**Analysis of large-scale human genetic datasets to identify novel risk factors and therapeutic targets for cardiometabolic disease**

by

**Connor Emdin**

Submitted in Partial Fulfillment of the Requirements for the M.D. Degree
with Honors in a Special Field at Harvard Medical School

February 10th, 20120

## Abstract

Through the analysis of large-scale human genetic datasets, I identify five therapeutic targets and risk factors for cardiometabolic disease. First, using Mendelian randomization, I demonstrate that body fat distribution is a causal risk factor for coronary artery disease and type 2 diabetes, with a similar magnitude of effect on disease risk as body mass index. Second, exploiting the genetic association between body fat distribution and type 2 diabetes, I identify a series of damaging variants in the receptor ALK7 that reduce abdominal adiposity and protect against type 2 diabetes. These findings suggest that pharmacologic ALK7 antagonism may be useful in the treatment of type 2 diabetes. Third, I show that genetic nitric oxide signaling protects against cardiovascular disease and improves renal function, suggesting that nitric oxide signaling agents such as PDE5A inhibitors could be repurposed for the treatment of cardiovascular and renal disease. Fourth, through the analysis of rare predicted loss-of-function variants in UK Biobank, I identify that deficiency in GPR151, a G-protein coupled receptor, and PDE3B, an intracellular enzyme, protects against obesity and coronary artery disease, respectively. These findings suggest that pharmacologic GPR151 inhibition may be a novel therapeutic approach to weight loss. Finally, I show that healthy lifestyle can mitigate inherited genetic predisposition to risk of cardiovascular disease, identifying a non-pharmacologic method of reducing genetic risk for coronary artery disease.

## Publications Contributing to Thesis

1. **Emdin CA***, Khera AV*, Natarajan P, Klarin D, Zekavat S, Hsiao AJ, Kathiresan S. (2017) Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes and coronary heart disease. *JAMA*. PMID: 28196256
2. **Emdin CA,** Khera AV, Aragam K, Haas M, Chaffin M, Klarin D, Natarajan P, et al. Kathiresan S. (2018) DNA Sequence Variation in *ACVR1C* Encoding the Activin-Receptor Like Kinase 7 Influences Body Fat Distribution and Protects Against Type 2 Diabetes. *Diabetes*. PMID: 30389748
3. **Emdin CA***, Khera AV*, Chaffin M, Klarin D, Natarajan P, Aragam K, Haas M, Bick A, et al., Gabriel S, Kathiresan S. Analysis of predicted loss of function variants in UK Biobank identifies variants protective for disease. *Nature Communications*. PMID: 29691411
4. **Emdin CA,** Khera AV, Klarin D, Natarajan P, Zekavat SM, Nomura A, Haas ME, Aragam K, Ardissino D, Wilson JG, Schunkert H, McPherson R, Watkins H, Elosua R, Bown MJ, Samani N, Baber U, Erdmann J, Gormley P, Palotie A, Stitziel N, Gupta N, Danesh JN, Saleheen D, Gabriel SB, Kathiresan S. (2017) Phenotypic Consequences of a Genetic Predisposition to Enhanced Nitric Oxide Signaling. *Circulation*. PMID: 28982690
5. Khera AV*, **Emdin CA***, Natarajan P, Bick AG, Cook N, Chasman DI, Baber U, Fuster V, Ridker PM, Kathiresan S. (2016) Genetic risk, adherence to a healthy lifestyle and risk of coronary artery disease. *New England Journal of Medicine*. PMID: 27959714 (Covered by NY Times).

## Summary of Contributions

For publications 1 through 4, I conceived of idea and design of the study, conducted the analysis and drafted the manuscript. For publication 5, the analysis was led by Dr. Amit Khera. I contributed as a co-first author, contributing to the design of the study and conducting the analysis of ARIC.

## Table of Contents

## 1. Introduction

More than 90% of potential therapeutics that enter clinical development are either ineffective or are too unsafe to be approved by regulatory agencies.[1] Although virtually all drugs have to be tested in preclinical models prior to human clinical trials, animal models are often poor models of human disease.[1] Consequently, the majority of drugs shown to be safe and effective in an animal model of human disease that enter clinical trials are found to be either ineffective or unsafe for use in human disease.[1]

In contrast to animal models, identification of naturally-occurring human genetic variation associated with disease can identify new therapeutic targets and pathways of clear relevance to human disease.[1] Indeed, therapeutic targets with human genetic support are more than twice as likely to result in successful development of a pharmaceutical than targets without genetic support.[2] Identification of novel protective variants for coronary artery disease and type 2 diabetes, such as variation in *ANGPTL3, ANGPTL4,* and *APOC3,* has catalyzed the development of new therapeutics targeting these genes for treatment of metabolic disorders.[3-6]

In this thesis, I analyze large-scale human genetic datasets to identify five novel risk factors and therapeutic targets for cardiometabolic disease. First, using Mendelian randomization, I demonstrate that body fat distribution is a causal risk factor for coronary artery disease and type 2 diabetes, with a similar magnitude of effect on disease risk as body mass index. Second, exploiting the genetic association between body fat distribution and type 2 diabetes, I identify a series of damaging variants in the receptor ALK7 that reduce abdominal adiposity and protect against type 2 diabetes. These findings suggest that pharmacologic ALK7 antagonism may be useful in the treatment of type 2 diabetes. Third, I show that genetic nitric

oxide signaling protects against cardiovascular disease and improves renal function, suggesting that nitric oxide signaling agents such as PDE5A inhibitors could be repurposed for the treatment of cardiovascular and renal disease. Fourth, through the analysis of individuals who carry rare predicted loss-of-function variants in UK Biobank, I identify loss-of-function variants in GPR151 and PDE5A that protect against obesity and coronary artery disease, respectively. These findings suggest that pharmacologic GPR151 inhibition may be a novel therapeutic approach to weight loss. Finally, I show that healthy lifestyle can mitigate inherited genetic predisposition to risk of cardiovascular disease, identifying a non-pharmacologic method of reducing genetic risk for coronary

## 2. Genetic association of waist-to-hip ratio with cardiometabolic disease

Obesity, typically defined on the basis of body mass index (BMI), is a leading cause of type 2 diabetes and coronary heart disease (CHD) in the population.[3,4] However, for any given BMI, body fat distribution can vary substantially; some individuals store proportionally more fat around their visceral organs (abdominal adiposity) than on their thighs and hip.[5] Waist-to-hip ratio adjusted for BMI (WHRadjBMI) is a surrogate measure of abdominal adiposity and has been correlated with direct imaging assessments of abdominal fat in observational studies.[6,7]

In observational studies, abdominal adiposity has been associated with cardiometabolic disease[8,9]; however, whether this association is causal remains unclear. For example, unmeasured lifestyle factors[10] might confound observational studies that link WHRadjBMI with type 2 diabetes and CHD. Furthermore, reverse causality could similarly lead to a statistically robust but non-causal relationship. For example, individuals with subclinical CHD might develop abdominal adiposity due to an inability to exercise.

Mendelian randomization is a human genetics tool that leverages the random assortment of genetic variants at time of conception to facilitate causal inference.[11] Because genetic predisposition to abdominal adiposity is determined by DNA sequence variants, it is less likely to be affected by confounding or reverse causality. In this study, I use a Mendelian randomization approach to determine whether a genetic predisposition to increased WHRadjBMI is associated with cardiometabolic quantitative traits, type 2 diabetes, and CHD.

### *Methods*

*Study Design and Instruments*

Observational epidemiology studies test association of an exposure (e.g., WHRadjBMI) with an outcome (e.g., CHD). However, unobserved confounders may affect both exposure and

outcome, thus biasing the observed association (**Figure 1**). Because genetic variants are both randomly assorted in a population and assigned at conception, genetic variants are largely unassociated with confounders and can be used as instrumental variables to estimate the causal association of an exposure (WHRadjBMI) with an outcome.[11] This Mendelian randomization approach has three assumptions.[12] First, genetic variants used as an instrument must be associated with the exposure of interest (e.g., WHRadjBMI, Assumption 1 in **Figure 1**). Second, genetic variants must not be associated with confounders (Assumption 2 in **Figure 1**). Third, genetic variants must not be associated with outcome independently of the exposure (Assumption 3 in **Figure 1**). The second and third assumptions are collectively known as independence from pleiotropy. Mendelian randomization can be extended to conduct a mediation analysis, estimating the association of an exposure (WHRadjBMI) with an outcome (CHD) that occurs through a given mediator (M in **Figure 1**) and that does not occur through a mediator (U in **Figure 1**).

A Mendelian randomization study using publicly-available summary-level data from large-scale genome-wide association studies (GWAS; both cross-sectional and case control datasets) as well as individual-level data from the UK Biobank (a cross-sectional dataset) was conducted (**Figure 2**).[13-19] The primary exposure was a polygenic risk score for WHRadjBMI. A recent large-scale GWAS from the Genome-wide Investigation of Anthropometric Traits (GIANT) consortium identified 48 single nucleotide polymorphisms (SNPs; genetic variants) associated with WHRadjBMI.[15] Combining these 48 SNPs into a weighted polygenic risk score enabled quantification of the genetic predisposition to increased WHRadjBMI for each individual.

*Data sources and study participants*

Summary-level data from six GWAS consortia were used.[13-19] For waist-to-hip ratio (WHR), BMI, waist circumference, hip circumference, and WHRadjBMI, data from the GIANT consortium was used;[15,16] this study included 322 154 individuals of European descent for BMI and 210 088 individuals of European descent for WHRadjBMI, WHR, waist circumference and hip circumference. The results from five additional GWAS examining blood lipids, glycemic traits, renal function, type 2 diabetes, and CHD, and predominantly containing individuals of European descent, were also assessed.[13,14,19-22] Summary results for type 2 diabetes and CHD were derived from studies of 149 821 individuals (DIAGRAM[19]) and 184 305 individuals (CARDIOGRAM[21]), respectively. Informed consent was obtained from all participants of contributing studies. Contributing studies received ethical approval from their respective institutional review boards.

Individual-level data from 111 986 individuals of European ancestry from the UK Biobank, collected from 2007-2011, was also used (**Table 1**). UK Biobank received ethical approval from the Research Ethics Committee (reference number 11/NW/0382). Analysis of UK Biobank was approved by the Partners Health Care Institutional Review Board (protocol 2013P001840). Informed consent was obtained from all participants by UK Biobank. WHRadjBMI was derived in UK Biobank through inverse normal transformation of WHR after adjustment for age, sex and BMI (as in the GIANT collaboration[15]). Type 2 diabetes and CHD were both ascertained at baseline by self-report, followed by a verbal interview with a trained nurse to confirm the diagnosis. Type 2 diabetes was defined as report of type 2 diabetes, report of type 2 diabetes unspecified, or current use of insulin medication. CHD was defined as report

of previous myocardial infarction or diagnosis of angina or hospitalization for myocardial infarction (ICD codes I21-I23).

In addition to the primary outcomes of type 2 diabetes and CHD, a phenome-wide association study (an analysis of the association of a genetic variant or polygenic risk score with a broad range of diseases and/or outcomes) for 35 additional diseases, including endocrine, renal, urological, gastrointestinal, neurological, musculoskeletal, respiratory and cancer disorders was conducted in UK Biobank to attempt to identify whether the polygenic risk score for WHRadjBMI is associated with any additional disorders.

*Statistical analysis*

For analyses of both summary-level data and UK Biobank, a weighted polygenic risk score was derived based on the magnitude of association each SNP with WHRadjBMI in the previously published GIANT analysis.[20] The association of polygenic risk score with each continuous trait and dichotomous outcome was then calculated after standardization to a one standard deviation (SD) predicted change in WHRadjBMI.

For the summary level data, this approach is equivalent to an inverse-variance weighted fixed effects meta-analysis of the association of each SNP with the trait or outcome of interest (e.g. CHD), divided by the association of each SNP with WHRadjBMI.[23,28] Explicitly, if x is the association of each SNP with the outcome of interest, and w the association of each SNP with WHRadjBMI, then the estimated genetic association of WHRadjBMI with the outcome was calculated as a fixed effects meta-analysis of x/w for all SNPs.

To validate that the polygenic risk score for WHRadjBMI was a strong instrument for WHRadjBMI (Assumption 1 in **Figure 1**), an F-statistic for the instrument was calculated in UK Biobank. An F-statistic is a measure of the significance of an instrument (the polygenic risk score) for prediction of the exposure (WHRadjBMI), controlling for additional covariates (age,

sex, ten principal components of ancestry and a dummy variable for the array type used in genotyping). An F-statistic greater than 10 is evidence of a strong instrument.[24]

For individual level data from UK Biobank, logistic regression was used to determine association of a polygenic risk score for WHRadjBMI and dichotomous outcomes (type 2 diabetes, CHD and 35 additional diseases). [25] Linear regression was used for continuous traits in UK Biobank (anthropometric traits and blood pressure). All UK Biobank analyses included adjustment for age, sex, ten principal components of ancestry and a dummy variable for the array type used in genotyping. The inclusion of principal components of ancestry as covariates is commonly implemented to correct for population stratification according to ancestral background.[26]

In order to test Assumption 2 (independence of polygenic risk score for WHRadjBMI from potential confounders, Figure 1), the relationship of the polygenic risk score to smoking, alcohol use, physical activity, vegetable consumption, red meat consumption and breastfeeding status as a child was determined among individuals in the UK Biobank. Test for trend was performed across quartiles of the polygenic risk score for WHRadjBMI using logistic regression with each potential confounder as the outcome. For comparison, individuals in UK Biobank were stratified into quartiles by observational WHRadjBMI and test for trend performed using logistic regression.

I conducted five sensitivity analyses to test the robustness of our results. Three additional polygenic risk scores were used, including one that included variants which were not significantly associated with BMI, a second that included variants which were significantly associated with gene expression in adipose tissue and a third that included variants which were

significantly associated with increased WHRadjBMI in women but not in men. The association of genetic variants with BMI was adjusted for and median regression was used.[12]

The threshold of statistical significance for type 2 diabetes and CHD (main outcome measures) was $p< 0.025$ (0.05/2=0.025). The threshold of significance for the analysis of fifteen traits was $p<0.0033$ (0.05/15=0.0033). The threshold of significance in the phenome-wide association analysis was $p<0.0014$ (0.05/35 = 0.0014).

Analyses were performed using R version 3.2.3 software (The R Project for Statistical Computing, Vienna, Austria) and Stata version 12 (StataCorp, Texas, United States).

### *Results*

The characteristics of UK Biobank participants are provided in **Table 1**. The mean age was 56.9 (SD 7.9), the mean systolic blood pressure 143.6 mm Hg (SD 21.8) and the mean diastolic blood pressure 84.5 mm Hg (SD 11.8). 5639 (5.0%) participants had CHD while 5690 (5.1%) individuals had type 2 diabetes. A 48-SNP polygenic risk score for WHRadjBMI score was a strong instrumental variable (F-statistic of 1713), statistically accounting for 1.5% of variance in WHRadjBMI in UK Biobank, thus validating Assumption 1 in **Figure 1**.

In order to test Assumption 2 (independence of polygenic risk score for WHRadjBMI from potential confounders, **Figure 1**), the relationship of the polygenic risk score to smoking, alcohol use, physical activity, vegetable consumption, red meat consumption and breastfeeding status as a child was determined among individuals in the UK Biobank. In each case, no significant relationship was noted (**Supp. Table 1**). For comparison, a similar analysis that categorized individuals according to observed WHRadjBMI (instead of genetic predisposition to WHRadjBMI) was conducted. In this observational epidemiology analysis, WHRadjBMI was strongly associated with each potential confounder (**Supp. Table 2**).

A one SD increase in WHRadjBMI due to the polygenic risk score was associated with a 1.0 kg/m$^2$ decrease in BMI (CI 0.9, 1.1), a 2 cm increase in waist circumference (CI 1.5, 2.4), a 4.1 cm decrease in hip circumference (CI 3.8, 4.4) and a 0.068 increase in WHR (CI 0.066, 0.07; Figure 3). A one SD increase in WHRadjBMI due to the polygenic risk score was associated with higher total cholesterol (5.6 CI 3.9, 7.3 mg/dl), higher LDL cholesterol (5.7 CI 4.1, 7.2 mg/dl), higher triglycerides (27 CI 25, 30 mg/dl) and lower HDL cholesterol (6.0 CI 5.3, 6.6 mg/dl). A one SD increase in WHRadjBMI due to the polygenic risk score was associated with higher log-transformed fasting insulin levels (0.065 CI 0.052, 0.078), higher two-hour glucose levels (4.1 CI 1.6, 6.5 mg/dl), and higher systolic blood pressure (2.1 CI 1.2, 3 mmHg).

A one SD increase in WHRadjBMI due to polygenic risk score was associated with a higher risk of type 2 diabetes (OR 1.77 CI 1.57, 2.00; absolute risk increase per 1000 participant years (ARI) 6.0 CI 4.4, 7.8; $p=7.3*10^{-21}$; number of participants with type 2 diabetes outcome, 40 530; **Figure 4**). A one SD increase in WHRadjBMI due to polygenic risk score was also associated with higher risk of CHD (OR 1.46, CI 1.32, 1.62; ARI 1.8 CI 1.3, 2.4; $p=9.9*10^{-14}$; number of participants with CHD outcome, 66 440; **Figure 4**).

Five sensitivity analyses of the genetic association of WHRadjBMI with cardiometabolic traits, type 2 diabetes and CHD were conducted to examine if results were influenced by pleiotropy (i.e., a violation of Assumptions 2 or 3 in **Figure 1**). Similar associations were observed with a score composed of variants not associated with BMI, a score composed of variants expressed with adipose tissue and using median regression. In the fifth sensitivity analysis, eight SNPs associated with elevated WHRadjBMI in women but not men were combined in an additive risk score. If WHRadjBMI causes CHD (rather than results being due to pleiotropy), then a risk score that increases WHRadjBMI in women but not in men should

increase risk of CHD in women but not in men. A numerically greater magnitude of association with type 2 diabetes and CHD was noted in women as compared to men (**Supp. Figure 1**), consistent with a causal effect of WHRadjBMI on type 2 diabetes and CHD.

Using the polygenic risk score of 48 SNPs associated with WHRadjBMI, a phenome wide association study of 35 additional diseases in UK Biobank was conducted (**Figure 5**). There was no significant association of WHRadjBMI with any of these diseases at the Bonferroni-adjusted level of significance ($p<0.0014$).

### *Discussion*

Using a Mendelian randomization analysis, I tested if human genetic evidence supported a causal relationship of WHRadjBMI, a measure of abdominal adiposity, with type 2 diabetes and CHD. Genetic predisposition to higher WHRadjBMI was associated with increased levels of quantitative risk factors (lipids, insulin, glucose, and systolic blood pressure) as well as a higher risk for type 2 diabetes (OR 1.77 CI 1.57, 2.00 per SD higher WHRadjBMI) and CHD (OR 1.46 CI 1.32, 1.62 per SD higher WHRadjBMI).

These results permit several conclusions. First, these findings lend human genetic support to previous observations associating abdominal adiposity with cardiometabolic disease.[8,9] In the INTERHEART acute myocardial infarction case-control study, a one SD higher WHR was associated with elevated odds of myocardial infarction (OR 1.37 CI 1.33, 1.40), after adjustment for BMI and other covariates.[8] However, residual confounding or reverse causality may have contributed to these associations. Indeed, in this study, observational WHRadjBMI was strongly associated with potential confounders, illustrating a limitation of observational epidemiology. Here, these prior findings were extended by testing a polygenic risk score that appeared independent of measured confounders.

Second, WHRadjBMI might prove useful as a biomarker for the development of therapies to prevent type 2 diabetes and CHD. Although a substantial focus of drug development has been towards therapeutics to reduce overall adiposity[27], there has been little effort towards the development of therapies that modify body fat distribution to reduce abdominal adiposity. Ongoing research to understand the mechanistic links between the numerous genetic loci that influence WHRadjBMI may lead to novel therapeutics strategies to reduce abdominal adiposity and reduce the risk of type 2 diabetes and CHD. In the next analysis, one such locus is identified.

## 3. DNA sequence variation in *ACVR1C* encoding the Activin-Receptor Like Kinase 7 influences body fat distribution and protects against type 2 diabetes

Body fat distribution strongly influences the development of type 2 diabetes.[8,28,29] In the above Mendelian randomization study of 296 291 individuals, I found that a genetic predisposition to increased abdominal fat distribution was associated with elevated triglyceride levels, elevated blood pressure and an increased risk of coronary artery disease, independent of overall adiposity.[28] Furthermore, a genetic predisposition to increased abdominal fat distribution was strongly associated with the development of type 2 diabetes. For each one standard deviation genetic increase in waist-to-hip ratio adjusted for body mass index (WHRadjBMI, a measure of body fat distribution), risk of type 2 diabetes increased by 77%.[28] These findings were replicated in a separate Mendelian randomization study.[29]

These results suggest the hypothesis that genetic variants that influence body fat distribution may also influence the risk of type 2 diabetes.[28] Here, I test this hypothesis by analyzing genetic variation in more than 400000 individuals in UK Biobank to identify novel genetic variants that lower WHRadjBMI and protect against type 2 diabetes. Below, I demonstrate that variants predicted to lead to loss of function of the gene *ACVR1C*, encoding the activin-receptor like kinase 7, influence body fat distribution and protect against type 2 diabetes.

### *Methods*

*Study Design*

In my discovery analysis, I analyzed the association of 614042 coding variants with WHRadjBMI in 405569 individuals in UK Biobank. Coding variants were defined as missense variants or variants predicted to result in loss of function of the protein: (1) nonsense mutations that resulted in early termination of a protein; (2) frameshift mutations due to insertions or

deletions of DNA; or (3) splice-site mutations which result in an incorrectly spliced protein. Only variants imputed with a quality score (info score) > 0.3 were analysed. Threshold for significance was defined as $p<5*10^{-8}$ (genome wide significance) and analysis was performed using PLINK2 software.[30] I reported novel variants as those located more than 1 megabase away from prior identified loci in the GIANT consortium.[15] I attempted to replicate the association of novel variants with WHRadjBMI using data from the GIANT consortium, when the variant was available in the GIANT consortium.[15]

Upon identification of variants in *ACVR1C* as significantly associated with WHRadjBMI, a conditional analysis was conducted to identify additional variants significantly associated ($p<0.0001$) with WHRadjBMI within the *ACVR1C* locus ($\pm$ 250 kb of the lead variant rs55920843). To replicate observed associations of *ACVR1C* variants with WHRadjBMI,Iexamined whether carriers of these variants had reduced WHRadjBMI in a meta-analysis of the GIANT consortium and two independent cohorts (Atherosclerosis Risk in Communities and Framingham Heart studies). I also examined whether these variants were associated with direct imaging-based measurements of abdominal fat in 4215 participants who underwent dual X-ray absorptiometry imaging in UK Biobank.

To test whether identified *ACVR1C* variants were also associated with risk of type 2 diabetes, I pooled data from the DIAGRAM Consortium (ExTexT2D exome chip analysis[31]) with UK Biobank. To test whether variation leading to loss of *ACVR1C* function protects against type 2 diabetes,Ianalyzed the sequences of the 9 exons of *ACVR1C* in 55516 participants [31672 from the Myocardial Infarction Genetics Consortium (MIGen)[32,33], 5388 from the Atherosclerosis Risk in Communities study and 18456 from the T2D Genes Consortium[34]] and examined if predicted damaging variants in the gene associate with risk of type 2 diabetes.

A phenome-wide association study of *ACVR1C* in UK Biobank was performed using an *ACVR1C* gene risk score.[35] Three metabolic traits available in UK Biobank (urinary albumin-to-creatinine ratio, systolic blood pressure and diastolic blood pressure) were also analyzed. A p-value of 0.001 (0.05/34) was used for significance in this analysis.

*Data Sources*

For the analysis of WHRadjBMI, individual-level data from 405569 unrelated individuals from the UK Biobank (335,660 individuals of European ancestry and 69,909 individuals of Non-European ancestry) was analyzed. UK Biobank received ethical approval from the Research Ethics Committee (reference number 11/NW/0382). Analysis of UK Biobank was approved by the Partners Health Care Institutional Review Board (protocol 2013P001840). Informed consent was obtained from all participants by UK Biobank. For replication, data for WHRadjBMI from the GIANT consortium (in which the Ile482Val variant was available) was pooled with data from the Atherosclerosis Risk in Communities and Framingham Heart Study datasets (in which Asn150His, Ile195Thr and rs72927479 variants were available). The GIANT consortium consisted of 224459 participants (210088 of European ancestry and 14371 of non-European ancestry) genotyped using the MetaboChip.[15] Atherosclerosis Risk in Communities study is a community based study of 15792 white and black participants, aged 45 to 64 years.[36] The Atherosclerosis Risk in Communities dataset consisted of 10122 individuals (8015 of European ancestry and 2107 of Non-European ancestry) who were genotyped and imputed, as previously described.[37] For 5388 participants, exome sequences were also available for analysis. In the Framingham Heart Study, a community based study of 10092 individuals of predominantly European ancestry, genotyped data was available from 6073 individuals of European ancestry.

For the analysis of type 2 diabetes, estimates from UK Biobank were pooled using inverse variance weighted fixed effects meta-analysis with estimates from the DIAGRAM ExTexT2D exome chip analysis of 452244 participants (81412 diabetes cases and 370832 controls).[31] In UK Biobank, type 2 diabetes was defined as (1) self-report of type 2 diabetes, followed by a verbal interview with a trained nurse to confirm the diagnosis or (2) hospitalization for ICD code E11. Because the ExTexT2D analysis included 120286 participants from UK Biobank[31], these individuals were excluded from the analysis of type 2 diabetes in UK Biobank to prevent analysis of overlapping samples.

Sequence data for *ACVR1C* were extracted from exome sequencing performed in the MIGen Consortium as previously described.[32,33] The Burrows–Wheeler Aligner algorithm was used to align reads from participants to the reference genome (hg19). The GATK HaploTypeCaller was used to jointly call variants. Metrics including Variant Quality Score Recalibration (VQSR), quality over depth, and strand bias were then used to filter variants. The Jackson Heart Study was excluded from analysis of MIGen as it was included in the T2D Genes consortium. Exome sequences from 5388 participants in ARIC were analysed as previously described.[38] Phenotype and genotype data were retrieved from the National Center for Biotechnology Information dbGAP server (accession: phs000090.v3.p1 and phs000572.v6.p4). Exome sequencing was performed in the T2D Genes consortium as previously described.[34] To analyse exome sequences from the T2D Genes consortium, the online Genetic Association Interactive Test in the T2D Knowledge portal was used.[34]

*Statistical Analysis*

In UK Biobank, WHRadjBMI was derived through inverse normal transformation of waist-to-hip ratio after adjustment for age, sex and BMI (as in the GIANT collaboration[15]). In UK Biobank, linear regression was used to estimate the association of variants with WHRadjBMI. All UK Biobank analyses included adjustment for age, sex, 10 principal components of ancestry, and a dummy variable for the array type used in genotyping. Logistic regression was used to estimate the association of variants with type 2 diabetes. Estimates of the association of each variant with type 2 diabetes in UK Biobank were pooled with estimates from the ExTexT2D consortium using inverse-variant weighted fixed effects meta-analysis.

To estimate the overall association of variation in *ACVR1C* with WHRadjBMI, I pooled across all variants using a gene risk score, weighted by the square root of allele frequency (estimating a weighted mean effect of *ACVR1C* variants on WHRadjBMI).[39] To estimate the overall association of variation in *ACVR1C* with type 2 diabetes, I pooled across all variants in a gene risk score, weighted by the association of each variant with WHRadjBMI.[40]

For analysis of exome sequencing data, logistic regression was performed with adjustment for sex, five principal components of ancestry and a dummy variable for each cohort to estimate the association of predicted damaging variants with type 2 diabetes. Estimates from the MIGen consortium were pooled with estimates from the T2D Genes consortium using inverse-variance weighted fixed effects meta-analysis.

For the phenome-wide association study, all four *ACVR1C* variants were pooled in a gene-risk score in UK Biobank, as previously described.[38,41] For each individual in UK Biobank, the *ACVR1C* variants associated with lower WHRadjBMI were weighted by their effect on WHRadjBMI and summed. The association of this gene-risk score with 31 different diseases in UK Biobank and three metabolic traits was tested using logistic regression with adjustment for

age, sex, 10 principal components of ancestry, and a dummy variable for the array type used in genotyping.

Analyses were performed using R version 3.2.3 (R Project for Statistical Computing).

## *Results*

*Exome wide association study of body fat distribution in UK Biobank*

Among 405569 participants in UK Biobank, 54% were female, the median age was 57 and the median waist-to-hip ratio, measured at enrollment, was 0.87 (**Table 2**). One standard deviation in waist-to-hip ratio corresponded to an absolute change of 0.09. In an analysis of 614012 coding variants in UK Biobank, no evidence of genomic inflation was observed (lambda 1.08).

The lead novel low frequency variant associated with lower WHRadjBMI lay within the gene *ACVR1C*. *ACVR1C* Asn150His (allele frequency 1.1%), associated with 0.09 standard deviations (SD) lower WHRadjBMI ($p=3.4*10^{-17}$). An independent missense variant in ACVR1C, Ile195Thr (AF 0.4%), also associated with lower WHRadjBMI (0.15 SD, $p=1.0*10^{-9}$).

Upon conditioning on these two variants, I identified an additional coding variant: Ile482Val (AF 7%), which associated with 0.019 SD lower WHRadjBMI ($p=1.6*10^{-5}$) and rs72927479, a non-coding variant, for which the minor G allele (AF 5%) was associated with 0.035 SD lower WHRadjBMI ($p=2.6*10^{-12}$, **Table 3**). Despite being an independent signal for WHRadjBMI, the non-coding variant rs72927479 is nominally correlated with Ile482Val ($r^2=0.06$ in UK Biobank). No other variants were correlated with one another in UK Biobank (all $r^2<0.001$).

Pooling across all four variants with weighting by square root of allele frequency, *ACVR1C* variation was associated with lower WHRadjBMI in UK Biobank (-0.07 SD, $p=2.6*10^{-}$

[35]). To replicate this finding, I pooled data from the GIANT consortium (in which the Ile482Val variant was available, n=224156) with data from the Atherosclerosis Risk in Communities and Framingham Heart Study datasets (in which Asn150His, Ile195Thr and rs72927479 variants were available, n=13704). Pooling across all four variants in these three replication studies, variation in *ACVR1C* was associated with reduced WHRadjBMI (-0.07 SD, p=0.0005).

Examining other anthropometric traits in UK Biobank, variation in *ACVR1C* was associated with elevated hip circumference (0.035 SD, $p=3*10^{-5}$), nominally elevated BMI (0.02 SD, p=0.002) and was unassociated with waist circumference (-0.007 SD, p=0.20) or height (0.005 SD, p=0.23). In the MAGIC consortium analysis of insulin resistance (HOMA-IR), Ile482Val (the only *ACVR1C* variant available for analysis), was unassociated with HOMA-IR (-1.1%, p=0.17).[42]

I examined whether *ACVR1C* variation also associated with direct imaging measurement of abdominal obesity. 4215 participants in UK Biobank underwent dual energy X-ray absorptiometry (DEXA) imaging to estimate abdominal fat mass. Carriers of *ACVR1C* variants had lower percent abdominal fat (p=0.008, **Figure 6**). When one outlying individual with four *ACVR1C* variants was excluded (**Figure 6**), carriers of *ACVR1C* continued to have significantly lower percent abdominal fat (p=0.009).

*Association of variants in ACVR1C with type 2 diabetes*

Genetic predisposition to increased WHRadjBMI strongly predisposes to type 2 diabetes.[28] I therefore examined whether variants in *ACVR1C* that lower WHRadjBMI adiposity protect against type 2 diabetes. In a combined analysis of UK Biobank and the DIAGRAM consortium, all four *ACVR1C* variants were found to independently protect against type 2 diabetes (OR 0.88, $p=8.7*10^{-5}$ for Asn150His; OR 0.79, p=0.005 for Ile195Thr; OR 0.95,

p=4.8*10$^{-6}$ for Ile482Val; OR 0.93, p=0.0006 for rs72927479; **Table 3**). Pooling across all four variants, a 0.2 SD decrease in WHRadjBMI through *ACVR1C* was associated with a 30% lower risk of type 2 diabetes (OR 0.70 CI 0.63, 0.77; p = 5.6*10$^{-13}$, **Figure 7**). When I excluded the non-coding variant rs72927479 which is nominally correlated with Ile482Val ($r^2$ = 0.06 in UK Biobank), the *ACVR1C* gene risk score remained associated with risk of type 2 diabetes (OR 0.71 CI 0.64, 0.79, p=1.8*10$^{-10}$)

In three independent datasets with exome sequence data – the MIGen Consortium, ARIC and the T2D Genes Consortium, I examined whether variants that lead to loss of *ACVR1C* gene function protect against type 2 diabetes. The 9 exons of the *ACVR1C* gene were sequenced in 55516 individuals. 105 predicted damaging variants were identified. Among 16452 type 2 diabetes cases, the frequency of predicted damaging variants in *ACVR1C* was 0.1% (17) compared to 0.2% (88) among 39064 diabetes-free controls. Overall, carrying a predicted damaging variant in *ACVR1C* was associated with 54% lower risk of type 2 diabetes (OR 0.46 CI 0.27, 0.81; p=0.006; Figure 8). When I excluded 47 carriers of I195T (annotated as a predicted damaging variant by five of five algorithms) in the exome sequencing analysis, carriers of predicted damaging variants in *ACVR1C* remained protected from type 2 diabetes (OR 0.48 CI 0.24, 0.97, p=0.04).

To further examine whether loss of *ACVR1C* function lowers WHRadjBMI, I examined whether the non-coding variant rs72927479 associates with *ACVR1C* expression. The minor allele of rs72927479 (G, frequency 5%) associated with lower WHRadjBMI (beta -0.035, p=2.6*10$^{-12}$) and type 2 diabetes (OR 0.93, p=0.0006). In the Genotype-Tissue Expression dataset[43], the minor allele of rs72927479 associated with reduced expression of *ACVR1C* in subcutaneous adipose tissue (p=0.02) and pancreas (p=0.02).

*Phenome-wide association study of ACVR1C in UK Biobank*

To anticipate whether *ACVR1C* inhibition may be associated with on-target adverse effects, I conducted a phenome-wide association study of 31 disease phenotypes in UK Biobank. I did not observe any significant associations between the *ACVR1C* gene risk score and the 31 diseases analyzed (Figure 9). The *ACVR1C* gene risk score was unassociated with coronary artery disease (OR 1.01, p=0.86). I also examined whether *ACVR1C* variation associates with three metabolic traits currently available in UK Biobank: urinary albumin, systolic blood pressure and diastolic blood pressure. While the *ACVR1C* gene risk score did not significantly associate with urinary albumin levels, it associated with significantly lower diastolic blood pressure (-0.6 mm Hg, p=0.0004) and nominally lower systolic blood pressure (-0.6 mm Hg, p=0.03).

## *Discussion*

In this study, four genetic variants in *ACVR1C,* ranging in frequency from 0.2% to 7.2%, independently protected against abdominal obesity and type 2 diabetes. Furthermore, damaging variants in *ACVR1C* protected against type 2 diabetes in an analysis of exome sequences from 55516 individuals. An *ACVR1C* gene risk score did not associate with any of 31 additional diseases in UK Biobank but did nominally associate with lower blood pressure.

These results permit several conclusions. First, pharmaceutical inhibition of *ACVR1C* may be useful in the treatment of type 2 diabetes. *ACVR1C* encodes the activin-receptor like kinase 7 (ALK7), a transforming growth factor-beta family receptor highly expressed on pancreatic islet cells[44,45] and adipocytes[46]. Overexpression of AKL7 induces growth inhibition and apoptosis of pancreatic beta cells[45,47], suggesting that it is a negative regulator of beta cell mass. *ACVR1C* deficient mice have been reported to have reduced body fat when fed a high-fat-diet[46,48]

and have improved glucose tolerance and insulin sensitivity when obese.[46] Chemical inhibition of *ACVR1C* has also been shown to reduce fat accumulation and increase lipolysis in mice.[49] In combination with the human genetic results presented here, these findings suggest that inhibition of *ACVR1C* may be a novel therapeutic mechanism for the treatment of abdominal obesity and type 2 diabetes.

Second, the lack of association of the *ACVR1C* gene risk score with 31 different diseases in UK Biobank suggests that therapeutic *ACVR1C* inhibition may not have adverse on-target effects. However, due to the small number of cases for many of the analyzed diseases (e.g. 1707 for cervical cancer), modest associations of *ACVR1C* variation with the analyzed phenotypes cannot excluded. In the Exome Aggregation Consortium, *ACVR1C* is tolerant of loss of function variants, with 18 of 66720 individuals of European ancestry carrying an early stop codon (Leu32Ter) in *ACVR1C*. [50] In combination with the phenome-wide association study presented here, these findings suggest that *ACVR1C* could be safely inhibited. The association of the *ACVR1C* gene risk score with nominally lower diastolic and systolic blood pressure suggests that *ACVR1C* inhibition may have the additional benefit of lowering blood pressure. However, this finding, which did not reach genome-wide significance, requires replication in independent datasets.

In summary, variants leading to loss of *ACVR1C* gene function reduce abdominal adiposity and protect against type 2 diabetes. These findings identify *ACVR1C* as a therapeutic target for type 2 diabetes.

## 4. Phenotypic consequences of a genetic predisposition to enhanced nitric oxide signaling

Nitric oxide signaling is a key regulator of vascular tone, blood pressure and platelet aggregation.[51,52] Endothelial nitric oxide synthase (eNOS), encoded by the gene *NOS3*, generates nitric oxide in the vascular endothelium (**Figure 10**).[53] Nitric oxide acts as a signaling molecule to activate soluble guanylyl cyclase (sGC), a heterodimeric protein with one subunit encoded by the gene *GUCY1A3*.[53,54] Cyclic guanosine monophosphate (cGMP) produced by sGC then activates downstream signaling molecules, leading to vasodilation, blood pressure lowering, inhibition of platelet aggregation and other cardiometabolic effects.[55,56] *NOS3* and *GUCY1A3* are thus key mediators of nitric oxide signaling and its downstream effects (Figure 10).

Previous studies have noted increased blood pressure and atherosclerotic burden in $eNOS^{-/-}$ knockout mice,[57,58] while loss of sGC promotes thrombus formation.[59] Common noncoding variants in the *NOS3* and *GUCY1A3* loci associate with blood pressure and coronary heart disease (CHD) in genome wide association studies, consistent with a role of nitric oxide signaling in regulation of arterial blood pressure.[21,60,61] Furthermore, a rare coding variant in *GUCY1A3* was identified as associating with early-onset myocardial infarction in a German family.[59] These findings in mice and humans suggest that stimulation of nitric oxide signaling may be a useful therapeutic strategy for prevention of cardiovascular diseases.

I sought to use common and rare DNA sequence variants in *NOS3* and *GUCY1A3* to: (1) determine the effects of a genetic predisposition to enhanced nitric oxide signaling on a range of cardiometabolic and other diseases; (2) assess whether the effect of nitric oxide signaling on CHD is primarily mediated through blood pressure; and (3) determine if rare inactivating

variants in *NOS3* and *GUCY1A3* that are predicted to reduce nitric oxide signaling associate with higher risk for CHD.

## *Methods*

*Study design, data sources and study participants*

I used individual-level data from 335,464 individuals of European ancestry from the UK Biobank, a large population-based cohort. I supplemented this individual-level data with seven genome-wide association study (GWAS) consortia examining blood lipids, anthropometric traits, glycaemic traits, diabetes, CHD, migraine and renal dysfunction, all predominantly containing individuals of European descent. Finally, I used gene sequence data from 27,815 participants from the Myocardial Infarction Genetics Consortium and 16,857 participants from the T2D GENES study[34] to examine whether rare variants in the *NOS3-GUCY1A3* pathway associate with blood pressure and CHD.

In my primary analysis, I examined the effect of a genetic predisposition to enhanced nitric oxide signaling on nine different cardiometabolic diseases: CHD, stroke, heart failure, atrial fibrillation, aortic stenosis, peripheral vascular disease, venous thromboembolism, diabetes and chronic kidney disease. I additionally examined the effect of a genetic predisposition to enhanced nitric oxide signaling on 16 quantitative traits: systolic blood pressure, diastolic blood pressure, waist-to-hip ratio adjusted for body mass index[35], body mass index, total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, fasting glucose, fasting insulin, two-hour glucose, hemoglobin A1C, serum creatinine-estimated glomerular filtration rate (GFR), cystatin-C-estimated GFR, forced expiratory volume in 1 second (FEV1) and the ratio of forced expiratory volume in 1 second to forced vital capacity (FEV1/FVC). All traits were standardized (that is, reported in standard

deviation (SD) units) to facilitate comparisons among traits. Using the UK Biobank cohort, we conducted a phenome-wide association study for 26 additional diseases, including endocrine, renal, urological, gastrointestinal, neurological, musculoskeletal, respiratory and neoplastic disorders.

Analysis of the UK Biobank data was approved by the Partners Health Care institutional review board (protocol 2013P001840; application 7089). Informed consent was obtained from all participants by the UK Biobank.

*Common variants in NOS3 and GUCY1A3*

We leveraged common variants in the *NOS3* and *GUCY1A3* loci to characterize the effects of a genetic predisposition to enhanced nitric oxide signaling. Two common variants were selected as instruments for a genetic predisposition to enhanced nitric oxide signaling: a promoter variant of *NOS3* (rs3918226; minor allele frequency 8%) and an intronic variant of *GUCY1A3* (rs7692387; minor allele frequency 19%). These variants were selected because: (1) they have been robustly associated with blood pressure (a downstream effect of nitric oxide signaling) and (2) they are located within the endothelial nitrate synthase-soluble guanylyl cyclase nitric oxide signaling pathway.[62,63] Furthermore, the minor allele of rs7692387 has also recently been characterized to reduce *GUCY1A3* expression via disruption of a ZEB1 transcription factor site.[64] I examined the effect of these variants on *NOS3* and *GUCY1A3* expression levels in aortic and lung tissue in the Genotype-Tissue Expression project database[43] and the effect of these variants on mean arterial pressure in UK Biobank.

To examine the effects of increased nitric oxide signaling on cardiometabolic and other traits, I pooled rs3918226 and rs7692387 into an additive genetic score (individuals had 0 to 4 risk alleles). I derived this score by multiplying the number of risk alleles by the association of

each allele with mean arterial pressure in UK Biobank (0.68 mm Hg for rs3918226 and 0.32 mm Hg for rs7692387). For example, if an individual had two risk alleles for rs3918226 and one risk allele for rs7692387, the genetic score was calculated as 2*0.68 mm Hg + 0.32 mm Hg = 1.68 mm Hg. Individuals missing one variant [either rs3918226 or rs7692387, n= 6036 (1.8%)] were imputed at the mean allele frequency for the variant prior to calculation of the genetic score. I standardized the nitric oxide signaling genetic score to a 5 mm Hg reduction in mean arterial pressure, corresponding to the effect of 1.5 mg of riociguat (a pharmacologic stimulator of soluble guanylate cyclase) on mean arterial pressure in a recent randomized controlled trial.[65] The standardization was performed by dividing the genetic nitric oxide score by 5 mm Hg (e.g. 1.68 mm Hg/5 mm Hg = 0.34). We also report estimates standardized to a 2.5 mm Hg and 10 mm Hg lower mean arterial pressure to clarify the expected reduction in risk of cardiometabolic outcomes with different levels of nitric oxide signaling.

*Rare predicted loss-of-function variants in NOS3 and GUCY1A3*

Common non-coding variants may influence the expression of several nearby genes.[66] Therefore, to provide complementary evidence that the *NOS3-GUCY1A3* pathway influences blood pressure and CHD risk, we examined whether rare (minor allele frequency<1%) predicted loss-of-function variants in *NOS3* and *GUCY1A3* are associated with systolic blood pressure, diastolic blood pressure and CHD. Predicted loss-of-function variants were defined as: (1) insertions or deletions of DNA that modify the reading frame of protein translation (frameshift); (2) point mutations at conserved splice site regions which alter the splicing process (splice-site); or (3) point mutations that change an amino acid codon to a stop codon, leading to truncation of a protein (nonsense).

For blood pressure, I tested whether presence of a predicted loss-of-function variant in

*NOS3* and *GUCY1A3* was associated systolic blood pressure or diastolic blood pressure in the Type 2 Diabetes Genetics Exome Sequencing study (n=16,857). I examined whether predicted loss-of-function variants were associated with systolic and diastolic blood pressure using the Genetic Association Interactive Tool on the Type 2 Diabetes Knowledge Portal.[34] I used linear regression, adjusted for age, sex and five principal components of ancestry.

For CHD, I tested whether presence of a predicted loss-of-function variant was associated with CHD in the Myocardial Infarction Genetics Consortium (n = 27,815) study using logistic regression, adjusted for sex, five principal components of ancestry and a dummy variable for each cohort. To examine whether variants in the *NOS3-GUCY1A3* pathway associate with CHD risk, I pooled the effect of predicted loss-of-function variants in *NOS3* and *GUCY1A3* on CHD using inverse variance weighted fixed effects meta-analysis.

*Statistical analysis*

For UK Biobank, I estimated the association of the nitric oxide signaling genetic score (standardized to a 5 mm Hg decrease in mean arterial pressure) with each outcome using a logistic regression model adjusting for age, sex, ten principal components of ancestry and a dummy variable for array type. For the summary-level data, this approach is equivalent to an inverse variance weighted fixed effects meta-analysis of the effect of each variant on traits or outcome of interest per 5 mm Hg lower mean arterial pressure. Tests for interaction between UK Biobank and summary-level estimates were calculated as the difference in log-transformed relative risks, as previously described.[67]

For our primary outcomes (nine cardiometabolic diseases), I set a Bonferroni adjusted level of significance of p=0.05/9=0.0056. For the secondary analyses of 16 cardiometabolic and pulmonary traits and the phenome wide association study of 26 phenotypes, I set a level of significance of p=0.05/42=0.001.

To examine whether an observed reduction in risk of CHD was caused by reduced blood pressure, a mediation analysis was conducted. An estimate of the causal effect of systolic blood pressure on CHD risk was derived from a recent genome wide association study (OR 1.21 CI 1.17, 1.24 per 5 mm Hg higher systolic blood pressure).[61] This effect was then multiplied by the decrease in systolic blood pressure due to nitric oxide signaling to estimate the decrease in CHD risk mediated by systolic blood pressure. I then subtracted this estimate from the overall estimate of the nitric oxide genetic score with CHD to derive the remaining proportion of CHD risk unaccounted for by a decrease in systolic blood pressure. For example, if the OR for CHD for the genetic score was 0.5, but the OR for CHD from the systolic blood pressure decrease was 0.75, the OR for CHD independent of systolic blood pressure was calculated as exp(log(0.5) – log(0.75))=0.67.

All analyses were performed using R version 3.2.3 software (The R Project for Statistical Computing, Vienna, Austria).

### *Results*

In the Genome Tissue Expression project database[43], the C allele of rs3918226 was associated with increased *NOS3* expression in lung tissue (48% higher expression, p=0.002) but was not significantly associated with *NOS3* expression in aortic tissue (23% higher expression, p=0.21, **Figure 11**). The A allele of rs7692387 was associated with increased *GUCY1A3* expression in aortic tissue (20% higher expression, p=0.012) but was not significantly associated with *GUCY1A3* expression in lungs (7% higher expression, p=0.20). As expected for variants that enhance nitric oxide signaling, both rs3918226 and rs7692387 were associated with lower mean arterial pressure among UK Biobank participants [(0.68 mm Hg lower mean arterial

pressure ($p=1.3*10^{-26}$) and 0.32 mm Hg lower mean arterial pressure ($p=8.3*10^{-14}$), respectively) (Figure 2)], replicating previously reported associations of these variants with blood pressure at genome wide significance.[62,63]

When combined into a nitric oxide signaling genetic score, and standardized to a 5 mm Hg reduction in mean arterial pressure, a genetic predisposition to enhanced nitric oxide signaling was nominally associated with improved renal function, as assessed by both cystatin-C-estimated GFR (7.6 ml/in; CI 2.9, 12 ml/min; p=0.0015) and creatinine-estimated GFR (2.8 ml/min; CI 0.57, 5.1 ml/min; p=0.014; **Figure 12**).

The genetic nitric oxide signaling score was significantly associated with three of the nine primary cardiometabolic outcomes (**Figure 12**). A genetic predisposition to enhanced nitric oxide signaling was associated with reduced risk of CHD (OR 0.36 CI 0.29, 0.46; $p= 7.0*10^{-17}$ in the CARDIoGRAMPlusC4D Consortium data. We replicated this association in UK Biobank participants (OR 0.39 CI 0.29, 0.52; $p=1.1*10^{-10}$), with an overall pooled 63% reduction in CHD risk (OR 0.37 CI 0.31, 0.45; $p=5.5*10^{-26}$; **Figure 13**).

Beyond CHD, the nitric oxide signaling genetic score was associated with a reduced risk of peripheral arterial disease (OR 0.42 CI 0.26, 0.68; p=0.0005). This association persisted in a sensitivity analysis that excluded individuals with concomitant CHD (OR 0.41 CI 0.23, 0.74, p=0.003). The genetic score was also associated with a reduced risk of stroke (OR 0.53 CI 0.37, 0.76; p=0.0006) and a nominally ($p<0.05$) reduced risk of chronic kidney disease (OR 0.46 CI 0.25, 0.83; p =0.011), heart failure (OR 0.59 CI 0.38, 0.92; p=0.02) and diabetes (OR 0.81 CI 0.67, 0.99; p=0.037). In a phenome wide association study of 26 different diseases, enhanced nitric oxide signaling was not significantly associated with any other disease, including a variety of gastrointestinal diseases, musculoskeletal diseases and cancers (**Figure 12**). In sensitivity

analyses, when standardized to a 2.5 mm Hg or 10 mm Hg lower mean arterial reduction, the nitric oxide signaling genetic score, was associated with a 39% (OR 0.61 CI 0.56, 0.67) and 86% lower risk of CHD (OR 0.14 CI 0.10, 0.20), respectively.

When an interaction term between the *NOS3* and *GUCY1A3* genetic loci was included, no evidence of an interaction in the association of nitric oxide signaling with systolic blood pressure (p interaction = 0.76) or diastolic blood pressure (p interaction = 0.49) was observed. Similarly, no evidence of an interaction between the *NOS3* and *GUCY1A3* loci was observed for CHD (p interaction = 0.53).

We performed mediation analysis to determine whether the degree to which the change in blood pressure associated with the nitric oxide signaling score explained the protective effect on CHD. Adjustment for the effect on systolic blood pressure led only to a modest attenuation of the association of the nitric oxide signaling genetic score with CHD (OR 0.37 CI 0.31, 0.45 prior to adjustment, OR 0.46 CI 0.38, 0.55 after adjustment), suggesting that much of the decrease in CHD risk through increased nitric oxide signaling seems to be through pathways other than blood pressure. Indeed, the effect size of rs3918226 (*NOS3*) and rs7692387 (*GUCY1A3*) on CHD deviated substantially from an estimate based on just the systolic blood pressure effects of these variants (**Figure 13**).

In contrast to common variants that promote increased nitric oxide signaling, I sought to test the hypothesis that rare inactivating variants in *NOS3* or *GUCY1A3* would be associated with increased blood pressure and risk of CHD. 27 participants with predicted loss-of-function variants in *NOS3* or *GUCY1A3* were identified in the T2D GENES study. Presence of a predicted loss-of-function variant in *NOS3* or *GUCY1A3* was associated with increased systolic blood pressure (22.8 mm Hg CI 11.7, 33.9, $p = 5.6*10^{-5}$, respectively; **Figure 14**) and diastolic pressure

(9.7 mm Hg; CI 3.5, 15.9 mm Hg, p= 0.002). 27 individuals with predicted loss-of-function variants were identified in Myocardial Infarction Genetics Consortium studies. Predicted loss-of-function variants in the *NOS3-GUCY1A3* pathway were associated with a three-fold higher risk of CHD (OR 3.03 CI 1.29, 7.12; p=0.01; **Figure 14**).

### *Discussion*

A genetic predisposition to enhanced nitric oxide signaling was associated with reduced blood pressure, improved renal and pulmonary function, and significantly reduced risks of CHD (OR 0.37 CI 0.31, 0.45), peripheral arterial disease (OR 0.42 CI 0.26, 0.68) and stroke (OR 0.53 CI 0.37, 0.76). Mediation analysis suggested that this protective effect is mediated only in part by blood-pressure related pathways. In contrast, mutations predicted to truncate *NOS3* or *GUCY1A3* associated with higher blood pressure and an approximately three-fold higher risk of CHD.

These results permit several conclusions. First, stimulation of nitric oxide signaling may prevent atherosclerotic cardiovascular disease. The use of an oral soluble guanylyl cyclase stimulator has proven effective in the treatment of pulmonary hypertension[65], reinforcing the potential to target this pathway using a small molecule approach. The 63% and 58% reductions in CHD and peripheral arterial disease observed in this study are likely to be of greater magnitude than what would be observed in a randomized trial of a nitric oxide signaling stimulator, as genetic estimates represent the effect of increased nitric oxide signaling over a lifetime rather than an intervention later in life and for a more limited duration.[11] However, the significant and large risk reductions in cardiovascular disease observed in this study lend support to efforts to target nitric oxide signaling in the prevention or treatment of atherosclerotic cardiovascular disease.[51,68,69]

Second, stimulation of nitric oxide signaling may represent a pathway for CHD prevention that is independent of current approaches. A genetic predisposition to nitric oxide signaling was not associated with lipid, glycemic, or anthropometric traits. Furthermore, the majority of the reduction in risk of CHD with increased nitric oxide signaling appeared to be independent of the variant's impact on blood pressure. Consistent with these findings, a recent functional characterization of the lead *GUCY1A3* variant rs7692387 observed carriers of the risk allele of this variant to have elevated levels of platelet aggregation and increased vascular smooth muscle cell migration upon stimulation of soluble guanylyl cyclase.[64] These findings suggested that PDE5A inhibitors including sildenafil may prove useful in the treatment of cardiovascular disease.

Third, these results suggest that stimulation of nitric oxide signaling may improve renal function. A genetic predisposition to enhanced nitric oxide signaling was associated with higher glomerular filtration rate as determined by either cystatin C or creatinine, a finding that awaits further confirmation in larger studies.

This study has several limitations. First, I used common variants in the *GUCY1A3* and *NOS3* loci to estimate the phenotypic effects of increased nitric oxide signaling. It is possible that these variants may be in linkage disequilibrium with other variants that have phenotypic effects independent of nitric oxide signaling pathways. However, the common variants were associated with direct measurement of *GUCY1A3* and *NOS3* expression in vascular and pulmonary tissues, and we observed consistent effects of rare, predicted loss-of-function variants in *GUCY1A3* and *NOS3* on blood pressure and CHD, suggesting that the common variants mediated their effects through the *NOS3-GUCY1A3* nitric oxide signaling pathway. Second, the phenome wide association study to determine the full spectrum of associations in the UK Biobank may have

been underpowered to detect associations for some diseases. Third, the rare variant analysis was restricted to variants predicted to lead to loss of *NOS3* or *GUCY1A3* function; the prevalence of missense mutations that impair the NOS3-GUCY1A3 pathway may be much larger than the prevalence of rare, predicted loss-of-function variants observed in this study. Finally, the majority of participants in this study were of European ancestry and as such, these observations need validation in ancestries outside of Europe as ethnic differences in nitric oxide mediated responses have been previously reported.[70]

In conclusion, a genetic predisposition to enhanced nitric oxide signaling was associated with reduced risks of CHD and peripheral arterial disease and improved renal and pulmonary function. Stimulation of nitric oxide signaling may prove useful for the prevention and treatment of a range of diseases.

## 5. Analysis of predicted loss of function variants in UK Biobank identifies variants protective for disease

A focused investigation of predicted loss-of-function (pLOF) variants provides several advantages when compared with analysis of other types of variants. First, analysis of pLOF variants may allow for the direct identification of a gene rather than a locus containing many candidate genes.[71] Second, pLOF variants provide directionality of effect, unlike non-coding regulatory variants which may increase or decrease expression of a given gene. Third, identification of pLOF variants which protect against disease may aid with prioritization of therapeutic target genes (e.g., the recent development of inhibitors of PCSK9 or ANGPTL3 which mimic human pLOF mutations protective against cardiovascular disease).[72-76]

In this study, I analysed 3,759 pLOF variants in UK Biobank and other datasets to identify 18 new low-frequency or rare (allele frequency <5%) pLOF variant-phenotype associations. I find that the p.Arg95Ter variant in the gene *GPR151* protects against obesity and type 2 diabetes. At the gene phosphodiesterase 3B (*PDE3B*), the p.Arg783Ter variant was associated with elevated height and improved body fat distribution. Outside of metabolic disease, I find that the splice acceptor variant c.662-2A>G in the gene *GSDMB* and the splice acceptor variant c.487-1G>C in the gene *IL33* protect against asthma and other allergic diseases, and that the splice donor variant c.1641+1G>C in the gene *IFIH1* protects against hypothyroidism. These findings prioritize genes where pharmacologic mimics of pLOF variants may lower risk for disease.

### *Methods*

*Gene and variant annotation*

Variants in hg19 coordinates were annotated with information from Ensembl release 82 using Variant Effect Predictor (VEP).[77] Only pLOFs, defined as premature stop (nonsense),

canonical splice-sites (splice-donor or splice-acceptor) or insertion/deletion variants that shifted frame (frameshift) were annotated as predicted loss of function (pLOF), using the "--pick-allele" annotation. PLOFs as defined by VEP were then merged with publicly available data from the Exome Aggregation Consortium (ExAC), Version 0.3.1, to confirm consistency in variant annotation.[50] I identified 3,759 pLOF variants in UK Biobank with an info score greater than 0.3. I used a Bonferroni corrected *P* value of $5.5*10^{-7}$ to denote significance [0.05/(3,759 variants*24 outcomes) = $5.5*10^{-7}$] in our primary pLOF analysis.

*Study Design*

In 405,569 individuals in UK Biobank (335,660 individuals of European ancestry and 69,909 individuals of Non-European ancestry), I analyzed the association of 3,759 pLOF variants with twenty-four phenotypes: cardiovascular, metabolic and pulmonary phenotypes: six metabolic traits (body mass index, waist-to-hip ratio, height, systolic blood pressure, diastolic blood pressure, forced expiratory volume to forced vital capacity ratio), six cardiometabolic diseases (coronary artery disease, type 2 diabetes, atrial fibrillation, stroke, heart failure, venous thromboembolism) and twelve diseases with more than five thousand cases (allergic rhinitis, asthma, anxiety, breast cancer, cataract, cholethiasis, depression, hypothyroidism, gastric reflux, osteoporosis, osteoarthritis, psoriasis). All six metabolic traits were inverse normalized prior to analysis, with adjustment for age and sex. Forced expiratory volume to forced vital capacity ratio was additionally adjusted for height. To adjust for the presence of antihypertensive medication, we added 15 mm Hg to systolic blood pressure and 10 mm Hg to diastolic blood pressure of individuals on antihypertensive medication at baseline, as in the International Consortium for Blood Pressure genome-wide association study.[20]

In UK Biobank, analysis was performed separately in unrelated individuals of European and Non-European ancestry. Estimates for variants were then pooled using inverse-variance weighted-fixed effects meta-analysis. For coronary artery disease, estimates for variants from UK Biobank were additionally pooled with the effect of variants in the CARDIOGRAM Exome consortium using inverse variance weighted-fixed effects meta-analysis.[18] For height, estimates of variants in UK Biobank were pooled with the GIANT Height Exome consortium using inverse variance weighted-fixed effects meta-analysis.[78]

*Genotyping and Quality Control*

Phasing and imputation were performed centrally, by UK Biobank, using a reference panel combining UK10k and 1000 Genomes samples. 39,235,157 variants included in the Haplotype Reference Consortium were imputed. As recommended by UK Biobank, we excluded any samples with an imputation quality < 0.3 as well as pLOF variants which were not included in the Haplotype Reference Consortium. Mitochondrial genetic data and sex chromosomes were excluded from this analysis. Individual level genetic data was available from individuals in UK Biobank, after excluding one related individual of each related pair of individuals, individuals whose genetic sex did not match self-reported sex and individuals with an excess of missing genotype calls or more heterozygosity than expected.

I analysed 3,759 variants identified as pLOF variants in UK Biobank. PLINK 2 software was used to examine the association of these variants with traits and disease in UK Biobank under the assumption of additive effects (https://www.cog-genomics.org/plink/2.0/). Adjustment was performed for age, sex, ten principal components of ancestry and array type. Quantile-quantile analysis was used to examine for the presence of population stratification. No evidence of inflation was observed (inflation factors < 1.1).

A locus-wide conditional analysis (± 500 kb of the pLOF variant) was performed to determine the extent to which the identified pLOF variant signal was independent of other genetic variation at the locus. I iteratively performed association analyses conditional on the top variants at each locus, until no variants were below the Bonferroni-adjusted threshold for significance ($p<5.5*10^{-7}$). A statistically significant and independent signal for the pLOF variant provides increased confidence for a causal association.

*Analysis of PDE3B association with coronary artery disease*

I aimed to analyse the association of pLOF variants in *PDE3B* with coronary artery disease in a combined analysis of UK Biobank and the Myocardial Infarction Genetics Consortium (MIGen). Replication was performed in MIGen rather than the CARDIOGRAM Exome consortium as rs150090666 was not included in the exome chip analysis of the CARDIOGRAM Exome consortium.[18] Estimates of the association of rs150090666 with coronary artery disease in UK Biobank were derived as described above, using logistic regression with adjustment for age, sex, ten principal components of ancestry and a dummy variable for array type. An additional pLOF variant, rs535108921, present in UK Biobank, was also analysed for association with coronary artery disease, as above.

pLOFs variants in *PDE3B* were identified in the MIGen Consortium using exome sequencing or whole genome sequencing, as previously described.[32,33,79] Studies included in the MIGen consortium were: 1) the Italian Atherosclerosis Thrombosis and Vascular Biology (ATVB) study (dbGaP Study Accession phs000814.v1.p1); 2) the Exome Sequencing Project Early-Onset Myocardial Infarction (ESP-EOMI) study(9); 3) a nested case-control cohort from the Jackson Heart Study (JHS); 4) the South German Myocardial Infarction study (dbGaP Study Accession phs000916.v1.p1); 5) the Ottawa Heart Study (OHS) (dbGaP Study Accession

phs000806.v1.p1); 6) the Precocious Coronary Artery Disease (PROCARDIS) study (dbGaP Study Accession phs000883.v1.p1) ; 7) the Pakistan Risk of Myocardial Infarction Study (PROMIS) (dbGaP Study Accession phs000917.v1.p1); 8) the Registre Gironi del COR (Gerona Heart Registry or REGICOR) study (dbGaP Study Accession phs000902.v1.p1); 9) the Leicester Myocardial Infarction study (dbGaP Study Accession phs001000.v1.p1); 10) the BioImage study (dbGaP Study Accession phs001058.v1.p1); 11) the North German Myocardial Infarction study (dbGaP Study Accession phs000990.v1.p1); 12) Multi-Ethnic Study of Atherosclerosis (dbGaP Study Accession: phs000209.v2.p1); 13) Variation In Recovery: Role of Gender on Outcomes of Young AMI cohort; and 14) Taiwan Metabochip Consortium.

The Burrows–Wheeler Aligner algorithm was used to align reads from participants to the reference genome (hg19). The GATK HaploTypeCaller was used to jointly call variants. Metrics including Variant Quality Score Recalibration (VQSR), quality over depth, and strand bias were then used to filter variants. We excluded samples which were related to other samples, which had high ratios of heterozygous to non-reference homozygous genotypes, which had high missing genotypes, which had a discordant genetic gender relative to reports gender, and samples which were discordant relative to genotype data.

After variant calling and quality control, the Variant Effect Predictor[77,80] was used to annotate variants which were pLOF: (1) nonsense mutations that resulted in early termination of *PDE3B* (2) frameshift mutations due to insertions or deletions of DNA; or (3) splice-site mutations which result in an incorrectly spliced protein. For analysis of rare pLOF variants, we pooled rare pLOF variants in MIGen, testing for the association of a pLOF with coronary artery disease using logistic regression, after adjustment for age, sex, cohort and five principle

components of ancestry. I meta-analysed the association of pLOFs with coronary artery disease in MIGen combined with UK Biobank.

*Replication of IL33 finding*

To replicate the association of rs146597587, a splice site variant in *IL33*, with asthma, I pooled estimates of the association of rs146597587 with asthma from Partners Biobank, from the Vanderbilt eMERGE network and from the Women's Genome Health Study. In Partners Biobank, rs146597587 was imputed (info score of 0.77) in 2,542 individuals. The association of rs146597587 with asthma (hospitalization for ICD9 code 493) was estimated using logistic regression, adjusted for age, sex and five principal components of ancestry. In the Vanderbilt eMERGE network, rs146597587 was genotyped in 25,363 individuals using the Illumina Exome BeadChip. The association of rs146597587 with asthma (hospitalization for ICD9 code 493) was estimated using logistic regression, adjusted for age, sex and principal components of ancestry. In Women's Genome Health Study, rs14659758 was genotyped in 22,618 individuals using the Illumina Exome. The association of rs14659758 with asthma (hospitalization for ICD9 code 493 or ICD10 code J45) was estimated using logistic regression, adjusted for age and principal components of ancestry.

## ***Results***

I identified 18 new low-frequency or rare (<5%) pLOF variants associated with traits and diseases in UK Biobank (**Table 4**). Variants identified within the same locus in prior genomewide association studies are provided (**Table 4**). Below, I highlight several new associations.

In *GPR151* (encoding G-protein coupled receptor 151), the p.Arg95Ter variant (rs114285050, allele frequency 0.8% in European ancestry) was associated with reduced body-

mass index (BMI; beta -0.07 standard deviations, -0.36 kg/m$^2$, $P$=4.9*10$^{-8}$). I replicated this association in an independent cohort, the Myocardial Infarction Genetics Consortium (MIGen), where p.Arg95Ter carriers had reduced BMI (beta -0.14, $P$=0.04; pooled beta -0.07, $P$=9.8*10$^{-9}$; **Figure 15**). UK Biobank participants who carry one copy of p.Arg95Ter were at 12% lower odds of clinical obesity (BMI≥30 kg/m$^2$). As obesity is a causal risk factor for type 2 diabetes and coronary artery disease, we examined whether p.Arg95Ter may provide protection against both diseases. p.Arg95Ter was associated with 14% lower odds of type 2 diabetes (OR 0.86; $P$=0.006) and 9% lower odds of coronary artery disease (OR 0.91; $P$=0.01; **Figure 15**). Although *GPR151* is a G-protein coupled receptor of unknown function whose expression is limited to the central nervous system, recent studies tracing the lineage of neurons expressing *GRP151* have localized connections to hypothalamic neurons, a region of the brain important in the control of appetite.[81]

At *PDE3B* encoding the gene phosphodiesterase 3B, p.Arg783Ter (rs150090666, allele frequency 0.0006 in European ancestry) associated with elevated height (beta 0.24, $P$ = 9.3*10$^{-9}$). Targeted deletion of *Pde3b* in mice leads to white adipose tissue gaining characteristics of brown adipose tissue[82], a reduction in adipocyte size[83], smaller fat deposits[84] and reduced atherosclerosis[85]. We therefore studied the association of *PDE3B* p.Arg783Ter with metabolic phenotypes in UK Biobank and/or MIGen, where 36,581 individuals have been sequenced for the 16 exons of the *PDE3B* gene. In UK Biobank, which lacks direct measurements of blood lipids, carriers of p.Arg783Ter carriers were at reduced risk of physician-diagnosis of hypercholesterolemia (OR 0.52, $P$=0.0002). Pooling UK Biobank and MIGen, pLOF carriers in *PDE3B* had reduced WHRadjBMI (beta -0.15, $P$=0.0005). As genetic predisposition to improved body fat distribution has been associated with a lower risk of coronary artery disease[28], we examined whether loss of *PDE3B* function protects against coronary artery disease. We aggregated rare *PDE3B* pLOFs in

cases and compared this count with that controls. Across 14,805 cases in UK Biobank, the carrier frequency of pLOF in cases was 0.1% and in controls 0.2%. Across 20,186 cases in MIGen, the carrier frequency of pLOF was 0.05% and 0.1% in controls. Collectively, carrier status for *PDE3B* pLOFs was associated with reduced risk for coronary artery disease (OR 0.65 CI 0.43, 0.97; p=0.03; **Figure 16**).

We identified several pLOF variants that associated with lower risk of asthma (**Table 4**). At *GSDMB* encoding the gene Gasdermin B, splice acceptor variant c.662-2A>G (rs11078928, allele frequency 46% in European ancestry) protected against asthma (OR 0.90 CI 0.89, 0.91, $P=6.7*10^{-50}$). This variant is in tight linkage disequilibrium ($r^2=0.99$) with a previously identified non-coding variant in the *GSDMB* locus (rs2305480) associated with lower risk of asthma ($P=9.6*10^{-8}$).[22] *GSDMB* c.662-2A>G is associated with lower expression of *GSDMB* transcripts.[86] Furthermore, overexpression of *GSDMB* causes airway remodeling and asthma symptoms in a mouse model[87], suggesting that loss of GSDMB function may protect against asthma.

At the *IL33* gene, a splice acceptor site variant c.487-1G>C (rs146597587, allele frequency 0.004 in European ancestry) was observed to protect against asthma (OR 0.58 CI 0.51, 0.66, $P=7.8*10^{-17}$). This variant was recently identified as associated with lower blood eosinophil concentration at genome-wide significance and with lower risk of asthma at more modest levels of significance ($P=1.8x10^{-4}$).[88] To further replicate the association of *IL33* c.487-1G>C with asthma, we examined the association of *IL33* c.487-1G>C with asthma in individuals from three additional studies (Partners Biobank, the Vanderbilt eMERGE network, and the Women's Genome Health Study). *IL33* c.487-1G>C was associated with a protective effect of asthma in each data set. Overall, *IL33* c.487-1G>C was associated with 43% lower odds of asthma (OR 0.57 CI 0.51, 0.65, $P=9.6*10^{-19}$, **Figure 17**), suggesting that IL33 inhibition may be a useful

approach for treatment of asthma. Of note, an inhibitor of IL33 is currently under development for treatment of asthma.[89]

A splice donor variant in *IFIH1* (interferon induced with helicase C domain 1), c.1641+1G>C (rs35337543, allele frequency 1.5% in European ancestry), is associated with a reduced risk of hypothyroidism in UK Biobank participants (OR 0.77 CI 0.70, 0.85; $P=5*10^{-9}$; **Table 4**). A gene-based test combining four additional pLOF variants in *IFIH1* (rs35732034, rs201026962, rs35744605, rs148590996) similarly demonstrated protection against hypothyroidism in UK Biobank (OR 0.79 CI 0.72, 0.86; $P=4.4*10^{-8}$). Carriers of pLOF variants in *IFIH1* were also protected against hyperthyroidism (OR 0.84 CI 0.73, 0.96; *P*=0.01).

In addition, an exploratory analysis demonstrated a nominally lower risk of coronary artery disease among *IFIH1* pLOF carriers. Pooling UK Biobank and MIGen, *IFIH1* pLOF carriers were protected against coronary artery disease (OR 0.92 CI 0.87, 0.98; p=0.009). To examine whether this may be a chance finding, we examined whether the common *IFIH1* missense variant rs1990760, previously identified as associated with autoimmune disorders, also associated with risk of coronary artery disease. The T allele of rs1990760 (frequency 41%) associated with a reduced risk of hypothyroidism (OR 0.92 CI 0.90, 0.94; $p = 9.3*10^{-17}$) in UK Biobank. Pooling UK Biobank and CARDIOGRAM Exome, the T allele of rs1990760 also associated with a lower risk of coronary artery disease (OR 0.97 CI 0.96, 0.99; $p=2.5*10^{-5}$), providing complementary evidence that *IFIH1* may influence coronary artery disease risk.

Presence of homozygote individuals for pLOF variants in target genes may provide an in vivo demonstration of safety of pharmacologic inhibition of target genes. We therefore examined whether homozygotes for these pLoF variants were present in UK Biobank and in the gnomAD database.[50] At least one individual homozygous for a pLOF variant was identified in UK Biobank

or the gnomAD database for the genes *GPR151*, *GSDMB*, *IL33* and *IFIH1*, and *PDE3B* (**Supp. Table 3**).

## *Discussion*

In this study, we identified pLOF variants that protect against obesity (*GPR151*), coronary artery disease (*PDE3B*), asthma (*GSDMB*, *IL33*) and autoimmune disorders (*IFIH1*), prioritizing genes and pathways where pharmacologic attempts to mimic these protective mutations might ameliorate disease.

Of note, cilostazol is an approved medicine that is a non-selective pharmacologic inhibitor of both phosphodiesterase 3B and the related isoform phosphodiesterase 3A.[90] In a small 211 participant randomized trial, cilostazol significantly reduced restenosis after percutaneous coronary balloon angioplasty.[91] The association of *PDE3B* pLOFs with improved body fat distribution, reduced risk of hypercholesterolemia and reduced risk of coronary artery disease suggests that selective inhibition of PDE3B may be useful for multiple features of metabolic syndrome.

Although our restriction of the present analysis to pLOF variants increases the likelihood of identifying causal variants substantially, it remains possible that a highly correlated nearby variant could be driving the association in some cases. Future functional studies may permit additional validation of causal variants. [92]

## 6. Genetic risk, adherence to a healthy lifestyle and risk of coronary artery disease

*Note: This analysis was led by Amit V Khera. I contributed as a co-first author, contributing to the design of the study and conducting the analysis of ARIC.*

A large body of evidence has also demonstrated that individuals adherent to a healthy lifestyle have markedly lower rates of incident cardiovascular events.[92] The promotion of healthy lifestyle behaviors, which include not smoking, avoiding obesity, regular physical activity, and a healthy diet pattern, is the foundation of current strategy to improve cardiovascular health in the population.[92]

Since 2007, genome-wide association analyses have associated more than fifty independent loci with risk of coronary artery disease. [93] These risk alleles, when aggregated into a polygenic risk score, are predictive of prospectively determined coronary events and provide a continuous and quantitative metric of genetic risk. [94]

Many assume that genetic predisposition to coronary artery disease operates in a deterministic fashion. [95] However, genetic risk might be attenuated by a favorable lifestyle. Here, we tested the hypothesis that both and baseline adherence to a healthy lifestyle are independent axes of risk that each associate with incident coronary events and prevalent subclinical atherosclerotic burden. Secondly, among those with high genetic risk, we determined the extent to which a healthy lifestyle is associated with a reduced risk of coronary artery disease.

### *Methods*

*Study Populations*

The Atherosclerosis Risk in Communities (ARIC) study is a prospective cohort which enrolled white and black participants aged 45 to 64 years starting in 1987. Genotyping was performed using the Affymetrix 6.0 array (Affymetrix, Santa Clara, California). Genotype and clinical data were retrieved from the National Center for Biotechnology Information dbGAP

server (accession: phs000280.v3.p1). Individuals with prevalent coronary artery disease at time of enrollment were excluded. The Women's Genome Health Study (WGHS) is a prospective cohort for genome-wide genetic analysis among female health professionals derived from the Women's Health Study, a clinical trial initiated in 1992 to evaluate the efficacy of aspirin and vitamin E in the primary prevention of cardiovascular disease. Genotyping was performed using the Illumina HumanExome BeadChip v1.0 (Illumina, San Diego, California). The Malmö Diet and Cancer Study (MDCS) is a prospective cohort that enrolled participants in Malmö, Sweden aged 44 to 73 years starting in 1991.Genotyping was performed using a multiplex method as previously described.Participants with prevalent coronary disease at baseline were excluded. The BioImage Study (ClinicalTrials.gov Identifier: NCT00738725) enrolled asymptomatic individuals at risk for cardiovascular disease aged 55 to 80 years beginning in 2008 and included cardiac computed tomography to quantify coronary artery calcium (CAC), a surrogate measure of subclinical coronary atherosclerosis, using the Agatston method.Genotyping was performed using the Illumina HumanExome Bead-Chip Array v1.1.

A genetic risk score was derived on the basis of up to 50 single nucleotide polymorphisms (SNP) that achieved genome-wide significance for association with coronary artery disease in previous studies. In the absence of available genotype, a proxy ($r^2 > 0.8$) SNP was implemented where possible. Individual participant scores were created by summing the number of risk alleles at each SNP multiplied by the literature-based effect size.Individuals missing more than two SNPs were excluded from analysis; for the remainder, missing values were imputed to the population mean. Population genetic substructure was assessed by calculation of principal components of ancestry.

Four healthy lifestyle factors were adapted from the strategic goals of the American Heart Association: 1) no current smoking; 2) absence of obesity (BMI <30 kg/m$^2$); 3) physical activity at least once weekly; and 4) a healthy diet pattern.[96] A healthy diet pattern was calculated based on adherence to at least half of the following characteristics recently endorsed in an evidence-based review, namely: consumption of more fruits, nuts, vegetables, whole grains, fish, and dairy products and less refined grains, processed meats, unprocessed red meats, sugar sweetened beverages, trans fats (WGHS only), and sodium (WGHS only).Because a detailed food frequency questionnaire was not performed in the BioImage study, diet scores in this cohort focused on self-reported consumption of fruit, vegetables, and fish.

We studied a composite coronary artery disease endpoint including myocardial infarction, coronary revascularization, and death from coronary causes. Endpoint adjudication was performed by committee review of medical records for reported endpoints within each cohort. Due to few validated coronary endpoints in the BioImage cohort, a cross-sectional analysis of baseline coronary artery calcium scores was performed.

*Statistical Analysis*

We used Cox proportional hazard models to test the association of genetic and lifestyle factors with incident coronary events. Hazard ratios for those at high (quintile 5) genetic risk were compared to those at intermediate (quintiles 2-4) or low (quintile 1) genetic risk, as previously performed. Similarly, a favorable lifestyle (3-4 healthy lifestyle factors) was compared to an intermediate (2 healthy lifestyle factors) and unfavorable (0-1 healthy lifestyle factor) lifestyle. The primary analyses included adjustment for age, gender, self-reported education level, and the first five principal components of ancestry (unavailable in MDCS). All WGHS analyses were additionally adjusted for initial trial randomization to aspirin versus

placebo and Vitamin E versus placebo. Cohort-specific findings were combined using random effects meta-analysis. 10-year event rates were calculated using Cox regression and standardized to the mean of all predictor variables within each population. Due to a skewed distribution of coronary artery calcification scores in BioImage, linear regression was performed on natural log-transformed calcification score with an offset of 1 [ln(CAC+1)]. Predicted values were then reverse-transformed to calculate standardized calcification scores within genetic and lifestyle risk categories.

Analyses were performed using R version 3.1 software (The R Project for Statistical Computing, Vienna, Austria) and associated "survival" and "ggplot2" packages.

### *Results*

The prospective cohort study populations included 7,814 white ARIC cohort participants, 21,222 initially healthy white women from the WGHS, and 22,389 participants of the MDCS cohort with genotype and covariate data available. 1,230 coronary events were observed over a median follow-up of 18.8 years in the ARIC study, 971 coronary events occurred over a median follow-up of 20.5 years in the WGHS cohort, and 2,902 coronary events occurred over a median follow-up of 19.4 years in MDCS.

A gradient of risk was noted across quintiles of genetic risk, such that those at high genetic risk, defined as those in the top quintile of the genetic risk score, were at 75% (95%CI 46 – 210%), 94% (CI 58 – 139%), and 98% (CI 76 – 123%) greater risk compared to those at low genetic risk (lowest quintile) in the ARIC, WGHS, and MDCS cohorts respectively (**Figure 18**). Survey-based assessment of a family history of coronary artery disease was an imperfect surrogate for genotype-defined risk, although the prevalence of self-reported family history tended to be higher in those with high versus low genetic risk in each cohort. LDL cholesterol

levels were modestly increased across categories of genetic risk within each cohort ($p< 0.001$); by contrast, genetic risk categories were independent of age, gender, prevalent hypertension or diabetes, body-mass index, smoking status, HDL cholesterol, triglyceride levels, C-Reactive protein (available only in WGHS and MDCS), and 10-year cardiovascular risk as predicted by the ACC/AHA Pooled Cohorts Equation.

Each of the four healthy lifestyle factors was associated with a decreased risk of coronary events in a combined analysis of the prospective cohorts: no current smoking (HR 0.56; 95%CI 0.47 – 0.66; $p< 0.001$), absence of obesity (HR 0.66; 95%CI 0.58 – 0.76; $p< 0.001$), regular physical activity (HR 0.88; 95%CI 0.80 – 0.97; $p= 0.01$); and healthy diet (HR 0.91; 95%CI 0.83 – 0.99; $p= 0.04$). Coronary risk increased in those with fewer healthy lifestyle factors within each cohort. Each cohort was divided into three lifestyle risk categories: favorable (at least three of four healthy lifestyle factors), intermediate (two healthy lifestyle factors), or unfavorable (zero or one healthy lifestyle factor) lifestyle. Those with an unfavorable lifestyle had higher rates of baseline hypertension and diabetes, higher body-mass-indices, and less favorable levels of circulating lipids. An unfavorable lifestyle was associated with a 71% (95%CI 47 – 98%), 127% (95%CI 92 – 167%), and 77% (95%CI 61 – 95) increased risk of coronary events when compared to those with a favorable lifestyle in the ARIC, WGHS, and MDCS cohorts respectively (**Figure 18**).

Within each category of genetic risk, lifestyle factors were strong predictors of coronary events (**Figure 19**). Adherence to a favorable lifestyle, as compared to an unfavorable lifestyle, was associated with a 45%, 47%, and 46% relative risk reduction among those with low, intermediate, and high genetic risk respectively. In absolute terms, among those with high genetic risk, standardized 10-year event rates decreased from 10.7 to 5.1%, 4.6 to 2.0%, and 8.2

to 5.3% in ARIC, WGHS, and MDCS, respectively, based on a favorable lifestyle (Figure 4). Similarly, a low genetic risk was largely offset by an unfavorable lifestyle. Standardized 10-year event rates increased from 3.1 to 5.8% in ARIC, 1.2 to 1.8% in WGHS, and 2.6 to 4.7% in MDCS among those with an unfavorable as compared to favorable lifestyle. These trends were observed across all categories of genetic risk.

Despite a paucity of well-validated genetic loci in black populations, we observed similar findings in 2,269 black participants of the ARIC study (350 incident coronary events). Those at high genetic risk were at 65% increased risk of coronary events (95%CI 16 – 134%; p= 0.006) compared to those at low genetic risk. Furthermore, an unfavorable lifestyle was associated with a 70% increased coronary risk (95%CI 120 – 139%; p= 0.003). As with white participants, risk of coronary events tended to increase with a less favorable lifestyle within categories of low and intermediate genetic risk. This pattern was not apparent among those with a high genetic risk, potentially related to decreased power due to a small number of incident events. Additional data is needed to confirm consistency of effect in non-European populations.

A cross-sectional analysis of 4,260 white participants of the BioImage Study demonstrated that both genetic and lifestyle factors were associated with coronary artery calcification, a measure of subclinical coronary atherosclerosis burden. The standardized calcification score was 46 Agatston units (95%CI 39 - 54) in those with high genetic risk compared to 21 Agatston units (95%CI 18 - 25) in those at low genetic risk; p< 0.001. Calcification score was similarly increased in those with an unfavorable versus favorable lifestyle; 46 (95%CI 40 – 53) versus 28 Agatston units (95%CI 25 - 31), p< 0.001. Within each subgroup of genetic risk a significant trend was observed towards decreased atherosclerosis burden in those more adherent to a healthy lifestyle (**Figure 20**).

### *Discussion*

We assessed both genetic and lifestyle risk for a common, complex disease – coronary atherosclerosis – in three cohorts encompassing 55,685 individuals with 5,103 incident coronary events and 4,260 coronary artery calcium measurements. Previous studies focused on either genetic or lifestyle factors have demonstrated risk gradients across categories of each. Here, we provide new quantitative insights into the interplay between genetic and lifestyle risk factors for coronary artery disease. High genetic risk was independent of healthy lifestyle behaviors and associated with a 91% increase in risk of coronary events and substantially increased coronary artery calcium scores. Within any genetic risk category, adherence to a healthy lifestyle was associated with a significantly decreased risk of clinical coronary events as well as subclinical coronary atherosclerosis burden.

These results permit several conclusions. First, inherited DNA variation and lifestyle factors contribute to coronary atherosclerosis susceptibility in an additive fashion. Our finding that a polygenic risk score has robust associations with incident coronary events is well-aligned with previous studies of both primary and secondary prevention populations. [94, 97] These findings reinforce longstanding beliefs that genetic variants identifiable from birth alter coronary risk. Importantly, aside from slight differences in LDL cholesterol levels and family history of coronary artery disease, genetic risk was independent of traditionally measured risk factors.

Second, a healthy lifestyle is associated with similar relative risk reductions in event rates across each stratum of genetic risk. Although the absolute risk reduction associated with adherence to a highest lifestyle was greatest in the high genetic risk group, as was recently observed in randomized trials involving statin therapy,our results support public health efforts emphasizing a healthy lifestyle *in everyone*. An alternate approach is to target intensive lifestyle modification to those at high genetic risk, with the expectation that disclosure of genetic risk can

motivate behavioral change. However, studies to date that have sought to influence behavior by communicating DNA-based risk have been largely negative.[98] Participants in the recently published MIGENES study who received their ten-year risk of coronary events incorporating a genetic risk score were more likely to initiate statin therapy but demonstrated no change in dietary fat intake or physical activity.[98] Whether genetic risk disclosure can impact other clinically meaningful health outcomes remains to be determined.

Third, patients may equate DNA-based risk estimated with determinism, a perceived lack of control over the ability to improve outcomes; however, our results provide evidence to counter this determinism and, rather, emphasize that lifestyle factors may powerfully modify disease risk regardless of genetic risk profile. Indeed, alternative analysis approaches that incorporate more stringent cutoffs or weight the relative impact for each healthy lifestyle factor may lead to an even more pronounced gradient of coronary risk. In absolute terms, adherence to a healthy lifestyle was associated with a 5.5%, 2.6%, and 2.9% reduction in standardized ten-year incidence of coronary events in the high genetic risk participants of the ARIC, WGHS, and MDCS studies respectively. Similarly, among BioImage Study participants with high genetic risk, coronary artery calcification was 32 Agatston units lower in those with a favorable lifestyle.

Several limitations to our study should be acknowledged. First, adherence to a healthy lifestyle was not randomized and as such, the association of lifestyle factors with risk for coronary events may not reflect a causal relationship. Second, lifestyle was assessed at baseline within each cohort using slightly different methodologies. Behavioral changes before or after ascertainment or competing risks of other illnesses may have further impacted lifestyle risk estimates. Third, we included a set of up to 50 genetic polymorphisms with robust associations with coronary artery disease; inclusion of even more polymorphisms with validated associations

with coronary artery disease or a causal risk factor may prove useful in future analyses. Lastly, although we provide new evidence confirming a relationship between a polygenic risk score and coronary events in black participants, our primary analysis was restricted to white individuals. As more robust ethnicity-specific genetic association data becomes available, the generalizability of our findings should be tested in more diverse populations.

In conclusion, on the basis of quantifying both genetic and lifestyle risk in 55,685 individuals, we demonstrate that adherence to a healthy lifestyle is associated with substantially reduced risk of coronary artery disease within any given category of genetic risk. These findings support policy initiatives that promote healthy lifestyle behaviors in the entire population, whether low or high genetic risk.

## 7. Summary of findings

Through the analysis of large-scale human genetic datasets, I identify five novel risk factors and therapeutic targets for cardiometabolic disease. First, using Mendelian randomization, I demonstrate that body fat distribution is a causal risk factor for coronary artery disease and type 2 diabetes, with a similar magnitude of effect on disease risk as body mass index. Second, exploiting the genetic association between body fat distribution and type 2 diabetes, I identify a series of damaging variants in the receptor ALK7 that reduce abdominal adiposity and protect against type 2 diabetes. These findings suggest that pharmacologic inhibition of ALK7 may be useful in the treatment of type 2 diabetes. Third, I show that genetic nitric oxide signaling protects against cardiovascular disease and improves renal function, suggesting that nitric oxide signaling agents such as PDE5A inhibitors may be useful in the treatment of cardiovascular and renal disease. Fourth, through the analysis of rare predicted loss-of-function variants in UK Biobank, I identify loss-of-function variants in GPR151 and PDE5A that protect against obesity and coronary artery disease, respectively. These findings suggest that GPR151 inhibition may be a novel therapeutic approach to weight loss. Finally, I show that healthy lifestyle can mitigate inherited genetic predisposition to risk of cardiovascular disease, identifying a non-pharmacologic method of reducing genetic risk for coronary artery disease.

## References


1. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*. 2013;12(8):581-594. doi:10.1038/nrd4051.

2. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. Marchini J, ed. *PLoS Genet*. 2019;15(12):e1008489. doi:10.1371/journal.pgen.1008489.

3. Nordestgaard BG, Palmer TM, Benn M, et al. The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach. Minelli C, ed. *PLoS medicine*. 2012;9(5):e1001212. doi:10.1371/journal.pmed.1001212.

4. Prospective Studies Collaboration, Whitlock G, Lewington S, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. 2009;373(9669):1083-1096. doi:10.1016/S0140-6736(09)60318-4.

5. Després J-P. Body fat distribution and risk of cardiovascular disease: an update. *Circulation*. 2012;126(10):1301-1313. doi:10.1161/CIRCULATIONAHA.111.067264.

6. Ashwell M, Cole TJ, Dixon AK. Obesity: new insight into the anthropometric classification of fat distribution shown by computed tomography. *Br Med J (Clin Res Ed)*. 1985;290(6483):1692-1694.

7. Seidell JC, Björntorp P, Sjöström L, Sannerstedt R, Krotkiewski M, Kvist H. Regional distribution of muscle and fat mass in men--new insight into the risk of abdominal obesity using computed tomography. *Int J Obes*. 1989;13(3):289-303.

8. Yusuf S, Hawken S, Ounpuu S, et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. 2005;366(9497):1640-1649. doi:10.1016/S0140-6736(05)67663-5.

9. Vazquez G, Duval S, Jacobs DR, Silventoinen K. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiol Rev*. 2007;29(1):115-128. doi:10.1093/epirev/mxm008.

10. Han TS, Bijnen FC, Lean ME, Seidell JC. Separate associations of waist and hip circumference with lifestyle factors. *Int J Epidemiol*. 1998;27(3):422-430.

11. Smith GD, Ebrahim S. "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.

12. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40(4):304-314. doi:10.1002/gepi.21965.

13. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283. doi:10.1038/ng.2797.

14. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet*. 2012;44(9):991-1005. doi:10.1038/ng.2385.

15. Shungin D, Winkler TW, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518(7538):187-196. doi:10.1038/nature14132.

16. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1038/nature14177.

17. Pattaro C, Teumer A, Gorski M, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun*. 2016;7:10023. doi:10.1038/ncomms10023.

18. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med*. 2016;374(12):1134-1144. doi:10.1056/NEJMoa1507652.

19. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44(9):981-990. doi:10.1038/ng.2383.

20. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478(7367):103-109. doi:10.1038/nature10405.

21. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121-1130. doi:10.1038/ng.3396.

22. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*. 2010;363(13):1211-1221. doi:10.1056/NEJMoa0906312.

23. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665. doi:10.1002/gepi.21758.

24. Stock JH, Wright JH, Yogo M. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*. 2012;20(4):518-529. doi:10.1198/073500102288618658.

25. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. 2012;380(9841):572-580. doi:10.1016/S0140-6736(12)60312-2.

26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909. doi:10.1038/ng1847.

27. Rodgers RJ, Tschöp MH, Wilding JPH. Anti-obesity drugs: past, present and future. *Dis Model Mech*. 2012;5(5):621-626. doi:10.1242/dmm.009621.

28. Emdin C, Khera AV, Natarajan P, et al. Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA : the journal of the American Medical Association*. 2017;317(6):626-634. doi:10.1001/jama.2016.21042.

29. Dale CE, Fatemifar G, Palmer TM, et al. Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. *Circulation*. 2017;135(24):2373-2388. doi:10.1161/CIRCULATIONAHA.116.026560.

30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8.

31. Mahajan A, Wessel J, Willems SM, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet*. 2018;50(4):559-571. doi:10.1038/s41588-018-0084-1.

32. Do R, Stitziel NO, Won H-H, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*. 2015;518(7537):102-106. doi:10.1038/nature13917.

33. Khera AV, Won H-H, Peloso GM, et al. Diagnostic Yield of Sequencing Familial Hypercholesterolemia Genes in Patients with Severe Hypercholesterolemia. *J Am Coll Cardiol*. March 2016. doi:10.1016/j.jacc.2016.03.520.

34. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536(7614):41-47. doi:10.1038/nature18642.

35. Emdin C, Khera AV, Klarin D, et al. Phenotypic Consequences of a Genetic Predisposition to Enhanced Nitric Oxide Signaling. *Circulation*. 2018;137(3):222-232. doi:10.1161/CIRCULATIONAHA.117.028021.

36. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *American journal of epidemiology*. 1989;129(4):687-702.

37. Khera AV, Chaffin M, Aragam K, et al. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. *bioRxiv*. November 2017:218388. doi:10.1101/218388.

38. Emdin C, Khera AV, Natarajan P, et al. Phenotypic Characterization of Genetically Lowered Human Lipoprotein(a) Levels. *J Am Coll Cardiol*. 2016;68(25):2761-2772. doi:10.1016/j.jacc.2016.10.033.

39. Park J-H, Gail MH, Weinberg CR, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci USA*. 2011;108(44):18026-18031. doi:10.1073/pnas.1114759108.

40. Liu DJ, Peloso GM, Yu H, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet*. 2017;49(12):1758-1766. doi:10.1038/ng.3977.

41. Emdin C, Klarin D, Natarajan P, et al. Genetic Variation at the Sulfonylurea Receptor, Type 2 Diabetes, and Coronary Heart Disease. *Diabetes*. 2017;66(8):2310-2315. doi:10.2337/db17-0149.

42. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42(2):105-116. doi:10.1038/ng.520.

43. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-660. doi:10.1126/science.1262110.

44. Bondestam J, Huotari MA, Morén A, et al. cDNA cloning, expression studies and chromosome mapping of human type I serine/threonine kinase receptor ALK7 (ACVR1C). *Cytogenet Cell Genet*. 2001;95(3-4):157-162. doi:10.1159/000059339.

45. Zhang N, Kumar M, Xu G, et al. Activin receptor-like kinase 7 induces apoptosis of pancreatic beta cells and beta cell lines. *Diabetologia*. 2006;49(3):506-518. doi:10.1007/s00125-005-0095-1.

46. Yogosawa S, Mizutani S, Ogawa Y, Izumi T. Activin receptor-like kinase 7 suppresses lipolysis to accumulate fat in obesity through downregulation of peroxisome proliferator-activated receptor γ and C/EBPα. *Diabetes*. 2013;62(1):115-123. doi:10.2337/db12-0295.

47. Zhao F, Huang F, Tang M, et al. Nodal induces apoptosis through activation of the ALK7 signaling pathway in pancreatic INS-1 β-cells. *American Journal of Physiology - Endocrinology and Metabolism*. 2012;303(1):E132-E143. doi:10.1152/ajpendo.00074.2012.

48. Andersson O, Korach-Andre M, Reissmann E, Ibáñez CF, Bertolino P. Growth/differentiation factor 3 signals through ALK7 and regulates accumulation of adipose tissue and diet-induced obesity. *Proc Natl Acad Sci USA*. 2008;105(20):7252-7256. doi:10.1073/pnas.0800272105.

49. Guo T, Marmol P, Moliner A, et al. Adipocyte ALK7 links nutrient overload to catecholamine resistance in obesity. *Elife*. 2014;3:e03245. doi:10.7554/eLife.03245.

50. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. doi:10.1038/nature19057.

51. Lundberg JO, Gladwin MT, Weitzberg E. Strategies to increase nitric oxide signalling in cardiovascular disease. *Nat Rev Drug Discov*. 2015;14(9):623-641. doi:10.1038/nrd4623.

52. Steinhorn BS, Loscalzo J, Michel T. Nitroglycerin and Nitric Oxide--A Rondo of Themes in Cardiovascular Therapeutics. *N Engl J Med*. 2015;373(3):277-280. doi:10.1056/NEJMsr1503311.

53. Förstermann U, Closs EI, Pollock JS, et al. Nitric oxide synthase isozymes. Characterization, purification, molecular cloning, and functions. *Hypertension*. 1994;23(6 Pt 2):1121-1131.

54. Mergia E, Friebe A, Dangel O, Russwurm M, Koesling D. Spare guanylyl cyclase NO receptors ensure high NO sensitivity in the vascular system. *J Clin Invest*. 2006;116(6):1731-1737. doi:10.1172/JCI27657.

55. Denninger JW, Marletta MA. Guanylate cyclase and the .NO/cGMP signaling pathway. *Biochim Biophys Acta*. 1999;1411(2-3):334-350.

56. Nong Z, Hoylaerts M, Van Pelt N, Collen D, Janssens S. Nitric oxide inhalation inhibits platelet aggregation and platelet-mediated pulmonary thrombosis in rats. *Circ Res*. 1997;81(5):865-869.

57. Knowles JW, Reddick RL, Jennette JC, Shesely EG, Smithies O, Maeda N. Enhanced atherosclerosis and kidney dysfunction in eNOS(-/-)Apoe(-/-) mice are ameliorated by enalapril treatment. *J Clin Invest*. 2000;105(4):451-458. doi:10.1172/JCI8376.

58. Kuhlencordt PJ, Gyurko R, Han F, et al. Accelerated atherosclerosis, aortic aneurysm formation, and ischemic heart disease in apolipoprotein E/endothelial nitric oxide synthase double-knockout mice. *Circulation*. 2001;104(4):448-454.

59. Erdmann J, Stark K, Esslinger UB, et al. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature*. 2013;504(7480):432-436. doi:10.1038/nature12722.

60. Salvi E, Kutalik Z, Glorioso N, et al. Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension*. 2012;59(2):248-255. doi:10.1161/HYPERTENSIONAHA.111.181990.

61. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet*. 2016;48(10):1171-1184. doi:10.1038/ng.3667.

62. Johnson T, Gaunt TR, Newhouse SJ, et al. Blood pressure loci identified with a gene-centric array. *Am J Hum Genet*. 2011;89(6):688-700. doi:10.1016/j.ajhg.2011.10.013.

63. Surendran P, Drenos F, Young R, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet*. 2016;48(10):1151-1161. doi:10.1038/ng.3654.

64. Kessler T, Wobst J, Wolf B, et al. Functional Characterization of the GUCY1A3 Coronary Artery Disease Risk Locus. *Circulation*. May 2017:CIRCULATIONAHA.116.024152. doi:10.1161/CIRCULATIONAHA.116.024152.

65. Ghofrani H-A, Galiè N, Grimminger F, et al. Riociguat for the treatment of pulmonary arterial hypertension. *N Engl J Med*. 2013;369(4):330-340. doi:10.1056/NEJMoa1209655.

66. Smemo S, Tena JJ, Kim K-H, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507(7492):371-375. doi:10.1038/nature13138.

67. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219.

68. Wobst J, Kessler T, Dang TA, Erdmann J, Schunkert H. Role of sGC-dependent NO signalling and myocardial infarction risk. *J Mol Med*. 2015;93(4):383-394. doi:10.1007/s00109-015-1265-3.

69. Wobst J, Ameln von S, Wolf B, et al. Stimulators of the soluble guanylyl cyclase: promising functional insights from rare coding atherosclerosis-related GUCY1A3 variants. *Basic Res Cardiol*. 2016;111(4):51. doi:10.1007/s00395-016-0570-5.

70. Cardillo C, Kilcoyne CM, Cannon RO, Panza JA. Attenuation of cyclic nucleotide-mediated smooth muscle relaxation in blacks as a cause of racial differences in vasodilator function. *Circulation*. 1999;99(1):90-95.

71. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93(5):779-797. doi:10.1016/j.ajhg.2013.10.012.

72. Abifadel M, Varret M, Rabès J-P, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*. 2003;34(2):154-156. doi:10.1038/ng1161.

73. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med*. 2006;354(12):1264-1272. doi:10.1056/NEJMoa054013.

74. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med*. March 2017:NEJMoa1615664. doi:10.1056/NEJMoa1615664.

75. Dewey FE, Gusarova V, Dunbar RL, et al. Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease. *N Engl J Med*. 2017;377(3):211-221. doi:10.1056/NEJMoa1612790.

76. Graham MJ, Lee RG, Brandt TA, et al. Cardiovascular and Metabolic Effects of ANGPTL3 Antisense Oligonucleotides. *N Engl J Med*. 2017;377(3):222-232. doi:10.1056/NEJMoa1701329.

77. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26(16):2069-2070. doi:10.1093/bioinformatics/btq330.

78. Marouli E, Graff M, Medina-Gomez C, et al. Rare and low-frequency coding variants alter human adult height. *Nature*. 2017;542(7640):186-190. doi:10.1038/nature21039.

79. Khera AV, Chaffin M, Zekavat SM, et al. Derivation and Validation of a 6.6 Million Variant Polygenic Score for Coronary Artery Disease. *In Press*.

80. Karczewski KJ. LOFTEE (Loss-Of-Function Transcript Effect Estimator). https://github.com/konradjk/loftee. Published 2015. Accessed May 27, 2016.

81. Broms J, Grahm M, Haugegaard L, Blom T, Meletis K, Tingström A. Monosynaptic retrograde tracing of neurons expressing the G-protein coupled receptor Gpr151 in the mouse brain. *J Comp Neurol*. 2017;525(15):3227-3250. doi:10.1002/cne.24273.

82. Guirguis E, Hockman S, Chung YW, et al. A role for phosphodiesterase 3B in acquisition of brown fat characteristics by white adipose tissue in male mice. *Endocrinology*. 2013;154(9):3152-3167. doi:10.1210/en.2012-2185.

83. Choi YH, Park S, Hockman S, et al. Alterations in regulation of energy homeostasis in cyclic nucleotide phosphodiesterase 3B-null mice. *J Clin Invest*. 2006;116(12):3240-3251. doi:10.1172/JCI24867.

84. Chung YW, Ahmad F, Tang Y, et al. White to beige conversion in PDE3B KO adipose tissue through activation of AMPK signaling and mitochondrial function. *Sci Rep*. 2017;7:40445. doi:10.1038/srep40445.

85. Ahmad F, Chung YW, Tang Y, et al. Phosphodiesterase 3B (PDE3B) regulates NLRP3 inflammasome in adipose tissue. *Sci Rep*. 2016;6(1):28056. doi:10.1038/srep28056.

86. Morrison FS, Locke JM, Wood AR, et al. The splice site variant rs11078928 may be associated with a genotype-dependent alteration in expression of GSDMB transcripts. *BMC Genomics*. 2013;14(1):627. doi:10.1186/1471-2164-14-627.

87. Das S, Miller M, Beppu AK, et al. GSDMB induces an asthma phenotype characterized by increased airway responsiveness and remodeling without lung inflammation. *Proc Natl Acad Sci USA*. 2016;113(46):13132-13137. doi:10.1073/pnas.1610433113.

88. Smith D, Helgason H, Sulem P, et al. A rare IL33 loss-of-function mutation reduces blood eosinophil counts and protects from asthma. Lappalainen T, ed. *PLoS Genet*. 2017;13(3):e1006659. doi:10.1371/journal.pgen.1006659.

89. Gangwar RS, Landolina N, Arpinati L, Levi-Schaffer F. Mast cell and eosinophil surface receptors as targets for anti-allergic therapy. *Pharmacol Ther*. 2017;170:37-63. doi:10.1016/j.pharmthera.2016.10.010.

90. Chung YW, Lagranha C, Chen Y, et al. Targeted disruption of PDE3B, but not PDE3A, protects murine heart from ischemia/reperfusion injury. *Proc Natl Acad Sci USA*. 2015;112(17):E2253-E2262. doi:10.1073/pnas.1416230112.

91. Tsuchikane E, Fukuhara A, Kobayashi T, et al. Impact of cilostazol on restenosis after percutaneous coronary balloon angioplasty. *Circulation*. 1999;100(1):21-26.

92. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(25 Suppl 2):S1-S45. doi:10.1161/01.cir.0000437738.63853.7a.

93. Klarin D, Zhu QM, Emdin C, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet*. 2017;49(9):1392-1397. doi:10.1038/ng.3914.

94. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z.

95. WHITE PD. Genes, the heart and destiny. *N Engl J Med*. 1957;256(21):965-969. doi:10.1056/NEJM195705232562101.

96. American Heart Association Nutrition Committee, Lichtenstein AH, Appel LJ, et al. Diet and lifestyle recommendations revision 2006: a scientific statement from the American Heart Association Nutrition Committee. *Circulation*. 2006;114(1):82-96. doi:10.1161/CIRCULATIONAHA.106.176158.

97. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. 2015;385(9984):2264-2271. doi:10.1016/S0140-6736(14)61730-X.

98. Kullo IJ, Jouni H, Austin EE, et al. Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the

MI-GENES Clinical Trial). *Circulation*. 2016;133(12):1181-1188. doi:10.1161/CIRCULATIONAHA.115.020109.

## Tables

**Table 1.** Characteristics of UK Biobank participants analyzed in the Mendelian randomization study of waist-to-hip ratio with cardiometabolic disease.

| | |
|---|---|
| N Individuals | 111986 |
| Age ± SD, years | 56.9 ± 7.9 |
| Male, n (%) | 53141 (47.5%) |
| UK BiLEVE Array, n (%) | 38505 (34.4%) |
| SBP ± SD, mm Hg* | 143.6 ± 21.8 |
| DBP ± SD, mm Hg* | 84.5 ± 11.8 |
| BMI ± SD, kg/m2 | 27.5 ± 4.8 |
| Waist-to-Hip Ratio ± SD | 0.875 ± 0.09 |
| Diabetes Mellitus, n (%) | 5690 (5.1%) |
| Coronary Heart Disease, n (%) | 5639 (5.0%) |

*7681 individuals were missing a BP measurement at baseline. Reported measurements are after adjustment for treatment (addition of 15 mm Hg to SBP and 10 mm Hg to DBP)
Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; BP, blood pressure; SD, standard deviation; BMI, body mass index.
UK BiLEVE Array refers to participants who were genotyped using the UK BiLEVE array rather than the UK Biobank Axiom Array.

**Table 2.** Baseline characteristics of participants in UK Biobank used for discovery of low-frequency coding variation associated with waist-to-hip ratio.

| N Individuals | All Participants (405569) | Type 2 Diabetes Cases (20458) | Type 2 Diabetes Free Controls (385111) |
|---|---|---|---|
| Age ± SD, years | 57 ± 8.1 | 61 ± 6.9 | 57 ± 8.1 |
| Female, n (%) | 218376 (54%) | 7801 (38%) | 210575 (55%) |
| UK BiLEVE Array, n (%) | 48625 (12.0%) | 17361 (15%) | 3097 (12%) |
| Body Mass Index ± SD, kg/m2 | 27 ± 4.8 kg/m2 | 32 ± 5.9 kg/m2 | 27 ± 4.6 kg/m2 |
| Waist-to-Hip Ratio ± SD | 0.87 ± 0.09 | 0.95 ± 0.08 | 0.87 ± 0.09 |

Abbreviations: SD, standard deviation;

**Table 3.** Association of variants in *ACVR1C* with waist-to-hip ratio adjusted for body mass index and with type 2 diabetes. Estimates for WHRadjBMI were derived through linear regression analysis in UK Biobank. Estimates for type 2 diabetes were derived through meta-analysis of UK Biobank and the DIAGRAM ExTexT2D consortium.

| | | ***WHRadjBMI*** | | ***Type 2 Diabetes*** | |
|---|---|---|---|---|---|
| **Variant** | **Minor Allele Frequency (%)** | **Beta (CI)** | **p-value** | **OR (CI)** | **p-value** |
| Asn150His | 1.1 | -0.089 (-0.11, -0.067) | $3.4*10^{-17}$ | 0.88 (0.83, 0.94) | $8.7*10^{-5}$ |
| Ile195Thr | 0.2 | -0.15 (-0.09, 0.19) | $1.0*10^{-9}$ | 0.79 (0.67, 0.93) | 0.005 |
| Ile482Val | 7.2 | -0.019 (-0.01, -0.027) | $1.6*10^{-5}$ | 0.95 (0.93, 0.97) | $4.8*10^{-6}$ |
| rs72927479 | 5.1 | -0.035 (-0.045, -0.025) | $2.6*10^{-12}$ | 0.93 (0.89, 0.97) | $6.0*10^{-4}$ |

**Table 4.** Predicted loss of function variants with minor allele frequency <5% which are significantly associated with traits or diseases in UK Biobank.

| Outcome | Gene | pLOF Variant | Location | EA | RA | Amino Acid Change | Frequency % | Beta | SE | P-value | Novel? | MHC Locus? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Asthma | FLG | rs61816761 | 1:152285861 | A | G | p.Arg501Ter | 1.51 | 0.21 | 0.03 | $1.51*10^{-15}$ | Yes | No |
| Asthma | HLA-DQB1 | rs28688207 | 6:32628660 | C | T | Splice Acceptor c.773-1A>G | 3.14 | -0.17 | 0.02 | $3.11*10^{-15}$ | Yes | Yes |
| Asthma | IL33 | rs146597587 | 9:6255967 | C | G | Splice Acceptor c.487-1G>C | 0.44 | -0.54 | 0.06 | $7.79*10^{-17}$ | No[65] | No |
| BMI | GPR151 | rs114285050 | 5:145895394 | A | G | p.Arg95Ter | 0.78 | -0.07 | 0.01 | $4.89*10^{-8}$ | Yes | No |
| BMI | PKHD1L1 | rs533623778 | 8:110523131 | T | C | p.Arg769Ter | $1.0*10^{-4}$ | 5.30 | 0.99 | $9.45*10^{-8}$ | Yes | No |
| DBP | ENPEP | rs33966350 | 4:111431444 | A | G | p.Trp413Ter | 1.19 | 0.06 | 0.01 | $8.12*10^{-10}$ | No[64] | No |
| DBP | BTN3A2 | rs58367598 | 6:26370833 | G | T | Splice Donor c.715+2T>A | 3.75 | 0.03 | 0.01 | $2.03*10^{-8}$ | Yes | Yes |
| DBP | TMC3 | rs150843673 | 15:81624929 | T | G | p.Ser1045Ter | 2.14 | 0.05 | 0.01 | $8.16*10^{-9}$ | Yes | No |
| Height | PDE11A | rs76308115 | 2:178879181 | A | G | p.Arg57Ter | 0.52 | 0.07 | 0.01 | $6.20*10^{-11}$ | Yes | No |
| Height | CLHC1 | rs114931154 | 2:55407644 | T | A | Splice Donor c.1384+2T>A | 1.26 | -0.05 | 0.01 | $1.54*10^{-11}$ | Yes | No |
| Height | CCDC66 | rs150364083 | 3:56628033 | T | C | p.Arg427Ter | 0.58 | 0.05 | 0.01 | $2.09*10^{-7}$ | Yes | No |
| Height | DAP | rs201354802 | 5:10761153 | A | C | p.Glu10Ter | 0.24 | 0.13 | 0.02 | $1.68*10^{-8}$ | Yes | No |
| Height | TRIM40 | rs115651142 | 6:30115320 | T | G | Splice Donor c.602+1G>T | 0.63 | -0.08 | 0.01 | $1.16*10^{-9}$ | Yes | Yes |
| Height | MICA | rs181430930 | 6:31378575 | A | G | Splice Donor c.286+1G>A | 0.26 | -0.12 | 0.02 | $7.87*10^{-8}$ | Yes | Yes |
| Height | PDE3B | rs150090666 | 11:14865399 | T | C | p.Arg783Ter | 0.06 | 0.24 | 0.04 | $9.32*10^{-9}$ | Yes | No |
| Height | APOLD1 | rs202116412 | 12:12879031 | A | G | Splice Donor c.96+1G>A | 0.03 | 0.12 | 0.02 | $3.06*10^{-8}$ | Yes | No |
| Hypothyroidism | IFIH1 | rs35337543 | 2:163136505 | G | C | Splice Donor c.1641+1G>A | 1.45 | -0.27 | 0.04 | $2.95*10^{-9}$ | Yes | No |
| Psoriasis | ZKSCAN3 | rs73387810 | 6:28318166 | A | G | Splice Donor c.-63+1G>A | 0.86 | 0.55 | 0.08 | $4.18*10^{-11}$ | Yes | Yes |
| Psoriasis | EGFL8 | rs141826798 | 6:32134395 | G | C | p.Arg74Ter | 0.53 | 0.90 | 0.08 | $2.19*10^{-26}$ | Yes | Yes |
| SBP | ENPEP | rs33966350 | 4:111431444 | A | G | p.Trp413Ter | 1.19 | 0.06 | 0.01 | $3.46*10^{-9}$ | No[64] | No |
| SBP | GEM | rs138582164 | 8:95264265 | A | G | p.Arg199Ter | 0.04 | 0.30 | 0.06 | $1.93*10^{-7}$ | No[67] | No |
| WHRadjBMI | PYGM | rs116987552 | 11:64527223 | A | G | p.Arg50Ter | 0.39 | 0.09 | 0.02 | $1.32*10^{-7}$ | Yes | No |

Beta in terms of standard deviations and reported for the effect allele. pLOF, predicted loss-of-function; EA, effect allele; RA, reference allele; BMI, body mass index; DBP, diastolic blood pressure; SBP, systolic blood pressure; WHRadjBMI, waist-to-hip ratio adjusted for body mass index.

## Figures

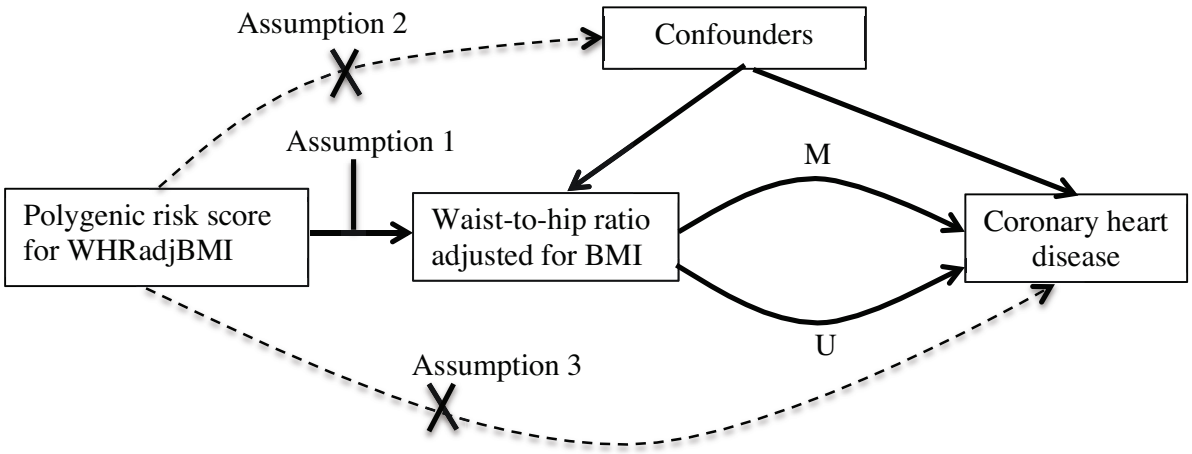



**Figure 1.** Assumptions of a Mendelian randomization analysis. Genetic variants, which are assigned at birth and largely randomly assorted in a population, can be used as instrumental variables to estimate the causal association of an exposure (e.g., WHRadjBMI) with an outcome of interested (e.g., coronary heart disease). This approach rests on three assumptions, denoted with Assumption 1 through Assumption 3 above. First, the genetic variants must be associated with the exposure (Assumption 1). Second, the genetic variants must not be associated with confounders (Assumption 2). Third, the genetic variants must influence risk of the outcome through the exposure and not through other pathways (Assumption 3). Mendelian randomization can be extended to estimate the association of exposure with outcome that is mediated (M in Figure 1) by a given a mediator (e.g., triglycerides) and that is not mediated (U in Figure 1) by that mediator.

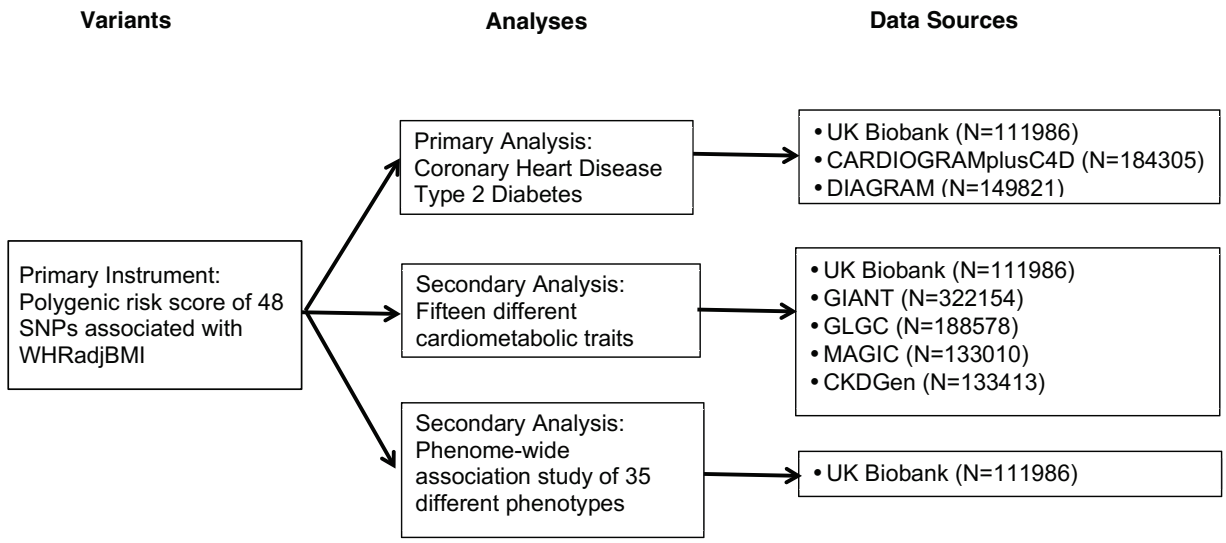



**Figure 2.** Study design. Shown are variants in the primary instrument used to estimate the association of waist-to-hip ratio adjusted for body mass index with cardiometabolic quantitative traits, type 2 diabetes, and coronary heart disease; sources of data for analysis including UK Biobank and publicly-available genome-wide association studies. Abbreviations: WHRadjBMI, waist-to-hip ratio adjusted for body mass index; BMI, body mass index; SNP; single nucleotide polymorphism; CARDIOGRAMplusC4D, Coronary ARtery DIsease Genome-wide Replication and Meta-analysis plus The Coronary Artery Disease Genetics consortium[19]; DIAGRAM, DIAbetes Genetics Replication And Meta-analysis[17]; GIANT, Genetic Investigation of ANthropometric Traits[13,14]; GLGC, Global Lipids Genetics Consortium[11]; MAGIC, Meta-Analyses of Glucose and Insulin-related traits Consortium[12]; CKDGen, Chronic Kidney Disease Genetics Consortium[15].

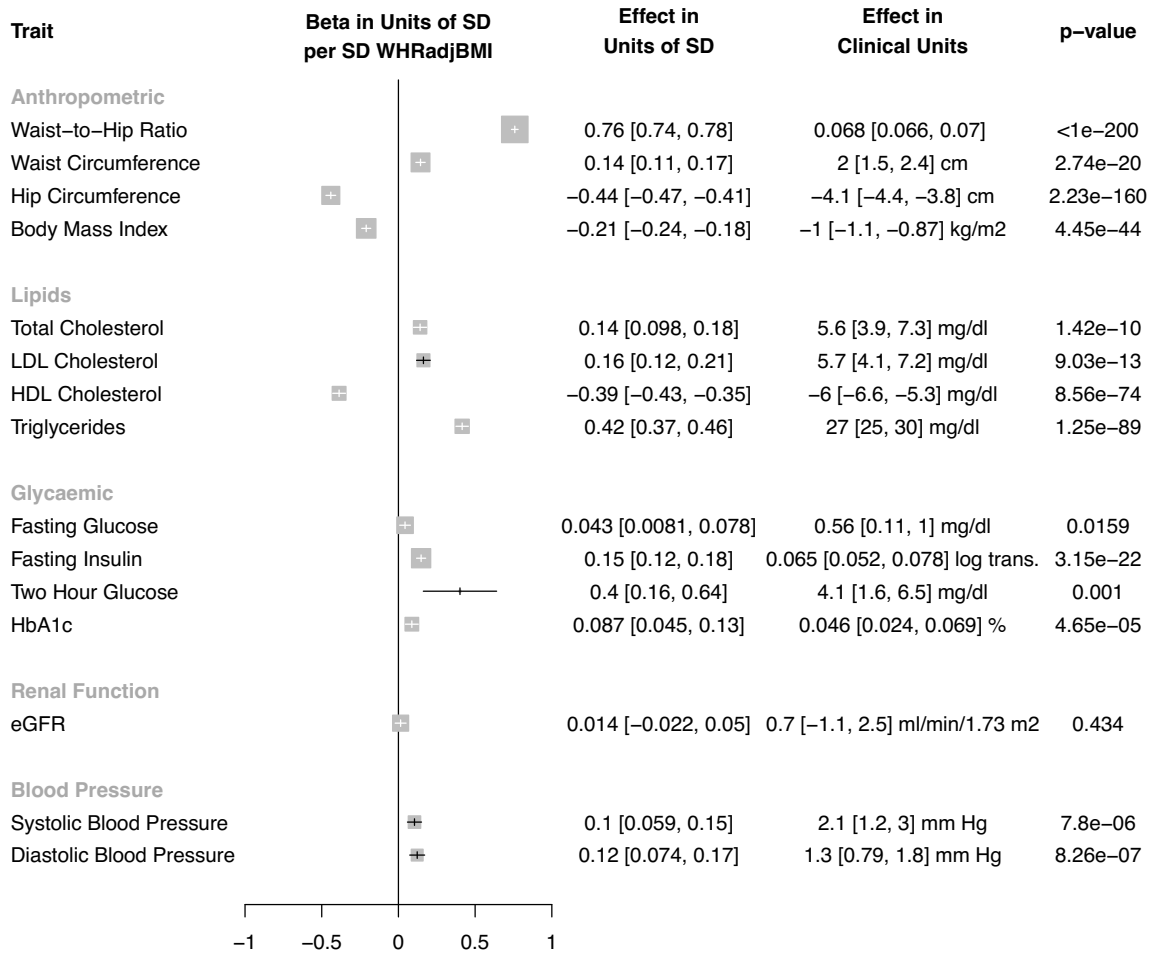



**Figure 3.** Association of 48 SNP polygenic risk score for WHRadjBMI with cardiometabolic quantitative traits. Results are standardized to a one SD increase in WHRadjBMI due to polygenic risk score. For systolic blood pressure, a one SD genetic increase WHRadjBMI is associated with a 2.1 mm Hg higher systolic blood pressure (95% CI 1.2, 3.0), or a 0.1 standard deviation increase in systolic blood pressure (CI 0.059, 0.15). For anthropometric traits, estimates from Genetic Investigation of Anthropometric Traits (GIANT – derived using inverse variance weighted fixed effects meta-analysis[13,14]) were pooled with UK Biobank (derived instrumental variables regression adjusting for age, sex, ten principal components of ancestry and array type) using inverse variance weighted fixed effects meta-analysis. For lipids, glycaemic and renal function traits, estimates were derived from genome-wide association studies (Global Lipids Genetics[11], Meta-analyses of Glucose and Insulin-related Traits[12] and Chronic Kidney Genetics Consortia[15], respectively). For blood pressure, estimates were derived from UK Biobank. Two hour glucose refers to measured blood glucose levels two hour after consumption of dissolved glucose. 95% confidence intervals are reported beside associations. The threshold of significance was p<0.0033 (0.05/15=0.0033). Error bars refer to the 95% confidence interval for each estimate. Size of data marker is inversely proportional to variance of estimate.

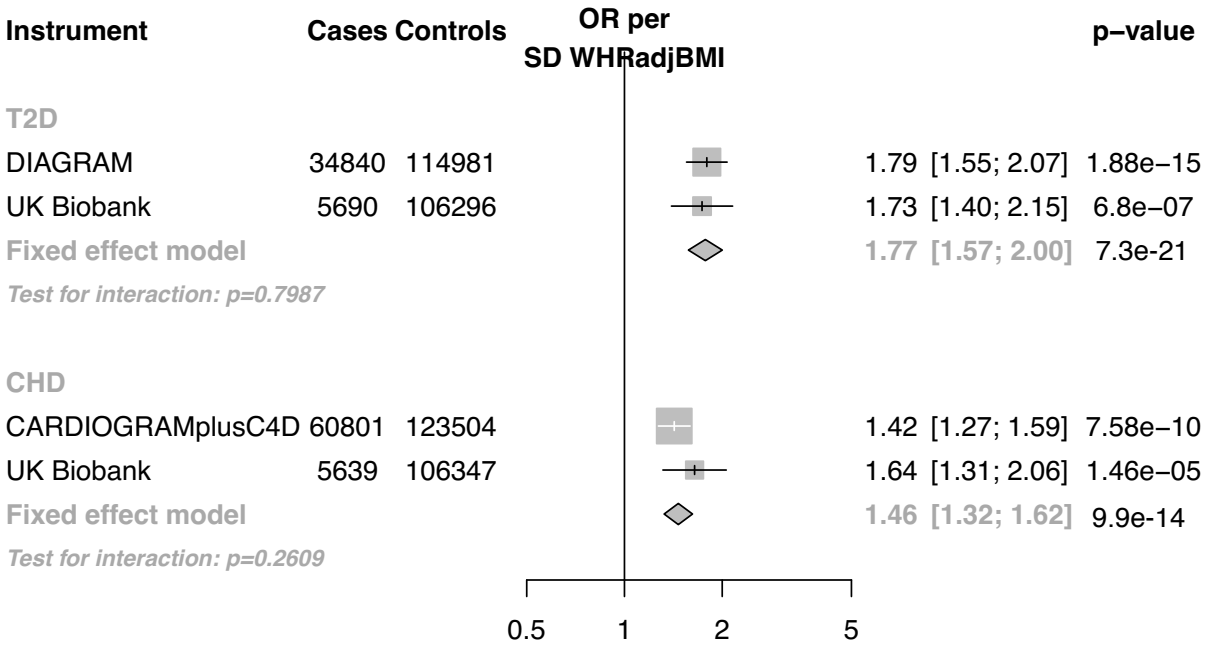



**Figure 4.** Association of 48 SNP polygenic risk score for WHRadjBMI with type 2 diabetes and coronary heart disease. Results are standardized to a one SD increase in WHRadjBMI due to polygenic risk score. Estimates were independently derived in genome-wide association studies (CARDIOGRAMplusC4D for coronary heart disease and DIAGRAM for type 2 diabetes) and UK Biobank. The threshold of significance was $p < 0.025$ (0.05/2=0.025). Error bars refer to the 95% confidence interval for each estimate. Size of data marker is inversely proportional to variance of estimate. Abbreviations: OR, odds ratio; SD, standard deviation; SNPs, single nucleotide polymorphisms; WHRadjBMI, waist-to-hip ratio adjusted for body mass index; CHD, coronary heart disease; T2D, type 2 diabetes; CARDIOGRAMplusC4D, Coronary ARtery DIsease Genome-wide Replication and Meta-analysis plus The Coronary Artery Disease Genetics consortium; DIAGRAM, DIAbetes Genetics Replication And Meta-analysis.

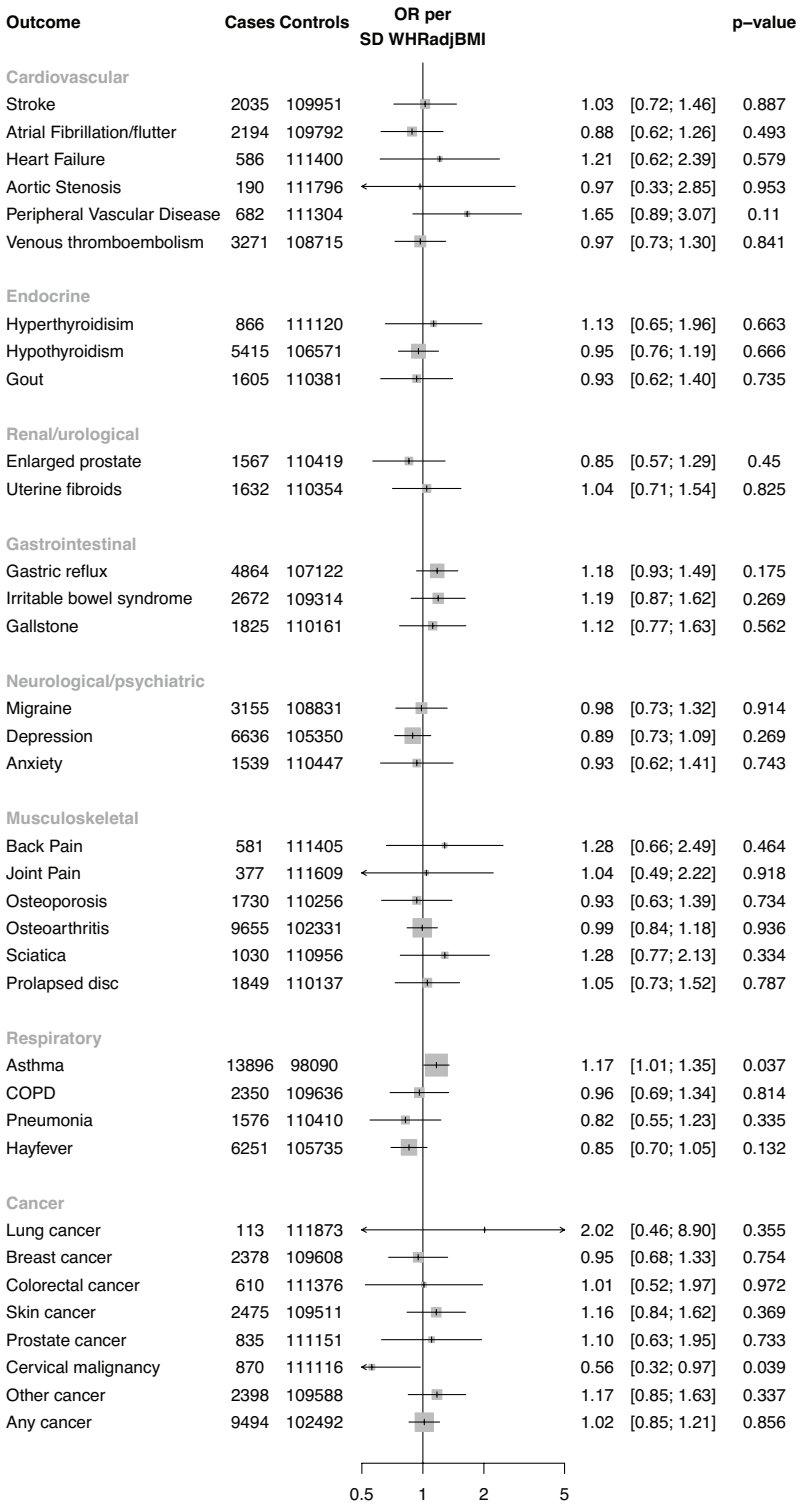



**Figure 5.** A phenome-wide association study testing if 48 SNP polygenic risk score for WHRadjBMI is associated with a range of disease phenotypes. Results are standardized to a one SD increase in WHRadjBMI due to polygenic risk score. All estimates were derived in UK Biobank using instrumental variables regression (adjusting for age, sex and ten principal components of ancestry).

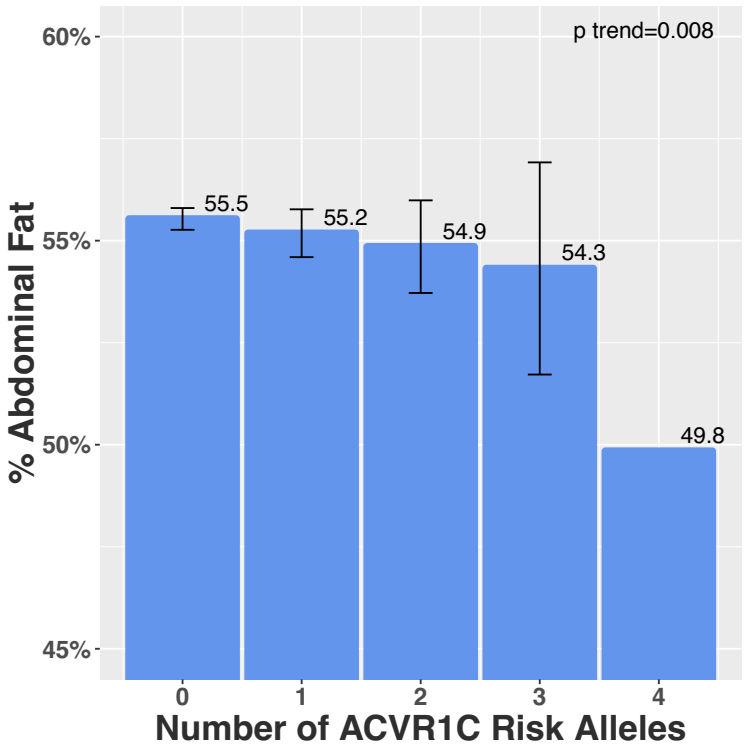



**Figure 6.** Across four *ACVR1C* genetic variants, association of number of WHRadjBMI-lowering alleles with mean directly measured abdominal fat (percent abdominal fat of total body fat) in 4215 participants who underwent dual X-ray absorptiometry scan in UK Biobank, adjusted for age, sex, ten principal components of ancestry and array type. Number of participants in each group is displayed in white for each bar.

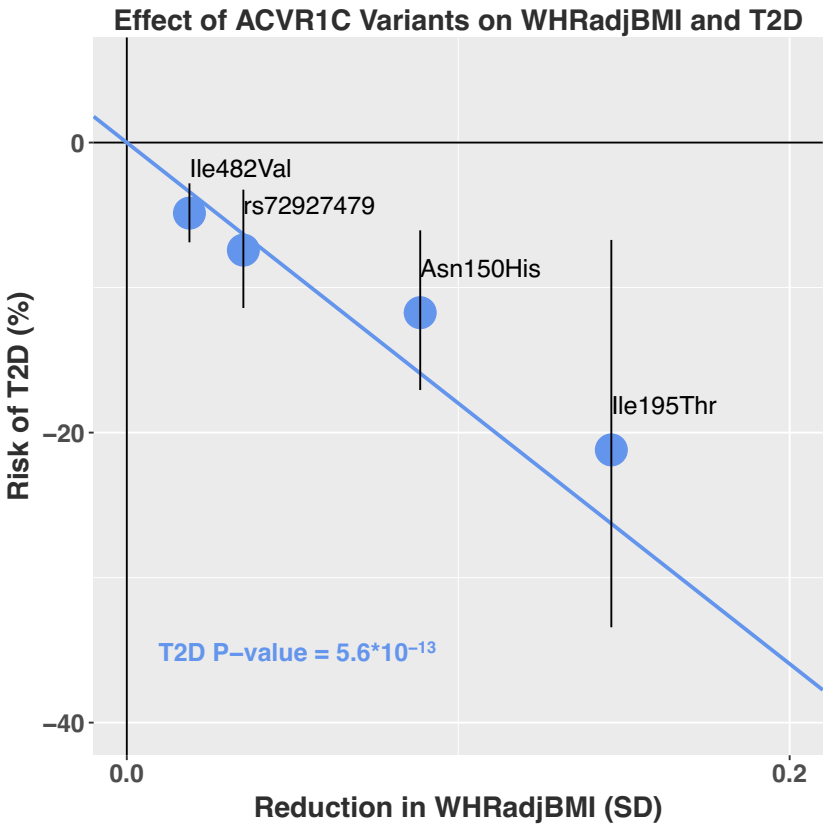



**Figure 7.** Association of four variants in *ACVR1C* with waist-to-hip ratio adjusted for body mass index (WHRadjBMI, x-axis) and type 2 diabetes (T2D, y-axis).

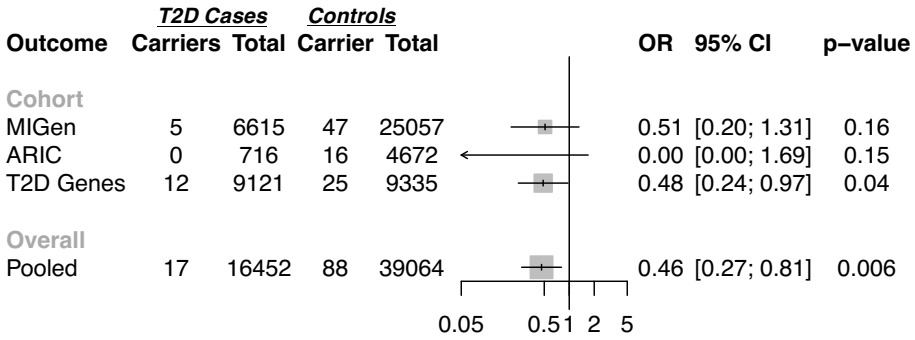



**Figure 8.** Association of predicted damaging variants in *ACVR1C* with type 2 diabetes from sequences in the Myocardial Infarction Genetics Consortium (MIGen), the Atherosclerosis Risk in Communities study (ARIC) and the Type 2 Diabetes Genes Consortium (T2D Genes).

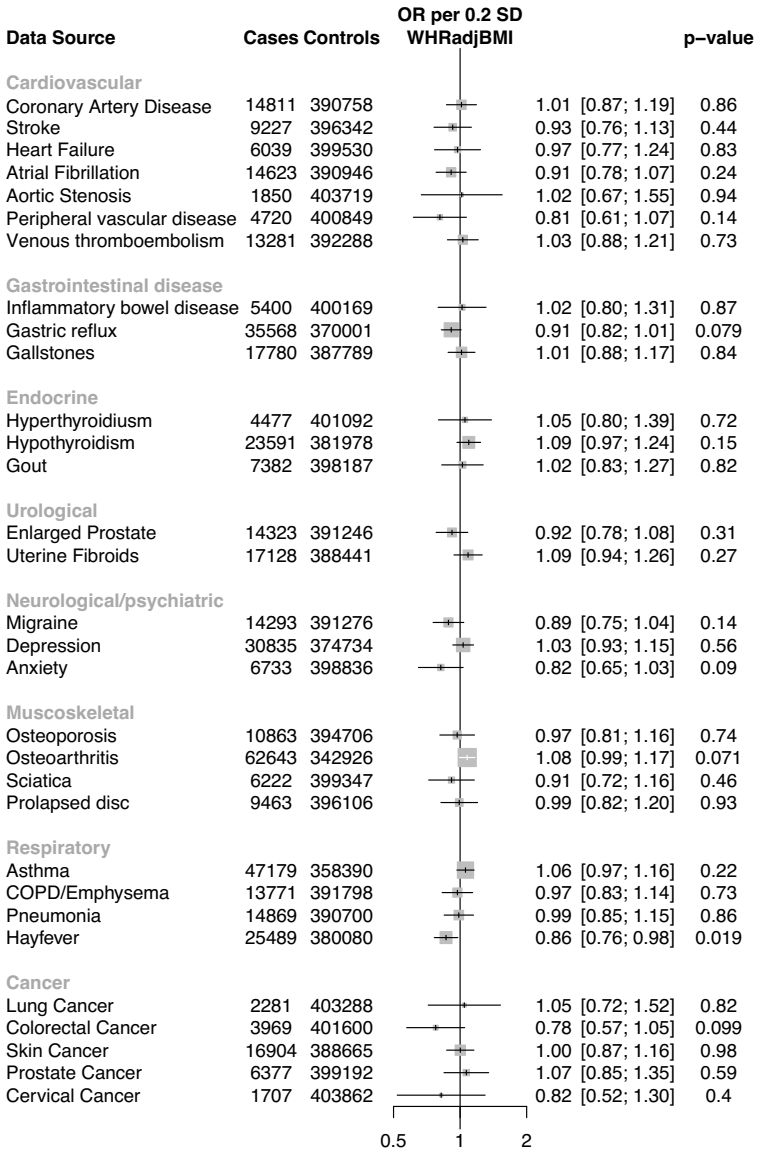



**Figure 9.** Association of *ACVR1C* gene risk score with 31 disease phenotypes in UK Biobank.

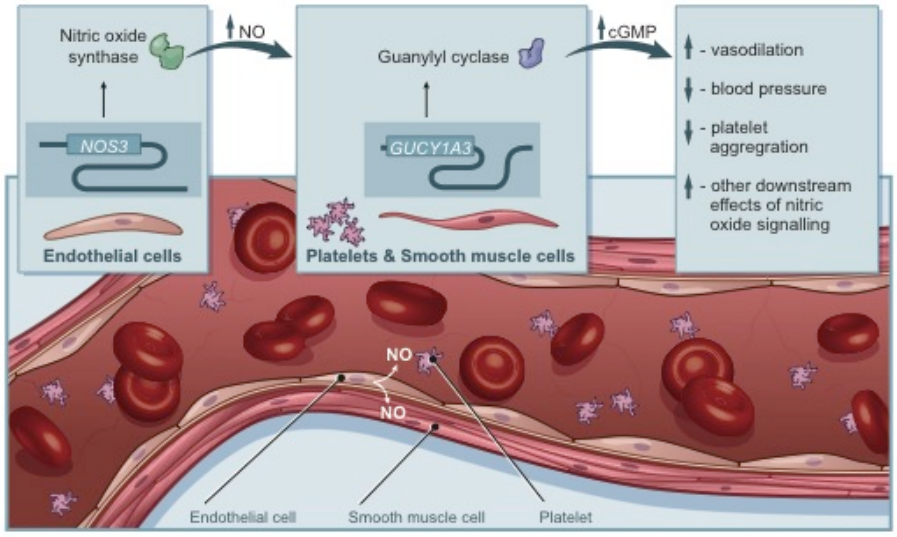



**Figure 10.** Role of *NOS3* and *GUCY1A3* in nitric oxide signaling. Endothelial nitric oxide synthase, encoded by the gene *NOS3*, generates nitric oxide in the vascular endothelium. Nitric oxide acts as a signaling molecule to activate soluble guanylyl cyclase, a heterodimeric protein with one subunit encoded by the gene *GUCY1A3*. Cyclic guanosine monophosphate produced by soluble guanylyl cyclase then activates downstream signaling molecules, leading to vasodilation, blood pressure lowering, inhibition of platelet aggregation and other cardiometabolic effects.

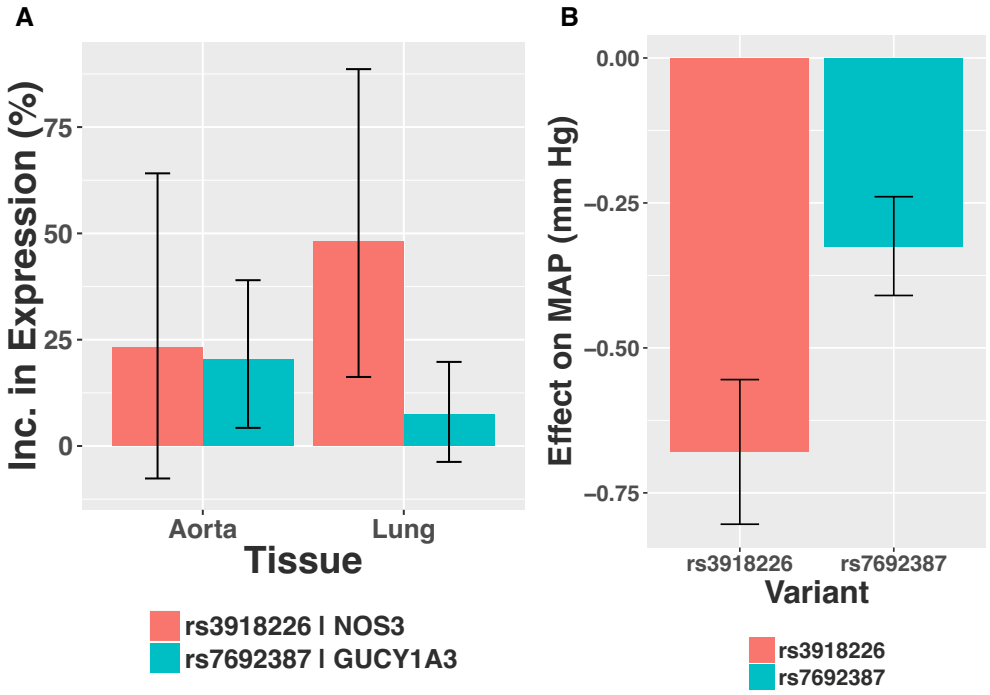



**Figure 11.** Association of rs3918226 and rs7692387 with (A) *NOS3* or *GUCY1A3* expression levels in aortic and lung tissue and (B) mean arterial pressure among UK Biobank participants. Effect of variants on expression levels were obtained from the Genotype-Tissue Expression project. Effect of variants on mean arterial pressure were derived in UK Biobank using linear regression adjusted for age, sex, ten principal components of ancestry and array type (least-squares means estimates). rs3918226 was associated with significantly elevated NOS3 expression in lung. rs7692387 was associated with significantly elevated *GUCY1A3* expression in aorta. MAP, mean arterial pressure.

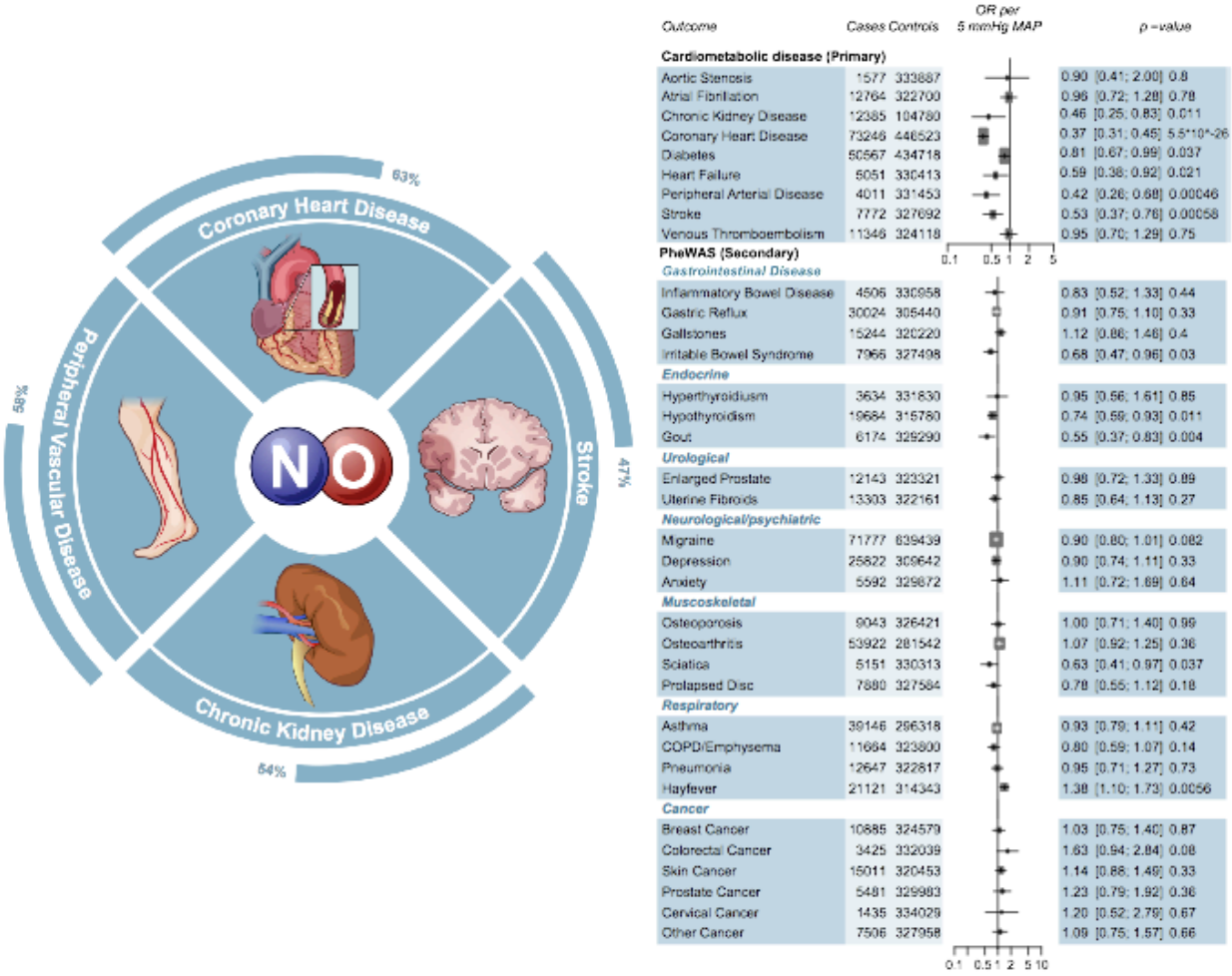



**Figure 12.** Association of the nitric oxide signaling genetic score with cardiometabolic (primary) and other diseases (secondary). Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components and array type, with the exception of chronic kidney disease, which was derived using summary statistics from CKDGen. Estimates for coronary heart disease, diabetes and migraine additionally included summary estimates from CARDIOGRAM, DIAGRAM and IHGC, and were pooled using inverse variance weighted fixed effects meta-analysis. OR, odds ratio; SD, standard deviation; COPD, chronic obstructive pulmonary disease. CARDIOGRAM, Coronary ARtery DIsease Genome wide Replication and Meta-analysis; DIAGRAM, DIAbetes Genetics Replication And Meta-analysis; PheWAS, Phenome Wide Association Study. Significant p-values are bolded.

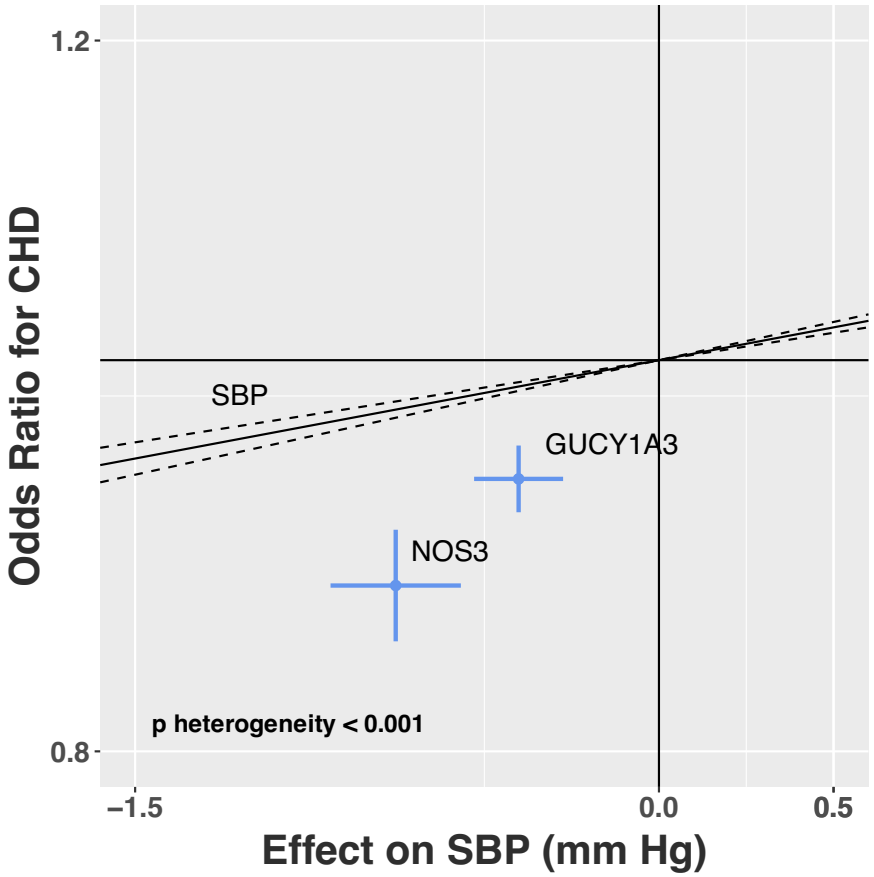
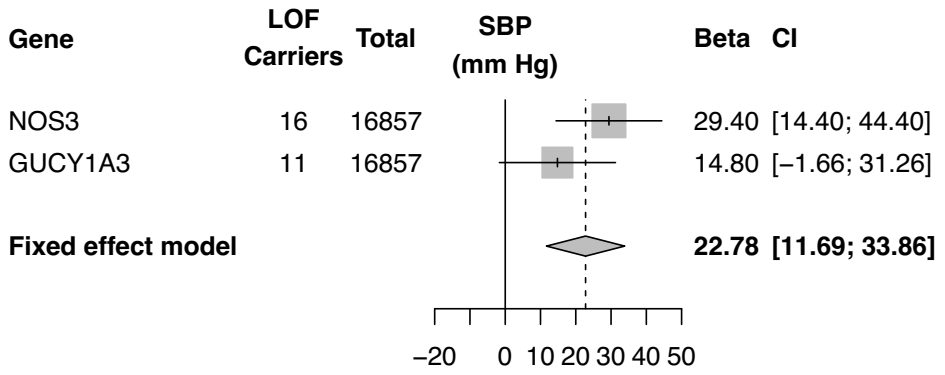



**Figure 13.** Association of common variants in the *NOS3* (rs3918226) and *GUCY1A3* (rs7692387) loci with systolic blood pressure and coronary heart disease. Solid line represents the estimated effect of systolic blood pressure on coronary heart disease from 54 distinct blood pressure loci from GWAS. A test for heterogeneity comparing the association for the genetic nitric oxide score to other blood pressure loci was significant ($p<0.001$). Estimates for systolic blood pressure was derived from UK Biobank with adjustment for age, sex and ten principal components. Estimates for coronary heart disease were derived from inverse variance fixed effects meta-analysis of CARDIOGRAM and UK Biobank. Abbreviations: CHD, coronary heart disease; SBP, systolic blood pressure.

A.




B.

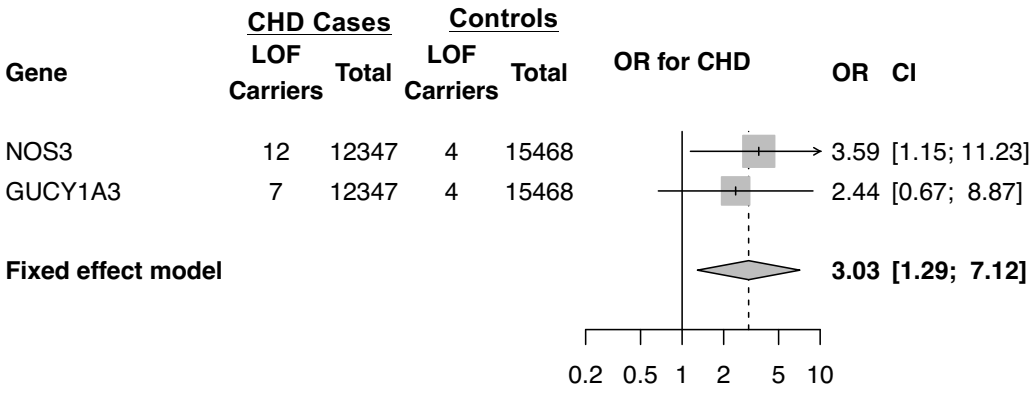
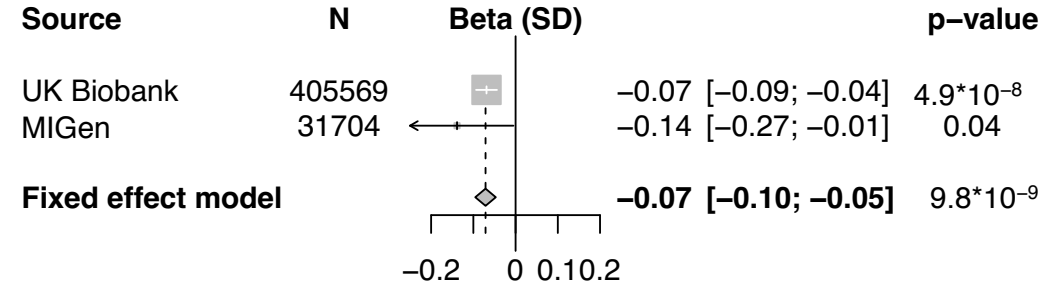
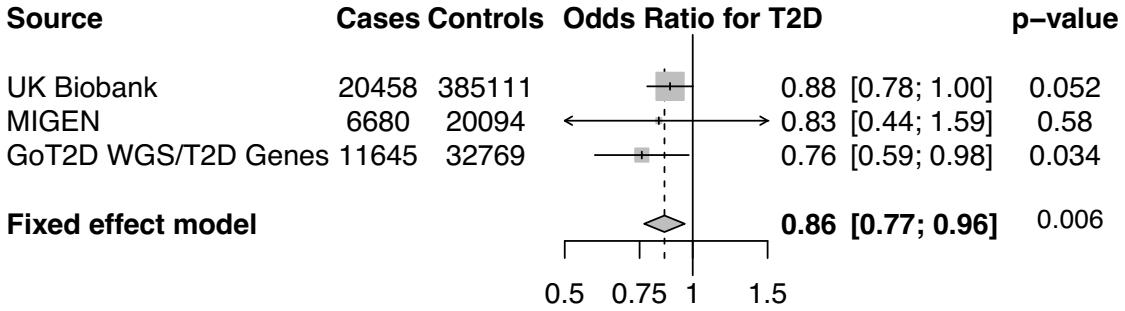



**Figure 14.** Association of rare, predicted loss-of-function variants in the *NOS3-GUCY1A3* nitric oxide signaling pathway with systolic blood pressure and coronary heart disease. (A) Estimates for systolic blood pressure and diastolic blood pressure from T2D GENES study were derived using linear regression with adjustment for five principal components of ancestry. (B) Estimates for CHD from the Myocardial Infarction Genetics Consortium were derived using logistic regression, with adjustment for sex, cohort and five principal components of ancestry. Abbreviations: LOF, Loss-of-function; OR, odds ratio; CHD, coronary heart disease; SBP, systolic blood pressure

A. Body mass index




B. Type 2 diabetes




C. Coronary artery disease

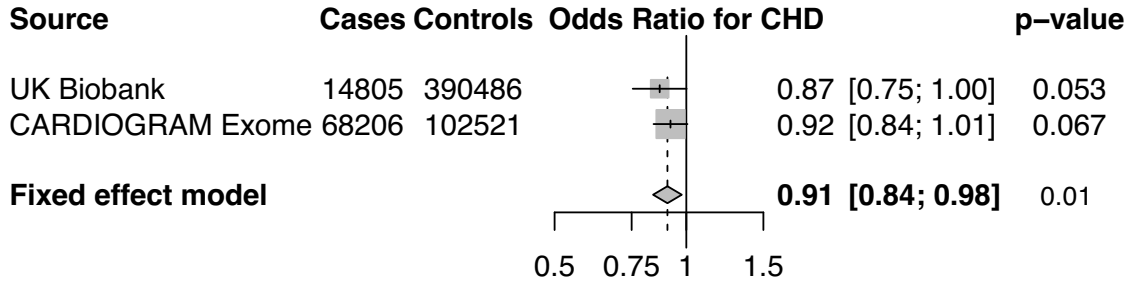



**Figure 15.** Association of a loss of function variant (p.Arg95Ter) in *GPR151* with (A) body mass index, (B) type 2 diabetes and (C) coronary artery disease. Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components of ancestry and array type.

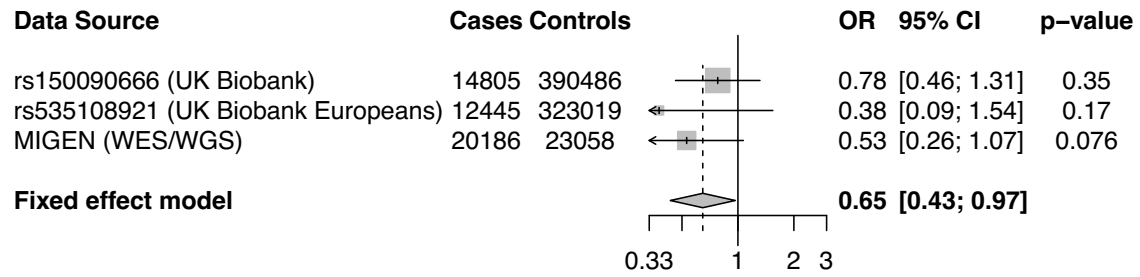



**Figure 16.** Association of predicted loss of function variants in *PDE3B* with coronary artery disease. Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components of ancestry and array type. Estimates were derived in MIGEN (Myocardial Infarction Genetics Consortium) using logistic regression, adjusted for sex and five principal components of ancestry.

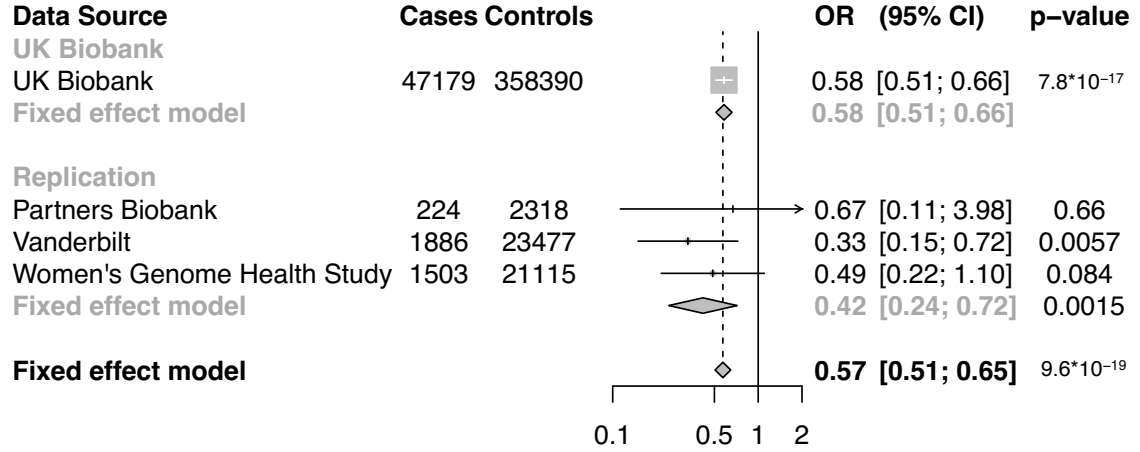



**Figure 17.** Association of *IL33* c.487-1G>C with asthma in UK Biobank, Partners Biobank, Vanderbilt eMERGE network and Women's Genome Health Study. UK Biobank estimates were derived using logistic regression, adjusted for age, sex, ten principal components of ancestry and array type. Partners Biobank and Vanderbilt estimates were derived using logistic regression, adjusted for age, sex and principal components of ancestry. Women's Genome Health Study estimates were derived using logistic regression, adjusted for age and principal components of ancestry.

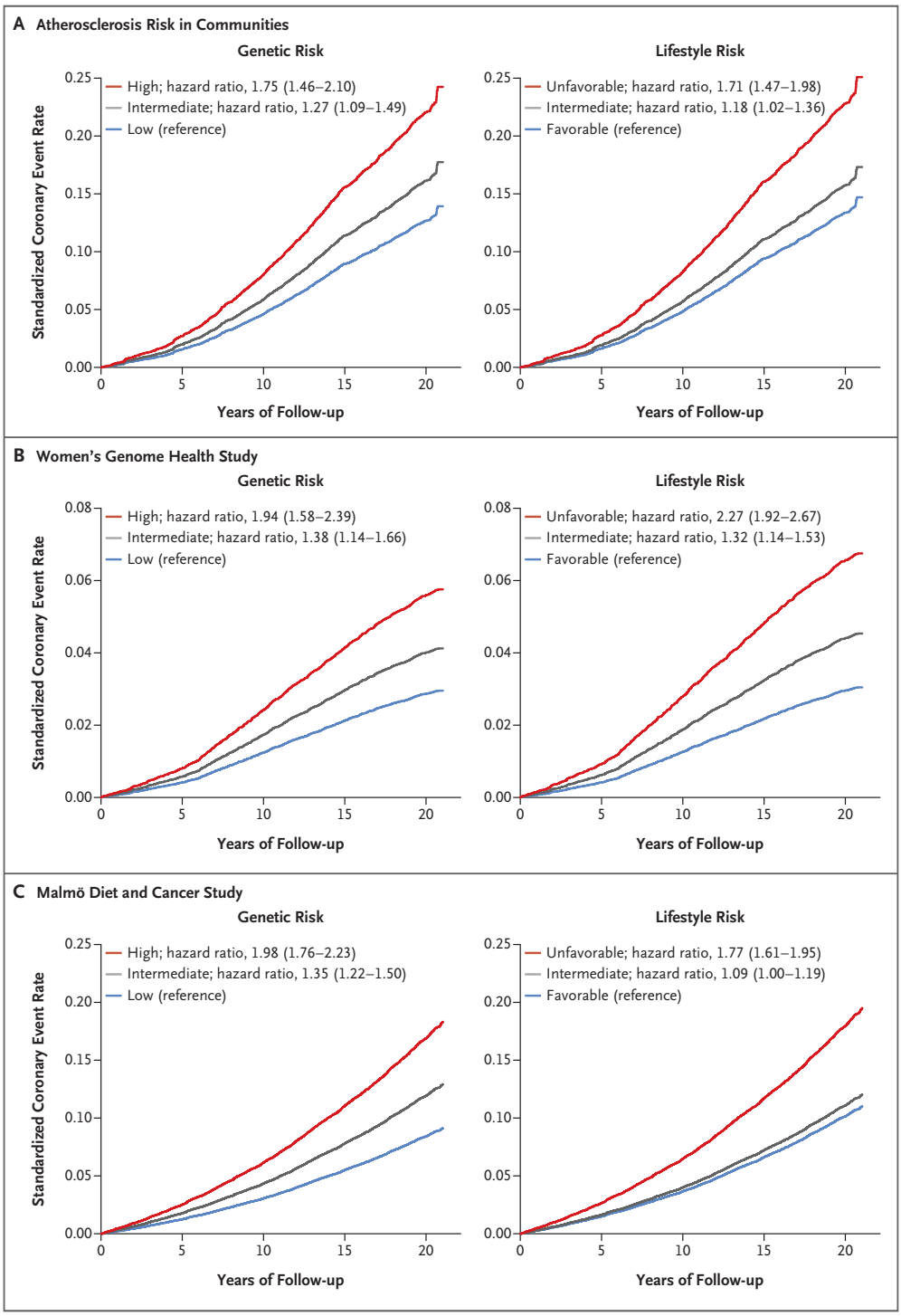



**Figure 18.** Standardized cumulative hazard plots by genetic and lifestyle risk category.

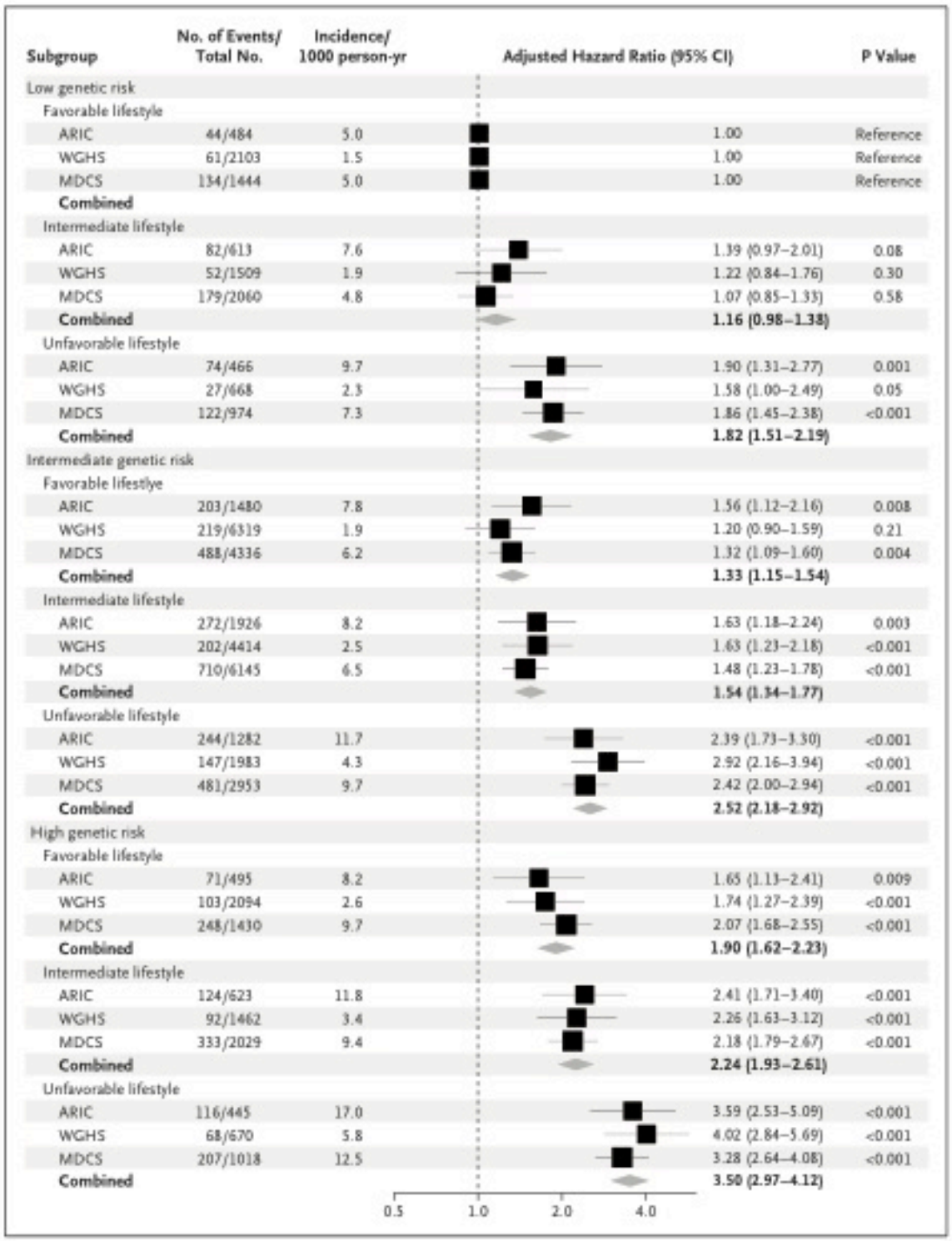



**Figure 19.** Risk of coronary events according to genetic and lifestyle category in prospective cohorts. Cox regression models were adjusted for age, gender, education level, and principal components of ancestry.

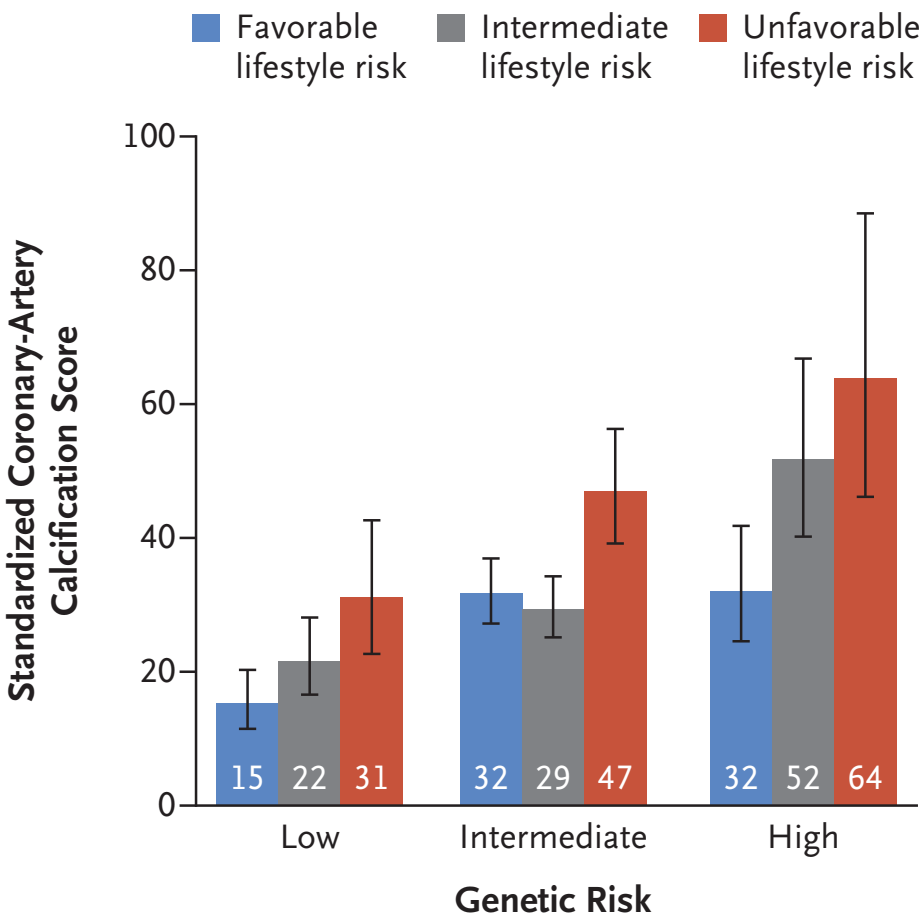



**Figure 20.** Coronary artery calcification score according to subgroups of lifestyle and genetic risk.

## Supplementary Appendix

**Supplementary Table 1.** Association of WHRadjBMI polygenic risk score with potential confounders in UK Biobank.

| | Quartiles of WHRadjBMI Polygenic Risk Score | | | | |
|---|---|---|---|---|---|
| | **Q1** | **Q2** | **Q3** | **Q4** | **Test for Trend** |
| Quartile Range | 0.73 to 1.12 | >1.12 to 1.2 | >1.2 to 1.28 | >1.28 to 1.73 | |
| Number of Participants | 27997 | 27996 | 27996 | 27997 | |
| Current smoker, n (%) | 3426 (12%) | 3366 (12%) | 3368 (12%) | 3451 (12%) | p=0.511 |
| Past smoker, n (%) | 10155 (40%) | 10188 (40%) | 10193 (40%) | 10227 (41%) | p=0.299 |
| Moderate exercise ± SD, mean number of days per week | 3.60 ± 2.33 | 3.63 ± 2.34 | 3.61 ± 2.33 | 3.60 ± 2.35 | p=0.439 |
| Intense exercise ± SD, mean number of days per week | 1.81 ± 1.93 | 1.80 ± 1.96 | 1.80 ± 1.94 | 1.81 ± 1.96 | p=0.936 |
| Daily alcohol consumption, n (%) | 6052 (22%) | 6032 (22%) | 6056 (22%) | 6020 (22%) | p=0.732 |
| Six tablespoons of vegetables or more per day, n(%) | 8175 (32%) | 8203 (32%) | 8322 (32%) | 8201 (32%) | p=0.579 |
| Red meat consumption three or more times per week, n(%) | 6215 (22%) | 6191 (22%) | 6223 (22%) | 6269 (23%) | p=0.612 |
| Breastfed as baby, n(%) | 15117 (71%) | 15052 (71%) | 14976 (71%) | 14971 (71%) | p=0.159 |

Abbreviations: SD = standard deviation

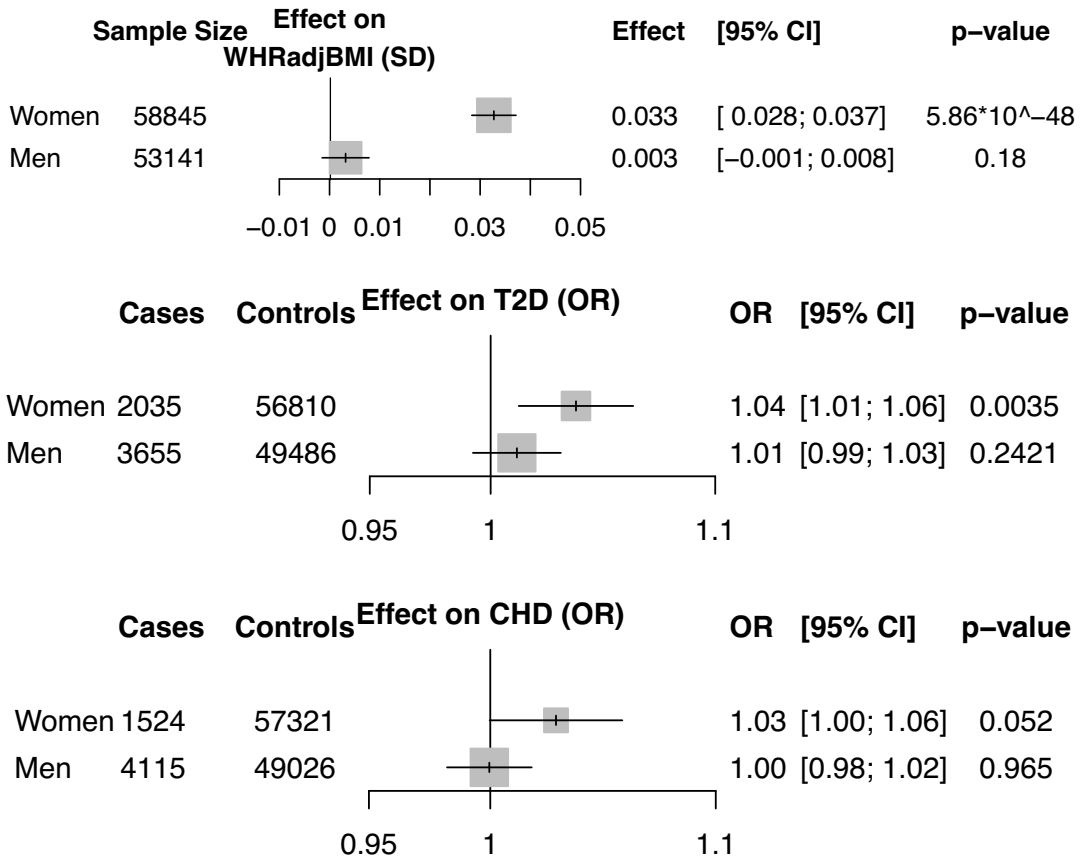
**Supplementary Table 2.** Association of observational WHRadjBMI with potential confounders in UK Biobank.

| | Quartiles of WHRadjBMI | | | | |
|---|---|---|---|---|---|
| | **Q1** | **Q2** | **Q3** | **Q4** | **Test for Trend** |
| Quartile Range | -6.97 to -0.662 | -0.661 to -0.0441 | -0.0442 to 0.626 | 0.627 to 8.2 | |
| Number of Participants | 27997 | 27996 | 27996 | 27997 | |
| Current smoker, n (%) | 2226 (8%) | 2973 (11%) | 3627 (13%) | 4785 (17%) | $p<1*10^{-100}$ |
| Past smoker, n (%) | 9221 (35%) | 9857 (38%) | 10506 (42%) | 11179 (47%) | $p<1*10^{-100}$ |
| Moderate exercise ± SD, mean number of days per week | 3.72 ± 2.32 | 3.68 ± 2.31 | 3.59 ± 2.34 | 3.44 ± 2.38 | $p=2.8*10^{-49}$ |
| Intense exercise ± SD, mean number of days per week | 1.98 ± 1.97 | 1.90 ± 1.95 | 1.77 ± 1.93 | 1.58 ± 1.90 | $p<1*10^{-100}$ |
| Daily alcohol consumption, n (%) | 5530 (20%) | 5965 (21%) | 6185 (22%) | 6480 (23%) | $p=4.0*10^{-24}$ |
| Six tablespoons of vegetables or more per day, n(%) | 8625 (33%) | 8341 (32%) | 8139 (31%) | 7796 (30%) | $p=1.7*10^{-11}$ |
| Red meat consumption three or more times per week, n(%) | 5931 (21%) | 6081 (22%) | 6233 (22%) | 6653 (24%) | $p=1.3*10^{-15}$ |
| Breastfed as baby, n (%) | 15492 (72%) | 15252 (71%) | 14984 (71%) | 14388 (70%) | $p=1.2*10^{-10}$ |

Abbreviations: SD = standard deviations

**Supplementary Table 3.** Frequency of loss of function homozygotes in UK Biobank and in gnoMAD.

| | | Number of Homozygotes | |
|---|---|---|---|
| Gene | Variant | UK Biobank | gnoMAD |
| *GPR151* | rs114285050 | 30 of 405569 | 13 of 138592 |
| *GSDMB* | rs11078928 | 85535 of 405569 | 14909 of 90689 |
| *IL33* | rs146597587 | 5 of 405569 | 1 of 138239 |
| *IFIH1* | rs35732034 | 41 of 405569 | 7 of 131469 |
| *PDE3B* | rs150090666 | 0 of 405569 | 1 of 138583 |




**Supplementary Figure 1.** Per allele association of a sexually dimorphic instrument testing if waist-to-hip ratio adjusted for body mass index associates with type 2 diabetes and/or coronary heart disease. Estimates were derived only in UK Biobank using instrumental variables regression adjusting for age, sex, ten principal components of ancestry and an interaction term between sex and the genetic instrument. Abbreviations: OR, odds ratio; SD, standard deviation; WHRadjBMI, waist-to-hip ratio adjusted for body mass index; CHD, coronary heart disease; T2D, type 2 diabetes.

**Published Manuscripts**

JAMA | Original Investigation

# Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease

Connor A. Emdin, DPhil; Amit V. Khera, MD; Pradeep Natarajan, MD; Derek Klarin, MD; Seyedeh M. Zekavat, BSc; Allan J. Hsiao, MPhil; Sekar Kathiresan, MD

Editorial page 589

Supplemental content

CME Quiz at jamanetworkcme.com

**IMPORTANCE** In observational studies, abdominal adiposity has been associated with type 2 diabetes and coronary heart disease (CHD). Whether these associations represent causal relationships remains uncertain.

**OBJECTIVE** To test the association of a polygenic risk score for waist-to-hip ratio (WHR) adjusted for body mass index (BMI), a measure of abdominal adiposity, with type 2 diabetes and CHD through the potential intermediates of blood lipids, blood pressure, and glycemic phenotypes.

**DESIGN, SETTING, AND PARTICIPANTS** A polygenic risk score for WHR adjusted for BMI, a measure of genetic predisposition to abdominal adiposity, was constructed with 48 single-nucleotide polymorphisms. The association of this score with cardiometabolic traits, type 2 diabetes, and CHD was tested in a mendelian randomization analysis that combined case-control and cross-sectional data sets. Estimates for cardiometabolic traits were based on a combined data set consisting of summary results from 4 genome-wide association studies conducted from 2007 to 2015, including up to 322 154 participants, as well as individual-level, cross-sectional data from the UK Biobank collected from 2007-2011, including 111 986 individuals. Estimates for type 2 diabetes and CHD were derived from summary statistics of 2 separate genome-wide association studies conducted from 2007 to 2015 and including 149 821 individuals and 184 305 individuals, respectively, combined with individual-level data from the UK Biobank.

**EXPOSURES** Genetic predisposition to increased WHR adjusted for BMI.

**MAIN OUTCOMES AND MEASURES** Type 2 diabetes and CHD.

**RESULTS** Among 111 986 individuals in the UK Biobank, the mean age was 57 (SD, 8) years, 58 845 participants (52.5%) were women, and mean WHR was 0.875. Analysis of summary-level genome-wide association study results and individual-level UK Biobank data demonstrated that a 1-SD increase in WHR adjusted for BMI mediated by the polygenic risk score was associated with 27-mg/dL higher triglyceride levels, 4.1-mg/dL higher 2-hour glucose levels, and 2.1-mm Hg higher systolic blood pressure (each $P < .001$). A 1-SD genetic increase in WHR adjusted for BMI was also associated with a higher risk of type 2 diabetes (odds ratio, 1.77 [95% CI, 1.57-2.00]; absolute risk increase per 1000 participant-years, 6.0 [95% CI, CI, 4.4-7.8]; number of participants with type 2 diabetes outcome, 40 530) and CHD (odds ratio, 1.46 [95% CI, 1.32-1.62]; absolute risk increase per 1000 participant-years, 1.8 [95% CI, 1.3-2.4]; number of participants with CHD outcome, 66 440).

**CONCLUSIONS AND RELEVANCE** A genetic predisposition to higher waist-to-hip ratio adjusted for body mass index was associated with increased risk of type 2 diabetes and coronary heart disease. These results provide evidence supportive of a causal association between abdominal adiposity and these outcomes.

*JAMA*. 2017;317(6):626-634. doi:10.1001/jama.2016.21042

**Author Affiliations:** Center for Genomic Medicine and Cardiology Division, Massachusetts General Hospital, Harvard Medical School, Boston (Emdin, Khera, Natarajan, Klarin, Zekavat, Kathiresan); Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts (Emdin, Khera, Natarajan, Klarin, Zekavat, Kathiresan); Department of Economics, Massachusetts Institute of Technology, Cambridge (Hsiao).

**Corresponding Author:** Sekar Kathiresan, MD, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge St, CPZN 5.830, Boston, MA 02114 (skathiresan1@mgh.harvard.edu).

Obesity, typically defined on the basis of body mass index (BMI), is a leading cause of type 2 diabetes and coronary heart disease (CHD) in the population.[1,2] However, for any given BMI, body fat distribution can vary substantially; some individuals store proportionally more fat around their visceral organs (abdominal adiposity) than on their thighs and hip.[3] Waist-to-hip ratio (WHR) adjusted for BMI is a surrogate measure of abdominal adiposity and has been correlated with direct imaging assessments of abdominal fat.[4,5]

In observational studies, abdominal adiposity has been associated with cardiometabolic disease[6,7]; however, whether this association is causal remains unclear. For example, unmeasured lifestyle factors[8] might confound observational studies that link WHR adjusted for BMI with type 2 diabetes and CHD. Furthermore, reverse causality could similarly lead to a statistically robust but noncausal relationship. For example, individuals with subclinical CHD might develop abdominal adiposity because of an inability to exercise.

Mendelian randomization is a human genetics tool that leverages the random assortment of genetic variants at time of conception to facilitate causal inference.[9] Because genetic predisposition to abdominal adiposity is determined by DNA sequence variants, it is less likely to be affected by confounding or reverse causality. In this study, a mendelian randomization approach was used to determine whether a genetic predisposition to increased WHR adjusted for BMI is associated with cardiometabolic quantitative traits, type 2 diabetes, and CHD.

### Key Points

**Question** Is genetic evidence consistent with a causal relationship among waist-to-hip ratio adjusted for body mass index (a measure of abdominal adiposity), type 2 diabetes, and coronary heart disease?

**Findings** In this mendelian randomization study, a polygenic risk score for increased waist-to-hip ratio adjusted for body mass index was significantly associated with adverse cardiometabolic traits and higher risks for both type 2 diabetes and coronary heart disease.

**Meaning** These results provide evidence supportive of a causal association between abdominal adiposity and the development of type 2 diabetes and coronary heart disease.

## Methods

### Study Design and Instruments

Observational epidemiology studies test association of an exposure (eg, WHR adjusted for BMI) with an outcome (eg, CHD). However, unobserved confounders may affect both exposure and outcome, thus biasing the observed association (**Figure 1**; eMethods A in the Supplement). Because genetic variants are both randomly assorted in a population and assigned at conception, they are largely unassociated with confounders and can be used as instrumental variables to estimate the causal association of an exposure (WHR adjusted for BMI) with an outcome.[9]

This mendelian randomization approach has 3 assumptions.[10] First, genetic variants used as an instrument must be associated with the exposure of interest (eg, WHR adjusted for BMI) (assumption 1 in Figure 1). Second, genetic variants must not be associated with confounders (assumption 2 in Figure 1). Third, genetic variants must not be associated with outcome independently of the exposure (assumption 3 in Figure 1). The second and third assumptions are collectively known as independence from pleiotropy. Mendelian randomization can be extended to conduct a mediation analysis, estimating the proportion of an observed association of an exposure (WHR adjusted for BMI) with an outcome (CHD) that occurs through a given mediator (Figure 1).

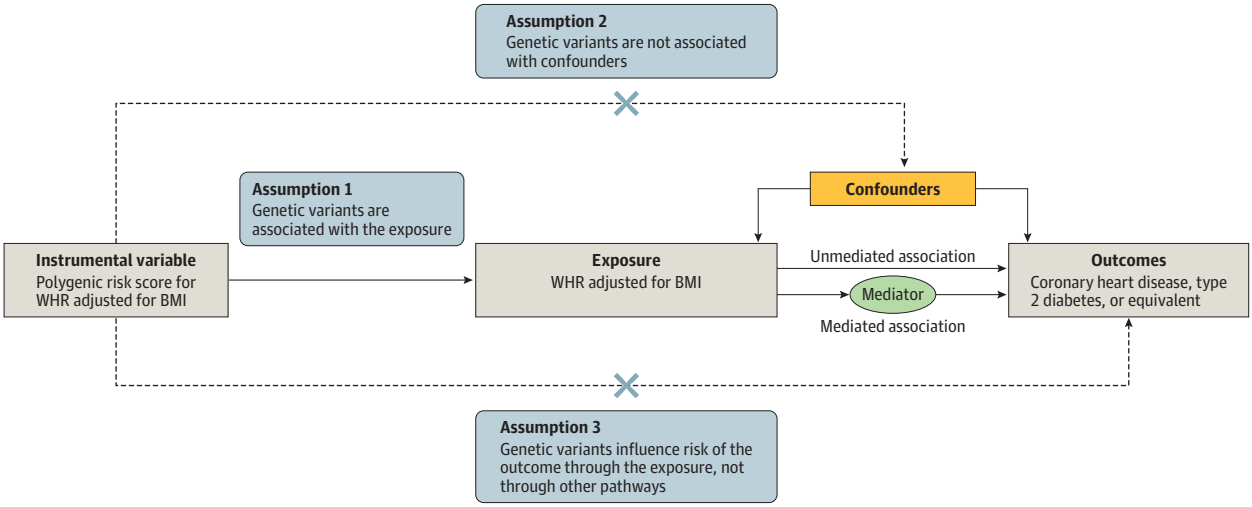
A mendelian randomization study using publicly available, summary-level data from large-scale genome-wide association studies (GWASs) (both cross-sectional and case-control data sets) as well as individual-level data from the UK Biobank (a cross-sectional data set) was conducted (**Figure 2**).[12-18] The primary instrument was a polygenic risk score for WHR adjusted for BMI. A recent large-scale GWAS from the Genome-Wide Investigation of Anthropometric Traits (GIANT) Consortium identified 48 single-nucleotide polymorphisms (SNPs), or genetic variants, associated with WHR adjusted for BMI (eTable 1 in the Supplement).[14] Combining these 48 SNPs into a weighted polygenic risk score enabled quantification of the genetic predisposition to increased WHR adjusted for BMI for each individual.

### Data Sources and Study Participants

Summary-level data from 6 GWAS consortia were used (GWASs conducted from 2007 to 2015) (eTable 3; eMethods B in the Supplement).[12-18] For WHR, BMI, waist circumference, hip circumference, and WHR adjusted for BMI, data from the GIANT Consortium was used (GWASs conducted from 2007 to 2013)[14,15]; this study included 322 154 individuals of European descent for BMI and 210 088 individuals of European descent for waist circumference, hip circumference, WHR, and WHR adjusted for BMI. The results from 5 additional GWAS (conducted from 2007 to 2015) examining blood lipids, glycemic traits, renal function, type 2 diabetes, and CHD, and predominantly including individuals of European descent, were also assessed.[11,13,16,17,19,20] Summary results for type 2 diabetes and CHD were derived from studies of 149 821 individuals (Diabetes Genetics Replication and Meta-analysis [DIAGRAM][13]) and 184 305 individuals (Coronary Artery Disease Genome-Wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics Consortium [CARDIOGRAMplusC4D][11]), respectively. Informed consent was obtained from all participants of contributing studies. Contributing studies received ethical approval from their respective institutional review boards.

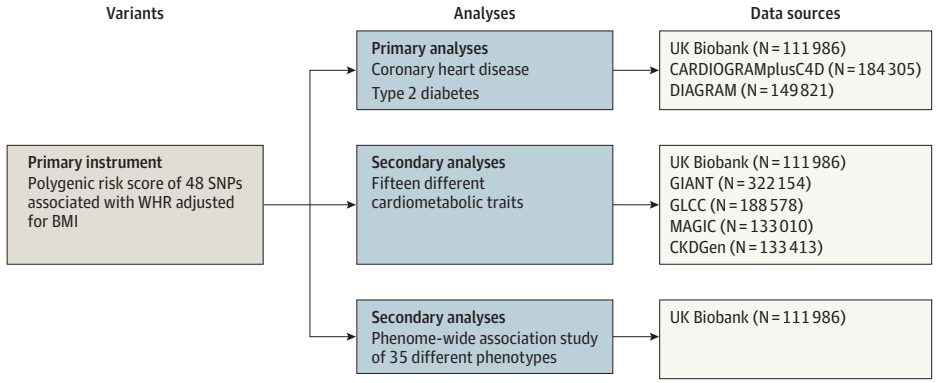
Individual-level data from 111 986 individuals of European ancestry from the UK Biobank, collected from 2007 to 2011, were also used (**Table**; eMethods C in the Supplement). The UK Biobank received ethical approval from the research ethics committee (reference number 11/NW/0382). Analysis of the UK Biobank data was approved by the Partners Health Care institutional review board (protocol 2013P001840). Informed consent was obtained from all participants by the UK Biobank.

Figure 1. Assumptions of a Mendelian Randomization Analysis




Genetic variants, which are assigned at birth and largely randomly assorted in a population, can be used as instrumental variables to estimate the causal association of an exposure (eg, waist-to-hip ratio [WHR] adjusted for body mass index [BMI]) with an outcome of interest (eg, coronary heart disease). This approach rests on 3 assumptions. First, the genetic variants must be associated with the exposure (assumption 1). Second, the genetic variants must not be associated with confounders (assumption 2). Third, the genetic variants must influence risk of the outcome through the exposure and not through other pathways (assumption 3). Mendelian randomization can be extended to estimate the association of exposure with outcome that is mediated by a given a mediator (eg, triglycerides).

Figure 2. Study Design




A polygenic score of 48 single-nucleotide polymorphisms was used as an instrument to estimate the causal association of waist-to-hip ratio (WHR) adjusted for body mass index (BMI) with cardiometabolic quantitative traits, type 2 diabetes, and coronary heart disease; sources of data for analysis included the UK Biobank and publicly available genome-wide association studies. CARDIOGRAMplusC4D indicates Coronary Artery DIsease Genome-wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics Consortium[11]; CKDGen, Chronic Kidney Disease Genetics Consortium[12]; DIAGRAM, Diabetes Genetics Replication and Meta-analysis[13]; GIANT, Genetic Investigation of Anthropometric Traits[14,15]; GLGC, Global Lipids Genetics Consortium[16]; MAGIC, Meta-analyses of Glucose and Insulin-Related Traits Consortium[17]; SNP, single-nucleotide polymorphism.

WHR adjusted for BMI was derived in the UK Biobank through inverse normal transformation of WHR after adjustment for age, sex, and BMI (as in the GIANT Consortium[14]). Type 2 diabetes and CHD were both ascertained at baseline by self-report, followed by a verbal interview with a trained nurse to confirm the diagnosis (eTable 4 in the Supplement). Type 2 diabetes was defined as report of type 2 diabetes, report of type 2 diabetes unspecified, or current use of insulin medication. CHD was defined as report of previous myocardial infarction or diagnosis of angina or hospitalization for myocardial infarction (*International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* codes I21-I23).

In addition to the primary outcomes of type 2 diabetes and CHD, a phenome-wide association study (an analysis of the association of a genetic variant or polygenic risk score with a broad range of diseases, outcomes, or both) for 35 additional diseases, including endocrine, renal, urologic, gastrointestinal, neurologic, musculoskeletal, respiratory, and cancer disorders, was conducted in the UK Biobank to attempt to iden-

tify whether the polygenic risk score for WHR adjusted for BMI is associated with any additional disorders (eTable 4 in the Supplement).

## Statistical Analysis

For analyses of both summary-level data and UK Biobank data, a weighted polygenic risk score was derived based on the magnitude of association of each SNP with WHR adjusted for BMI in the previously published GIANT analysis.[19] The association of polygenic risk score with each continuous trait and dichotomous outcome was then calculated after standardization to a 1-SD predicted change in WHR adjusted for BMI.

For the summary-level data, this approach is equivalent to an inverse-variance-weighted fixed-effects meta-analysis of the association of each SNP with the trait or outcome of interest (eg, CHD), divided by the association of each SNP with WHR adjusted for BMI.[21] Explicitly, if $x$ is the association of each SNP with the outcome of interest, and $w$ the association of each SNP with WHR adjusted for BMI, then the estimated genetic association of WHR adjusted for BMI with the outcome was calculated as a fixed-effects meta-analysis of $x/w$ for all SNPs.

To validate that the polygenic risk score for WHR adjusted for BMI was a strong instrument for WHR adjusted for BMI (assumption 1 in Figure 1), an F statistic for the instrument was calculated in the UK Biobank. An F statistic is a measure of the significance of an instrument (the polygenic risk score) for prediction of the exposure (WHR adjusted for BMI), controlling for additional covariates (age, sex, 10 principal components of ancestry, and a dummy variable for the array type used in genotyping). An F statistic greater than 10 is evidence of a strong instrument.[22]

For individual-level data from the UK Biobank, logistic regression was used to determine association of a polygenic risk score for WHR adjusted for BMI and dichotomous outcomes (type 2 diabetes, CHD, and 35 additional diseases) (eMethods C in the Supplement).[23] Linear regression was used for continuous traits (anthropometric traits and blood pressure) in the UK Biobank. All UK Biobank analyses included adjustment for age, sex, 10 principal components of ancestry, and a dummy variable for the array type used in genotyping. The inclusion of principal components of ancestry as covariates is commonly implemented to correct for population stratification according to ancestral background.[24]

To test assumption 2 (independence of polygenic risk score for WHR adjusted for BMI from potential confounders) (Figure 1), the relationship of the polygenic risk score to smoking, alcohol use, physical activity, vegetable consumption, red meat consumption, and breastfeeding status as a child was determined among individuals in the UK Biobank. Test for trend was performed across quartiles of the polygenic risk score for WHR adjusted for BMI using logistic regression, with each potential confounder as the outcome. For comparison, individuals in the UK Biobank were stratified into quartiles by observational WHR adjusted for BMI and test for trend performed using logistic regression.

Five additional sensitivity analyses were conducted to test the robustness of the results (eMethods D in the Supplement).

Table. Characteristics of UK Biobank Participants

| Characteristic | No. (%) or Mean (SD) |
|---|---|
| No. Individuals | 111 986 |
| Age, mean (SD), y | 56.9 (7.9) |
| Men, No. (%) | 53 141 (47.5) |
| UK BiLEVE array, No. (%)[a] | 38 505 (34.4) |
| Blood pressure, mean (SD), mm Hg[b] | |
| Systolic | 143.6 (21.8) |
| Diastolic | 84.5 (11.8) |
| Body mass index, mean (SD)[c] | 27.5 (4.8) |
| Waist-to-hip ratio, mean (SD) | 0.875 (0.09) |
| Diabetes mellitus, No. (%) | 5690 (5.1) |
| Coronary heart disease, No. (%) | 5639 (5.0) |

[a] Participants genotyped using the UK BiLEVE array rather than the UK Biobank Axiom array.

[b] Baseline blood pressure was missing for 7681 individuals. Reported measurements are after adjustment for treatment (addition of 15 mm Hg to systolic blood pressure and 10 mm Hg to diastolic blood pressure).

[c] Calculated as weight in kilograms divided by height in meters squared.

Three additional polygenic risk scores were used, including one that included variants not significantly associated with BMI, a second that included variants significantly associated with gene expression in adipose tissue, and a third that included variants significantly associated with increased WHR adjusted for BMI in women but not in men. The association of genetic variants with BMI was adjusted for, and median regression was used (eMethods D in the Supplement).[10] The rationale for these sensitivity analyses is provided in eMethods D.

Absolute increases associated with WHR adjusted for BMI for type 2 diabetes and CHD were calculated using the United States population incidence of type 2 diabetes and CHD (eMethods E in the Supplement). Tests for nonlinear associations of a genetic predisposition to increased WHR adjusted for BMI with type 2 diabetes and CHD were performed using nonlinear instrumental variable estimation (eMethods F in the Supplement).[25]

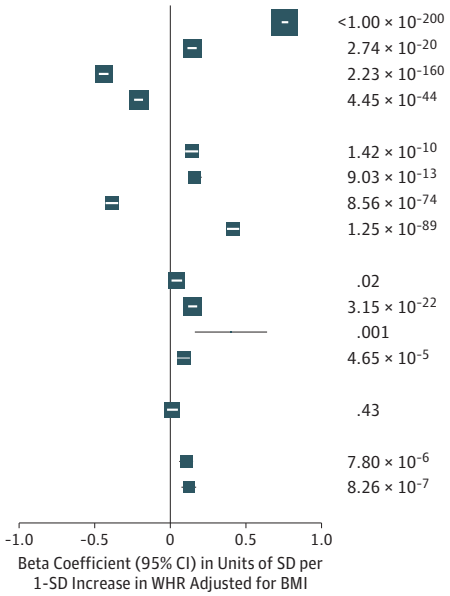
The threshold of statistical significance for type 2 diabetes and CHD (main outcome measures) was $P < .025$ (.05/2 = .025). The threshold of significance for the analysis of 15 traits was $P < .0033$ (.05/15 = .0033). The threshold of significance in the phenome-wide association analysis was $P < .0014$ (.05/35 = .0014).

## Mediation Analysis

Among continuous traits, the polygenic risk score for WHR adjusted for BMI was most strongly associated with plasma triglyceride levels. The extent to which the polygenic risk score association with CHD was mediated by plasma triglycerides was tested using mediation analysis, conducted post hoc after triglyceride level was identified as the cardiometabolic trait most strongly associated with WHR adjusted for BMI. An estimate of the genetic association of triglyceride level on CHD risk, previously derived by Do et al[26] (odds ratio [OR], 1.52 per 1-SD increase in triglyceride level),[26] was used to calculate the predicted magnitude of increased CHD risk based on the observed association of the WHR adjusted for

Figure 3. Association of 48-SNP Polygenic Risk Score for WHR Adjusted for BMI With Cardiometabolic Quantitative Traits

| Trait | Association in Clinical Units (95% CI)[a] | Association in Units of SD (95% CI) | P Value |
|---|---|---|---|
| Anthropometric | | | |
| WHR | 0.068 (0.066 to 0.07) | 0.76 (0.74 to 0.78) | $<1.00 \times 10^{-200}$ |
| Waist circumference, cm | 2 (1.5 to 2.4) | 0.14 (0.11 to 0.17) | $2.74 \times 10^{-20}$ |
| Hip circumference, cm | -4.1 (-4.4 to -3.8) | -0.44 (-0.47 to -0.41) | $2.23 \times 10^{-160}$ |
| BMI[b] | -1.0 (-1.1 to -0.9) | -0.21 (-0.24 to -0.18) | $4.45 \times 10^{-44}$ |
| Lipids | | | |
| Total cholesterol, mg/dL | 5.6 (3.9 to 7.3) | 0.14 (0.10 to 0.18) | $1.42 \times 10^{-10}$ |
| LDL-C, mg/dL | 5.7 (4.1 to 7.2) | 0.16 (0.12 to 0.21) | $9.03 \times 10^{-13}$ |
| HDL-C, mg/dL | -6.0 (-6.6 to -5.3) | -0.39 (-0.43 to -0.35) | $8.56 \times 10^{-74}$ |
| Triglycerides, mg/dL | 27 (25 to 30) | 0.42 (0.37 to 0.46) | $1.25 \times 10^{-89}$ |
| Glycemic | | | |
| Fasting glucose, mg/dL | 0.56 (0.11 to 1.0) | 0.04 (0.01 to 0.08) | .02 |
| Fasting insulin, log (pmol/L) | 0.07 (0.05 to 0.08) | 0.15 (0.12 to 0.18) | $3.15 \times 10^{-22}$ |
| Two-hour glucose, mg/dL | 4.1 (1.6 to 6.5) | 0.40 (0.16 to 0.64) | .001 |
| $HbA_{1c}$, % | 0.05 (0.02 to 0.07) | 0.09 (0.05 to 0.13) | $4.65 \times 10^{-5}$ |
| Renal function | | | |
| eGFR, mL/min/1.73 $m^2$ | 0.70 (−1.1 to 2.5) | 0.01 (-0.02 to 0.05) | .43 |
| Blood pressure | | | |
| Systolic, mm Hg | 2.1 (1.2 to 3.0) | 0.10 (0.06 to 0.15) | $7.80 \times 10^{-6}$ |
| Diastolic, mm Hg | 1.3 (0.8 to 1.8) | 0.12 (0.07 to 0.17) | $8.26 \times 10^{-7}$ |




Results are standardized to a 1-SD increase in waist-to-hip ratio (WHR) adjusted for body mass index (BMI) due to polygenic risk score. For systolic blood pressure, a 1-SD genetic increase in WHR adjusted for BMI is associated with a 2.1-mm Hg higher systolic blood pressure (95% CI, 1.2-3.0) or a 0.1-SD increase in systolic blood pressure (95% CI, 0.059-0.15). For anthropometric traits, estimates from Genetic Investigation of Anthropometric Traits (GIANT) derived using inverse variance–weighted fixed-effects meta-analysis[14,15]) were pooled with data from the UK Biobank (derived instrumental variables regression adjusting for age, sex, 10 principal components of ancestry, and array type) using inverse variance–weighted fixed-effects meta-analysis. For lipids, glycemic, and renal function traits, estimates were derived from genome-wide association studies (Global Lipids Genetics,[16] Meta-analyses of Glucose and Insulin-Related Traits,[17] and Chronic Kidney Genetics Consortia,[12] respectively). For blood pressure, estimates were derived from UK Biobank. Two-hour glucose refers to measured blood glucose levels 2 hours after consumption of dissolved glucose. The threshold of significance was $P < .0033$ (.05/15 = .0033). Size of data markers is inversely proportional to variance of estimate. To convert total cholesterol, LDL-C, and HDL-C values to mmol/L, multiply by 0.0259; triglyceride values to mmol, multiply by 0.0113; and glucose values to mmol/L, multiply by 0.0555. eGFR indicates estimated glomerular filtration rate; $HbA_{1c}$, hemoglobin $A_{1c}$; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; OR, odds ratio; WHR, waist-to-hip ratio.

[a] Units reported in column 1.

[b] Calculated as weight in kilograms divided by height in meters squared.

BMI polygenic risk score with triglyceride level (estimated using linear regression). To derive the remaining proportion of CHD risk unaccounted for by an increase in triglyceride levels, the magnitude of association of the change in triglyceride level with CHD was subtracted from the estimate of the genetic association of WHR adjusted for BMI with CHD (estimated using logistic regression).

Analyses were performed using R version 3.2.3 (R Project for Statistical Computing) and Stata version 12 (StataCorp).

## Results

The characteristics of UK Biobank participants are reported in the Table. The mean age was 56.9 (SD, 7.9) years, mean systolic blood pressure was 143.6 mm Hg (SD, 21.8), and mean diastolic blood pressure was 84.5 mm Hg (SD, 11.8); 5639 participants (5.0%) had CHD, and 5690 (5.1%) had type 2 diabetes.

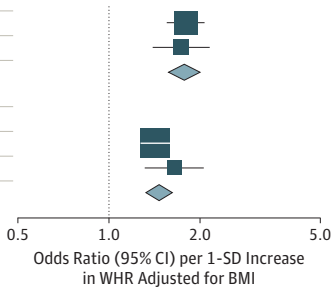
A 48-SNP polygenic risk score for WHR adjusted for BMI was a strong instrumental variable (F = 1713), statistically accounting for 1.5% of variance in WHR adjusted for BMI in the UK Biobank, thus validating assumption 1 in Figure 1.

To test assumption 2 (independence of polygenic risk score for WHR adjusted for BMI from potential confounders, Figure 1), the relationship of the polygenic risk score to smoking, alcohol use, physical activity, vegetable consumption, red meat consumption, and breastfeeding status as a child was determined among individuals in the UK Biobank. In each case, no significant relationship was noted (eTable 5 in the Supplement). For comparison, a similar analysis that categorized individuals according to observed WHR adjusted for BMI (instead of genetic predisposition to WHR adjusted for BMI) was conducted (eTable 6 in the Supplement). In this observational epidemiology analysis, WHR adjusted for BMI was associated with each potential confounder.

A 1-SD increase in WHR adjusted for BMI due to the polygenic risk score was associated with a 1-point decrease in BMI (95% CI, 0.87-1.1), a 2-cm increase in waist circumference (95% CI, 1.5-2.4), a 4.1-cm decrease in hip circumference (95% CI, 3.8-4.4), and an increase of 0.068 in WHR (95% CI, 0.066-0.070)( **Figure 3**). A 1-SD increase in WHR adjusted for BMI due to the polygenic risk score was associated with higher total cholesterol level (5.6 [95% CI, 3.9-7.3] mg/dL [0.15 {95% CI, 0.10-0.19} mmol/L]), higher low-density

Figure 4. Association of 48-SNP Polygenic Risk Score for WHR Adjusted for BMI With Type 2 Diabetes and Coronary Heart Disease

| Instrument | Cases, No. | Controls, No. | Odds Ratio (95% CI) per 1-SD Increase in WHR Adjusted for BMI | P Value |
|---|---|---|---|---|
| Type 2 diabetes | | | | |
| DIAGRAM | 34 840 | 114 981 | 1.79 (1.55-2.07) | $1.88 \times 10^{-15}$ |
| UK Biobank | 5690 | 106 296 | 1.73 (1.40-2.15) | $6.80 \times 10^{-7}$ |
| Fixed-effects model<br>P = .80 for interaction | | | 1.77 (1.57-2.00) | $7.30 \times 10^{-21}$ |
| Coronary heart disease | | | | |
| CARDIOGRAMplusC4D | 60 801 | 123 504 | 1.42 (1.27-1.59) | $7.58 \times 10^{-10}$ |
| UK Biobank | 5639 | 106 347 | 1.64 (1.31-2.06) | $1.46 \times 10^{-5}$ |
| Fixed-effects model<br>P = .26 for interaction | | | 1.46 (1.32-1.62) | $9.90 \times 10^{-14}$ |




Results are standardized to a 1-SD increase in waist-to-hip ratio adjusted for body mass index due to polygenic risk score. Estimates were independently derived in genome-wide association studies (CARDIOGRAMplusC4D for coronary heart disease and DIAGRAM for type 2 diabetes) and the UK Biobank. The threshold of significance was $P < .025$ (0.05/2 = 0.025). Size of data markers is inversely proportional to variance of estimate. CARDIOGRAMplusC4D indicates Coronary Artery DIsease Genome-Wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics Consortium; DIAGRAM, Diabetes Genetics Replication and Meta-analysis.

lipoprotein cholesterol level (5.7 [95% CI, 4.1-7.2] mg/dL [0.15 {95% CI, 0.11-0.19} mmol/L]), higher triglyceride level (27 [95% CI, CI, 25-30] mg/dL [0.31 {95% CI, 0.28-0.34} mmol/L]), and lower high-density lipoprotein cholesterol level (6.0 [95% CI, 5.3-6.6] mg/dL [0.16 {0.14-0.17} mmol/L]). A 1-SD increase in WHR adjusted for BMI due to the polygenic risk score was associated with higher log-transformed fasting insulin levels (0.07 [95% CI, 0.05-0.08] log[pmol/L]), higher 2-hour glucose levels (4.1 [95% CI, 1.6-6.5] mg/dL [0.23 {95% CI, 0.09-0.36} mmol/L), and higher systolic blood pressure (2.1 [95% CI, 1.2-3.0] mm Hg).

A 1-SD increase in WHR adjusted for BMI due to the polygenic risk score was associated with a higher risk of type 2 diabetes (OR, 1.77 [95% CI, 1.57-2.00]; absolute risk increase per 1000 participant-years, 6.0 [95% CI, 4.4-7.8]; $P = 7.30 \times 10^{-21}$; number of participants with type 2 diabetes outcome, 40 530) (**Figure 4**). A 1-SD increase in WHR adjusted for BMI due to the polygenic risk score was also associated with higher risk of CHD (OR, 1.46 [95% CI, 1.32-1.62]; absolute risk increase per 1000 participant-years, 1.8 [95% CI, 1.3-2.4]; $P = 9.90 \times 10^{-14}$; number of participants with CHD outcome, 66 440) (Figure 4).

Five sensitivity analyses (eMethods D, eFigures 1-9 in the Supplement) of the genetic association of WHR adjusted for BMI with cardiometabolic traits, type 2 diabetes, and CHD were conducted to examine if results were influenced by pleiotropy (ie, a violation of assumptions 2 or 3 in Figure 1). Four of the 5 sensitivity analyses were consistent with the results not being influenced by pleiotropy (eFigures 1-7 in the Supplement). In the fifth sensitivity analysis, 8 SNPs associated with increased WHR adjusted for BMI in women but not men were combined in an additive risk score. If increased WHR adjusted for BMI causes CHD (rather than results being due to pleiotropy), then a risk score that increases WHR adjusted for BMI in women but not in men should increase risk of CHD in women but not in men. Although a numerically greater magnitude of association with type 2 diabetes and CHD was noted in women as compared with men, no significant difference was observed ($P = 0.10$ and $P = 0.11$, respectively, for interaction) (eFigures 8 and 9, eMethods D in the Supplement).

Using the polygenic risk score of 48 SNPs associated with WHR adjusted for BMI, a phenome-wide association study of 35 additional diseases in the UK Biobank was conducted (**Figure 5**). There was no significant association of WHR adjusted for BMI with any of these diseases at the Bonferroni-adjusted level of significance ($P < .0014$).

In mediation analysis, the association of polygenic risk score for WHR adjusted for BMI with CHD was attenuated from an OR of 1.46 (95% CI, 1.32-1.62) to an OR of 1.23 (95% CI, 1.11-1.36), after accounting for the association of the polygenic risk score with triglyceride level (eFigure 10 in the Supplement).

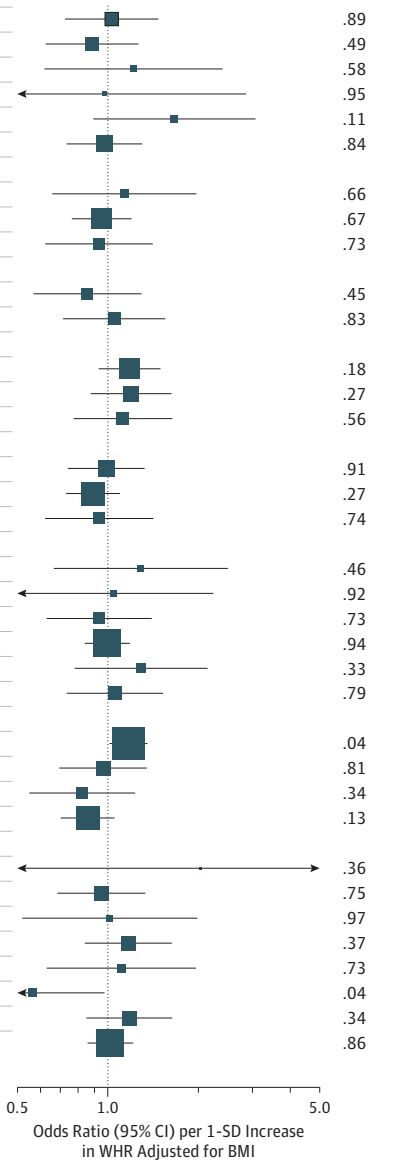
## Discussion

Mendelian randomization analyses tested if human genetic evidence supported a causal relationship of WHR adjusted for BMI (a measure of abdominal adiposity) with type 2 diabetes and CHD. Genetic predisposition to higher WHR adjusted for BMI was associated with increased levels of quantitative risk factors (lipids, insulin, glucose, and systolic blood pressure) as well as a higher risk for type 2 diabetes (OR, 1.77 [95% CI, 1.57-2.00] per 1-SD higher WHR adjusted for BMI) and CHD (OR, 1.46 [95% CI, 1.32-1.62] per 1-SD higher WHR adjusted for BMI).

These results permit several conclusions. First, these findings lend human genetic support to previous observations associating abdominal adiposity with cardiometabolic disease.[6,7] In the INTERHEART acute myocardial infarction case-control study, a 1-SD higher WHR was associated with increased odds of myocardial infarction (OR, 1.37 [95% CI, 1.33-1.40]) after adjustment for BMI and other covariates.[6] However, residual confounding or reverse causality may have contributed to these associations. Indeed, in this study, observational WHR adjusted for BMI was strongly associated with potential confounders, illustrating a limitation of observational

Figure 5. Phenome-Wide Association Study Testing if 48-SNP Polygenic Risk Score for WHR Adjusted for BMI Is Associated With a Range of Disease Phenotypes

| Outcome | Cases, No. | Controls, No. | Odds Ratio (95% CI) per 1-SD Increase in WHR Adjusted for BMI | *P* Value |
|---|---|---|---|---|
| Cardiovascular | | | | |
| Stroke | 2035 | 109 951 | 1.03 (0.72-1.46) | .89 |
| Atrial fibrillation or flutter | 2194 | 109 792 | 0.88 (0.62-1.26) | .49 |
| Heart failure | 586 | 111 400 | 1.21 (0.62-2.39) | .58 |
| Aortic stenosis | 190 | 111 796 | 0.97 (0.33-2.85) | .95 |
| Peripheral vascular disease | 682 | 111 304 | 1.65 (0.89-3.07) | .11 |
| Venous thromboembolism | 3271 | 108 715 | 0.97 (0.73-1.30) | .84 |
| Endocrine | | | | |
| Hyperthyroidism | 866 | 111 120 | 1.13 (0.65-1.96) | .66 |
| Hypothyroidism | 5415 | 106 571 | 0.95 (0.76-1.19) | .67 |
| Gout | 1605 | 110 381 | 0.93 (0.62-1.40) | .73 |
| Renal or urologic | | | | |
| Enlarged prostate | 1567 | 110 419 | 0.85 (0.57-1.29) | .45 |
| Uterine fibroids | 1632 | 110 354 | 1.04 (0.71-1.54) | .83 |
| Gastrointestinal | | | | |
| Gastric reflux | 4864 | 107 122 | 1.18 (0.93-1.49) | .18 |
| Irritable bowel syndrome | 2672 | 109 314 | 1.19 (0.87-1.62) | .27 |
| Gallstones | 1825 | 110 161 | 1.12 (0.77-1.63) | .56 |
| Neurologic or psychiatric | | | | |
| Migraine | 3155 | 108 831 | 0.98 (0.73-1.32) | .91 |
| Depression | 6636 | 105 350 | 0.89 (0.73-1.09) | .27 |
| Anxiety | 1539 | 110 447 | 0.93 (0.62-1.41) | .74 |
| Musculoskeletal | | | | |
| Back pain | 581 | 111 405 | 1.28 (0.66-2.49) | .46 |
| Joint pain | 377 | 111 609 | 1.04 (0.49-2.22) | .92 |
| Osteoporosis | 1730 | 110 256 | 0.93 (0.63-1.39) | .73 |
| Osteoarthritis | 9655 | 102 331 | 0.99 (0.84-1.18) | .94 |
| Sciatica | 1030 | 110 956 | 1.28 (0.77-2.13) | .33 |
| Prolapsed disk | 1849 | 110 137 | 1.05 (0.73-1.52) | .79 |
| Respiratory | | | | |
| Asthma | 13 896 | 98 090 | 1.17 (1.01-1.35) | .04 |
| COPD | 2350 | 109 636 | 0.96 (0.69-1.34) | .81 |
| Pneumonia | 1576 | 110 410 | 0.82 (0.55-1.23) | .34 |
| Hay fever | 6251 | 105 735 | 0.85 (0.70-1.05) | .13 |
| Cancer | | | | |
| Lung | 113 | 111 873 | 2.02 (0.46-8.90) | .36 |
| Breast | 2378 | 109 608 | 0.95 (0.68-1.33) | .75 |
| Colorectal | 610 | 111 376 | 1.01 (0.52-1.97) | .97 |
| Skin | 2475 | 109 511 | 1.16 (0.84-1.62) | .37 |
| Prostate | 835 | 111 151 | 1.10 (0.63-1.95) | .73 |
| Cervical | 870 | 111 116 | 0.56 (0.32-0.97) | .04 |
| Other | 2398 | 109 588 | 1.17 (0.85-1.63) | .34 |
| Any | 9494 | 102 492 | 1.02 (0.85-1.21) | .86 |




Results are standardized to a 1-SD increase in waist-to-hip ratio adjusted for body mass index due to polygenic risk score. All estimates were derived in UK Biobank using instrumental variables regression (adjusting for age, sex, and 10 principal components of ancestry). The threshold for significance was $P < .0014$ (0.05/35 = 0.0014). Size of data markers is inversely proportional to variance of estimate. COPD indicates chronic obstructive pulmonary disease; OR, odds ratio; SNP, single-nucleotide polymorphism.

epidemiology. Here, these prior findings were extended by testing a polygenic risk score that appeared independent of measured confounders (eTable 5 in the Supplement). Elevated levels of triglyceride-rich lipoproteins, a risk factor for CHD with genetic and experimental evidence for causality,[26,27] appeared to mediate a substantial proportion of the increased risk for CHD.

Second, these results suggest that body fat distribution, beyond simple measurement of BMI, could explain part of the variation in risk of type 2 diabetes and CHD noted across individuals and subpopulations. For example, increased abdominal adiposity at a given BMI has been proposed as an explanation for the excess risk of CHD observed in South Asians.[28] Similarly, greater abdominal adipose tissue at a given BMI has been proposed to underlie the excess risk of CHD at a given BMI among men compared with women.[29] In the INTERHEART study, which observed a similar strength of association of WHR adjusted for BMI with myocardial infarc-

tion as the genetic estimate reported here, 33.7% of myocardial infarctions were attributed to increased WHR compared with 10.8% of infarctions attributed to overweight and obesity (BMI >25).[6] When combined with the evidence supportive of causality reported here, these results raise the potential that abdominal adiposity, independent of elevated BMI, is a major driver of global CHD burden.

Third, WHR adjusted for BMI might prove useful as a biomarker for the development of therapies to prevent type 2 diabetes and CHD. Although a substantial focus of drug development has been toward therapeutics to reduce overall adiposity,[30] there has been little effort toward the development of therapies that modify body fat distribution to reduce abdominal adiposity. Ongoing research to understand the mechanistic links between the numerous genetic loci that influence WHR adjusted for BMI may lead to novel therapeutic strategies to reduce abdominal adiposity and reduce the risk of type 2 diabetes and CHD.

The mendelian randomization approach used in this study rests on 2 major principles (Figure 1). First, it requires a strong link between the genetic variants used as an instrument and the exposure (WHR adjusted for BMI, assumption 1 in Figure 1). The 48-SNP polygenic risk score explained 1.5% of variance in WHR adjusted for BMI and had an F statistic of 1713 in the UK Biobank, classifying it as a strong instrument with negligible weak instrument bias.[31] Second, mendelian randomization assumes the absence of pleiotropy, that is, it assumes that the genetic variants used as an instrument affect the outcome (CHD) through the exposure (WHR adjusted for BMI) and not through any other pathway or confounding factors (assumptions 2 and 3 in Figure 1). Although it is not possible to directly test whether pleiotropy is present in any mendelian randomization study,[32] a number of steps were taken in this study to reduce the risk of pleiotropy, including use of 3 different genetic instruments, use of weighted median regression, and use of an instrument associated with higher WHR adjusted for BMI in women but not men. Results from 4 of 5 of these sensitivity analyses were consistent with the primary results. Tests for interaction using sex-specific instruments for CHD and diabetes were directionally consistent with expectation but did not demonstrate significant heterogeneity of effect by sex. This analysis required individual-level data available only in UK Biobank participants and may have been underpowered to detect a difference. Future research that explores such sex-specific instruments in larger data sets may prove more conclusive.

This study has several limitations. First, although a number of approaches were used in an attempt to rule out pleiotropy, it is possible that these results represent a shared genetic basis between WHR adjusted for BMI and CHD rather than a causal relationship. Second, prevalent events largely derived from a verbal interview with a nurse were used for the phenome-wide association study of 35 different disorders. Although these events are likely to be of greater specificity and sensitivity than coded mortality data, they have not been independently validated. Third, the phenome-wide association study may have been underpowered to detect an association of genetic WHR adjusted for BMI with outcomes other than type 2 diabetes and CHD. Fourth, this analysis was restricted to individuals of European ancestry; the association of genetic WHR adjusted for BMI with type 2 diabetes and CHD may differ by ethnicity or genetic ancestry.

## Conclusions

A genetic predisposition to higher WHR adjusted for BMI was associated with increased risk of type 2 diabetes and CHD. These results provide evidence supportive of a causal association between abdominal adiposity and these outcomes.

### ARTICLE INFORMATION

**Author Contributions:** Dr Emdin had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Emdin and Khera contributed equally.
*Concept and design:* Emdin, Khera, Kathiresan.
*Acquisition, analysis, or interpretation of data:* All authors.
*Drafting of the manuscript:* Emdin, Khera, Natarajan, Kathiresan.
*Critical revision of the manuscript for important intellectual content:* Emdin, Khera, Klarin, Zekavat, Hsiao, Kathiresan.
*Statistical analysis:* Emdin, Khera, Natarajan, Hsiao, Kathiresan.
*Administrative, technical, or material support:* Klarin, Kathiresan.

**Conflict of Interest Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Khera reported receiving personal fees from Merck and Amarin Pharmaceuticals. Dr Kathiresan reported receiving grants from Bayer Healthcare, Amarin, and Regeneron; serving on scientific advisory boards for Catabasis, Regeneron Genetics Center, Merck, Celera, and Genomics PLC; receiving personal fees from Novartis, Sanofi, AstraZeneca, Alnylam, Eli Lilly, Lerink Partners, Noble Insights, Bayer, and Ionis; receiving consulting fees from Regeneron, Merck, Quest Diagnostics, Novartis, Amgen, Genentech, Corvidia, Genomics PLC, Ionis Pharmaceuticals, and Eli Lilly; and holding equity in Catabasis and San Therapeutics. No other authors reported disclosures.

**Funding/Support:** Dr Emdin is funded by the Rhodes Trust. Dr Khera is funded by an ACCF/Merck Cardiovascular Research Fellowship and a John S. LaDue Memorial Fellowship in Cardiology. Dr Klarin is supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) under award T32 HL007734. Dr Kathiresan is supported by the Ofer and Shelly Nemirovsky Research Scholar award from the Massachusetts General Hospital, the Donovan Family Foundation, and R01HL127564 from the NIH.

**Role of the Funder/Sponsor:** No funders or sponsors were involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

**Additional Information:** This project was conducted using the UK Biobank resource (project ID 7089).

### REFERENCES

1. Nordestgaard BG, Palmer TM, Benn M, et al. The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach. *PLoS Med*. 2012;9(5):e1001212.

2. Whitlock G, Lewington S, Sherliker P, et al; Prospective Studies Collaboration. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet*. 2009;373(9669):1083-1096.

3. Després J-P. Body fat distribution and risk of cardiovascular disease: an update. *Circulation*. 2012;126(10):1301-1313.

4. Ashwell M, Cole TJ, Dixon AK. Obesity: new insight into the anthropometric classification of fat distribution shown by computed tomography. *Br Med J (Clin Res Ed)*. 1985;290(6483):1692-1694.

5. Seidell JC, Björntorp P, Sjöström L, Sannerstedt R, Krotkiewski M, Kvist H. Regional distribution of muscle and fat mass in men—new insight into the

risk of abdominal obesity using computed tomography. *Int J Obes*. 1989;13(3):289-303.

**6**. Yusuf S, Hawken S, Ounpuu S, et al; INTERHEART Study Investigators. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet*. 2005;366(9497):1640-1649.

**7**. Vazquez G, Duval S, Jacobs DR Jr, Silventoinen K. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiol Rev*. 2007;29(1):115-128.

**8**. Han TS, Bijnen FC, Lean ME, Seidell JC. Separate associations of waist and hip circumference with lifestyle factors. *Int J Epidemiol*. 1998;27(3):422-430.

**9**. Smith GD, Ebrahim S. "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.

**10**. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304-314.

**11**. Nikpay M, Goel A, Won HH, et al; CARDIoGRAMplusC4D Consortium. A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121-1130.

**12**. Pattaro C, Teumer A, Gorski M, et al; ICBP Consortium; AGEN Consortium; CARDIOGRAM; CHARGe-Heart Failure Group; ECHOGen Consortium. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun*. 2016;7:10023.

**13**. Morris AP, Voight BF, Teslovich TM, et al; Wellcome Trust Case Control Consortium; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; Asian Genetic Epidemiology Network-Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44(9):981-990.

**14**. Shungin D, Winkler TW, Croteau-Chonka DC, et al; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; ReproGen Consortium. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518(7538):187-196.

**15**. Locke AE, Kahali B, Berndt SI, et al; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.

**16**. Willer CJ, Schmidt EM, Sengupta S, et al; Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283.

**17**. Scott RA, Lagou V, Welch RP, et al; DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet*. 2012;44(9):991-1005.

**18**. Stitziel NO, Stirrups KE, Masca NG, et al; Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding Variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N Engl J Med*. 2016;374(12):1134-1144.

**19**. Ehret GB, Munroe PB, Rice KM, et al; International Consortium for Blood Pressure Genome-Wide Association Studies; CARDIoGRAM Consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen Consortium; CHARGE-HF Consortium. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478(7367):103-109.

**20**. Moffatt MF, Gut IG, Demenais F, et al; GABRIEL Consortium. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*. 2010;363(13):1211-1221.

**21**. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665.

**22**. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat*. 2002;20(4):518-529.

**23**. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet*. 2012;380(9841):572-580.

**24**. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909.

**25**. Burgess S, Davies NM, Thompson SG; EPIC-InterAct Consortium. Instrumental variable analysis with a nonlinear exposure-outcome relationship. *Epidemiology*. 2014;25(6):877-885.

**26**. Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*. 2013;45(11):1345-1352.

**27**. Nordestgaard BG, Varbo A. Triglycerides and cardiovascular disease. *Lancet*. 2014;384(9943):626-635.

**28**. Gupta M, Singh N, Verma S. South Asians and cardiovascular risk: what clinicians should know. *Circulation*. 2006;113(25):e924-e929.

**29**. Barrett-Connor E. Sex differences in coronary heart disease: why are women so superior? the 1995 Ancel Keys Lecture. *Circulation*. 1997;95(1):252-264.

**30**. Rodgers RJ, Tschöp MH, Wilding JPH. Anti-obesity drugs: past, present and future. *Dis Model Mech*. 2012;5(5):621-626.

**31**. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol*. 2011;40(3):740-752.

**32**. von Hinke S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as instrumental variables. *J Health Econ*. 2016;45:131-148.

# DNA Sequence Variation in *ACVR1C* Encoding the Activin Receptor-Like Kinase 7 Influences Body Fat Distribution and Protects Against Type 2 Diabetes

Connor A. Emdin,[1,2,3] Amit V. Khera,[1,2,3] Krishna Aragam,[1,2,3] Mary Haas,[1,2,3] Mark Chaffin,[3] Derek Klarin,[1,3,4] Pradeep Natarajan,[1,2,3] Alexander Bick,[1,3] Seyedeh M. Zekavat,[1,3,5] Akihiro Nomura,[1,2,3] Diego Ardissino,[6,7] James G. Wilson,[8] Heribert Schunkert,[9] Ruth McPherson,[10] Hugh Watkins,[11,12] Roberto Elosua,[13,14,15] Matthew J. Bown,[16] Nilesh J. Samani,[16] Usman Baber,[17] Jeanette Erdmann,[18,19] Namrata Gupta,[3] John Danesh,[20,21,22] Danish Saleheen,[23,24] Stacey Gabriel,[3] and Sekar Kathiresan[1,2,3]

*Diabetes* 2019;68:226–234 | https://doi.org/10.2337/db18-0857

**A genetic predisposition to higher waist-to-hip ratio adjusted for BMI (WHRadjBMI), a measure of body fat distribution, associates with increased risk for type 2 diabetes. We conducted an exome-wide association study of coding variation in UK Biobank (405,569 individuals) to identify variants that lower WHRadjBMI and protect against type 2 diabetes. We identified four variants in the gene *ACVR1C* (encoding the activin receptor-like kinase 7 receptor expressed on adipocytes and pancreatic β-cells), which independently associated with reduced WHRadjBMI: Asn150His (−0.09 SD, $P = 3.4 \times 10^{-17}$), Ile195Thr (−0.15 SD, $P = 1.0 \times 10^{-9}$), Ile482Val (−0.019 SD, $P = 1.6 \times 10^{-5}$), and rs72927479 (−0.035 SD, $P = 2.6 \times 10^{-12}$). Carriers of these variants exhibited reduced percent abdominal fat in DEXA imaging. Pooling across all four variants, a 0.2 SD decrease in WHRadjBMI through *ACVR1C* was associated with a 30% lower risk of type 2 diabetes (odds ratio [OR] 0.70, 95% CI 0.63, 0.77; $P = 5.6 \times 10^{-13}$). In an analysis of exome sequences from 55,516 individuals, carriers of predicted damaging variants in *ACVR1C* were at 54% lower risk of type 2 diabetes (OR 0.46, 95% CI 0.27, 0.81; $P = 0.006$). These findings indicate that variants predicted to lead to loss of *ACVR1C* gene function influence body fat distribution and protect from type 2 diabetes.**

[1]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA
[2]Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA
[3]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA
[4]Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA
[5]Yale Medical School and Department of Computational Biology and Bioinformatics, Yale University, New Haven, CT
[6]Division of Cardiology, Azienda Ospedaliero–Universitaria di Parma, Parma, Italy
[7]Associazione per lo Studio della Trombosi in Cardiologia, Pavia, Italy
[8]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS
[9]Deutsches Herzzentrum München, Technische Universität München, and Deutsches Zentrum für Herz-Kreislauf-Forschung, München, Germany
[10]University of Ottawa Heart Institute, Ottawa, Ontario, Canada
[11]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, U.K.
[12]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.
[13]Cardiovascular Epidemiology and Genetics, Hospital del Mar Medical Research Institute, Barcelona, Spain
[14]CIBER Enfermedades Cardiovasculares (CIBERCV), Barcelona, Spain
[15]Facultat de Medicina, Universitat de Vic-Central de Cataluña, Vic, Spain
[16]Department of Cardiovascular Sciences, University of Leicester, and NIHR Leicester Biomedical Research Centre, Leicester, U.K.
[17]The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY
[18]Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany
[19]DZHK (German Research Centre for Cardiovascular Research) partner site Hamburg/Lübeck/Kiel, Lübeck, Germany
[20]Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, U.K.
[21]Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.
[22]National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge. Cambridge, U.K.
[23]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
[24]Center for Non-Communicable Diseases, Karachi, Pakistan

Corresponding author: Sekar Kathiresan, skathiresan1@mgh.harvard.edu

Received 14 August 2018 and accepted 21 October 2018

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db18-0857/-/DC1.

Discovery of genetic variants that protect against disease can identify novel mechanisms of disease and novel therapeutic targets (1). For example, the discovery of low-frequency coding variants in *PCSK9*, *ANGPTL3*, and *APOC3* that lower blood lipid levels and protect against coronary artery disease catalyzed the development of novel therapeutics for coronary artery disease (2–10). *PCSK9* inhibitors are now approved for treatment of coronary artery disease (4), while inhibitors of *ANGPTL3* (7) and *APOC3* (11) are in clinical development.

Body fat distribution strongly influences the development of type 2 diabetes (12–14). In a Mendelian randomization study of 296,291 individuals, we previously found that a genetic predisposition to increased abdominal fat distribution was associated with elevated triglyceride levels, elevated blood pressure, and an increased risk of coronary artery disease, independent of overall adiposity (12). Furthermore, a genetic predisposition to increased abdominal fat distribution was strongly associated with the development of type 2 diabetes. For each 1 SD genetic increase in waist-to-hip ratio adjusted for BMI (WHRadjBMI) (a measure of body fat distribution), risk of type 2 diabetes increased by 77% (12). These findings were replicated in a separate Mendelian randomization study (14).

These results suggest the hypothesis that genetic variants that influence body fat distribution may also influence the risk of type 2 diabetes (12). Here, we test this hypothesis by analyzing genetic variation in more than 400,000 individuals in UK Biobank to identify novel genetic variants that lower WHRadjBMI and protect against type 2 diabetes. Below, we demonstrate that variants predicted to lead to loss of function of the gene *ACVR1C*, which encodes the activin receptor-like kinase 7 (ALK7), influence body fat distribution and protect against type 2 diabetes.

## RESEARCH DESIGN AND METHODS

### Study Design

In our discovery analysis, we analyzed the association of 614,042 coding variants with WHRadjBMI in 405,569 individuals in UK Biobank. Coding variants were defined as missense variants or variants predicted to result in loss of function of the protein: *1*) nonsense mutations that resulted in early termination of a protein, *2*) frameshift mutations due to insertions or deletions of DNA, or *2*) splice-site mutations that result in an incorrectly spliced protein. Only variants imputed with a quality score (info score) >0.3 were analyzed. Threshold for significance was defined as $P < 5 \times 10^{-8}$ (genome-wide significance), and analysis was performed using PLINK2 software (15). We reported novel variants as those located more than one megabase away from previously identified loci in the Genetic Investigation of ANthropometric Traits (GIANT) Consortium (16). We reported variants located outside of the MHC locus separately from those within the MHC locus, as variants within the MHC locus typically tag HLA risk alleles and are thus associated with phenotypes due to linkage disequilibrium with HLA alleles (17). We attempted to replicate the association of novel variants with WHRadjBMI using data from the GIANT Consortium, when the variant was available in the GIANT Consortium (16).

Upon identification of variants in *ACVR1C* as significantly associated with WHRadjBMI, a conditional analysis was conducted to identify additional variants significantly associated ($P < 0.0001$) with WHRadjBMI within the *ACVR1C* locus (±250 kb of the lead variant rs55920843). To replicate observed associations of *ACVR1C* variants with WHRadjBMI, we examined whether carriers of these variants had reduced WHRadjBMI in a meta-analysis of the GIANT Consortium and two independent cohorts (Atherosclerosis Risk in Communities [ARIC] and Framingham Heart Study). We also examined whether these variants were associated with direct imaging-based measurements of abdominal fat in 4,215 participants who underwent DEXA imaging in UK Biobank.

To test whether identified *ACVR1C* variants were also associated with risk of type 2 diabetes, we pooled data from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (ExTexT2D exome chip analysis [18]) with UK Biobank. To test whether variation leading to loss of *ACVR1C* function protects against type 2 diabetes, we analyzed the sequences of the nine exons of *ACVR1C* in 55,516 participants (31,672 from the Myocardial Infarction Genetics Consortium [MIGen] [19,20], 5,388 from the ARIC study, and 18,456 from the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples [T2D-GENES] Consortium [21]) and examined whether predicted damaging variants in the gene associate with risk of type 2 diabetes.

A phenome-wide association study of *ACVR1C* in UK Biobank was performed using an *ACVR1C* gene risk score (22). Definitions for 31 different diseases analyzed in the phenome-wide association study are provided (Supplementary Table 1). Three metabolic traits available in UK Biobank (urinary albumin-to-creatinine ratio, systolic blood pressure, and diastolic blood pressure) were also analyzed. A *P* value of 0.001 (0.05/34) was used for significance in this analysis.

### Data Sources

For the analysis of WHRadjBMI, individual-level data from 405,569 unrelated individuals from the UK Biobank (335,660 individuals of European ancestry and 69,909 individuals of non-European ancestry) were analyzed. UK Biobank received ethical approval from the Research Ethics Committee (reference number 11/NW/0382). Analysis of UK Biobank was approved by the Partners Health Care Institutional Review Board (protocol 2013P001840). Informed consent was obtained from all participants by UK Biobank. For replication, data for WHRadjBMI from the GIANT Consortium (in which the Ile482Val variant was available) were pooled with data from the ARIC and

Framingham Heart Study data sets (in which Asn150His, Ile195Thr, and rs72927479 variants were available). The GIANT Consortium consisted of 224,459 participants (210,088 of European ancestry and 14,371 of non-European ancestry) genotyped using the MetaboChip (16). The ARIC study is a community-based study of 15,792 white and black participants, aged 45 to 64 years (23). The ARIC data set consisted of 10,122 individuals (8,015 of European ancestry and 2,107 of non-European ancestry) who were genotyped and imputed, as previously described (24). For 5,388 participants, exome sequences were also available for analysis. In the Framingham Heart Study, a community–based study of 10,092 individuals of predominantly European ancestry, genotyped data were available from 6,073 individuals of European ancestry.

For the analysis of type 2 diabetes, estimates from UK Biobank were pooled using inverse variance–weighted fixed-effects meta-analysis with estimates from the DIAGRAM ExTexT2D exome chip analysis of 452,244 participants (81,412 case subjects with diabetes and 370,832 control subjects) (18). In UK Biobank, type 2 diabetes was defined as *1*) self-report of type 2 diabetes, followed by a verbal interview with a trained nurse to confirm the diagnosis, or *2*) hospitalization for ICD code E11. Because the ExTexT2D analysis included 120,286 participants from UK Biobank (18), these individuals were excluded from the analysis of type 2 diabetes in UK Biobank to prevent analysis of overlapping samples.

Sequence data for *ACVR1C* were extracted from exome sequencing performed in MIGen as previously described (19,20). The Burrows–Wheeler aligner algorithm was used to align reads from participants to the reference genome (hg19). The GATK HaplotypeCaller was used to jointly call variants. Metrics including variant quality score recalibration, quality over depth, and strand bias were then used to filter variants. The Jackson Heart Study (JHS) was excluded from analysis of MIGen, as it was included in the T2D-GENES Consortium. Exome sequences from 5,388 participants in ARIC were analyzed as previously described (25). Phenotype and genotype data were retrieved from the National Center for Biotechnology Information dbGaP server (accession phs000090.v3.p1 and phs000572.v6.p4). Exome sequencing was performed in the T2D-GENES Consortium as previously described (21). To analyze exome sequences from the T2D-GENES Consortium, the online Genetic Association Interactive Tool in the Type 2 Diabetes Knowledge Portal was used (21).

Studies included in MIGen were *1*) the Italian Atherosclerosis, Thrombosis, and Vascular Biology (ATVB) study (dbGaP study accession phs000814.v1.p1); *2*) the Exome Sequencing Project Early-Onset Myocardial Infarction (ESP-EOMI) study (9); *3*) a nested case-control cohort from the JHS; *4*) the South German Myocardial Infarction study (dbGaP study accession phs000916.v1.p1); *5*) the Ottawa Heart Study (OHS) (dbGaP study accession phs000806.v1.p1); *6*) the Precocious Coronary Artery Disease (PROCARDIS) study (dbGaP Study Accession phs000883.v1.p1) ; *7*) the Pakistan Risk of Myocardial Infarction Study (PROMIS) (dbGaP study accession phs000917.v1.p1); *8*) the Registre Gironí del COR (Gerona Heart Registry or REGICOR) study (dbGaP study accession phs000902.v1.p1); *9*) the Leicester Myocardial Infarction study (dbGaP study accession phs001000.v1.p1); *10*) the BioImage study (dbGaP study accession phs001058.v1.p1); and *11*) the North German Myocardial Infarction study (dbGaP Study Accession phs000990.v1.p1).

Predicted damaging *ACVR1C* variants in the exome sequencing analysis were defined as those that resulted in loss of function of the protein (nonsense mutations that resulted in early termination of *ACVR1C*, frameshift mutations due to insertions or deletions of DNA, or splice-site mutations that result in an incorrectly spliced protein) or those labeled as damaging by each of five different algorithms (LRT score, MutationTaster, PolyPhen-2 HumDiv, PolyPhen-2 HumVar, and SIFT), as previously described (19,26). The Variant Effect Predictor algorithm was used to annotate predicted damaging variants (27).

### Statistical Analysis

In UK Biobank, WHRadjBMI was derived through inverse normal transformation of waist-to-hip ratio after adjustment for age, sex, and BMI (as in the GIANT Consortium [16]). In UK Biobank, linear regression was used to estimate the association of variants with WHRadjBMI. All UK Biobank analyses included adjustment for age, sex, 10 principal components of ancestry, and a dummy variable for the array type used in genotyping. Logistic regression was used to estimate the association of variants with type 2 diabetes. Estimates of the association of each variant with type 2 diabetes in UK Biobank were pooled with estimates from the ExTexT2D Consortium using inverse variant–weighted fixed-effects meta-analysis.

In the primary analysis of 405,569 individuals for WHRadjBMI, we had 80% power to detect a minimum effect of 0.05 SD with a minor allele frequency of 5% at genome-wide significance ($P < 5 \times 10^{-8}$). In the primary analysis of 95,978 case subjects with type 2 diabetes and 646,985 control subjects, we had 80% power to detect a minimum odds ratio (OR) of 1.05 at $P < 5 \times 10^{-8}$. With a minor allele frequency of 1%, we had 80% power to detect a minimum effect of 0.1 SD for WHRadjBMI and an OR of 1.10 for type 2 diabetes at $P < 5 \times 10^{-8}$.

To estimate the overall association of variation in *ACVR1C* with WHRadjBMI, we pooled across all variants using a gene risk score, weighted by the square root of allele frequency (estimating a weighted mean effect of *ACVR1C* variants on WHRadjBMI) (28). To estimate the overall association of variation in *ACVR1C* with type 2 diabetes, we pooled across all variants in a gene risk score, weighted by the association of each variant with WHRadjBMI (29).

For analysis of exome sequencing data, logistic regression was performed with adjustment for sex, five principal components of ancestry, and a dummy variable for each

**Table 1—Baseline characteristics of participants in UK Biobank**

| | All participants (*N* = 405,569) | Type 2 diabetes case subjects (*N* = 20,458) | Control subjects without diabetes (*N* = 385,111) |
|---|---|---|---|
| Age, years | 57 ± 8.1 | 61 ± 6.9 | 57 ± 8.1 |
| Female | 218,376 (54) | 7,801 (38) | 210,575 (55) |
| UK BiLEVE array | 48,625 (12.0) | 3,097 (15) | 45,528 (12) |
| BMI, kg/m$^2$ | 27 ± 4.8 | 32 ± 5.9 | 27 ± 4.6 |
| Waist-to-hip ratio | 0.87 ± 0.09 | 0.95 ± 0.08 | 0.87 ± 0.09 |

Data are mean ± SD or *n* (%).

cohort to estimate the association of predicted damaging variants with type 2 diabetes. Estimates from MIGen were pooled with estimates from the T2D-GENES Consortium using inverse variance–weighted fixed-effects meta-analysis.

For the phenome-wide association study, all four *ACVR1C* variants were pooled in a gene risk score in UK Biobank, as previously described (25,30). For each individual in UK Biobank, the *ACVR1C* variants associated with lower WHRadjBMI were weighted by their effect on WHRadjBMI and summed. The association of this gene risk score with 31 different diseases in UK Biobank and three metabolic traits was tested using logistic regression with adjustment for age, sex, 10 principal components of ancestry, and a dummy variable for the array type used in genotyping. Although exploratory due to the low number of case subjects for certain diseases (e.g., 1,707 case subjects for cervical cancer), we conducted this analysis to detect possible adverse associations of *ACVR1C* with diseases that could allow for prediction of adverse effects of pharmacologic inhibition of *ACVR1C*.

Analyses were performed using R version 3.2.3 (R Project for Statistical Computing).

## RESULTS

### Exome-Wide Association Study of Body Fat Distribution in UK Biobank

Among 405,569 participants in UK Biobank, 54% were female, the median age was 57 years, and the median waist-to-hip ratio, measured at enrollment, was 0.87 (Table 1). One SD in waist-to-hip ratio corresponded to an absolute change of 0.09. In an analysis of 614,012 coding variants in UK Biobank, no evidence of genomic inflation was observed (λ 1.08) (Supplementary Fig. 1).

We identified 16 low-frequency variants (<5%) associated with WHRadjBMI outside of known loci (Supplementary Table 2) (16). We identified an additional 43 novel common variants (frequency >5%) associated with WHRadjBMI (Supplementary Table 3). We identified 94 independent variants in total, including variants at known loci (Supplementary Data). Of 59 novel coding variants, 34 variants were available in the GIANT Consortium for replication (16). A strong correlation in the effect sizes of the association of variants with WHRadjBMI in UK Biobank and GIANT was observed ($R^2$ = 0.87). A total of 24 variants were independently associated with WHRadjBMI in GIANT ($P < 0.05$) (Supplementary Table 4). Of the 10 variants that were not significantly associated with WHRadjBMI in GIANT ($P > 0.05$), only 3 variants exhibited evidence of heterogeneity between estimates in UK Biobank and GIANT, with the remaining 7 variants showing similar estimates of association in UK Biobank and GIANT (although not reaching significance in GIANT) (Supplementary Table 4).

The lead novel low-frequency variant was a missense variant in *PNPLA2* (rs140201358, Asn252Lys) that associated with elevated WHRadjBMI (0.09 SD, $P = 3.2 \times 10^{-19}$). *PNPLA2* encodes adipocyte triglyceride lipase, which hydrolyzes triglycerides in adipose tissue to mobilize fat stores (31). Two low-frequency missense variants in *ABHD15*, which encodes alpha/beta hydrolase domain-containing protein 15, were found to be associated with elevated WHRadjBMI. *ABHD15* is also highly expressed in adipocytes and has been reported to mediate insulin-induced suppression of lipolysis in adipocytes (32). We also identified a low-frequency variant in a known locus (*CALCRL* Leu87Pro, 0.1% frequency) associated with lower WHRadjBMI (−0.14 SD, $P = 1.9 \times 10^{-11}$). A common noncoding variant in the *CALCRL* locus, encoding calcitonin receptor-like receptor, was previously identified in the GIANT Consortium as associated with WHRadjBMI (16). The identification of an independent low-frequency missense variant in the gene suggests that *CALCRL* may be the causal gene at this locus.

### Variation Leading to Lower WHRadjBMI: The *ACVR1C* Locus

We next focused on novel variant alleles leading to lower WHRadjBMI. The lead novel low-frequency variant associated with lower WHRadjBMI lay within the gene *ACVR1C*. *ACVR1C* Asn150His (allele frequency 1.1%) associated with 0.09 SD lower WHRadjBMI ($P = 3.4 \times 10^{-17}$). An independent missense variant in *ACVR1C*, Ile195Thr (AF 0.4%), also associated with lower WHRadjBMI (0.15 SD, $P = 1.0 \times 10^{-9}$).

Upon conditioning on these two variants, we identified an additional coding variant: Ile482Val (AF 7%), which associated with 0.019 SD lower WHRadjBMI ($P = 1.6 \times 10^{-5}$), and rs72927479, a noncoding variant, for which the minor G allele (AF 5%) was associated with 0.035 SD lower WHRadjBMI ($P = 2.6 \times 10^{-12}$) (Table 2). Despite being an

independent signal for WHRadjBMI, the noncoding variant rs72927479 is nominally correlated with Ile482Val ($r^2$ = 0.06 in UK Biobank). No other variants were correlated with one another in UK Biobank (all $r^2 < 0.001$).

Pooling across all four variants with weighting by square root of allele frequency, *ACVR1C* variation was associated with lower WHRadjBMI in UK Biobank (−0.07 SD, $P = 2.6 \times 10^{-35}$). To replicate this finding, we pooled data from the GIANT Consortium (in which the Ile482Val variant was available, $n$ = 224,156) with data from ARIC and Framingham Heart Study data sets (in which Asn150His, Ile195Thr, and rs72927479 variants were available, $n$ = 13,704). Pooling across all four variants in these three replication studies, variation in *ACVR1C* was associated with reduced WHRadjBMI (−0.07 SD, $P$ = 0.0005) (Supplementary Table 5).

Examining other anthropometric traits in UK Biobank, variation in *ACVR1C* was associated with elevated hip circumference (0.035 SD, $P = 3 \times 10^{-8}$) and nominally elevated BMI (0.02 SD, $P$ = 0.002) and was unassociated with waist circumference (−0.007 SD, $P$ = 0.20) or height (0.005 SD, $P$ = 0.23). In the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) analysis of insulin resistance (HOMA-IR), Ile482Val (the only *ACVR1C* variant available for analysis) was unassociated with HOMA-IR (−1.1%, $P$ = 0.17) (33).

We examined whether *ACVR1C* variation also associated with direct imaging measurement of abdominal obesity. A total of 4,215 participants in UK Biobank underwent DEXA imaging to estimate abdominal fat mass. Carriers of *ACVR1C* variants had lower percent abdominal fat ($P$ = 0.008) (Fig. 1). When one outlying individual with four *ACVR1C* variants was excluded (Fig. 1), carriers of *ACVR1C* continued to have significantly lower percent abdominal fat ($P$ = 0.009).

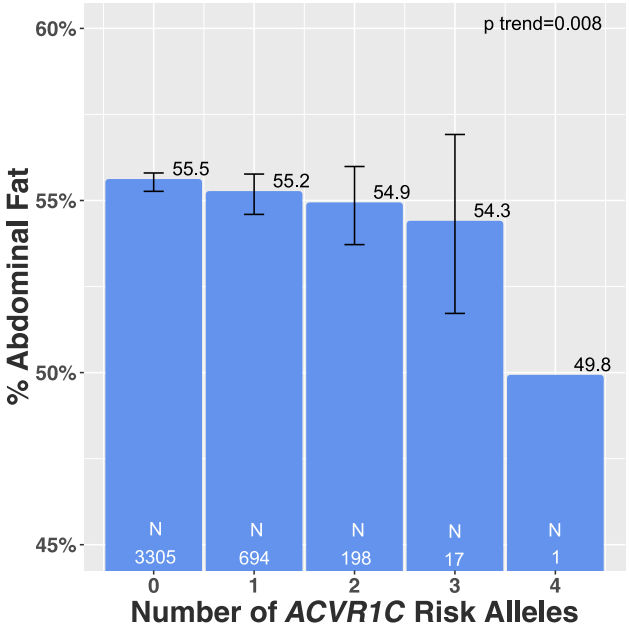



**Figure 1**—Across four *ACVR1C* genetic variants, association of number of WHRadjBMI-lowering alleles with mean directly measured abdominal fat (percent abdominal fat of total body fat) in 4,215 participants who underwent DEXA scan in UK Biobank, adjusted for age, sex, 10 principal components of ancestry, and array type. Number of participants in each group is displayed in white for each bar.

### Association of Variants in *ACVR1C* With Type 2 Diabetes

Genetic predisposition to increased WHRadjBMI strongly predisposes to type 2 diabetes (12). We therefore examined whether variants in *ACVR1C* that lower WHRadjBMI adiposity protect against type 2 diabetes. In a combined analysis of UK Biobank and the DIAGRAM Consortium, all four *ACVR1C* variants were found to independently protect against type 2 diabetes (OR 0.88, $P = 8.7 \times 10^{-5}$ for Asn150His; OR 0.79, $P$ = 0.005 for Ile195Thr; OR 0.95, $P = 4.8 \times 10^{-6}$ for Ile482Val; OR 0.93, $P$ = 0.0006 for rs72927479) (Table 1). Pooling across all four variants, a 0.2 SD decrease in WHRadjBMI through *ACVR1C* was associated with a 30% lower risk of type 2 diabetes (OR 0.70, 95% CI 0.63, 0.77; $P = 5.6 \times 10^{-13}$) (Fig. 2). When we excluded the noncoding variant rs72927479, which is nominally correlated with Ile482Val ($r^2$ = 0.06 in UK Biobank), the *ACVR1C* gene risk score remained associated with risk of type 2 diabetes (OR 0.71, 95% CI 0.64, 0.79; $P = 1.8 \times 10^{-10}$)

In three independent data sets with exome sequence data—MIGen, ARIC, and the T2D-GENES Consortium—we examined whether variants predicted to damage *ACVR1C* gene function protect against type 2 diabetes. The nine exons of the *ACVR1C* gene were sequenced in

**Table 2—Association of variants in *ACVR1C* with WHRadjBMI and with type 2 diabetes**

| Variant | Minor allele frequency (%) | WHRadjBMI | | Type 2 diabetes | |
|---|---|---|---|---|---|
| | | β (95% CI) | *P* value | OR (95% CI) | *P* value |
| Asn150His | 1.1 | −0.089 (−0.11, −0.067) | $3.4 \times 10^{-17}$ | 0.88 (0.83, 0.94) | $8.7 \times 10^{-5}$ |
| Ile195Thr | 0.2 | −0.15 (−0.09, 0.19) | $1.0 \times 10^{-9}$ | 0.79 (0.67, 0.93) | 0.005 |
| Ile482Val | 7.2 | −0.019 (−0.01, −0.027) | $1.6 \times 10^{-5}$ | 0.95 (0.93, 0.97) | $4.8 \times 10^{-6}$ |
| rs72927479 | 5.1 | −0.035 (−0.045, −0.025) | $2.6 \times 10^{-12}$ | 0.93 (0.89, 0.97) | $6.0 \times 10^{-4}$ |

Estimates for WHRadjBMI were derived through linear regression analysis in UK Biobank. Estimates for type 2 diabetes were derived through meta-analysis of UK Biobank and the DIAGRAM ExTexT2D Consortium.

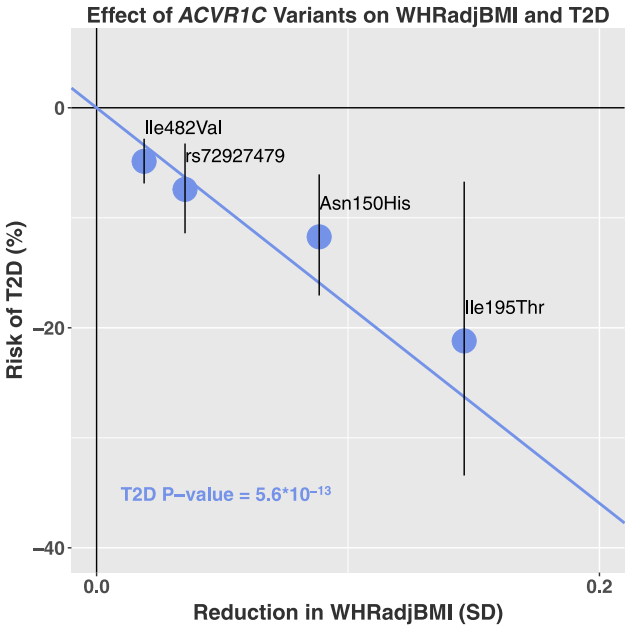



**Figure 2**—Association of four variants in *ACVR1C* with WHRadjBMI (*x*-axis) and type 2 diabetes (T2D) (*y*-axis).

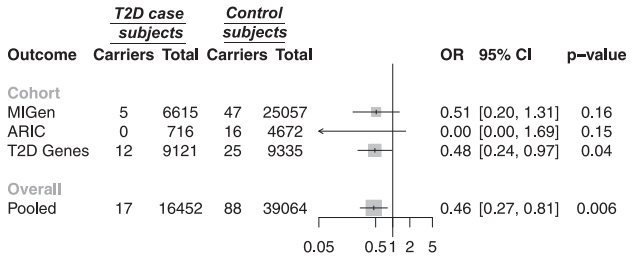



**Figure 3**—Association of predicted damaging variants in *ACVR1C* with type 2 diabetes (T2D) from sequences in MIGen, ARIC, and the T2D-GENES Consortium (T2D Genes).

55,516 individuals. A total of 105 predicted damaging variants were identified (Supplementary Tables 6–8). Among 16,452 case subjects with type 2 diabetes, the frequency of predicted damaging variants in *ACVR1C* was 0.1% (17) compared with 0.2% (88) among 39,064 control subjects without diabetes. Overall, carrying a predicted damaging variant in *ACVR1C* was associated with 54% lower risk of type 2 diabetes (OR 0.46, 95% CI 0.27, 0.81; $P$ = 0.006) (Fig. 3). When we excluded 47 carriers of I195T (annotated as a predicted damaging variant by five of five algorithms) in the exome sequencing analysis, carriers of predicted damaging variants in *ACVR1C* remained protected from type 2 diabetes (OR 0.48, 95% CI 0.24, 0.97; $P$ = 0.04).

To further examine whether loss of *ACVR1C* function lowers WHRadjBMI, we examined whether the noncoding variant rs72927479 associates with *ACVR1C* expression. The minor allele of rs72927479 (G, frequency 5%) associated with lower WHRadjBMI (β −0.035, $P = 2.6 \times 10^{-12}$) and type 2 diabetes (OR 0.93, $P$ = 0.0006). In the Genotype-Tissue Expression (GTEx) data set (34), the minor allele of rs72927479 associated with reduced expression of *ACVR1C* in subcutaneous adipose tissue ($P$ = 0.02) and pancreas ($P$ = 0.02) (Supplementary Fig. 2).

### Phenome-Wide Association Study of *ACVR1C* in UK Biobank

To anticipate whether *ACVR1C* inhibition may be associated with on-target adverse effects, we conducted a phenome-wide association study of 31 disease phenotypes in UK Biobank. We did not observe any significant associations between the *ACVR1C* gene risk score and the 31 diseases analyzed (Fig. 4). The *ACVR1C* gene risk score was unassociated with coronary artery disease (OR 1.01, $P$ = 0.86). We also examined whether *ACVR1C* variation associates with three metabolic traits currently available in UK Biobank: urinary albumin, systolic blood pressure, and diastolic blood pressure. While the *ACVR1C* gene risk score did not significantly associate with urinary albumin levels, it associated with significantly lower diastolic blood pressure (−0.6 mmHg, $P$ = 0.0004) and nominally lower systolic blood pressure (−0.6 mmHg, $P$ = 0.03).

## DISCUSSION

In this study, four genetic variants in *ACVR1C,* ranging in frequency from 0.2 to 7.2%, independently associated with lower WHRadjBMI and protected against type 2 diabetes. Furthermore, damaging variants in *ACVR1C* protected against type 2 diabetes in an analysis of exome sequences from 55,516 individuals. An *ACVR1C* gene risk score did not associate with any of 31 additional diseases in UK Biobank but did nominally associate with lower blood pressure.

These results permit several conclusions. First, pharmaceutical inhibition of *ACVR1C* may be useful in the treatment of type 2 diabetes. *ACVR1C* encodes ALK7, a transforming growth factor-β family receptor highly expressed on pancreatic islet cells (35,36) and adipocytes (37). Overexpression of AKL7 induces growth inhibition and apoptosis of pancreatic β-cells (36,38), suggesting that it is a negative regulator of β-cell mass. A number of findings from model systems also suggest that ALK7 may be useful as a therapeutic target for abdominal obesity, type 2 diabetes, and other metabolic diseases. *ACVR1C*-deficient mice have been reported to have reduced body fat when fed a high-fat diet (37,39) and have improved glucose tolerance and insulin sensitivity when obese (37). Chemical inhibition of *ACVR1C* has also been shown to reduce fat accumulation and increase lipolysis in mice (40). Rats with streptozocin-induced diabetes show elevated ALK7 expression, and shRNA knockdown of *ACVR1C* reduces arterial stiffness in this model (41). In combination with the human genetic results presented here, these findings suggest that inhibition of *ACVR1C* may prove useful to modify body fat distribution and lower risk for type 2 diabetes.

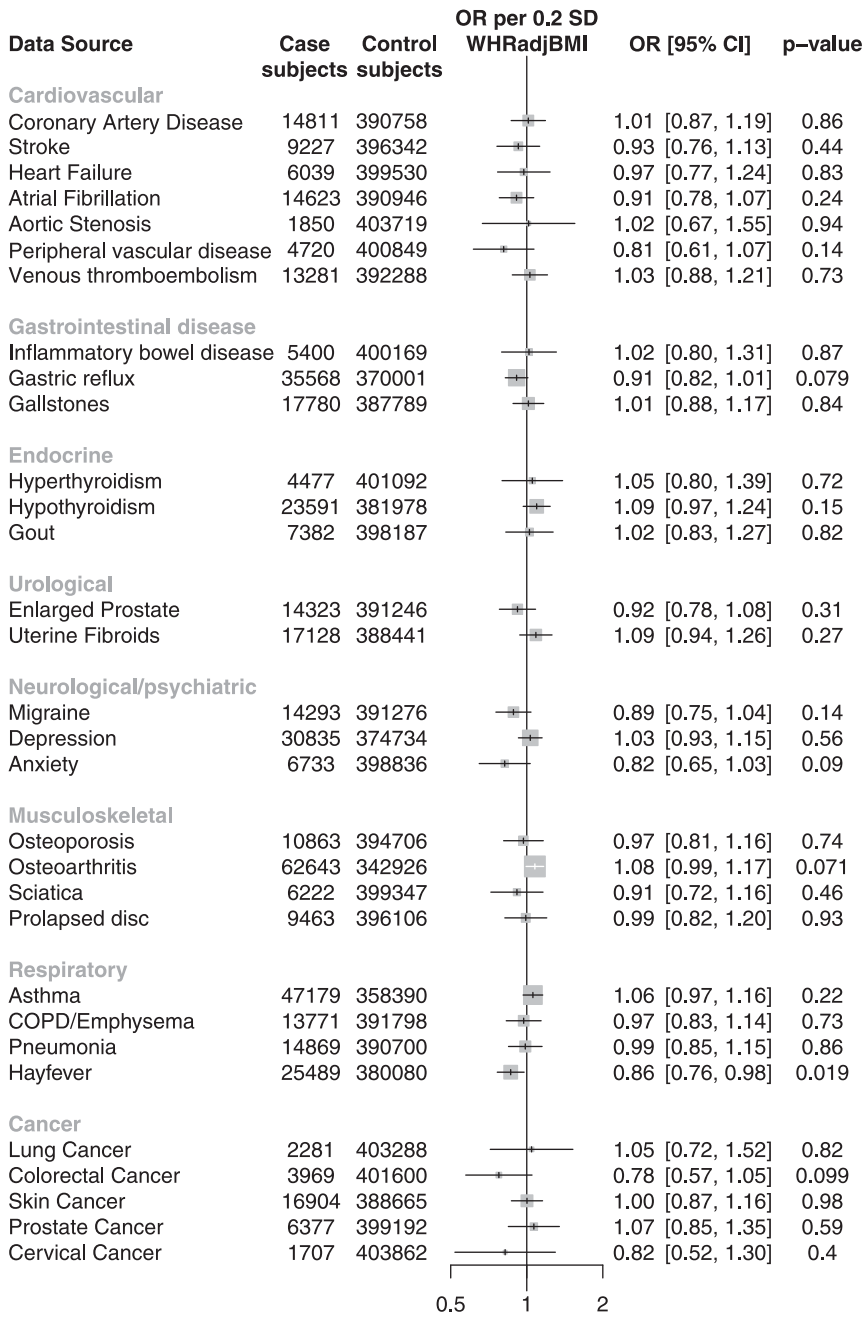



**Figure 4**—Association of *ACVR1C* gene risk score with 31 disease phenotypes in UK Biobank. COPD, chronic obstructive pulmonary disease.

Second, the lack of association of the *ACVR1C* gene risk score with 31 different diseases in UK Biobank suggests that therapeutic *ACVR1C* inhibition may not have adverse on-target effects. In the Exome Aggregation Consortium (ExAC), *ACVR1C* is tolerant of loss-of-function variants, with 18 of 66,720 individuals of European ancestry carrying an early stop codon (Leu32Ter) in *ACVR1C* (42). In combination with the phenome-wide association study presented here, these findings suggest that *ACVR1C* could be safely inhibited. However, due to the small number of case subjects for many of the analyzed diseases and the multiple diseases tested in the phenome-wide association study, modest associations of *ACVR1C* variation with the analyzed phenotypes cannot excluded. Furthermore, many common adverse effects of therapeutics, such as elevations in liver function enzymes, could not be analyzed due to a lack of available data in UK Biobank. In particular, ALK7-deficient mice have been reported to have lengthened QT intervals (43), a phenotype that was unavailable for analysis in UK Biobank. The association of the *ACVR1C* gene risk score with nominally lower diastolic and systolic blood pressure suggests that *ACVR1C*

inhibition may have the additional benefit of lowering blood pressure. However, this finding, which did not reach genome-wide significance, requires replication in independent data sets.

The primary strength of the analysis is the use of multiple data sources to replicate the association of *ACVR1C* variation with WHRadjBMI and type 2 diabetes. We demonstrated associations of *ACVR1C* with WHRadjBMI and type 2 diabetes to greater than genome-wide significance ($P = 2.6 \times 10^{-35}$ for WHRadjBMI and $P = 5.6 \times 10^{-13}$ for type 2 diabetes). We further showed that variants leading to loss of *ACVR1C* function protected against diabetes in an analysis of predicted damaging variants in exome sequences of 55,516 individuals. A primary limitation of the analysis is that we did not experimentally characterize the analyzed variants. The consistent protective associations observed with three different *ACVR1C* missense variants, with a noncoding variant that reduces *ACVR1C* expression, and with variants predicted to truncate the *ACVR1C* protein or predicted to damage *ACVR1C* function by five different algorithms suggest that variants leading to loss of *ACVR1C* function protect against type 2 diabetes. However, experimental demonstration that *ACVR1C* variants that lower WHRadjBMI and protect against type 2 diabetes actually reduce ALK7 receptor function is necessary before it can be concluded that *ACVR1C* deficiency will protect against type 2 diabetes. A second limitation is that pharmaceutical *ACVR1C* inhibition may be associated with off-target effects that cannot be characterized in a human genetic study.

In summary, variants predicted to damage *ACVR1C* gene function lower WHRadjBMI and protect against type 2 diabetes. These findings provide human genetic validation for the *ACVR1C* gene as a therapeutic target for type 2 diabetes.

**Funding and Duality of Interest.** This research has been conducted using the UK Biobank resource, application 7089. This work was funded by the National Institutes of Health (R01 HL127564 to S.K.), which had no involvement in the design and conduct of the study; the collection, analysis, and interpretation of the data; or the preparation, review, and approval of the manuscript. This project was also conducted using the Type 2 Diabetes Knowledge Portal resource, which is funded by the Accelerating Medicines Partnership. The REGICOR study was supported by the Spanish Ministry of Economy and Innovation through the Carlos III Health Institute (Red Investigación Cardiovascular RD12/0042, PI09/90506), European Regional Development Fund (ERDF), and the Catalan Research and Technology Innovation Interdepartmental Commission (2014SGR240). Samples for the Leicester cohort were collected as part of projects funded by the British Heart Foundation (British Heart Foundation Family Heart Study, British Heart Foundation grant RG2000010, United Kingdom Aneurysm Growth Study [UKAGS], British Heart Foundation grant CS/14/2/30841) and the National Institute for Health Research (NIHR) (NIHR Leicester Cardiovascular Biomedical Research Unit Biomedical Research Informatics Centre for Cardiovascular Science, IS_BRU_0211_20033). N.J.S. is supported by the British Heart Foundation and is an NIHR Senior Investigator. The Northern German Myocardial Infarction Study is supported by the German Federal Ministry of Education and Research (BMBF) in the context of the e:Med program (e:AtheroSysMed) and the FP7 European Union project CVgenes@target (261123). Additional grants were received from the Fondation Leducq (CADgenomics: Understanding Coronary Artery Disease Genes, 12CVD02). This study was also supported through the Deutsche Forschungsgemeinschaft cluster of excellence "Inflammation at Interfaces" and SFB 1123. The Italian ATVB study was supported by a grant from RFPS-2007-3-644382 and Programma di ricerca Regione-Università 2010-2012 Area 1–Strategic Programmes–Regione Emilia-Romagna. Funding for ESP was provided by RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO), and RC2 HL102924 (WHISP). Exome sequencing was performed through RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). The JHS is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, and HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities. J.G.W. is supported by U54GM115428 from the National Institute of General Medical Sciences. Exome sequencing in ATVB, PROCARDIS, Ottawa, PROMIS, South German Myocardial Infarction Study, and the JHS was supported by 5U54HG003067 (to S.G.). A.V.K. is supported by a K08 from the National Human Genome Research Institute (K08HG010155) and a Junior Faculty Award from the National Lipid Association and has received consulting fees from Amarin. P.N. reports funding from the John S. LaDue Memorial Fellowship at Harvard Medical School and has received consulting fees from Amarin. S.K. is supported by a research scholar award from Massachusetts General Hospital, the Donovan Family Foundation, and R01 HL127564; has received a research grant from Bayer Healthcare and consulting fees from Merck, Novartis, Sanofi, AstraZeneca, Alnylam Pharmaceuticals, Leerink Partners, Noble Insights, MedGenome, Aegerion Pharmaceuticals, Regeneron Pharmaceuticals, Quest Diagnostics, Color Genomics, Genomics PLC, and Eli Lilly and Company; and holds equity in San Therapeutics, Catabasis Pharmaceuticals, and Endcadia. No other potential conflicts of interest relevant to this article were reported.

**Author Contributions.** C.A.E. and S.K. contributed to study concept and design. C.A.E., A.V.K., M.C., D.S., S.G., and S.K. contributed to acquisition, analysis, or interpretation of data. C.A.E., A.V.K., and S.K. drafted the manuscript. S.K. provided administrative, technical, and material support. All authors critically revised the manuscript for important intellectual content. C.A.E. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## References

1. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 2013;93:779–797
2. Abifadel M, Varret M, Rabès J-P, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 2003;34:154–156
3. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264–1272
4. Sabatine MS, Giugliano RP, Keech AC, et al.; FOURIER Steering Committee and Investigators. Evolocumab and clinical outcomes in patients with cardiovascular disease. N Engl J Med 2017;376:1713–1722
5. Stitziel NO, Khera AV, Wang X, et al.; PROMIS and Myocardial Infarction Genetics Consortium Investigators. ANGPTL3 deficiency and protection against coronary artery disease. J Am Coll Cardiol 2017;69:2054–2063
6. Dewey FE, Gusarova V, Dunbar RL, et al. Genetic and pharmacologic inactivation of ANGPTL3 and cardiovascular disease. N Engl J Med 2017;377:211–221
7. Graham MJ, Lee RG, Brandt TA, et al. Cardiovascular and metabolic effects of ANGPTL3 antisense oligonucleotides. N Engl J Med 2017;377:222–232
8. Pollin TI, Damcott CM, Shen H, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. Science 2008;322:1702–1705
9. Crosby J, Peloso GM, Auer PL, et al.; TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. N Engl J Med 2014;371:22–31
10. Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjærg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. N Engl J Med 2014;371:32–41

11. Gaudet D, Alexander VJ, Baker BF, et al. Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia. N Engl J Med 2015;373438–447
12. Emdin C, Khera AV, Natarajan P, et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. JAMA 2017;317:626–634
13. Yusuf S, Hawken S, Ounpuu S, et al.; INTERHEART Sudy Investigators. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. Lancet 2005;366:1640–1649
14. Dale CE, Fatemifar G, Palmer TM, et al.; UCLEB Consortium; METASTROKE Consortium. Causal associations of adiposity and body fat distribution with coronary heart disease, stroke subtypes, and type 2 diabetes mellitus: a Mendelian randomization analysis. Circulation 2017;135:2373–2388
15. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7
16. Shungin D, Winkler TW, Croteau-Chonka DC, et al.; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; ReproGen Consortium. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015;518:187–196
17. DeBoever C, Tanigawa Y, Lindholm ME, et al. Medical relevance of protein-truncating variants across 337,205 individuals in the UK Biobank study. Nat Commun 2018;9:1612
18. Mahajan A, Wessel J, Willems SM, et al.; ExomeBP Consortium; MAGIC Consortium; GIANT Consortium. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet 2018;50:559–571
19. Do R, Stitziel NO, Won H-H, et al.; NHLBI Exome Sequencing Project. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature 2015;518:102–106
20. Khera AV, Won H-H, Peloso GM, et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. J Am Coll Cardiol 2016;67:2578–2589
21. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature 2016;536:41–47
22. Emdin CA, Khera AV, Klarin D, et al.; American Heart Association. Phenotypic consequences of a genetic predisposition to enhanced nitric oxide signaling. Circulation 2018;137:222–232
23. ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. Am J Epidemiol 1989;129:687–702
24. Khera AV, Chaffin M, Aragam K, et al. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. Nat Genet 2018;50:1219–1224
25. Emdin CA, Khera AV, Natarajan P, et al.; CHARGE–Heart Failure Consortium; CARDIoGRAM Exome Consortium. Phenotypic characterization of genetically lowered human lipoprotein(a) levels. J Am Coll Cardiol 2016;68:2761–2772
26. Khera AV, Won H-H, Peloso GM, et al.; Myocardial Infarction Consortium; DiscovEHR Study Group, CARDIoGRAM Exome Consortium; Global Lipids Genetics Consortium. Association of rare and common variation in the lipoprotein lipase gene with coronary artery disease. JAMA 2017;317:937–946
27. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010;26:2069–2070
28. Park J-H, Gail MH, Weinberg CR, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc Natl Acad Sci U S A 2011;108:18026–18031
29. Liu DJ, Peloso GM, Yu H, et al.; Charge Diabetes Working Group; EPIC-InterAct Consortium; EPIC-CVD Consortium; GOLD Consortium; VA Million Veteran Program. Exome-wide association study of plasma lipids in >300,000 individuals. Nat Genet 2017;49:1758–1766
30. Emdin CA, Klarin D, Natarajan P, Florez JC, Kathiresan S, Khera AV; CARDIOGRAM Exome Consortium. Genetic variation at the sulfonylurea receptor, type 2 diabetes, and coronary heart disease. Diabetes 2017;66:2310–2315
31. Zimmermann R, Strauss JG, Haemmerle G, et al. Fat mobilization in adipose tissue is promoted by adipose triglyceride lipase. Science 2004;306:1383–1386
32. Xia W, Pessentheiner AR, Hofer DC, et al. Loss of ABHD15 impairs the anti-lipolytic action of insulin by altering PDE3B stability and contributes to insulin resistance. Cell Reports 2018;23:1948–1961
33. Dupuis J, Langenberg C, Prokopenko I, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42:105–116
34. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648–660
35. Bondestam J, Huotari MA, Morén A, et al. cDNA cloning, expression studies and chromosome mapping of human type I serine/threonine kinase receptor ALK7 (ACVR1C). Cytogenet Cell Genet 2001;95:157–162
36. Zhang N, Kumar M, Xu G, et al. Activin receptor-like kinase 7 induces apoptosis of pancreatic beta cells and beta cell lines. Diabetologia 2006;49:506–518
37. Yogosawa S, Mizutani S, Ogawa Y, Izumi T. Activin receptor-like kinase 7 suppresses lipolysis to accumulate fat in obesity through downregulation of peroxisome proliferator-activated receptor $\gamma$ and C/EBP$\alpha$. Diabetes 2013;62:115–123
38. Zhao F, Huang F, Tang M, et al. Nodal induces apoptosis through activation of the ALK7 signaling pathway in pancreatic INS-1 $\beta$-cells. Am J Physiol Endocrinol Metab 2012;303:E132–E143
39. Andersson O, Korach-Andre M, Reissmann E, Ibáñez CF, Bertolino P. Growth/differentiation factor 3 signals through ALK7 and regulates accumulation of adipose tissue and diet-induced obesity. Proc Natl Acad Sci U S A 2008;105:7252–7256
40. Guo T, Marmol P, Moliner A, et al. Adipocyte ALK7 links nutrient overload to catecholamine resistance in obesity. eLife 2014;3:e03245
41. Li W-B, Zhao J, Liu L, et al. Silencing of activin receptor-like kinase 7 alleviates aortic stiffness in type 2 diabetic rats. Acta Diabetol 2015;52:717–726
42. Lek M, Karczewski KJ, Minikel EV, et al.; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–291
43. Ying S, Cao H, Hu H, Wang X, Tang Y, Huang C. Alk7 depleted mice exhibit prolonged cardiac repolarization and are predisposed to ventricular arrhythmia. PLoS One 2016;11:e0149205

# Phenotypic Consequences of a Genetic Predisposition to Enhanced Nitric Oxide Signaling

Editorial, see p 233

Connor A. Emdin, DPhil
et al

**BACKGROUND:** Nitric oxide signaling plays a key role in the regulation of vascular tone and platelet activation. Here, we seek to understand the impact of a genetic predisposition to enhanced nitric oxide signaling on risk for cardiovascular diseases, thus informing the potential utility of pharmacological stimulation of the nitric oxide pathway as a therapeutic strategy.

**METHODS:** We analyzed the association of common and rare genetic variants in 2 genes that mediate nitric oxide signaling (Nitric Oxide Synthase 3 [*NOS3*] and Guanylate Cyclase 1, Soluble, Alpha 3 [*GUCY1A3*]) with a range of human phenotypes. We selected 2 common variants (rs3918226 in *NOS3* and rs7692387 in *GUCY1A3*) known to associate with increased *NOS3* and *GUCY1A3* expression and reduced mean arterial pressure, combined them into a genetic score, and standardized this exposure to a 5 mm Hg reduction in mean arterial pressure. Using individual-level data from 335 464 participants in the UK Biobank and summary association results from 7 large-scale genome-wide association studies, we examined the effect of this nitric oxide signaling score on cardiometabolic and other diseases. We also examined whether rare loss-of-function mutations in *NOS3* and *GUCY1A3* were associated with coronary heart disease using gene sequencing data from the Myocardial Infarction Genetics Consortium (n=27 815).

**RESULTS:** A genetic predisposition to enhanced nitric oxide signaling was associated with reduced risks of coronary heart disease (odds ratio, 0.37; 95% confidence interval [CI], 0.31–0.45; $P=5.5*10^{-26}$], peripheral arterial disease (odds ratio 0.42; 95% CI, 0.26–0.68; $P=0.0005$), and stroke (odds ratio, 0.53; 95% CI, 0.37–0.76; $P=0.0006$). In a mediation analysis, the effect of the genetic score on decreased coronary heart disease risk extended beyond its effect on blood pressure. Conversely, rare variants that inactivate the *NOS3* or *GUCY1A3* genes were associated with a 23 mm Hg higher systolic blood pressure (95% CI, 12–34; $P=5.6*10^{-5}$) and a 3-fold higher risk of coronary heart disease (odds ratio, 3.03; 95% CI, 1.29–7.12; $P=0.01$).

**CONCLUSIONS:** A genetic predisposition to enhanced nitric oxide signaling is associated with reduced risks of coronary heart disease, peripheral arterial disease, and stroke. Pharmacological stimulation of nitric oxide signaling may prove useful in the prevention or treatment of cardiovascular disease.

The full author list is available on page 229.

**Correspondence to:** Sekar Kathiresan, MD, Center for Genomic Medicine Massachusetts General Hospital, 185 Cambridge Street, CPZN 5.830, Boston, MA 02114. E-mail skathiresan1@mgh.harvard.edu

Sources of Funding, see page 229

**Key Words:** cardiovascular disease ■ genetics ■ nitric oxide ■ nitric oxide synthase

## Clinical Perspective

### What Is New?

- A genetic predisposition to enhanced nitric oxide signaling, as mediated by common variants in the *NOS3* and *GUCY1A3* genes, was associated with lower blood pressure and a reduced risk of coronary heart disease, peripheral arterial disease, and stroke.
- Rare variants that inactivate either the *NOS3* or *GUCY1A3* gene were associated with increased blood pressure and a higher risk of coronary heart disease.

### What Are the Clinical Implications?

- The results suggest that the pharmacological stimulation of nitric oxide signaling may be an effective therapy for the prevention or treatment of atherosclerotic cardiovascular disease.

Nitric oxide signaling is a key regulator of vascular tone, blood pressure, and platelet aggregation.[1,2] Endothelial nitric oxide synthase, encoded by the gene *NOS3*, generates nitric oxide in the vascular endothelium (Figure 1).[3] Nitric oxide acts as a signaling molecule to activate soluble guanylyl cyclase, a heterodimeric protein with 1 subunit encoded by the gene *GUCY1A3*.[3,4] Cyclic guanosine monophosphate produced by soluble guanylyl cyclase then activates downstream signaling molecules, leading to vasodilation, blood pressure lowering, inhibition of platelet aggregation, and other cardiometabolic effects.[5,6] *NOS3* and *GUCY1A3* are thus key mediators of nitric oxide signaling and its downstream effects (Figure 1).

Previous studies have noted increased blood pressure and atherosclerotic burden in endothelial nitric oxide synthase$^{-/-}$ knockout mice,[7,8] whereas loss of soluble guanylyl cyclase promotes thrombus formation.[9] Common noncoding variants in the *NOS3* and *GUCY1A3* loci associate with blood pressure and coronary heart disease (CHD) in genome-wide association studies, consistent with a role of nitric oxide signaling in the regulation of arterial blood pressure.[10–12] Furthermore, a rare coding variant in *GUCY1A3* was identified as associating with early onset myocardial infarction in a German family.[9] These findings in mice and humans suggest that stimulation of nitric oxide signaling may be a useful therapeutic strategy for the prevention of cardiovascular diseases.

DNA sequence variants in therapeutic target genes represent naturally occurring, lifelong variation of the gene (ie, experiments of nature). Consequently, if a genetic predisposition to enhanced nitric oxide signaling associates with a reduced risk of cardiovascular and other diseases, these results would support the therapeutic hypothesis that pharmacological stimulation of nitric oxide signaling (eg, through soluble guanylyl cyclase stimulation[13]) will prevent or treat cardiovascular disease. Furthermore, if the association of nitric oxide signaling with CHD risk is partially mediated through blood pressure-independent pathways, stimulation of nitric oxide signaling may represent an approach for CHD prevention independent of current blood pressure-lowering therapies.

We therefore sought to use common and rare DNA sequence variants in *NOS3* and *GUCY1A3* to: (1) determine the effects of a genetic predisposition to enhanced nitric oxide signaling on a range of cardiometabolic and other diseases, (2) assess whether the effect of nitric oxide signaling on CHD is primarily mediated through blood pressure, and (3) determine whether rare inactivating variants in *NOS3* and *GUCY1A3* that are predicted to reduce nitric oxide signaling associate with higher risk for CHD.

## METHODS

The data, analytic methods, and study materials are available to other researchers for the purposes of reproducing the results or replicating the procedure, with access maintained by UK Biobank.[14]

### Study Design, Data Sources, and Study Participants

The study design is shown in Figure I in the online-only Data Supplement. We used individual-level data from 335 464 individuals of European ancestry from the UK Biobank, a large population-based cohort (Methods A in the online-only Data Supplement). Characteristics of individuals in UK Biobank are provided (Table I in the online-only Data Supplement). We supplemented this individual-level data with 7 genome-wide association study consortia examining blood lipids, anthropometric traits, glycemic traits, diabetes mellitus, CHD, migraine, and renal dysfunction, all predominantly containing individuals of European descent (Methods B and Table II in the online-only Data Supplement). Finally, we used gene sequence data from 27 815 participants from the Myocardial Infarction Genetics Consortium and 16 857 participants from the T2D GENES study (Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples)[15] to examine whether rare variants in the *NOS3-GUCY1A3* pathway associate with blood pressure and CHD (Methods C in the online-only Data Supplement).

In our primary analysis, we examined the effect of a genetic predisposition to enhanced nitric oxide signaling on 9 different cardiometabolic diseases: CHD, stroke, heart failure, atrial fibrillation, aortic stenosis, peripheral vascular disease, venous thromboembolism, diabetes mellitus, and chronic kidney disease (Table III in the online-only Data Supplement). We additionally examined the effect of a genetic predisposition to enhanced nitric oxide signaling on 16 quantitative traits (Methods A and B in the online-only Data Supplement): systolic blood pressure, diastolic blood pressure, waist-to-hip ratio adjusted for body mass index,[16] body mass index, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, fasting glucose, fasting

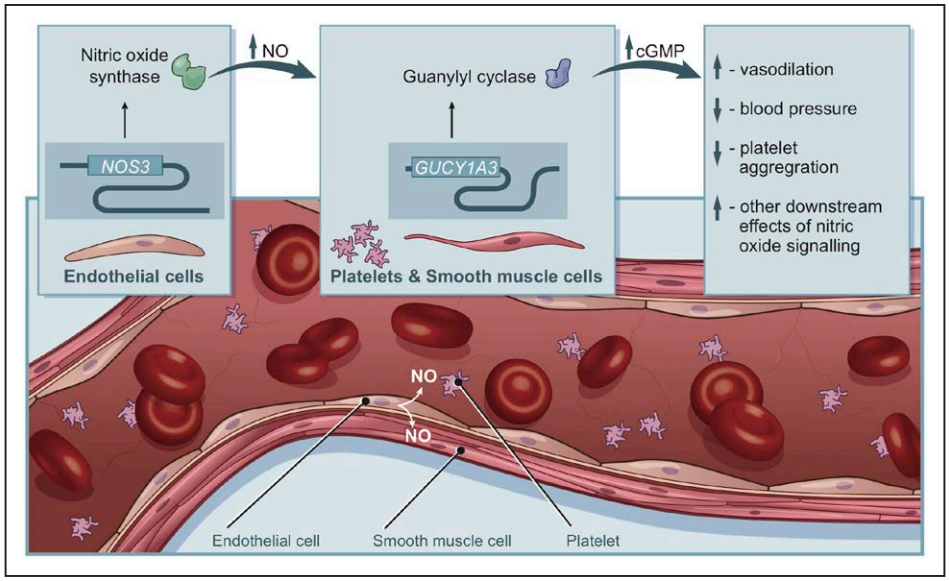



**Figure 1. Role of *NOS3* and *GUCY1A3* in nitric oxide signaling.**
Endothelial nitric oxide synthase, encoded by the gene *NOS3*, generates nitric oxide in the vascular endothelium. Nitric oxide acts as a signaling molecule to activate soluble guanylyl cyclase, a heterodimeric protein with 1 subunit encoded by the gene *GUCY1A3*. Cyclic guanosine monophosphate produced by soluble guanylyl cyclase then activates downstream signaling molecules, leading to vasodilation, blood pressure lowering, inhibition of platelet aggregation, and other cardiometabolic effects. cGMP indicates cyclic guanosine monophosphate; and NO, nitric oxide.

insulin, 2-hour glucose, hemoglobin A1C, serum creatinine-estimated glomerular filtration rate, cystatin-C-estimated glomerular filtration rate, forced expiratory volume in 1 second, and the ratio of forced expiratory volume in 1 second to forced vital capacity. All traits were standardized (ie, reported in standard deviation [SD] units) to facilitate comparisons among traits (Methods B in the online-only Data Supplement). Using the UK Biobank cohort, we conducted a phenome-wide association study for 26 additional diseases, including endocrine, renal, urologic, gastrointestinal, neurological, musculoskeletal, respiratory, and neoplastic disorders (Table III in the online-only Data Supplement).

Analysis of the UK Biobank data was approved by the Partners Health Care institutional review board (protocol 2013P001840; application 7089). Informed consent was obtained from all participants by the UK Biobank.

## Common Variants in *NOS3* and *GUCY1A3*

We leveraged common variants in the *NOS3* and *GUCY1A3* loci to characterize the effects of a genetic predisposition to enhanced nitric oxide signaling. Two common variants were selected as instruments for a genetic predisposition to enhanced nitric oxide signaling: a promoter variant of *NOS3* (rs3918226; minor allele frequency 8%) and an intronic variant of *GUCY1A3* (rs7692387; minor allele frequency 19%) (Table IV in the online-only Data Supplement). These variants were selected because they: (1) have been robustly associated with blood pressure (a downstream effect of nitric oxide signaling); and (2) are located within the endothelial nitric oxide synthase-soluble guanylyl cyclase signaling pathway.[17,18] Furthermore, the minor allele of rs7692387 has also recently been characterized to reduce *GUCY1A3* expression via disruption of a ZEB1 transcription factor site.[19] We examined the effect of these variants on *NOS3* and *GUCY1A3* expression levels in aortic and lung tissue in the Genotype-Tissue Expression project database[20] (Methods D in the online-only Data Supplement) and the effect of these variants on mean arterial pressure in the UK Biobank.

To examine the effects of increased nitric oxide signaling on cardiometabolic and other traits, we pooled rs3918226 and rs7692387 into an additive genetic score (individuals had 0–4 risk alleles). We derived this score by multiplying the number of risk alleles by the association of each allele with mean arterial pressure in the UK Biobank (0.68 mm Hg for rs3918226 and 0.32 mm Hg for rs7692387). For example, if an individual had 2 risk alleles for rs3918226 and 1 risk allele for rs7692387, the genetic score was calculated as 2*0.68 mm Hg + 0.32 mm Hg = 1.68 mm Hg. Individuals missing 1 variant (either rs3918226 or rs7692387; n=6036 [1.8%]) were imputed at the mean allele frequency for the variant prior to calculation of the genetic score. We standardized the nitric oxide signaling genetic score to a 5 mm Hg reduction in mean arterial pressure, corresponding to the effect of 1.5 mg of riociguat (a pharmacological stimulator of soluble guanylate cyclase) on mean arterial pressure in a recent randomized controlled trial.[13] The standardization was performed by dividing the genetic nitric oxide score by 5 mm Hg (eg, 1.68 mm Hg/5 mm Hg=0.34). We also report estimates standardized to a 2.5 mm Hg and 10 mm Hg lower mean arterial pressure to clarify the expected reduction in risk of cardiometabolic outcomes with different levels of nitric oxide signaling.

## Rare Predicted Loss-of-Function Variants in *NOS3* and *GUCY1A3*

Common noncoding variants may influence the expression of several nearby genes.[21] Therefore, to provide complementary

evidence that the *NOS3–GUCY1A3* pathway influences blood pressure and CHD risk, we examined whether rare (minor allele frequency <1%) predicted loss-of-function variants in *NOS3* and *GUCY1A3* are associated with systolic blood pressure, diastolic blood pressure, and CHD (Methods C in the online-only Data Supplement). Predicted loss-of-function variants were defined as: (1) insertions or deletions of DNA that modify the reading frame of protein translation (frameshift); (2) point mutations at conserved splice site regions that alter the splicing process (splice-site); or (3) point mutations that change an amino acid codon to a stop codon, leading to the truncation of a protein (nonsense).

For blood pressure, we tested whether presence of a predicted loss-of-function variant in *NOS3* and *GUCY1A3* was associated with systolic blood pressure or diastolic blood pressure in the T2D GENES study (n=16 857). We examined whether predicted loss-of-function variants were associated with systolic and diastolic blood pressure using the Genetic Association Interactive Tool on the Type 2 Diabetes Knowledge Portal.[15] We used linear regression and adjusted for age, sex, and 5 principal components of ancestry.

For CHD, we tested whether presence of a predicted loss-of-function variant was associated with CHD in the Myocardial Infarction Genetics Consortium (n=27 815) study using logistic regression, adjusted for sex, 5 principal components of ancestry, and a dummy variable for each cohort. To examine whether variants in the *NOS3–GUCY1A3* pathway associate with CHD risk, we pooled the effect of predicted loss-of-function variants in *NOS3* and *GUCY1A3* on CHD using inverse variance weighted fixed effects meta-analysis.

## Statistical Analysis

For the UK Biobank, we estimated the association of the nitric oxide signaling genetic score (standardized to a 5 mm Hg decrease in mean arterial pressure) with each outcome using a logistic regression model adjusting for age, sex, 10 principal components of ancestry, and a dummy variable for array type. For the summary-level data, this approach is equivalent to an inverse variance weighted fixed effects meta-analysis of the effect of each variant on traits or outcome of interest per 5 mm Hg lower mean arterial pressure. Tests for interaction between UK Biobank and summary-level estimates were calculated as the difference in log-transformed relative risks, as previously described.[22]

For our primary outcomes (9 cardiometabolic diseases), we set a Bonferroni adjusted level of significance of $P$=0.05/9=0.0056. For our secondary analyses of 16 cardiometabolic and pulmonary traits and our phenome-wide association study of 26 phenotypes, we set a level of significance of $P$=0.05/42=0.001.

To examine whether an observed reduction in risk of CHD was caused by reduced blood pressure, a mediation analysis was conducted. An estimate of the causal effect of systolic blood pressure on CHD risk was derived from a recent genome-wide association study (odds ratio [OR], 1.21; 95% confidence interval [CI], 1.17–1.24 per 5 mm Hg higher systolic blood pressure) (Methods E in the online-only Data Supplement).[11] This effect was then multiplied by the decrease in systolic blood pressure because of nitric oxide signaling to estimate the decrease in CHD risk mediated by systolic blood pressure. We then subtracted this estimate from the overall estimate of the nitric oxide genetic score with CHD to derive the remaining proportion of CHD risk unaccounted for by a decrease in systolic blood pressure. For example, if the odds ratio for CHD for the genetic score was 0.5 but the odds ratio for CHD from the systolic blood pressure decrease was 0.75, the OR for CHD independent of systolic blood pressure was calculated as exp(log(0.5) – log(0.75))=0.67.

All analyses were performed using R version 3.2.3 software (The R Project for Statistical Computing).

# RESULTS

In the Genome Tissue Expression project database,[20] the C allele of rs3918226 was associated with increased *NOS3* expression in lung tissue (48% higher expression; $P$=0.002) but was not significantly associated with *NOS3* expression in aortic tissue (23% higher expression; $P$=0.21) (Figure 2). The A allele of rs7692387 was associated with increased *GUCY1A3* expression in aortic tissue (20% higher expression; $P$=0.012) but was not significantly associated with *GUCY1A3* expression in lungs (7% higher expression, $P$=0.20). As expected for variants that enhance nitric oxide signaling, both rs3918226 and rs7692387 were associated with lower mean arterial pressure among UK Biobank participants (0.68 mm Hg lower mean arterial pressure [$P$=1.3*10$^{-26}$] and 0.32 mm Hg lower mean arterial pressure [$P$=8.3*10$^{-14}$], respectively) (Figure 2), replicating previously reported associations of these variants with blood pressure at genome-wide significance.[17,18]

When combined into a nitric oxide-signaling genetic score and standardized to a 5 mm Hg reduction in mean arterial pressure, a genetic predisposition to enhanced nitric oxide signaling was nominally associated with improved renal function, as assessed by both cystatin C-estimated glomerular filtration rate (7.6 mL/min; 95% CI, 2.9–12 mL/min; $P$=0.0015) and creatinine-estimated glomerular filtration rate (2.8 mL/min; 95% CI, 0.57–5.1 mL/min; $P$=0.014) (Figure 3).

Enhanced nitric oxide signaling was significantly associated with higher forced expiratory volume in 1 second (0.09 L; 95% CI, 0.05–0.12; $P$=5.6*10$^{-7}$). We examined whether the association of a genetic predisposition to enhanced nitric oxide signaling with pulmonary function differed by baseline pulmonary function. No evidence of a trend by baseline pulmonary function for forced expiratory volume in 1 second was observed ($P$=0.34, respectively) (Figure II in the online-only Data Supplement). Enhanced nitric oxide signaling was not significantly associated with any other cardiometabolic traits, including glycemic traits or blood lipids (Figure 3).

The genetic nitric oxide signaling score was significantly associated with 3 of the 9 primary cardiometabolic outcomes. A genetic predisposition to enhanced nitric oxide signaling was associated with reduced risk

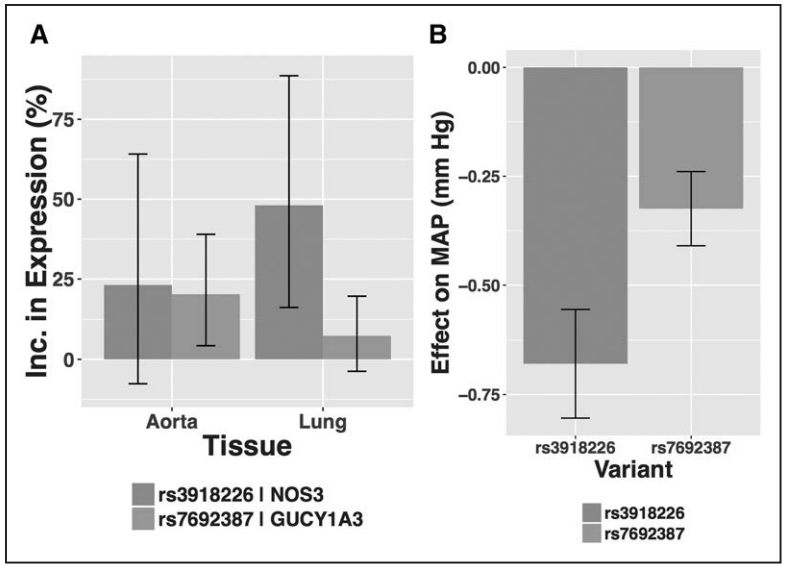



**Figure 2. Association of *NOS3* and *GUCY1A3* variants with gene expression and blood pressure.**
Association of rs3918226 and rs7692387 with (**A**) *NOS3* or *GUCY1A3* expression levels in aortic and lung tissue, and (**B**) mean arterial pressure among UK Biobank participants. Effect of variants on expression levels were obtained from the Genotype-Tissue Expression project. Effect of variants on mean arterial pressure were derived in the UK Biobank using linear regression adjusted for age, sex, 10 principal components of ancestry, and array type (least squares means estimates). rs3918226 was associated with significantly elevated NOS3 expression in the lung. rs7692387 was associated with significantly elevated *GUCY1A3* expression in the aorta. MAP indicates mean arterial pressure.

of CHD (OR, 0.36; 95% CI, 0.29–0.46; $P$=7.0*10$^{-17}$) (Figure III in the online-only Data Supplement) in the CARDIoGRAMPlusC4D Consortium data. We replicated this association in UK Biobank participants (OR, 0.39; 95% CI, 0.29–0.52; $P$=1.1*10$^{-10}$), with an overall pooled 63% reduction in CHD risk (OR, 0.37; 95% CI, 0.31–0.45; $P$=5.5*10$^{-26}$) (Figure 4).

Beyond CHD, the nitric oxide signaling genetic score was associated with a reduced risk of peripheral arterial disease (OR, 0.42; 95% CI, 0.26–0.68; $P$=0.0005). This association persisted in a sensitivity analysis that excluded individuals with concomitant CHD (OR, 0.41; 95% CI, 0.23–0.74; $P$=0.003). The genetic score was also associated with a reduced risk of stroke (OR, 0.53; 95% CI, 0.37–0.76; $P$=0.0006) and a nominally ($P$<0.05) reduced risk of chronic kidney disease (OR, 0.46; 95% CI, 0.25–0.83; $P$=0.011), heart failure (OR, 0.59; 95% CI, 0.38–0.92; $P$=0.02), and diabetes mellitus (OR, 0.81;

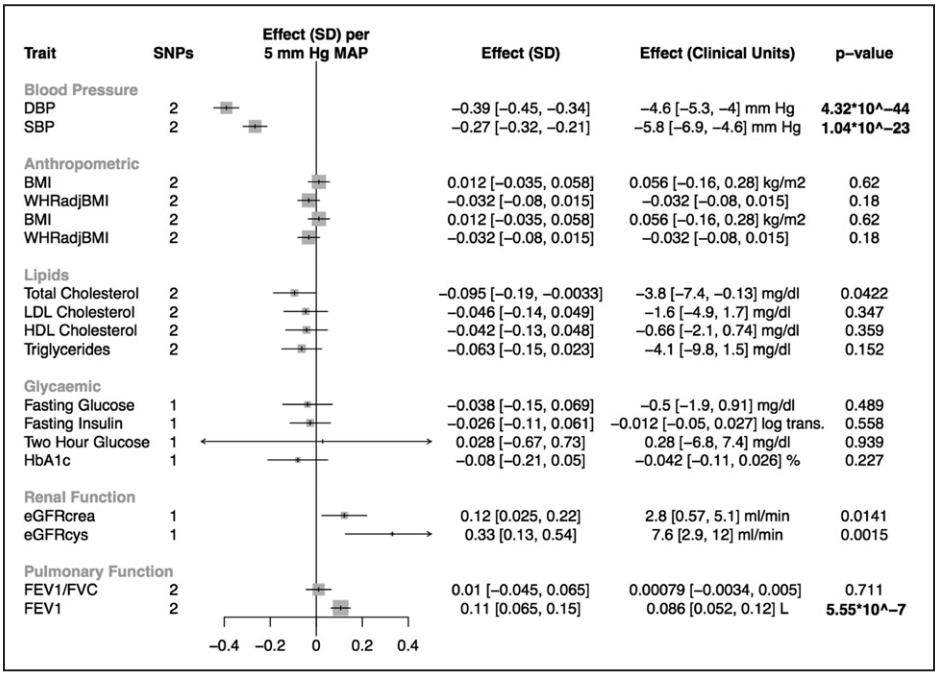
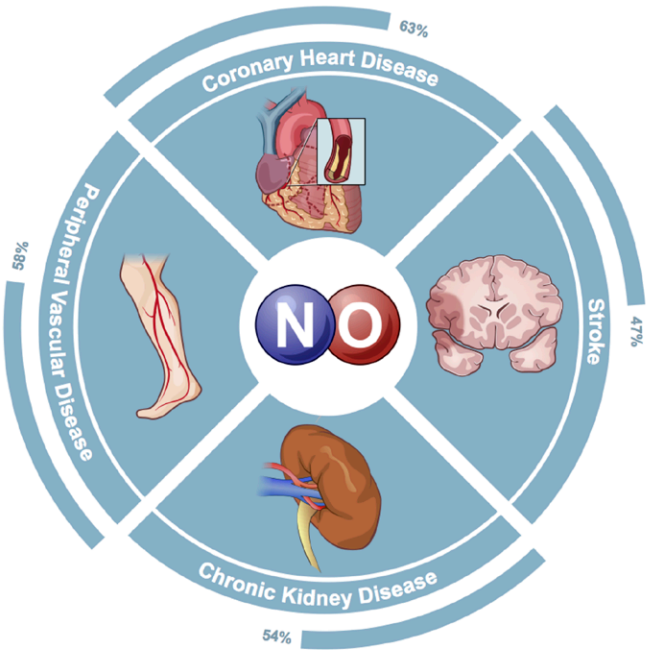



**Figure 3. Association of the nitric oxide signaling genetic score with cardiometabolic traits (secondary outcomes).**
BMI indicates body mass index; eGFRcrea, serum creatinine-estimated glomerular filtration rate; eGFRcys, cystatin-C-estimated glomerular filtration rate; DBP, diastolic blood pressure; FEV1, forced expiratory volume in 1 second; FVC, forced vital capacity; HbA1c, hemoglobin A1c; HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol; SBP, systolic blood pressure; SD, standard deviation; and WHRadjBMI, waist-to-hip ratio adjusted for body mass index. *P* values that reach a Bonferroni-adjusted level of statistical significance ($P$<0.001) are in boldface.




| Outcome | Cases | Controls | OR per 5 mmHg MAP | p-value |
|---|---|---|---|---|
| **Cardiometabolic disease (Primary)** | | | | |
| Aortic Stenosis | 1577 | 333887 | 0.90 [0.41; 2.00] | 0.8 |
| Atrial Fibrillation | 12764 | 322700 | 0.96 [0.72; 1.28] | 0.78 |
| Chronic Kidney Disease | 12385 | 104780 | 0.46 [0.25; 0.83] | 0.011 |
| Coronary Heart Disease | 73246 | 446523 | 0.37 [0.31; 0.45] | 5.5*10^-26 |
| Diabetes | 50567 | 434718 | 0.81 [0.67; 0.99] | 0.037 |
| Heart Failure | 5051 | 330413 | 0.59 [0.38; 0.92] | 0.021 |
| Peripheral Arterial Disease | 4011 | 331453 | 0.42 [0.26; 0.68] | 0.00046 |
| Stroke | 7772 | 327692 | 0.53 [0.37; 0.76] | 0.00058 |
| Venous Thromboembolism | 11346 | 324118 | 0.95 [0.70; 1.29] | 0.75 |
| **PheWAS (Secondary)** | | | | |
| *Gastrointestinal Disease* | | | | |
| Inflammatory Bowel Disease | 4506 | 330958 | 0.83 [0.52; 1.33] | 0.44 |
| Gastric Reflux | 30024 | 305440 | 0.91 [0.75; 1.10] | 0.33 |
| Gallstones | 15244 | 320220 | 1.12 [0.86; 1.46] | 0.4 |
| Irritable Bowel Syndrome | 7966 | 327498 | 0.68 [0.47; 0.96] | 0.03 |
| *Endocrine* | | | | |
| Hyperthyroidism | 3634 | 331830 | 0.95 [0.56; 1.61] | 0.85 |
| Hypothyroidism | 19684 | 315780 | 0.74 [0.59; 0.93] | 0.011 |
| Gout | 6174 | 329290 | 0.55 [0.37; 0.83] | 0.004 |
| *Urological* | | | | |
| Enlarged Prostate | 12143 | 323321 | 0.98 [0.72; 1.33] | 0.89 |
| Uterine Fibroids | 13303 | 322161 | 0.85 [0.64; 1.13] | 0.27 |
| *Neurological/psychiatric* | | | | |
| Migraine | 71777 | 639439 | 0.90 [0.80; 1.01] | 0.082 |
| Depression | 25822 | 309642 | 0.90 [0.74; 1.11] | 0.33 |
| Anxiety | 5592 | 329872 | 1.11 [0.72; 1.69] | 0.64 |
| *Muscoskeletal* | | | | |
| Osteoporosis | 9043 | 326421 | 1.00 [0.71; 1.40] | 0.99 |
| Osteoarthritis | 53922 | 281542 | 1.07 [0.92; 1.25] | 0.36 |
| Sciatica | 5151 | 330313 | 0.63 [0.41; 0.97] | 0.037 |
| Prolapsed Disc | 7880 | 327584 | 0.78 [0.55; 1.12] | 0.18 |
| *Respiratory* | | | | |
| Asthma | 39146 | 296318 | 0.93 [0.79; 1.11] | 0.42 |
| COPD/Emphysema | 11664 | 323800 | 0.80 [0.59; 1.07] | 0.14 |
| Pneumonia | 12647 | 322817 | 0.95 [0.71; 1.27] | 0.73 |
| Hayfever | 21121 | 314343 | 1.38 [1.10; 1.73] | 0.0056 |
| *Cancer* | | | | |
| Breast Cancer | 10885 | 324579 | 1.03 [0.75; 1.40] | 0.87 |
| Colorectal Cancer | 3425 | 332039 | 1.63 [0.94; 2.84] | 0.08 |
| Skin Cancer | 15011 | 320453 | 1.14 [0.88; 1.49] | 0.33 |
| Prostate Cancer | 5481 | 329983 | 1.23 [0.79; 1.92] | 0.36 |
| Cervical Cancer | 1435 | 334029 | 1.20 [0.52; 2.79] | 0.67 |
| Other Cancer | 7506 | 327958 | 1.09 [0.75; 1.57] | 0.66 |



**Figure 4. Association of the nitric oxide-signaling genetic score with cardiometabolic (primary) and other diseases (secondary).**

Estimates were derived in the UK Biobank using logistic regression, adjusted for age, sex, 10 principal components, and array type, with the exception of chronic kidney disease, which was derived using summary statistics from CKDGen. Estimates for coronary heart disease, diabetes mellitus, and migraine additionally included summary estimates from CARDIOGRAM (Coronary Artery Disease Genome Wide Replication and Meta-Analysis), DIAGRAM (Diabetes Genetics Replication and Meta-Analysis), and IHGC (International Headache Genetics Consortium) and were pooled using inverse variance weighted fixed effects meta-analysis. COPD indicates chronic obstructive pulmonary disease; MAP, mean arterial pressure; OR, odds ratio; PheWAS, phenome-wide association study; and SD, standard deviation. Significant *P* values are in boldface.

95% CI, 0.67–0.99; *P*=0.037). In a phenome-wide association study of 26 different diseases, enhanced nitric oxide signaling was not significantly associated with any other disease, including a variety of gastrointestinal diseases, musculoskeletal diseases, and cancers (Figure 4). In sensitivity analyses, when standardized to a 2.5 or 10 mm Hg lower mean arterial reduction, the nitric oxide signaling genetic score was associated with a 39% (OR, 0.61; 95% CI, 0.56–0.67) and 86% lower risk of CHD (OR, 0.14; 95% CI, 0.10–0.20), respectively (Figure IV in the online-only Data Supplement).

When an interaction term between the *NOS3* and *GUCY1A3* genetic loci was included, no evidence of an interaction in the association of nitric oxide signaling with systolic blood pressure (*P* interaction=0.76) or diastolic blood pressure (*P* interaction=0.49) was observed. Similarly, no evidence of an interaction between the *NOS3* and *GUCY1A3* loci was observed for CHD (*P* interaction=0.53).

We performed mediation analysis to determine whether the degree to which the change in blood pressure associated with the nitric oxide signaling score explained the protective effect on CHD. Adjustment for the effect on systolic blood pressure led only to a modest attenuation of the association of the nitric oxide signaling genetic score with CHD (OR, 0.37; 95% CI, 0.31–0.45 prior to adjustment; OR, 0.46; 95% CI, 0.38–0.55 after adjustment) (Figure V in the online-

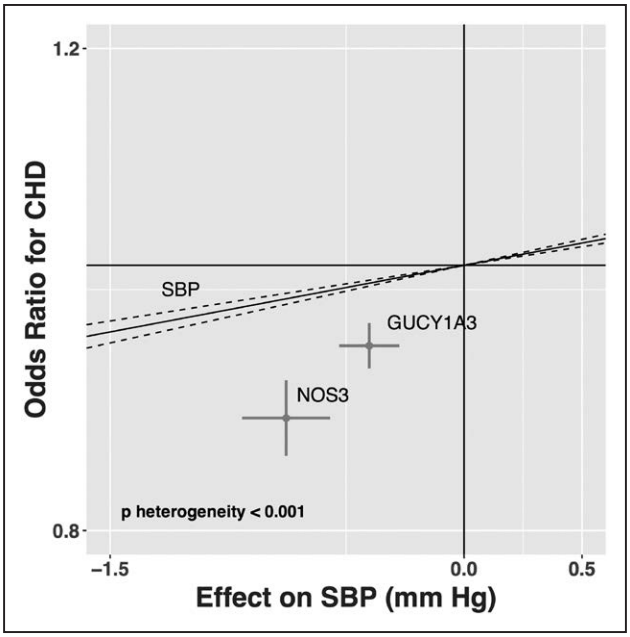



**Figure 5. Association of common variants in the *NOS3* (rs3918226) and *GUCY1A3* (rs7692387) loci with systolic blood pressure and coronary heart disease.**
Solid line represents the estimated effect of systolic blood pressure on coronary heart disease from 54 distinct blood pressure loci from GWAS. A test for heterogeneity comparing the association for the genetic nitric oxide score to other blood pressure loci was significant ($P$<0.001). Estimates for systolic blood pressure were derived from the UK Biobank with adjustment for age, sex, and 10 principal components. Estimates for coronary heart disease were derived from inverse variance fixed effects meta-analyses of CARDIOGRAM (Coronary Artery Disease Genome Wide Replication and Meta-Analysis)and the UK Biobank. CHD indicates coronary heart disease; and SBP, systolic blood pressure.

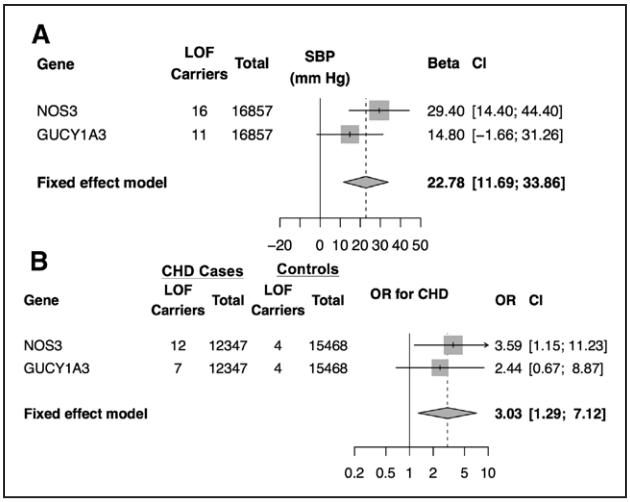



**Figure 6. Association of rare, predicted loss-of-function variants in the *NOS3*–*GUCY1A3* nitric oxide-signaling pathway with systolic blood pressure and coronary heart disease.**
**A**, Estimates for systolic blood pressure and diastolic blood pressure from T2D GENES study (Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples) were derived using linear regression with adjustment for 5 principal components of ancestry. **B**, Estimates for CHD from the Myocardial Infarction Genetics Consortium were derived using logistic regression, with adjustment for sex, cohort, and 5 principal components of ancestry. CHD indicates coronary heart disease; CI indicates confidence interval; LOF, loss of function; OR, odds ratio; and SBP, systolic blood pressure.

only Data Supplement), suggesting that much of the decrease in CHD risk through increased nitric oxide signaling seems to be through pathways other than blood pressure. Indeed, the effect size of rs3918226 (*NOS3*) and rs7692387 (*GUCY1A3*) on CHD deviated substantially from an estimate based on just the systolic blood pressure effects of these variants (Figure 5).

In contrast to common variants that promote increased nitric oxide signaling, we sought to test the hypothesis that rare inactivating variants in *NOS3* or *GUCY1A3* would be associated with increased blood pressure and risk of CHD. Twenty-seven participants with predicted loss-of-function variants in *NOS3* or *GUCY1A3* were identified in the T2D GENES study (Table V in the online-only Data Supplement). Presence of a predicted loss-of-function variant in *NOS3* or *GUCY1A3* was associated with increased systolic blood pressure (22.8 mm Hg; 95% CI, 11.7–33.9; $P$=5.6*10$^{-5}$, respectively) (Figure 6) and diastolic pressure (9.7 mm Hg; 95% CI 3.5–15.9; $P$=0.002) (Figure VI in the online-only Data Supplement). Twenty-seven individuals with predicted loss-of-function variants were identified in Myocardial Infarction Genetics Consortium studies (Table VI in the online-only Data Supplement). Predicted loss-of-function variants in the *NOS3*–*GUCY1A3* pathway were associated with a 3-fold higher risk of CHD (OR, 3.03; 95% CI, 1.29–7.12; $P$=0.01) (Figure 6).

## DISCUSSION

A genetic predisposition to enhanced nitric oxide signaling was associated with reduced blood pressure, improved renal and pulmonary function, and significantly reduced risks of CHD (OR, 0.37; 95% CI, 0.31–0.45), peripheral arterial disease (OR, 0.42; 95% CI, 0.26–0.68), and stroke (OR, 0.53; 95% CI, 0.37–0.76). Mediation analysis suggested that this protective effect is mediated only in part by blood pressure-related pathways. In contrast, mutations predicted to truncate *NOS3* or *GUCY1A3* are associated with higher blood pressure and a ≈3-fold higher risk of CHD.

These results permit several conclusions. First, stimulation of nitric oxide signaling may prevent atherosclerotic cardiovascular disease. The use of an oral soluble guanylyl cyclase stimulator has proven effective in the treatment of pulmonary hypertension,[13] reinforcing the

potential to target this pathway using a small-molecule approach. The 63% and 58% reductions in CHD and peripheral arterial disease, respectively, observed in this study are likely to be of greater magnitude than what would be observed in a randomized trial of a nitric oxide signaling stimulator because genetic estimates represent the effect of increased nitric oxide signaling over a lifetime rather than an intervention later in life and for a more limited duration.[23] However, the significant and large risk reductions in cardiovascular disease observed in this study lend support to efforts to target nitric oxide signaling in the prevention or treatment of atherosclerotic cardiovascular disease.[1,24,25]

Second, stimulation of nitric oxide signaling may represent a pathway for CHD prevention that is independent of current approaches. A genetic predisposition to nitric oxide signaling was not associated with lipid, glycemic, or anthropometric traits. Furthermore, the majority of the reduction in risk of CHD with increased nitric oxide signaling appeared to be independent of the variant's impact on blood pressure. Consistent with these findings, a recent functional characterization of the lead *GUCY1A3* variant rs7692387 observed carriers of the risk allele of this variant to have elevated levels of platelet aggregation and increased vascular smooth muscle cell migration on stimulation of soluble guanylyl cyclase.[19]

Third, these results suggest that stimulation of nitric oxide signaling may improve renal and pulmonary function. A genetic predisposition to enhanced nitric oxide signaling was associated with a higher glomerular filtration rate as determined by either cystatin C or creatinine, a finding that awaits further confirmation in larger studies. Increased nitric oxide signaling was also associated with improved pulmonary function. Although medications to stimulate nitric oxide signaling are effective and approved for treatment of pulmonary hypertension,[1,13] these results suggest that the stimulation of nitric oxide signaling may prove useful for other populations with reduced pulmonary function.

Our study has several limitations. First, we used common variants in the *GUCY1A3* and *NOS3* loci to estimate the phenotypic effects of increased nitric oxide signaling. It is possible that these variants may be in linkage disequilibrium with other variants that have phenotypic effects independent of nitric oxide-signaling pathways. However, the common variants were associated with direct measurement of *GUCY1A3* and *NOS3* expression in vascular and pulmonary tissues, and we observed consistent effects of rare predicted loss-of-function variants in *GUCY1A3* and *NOS3* on blood pressure and CHD, suggesting that the common variants mediated their effects through the *NOS3–GUCY1A3* nitric oxide-signaling pathway. Second, the phenome-wide association study to determine the full spectrum of associations in the UK Biobank may have been underpowered to detect associations for some diseases. Third, our rare variant analysis was restricted to variants predicted to lead to loss of *NOS3* or *GUCY1A3* function; the prevalence of missense mutations that impair the *NOS3–GUCY1A3* pathway may be much larger than the prevalence of rare predicted loss-of-function variants observed in this study. Finally, the majority of participants in this study were of European ancestry, and, as such, these observations need validation in ancestries outside of Europe because ethnic differences in nitric oxide-mediated responses have been previously reported.[26]

In conclusion, a genetic predisposition to enhanced nitric oxide signaling was associated with reduced risks of CHD and peripheral arterial disease and improved renal and pulmonary function. Stimulation of nitric oxide signaling may prove useful for the prevention and treatment of a range of diseases.

## AUTHORS

Connor A. Emdin, DPhil; Amit V. Khera, MD; Derek Klarin, MD; Pradeep Natarajan, MD; Seyedeh M. Zekavat, BSc; Akihiro Nomura, MD; Mary Haas, PhD; Krishna Aragam, MD; Diego Ardissino, MD; James G. Wilson, MD; Heribert Schunkert, MD; Ruth McPherson, MD; Hugh Watkins, MD, PhD; Roberto Elosua, MD, PhD; Matthew J. Bown, MD; Nilesh J. Samani, FRCP; Usman Baber, MD; Jeanette Erdmann, PhD; Padhraig Gormley, PhD; Aarno Palotie, MD, PhD; Nathan O. Stitziel, MD, PhD; Namrata Gupta, PhD; John Danesh, DPhil; Danish Saleheen, MBBS, PhD; Stacey Gabriel, PhD; Sekar Kathiresan, MD

## SOURCES OF FUNDING

This work was funded by the National Institutes of Health (R01 HL127564 to S.K.), which had no involvement in the design and conduct of the study; the collection, analysis, and interpretation of the data; or the preparation, review, and approval of the manuscript. This project was conducted using the UK Biobank resource (project ID 7089). This project was also conducted using the Type 2 Diabetes Knowledge Portal resource, which is funded by the Accelerating Medicines Partnership. The REGICOR study (Registre Gironi del COR, Gerona Heart Registry) was supported by the Spanish Ministry of Economy and Innovation through the Carlos III Health Institute (Red Investigación Cardiovascular RD12/0042, PI09/90506), European Funds for Development (ERDF-FEDER), and the Catalan Research and Technology Innovation Interdepartmental Commission (2014SGR240). Samples for the Leicester cohort were collected as part of projects funded by the British Heart Foundation (British Heart Foundation Family Heart Study, RG2000010; UK Aneurysm Growth Study, CS/14/2/30841) and the National Institute for Health Research (NIHR Leicester Cardiovascular Biomedical Research Unit Biomedical Research Informatics Center for Cardiovascular Science, IS_BRU_0211_20033). Dr Samani is supported by the British Heart Foundation and is a NIHR senior investigator. The Munich MI Study is supported by the German Federal Ministry

of Education and Research (BMBF) in the context of the e:Med program (e:AtheroSysMed) and the FP7 European Union project CVgenes@target (261123). Additional grants were received from the Fondation Leducq (CADgenomics: Understanding Coronary Artery Disease Genes, 12CVD02).This study was also supported through the Deutsche Forschungsgemeinschaft cluster of excellence "Inflammation at Interfaces" and SFB 1123. The Italian Atherosclerosis, Thrombosis, and Vascular Biology study (ATVB) was supported by a grant from RFPS-2007-3-644382 and Programma di ricerca Regione-Università 2010 to 2012 Area 1–Strategic Programmes–Regione Emilia-Romagna. Funding for the exome-sequencing project was provided by RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO), and RC2 HL102924 (WHISP). Exome sequencing was performed through RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). The Jackson Heart Study is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, and HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities. Dr Wilson is supported by U54GM115428 from the National Institute of General Medical Sciences. Exome sequencing in ATVB, PROCARDIS, Ottawa, PROMIS, Southern German Myocardial Infarction Study, and the Jackson Heart Study was supported by 5U54HG003067 (to Dr Gabriel).

## DISCLOSURES

Dr Khera is supported by a John S. LaDue Memorial Fellowship at Harvard Medical School and a KL2/Catalyst Medical Research Investigator Training award from Harvard Catalyst funded by the National Institutes of Health (TR001100) and has received consulting fees from Merck and Amarin. Dr Natarajan reports funding from the John S. LaDue Memorial Fellowship at Harvard Medical School and has received consulting fees from Amarin. Dr Stitziel reports funding from K08HL114642 and R01HL131961 and has received a research grant from AstraZeneca and consulting fees from Regeneron. Dr Kathiresan is supported by a research scholar award from Massachusetts General Hospital, the Donovan Family Foundation, and R01 HL127564; has received grants from Bayer Healthcare, Aegerion Pharmaceuticals, and Regeneron Pharmaceuticals; and has received consulting fees from Merck, Novartis, Sanofi, AstraZeneca, Alnylam Pharmaceuticals, Leerink Partners, Noble Insights, Quest Diagnostics, Genomics PLC, and Eli Lilly and Company; and holds equity in San Therapeutics and Catabasis Pharmaceuticals. The other authors report no conflicts of interest.

## AFFILIATIONS

Center for Genomic Medicine (C.A.E., A.V.K., D.K., P.N., S.M.Z., A.N., M.H., K.A., A.P., N.G., S.G., S.K.), and Department of Medicine, Cardiology Division (C.A.E., A.V.K., P.N., S.M.Z., A.N., P.G., A.P., M.H., K.A., S.K.), Massachusetts General Hospital, Harvard Medical School, Boston, MA. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA (C.A.E., A.V.K., D.K., P.N., S.M.Z., A.N., M.H., P.G., A.P., K.A., S.K.). Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA (D.K.). Division of Cardiology, Azienda Ospedaliero–Universitaria di Parma, Parma, Italy (D.A.). Associazione per lo Studio Della Trombosi in Cardiologia, Pavia, Italy (D.A.). Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson (J.G.W.). Deutsches Herzzentrum München, Technische Universität München, Deutsches Zentrum für Herz-Kreislauf-Forschung, München, Germany (H.S.). University of Ottawa Heart Institute, Ontario, Canada (R.M.). Wellcome Trust Centre for Human Genetics, University of Oxford, United Kingdom (H.W.). Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, United Kingdom (H.W.). Cardiovascular Epidemiology and Genetics, Hospital del Mar Research Institute, Barcelona, Spain (R.E). CIBER Enfermedades Cardiovasculares, Barcelona, Spain (R.E.). Facultat de Medicina, Universitat de Vic-Central de Cataluña, Spain (R.E.). Department of Cardiovascular Sciences, University of Leicester, National Institute for Health Research, Leicester Biomedical Research Centre, United Kingdom (M.J.B., N.J.S.). The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York (U.B.). Institute for Integrative and Experimental Genomics, University of Lübeck, Germany (J.E.). Departments of Medicine and Genetics, Cardiovascular Division, McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO (N.O.S.). Department of Public Health and Primary Care, Cardiovascular Epidemiology Unit, University of Cambridge, United Kingdom (J.D.). Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom (J.D). National Institute of Health Research Blood and Transplant, Research Unit in Donor Health and Genomics, University of Cambridge, United Kingdom (J.D.). Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA (D.S.). Center for Non-Communicable Diseases, Karachi, Pakistan (D.S.).

## FOOTNOTES

Received April 10, 2017; accepted September 25, 2017.

The online-only Data Supplement, podcast, and transcript are available with this article at http://circ.ahajournals.org/lookup/suppl/doi:10.1161/CIRCULATIONAHA.117.028021/-/DC1.

*Circulation* is available at http://circ.ahajournals.org.

## REFERENCES

1. Lundberg JO, Gladwin MT, Weitzberg E. Strategies to increase nitric oxide signalling in cardiovascular disease. *Nat Rev Drug Discov*. 2015;14:623–641. doi: 10.1038/nrd4623.
2. Steinhorn BS, Loscalzo J, Michel T. Nitroglycerin and nitric oxide: a rondo of themes in cardiovascular therapeutics. *N Engl J Med*. 2015;373:277–280. doi: 10.1056/NEJMsr1503311.
3. Förstermann U, Closs EI, Pollock JS, Nakane M, Schwarz P, Gath I, Kleinert H. Nitric oxide synthase isozymes: characterization, purification, molecular cloning, and functions. *Hypertension*. 1994;23(6 Pt 2):1121–1131.
4. Mergia E, Friebe A, Dangel O, Russwurm M, Koesling D. Spare guanylyl cyclase NO receptors ensure high NO sensitivity in the vascular system. *J Clin Invest*. 2006;116:1731–1737. doi: 10.1172/JCI27657.
5. Denninger JW, Marletta MA. Guanylate cyclase and the.NO/cGMP signaling pathway. *Biochim Biophys Acta*. 1999;1411:334–350.
6. Nong Z, Hoylaerts M, Van Pelt N, Collen D, Janssens S. Nitric oxide inhalation inhibits platelet aggregation and platelet-mediated pulmonary thrombosis in rats. *Circ Res*. 1997;81:865–869.

7. Knowles JW, Reddick RL, Jennette JC, Shesely EG, Smithies O, Maeda N. Enhanced atherosclerosis and kidney dysfunction in eNOS(-/-)Apoe(-/-) mice are ameliorated by enalapril treatment. *J Clin Invest*. 2000;105:451–458. doi: 10.1172/JCI8376.
8. Kuhlencordt PJ, Gyurko R, Han F, Scherrer-Crosbie M, Aretz TH, Hajjar R, Picard MH, Huang PL. Accelerated atherosclerosis, aortic aneurysm formation, and ischemic heart disease in apolipoprotein E/endothelial nitric oxide synthase double-knockout mice. *Circulation*. 2001;104:448–454.
9. Erdmann J, Stark K, Esslinger UB, Rumpf PM, Koesling D, de Wit C, Kaiser FJ, Braunholz D, Medack A, Fischer M, Zimmermann ME, Tennstedt S, Graf E, Eck S, Aherrahrou Z, Nahrstaedt J, Willenborg C, Bruse P, Brænne I, Nöthen MM, Hofmann P, Braund PS, Mergia E, Reinhard W, Burgdorf C, Schreiber S, Balmforth AJ, Hall AS, Bertram L, Steinhagen-Thiessen E, Li SC, März W, Reilly M, Kathiresan S, McPherson R, Walter U, Ott J, Samani NJ, Strom TM, Meitinger T, Hengstenberg C, Schunkert H; CARDIoGRAM. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature*. 2013;504:432–436. doi: 10.1038/nature12722.
10. Salvi E, Kutalik Z, Glorioso N, Benaglio P, Frau F, Kuznetsova T, Arima H, Hoggart C, Tichet J, Nikitin YP, Conti C, Seidlerova J, Tikhonoff V, Stolarz-Skrzypek K, Johnson T, Devos N, Zagato L, Guarrera S, Zaninello R, Calabria A, Stancanelli B, Troffa C, Thijs L, Rizzi F, Simonova G, Lupoli S, Argiolas G, Braga D, D'Alessio MC, Ortu MF, Ricceri F, Mercurio M, Descombes P, Marconi M, Chalmers J, Harrap S, Filipovsky J, Bochud M, Iacoviello L, Ellis J, Stanton AV, Laan M, Padmanabhan S, Dominiczak AF, Samani NJ, Melander O, Jeunemaitre X, Manunta P, Shabo A, Vineis P, Cappuccio FP, Caulfield MJ, Matullo G, Rivolta C, Munroe PB, Barlassina C, Staessen JA, Beckmann JS, Cusi D. Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension*. 2012;59:248–255. doi: 10.1161/HYPERTENSIONAHA.111.181990.
11. Ehret GB, Ferreira T, Chasman DI, Jackson AU, Schmidt EM, Johnson T, Thorleifsson G, Luan J, Donnelly LA, Kanoni S, Petersen AK, Pihur V, Strawbridge RJ, Shungin D, Hughes MF, Meirelles O, Kaakinen M, Bouatia-Naji N, Kristiansson K, Shah S, Kleber ME, Guo X, Lyytikäinen LP, Fava C, Eriksson N, Nolte IM, Magnusson PK, Salfati EL, Rallidis LS, Theusch E, Smith AJP, Folkersen L, Witkowska K, Pers TH, Joehanes R, Kim SK, Lataniotis L, Jansen R, Johnson AD, Warren H, Kim YJ, Zhao W, Wu Y, Tayo BO, Bochud M, Absher D, Adair LS, Amin N, Arking DE, Axelsson T, Baldassarre D, Balkau B, Bandinelli S, Barnes MR, Barroso I, Bevan S, Bis JC, Bjornsdottir G, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Bornstein SR, Brown MJ, Burnier M, Cabrera CP, Chambers JC, Chang IS, Cheng CY, Chines PS, Chung RH, Collins FS, Connell JM, Döring A, Dallongeville J, Danesh J, de Faire U, Delgado G, Dominiczak AF, Doney ASF, Drenos F, Edkins S, Eicher JD, Elosua R, Enroth S, Erdmann J, Eriksson P, Esko T, Evangelou E, Evans A, Fall T, Farrall M, Felix JF, Ferrières J, Ferrucci L, Fornage M, Forrester T, Franceschini N, Duran OHF, Franco-Cereceda A, Fraser RM, Ganesh SK, Gao H, Gertow K, Gianfagna F, Gigante B, Giulianini F, Goel A, Goodall AH, Goodarzi MO, Gorski M, Gräßler J, Groves C, Gudnason V, Gyllensten U, Hallmans G, Hartikainen AL, Hassinen M, Havulinna AS, Hayward C, Hercberg S, Herzig KH, Hicks AA, Hingorani AD, Hirschhorn JN, Hofman A, Holmen J, Holmen OL, Hottenga JJ, Howard P, Hsiung CA, Hunt SC, Ikram MA, Illig T, Iribarren C, Jensen RA, Kähönen M, Kang H, Kathiresan S, Keating BJ, Khaw KT, Kim YK, Kim E, Kivimaki M, Klopp N, Kolovou G, Komulainen P, Kooner JS, Kosova G, Krauss RM, Kuh D, Kutalik Z, Kuusisto J, Kvaløy K, Lakka TA, Lee NR, Lee IT, Lee WJ, Levy D, Li X, Liang KW, Lin H, Lin L, Lindström J, Lobbens S, Männistö S, Müller G, Müller-Nurasyid M, Mach F, Markus HS, Marouli E, McCarthy MI, McKenzie CA, Meneton P, Menni C, Metspalu A, Mijatovic V, Moilanen L, Montasser ME, Morris AD, Morrison AC, Mulas A, Nagaraja R, Narisu N, Nikus K, O'Donnell CJ, O'Reilly PF, Ong KK, Paccaud F, Palmer CD, Parsa A, Pedersen NL, Penninx BW, Perola M, Peters A, Poulter N, Pramstaller PP, Psaty BM, Quertermous T, Rao DC, Rasheed A, Rayner NWNWR, Renström F, Rettig R, Rice KM, Roberts R, Rose LM, Rossouw J, Samani NJ, Sanna S, Saramies J, Schunkert H, Sebert S, Sheu WH, Shin YA, Sim X, Smit JH, Smith AV, Sosa MX, Spector TD, Stančáková A, Stanton A, Stirrups KE, Stringham HM, Sundstrom J, Swift AJ, Syvänen AC, Tai ES, Tanaka T, Tarasov KV, Teumer A, Thorsteinsdottir U, Tobin MD, Tremoli E, Uitterlinden AG, Uusitupa M, Vaez A, Vaidya D, van Duijn CM, van Iperen EPA, Vasan RS, Verwoert GC, Virtamo J, Vitart V, Voight BF, Vollenweider P, Wagner A, Wain LV, Wareham NJ, Watkins H, Weder AB, Westra HJ, Wilks R, Wilsgaard T, Wilson JF, Wong TY, Yang TP, Yao J, Yengo L, Zhang W, Zhao JH, Zhu X, Bovet P, Cooper RS, Mohlke KL, Saleheen D, Lee JY, Elliott P, Gierman HJ, Willer CJ, Franke L, Hovingh GK, Taylor KD, Dedoussis G, Sever P, Wong A, Lind L, Assimes TL, Njølstad I, Schwarz PE, Langenberg C, Snieder H, Caulfield MJ, Melander O, Laakso M, Saltevo J, Rauramaa R, Tuomilehto J, Ingelsson E, Lehtimäki T, Hveem K, Palmas W, März W, Kumari M, Salomaa V, Chen YI, Rotter JI, Froguel P, Jarvelin MR, Lakatta EG, Kuulasmaa K, Franks PW, Hamsten A, Wichmann HE, Palmer CNA, Stefansson K, Ridker PM, Loos RJF, Chakravarti A, Deloukas P, Morris AP, Newton-Cheh C, Munroe PB; CHARGE-EchoGen consortium; CHARGE-HF consortium; Wellcome Trust Case Control Consortium. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet*. 2016;48:1171–1184. doi: 10.1038/ng.3667.
12. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–1130.
13. Ghofrani HA, Galiè N, Grimminger F, Grünig E, Humbert M, Jing ZC, Keogh AM, Langleben D, Kilama MO, Fritsch A, Neuser D, Rubin LJ; PATENT-1 Study Group. Riociguat for the treatment of pulmonary arterial hypertension. *N Engl J Med*. 2013;369:330–340. doi: 10.1056/NEJMoa1209655.
14. 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. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779. doi: 10.1371/journal.pmed.1001779.
15. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JRB, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Tajes JF, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Müller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SCJ, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim BJ, Kim Y, Kim YJ, Kwon MS, Lee J, Lee S, Lin KH, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han BG, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MCY, Palmer ND, Balkau B, Stančáková A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JMM, Lee HM, Lu Y, Kwak SH, Zhao W, Danesh J, Lam VKL, Park KS, Saleheen D, So WY, Tam CHT, Afzal U, Aguilar D, Arya R, Aung T, Chan E, Navarro C, Cheng CY, Palli D, Correa A, Curran JE, Rybin D, Farook VS, Fowler SP, Freedman BI, Griswold M, Hale DE, Hicks PJ, Khor CC, Kumar S, Lehne B, Thuillier D, Lim WY, Liu J, van der Schouw YT, Loh M, Musani SK, Puppala S, Scott WR, Yengo L, Tan ST, Taylor HA Jr, Thameem F, Wilson G Sr, Wong TY, Njølstad PR, Levy JC, Mangino M, Bonnycastle LL, Schwarzmayr T, Fadista J, Surdulescu GL, Herder C, Groves CJ, Wieland T, Bork-Jensen J, Brandslund I, Christensen C, Koistinen HA, Doney ASF, Kinnunen L, Esko T, Farmer AJ, Hakaste L, Hodgkiss D, Kravic J, Lyssenko V, Hollensted M, Jørgensen ME, Jørgensen T, Ladenvall C, Justesen JM, Käräjämäki A, Kriebel J, Rathmann W, Lannfelt L, Lauritzen T, Narisu N, Linneberg A, Melander O, Milani L, Neville M, Orho-Melander M, Qi L, Qi Q, Roden M, Rolandsson O, Swift A, Rosengren AH, Stirrups K, Wood AR, Mihailov E, Blancher C, Carneiro MO, Maguire J, Poplin R, Shakir K, Fennell T, DePristo M, de Angelis MH, Deloukas P, Gjesing AP, Jun G, Nilsson P, Murphy J, Onofrio R, Thorand B, Hansen T, Meisinger C, Hu FB, Isomaa B, Karpe F, Liang L, Peters A, Huth C, O'Rahilly SP, Palmer CNA, Pedersen O, Rauramaa R, Tuomilehto J, Salomaa V, Watanabe RM, Syvänen AC, Bergman RN, Bharadwaj D, Bottinger EP, Cho YS, Chandak GR, Chan JCN, Chia KS, Daly MJ, Ebrahim SB, Langenberg C, Elliott P, Jablonski KA, Lehman DM, Jia W, Ma RCW, Pollin TI, Sandhu M, Tandon N, Froguel P, Barroso I, Teo YY, Zeggini E, Loos RJF, Small KS, Ried JS, DeFronzo RA, Grallert H, Glaser B, Metspalu A, Wareham NJ, Walker M, Banks E, Gieger C, Ingelsson E, Im HK, Illig T, Franks PW, Buck G, Trakalo J, Buck D, Prokopenko I, Mägi R, Lind L, Farjoun Y, Owen KR, Gloyn AL, Strauch K, Tuomi T, Kooner JS, Lee JY, Park T, Donnelly P, Morris AD, Hattersley AT, Bowden DW, Collins FS, Atzmon G, Chambers JC, Spector TD, Laakso M, Strom TM, Bell GI, Blangero J, Duggirala R, Tai ES, McVean G, Hanis CL, Wilson JG, Seielstad M, Frayling TM, Meigs JB, Cox NJ, Sladek R, Lander ES, Gabriel S, Burtt NP, Mohlke KL, Meitinger T, Groop L, Abecasis G, Florez JC, Scott LJ, Morris AP, Kang HM, Boehnke M, Altshuler D, McCarthy MI. The genetic architecture of type 2 diabetes. *Nature*. 2016;536:41–47. doi: 10.1038/nature18642.
16. Emdin CA, Khera AV, Natarajan P, Klarin D, Zekavat SM, Hsiao AJ, Kathiresan S. Genetic association of waist-to-hip ratio with cardiometabolic traits,

type 2 diabetes, and coronary heart disease. *JAMA*. 2017;317:626–634. doi: 10.1001/jama.2016.21042.
17. Johnson T, Gaunt TR, Newhouse SJ, Padmanabhan S, Tomaszewski M, Kumari M, Morris RW, Tzoulaki I, O'Brien ET, Poulter NR, Sever P, Shields DC, Thom S, Wannamethee SG, Whincup PH, Brown MJ, Connell JM, Dobson RJ, Howard PJ, Mein CA, Onipinla A, Shaw-Hawkins S, Zhang Y, Davey Smith G, Day IN, Lawlor DA, Goodall AH, Fowkes FG, Abecasis GR, Elliott P, Gateva V, Braund PS, Burton PR, Nelson CP, Tobin MD, van der Harst P, Glorioso N, Neuvrith H, Salvi E, Staessen JA, Stucchi A, Devos N, Jeunemaitre X, Plouin PF, Tichet J, Juhanson P, Org E, Putku M, Sõber S, Veldre G, Viigimaa M, Levinsson A, Rosengren A, Thelle DS, Hastie CE, Hedner T, Lee WK, Melander O, Wahlstrand B, Hardy R, Wong A, Cooper JA, Palmen J, Chen L, Stewart AF, Wells GA, Westra HJ, Wolfs MG, Clarke R, Franzosi MG, Goel A, Hamsten A, Lathrop M, Peden JF, Seedorf U, Watkins H, Ouwehand WH, Sambrook J, Stephens J, Casas JP, Drenos F, Holmes MV, Kivimaki M, Shah S, Shah T, Talmud PJ, Whittaker J, Wallace C, Delles C, Laan M, Kuh D, Humphries SE, Nyberg F, Cusi D, Roberts R, Newton-Cheh C, Franke L, Stanton AV, Dominiczak AF, Farrall M, Hingorani AD, Samani NJ, Caulfield MJ, Munroe PB; Cardiogenics Consortium; Global BPgen Consortium. Blood pressure loci identified with a gene-centric array. *Am J Hum Genet*. 2011;89:688–700. doi: 10.1016/j.ajhg.2011.10.013.
18. Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, Grarup N, Sim X, Barnes DR, Witkowska K, Staley JR, Tragante V, Tukiainen T, Yaghootkar H, Masca N, Freitag DF, Ferreira T, Giannakopoulou O, Tinker A, Harakalova M, Mihailov E, Liu C, Kraja AT, Fallgaard Nielsen S, Rasheed A, Samuel M, Zhao W, Bonnycastle LL, Jackson AU, Narisu N, Swift AJ, Southam L, Marten J, Huyghe JR, Stančáková A, Fava C, Ohlsson T, Matchan A, Stirrups KE, Bork-Jensen J, Gjesing AP, Kontto J, Perola M, Shaw-Hawkins S, Havulinna AS, Zhang H, Donnelly LA, Groves CJ, Rayner NW, Neville MJ, Robertson NR, Yiorkas AM, Herzig KH, Kajantie E, Zhang W, Willems SM, Lannfelt L, Malerba G, Soranzo N, Trabetti E, Verweij N, Evangelou E, Moayyeri A, Vergnaud AC, Nelson CP, Poveda A, Varga TV, Caslake M, de Craen AJ, Trompet S, Luan J, Scott RA, Harris SE, Liewald DC, Marioni R, Menni C, Farmaki AE, Hallmans G, Renström F, Huffman JE, Hassinen M, Burgess S, Vasan RS, Felix JF, Uria-Nickelsen M, Malarstig A, Reily DF, Hoek M, Vogt T, Lin H, Lieb W, Traylor M, Markus HF, Highland HM, Justice AE, Marouli E, Lindström J, Uusitupa M, Komulainen P, Lakka TA, Rauramaa R, Polasek O, Rudan I, Rolandsson O, Franks PW, Dedoussis G, Spector TD, Jousilahti P, Männistö S, Deary IJ, Starr JM, Langenberg C, Wareham NJ, Brown MJ, Dominiczak AF, Connell JM, Jukema JW, Sattar N, Ford I, Packard CJ, Esko T, Mägi R, Metspalu A, de Boer RA, van der Meer P, van der Harst P, Gambaro G, Ingelsson E, Lind L, de Bakker PI, Numans ME, Brandslund I, Christensen C, Petersen ER, Korpi-Hyövälti E, Oksa H, Chambers JC, Kooner JS, Blakemore AI, Franks S, Jarvelin MR, Husemoen LL, Linneberg A, Skaaby T, Thuesen B, Karpe F, Tuomilehto J, Doney AS, Morris AD, Palmer CN, Holmen OL, Hveem K, Willer CJ, Tuomi T, Groop L, Käräjämäki A, Palotie A, Ripatti S, Salomaa V, Alam DS, Shafi Majumder AA, Di Angelantonio E, Chowdhury R, McCarthy MI, Poulter N, Stanton AV, Sever P, Amouyel P, Arveiler D, Blankenberg S, Ferrières J, Kee F, Kuulasmaa K, Müller-Nurasyid M, Veronesi G, Virtamo J, Deloukas P, Elliott P, Zeggini E, Kathiresan S, Melander O, Kuusisto J, Laakso M, Padmanabhan S, Porteous D, Hayward C, Scotland G, Collins FS, Mohlke KL, Hansen T, Pedersen O, Boehnke M, Stringham HM, Frossard P, Newton-Cheh C, Tobin MD, Nordestgaard BG, Caulfield MJ, Mahajan A, Morris AP, Tomaszewski M, Samani NJ, Saleheen D, Asselbergs FW, Lindgren CM, Danesh J, Wain LV, Butterworth AS, Howson JM, Munroe PB; CHARGE-Heart Failure Consortium; EchoGen Consortium; METASTROKE Consortium; GIANT Consortium; EPIC-InterAct Consortium; Lifelines Cohort Study; Wellcome Trust Case Control Consortium; Understanding Society Scientific Group; EPIC-CVD Consortium; CHARGE+ Exome Chip Blood Pressure Consortium; T2D-GENES Consortium; GoT2DGenes Consortium; ExomeBP Consortium; CHD Exome+ Consortium. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet*. 2016;48:1151–1161. doi: 10.1038/ng.3654.
19. Kessler T, Wobst J, Wolf B, Eckhold J, Vilne B, Hollstein R, von Ameln S, Dang TA, Sager HB, Moritz Rumpf P, Aherrahrou R, Kastrati A, Björkegren JLM, Erdmann J, Lusis AJ, Civelek M, Kaiser FJ, Schunkert H. Functional characterization of the GUCY1A3 coronary artery disease risk locus. *Circulation*. 2017;136:476–489. doi: 10.1161/CIRCULATIONAHA.116.024152.
20. GTEx Consortium. Human genomics: the Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648–660.
21. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung HK, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui CC, Gomez-Skarmeta JL, Nóbrega MA. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507:371–375. doi: 10.1038/nature13138.
22. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326:219.
23. Smith GD, Ebrahim S. "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1–22.
24. Wobst J, Kessler T, Dang TA, Erdmann J, Schunkert H. Role of sGC-dependent NO signalling and myocardial infarction risk. *J Mol Med (Berl)*. 2015;93:383–394. doi: 10.1007/s00109-015-1265-3.
25. Wobst J, von Ameln S, Wolf B, Wierer M, Dang TA, Sager HB, Tennstedt S, Hengstenberg C, Koesling D, Friebe A, Braun SL, Erdmann J, Schunkert H, Kessler T. Stimulators of the soluble guanylyl cyclase: promising functional insights from rare coding atherosclerosis-related GUCY1A3 variants. *Basic Res Cardiol*. 2016;111:51. doi: 10.1007/s00395-016-0570-5.
26. Cardillo C, Kilcoyne CM, Cannon RO III, Panza JA. Attenuation of cyclic nucleotide-mediated smooth muscle relaxation in blacks as a cause of racial differences in vasodilator function. *Circulation*. 1999;99:90–95.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-03911-8 OPEN

# Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease

Connor A. Emdin [1,2,3], Amit V. Khera et al.#

Less than 3% of protein-coding genetic variants are predicted to result in loss of protein function through the introduction of a stop codon, frameshift, or the disruption of an essential splice site; however, such predicted loss-of-function (pLOF) variants provide insight into effector transcript and direction of biological effect. In >400,000 UK Biobank participants, we conduct association analyses of 3759 pLOF variants with six metabolic traits, six cardiometabolic diseases, and twelve additional diseases. We identified 18 new low-frequency or rare (allele frequency < 5%) pLOF variant-phenotype associations. pLOF variants in the gene *GPR151* protect against obesity and type 2 diabetes, in the gene *IL33* against asthma and allergic disease, and in the gene *IFIH1* against hypothyroidism. In the gene *PDE3B*, pLOF variants associate with elevated height, improved body fat distribution and protection from coronary artery disease. Our findings prioritize genes for which pharmacologic mimics of pLOF variants may lower risk for disease.

Correspondence and requests for materials should be addressed to S.K. (email: skathiresan1@mgh.harvard.edu).
#A full list of authors and their affliations appears at the end of the paper.

A focused investigation of predicted loss-of-function (pLOF) variants provides several advantages when compared with analysis of other types of variants. First, analysis of pLOF variants may allow for the direct identification of a gene rather than a locus containing many candidate genes[1]. Second, pLOF variants provide directionality of effect, unlike non-coding regulatory variants which may increase or decrease expression of a given gene. Third, identification of pLOF variants which protect against disease may aid with prioritization of therapeutic target genes (e.g., the recent development of inhibitors of PCSK9 or ANGPTL3 which mimic human pLOF mutations protective against cardiovascular disease)[2–6].

Here, we analyse pLOF variants in UK Biobank and other datasets to identify genes for which pharmacologic inhibition may protect against disease. We identify associations of pLOF variants with cardiometabolic and immune disease, prioritizing the genes *GPR151, IL33, GSDMB, IFIH1,* and *PDE3B* as potential therapeutic targets.

## Results

**Analysis of loss-of-function variants**. In 405,569 individuals in UK Biobank (335,660 individuals of European ancestry and 69,909 individuals of Non-European ancestry, Supplementary Table 1), we analyzed the association of 3759 pLOF variants with six metabolic traits [body mass index (BMI), waist-to-hip ratio adjusted for body mass index (WHRadjBMI), height, systolic blood pressure (SBP), diastolic blood pressure (DBP), forced expiratory volume to forced vital capacity ratio (FEV1/FVC)], six cardiometabolic diseases (coronary artery disease, type 2 diabetes, atrial fibrillation, stroke, heart failure, venous thromboembolism) and twelve diseases with more than 5000 cases (allergic rhinitis, asthma, anxiety, breast cancer, cataract, cholelithiasis, depression, hypothyroidism, gastric reflux, osteoporosis, osteoarthritis, and psoriasis). The Variant Effect Predictor and associated LOFTEE plugin[7,8] algorithms were used to annotate variants which were pLOF (1) nonsense mutations that resulted in early termination of a protein; (2) frameshift mutations due to insertions or deletions of DNA; or (3) splice-site mutations which result in an incorrectly spliced protein. For coronary artery disease, we pooled UK Biobank estimates with estimates from the CARDIoGRAM Exome consortium[9]. For height, we pooled UK Biobank estimates with estimates from the GIANT Exome Consortium[10]. Variants with a $P < 5.5 \times 10^{-7}$ [0.05/(24 outcomes × 3759 variants) in the combined analysis were considered significant. Quantile–quantile analysis was used to examine for the presence of population stratification. No evidence of inflation was observed (inflation factors <1.1; Supplementary Fig. 1).

We identified 18 new low-frequency or rare (<5%) pLOF variants associated with traits and diseases in UK Biobank (Table 1 and Supplementary Table 2). We also discovered 26 new common frequency (≥5%) pLOF variants associated with UK Biobank phenotypes (Supplementary Table 3). Variants identified within the same locus in prior genome-wide association studies (+500 Kb; 1 Mb total) for the same phenotype are provided (Supplementary Tables 4 and 5).

A locus-wide conditional analysis (±500 kb of the pLOF variant) was performed to determine the extent to which the identified pLOF variant signal was independent of other genetic variation at the locus. Independence of pLOF variants may provide increased confidence for a causal association. Independent variants at the loci of rare and low frequency pLOFs are reported in Supplementary Data 1 and independent variants at the loci of common frequency pLOFs are reported in Supplementary Data 2. Of the 16 low frequency pLOF variants outside of the MHC locus, 14 were identified as independent variants in the locus-wide conditional analysis. These include the *IL33* variant rs146597587, the *GPR151* variant rs114285050, the *IFIH1* variant rs35337543, and the *PDE3B* variant rs15009066 which are further analyzed in the text of this manuscript. Below, we highlight several of these associations.

***GPR151*, obesity, and type 2 diabetes**. In *GPR151* (encoding G-protein coupled receptor 151), the p.Arg95Ter variant (rs114285050, allele frequency 0.8% in European ancestry) was associated with reduced BMI (beta −0.07 standard deviations, $-0.36\ kg/m^2$, $P = 4.9 \times 10^{-8}$). We replicated this association in an independent cohort, the Myocardial Infarction Genetics Consortium (MIGen), where p.Arg95Ter carriers had reduced

**Table 1 Predicted loss-of-function variants with minor allele frequency <5% which are significantly associated with traits or diseases in UK Biobank**

| Outcome | Gene | pLOF variant | Location | EA | RA | AA change | Freq (%) | Beta | SE | *P*-value | Novel? | MHC locus? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Asthma | FLG | rs61816761 | 1:152285861 | A | G | p.Arg501Ter | 1.51 | 0.21 | 0.03 | $1.51 \times 10^{-15}$ | Yes | No |
| Asthma | HLA-DQB1 | rs28688207 | 6:32628660 | C | T | Splice Acceptor c.773−1A > G | 3.14 | −0.17 | 0.02 | $3.11 \times 10^{-15}$ | Yes | Yes |
| Asthma | IL33 | rs146597587 | 9:6255967 | C | G | Splice Acceptor c.487−1G > C | 0.44 | −0.54 | 0.06 | $7.79 \times 10^{-17}$ | No[15] | No |
| BMI | GPR151 | rs114285050 | 5:145895394 | A | G | p.Arg95Ter | 0.78 | −0.07 | 0.01 | $4.89 \times 10^{-8}$ | Yes | No |
| BMI | PKHD1L1 | rs533623778 | 8:110523131 | T | C | p.Arg769Ter | $1.0 \times 10^{-4}$ | 5.30 | 0.99 | $9.45 \times 10^{-8}$ | Yes | No |
| DBP | ENPEP | rs33966350 | 4:111431444 | A | G | p.Trp413Ter | 1.19 | 0.06 | 0.01 | $8.12 \times 10^{-10}$ | No[39] | No |
| DBP | BTN3A2 | rs58367598 | 6:26370833 | G | T | Splice Donor c.715 + 2T > G | 3.75 | 0.03 | 0.01 | $2.03 \times 10^{-8}$ | Yes | Yes |
| DBP | TMC3 | rs150843673 | 15:81624929 | T | G | p.Ser1045Ter | 2.14 | 0.05 | 0.01 | $8.16 \times 10^{-9}$ | Yes | No |
| Height | PDE11A | rs76308115 | 2:178879181 | A | G | p.Arg57Ter | 0.52 | 0.07 | 0.01 | $6.20 \times 10^{-11}$ | Yes | No |
| Height | CLHC1 | rs114931154 | 2:55407644 | T | A | Splice Donor c.1384 + 2T > A | 1.26 | −0.05 | 0.01 | $1.54 \times 10^{-11}$ | Yes | No |
| Height | CCDC66 | rs150364083 | 3:56628033 | T | C | p.Arg427Ter | 0.58 | 0.05 | 0.01 | $2.09 \times 10^{-7}$ | Yes | No |
| Height | DAP | rs201354802 | 5:10761153 | A | C | p.Glu10Ter | 0.24 | 0.13 | 0.02 | $1.68 \times 10^{-8}$ | Yes | No |
| Height | TRIM40 | rs115651142 | 6:30115320 | T | G | Splice Donor c.602 + 1G > T | 0.63 | −0.08 | 0.01 | $1.16 \times 10^{-9}$ | Yes | Yes |
| Height | MICA | rs181430930 | 6:31378575 | A | G | Splice Donor c.286 + 1G > A | 0.26 | −0.12 | 0.02 | $7.87 \times 10^{-8}$ | Yes | Yes |
| Height | PDE3B | rs150090666 | 11:14865399 | T | C | p.Arg783Ter | 0.06 | 0.24 | 0.04 | $9.32 \times 10^{-9}$ | Yes | No |
| Height | APOLD1 | rs202116412 | 12:12879031 | A | G | Splice Donor c.96 + 1G > A | 0.03 | 0.12 | 0.02 | $3.06 \times 10^{-8}$ | Yes | No |
| Hypothyroidism | IFIH1 | rs35337543 | 2:163136505 | G | C | Splice Donor c.1641 + 1G > C | 1.45 | −0.27 | 0.04 | $2.95 \times 10^{-9}$ | Yes | No |
| Psoriasis | ZKSCAN3 | rs73387810 | 6:28318166 | A | G | Splice Donor c.−63 + 1G > A | 0.86 | 0.55 | 0.08 | $4.18 \times 10^{-11}$ | Yes | Yes |
| Psoriasis | EGFL8 | rs141826798 | 6:32134395 | G | C | p.Arg74Ter | 0.53 | 0.90 | 0.08 | $2.19 \times 10^{-26}$ | Yes | Yes |
| SBP | ENPEP | rs33966350 | 4:111431444 | A | G | p.Trp413Ter | 1.19 | 0.06 | 0.01 | $3.46 \times 10^{-9}$ | No[39] | No |
| SBP | GEM | rs138582164 | 8:95264265 | A | G | p.Arg199Ter | 0.04 | 0.30 | 0.06 | $1.93 \times 10^{-7}$ | No[40] | No |
| WHRadjBMI | PYGM | rs116987552 | 11:64527223 | A | G | p.Arg50Ter | 0.39 | 0.09 | 0.02 | $1.32 \times 10^{-7}$ | Yes | No |

Beta in terms of standard deviations and reported for the effect allele
*pLOF* predicted loss-of-function, *EA* effect allele, *RA* reference allele, *AA Change* amino acid change, *Freq(%)* Frequency(%); *BMI* body mass index, *DBP* diastolic blood pressure, *SBP* systolic blood pressure, *WHRadjBMI* waist-to-hip ratio adjusted for body mass index

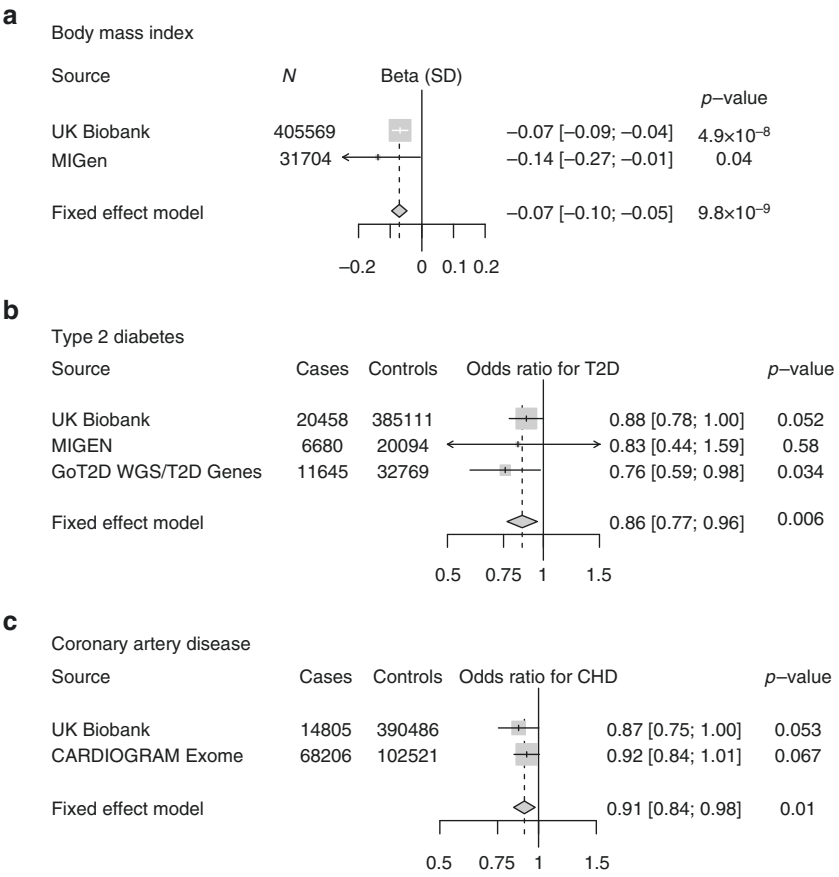



**Fig. 1** Association of a loss-of-function variant (p.Arg95Ter) in *GPR151* with **a** body mass index, **b** type 2 diabetes, and **c** coronary artery disease. Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components of ancestry, and array type

BMI (beta −0.14, $P = 0.04$; pooled beta −0.07, $P = 9.8 \times 10^{-9}$; Fig. 1). UK Biobank participants who carry one copy of p.Arg95Ter were at 12% lower odds of clinical obesity (BMI $\geq$ 30 kg/m$^2$). As obesity is a causal risk factor for type 2 diabetes and coronary artery disease, we examined whether p.Arg95Ter may provide protection against both diseases. p.Arg95Ter was associated with 14% lower odds of type 2 diabetes (OR 0.86; $P = 0.006$) and 9% lower odds of coronary artery disease (OR 0.91; $P = 0.01$; Fig. 1). Although *GPR151* encodes a G-protein coupled receptor of unknown function whose expression is limited to the central nervous system, recent studies tracing the lineage of neurons expressing *GRP151* have localized connections to hypothalamic neurons, a region of the brain important in the control of appetite[11].

***IL33, GSDMB*, and asthma**. We identified several pLOF variants that associated with lower risk of asthma (Table 1). At *GSDMB* encoding gasdermin B, splice acceptor variant c.662-2A > G (rs11078928, allele frequency 46% in European ancestry) protected against asthma (OR 0.90 CI 0.89, 0.91, $P = 6.7 \times 10^{-50}$; Supplementary Table 3). This variant is in tight linkage disequilibrium ($r^2 = 0.99$) with a previously identified non-coding variant in the *GSDMB* locus (rs2305480) associated with lower risk of asthma ($P = 9.6 \times 10^{-8}$)[12]. *GSDMB* c.662-2A > G is associated with lower expression of *GSDMB* transcripts[13]. Furthermore, overexpression of *GSDMB* causes airway remodeling and asthma symptoms in a mouse model[14], suggesting that loss of GSDMB function may protect against asthma.

At the *IL33* gene, a splice acceptor site variant c.487-1G > C (rs146597587, allele frequency 0.004 in European ancestry) was observed to protect against asthma (OR 0.58 CI 0.51, 0.66, $P = 7.8 \times 10^{-17}$). This variant was recently identified as associated with lower blood eosinophil concentration at genome-wide significance and with lower risk of asthma at more modest levels of significance ($P = 1.8 \times 10^{-4}$)[15]. To further replicate the association of *IL33* c.487-1G > C with asthma, we examined the association of *IL33* c.487-1G > C with asthma in individuals from three additional studies (Partners Biobank, the Vanderbilt eMERGE network, and the Women's Genome Health Study). *IL33* c.487-1G > C was associated with a protective effect of asthma in each data set. Overall, *IL33* c.487-1G > C was associated with 43% lower odds of asthma (OR 0.57 CI 0.51, 0.65, $P = 9.6 \times 10^{-19}$, Fig. 2), suggesting that IL33 inhibition may be a useful approach for treatment of asthma. Of note, an inhibitor of IL33 is currently under development for treatment of asthma[16].

Asthma is often associated with other allergic phenotypes—atopic dermatitis, food allergy, and allergic rhinitis[17]. We therefore examined whether asthma-associated pLOF variants in *GSDMB* and *IL33* associate with a lower risk of other atopic disorders in UK Biobank. Both pLOF variants also protected against allergic rhinitis (Fig. 3). In contrast, nominal associations with atopic dermatitis and food allergy were not observed, although point estimates for food allergy were similar to asthma.

***IFIH1* and autoimmune disorders**. A splice donor variant in *IFIH1* (interferon induced with helicase C domain 1), c.1641 + 1G > C (rs35337543, allele frequency 1.5% in European ancestry), is associated with a reduced risk of hypothyroidism in UK

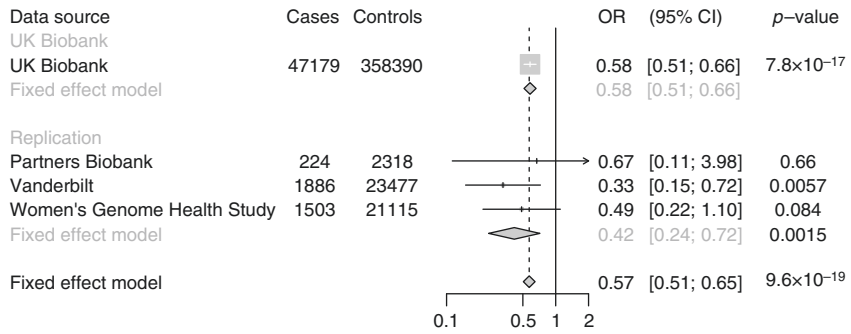



**Fig. 2** Association of *IL33* c.487-1G > C with asthma in UK Biobank, Partners Biobank, Vanderbilt eMERGE network and Women's Genome Health Study. UK Biobank estimates were derived using logistic regression, adjusted for age, sex, ten principal components of ancestry, and array type. Partners Biobank and Vanderbilt estimates were derived using logistic regression, adjusted for age, sex, and principal components of ancestry. Women's Genome Health Study estimates were derived using logistic regression, adjusted for age and principal components of ancestry

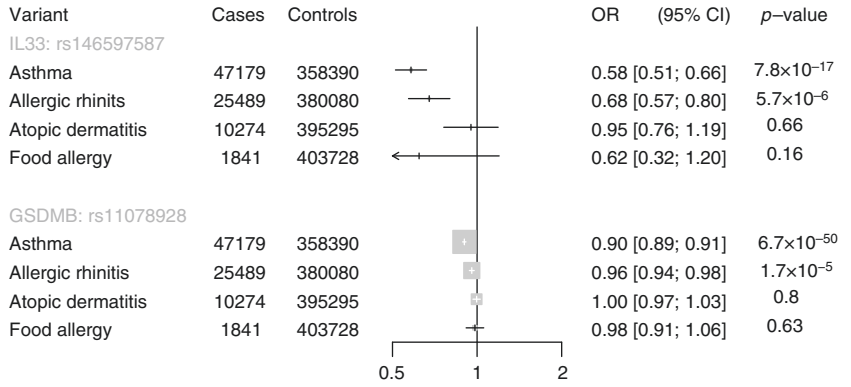



**Fig. 3** Association of predicted loss-of-function variant in *GSDMB* and *IL33* with allergic disease in UK Biobank. Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components of ancestry, and array type

Biobank participants (OR 0.77 CI 0.70, 0.85; $P = 5 \times 10^{-9}$; Table 1). A gene-based test combining four additional pLOF variants in *IFIH1* (rs35732034, rs201026962, rs35744605, rs148590996) similarly demonstrated protection against hypothyroidism in UK Biobank (OR 0.79 CI 0.72, 0.86; $P = 4.4 \times 10^{-8}$). Carriers of pLOF variants in *IFIH1* were also protected against hyperthyroidism (OR 0.84 CI 0.73, 0.96; $P = 0.01$; Fig. 4).

Common variants in the *IFIH1* locus have previously been associated with psoriasis[18] and vitiligo[19], while rare pLOF variants in *IFIH1* have been associated with a reduced risk of type 1 diabetes[20]. We therefore examined whether carriers of pLOF variants in *IFIH1* in UK Biobank were protected against these diseases. Carriers of pLOF variants in *IFIH1* were protected against type 1 diabetes, psoriasis, and vitiligo in UK Biobank (Fig. 4). These results suggest that *IF1H1* pLOF variants alter risk for a range of autoimmune diseases.

In addition, an exploratory analysis demonstrated a nominally lower risk of coronary artery disease among *IFIH1* pLOF carriers. Pooling UK Biobank and MIGen, *IFIH1* pLOF carriers were protected against coronary artery disease (OR 0.92 CI 0.87, 0.98; $P = 0.009$). To examine whether this may be a chance finding, we examined whether the common *IFIH1* missense variant rs1990760, previously identified as associated with autoimmune disorders, also associated with risk of coronary artery disease. The T allele of rs1990760 (frequency 41%) associated with a reduced risk of hypothyroidism (OR 0.92 CI 0.90, 0.94; $P = 9.3 \times 10^{-17}$) in UK Biobank. Pooling UK Biobank and CARDIOGRAM Exome, the T allele of rs1990760 also associated with a lower risk of coronary artery disease (OR 0.97 CI 0.96, 0.99; $P = 2.5 \times 10^{-5}$), providing complementary evidence that *IFIH1* may influence coronary artery disease risk.

***PDE3B* and body fat distribution**. At *PDE3B* encoding the gene phosphodiesterase 3B, p.Arg783Ter (rs150090666, allele frequency 0.0006 in European ancestry) associated with elevated height (beta 0.24, $P = 9.3 \times 10^{-9}$). Targeted deletion of *Pde3b* in mice leads to white adipose tissue gaining characteristics of brown adipose tissue[21], a reduction in adipocyte size[22], smaller fat deposits[23] and reduced atherosclerosis[24]. We therefore studied the association of *PDE3B* p.Arg783Ter with metabolic phenotypes in UK Biobank and/or MIGen, where 36,581 individuals have been sequenced for the 16 exons of the *PDE3B* gene. In UK Biobank, which lacks direct measurements of blood lipids, carriers of p.Arg783Ter carriers were at reduced risk of physician-diagnosis of hypercholesterolemia (OR 0.52, $P = 0.0002$). Pooling UK Biobank and MIGen, pLOF carriers in *PDE3B* had reduced WHRadjBMI (beta −0.15, $P = 0.0005$). As genetic predisposition to improved body fat distribution has been associated with a lower risk of coronary artery disease[25], we examined whether loss of *PDE3B* function protects against coronary artery disease. We aggregated rare *PDE3B* pLOFs in cases and compared this count with that controls. Across 14,805 cases in UK Biobank, the carrier frequency of pLOF in cases was 0.1% and in controls 0.2%. Across 20,186 cases in MIGen, the carrier frequency of pLOF was 0.05% and 0.1% in controls. Collectively, carrier status for *PDE3B*

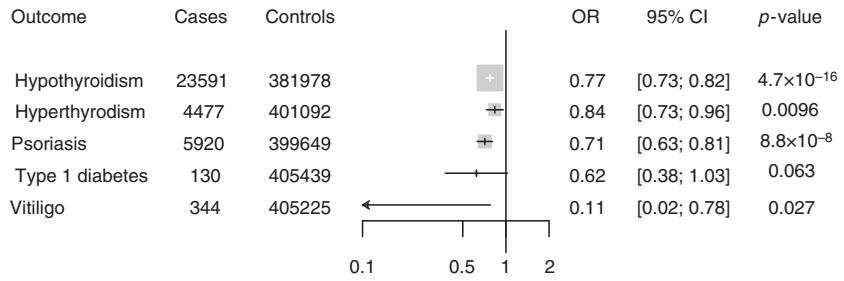



**Fig. 4** Association of predicted loss-of-function variants in *IFIH1* with thyroid disorders, type 1 diabetes, psoriasis, and vitiligo in UK Biobank. Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components of ancestry, and array type

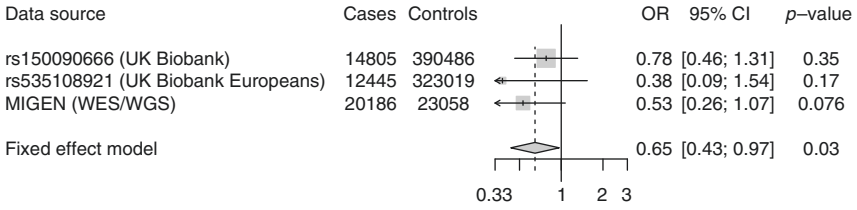



**Fig. 5** Association of predicted loss-of-function variants in *PDE3B* with coronary artery disease. Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components of ancestry, and array type. Estimates were derived in MIGEN (Myocardial Infarction Genetics Consortium) using logistic regression, adjusted for sex and five principal components of ancestry

pLOFs was associated with reduced risk for coronary artery disease (OR 0.65 CI 0.43, 0.97; $P = 0.03$; Fig. 5).

Presence of homozygote individuals for pLOF variants in target genes may provide an in vivo demonstration of safety of pharmacologic inhibition of target genes. We therefore examined whether homozygotes for these pLoF variants were present in UK Biobank and in the gnomAD database[26]. At least one individual homozygous for a pLOF variant was identified in UK Biobank or the gnomAD database for the genes *GPR151*, *GSDMB*, *IL33* and *IFIH1*, and *PDE3B* (Supplementary Table 6).

## Discussion

In this study, we identified pLOF variants that protect against obesity (*GPR151*), asthma (*GSDMB*, *IL33*), autoimmune disorders (*IFIH1*), and coronary artery disease (*PDE3B*), prioritizing genes and pathways for which pharmacologic attempts to mimic these protective mutations might ameliorate disease.

Identification of protective loss-of-function variants has led to the development of therapeutics. Most notably, the discovery of missense and loss-of-function variants in *PCSK9* that lower LDL cholesterol and protect against coronary artery disease suggested that inhibition of PCSK9 may be a useful therapeutic strategy for prevention and treatment of cardiovascular disease[3,27]. These genetic studies were validated by a large-scale randomized trial demonstrating that a monoclonal antibody directed against *PCSK9* reduced the risk of recurrent cardiovascular events[4]. More recently, the discovery of loss-of-function variants in *ANGPTL3* that lower blood triglyceride levels and protect against coronary artery disease has suggested that ANGPTL3 inhibition may reduce blood triglyceride levels and risk of coronary artery disease[28,29]. ANGTPL3 inhibitors are in clinical development and have now been demonstrated to reduce blood triglyceride levels[5,6].

Our findings identify putative therapeutic targets that, similar to PCSK9 and ANGPTL3, may be useful for prevention of disease. For obesity and type 2 diabetes, these results hightlight GPR151 as a potential therapeutic target. *GPR151* encodes a largely uncharacterized G protein-coupled receptor. The mechanism by which it influences risk of obesity and type 2 diabetes is unclear. However, it is highly expressed in the hypothalamus, a region of the brain known to be involved in appetite regulation[11]. Indeed, genetic variation in *MC4R*, which encodes the melanocortin 4 receptor, strongly influences obesity risk at a population level[30]. Similar to *GPR151*, *MC4R*is highly expressed in the hypothalamus and is involved in appetite regulation[31].

PDE3B inhibition may be a useful therapeutic strategy to improve body fat distribution and reduce risk of coronary artery disease. Unlike *GPR151*, for which no pharmacologic inhibitor is currently in clinical use, an inhibitor of *PDE3B* is available. Cilostazol is a non-selective pharmacologic inhibitor of both phosphodiesterase 3B and the related isoform phosphodiesterase 3A[32]. In a small randomized trial including 211 participant, cilostazol significantly reduced restenosis after percutaneous coronary balloon angioplasty[33]. The association of *PDE3B* pLOFs with improved body fat distribution, reduced risk of hypercholesterolemia and reduced risk of coronary artery disease suggests that selective inhibition of PDE3B may be useful for multiple features of metabolic syndrome.

We identified pLOF variants in *IL33* (encoding interleukin 33) and *GSDMB* (encoding gasdermin B) as associated with a lower risk of asthma. The *IL33* variant rs146597587 was recently found to be associated with lower blood eosinophil concentration at genome-wide significance and with lower risk of asthma at more modest levels of significance ($P = 1.8 \times 10^{-4}$)[15]. Consistent with these findings, induction of antibodies against IL33 by vaccination induces protection against airway inflammation in a mouse model of asthma[34]. An inhibitor of IL33 is currently under development for treatment of asthma[16]. In contrast to IL33, no inhibitor for GSDMB is in clinical development. These findings suggest that IL33 and GSDMB inhibition may both be useful therapeutic strategies for treatment of asthma and allergic disease.

Although our restriction of the present analysis to pLOF variants increases the likelihood of identifying causal variants substantially, it remains possible that a highly correlated nearby

variant could be driving the association in some cases. Future functional studies may permit additional validation of causal variants.

In summary, we associated pLOF variants with a range of biomarker and disease phenotypes in a large, national biobank and identified several new genes in which pLOF variants protect against disease, prioritizing these genes for therapeutic targeting. More generally, large-scale analysis of pLOF variants is emerging as a useful tool for therapeutic target identification and validation.

## Methods

**Gene and variant annotation**. Variants in hg19 coordinates were annotated with information from Ensembl release 82 using Variant Effect Predictor (VEP)[7]. Only pLOFs, defined as premature stop (nonsense), canonical splice-sites (splice-donor or splice-acceptor) or insertion/deletion variants that shifted frame (frameshift) were annotated as predicted loss-of-function (pLOF), using the "--pick-allele" annotation. PLOFs as defined by VEP were then merged with publicly available data from the Exome Aggregation Consortium (ExAC), Version 0.3.1, to confirm consistency in variant annotation[26].

We identified 3759 pLOF variants in UK Biobank with an info score greater than 0.3 (Supplementary Table 1). We used a Bonferroni corrected $P$ value of $5.5 \times 10^{-7}$ to denote significance [$0.05/(3759 \text{ variants} \times 24 \text{ outcomes}) = 5.5 \times 10^{-7}$] in our primary pLOF analysis.

**Study design**. We analyzed the association of pLOF variants with 24 phenotypes: cardiovascular, metabolic and pulmonary phenotypes: six metabolic traits (body mass index, waist-to-hip ratio, height, systolic blood pressure, diastolic blood pressure, and forced expiratory volume to forced vital capacity ratio), six cardiometabolic diseases (coronary artery disease, type 2 diabetes, atrial fibrillation, stroke, heart failure, and venous thromboembolism) and 12 diseases with more than 5000 cases (allergic rhinitis, asthma, anxiety, breast cancer, cataract, cholethiasis, depression, hypothyroidism, gastric reflux, osteoporosis, osteoarthritis, and psoriasis; Supplementary Table 4). All six metabolic traits were inverse normalized prior to analysis, with adjustment for age and sex. Forced expiratory volume to forced vital capacity ratio was additionally adjusted for height. To adjust for the presence of antihypertensive medication, we added 15 mm Hg to systolic blood pressure and 10 mm Hg to diastolic blood pressure of individuals on antihypertensive medication at baseline, as in the International Consortium for Blood Pressure genome-wide association study[35]. Definitions for disease outcomes in UK Biobank are provided (Supplementary Table 7).

In UK Biobank, analysis was performed separately in unrelated individuals of European and Non-European ancestry. Estimates for variants were then pooled using inverse-variance weighted-fixed effects meta-analysis. For coronary artery disease, estimates for variants from UK Biobank were additionally pooled with the effect of variants in the CARDIOGRAM Exome consortium using inverse variance weighted-fixed effects meta-analysis[9]. For height, estimates of variants in UK Biobank were pooled with the GIANT Height Exome consortium using inverse variance weighted-fixed effects meta-analysis[10].

**Genotyping and quality control**. Phasing and imputation were performed centrally, by UK Biobank, using a reference panel combining UK10k and 1000 Genomes samples. 39,235,157 variants included in the Haplotype Reference Consortium were imputed. As recommended by UK Biobank, we excluded any samples with an imputation quality <0.3 as well as pLOF variants which were not included in the Haplotype Reference Consortium. Mitochondrial genetic data and sex chromosomes were excluded from this analysis. Individual level genetic data was available from individuals in UK Biobank, after excluding one related individual of each related pair of individuals, individuals whose genetic sex did not match self-reported sex and individuals with an excess of missing genotype calls or more heterozygosity than expected.

We analysed 3759 variants identified as pLOF variants in UK Biobank. PLINK 2 software was used to examine the association of these variants with traits and disease in UK Biobank under the assumption of additive effects (https://www.cog-genomics.org/plink/2.0/). Adjustment was performed for age, sex, ten principal components of ancestry, and array type.

**Conditional analysis**. A locus-wide conditional analysis (±500 kb of the pLOF variant) was performed to determine the extent to which the identified pLOF variant signal was independent of other genetic variation at the locus. We iteratively performed association analyses conditional on the top variants at each locus, until no variants were below the Bonferroni-adjusted threshold for significance ($P < 5.5 \times 10^{-7}$). A statistically significant and independent signal for the pLOF variant provides increased confidence for a causal association.

**Analysis of PDE3B association with coronary artery disease**. We aimed to analyse the association of pLOF variants in *PDE3B* with coronary artery disease in a combined analysis of UK Biobank and the Myocardial Infarction Genetics Consortium (MIGen). Replication was performed in MIGen rather than the CARDIOGRAM Exome consortium as rs150090666 was not included in the exome chip analysis of the CARDIOGRAM Exome consortium[9]. Estimates of the association of rs150090666 with coronary artery disease in UK Biobank were derived as described above, using logistic regression with adjustment for age, sex, ten principal components of ancestry, and a dummy variable for array type. An additional pLOF variant, rs535108921, present in UK Biobank, was also analysed for association with coronary artery disease, as above.

pLOFs variants in *PDE3B* were identified in the MIGen Consortium using exome sequencing or whole genome sequencing, as previously described[36–38]. Studies included in the MIGen consortium were: (1) the Italian Atherosclerosis Thrombosis and Vascular Biology (ATVB) study (dbGaP Study Accession phs000814.v1.p1); (2) the Exome Sequencing Project Early-Onset Myocardial Infarction (ESP-EOMI) study(9); (3) a nested case-control cohort from the Jackson Heart Study (JHS); (4) the South German Myocardial Infarction study (dbGaP Study Accession phs000916.v1.p1); (5) the Ottawa Heart Study (OHS) (dbGaP Study Accession phs000806.v1.p1); (6) the Precocious Coronary Artery Disease (PROCARDIS) study (dbGaP Study Accession phs000883.v1.p1); (7) the Pakistan Risk of Myocardial Infarction Study (PROMIS) (dbGaP Study Accession phs000917.v1.p1); (8) the Registre Gironi del COR (Gerona Heart Registry or REGICOR) study (dbGaP Study Accession phs000902.v1.p1); (9) the Leicester Myocardial Infarction study (dbGaP Study Accession phs001000.v1.p1); (10) the BioImage study (dbGaP Study Accession phs001058.v1.p1); (11) the North German Myocardial Infarction study (dbGaP Study Accession phs000990.v1.p1); (12) Multi-Ethnic Study of Atherosclerosis (dbGaP Study Accession: phs000209.v2.p1); (13) Variation In Recovery: Role of Gender on Outcomes of Young AMI cohort; and (14) Taiwan Metabochip Consortium.

The Burrows–Wheeler Aligner algorithm was used to align reads from participants to the reference genome (hg19). The GATK HaploTypeCaller was used to jointly call variants. Metrics including Variant Quality Score Recalibration (VQSR), quality over depth, and strand bias were then used to filter variants. We excluded samples which were related to other samples, which had high ratios of heterozygous to non-reference homozygous genotypes, which had high missing genotypes, which had a discordant genetic gender relative to reports gender, and samples which were discordant relative to genotype data.

After variant calling and quality control, the Variant Effect Predictor[7,8] was used to annotate variants which were pLOF: (1) nonsense mutations that resulted in early termination of *PDE3B* (2) frameshift mutations due to insertions or deletions of DNA; or (3) splice-site mutations which result in an incorrectly spliced protein (Supplementary Table 8). For analysis of rare pLOF variants, we pooled rare pLOF variants in MIGen, testing for the association of a pLOF with coronary artery disease using logistic regression, after adjustment for age, sex, cohort, and five principle components of ancestry. We meta-analysed the association of pLOFs with coronary artery disease in MIGen combined with UK Biobank.

**Replication of IL33 finding**. To replicate the association of rs146597587, a splice site variant in *IL33*, with asthma, we pooled estimates of the association of rs146597587 with asthma from Partners Biobank, from the Vanderbilt eMERGE network and from the Women's Genome Health Study. In Partners Biobank, rs146597587 was imputed (info score of 0.77) in 2542 individuals. The association of rs146597587 with asthma (hospitalization for ICD9 code 493) was estimated using logistic regression, adjusted for age, sex, and five principal components of ancestry. In the Vanderbilt eMERGE network, rs146597587 was genotyped in 25,363 individuals using the Illumina Exome BeadChip. The association of rs146597587 with asthma (hospitalization for ICD9 code 493) was estimated using logistic regression, adjusted for age, sex, and principal components of ancestry. In Women's Genome Health Study, rs14659758 was genotyped in 22,618 individuals using the Illumina Exome. The association of rs14659758 with asthma (hospitalization for ICD9 code 493 or ICD10 code J45) was estimated using logistic regression, adjusted for age and principal components of ancestry.

**Data availability**. All individual-level data from UK Biobank can be accessed by applying to the UK Biobank Central Access Committee (http://www.ukbiobank.ac.uk/register-apply/).

Received: 3 October 2017 Accepted: 21 March 2018
Published online: 24 April 2018

## References

1. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**, 779–797 (2013).

2. Abifadel, M. et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).
3. Cohen, J. C., Boerwinkle, E., Mosley, T. H. & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
4. Sabatine, M. S. et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N. Engl. J. Med.* **376**, 1713-1722 (2017).
5. Dewey, F. E. et al. Genetic and pharmacologic inactivation of ANGPTL3 and cardiovascular disease. *N. Engl. J. Med.* **377**, 211–221 (2017).
6. Graham, M. J. et al. Cardiovascular and metabolic effects of ANGPTL3 antisense oligonucleotides. *N. Engl. J. Med.* **377**, 222–232 (2017).
7. McLaren, W. et al. Deriving the consequences of genomic variants with the ensemble API and SNP effect predictor. *Bioinformatics* **26**, 2069–2070 (2010).
8. Karczewski, K. J. LOFTEE (Loss-Of-Function Transcript Effect Estimator). https://github.com/konradjk/loftee (2015).
9. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N. Engl. J. Med.* **374**, 1134–1144 (2016).
10. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
11. Broms, J. et al. Monosynaptic retrograde tracing of neurons expressing the G-protein coupled receptor Gpr151 in the mouse brain. *J. Comp. Neurol.* **525**, 3227–3250 (2017).
12. Moffatt, M. F. et al. A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363**, 1211–1221 (2010).
13. Morrison, F. S. et al. The splice site variant rs11078928 may be associated with a genotype-dependent alteration in expression of GSDMB transcripts. *BMC Genom.* **14**, 627 (2013).
14. Das, S. et al. GSDMB induces an asthma phenotype characterized by increased airway responsiveness and remodeling without lung inflammation. *Proc. Natl Acad. Sci. USA* **113**, 13132–13137 (2016).
15. Smith, D. et al. A rare IL33 loss-of-function mutation reduces blood eosinophil counts and protects from asthma. *PLOS Genet.* **13**, e1006659 (2017).
16. Gangwar, R. S., Landolina, N., Arpinati, L. & Levi-Schaffer, F. Mast cell and eosinophil surface receptors as targets for anti-allergic therapy. *Pharmacol. Ther.* **170**, 37–63 (2017).
17. Spergel, J. M. Epidemiology of atopic dermatitis and atopic march in children. *Immunol. Allergy Clin. North. Am.* **30**, 269–280 (2010).
18. Stuart, P. E. et al. Genome-wide association analysis of psoriatic arthritis and cutaneous psoriasis reveals differences in their genetic architecture. *Am. J. Hum. Genet.* **97**, 816–836 (2015).
19. Jin, Y. et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat. Genet.* **48**, 1418–1424 (2016).
20. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
21. Guirguis, E. et al. A role for phosphodiesterase 3B in acquisition of brown fat characteristics by white adipose tissue in male mice. *Endocrinology* **154**, 3152–3167 (2013).
22. Choi, Y. H. et al. Alterations in regulation of energy homeostasis in cyclic nucleotide phosphodiesterase 3B-null mice. *J. Clin. Invest.* **116**, 3240–3251 (2006).
23. Chung, Y. W. et al. White to beige conversion in PDE3B KO adipose tissue through activation of AMPK signaling and mitochondrial function. *Sci. Rep.* **7**, 40445 (2017).
24. Ahmad, F. et al. Phosphodiesterase 3B (PDE3B) regulates NLRP3 inflammasome in adipose tissue. *Sci. Rep.* **6**, 28056 (2016).
25. Emdin, C. et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA: J. Am. Med. Assoc.* **317**, 626–634 (2017).
26. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
27. Cohen, J. et al. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat. Genet.* **37**, 161–165 (2005).
28. Musunuru, K. et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **363**, 2220–2227 (2010).
29. Stitziel, N. O. et al. ANGPTL3 deficiency and protection against coronary artery disease. *J. Am. Coll. Cardiol.* **69**, 2054–2063 (2017).
30. Turcot, V. et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.* **50**, 26–41 (2018).
31. Shah, B. P. et al. MC4R-expressing glutamatergic neurons in the paraventricular hypothalamus regulate feeding and are synaptically connected to the parabrachial nucleus. *Proc. Natl Acad. Sci. USA* **111**, 13193–13198 (2014).
32. Chung, Y. W. et al. Targeted disruption of PDE3B, but not PDE3A, protects murine heart from ischemia/reperfusion injury. *Proc. Natl Acad. Sci. USA* **112**, E2253–E2262 (2015).
33. Tsuchikane, E. et al. Impact of cilostazol on restenosis after percutaneous coronary balloon angioplasty. *Circulation* **100**, 21–26 (1999).
34. Lei, Y. et al. Vaccination against IL-33 inhibits airway hyperresponsiveness and inflammation in a house dust mite model of asthma. *PLOS One* **10**, e0133774 (2015).
35. International Consortium for Blood Pressure Genome-Wide Association Studies. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
36. Do, R. et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102–106 (2015).
37. Khera, A. V. et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J. Am. Coll. Cardiol.* **67**, 2578-2589 (2016).
38. Khera, A. V. et al. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. Preprint at *bioRxiv* https://doi.org/10.1101/218388 (2017).
39. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
40. Warren, H. R. et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49**, 403-415 (2017).


## Acknowledgements

This research has been conducted using the UK Biobank resource, application 7089. The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913) with funding for genotyping provided by Amgen. The VIRGO study was supported by grant R01 HL081153-01A1K from the National Heart, Lung, and Blood Institute. The TAICHI study was supported by the National Health Research Institutes, Taiwan (PH-099-PP-03, PH-100-PP-03, PH-101-PP-03), the National Science Council, Taiwan (Grant Nos NSC 101-2314-B-075A-006-MY3, MOST 104-2314-B-075A-006-MY3, MOST 104-2314-B-075A-007, MOST 105-2314-B-075A-003), the Taichung Veterans General Hospital, Taiwan (TCVGH-1020101C, TCVGH-1020102D, TCVGH-1023102B, TCVGH-1023107D, TCVGH-1030101C, TCVGH-1030105D, TCVGH-1033503C, TCVGH-1033102B, TCVGH-1033108D, TCVGH-1040101C, TCVGH-1040102D, TCVGH-1043504C, TCVGH-1043104B), and the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881. The MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Whole genome sequencing of the VIRGO and TAICHI cohorts was funded by grant 5UM1HG008895-02 from the National Human Genome Research Institute's Center for Common Disease Genomics. Whole genome sequencing of the MESA cohort was funded through the Trans-Omics for Precision Medicine (TOPMed) Program of the National Heart, Lung, and Blood Institute. General study coordination was provided by the TOPMed Data Coordinating Center (3R01HL-12393-02S1). The contributions of the investigators of the NHLBI TOPMed Consortium (https://www.nhlbiwgs.org/topmed-banner-authorship) are gratefully acknowledged. The Atherosclerosis Risk in Communities study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).

## Author contributions

C.A.E., A.V.K., D.K., and S.K. conceived and designed the study. C.A.E., A.V.K., M.C., H. K., D.S., S.G., and S.K. acquired, analysed, and interpreted data. C.A.E., A.V.K., and S.K. drafted the manuscript. All authors revised the manuscript for important intellectual content. S.K. supervised the study.

## Additional information

**Supplementary Information** accompanies this paper at https://doi.org/10.1038/s41467-018-03911-8.

**Competing interests:** A.V.K. is supported by a John S. LaDue Memorial Fellowship at Harvard Medical School, and a KL2/Catalyst Medical Research Investigator Training award from Harvard Catalyst funded by the National Institutes of Health (NIH)

(TR001100) and has received consulting fees from Merck and Amarin. P.N. reports funding from the John S. LaDue Memorial Fellowship at Harvard Medical School and has received consulting fees from Amarin. S.K. is supported by a research scholar award from Massachusetts General Hospital, the Donovan Family Foundation, and R01 HL127564; he has received a research grant from Bayer Healthcare; and consulting fees from Merck, Novartis, Sanofi, AstraZeneca, Alnylam Pharmaceuticals, Leerink Partners, Noble Insights, MedGenome, Aegerion Pharmaceuticals, Regeneron Pharmaceuticals, Quest Diagnostics, Genomics PLC, and Eli Lilly and Company; and holds equity in San Therapeutics and Catabasis Pharmaceuticals. The remaining authors declare no competing interests.



Connor A. Emdin[1,2,3], Amit V. Khera[1,2,3], Mark Chaffin[1,2,3], Derek Klarin[1,3,4], Pradeep Natarajan[1,2,3], Krishna Aragam[1,2,3], Mary Haas[1,2,3], Alexander Bick[1,3], Seyedeh M. Zekavat[1,3,5], Akihiro Nomura[1,2,3], Diego Ardissino[6,7], James G. Wilson[8], Heribert Schunkert[9], Ruth McPherson[10], Hugh Watkins[11,12], Roberto Elosua[13,14,15], Matthew J. Bown[16], Nilesh J. Samani[16], Usman Baber[17], Jeanette Erdmann[18], Namrata Gupta[3], John Danesh[19,20,21], Daniel Chasman[22], Paul Ridker[22], Joshua Denny[23], Lisa Bastarache[23], Judith H. Lichtman[24], Gail D'Onofrio[25], Jennifer Mattera[26], John A. Spertus[27], Wayne H.-H. Sheu[28], Kent D. Taylor[29], Bruce M. Psaty[30,31], Stephen S. Rich[32], Wendy Post[33], Jerome I. Rotter[29], Yii-Der Ida Chen[29], Harlan Krumholz[26], Danish Saleheen[34,35], Stacey Gabriel[3] & Sekar Kathiresan[1,2,3]

[1]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [2]Department of Medicine, Massachusetts General Hospital, Cardiology Division, Harvard Medical School, Boston, MA 02114, USA. [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA. [4]Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [5]Department of Computational Biology & Bioinformatics, Yale Medical School, Yale University, New Haven, MA 06510, USA. [6]Division of Cardiology, Azienda Ospedaliero-Universitaria di Parma, Parma 43121, Italy. [7]Associazione per lo Studio Della Trombosi in Cardiologia, Pavia 27100, Italy. [8]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA. [9]Deutsches Herzzentrum München, Technische Universität München, Deutsches Zentrum für Herz-Kreislauf-Forschung, München 80333, Germany. [10]University of Ottawa Heart Institute, Ottawa, ON K1Y4W7, Canada. [11]Radcliffe Department of Medicine, Division of Cardiovascular Medicine, University of Oxford, Oxford OX1 2JD, UK. [12]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX1 2JD, UK. [13]Cardiovascular Epidemiology and Genetics, Hospital del Mar Research Institute, Barcelona 08003, Spain. [14]CIBER Enfermedades Cardiovasculares (CIBERCV), Barcelona 28029, Spain. [15]Facultat de Medicina, Universitat de Vic-Central de Cataluña, Barcelona, VIC 08500, Spain. [16]Department of Cardiovascular Sciences, University of Leicester, and NIHR Leicester Biomedical Research Centre, Leicester LE1 7RH, UK. [17]The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York 10029 NY, USA. [18]Institute for Integrative and Experimental Genomics, University of Lübeck, Lübeck 23562, Germany. [19]Department of Public Health and Primary Care, Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge CB2 0SR, UK. [20]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [21]National Institute of Health Research Blood and Transplant; Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge CB2 1TN, UK. [22]Center for Cardiovascular Disease Prevention, Brigham and Women's Hospital, Boston 02115, USA. [23]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37235, USA. [24]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT 06510, USA. [25]Department of Emergency Medicine, Yale University, New Haven, CT 06520, USA. [26]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT 06510, USA. [27]Department of Biomedical & Saint Luke's Mid America Heart Institute and the Health Informatics, Division of Endocrinology and Metabolism, University of Missouri-Kansas City, Kansas City, MO 64110, USA. [28]Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 40705, Taiwan. [29]The Institute for Translational Genomics and Population Sciences LABioMed and Department of Pediatrics at Harbor-UCLA Medical Center, Torrance, CA 90095, USA. [30]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle 98195 WA, USA. [31]Cardiovascular Health Research Unit, Kaiser Permanente Washington Health Research Institute, 98101 Seattle, WA, USA. [32]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA. [33]Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [34]Center for Non-Communicable Diseases, Karachi 74800, Pakistan. [35]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. These authors contributed equally: Connor A. Emdin, Amit V. Khera.

ORIGINAL ARTICLE

# Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease

Amit V. Khera, M.D., Connor A. Emdin, D.Phil., Isabel Drake, Ph.D., Pradeep Natarajan, M.D., Alexander G. Bick, M.D., Ph.D., Nancy R. Cook, Ph.D., Daniel I. Chasman, Ph.D., Usman Baber, M.D., Roxana Mehran, M.D., Daniel J. Rader, M.D., Valentin Fuster, M.D., Ph.D., Eric Boerwinkle, Ph.D., Olle Melander, M.D., Ph.D., Marju Orho-Melander, Ph.D., Paul M Ridker, M.D., and Sekar Kathiresan, M.D.

## ABSTRACT

### BACKGROUND

Both genetic and lifestyle factors contribute to individual-level risk of coronary artery disease. The extent to which increased genetic risk can be offset by a healthy lifestyle is unknown.

### METHODS

Using a polygenic score of DNA sequence polymorphisms, we quantified genetic risk for coronary artery disease in three prospective cohorts — 7814 participants in the Atherosclerosis Risk in Communities (ARIC) study, 21,222 in the Women's Genome Health Study (WGHS), and 22,389 in the Malmö Diet and Cancer Study (MDCS) — and in 4260 participants in the cross-sectional BioImage Study for whom genotype and covariate data were available. We also determined adherence to a healthy lifestyle among the participants using a scoring system consisting of four factors: no current smoking, no obesity, regular physical activity, and a healthy diet.

### RESULTS

The relative risk of incident coronary events was 91% higher among participants at high genetic risk (top quintile of polygenic scores) than among those at low genetic risk (bottom quintile of polygenic scores) (hazard ratio, 1.91; 95% confidence interval [CI], 1.75 to 2.09). A favorable lifestyle (defined as at least three of the four healthy lifestyle factors) was associated with a substantially lower risk of coronary events than an unfavorable lifestyle (defined as no or only one healthy lifestyle factor), regardless of the genetic risk category. Among participants at high genetic risk, a favorable lifestyle was associated with a 46% lower relative risk of coronary events than an unfavorable lifestyle (hazard ratio, 0.54; 95% CI, 0.47 to 0.63). This finding corresponded to a reduction in the standardized 10-year incidence of coronary events from 10.7% for an unfavorable lifestyle to 5.1% for a favorable lifestyle in ARIC, from 4.6% to 2.0% in WGHS, and from 8.2% to 5.3% in MDCS. In the BioImage Study, a favorable lifestyle was associated with significantly less coronary-artery calcification within each genetic risk category.

### CONCLUSIONS

Across four studies involving 55,685 participants, genetic and lifestyle factors were independently associated with susceptibility to coronary artery disease. Among participants at high genetic risk, a favorable lifestyle was associated with a nearly 50% lower relative risk of coronary artery disease than was an unfavorable lifestyle. (Funded by the National Institutes of Health and others.)

From the Center for Human Genetic Research and Cardiology Division, Massachusetts General Hospital (A.V.K., P.N., S.K.), and the Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital (N.R.C., D.I.C., P.M.R.), Boston, and the Program in Medical and Population Genetics, Broad Institute, Cambridge (A.V.K., C.A.E., A.G.B., S.K.) — all in Massachusetts; the Department of Clinical Sciences, Lund University, Malmö, Sweden (I.D., O.M., M.O.-M.); the Cardiovascular Institute, Mount Sinai Medical Center, Icahn School of Medicine at Mount Sinai, New York (U.B., R.M., V.F.); Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia (D.J.R.); and the University of Texas Health Science Center School of Public Health, Houston (E.B.). Address reprint requests to Dr. Kathiresan at the Center for Human Genetics Research, Massachusetts General Hospital, 185 Cambridge St., CPZN 5.252, Boston, MA 02114, or at skathiresan1@mgh.harvard.edu.

Drs. Khera and Emdin contributed equally to this article.

This article was published on November 13, 2016, at NEJM.org.

N Engl J Med 2016;375:2349-58.
DOI: 10.1056/NEJMoa1605086

Both genetic and lifestyle factors are key drivers of coronary artery disease, a complex disorder that is the leading cause of death worldwide.[1] A familial pattern in the risk of coronary artery disease was first described in 1938 and was subsequently confirmed in large studies involving twins and prospective cohorts.[2-6] Since 2007, genomewide association analyses have identified more than 50 independent loci associated with the risk of coronary artery disease.[7-15] These risk alleles, when aggregated into a polygenic risk score, are predictive of incident coronary events and provide a continuous and quantitative measure of genetic susceptibility.[16-24]

Much evidence has also shown that persons who adhere to a healthy lifestyle have markedly reduced rates of incident cardiovascular events.[25-30] The promotion of healthy lifestyle behaviors, which include not smoking, avoiding obesity, regular physical activity, and a healthy diet pattern, underlies the current strategy to improve cardiovascular health in the general population.[31]

Many observers assume that a genetic predisposition to coronary artery disease is deterministic.[32] However, genetic risk might be attenuated by a favorable lifestyle. Here, we analyzed data for participants in three prospective cohorts and one cross-sectional study to test the hypothesis that both genetic factors and baseline adherence to a healthy lifestyle contribute independently to the risk of incident coronary events and the prevalent subclinical burden of atherosclerosis. We then determined the extent to which a healthy lifestyle is associated with a reduced risk of coronary artery disease among participants with a high genetic risk.

## Methods

### Study Populations

The Atherosclerosis Risk in Communities (ARIC) study is a prospective cohort that enrolled white participants and black participants between the ages of 45 and 64 years, starting in 1987.[33] For data from this study, we retrieved genotype and clinical data from the National Center for Biotechnology Information dbGAP server (accession number, phs000280.v3.p1). The Women's Genome Health Study (WGHS) is a prospective cohort of female health professionals derived from the Women's Health Study, a clinical trial initiated in 1992 to evaluate the efficacy of aspirin and vitamin E in the primary prevention of cardiovascular disease.[34] The Malmö Diet and Cancer Study (MDCS) is a prospective cohort that enrolled participants between the ages of 44 and 73 years in Malmö, Sweden, starting in 1991.[35] In this study, participants with prevalent coronary disease at baseline were excluded. The BioImage Study enrolled asymptomatic participants between the ages of 55 and 80 years who were at risk for cardiovascular disease, beginning in 2008. This study included quantification of subclinical coronary artery disease in Agatston units, a metric that combines the area and density of observed coronary-artery calcification.[36]

### Polygenic Risk Score

We derived a polygenic risk score from an analysis of up to 50 single-nucleotide polymorphisms (SNPs) that had achieved genomewide significance for association with coronary artery disease in previous studies. Details regarding the cohort-specific genotyping platform and risk scores are provided in Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org.[11-14] An example of the calculation of the polygenic risk score is provided in Table S2 in the Supplementary Appendix. Individual participant scores were created by adding up the number of risk alleles at each SNP and then multiplying the sum by the literature-based effect size.[17] The genetic substructure of the population was assessed by calculating the principal components of ancestry.[37]

### Healthy Lifestyle Factors

We adapted four healthy lifestyle factors from the strategic goals of the American Heart Association (AHA) — no current smoking, no obesity (body-mass index [the weight in kilograms divided by the square of the height in meters], <30), physical activity at least once weekly, and a healthy diet pattern.[31] A healthy diet pattern was ascertained on the basis of adherence to at least half of the following recently endorsed characteristics[38]: consumption of an increased amount of fruits, nuts, vegetables, whole grains, fish, and dairy products and a reduced amount of refined grains, processed meats, unprocessed red meats, sugar-sweetened beverages, trans fats (WGHS

only), and sodium (WGHS only). Because a detailed food-frequency questionnaire was not performed in the BioImage Study, diet scores in that cohort focused on self-reported consumption of fruits, vegetables, and fish. Additional details regarding cohort-specific metrics for lifestyle factors are provided in Table S3 in the Supplementary Appendix.

### STUDY END POINTS

The primary study end point for the prospective cohort populations was a composite of coronary artery disease events that included myocardial infarction, coronary revascularization, and death from coronary causes. End-point adjudication was performed by a committee review of medical records within each cohort. In the BioImage Study, a cross-sectional analysis of baseline scores for coronary-artery calcification was performed.

### STATISTICAL ANALYSIS

We used Cox proportional-hazard models to test the association of genetic and lifestyle factors with incident coronary events. We compared hazard ratios for participants at high genetic risk (i.e., highest quintile of polygenic scores) with those at intermediate risk (quintiles 2 to 4) or low risk (lowest quintile), as described previously.[22,23] Similarly, we compared a favorable lifestyle (which was defined as the presence of at least three of the four healthy lifestyle factors) with an intermediate lifestyle (two healthy lifestyle factors) or an unfavorable lifestyle (no or only one healthy lifestyle factor). The primary analyses included adjustment for age, sex, self-reported education level, and the first five principal components of ancestry (unavailable in MDCS). In addition, WGHS analyses were adjusted for initial trial randomization to aspirin versus placebo and vitamin E versus placebo. We used Cox regression to calculate 10-year event rates, which were standardized to the mean of all predictor variables within each population. Because of a skewed distribution of scores for coronary-artery calcification in the BioImage Study, linear regression was performed on natural log-transformed calcification scores with an offset of 1. Predicted values were then reverse-transformed to calculate standardized scores, with higher values indicating an increased burden of coronary atherosclerosis. All the analyses were performed with the use of R software, version 3.1 (R Project for Statistical Computing).

## RESULTS

The populations in the prospective cohort studies included 7814 of 11,478 white participants in the ARIC cohort, 21,222 of 23,294 white women in the WGHS cohort, and 22,389 of 30,446 participants in the MDCS cohort for whom genotype and covariate data were available (Table 1). During follow-up, 1230 coronary events were observed in the ARIC cohort (median follow-up, 18.8 years), 971 coronary events in the WGHS cohort (median follow-up, 20.5 years), and 2902 coronary events in the MDCS cohort (median follow-up, 19.4 years) (Table S4 in the Supplementary Appendix). Categories of genetic and lifestyle risk were mutually independent within each cohort (Fig. S1 in the Supplementary Appendix).

Polygenic risk scores approximated a normal distribution within each cohort (Fig. S2 in the Supplementary Appendix). A risk gradient was noted across quintiles of genetic risk such that the participants at high genetic risk (i.e., in the top quintile of the polygenic scores) were at significantly higher risk of coronary events than those at low genetic risk (i.e., in the lowest quintile), with adjusted hazard ratios of 1.75 (95% confidence interval [CI], 1.46 to 2.10) in the ARIC cohort, 1.94 (95% CI, 1.58 to 2.39) in the WGHS cohort, and 1.98 (95% CI, 1.76 to 2.23) in the MDCS cohort (Fig. 1, and Table S5 and Fig. S3 in the Supplementary Appendix). Across all three cohorts, the relative risk of incident coronary events was 91% higher among participants at high genetic risk than among those at low genetic risk (hazard ratio, 1.91; 95% CI, 1.75 to 2.09). A family history of coronary artery disease was an imperfect surrogate for genotype-defined risk, although the prevalence of such a self-reported family history tended to be higher among participants at high genetic risk than among those at low genetic risk. Levels of low-density lipoprotein (LDL) cholesterol were modestly increased across categories of genetic risk within each cohort. By contrast, genetic risk categories were independent of other cardiometabolic risk factors and 10-year cardiovascular risk as predicted by the pooled cohorts equation of the American

**Table 1. Characteristics of the Participants at Baseline.***

| Characteristic | Atherosclerosis Risk in Communities (N=7814) | Women's Genome Health Study (N=21,222) | Malmö Diet and Cancer Study (N=22,389) | BioImage Study (N=4260) |
|---|---|---|---|---|
| Age — yr | 54±5.7 | 54.2±7.1 | 58.0±7.7 | 69.1±6.0 |
| Male sex — no. (%) | 3555 (45) | 0 | 8,515 (38) | 1879 (44) |
| Clinical history — no./total no. (%) | | | | |
| Hypertension | 2020/7784 (26) | 5164/21,217 (24) | 13,553/22,389 (61) | 2576/4258 (60) |
| Diabetes mellitus | 632/7799 (8) | 519/21,222 (2) | 904/22,389 (4) | 522/4257 (12) |
| Family history of premature coronary artery disease† | 751/6812 (11) | 2476/19,121 (13) | 7,225/22,389 (32) | 1717/4054 (42) |
| Body-mass index‡ | 26.9±4.8 | 25.9±4.9 | 25.7±3.9 | 28.8±5.5 |
| Lipid levels — mg/dl§ | | | | |
| LDL cholesterol | 136.7±38.7 | 124±34 | 161.2±38.6 | 113±33 |
| HDL cholesterol | 37.6±10.9 | 54±15 | 53.7±14.7 | 56±16 |
| Median triglycerides (IQR) | 110 (79–156) | 119 (84–176) | 102 (76–143) | 148 (107–210) |
| Use of lipid-lowering medication — no. (%) | 45 (1) | 690 (3) | 488 (2) | 1467 (34) |
| Healthy lifestyle factors — no. (%) | | | | |
| No current smoking | 5873 (75) | 18,784 (89) | 16,162 (72) | 3887 (91) |
| No obesity | 6093 (78) | 17,566 (83) | 19,507 (87) | 2729 (64) |
| Regular physical activity | 2743 (35) | 9,256 (44) | 9,093 (41) | 1967 (46) |
| Healthy diet | 1515 (19) | 7,251 (34) | 2,795 (12) | 610 (14) |
| Healthy lifestyle score — no. (%) | | | | |
| 3 or 4 healthy lifestyle factors | 2459 (31) | 10,516 (50) | 7,210 (32) | 1564 (37) |
| 2 healthy lifestyle factors | 3162 (40) | 7,385 (35) | 10,234 (46) | 1598 (38) |
| 0 or 1 healthy lifestyle factor | 2193 (28) | 3,321 (16) | 4,945 (22) | 1098 (26) |
| Genetic risk category — no. (%) | | | | |
| Low | 1563 (20) | 4,280 (20) | 4,478 (20) | 846 (20) |
| Intermediate | 4688 (60) | 12,716 (60) | 13,434 (60) | 2557 (60) |
| High | 1563 (20) | 4,226 (20) | 4,477 (20) | 857 (20) |

* Plus–minus values are means ±SD. P values for the differences between the study groups in each individual cohort at baseline are provided in the tables in the Supplementary Appendix. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. To convert the values for triglycerides to millimoles per liter, multiply by 0.01129. HDL denotes high-density lipoprotein, IQR interquartile range, and LDL low-density lipoprotein.

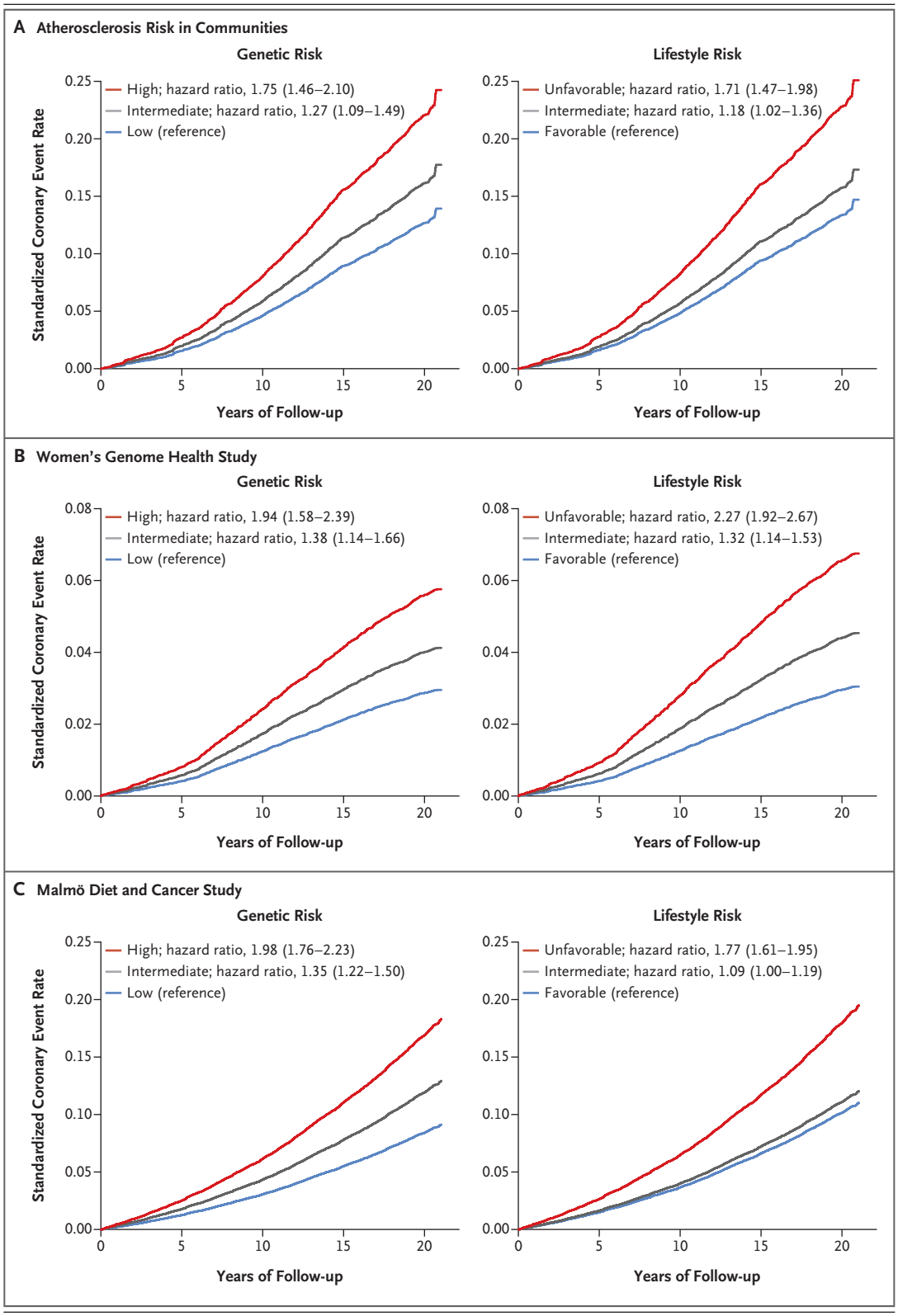
† A family history of premature coronary artery disease refers to a self-reported parental history of myocardial infarction before the age of 60 years. In the BioImage Study and the Malmö Diet and Cancer Study (MDCS), participants were asked about a parental history of myocardial infarction without an age restriction.

‡ The body-mass index is the weight in kilograms divided by the square of the height in meters.

§ Lipid levels were available in a subgroup of 4303 participants in the MDCS study.

**Figure 1 (facing page). Standardized Coronary Events Rates, According to Genetic and Lifestyle Risk in the Prospective Cohorts.**

Shown are the standardized rates of coronary events, according to the genetic risk and lifestyle risk of participants in the Atherosclerosis Risk in Communities (ARIC) cohort, the Women's Genome Health Study (WGHS) cohort, and the Malmö Diet and Cancer Study (MDCS) cohort. The 95% confidence intervals for the hazard ratios are provided in parentheses. Cox regression models were adjusted for age, sex (in ARIC and MDCS), randomization to receive vitamin E or aspirin (in WGHS), education level, and principal components of ancestry (in ARIC and WGHS). Standardization was performed to cohort-specific population averages for each covariate.



College of Cardiology–AHA (Tables S6 through S9 in the Supplementary Appendix).

Each of the four healthy lifestyle factors was associated with a decreased risk of coronary events in a combined analysis of the prospective cohorts: no current smoking (hazard ratio, 0.56; 95% CI, 0.47 to 0.66), no obesity (hazard ratio, 0.66; 95% CI, 0.58 to 0.76), regular physical activity (hazard ratio, 0.88; 95% CI, 0.80 to 0.97), and healthy diet (hazard ratio, 0.91; 95% CI, 0.83 to 0.99) (Table S10 in the Supplementary Appendix). Coronary risk increased among participants with fewer healthy lifestyle factors within each cohort (Table S11 in the Supplementary Appendix).

Each cohort was divided into three lifestyle risk categories: favorable (at least three of the four healthy lifestyle factors), intermediate (two healthy lifestyle factors), or unfavorable (no or only one healthy lifestyle factor). Participants with an unfavorable lifestyle had higher rates of baseline hypertension and diabetes, a higher body-mass index, and less favorable levels of circulating lipids than did those with a favorable lifestyle (Tables S12, S13, and S14 in the Supplementary Appendix). An unfavorable lifestyle was associated with a higher risk of coronary events than a favorable lifestyle, with an adjusted hazard ratio of 1.71 (95% CI, 1.47 to 1.98) in the ARIC cohort, 2.27 (95% CI, 1.92 to 2.67) in the WGHS cohort, and 1.77 (95% CI, 1.61 to 1.95) in the MDCS cohort (Fig. 1, and Fig. S3 in the Supplementary Appendix).

Within each category of genetic risk, lifestyle factors were strong predictors of coronary events (Fig. 2). Adherence to a favorable lifestyle, as compared with an unfavorable lifestyle, was associated with a 45% lower relative risk among participants at low genetic risk, a 47% lower relative risk among those at intermediate genetic risk, and a 46% lower relative risk (hazard ratio, 0.54; 95% CI, 0.47 to 0.63) among those at high genetic risk. Among participants at high genetic risk, the standardized 10-year coronary event rates were 10.7% among those with an unfavorable lifestyle and 5.1% among those with a favorable lifestyle in the ARIC cohort, 4.6% and 2.0%, respectively, in the WGHS cohort, and 8.2% and 5.3% in the MDCS cohort (Fig. 3). Similarly, a low genetic risk was largely offset by an unfavorable lifestyle. Among participants at low genetic risk, standardized 10-year coronary event rates were 5.8% among those with an unfavorable lifestyle and 3.1% among those with a favorable lifestyle in the ARIC cohort, 1.8% and 1.2%, respectively, in the WGHS cohort, and 4.7% and 2.6% in the MDCS cohort. Similar patterns were noted after the exclusion of coronary revascularization from the composite end point (Fig. S4 in the Supplementary Appendix). Adjustment for traditional risk factors attenuated estimates, although the decreased risk among participants with a favorable lifestyle within each genetic risk category remained apparent (Table S15 and Fig. S5 in the Supplementary Appendix).

Despite a paucity of well-validated genetic loci in black populations, we observed similar findings among black participants and white participants in the ARIC cohort (Fig. S6 in the Supplementary Appendix). However, additional data are needed to confirm the consistency of the effect in populations of African ancestry.

A cross-sectional analysis of 4260 of 4301 white participants with available data from the BioImage Study showed that both genetic and lifestyle factors were associated with coronary-artery calcification (stratified according to the baseline characteristics in Tables S16 and S17 in the Supplementary Appendix). The standardized calcification score was 46 Agatston units (95% CI, 39 to 54) among participants at high genetic risk, as compared with 21 Agatston units (95% CI, 18 to 25) among those at low genetic risk ($P<0.001$). The calcification score was similarly higher among participants with an unfavorable lifestyle than among those with a favorable lifestyle: 46 Agatston units (95% CI, 40 to 53) versus 28 Agatston units (95% CI, 25 to 31) ($P<0.001$). Within each subgroup of genetic risk, a significant trend was observed toward decreased coronary-artery calcification among participants who were more adherent to a healthy lifestyle (Fig. 4).

## DISCUSSION

In this study, we have provided quantitative data about the interplay between genetic and lifestyle risk factors for coronary artery disease in three prospective cohorts and one cross-sectional study. High genetic risk was independent of healthy lifestyle behaviors and was associated with an increased risk (hazard ratio, 1.91) of coronary events and a substantially increased burden of coronary-artery calcification. However, within any genetic risk category, adherence to a healthy

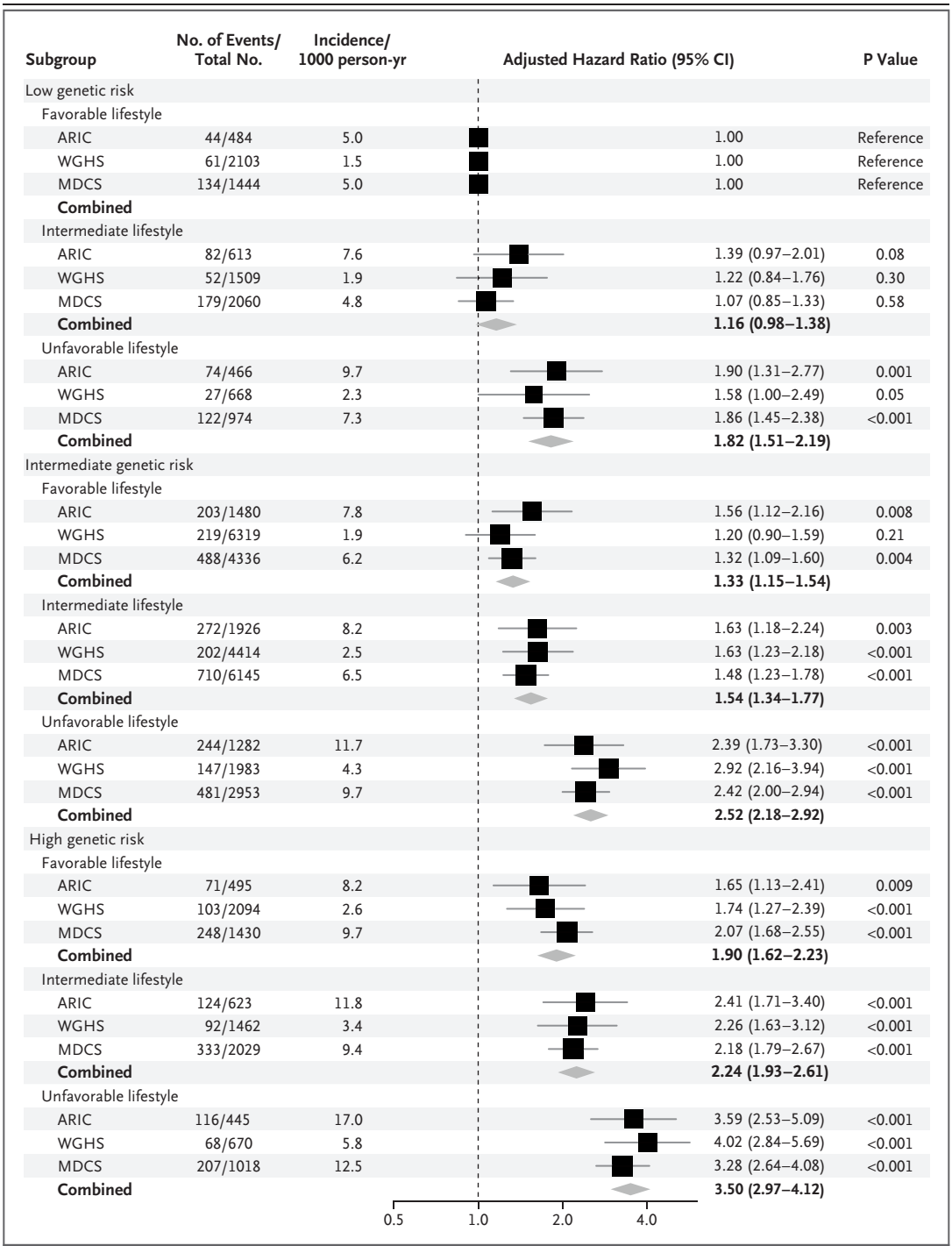



**Figure 2. Risk of Coronary Events, According to Genetic and Lifestyle Risk in the Prospective Cohorts.**

Shown are adjusted hazard ratios for coronary events in each of the three prospective cohorts, according to genetic risk and lifestyle risk. In these comparisons, participants at low genetic risk with a favorable lifestyle served as the reference group. There was no evidence of a significant interaction between genetic and lifestyle risk factors (P=0.38 for interaction in the Atherosclerosis Risk in Communities (ARIC) cohort, P=0.31 in the Women's Genome Health Study (WGHS) cohort, and P=0.24 in the Malmö Diet and Cancer Study (MDCS) cohort). Unadjusted incidence rates are reported per 1000 person-years of follow-up. A random-effects meta-analysis was used to combine cohort-specific results.

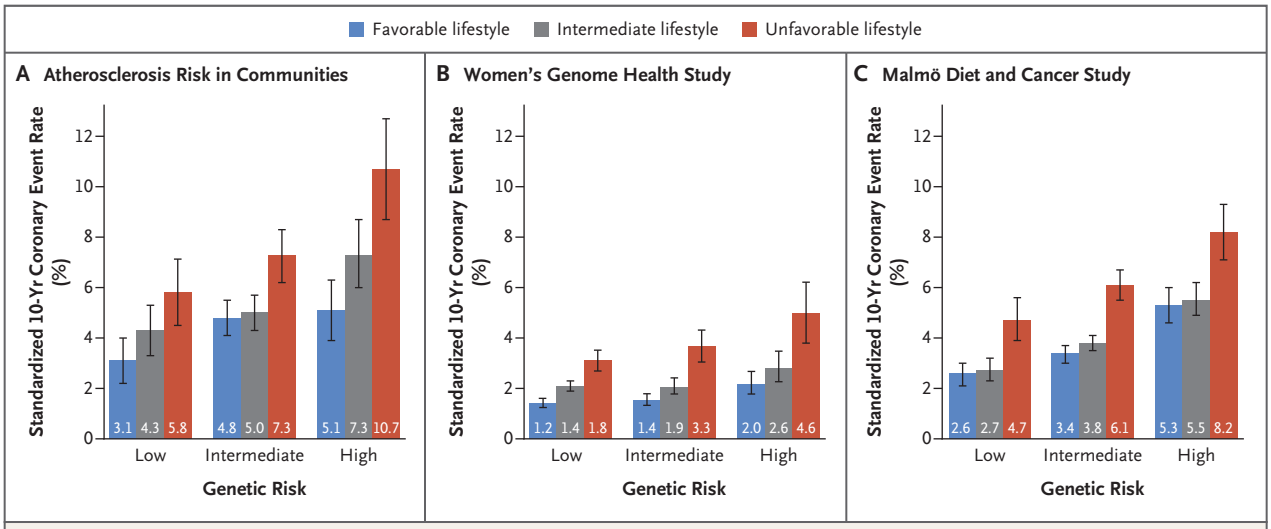



**Figure 3. 10-Year Coronary Event Rates, According to Lifestyle and Genetic Risk in the Prospective Cohorts.**

Shown are standardized 10-year cumulative incidence rates for coronary events in the three prospective cohorts, according to lifestyle and genetic risk. Standardization was performed to cohort-specific population averages for each covariate. The I bars represent 95% confidence intervals.

lifestyle was associated with a significantly decreased risk of both clinical coronary events and subclinical burden of coronary artery disease.

The results of this analysis support three noteworthy conclusions. First, our data indicate that inherited DNA variation and lifestyle factors contribute independently to a susceptibility to coronary artery disease. Our finding that a polygenic risk score has robust associations with incident coronary events is well aligned with previous studies of both primary and secondary prevention populations.[16-24] Such findings support long-standing beliefs that genetic variants that are identifiable from birth alter coronary risk.[2-4] Aside from slight differences in LDL cholesterol levels and a family history of coronary artery disease, genetic risk was independent of traditionally measured risk factors.

Second, a healthy lifestyle was associated with similar relative risk reductions in event rates across each stratum of genetic risk. Although the absolute risk reduction that was associated with adherence to a healthy lifestyle was greatest in the group at high genetic risk, our results support public health efforts that emphasize a healthy lifestyle for everyone. An alternative approach is to target intensive lifestyle modification to those at high genetic risk, with the expectation that disclosure of genetic risk can motivate behavioral change. However, whether the provision of such information can improve cardiovascular outcomes remains to be determined.

Third, patients may equate DNA-based risk estimates with determinism, a perceived lack of control over the ability to improve outcomes.[32] However, our results provide evidence that lifestyle factors may powerfully modify risk regardless of the patient's genetic risk profile. Indeed, alternative analytic approaches that incorporate more stringent cutoffs or weight the relative effect for each healthy lifestyle factor may lead to an even more pronounced coronary risk gradient.

Our study has several limitations. First, because adherence to a healthy lifestyle was not randomized, the association of lifestyle factors with the risk of coronary events cannot be taken as a causal relationship. Second, investigators in each cohort used slightly different methods to assess lifestyle at baseline. Behavioral changes before or after ascertainment or competing risks of other illnesses may have had an effect on risk estimates. Third, although we included up to 50 previously validated genetic polymorphisms in the polygenic risk score, the inclusion of even more variants may prove useful in future analyses.[24] Finally, even though we provide evidence con-

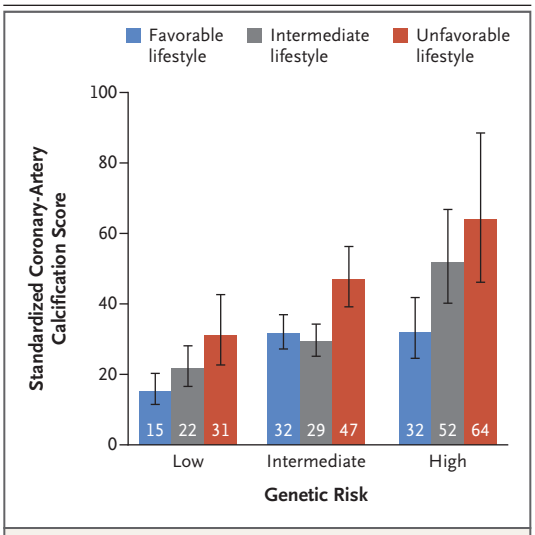



**Figure 4. Coronary-Artery Calcification Score in the BioImage Study, According to Lifestyle and Genetic Risk.**

Among the participants in the BioImage Study, a standardized score for coronary-artery calcification was determined by means of linear regression after adjustment for age, sex, education level, and principal components of ancestry. Standardization was performed on the basis of study averages for each covariate. Average standardized coronary-artery calcification scores are expressed in Agatston units, with higher scores indicating an increased burden of coronary atherosclerosis. The I bars represent 95% confidence intervals.

firming a relationship between a polygenic risk score and coronary events among black participants, the generalizability of our findings should be tested in more diverse populations as more robust ethnicity-specific data regarding genetic association become available.

In conclusion, after quantifying both genetic and lifestyle risk among 55,685 participants in three prospective cohorts and one cross-sectional study, we found that adherence to a healthy lifestyle was associated with a substantially reduced risk of coronary artery disease within each category of genetic risk.

The views expressed in this article are those of the authors and do not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the National Institutes of Health.

Supported by a grant from the American College of Cardiology–Merck Research Fellowship, a John S. Ladue Memorial Fellowship from Harvard Medical School, and a KL2/Catalyst Medical Research Investigator Training award from Harvard Catalyst funded by the National Institutes of Health (NIH) (TR001100, to Dr. Khera). Dr. Kathiresan is supported by an Ofer and Shelly Nemirovsky Research Scholar Award from Massachusetts General Hospital and grants from the NIH (HL127564 and UM1HG008895). Information on support for the cohort studies that are reviewed here is provided in the Supplementary Appendix.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the investigators and participants in the ARIC, WGHS, MDCS, and BioImage studies for their contributions to this study.

## REFERENCES

1. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2095-128.
2. Müller C. Xanthomata, hypercholesterolemia, angina pectoris. Acta Med Scand 1938;89:75-84.
3. Gertler MM, Garn SM, White PD. Young candidates for coronary heart disease. J Am Med Assoc 1951;147:621-5.
4. Slack J, Evans KA. The increased risk of death from ischaemic heart disease in first degree relatives of 121 men and 96 women with ischaemic heart disease. J Med Genet 1966;3:239-57.
5. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. N Engl J Med 1994;330:1041-6.
6. Lloyd-Jones DM, Nam BH, D'Agostino RB Sr, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA 2004;291:2204-11.
7. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med 2007; 357:443-53.
8. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 2007;316:1491-3.
9. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. Science 2007;316:1488-91.
10. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334-41.
11. Erdmann J, Grosshennig A, Braund PS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet 2009;41:280-2.
12. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet 2011;43:339-44.
13. IBC 50K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. PLoS Genet 2011;7(9):e1002260.
14. The CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013;45:25-33.
15. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015; 47:1121-30.
16. Kathiresan S, Melander O, Anevski D, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med 2008;358:1240-9.
17. Ripatti S, Tikkanen E, Orho-Melander M, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet 2010;376:1393-400.
18. Paynter NP, Chasman DI, Paré G, et al. Association between a literature-based genetic risk score and cardiovascular events in women. JAMA 2010;303:631-7.
19. Thanassoulis G, Peloso GM, Pencina MJ, et al. A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham

Heart Study. Circ Cardiovasc Genet 2012;5:113-21.
**20.** Brautbar A, Pompeii LA, Dehghan A, et al. A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. Atherosclerosis 2012;223:421-6.
**21.** Ganna A, Magnusson PK, Pedersen NL, et al. Multilocus genetic risk scores for coronary heart disease prediction. Arterioscler Thromb Vasc Biol 2013;33:2267-72.
**22.** Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet 2015;385:2264-71.
**23.** Tada H, Melander O, Louie JZ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. Eur Heart J 2016;37:561-7.
**24.** Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. Eur Heart J 2016 September 21 (Epub ahead of print).
**25.** Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. N Engl J Med 2000;343:16-22.
**26.** Folsom AR, Yatsuya H, Nettleton JA, Lutsey PL, Cushman M, Rosamond WD. Community prevalence of ideal cardiovascular health, by the American Heart Association definition, and relationship with cardiovascular disease incidence. J Am Coll Cardiol 2011;57:1690-6.
**27.** Yang Q, Cogswell ME, Flanders WD, et al. Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. JAMA 2012;307:1273-83.
**28.** Xanthakis V, Enserro DM, Murabito JM, et al. Ideal cardiovascular health: associations with biomarkers and subclinical disease and impact on incidence of cardiovascular disease in the Framingham Offspring Study. Circulation 2014;130:1676-83.
**29.** Chomistek AK, Chiuve SE, Eliassen AH, Mukamal KJ, Willett WC, Rimm EB. Healthy lifestyle in the primordial prevention of cardiovascular disease among young women. J Am Coll Cardiol 2015;65:43-51.
**30.** Akesson A, Larsson SC, Discacciati A, Wolk A. Low-risk diet and lifestyle habits in the primary prevention of myocardial infarction in men: a population-based prospective cohort study. J Am Coll Cardiol 2014;64:1299-306.
**31.** Lloyd-Jones DM, Hong Y, Labarthe D, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. Circulation 2010;121:586-613.
**32.** White PD. Genes, the heart and destiny. N Engl J Med 1957;256:965-9.
**33.** The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 1989;129:687-702.
**34.** Ridker PM, Chasman DI, Zee RY, et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. Clin Chem 2008;54:249-55.
**35.** Berglund G, Elmstähl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study: design and feasibility. J Intern Med 1993;233:45-51.
**36.** Baber U, Mehran R, Sartori S, et al. Prevalence, impact, and predictive value of detecting subclinical coronary and carotid atherosclerosis in asymptomatic adults: the BioImage study. J Am Coll Cardiol 2015;65:1065-74.
**37.** Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-9.
**38.** Mozaffarian D. Dietary and policy priorities for cardiovascular disease, diabetes, and obesity: a comprehensive review. Circulation 2016;133:187-225.