# Supplementary Information

## Genetics of 35 blood and urine biomarkers in the UK Biobank

Sinnott-Armstrong*, Tanigawa*, et al

# Table of contents



Corresponding authors: Nasa Sinnott-Armstrong (nasa@stanford.edu), Yosuke Tanigawa (ytanigaw@stanford.edu), and Manuel A. Rivas (mrivas@stanford.edu)

# Supplementary Note

## Population structure analysis

Our definitions of the initial population assignment to self-identified White British, self-identified non-British White, African, East Asian, and South Asian groups are as follows (**Supplementary Figure 1A**):

- Self-identified White British: -20 <= PC1 <= 40 and -25 <= PC2 <= 10 and in_white_British_ancestry_subset == 1 (in sample QC file);
- Self-identified non-British White: -20 <= PC1 <= 40, -25 <= PC2 <= 10, has a self-reported ancestry of White, and does not identify themselves as White British;
- African: 260 <= PC1, 50 <= PC2, and does not identify themselves as any of the following: Asian, White, Mixed, or Other ethnic groups;
- South Asian: 40 <= PC1 <= 120, -170 <= PC2 <= -80, and does not identify themselves as any of the following: Black, White, Mixed, or Other ethnic groups;
- East Asian: 130 <= PC1 <= 170, PC2 <= -230, and does not identify themselves as any of the following: Black, White, Mixed, or Other ethnic groups.

To further refine the population definition, we computed population-specific genotype PCs for non-British white, African, South Asian, and East Asian initial population assignments. Specifically, we used autosomal common (population-specific MAF > 5%) biallelic variants outside of the outside of major histocompatibility complex (MHC) region (chr6:25477797-36448354) that has Hardy-Weinberg p-value greater than $1x10^{-10}$, performed LD pruning on those variants with plink 2.0 (--indep-pairwise 50 5 .5 option), and characterized the genotype PCs with plink 2.0 (--pca 10 var-wts approx vzs option). We used PLINK v2.00a2LM AVX2 Intel (31 Jul 2019) for those analyses. After the manual inspection of those population-specific PCs (**Supplementary Figures 1B-E**), we applied the following additional refinement filters:

- South Asian: -0.02 <= population-specific PC1 <= 0.03, -0.05 <= population-specific PC2 <= 0.02;
- East Asian: -0.01 <= population-specific PC1 <= 0.02, -0.02 <= population-specific PC2 <= 0.

This reduced the number of individuals in South Asian and East Asian populations from 7,962 and 1,772 to 7,885 and 1,154 individuals, respectively.

For the refined populations, we recomputed the population-specific PCs using the same procedure described above (**Supplementary Figures 1F-G**). We used those population-specific PCs for the association analysis of non-British white, African, South Asian, and East Asian populations.

## Variant annotation and quality control

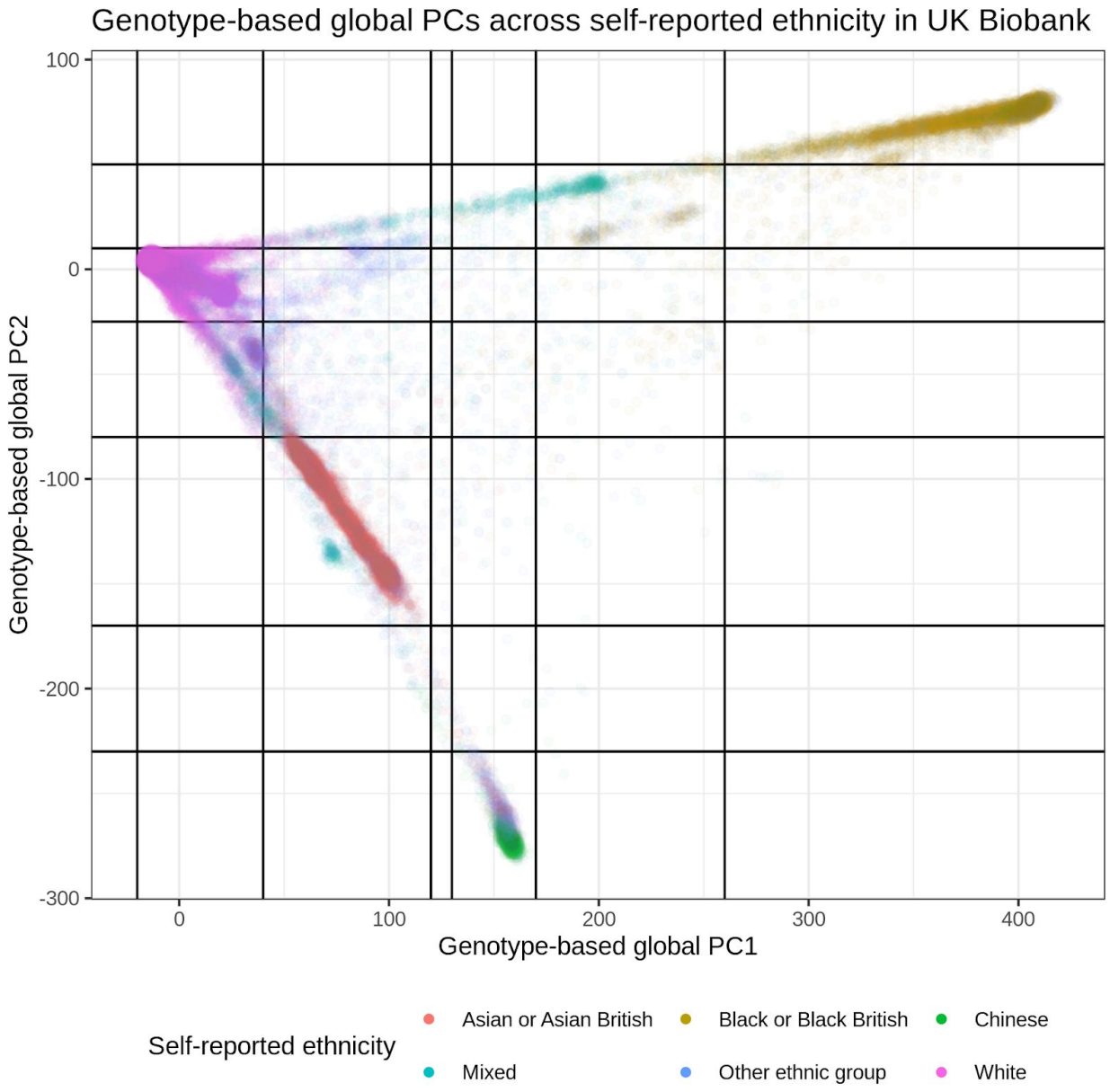
We annotated the directly-genotyped variants using the VEP (April 2017 version) with LOFTEE plugin (https://github.com/konradjk/loftee, version v0.3-beta) and variant quality

control by comparing allele frequencies in the UK Biobank and gnomAD (gnomad.exomes.r2.0.1.sites.vcf.gz) as previously described[76,96]. For directly-genotyped variants, we focused on variants outside of the major histocompatibility complex (MHC) region (hg19 chr6:25477797-36448354) and applied the following filtering criteria[27]:

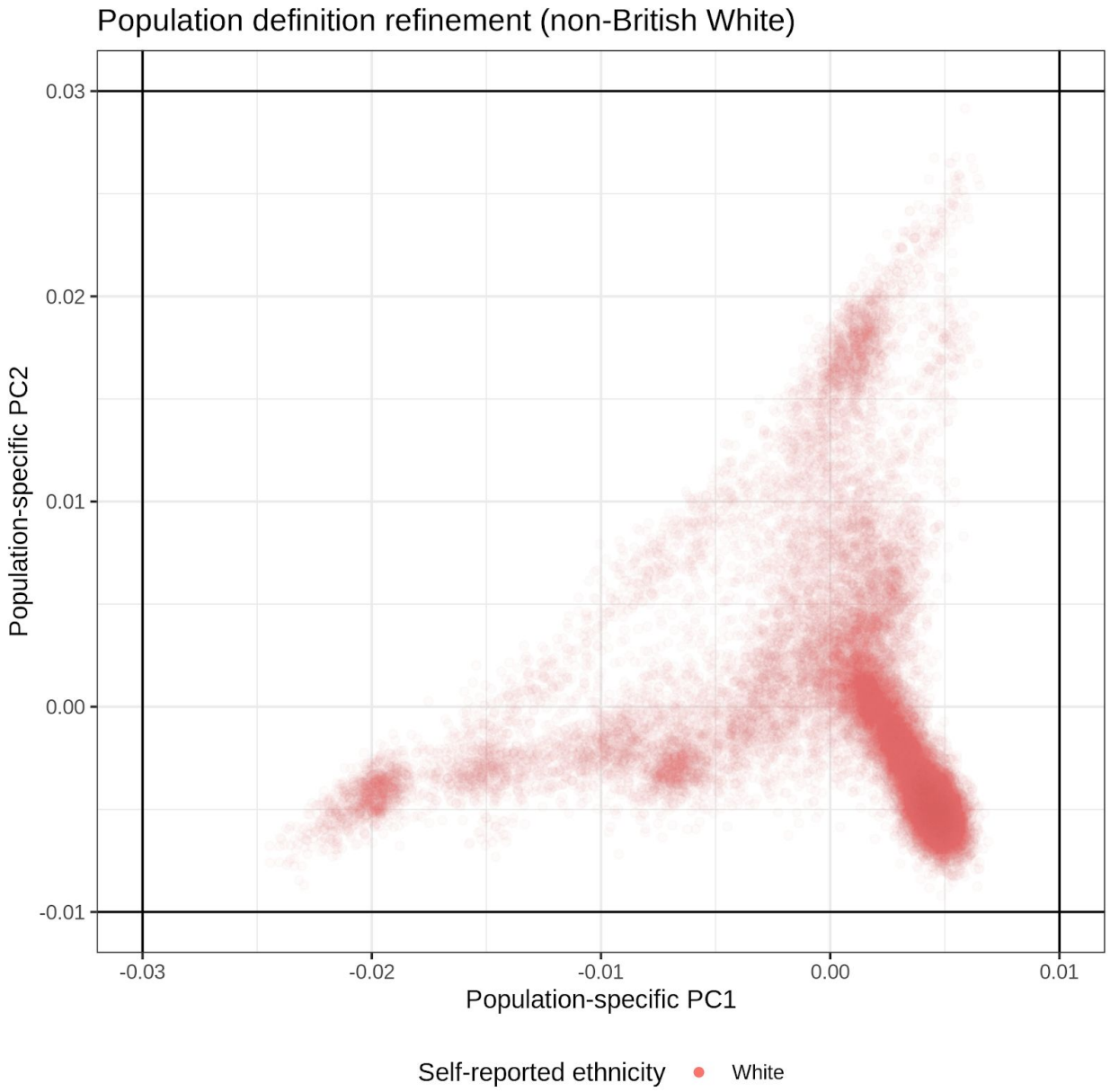
1. The missingness of the variant is less than 1%, considering that two genotyping arrays (the UK BiLEVE array and the UK Biobank array) cover a slightly different set of variants[10].
2. Minor-allele frequency is greater than 0.01%, given the recent reports casting questions on the reliability of ultra low-frequency variants[97,98].
3. Hardy-Weinberg disequilibrium test p-value is less than $1.0 \times 10^{-7}$
4. Manual cluster plot inspection. We investigated the cluster plots for a subset of variants and removed 11 variants that have unreliable genotype calls[76].
5. Passed the comparison of minor allele frequency with the gnomAD dataset as described before[76].

For the imputed variants, we focused on the HRC-imputed SNPs in version 3 of the UK Biobank imputed genotype data release (~97 million) and subsequently applied a filter on minor allele frequency and on the INFO score. Within each population, SNPs with an assigned variant ID from UK Biobank (~81 million) were filtered to only those with an INFO score greater than 0.3 and MAF greater than 0.1% before meta-analysis (~16 million to ~29 million variants, varying by trait and by population). After meta-analysis, variants with a MAF greater than 1% in the White British (~9 million) were kept for downstream analysis.

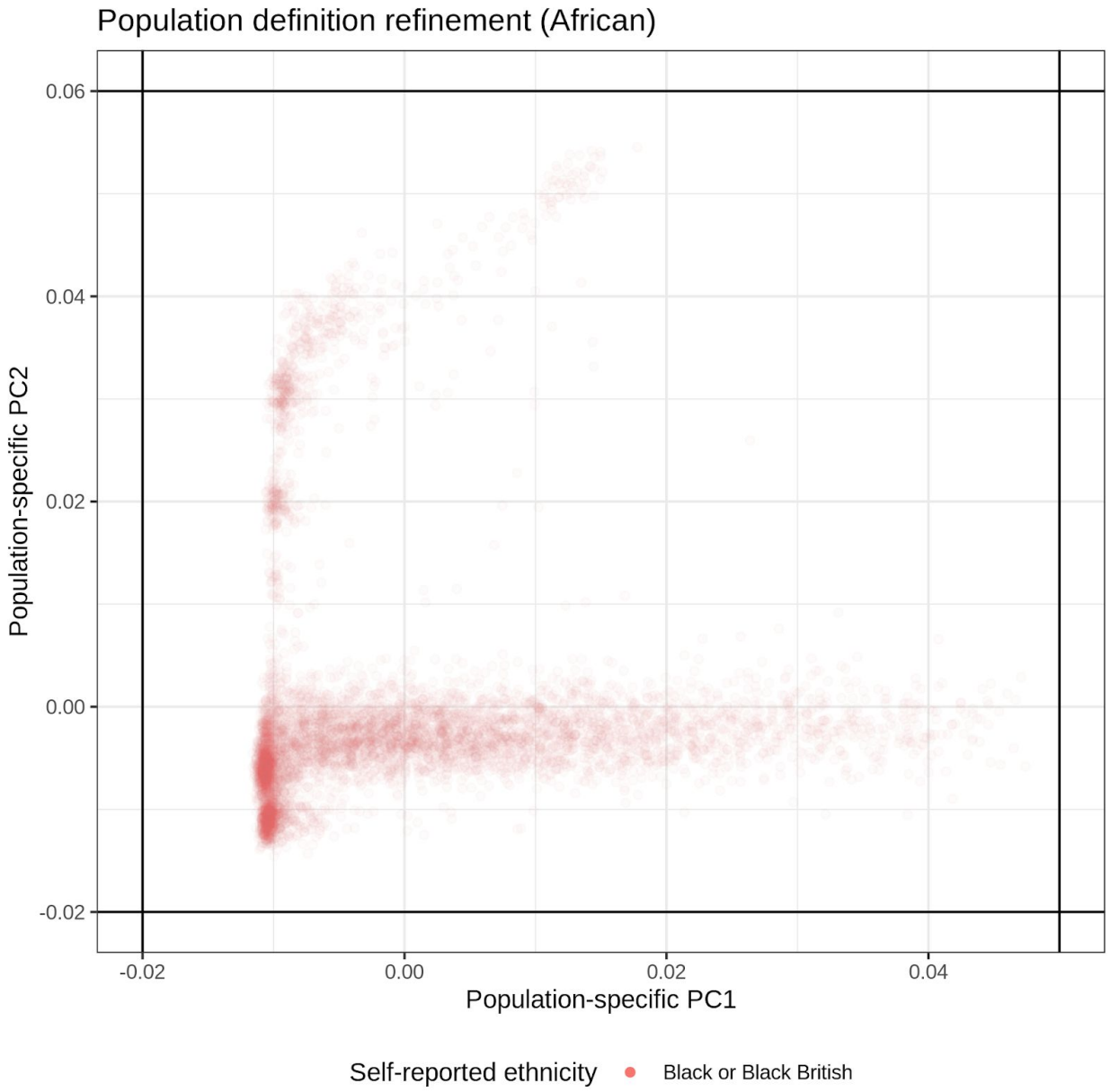
(A)


Genotype-based global PCs across self-reported ethnicity in UK Biobank
Genotype-based global PC2
100
0
-100
-200
-300
0
100
200
300
400
Genotype-based global PC1
Self-reported ethnicity
Asian or Asian British
Black or Black British
Chinese
Mixed
Other ethnic group
White

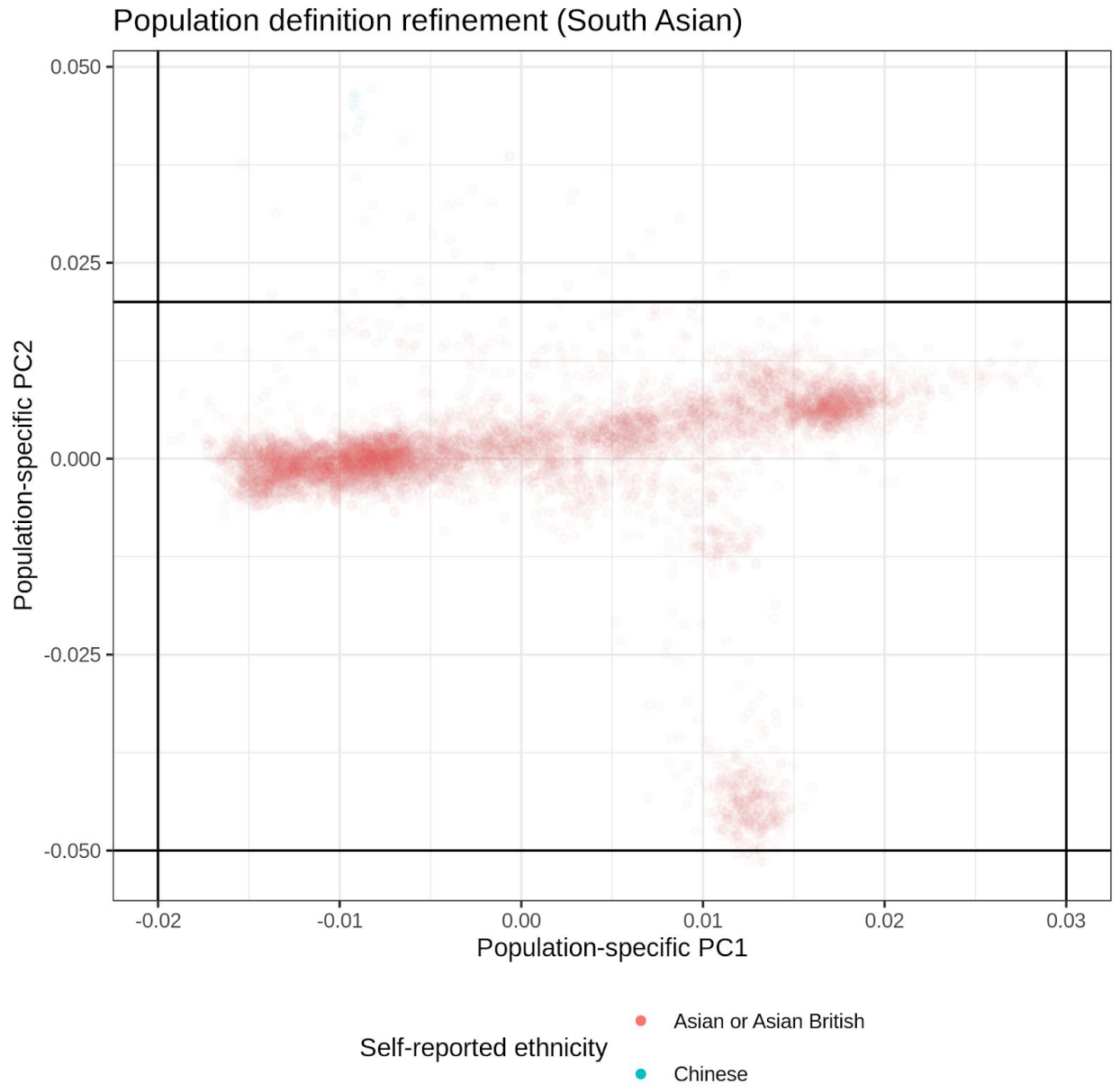
(B)


Population definition refinement (non-British White)
0.03
0.02
0.01
0.00
-0.01
Population-specific PC2
-0.03
-0.02
-0.01
0.00
0.01
Population-specific PC1
Self-reported ethnicity
White

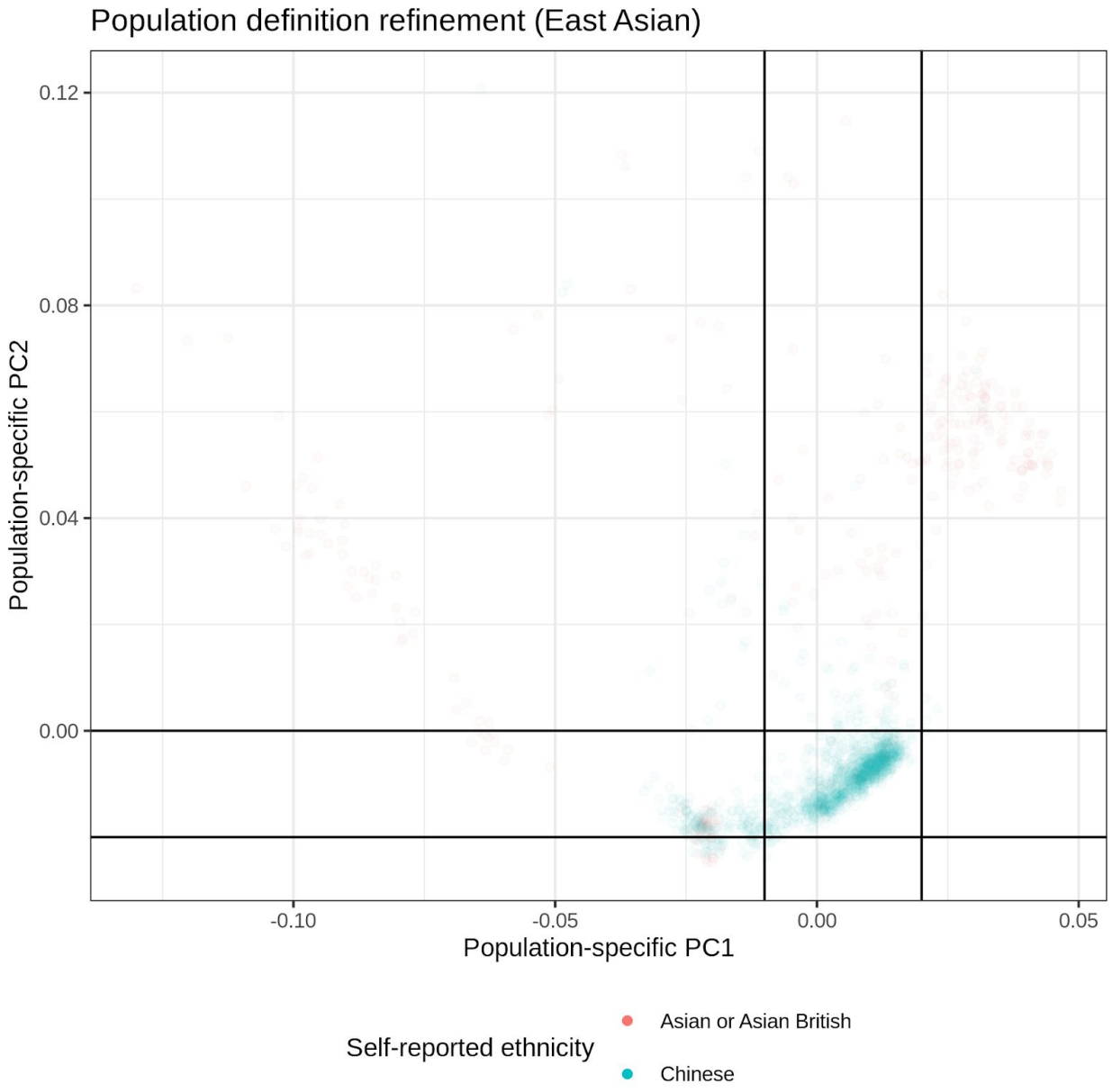
(C)


Population definition refinement (African)
0.06
0.04
0.02
0.00
-0.02
Population-specific PC2
-0.02
0.00
0.02
0.04
Population-specific PC1
Self-reported ethnicity
Black or Black British

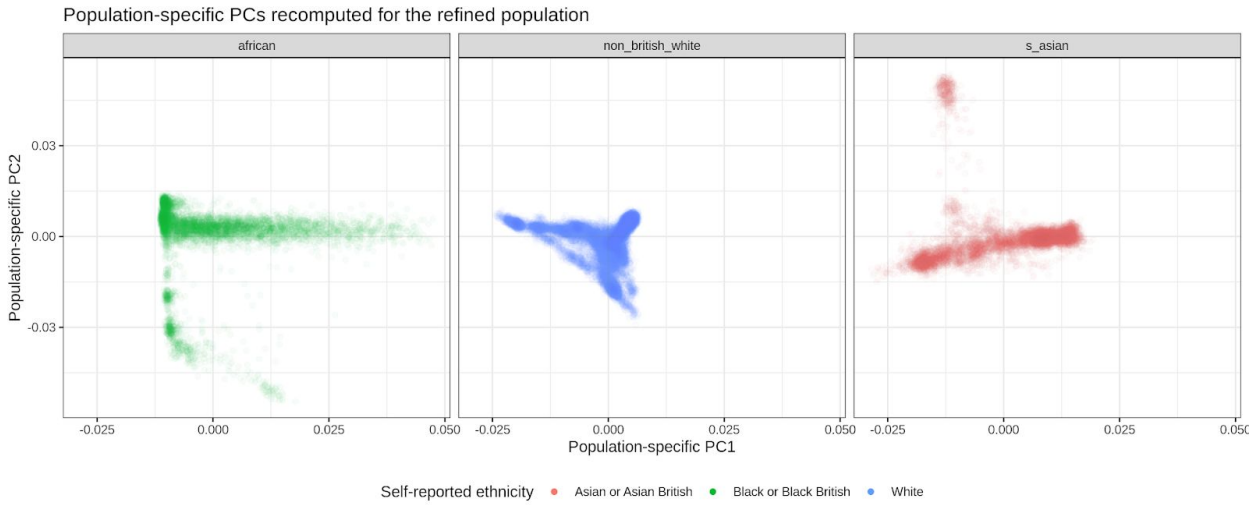
(D)


Population definition refinement (South Asian)
0.050
0.025
0.000
-0.025
-0.050
Population-specific PC2
-0.02
-0.01
0.00
0.01
0.02
0.03
Population-specific PC1
Self-reported ethnicity
Asian or Asian British
Chinese

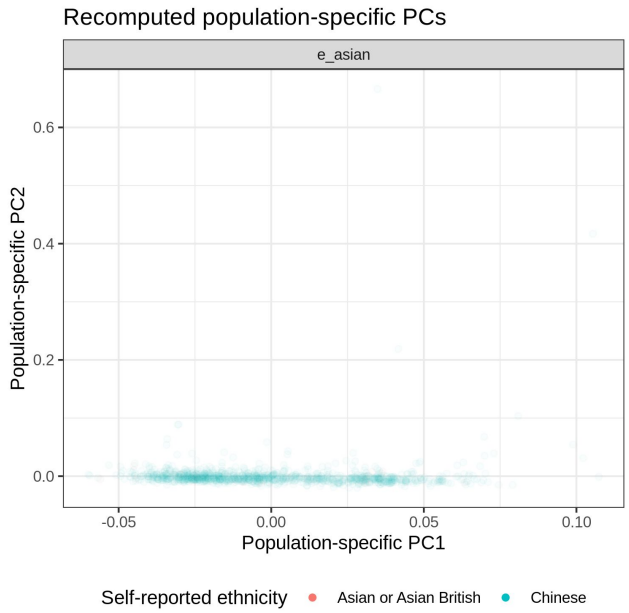
(E)


Population definition refinement (East Asian)
0.12
0.08
0.04
0.00
Population-specific PC2
-0.10
-0.05
0.00
0.05
Population-specific PC1
Self-reported ethnicity
Asian or Asian British
Chinese

(F)




(G)




**Supplementary Figure 1. Population structure analysis for UK Biobank with a combination of principal component analysis and self-reported ethnicities.** (A) The 487,409 subjects in the UK Biobank with the genotype data are projected to the genotype-based principal components and shown on the global PC1 vs. PC2. Color represents the self-reported ethnicity. The thresholds used in the population definitions are shown as black vertical or horizontal lines (Methods). (B-E) Population refinement based on population-specific PCs (Methods) for (B) non-British white, (C) African, (D) South Asian, and (E) East Asian individuals. The color represents the self-reported ethnicity. The threshold used to refine the population definitions are shown as black vertical or horizontal lines. (F, G) The population-specific PCs were recomputed for the refined definitions of (F) non-British white, African, and South Asian, and (G) East Asian individuals (Methods). The first two population-specific PCs are shown in the plot. The color represents the self-reported ethnicity.

## Phenotype distributions

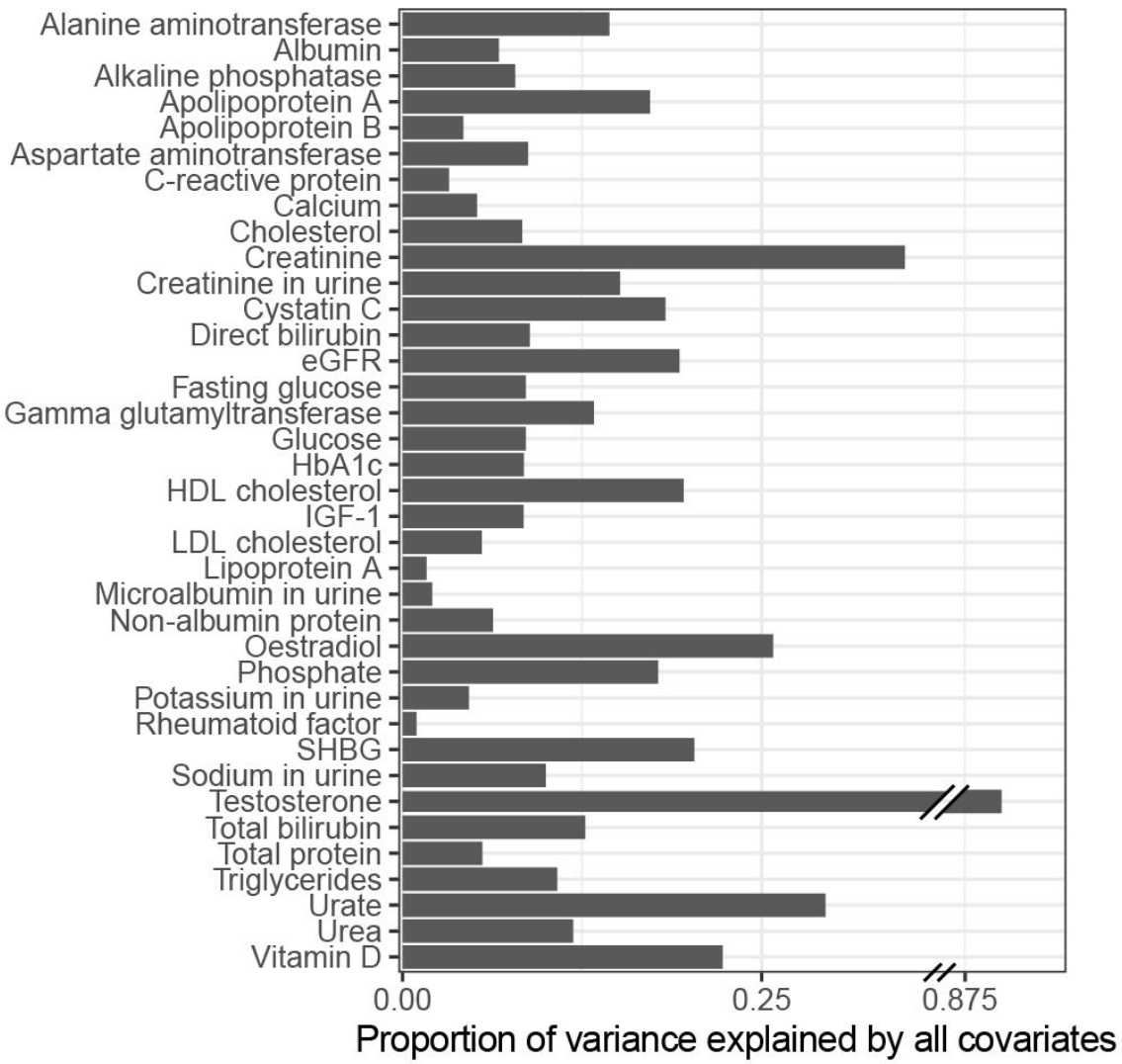
Within the UK Biobank, we empirically estimated the adjustments for statin treatment effect on the biomarkers. In UK Biobank, about 20,000 individuals returned for a repeated assessment (imaging assessments did not include biomarker level measurement). Of those, 1,705 either started or stopped a statin between enrollment and that second visit. We utilized the N=1,427 people who were on statins at the second visit but had not been on them at enrollment. Our empirical estimates of the effect size (in log relative scale) and their corresponding p-values are summarized in Supplementary Tables 1B-1C. Given the significance of the associations, we applied the statin adjustment only for LDL cholesterol, total cholesterol, and apolipoprotein B. The statin adjustment factors were listed in Supplementary Table 1A.

We note that there are pre-existing estimates in the literature for LDL [3] and that LDL is typically adjusted by 0.7 and total cholesterol by 0.8, similar to our empirical estimates.

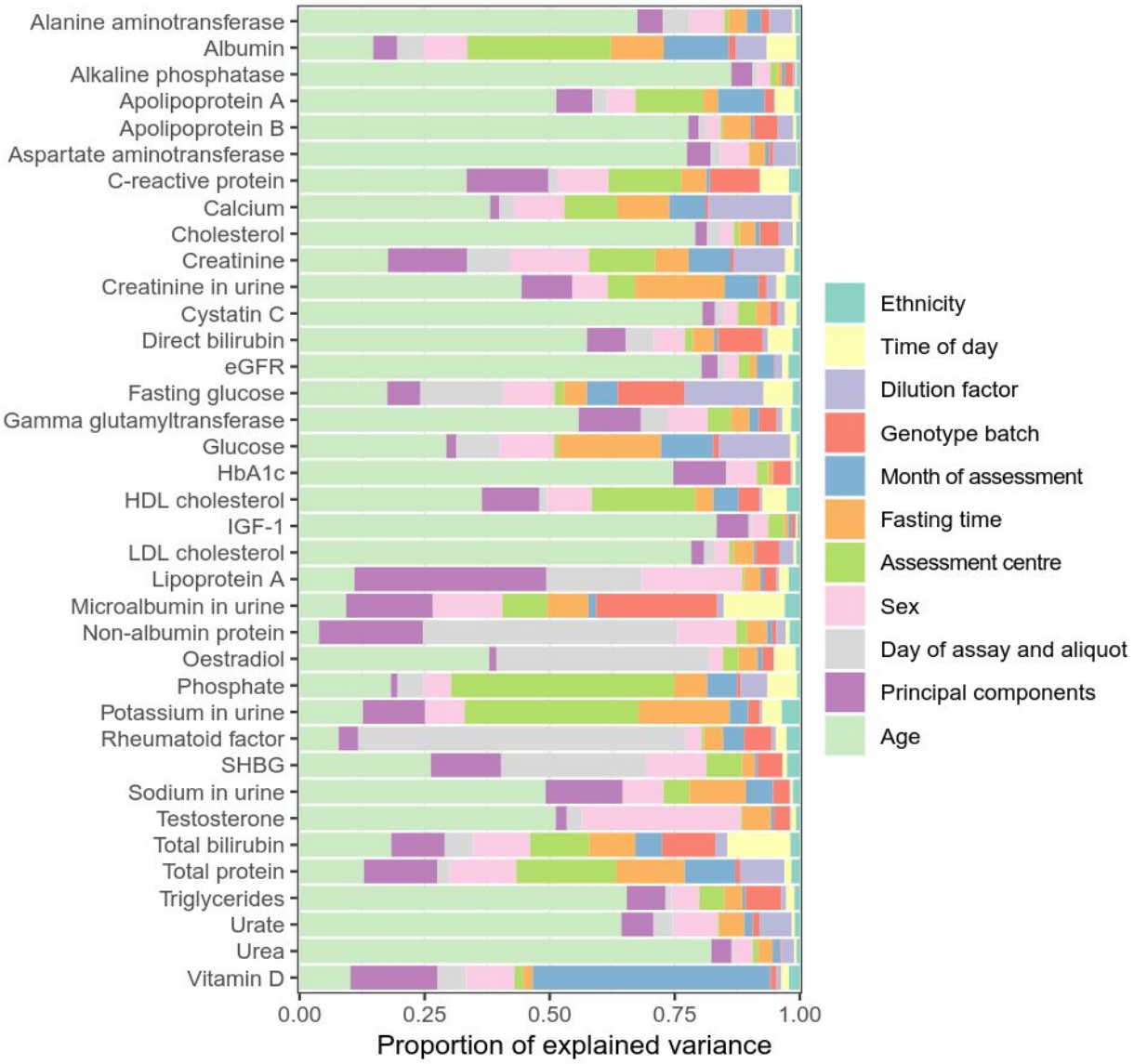
**Supplementary Table 1A.** Estimated adjustment factor based on statin usage.

**Supplementary Table 1B.** Estimated adjustment based on statin usage (BMI adjusted).

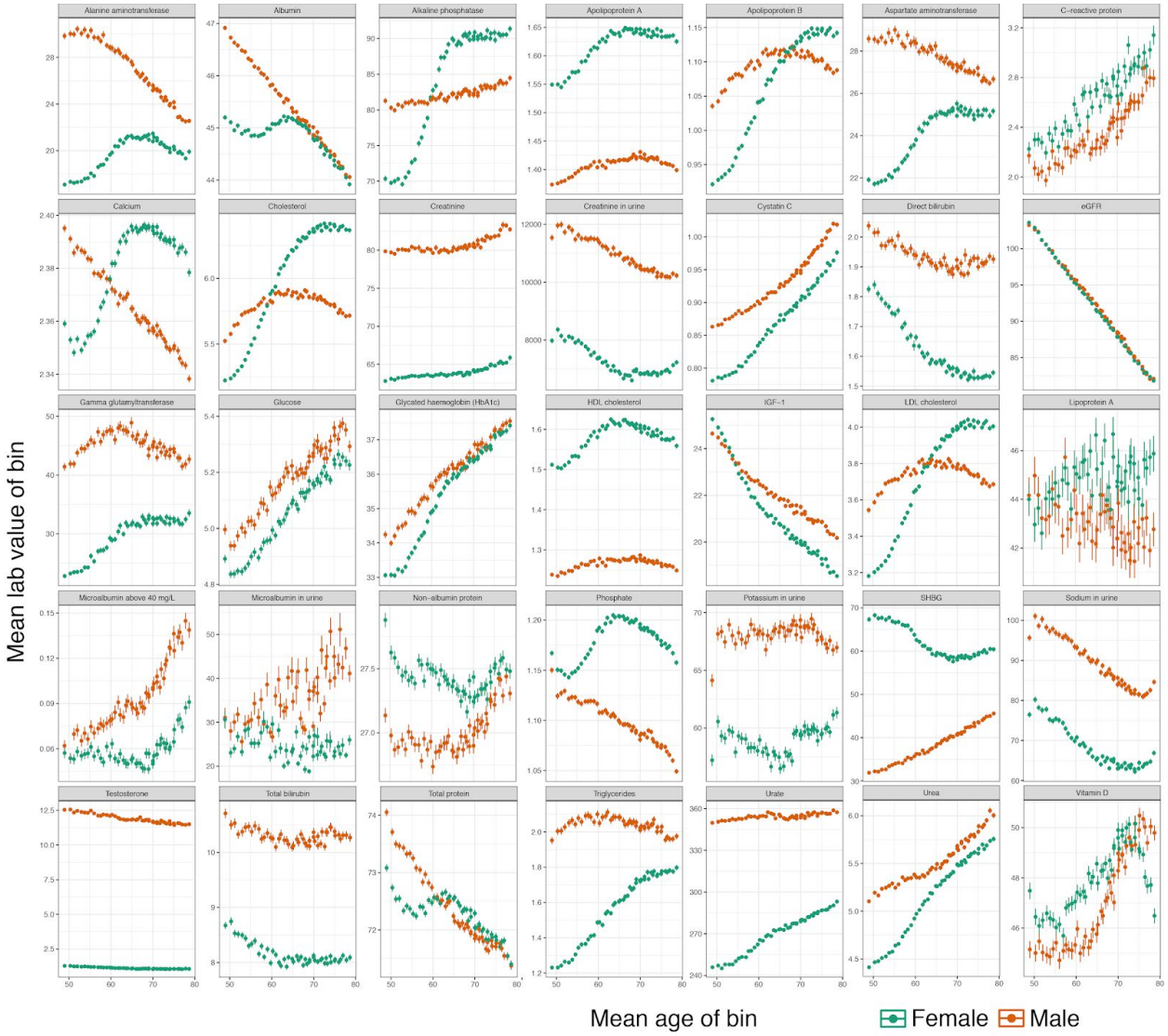
**Supplementary Table 1C.** Estimated adjustment based on statin usage (not BMI adjusted).




**Supplementary Figure 2A. Proportion of variance explained by all covariates across the 37 raw laboratory phenotypes.** (x-axis) Regression estimate of the proportion of variance explained by all covariates in a linear model for 37 raw laboratory phenotypes including Fasting glucose defined if fasting time between 8 and 24 hours according to Data Field 74 in UK Biobank Data Showcase (y-axis). Blue bar plots indicate estimate before medication adjustment and red bar plots indicate estimate after medication adjustment




**Supplementary Figure 2B. Breakdown of variance explained per covariate in the log-transformed phenotypes.** (x-axis) Fraction of explained variance (y-axis) Biomarker being evaluated. Interaction terms were split equally among constituent terms.




**Supplementary Figure 2C. Phenotype distributions of all raw laboratory tests by age and sex.** (x-axis) Age of individuals within a pentacontile were averaged. (y-axis) The corresponding average value +/- 1 SD of each laboratory test measurement for all individuals with available data in the study. Color indicates the reported sex of the individuals (orange = male, turquoise = female).

**Supplementary Table 2.** Phenotype distributions for each of the raw and log-transformed phenotypes.

**Supplementary Table 3.** Genetic correlations between BMI-adjusted and BMI-unadjusted measurements.

**Supplementary Table 4A.** Description of 35 measured and derived lab phenotypes used in the genetic analysis.

**Supplementary Table 4B.** Effects of different sets of covariates in covariate adjustment with residualized models.

## Statin identification and LDL adjustment

We reviewed the medications taken by one or more participants in the UK Biobank (UK Biobank Field ID: 20003) and identified 13 medication codes corresponding to statins (UK Biobank Data-coding ID 4: 1141146234, atorvastatin; 1141192414, crestor 10mg tablet; 1140910632, eptastatin; 1140888594, fluvastatin; 1140864592, lescol 20mg capsule; 1141146138, lipitor 10mg tablet; 1140861970, lipostat 10mg tablet; 1140888648, pravastatin; 1141192410, rosuvastatin; 1141188146, simvador 10mg tablet; 1140861958, simvastatin; 1140881748, zocor 10mg tablet; 1141200040, zocor heart-pro 10mg tablet). We then identified participants (n = 1,427) with biomarker measurements who were not taking a statin upon enrollment (years 2006-2010), but who were taking a statin at the time of the first repeat assessment visit (years 2012-2013). For each participant, we divided their on-statin biomarker measurement by their pre-statin biomarker measurement. The mean of this value was considered to be the statin correction factor within the UK Biobank. For all individuals who were taking statins upon enrollment, we divided their on-statin measurement by the correction factor to yield an adjusted biomarker measurement value. For all traits, we calculated a p-value from a regression testing whether the log ratio of pre- and on-statin values were significantly different from 0 after adjusting for sex, Townsend Deprivation Index at baseline, the top 20 global PCs of the genotype matrix, age at baseline, BMI at baseline, age difference between baseline and followup, BMI difference between baseline and followup, and baseline age, age difference, and BMI difference by sex interactions. For the model excluding BMI, all BMI terms were excluded.

Only traits with a significant non-zero effect were evaluated with adjustment for statins. The following list of 7 statins were identified in the UK Biobank for the purposes of adjusting by the estimated factor: simvastatin, fluvastatin, pravastatin, atorvastatin, rosuvastatin, lipid-lowering drug, lipitor 10mg tablet.

## Covariate correction

Log-transformed UK Biobank measurements for all reported individuals (excluding out of range and QC failed measurements) were fit with linear regression against covariates. Trait measurements are first log-transformed, then adjusted for genotype principal components (the top 40 principal components of the UK Biobank-provided genotype-based global PCs), age indicator variables (one for each integer age), sex, 5-year age indicators by sex interactions, self-identified ethnicity, self-identified ethnicity by sex interactions, fasting time (one indicator per fasting time, except a single indicator for >18h and for 0 or 1 hours), estimated sample dilution factor (icosatiles), assessment center indicators, genotyping batch indicators, icosatiles of time of sampling during the day, month of assessment (indicators for each month of participation, with the exception that all of 2006 and August through October of 2010 were assigned a single indicator), and day of assay (one indicator per day assay was performed). The residuals were used for downstream analysis unless specified otherwise.

## Genetics of biomarkers

## Heritability estimates

### LD Score regression

We used LD score regression version 1.0.1 for the analysis. We used the default LD Scores from the 489 unrelated European individuals in 1000 Genomes as our reference[24,99]. We converted our White British summary statistics to LDSC format using munge_sumstats, munging against the set of 1000 Genomes Phase 1 variants with calls of an ancestral allele in 1000 Genomes Phase 3. We ran ldsc.py with the following parameters:

```
ldsc.py --h2 <trait summary statistics> --ref-ld-chr
<ldsc/1000G.EUR.QC/> --w-ld-chr
<ldsc/weights_hm3_no_hla/weights.>
```

### HESS

We performed standard stage 1 fitting[25], then removed all regions which contained no SNPs with MAF > 5% (5/~1700 bins genome-wide) and generated stage 2 estimates from the resulting matrices. We used the same munged White British summary statistics described above, which were generated using a modified version of the munge_sumstats.py which also outputs chromosome and position. We confirmed heritability estimates of select associations using GCTA-GREML[100] and genotyped array variants on a subset of individuals (data not shown) to ensure estimates were comparable to this model.

## GWAS of coding variants in genotyping array

Univariate association analyses for single variants were applied to the 35 phenotypes independently using PLINK v2.00aLM (2 April 2019). Specifically, we applied generalized linear model association analysis for each of the quantitative phenotypes after adjusting the covariates.

## GWAS of imputed variants

We employed a GWAS with covariates of the population-specific PCs and the genotyping array on the residuals computed above. Variants were the full set of HRC-imputed SNPs in the version 3 UK Biobank data release. This was run using plink v2.00aLM with the following parameters:

```
--glm
cols=chrom,pos,ref,alt,altfreq,firth,test,nobs,orbeta,se,ci,t,p
hide-covar --pgen <UKBB imputed PGEN> --remove <out of population
or related individuals> --geno 0.1 --hwe 1e-50 midp;
```

## Association and Bayesian model averaging analyses for HLA alleles

The HLA data from the UK Biobank contains all HLA loci (one line per person) in a specific order (A, B, C, DRB5, DRB4, DRB3, DRB1, DQB1, DQA1, DPB1, DPA1). We downloaded these values, which were imputed via the HLA:IMP*2 program (Resource 182); the UK Biobank reports one value per imputed allele, and only the best-guess alleles are reported. Out of the 362 alleles reported in UKB, we used 175 alleles that were present in >0.1% of the population surveyed.

We performed association analysis for our 35 phenotypes and 175 HLA alleles using PLINK v2.00aLM (2 April 2019). We included only self-identified White British individuals (n = 337,151) and used generalized linear models with BY corrected p-value threshold of 0.05[87].

To further eliminate potentially spurious associations due to the pervasive LD in the HLA alleles and identify causal alleles for each phenotype, we applied Bayesian Model Averaging (BMA) for the phenotypes with at least two significant (BY-corrected p-value < 0.05) allelotype associations. BMA is a model selection procedure to identify the causal configuration of allelotypes based on Bayesian information criterion (BIC) of each model given a set of associated alleles, the allele dosage information, and the covariate-adjusted phenotypes. To make the computation tractable, we selected at most 10 significant allelotype associations for each phenotype, given that the search space of causal configuration is exponential to the number of associated alleles, and focused on models whose posterior model probability was within a factor of 1/5 of that of the best model. With those filters, we applied the BMA procedure implemented in the 'bma' R package [101] version 3.18.12 across 33 phenotypes, with 56 alleles included in at least one analysis, where we used Gaussian link function and a uniform prior of 0.5 as the prior weight for each allelotype. As a measure of confidence in the association between allelotype and phenotype, we computed the allelotype inclusion probability across models as the posterior probabilities and reported the associations with posterior inclusion probabilities > 0.8.

We reported allele, phenotype, posterior mean effect size, the standard deviation of said effect size, and the posterior probability that the effect is not equal to 0.

## GWAS of copy number variants

CNVs were called by applying PennCNV v1.0.4 on raw signal intensity data from each array within each genotyping batch as previously described[34]. In total, we conduct association tests for 8,274 non-rare (MAF > 0.01%) CNVs and 23,598 genes. Genotypes for gene-level burden tests were treated as an indicator variable for the presence of any CNV which overlaps within 10*kb* of the gene region.

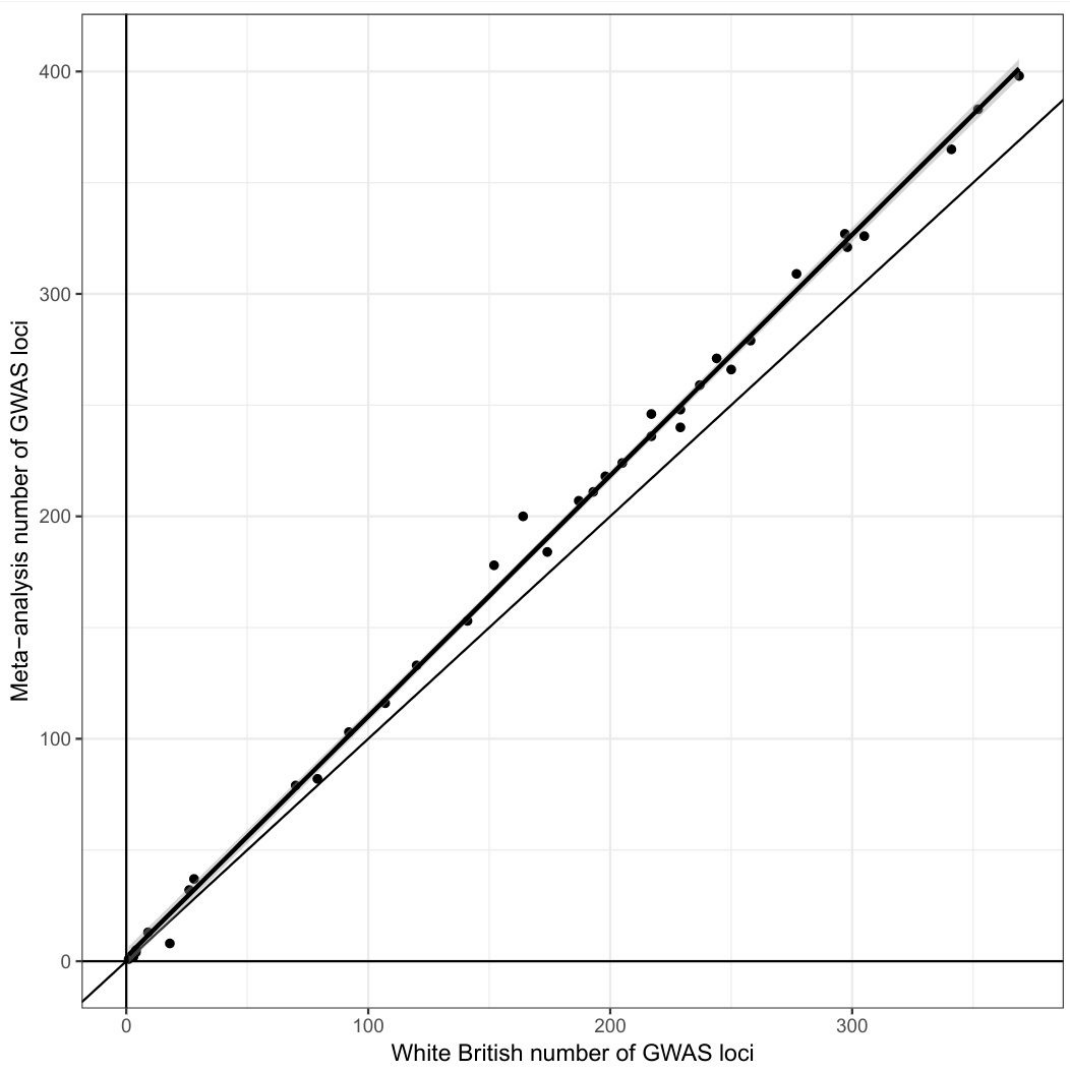
We computed generalized linear models using PLINK v2.00aLM (31 Mar 2018) --glm with age and sex as covariates, and with phenotype quantile normalization (--pheno-quantile-normalize option). For burden tests, we added the number and the total length of CNV as covariates. See the "GWAS on genetic variants on genotyping array" section for further description of PLINK's

implementation of these model specifications.

These summary statistics were additionally meta-analyzed with METAL. After the meta-analysis was performed (see “Meta-analysis” subsection below), the association statistics were clumped and filtered using the following arguments to PLINK 1.90b6[90]:

```
--clump <CNV summary statistics> --clump-r2 0.01
--clump-field P-value  --clump-snp-field MarkerName
--clump-kb 10000 --clump-p1 1e-6 --clump-p2 1e-6
--keep <unrelated white british>
```

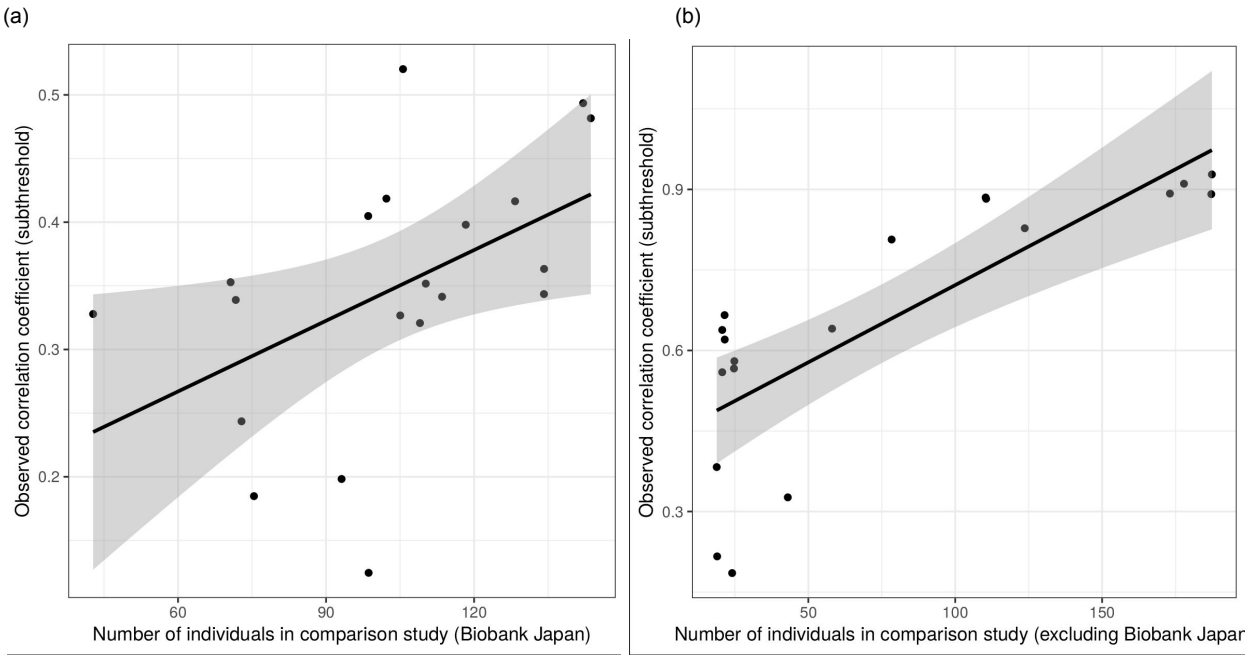
Meta-analysis across 4 ancestry groups in UK Biobank




**Supplementary Figure 3. Effect of meta-analysis of White British individuals with other UK Biobank participants on the discovery of genetic associations.** (x-axis) Number of lead variants (before 1cM clumping) in White British (y-axis) Number of lead variants (before 1cM clumping) in meta-analysis. Each point represents one biomarker trait. Meta-analysis substantially increases power for variant discovery.

## Comparison of effect sizes with published studies

**Supplementary Table 5.** Comparison of estimated effect sizes between UK Biobank and other studies.




**Supplementary Figure 4. Correlation of effect sizes between UK Biobank and previous GWAS is predicted primarily by sample size.** (x-axis) Sample size in thousands of comparative studies. (y-axis) Observed Spearman's correlation coefficient. All variants associated $p < 1 \times 10^{-6}$ (subthreshold; two-sided t-test) in either study are included. (a) Comparisons from the Biobank Japan study [17] (b) Comparisons from other studies of mostly European ancestry (see Supplementary Table 5 for full list). Grey bands represent 95% confidence intervals on the regression fit, and lines represent the estimated mean regression fit on the basis of the observed data.

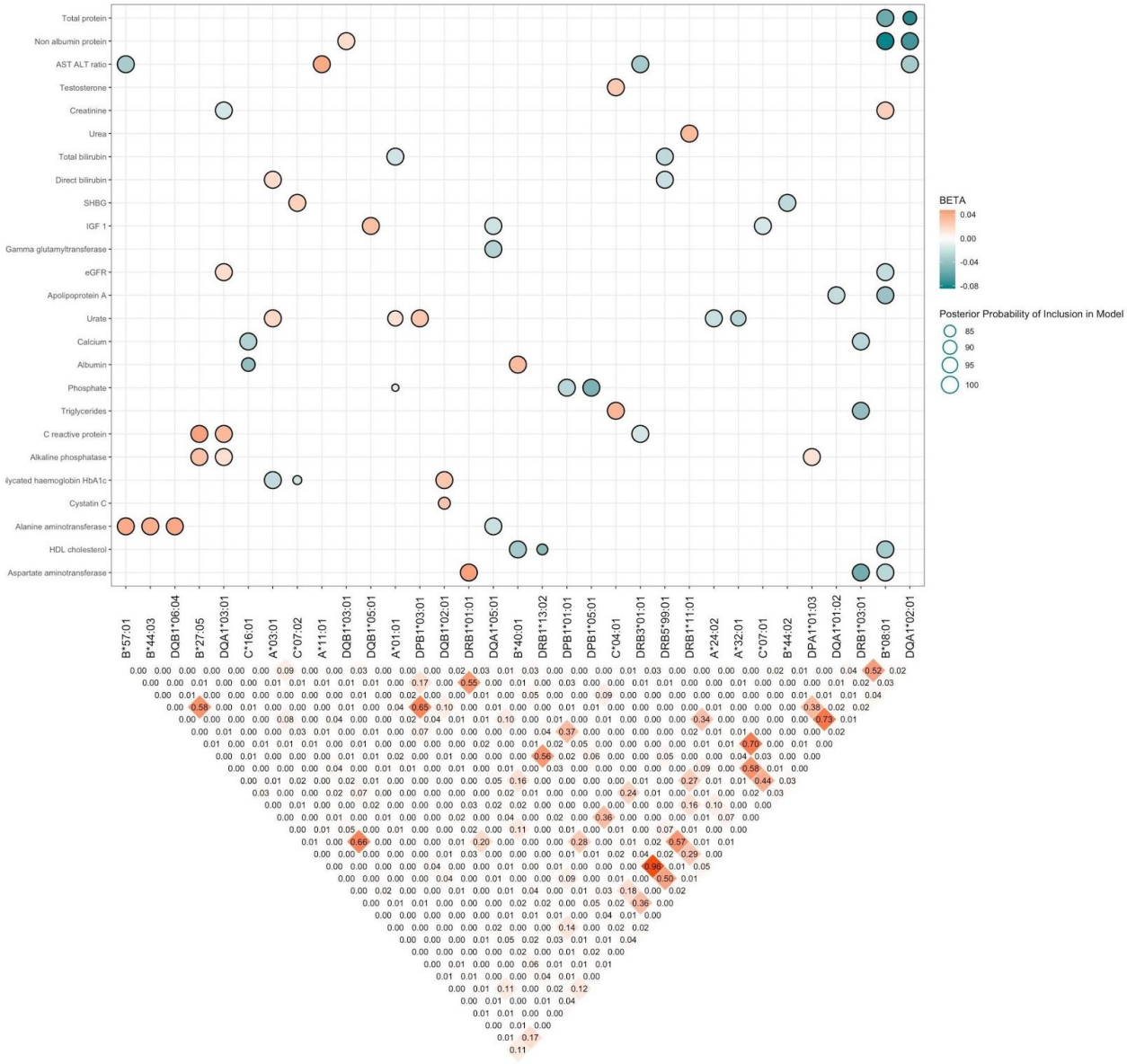
## Biomarker associated variants prioritize therapeutic targets

**Supplementary Table 6.** Association results for protein-truncating variants across the 35 lab phenotypes (within UK Biobank meta-analysis $p < 5 \times 10^{-9}$).

**Supplementary Table 7.** Association results for protein-altering variants across the 35 lab phenotypes (within UK Biobank meta-analysis $p < 5 \times 10^{-9}$).

**Supplementary Table 8.** Association results for non-coding variants across the 35 lab phenotypes (within UK Biobank meta-analysis $p < 5 \times 10^{-9}$).

**Supplementary Table 9.** HLA alleles found to be associated with the 35 lab phenotypes via both PLINK association tests and Bayesian Model Averaging (BMA).

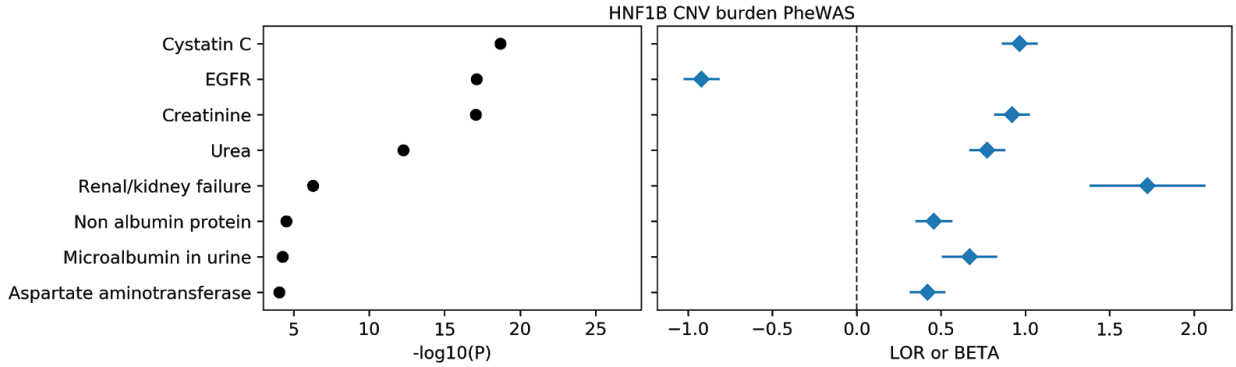



**Supplementary Figure 5. Posterior effect sizes, probabilities of Bayesian Model Averaging model inclusion, and linkage disequilibrium for HLA alleles on 35 different laboratory test phenotypes.** *y*-axis indicates phenotype and *x*-axis indicates allele. Phenotypes along the *y*-axis are hierarchically clustered with respect to effect size estimates of adjacent alleles. Above - the size of each dot corresponds to the posterior probability that the HLA allele is included as a variable across all plausible models as deemed by BIC measures from BMA, and the color of each dot corresponds to the size and direction of the effect of the allele on the phenotype as found by PLINK. Only the top 10 significant PLINK hits per phenotype were considered for the analysis, and the figure only enumerates those associations that were found to have both a PLINK association p-value <= 0.05/10000 (two-sided linear regression) and a BMA posterior probability >= 0.8. Below - LD measures ($R^2$ values, as determined and visualized by the *gaston* package) across HLA alleles. The HLA association summary statistics is available in **Supplementary Table 9**.

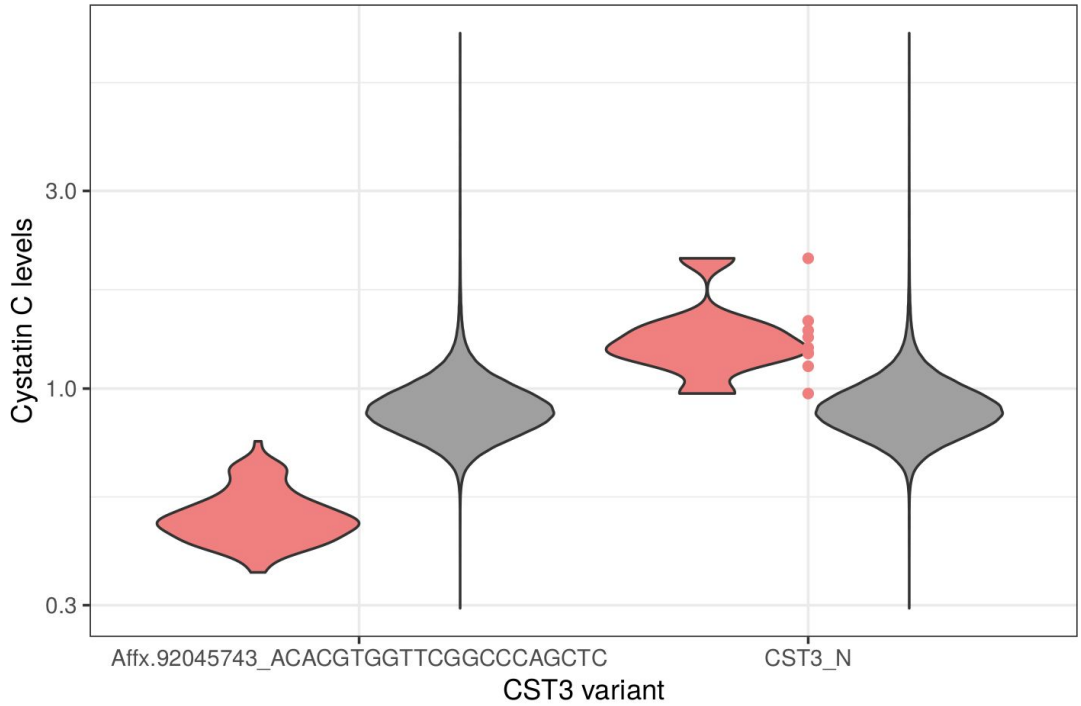
## CNVs influencing lab phenotypes

**Supplementary Table 10A.** Single and burden copy number variation associated with the 35 lab phenotypes.

**Supplementary Table 10B.** *HNF1B* CNV associations with kidney failure, conditional on diabetes status.




**Supplementary Figure 6A. PheWAS of rare CNVs affecting *HNF1B*.** X-axis log-odds ratio (LOR, for disease outcomes) or BETA (for biomarkers) and standard errors and -log10(P) for each trait having association with *HNF1B* CNVs at $p < 1x10^{-4}$ (linear regression for continuous traits and logistic/Firth regression for dichotomous traits, two-sided in all cases). Associations for all traits run as in previous analysis [102], except for biomarker traits described here.




**Supplementary Figure 6B. Comparison of PTV and duplication at the *CST3* locus on Cystatin C levels.** Violin plots across carrier individuals (red) are compared with other participants (gray), for both a rare PTV at *CST3* (left, Affx.92045743, n=105) and a burden of duplication of the *CST3* gene (right, *CST3*, n=9, points are individuals). As

expected, the PTV results in approximately halved mean Cystatin 3 levels, while individuals with a duplication have approximately 50% higher mean Cystatin 3 levels.

## Global and local heritability of biomarkers

**Supplementary Table 11A.** LD Score regression-based estimated SNP heritability and intercept.

**Supplementary Table 11B.** HESS estimated SNP heritability and cumulative fraction of heritability.

## Targeted phenome-wide association study

**Supplementary Table 12.** List of phenotypes used in the PheWAS analysis.

**Supplementary Table 13.** Phenome-wide association for the targeted variants ($p < 1x10^{-7}$).

## Fine-mapping of common associated variants

**Supplementary Table 14A.** List of fine-mapped variants with >99% posterior inclusion probability.

**Supplementary Table 14B.** Residual phenotypic variance explained by fine-mapped variants with >99% posterior inclusion probability.

## Causal inference between biomarkers, diseases, and medically relevant phenotypes

**Supplementary Table 15.** Disease and medically relevant phenotypes used for Mendelian randomization analysis.

**Supplementary Table 16.** Causal inference results using MR.

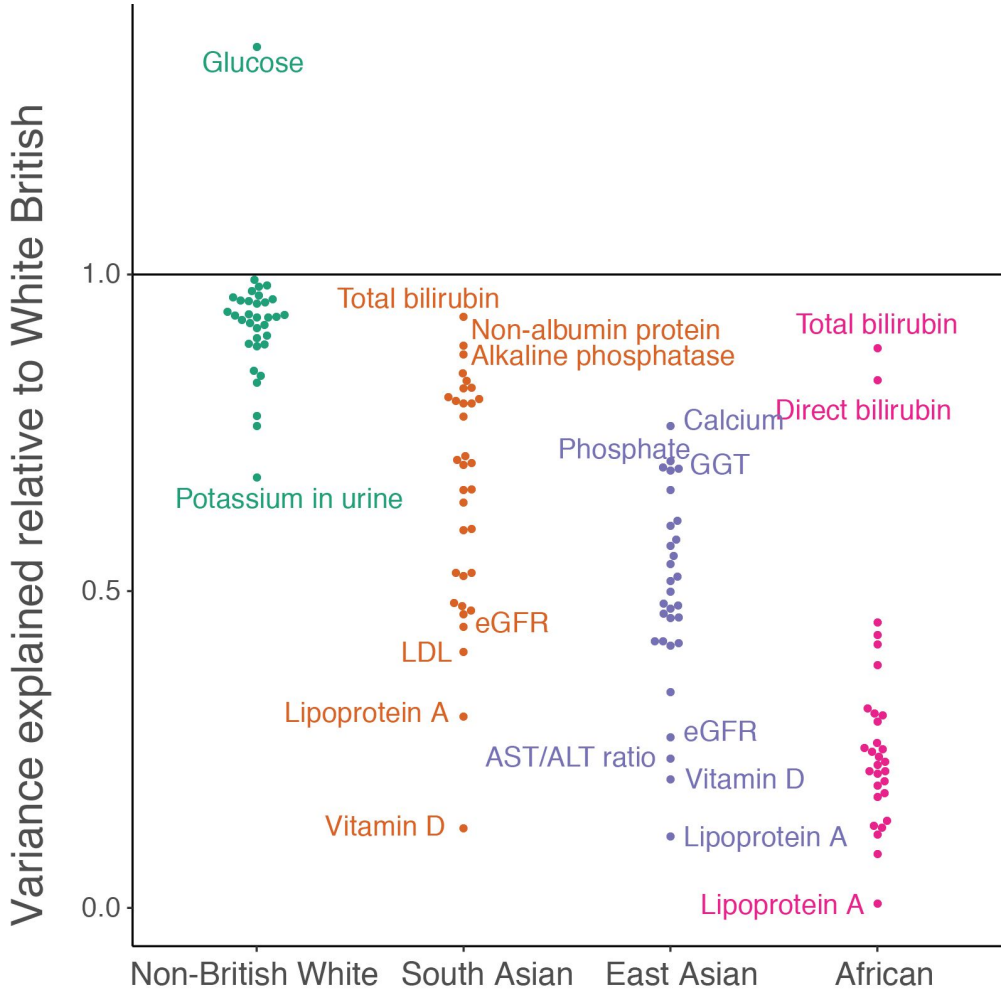
## Polygenic prediction of biomarkers within and across populations

### Evaluation of snpnet PRS models with MESA cohort

MESA SHARE genotypes from all populations were imputed using the Haplotype Reference Consortium reference panel using the Sanger Imputation Server with EAGLE pre-phasing[73,103]. HRC filtering and pre-checking SNPs was applied before imputation[104]. Following imputation, the PRSs were scored using PLINK separately for each chromosome using the allele sum option and scores were then summed across chromosomes before being used for prediction. In evaluation, individual biomarkers from exam 1 or exam 2 were matched to the corresponding measurements in UK Biobank, including biomarkers part of ancillary studies. The evaluation was run within the individuals identified as white in MESA Exam 1. Cholesterol measurements were divided by 0.8 for individuals reporting statin use at exam 1 (as reported in the sttn1c variable) and LDL measurements were divided by 0.7 for individuals reporting statin use at exam 1 similarly. Then, biomarker measurements were log-transformed and adjusted for the exam 1 variables gender1 and age1c.

**Supplementary Table 17A.** Predictive performance of polygenic risk scores for 35 lab phenotypes from genetic data and covariates within and across populations.

**Supplementary Table 17B.** Effect of adjustment for covariates on population-specific polygenic risk score predictive performance.




**Supplementary Figure 7. Variance explained for polygenic scores of individual biomarkers across populations.** The variance explained in a held-out subset of White British was used to normalize each biomarker and prediction was applied in non-British White, South Asian, East Asian, and African ancestry individuals (Methods). For each population, outlier points are labeled. Full list of all results available in Supplementary Table 17A.

**Supplementary Table 18.** Validation of polygenic risk scores in the Multi-Ethnic Study of Atherosclerosis (MESA).

**Supplementary Table 19.** Enrichment in tails of PRS for multiple related traits.

## Multiple regression with PRSs for biomarkers improves prediction of traits and diseases

We also considered a model of myocardial infarction adjusted for self-reported family history of heart disease (referred to as “Heart Disease”). For the family history model, individuals were adjusted by a binary indicator of whether they reported their mother or father as having heart

disease, and ambiguous individuals (for which data from at least one parent was missing, and no parents were reported as having a history of heart disease) were excluded.

For Figure 5A, we generated the corresponding polygenic scores of the multi-PRS in the test set and calculated the prevalence of chronic kidney disease in each non-overlapping quantile of the corresponding multi-polygenic scores.

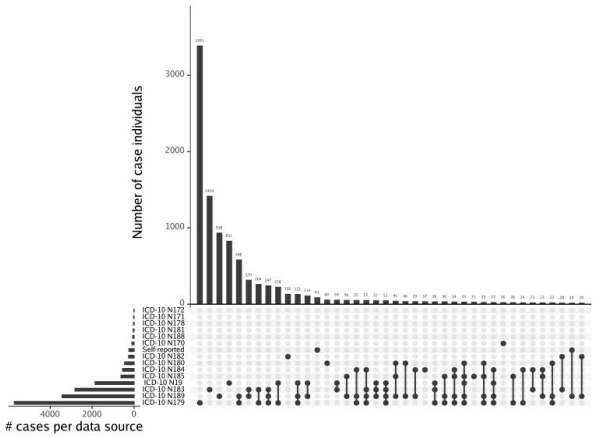
**Supplementary Table 20.** Prediction accuracy and odds ratios for multi-PRS models including those with pre-existing PRSs added.

**Supplementary Table 21.** Cross-population prediction of complex traits with multi-PRS models.

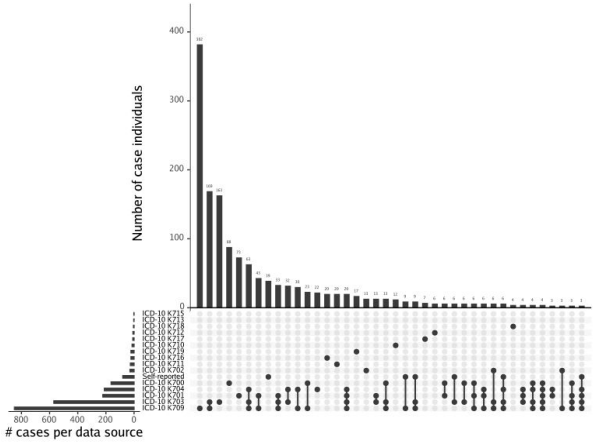
**Supplementary Table 22.** Weights on standardized PRSs for generating the baseline models only including trait polygenic scores and not biomarkers.

**Supplementary Table 23.** Weights on standardized PRSs for generating the multi-PRS models.

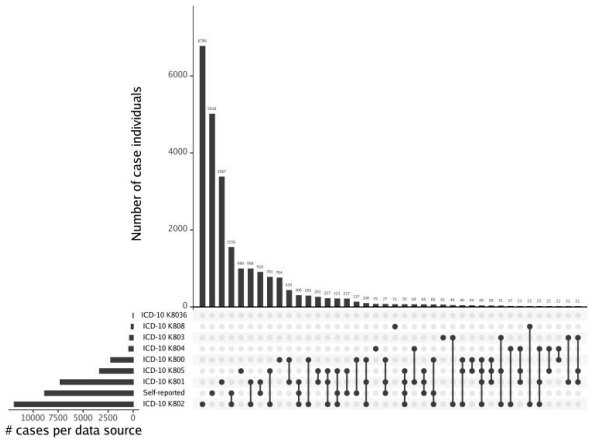
(A) CKD



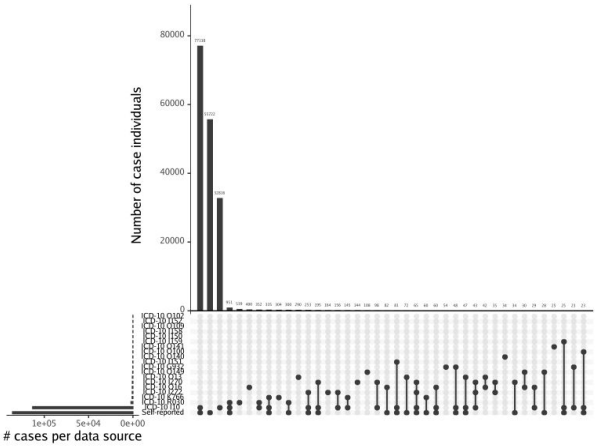

(B) Alcoholic cirrhosis



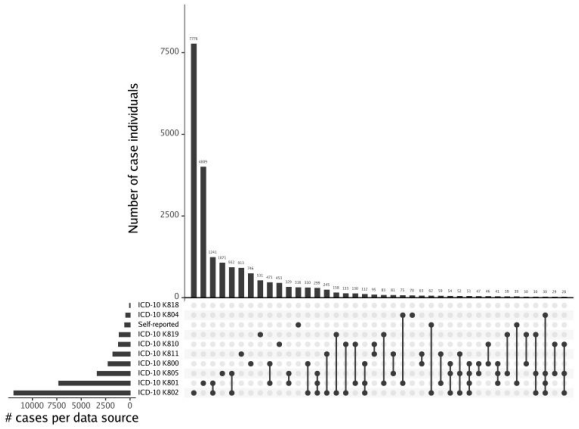

(C) Gallstones



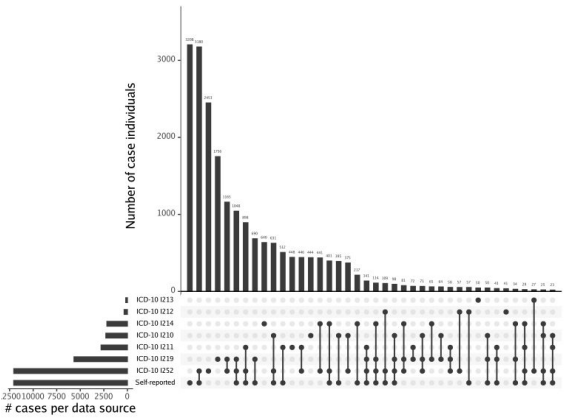

(D) Hypertension

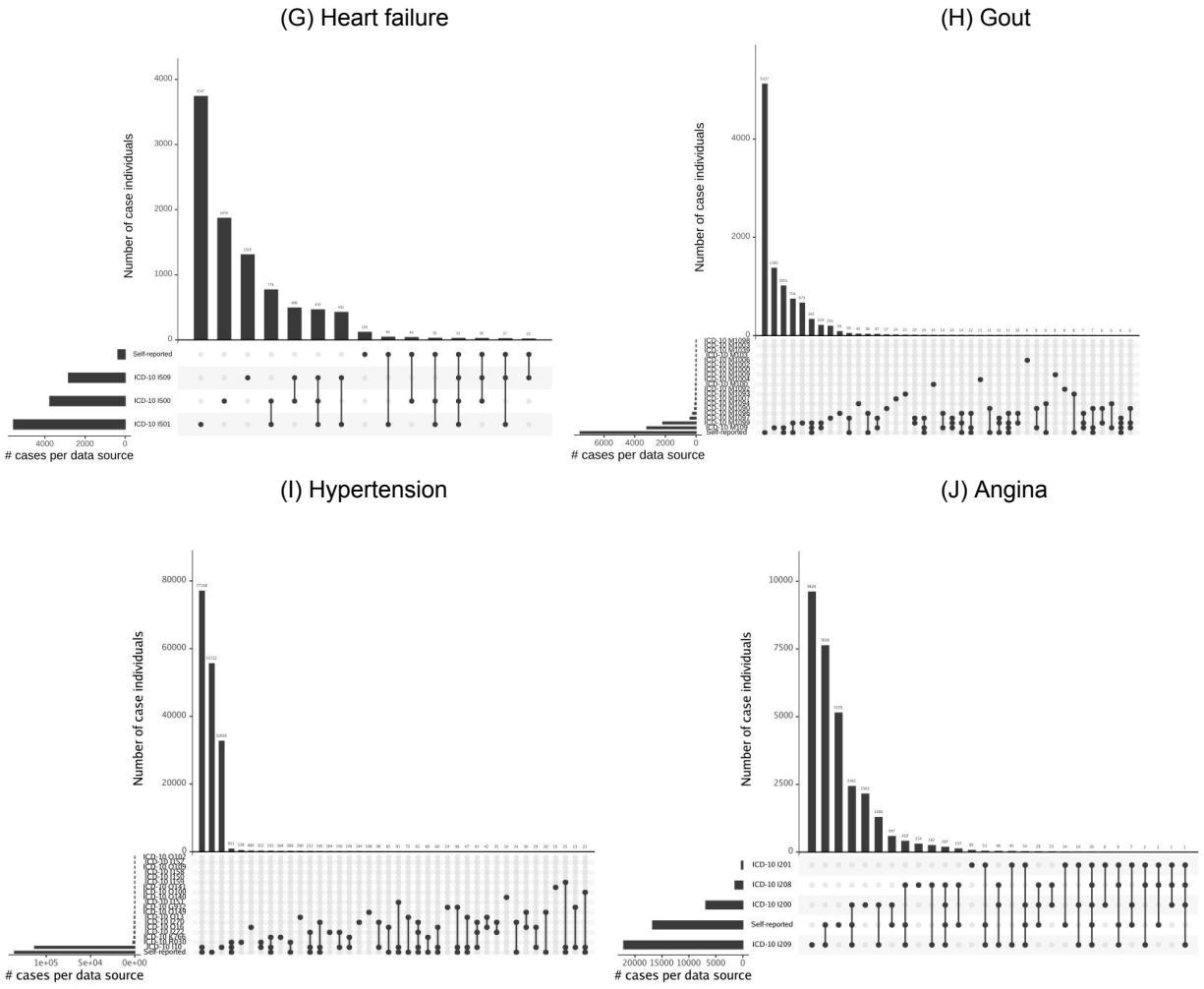



(E) Cholecystitis

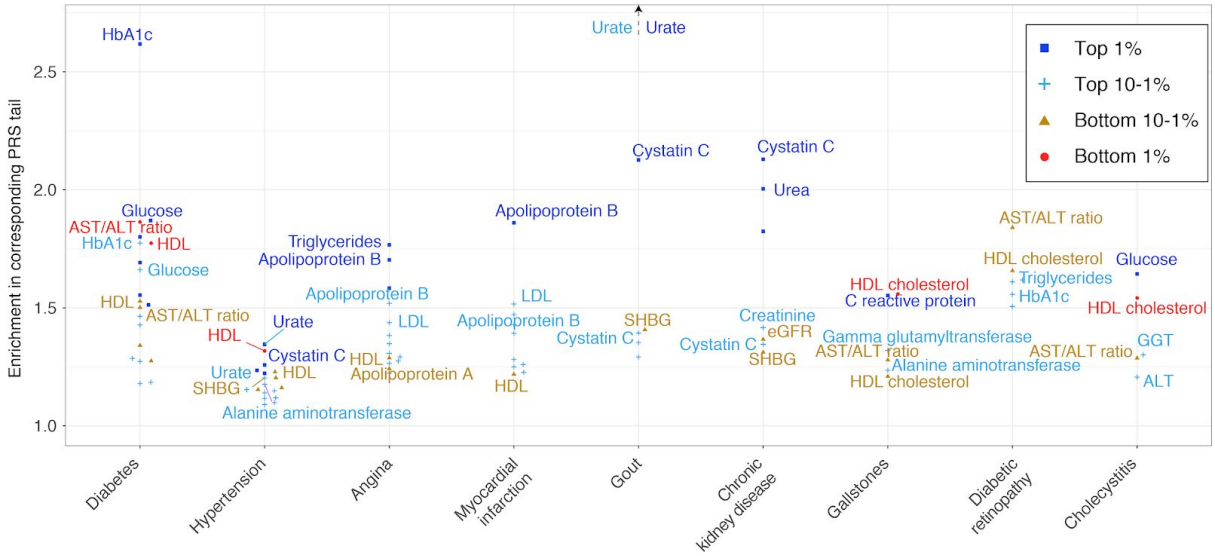



(F) Myocardial infarction



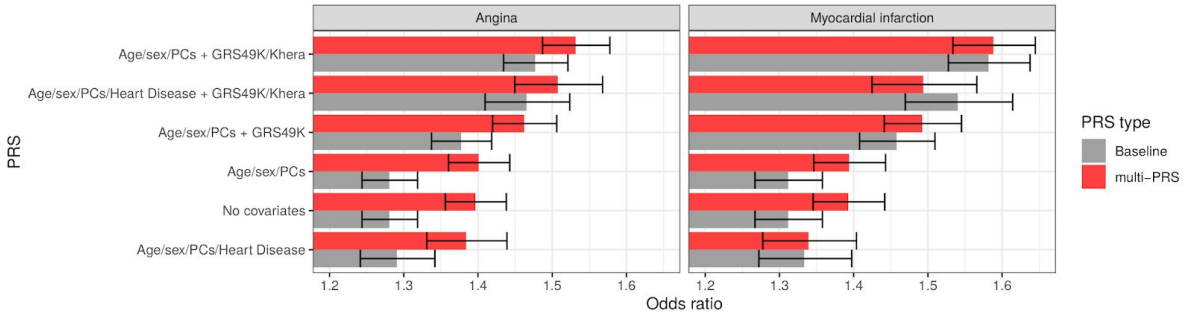



**Supplementary Figure 8. The breakdown of the data sources used for the definition of evaluated diseases in UK Biobank.** Chronic kidney disease (CKD, panel A) endpoints was defined based on the combination of self-reported renal failure (coded as "1192" in UKB Data coding ID 6) and ICD-10 code (N17 [“Acute kidney failure”], N18 [“Chronic kidney disease (CKD)”], N19 [“Unspecified kidney failure”], and its sub-concepts) from hospital inpatient data are used for the chronic kidney disease definition in UK Biobank. For each other disease, the most common 40 combinations of phenotyping sources are shown in the plot: (b) alcoholic cirrhosis, (c) gallstones, (d) hypertension, (e) cholecystitis, (f) myocardial infarction, (g) heart failure, (h) gout, (i) hypertension, and (j) angina.

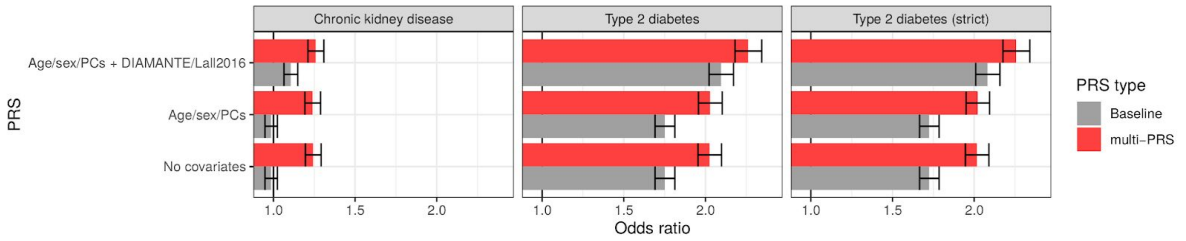



**Supplementary Figure 9. Traits enriched in multiple polygenic scores.** Plot of observed enrichment in PRS tails across the nine traits enriched in at least three distinct traits. Only FDR adjusted significant associations are shown.

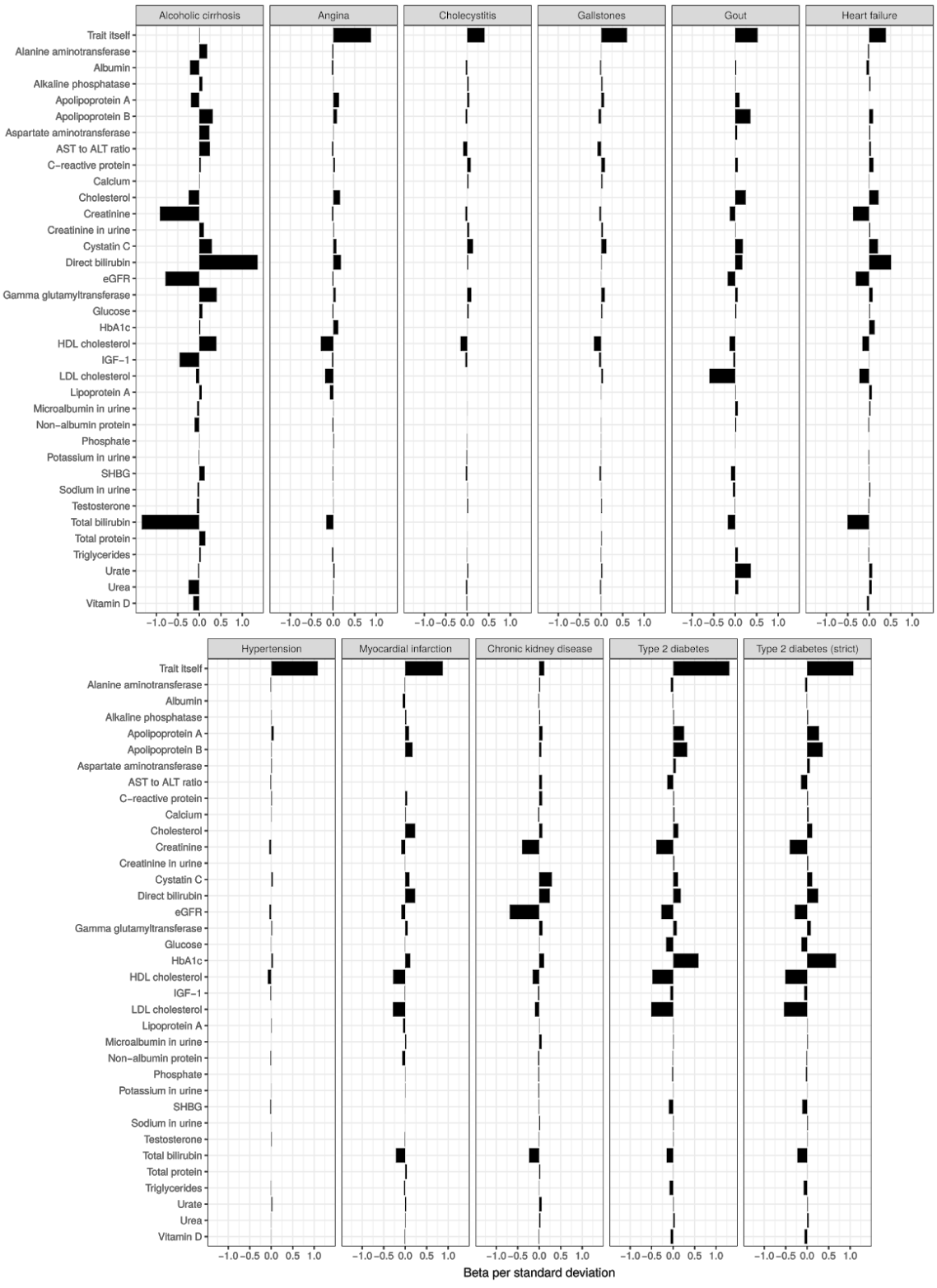
(A) Myocardial infarction




(B) Chronic kidney disease and diabetes




**Supplementary Figure 10. Extending multi-PRS with additional trait polygenic scores. (A) Myocardial infarction.** Odds ratios for angina and myocardial infarction using the multi-PRS including biomarkers (red) or multi-PRS of just the trait polygenic score and existing scores (GRS49K [PMID 27655226] and Khera et al. [PMID 30104762] SNPs; grey). Error bars represent 95% confidence intervals and bar endpoints represent the mean odds ratio estimate from the logistic regression in the test set. Angina n = 4983 cases and n = 89409 total; myocardial infarction n = 3495 cases and n = 89409 total. **(B) Chronic kidney disease and diabetes.** Odds ratios for diabetes and chronic kidney disease using the multi-PRS including biomarkers (red) or multi-PRS of just the trait polygenic score and existing scores (DIAMANTE [PMID 30297969] and Läll et al. 2016 [PMID 27513194]; grey). Error bars represent 95% confidence intervals and bar endpoints represent the mean odds ratio estimate from the logistic regression in the test set. Type 2 diabetes (strict) refers to training a polygenic score with type 2 diabetes controls with HbA1c > 39 mmol/mol excluded. Type 2 diabetes n = 3612 cases and n = 76257 total; CKD n = 2780 cases and n = 89409 total.

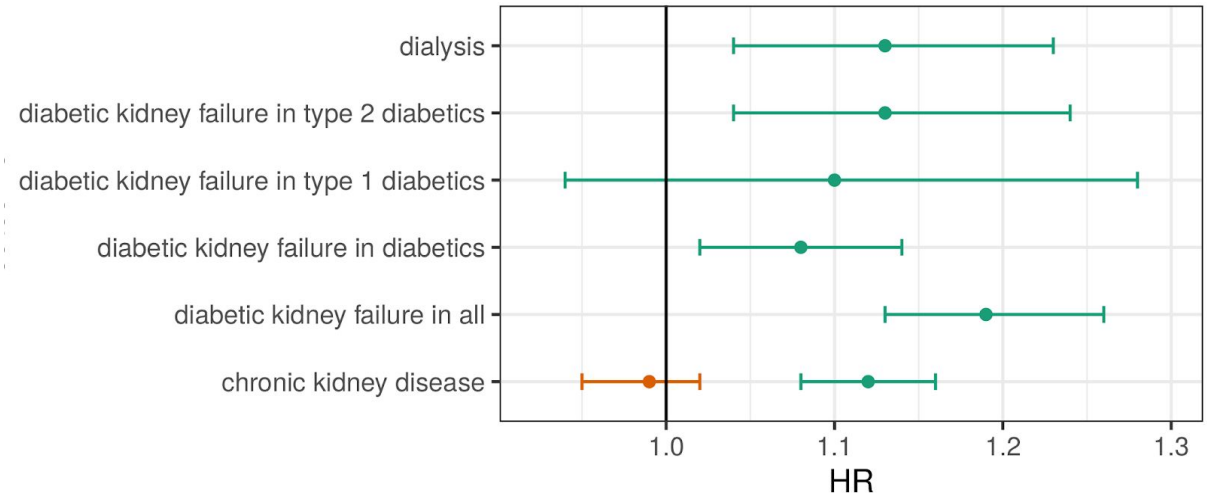



**Supplementary Figure 11. Visualization of multi-PRS weights.** Betas per standard deviation fit for each of the multi-PRSs which show, for each biomarker and the trait baseline score, the beta (log odds) of the given outcome for each standard deviation change in that score. Type 2 diabetes (strict) refers to training a polygenic score with type 2 diabetes controls with HbA1c > 39 mmol/mol excluded.

**Supplementary Table 24.** Description of case definition in FinnGen derived from ICD codes and registry data.

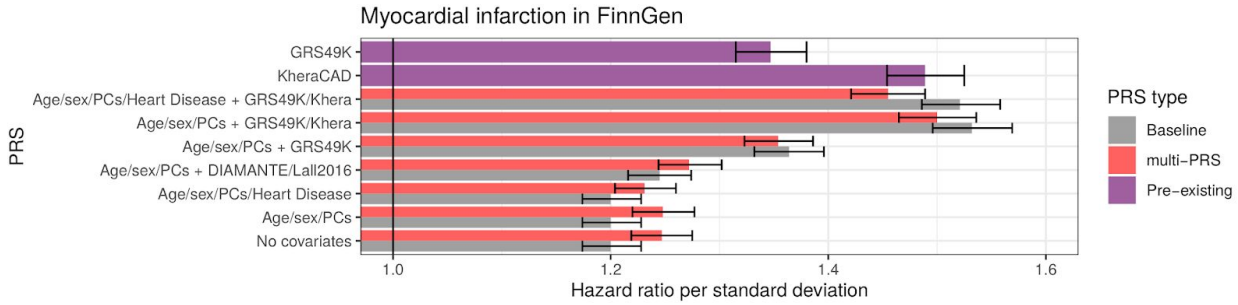
**Supplementary Table 25.** Hazard ratios of multi-PRS models evaluated in the FinnGen cohort.

**Supplementary Table 26.** Hazard ratios of incident cases for models including pre-existing polygenic scores for both type 2 diabetes and myocardial infarction in FinnGen.

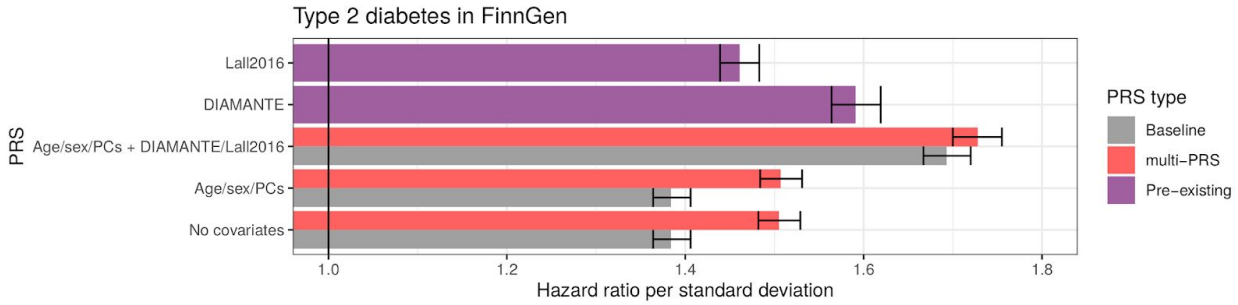



**Supplementary Figure 12. Extended diabetes status predictions of kidney failure multi-PRS.** Hazard ratios (HR) for incidence of various outcomes using the kidney failure multi-PRS (green) or snpnet PRS for kidney failure (orange). Error bars represent 95% confidence intervals and points represent the mean hazard ratio estimate from the regression fit. Number of individuals with each diagnosis, statistical significance, and covariates described in Supplementary Table 23 and Methods.

(A) Myocardial infarction

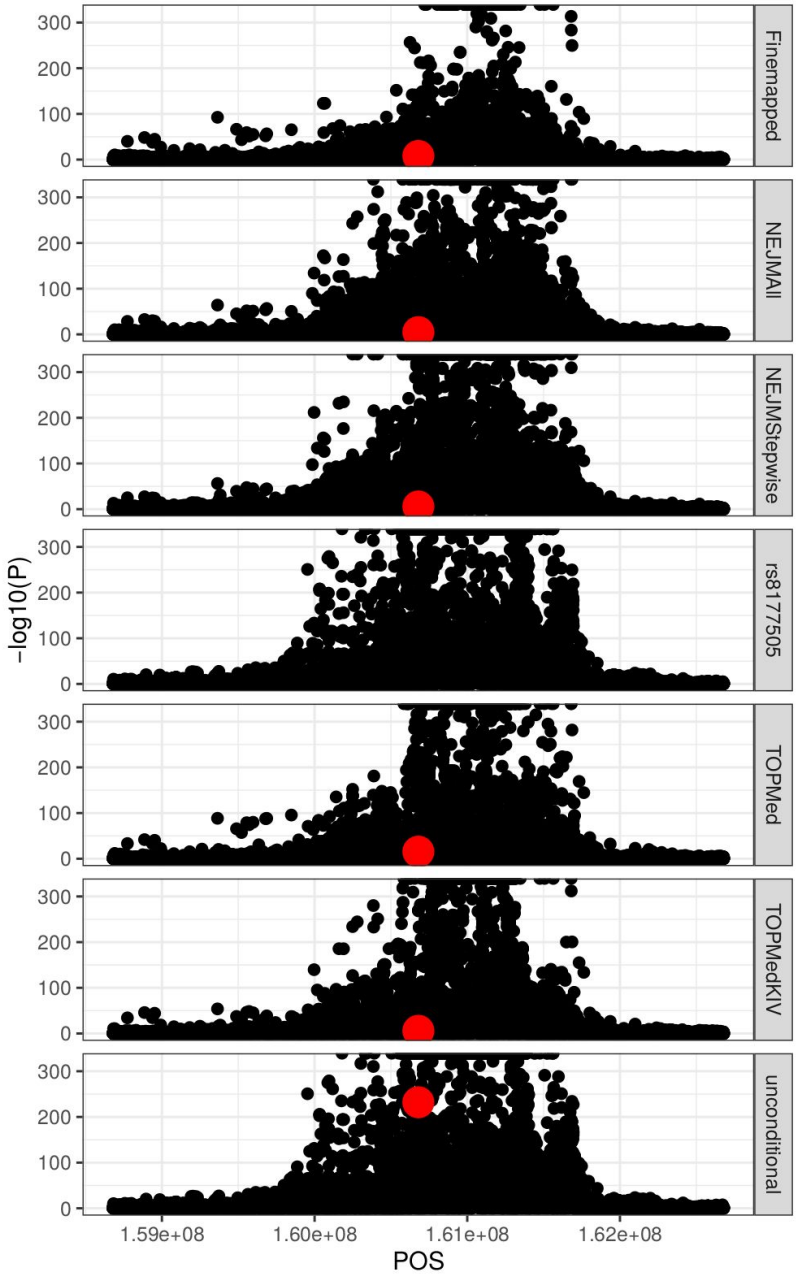



(B) Type 2 diabetes




**Supplementary Figure 13. Extending multi-PRS with additional trait polygenic scores in FinnGen. (A) Myocardial infarction.** Hazard ratios for myocardial infarction using the multi-PRS including biomarkers (red) or multi-PRS of just the trait polygenic score and existing scores (grey; including just GRS49K and Khera et al. SNPs; purple). Error bars represent 95% confidence intervals and bar endpoints represent the mean odds ratio estimate from the regression fit. Myocardial infarction (7612 cases and 122161 controls) is the outcome evaluated. **(B) Type 2 diabetes.** Hazard ratios for type 2 diabetes using the multi-PRS including biomarkers (red) or multi-PRS of just the trait polygenic score and existing scores (grey; including just DIAMANTE and Lall et al. 2016 SNPs, purple). Error bars represent 95% confidence intervals and bar endpoints represent the mean odds ratio estimate from the regression fit. Type 2 diabetes (strict; 16877 cases and 129800 controls) is the outcome evaluated. .

## Conditional effects at rs8177505 and the *LPA* locus




**Supplementary Figure 14. Extended comparison of conditional effects at rs8177505 and the *LPA* locus.** The putative LpA-associated variant rs8177505 is shown in red, including the marginal effect "unconditional" (rs8177505 beta = 1.44208, two-sided linear regression p = 9.46191 x $10^{-233}$); the effect sizes conditioned on the genotype at rs8177505 ("rs8177505"); the effect sizes conditioned on all the >99% posterior probability fine-mapped SNPs from the study ("Finemapped" -- rs8177505 beta = 0.213102, two-sided linear regression p = 1.1 x $10^{-8}$); for the list of all variants and stepwise independent variants from a previous study ("NEJMAll" and "NEJMStepwise" [105] -- rs8177505 beta = 0.157889 and 0.175442, two-sided linear regression p = 1.3 x $10^{-5}$ and 1.3 x $10^{-6}$); and for the list of independent and Kringle IV repeat associated variants from another study ("TOPMed" and "TOPMedKIV" [106] -- rs8177505 beta = 0.27945 and 0.19539, two-sided linear regression p = 3.3 x $10^{-16}$ and 1.1 x $10^{-6}$).

# Additional datasets generated in this work

Summary-level data generated in this work is available at NIH's instance of figshare[78].

**The meta-analyzed GWAS summary statistics for 35 lab biomarkers.** For each biomarker trait, we provide two table files for meta-analyzed GWAS summary statistics, corresponding to the one from the directly genotyped dataset and the imputed dataset. The columns are: the position of the genetic variant (CHROM and POS), the variant identifier (MarkerName), the reference and the alternate allele (REF and ALT), the effect size estimates and its standard error (Effect and StdErr) reported with respect to the alternate allele as the risk allele, the meta-analyzed p-value of the association (P-value), the direction of the association (Direction) in the four populations (white British, non-British white, African, and South Asian), and the METAL heterogeneity test statistics (HetISq, HetChiSq, HetDf, and HetPVal). This dataset is available at NIH's instance of figshare: https://doi.org/10.35092/yhjc.12355382.

**The output from the FINEMAP analysis for 35 lab biomarkers.** For each trait, we provide a tar archive file which contains the full output from FINEMAP for the regions with at least one genome-wide significant associations ($p < 5 \times 10^{-9}$) from the multi-ethnic GWAS meta-analysis within UK Biobank. Specifically, we provide bdose, config, cred, ld, master, snp, and z files used in FINEMAP software, whose file formats are described in its documentation (http://christianbenner.com/). This dataset is available at NIH's instance of figshare: https://doi.org/10.35092/yhjc.12344351.

**The snpnet polygenic risk score coefficients for 35 lab biomarkers.** The coefficients (weights) of the polygenic risk score are provided in a table file with the following set of columns: the position of the genetic variant (CHROM and POS), the variant identifier (ID), the reference and the alternate allele (REF and ALT), the coefficients (weights) of the PRS (BETA). The BETA is always reported for the alternate allele. This dataset is available at NIH's instance of figshare: https://doi.org/10.35092/yhjc.12298838.

**The multi-PRS risk score coefficients for 10 disease outcomes.** The multi-PRS risk score coefficients for the following 10 disease endpoints are included in this dataset: angina, alcoholic cirrhosis, gallstones, hypertension, cholecystitis, kidney failure, heart failure, myocardial infarction, gout, and type 2 diabetes (T2D). The coefficients (weights) of the polygenic risk score are provided in a table file with the following set of columns: the position of the genetic variant (CHROM and POS), the variant identifier (ID), the reference, the alternate allele, and the risk alleles (REF, ALT, and A1), the coefficients (weights) of the PRS weights (weights_<trait>). This dataset is available at NIH's instance of figshare: https://doi.org/10.35092/yhjc.12355424.

# Contributors of FinnGen

## Steering Committee

| | |
|---|---|
| Aarno Palotie | Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland |
| Mark Daly | Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland |

## Pharmaceutical companies

| | |
|---|---|
| Howard Jacob | Abbvie, Chicago, IL, United States |
| Athena Matakidou | Astra Zeneca, Cambridge, United Kingdom |
| Heiko Runz | Biogen, Cambridge, MA, United States |
| Sally John | Biogen, Cambridge, MA, United States |
| Robert Plenge | Celgene, Summit, NJ, United States |
| Mark McCarthy | Genentech, San Francisco, CA, United States |
| Julie Hunkapiller | Genentech, San Francisco, CA, United States |
| Meg Ehm | GlaxoSmithKline, Brentford, United Kingdom |
| Dawn Waterworth | GlaxoSmithKline, Brentford, United Kingdom |
| Caroline Fox | Merck, Kenilworth, NJ, United States |

| | |
|---|---|
| Anders Malarstig | Pfizer, New York, NY, United States |
| Kathy Klinger | Sanofi, Paris, France |
| Kathy Call | Sanofi, Paris, France |

### University of Helsinki & Biobanks

| | |
|---|---|
| Tomi Mäkelä | HiLIFE, University of Helsinki, Finland, Finland |
| Jaakko Kaprio | Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland, Finland |
| Petri Virolainen | Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland |
| Kari Pulkki | Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland |
| Terhi Kilpi | THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland |
| Markus Perola | THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland |
| Jukka Partanen | Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland |
| Anne Pitkäranta | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Riitta Kaarteenaho | Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland |
| Seppo Vainio | Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland |
| Kimmo Savinainen | Finnish Clinical Biobank Tampere/University of Tampere/Pirkanmaa Hospital District, Tampere, Finland |
| Veli-Matti Kosma | Biobank of Eastern Finland/ University of Eastern Finland/ Northern Savo Hospital District, Kuopio, Finland |
| Urho Kujala | Central Finland Biobank/University of Jyväskylä/ Central Finland Health Care District, Jyväskylä, Finland |

### Other Experts/ Non-Voting Members

| | |
|---|---|
| Outi Tuovila | Business Finland, Helsinki, Finland |
| Minna Hendolin | Business Finland, Helsinki, Finland |
| Raimo Pakkanen | Business Finland, Helsinki, Finland |

## Scientific Committee

### Pharmaceutical companies

| | |
|---|---|
| Jeff Waring | Abbvie, Chicago, IL, United States |
| Bridget Riley-Gillis | Abbvie, Chicago, IL, United States |
| Athena Matakidou | Astra Zeneca, Cambridge, United Kingdom |
| Heiko Runz | Biogen, Cambridge, MA, United States |
| Jimmy Liu | Biogen, Cambridge, MA, United States |
| Shameek Biswas | Celgene, Summit, NJ, United States |
| Julie Hunkapiller | Genentech, San Francisco, CA, United States |
| Dawn Waterworth | GlaxoSmithKline, Brentford, United Kingdom |
| Meg Ehm | GlaxoSmithKline, Brentford, United Kingdom |
| Dorothee Diogo | Merck, Kenilworth, NJ, United States |
| Caroline Fox | Merck, Kenilworth, NJ, United States |
| Anders Malarstig | Pfizer, New York, NY, United States |
| Catherine Marshall | Pfizer, New York, NY, United States |
| Xinli Hu | Pfizer, New York, NY, United States |
| Kathy Call | Sanofi, Paris, France |

| | |
|---|---|
| Kathy Klinger | Sanofi, Paris, France |
| Matthias Gossel | Sanofi, Paris, France |

### University of Helsinki & Biobanks

| | |
|---|---|
| Samuli Ripatti | Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland |
| Johanna Schleutker | Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland |
| Markus Perola | THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland |
| Mikko Arvas | Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland |
| Olli Carpén | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Reetta Hinttala | Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland |
| Johannes Kettunen | Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland |
| Reijo Laaksonen | Finnish Clinical Biobank Tampere**/**University of Tampere/ Pirkanmaa Hospital District, Tampere, Finland |
| Arto Mannermaa | Biobank of Eastern Finland/University of Eastern Finland/ Northern Savo Hospital District, Kuopio, Finland |
| Juha Paloneva | Central Finland Biobank/University of Jyväskylä/ Central Finland Health Care District, Jyväskylä, Finland |
| Urho Kujala | Central Finland Biobank/University of Jyväskylä/ Central Finland Health Care District, Jyväskylä, Finland |

### Other Experts/ Non-Voting Members

| | |
|---|---|
| Outi Tuovila | Business Finland, Helsinki, Finland |
| Minna Hendolin | Business Finland, Helsinki, Finland |
| Raimo Pakkanen | Business Finland, Helsinki, Finland |

## Clinical Expert Groups

### Neurology

| | |
|---|---|
| Hilkka Soininen | Northern Savo Hospital District, Kuopio, Finland |
| Valtteri Julkunen | Northern Savo Hospital District, Kuopio, Finland |
| Anne Remes | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Reetta Kälviäinen | Northern Savo Hospital District, Kuopio, Finland |
| Mikko Hiltunen | Northern Savo Hospital District, Kuopio, Finland |
| Jukka Peltola | Pirkanmaa Hospital District, Tampere, Finland |
| Pentti Tienari | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Juha Rinne | Hospital District of Southwest Finland, Turku, Finland |
| Adam Ziemann | Abbvie, Chicago, IL, United States |
| Jeffrey Waring | Abbvie, Chicago, IL, United States |
| Sahar Esmaeeli | Abbvie, Chicago, IL, United States |
| Nizar Smaoui | Abbvie, Chicago, IL, United States |
| Anne Lehtonen | Abbvie, Chicago, IL, United States |
| Susan Eaton | Biogen, Cambridge, MA, United States |
| Heiko Runz | Biogen, Cambridge, MA, United States |
| Sanni Lahdenperä | Biogen, Cambridge, MA, United States |
| Shameek Biswas | Celgene, Summit, NJ, United States |

| | |
|---|---|
| John Michon | Genentech, San Francisco, CA, United States |
| Geoff Kerchner | Genentech, San Francisco, CA, United States |
| Julie Hunkapiller | Genentech, San Francisco, CA, United States |
| Natalie Bowers | Genentech, San Francisco, CA, United States |
| Edmond Teng | Genentech, San Francisco, CA, United States |
| John Eicher | Merck, Kenilworth, NJ, United States |
| Vinay Mehta | Merck, Kenilworth, NJ, United States |
| Padhraig Gormley | Merck, Kenilworth, NJ, United States |
| Kari Linden | Pfizer, New York, NY, United States |
| Christopher Whelan | Pfizer, New York, NY, United States |
| Fanli Xu | GlaxoSmithKline, Brentford, United Kingdom |
| David Pulford | GlaxoSmithKline, Brentford, United Kingdom |

## Gastroenterology

| | |
|---|---|
| Martti Färkkilä | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Sampsa Pikkarainen | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Airi Jussila | Pirkanmaa Hospital District, Tampere, Finland |
| Timo Blomster | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Mikko Kiviniemi | Northern Savo Hospital District, Kuopio, Finland |
| Markku Voutilainen | Hospital District of Southwest Finland, Turku, Finland |
| Bob Georgantas | Abbvie, Chicago, IL, United States |
| Graham Heap | Abbvie, Chicago, IL, United States |
| Jeffrey Waring | Abbvie, Chicago, IL, United States |
| Nizar Smaoui | Abbvie, Chicago, IL, United States |
| Fedik Rahimov | Abbvie, Chicago, IL, United States |
| Anne Lehtonen | Abbvie, Chicago, IL, United States |
| Keith Usiskin | Celgene, Summit, NJ, United States |
| Joseph Maranville | Celgene, Summit, NJ, United States |
| Tim Lu | Genentech, San Francisco, CA, United States |
| Natalie Bowers | Genentech, San Francisco, CA, United States |
| Danny Oh | Genentech, San Francisco, CA, United States |
| John Michon | Genentech, San Francisco, CA, United States |
| Vinay Mehta | Merck, Kenilworth, NJ, United States |
| Kirsi Kalpala | Pfizer, New York, NY, United States |
| Melissa Miller | Pfizer, New York, NY, United States |
| Xinli Hu | Pfizer, New York, NY, United States |
| Linda McCarthy | GlaxoSmithKline, Brentford, United Kingdom |

## Rheumatology

| | |
|---|---|
| Kari Eklund | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Antti Palomäki | Hospital District of Southwest Finland, Turku, Finland |
| Pia Isomäki | Pirkanmaa Hospital District, Tampere, Finland |
| Laura Pirilä | Hospital District of Southwest Finland, Turku, Finland |
| Oili Kaipiainen-Seppänen | Northern Savo Hospital District, Kuopio, Finland |
| Johanna Huhtakangas | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Bob Georgantas | Abbvie, Chicago, IL, United States |
| Jeffrey Waring | Abbvie, Chicago, IL, United States |
| Fedik Rahimov | Abbvie, Chicago, IL, United States |
| Apinya Lertratanakul | Abbvie, Chicago, IL, United States |
| Nizar Smaoui | Abbvie, Chicago, IL, United States |

| | |
|---|---|
| Anne Lehtonen | Abbvie, Chicago, IL, United States |
| David Close | Astra Zeneca, Cambridge, United Kingdom |
| Marla Hochfeld | Celgene, Summit, NJ, United States |
| Natalie Bowers | Genentech, San Francisco, CA, United States |
| John Michon | Genentech, San Francisco, CA, United States |
| Dorothee Diogo | Merck, Kenilworth, NJ, United States |
| Vinay Mehta | Merck, Kenilworth, NJ, United States |
| Kirsi Kalpala | Pfizer, New York, NY, United States |
| Nan Bing | Pfizer, New York, NY, United States |
| Xinli Hu | Pfizer, New York, NY, United States |
| Jorge Esparza Gordillo | GlaxoSmithKline, Brentford, United Kingdom |
| Nina Mars | Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland |

## Pulmonology

| | |
|---|---|
| Tarja Laitinen | Pirkanmaa Hospital District, Tampere, Finland |
| Margit Pelkonen | Northern Savo Hospital District, Kuopio, Finland |
| Paula Kauppi | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Hannu Kankaanranta | Pirkanmaa Hospital District, Tampere, Finland |
| Terttu Harju | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Nizar Smaoui | Abbvie, Chicago, IL, United States |
| David Close | Astra Zeneca, Cambridge, United Kingdom |
| Steven Greenberg | Celgene, Summit, NJ, United States |
| Hubert Chen | Genentech, San Francisco, CA, United States |
| Natalie Bowers | Genentech, San Francisco, CA, United States |
| John Michon | Genentech, San Francisco, CA, United States |
| Vinay Mehta | Merck, Kenilworth, NJ, United States |
| Jo Betts | GlaxoSmithKline, Brentford, United Kingdom |
| Soumitra Ghosh | GlaxoSmithKline, Brentford, United Kingdom |

## Cardiometabolic Diseases

| | |
|---|---|
| Veikko Salomaa | The National Institute of Health and Welfare, Helsinki, Finland |
| Teemu Niiranen | The National Institute of Health and Welfare, Helsinki, Finland |
| Markus Juonala | Hospital District of Southwest Finland, Turku, Finland |
| Kaj Metsärinne | Hospital District of Southwest Finland, Turku, Finland |
| Mika Kähönen | Pirkanmaa Hospital District, Tampere, Finland |
| Juhani Junttila | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Markku Laakso | Northern Savo Hospital District, Kuopio, Finland |
| Jussi Pihlajamäki | Northern Savo Hospital District, Kuopio, Finland |
| Juha Sinisalo | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Marja-Riitta Taskinen | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Tiinamaija Tuomi | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Jari Laukkanen | Central Finland Health Care District, Jyväskylä, Finland |
| Ben Challis | Astra Zeneca, Cambridge, United Kingdom |
| Andrew Peterson | Genentech, San Francisco, CA, United States |
| Julie Hunkapiller | Genentech, San Francisco, CA, United States |
| Natalie Bowers | Genentech, San Francisco, CA, United States |
| John Michon | Genentech, San Francisco, CA, United States |
| Dorothee Diogo | Merck, Kenilworth, NJ, United States |

| | |
|---|---|
| Audrey Chu | Merck, Kenilworth, NJ, United States |
| Vinay Mehta | Merck, Kenilworth, NJ, United States |
| Jaakko Parkkinen | Pfizer, New York, NY, United States |
| Melissa Miller | Pfizer, New York, NY, United States |
| Anthony Muslin | Sanofi, Paris, France |
| Dawn Waterworth | GlaxoSmithKline, Brentford, United Kingdom |

## Oncology

| | |
|---|---|
| Heikki Joensuu | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Tuomo Meretoja | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Olli Carpén | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Lauri Aaltonen | University of Helsinki, Helsinki, Finland |
| Annika Auranen | Pirkanmaa Hospital District , Tampere, Finland |
| Peeter Karihtala | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Saila Kauppila | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Päivi Auvinen | Northern Savo Hospital District, Kuopio, Finland |
| Klaus Elenius | Hospital District of Southwest Finland, Turku, Finland |
| Relja Popovic | Abbvie, Chicago, IL, United States |
| Jeffrey Waring | Abbvie, Chicago, IL, United States |
| Bridget Riley-Gillis | Abbvie, Chicago, IL, United States |
| Anne Lehtonen | Abbvie, Chicago, IL, United States |
| Athena Matakidou | Astra Zeneca, Cambridge, United Kingdom |
| Jennifer Schutzman | Genentech, San Francisco, CA, United States |
| Julie Hunkapiller | Genentech, San Francisco, CA, United States |
| Natalie Bowers | Genentech, San Francisco, CA, United States |
| John Michon | Genentech, San Francisco, CA, United States |
| Vinay Mehta | Merck, Kenilworth, NJ, United States |
| Andrey Loboda | Merck, Kenilworth, NJ, United States |
| Aparna Chhibber | Merck, Kenilworth, NJ, United States |
| Heli Lehtonen | Pfizer, New York, NY, United States |
| Stefan McDonough | Pfizer, New York, NY, United States |
| Marika Crohns | Sanofi, Paris, France |
| Diptee Kulkarni | GlaxoSmithKline, Brentford, United Kingdom |

## Opthalmology

| | |
|---|---|
| Kai Kaarniranta | Northern Savo Hospital District, Kuopio, Finland |
| Joni A Turunen | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Terhi Ollila | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Sanna Seitsonen | Hospital District of Helsinki and Uusimaa, Helsinki, Finland |
| Hannu Uusitalo | Pirkanmaa Hospital District, Tampere, Finland |
| Vesa Aaltonen | Hospital District of Southwest Finland, Turku, Finland |
| Hannele Uusitalo-Järvinen | Pirkanmaa Hospital District, Tampere, Finland |
| Marja Luodonpää | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Nina Hautala | Northern Ostrobothnia Hospital District, Oulu, Finland |
| Heiko Runz | Biogen, Cambridge, MA, United States |
| Erich Strauss | Genentech, San Francisco, CA, United States |
| Natalie Bowers | Genentech, San Francisco, CA, United States |
| Hao Chen | Genentech, San Francisco, CA, United States |
| John Michon | Genentech, San Francisco, CA, United States |

Anna Podgornaia Merck, Kenilworth, NJ, United States
Vinay Mehta Merck, Kenilworth, NJ, United States
Dorothee Diogo Merck, Kenilworth, NJ, United States
Joshua Hoffman GlaxoSmithKline, Brentford, United Kingdom

### Dermatology

Kaisa Tasanen Northern Ostrobothnia Hospital District, Oulu, Finland
Laura Huilaja Northern Ostrobothnia Hospital District, Oulu, Finland
Katariina Hannula-Jouppi Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Teea Salmi Pirkanmaa Hospital District, Tampere, Finland
Sirkku Peltonen Hospital District of Southwest Finland, Turku, Finland
Leena Koulu Hospital District of Southwest Finland, Turku, Finland
Ilkka Harvima Northern Savo Hospital District, Kuopio, Finland
Kirsi Kalpala Pfizer, New York, NY, United States
Ying Wu Pfizer, New York, NY, United States
David Choy Genentech, San Francisco, CA, United States
John Michon Genentech, San Francisco, CA, United States
Nizar Smaoui Abbvie, Chicago, IL, United States
Fedik Rahimov Abbvie, Chicago, IL, United States
Anne Lehtonen Abbvie, Chicago, IL, United States
Dawn Waterworth GlaxoSmithKline, Brentford, United Kingdom

## Analysis working group

Justin Wade Davis Abbvie, Chicago, IL, United States
Bridget Riley-Gillis Abbvie, Chicago, IL, United States
Danjuma Quarless Abbvie, Chicago, IL, United States
Slavé Petrovski Astra Zeneca, Cambridge, United Kingdom
Jimmy Liu Biogen, Cambridge, MA, United States
Chia-Yen Chen Biogen, Cambridge, MA, United States
Paola Bronson Biogen, Cambridge, MA, United States
Robert Yang Celgene, Summit, NJ, United States
Joseph Maranville Celgene, Summit, NJ, United States
Shameek Biswas Celgene, Summit, NJ, United States
Diana Chang Genentech, San Francisco, CA, United States
Julie Hunkapiller Genentech, San Francisco, CA, United States
Tushar Bhangale Genentech, San Francisco, CA, United States
Natalie Bowers Genentech, San Francisco, CA, United States
Dorothee Diogo Merck, Kenilworth, NJ, United States
Emily Holzinger Merck, Kenilworth, NJ, United States
Padhraig Gormley Merck, Kenilworth, NJ, United States
Xulong Wang Merck, Kenilworth, NJ, United States
Xing Chen Pfizer, New York, NY, United States
Åsa Hedman Pfizer, New York, NY, United States
Kirsi Auro GlaxoSmithKline, Brentford, United Kingdom
Clarence Wang Sanofi, Paris, France
Ethan Xu Sanofi, Paris, France
Franck Auge Sanofi, Paris, France
Clement Chatelain Sanofi, Paris, France
Mitja Kurki Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland/Broad Institute, Cambridge, MA, United States

Samuli Ripatti Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mark Daly Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Juha Karjalainen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland/Broad Institute, Cambridge, MA, United States
Aki Havulinna Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Anu Jalanko Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Kimmo Palin of Helsinki, Helsinki, Finland
Priit Palta Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Pietro Della Briotta Parolo Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Wei Zhou Broad Institute, Cambridge, MA, United States
Susanna Lemmelä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Manuel Rivas University of Stanford, Stanford, CA, United States
Jarmo Harju Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Aarno Palotie Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Arto Lehisto Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Andrea Ganna Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Vincent Llorens Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Antti Karlsson Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland
Kati Kristiansson THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Mikko Arvas Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Kati Hyvärinen Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Jarmo Ritari Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Tiina Wahlfors Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland
Miika Koskinen Hospital District of Helsinki and Uusimaa, Helsinki, Finland BB/HUS/Univ Hosp Districts
Olli Carpén Hospital District of Helsinki and Uusimaa, Helsinki, Finland BB/HUS/Univ Hosp Districts
Johannes Kettunen Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland
Katri Pylkäs Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland
Marita Kalaoja Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland
Minna Karjalainen Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland
Tuomo Mantere Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland
Eeva Kangasniemi Finnish Clinical Biobank Tampere/University of Tampere/Pirkanmaa Hospital District, Tampere, Finland
Sami Heikkinen Biobank of Eastern Finland/University of Eastern Finland/ Northern Savo Hospital District, Kuopio, Finland
Arto Mannermaa Biobank of Eastern Finland/University of Eastern Finland/ Northern Savo Hospital District, Kuopio, Finland

Eija Laakkonen Central Finland Biobank/University of Jyväskylä/Central Finland Health Care District, Jyväskylä, Finland
Juha Kononen Central Finland Biobank/University of Jyväskylä/ Central Finland Health Care District, Jyväskylä, Finland

## Biobank directors

Lila Kallio Auria Biobank, Turku, Finland
Sirpa Soini THL Biobank, Helsinki, Finland
Jukka Partanen Blood Service Biobank, Helsinki, Finland
Kimmo Pitkänen Helsinki Biobank, Helsinki, Finland
Seppo Vainio Northern Finland Biobank Borealis, Oulu, Finland
Kimmo Savinainen Tampere Biobank, Tampere, Finland
Veli-Matti Kosma Biobank University University of Eastern Finland, Kuopio, Finland
Teijo Kuopio Central Finland Biobank, Jyväskylä, Finland

## FinnGen Teams

### Administration

Anu Jalanko Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Risto Kajanne Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Ulrike Lyhs Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

### Communication

Mari Kaunisto Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

### Sample Collection Coordination

Anu Loukola Hospital District of Helsinki and Uusimaa, Helsinki, Finland

### Sample Logistics

Päivi Laiho THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Tuuli Sistonen THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Essi Kaiharju THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Markku Laukkanen THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Elina Järvensivu THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Sini Lähteenmäki THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Lotta Männikkö THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Regis Wong THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland

### Registry Data Operations

Kati Kristiansson THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Hannele Mattsson THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Susanna Lemmelä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Tero Hiekkalinna THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland
Manuel González Jiménez THL Biobank/The National Institute of Health and Welfare, Helsinki, Finland

### Genotyping

Kati Donner Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

### Sequencing Informatics

Priit Palta Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Kalle Pärn Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Javier Nunez-Fontarnau Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

### Data Management and IT Infrastructure

Jarmo Harju Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Elina Kilpeläinen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Timo P. Sipilä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Georg Brein Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Oluwaseun A. Dada Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Ghazal Awaisa Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Anastasia Shcherban Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Tuomas Sipilä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

### Analysis

Mitja Kurki Institute for Molecular Medicine Finland, HiLIFE-Helsinki Institute of Life Science, of Helsinki, Helsinki, Finland/Broad Institute, Cambridge, MA, United States
Juha Karjalainen Institute for Molecular Medicine Finland, HiLIFE-Helsinki Institute of Life Science, of Helsinki, Helsinki, Finland/Broad Institute, Cambridge, MA, United States
Pietro della Briotta Parolo Institute for Molecular Medicine Finland, HiLIFE-Helsinki Institute of Life Science, of Helsinki, Helsinki, Finland
Wei Zhou Broad Institute, Cambridge, MA, United States
Masahiro Kanai Broad Institute, Cambridge, MA, United States

### Clinical Endpoint Development

Hannele Laivuori Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Aki Havulinna Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Susanna Lemmelä Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Tuomo Kiiskinen Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

### Trajectory Team

Tarja Laitinen Pirkanmaa Hospital District/University of Tampere, Tampere, Finland
Harri Siirtola University of Tampere, Tampere, Finland
Javier Gracia Tabuenca University of Tampere, Tampere, Finland

# Supplementary Acknowledgments

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i–Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of

excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.

Data on glycaemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org.

Data on coronary artery disease / myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG'. For the Exome chip study please acknowledge the source of the data as follows: 'Data on coronary artery disease / myocardial infarction have been contributed by the Myocardial Infarction Genetics and CARDIoGRAM Exome investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG'. For the Chromosome X-CAD please acknowledge the source of the data as follows: ‘Data on the X chromosomal analysis of coronary artery disease / myocardial infarction have been contributed by the referenced authors.’ For the UK Biobank meta-analysis please acknowledge the source of the data as follows: 'Data on coronary artery disease / myocardial infarction have been contributed by the CARDIoGRAMplusC4D and UK Biobank CardioMetabolic Consortium CHD working group who used the UK Biobank Resource (application number 9922). Data have been downloaded from www.CARDIOGRAMPLUSC4D.ORG.

We acknowledge the participants and researchers in the Biobank Japan Project.

# Supplementary References

96. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

97. Van Hout, C. V. *et al.* Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv* (2019) doi:10.1101/572347.

98. Weedon, M. N. *et al.* Assessing the analytical validity of SNP-chips for detecting very rare pathogenic variants: implications for direct-to-consumer genetic testing. *bioRxiv* (2019) doi:10.1101/696799.

99. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* vol. 47 1236–1241 (2015).

100. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

101. Raftery, A. E., Painter, I. S. & Volinsky, C. T. BMA: an R package for Bayesian model averaging. *The Newsletter of the R Project Volume* **5**, 2 (2005).

102. Aguirre, M., Rivas, M. & Priest, J. Phenome-wide burden of copy number variation in UK Biobank. *Genomics* (2019).

103. Vargas, J. D. *et al.* Common genetic variants and subclinical atherosclerosis: The Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis* **245**, 230–236 (2016).

104. Robertson, N., Rayner, N. W. & McCarthy, M. I. ScatterShot web application for UK Biobank, McCarthy Group Tools. https://www.well.ox.ac.uk/~wrayner/tools/.

105. Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* **361**, 2518–2528 (2009).

106. Zekavat, S. M. *et al.* Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. *Nat. Commun.* **9**, 2606 (2018).