European Heart Journal

Genetically modulated educational attainment and coronary disease risk

--Manuscript Draft--

Manuscript Number:	EURHEARTJ-D-17-01705R3
Full Title:	Genetically modulated educational attainment and coronary disease risk
Article Type:	Basic Science paper
Keywords:	atherosclerosis; educational attainment; coronary artery disease; genetic risk score; Genetics; genome-wide association studies
Corresponding Author:	Heribert Schunkert, MD Deutsches Herzzentrum München and DZHK (German Centre for Cardiovascular Research) Munich, GERMANY
Order of Authors (with Contributor Roles):	Lingyao Zeng, Ph.D.
	Ioanna Ntalla, Ph.D.
	Thorsten Kessler, M.D.
	Adnan Kastrati, M.D.
	Jeanette Erdmann, Ph.D.
	John Danesh, M.D.
	Hugh Watkins, M.D.
	Nilesh J. Samani, M.D.
	Panos Deloukas, Ph.D.
	Heribert Schunkert, M.D.
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Deutsches Herzzentrum München and DZHK (German Centre for Cardiovascular Research)
Corresponding Author's Secondary Institution:	Partner site Munich Heart Alliance
First Author:	Lingyao Zeng, Ph.D.
Order of Authors Secondary Information:	
Abstract:	Aims: Genetic disposition and lifestyle factors are understood as independent components underlying the risk of multiple diseases. In this study, we aim to investigate the interplay between genetics, educational attainment - an important denominator of lifestyle - and coronary artery disease (CAD) risk. Methods and Results: Based on the effect sizes of 74 genetic variants associated with educational attainment, we calculated a ‘genetic education score’ in 13,080 cases and 14,471 controls and observed an inverse correlation between the score and risk of CAD (p=1.52x10-8; odds ratio [OR] 0.79 (95% confidence interval [CI] 0.73-0.85) for the higher compared to the lowest score quintile). We replicated in 146,514 individuals from UK Biobank (p=1.85x10-6) and also found strong associations between the ‘genetic education score’ with ‘modifiable’ risk factors including smoking (p=5.36x10-23), body mass index (p=1.66x10-30), and hypertension (p=3.86x10-8). Interestingly, these associations were only modestly attenuated by adjustment for years spent in school. By contrast, a model adjusting for BMI and smoking abolished the association signal between the ‘genetic education score’ and CAD risk suggesting an intermediary role of these two risk factors. Mendelian randomization analyses performed with summary statistics from large genome-wide meta-analyses and sensitivity analysis using 1271 variants affecting educational attainment (OR 0.68 for the higher compared to the lowest score quintile; 95% CI 0.63-0.74; p=3.99x10-21) further strengthened

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

	these findings. Conclusion: Genetic variants known to affect educational attainment may have implications for a health-conscious lifestyle later in life and subsequently affect the risk of coronary artery disease.
Suggested Reviewers:	
Opposed Reviewers:	
Additional Information:	
Question	**Response**
Did you cite ESC guidelines where appropriate?	yes
Please enter the names of the authors who Performed statistical analysis	Lingyao Zeng, Ioanna Ntalla
Please enter the names of the authors who Acquired the data	Thorsten Kessler, Adnan Kastrati, Jeanette Erdmann, The UK Biobank CardioMedabolic Consortium CHD working group, John Danesh, Hugh Watkins
Please enter the names of the authors who Conceived and designed the research	Lingyao Zeng, Heribert Schunkert
Please enter the names of the authors who Drafted the manuscript	Heribert Schunkert, Lingyao Zeng, Panos Deloukas, Ioanna Ntalla
Please enter the names of the authors who Made critical revision of the manuscript for key intellectual content	Nilesh Samani, Jeanette Erdmann
Please enter the names of the authors who did anything else on the manuscript other than what we have listed:	N/A
Each person listed as co-author has been entered as contributing to at least one part of the manuscript	Yes
As Corresponding Author, I take full responsibility for all information declared in this notification.	Yes
As Corresponding Author, I agree to be the principal correspondent with the Editorial Office, review the edited manuscript and proof, and make decisions about releasing manuscript information to the media, federal agencies, etc.	Yes
All persons who have made substantial contributions to the manuscript (e.g. data acquisition, analysis, or writing / editing assistance), but who do not fulfill authorship criteria, are named with their specific contributions in the Acknowledgements Section of the	Yes

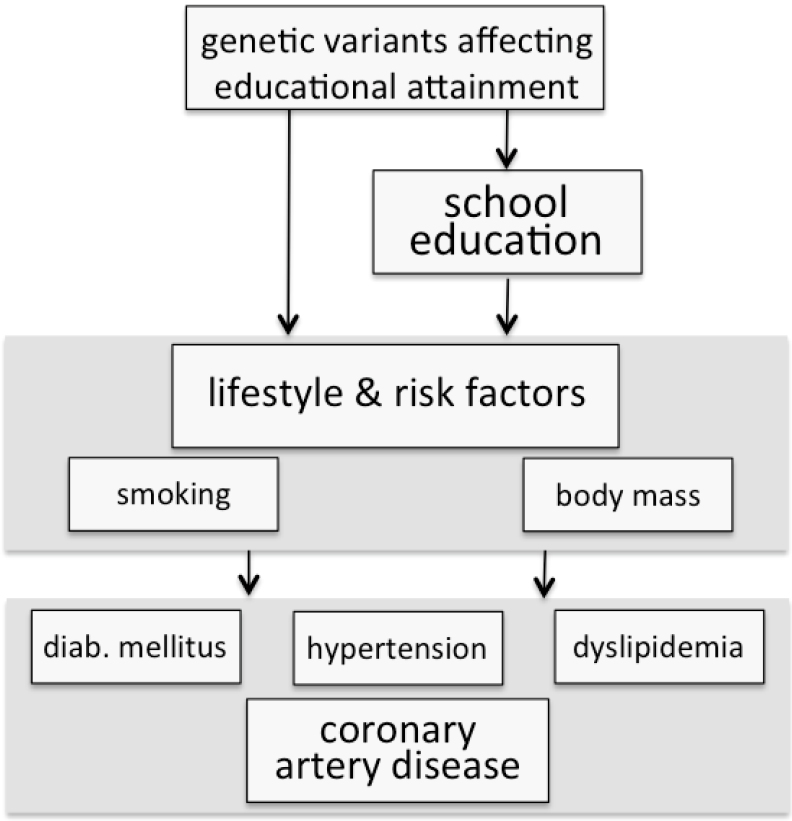
manuscript.	
All persons named in the Acknowledgements Section have provided the Corresponding Author with written permission to be named in the manuscript.	Yes
This manuscript represents valid and substantiated work.	Yes
If an Acknowledgements Section is not included in the paper then no other persons have made substantial contributions to this manuscript.	Yes
If asked, I will provide or fully cooperate in obtaining and providing the original data on which the manuscript is based so the editors or their designates can examine it.	Yes
The paper under question is being submitted by an ESC Working Group.	No
Did you obtain approval for the animal experiments in your study?	N/A
Total Word Count:	10601
Word Count Manuscript-only (excluding references):	4419
Has the trial in question been registered effectively with an appropriate clinical trials registry?	N/A
TWITTER message (Please submit a catchy Twitter message of max. 280 characters, which we would use to promote this submission in the event of acceptance - Max 280 characters).	Some gene variants not only affect our success in school but also the risk of having a heart attack
First Author Secondary Information:	

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Translational Perspective

Genetic disposition and lifestyle factors are understood as independent components underlying the risk of multiple diseases. An important denominator of lifestyle is educational attainment. Here we utilized genetic variants affecting educational attainment as an instrument for studying the interplay between genetics, lifestyle and coronary artery disease (CAD). We found strong associations between the 'genetic education score' and CAD and its risk factors which, interestingly, are only partially explained by years spend in school. Our study shows that genetic variants known to affect educational attainment may have broad implications for a health-conscious lifestyle later in life and the risk of CAD.

'Take-home figure':



'One-sentence Summary': In the present study, we have found that genetic factors known to affect educational attainment may in parallel influence the prevalence of CAD possibly through its influence on the lifestyle-related risk factors.

Genetically modulated educational attainment and coronary disease risk

Lingyao Zeng[1,15], PhD; Ioanna Ntalla[2,3,15], PhD; Thorsten Kessler[1,4], MD; Adnan Kastrati[1,4], MD; Jeanette Erdmann[5,6], PhD; The UK Biobank CardioMetabolic Consortium CHD working group; John Danesh[7,8,9], MD; Hugh Watkins [10,11], MD; Nilesh J. Samani[12,13], MD; Panos Deloukas[2,3,14,15], PhD; Heribert Schunkert[1,4,15], MD

1. Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, 80636 Munich, Germany
2. William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, London EC1M 6BQ, UK
3. Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK
4. Deutsches Zentrum für Herz- und Kreislauferkrankungen (DZHK), partner site Munich Heart Alliance, 80636 Munich, Germany
5. Institute for Cardiogenetics and University Heart Center Luebeck, University of Lübeck, 23562 Lübeck, Germany
6. DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, 23562 Lübeck, Germany
7. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
8. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
9. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1RQ, UK
10. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK
11. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
12. Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK
13. National Institute for Health Research Leicester Cardiovascular Biomedical Research Centre, Leicester LE3 9QP, UK
14. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia
15. These authors contributed equally to this work

Corresponding author:

Prof. Dr. Heribert Schunkert

Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen Technische Universität München Lazarettstr. 36 80636 Munich, Germany
Phone: +49 89/1218-4073, Fax: +49 89/1218-4013, E-Mail: schunkert@dhm.mhn.de

Introduction

Epidemiological studies have repeatedly observed an inverse association between years of school education and coronary artery disease (CAD) risk[1, 2]. A number of detrimental lifestyle factors including smoking, unhealthy diet and less recreational physical activity – observed with less educational attainment – have been considered for explaining this association[3]. However, the correlation between length of school education and CAD risk is difficult to untangle[3]. Indeed, compulsory programmes that increased the years of school education were without measurable effects on CAD risk[4, 5].

The genetics of CAD as well as those of educational attainment have been widely researched in recent years[6-12]. Both, are considered as complex traits with a strong genetic component[6-9]. In parallel, *Mendelian randomisation* has evolved as a valuable tool for investigation of causal relationships between risk factors and complex traits[13]. This raised our interest to explore the genetic impact of educational attainment on coronary disease risk.

In the present study, we specifically analyzed whether there is a genetic component underlying the relationship between higher educational attainment and lower CAD risk (Extended Data Figure 1A, blue block). We also aimed to study the potential intermediate role of lifestyle-related risk factors in linking (the genetic basis of) educational attainment with the prevalence of CAD in the European population (Extended Data Figure 1B, green block).

Methods

Study design

The study addresses two major questions. First, we explored the association between the genetic component of educational attainment and CAD as well as its related risk factors. As complex traits are considered to have polygenic genetic architecture, we resorted to the genetic risk score approach to get a cumulative surrogate estimate for the genetic component of educational attainment, utilizing (n=74 and n=1271) independent single nucleotide polymorphisms (SNPs) reported by two recent published genome-wide association studies (GWAS) for educational attainment with the measurement as the years spent in school (EduYears)[12, 14]. The median effect size of a single SNP corresponds to 1.7 weeks of schooling, and a combined polygenic score explains around 11% of the variance in EduYears[12]. Based on genotype data from multiple cohorts we generated for each individual a 'genetic education score' which reflects the integrated effect of all SNPs affecting EduYears and associated this score with the susceptibility of CAD and other cardiovascular traits by regression analyses.

Definition of educational attainment

Educational attainment is the visible output of education systems and a measure of their success, which varies among different countries, as is shown by the European statistics about education (*http://ec.europa.eu/eurostat/statistics-explained/index.php/Educational_attainment_statistics*). In UKBB, by far the largest cohort explored in our study, educational attainment was available on all cases and controls for subsequent analyses. This information was not available in the nine CAD case-control studies used for discovery. As a standard, the International Standard Classification of Education (ISCED) is nowadays taken as a measure of educational attainment. The details of ISCED codes obtained from UK Biobank phenotypes are described in the Supplementary Text. In the context of genetic association, Okbay et al.[14] have examined two possible indicator variables of educational attainment, namely *EduYear*, i.e. a continuous variable measuring the number of years of schooling completed, and *College*, i.e. a binary variable measuring the successful completion of college education. They revealed that EduYears was

better powered for detecting associations. This EduYear measurement has also been adopted in the latest GWAS for educational attainment[12]. Hence, in the present study, we also calculated real EduYears - a continuous variable measuring the number of years of schooling completed to represent educational attainment - using the same methods that were used in the report that led to the identification of the reported EduYear-SNPs[14].

Genetic risk score analysis

Individual-level genotype data were collected from nine CAD case-control studies[15-21] for discovery. All subjects were of European origin, most coming from Germany and England, and gave written informed consent before participating. Individual-level genotype data were also collected from UK Biobank[22] for replication. All data were utilized for sensitivity analysis with the samples size of UK Biobank increased from 150k to 500k of the latest release. Details on the participating studies and pre-processing methods used are shown in the Supplement Text.

Based on the 74 EduYear SNPs reported by Okbay et al[14] (Extended Data Table 1), a weighted genetic risk score was calculated to evaluate cumulatively the genetic underpinnings of educational attainment and their effect on the risk of CAD. We gave a value from minimum 0 to maximum 2 for every SNP for every individual according to the sum of the posterior probabilities from the imputation files to indicate the number of EduYear-increasing alleles and multiplied the number of alleles by the reported effect sizes. Then we totalled these values for each individual across all 74 SNPs to generate a weighted genetic risk score (wGRS) of EduYears, namely 'genetic education score'. Afterwards, all the individuals were grouped into quintiles based on their 'genetic education score'. Likewise, in the sensitivity analysis we constructed a second 'genetic education score' based on the 1216 SNPs* independent of the initial 74 SNPs* reported by Lee et al[12], as well as the full 1271 SNPs* (*out of the initially reported 74[14] some SNPs were not found in the list of 1271 SNPs shown to be genome-wide significantly associated with EduYears in the most recent analysis[12] Extended Data Table 5).

Logistic regression was performed to evaluate the effect size of the wGRS on the risk of CAD in each study. The wGRS was modelled as a continuous variable and standardized into Z-scores (centred and scaled to have a mean of 0 and standard deviation (SD) of 1) (Extended Data Figure 2). As population stratification or batch effects, which could bias prediction accuracy in the genetic risk score analysis[23], should be considered as covariates, we included in the present study, the top two (for nine case-control studies) or five (for UK Biobank) principle components based on autosomal genotypes in order to adjust for the possible presence of population stratification. In all regression analysis for UK Biobank, genotype array (UK Biobank Axiom array vs UK BiLEVE array) was included as an additional covariate to account for the differences between the two GWAS arrays used for genotyping of participants. The regression was performed for each study separately and afterwards a fixed-effect meta-analysis was performed to combine the effects across all studies (Figure 1, Extended Data Figure 3A).

Using the pooled genotype data of nine CAD case-control cohorts, same logistic regression was performed with the cohort center and top 10 principle components included as additional covariates, so as to estimate the variance explained by the 'genetic education score' on CAD onset in the measure of the incremental McFadden's pseudo R^2. The incremental part was calculated as the pseudo R^2 difference between the regression with and without the score included. The 95% confidence intervals for pseudo R^2 were estimated via 1,000 times bootstrapping.

The 'genetic education score' reflects effects of SNPs that were identified for their association with EduYears in the first place[14]. However, these SNPs may have other (pleiotropic) effects, which may also come into place when the 'genetic education score' is associated with CAD risk. Therefore, the real number of years spend in school (EduYears) was extracted for each individual and included as adjustment to estimate the effect size of the 'genetic education score' on the susceptibility of CAD and multiple traits (Extended Data Figure 1, blue block). With the aim to further characterize the interplay among the genetic basis of educational attainment, the risk of CAD and its related lifestyle risk factors, we included additional adjustments of possible confounders such as lifestyle and risk factors, to check the (in)dependence of the effect of the genetic component of educational attainment on CAD risk.

Using UK Biobank data, regression analyses were also performed to evaluate the effect size of the wGRS on multiple cardiovascular risk factors, namely hypertension, hypercholesterolemia, type 2 diabetes, smoking (ever smokers vs. never smokers), and BMI. Furthermore, the same logistic regression (risk factors measured in binary value) or linear regression (risk factor measured in continuous values, i.e. BMI here) model was performed with adjustment for EduYears, BMI and smoking (ever smokers vs. never smokers [reference group]). The definition of each risk factor is described in Supplementary Text. The same regression of 'genetic education score' and CAD and its risk factors was performed with EduYears (defined in Suplementary Text) included as a covariate. We also tested the association between a 'genetic CAD score' and real EduYears in UK Biobank to investigate the possibility for a reverse causation. (Extended Data Table 2)

Mendelian randomization analysis

Mendelian randomization (MR) analyses were performed in order to investigate the genetic causal effect between educational attainment and CAD or cardiovascular risk factors. The detailed list of the 74 SNPs and their reported effect sizes of EduYear are recorded in Extended Data Table 3. The effect sizes of all these EduYear-SNPs for CAD were extracted from the summary statistics of the CARDIoGRAMplusC4D meta-analysis[24]. Summary statistics of the EduYear-SNPs for cardiovascular risk factors were extracted from various consortia, including GIANT (BMI)[25], TAG (smoking behaviour)[26], and GLGC (LDL-cholesterol, HDL-cholesterol, triglycerides, total cholesterol level)[27]. A description of the sample-size used in each GWAS and the corresponding phenotypes are shown in Supplementary Text.

In all traits the effect size for each SNP was first aligned to the reported EduYear-increasing allele. Then the genetic causal effect was estimated by regressing the SNP-EduYear effect (exposure) to the SNP effect of CAD or risk factors (outcome) in several methods. Inverse-variance-weighted fixed-effect meta-analysis (IVW) is the most classic one in MR to combine individual-SNP beta estimates[28]. Unfortunately, despite of its efficiency IVW estimate will be biased when there exists genetic variant as an invalid instrumental variable (IV). In light of this, modern MR methods such as MR-Egger[29] and weighted

median[30] have been developed to supplement the IVW performance to account for violations of IV assumptions and adjust for potential pleiotropic effects. Therefore, we employed all three methods as recommended by Bowden et al[30], with MR-egger which has substantially less efficiency and low power to suggest whether a causal effect is present or not, and with weighted median which has generally greater power and efficiency to obtain the unbiased estimates.

Furthermore, we performed sensitivity analysis by applying various filters on the original selection of 74 SNPs, i.e. to exclude the SNPs at loci known to affect the risk of CAD, as well as loci known to affect the risk of cardiovascular risk factors (Extended Data Table 3).

Results

We studied 13,080 CAD cases and 14,471 controls from 9 genome-wide association studies (GWAS) on the basis of previously published array data[15-20, 24]. All participants were of Western European decent. The majority of participants came from the UK or Germany (77.8%). Two multinational studies, MIGen and Cardiogenics, also contributed to our sample with 12.1%, 5.5%, 2.3%, 1.9%, and 0.4% of individuals coming from Italy, USA, Spain, Northern Europe, and France, respectively. A detailed cohort description can be found in the Supplementary Text. For each subject we generated individually a weighted 'genetic education score', based on 74 single nucleotide polymorphisms (SNPs) that have been genome-wide significantly associated with educational attainment (EduYear-SNPs) through a GWAS on this trait[14]. The score was normally distributed in the participants (Extended Data Figure 2). Summary statistics for the association of the 'genetic education score' with CAD in each study are shown in Figure 1. As a result of meta-analysis, a higher 'genetic education score' correlated with a lower odds of coronary disease (p=1.52x10^{-8} for fixed-effect meta-analysis). Next, all individuals were grouped into quintiles of the score. Figure 2 shows a constant decline of CAD risk with increasing quintiles. Individuals in the highest quintile had about 21% lower odds than those in the lowest quintile, with individuals in the intermediate quintiles ranging in between (p=7.66x10^{-9} for trend test across quintiles, Figure 2). Sex-stratified analysis

confirmed this inverse correlation in both males and females (Extended Data Table 4). For sensitivity analysis we extracted a new list of 1271 SNPs reported by a larger scale GWAS of educational attainment published during the revision process, and studied a second 'genetic education score' based on n=1216 (out of 1271) SNPs who were all in low linkage disequilibrium (r^2<0.5) with the 74 EduYear-SNPs which we had studied initially (Extended Data Table 5). We observed an even enlarged inverse association between the second 'genetic education score' and CAD risk, with individuals in the highest quintile had about 31% lower odds than those in the lowest quintile (Extended Data Table 6). Quantitatively, the full 'genetic education score' in our data based on 1271 SNPs explains 0.37 % variance [95% CI 0.27-0.51] of CAD onset (Methods).

We then set to replicate and extend our findings in a large population cohort, namely the UK Biobank[22] (Supplementary Text). We successfully replicated the primary association between 'genetic education score' and CAD considering 13,183 CAD cases and 133,203 controls of European ancestry in this study (p=1.85x10^{-6}, Table 1; p=7.34x10^{-13} for meta-analysis with the discovery set, Extended Data Figure 3A and Methods). Sensitivity analysis with the second 'genetic education score' based on a more recent release of the UK Biobank data (38,489 CAD cases and 416,951 controls) also replicated our initial finding (p=1.02x10^{-122}; p=3.88x10^{-132} for meta-analysis with the discovery set, Extended Data Table 6).

We also looked at the countries separately, in that we meta-analysed data from Germany (six German MI family studies), the UK (WTCCC, UK Biobank) and a mixture of Western countries (MIGen, Cardiogenics). We observed directionally identical effects in these three groups despite substantial differences in the educational systems (*http://ec.europa.eu/eurostat/statistics-explained/index.php/Educational_attainment_statistics*) (p=1.05x10^{-6} for the UK; p=5.30x10^{-5} for Germany; p=2.40x10^{-5} for the mixture of several other Western countries, Extended Data Figure 3B).

Phenotypically, real EduYears was also significantly associated with CAD (odds ratio 0.72 for 1 SD increase of education years, 95% CI 0.71, 0.74, p=2.65x10^{-281}) (Extended Data Table 7). However, adjustment for real EduYears had only a small effect on the association between the 'genetic education

score' and CAD (Table 1), suggesting that the 'genetic education score' mediates its effects on CAD risk not entirely through its effects on years spend in school.

We next asked whether there also exists an effect of 'genetic education score' on traditional CAD risk factors in UK Biobank (with the exception of lipid levels, which were not available), and whether the effects were likewise partially independent of the phenotype EduYears. We found strong associations of the 'genetic education score' with hypertension, body mass index (BMI), and smoking whereas other risk factors tested gave nominally significant signals (Table 1). Like with CAD, the phenotype real EduYears was also significantly associated with most cardiovascular risk factors (Extended Data Table 7). However, adjustment for this measure only marginally attenuated the association signal between the 'genetic education score' and the most strongly associated risk factors (e.g. BMI and smoking status; Table 1), again suggesting that school education by itself does not mediate exclusively the association between the 'genetic education score' and risk factors (i.e. higher BMI and smoking) (Extended Data Figure 4, black lines).

Given the strong association of 'genetic education score' with BMI and smoking, even after adjustment for real EduYears in UK Biobank (Table 1), we next tested the effect of the 'genetic education score' on CAD risk, hypertension, hypercholesterolemia, and type 2 diabetes with additional adjustment for these two risk factors. The smoking and BMI adjusted model resulted in a marked attenuation of the association signals for the 'genetic education score' with CAD (OR 0.992 [0.973-1.011], p=0.407), and also with the other risk factors (Extended Data Table 8), suggesting that the effect of the 'genetic education score' on CAD risk may be mainly mediated through its effects on these two risk factors.

In order to investigate the possibility for a reverse causation we also tested the association between a 'genetic CAD score'[31] and real EduYears in UK Biobank (Supplementary Text and Methods) but found no such signal (p=0.58), making reverse causation unlikely.

To substantiate our previous observations that the genetic component of educational attainment has significant impact on CAD risk and that this effect may be mediated through its effects on

intermediate risk factors, we also performed Mendelian randomisation (MR) analyses. We used summary statistics from well-powered GWAS for CAD[24], BMI[25], smoking behaviour[26], LDL-cholesterol, HDL-cholesterol, triglycerides, total cholesterol level[27], to regress the SNP-education effect (exposure) with the SNP-effects on CAD and its risk factors (outcome), and then several modern methods were employed to combine individual-SNP beta estimates (Supplementary Text and Methods). Using the inverse-variance weighted (IVW) and weighted median methods, we observed that risk of CAD decreased by about a third per 1-standard deviation (SD) increment in the education years, with odds ratios of 0.64 (95% CI: 0.55, 0.76, p=9.5 x10^{-7}) and 0.67 (95% CI: 0.55-0.82, p=0.00013) respectively. Among the tested risk factors, BMI (p=0.04 for IVW, p=0.01 for weighted median) and triglycerides (p=0.03 for IVW, and p=0.06 for weighted median) were also significantly associated with the 'genetic education score' in both methods. Smoking behaviours measured as either number of cigarettes smoked per day or ever-smoked in the public available data collection from 2010[26], showed marginal genetic association with educational attainment (*p-value* about 0.06) in either method, which could be due to lack of power (Extended Data Table 9).

For MR sensitivity analyses we removed from the 74 EduYear-SNPs which had marginal associations (p<0.01 in the published summary statistics) with CAD[24] (4 SNPs), BMI[25] (8 SNPs), triglycerides[27] (5 SNPs), and smoking behaviours[26] (3 SNPs) (Extended Data Table 3). The final 56 EduYear-SNPs still showed a causal effect between higher educational attainment (1-SD increase in education years) and 28% lower risk of CAD (odds ratio for weighted median method of 0.72, 95% CI 0.57, 0.90, p=4.3x10^{-3}), which further endorsed the effect found in the initial MR analysis (Extended Data Figure 5).

Discussion

Educational attainment is well known for its inverse association with cardiovascular diseases[1, 2,3]. Here we show that a genetic score based on cumulative effects of variants associated with the number of years

spend in school[14] is associated with multiple cardiovascular risk factors and the manifestation of CAD. Interestingly, these associations were only ameliorated by adjustment for real years of schooling suggesting that the genetic component of educational attainment may have more complex phenotypic consequences that all contribute to the observed statistical findings (Figure 3 and Extended Data Figure 4A). Furthermore, we show that the genetic effect between educational attainment and CAD risk is partially mediated via specific lifestyle and risk factors (BMI and smoking status) (Figure 3).

Notably, our findings imply that the health benefits related to higher educational attainment or more years in education *per se* might have been over-estimated, given that the associations observed for the 'genetic education score' with CAD risk as well as a number of health-related outcomes are not exclusively mediated by EduYears. Our data suggest these health-related outcomes are in part modulated by certain genetic variants, which lead, on the one hand, to more EduYears, and on the other hand, to characteristics that are independent of EduYears but go along with a healthier lifestyle (Figure 3). This notion possibly explains why previous compulsory programmes that increased the years of school education were not showing a measurable reduction of CAD risk [4, 5]. However, we do not want to conclude that a causal relation between EduYears and CAD risk does not exist. Indeed, the effect size of the 'genetic education score' and CAD was markedly weakened – but nevertheless remained to be significant – after adjustment for actual EduYears.

Indeed, the 'genetic education score' may have broader implications since it may relate to a number of socioeconomic measures in respective individuals as well as the parental generation (Extended Figure 4B illustrates some hypothetical consequences)[1, 2,3,8, 32-40]. Given such complexity of phenotypes studied here, our findings fall short delineating the precise mechanism for each SNP. However, we uncover that smoking and obesity may be key intermediary factors for the link between the 'genetic education score' as a whole and CAD risk, as they not only blunted the association but also appeared to be very robust in mediating the effects of the score on other cardiovascular risk factors, including hypertension, hypercholesterolemia, and type 2 diabetes (Extended Data Table 8). Given these and probably more intermediate factors to be discovered we obviously do not aim to implement an actionable

genetic test in a clinical environment but rather aim to claim a genetic link between educational attainment, intermediate risk factors, and CAD risk.

Since both intermediate risk factors – smoking and obesity – reflect modifiable health-relevant lifestyle decisions, we hypothesize that addressing these factors in Western societies could attenuate largely the inequality of CAD risk related to the EduYear-SNPs. Our findings therefore further suggest that genetics contribute to the variability of health-relevant lifestyles in Western societies and thereby have an effect on CAD risk later in life (Figure 3). Consistent with our findings, e.g. Marioni et al reported a higher life expectancy in parents of subjects with a higher 'genetic education score'[37] and Arden et al revealed that the association between a longer lifespan and intelligence is mostly genetic[38]. In synthesis these observations suggest that genetic components and their underlying biological traits may influence the length of educational attainment but also the decision-making process in lifestyle choices that underpin cardiovascular risk factors such as smoking and obesity[33]. Our findings also emphasize that future studies on factors contributing to CAD risk should pay more attention to a standardized data collection on educational attainment as a potential factor.

We focused our analyses on CAD and cardiovascular outcomes. However, the lifestyle-related intermediary phenotypes such as smoking and obesity may have implications for the risk of other complex conditions as well, e.g. malignant, pulmonary or infectious diseases. In an exploratory analysis we also observed a strong association between the 'genetic education score' and chronic obstructive pulmonary disease (COPD) in UK Biobank but no statistically significant association with peripheral arterial disease, lung cancer or stroke (Extended Data Figure 6). Thus, the health implications of the EduYear-SNPs may be even be broader.

Our study has a number of limitations. Firstly, the level of educational attainment and the number of years spend in school are different among countries. Likewise, lifestyle, cultural background and environmental exposures also differ between countries[41]. The EduYear-SNPs utilized in this study were identified from a GWAS meta-analysis largely based on individuals of European descent[14]. To avoid any

extrapolation bias we have restricted our analysis to individuals from the UK, Germany and other Western European countries, which limits the interpretation for other ethnic groups or countries. Secondly, it is likely that a number of further genetic and socioeconomic factors[8] are involved in the interplay of educational attainment, intermediary factors and prevalence of CAD (Extended Data Figure 4B). Specifically, environmental exposures may differ between countries and could also be confounders of the genetic effects of education on CAD risk, which we lacked investigating in the current study. Moreover, we arbitrarily restricted our analysis to EduYear-SNPs with established genome-wide significance for association with educational attainment. It is likely that larger meta-analyses will identify more genetic variants affecting school attainment as well as other socioeconomic factors[42], such that the genetic effects unravelled here are likely to underestimate the true effect. Thirdly, many of the lifestyle and risk factors reported in the UK Biobank are self-reported and not externally validated, which might have underestimated some of the effects. Calculating years of school attainment for participants of the UK Biobank may also have some limitations[14]. We employed the same methods that were used in the report that led to the identification of the EduYear-SNPs[14]. The fact that years of school attainment were also strongly related to CAD and its risk factors supports the validity of this calculation. Next, the currently available functional information on the EduYear-SNPs is fairly basic. It includes the genomic position, allele frequencies, genes in the vicinity and signals from GWAS studies on other traits including CAD and risk factors (Extended Data Table 3). It has been inferred that the 'genetic education score' may reflect motivation[39], cognitive ability[35, 36], as well as an array of biological pathways[14] which also could influence lifetime cardiovascular risk. However, we still have no precise information on the mechanisms by which the EduYear-SNPs led to differences in educational attainment or even more complex traits such as smoking and therefore coronary disease risk. We annotated these SNPs to physically nearby genes and putative functions (Extended Data Table 1), with the hope to provide some information on the biological context and the diversity of mechanisms underlying the SNPs affecting educational attainment. Future research will need to unravel how these variants affect educational attainment or the prevalence of smoking and obesity.

Finally, association findings in such complex settings – even if they are based on genetic variants which allow inferring directionality in terms of a Bayesian approach – may reflect reverse causation. In this sense a reflection of the Bradford Hill Criteria[43] may be of help. The associations reported here show 1.) substantial **strength** (effect size), i.e. over 20% difference between the highest/lowest quintile, 2.) **consistency** (reproducibility) as they were reproduced in an independent sample, and with an independent set of SNPs. 3.) **specificity**, i.e. causation was built on the Mendelian randomization approach, 4.) **temporality**, i.e. genetics come first - by nature, 5.) a **biological gradient**, i.e. there is a stepwise effect across the quintiles, 6.) **plausibility,** i.e. the chain of events between SNPs —> school attainment —> smoking/obesity —> CAD is clearly plausible, 7.) **coherence** with the literature on educational attainment and cardiovascular risk, and 8.) **analogy** in that the effect size of 'educational attainment' on 'CAD risk' is analogous with that of other risk factors. We have to admit, however, that 9.) **experimental validation** cannot be obtained in this setting. Rather, we want to suppose that with the event of large-scale GWAS statistical significance—not necessarily the magnitude of association—is the accepted benchmark for judging the strength of an observed association between a genetic variant and a phenotype, and thus its potential causality [44].

Genetic and lifestyle factors are often contrasted regarding their influence on coronary disease risk[31]. An important determinant of lifestyle and related health outcomes is educational attainment. Our study revealed that the genetic basis of educational attainment – like educational attainment itself[6-9, 14, 31, 37] – is a partly independent element linked with lifestyle factors affecting CAD risk. Thus, our data give rise to the hypothesis that the attitude towards a health-conscious lifestyle includes an inherited component affecting educational attainment and decision making later in life, a finding which may have broad implications for battling cardiovascular risks in Western societies.

Acknowledgements

We wish to thank Prof. Dr. Bertram Müller-Myhsok, Munich, and Prof. Dr. John R. Thompson, Leicester for statistical advice.

This work was supported by grants from the Fondation Leducq (CADgenomics: Understanding CAD Genes, 12CVD02), the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (e:AtheroSysMed, grant 01ZX1313A-2014), and the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no HEALTH-F2-2013-601456 (CVgenes-at-target). Further grants were received from the DFG as part of the Sonderforschungsbereich CRC 1123 (B2). T.K. was supported by a DZHK Rotation Grant. Analysis in UK Biobank was funded by British Heart Foundation (BHF) grant RG/14/5/30893 to P.D. and forms part of the research themes contributing to the translational research portfolio of Barts Cardiovascular Biomedical Research Centre which is funded by the National Institute for Health Research (NIHR).

Conflict of Interest

Dr. Danesh reports grants from UK Medical Research Council, grants from British Heart Foundation, grants from UK National Institute of Health Research, grants from European Commission, during the conduct of the study; personal fees and non-financial support from Merck Sharpe & Dohme UK Atherosclerosis, personal fees and non-financial support from Novartis Cardiovascular and Metabolic Advisory Board, personal fees and non-financial support from Pfizer Population Research Advisory Panel, grants from British Heart Foundation, grants from European Research Council, grants from Merck, grants from National Institute of Health Research, grants from NHS Blood and Transplant, grants from Novartis, grants from Pfizer, grants from UK Medical Research Council, grants from Wellcome Trust, outside the submitted work.

References

1. Manrique-Garcia E, Sidorchuk A, Hallqvist J, Moradi T. Socioeconomic position and incidence of acute myocardial infarction: a meta-analysis. Journal of epidemiology and community health 2011;**65**(4):301-9.
2. Veronesi G, Ferrario MM, Kuulasmaa K, Bobak M, Chambless LE, Salomaa V, Soderberg S, Pajak A, Jorgensen T, Amouyel P, Arveiler D, Drygas W, Ferrieres J, Giampaoli S, Kee F, Iacoviello L, Malyutina S, Peters A, Tamosiunas A, Tunstall-Pedoe H, Cesana G. Educational class inequalities in the incidence of coronary heart disease in Europe. Heart (British Cardiac Society) 2016;**102**(12):958-65.
3. Brunello G, Fort M, Schneeweis N, Winter-Ebmer R. The Causal Effect of Education on Health: What is the Role of Health Behaviors? Health economics 2016;**25**(3):314-36.
4. Lager AC, Torssander J. Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. Proceedings of the National Academy of Sciences of the United States of America 2012;**109**(22):8461-6.
5. Clark D, Royer H. The Effect of Education on Adult Mortality and Health: Evidence from Britain. American Economic Review 2013;**103**(6):2087-2120.
6. Benjamin DJ, Cesarini D, van der Loos MJHM, Dawes CT, Koellinger PD, Magnusson PKE, Chabris CF, Conley D, Laibson D, Johannesson M, Visscher PM. The genetic architecture of economic and political preferences. Proceedings of the National Academy of Sciences of the United States of America 2012;**109**(21):8026-8031.
7. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra HJ, Shakhbazov K, Abdellaoui A, Agrawal A, Albrecht E, Alizadeh BZ, Amin N, Barnard J, Baumeister SE, Benke KS, Bielak LF, Boatman JA, Boyle PA, Davies G, de Leeuw C, Eklund N, Evans DS, Ferhmann R, Fischer K, Gieger C, Gjessing HK, Hagg S, Harris JR, Hayward C, Holzapfel C, Ibrahim-Verbaas CA, Ingelsson E, Jacobsson B, Joshi PK, Jugessur A, Kaakinen M, Kanoni S, Karjalainen J, Kolcic I, Kristiansson K, Kutalik Z, Lahti J, Lee SH, Lin P, Lind PA, Liu Y, Lohman K, Loitfelder M, McMahon G, Vidal PM, Meirelles O, Milani L, Myhre R, Nuotio ML, Oldmeadow CJ, Petrovic KE, Peyrot WJ, Polasek O, Quaye L, Reinmaa E, Rice JP, Rizzi TS, Schmidt H, Schmidt R, Smith AV, Smith JA, Tanaka T, Terracciano A, van der Loos MJ, Vitart V, Volzke H, Wellmann J, Yu L, Zhao W, Allik J, Attia JR, Bandinelli S, Bastardot F, Beauchamp J, Bennett DA, Berger K, Bierut LJ, Boomsma DI, Bultmann U, Campbell H, Chabris CF, Cherkas L, Chung MK, Cucca F, de Andrade M, De Jager PL, De Neve JE, Deary IJ, Dedoussis GV, Deloukas P, Dimitriou M, Eiriksdottir G, Elderson MF, Eriksson JG, Evans DM, Faul JD, Ferrucci L, Garcia ME, Gronberg H, Guethnason V, Hall P, Harris JM, Harris TB, Hastie ND, Heath AC, Hernandez DG, Hoffmann W, Hofman A, Holle R, Holliday EG, Hottenga JJ, Iacono WG, Illig T, Jarvelin MR, Kahonen M, Kaprio J, Kirkpatrick RM, Kowgier M, Latvala A, Launer LJ, Lawlor DA, Lehtimaki T, Li J, Lichtenstein P, Lichtner P, Liewald DC, Madden PA, Magnusson PK, Makinen TE, Masala M, McGue M, Metspalu A, Mielck A, Miller MB, Montgomery GW, Mukherjee S, Nyholt DR, Oostra BA, Palmer LJ, Palotie A, Penninx BW, Perola M, Peyser PA, Preisig M, Raikkonen K, Raitakari OT, Realo A, Ring SM, Ripatti S, Rivadeneira F, Rudan I, Rustichini A, Salomaa V, Sarin AP, Schlessinger D, Scott RJ, Snieder H, St Pourcain B, Starr JM, Sul JH, Surakka I, Svento R, Teumer A, Tiemeier H, van Rooij FJ, Van Wagoner DR, Vartiainen E, Viikari J, Vollenweider P, Vonk JM, Waeber G, Weir DR, Wichmann HE, Widen E, Willemsen G, Wilson JF, Wright AF, Conley D, Davey-Smith G, Franke L, Groenen PJ, Hofman A, Johannesson M, Kardia SL, Krueger RF, Laibson D, Martin NG, Meyer MN, Posthuma D, Thurik AR, Timpson NJ, Uitterlinden AG, van Duijn CM, Visscher PM, Benjamin DJ, Cesarini D, Koellinger PD. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science 2013;**340**(6139):1467-71.
8. Rietveld CA, Conley D, Eriksson N, Esko T, Medland SE, Vinkhuyzen AA, Yang J, Boardman JD, Chabris CF, Dawes CT, Domingue BW, Hinds DA, Johannesson M, Kiefer AK, Laibson D, Magnusson PK, Mountain JL, Oskarsson S, Rostapshova O, Teumer A, Tung JY, Visscher PM, Benjamin DJ, Cesarini D, Koellinger PD. Replicability and robustness of genome-wide-association studies for behavioral traits. Psychological science 2014;**25**(11):1975-86.
9. Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE, Ritchie SJ, Luciano M, Fawns-Ritchie C, Lyall D, Cullen B, Cox SR, Hayward C, Porteous DJ, Evans J, McIntosh AM, Gallacher J, Craddock N, Pell JP, Smith DJ, Gale CR, Deary IJ. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Molecular psychiatry 2016;**21**(6):758-67.
10. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang SJ, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikainen LP, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han BG,

Huang J, Jalilzadeh S, Kessler T, Konig IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki ML, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon FU, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim BJ, Kooner JS, Kullo IJ, Lehtimaki T, Loos RJF, Melander O, Metspalu A, Marz W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics 2015;**47**(10):1121-1130.

11. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, Zeng L, Ntalla I, Lai FY, Hopewell JC, Giannakopoulou O, Jiang T, Hamby SE, Di Angelantonio E, Assimes TL, Bottinger EP, Chambers JC, Clarke R, Palmer CNA, Cubbon RM, Ellinor P, Ermel R, Evangelou E, Franks PW, Grace C, Gu D, Hingorani AD, Howson JMM, Ingelsson E, Kastrati A, Kessler T, Kyriakou T, Lehtimaki T, Lu X, Lu Y, Marz W, McPherson R, Metspalu A, Pujades-Rodriguez M, Ruusalepp A, Schadt EE, Schmidt AF, Sweeting MJ, Zalloua PA, AlGhalayini K, Keavney BD, Kooner JS, Loos RJF, Patel RS, Rutter MK, Tomaszewski M, Tzoulaki I, Zeggini E, Erdmann J, Dedoussis G, Bjorkegren JLM, Schunkert H, Farrall M, Danesh J, Samani NJ, Watkins H, Deloukas P. Association analyses based on false discovery rate implicate new loci for coronary artery disease.
Nat Genet 2017 Sep; **49**(9):1385-1391.

12. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linner R, Fontana MA, Kundu T, Lee C, Li H, Li R, Royer R, Timshel PN, Walters RK, Willoughby EA, Yengo L, Alver M, Bao Y, Clark DW, Day FR, Furlotte NA, Joshi PK, Kemper KE, Kleinman A, Langenberg C, Magi R, Trampush JW, Verma SS, Wu Y, Lam M, Zhao JH, Zheng Z, Boardman JD, Campbell H, Freese J, Harris KM, Hayward C, Herd P, Kumari M, Lencz T, Luan J, Malhotra AK, Metspalu A, Milani L, Ong KK, Perry JRB, Porteous DJ, Ritchie MD, Smart MC, Smith BH, Tung JY, Wareham NJ, Wilson JF, Beauchamp JP, Conley DC, Esko T, Lehrer SF, Magnusson PKE, Oskarsson S, Pers TH, Robinson MR, Thom K, Watson C, Chabris CF, Meyer MN, Laibson DI, Yang J, Johannesson M, Koellinger PD, Turley P, Visscher PM, Benjamin DJ, Cesarini D. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 2018;**50**(8):1112-1121.

13. Jansen H, Samani NJ, Schunkert H. Mendelian randomization studies in coronary artery disease. European heart journal 2014;**35**(29):1917-24.

14. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen GB, Emilsson V, Meddens SF, Oskarsson S, Pickrell JK, Thom K, Timshel P, de Vlaming R, Abdellaoui A, Ahluwalia TS, Bacelis J, Baumbach C, Bjornsdottir G, Brandsma JH, Pina Concas M, Derringer J, Furlotte NA, Galesloot TE, Girotto G, Gupta R, Hall LM, Harris SE, Hofer E, Horikoshi M, Huffman JE, Kaasik K, Kalafati IP, Karlsson R, Kong A, Lahti J, van der Lee SJ, deLeeuw C, Lind PA, Lindgren KO, Liu T, Mangino M, Marten J, Mihailov E, Miller MB, van der Most PJ, Oldmeadow C, Payton A, Pervjakova N, Peyrot WJ, Qian Y, Raitakari O, Rueedi R, Salvi E, Schmidt B, Schraut KE, Shi J, Smith AV, Poot RA, St Pourcain B, Teumer A, Thorleifsson G, Verweij N, Vuckovic D, Wellmann J, Westra HJ, Yang J, Zhao W, Zhu Z, Alizadeh BZ, Amin N, Bakshi A, Baumeister SE, Biino G, Bonnelykke K, Boyle PA, Campbell H, Cappuccio FP, Davies G, De Neve JE, Deloukas P, Demuth I, Ding J, Eibich P, Eisele L, Eklund N, Evans DM, Faul JD, Feitosa MF, Forstner AJ, Gandin I, Gunnarsson B, Halldorsson BV, Harris TB, Heath AC, Hocking LJ, Holliday EG, Homuth G, Horan MA, Hottenga JJ, de Jager PL, Joshi PK, Jugessur A, Kaakinen MA, Kahonen M, Kanoni S, Keltigangas-Jarvinen L, Kiemeney LA, Kolcic I, Koskinen S, Kraja AT, Kroh M, Kutalik Z, Latvala A, Launer LJ, Lebreton MP, Levinson DF, Lichtenstein P, Lichtner P, Liewald DC, LifeLines Cohort S, Loukola A, Madden PA, Magi R, Maki-Opas T, Marioni RE, Marques-Vidal P, Meddens GA, McMahon G, Meisinger C, Meitinger T, Milaneschi Y, Milani L, Montgomery GW, Myhre R, Nelson CP, Nyholt DR, Ollier WE, Palotie A, Paternoster L, Pedersen NL, Petrovic KE, Porteous DJ, Raikkonen K, Ring SM, Robino A, Rostapshova O, Rudan I, Rustichini A, Salomaa V, Sanders AR, Sarin AP, Schmidt H, Scott RJ, Smith BH, Smith JA, Staessen JA, Steinhagen-Thiessen E, Strauch K, Terracciano A, Tobin MD, Ulivi S, Vaccargiu S, Quaye L, van Rooij FJ, Venturini C, Vinkhuyzen AA, Volker U, Volzke H, Vonk JM, Vozzi D, Waage J, Ware EB, Willemsen G, Attia JR, Bennett DA, Berger K, Bertram L, Bisgaard H, Boomsma DI, Borecki IB, Bultmann U, Chabris CF, Cucca F, Cusi D, Deary IJ, Dedoussis GV, van Duijn CM, Eriksson JG, Franke B, Franke L, Gasparini P, Gejman PV, Gieger C, Grabe HJ, Gratten J, Groenen PJ, Gudnason V, van der Harst P, Hayward C, Hinds DA, Hoffmann W, Hypponen E, Iacono WG, Jacobsson B, Jarvelin MR, Jockel KH, Kaprio J, Kardia SL, Lehtimaki T, Lehrer SF, Magnusson PK, Martin NG, McGue M, Metspalu A, Pendleton N, Penninx BW, Perola M, Pirastu N, Pirastu M, Polasek O, Posthuma D, Power C, Province MA, Samani NJ, Schlessinger D, Schmidt R, Sorensen TI,

Spector TD, Stefansson K, Thorsteinsdottir U, Thurik AR, Timpson NJ, Tiemeier H, Tung JY, Uitterlinden AG, Vitart V, Vollenweider P, Weir DR, Wilson JF, Wright AF, Conley DC, Krueger RF, Davey Smith G, Hofman A, Laibson DI, Medland SE, Meyer MN, Yang J, Johannesson M, Visscher PM, Esko T, Koellinger PD, Cesarini D, Benjamin DJ. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 2016;**533**(7604):539-42.

15. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. The New England journal of medicine 2007;**357**(5):443-53.

16. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet DA, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, El Mokhtari NE, Schafer A, Marz W, Renner W, Bugert P, Kluter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet 2009;**41**(3):280-2.

17. Erdmann J, Willenborg C, Nahrstaedt J, Preuss M, Konig IR, Baumert J, Linsel-Nitschke P, Gieger C, Tennstedt S, Belcredi P, Aherrahrou Z, Klopp N, Loley C, Stark K, Hengstenberg C, Bruse P, Freyer J, Wagner AK, Medack A, Lieb W, Grosshennig A, Sager HB, Reinhardt A, Schafer A, Schreiber S, El Mokhtari NE, Raaz-Schrauder D, Illig T, Garlichs CD, Ekici AB, Reis A, Schrezenmeir J, Rubin D, Ziegler A, Wichmann HE, Doering A, Meisinger C, Meitinger T, Peters A, Schunkert H. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. European heart journal 2011;**32**(2):158-68.

18. Stitziel NO, Won HH, Morrison AC, Peloso GM, Do R, Lange LA, Fontanillas P, Gupta N, Duga S, Goel A, Farrall M, Saleheen D, Ferrario P, Konig I, Asselta R, Merlini PA, Marziliano N, Notarangelo MF, Schick U, Auer P, Assimes TL, Reilly M, Wilensky R, Rader DJ, Hovingh GK, Meitinger T, Kessler T, Kastrati A, Laugwitz KL, Siscovick D, Rotter JI, Hazen SL, Tracy R, Cresci S, Spertus J, Jackson R, Schwartz SM, Natarajan P, Crosby J, Muzny D, Ballantyne C, Rich SS, O'Donnell CJ, Abecasis G, Sunaev S, Nickerson DA, Buring JE, Ridker PM, Chasman DI, Austin E, Kullo IJ, Weeke PE, Shaffer CM, Bastarache LA, Denny JC, Roden DM, Palmer C, Deloukas P, Lin DY, Tang ZZ, Erdmann J, Schunkert H, Danesh J, Marrugat J, Elosua R, Ardissino D, McPherson R, Watkins H, Reiner AP, Wilson JG, Altshuler D, Gibbs RA, Lander ES, Boerwinkle E, Gabriel S, Kathiresan S. Inactivating mutations in NPC1L1 and protection from coronary heart disease. The New England journal of medicine 2014;**371**(22):2072-82.

19. WTCC Consortium.. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;**447**(7145):661-78.

20. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Ardissino D, Merlini PA, Berzuini C, Bernardinelli L, Peyvandi F, Tubaro M, Celli P, Ferrario M, Fetiveau R, Marziliano N, Casari G, Galli M, Ribichini F, Rossi M, Bernardi F, Zonzin P, Piazza A, Mannucci PM, Schwartz SM, Siscovick DS, Yee J, Friedlander Y, Elosua R, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Kathiresan S, Meigs JB, Williams G, Nathan DM, MacRae CA, O'Donnell CJ, Salomaa V, Havulinna AS, Peltonen L, Melander O, Berglund G, Voight BF, Kathiresan S, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Musunuru K, Daly MJ, Purcell S, Voight BF, Purcell S, Nemesh J, Korn JM, McCarroll SA, Schwartz SM, Yee J, Kathiresan S, Lucas G, Subirana I, Elosua R, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burtt N, Gabriel SB, Samani NJ, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall A, Schunkert H, Erdmann J, Linsel-Nitschke P, Lieb W, Ziegler A, Konig I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Schunkert H, Samani NJ, Erdmann J, Ouwehand W, Hengstenberg C, Deloukas P, Scholz M, Cambien F, Reilly MP, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Knouff CW, Waterworth DM, Walker MC, Mooser V, Epstein SE, Rader DJ, Scheffold T, Berger K, Stoll M, Huge A, Girelli D, Martinelli N, Olivieri O, Corrocher R, Morgan T, Spertus JA, McKeown P, Patterson CC, Schunkert H, Erdmann E, Linsel-Nitschke P, Lieb W, Ziegler A, Konig IR, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Holm H, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Engert JC, Do R, Xie C, Anand S, Kathiresan S, Ardissino D, Mannucci PM, Siscovick D, O'Donnell CJ, Samani NJ, Melander O, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Altshuler D.

Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;**41**(3):334-41.
21. CARDIoGRAM Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, Konig IR, Cazier JB, Johansson A, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikainen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Consortium D, Consortium C, Doney AS, El Mokhtari N, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Muller-Nurasyid M, Mu TC, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schafer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, van der Schoot CE, Wagner PJ, Wellcome Trust Case Control C, Wells GA, Wild PS, Yang TP, Amouyel P, Arveiler D, Basart H, Boehnke M, Boerwinkle E, Brambilla P, Cambien F, Cupples AL, de Faire U, Dehghan A, Diemert P, Epstein SE, Evans A, Ferrario MM, Ferrieres J, Gauguier D, Go AS, Goodall AH, Gudnason V, Hazen SL, Holm H, Iribarren C, Jang Y, Kahonen M, Kee F, Kim HS, Klopp N, Koenig W, Kratzer W, Kuulasmaa K, Laakso M, Laaksonen R, Lee JY, Lind L, Ouwehand WH, Parish S, Park JE, Pedersen NL, Peters A, Quertermous T, Rader DJ, Salomaa V, Schadt E, Shah SH, Sinisalo J, Stark K, Stefansson K, Tregouet DA, Virtamo J, Wallentin L, Wareham N, Zimmermann ME, Nieminen MS, Hengstenberg C, Sandhu MS, Pastinen T, Syvanen AC, Hovingh GK, Dedoussis G, Franks PW, Lehtimaki T, Metspalu A, Zalloua PA, Siegbahn A, Schreiber S, Ripatti S, Blankenberg SS, Perola M, Clarke R, Boehm BO, O'Donnell C, Reilly MP, Marz W, Collins R, Kathiresan S, Hamsten A, Kooner JS, Thorsteinsdottir U, Danesh J, Palmer CN, Roberts R, Watkins H, Schunkert H, Samani NJ. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013;**45**(1):25-33.
22. 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**(3):e1001779.
23. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nature reviews Genetics 2013;**14**(7):507-15.
24. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang SJ, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikainen LP, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han BG, Huang J, Jalilzadeh S, Kessler T, Konig IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki ML, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon FU, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim BJ, Kooner JS, Kullo IJ, Lehtimaki T, Loos RJ, Melander O, Metspalu A, Marz W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015;**47**(10):1121-30.
25. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Magi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Hua Zhao J, Zhao W, Chen J, Fehrmann R, Hedman AK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Mateo Leach I, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stancakova A, Strawbridge RJ, Ju Sung Y, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Arnlov J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Bluher M, Bohringer S, Bonnycastle LL, Bottcher

Y, Boyd HA, Bruinenberg M, Caspersen IH, Ida Chen YD, Clarke R, Daw EW, de Craen AJ, Delgado G, Dimitriou M, Doney AS, Eklund N, Estrada K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe HJ, Grallert H, Grammer TB, Grassler J, Gronberg H, Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga JJ, James AL, Jeff JM, Johansson A, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindstrom J, Sin Lo K, Lobbens S, Lorbeer R, Lu Y, Mach F, Magnusson PK, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Muller G, Muller-Nurasyid M, Musk AW, Nagaraja R, Nothen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Vernon Smith A, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, Stringham HM, Sundstrom J, Swertz MA, Swift AJ, Syvanen AC, Tan ST, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh HW, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gadin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee JY, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JR, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka T, Van't Hooft FM, Vinkhuyzen AA, Westra HJ, Zheng W, Zondervan KT, Heath AC, Arveiler D, Bakker SJ, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Cupples LA, Cusi D, Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall M, Felix SB, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth G, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Illig T, Jacobs KB, Jarvelin MR, Jockel KH, Johansen B, Jousilahti P, Jukema JW, Jula AM, Kaprio J, Kastelein JJ, Keinanen-Kiukaanniemi SM, Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimaki T, Lyssenko V, Mannisto S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PA, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PE, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tonjes A, Tregouet DA, Tremblay A, Tremoli E, Virtamo J, Vohl MC, Volker U, Waeber G, Willemsen G, Witteman JC, Zillikens MC, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PI, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CN, Pedersen NL, Perola M, Perusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJ, Speliotes EK. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;**518**(7538):197-206.

26. TAG Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010;**42**(5):441-7.

27. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog HM, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkila K, Hypponen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J, Lyytikainen LP, Magnusson PK, Mangino M, Mihailov E, Montasser ME, Muller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney AS, Doring A, Elliott P, Epstein SE, Eyjolfsson GI, Gigante B, Goodarzi MO, Grallert H, Gravito ML, Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T, Jones MR, Kaleebu P, Kastelein JJ, Khaw KT,

Kim E, Klopp N, Komulainen P, Kumari M, Langenberg C, Lehtimaki T, Lin SY, Lindstrom J, Loos RJ, Mach F, McArdle WL, Meisinger C, Mitchell BD, Muller G, Nagaraja R, Narisu N, Nieminen TV, Nsubuga RN, Olafsson I, Ong KK, Palotie A, Papamarkou T, Pomilla C, Pouta A, Rader DJ, Reilly MP, Ridker PM, Rivadeneira F, Rudan I, Ruokonen A, Samani N, Scharnagl H, Seeley J, Silander K, Stancakova A, Stirrups K, Swift AJ, Tiret L, Uitterlinden AG, van Pelt LJ, Vedantam S, Wainwright N, Wijmenga C, Wild SH, Willemsen G, Wilsgaard T, Wilson JF, Young EH, Zhao JH, Adair LS, Arveiler D, Assimes TL, Bandinelli S, Bennett F, Bochud M, Boehm BO, Boomsma DI, Borecki IB, Bornstein SR, Bovet P, Burnier M, Campbell H, Chakravarti A, Chambers JC, Chen YD, Collins FS, Cooper RS, Danesh J, Dedoussis G, de Faire U, Feranil AB, Ferrieres J, Ferrucci L, Freimer NB, Gieger C, Groop LC, Gudnason V, Gyllensten U, Hamsten A, Harris TB, Hingorani A, Hirschhorn JN, Hofman A, Hovingh GK, Hsiung CA, Humphries SE, Hunt SC, Hveem K, Iribarren C, Jarvelin MR, Jula A, Kahonen M, Kaprio J, Kesaniemi A, Kivimaki M, Kooner JS, Koudstaal PJ, Krauss RM, Kuh D, Kuusisto J, Kyvik KO, Laakso M, Lakka TA, Lind L, Lindgren CM, Martin NG, Marz W, McCarthy MI, McKenzie CA, Meneton P, Metspalu A, Moilanen L, Morris AD, Munroe PB, Njolstad I, Pedersen NL, Power C, Pramstaller PP, Price JF, Psaty BM, Quertermous T, Rauramaa R, Saleheen D, Salomaa V, Sanghera DK, Saramies J, Schwarz PE, Sheu WH, Shuldiner AR, Siegbahn A, Spector TD, Stefansson K, Strachan DP, Tayo BO, Tremoli E, Tuomilehto J, Uusitupa M, van Duijn CM, Vollenweider P, Wallentin L, Wareham NJ, Whitfield JB, Wolffenbuttel BH, Ordovas JM, Boerwinkle E, Palmer CN, Thorsteinsdottir U, Chasman DI, Rotter JI, Franks PW, Ripatti S, Cupples LA, Sandhu MS, Rich SS, Boehnke M, Deloukas P, Kathiresan S, Mohlke KL, Ingelsson E, Abecasis GR. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;**45**(11):1274-83.

28. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;**37**(7):658-65.

29. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology 2015;**44**(2):512-25.

30. 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):304-14.

31. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, Fuster V, Boerwinkle E, Melander O, Orho-Melander M, Ridker PM, Kathiresan S. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. The New England journal of medicine 2016;**375**(24):2349-2358.

32. Davies NM, Hemani G, Timpson NJ, Windmeijer F, Davey Smith G. The role of common genetic variation in educational attainment and income: evidence from the National Child Development Study. Scientific reports 2015;**5**:16509.

33. Marioni RE, Yang J, Dykiert D, Mottus R, Campbell A, Davies G, Hayward C, Porteous DJ, Visscher PM, Deary IJ. Assessing the genetic overlap between BMI and cognitive function. Molecular psychiatry 2016;**21**(10):1477-82.

34. Trzaskowski M, Harlaar N, Arden R, Krapohl E, Rimfeld K, McMillan A, Dale PS, Plomin R. Genetic influence on family socioeconomic status and children's intelligence. Intelligence 2014;**42**(100):83-88.

35. Deary IJ, Strand S, Smith P, Fernandes C. Intelligence and educational achievement. Intelligence 2007;**35**(1):13-21.

36. Strenze T. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. Intelligence 2007;**35**(5):401-426.

37. Marioni RE, Ritchie SJ, Joshi PK, Hagenaars SP, Okbay A, Fischer K, Adams MJ, Hill WD, Davies G, Social Science Genetic Association C, Nagy R, Amador C, Lall K, Metspalu A, Liewald DC, Campbell A, Wilson JF, Hayward C, Esko T, Porteous DJ, Gale CR, Deary IJ. Genetic variants linked to education predict longevity. Proceedings of the National Academy of Sciences of the United States of America 2016;**113**(47):13366-13371.

38. Arden R, Luciano M, Deary IJ, Reynolds CA, Pedersen NL, Plassman BL, McGue M, Christensen K, Visscher PM. The association between intelligence and lifespan is mostly genetic. International journal of epidemiology 2016;**45**(1):178-85.

39. Schoon I. A Transgenerational Model of Status Attainment: the Potential Mediating Role of School Motivation and Education. National Institute Economic Review 2008;**205**(1):72-82.

40. Dubow EF, Boxer P, Huesmann LR. Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations. Merrill-Palmer quarterly (Wayne State University Press) 2009;**55**(3):224-249.

41. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schafer H. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. Eur J Hum Genet 2008;**16**(10):1164-72.

42. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, Bergeron A, Cutler D. The Association Between Income and Life Expectancy in the United States, 2001-2014. JAMA 2016;**315**(16):1750-66.

43. Hill AB. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? Proceedings of the Royal Society of Medicine 1965;**58**:295-300.
44. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. Journal of clinical epidemiology 2016;**69**:208-16.

Figure legends

Figure 1. 'Genetic education score' and risk of coronary artery disease.

A weighted genetic risk score (wGRS) based on 74 genetic variants affecting length of education reported by Okbay et al[14] ("genetic education score") was calculated in individuals from nine case-control studies for CAD (total n=27,551). The calculation of the score is described in Methods, and the description for these 9 studies is provided in Supplementary Text. Logistic regression was performed to evaluate the effect size of the wGRS on the risk of CAD in each study separately and afterwards meta-analysis was performed to combine the effects across nine studies. Forest plot displays a consistent inverse correlation across studies between the 'genetic education score' and odds of CAD.

Figure 2. Inverse relationship of genetically determined educational attainment and risk of coronary artery disease.

Individuals from each of the nine studies were grouped into quintiles based on their weighted genetic risk score for school attainment, with quintile 1 indicating the lowest genetic score and quintile 5 the highest. Odds ratios, shown with and 95% confidence intervals, for CAD were 20.8% lower in the quintile with the highest genetically determined educational attainment as compared to those with the lowest 'genetic

education score'. The distribution of all cases (red bars) with CAD is decreasing with an increasing 'genetic education score', while that of all controls (blue bars) has an opposite trend. P-value=7.66×10^{-9} was obtained from Cochran–Armitage trend test.

Figure 3. Interpretation of findings.

Educational attainment is partially modulated by common genetic variants[6-9, 14] and inversely associated with cardiovascular diseases[1, 2]. In the present study, we have found that the common genetic variants that influence length[8] of school education are also associated with cardiovascular risk factors and the manifestation of CAD. Importantly, the associations between genetic variants with CAD and its risk factors remained to be significant after adjustment for actual length of education. Thus, genetic factors may also influence decision making for a health-conscious lifestyle later in life and affect the prevalence of CAD through its risk factors.

Text tables

Table 1. Associations between the 'genetic education score' and cardiovascular conditions in UK Biobank.

	Odds ratio [95% CI]	p	realEduYear-adj
Coronary artery disease	0.96 [0.94, 0.97]	1.85E-06	no
	0.98 [0.96, 1.00]	0.014	yes
BMI	-0.15 [0.01] *****	1.66E-30	no
	-0.12 [0.01] *****	7.54E-20	yes
Smoking	0.95 [0.93, 0.96]	5.36E-23	no
	0.96 [0.95, 0.97]	7.68E-12	yes
Hypercholesterolemia	0.98 [0.96, 0.99]	3.42E-04	no
	0.99 [0.98, 1.01]	0.319	yes

Hypertension	0.97 [0.96, 0.98]	3.86E-08	no
	0.99 [0.98, 1.00]	0.007	yes
Type 2 diabetes	0.96 [0.93, 0.98]	1.73E-04	no
	0.98 [0.95, 1.00]	0.045	yes

Regression models of the 'genetic education score' on coronary disease risk and cardiovascular risk factors without (no - unadjusted) and with adjustment (yes) for years of school education completed (real EduYear-adjusted). All models were also adjusted for the first 5 principal components and genotyping array.

* Odds ratio [95% CI] is reported for each binary phenotype (except for BMI) as per one SD increase in the 'genetic education score'. The slope [SE] of the linear regression is reported for the continuous phenotype – BMI per one SD increase in the `genetic education score´.

Smoking: ever smokers vs. never smokers. BMI in kg/m^2. CI: confidence interval. SE: standard error.

Figures

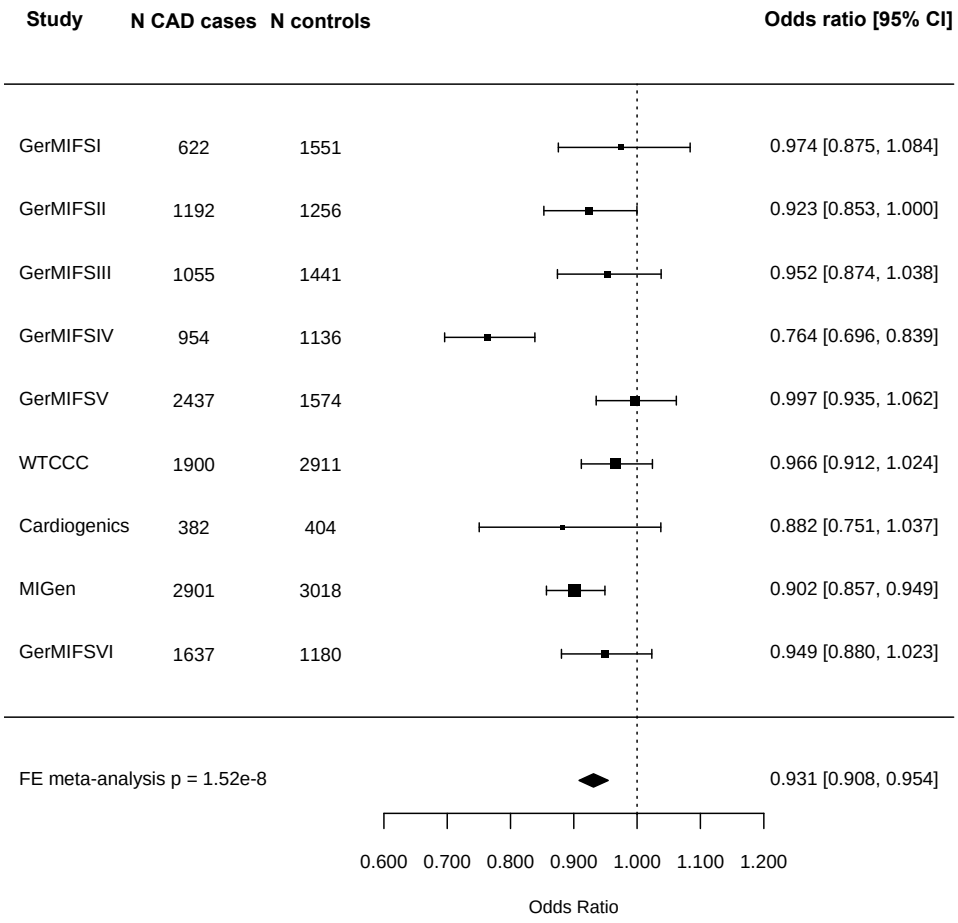


Figure 1. 'Genetic education score' and risk of coronary artery disease.

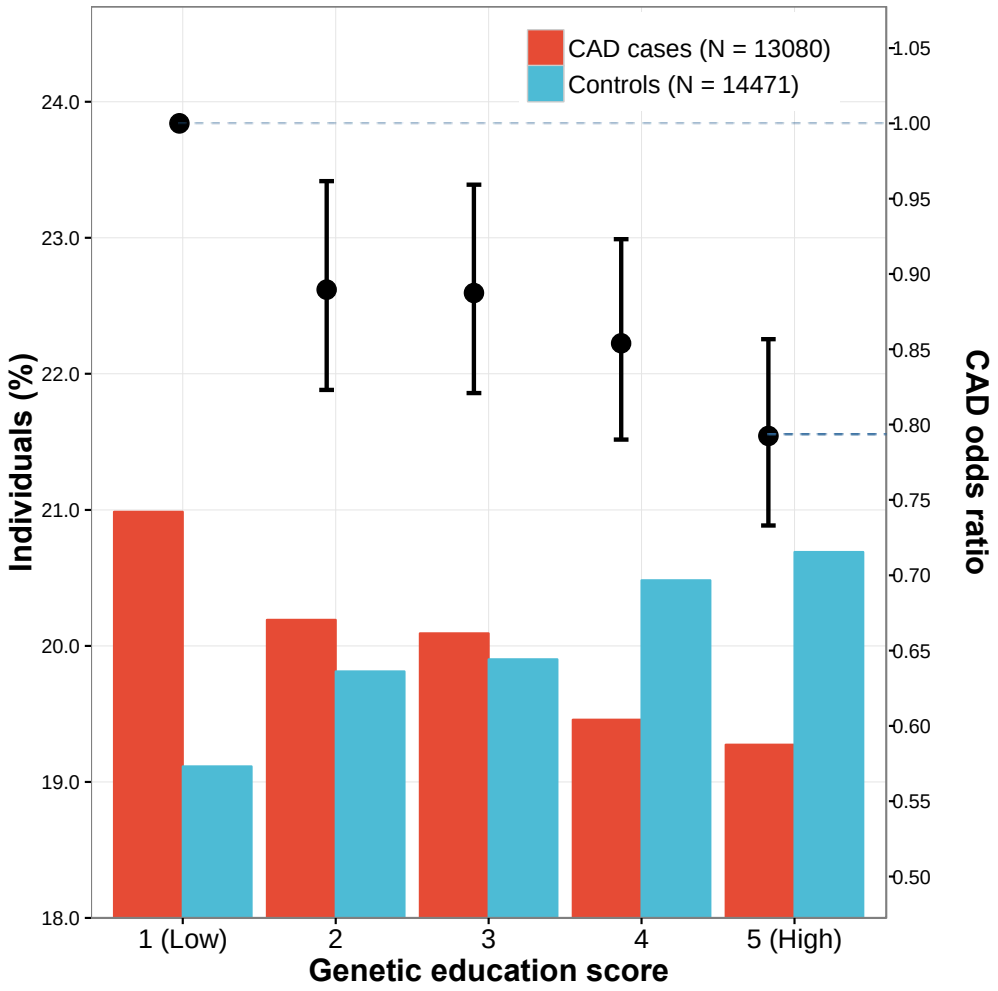


Figure 2. Inverse relationship of genetically determined educational attainment and risk of coronary artery disease.

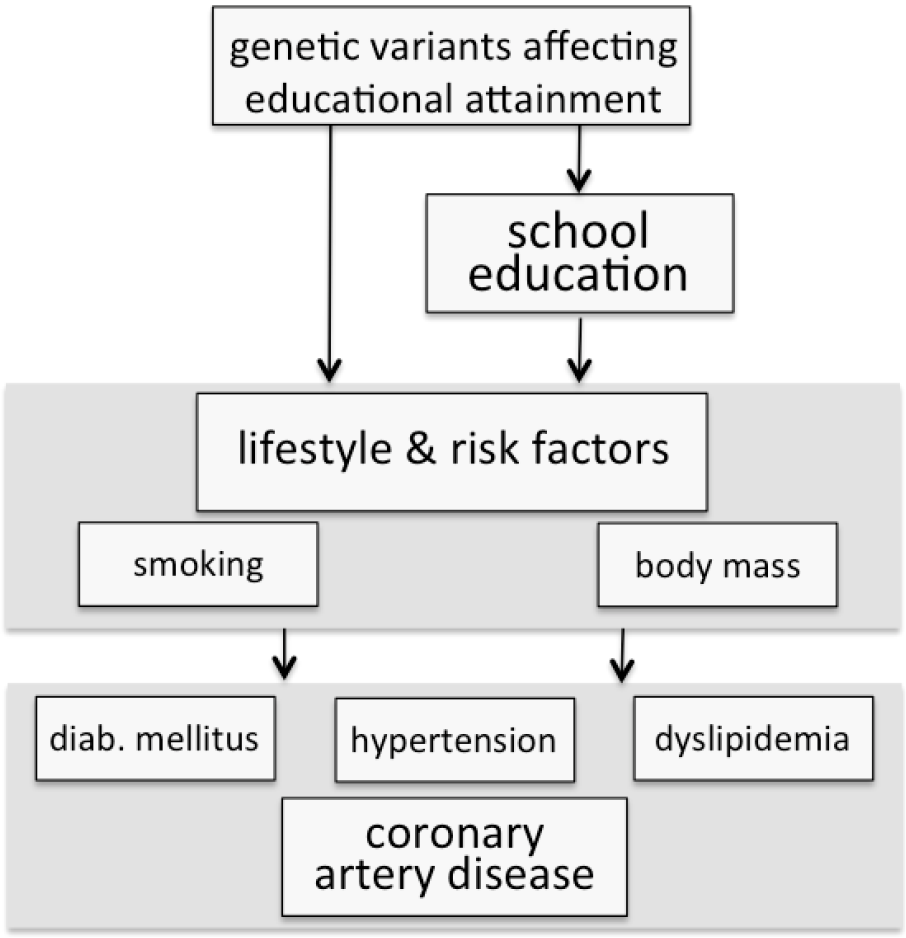


Figure 3. Interpretation of findings.

Supplementary Text

Genetically modulated educational attainment and coronary disease risk

Table of contents

1. Description of data sets and cohorts

1.1 CAD case-controls studies

1.2 UK-Biobank

1.3 Consortium for coronary risk factors

2. Phenotypic association analysis

2.1 Definition of phenotypes

2.2 phenotypic associations

2.3 'genetic CAD score' and EduYears

3. Processing of individual-level genotypes

3.1 Genotyping and imputation

3.2 Quality control

4. Selection of genetic variants

4.1 Genetic variants for educational attainment

4.2 Genetic variants for coronary artery disease

4.3 Genetic variants proxies for risk factors

Supplementary reference

1. Description of data sets and cohorts

1.1 CAD case-controls studies

Individual level data were obtained from nine case-control studies for coronary artery disease. Most individuals came from Germany: the German Myocardial Infarction Family Studies (GerMIFS) I[1], II[2], III[3] (KORA), IV[4], V[5], VI[6], and/or England: Wellcome Trust Case Control Consortium (WTCCC)[7] and/or France: Cardiogenics[8]. Others included subjects from Italy and the United States: Myocardial Infarction Genetics Consortium (MIGen)[9]. All subjects were of Western European descent and gave a broad written informed consent before participating in these studies on the understanding of genetic underpinnings of cardiovascular disease. All individuals provided informed consent that specifically addresses that the materials will be used for studying the effect of genetic variants on coronary risk. And all respective studies have obtained IRB approval from their local Ethical Committees. In the all German MI Family studies, Cardiogenics, WTCCC and MIGen, the information on CAD manifestation was validated by medical records. The detail assessment of CAD was given in the respective references.

Genome-wide genotype data and associated phenotype data for GerMIFS I-VI were collected by our group. Data for MIGen were obtained via the database of Genotypes And Phenotypes (dbgap)[10] (project ID #49717-3). Data for WTCCC and Cardiogenics were obtained via the Leducq network "CADgenomics" (https://www.fondationleducq.org/network/understanding-coronary-artery-disease-genes/). A summary of individual statistics is shown below and more detailed cohort descriptions could be sourced from the corresponding references.

	CAD cases		Controls	
Study	**N**	**Female N (%)**	**N**	**Female N (%)**
GerMIFSI	622	207 (33.2)	1551	795 (51.3)
GerMIFSII	1192	244 (20.5)	1256	604 (48.1)
GerMIFSIII	1055	212 (20.1)	1441	696 (48.3)
GerMIFSIV	954	336 (35.2)	1136	697 (61.4)
GerMIFSV	2437	593 (24.3)	1574	827 (52.5)
GerMIFSVI	1637	492 (30.1)	1180	607 (51.4)
Cardiogenics	382	49 (12.8)	404	239 (59.1)
WTCCC	1900	395 (20.8)	2911	1481 (50.9)
MIGen	2901	646 (22.3)	3018	733 (24.3)

1.2 UK-Biobank

UK Biobank (UKBB) was established to improve understanding of the causes of common diseases including CAD (www.**ukbiobank**.ac.**uk**/) and completed the recruitment of 502,713 (94% of self-reported European ancestry) individuals aged 40-69 years across England, Scotland and Wales between 2005 and 2010 (94% of self-reported European ancestry). UKBB adopted the current worldwide practice of consenting subjects using a 'broad informed consent' from (http://www.ukbiobank.ac.uk/wp-content/uploads/2011/06/Consent_form.pdf) to cater both for technological advances as well as the ability to maximise the use of data i.e. by addressing a broad spectrum of scientific questions. This consent covers the present research question. UKBB covers the thousands of phenotypes collected. In addition to self-reported disease outcomes as well as extensive health and life-style questionnaire data, UKBB participants are being tracked through their NHS records and national registries (including cause of death and Hospital Episode Statistics [HES]). In July 2015, UKBB released genotype data imputed to the 1000 Genomes panel for 152,249 participants profiled with a SNP array harboring 820,967 variants comprising a backbone of common variants optimized for imputation, a validated subset of rare coding variants from the Exomechip array (http://genome .sph.umich.edu/wiki/Exome_Chip_Design) and a set of likely functional variants or their proxies

(e.g. GWAS catalogue). Analyses were performed on 146,514 participants of European ancestry after standard quality control.

During the revision process of this manuscript, a latest version of UKBB data were released which increased the sample size from 150k to 500k. We utilized this larger set of data in the sensitivity analysis, including 455,440 participants of European ancestry after standard quality control.

1.3 Consortia for coronary risk factors

The CARDIoGRAMplusC4D consortium[4] reported summary statistics of genome-wide association studies for coronary artery disease (CAD) for participants of multiple European states, the US, Iceland, Australia as well as some other countries. The ethnic origin of the vast majority of individuals in this consortium is European. Summary statistics for other CAD risk factors were obtained from other consortia, including GLGC (LDL-c, HDL-c, triglyceride, total cholesterol level), GIANT [11] (body-mass index), and TAG (smoking behaviour) [12]. A description of consortium size and the corresponding phenotypes is shown here.

Trait name	Consortium	Sample size	PMID	Publish year	Website for Consortium or Data
Coronary artery disease	CARDIoGRAM-plusC4D	184,305	26343387	2015	http://www.cardiogramplusc4d.org
LDL cholesterol	GLGC	188,577	24097068	2013	http://csg.sph.umich.edu/abecasis/public/lipids2013/
HDL cholesterol					
Triglycerides					
Total cholesterol					

Body mass index	GIANT	339,224	25673413	2015	http://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page
Ever vs never smoked	TAG	74,053	20418890	2010	https://www.med.unc.edu/pgc/results-and-downloads
Cigarettes smoked per day					

2. Phenotypic association analysis in UK Biobank

2.1 Definition of phenotypes

All conditions were defined by either self-reported, hospital episode and/or death registry data. In UK Biobank self-reported data, CAD cases were defined if they reported having 'vascular/heart problems diagnosed by doctor' or 'non-cancer illnesses' as angina or heart attack. Self-reported operation included PTCA, or coronary artery bypass grafting (CABG). In hospital episode data and death registry data including primary and secondary diagnoses and operations, MI was defined as hospital admission or cause of death due to ICD9 410–412, ICD10 I21–I24, I25.2; PTCA was defined as hospital admission for PTCA (OPCS-4 K49, K50.1, K75); CABG was defined as hospital admission for CABG (OPCS-4 K40–K46); and angina or chronic IHD was defined as hospital admission or death due to ICD9 413, 414.0, 414.8, 414.9, ICD10 I20, I25.1, I25.5–I25.9. Exclusions were made for aneurysm and atherosclerotic cardiovascular disease using hospital admissions or cause of death codes ICD9 414.1, ICD 10 I25.0, I25.3, I25.4 (and not having MI, PTCA, CABG, Angina or chronic IHD as defined above). Controls were defined as patients who were not a CAD case after exclusions.

Individuals were classified as hypertensives if their mean SBP ≥140 mm Hg or mean DBP ≥90 mm Hg (from the two blood pressure measurements from initial assessment), or if they were reported

taking blood pressure lowering medication; otherwise, they were classified as non-hypertensive. In detail, ICD 10 and ICD 9 codes used to define cases of CAD, type 2 diabetes, stroke, peripheral arterial disease, and chronic obstructive pulmonary disease, and lung cancer are given in the table below. For each disease, individuals who were not identified as disease cases were defined as controls. Smoking status was defined using self-reported data – current or ex-smokers were defined as ever smokers (cases), while individuals who reported no smoking were classified as non-ever smokers (controls). BMI was taken from initial assessment data only.

Hypersholesterolemia was defined using self-reported data, i.e, individuals who reported either having high cholesterol or taking cholesterol lowering medication were defined as cases. Individuals with familial hypercholesterolemia (codes ICD10 E78.0, ICD9 272.0) were excluded. Controls were defined as individuals who were not a hypercholesterolemia case after exclusions.

Phenotypically, educational attainment was measured as length of school education (EduYears) using the same methods that were used in the report that led to the identification of the EduYear-SNPs[13], that is, a continuous variable measuring the number of years of schooling completed. Briefly, using the mapping shown in Supplementary Table 1.2. of Okbay et al. *years of schooling / years-of-education equivalent for each ISCED category* were calculated as: pre-primary education 1/0; primary education or first stage of basic education 7/1; lower secondary or second stage of basic education 10/2; (upper) secondary education 13/3; post-secondary non-tertiary education 15/4; first stage of tertiary education (not leading directly to an advanced research qualification) 19/5; second stage of tertiary education (leading to an advanced research qualification, e.g. a Ph.D.) 22/6.

Disease	**ICD 10 codes**	**ICD 9 codes**	**Operation codes**

Coronary artery disease	I20 - 25 (excluding I25.0, I25.3, I25.4)	410 - 414 (excluding 414.1)	K40 – K46, K49, K50.1, K75
Type 2 diabetes	E11	25000, 25010	-
Stroke	I60, I61, I63	430, 431, 435	-
Peripheral arterial disease	I73.9	443.9	-
Chronic obstructive pulmonary disease	J44	491.2, 496.9	-
Chronic obstructive pulmonary disease – extended	J43.9, J40, J44, J45.9, J46, J47	430, 431, 435, 491.2, 496	-
Lung cancer	C34	162	-

2.2 phenotypic association

Regression analyses were performed to evaluate the phenotypic association between EduYears and risk of CAD and its risk factors (in section 2.1). Prior analysis real years of school education were standardized into Z-scores (centred and scaled to have a mean of 0 and standard deviation (SD) of 1).

2.3 'genetic CAD score' and EduYears

Based on the 50 CAD SNPs reported by Khera et al[24] (Extended Data Table 2) a weighted genetic risk score was calculated to evaluate the cumulative genetic effect of CAD risk on EduYears. Genetic risk score for CAD was estimated in the same way as the EduYear-wGRS. We gave a value from minimum 0 to maximum 2 for every SNP for every individual according to the sum of the posterior probabilities from the imputation files to indicate the number of CAD-increasing alleles and multiplied the number of alleles by

the reported effect sizes. Then we totalled these values for each individual across all 50 SNPs to generate a weighted genetic risk score of CAD.

Linear regression was performed to evaluate the effect size of the weighted CAD genetic risk score on real EduYears. The genetic risk score was modelled as a continuous variable and standardized into Z-scores (centred and scaled to have a mean of 0 and standard deviation (SD) of 1). The top five principle components based on autosomal genotypes and the genotyping array were included in the regression model as covariates.

3. Processing of individual-level genotypes

3.1 Genotyping and imputation

Genotyping was performed using a range of common, commercially available genotyping arrays. The 1000 Genomes Phase I integrated variant (v3) set released in NCBI build 37 (hg19) coordinates with reference data from March 2012 (updated August 2012) was utilized as the reference panel for imputation in all nine case-control studies of coronary disease.

UK Biobank released genotypes imputed to the 1000 Genomes panel for 152,249 participants profiled with a SNP array harboring 820,967 variants comprising common variants optimized for imputation, validated rare coding variants and sets of phenotype-associated variants or their proxies (e.g. GWAS catalogue).

3.2 Quality control

The following pre-imputation QC criteria were taken in all nine case-control studies of coronary disease: individual call rate >= 0.98, SNP call rate > 0.98, minor allele frequency (MAF) > 0.01, concordant recorded and genotype-derived gender, population outliers excluded (deviate beyond mean ± 5*SD for top two principle components), IBD PI_HAT < 0.125 (individuals distant away than third-degree relatives), heterozygosity rate within mean ± 3*SD, and deviation from Hardy-Weinberg Equilibrium(HWE) p > 1e-6. As an essential statistical technique to estimate genotypes that were not directly assayed, and in order to harmonize our cohort data collected from different genotyping array platforms we imputed the data prior to the meta-analysis. After genotype QC, haplotypes were then pre-phased from genotypes with SHAPEIT2 haplotype estimation tool to generate the best guess haplotypes based on the given genotypes. Then the best guess haplotypes were forwarded to IMPUTE2 for imputation. For each bi-allelic variant [A/B] for each individual, the main output of IMPUTE2 reported the three genotypes AA, AB and BB in the form of their probabilities accounting for the genotype imputation uncertainty, instead of giving an fixed designation. Finally the following post-imputation QC criteria were taken: SNP call rate > 0.98, MAF > 0.05, Hardy-Weinberg p > 1e-5.

4. Selection of genetic variants

4.1 Genetic variants for educational attainment

A list of 74 lead SNPs of each genetic locus for educational attainment (EduYear-SNPs) was obtained from Okbay et al. of the latest GWAS publication for educational attainment[13], where the authors have generated a list of 74 variants, and their effect-alleles, frequencies, and corresponding effect sizes and p-values from meta-analyses analyses. For each SNP two alleles were extracted

from the 1000G Reference Genome. Then for all SNPs the reported effects were aligned to the EduYear-increasing alleles. The detailed list is compiled and shown in Extended Data Table 1.

4.2 Genetic variants for coronary artery disease

A list of 50 lead SNPs of each genetic locus for CAD risk was compiled from Khera et al.[14] where the authors have used to compose the genetic risk score of CAD. Effect-alleles, frequencies, and corresponding effect sizes were obtained from Supplementary Table 1 of Khera et al. The detailed list is compiled and shown in Extended Data Table 2. For each SNP two alleles were extracted from the 1000G Reference Genome. Then for all SNPs the reported effects were aligned to the CAD risk increasing alleles.

4.3 Genetic variants proxies for risk factors

In Mendelian Randomization analysis summary statistics were extracted for CAD and each trait respectively for 74 EduYear-SNPs in the corresponding genome-wide association studies. However, most of these GWAS meta-analysis were originally performed based on HapMap-imputed genotypes, thus not for all of these EduYear-SNPs summary statistics for various traits were reported. Therefore, for SNPs not available in the reported summary statistics we identified SNP proxies with an LD $r^2 > 0.5$ in the reference European genomes (both 1000G pilot 1 and Phase I v3). Extended Data Table 3 is showing the full SNP availability in each reported trait and their proxies used in the analyses.

Supplementary reference

1 Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* **357**, 443-453, doi:10.1056/NEJMoa072366 (2007).
2 Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* **41**, 280-282, doi:10.1038/ng.307 (2009).
3 Erdmann, J. *et al.* Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur Heart J* **32**, 158-168, doi:10.1093/eurheartj/ehq405 (2011).
4 Consortium, C. D. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130, doi:10.1038/ng.3396 (2015).
5 Stitziel, N. O. *et al.* Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med* **371**, 2072-2082, doi:10.1056/NEJMoa1405386 (2014).
6 Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature genetics*, doi:10.1038/ng.3913 (2017).
7 Consortium., W. T. C. C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678, doi:10.1038/nature05911 (2007).
8 Consortium, C. A. D. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25-33, doi:10.1038/ng.2480 (2013).
9 Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-341, doi:10.1038/ng.327 (2009).
10 Tryka, K. A. *et al.* NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res* **42**, D975-979, doi:10.1093/nar/gkt1211 (2014).
11 Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206, doi:10.1038/nature14177 (2015).
12 Consortium, T. a. G. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-447, doi:10.1038/ng.571 (2010).
13 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542, doi:10.1038/nature17671 (2016).
14 Khera, A. V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med* **375**, 2349-2358, doi:10.1056/NEJMoa1605086 (2016).

Extended Data Figure and Legends

Genetically modulated educational attainment and coronary disease risk

A.

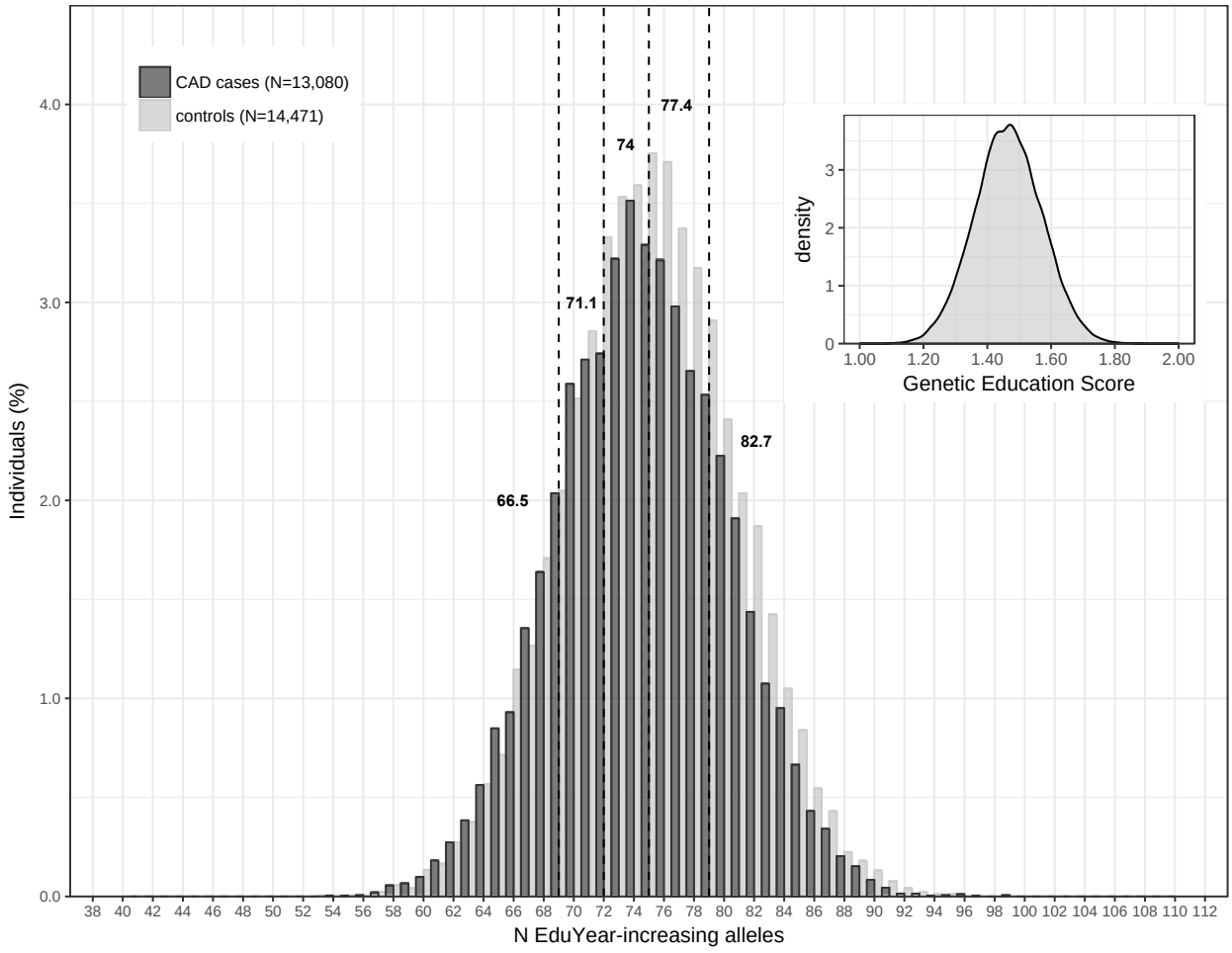
Question: whether?	Approach				Data
		Outcome	Exposure	Counfounders	
Is the genetic component of educational attainment associated with CAD?	genetic risk score and regression analysis	CAD	'genetic education score'		GerMIFSI GerMIFSII GerMIFSIII GerMIFSIV GerMIFSV GerMIFSVI WTCCC Cardiogenics MIGen UKBB
... and even partly independent of the actual 'years of school education'?				'years of school education'	UKBB
Is the genetic component of educational attainment associated with specific lifestyle & risk factors of CAD?	genetic risk score and regression analysis	lifestyle & risk factors			
... and even partly independent of the actual 'years of school education'?				'years of school education'	

B.

Question: how?	Approach				Data
		Outcome	Exposure	Counfounders	
Is the impact of the genetic component of educational attainment on CAD risk due to the intermediate impact on specific lifestyle & risk factors? - Association	genetic risk score and regression analysis	CAD	'genetic education score'	'years of school education' + possible intermediate factors	UKBB
Is the impact of the genetic component of educational attainment on CAD risk due to the intermediate impact on specific lifestyle & risk factors? - Causality	Mendelian randomization	Outcome	Exposure	Genetic Instrumental variables	Summary statistics obtained from various large GWAS consortia
		CAD	EduYear	EduYear-SNPs	
		lifestyle & risk factors			
		CAD		EduYear-SNPs excluding those associated with CAD and other causal risk factors	

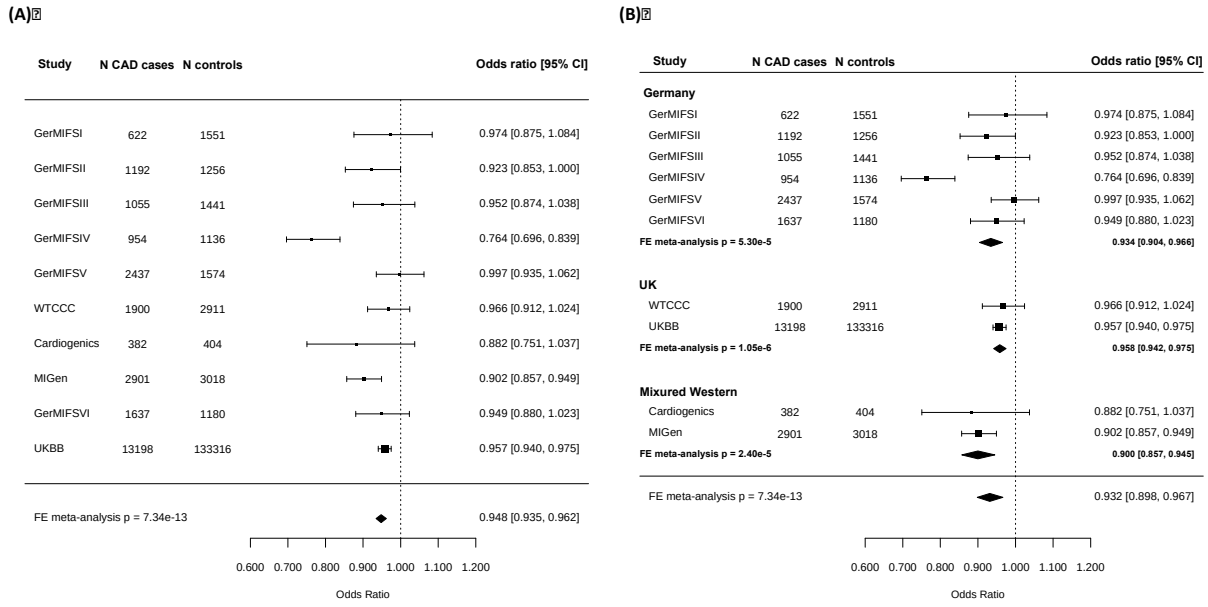
Extended Data Figure 1. Study design. (A). We studied the genetic component underlying the relationship between higher educational attainment and lower CAD risk (blue block) in multiple cohort data (GerMIFS I-VI, WTCCC, Cardiogenics, MIGen at discovery stage and UK Biobank (UKBB) at replication stage) and associated the 'genetic education score' with the susceptibility of CAD and other traits, which allowed to estimate the respective effect sizes. As the 'genetic education score' certainly affects the (exogenous) phenotype 'years of school education' (real EduYears) further adjustment for this phenotype (which is only available in the UKBB cohort) was needed to examine effects of the 'genetic education score' that are not directly mediated by school education. **(B).** With the aim to further characterize the interplay between the genetic basis of educational attainment, the prevalence of CAD and its related lifestyle risk factors (green

block), we performed similar regression analysis with additional inclusion of possible intermediate lifestyle and risk factors, to check the (inter)dependence of the 'genetic education score', lifestyle risk factors and CAD risk. Furthermore, we employed genetic Mendelian randomization approach to investigate the causality of educational attainment on CAD and/or its related risk factors.

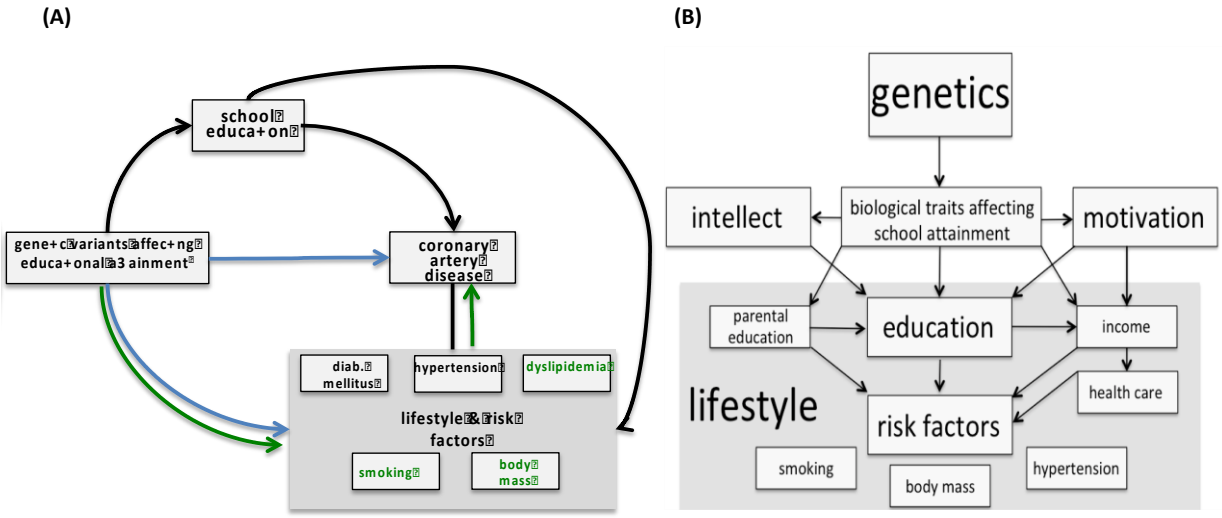


Extended Data Figure 2. Distribution of genetic risk score of EduYear in participants from nine case-controls studies for coronary artery disease.

Gaussian distribution showing the number of EduYear-increasing alleles in a group of 13,080 cases with coronary disease and 14,471 controls from nine large case-control studies. Individuals were divided in quintiles (dashed lines) in each study based on their 'genetic education score'. The number above each quintile represents the average number of EduYear-increasing alleles in each quintile.



Extended Data Figure 3. Forest plot for meta-analysis including all studies. In addition to the nine studies from which we have initially identified the inverse association between the 'genetic education score' and risk of CAD, a consistent inverse correlation with similar logistic regression was also found in UK Biobank for replication. Afterwards a meta-analysis was performed to combine the effects across all ten studies including UK Biobank. (A) The same effect is replicated and obtained with an even more significant fixed-effect meta-analysis p-value ($p=7.34 \times 10^{-13}$ compared to $p=1.52 \times 10^{-8}$ in the discovery studies). (B) We also categorized all studies according to countries, and observed directionally identical effects ($p=1.05 \times 10^{-6}$ for the UK; $p=5.30 \cdot \times 10^{-5}$ for Germany; $p=2.40 \times 10^{-5}$ for the mixture of several Western countries).

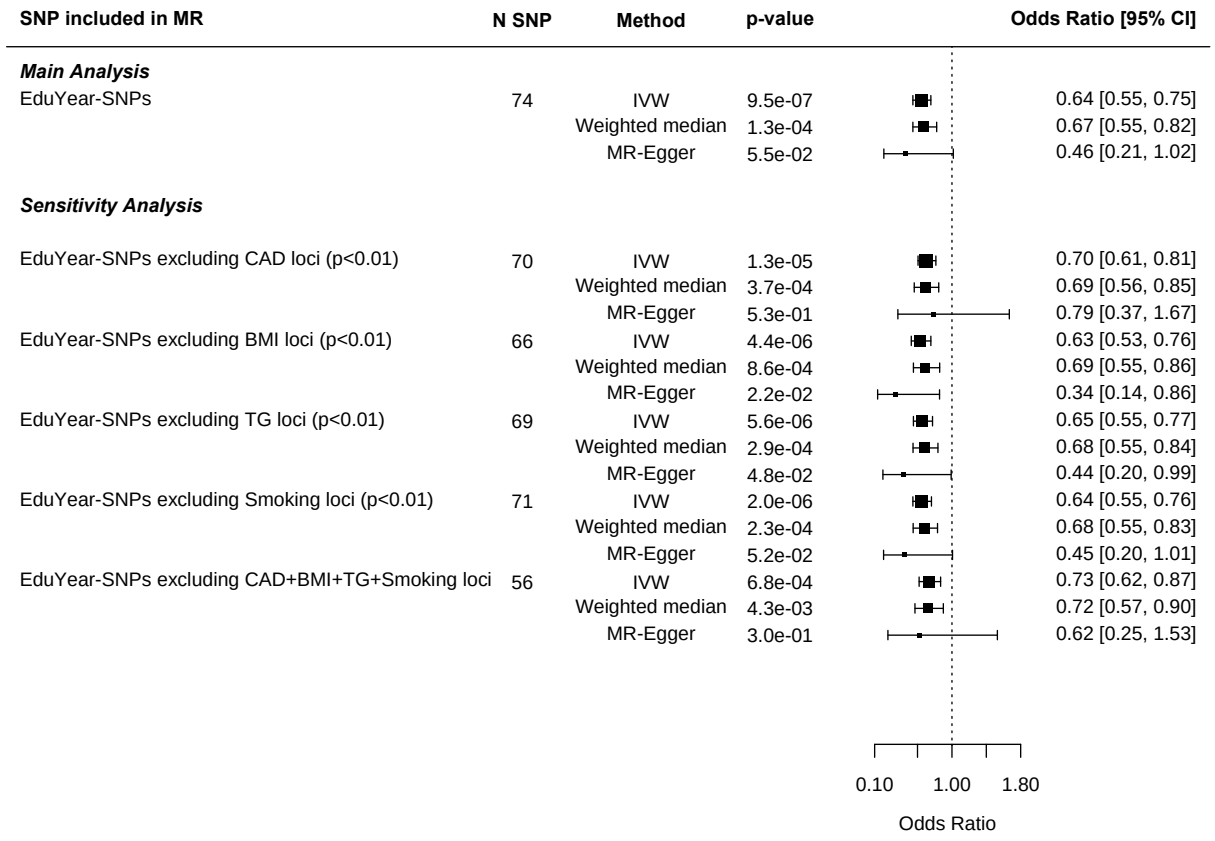


Extended Data Figure 4. Interpretation of findings.

A. It has been known that educational attainment is partially modulated by common genetic variants and inversely associated with cardiovascular diseases (black lines). In the present study, we have firstly found that the common genetic variants that influence length of school education are associated with the manifestation of CAD (straight blue line in the middle). Further analyses indicated the common genetic variants that influence length of school education are in parallel associated with specific lifestyle and risk factors (blue line in the lower part) (e.g. BMI, smoking status, and triglycerides). Importantly, these associations (blue lines) are partly independent on the actual length of education (black lines). Finally, several lifestyle related factors were deduced as intermediate factors that mediate the genetic component increasing in CAD risk (green elements and lines). In summary, genetic factors may also influence decision making for a health-conscious lifestyle later in life and affect thereby the prevalence of CAD through its risk factors.

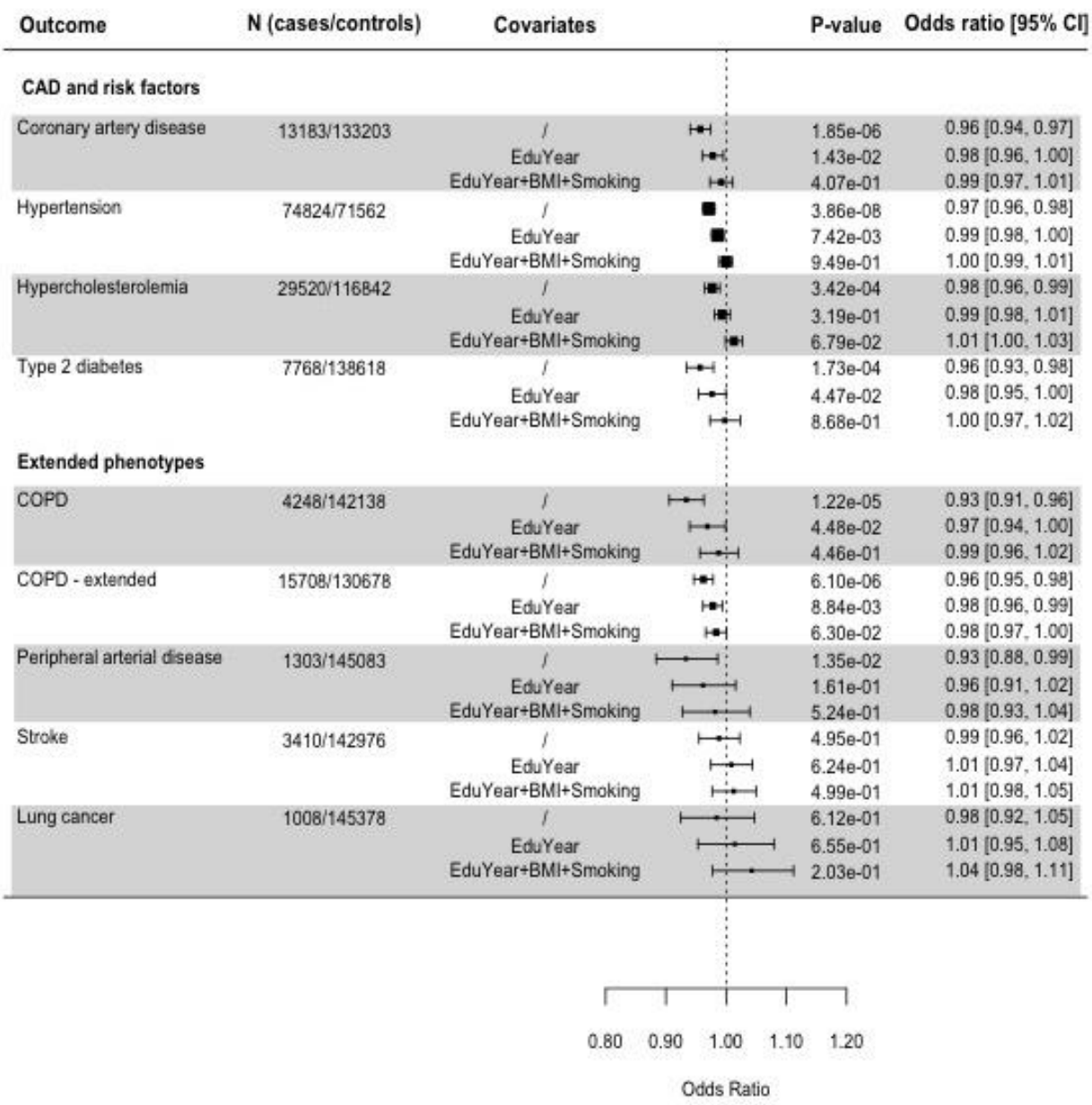
B. We are well aware of that despite of our observation that the signal between the ‘genetic education score’ and CAD risk decreases when smoking and BMI are included as co-variables, we can only demonstrate plausible deductions (the green elements and lines in Panel A) instead of causality, due to the fact that the ‘genetic education score’ may have broader implications. It is likely that a number of confounders such as intelligence, behaviour patterns, socioeconomic factors in respective individuals as

well as the parental generation are involved in the interplay of educational attainment, CAD factors and prevalence of CAD. Here in Panel B we illustrate these potential and plausible links.



Extended Data Figure 5. Genetic causality of educational attainment and coronary artery disease as well as its risk factors

Mendelian randomization (MR) analysis was performed to investigate the genetic causality of educational attainment and coronary artery disease (CAD) as well as its risk factors. Several MR methods were employed to combine individual-SNP beta estimates retrieved from various genome-wide meta-analyses (details described in Methods and Supplementary Text). For sensitivity analyses we excluded from the 74 EduYear-SNPs which had marginal associations ($p<0.01$ in the published summary statistics) with CAD or an examined risk factor (Extended Data Table 3). A 1-SD increase in education years and 28% lower risk of CAD (odds ratio for weighted median method of 0.72, 95% CI 0.57, 0.90, $p=4.3\times10^{-3}$) was observed even after all confounder-influential SNPs have been excluded.



Extended Data Figure 6. Association between the 'genetic education score' and extended phenotypes in UK Biobank.

Regression models of the 'genetic education score' on extended phenotypes with adjustment for real years of school education completed (EduYears), BMI or smoking. In addition to the covariates shown here, all

models were also adjusted for the first 5 principal components based on autosomal genotypes plus genotype array platform.

Word count

Manuscripts (including legend for main displaying items) : 4876

References: 5725 (several Consortium papers have very long author list)

The authors do hereby declare that all illustrations and figures in the manuscript are entirely original and do not require reprint permission.

Abstract

Aims: Genetic disposition and lifestyle factors are understood as independent components underlying the risk of multiple diseases. In this study, we aim to investigate the interplay between genetics, educational attainment - an important denominator of lifestyle - and coronary artery disease (CAD) risk.

Methods and Results: Based on the effect sizes of 74 genetic variants associated with educational attainment, we calculated a 'genetic education score' in 13,080 cases and 14,471 controls and observed an inverse correlation between the score and risk of CAD (p=1.52x10^{-8}; odds ratio [OR] 0.79 (95% confidence interval [CI] 0.73-0.85) for the higher compared to the lowest score quintile). We replicated in 146,514 individuals from UK Biobank (p=1.85x10^{-6}) and also found strong associations between the 'genetic education score' with 'modifiable' risk factors including smoking (p=5.36x10^{-23}), body mass index (p=1.66x10^{-30}), and hypertension (p=3.86x10^{-8}). Interestingly, these associations were only modestly attenuated by adjustment for years spent in school. By contrast, a model adjusting for BMI and smoking abolished the association signal between the 'genetic education score' and CAD risk suggesting an intermediary role of these two risk factors. Mendelian randomization analyses performed with summary statistics from large genome-wide meta-analyses and sensitivity analysis using 1271 variants affecting educational attainment (OR 0.68 for the higher compared to the lowest score quintile; 95% CI 0.63-0.74; p=3.99x10^{-21}) further strengthened these findings.

Conclusion: Genetic variants known to affect educational attainment may have implications for a health-conscious lifestyle later in life and subsequently affect the risk of coronary artery disease.

Keywords: atherosclerosis / school education / coronary artery disease / genome-wide association studies / genetics

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
coi_Zeng.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
coi_INtalla.pdf

Click here to access/download

ICMJE Conflicts of Interest form (1 for each author listed)

coi_TK.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
coi_Erdmann.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
COI_Kastrati.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
coi_njs.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
coi_deloukas.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
coi_disclosure_HS.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
COI_Danesh.pdf

Click here to access/download
ICMJE Conflicts of Interest form (1 for each author listed)
coi_Watkins.pdf

Extended Data Tables

Click here to access/download

Supplemental files not built into the pdf

R3_ExtendedDataTables_20181022.xlsx