Cross-Population Genetic Variation of Loci Identified by Genome-Wide Association Studies conducted in British participants of European-descent from the UK Biobank

Antonella De Lillo[1], Salvatore D'Antona[1], Maria Fuciarelli[1], Renato Polimanti[2,3*]

[1]Department of Biology, University of Rome Tor Vergata, Rome, Italy

[2]Department of Psychiatry, Yale University School of Medicine, West Haven, CT, USA

[3]VA CT Healthcare Center, West Haven, CT, USA

[*]Corresponding author: Dr. Renato Polimanti. Department of Psychiatry, Yale University School of Medicine and VA CT Healthcare Center, VA CT 116A2, 950 Campbell Avenue, West Haven, CT 06516, USA. Tel: +1 203 932 5711 x5745; Fax: +1 203 937-3897; E-mail: renato.polimanti@yale.edu; ORCID: 0000-0003-0745-6046

Abstract

To provide novel insight regarding the inter-population diversity of loci associated with complex traits, we integrated genome-wide data from UK Biobank (UKB) and 1,000 Genomes Project (1KG) data representative of the genetic diversity among worldwide populations. We investigated genome-wide data of 4,359 traits from 361,194 UKB participants of European descent. Using 1KG data, we explored the allele frequency differences and linkage disequilibrium (LD) structure of UKB genome-wide significant (GWS) loci across worldwide populations. Functional annotation data were used to identify regulatory elements and evaluate the tagging properties of GWS variants. No significant difference was observed in allele frequency between UKB and 1KG GBR (British in England and Scotland). Considering other population groups, we identified genome-wide significant alleles with frequencies different from what expected by chance: UKB vs. 1KG Europeans without GBR (rs74945666; allele=T [0.908 vs. 0.03], standing height $p_{GWAS}=1.48\times10^{-17}$), UKB vs. 1KG African (rs556562; allele=A [0.942 vs. 0.083], platelet count $p_{GWAS}=4.84\times10^{-15}$), UKB vs. 1KG Admixed Americans (rs1812378; allele=T [0.931 vs. 0.089], standing height $p_{GWAS}=4.23\times10^{-12}$), UKB vs. 1KG East Asian (rs55881864; allele=T [0.911 vs. 0.001], monocyte count $p_{GWAS}=7.29\times10^{-13}$), and UKB vs. South Asian (rs74945666; allele=T [0.908 vs. 0.061], standing height $p_{GWAS}=1.48\times10^{-17}$). LD-structure analysis and computational prediction showed differences in how these alleles tag functional elements across human populations. In conclusion, the human diversity of certain GWS loci appear to be affected by local adaptation while in other cases the associations may be biased by residual population stratification.

Keywords: Ancestry, complex traits, 1,000 Genomes Project, GWAS, phenome, UK Biobank

Introduction

Genome-wide association studies (GWAS) are a powerful tool to identify genetic variants associated with human traits and diseases (Visscher et al. 2017). Since the first GWAS conducted in 2005 (Klein et al. 2005), 4,671 GWAS reporting >19,813 associations have been listed in the GWAS Catalog (Buniello et al. 2019) as of August 13, 2020. This unprecedented amount of information has revolutionized our understanding of the predisposition to complex phenotypes, demonstrating that a large portion of the heritability of complex traits resides in common genetic variation (i.e., polymorphisms in the human genome that show a minor allele frequency (MAF) greater than 1%) (Visscher et al. 2017). In recent years, the investigations of massive cohorts from 100,000 to more than 1,000,000 participants were possible because of large collaborative projects combining numerous studies (Colodro-Conde et al. 2017; Kim et al. 2017; Sullivan et al. 2018; Thompson et al. 2014), the availability of biobanks enrolling an unprecedented number of participant (Fan et al. 2008; Kubo and Guest 2017; Sudlow et al. 2015), and collaboration with direct-to-consumer genetic testing companies (Check Hayden 2017). These large-scale GWAS identifying ever-greater numbers of risk loci with ever-smaller individual effects demonstrated that the genetic architecture of common diseases is highly polygenic and their heritability is likely due to the contribution of several thousand (or even more) risk loci across the human genome (Evangelou et al. 2018; Karlsson Linner et al. 2019; Lee et al. 2018; Timmers et al. 2019). One of the main GWAS promises is that the knowledge gained can be used to develop genetic instruments useful to predict disease risk, treatment response, and disease prognosis. Leveraging data generated by large-scale GWAS, a growing number of studies are developing approaches to test the utility of polygenic information with respect to the human phenotypic spectrum (Inouye et al. 2018; Khera et al. 2019; Sparano et al. 2019; Weigl et al. 2018). Although these successful experiments strongly support the movement towards the application of GWAS data to develop new strategies to prevent and treat human diseases, important challenges remain. Among them, one of the most pressing is related to the limited ancestry and ethnic diversity

of large-scale GWAS that have created a large gap between the genetic data available for populations of European descent and non-European human groups (Sirugo et al. 2019). Applying GWAS data generated from European-ancestry cohorts to non-European individuals raise serious issues, including much lower predictive power than that observed in comparisons between like populations (Martin et al. 2019; Mostafavi et al. 2019) and possible biases (e.g., reflecting an accounted population stratification rather than the phenotype of interest) due to the genetic diversity among human populations (Duncan et al. 2019; Martin et al. 2017). The most reliable solution to this problem is to conduct large-scale GWAS in populations with non-European ancestry. Ongoing efforts such as the Million Veteran Program (Gaziano et al. 2016) and the AllofUS Research Program (Sankar and Parker 2017) are investigating multiple ancestry groups representative of the US population to reduce this gap. Although these kinds of projects are expected to eliminate the population disparities in human genetic research, this is likely to be a long-term outcome. To date, to contribute to a more comprehensive understanding of human genetic diversity, we can leverage the data available, combining large-scale genome-wide association datasets generated from cohorts including mainly participants of European descent with reference panels representative of the genetic diversity among worldwide populations (Daub et al. 2013; Hofer et al. 2009; Iorio et al. 2017; Polimanti et al. 2015). In the present study, we focused our attention on the UK Biobank (UKB). This large cohort including more than 500,000 participants with approximately 90% of them as British individuals of European descent (Bycroft et al. 2018). Based on UKB participants of European descent, GWAS have been conducted with respect to the human phenome spectrum, identifying a large number of risk loci surviving the genome-wide significance threshold ($p<5\times10^{-8}$). Using 1,000 Genomes Project (1KG) data, we explored the diversity of these loci, comparing allele frequency differences across worldwide populations. The results obtained showed that allele frequency differences in certain risk loci are significantly different from that expected from randomly selected variants with similar genomic characteristics (i.e., minor allele frequency (MAF), gene density, distance to nearest gene, and linkage disequilibrium (LD) proxies). In some cases, these population differences appear to be due to the

evolutionary events related to local adaptation (i.e., adaptation in response to selective pressure related to the local environment), while other cases may be related to the residual effect of population stratification in UKB GWAS.

Materials and Methods

UK Biobank

The present study was conducted leveraging UKB genome-wide association data. UKB is a large population-based prospective study to explore different life-threatening disorders using information about environment and genes in order to improve diagnosis and treatment (Sudlow et al. 2015). A wide variety of phenotypic information, including socio-demographic and lifestyle factors, electronic health records data, and physiological conditions have been collected for more than 500,000 UKB participants (Bycroft et al. 2018). The genotypes of the whole cohort were defined by applying a bespoke genome-wide DNA microarray that contains about 850,000 genetic variants (including rare, intermediate, and common variants) (Allen et al. 2014). Genetic data were then used to generate genome-wide association datasets that can be employed to explore the genetics of complex traits. The genome-wide datasets used in the present study were derived from the analysis of 361,194 unrelated British participants of European descent. Genome association analyses for over 4,000 phenotypes was conducted using appropriate regression models available in Hail (available at https ://github.com/hail-is/hail) including the first 20 ancestry principal components, sex, age, age^2, sex × age, and sex × age^2 as covariates. The principal components included in the regression model were generated by the UKB investigators using fastPCA algorithm (Galinsky et al. 2016) and considering unrelated subjects and genetic markers pruned for linkage disequilibrium (Bycroft et al. 2018). Details regarding QC criteria, GWAS methods, and the original data are available at https://github.com/Nealelab/UK_Biobank_GWAS/tree/master/imputed-v2-gwas.

1000 Genomes Project Phase3

To dissect the genetic differences of UKB participants with respect to other European samples and other worldwide populations, we used data derived from 1KG Phase3. The 1KG project aims to provide information about common and rare human genetic variation by applying whole-genome sequencing to a large cohort of individuals derived from different populations (Genomes Project et al. 2010; Genomes Project et al. 2012; Genomes Project et al. 2015). The 1KG Phase 3 of the project includes data about 2,504 individuals sampled from 26 populations representative of Africa (AFR), East Asia (EAS), Europe (EUR), South Asia (SAS), and the Americas (admixed; AMR) (Genomes Project et al. 2015). Details regarding alignment, mapping algorithm, SNP (single nucleotide polymorphism) calling, and the data of the project are available at https://www.internationalgenome.org/analysis.

Variants filtering and clumping

We considered genetic association results generated from 361,194 UKB participants of European descent tested with respect to 4,359 phenotypic outcomes including physiological, health, and lifestyle conditions (Supplementary File 1). We focused our attention on variants with a GWAS p-value significance threshold of $P \leq 5\times10^{-8}$ and MAF $\geq$ 5 %. Furthermore, to control the potential inflation in the test statistics, as suggested by the investigators that generated the data (details at http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas), we selected high-confidence associations results generated from variants with at least 25 minor alleles in the smaller group between case or control. To find independent association signals among variants selected, we conducted a P-value-informed clumping with a LD cut-off of $R^2 = 0.1$ within a 1000 kb window.

Allele frequency differences among human populations

We calculated the allele frequency of the index variants identified from the LD-clumping in AFR, EAS, EUR, SAS, and the AMR 1KG superpopulations. Specifically, we tested the following comparisons: i) UKB *vs*. 1KG GBR (British in England and Scotland) reference sample; ii) UKB *vs*. 1KG EUR reference panel (excluding GBR sample); iii) UKB *vs*. each of the non-European 1KG superpopulations (AFR, AMR, EAS, and SAS). For the subsequent analyses, we considered the loci showing allele frequency differences in the top 1% of all index variants investigated with respect to each comparison conducted.

Comparisons with respect to randomly-selected variants matched by genomic characteristics

To verify whether the allele frequency of each variant identified was different from what expected by chance, we generated a control set of matched variants using SNPsnap tool (Pers et al. 2015). This permitted us to identify sets of randomly selected variants SNPs matched to the index variants on the basis of four genomic characteristics: i) MAF, ii) LD proxies, iii) distance to nearest gene, and iv) gene density. Thus, variants identified in the first percentile were used as inputs considering the following parameters: 1KG EUR population (which is the closest reference panel among those available in SNPsnap); LD distance cut-off of R^2=0.5; ±5% point deviation; ±50% of gene density relative deviation; ±50% of relative deviation of the distance to nearest gene; ±50% of relative deviation of LD proxies. For each index variant identified in the initial screening described in the section above, we extracted up to 10,000 matched SNP, excluding the HLA region due to its complex LD structure. Based on the corresponding randomly-selected genomically-matched sets, we calculated empirical p values for each index variant tested and considered type I error rate at 1% as the significance threshold. Finally, we checked whether the significative index variants showed

allele frequency mismatches and mismapping using previously generated data available at http://kunertgraf.com/data/biobank.html (Kunert-Graf et al. 2020).

Cross-Ancestry LD comparison and Functional Annotation

For the index variants with empirical p values surviving statistical significance, we conducted computational analyses to explore their functional consequences. Using LDlink (Machiela and Chanock 2015, 2018), we tested the effect of the LD structure variability across human populations on the ability of differentiated index variants to tag (measured as LD R^2) functional variants in the surrounding regions (±500Kb). RegulomeDB (Boyle et al. 2012) was used to score the regulatory effect of the tagged variants on the basis of high-throughput, experimental data sets as well as computational predictions and manual annotations. LD R^2>0.50 and RegulomeDB score = 1a-f (Supplementary File 2) were used as criteria to identify functional tag SNPs.

Enrichment analysis for significant phenotypic traits

To test whether traits related to differentiated loci were overrepresented with respect to certain phenotypic domains, we performed χ^2 test comparing whether the proportions of the phenotypic distribution observed with respect to the identified loci are significantly different from the ones of the overall distribution observed across the 4,000+ UKB phenotypes analyzed.

Pan-UK Biobank data

To investigate the loci identified in non-European ancestral groups, we used the newly-released Pan-UKB genome-wide association statistics related to 7,221 phenotypes: 6,636 of AFR individuals; 980 AMR individuals; 8,876 individuals of Central/South Asian ancestry (CSA); 2,709

EAS individuals. A detailed description of the methods used to generate these data is available at https://pan.ukbb.broadinstitute.org/. Using these data, we investigated whether the EUR associations of the index variants were also concordant in AFR, AMR, CSA, and EAS. Pan-UKB data are available at https://pan.ukbb.broadinstitute.org/downloads.

Results

Based on genome-wide significant associations ($p< P \leq 5\times10\text{-}^{8}$) across the UKB phenotypic spectrum assessed (4,359 traits), we identified a total of 15,327 LD-independent risk alleles. Among these, we identified 154 index variants showing allelic frequency differences in the top 1% with respect to the three comparisons conducted: i) UKB *vs*. 1KG GBR; ii) UKB *vs*. 1KG EUR (excluding GBR sample); iii) UKB *vs*. each of the non-European 1KG superpopulations (AFR, AMR, EAS, and SAS) (Figure 1; Supplementary File 3). To test whether the allele frequency differences were significantly different from what expected by chance, we generated a control set of 10,000 variants matched by genomic characteristics (i.e., gene density, distance to the nearest gene, and the number of LD proxies) for each of the index variants (Supplementary File 4). For all significative index variants, we reported their phenotypic associations and those related to the variants in LD with them in Supplementary File 5. In line with the fact that both samples are representative of the genetic variability of British populations, no significant difference was observed in the allele frequency of index variants between the UKB cohort and 1KG GBR panel (Supplementary File 4). Conversely, when comparing UKB with other population groups, allele frequency differences were observed in loci associated with several traits. The differentiated loci appear to be associated mainly with observed that anthropometric traits and hematologic parameters. Across multiple populations comparisons, we observed that the phenotypic enrichments were significantly different from what expected by chance ($5.39\times10^{-7}<p<2.75\times10^{-79}$; Figure 2).

UKB British participants vs. 1,000 Genomes Project non-British Europeans

Considering UKB *vs*. 1KG EUR reference sample (excluding GBR sample), we identified several significant different risk loci associated with different traits (Table 1; Figure 2). Among them, we observed some traits related to anthropometric measurements and white blood cells and platelet parameters: *standing height* (rs74945666, allele=T [0.909 vs. 0.030], p=1.48×10^{-17}); *Heel Broadband ultrasound attenuation direct entry* (rs200033476, allele= C [0.917 vs. 0.018], p=4.01×10^{-9}); *immature reticulocyte fraction* (rs34690548, allele= CA [0.904 vs. 0.0303], p=2.85×10^{-320}); *eosinophil percentage* (rs200725444, allele= A [0.927 vs. 0.047], p=8.60×10^{-14}); *platelet count* (rs201088941, allele= TA [0.922 vs. 0.029], p=6.72×10^{-10}). Because of the similar LD structure, we did not observe differences between UKB and EURnoGBR populations with respect to the ability of the index variants to tag functional elements (Supplementary File 6).

UKB British participants vs. 1,000 Genomes Project Africans

Comparing UKB with 1KG AFR superpopulation, we identified loci associated with several conditions (Table 1, Figure 2, Supplementary File 5). Particularly, these were related to anthropometric traits: *leg impedance* (rs3749748, allele=T [0.944 vs. 0.017], p=1.75×10^{-123}); *standing height* (rs157573, allele=A [0.876 vs. 0.015], p=8.06×10^{-33},; rs35497246, allele=C [0.058 vs. 0.893], p=2.53×10^{-13}; rs1812378, allele=T [0.931 vs. 0.031], p=4.23×10^{-12}; rs42525, allele=C [0.877 vs. 0.021], p=1.90×10^{-10}; rs625670, allele=A [0.998 vs. 0.002], p=4.29×10^{-8}); *arm impedance* (rs1881131, allele=A [0.942 vs. 0.031], p=3.34×10^{-16}); *whole body water mass* (rs475591, allele=T [0.998 vs. 0.110], p=1.56×10^{-12}); *Heel Broadband ultrasound attenuation direct entry* (rs200033476, allele=C [0.917 vs. 0.046], p=4.01×10^{-9}). Additionally, we observed several associations with hematologic parameters: *lymphocyte count* (rs451367, allele=T [0.950 vs. 0.014], p=1.08×10^{-27}; rs3748022, allele=T, p=9.19×10^{-13}); *immature reticulocyte fraction* (rs603620, allele=A [0.948 vs. 0.019], p=1.55×10^{-27}); *monocyte percentage* (rs456798, allele=T [0.059 vs. 0.918], p=3.70×10^{-22},;

rs625465, allele=G [0.932 vs. 0.082], p=2.27×10^{-8}); *mean platelet (thrombocyte) volume* (rs171042; p=3.97×10^{-18}); r*ed blood cell (erythrocyte) distribution width* (rs374361, allele=T [0.900 vs. 0.012], p=3.03×10^{-15}; rs55959450, allele=C [0.947 vs. 0.013], p=1.84×10^{-9}); *platelet count* (Phesant:30080_irnt; rs556562, allele=A [0.943 vs. 0.084], p=4.84×10^{-15}); *eosinophil percentage* (rs200725444, allele=A [0.927 vs. 0.007], p=8.60×10^{-14}; *monocyte count* (rs55881864, allele=T [0.912 vs. 0.006], p=7.29×10^{-13}); *reticulocyte percentage* (rs57236847, allele=G [0.891 vs. 0.037], p=2.53×10^{-8}). We observed also other traits that are related to anthropometric and hematologic phenotypes: *palmar fascial fibromatosis* (rs651985, allele= G [0.903 vs. 0.044], p=2.29×10^{-42}); *systolic blood pressure, automated reading* (rs604723, allele= T [0.899 vs. 0.008], p=6.73×10^{-40}; rs55815739, allele=A, p=2.23×10^{-8}); *6mm asymmetry index* (; rs55971426, allele=G [0.887 vs. 0.007], p=1.60×10^{-10}).

Regarding cross-ancestry LD analysis, we observed that several index variants showed different tagging properties with respect to functional elements. Indeed, while only rs3749748 tag functional elements in both populations (Supplementary File 7, Supplementary File 8-Figure S8.1), several index variants (i.e., rs157573, rs451367, rs475591, rs625465) are in LD with functional loci in EUR populations but not in AFR populations (Supplementary File 7, Supplementary File 8-Figure S8.2-5).

UKB British participants vs. 1,000 Genomes Project Admixed Americans

The allele frequency differences between UKB and 1KG AMR are related to loci mainly associated to hematologic traits (Figure 2; Table 1): *immature reticulocyte fraction* (rs34690548, allele=CA [0.904 vs. 0.052], p=2.85×10^{-320}); *reticulocyte percentage* (rs321600, allele=A [0.997 vs. 0.123], p=1.54×10^{-30}); *lymphocyte count* (rs451367, allele=T [0.950 vs. 0.091], p=1.08×10^{-27}); *neutrophill count* (rs571497, allele=A [0.943 vs. 0.109], p=2.58×10^{-23}; rs4544340, allele=T [0.949 vs. 0.087], p=1.86×10^{-10}); *mean platelet (thrombocyte) volume* (rs171042, allele=T [0.951 vs. 0.127 p=3.97×10^{-}

[18]); *eosinophil percentage* (rs200725444, allele=A [0.927 vs. 0.055], p=8.60×10^{-14}); *platelet distribution width* (rs1875103, allele=T [0.998 vs. 0.075], p=4.43×10^{-12}); *red blood cell (erythrocyte) distribution width* (rs55959450, allele=C [0.947 vs. 0.074], p=1.84×10^{-9}); *standing height* (rs1812378, allele=T [0.931 vs. 0.089], p=4.23×10^{-12}); *heel broadband ultrasound attenuation, direct entry* (rs200033476, allele=C [0.917 vs. 0.012], p=4.01×10^{-9}); *systolic blood pressure, automated reading* (rs55815739, allele=A [0.928 vs. 0.052], p=2.23×10^{-8}). Comparing the LD structure of UKB and AMR populations, we observed two variants (rs571497, rs4544340) tagging functional elements in both populations (Supplementary File 9, Supplementary File 10-Figure S10.1-2). Conversely, rs451367 associated with *lymphocyte count* is in LD (R^2=0.61) with a functional SNP (rs4808485; RegulomeDB=1a) in British individuals but not in AMR populations (Supplementary File 9; Supplementary File 10-FigureS10.3).

UKB British participants vs. 1,000 Genomes Project East Asians

We observed allele frequencies differences between UKB *vs.* 1KG EAS in loci associated with parameters (Table 1, Figure 2): *monocyte count* (rs3732378, allele=A [0.941 vs. 0.029], p=1.64×10^{-67}; rs55881864, allele=T [0.912 vs. 0.001], p=7.29×10^{-13}); *immature reticulocyte fraction* (rs6014986, allele=A [0.911 vs. 0.016], p=3.05×10^{-43}; rs603620, allele=A [0.948 vs. 0.003], p=1.55×10^{-27},); *eosinophil percentage* (rs34495, allele=T [0.916 vs. 0.043], p=1.14×10^{-15}; rs200725444, allele=A [0.927 vs. 0.024], p=8.60×10^{-14}); *reticulocyte percentage* (rs321600, allele=A [0.997 vs. 0.133], p=1.54×10^{-30}); *neutrophil count* (rs571497, allele=A [0.943 vs 0.001], p=2.58×10^{-23}); *platelet distribution width* (rs1875103, allele=T [0.999 vs. 0], p=4.43×10^{-12}); *platelet crit* (rs9932254, allele=C [0.925 vs. 0.025], p=8.55×10^{-11}); r*ed blood cell (erythrocyte) distribution width* (rs55959450, allele=C [0.947 vs. 0.003], p=1.84×10^{-9}). Similarly to the other ancestry comparisons, several UKB-EAS differentiated loci are associated with anthropometric traits: *arm impedance* (rs1881131, allele=A [0.943 vs. 0.002], p=3.34×10^{-16}); *whole body water mass* (allele=T [0.998 vs.

0.086], p=1.56×10^{-12}); *standing height* (rs2861745, allele=G [0.876 vs. 0], p=8.19×10^{-11}; rs42525, allele=C [0.877 vs. 0.024], p=1.90×10^{-10}; rs625670, allele=A [0.998 vs. 0], p=4.29×10^{-8}); *Heel Broadband ultrasound attenuation* (rs200033476, allele=C [0.917 vs. 0.002], p=4.01×10^{-9})). Finally, additional variants showing allele frequency differences were related to: *palmar fascial fibromatosis* (rs651985, allele=G [0.903 vs. 0], p=2.29×10^{-42}; rs55971426l, allele=G [0.887 vs. 0.018] p=1.60×10^{-10}); *spherical power* (rs56207218, allele=C [0.896 vs. 0.009], p=1.23×10^{-9}); *systolic blood pressure, automated reading* (rs55815739, allele=A [0.923 vs. 0.003], p=2.23×10^{-8};).

Comparing UKB and EAS LD structures, we observed that certain index variants tag different functional SNPs depending on the population considered (Supplementary File 11; Supplementary File 12-FigureS12.1-3). Conversely, rs571497 and rs56207218, associated with *Neutrophil count* and *Spherical power* respectively, are in LD (R^2>0.5) with functional elements in both populations (Supplementary File 11; Supplementary File 12-FigureS12.4-5).

UK Biobank British participants vs. 1KG South Asians

Similarly, to what observed in the other ancestry comparisons, allele frequency differences between UKB and SAS were observed in variants associated with anthropometric traits and hematologic parameters. These included *immature reticulocyte fraction* (rs34690548, allele=CA [0.904 vs. 0.025], p=2.85×10^{-320}); *standing height* (rs74945666, allele=T [0.909 vs. 0.061], p=1.48×10^{-17}); *eosinophil percentage* (rs200725444, allele=A [0.927 vs. 0.030], p=8.60×10^{-14}); *Heel Broadband ultrasound attenuation* (rs200033476, allele=C [0.917 vs. 0.013], p=4.01×10^{-9}); (Figure 2; Table 1). The UKB-SAS differentiated loci did not show evidence of regulatory function or tagging of regulatory SNPs in any of the two populations (Supplementary File 13).

Cross-ancestry association analysis in non-European UK Biobank participants

Considering Pan-UK Biobank data related to non-European populations, we tested whether the differentiated variants and their functional tagged SNPs were associated with their related phenotypic

traits in AFR, AMR, EAS and, CSA participants from UKB. Due to the dramatic difference in sample size between UKB participants of European descent (N=361,194) and UKB participants of non-European descent (980<N<8,876), only two variants differentiated between UKB and AFR were nominally replicated in UKB-AFR participants with respect to their related conditions (rs171042, mean platelet volume; rs374361, red blood cell distribution width) (Supplementary File 14).

Discussion

To provide a more comprehensive understanding of the genetics of complex traits across worldwide populations, we assessed loci associated with complex traits UKB participants of European descent that present allele frequency differences in other human groups worldwide populations leveraging 1KG reference data (Rees et al. 2020). As expected, there was no significant difference in the allele frequency of index variants between UKB cohort and 1KG GBR population, confirming that both samples are presentative of the genetic structure of the British population. Conversely, certain loci associated with complex traits in UKB participants of European descent showed allele frequency differences significantly different from what expected by chance when compared with non-British European populations (EURnoGBR) and with AFR, AMR, EAS, and SAS ancestries. Comparing the LD structure across these human groups, we observed that differentiated loci can tag differently regulatory elements, changing the functional meaning of genome-wide significant variants observed in UKB participants of European descent when analyzed in the context of other ancestral groups.

Considering the traits related to differentiated loci, we observed significant overrepresentation for anthropometric traits and hematologic parameters across multiple ancestry comparisons ($5.39\times10^{-7}<p<2.75\times10^{-79}$; Figure 2). These phenotypic categories are well-known to be differentiated across human populations due to evolutionary pressures and human demographic history (Guo et al. 2018). Several studies investigated the underlying mechanisms that shaped the genetic architecture of

anthropometric measures among human populations (Berg et al. 2019; Guo et al. 2018; Park et al. 2016; Polimanti et al. 2016; Turchin et al. 2012; Wood et al. 2014). In particular, gradients were observed polygenic height scores within European populations (north to south) and across Eurasia (east to west) (Berg et al. 2019; Turchin et al. 2012). Several hypotheses have been made regarding the presence of evolutionary pressures shaping the genetic architecture of height and other anthropometric traits (Guo et al. 2018). However, Sohail et al. (2019) demonstrated that the signature of polygenic adaptation on height is overestimated due to GWAS uncorrected stratification. Comparing results obtained from UKB and GIANT (Genetic Investigation of ANthropometric Traits) consortium, population-level differences in genetic height showed robust evidence only at highly significant SNPs while less significant P values were affected by residual population stratification. The findings provided by Sohail et al. (2019) indicate that previous analyses cannot distinguish the proportion of the population differences of genetic height due to evolutionary pressures vs. population stratification biases. In our analyses, we considered genome-wide significant variants ($p<5\times10^{-8}$) identified from UKB participants of European descent. In line with the study of Sohail et al. (2019), we expect that the variants investigated in the present study are less affected by population stratification. Accordingly, the observation that loci differentiated between UKB and 1KG reference populations (data independent from UKB) are enriched for anthropometric traits may support the involvement of evolutionary pressure and population demographic history in shaping the genetic architecture of anthropometric traits.

The second strong enrichment for loci differentiated between UKB and worldwide populations is related to hematologic parameters including traits related to red blood cell (RBC), white blood cell (WBC), and platelet. Similarly to anthropometric traits, several studies assessed genetic variation of hematologic phenotypes traits across human populations, observing strong differences in their geographic distribution (Beutler and West 2005; Chambers et al. 2009; Chen et al. 2020; Eicher et al.

2016; Ganesh et al. 2009; Hodonsky et al. 2017; Kamatani et al. 2010; Rappoport et al. 2019; Schick et al. 2016). This inter-population genetic variability is probably linked to the evolutionary pressures of infectious diseases (Astle et al. 2016; Dominguez-Andres and Netea 2019; Polimanti et al. 2016; Raffield et al. 2018). Risk alleles associated with RBC traits show high frequencies in African malaria-prone regions where there is a high prevalence of anemia and microcytosis (Barrera-Reyes and Tejero 2019; Dominguez-Andres and Netea 2019; Raffield et al. 2018). WBC and platelet-associated loci appear also differentiated across human populations (Chen et al. 2020; Eicher et al. 2016; Rappoport et al. 2019; Schick et al. 2016). Examples of this are i) the *Duffy/DARC* null variant in AFR individuals that is associated with low WBC and neutrophil counts and confers a selective advantage against malaria (Rappoport et al. 2019); ii) *GATA2* genetic variation that reflects differences in eosinophil and basophil counts in Japanese population and monocyte and basophil counts in Europeans (Okada and Kamatani 2012), and iii) the presence of population-specific risk could partially account for the high platelet counts observed in Hispanic/Latinos (lower respect to other human population) (Schick et al. 2016).

Finally, differentiated loci that showed genome-wide significant associations in UKB participants of European descent were not replicated in non-European UKB participants (i.e., AFR, AMR, EAS, and CSA) independently from their tagging of functional elements across populations. Due to the dramatic change in sample size (N=361,194 vs. 980<N<8,876), this lack of replication is likely due to the strong reduction of statistical power in the non-European association analyses. Unfortunately, this is in line with the well-known issue related to the lack of non-European genome-wide data (Sirugo et al. 2019). As mentioned previously, many factors influence how causal variants are captured by tagging SNPs identified in a single population (Rees et al. 2020; Sirugo et al. 2019). We showed that loci associated with complex traits and differentiated across human populations can show different cross-ancestry LD tagging properties that can affect the functional meaning of the variant tested in the context of the ancestry group investigated. Thus, a large amount of genetic data of diverse

populations are needed to provide a more comprehensive understanding of the molecular mechanisms at the basis of complex diseases.

In conclusion, this study provided novel evidence regarding the predisposition to complex traits in the context of human genetic variation. We observed that loci differentiated are enriched for traits that may be shaped by human evolutionary history (i.e., anthropometric traits and hematologic parameters). Additionally, we showed how the LD structure of human populations can affect the functional meaning of loci known to be associated with a specific ancestry group. Finally, although our data contribute to increasing our knowledge regarding cross-ancestry genetic predisposition to complex traits, they also clearly indicate that there is an urgent need for greater population diversity in genome-wide studies.

ACKNOWLEDGMENTS

We thank the participants and investigators of the UK Biobank, the Neale lab for generating the genome-wide data used in the present study, and Dr. Riccardo Pennacchi for the computational support. R.P. acknowledges support from the National Institutes of Health via R21 DA047527 and R21 DC018098. The sources of funding had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Conflict of interest: The authors reported no biomedical financial interests or potential conflicts of interest.

Ethical approval: This study was conducted using summary association data generated by previous studies. Owing to the use of previously collected, deidentified, aggregated data, this study did not require institutional review board approval.

Data Availability: Data supporting the findings of this study are available within this article and its additional files. UK Biobank GWAS summary association data are available at https://github.com/Nealelab/UK_Biobank_GWAS/tree/master/imputed-v2-gwas

References

Allen NE, Sudlow C, Peakman T, Collins R, Biobank UK (2014) UK biobank data: come and get it. Sci Transl Med 6: 224ed4. doi: 10.1126/scitranslmed.3008601

Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, Lambourne JJ, Sivapalaratnam S, Downes K, Kundu K, Bomba L, Berentsen K, Bradley JR, Daugherty LC, Delaneau O, Freson K, Garner SF, Grassi L, Guerrero J, Haimel M, Janssen-Megens EM, Kaan A, Kamat M, Kim B, Mandoli A, Marchini J, Martens JHA, Meacham S, Megy K, O'Connell J, Petersen R, Sharifi N, Sheard SM, Staley JR, Tuna S, van der Ent M, Walter K, Wang SY, Wheeler E, Wilder SP, Iotchkova V, Moore C, Sambrook J, Stunnenberg HG, Di Angelantonio E, Kaptoge S, Kuijpers TW, Carrillo-de-Santa-Pau E, Juan D, Rico D, Valencia A, Chen L, Ge B, Vasquez L, Kwan T, Garrido-Martin D, Watt S, Yang Y, Guigo R, Beck S, Paul DS, Pastinen T, Bujold D, Bourque G, Frontini M, Danesh J, Roberts DJ, Ouwehand WH, Butterworth AS, Soranzo N (2016) The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell 167: 1415-1429 e19. doi: 10.1016/j.cell.2016.10.042

Barrera-Reyes PK, Tejero ME (2019) Genetic variation influencing hemoglobin levels and risk for anemia across populations. Ann N Y Acad Sci 1450: 32-46. doi: 10.1111/nyas.14200

Berg JJ, Zhang X, Coop G (2019) Polygenic Adaptation has Impacted Multiple Anthropometric Traits. bioRxiv: 167551. doi: 10.1101/167551

Beutler E, West C (2005) Hematologic differences between African-Americans and whites: the roles of iron deficiency and alpha-thalassemia on hemoglobin levels and mean corpuscular volume. Blood 106: 740-5. doi: 10.1182/blood-2005-02-0713

Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22: 1790-7. doi: 10.1101/gr.137323.112

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47: D1005-D1012. doi: 10.1093/nar/gky1120

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562: 203-209. doi: 10.1038/s41586-018-0579-z

Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L, Freimer NB, Srai SK, Maxwell PH, Sternberg MJ, Ruokonen A, Abecasis G, Jarvelin MR, Scott J, Elliott P, Kooner JS (2009) Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. Nat Genet 41: 1170-2. doi: 10.1038/ng.462

Check Hayden E (2017) The rise and fall and rise again of 23andMe. Nature 550: 174-177. doi: 10.1038/550174a

Chen M-H, Raffield LM, Mousas A, Sakaue S, Huffman JE, Jiang T, Akbari P, Vuckovic D, Bao EL, Moscati A, Zhong X, Manansala R, Laplante V, Chen M, Lo KS, Qian H, Lareau CA, Beaudoin M, Akiyama M, Bartz TM, Ben-Shlomo Y, Beswick A, Bork-Jensen J, Bottinger EP, Brody JA, van Rooij FJA, Chitrala K, Cho K, Choquet H, Correa A, Danesh J, Di Angelantonio E, Dimou N, Ding J, Elliott P, Esko T, Evans MK, Floyd JS, Broer L, Grarup N, Guo MH, Greinacher A, Haessler J, Hansen T, Howson JMM, Huang W, Jorgenson E, Kacprowski T, Kähönen M, Kamatani Y, Kanai M, Karthikeyan S, Koskeridis F, Lange LA,

Lehtimäki T, Lerch MM, Linneberg A, Liu Y, Lyytikäinen L-P, Manichaikul A, Matsuda K, Mohlke KL, Mononen N, Murakami Y, Nadkarni GN, Nauck M, Nikus K, Ouwehand WH, Pankratz N, Pedersen O, Preuss M, Psaty BM, Raitakari OT, Roberts DJ, Rich SS, Rodriguez BAT, Rosen JD, Rotter JI, Schubert P, Spracklen CN, Surendran P, Tang H, Tardif J-C, Ghanbari M, Völker U, Völzke H, Watkins NA, Zonderman AB, Million Veteran Program V, Wilson PWF, Li Y, Butterworth AS, Gauchat J-F, Chiang CWK, Li B, Loos RJF, Astle WJ, Evangelou E, Sankaran VG, Okada Y, et al. (2020) Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. bioRxiv: 2020.01.17.910497. doi: 10.1101/2020.01.17.910497

Colodro-Conde L, Cross SM, Lind PA, Painter JN, Gunst A, Jern P, Johansson A, Lund Maegbaek M, Munk-Olsen T, Nyholt DR, Ordonana JR, Paternoster L, Sanchez-Romera JF, Wright MJ, Medland SE (2017) Cohort Profile: Nausea and vomiting during pregnancy genetics consortium (NVP Genetics Consortium). Int J Epidemiol 46: e17. doi: 10.1093/ije/dyv360

Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L (2013) Evidence for polygenic adaptation to pathogens in the human genome. Mol Biol Evol 30: 1544-58. doi: 10.1093/molbev/mst080

Dominguez-Andres J, Netea MG (2019) Impact of Historic Migrations and Evolutionary Processes on Human Immunity. Trends Immunol 40: 1105-1119. doi: 10.1016/j.it.2019.10.001

Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B (2019) Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun 10: 3328. doi: 10.1038/s41467-019-11112-0

Eicher JD, Chami N, Kacprowski T, Nomura A, Chen MH, Yanek LR, Tajuddin SM, Schick UM, Slater AJ, Pankratz N, Polfus L, Schurmann C, Giri A, Brody JA, Lange LA, Manichaikul A, Hill WD, Pazoki R, Elliot P, Evangelou E, Tzoulaki I, Gao H, Vergnaud AC, Mathias RA, Becker DM, Becker LC, Burt A, Crosslin DR, Lyytikainen LP, Nikus K, Hernesniemi J, Kahonen M, Raitoharju E, Mononen N, Raitakari OT, Lehtimaki T, Cushman M, Zakai NA, Nickerson DA, Raffield LM, Quarells R, Willer CJ, Peloso GM, Abecasis GR, Liu DJ, Global Lipids Genetics C, Deloukas P, Samani NJ, Schunkert H, Erdmann J, Consortium CAE, Myocardial Infarction Genetics C, Fornage M, Richard M, Tardif JC, Rioux JD, Dube MP, de Denus S, Lu Y, Bottinger EP, Loos RJ, Smith AV, Harris TB, Launer LJ, Gudnason V, Velez Edwards DR, Torstenson ES, Liu Y, Tracy RP, Rotter JI, Rich SS, Highland HM, Boerwinkle E, Li J, Lange E, Wilson JG, Mihailov E, Magi R, Hirschhorn J, Metspalu A, Esko T, Vacchi-Suzzi C, Nalls MA, Zonderman AB, Evans MK, Engstrom G, Orho-Melander M, Melander O, O'Donoghue ML, Waterworth DM, Wallentin L, White HD, Floyd JS, Bartz TM, Rice KM, Psaty BM, Starr JM, Liewald DC, Hayward C, Deary IJ, et al. (2016) Platelet-Related Variants Identified by Exomechip Meta-analysis in 157,293 Individuals. Am J Hum Genet 99: 40-55. doi: 10.1016/j.ajhg.2016.05.005

Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, Ng FL, Evangelou M, Witkowska K, Tzanis E, Hellwege JN, Giri A, Velez Edwards DR, Sun YV, Cho K, Gaziano JM, Wilson PWF, Tsao PS, Kovesdy CP, Esko T, Magi R, Milani L, Almgren P, Boutin T, Debette S, Ding J, Giulianini F, Holliday EG, Jackson AU, Li-Gao R, Lin WY, Luan J, Mangino M, Oldmeadow C, Prins BP, Qian Y, Sargurupremraj M, Shah N, Surendran P, Theriault S, Verweij N, Willems SM, Zhao JH, Amouyel P, Connell J, de Mutsert R, Doney ASF, Farrall M, Menni C, Morris AD, Noordam R, Pare G, Poulter NR, Shields DC, Stanton A, Thom S, Abecasis G, Amin N, Arking DE, Ayers KL, Barbieri CM, Batini C, Bis JC, Blake T, Bochud M, Boehnke M, Boerwinkle E, Boomsma DI, Bottinger EP, Braund PS, Brumat M, Campbell A, Campbell H, Chakravarti A, Chambers JC, Chauhan G, Ciullo M, Cocca M, Collins F, Cordell HJ, Davies G, de Borst MH, de Geus EJ, Deary IJ, Deelen J, Del Greco MF, Demirkale CY, Dorr M, Ehret GB, Elosua R, Enroth S, Erzurumluoglu AM, Ferreira T, Franberg M, Franco OH, Gandin I, et al.

(2018) Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 50: 1412-1425. doi: 10.1038/s41588-018-0205-x

Fan CT, Lin JC, Lee CH (2008) Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island. Pharmacogenomics 9: 235-46. doi: 10.2217/14622416.9.2.235

Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, Price AL (2016) Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am J Hum Genet 98: 456-472. doi: 10.1016/j.ajhg.2015.12.022

Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, Kuhnel B, Aspelund T, Yang Q, Tanaka T, Jaffe A, Bis JC, Verwoert GC, Teumer A, Fox CS, Guralnik JM, Ehret GB, Rice K, Felix JF, Rendon A, Eiriksdottir G, Levy D, Patel KV, Boerwinkle E, Rotter JI, Hofman A, Sambrook JG, Hernandez DG, Zheng G, Bandinelli S, Singleton AB, Coresh J, Lumley T, Uitterlinden AG, Vangils JM, Launer LJ, Cupples LA, Oostra BA, Zwaginga JJ, Ouwehand WH, Thein SL, Meisinger C, Deloukas P, Nauck M, Spector TD, Gieger C, Gudnason V, van Duijn CM, Psaty BM, Ferrucci L, Chakravarti A, Greinacher A, O'Donnell CJ, Witteman JC, Furth S, Cushman M, Harris TB, Lin JP (2009) Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet 41: 1191-8. doi: 10.1038/ng.466

Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, O'Leary TJ (2016) Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70: 214-23. doi: 10.1016/j.jclinepi.2015.09.016

Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-73. doi: 10.1038/nature09534

Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56-65. doi: 10.1038/nature11632

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. Nature 526: 68-74. doi: 10.1038/nature15393

Guo J, Wu Y, Zhu Z, Zheng Z, Trzaskowski M, Zeng J, Robinson MR, Visscher PM, Yang J (2018) Global genetic differentiation of complex traits shaped by natural selection in humans. Nat Commun 9: 1865. doi: 10.1038/s41467-018-04191-y

Hodonsky CJ, Jain D, Schick UM, Morrison JV, Brown L, McHugh CP, Schurmann C, Chen DD, Liu YM, Auer PL, Laurie CA, Taylor KD, Browning BL, Li Y, Papanicolaou G, Rotter JI, Kurita R, Nakamura Y, Browning SR, Loos RJF, North KE, Laurie CC, Thornton TA, Pankratz N, Bauer DE, Sofer T, Reiner AP (2017) Genome-wide association study of red blood cell traits in Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos. PLoS Genet 13: e1006760. doi: 10.1371/journal.pgen.1006760

Hofer T, Ray N, Wegmann D, Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Ann Hum Genet 73: 95-108. doi: 10.1111/j.1469-1809.2008.00489.x

Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY, Kaptoge S, Brozynska M, Wang T, Ye S, Webb TR, Rutter MK, Tzoulaki I, Patel RS, Loos RJF, Keavney B, Hemingway H, Thompson J, Watkins H, Deloukas P, Di Angelantonio E, Butterworth AS, Danesh J, Samani NJ, Group UKBCCCW (2018) Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol 72: 1883-1893. doi: 10.1016/j.jacc.2018.07.079

Iorio A, De Angelis F, Di Girolamo M, Luigetti M, Pradotto LG, Mazzeo A, Frusconi S, My F, Manfellotto D, Fuciarelli M, Polimanti R (2017) Population diversity of the genetically

determined TTR expression in human tissues and its implications in TTR amyloidosis. BMC Genomics 18: 254. doi: 10.1186/s12864-017-3646-1

Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 42: 210-5. doi: 10.1038/ng.531

Karlsson Linner R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, Lebreton M, Tino SP, Abdellaoui A, Hammerschlag AR, Nivard MG, Okbay A, Rietveld CA, Timshel PN, Trzaskowski M, Vlaming R, Zund CL, Bao Y, Buzdugan L, Caplin AH, Chen CY, Eibich P, Fontanillas P, Gonzalez JR, Joshi PK, Karhunen V, Kleinman A, Levin RZ, Lill CM, Meddens GA, Muntane G, Sanchez-Roige S, Rooij FJV, Taskesen E, Wu Y, Zhang F, and Me Research T, e QC, International Cannabis C, Social Science Genetic Association C, Auton A, Boardman JD, Clark DW, Conlin A, Dolan CC, Fischbacher U, Groenen PJF, Harris KM, Hasler G, Hofman A, Ikram MA, Jain S, Karlsson R, Kessler RC, Kooyman M, MacKillop J, Mannikko M, Morcillo-Suarez C, McQueen MB, Schmidt KM, Smart MC, Sutter M, Thurik AR, Uitterlinden AG, White J, Wit H, Yang J, Bertram L, Boomsma DI, Esko T, Fehr E, Hinds DA, Johannesson M, Kumari M, Laibson D, Magnusson PKE, Meyer MN, Navarro A, Palmer AA, Pers TH, Posthuma D, Schunk D, Stein MB, Svento R, Tiemeier H, Timmers P, Turley P, Ursano RJ, Wagner GG, Wilson JF, Gratten J, Lee JJ, Cesarini D, Benjamin DJ, Koellinger PD, Beauchamp JP (2019) Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 51: 245-257. doi: 10.1038/s41588-018-0309-3

Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, Distefano M, Senol-Cosar O, Haas ME, Bick A, Aragam KG, Lander ES, Smith GD, Mason-Suares H, Fornage M, Lebo M, Timpson NJ, Kaplan LM, Kathiresan S (2019) Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell 177: 587-596 e9. doi: 10.1016/j.cell.2019.03.028

Kim Y, Han BG, Ko GESg (2017) Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. Int J Epidemiol 46: e20. doi: 10.1093/ije/dyv316

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308: 385-9. doi: 10.1126/science.1109557

Kubo M, Guest E (2017) BioBank Japan project: Epidemiological study. J Epidemiol 27: S1. doi: 10.1016/j.je.2016.11.001

Kunert-Graf J, Sakhanenko N, Galas D (2020) Allele Frequency Mismatches and Apparent Mismappings in UK Biobank SNP Data. bioRxiv: 2020.08.03.235150. doi: 10.1101/2020.08.03.235150

Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linner R, Fontana MA, Kundu T, Lee C, Li H, Li R, Royer R, Timshel PN, Walters RK, Willoughby EA, Yengo L, andMe Research T, Cogent, Social Science Genetic Association C, Alver M, Bao Y, Clark DW, Day FR, Furlotte NA, Joshi PK, Kemper KE, Kleinman A, Langenberg C, Magi R, Trampush JW, Verma SS, Wu Y, Lam M, Zhao JH, Zheng Z, Boardman JD, Campbell H, Freese J, Harris KM, Hayward C, Herd P, Kumari M, Lencz T, Luan J, Malhotra AK, Metspalu A, Milani L, Ong KK, Perry JRB, Porteous DJ, Ritchie MD, Smart MC, Smith BH, Tung JY, Wareham NJ, Wilson JF, Beauchamp JP, Conley DC, Esko T, Lehrer SF, Magnusson PKE, Oskarsson S, Pers TH, Robinson MR, Thom K, Watson C, Chabris CF, Meyer MN, Laibson DI, Yang J, Johannesson M, Koellinger PD, Turley P, Visscher PM, Benjamin DJ, Cesarini D (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50: 1112-1121. doi: 10.1038/s41588-018-0147-3

Machiela MJ, Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31: 3555-7. doi: 10.1093/bioinformatics/btv402

Machiela MJ, Chanock SJ (2018) LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. Bioinformatics 34: 887-889. doi: 10.1093/bioinformatics/btx561

Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE (2017) Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet 100: 635-649. doi: 10.1016/j.ajhg.2017.03.004

Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51: 584-591. doi: 10.1038/s41588-019-0379-x

Mostafavi H, Harpak A, Conley D, Pritchard JK, Przeworski M (2019) Variable prediction accuracy of polygenic scores within an ancestry group. bioRxiv: 629949. doi: 10.1101/629949

Okada Y, Kamatani Y (2012) Common genetic factors for hematological traits in humans. J Hum Genet 57: 161-9. doi: 10.1038/jhg.2012.2

Park H, Li X, Song YE, He KY, Zhu X (2016) Multivariate Analysis of Anthropometric Traits Using Summary Statistics of Genome-Wide Association Studies from GIANT Consortium. PLoS One 11: e0163912. doi: 10.1371/journal.pone.0163912

Pers TH, Timshel P, Hirschhorn JN (2015) SNPsnap: a Web-based tool for identification and annotation of matched SNPs. Bioinformatics 31: 418-20. doi: 10.1093/bioinformatics/btu655

Polimanti R, Yang BZ, Zhao H, Gelernter J (2016) Evidence of Polygenic Adaptation in the Systems Genetics of Anthropometric Traits. PLoS One 11: e0160654. doi: 10.1371/journal.pone.0160654

Polimanti R, Yang C, Zhao H, Gelernter J (2015) Dissecting ancestry genomic background in substance dependence genome-wide association studies. Pharmacogenomics 16: 1487-98. doi: 10.2217/pgs.15.91

Raffield LM, Ulirsch JC, Naik RP, Lessard S, Handsaker RE, Jain D, Kang HM, Pankratz N, Auer PL, Bao EL, Smith JD, Lange LA, Lange EM, Li Y, Thornton TA, Young BA, Abecasis GR, Laurie CC, Nickerson DA, McCarroll SA, Correa A, Wilson JG, Nhlbi Trans-Omics for Precision Medicine Consortium H, Hemostasis D, Structural Variation TWG, Lettre G, Sankaran VG, Reiner AP (2018) Common alpha-globin variants modify hematologic and other clinical phenotypes in sickle cell trait and disease. PLoS Genet 14: e1007293. doi: 10.1371/journal.pgen.1007293

Rappoport N, Simon AJ, Amariglio N, Rechavi G (2019) The Duffy antigen receptor for chemokines, ACKR1,- 'Jeanne DARC' of benign neutropenia. Br J Haematol 184: 497-507. doi: 10.1111/bjh.15730

Rees JS, Castellano S, Andres AM (2020) The Genomics of Human Local Adaptation. Trends Genet 36: 415-428. doi: 10.1016/j.tig.2020.03.006

Sankar PL, Parker LS (2017) The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genet Med 19: 743-750. doi: 10.1038/gim.2016.183

Schick UM, Jain D, Hodonsky CJ, Morrison JV, Davis JP, Brown L, Sofer T, Conomos MP, Schurmann C, McHugh CP, Nelson SC, Vadlamudi S, Stilp A, Plantinga A, Baier L, Bien SA, Gogarten SM, Laurie CA, Taylor KD, Liu Y, Auer PL, Franceschini N, Szpiro A, Rice K, Kerr KF, Rotter JI, Hanson RL, Papanicolaou G, Rich SS, Loos RJ, Browning BL, Browning SR, Weir BS, Laurie CC, Mohlke KL, North KE, Thornton TA, Reiner AP (2016) Genome-wide Association Study of Platelet Count Identifies Ancestry-Specific Loci in Hispanic/Latino Americans. Am J Hum Genet 98: 229-42. doi: 10.1016/j.ajhg.2015.12.003

Sirugo G, Williams SM, Tishkoff SA (2019) The Missing Diversity in Human Genetic Studies. Cell 177: 1080. doi: 10.1016/j.cell.2019.04.032

Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, Chiang CW, Hirschhorn J, Daly MJ, Patterson N, Neale B, Mathieson I, Reich D, Sunyaev SR (2019) Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. Elife 8. doi: 10.7554/eLife.39702

Sparano JA, Gray RJ, Ravdin PM, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE, Jr., Dees EC, Goetz MP, Olson JA, Jr., Lively T, Badve SS, Saphner TJ, Wagner LI, Whelan TJ, Ellis MJ, Paik S, Wood WC, Keane MM, Gomez Moreno HL, Reddy PS, Goggins TF, Mayer IA, Brufsky AM, Toppmeyer DL, Kaklamani VG, Berenberg JL, Abrams J, Sledge GW, Jr. (2019) Clinical and Genomic Risk to Guide the Use of Adjuvant Therapy for Breast Cancer. N Engl J Med 380: 2395-2405. doi: 10.1056/NEJMoa1904819

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R (2015) 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. doi: 10.1371/journal.pmed.1001779

Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Borglum AD, Breen G, Cichon S, Edenberg HJ, Faraone SV, Gelernter J, Mathews CA, Nievergelt CM, Smoller JW, O'Donovan MC, Psychiatric Genomics C (2018) Psychiatric Genomics: An Update and an Agenda. Am J Psychiatry 175: 15-27. doi: 10.1176/appi.ajp.2017.17030283

Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, Renteria ME, Toro R, Jahanshad N, Schumann G, Franke B, Wright MJ, Martin NG, Agartz I, Alda M, Alhusaini S, Almasy L, Almeida J, Alpert K, Andreasen NC, Andreassen OA, Apostolova LG, Appel K, Armstrong NJ, Aribisala B, Bastin ME, Bauer M, Bearden CE, Bergmann O, Binder EB, Blangero J, Bockholt HJ, Boen E, Bois C, Boomsma DI, Booth T, Bowman IJ, Bralten J, Brouwer RM, Brunner HG, Brohawn DG, Buckner RL, Buitelaar J, Bulayeva K, Bustillo JR, Calhoun VD, Cannon DM, Cantor RM, Carless MA, Caseras X, Cavalleri GL, Chakravarty MM, Chang KD, Ching CR, Christoforou A, Cichon S, Clark VP, Conrod P, Coppola G, Crespo-Facorro B, Curran JE, Czisch M, Deary IJ, de Geus EJ, den Braber A, Delvecchio G, Depondt C, de Haan L, de Zubicaray GI, Dima D, Dimitrova R, Djurovic S, Dong H, Donohoe G, Duggirala R, Dyer TD, Ehrlich S, Ekman CJ, Elvsashagen T, Emsell L, Erk S, Espeseth T, Fagerness J, Fears S, Fedko I, Fernandez G, Fisher SE, Foroud T, Fox PT, Francks C, Frangou S, Frey EM, Frodl T, Frouin V, Garavan H, Giddaluru S, Glahn DC, Godlewska B, Goldstein RZ, Gollub RL, Grabe HJ, et al. (2014) The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav 8: 153-82. doi: 10.1007/s11682-013-9269-5

Timmers PR, Mounier N, Lall K, Fischer K, Ning Z, Feng X, Bretherick AD, Clark DW, e QC, Agbessi M, Ahsan H, Alves I, Andiappan A, Awadalla P, Battle A, Bonder MJ, Boomsma D, Christiansen M, Claringbould A, Deelen P, van Dongen J, Esko T, Fave M, Franke L, Frayling T, Gharib SA, Gibson G, Hemani G, Jansen R, Kalnapenkis A, Kasela S, Kettunen J, Kim Y, Kirsten H, Kovacs P, Krohn K, Kronberg-Guzman J, Kukushkina V, Kutalik Z, Kahonen M, Lee B, Lehtimaki T, Loeffler M, Marigorta U, Metspalu A, van Meurs J, Milani L, Muller-Nurasyid M, Nauck M, Nivard M, Penninx B, Perola M, Pervjakova N, Pierce B, Powell J, Prokisch H, Psaty BM, Raitakari O, Ring S, Ripatti S, Rotzschke O, Rueger S, Saha A, Scholz M, Schramm K, Seppala I, Stumvoll M, Sullivan P, Teumer A, Thiery J, Tong L, Tonjes A, Verlouw J, Visscher PM, Vosa U, Volker U, Yaghootkar H, Yang J, Zeng B, Zhang F, Agbessi M, Ahsan H, Alves I, Andiappan A, Awadalla P, Battle A, Bonder MJ, Boomsma D, Christiansen M, Claringbould A, Deelen P, van Dongen J, Esko T, Fave M, Franke L, Frayling T, Gharib SA, Gibson G, Hemani G, Jansen R, et al. (2019) Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. Elife 8. doi: 10.7554/eLife.39856

Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Genetic Investigation of ATC, Hirschhorn JN (2012) Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat Genet 44: 1015-9. doi: 10.1038/ng.2368

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101: 5-22. doi: 10.1016/j.ajhg.2017.06.005

Weigl K, Chang-Claude J, Knebel P, Hsu L, Hoffmeister M, Brenner H (2018) Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. Clin Epidemiol 10: 143-152. doi: 10.2147/CLEP.S145636

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AA, Westra HJ, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlov J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Bluher M, Bolton JL, Bottcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46: 1173-86. doi: 10.1038/ng.3097

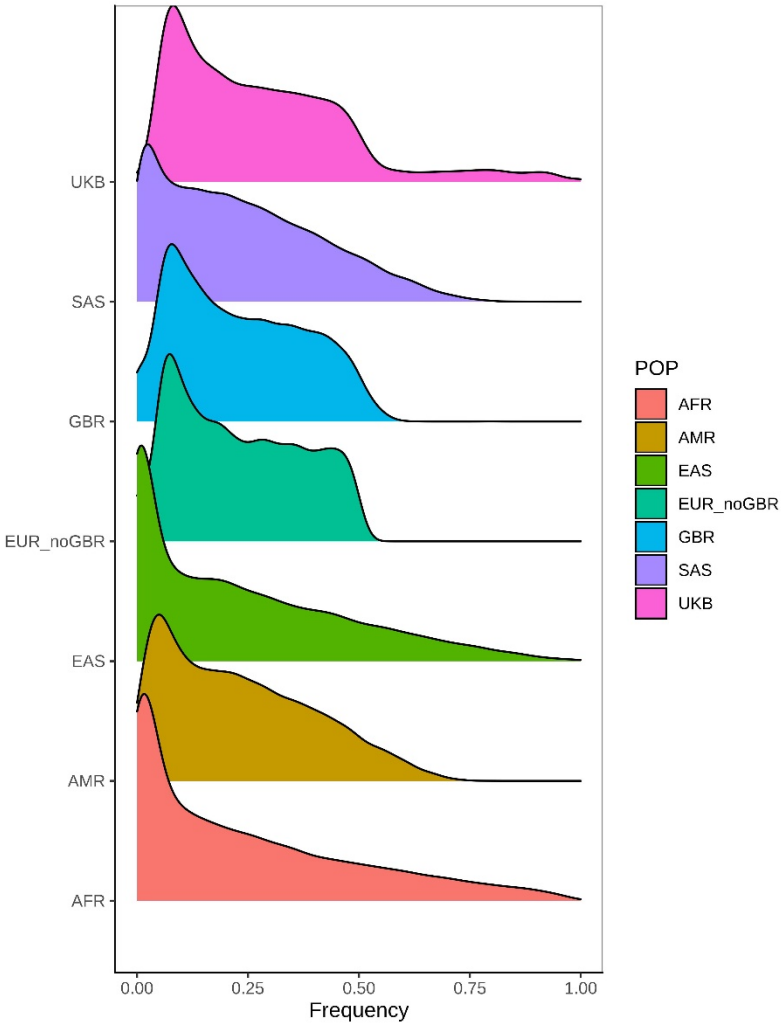

Figure 1: Allele frequency distribution of 15,327 indipendent risk alleles in all populations. AFR=Africa; AMR=Americas; EAS=East Asia, EUR=Europe; EUR_noGBR=Europe without GBR population; GBR= British in England and Scotland; SAS=South Asia; UKB=UK Biobank individuals.

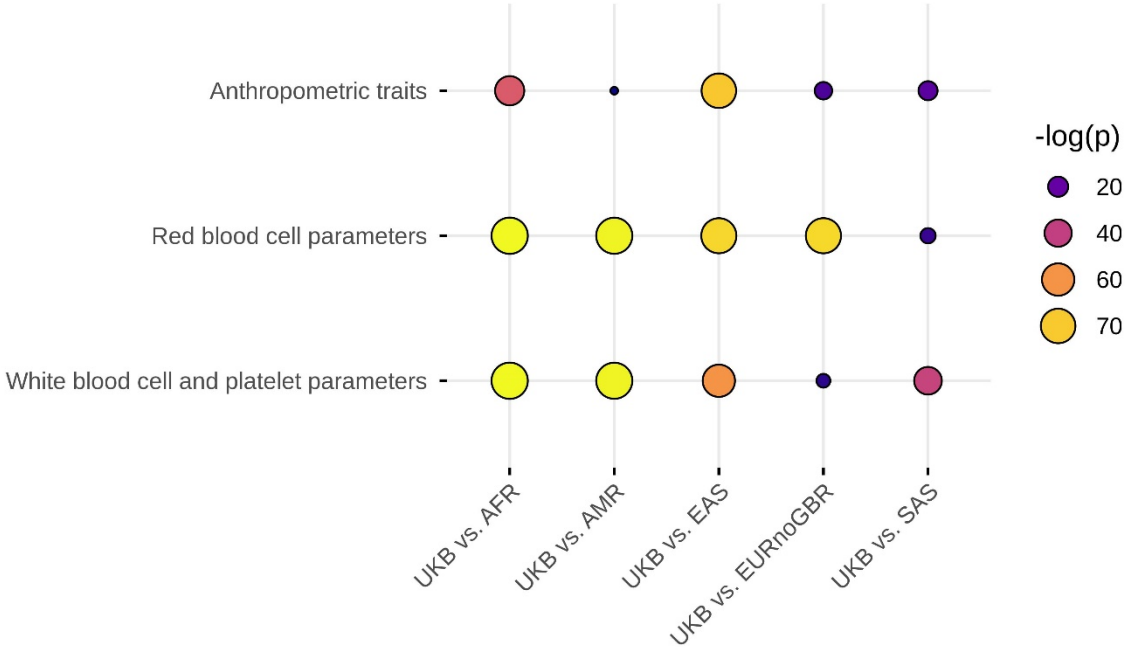

Figure 2: Results of χ^2 test related to significant phenotypic enrichment domains observed in each comparison (UKB vs. AFR; UKB vs. AMR; UKB vs. EAS; UKB vs. EURnoGBR; UKB vs. SAS). AFR=Africa; AMR=Americas; EAS=East Asia, EUR=Europe; EUR_noGBR=Europe without GBR population; GBR= British in England and Scotland; SAS=South Asia; UKB=UK Biobank individuals.

Table 1: Index variants identified in all tested groups. For each comparison, variants id (RS_ID), allele frequencies in both tested populations, phenotypic traits code derived from UK Biobank (code UKB), p value and, traits description and, number of phenotypes associated (derived from SNPs that are il LD with index variants) are reported.

	RS_ID	Allele (A1)	freq (A1) UKB	Freq (A1)	Percentile	code UKB	p_value	Phenotype	N
UKB vs. EURnoGBR	rs34690548	CA	0.904	0.028	1	30280_irnt	2.848E-320	Immature reticulocyte fraction	1
	rs74945666	T	0.909	0.030	0.001	50_irnt	1.48E-17	Standing height	1
	rs200725444	A	0.927	0.047	0.001	30210_irnt	8.60E-14	Eosinophill percentage	2
	rs201088941	TA	0.922	0.029	0.001	30080_irnt	6.72E-10	Platelet count	1
	rs200033476	C	0.917	0.018	0.001	3144_irnt	4.01E-09	Heel Broadband ultrasound attenuation direct entry	1
UKB vs. AFR	rs3749748	T	0.944	0.017	1	23108_irnt	1.7516E-123	Impedance of leg (left)	2
	rs651985	G	0.903	0.044	1	M13_DUPUTRYEN	2.29E-42	Palmar fascial fibromatosis [Dupuytren]	2
	rs604723	T	0.899	0.008	0.001	4080_irnt	6.73E-40	Systolic blood pressure, automated reading	4
	rs157573	A	0.876	0.015	1	50_irnt	8.06E-33	Standing height	5
	rs451367	T	0.950	0.014	1	30120_irnt	1.08E-27	Lymphocyte count	1
	rs603620	A	0.948	0.019	1	30280_irnt	1.55E-27	Immature reticulocyte fraction	5
	rs456798	T	0.059	0.917	1	30190_irnt	3.70E-22	Monocyte percentage	5
	rs171042	T	0.951	0.019	1	30100_irnt	3.97E-18	Mean platelet (thrombocyte) volume	1
	rs1881131	A	0.942	0.031	1	23110_irnt	3.34E-16	Impedance of arm (left)	5
	rs374361	T	0.900	0.012	1	30070_irnt	3.03E-15	Red blood cell (erythrocyte) distribution width	1
	rs556562	A	0.943	0.084	0.005	30080_irnt	4.84E-15	Platelet count	5
	rs200725444	A	0.927	0.007	0.001	30210_irnt	8.60E-14	Eosinophill percentage	1
	rs35497246	C	0.058	0.893	0.001	50_irnt	2.53E-13	Standing height	1
	rs55881864	T	0.912	0.006	0.004	30130_irnt	7.29E-13	Monocyte count	1
	rs3748022	T	0.894	0.018	1	30120_irnt	9.19E-13	Lymphocyte count	3
	rs475591	T	0.998	0.110	1	23102_irnt	1.56E-12	Whole body water mass	8
	rs1812378	T	0.931	0.031	0.002	50_irnt	4.23E-12	Standing height	1
	rs55971426	G	0.887	0.007	1	5158_irnt	1.60E-10	6mm asymmetry index (right)	2
	rs42525	C	0.877	0.021	0.001	50_irnt	1.90E-10	Standing height	1
	rs55959450	C	0.947	0.013	0.003	30070_irnt	1.84E-09	Red blood cell (erythrocyte) distribution width	2

	rs200033476	C	0.917	0.046	1	3144_irnt	4.01E-09	Heel Broadband ultrasound attenuation, direct entry	1
	rs55815739	A	0.929	0.018	1	4080_irnt	2.23E-08	Systolic blood pressure, automated reading	2
	rs625465	G	0.932	0.082	0.001	30190_irnt	2.27E-08	Monocyte percentage	1
	rs57236847	G	0.891	0.037	1	30240_irnt	2.53E-08	Reticulocyte percentage	1
	rs625670	A	0.998	0.002	1	50_irnt	4.29E-08	Standing height	1
UKB vs. AMR	rs34690548	CA	0.904	0.052	1	30280_irnt	2.85E-320	Immature reticulocyte fraction	1
	rs321600	A	0.997	0.122	1	30240_irnt	1.54E-30	Reticulocyte percentage	2
	rs451367	T	0.950	0.091	1	30120_irnt	1.08E-27	Lymphocyte count	1
	rs571497	A	0.943	0.109	0.003	30140_irnt	2.58E-23	Neutrophill count	1
	rs171042	T	0.951	0.127	1	30100_irnt	3.97E-18	Mean platelet (thrombocyte) volume	1
	rs200725444	A	0.927	0.055	0.001	30210_irnt	8.60E-14	Eosinophill percentage	1
	rs1812378	T	0.931	0.089	0.008	50_irnt	4.23E-12	Standing height	1
	rs1875103	T	0.999	0.075	1	30110_irnt	4.43E-12	Platelet distribution width	2
	rs4544340	T	0.950	0.086	0.001	30140_irnt	1.86E-10	Neutrophill count	1
	rs55959450	C	0.947	0.073	0.004	30070_irnt	1.84E-09	Red blood cell (erythrocyte) distribution width	1
	rs200033476	C	0.917	0.011	0.001	3144_irnt	4.01E-09	Heel Broadband ultrasound attenuation, direct entry	1
	rs55815739	A	0.929	0.102	1	4080_irnt	2.23E-08	Systolic blood pressure, automated reading	2
UKB vs. EAS	rs3732378	A	0.941	0.029	0.001	30130_irnt	1.64E-67	Monocyte count	2
	rs6014986	A	0.911	0.017	0.001	30280_irnt	3.05E-43	Immature reticulocyte fraction	6
	rs651985	G	0.903	0	0.001	M13_DUPUTRYEN	2.29E-42	Palmar fascial fibromatosis [Dupuytren]	2
	rs321600	A	0.997	0.133	1	30240_irnt	1.54E-30	Reticulocyte percentage	2
	rs603620	A	0.948	0.003	1	30280_irnt	1.55E-27	Immature reticulocyte fraction	5
	rs571497	A	0.943	0.001	1	30140_irnt	2.58E-23	Neutrophill count	1
	rs1881131	A	0.942	0.002	1	23110_irnt	3.34E-16	Impedance of arm (left)	5
	rs34495	T	0.916	0.043	1	30210_irnt	1.14E-15	Eosinophill percentage	1
	rs200725444	A	0.927	0.024	1	30210_irnt	8.60E-14	Eosinophill percentage	1
	rs55881864	T	0.912	0.001	0.004	30130_irnt	7.29E-13	Monocyte count	1
	rs475591	T	0.998	0.086	1	23102_irnt	1.56E-12	Whole body water mass	8
	rs1875103	T	0.999	0	1	30110_irnt	4.43E-12	Platelet distribution width	2
	rs2861745	G	0.876	0	0.004	50_irnt	8.19E-11	Standing height	4
	rs9932254	C	0.925	0.025	1	30090_irnt	8.55E-11	Platelet crit	1
	rs55971426	G	0.887	0.018	0.002	5158_irnt	1.60E-10	6mm asymmetry index (right)	2
	rs42525	C	0.877	0.024	0.002	50_irnt	1.90E-10	Standing height	1

	rs56207218	C	0.896	0.008	0.001	5085_irnt	1.23E-09	Spherical power (left)	1
	rs55959450	C	0.947	0.003	0.002	30070_irnt	1.84E-09	Red blood cell (erythrocyte) distribution width	1
	rs200033476	C	0.917	0.002	0.001	3144_irnt	4.01E-09	Heel Broadband ultrasound attenuation, direct entry	1
	rs55815739	A	0.929	0.003	1	4080_irnt	2.23E-08	Systolic blood pressure, automated reading	2
	rs625670	A	0.998	0	0.001	50_irnt	4.29E-08	Standing height	1
UKB vs. SAS	rs74945666	T	0.909	0.061	0.001	50_irnt	1.48E-17	Standing height	1
	rs200725444	A	0.927	0.030	1	30210_irnt	8.60E-14	Eosinophill percentage	1
	rs200033476	C	0.917	0.013	1	3144_irnt	4.01E-09	Heel Broadband ultrasound attenuation, direct entry	1
	rs34690548	CA	0.904	0.024	1	30280_irnt	2.84800E-320	Immature reticulocyte fraction	1