Norris, Y. Ohyagi, A. Peters, P. A. Peyser, O. Polasek, Q. Qi, D. Raven, D. F. Reilly, A. Reiner, F. Rivideneira, K. Roll, I. Rudan, C. Sabanayagam, K. Sandow, N. Sattar, A. Schurmann, J. Shi, H. M. Stringham, K. D. Taylor, T. M. Teslovich, B. Thuesen, P. Timmers, E. Tremoli, M. Y. Tsai, A. Uitterlinden, R. M. van Dam, D. van Heemst, A. van Hylckama Vlieg, J. V. van Vliet-Ostaptchouk, J. Vangipurapu, H. Vestergaard, T. Wang, K. W. van Dijk, T. Zemunik, G. R. Abecasis, L. S. Adair, C. A. Aguilar-Salinas, M. E. Alarcon-Riquelme, P. An, L. Aviles-Santa, D. M. Becker, L. J. Beilin, S. Bergmann, H. Bisgaard, C. Black, M. Boehnke, E. Boerwinkle, B. O. Bohm, K. Bonnelykke, D. I. Boomsma, E. P. Bottinger, T. A. Buchanan, M. Canouil, M. J. Caulfield, J. C. Chambers, D. I. Chasman, Y. I. Chen, C. Y. Cheng, F. S. Collins, A. Correa, F. Cucca, H. J. de Silva, G. Dedoussis, S. Elmstahl, M. K. Evans, E. Ferrannini, L. Ferrucci, J. C. Florez, P. W. Franks, T. M. Frayling, P. Froguel, B. Gigante, M. O. Goodarzi, P. Gordon-Larsen, H. Grallert, N. Grarup, S. Grimsgaard, L. Groop, V. Gudnason, X. Guo, A. Hamsten, T. Hansen, C. Hayward, S. R. Heckbert, B. L. Horta, W. Huang, E. Ingelsson, P. S. James, M. R. Jarvelin, J. B. Jonas, J. W. Jukema, P. Kaleebu, R. Kaplan, S. L. R. Kardia, N. Kato, S. M. Keinanen-Kiukaanniemi, B. J. Kim, M. Kivimaki, H. A. Koistinen, J. S. Kooner, A. Korner, P. Kovacs, D. Kuh, M. Kumari, Z. Kutalik, M. Laakso, T. A. Lakka, L. J. Launer, K. Leander, H. Li, X. Lin, L. Lind, C. Lindgren, S. Liu, R. J. F. Loos, P. K. E. Magnusson, A. Mahajan, A. Metspalu, D. O. Mook-Kanamori, T. A. Mori, P. B. Munroe, I. Njolstad, J. R. O'Connell, A. J. Oldehinkel, K. K. Ong, S. Padmanabhan, C. N. A. Palmer, N. D. Palmer, O. Pedersen, C. E. Pennell, D. J. Porteous, P. P. Pramstaller, M. A. Province, B. M. Psaty, L. Qi, L. J. Raffel, R. Rauramaa, S. Redline, P. M. Ridker, F. R. Rosendaal, T. E. Saaristo, M. Sandhu, J. Saramies, N. Schneiderman, P. Schwarz, L. J. Scott, E. Selvin, P. Sever, X. O. Shu, P. E. Slagboom, K. S. Small, B. H. Smith, H. Snieder, T. Sofer, T. I. A. Sorensen, T. D. Spector, A. Stanton, C. J. Steves, M. Stumvoll, L. Sun, Y. Tabara, E. S. Tai, N. J. Timpson, A. Tonjes, J. Tuomilehto, T. Tusie, M. Uusitupa, P. van der Harst, C. van Duijn, V. Vitart, P. Vollenweider, T. G. M. Vrijkotte, L. E. Wagenknecht, M. Walker, Y. X. Wang, N. J. Wareham, R. M. Watanabe, H. Watkins, W. B. Wei, A. R. Wickremasinghe, G. Willemsen, J. F. Wilson, T. Y. Wong, J. Y. Wu, A. H. Xiang, L. R. Yanek, L. Yengo, M. Yokota, E. Zeggini, W. Zheng, A. B. Zonderman, J. I. Rotter, A. L. Gloyn, M. I. McCarthy, J. Dupuis, J. B. Meigs, R. A. Scott, I. Prokopenko, A. Leong, C. T. Liu, S. C. J. Parker, K. L. Mohlke, C. Langenberg, E. Wheeler, A. P. Morris, I. Barroso; The Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC), The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).
26. P. Akbari, A. Gilani, O. Sosina, J. A. Kosmicki, L. Khrimian, Y. Y. Fang, T. Persaud, V. Garcia, D. Sun, A. Li, J. Mbatchou, A. E. Locke, C. Benner, N. Verweij, N. Lin, S. Hossain, K. Agostinucci, J. V. Pascale, E. Dirice, M. Dunn, C. Regeneron Genetics, E. H. R. C. Discov, W. E. Kraus, S. H. Shah, Y. I. Chen, J. I. Rotter, D. J. Rader, O. Melander, C. D. Still, T. Mirshahi, D. J. Carey, J. Berumen-Campos, P. Kuri-Morales, J. Alegre-Diaz, J. M. Torres, J. R. Emberson, R. Collins, S. Balasubramanian, A. Hawes, M. Jones, B. Zambrowicz, A. J. Murphy, C. Paulding, G. Coppola, J. D. Overton, J. G. Reid, A. R. Shuldiner, M. Cantor, H. M. Kang, G. R. Abecasis, K. Karalis, A. N. Economides, J. Marchini, G. D. Yancopoulos, M. W. Sleeman, J. Altarejos, G. Della Gatta, R. Tapia-Conyer, M. L. Schwartzman, A. Baras, M. A. R. Ferreira, L. A. Lotta, Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity. *Science* **373**, eabf8683 (2021).
27. L. Pottie, C. S. Adamo, A. Beyens, S. Lutke, P. Tapaneeyaphan, A. De Clercq, P. L. Salmon, R. De Rycke, A. Gezdirici, E. Y. Gulec, N. Khan, J. E. Urquhart, W. G. Newman, K. Metcalfe, S. Efthymiou, R. Maroofian, N. Anwar, S. Maqbool, F. Rahman, I. Altweijri, M. Alsaleh, S. M. Abdullah, M. Al-Owain, M. Hashem, H. Houlden, F. S. Alkuraya, P. Sips, G. Sengle, B. Callewaert, Bi-allelic premature truncating variants in LTBP1 cause cutis laxa syndrome. *Am. J. Hum. Genet.* **108**, 1095–1114 (2021).
28. J. M. Fernandez-Real, J. M. Mercader, F. J. Ortega, J. M. Moreno-Navarrete, P. Lopez-Romero, W. Ricart, Transferrin receptor-1 gene polymorphisms are associated with type 2 diabetes. *Eur. J. Clin. Invest* **40**, 600–607 (2010).
29. J. M. Fernandez-Real, A. Lopez-Bermejo, W. Ricart, Cross-talk between iron metabolism and diabetes. *Diabetes* **51**, 2348–2354 (2002).
30. M. Akiyama, K. Ishigaki, S. Sakaue, Y. Momozawa, M. Horikoshi, M. Hirata, K. Matsuda, S. Ikegawa, A. Takahashi, M. Kanai, S. Suzuki, D. Matsui, M. Naito, T. Yamaji, M. Iwasaki, N. Sawada, K. Tanno, M. Sasaki, A. Hozawa, N. Minegishi, K. Wakai, S. Tsugane, A. Shimizu, M. Yamamoto, Y. Okada, Y. Murakami, M. Kubo, Y. Kamatani, Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.* **10**, 4393 (2019).
31. F. Kamada, Y. Aoki, A. Narisawa, Y. Abe, S. Komatsuzaki, A. Kikuchi, J. Kanno, T. Niihori, M. Ono, N. Ishii, Y. Owada, M. Fujimura, Y. Mashimo, Y. Suzuki, A. Hata, S. Tsuchiya, T. Tominaga, Y. Matsubara, S. Kure, A genome-wide association study identifies RNF213 as the first Moyamoya disease gene. *J. Hum. Genet.* **56**, 34–40 (2011).
32. J. E. Campbell, Targeting the GIPR for obesity: To agonize or antagonize? Potential mechanisms. *Mol. Metab.* **46**, 101139 (2021).

33. A. A. Butler, The melanocortin system and energy balance. *Peptides* **27**, 281–290 (2006).
34. K. C. R. Salum, G. O. de Souza, G. M. Abreu, M. Campos Junior, F. B. Kohlrausch, J. R. I. Carneiro, J. F. Nogueira Neto, F. Magno, E. L. Rosado, L. Palhinha, C. M. Maya-Monteiro, G. M. K. de Cabello, P. H. Cabello, P. T. Bozza, V. M. Zembrzuski, A. C. P. da Fonseca, Identification of a rare and potential pathogenic MC4R variant in a brazilian patient with adulthood-onset severe obesity. *Front Genet.* **11**, 608840 (2020).
35. G. Wang, A. Sarkar, P. Carbonetto, M. Stephens, A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B (Stat. Methodol.)* **82**, 1273–1300 (2020).
36. N. Sulaiman, M. Yaseen Hachim, A. Khalique, A. K. Mohammed, S. Al Heialy, J. Taneera, EXOC6 (Exocyst Complex Component 6) is associated with the risk of type 2 diabetes and pancreatic beta-cell dysfunction. *Biology (Basel)* **11**, 388 (2022).
37. E. N. Fazio, M. Everest, R. Colman, R. Wang, C. L. Pin, Altered Glut-2 accumulation and β-cell function in mice lacking the exocrine-specific transcription factor, Mist1. *J. Endocrinol.* **187**, 407–418 (2005).
38. X. Ding, R. Iyer, C. Novotny, D. Metzger, H. H. Zhou, G. I. Smith, M. Yoshino, J. Yoshino, S. Klein, G. Swaminath, S. Talukdar, Y. Zhou, Inhibition of Grb14, a negative modulator of insulin signaling, improves glucose homeostasis without causing cardiac dysfunction. *Sci. Rep.* **10**, 3417 (2020).
39. R. Starr, D. J. Hilton, Negative regulation of the JAK/STAT pathway. *Bioessays* **21**, 47–52 (1999).
40. M. E. Miller, C. Z. Michaylira, J. G. Simmons, D. M. Ney, E. M. Dahly, J. K. Heath, P. K. Lund, Suppressor of cytokine signaling-2: A growth hormone-inducible inhibitor of intestinal epithelial cell proliferation. *Gastroenterology* **127**, 570–581 (2004).
41. A. Al-Araimi, A. Al Kharusi, A. Bani Oraba, M. M. Al-Maney, S. Al Sinawi, I. Al-Haddabi, F. Zadjali, Deletion of SOCS2 reduces post-colitis fibrosis via alteration of the TGFβ pathway. *Int. J. Mol. Sci.* **21**, 3073 (2020).
42. D. Metcalf, C. J. Greenhalgh, E. Viney, T. A. Willson, R. Starr, N. A. Nicola, D. J. Hilton, W. S. Alexander, Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature* **405**, 1069–1073 (2000).
43. A. B. Popejoy, S. M. Fullerton, Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
44. V. Tam, N. Patel, M. Turcotte, Y. Bosse, G. Pare, D. Meyre, Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).
45. S. W. Im, J. Chae, H. Y. Son, B. Cho, J. I. Kim, J. H. Park, A population-specific low-frequency variant of SLC22A12 (p.W258*) explains nearby genome-wide association signals for serum uric acid concentrations among Koreans. *PLOS ONE* **15**, e0231336 (2020).
46. E. Marouli, M. Graff, C. Medina-Gomez, K. S. Lo, A. R. Wood, T. R. Kjaer, R. S. Fine, Y. Lu, C. Schurmann, H. M. Highland, S. Rueger, G. Thorleifsson, A. E. Justice, D. Lamparter, K. E. Stirrups, V. Turcot, K. L. Young, T. W. Winkler, T. Esko, T. Karaderi, A. E. Locke, N. G. Masca, M. C. Ng, P. Mudgal, M. A. Rivas, S. Vedantam, A. Mahajan, X. Guo, G. Abecasis, K. K. Aben, L. S. Adair, D. S. Alam, E. Albrecht, K. H. Allin, M. Allison, P. Amouyel, E. V. Appel, D. Arveiler, F. W. Asselbergs, P. L. Auer, B. Balkau, B. Banas, L. E. Bang, M. Benn, S. Bergmann, L. F. Bielak, M. Bluher, H. Boeing, E. Boerwinkle, C. A. Boger, L. L. Bonnycastle, J. Bork-Jensen, M. L. Bots, E. P. Bottinger, D. W. Bowden, I. Brandslund, G. Breen, M. H. Brilliant, L. Broer, A. A. Burt, A. S. Butterworth, D. J. Carey, M. J. Caulfield, J. C. Chambers, D. I. Chasman, Y. I. Chen, R. Chowdhury, C. Christensen, A. Y. Chu, M. Cocca, F. S. Collins, J. P. Cook, J. Corley, J. C. Galbany, A. J. Cox, G. Cuellar-Partida, J. Danesh, G. Davies, P. I. de Bakker, G. J. de Borst, S. de Denus, M. C. de Groot, R. de Mutsert, I. J. Deary, G. Dedoussis, E. W. Demerath, A. I. den Hollander, J. G. Dennis, E. Di Angelantonio, F. Drenos, M. Du, A. M. Dunning, D. F. Easton, T. Ebeling, T. L. Edwards, P. T. Ellinor, P. Elliott, E. Evangelou, A. E. Farmaki, J. D. Faul, M. F. Feitosa, S. Feng, E. Ferrannini, M. M. Ferrario, J. Ferrieres, J. C. Florez, I. Ford, M. Fornage, P. W. Franks, R. Frikke-Schmidt, T. E. Galesloot, W. Gan, I. Gandin, P. Gasparini, V. Giedraitis, A. Giri, G. Girotto, S. D. Gordon, P. Gordon-Larsen, M. Gorski, N. Grarup, M. L. Grove, V. Gudnason, S. Gustafsson, T. Hansen, K. M. Harris, T. B. Harris, A. T. Hattersley, C. Hayward, L. He, I. M. Heid, K. Heikkila, O. Helgeland, J. Hernesniemi, A. W. Hewitt, L. J. Hocking, M. Hollensted, O. L. Holmen, G. K. Hovingh, J. M. Howson, C. B. Hoyng, P. L. Huang, K. Hveem, M. A. Ikram, E. Ingelsson, A. U. Jackson, J. H. Jansson, G. P. Jarvik, G. B. Jensen, M. A. Jhun, Y. Jia, X. Jiang, S. Johansson, M. E. Jorgensen, T. Jorgensen, P. Jousilahti, J. W. Jukema, B. Kahali, R. S. Kahn, M. Kahonen, P. R. Kamstrup, S. Kanoni, J. Kaprio, M. Karaleftheri, S. L. Kardia, F. Karpe, F. Kee, R. Keeman, L. A. Kiemeney, H. Kitajima, K. B. Kluivers, T. Kocher, P. Komulainen, J. Kontto, J. S. Kooner, C. Kooperberg, P. Kovacs, J. Kriebel, H. Kuivaniemi, S. Kury, J. Kuusisto, M. La Bianca, M. Laakso, T. A. Lakka, E. M. Lange, L. A. Lange, C. D. Langefeld, C. Langenberg, E. B. Larson, I. T. Lee, T. Lehtimaki, C. E. Lewis, H. Li, J. Li, R. Li-Gao, H. Lin, L. A. Lin, X. Lin, L. Lind, J. Lindstrom, A. Linneberg, Y. Liu, Y. Liu, A. Lophatananon, J. Luan, S. A. Lubitz, L. P. Lyytikainen, D. A. Mackey, P. A. Madden, A. K. Manning, S. Mannisto, G. Marenne, J. Marten, N. G. Martin, A. L. Mazul, K. Meidtner, A. Metspalu, P. Mitchell, K. L. Mohlke, D. O. Mook-Kanamori, A. Morgan, A. D. Morris, A. P. Morris, M. Muller-Nurasyid, P. B. Munroe, M. A. Nalls, M. Nauck, C. P. Nelson, M. Neville, S. F. Nielsen, K. Nikus, P. R. Njolstad, B. G. Nordestgaard, I. Ntalla, J. R. O'Connel, H. Oksa, L. M. Loohuis, R. A. Ophoff, K. R. Owen, C. J. Packard, S. Padmanabhan, C. N. Palmer, G. Pasterkamp, A. P. Patel, A. Pattie, O. Pedersen, P. L. Peissig, G. M. Peloso, C. E. Pennell, M. Perola, J. A. Perry, J. R. Perry, T. N. Person, A. Pirie, O. Polasek, D. Posthuma, O. T. Raitakari, A. Rasheed, R. Rauramaa, D. F. Reilly, A. P. Reiner, F. Renstrom, P. M. Ridker, J. D. Rioux, N. Robertson, A. Robino, O. Rolandsson, I. Rudan, K. S. Ruth, D. Saleheen, V. Salomaa, N. J. Samani, K. Sandow, Y. Sapkota, N. Sattar, M. K. Schmidt, P. J. Schreiner, M. B. Schulze, R. A. Scott, M. P. Segura-Lepe, S. Shah, X. Sim, S. Sivapalaratnam, K. S. Small, A. V. Smith, J. A. Smith, L. Southam, T. D. Spector, E. K. Speliotes, J. M. Starr, V. Steinthorsdottir, H. M. Stringham, M. Stumvoll, P. Surendran, L. M. 't Hart, K. E. Tansey, J. C. Tardif, K. D. Taylor, A. Teumer, D. J. Thompson, U. Thorsteinsdottir, B. H. Thuesen, A. Tonjes, G. Tromp, S. Trompet, E. Tsafantakis, J. Tuomilehto, A. Tybjaerg-Hansen, J. P. Tyrer, R. Uher, A. G. Uitterlinden, S. Ulivi, S. W. van der Laan, A. R. Van Der Leij, C. M. van Duijn, N. M. van Schoor, J. van Setten, A. Varbo, T. V. Varga, R. Varma, D. R. Edwards, S. H. Vermeulen, H. Vestergaard, V. Vitart, T. F. Vogt, D. Vozzi, M. Walker, F. Wang, C. A. Wang, S. Wang, Y. Wang, N. J. Wareham, H. R. Warren, J. Wessel, S. M. Willems, J. G. Wilson, D. R. Witte, M. O. Woods, Y. Wu, H. Yaghootkar, J. Yao, P. Yao, L. M. Yerges-Armstrong, R. Young, E. Zeggini, X. Zhan, W. Zhang, J. H. Zhao, W. Zhao, W. Zhao, H. Zheng, W. Zhou; EPIC-InterAct Consortium; CHD Exome+ Consortium; ExomeBP Consortium; T2D-Genes Consortium; GoT2D Genes Consortium; Global Lipids Genetics Consortium; ReproGen Consortium; MAGIC Investigators; J. I. Rotter, M. Boehnke, S. Kathiresan, M. I. McCarthy, C. J. Willer, K. Stefansson, I. B. Borecki, D. J. Liu, K. E. North, N. L. Heard-Costa, T. H. Pers, C. M. Lindgren, C. Oxvig, Z. Kutalik, F. Rivadeneira, R. J. Loos, T. M. Frayling, J. N. Hirschhorn, P. Deloukas, G. Lettre, Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
47. U. Styrkarsdottir, G. Thorleifsson, P. Sulem, D. F. Gudbjartsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, A. Oddsson, A. Helgason, O. T. Magnusson, G. B. Walters, M. L. Frigge, H. T. Helgadottir, H. Johannsdottir, K. Bergsteinsdottir, M. H. Ogmundsdottir, J. R. Center, T. V. Nguyen, J. A. Eisman, C. Christiansen, E. Steingrimsson, J. G. Jonasson, L. Tryggvadottir, G. I. Eyjolfsson, A. Theodors, T. Jonsson, T. Ingvarsson, I. Olafsson, T. Rafnar, A. Kong, G. Sigurdsson, G. Masson, U. Thorsteinsdottir, K. Stefansson, Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **497**, 517–520 (2013).
48. E. M. van Leeuwen, L. C. Karssen, J. Deelen, A. Isaacs, C. Medina-Gomez, H. Mbarek, A. Kanterakis, S. Trompet, I. Postmus, N. Verweij, D. J. van Enckevort, J. E. Huffman, C. C. White, M. F. Feitosa, T. M. Bartz, A. Manichaikul, P. K. Joshi, G. M. Peloso, P. Deelen, F. van Dijk, G. Willemsen, E. J. de Geus, Y. Milaneschi, B. W. Penninx, L. C. Francioli, A. Menelaou, S. L. Pulit, F. Rivadeneira, A. Hofman, B. A. Oostra, O. H. Franco, I. Mateo Leach, M. Beekman, A. J. de Craen, H. W. Uh, H. Trochet, L. J. Hocking, D. J. Porteous, N. Sattar, C. J. Packard, B. M. Buckley, J. A. Brody, J. C. Bis, J. I. Rotter, J. C. Mychaleckyj, H. Campbell, Q. Duan, L. A. Lange, J. F. Wilson, C. Hayward, O. Polasek, V. Vitart, I. Rudan, A. F. Wright, S. S. Rich, B. M. Psaty, I. B. Borecki, P. M. Kearney, D. J. Stott, L. A. Cupples; Genome of The Netherlands Consortium, J. W. Jukema, P. van der Harst, E. J. Sijbrands, J. J. Hottenga, A. G. Uitterlinden, M. A. Swertz, G. J. van Ommen, P. I. de Bakker, P. E. Slagboom, D. I. Boomsma, C. Wijmenga, C. M. van Duijn, Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nat Commun* **6**, 6065 (2015).
49. M. S. Sabatine, R. P. Giugliano, A. C. Keech, N. Honarpour, S. D. Wiviott, S. A. Murphy, J. F. Kuder, H. Wang, T. Liu, S. M. Wasserman, P. S. Sever, T. R. Pedersen; FOURIER Steering Committee and Investigators, Evolocumab and clinical outcomes in patients with cardiovascular disease. *N. Engl. J. Med.* **376**, 1713–1722 (2017).
50. S. Nurk, S. Koren, A. Rhie, M. Rautiainen, A. V. Bzikadze, A. Mikheenko, M. R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S. J. Hoyt, M. Diekhans, G. A. Logsdon, M. Alonge, S. E. Antonarakis, M. Borchers, G. G. Bouffard, S. Y. Brooks, G. V. Caldas, N. C. Chen, H. Cheng, C. S. Chin, W. Chow, L. G. de Lima, P. C. Dishuck, R. Durbin, T. Dvorkina, I. T. Fiddes, G. Formenti, R. S. Fulton, A. Fungtammasan, E. Garrison, P. G. S. Grady, T. A. Graves-Lindsay, I. M. Hall, N. F. Hansen, G. A. Hartley, M. Haukness, K. Howe, M. W. Hunkapiller, C. Jain, M. Jain, E. D. Jarvis, P. Kerpedjiev, M. Kirsche, M. Kolmogorov, J. Korlach, M. Kremitzki, H. Li, V. V. Maduro, T. Marschall, A. M. McCartney, J. McDaniel, D. E. Miller, J. C. Mullikin, E. W. Myers, N. D. Olson, B. Paten, P. Peluso, P. A. Pevzner, D. Porubsky, T. Potapova, E. I. Rogaev, J. A. Rosenfeld, S. L. Salzberg, V. A. Schneider, F. J. Sedlazeck, K. Shafin, C. J. Shew, A. Shumate, Y. Sims, A. F. A. Smit, D. C. Soto, I. Sovic, J. M. Storer, A. Streets, B. A. Sullivan, F. Thibaud-Nissen, J. Torrance, J. Wagner, B. P. Walenz, A. Wenger, J. M. D. Wood, C. Xiao, S. M. Yan, A. C. Young, S. Zarate, U. Surti, R. C. McCoy, M. Y. Dennis, I. A. Alexandrov, J. L. Gerton, R. J. O'Neill, W. Timp, J. M. Zook, M. C. Schatz, E. E. Eichler, K. H. Miga, A. M. Phillippy, The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
51. T. Wang, L. Antonacci-Fulton, K. Howe, H. A. Lawson, J. K. Lucas, A. M. Phillippy, A. B. Popejoy, M. Asri, C. Carson, M. J. P. Chaisson, X. Chang, R. Cook-Deegan, A. L. Felsenfeld, R. S. Fulton, E. P. Garrison, N. A. Garrison, T. A. Graves-Lindsay, H. Ji, E. E. Kenny, B. A. Koenig, D. Li, T. Marschall, J. F. McMichael, A. M. Novak, D. Purushotham,

V. A. Schneider, B. I. Schultz, M. W. Smith, H. J. Sofia, T. Weissman, P. Flicek, H. Li, K. H. Miga, B. Paten, E. D. Jarvis, I. M. Hall, E. E. Eichler, D. Haussler; the Human Pangenome Reference Consortium, The human pangenome project: A global resource to map genomic diversity. *Nature* **604**, 437–446 (2022).

52. L. Yengo, S. Vedantam, E. Marouli, J. Sidorenko, E. Bartell, S. Sakaue, M. Graff, A. U. Eliasen, Y. Jiang, S. Raghavan, J. Miao, J. D. Arias, S. E. Graham, R. E. Mukamel, C. N. Spracklen, X. Yin, S. H. Chen, T. Ferreira, H. H. Highland, Y. Ji, T. Karaderi, K. Lin, K. Lull, D. E. Malden, C. Medina-Gomez, M. Machado, A. Moore, S. Rueger, X. Sim, S. Vrieze, T. S. Ahluwalia, M. Akiyama, M. A. Allison, M. Alvarez, M. K. Andersen, A. Ani, V. Appadurai, L. Arbeeva, S. Bhaskar, L. F. Bielak, S. Bollepalli, L. L. Bonnycastle, J. Bork-Jensen, J. P. Bradfield, Y. Bradford, P. S. Braund, J. A. Brody, K. S. Burgdorf, B. E. Cade, H. Cai, Q. Cai, A. Campbell, M. Canadas-Garre, E. Catamo, J. F. Chai, X. Chai, L. C. Chang, Y. C. Chang, C. H. Chen, A. Chesi, S. H. Choi, R. H. Chung, M. Cocca, M. P. Concas, C. Couture, G. Cuellar-Partida, R. Danning, E. W. Daw, F. Degenhard, G. E. Delgado, A. Delitala, A. Demirkan, X. Deng, P. Devineni, A. Dietl, M. Dimitriou, L. Dimitrov, R. Dorajoo, A. B. Ekici, J. E. Engmann, Z. Fairhurst-Hunter, A. E. Farmaki, J. D. Faul, J. C. Fernandez-Lopez, L. Forer, M. Francescatto, S. Freitag-Wolf, C. Fuchsberger, T. E. Galesloot, Y. Gao, Z. Gao, F. Geller, O. Giannakopoulou, F. Giulianini, A. P. Gjesing, A. Goel, S. D. Gordon, M. Gorski, J. Grove, X. Guo, S. Gustafsson, J. Haessler, T. F. Hansen, A. S. Havulinna, S. J. Haworth, J. He, N. Heard-Costa, P. Hebbar, G. Hindy, Y. A. Ho, E. Hofer, E. Holliday, K. Horn, W. E. Hornsby, J. J. Hottenga, H. Huang, J. Huang, A. Huerta-Chagoya, J. E. Huffman, Y. J. Hung, S. Huo, M. Y. Hwang, H. Iha, D. D. Ikeda, M. Isono, A. U. Jackson, S. Jager, I. E. Jansen, I. Johansson, J. B. Jonas, A. Jonsson, T. Jorgensen, I. P. Kalafati, M. Kanai, S. Kanoni, L. L. Karhus, A. Kasturiratne, T. Katsuya, T. Kawaguchi, R. L. Kember, K. A. Kentistou, H. N. Kim, Y. J. Kim, M. E. Kleber, M. J. Knol, A. Kurbasic, M. Lauzon, P. Le, R. Lea, J. Y. Lee, H. L. Leonard, S. A. Li, X. Li, X. Li, J. Liang, H. Lin, S. Y. Lin, J. Liu, X. Liu, K. S. Lo, J. Long, L. Lores-Motta, J. Luan, V. Lyssenko, L. P. Lyytikainen, A. Mahajan, V. Mamakou, M. Mangino, A. Manichaikul, J. Marten, M. Mattheisen, L. Mavarani, A. F. McDaid, K. Meidtner, T. L. Melendez, J. M. Mercader, Y. Milaneschi, J. E. Miller, I. Y. Millwood, P. P. Mishra, R. E. Mitchell, L. T. Mollehave, A. Morgan, S. Mucha, M. Munz, M. Nakatochi, C. P. Nelson, M. Nethander, C. W. Nho, A. A. Nielsen, I. M. Nolte, S. S. Nongmaithem, R. Noordam, I. Ntalla, T. Nutile, A. Pandit, P. Christofidou, K. Parna, M. Pauper, E. R. B. Petersen, L. V. Petersen, N. Pitkanen, O. Polasek, A. Poveda, M. H. Preuss, S. Pyarajan, L. M. Raffield, H. Rakugi, J. Ramirez, A. Rasheed, D. Raven, N. W. Rayner, C. Riveros, R. Rohde, D. Ruggiero, S. E. Ruotsalainen, K. A. Ryan, M. Sabater-Lleal, R. Saxena, M. Scholz, A. Sendamarai, B. Shen, J. Shi, J. H. Shin, C. Sidore, C. M. Sitlani, R. C. Slieker, R. A. J. Smit, A. V. Smith, J. A. Smith, L. J. Smyth, L. Southam, V. Steinthorsdottir, L. Sun, F. Takeuchi, D. S. P. Tallapragada, K. D. Taylor, B. O. Tayo, C. Tcheandjieu, N. Terzikhan, P. Tesolin, A. Teumer, E. Theusch, D. J. Thompson, G. Thorleifsson, P. Timmers, S. Trompet, C. Turman, S. Vaccargiu, S. W. van der Laan, P. J. van der Most, J. B. van Klinken, J. van Setten, S. S. Verma, N. Verweij, Y. Veturi, C. A. Wang, C. Wang, L. Wang, Z. Wang, H. R. Warren, W. B. Wei, A. R. Wickremasinghe, M. Wielscher, K. L. Wiggins, B. S. Winsvold, A. Wong, Y. Wu, M. Wuttke, R. Xia, T. Xie, K. Yamamoto, J. Yang, J. Yao, H. Young, N. A. Yousri, L. Yu, L. Zeng, W. Zhang, X. Zhang, J. H. Zhao, W. Zhao, W. Zhou, M. E. Zimmermann, M. Zoledziewska, L. S. Adair, H. H. H. Adams, C. A. Aguilar-Salinas, F. Al-Mulla, D. K. Arnett, F. W. Asselbergs, B. O. Asvold, J. Attia, B. Banas, S. Bandinelli, D. A. Bennett, T. Bergler, D. Bharadwaj, G. Biino, H. Bisgaard, E. Boerwinkle, C. A. Boger, K. Bonnelykke, D. I. Boomsma, A. D. Borglum, J. B. Borja, C. Bouchard, D. W. Bowden, I. Brandslund, B. Brumpton, J. E. Buring, M. J. Caulfield, J. C. Chambers, G. R. Chandak, S. J. Chanock, N. Chaturvedi, Y. I. Chen, Z. Chen, C. Y. Cheng, I. E. Christophersen, M. Ciullo, J. W. Cole, F. S. Collins, R. S. Cooper, M. Cruz, F. Cucca, L. A. Cupples, M. J. Cutler, S. M. Damrauer, T. M. Dantoft, G. J. de Borst, L. de Groot, P. L. De Jager, D. P. V. de Kleijn, H. J. de Silva, G. V. Dedoussis, A. I. den Hollander, S. Du, D. F. Easton, P. J. M. Elders, A. H. Eliassen, P. T. Ellinor, S. Elmstahl, J. Erdmann, M. K. Evans, D. Fatkin, B. Feenstra, M. F. Feitosa, L. Ferrucci, I. Ford, M. Fornage, A. Franke, P. W. Franks, B. I. Freedman, P. Gasparini, C. Gieger, G. Girotto, M. E. Goddard, Y. M. Golightly, C. Gonzalez-Villalpando, P. Gordon-Larsen, H. Grallert, S. F. A. Grant, N. Grarup, L. Griffiths, V. Gudnason, C. Haiman, H. Hakonarson, T. Hansen, C. A. Hartman, A. T. Hattersley, C. Hayward, S. R. Heckbert, C. K. Heng, C. Hengstenberg, A. W. Hewitt, H. Hishigaki, C. B. Hoyng, P. L. Huang, W. Huang, S. C. Hunt, K. Hveem, E. Hypponen, W. G. Iacono, S. Ichihara, M. A. Ikram, C. R. Isasi, R. D. Jackson, M. R. Jarvelin, Z. B. Jin, K. H. Jockel, P. K. Joshi, P. Jousilahti, J. W. Jukema, M. Kahonen, Y. Kamatani, K. D. Kang, J. Kaprio, S. L. R. Kardia, F. Karpe, N. Kato, F. Kee, T. Kessler, A. V. Khera, C. C. Khor, L. Kiemeney, B. J. Kim, E. K. Kim, H. L. Kim, P. Kirchhof, M. Kivimaki, W. P. Koh, H. A. Koistinen, G. D. Kolovou, J. S. Kooner, C. Kooperberg, A. Kottgen, P. Kovacs, A. Kraaijeveld, P. Kraft, R. M. Krauss, M. Kumari, Z. Kutalik, M. Laakso, L. A. Lange, C. Langenberg, L. J. Launer, L. Le Marchand, H. Lee, N. R. Lee, T. Lehtimaki, H. Li, L. Li, W. Lieb, X. Lin, L. Lind, A. Linneberg, C. T. Liu, J. Liu, M. Loeffler, B. London, S. A. Lubitz, S. J. Lye, D. A. Mackey, R. Magi, P. K. E. Magnusson, G. M. Marcus, P. M. Vidal, N. G. Martin, W. Marz, F. Matsuda, R. W. McGarrah, M. McGue, A. J. McKnight, S. E. Medland, D. Mellstrom, A. Metspalu, B. D. Mitchell, P. Mitchell, D. O. Mook-Kanamori, A. D. Morris, L. A. Mucci, P. B. Munroe, M. A. Nalls, S. Nazarian, A. E. Nelson, M. J. Neville, C. Newton-Cheh, C. S. Nielsen, M. M. Nothen, C. Ohlsson, A. J. Oldehinkel, L. Orozco, K. Pahkala, P. Pajukanta, C. N. A. Palmer, E. J. Parra, C. Pattaro, O. Pedersen, C. E. Pennell, B. Penninx, L. Perusse, A. Peters, P. A. Peyser, D. J. Porteous, D. Posthuma, C. Power, P. P. Pramstaller, M. A. Province, Q. Qi, J. Qu, D. J. Rader, O. T. Raitakari, S. Ralhan, L. S. Rallidis, D. C. Rao, S. Redline, D. F. Reilly, A. P. Reiner, S. Y. Rhee, P. M. Ridker, M. Rienstra, S. Ripatti, M. D. Ritchie, D. M. Roden, F. R. Rosendaal, J. I. Rotter, I. Rudan, F. Rutters, C. Sabanayagam, D. Saleheen, V. Salomaa, N. J. Samani, D. K. Sanghera, N. Sattar, B. Schmidt, H. Schmidt, R. Schmidt, M. B. Schulze, H. Schunkert, L. J. Scott, R. J. Scott, P. Sever, E. J. Shiroma, M. B. Shoemaker, X. O. Shu, E. M. Simonsick, M. Sims, J. R. Singh, A. B. Singleton, M. F. Sinner, J. G. Smith, H. Snieder, T. D. Spector, M. J. Stampfer, K. J. Stark, D. P. Strachan, L. M. 't Hart, Y. Tabara, H. Tang, J. C. Tardif, T. A. Thanaraj, N. J. Timpson, A. Tonjes, A. Tremblay, T. Tuomi, J. Tuomilehto, M. T. Tusie-Luna, A. G. Uitterlinden, R. M. van Dam, P. van der Harst, N. Van der Velde, C. M. van Duijn, N. M. van Schoor, V. Vitart, U. Volker, P. Vollenweider, H. Volzke, N. H. Wacher-Rodarte, M. Walker, Y. X. Wang, N. J. Wareham, R. M. Watanabe, H. Watkins, D. R. Weir, T. M. Werge, E. Widen, L. R. Wilkens, G. Willemsen, W. C. Willett, J. F. Wilson, T. Y. Wong, J. T. Woo, A. F. Wright, J. Y. Wu, H. Xu, C. S. Yajnik, M. Yokota, J. M. Yuan, E. Zeggini, B. S. Zemel, W. Zheng, X. Zhu, J. M. Zmuda, A. B. Zonderman, J. A. Zwart; 23andMe Research Team; VA Million Veteran Program; DiscovEHR (DiscovEHR and MyCode Community Health Initiative); eMERGE (Electronic Medical Records and Genomics Network); Lifelines Cohort Study; PRACTICAL Consortium; Understanding Society Scientific Group; D. I. Chasman, Y. S. Cho, I. M. Heid, M. I. McCarthy, M. C. Y. Ng, C. J. O'Donnell, F. Rivadeneira, U. Thorsteinsdottir, Y. V. Sun, E. S. Tai, M. Boehnke, P. Deloukas, A. E. Justice, C. M. Lindgren, R. J. F. Loos, K. L. Mohlke, K. E. North, K. Stefansson, R. G. Walters, T. W. Winkler, K. L. Young, P. R. Loh, J. Yang, T. Esko, T. L. Assimes, A. Auton, G. R. Abecasis, C. J. Willer, A. E. Locke, S. I. Berndt, G. Lettre, T. M. Frayling, Y. Okada, A. R. Wood, P. M. Visscher, J. N. Hirschhorn, A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).
53. C. E. Breeze, S. Beck, S. I. Berndt, N. Franceschini, The missing diversity in human epigenomic studies. *Nat. Genet.* **54**, 737–739 (2022).
54. S. Moon, Y. J. Kim, S. Han, M. Y. Hwang, D. M. Shin, M. Y. Park, Y. Lu, K. Yoon, H. M. Jang, Y. K. Kim, T. J. Park, D. S. Song, J. K. Park, J. E. Lee, B. J. Kim, The Korea biobank array: Design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* **9**, 1382 (2019).
55. C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins, UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Med.* **12**, e1001779 (2015).
56. R. Poplin, V. Ruano-Rubio, M. A. DePristo, T. J. Fennell, M. O. Carneiro, G. A. Van der Auwera, D. E. Kling, L. D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M. J. Daly, B. Neale, D. G. MacArthur, E. Banks, Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv:201178 (2018). https://doi.org/10.1101/201178.
57. The International HapMap Consortium, A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
58. P. H. Sudmant, T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. H. Fritz, M. K. Konkel, A. Malhotra, A. M. Stutz, X. Shi, F. P. Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. J. Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E. W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll; The 1000 Genomes Project Consortium, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler, J. O. Korbel, An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
59. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alfoldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski; Genome Aggregation Database Consortium, B. M. Neale, M. J. Daly, D. G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
60. C. Fuchsberger, J. Flannick, T. M. Teslovich, A. Mahajan, V. Agarwala, K. J. Gaulton, C. Ma, P. Fontanillas, L. Moutsianas, D. J. McCarthy, M. A. Rivas, J. R. B. Perry, X. Sim, T. W. Blackwell, N. R. Robertson, N. W. Rayner, P. Cingolani, A. E. Locke, J. F. Tajes, H. M. Highland, J. Dupuis, P. S. Chines, C. M. Lindgren, C. Hartl, A. U. Jackson, H. Chen, J. R. Huyghe, M. van de Bunt,

R. D. Pearson, A. Kumar, M. Muller-Nurasyid, N. Grarup, H. M. Stringham, E. R. Gamazon, J. Lee, Y. Chen, R. A. Scott, J. E. Below, P. Chen, J. Huang, M. J. Go, M. L. Stitzel, D. Pasko, S. C. J. Parker, T. V. Varga, T. Green, N. L. Beer, A. G. Day-Williams, T. Ferreira, T. Fingerlin, M. Horikoshi, C. Hu, I. Huh, M. K. Ikram, B. J. Kim, Y. Kim, Y. J. Kim, M. S. Kwon, J. Lee, S. Lee, K. H. Lin, T. J. Maxwell, Y. Nagai, X. Wang, R. P. Welch, J. Yoon, W. Zhang, N. Barzilai, B. F. Voight, B. G. Han, C. P. Jenkinson, T. Kuulasmaa, J. Kuusisto, A. Manning, M. C. Y. Ng, N. D. Palmer, B. Balkau, A. Stancakova, H. E. Abboud, H. Boeing, V. Giedraitis, D. Prabhakaran, O. Gottesman, J. Scott, J. Carey, P. Kwan, G. Grant, J. D. Smith, B. M. Neale, S. Purcell, A. S. Butterworth, J. M. M. Howson, H. M. Lee, Y. Lu, S. H. Kwak, W. Zhao, J. Danesh, V. K. L. Lam, K. S. Park, D. Saleheen, W. Y. So, C. H. T. Tam, U. Afzal, D. Aguilar, R. Arya, T. Aung, E. Chan, C. Navarro, C. Y. Cheng, D. Palli, A. Correa, J. E. Curran, D. Rybin, V. S. Farook, S. P. Fowler, B. I. Freedman, M. Griswold, D. E. Hale, P. J. Hicks, C. C. Khor, S. Kumar, B. Lehne, D. Thuillier, W. Y. Lim, J. Liu, Y. T. van der Schouw, M. Loh, S. K. Musani, S. Puppala, W. R. Scott, L. Yengo, S. T. Tan, H. A. Taylor Jr., F. Thameem, G. Wilson Sr., T. Y. Wong, P. R. Njolstad, J. C. Levy, M. Mangino, L. L. Bonnycastle, T. Schwarzmayr, J. Fadista, G. L. Surdulescu, C. Herder, C. J. Groves, T. Wieland, J. Bork-Jensen, I. Brandslund, C. Christensen, H. A. Koistinen, A. S. F. Doney, L. Kinnunen, T. Esko, A. J. Farmer, L. Hakaste, D. Hodgkiss, J. Kravic, V. Lyssenko, M. Hollensted, M. E. Jorgensen, T. Jorgensen, C. Ladenvall, J. M. Justesen, A. Karajamaki, J. Kriebel, W. Rathmann, L. Lannfelt, T. Lauritzen, N. Narisu, A. Linneberg, O. Melander, L. Milani, M. Neville, M. Orho-Melander, L. Qi, Q. Qi, M. Roden, O. Rolandsson, A. Swift, A. H. Rosengren, K. Stirrups, A. R. Wood, E. Mihailov, C. Blancher, M. O. Carneiro, J. Maguire, R. Poplin, K. Shakir, T. Fennell, M. DePristo, M. H. de Angelis, P. Deloukas, A. P. Gjesing, G. Jun, P. Nilsson, J. Murphy, R. Onofrio, B. Thorand, T. Hansen, C. Meisinger, F. B. Hu, B. Isomaa, F. Karpe, L. Liang, A. Peters, C. Huth, S. P. O'Rahilly, C. N. A. Palmer, O. Pedersen, R. Rauramaa, J. Tuomilehto, V. Salomaa, R. M. Watanabe, A. C. Syvanen, R. N. Bergman, D. Bharadwaj, E. P. Bottinger, Y. S. Cho, G. R. Chandak, J. C. N. Chan, K. S. Chia, M. J. Daly, S. B. Ebrahim, C. Langenberg, P. Elliott, K. A. Jablonski, D. M. Lehman, W. Jia, R. C. W. Ma, T. I. Pollin, M. Sandhu, N. Tandon, P. Froguel, I. Barroso, Y. Y. Teo, E. Zeggini, R. J. F. Loos, K. S. Small, J. S. Ried, R. A. DeFronzo, H. Grallert, B. Glaser, A. Metspalu, N. J. Wareham, M. Walker, E. Banks, C. Gieger, E. Ingelsson, H. K. Im, T. Illig, P. W. Franks, G. Buck, J. Trakalo, D. Buck, I. Prokopenko, R. Magi, L. Lind, Y. Farjoun, K. R. Owen, A. L. Gloyn, K. Strauch, T. Tuomi, J. S. Kooner, J. Y. Lee, T. Park, P. Donnelly, A. D. Morris, A. T. Hattersley, D. W. Bowden, F. S. Collins, G. Atzmon, J. C. Chambers, T. D. Spector, M. Laakso, T. M. Strom, G. I. Bell, J. Blangero, R. Duggirala, E. S. Tai, G. McVean, C. L. Hanis, J. G. Wilson, M. Seielstad, T. M. Frayling, J. B. Meigs, N. J. Cox, R. Sladek, E. S. Lander, S. Gabriel, N. P. Burtt, K. L. Mohlke, T. Meitinger, L. Groop, G. Abecasis, J. C. Florez, L. J. Scott, A. P. Morris, H. M. Kang, M. Boehnke, D. Altshuler, M. I. McCarthy, The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).

61. The UK10K Consortium, K. Walter, J. L. Min, J. Huang, L. Crooks, Y. Memari, S. McCarthy, J. R. Perry, C. Xu, M. Futema, D. Lawson, V. Iotchkova, S. Schiffels, A. E. Hendricks, P. Danecek, R. Li, J. Floyd, L. V. Wain, I. Barroso, S. E. Humphries, M. E. Hurles, E. Zeggini, J. C. Barrett, V. Plagnol, J. B. Richards, C. M. Greenwood, N. J. Timpson, R. Durbin, N. Soranzo, The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
62. C. C. Laurie, K. F. Doheny, D. B. Mirel, E. W. Pugh, L. J. Bierut, T. Bhangale, F. Boehm, N. E. Caporaso, M. C. Cornelis, H. J. Edenberg, S. B. Gabriel, E. L. Harris, F. B. Hu, K. B. Jacobs, P. Kraft, M. T. Landi, T. Lumley, T. A. Manolio, C. McHugh, I. Painter, J. Paschall, J. P. Rice, K. M. Rice, X. Zheng, B. S. Weir, G. Investigators, Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* **34**, 591–602 (2010).
63. M. Nagasaki, J. Yasuda, F. Katsuoka, N. Nariai, K. Kojima, Y. Kawai, Y. Yamaguchi-Kabata, J. Yokozawa, I. Danjoh, S. Saito, Y. Sato, T. Mimori, K. Tsuda, R. Saito, X. Pan, S. Nishikawa, S. Ito, Y. Kuroki, O. Tanabe, N. Fuse, S. Kuriyama, H. Kiyomoto, A. Hozawa, N. Minegishi, J. D. Engel, K. Kinoshita, S. Kure, N. Yaegashi; ToMMo Japanese Reference Panel Project, M. Yamamoto, Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. *Nat. Commun.* **6**, 8018 (2015).
64. All Of Us Research Program, "All Of Us Q2 2022 Release Genomic Quality Report" (All Of Us, 2020 ; www.researchallofus.org/wp-content/themes/research-hub-wordpress-theme/media/2022/06/All%20Of%20Us%20Q2%202022%20Release%20Genomic%20Quality%20Report.pdf). [the easiest access to this source is via the URL].
65. G. Glusman, J. Caballero, D. E. Mauldin, L. Hood, J. C. Roach, Kaviar: An accessible system for testing SNV novelty. *Bioinformatics* **27**, 3216–3217 (2011).
66. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H. H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur; Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
67. S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, K. Sirotkin, dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
68. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W. M. Chen, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
69. J. S. Seo, A. Rhie, J. Kim, S. Lee, M. H. Sohn, C. U. Kim, A. Hastie, H. Cao, J. Y. Yun, J. Kim, J. Kuk, G. H. Park, J. Kim, H. Ryu, J. Kim, M. Roh, J. Baek, M. W. Hunkapiller, J. Korlach, J. Y. Shin, C. Kim, De novo assembly and phasing of a Korean human genome. *Nature* **538**, 243–247 (2016).
70. The 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, G. A. McVean, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
71. B. L. Browning, Y. Zhou, S. R. Browning, A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
72. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
73. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
74. J. H. Leland McInnes, James Melville, UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 (2018). https://doi.org/10.48550/arXiv.1802.03426.
75. P. R. Loh, P. Danecek, P. F. Palamara, C. Fuchsberger, Y. A. Reshef, H. K. Finucane, S. Schoenherr, L. Forer, S. McCarthy, G. R. Abecasis, R. Durbin, A. L. Price, Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
76. S. Das, L. Forer, S. Schonherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P. R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fuchsberger, Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
77. R. M. Kuhn, D. Haussler, W. J. Kent, The UCSC genome browser and associated tools. *Brief Bioinform.* **14**, 144–161 (2013).
78. S. Sakaue, M. Kanai, Y. Tanigawa, J. Karjalainen, M. Kurki, S. Koshiba, A. Narita, T. Konuma, K. Yamamoto, M. Akiyama, K. Ishigaki, A. Suzuki, K. Suzuki, W. Obara, K. Yamaji, K. Takahashi, S. Asai, Y. Takahashi, T. Suzuki, N. Shinozaki, H. Yamaguchi, S. Minami, S. Murayama, K. Yoshimori, S. Nagayama, D. Obata, M. Higashiyama, A. Masumoto, Y. Koretsune, FinnGen, K. Ito, C. Terao, T. Yamauchi, I. Komuro, T. Kadowaki, G. Tamiya, M. Yamamoto, Y. Nakamura, M. Kubo, Y. Murakami, K. Yamamoto, Y. Kamatani, A. Palotie, M. A. Rivas, M. J. Daly, K. Matsuda, Y. Okada, A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415–1424 (2021).
79. O. Weissbrod, F. Hormozdiari, C. Benner, R. Cui, J. Ulirsch, S. Gazal, A. P. Schoech, B. van de Geijn, Y. Reshef, C. Marquez-Luna, L. O'Connor, M. Pirinen, H. K. Finucane, A. L. Price, Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
80. K. Zhang, J. D. Hocker, M. Miller, X. Hou, J. Chiou, O. B. Poirion, Y. Qiu, Y. E. Li, K. J. Gaulton, A. Wang, S. Preissl, B. Ren, A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021).
81. J. K. Pickrell, Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
82. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R. Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M. Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander, J. M. Engreitz, Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
83. J. Cao, D. R. O'Day, H. A. Pliner, P. D. Kingsley, M. Deng, R. M. Daza, M. A. Zager, K. A. Aldinger, R. Blecher-Gonen, F. Zhang, M. Spielmann, J. Palis, D. Doherty, F. J. Steemers, I. A. Glass, C. Trapnell, J. Shendure, A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
84. The Tabula Sapiens Consortium, The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
85. W. Meuleman, A. Muratov, E. Rynes, J. Halow, K. Lee, D. Bates, M. Diegel, D. Dunn, F. Neri, A. Teodosiadis, A. Reynolds, E. Haugen, J. Nelson, A. Johnson, M. Frerker, M. Buckley, R. Sandstrom, J. Vierstra, R. Kaul, J. Stamatoyannopoulos, Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020).
86. C. A. Boix, B. T. James, Y. P. Park, W. Meuleman, M. Kellis, Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300–307 (2021).

87. The ENCODE Project Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X. O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lecuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigo, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
88. J. A. Castro-Mondragon, R. Riudavets-Puig, I. Rauluseviciute, R. B. Lemma, L. Turchi, R. Blanc-Mathieu, J. Lucas, P. Boddie, A. Khan, N. Manosalva Perez, O. Fornes, T. Y. Leung, A. Aguirre, F. Hammal, D. Schmelter, D. Baranasic, B. Ballester, A. Sandelin, B. Lenhard, K. Vandepoele, W. W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
89. D. Lee, D. U. Gorkin, M. Baker, B. J. Strober, A. L. Asoni, A. S. McCallion, M. A. Beer, A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet.* **47**, 955–961 (2015).
90. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
91. J. Yang, T. Ferreira, A. P. Morris, S. E. Medland; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, P. M. Visscher, Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
92. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
93. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
94. C. N. Foley, J. R. Staley, P. G. Breen, B. B. Sun, P. D. W. Kirk, S. Burgess, J. M. M. Howson, A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* **12**, 764 (2021).
95. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).
96. A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, H. Pfister, UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
97. W. Wen, W. Zheng, Y. Okada, F. Takeuchi, Y. Tabara, J. Y. Hwang, R. Dorajoo, H. Li, F. J. Tsai, X. Yang, J. He, Y. Wu, M. He, Y. Zhang, J. Liang, X. Guo, W. H. Sheu, R. Delahanty, X. Guo, M. Kubo, K. Yamamoto, T. Ohkubo, M. J. Go, J. J. Liu, W. Gan, C. C. Chen, Y. Gao, S. Li, N. R. Lee, C. Wu, X. Zhou, H. Song, J. Yao, I. T. Lee, J. Long, T. Tsunoda, K. Akiyama, N. Takashima, Y. S. Cho, R. T. Ong, L. Lu, C. H. Chen, A. Tan, T. K. Rice, L. S. Adair, L. Gui, M. Allison, W. J. Lee, Q. Cai, M. Isomura, S. Umemura, Y. J. Kim, M. Seielstad, J. Hixson, Y. B. Xiang, M. Isono, B. J. Kim, X. Sim, W. Lu, T. Nabika, J. Lee, W. Y. Lim, Y. T. Gao, R. Takayanagi, D. H. Kang, T. Y. Wong, C. A. Hsiung, I. C. Wu, J. M. Juang, J. Shi, B. Y. Choi, T. Aung, F. Hu, M. K. Kim, W. Y. Lim, T. D. Wang, M. H. Shin, J. Lee, B. T. Ji, Y. H. Lee, T. L. Young, D. H. Shin, B. Y. Chun, M. C. Cho, B. G. Han, C. M. Hwu, T. L. Assimes, D. Absher, X. Yan, E. Kim, J. Z. Kuo, S. Kwon, K. D. Taylor, Y. D. Chen, J. I. Rotter, L. Qi, D. Zhu, T. Wu, K. L. Mohlke, D. Gu, Z. Mo, J. Y. Wu, X. Lin, T. Miki, E. S. Tai, J. Y. Lee, N. Kato, X. O. Shu, T. Tanaka, Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum. Mol. Genet.* **23**, 5492–5504 (2014).
98. I. Lastres-Becker, S. Brodesser, D. Lutjohann, M. Azizov, J. Buchmann, E. Hintermann, K. Sandhoff, A. Schurmann, J. Nowock, G. Auburger, Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice. *Hum. Mol. Genet.* **17**, 1465–1481 (2008).
99. K. Bloom, A. W. Mohsen, A. Karunanidhi, D. El Demellawy, M. Reyes-Mugica, Y. Wang, L. Ghaloul-Gonzalez, C. Otsubo, K. Tobita, R. Muzumdar, Z. Gong, E. Tas, S. Basu, J. Chen, M. Bennett, C. Hoppel, J. Vockley, Investigating the link of ACAD10 deficiency to type 2 diabetes mellitus. *J. Inherit. Metab. Dis.* **41**, 49–57 (2018).
100. D. Zhang, M. Yang, D. Zhou, Z. Li, L. Cai, Y. Bao, H. Li, Z. Shan, J. Liu, D. Lv, Y. Liu, C. Xu, J. Ling, Y. Xu, S. Zhang, Q. Huang, Y. Shi, Y. Zhu, M. Lai, The polymorphism rs671 at ALDH2 associated with serum uric acid levels in Chinese Han males: A genome-wide association study. *Gene* **651**, 62–69 (2018).
101. A. Tan, J. Sun, N. Xia, X. Qin, Y. Hu, S. Zhang, S. Tao, Y. Gao, X. Yang, H. Zhang, S. T. Kim, T. Peng, X. Lin, L. Li, L. Mo, Z. Liang, D. Shi, Z. Huang, X. Huang, M. Liu, Q. Ding, J. M. Trent, S. L. Zheng, Z. Mo, J. Xu, A genome-wide association and gene-environment interaction study for serum triglycerides levels in a healthy Chinese male population. *Hum. Mol. Genet.* **21**, 1658–1664 (2012).
102. F. Takeuchi, M. Yokota, K. Yamamoto, E. Nakashima, T. Katsuya, H. Asano, M. Isono, T. Nabika, T. Sugiyama, A. Fujioka, N. Awata, K. Ohnaka, M. Nakatochi, H. Kitajima, H. Rakugi, J. Nakamura, T. Ohkubo, Y. Imai, K. Shimamoto, Y. Yamori, S. Yamaguchi, S. Kobayashi, R. Takayanagi, T. Ogihara, N. Kato, Genome-wide association study of coronary artery disease in the Japanese. *Eur. J. Hum. Genet.* **20**, 333–340 (2012).
103. J. M. Guo, A. J. Liu, P. Zang, W. Z. Dong, L. Ying, W. Wang, P. Xu, X. R. Song, J. Cai, S. Q. Zhang, J. L. Duan, J. L. Mehta, D. F. Su, ALDH2 protects against stroke by clearing 4-HNE. *Cell Res.* **23**, 915–930 (2013).
104. M. L. Hribal, I. Presta, T. Procopio, M. A. Marini, A. Stancakova, J. Kuusisto, F. Andreozzi, A. Hammarstedt, P. A. Jansson, N. Grarup, T. Hansen, M. Walker, N. Stefan, A. Fritsche, H. U. Haring, O. Pedersen, U. Smith, M. Laakso, G. Sesti, EUGENE2 Consortium, Glucose tolerance, insulin sensitivity and insulin release in European non-diabetic carriers of a polymorphism upstream of CDKN2A and CDKN2B. *Diabetologia* **54**, 795–802 (2011).
105. Y. Kong, R. B. Sharma, B. U. Nwosu, L. C. Alonso, Islet biology, the CDKN2A/B locus and type 2 diabetes risk. *Diabetologia* **59**, 1579–1593 (2016).
106. M. Moritani, S. Yamasaki, M. Kagami, T. Suzuki, T. Yamaoka, T. Sano, J. Hata, M. Itakura, Hypoplasia of endocrine and exocrine pancreas in homozygous transgenic TGF-β1. *Mol. Cell Endocrinol.* **229**, 175–184 (2005).
107. A. Bergström, S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanche, J. F. Deleuze, H. Cann, S. Mallick, D. Reich, M. S. Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, C. Tyler-Smith, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
108. L. Pagani, D. J. Lawson, E. Jagoda, A. Morseburg, A. Eriksson, M. Mitt, F. Clemente, G. Hudjashov, M. DeGiorgio, L. Saag, J. D. Wall, A. Cardona, R. Magi, M. A. Wilson Sayres, S. Kaewert, C. Inchley, C. L. Scheib, M. Jarve, M. Karmin, G. S. Jacobs, T. Antao, F. M. Iliescu, A. Kushniarevich, Q. Ayub, C. Tyler-Smith, Y. Xue, B. Yunusbayev, K. Tambets, C. B. Mallick, L. Saag, E. Pocheshkhova, G. Andriadze, C. Muller, M. C. Westaway, D. M. Lambert, G. Zoraqi, S. Turdikulova, D. Dalimova, Z. Sabitov, G. N. N. Sultana, J. Lachance, S. Tishkoff, K. Momynaliev, J. Isakova, L. D. Damba, M. Gubina, P. Nymadawa, I. Evseeva, L. Atramentova, O. Utevska, F. X. Ricaut, N. Brucato, H. Sudoyo, T. Letellier, M. P. Cox, N. A. Barashkov, V. Skaro, L. Mulahasanovic, D. Primorac, H. Sahakyan, M. Mormina, C. A. Eichstaedt, D. V. Lichman, S. Abdullah, G. Chaubey, J. T. S. Wee, E. Mihailov, A. Karunas, S. Litvinov, R. Khusainova, N. Ekomasova, V. Akhmetova, I. Khidiyatova, D. Marjanovic, L. Yepiskoposyan, D. M. Behar, E. Balanovska, A. Metspalu, M. Derenko, B. Malyarchuk, M. Voevoda, S. A. Fedorova, L. P. Osipova, M. M. Lahr, P. Gerbault, M. Leavesley, A. B. Migliano, M. Petraglia, O. Balanovsky, E. K. Khusnutdinova, E. Metspalu, M. G. Thomas, A. Manica, R. Nielsen, R. Villems, E. Willerslev, T. Kivisild, M. Metspalu, Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* **538**, 238–242 (2016).
109. W. S. Watkins, J. E. Feusier, J. Thomas, C. Goubert, S. Mallick, L. B. Jorde, The simons genome diversity project: A global analysis of mobile element diversity. *Genome Biol. Evol.* **12**, 779–794 (2020).
110. J. L. Rodriguez-Flores, K. Fakhro, F. Agosto-Perez, M. D. Ramstetter, L. Arbiza, T. L. Vincent, A. Robay, J. A. Malek, K. Suhre, L. Chouchane, R. Badii, A. Al-Nabet Al-Marri, C. Abi Khalil, M. Zirie, A. Jayyousi, J. Salit, A. Keinan, A. G. Clark, R. G. Crystal, J. G. Mezey, Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. *Genome Res.* **26**, 151–162 (2016).
111. J. Kim, J. A. Weber, S. Jho, J. Jang, J. Jun, Y. S. Cho, H. M. Kim, H. Kim, Y. Kim, O. Chung, C. G. Kim, H. Lee, B. C. Kim, K. Han, I. Koh, K. S. Chae, S. Lee, J. S. Edwards, J. Bhak, Ko-Variome: Korean National Standard Reference Variome database of whole genomes with comprehensive SNV, indel, CNV, and SV analyses. *Sci. Rep.* **8**, 5677 (2018).
112. PGP-UK Consortium, Personal Genome Project UK (PGP-UK): A research and citizen science hybrid project in support of personalized medicine. *BMC Med. Genomics* **11**, 108 (2018).
113. C. Zhang, Y. Lu, Q. Feng, X. Wang, H. Lou, J. Liu, Z. Ning, K. Yuan, Y. Wang, Y. Zhou, L. Deng, L. Liu, Y. Yang, S. Li, L. Ma, Z. Zhang, L. Jin, B. Su, L. Kang, S. Xu, Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biol.* **18**, 115 (2017).
114. M. Mondal, F. Casals, T. Xu, G. M. Dall'Olio, M. Pybus, M. G. Netea, D. Comas, H. Laayouni, Q. Li, P. P. Majumder, J. Bertranpetit, Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat. Genet.* **48**, 1066–1070 (2016).
115. J. Chen, J. S. Wu, T. Mize, M. Moreno, M. Hamid, F. Servin, B. Bashy, Z. Zhao, P. Jia, M. T. Tsuang, K. S. Kendler, M. Xiong, X. Chen, A frameshift variant in the CHST9 gene identified by family-based whole genome sequencing is associated with schizophrenia in chinese population. *Sci. Rep.* **9**, 12717 (2019).
116. S. S. Wong, K. M. Kim, J. C. Ting, K. Yu, J. Fu, S. Liu, R. Cristescu, M. Nebozhyn, L. Gong, Y. G. Yue, J. Wang, C. Ronghua, A. Loboda, J. Hardwick, X. Liu, H. Dai, J. G. Jin, X. S. Ye,

S. Y. Kang, I. G. Do, J. O. Park, T. S. Sohn, C. Reinhard, J. Lee, S. Kim, A. Aggarwal, Genomic landscape and genetic heterogeneity in gastric adenocarcinoma revealed by whole-genome sequencing. *Nat. Commun.* **5**, 5477 (2014).

117. J. A. Robinson, S. Belsare, S. Birnbaum, D. E. Newman, J. Chan, J. P. Glenn, B. Ferguson, L. A. Cox, J. D. Wall, Analysis of 100 high-coverage genomes from a pedigreed captive baboon colony. *Genome Res.* **29**, 848–856 (2019).
118. P. de Barros Damgaard, R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G. Barjamovic, S. Rasmussen, C. Zacho, N. Baimukhanov, V. Zaibert, V. Merz, A. Biddanda, I. Merz, V. Loman, V. Evdokimov, E. Usmanova, B. Hemphill, A. Seguin-Orlando, F. E. Yediay, I. Ullah, K. G. Sjögren, K. H. Iversen, J. Choin, C. de la Fuente, M. Ilardo, H. Schroeder, V. Moiseyev, A. Gromov, A. Polyakov, S. Omura, S. Y. Senyurt, H. Ahmad, C. McKenzie, A. Margaryan, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, O. I. Goriunova, V. I. Bazaliiskii, J. Novembre, A. W. Weber, L. Orlando, M. E. Allentoft, R. Nielsen, K. Kristiansen, M. Sikora, A. K. Outram, R. Durbin, E. Willerslev, The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
119. J. J. Lee, S. Park, H. Park, S. Kim, J. Lee, J. Lee, J. Youk, K. Yi, Y. An, I. K. Park, C. H. Kang, D. H. Chung, T. M. Kim, Y. K. Jeon, D. Hong, P. J. Park, Y. S. Ju, Y. T. Kim, Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. *Cell* **177**, 1842–1857.e21 (2019).
120. E. H. Wong, A. Khrunin, L. Nichols, D. Pushkarev, D. Khokhrin, D. Verbenko, O. Evgrafov, J. Knowles, J. Novembre, S. Limborska, A. Valouev, Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res.* **27**, 1–14 (2017).
121. F. J. Clemente, A. Cardona, C. E. Inchley, B. M. Peter, G. Jacobs, L. Pagani, D. J. Lawson, T. Antao, M. Vicente, M. Mitt, M. DeGiorgio, Z. Faltyskova, Y. Xue, Q. Ayub, M. Szpak, R. Magi, A. Eriksson, A. Manica, M. Raghavan, M. Rasmussen, S. Rasmussen, E. Willerslev, A. Vidal-Puig, C. Tyler-Smith, R. Villems, R. Nielsen, M. Metspalu, B. Malyarchuk, M. Derenko, T. Kivisild, A selective sweep on a deleterious mutation in CPT1A in arctic populations. *Am. J. Hum. Genet.* **95**, 584–589 (2014).
122. L. Gesang, L. Gusang, C. Dawa, G. Gesang, K. Li, Whole-Genome Sequencing Identifies the Egl Nine Homologue 3 (egln3/phd3) and Protein Phosphatase 1 Regulatory Inhibitor Subunit 2 (PPP1R2P1) Associated with High-Altitude Polycythemia in Tibetans at High Altitude. *Dis. Markers* **2019**, 5946461 (2019).
123. Y. S. Lee, Y. S. Cho, G. K. Lee, S. Lee, Y. W. Kim, S. Jho, H. M. Kim, S. H. Hong, J. A. Hwang, S. Y. Kim, D. Hong, I. J. Choi, B. C. Kim, B. C. Kim, C. H. Kim, H. Choi, Y. Kim, K. W. Kim, G. Kong, H. L. Kim, J. Bhak, S. H. Lee, J. S. Lee, Genomic profile analysis of diffuse-type gastric cancers. *Genome Biol.* **15**, R55 (2014).
124. C. Alkan, P. Kavak, M. Somel, O. Gokcumen, S. Ugurlu, C. Saygi, E. Dal, K. Bugra, T. Gungor, S. C. Sahinalp, N. Ozoren, C. Bekpen, Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe. *Asia and Africa. BMC Genomics* **15**, 963 (2014).
125. B. Zeng, P. Huang, P. Du, X. Sun, X. Huang, X. Fang, L. Li, Comprehensive study of germline mutations and double-hit events in esophageal squamous cell cancer. *Front. Oncol.* **11**, 637431 (2021).
126. S. K. Yoo, Y. S. Song, E. K. Lee, J. Hwang, H. H. Kim, G. Jung, Y. A. Kim, S. J. Kim, S. W. Cho, J. K. Won, E. J. Chung, J. Y. Shin, K. E. Lee, J. I. Kim, Y. J. Park, J. S. Seo, Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer. *Nat. Commun.* **10**, 2764 (2019).
127. W. Jaratlerdsiri, E. K. F. Chan, T. Gong, D. C. Petersen, A. M. F. Kalsbeek, P. A. Venter, P. D. Stricker, M. S. R. Bornman, V. M. Hayes, Whole-genome sequencing reveals elevated tumor mutational burden and initiating driver mutations in african men with treatment-naive, High-risk prostate cancer. *Cancer Res.* **78**, 6736–6746 (2018).
128. C. A. Eichstaedt, L. Pagani, T. Antao, C. E. Inchley, A. Cardona, A. Morseburg, F. J. Clemente, T. J. Sluckin, E. Metspalu, M. Mitt, R. Magi, G. Hudjashov, M. Metspalu, M. Mormina, G. S. Jacobs, T. Kivisild, Evidence of early-stage selection on EPAS1 and GPR126 genes in andean high altitude populations. *Sci. Rep.* **7**, 13042 (2017).
129. M. Raghavan, M. Steinrucken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, A. Albrechtsen, C. Valdiosera, M. C. Avila-Arcos, A. S. Malaspinas, A. Eriksson, I. Moltke, M. Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. V. Moreno-Mayar, T. S. Korneliussen, T. Pierre, M. Rasmussen, P. F. Campos, P. de Barros Damgaard, M. E. Allentoft, J. Lindo, E. Metspalu, R. Rodriguez-Varela, J. Mansilla, C. Henrickson, A. Seguin-Orlando, H. Malmstrom, T. Stafford Jr., S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergstrom, Y. Xue, V. Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Balloux, I. Leboreiro, J. L. Vera, H. Rangel-Villalobos, D. Pettener, D. Luiselli, L. G. Davis, E. Heyer, C. P. E. Zollikofer, M. S. P. de Leon, C. I. Smith, V. Grimes, K. A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F. Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevoda, O. L. Posukh, O. Balanovsky, M. Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S. Litvinov, B. Malyarchuk, M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J. Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S. Sandhu, M. Crawford, R. Villems, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S. Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbin, C. D. Bustamante, Y. S. Song, R. Nielsen, E. Willerslev, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, eaab3884 (2015).
130. P. Gelabert, M. Ferrando-Bernal, T. de-Dios, B. Mattorre, E. Campoy, A. Gorostiza, E. Patin, A. Gonzalez-Martin, C. Lalueza-Fox, Genome-wide data from the Bubi of Bioko Island clarifies the Atlantic fringe of the Bantu dispersal. *BMC Genomics* **20**, 179 (2019).
131. C. Tu, Z. Zeng, P. Qi, X. Li, C. Guo, F. Xiong, B. Xiang, M. Zhou, Q. Liao, J. Yu, Y. Li, X. Li, G. Li, W. Xiong, Identification of genomic alterations in nasopharyngeal carcinoma and nasopharyngeal carcinoma-derived Epstein-Barr virus by whole-genome sequencing. *Carcinogenesis* **39**, 1517–1528 (2018).
132. E. A. Vidal, T. C. Moyano, B. I. Bustos, E. Perez-Palma, C. Moraga, E. Riveras, A. Montecinos, L. Azocar, D. C. Soto, M. Vidal, A. Di Genova, K. Puschel, P. Nurnberg, S. Buch, J. Hampe, M. L. Allende, V. Cambiazo, M. Gonzalez, C. Hodar, M. Montecino, C. Munoz-Espinoza, A. Orellana, A. Reyes-Jara, D. Travisany, P. Vizoso, M. Moraga, S. Eyheramendy, A. Maass, G. V. De Ferrari, J. F. Miquel, R. A. Gutierrez, Whole genome sequence, variant discovery and annotation in mapuche-huilliche native south americans. *Sci. Rep.* **9**, 2132 (2019).
133. H. Fang, Y. Wu, H. Yang, M. Yoon, L. T. Jimenez-Barron, D. Mittelman, R. Robison, K. Wang, G. J. Lyon, Whole genome sequencing of one complex pedigree illustrates challenges with genomic medicine. *BMC Med. Genomics* **10**, 10 (2017).
134. B. Lorente-Galdos, O. Lao, G. Serra-Vidal, G. Santpere, L. F. K. Kuderna, L. R. Arauna, K. Fadhlaoui-Zid, V. N. Pimenoff, H. Soodyall, P. Zalloua, T. Marques-Bonet, D. Comas, Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. *Genome Biol.* **20**, 77 (2019).
135. J. I. Kim, Y. S. Ju, H. Park, S. Kim, S. Lee, J. H. Yi, J. Mudge, N. A. Miller, D. Hong, C. J. Bell, H. S. Kim, I. S. Chung, W. C. Lee, J. S. Lee, S. H. Seo, J. Y. Yun, H. N. Woo, H. Lee, D. Suh, S. Lee, H. J. Kim, M. Yavartanoo, M. Kwak, Y. Zheng, M. K. Lee, H. Park, J. Y. Kim, O. Gokcumen, R. E. Mills, A. W. Zaranek, J. Thakuria, X. Wu, R. W. Kim, J. J. Huntley, S. Luo, G. P. Schroth, T. D. Wu, H. Kim, K. S. Yang, W. Y. Park, H. Kim, G. M. Church, C. Lee, S. F. Kingsmore, J. S. Seo, A highly annotated whole-genome sequence of a Korean individual. *Nature* **460**, 1011–1015 (2009).
136. J. C. Chambers, J. Abbott, W. Zhang, E. Turro, W. R. Scott, S. T. Tan, U. Afzal, S. Afaq, M. Loh, B. Lehne, P. O'Reilly, K. J. Gaulton, R. D. Pearson, X. Li, A. Lavery, J. Vandrovcova, M. N. Wass, K. Miller, J. Sehmi, L. Oozageer, I. K. Kooner, A. Al-Hussaini, R. Mills, J. Grewal, V. Panoulas, A. M. Lewin, K. Northwood, G. S. Wander, F. Geoghegan, Y. Li, J. Wang, T. J. Aitman, M. I. McCarthy, J. Scott, S. Butcher, P. Elliott, J. S. Kooner, The South Asian genome. *PLOS ONE* **9**, e102645 (2014).
137. U. Kairov, A. Molkenov, S. Rakhimova, U. Kozhamkulov, A. Sharip, D. Karabayev, A. Daniyarov, J. H. Lee, J. D. Terwilliger, A. Akilzhanova, Z. Zhumadilov, Whole-genome sequencing data of Kazakh individuals. *BMC Res. Notes* **14**, 45 (2021).
138. C. Jeong, G. Alkorta-Aranburu, B. Basnyat, M. Neupane, D. B. Witonsky, J. K. Pritchard, C. M. Beall, A. Di Rienzo, Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat. Commun.* **5**, 3281 (2014).
139. M. ElHefnawi, S. Jeon, Y. Bhak, A. ElFiky, A. Horaiz, J. Jun, H. Kim, J. Bhak, Whole genome sequencing and bioinformatics analysis of two Egyptian genomes. *Gene* **668**, 129–134 (2018).
140. S. Y. Khan, F. Kabir, O. M'Hamdi, X. Jiao, M. A. Naeem, S. N. Khan, S. Riazuddin, J. F. Hejtmancik, S. A. Riazuddin, Whole genome sequencing data for two individuals of Pakistani descent. *Sci. Data* **5**, 180174 (2018).
141. G. Kang, H. Yun, C. H. Sun, I. Park, S. Lee, J. Kwon, I. Do, M. E. Hong, M. Van Vrancken, J. Lee, J. O. Park, J. Cho, K. M. Kim, T. S. Sohn, Integrated genomic analyses identify frequent gene fusion events and VHL inactivation in gastrointestinal stromal tumors. *Oncotarget* **7**, 6538–6551 (2016).
142. M. Kim, J. K. Rhee, H. Choi, A. Kwon, J. Kim, G. D. Lee, D. W. Jekarl, S. Lee, Y. Kim, T. M. Kim, Passage-dependent accumulation of somatic mutations in mesenchymal stromal cells during in vitro culture revealed by whole genome sequencing. *Sci. Rep.* **7**, 14508 (2017).
143. H. Dogan, H. Can, H. H. Otu, Whole genome sequence of a Turkish individual. *PLOS ONE* **9**, e85233 (2014).
144. H. S. Kim, S. Jeon, C. Kim, Y. K. Kim, Y. S. Cho, J. Kim, A. Blazyte, A. Manica, S. Lee, J. Bhak, Chromosome-scale assembly comparison of the Korean Reference Genome KOREF from PromethION and PacBio with Hi-C mapping information. *Gigascience* **8**, giz125 (2019).
145. L. Shi, Y. Guo, C. Dong, J. Huddleston, H. Yang, X. Han, A. Fu, Q. Li, N. Li, S. Gong, K. E. Lintner, Q. Ding, Z. Wang, J. Hu, D. Wang, F. Wang, L. Wang, G. J. Lyon, Y. Guan, Y. Shen, O. V. Evgrafov, J. A. Knowles, F. Thibaud-Nissen, V. Schneider, C. Y. Yu, L. Zhou, E. E. Eichler, K. F. So, K. Wang, Long-read sequencing and de novo assembly of a Chinese genome. *Nat. Commun.* **7**, 12065 (2016).
146. H. Bai, X. Guo, D. Zhang, N. Narisu, J. Bu, J. Jirimutu, F. Liang, X. Zhao, Y. Xing, D. Wang, T. Li, Y. Zhang, B. Guan, X. Yang, Z. Yang, S. Shuangshan, Z. Su, H. Wu, W. Li, M. Chen, S. Zhu, B. Bayinnamula, Y. Chang, Y. Gao, T. Lan, S. Suyalatu, H. Huang, Y. Su, Y. Chen, W. Li, X. Yang, Q. Feng, J. Wang, H. Yang, J. Wang, Q. Wu, Y. Yin, H. Zhou, The genome of a Mongolian

individual reveals the genetic imprints of Mongolians on modern human populations. *Genome Biol. Evol.* **6**, 3122–3136 (2014).

147. Y. S. Ju, W. C. Lee, J. Y. Shin, S. Lee, T. Bleazard, J. K. Won, Y. T. Kim, J. I. Kim, J. H. Kang, J. S. Seo, A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res.* **22**, 436–445 (2012).

Acknowledgments: We thank the members of SG10K for the contribution of data. This study was conducted with bioresource KBN-2020-029 and Clinical and Omics Data Archive (CODA) from National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea. We would also like to thank M.-H. Sohn (Precision Medicine Institute, Macrogen Inc.) for the excellent technical assistance. We would like to thank the Biobank Japan Project for collecting and providing the extensive dataset used in this study. **Funding:** This work was supported to J.-I.K. by the National Research Foundation of Korea grant funded by the Korean government (Ministry of Science and ICT, MSIT) (2020R1A2C3012524). This work was supported to S.-W.I. by the National Research Foundation of Korea grant funded by the Korean government (MSIT) (2022R1F1A1066135). This work was supported to H.-Y.S. by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2020R1I1A1A01074772). This work was supported to J.-I.K. by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2020R1A6A1A03047972). This work was supported to N.-J.K. by the World Class 300 Project (S2638360) of the Ministry of Trade, Industry and Energy, and Ministry of SMEs and Startups of the Republic of Korea. **Author contributions:** Conceptualization: J.C., S.K., J.K., S.-K.Y., H.-J.K., J.-S.S., and J.-I.K.; data curation: J.C., S.K., J.K., H.-Y.S., Young Jun Park, J.-H.P., and S.-W.I.; formal analysis: J.C., S.K., J.K., C.-U.K., and S.-W.I.; funding acquisition: H.-Y.S., J.-S.S., J.-I.K., and S.-W.I.; investigation: J.C., S.K., J.K., H.-Y.S., S.-K.Y., C.-U.K., Young Jun Park, and S.-W.I.; methodology: J.C., S.K., J.K., S.-K.Y., C.-U.K., and S.-W.I.; project administration: H.-J.K. and J.-I.K.; resources: S.K., S.-K.Y., N.-J.K., Young Joo Park, F.M., Y.M., M.K., B.B.J., J.-S.S., and J.-I.K.; software: J.C., S.K., J.K., S.-K.Y., C.-U.K., S.M., N.K., C.K., N.-J.K., and S.-W.I.; supervision: H.-J.K., J.-H.P., J.-S.S., J.-I.K., and S.-W.I.; validation: H.-Y.S., B. Cha, M.C.J., K.P, J.M.Y., B. Cho, J.-H.P., and J.-S.S.; visualization: J.C., S.K., J.K., and H.-Y.S.; writing: J.C., S.K., J.K., H.-Y.S., and S.-W.I. **Competing interests:** The authors affiliated with Macrogen and Psomagen are full-time employees at Macrogen Inc. and Psomagen Inc., respectively. All other authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Information for NARD2 imputation panel as well as web server for imputation is available at gmi.snu.ac.kr/imputation; full GWAS summary statistics and all codes for NARD2 project are available at Dryad (https://doi.org/10.5061/dryad.ncjsxkt11).

Submitted 11 January 2023
Accepted 6 July 2023
Published 9 August 2023
10.1126/sciadv.adg6319

ScienceAdvances

A whole-genome reference panel of 14,393 individuals for East Asian populations accelerates discovery of rare functional variants

Jaeyong Choi, Sungjae Kim, Juhyun Kim, Ho-Young Son, Seong-Keun Yoo, Chang-Uk Kim, Young Jun Park, Sungji Moon, Bukyoung Cha, Min Chul Jeon, Kyunghyuk Park, Jae Moon Yun, Belong Cho, Namcheol Kim, Changhoon Kim, Nak-Jung Kwon, Young Joo Park, Fumihiko Matsuda, Yukihide Momozawa, Michiaki Kubo, Biobank Japan Project, Hyun-Jin Kim, Jin-Ho Park, Jeong-Sun Seo, Jong-Il Kim, and Sun-Wha Im

Sci. Adv. **9** (32), eadg6319. DOI: 10.1126/sciadv.adg6319