CrossMark

ARTICLE

Genetic support for the causal role of insulin in coronary heart disease

Emmi Tikkanen[1,2] · Matti Pirinen[2] · Antti-Pekka Sarin[2] · Aki S. Havulinna[3] ·
Satu Männistö[3] · Juha Saltevo[4] · Marja-Liisa Lokki[5] · Juha Sinisalo[6] ·
Annamari Lundqvist[3] · Antti Jula[3] · Veikko Salomaa[3] · Samuli Ripatti[1,2,7]

Abstract

Aims/hypothesis Epidemiological studies have identified several traits associated with CHD, but few of these have been shown to be causal risk factors and thus suitable targets for treatment. Our aim was to evaluate the causal role of a large set of known CHD risk factors using single-nucleotide polymorphisms (SNPs) as instrumental variables.

Methods Based on published genome-wide association studies (GWASs), we estimated the associations between the established risk factors (blood lipids, obesity, glycaemic traits and BP) and CHD with two complementary approaches: (1) using summary statistics from GWASs to analyse the accordance of SNP effects on risk factors and on CHD; and (2) individual-level analysis where we constructed genetic risk scores (GRSs) in a large Finnish dataset ($N=26,554$, CHD events $n=4016$).

We used a weighted regression-based method for summary-level data to evaluate the causality of risk factors. The associations between the GRSs and CHD in the Finnish dataset were evaluated with logistic and conditional logistic regression models.

Results The summary-level data analysis revealed causal effects between glycaemic traits (insulin and glucose) and CHD. The individual-level data analysis supported the causal role of insulin, but not of glucose, on CHD. The GRS for insulin was associated with CHD in the Finnish cohort (OR 1.06 per SD in GRS, 95% CI 1.02, 1.10, $p=0.002$).

Conclusions/interpretation These results support the causal role of insulin in the pathogenesis of CHD. Efficient treatment and prevention of insulin resistance is essential to prevent future CHD events.

Keywords Epidemiology · Genetics · Insulin resistance · Insulin sensitivity

Electronic supplementary material The online version of this article (doi:10.1007/s00125-016-4081-6) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Emmi Tikkanen
emmi.tikkanen@helsinki.fi

✉ Samuli Ripatti
samuli.ripatti@fimm.fi

1 Department of Public Health, University of Helsinki, PO Box 20, Tukholmankatu 8 B, FIN-00014 Helsinki, Finland

2 Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland

3 Department of Health, National Institute for Health and Welfare, Helsinki, Finland

4 Department of Medicine, Central Finland Central Hospital, Jyväskylä, Finland

5 Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland

6 Heart and Lung Center, Helsinki University Central Hospital, Helsinki, Finland

7 Wellcome Trust Sanger Institute, Hinxton, UK

Abbreviations

2SLS	Two-stage least-squares
β_{Ins}	$\beta_{Insulin}$
β_{Glu}	$\beta_{Glucose}$
CVD	Cardiovascular disease
DBP	Diastolic BP
DILGOM	DIetary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome
GRS	Genetic risk score
GWAS	Genome-wide association study
ISI	Insulin sensitivity index

MAF	Minor allele frequency
MR	Mendelian randomisation
RCT	Randomised controlled trial
SBP	Systolic BP
SNP	Single-nucleotide polymorphism

Introduction

Prospective cohort studies have identified several risk factors for CHD, including high LDL- and low HDL-cholesterol, obesity, glycaemic traits and high BP. The slow progress of the disease and the potential for reverse causation and confounding make it difficult to infer the causal relationships between the risk factors and CHD using traditional epidemiological methods or intervention studies.

Randomised controlled trials (RCTs) are usually considered a gold-standard method to establish causality. For example, RCTs have shown that lowering LDL-cholesterol levels with statin treatment reduces cardiovascular events [1–3], which confirms the causal role of LDL-cholesterol in CHD. In contrast, even though low levels of HDL-cholesterol are associated with increased CHD risk, treatment that elevates HDL-cholesterol levels has not been successful in CHD prevention [4]. The causal role of many other CHD risk factors, such as glycaemic traits and obesity, still needs to be determined.

Genetic data can be used to infer causal relationships between risk factors and disease in a more cost-effective way compared with traditional RCTs. As genotypes are determined at conception, they are unaffected by reverse causation and confounding and, thus, genetic data can be used to reveal causal and mechanistic insights leading to CHD. Genome-wide association studies (GWASs) have uncovered genetic variants robustly associated with various phenotypes. There is a growing number of studies demonstrating how the genetic data can be used to assess causal relationships between risk factors and disease. For example, studies using summary-level data from GWASs have supported the causal role of blood triacylglycerols, but not HDL-cholesterol, on CHD [5, 6].

In this study, our objective was to simultaneously evaluate the causal role of several CHD risk factor traits by using genome-wide genetic data from large-scale GWASs [7–12]. We studied associations between the established risk factors (blood lipids, obesity, glycaemic traits and BP) and CHD with two complementary approaches: (1) using summary statistics from GWASs to analyse the accordance of SNP effects on risk factors and on CHD; and (2) individual-level analysis where we constructed genetic risk scores (GRSs) in a large Finnish dataset ($N = 26,554$, CHD events $n = 4016$).

Methods

Genetic marker selection

Single-nucleotide polymorphisms (SNPs) can be used as instrumental variables in Mendelian randomisation (MR) analysis to assess causality of risk factors on disease. The valid instrument should be associated with the outcome only through the risk factor of interest, i.e. it should not have pleiotropic effects through other risk factors. Weak intruments and pleiotropic effects are common challenges of genetic instruments. Inclusion of a high number of SNPs increases the power of the instrument, but it also increases the risk of pleiotropy. On the other hand, excluding SNPs based on observed associations with other traits to minimise pleiotropic effects could lead to weak intrument bias. Mechanistic SNP exclusions based on pre-defined statistical thresholds might also exclude SNPs that are not pleiotropic, but have effects for other traits on the same causal pathway than the risk factor of interest (Fig. 1), which does not violate the causal inference. In this study, rather than excluding SNPs to decrease the risk for pleiotropy, we applied methods that take pleiotropic effects into account at the modelling phase [5, 13, 14].

For our analyses, we selected all SNPs associated with established CHD risk factors and CHD at genome-wide significance level from the largest GWASs for each trait (for lipids $n = 188,577$, for BMI $n = 339,224$, for WHR $n = 224,459$, for glycaemic traits $n = 133,010$, for BP $n = 200,000$, for CHD $n = 184,305$) [7–12] (electronic supplementary material [ESM] Table 1). For each SNP, we obtained effect estimates for LDL- and HDL-cholesterol, triacylglycerols, BMI, WHR, fasting insulin and glucose, systolic and diastolic BP and CHD from the public databases of published GWASs. The effect sizes for lipids and obesity traits are in SD units from inverse normal-transformed traits. Fasting insulin results are in ln-transformed units for fasting insulin. Data on lipid traits have been contributed by Global Lipids Genetics
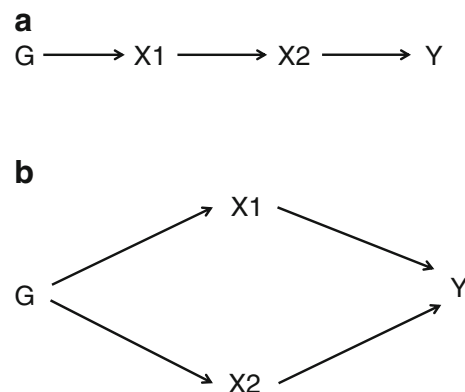

Fig. 1 Causal graph between genetic variant (G) and outcome (Y), when the variant is associated with two risk factors (X1, X2) in the same (**a**) and multiple (**b**) causal pathways

investigators, and have been downloaded from http://csg.sph. umich.edu//abecasis/public/lipids2013 (accessed 12 June 2016). Data on BMI and WHR have been contributed by Genome-wide Investigation of ANThropometric measures (GIANT) investigators and have been downloaded from www.broadinstitute.org/collaboration/giant/index. php/GIANT_consortium_data_files (accessed 12 June 2016). Data on glycaemic traits have been contributed by the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) investigators and have been downloaded from www.magicinvestigators.org (accessed 12 June 2016). Data on BP traits have been contributed by International Consortium for Blood Pressure and have been downloaded from www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study. cgi?study_id=phs000585.v1.p1 (accessed 12 June 2016). Data on coronary artery disease/myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.cardiogramplusc4d. org (accessed 12 June 2016).

Summary-level data analysis

To evaluate the causal effects of risk factor traits on CHD, we applied a statistical method that controls for pleiotropic effects, noted as 'weighted regression-based method' [14]. This method has been described in Do et al and Burgess et al [5, 14]; in short, we extracted the SNP effect estimates for LDL- and HDL-cholesterol, triacylglycerols, BMI, WHR, fasting insulin and glucose, systolic and diastolic BP, and CHD from the summary result files of reference GWASs. As some of the reference GWASs were conducted by using cardio-metabochip data, we were able to find 278 SNPs that had effect estimates available for all traits. In addition, we removed SNPs based on their high correlation with other SNPs ($r^2 > 0.5$), as strong linkage disequilibrium (LD) between the SNPs might result in too narrow CIs [13]. This resulted in 257 SNPs. Pairwise SNP correlations were estimated using the Finnish population cohort FINRISK. To further evaluate the pleiotropy assumption, we applied a sensitivity analyses with the R package gtx (version 0.0.8, https://cran.r-project.org/web/packages/gtx/index.html). In this method, SNPs are removed based on the null hypothesis that they affect disease risk only through their effects on the trait of interest (i.e. they have no pleiotropic effects). The details of this method are provided in Johnson [15]; in short, the SNPs are filtered stepwise until there is no significant heterogeneity ($p \leq 0.05$) of the SNP effects on CHD risk relative to the effects on the risk factor of interest. In addition, the effect of influential observations was evaluated using the Cook's D statistic.

The association between the SNP effects on the trait of interest (β_{Trait_1}) and CHD (β_{CHD}) were tested with a linear regression model. The effects of genetic variants on other risk factor traits were taken into account in the modelling so that the SNP effects on other traits ($\beta_{Trait\_2}, \ldots, \beta_{Trait\_n}$) were included as covariates in the linear regression model. The regression model parameters were estimated using weights defined by inverse standard errors of β_{CHD} to account for uncertainty of the β coefficients [14].

Individual-level data analysis

Study cohorts The prospective study population consists of 19,751 individuals from four FINRISK studies with risk factor and genome-wide genetic marker data. Each FINRISK study is a random sample drawn from the 25–64- or 25–74-year-old inhabitants in different regions of Finland. Extensive independent surveys were conducted in 1992, 1997, 2002 and 2007, and hospitalisation and mortality data for study participants were obtained from the Finnish National Hospital Discharge Register and the Finnish National Causes-of-Death Register. CHD was defined as myocardial infarction, unstable angina pectoris, coronary revascularisation, or death due to CHD (ICD-9 codes 410-414, 798 [www.icd9data.com/2007 /Volume1] and ICD-10 codes I20-I25, I46, R96, R98 [www. who.int/classifications/icd/en/]). The registry-based follow-up ended on 31 December 2013.

The data were genotyped in six batches with Illumina 610K, Illumina Omni Express, Illumina HumanCoreExome and Affymetrix 6.0. SNPs with Hardy–Weinberg equilibrium (HWE) p value $< 1 \times 10^{-6}$, call rate $< 95\%$ (99% for SNPs with minor allele frequency [MAF] $< 5\%$) were excluded. For Illumina 610K, Illumina Omni Express and Affymetrix 6.0 MAF $\leq 1\%$ were excluded. In HumanCoreExome data, monomorphic markers and singletons were excluded. Each subset was imputed with IMPUTE 2 [16] by using the 1000 genomes reference panel (September 2013 release). After imputation, data were merged with qctool software (version 1.4, www.well.ox.ac. uk/~gav/qctool/#overview). The FINRISK study was approved by the ethical committee of the National Public Health Institute, Helsinki, Finland. Participants gave written informed consent.

The DIetary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) study is a subcohort of FINRISK 2007. The cohort has detailed phenotypic information on metabolic traits (including the oral glucose tolerance test) and genetic data genotyped with the Illumina cardio-metabochip. In this study, 3971 individuals from the DILGOM study sample were used for sensitivity analysis of glycaemic traits.

The Corogene cohort is a case–control dataset including 2113 Finnish CHD patients and their matched controls [17]. CHD patients were collected from the Helsinki University Hospital in 2006–2008, where CHD was defined as coronary artery obstruction $> 50\%$ in at least one coronary artery. Controls were collected from the FINRISK participants living in the Helsinki-Vantaa region. For each CHD case, two

controls were matched by sex and birth year. The data were genotyped with Illumina 610K and the same pre-imputation quality control was applied as for FINRISK. Data were imputed with IMPUTE 2 by using the 1000 genomes reference panel (September 2013 release). The Corogene study was approved by the ethics committee of Helsinki University Hospital. Participants gave written informed consent.

The Genmets cohort is a case–control study for the metabolic syndrome, sampled from the Health 2000 survey [18]. Half of the study participants were defined to have the metabolic syndrome based on IDF criteria, and and half of the study participants were their matched controls. Prevalent CHD was determined at baseline based on self-reports and clinical diagnoses made by the field physicians. The genetic data were genotyped with Illumina 610K and the same quality control was applied as for FINRISK and Corogene cohorts. Data were imputed with IMPUTE 2 by using the 1000 genomes reference panel (September 2013 release). For the present study, a total of 1998 individuals were included for the analysis. The Health 2000 study was approved by the ethical committee of the National Public Health Institute, Helsinki, Finland. Participants gave written informed consent.

Statistical methods

Imputation quality for all selected SNPs was good (information score > 0.8), thus there was no need to exclude any SNPs based on the imputation quality. We calculated GRSs for each trait as a weighted sum of the risk alleles and divided by the number of SNPs in each score, after which GRSs were standardised (mean = 0, SD = 1). Effect estimates obtained from discovery GWASs were used as the weights [7–11]. There was a high correlation between systolic BP (SBP)-GRS and diastolic BP (DBP)-GRS ($r = 0.87$). Otherwise, the correlations were modest ($|r| < 0.41$, ESM Fig. 1).

The associations between the GRSs and CHD events were estimated with logistic regression and conditional logistic regression models. Analyses were conducted separately in FINRISK, Corogene and Genmets study cohorts. We first fitted the models individually for each GRS, and then combined them into a single model to reduce the effect of pleiotropic SNPs that are included in more than one GRS. DBP-GRS was not included in the joint model because of its high correlation with SBP-GRS. In FINRISK, we adjusted the models for age, sex, cohort indicator and genotyping batch. To avoid overlap with the Corogene cohort, those individuals used as controls in Corogene were excluded from the FINRISK dataset. In Corogene, we studied the association with conditional logistic regression analysis. In Genmets, we used logistic regression model adjusted for age, sex and metabolic syndrome case status. The results from the three datasets were combined with fixed-effects meta-analysis, as no significant heterogeneity was detected (ESM Table 2).

We then performed some additional analyses in the FINRISK cohort ($N = 19,078$, incident CHD events $n = 1376$), where the risk factor traits were measured at baseline (1992, 1997, 2002 and 2007). We estimated the associations between the GRSs and risk factor traits. The measurements of insulin and glucose were available only for a subset of the data ($n = 12,754$ for insulin, $n = 11,194$ for glucose). The trait distributions were inspected and outliers removed, and log-transformation was applied to normalise triacylglycerols and insulin. In the analyses for lipid traits, those individuals receiving lipid-lowering medication were excluded. For BP analyses, we added 15 mmHg for SBP and 10 mmHg for DBP for individuals receiving treatment for high BP. For glycaemic traits, diabetic individuals were excluded. Associations between the GRSs and risk factor traits were obtained with linear regression models adjusted for age, sex, cohort indicator and genotyping batch. To evaluate causal effects of risk factor traits on CHD, we applied the two-stage least-squares (2SLS) method that controls for pleiotropic effects of the GRSs. This is a two-step regression model, where the risk factors are first regressed on the GRSs and then the outcome is regressed on the fitted values of the risk factors [13].

All analyses were done with the R statistical package (version 3.0.2).

Results

Summary-level data analysis

The effect estimates of SNPs for the traits are denoted as β_{LDL}, β_{HDL}, β_{TG}, β_{BMI}, β_{WHR}, $\beta_{Insulin}$ (β_{Ins}), $\beta_{Glucose}$ (β_{Glu}), β_{SBP}, β_{DBP} and β_{CHD}. The results of the linear regression model $\beta_{CHD} \sim \beta_{LDL} + \beta_{HDL} + \beta_{TG} + \beta_{BMI} + \beta_{WHR} + \beta_{Ins}$ are shown in Table 1. The effects for glycaemic traits β_{Ins} and β_{Glu} and lipids β_{LDL} and β_{TG} were associated with β_{CHD} after accounting for all other risk factor related SNP effects. Interestingly, the effects for obesity traits β_{BMI} and β_{WHR} were also associated with β_{CHD} only when β_{Ins} was omitted from the model (Table 2). Excluding other βs did not modify the results for other predictors. This supports the previous findings that LDL-cholesterol and triacylglycerols are causally related to CHD, and suggests that insulin is a causal risk factor that might explain the epidemiological associations between obesity traits and CHD.

Our sensitivity analyses with gtx removed 68 SNPs based on the null hypothesis of no pleiotropic effects. With 189 SNPs, the effect size for β_{Ins} was 0.80 (95% CI 0.55, 1.05; $p = 2.5^{-9}$; Fig. 2a, b), which is comparable with the estimate of β_{Ins} in Table 1. Cook's D statistic indicated one influential SNP (rs3184504, ESM Fig. 2), but removing this from the model did not change the results (ESM Table 3). Finally, we

Table 1 Associations between the SNP effects on risk factor traits and SNP effects on CHD in a multivariable model

Effects of 257 SNPs (units)	β	95% CI	p value
β_{LDL} (SD)	0.418	0.335, 0.501	2×10^{-16}***
β_{HDL} (SD)	−0.021	−0.107, 0.065	0.6283
β_{TG} (SD)	0.173	0.048, 0.298	0.0073**
β_{BMI} (SD)	0.044	−0.099, 0.187	0.544
β_{WHR} (SD)	0.117	−0.073, 0.307	0.228
β_{Ins} ($\log_e$[pmol/l])	0.835	0.450, 1.220	3×10^{-5}***
β_{Glu} (mmol/l)	0.529	0.186, 0.872	0.0028**
β_{SBP} (mmHg)	0.009	−0.013, 0.032	0.4185
β_{DBP} (mmHg)	0.017	−0.018, 0.053	0.3405

βs represent the effect sizes for SNPs on LDL, HDL, TG, BMI, WHR, insulin, glucose, SBP, DBP and CHD in the published GWAS meta-analysis. Regression was performed with the predictor variable of the effect size on risk factor traits (β) and the outcome variable of the effect size on CHD

HDL, HDL-cholesterol; LDL, LDL-cholesterol; TG, triacylglycerols

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

performed the analysis without the SNPs with lower frequency (MAF < 0.05) as they might have imprecise effect estimates. This exclusion did not affect the results notably (ESM Table 4).

Individual-level data analysis

Study characteristics The characteristics of the study cohorts are shown in Table 3. Four FINRISK studies were initiated in 1992, 1997, 2002 and 2007, when the baseline information was collected. The mean age at baseline was 47.9 years (SD 13.3 years) and 47.0% of participants were men and 53.0% women. During the median follow-up of 14.5 years

Table 2 Associations between the SNP effects on risk factor traits and SNP effects on CHD in a multivariable model, without β_{Ins}

Effects of 257 SNPs (units)	β	95% CI	p value
β_{LDL} (SD)	0.409	0.323, 0.494	2×10^{-16}***
β_{HDL} (SD)	−0.047	−0.135, 0.041	0.2973
β_{TG} (SD)	0.192	0.063, 0.321	0.0039**
β_{BMI} (SD)	0.183	0.051, 0.315	0.0071**
β_{WHR} (SD)	0.243	0.057, 0.430	0.0113*
β_{Glu} (mmol/l)	0.692	0.346, 1.038	0.0001***
β_{SBP} (mmHg)	0.007	−0.016, 0.031	0.5442
β_{DBP} (mmHg)	0.023	−0.014, 0.060	0.2173

βs represent the effect sizes for SNPs on LDL, HDL, TG, BMI, WHR, glucose, SBP, DBP and CHD in the published GWAS meta-analysis. Regression was performed with the predictor variable of the effect size on risk factor traits (β) and the outcome variable of the effect size on CHD

HDL, HDL-cholesterol; LDL, LDL-cholesterol; TG, triacylglycerols

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

(interquartile range 10.75–18.25), 1420 (7.2%) incident CHD cases occurred. In total, 653 individuals had been diagnosed cardiovascular disease (CVD) at baseline. The Corogene study consisted of 2113 CHD events and 4440 matched controls (of whom 1576 were unique). The Genmets cohort consisted of 1998 individuals of whom 99 (5%) had CHD at baseline.

Associations between GRSs, risk factors and CHD Among nine GRSs for risk factor traits, BMI-, WHR- and insulin-GRSs had individually significant associations with CHD (Table 4). The estimates were largely similar in a multivariable model with all GRSs tested simultaneously (BMI-GRSs: OR 1.08 per SD in GRS, 95% CI 1.04, 1.13, $p = 1 \times 10^{-4}$; WHR-GRS: OR 1.06 per SD in GRS, 95% CI 1.02, 1.10, $p = 0.003$; insulin-GRS: OR 1.05 per SD in GRS, 95% CI 1.01, 1.09, $p = 0.025$).

The analyses with risk factor data in the FINRISK cohort showed that all the GRSs were associated with the corresponding risk factor trait ($p \leq 1 \times 10^{-8}$, ESM Table 5). However, the 2SLS model yielded non-significant results for all risk factors, probably because of lack of power and imprecise measurement of glycaemic traits due to inadequate fasting ($n = 12,754$ for insulin, $n = 11,194$ for glucose with, on average, 5 h fasting). The modest association between LDL-GRSs and CHD also suggests that the power of the GRSs to evaluate causal process is relatively weak; in fact, the GRSs only explained 0.3–7.3% of the trait variance (ESM Table 5).

As data for glycaemic traits were limited, we performed a sensitivity analyses using the DILGOM metabochip data with > 8 h fasting ($n = 3971$), from which we estimated the effect of measured insulin levels on incident CHD. The effect estimate of fasting insulin on CHD in the DILGOM cohort was higher (HR 1.32 per SD in GRS, 95% CI 1.14, 1.53) than in FINRISK (HR 1.18 per SD in GRS, 95% CI 1.12, 1.25). This result illustrates how incomplete fasting can bias the observational estimates of insulin. To further evaluate the ability of insulin-GRS to reflect insulin resistance, we studied associations between insulin-GRS and different measurements of insulin resistance in the DILGOM cohort. Insulin-GRS was associated with fasting and 2 h insulin (insulin 2 h after oral glucose tolerance test), HOMA-IR and insulin sensitivity index (ISI) (ESM Table 6).

Discussion

We have shown evidence that glycaemic traits (insulin and glucose) are causally related to CHD. Our results also support the previous findings that triacylglycerols and LDL-cholesterol are causal risk factors for CHD. Moreover, the SNP effects for insulin explained the effects for BMI and WHR, which suggests that the risk-increasing effect of obesity

Fig. 2 Effects of SNPs on insulin and risk for CHD. (**a**) Effects of all 257 independent SNPs. (**b**) Effects of 189 SNPs after removing the SNPs based on the heterogeneity test. Vertical grey lines show 95% confidence interval (CI) for each SNP. A solid red line shows the estimated effect for insulin on CHD with dashed lines representing the 95% CI
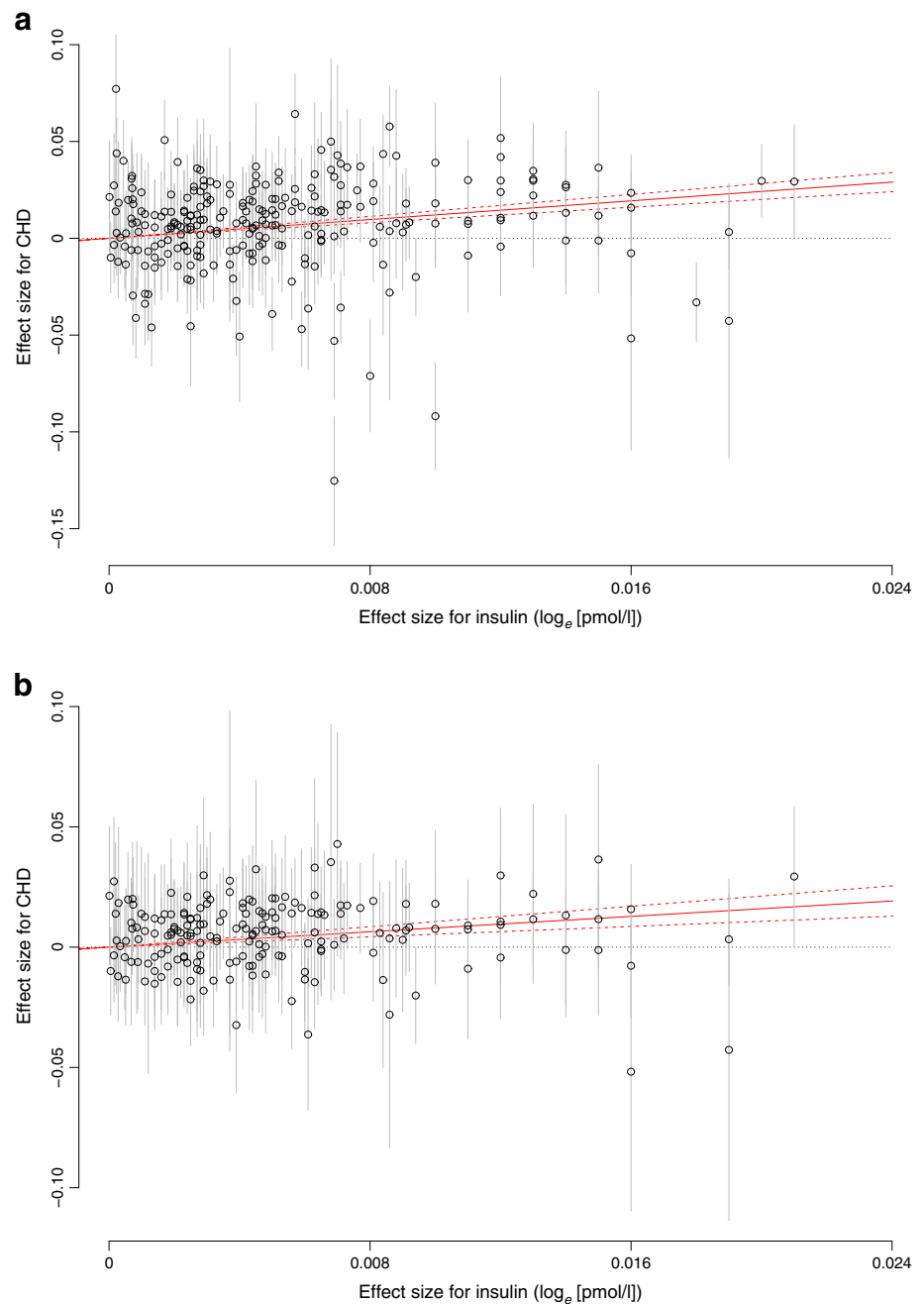
is at least partially mediated by insulin. This is in line with the idea that healthy obesity (as occurs in 10–25% of obese individuals) is due to preserved insulin sensitivity [19]. Our results from the individual-level data analysis supported the causal role of insulin, but not glucose, on CHD.

Insulin resistance usually coexists with obesity, but not always. For example, insulin resistance is common in some populations with normal BMI [20]. The pathological role of insulin resistance in CVD has been debated for decades. The Helsinki Policemen Study was one of the first prospective studies demonstrating associations between high plasma insulin and incident CHD [21–23], stroke [24] and all-cause

mortality [25]. Further, other markers of insulin resistance, namely HOMA-IR and ISI, have been shown to associate with CVD risk independently of the metabolic syndrome [26–28]. Other studies, however, have shown that fasting insulin levels predict the CVD risk only moderately and the effect is highly dependent on the measurement assay used [29].

Contradictory results could be at least partly explained by the limitations in observational epidemiological studies to assess causality. Observational studies are prone to confounding, reverse causality and multicollinearity, which could especially affect the traits that are measured with lower precision, such as insulin resistance. The advantage of using genetic

Table 3 Cohort characteristics

Characteristic	FINRISK	DILGOM	Corogene	Genmets
N	19,731	3971	6553	1998
Sex, n (%)				
Men	9283 (47.0)	1813 (45.7)	4572 (69.8)	975 (48.8)
Women	10,448 (53.0)	2158 (54.3)	1981 (30.2)	1023 (51.2)
Mean age, years (SD)	47.9 (13.3)	52.1 (13.6)	66.6 (10.9)	50.8 (11.1)
Blood lipids, mmol/l, (SD)				
TC	5.5 (1.1)	5.3 (1.0)	NA	6.0 (1.1)
LDL	3.4 (1.0)	3.2 (0.9)	NA	3.9 (1.2)
HDL	1.4 (0.4)	1.4 (0.4)	NA	1.3 (0.4)
TG	1.5 (1.0)	1.4 (0.9)	NA	1.6 (1.0)
BP, mmHg (SD)				
SBP	135.6 (19.9)	137.1 (20.5)	NA	134.4 (19.7)
DBP	80.4 (11.6)	79.2 (11.3)	NA	83.6 (10.7)
BMI, kg/m^2 (SD)	26.7 (4.7)	27.2 (4.9)	NA	27.3 (4.5)
WHR, cm/cm (SD)	0.9 (0.1)	0.9 (0.1)	NA	0.92 (0.1)
Glucose, mmol/l (SD)	5.7 (1.2)[a]	5.9 (1.0)	NA	5.4 (0.5)
Insulin, pmol/l (SD)	7.6 (9.8)[b]	6.9 (7.2)	NA	8.5 (6.2)
Prevalent diabetes, n (%)	1311 (6.6)	353 (8.9)	NA	NA
CHD, n (%)	1899 (9.6)	165 (4.2)	2113 (32.2)	99 (5.0)
Prevalent CVD, n (%)	653 (3.3)	191 (4.8)	NA	NA
Incident CHD, n (%)	1420 (7.2)	138 (3.5)	NA	NA
Follow-up, years	14.5	7	NA	NA

[a] Data available for 11,194 participants

[b] Data available for 12,754 participants

HDL, HDL-cholesterol; LDL, LDL-cholesterol; NA, not applicable; TG, triacylglycerols

markers in epidemiological studies compared with traditional risk factor traits is that it is possible to avoid most of these common problems. However, it is probable that SNPs associated with direct measurement of insulin resistance (rather than fasting insulin) would be more suitable genetic instruments for causal analysis. As genetic loci associated with insulin resistence will be identified, this hypothesis can be tested. Nevertheless, the correlation between fasting insulin and the gold-standard clamp measurement of insulin resistance is reasonably high, and the measurements correlate similarly with

Table 4 Associations between GRSs and CHD ($N = 25{,}534$, cases $n = 4016$)

GRS	No. of SNPs	Reference	OR[a]	95% CI	p	Significance after multiple testing correction[b]
LDL	67	7	1.045	1.006, 1.086	0.0245	NS
HDL	79	7	0.959	0.923, 0.996	0.0321	NS
TG	48	7	1.047	1.007, 1.088	0.0203	NS
BMI	96	10	1.084	1.043, 1.127	4×10^{-5}	***
WHR	48	9	1.068	1.028, 1.110	0.0007	**
Ins	20	11	1.061	1.021, 1.103	0.0024	*
Glu	20	11	0.996	0.959, 1.034	0.8220	NS
SBP	20	8	1.015	0.977, 1.055	0.4507	NS
DBP	19	8	1.033	0.995, 1.074	0.093	NS

[a] Per SD in GRS, adjusted for study-specific covariates

[b] Bonferroni correction, $p < 0.0056$

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Glu, glucose; HDL, HDL-cholesterol; Ins, insulin; LDL, LDL-cholesterol; NS, non-significant; TG, triacylglycerols

metabolic and cardiovascular risks [30]. Thus, fasting insulin is a useful marker of insulin resistance in large-scale studies where it is not possible to conduct the clamp measurement.

To overcome the limitations of observational studies to assess causality, many studies have used summary-level data from GWASs to infer causal relationships between risk factors and CHD [5, 6, 31, 32]. Yaghootkar et al [32] showed that a GRS comprising SNPs associated with fasting insulin and adverse metabolic profile, but lower BMI, is associated with CHD. This is consistent with our finding that insulin is associated with CHD independently of obesity. Our study has several strengths compared with these studies. First, these studies have only evaluated one or a few traits at a time (e.g. lipids [5], BMI [31]). We are not aware of any other study that takes genetic effects on multiple CHD risk factor traits into account simultaneously. Second, the method for pleiotropic SNP effects suggested by Do et al [5] has been criticised as it does not account for uncertainty of the effect estimates [13]. In our analyses, we applied a simple modification of this method that takes the uncertainty into account [14]. Third, apart from studies using only summary-level data, we performed complementary analysis in our study with the large individual-level dataset with consistent disease event definitions and genome-wide genetic data. In addition, by using the population-based FINRISK cohort with extensive baseline health information and follow-up data, we were able to estimate the relationships between the GRSs, risk factor traits and prospectively ascertained CHD events in the same dataset. The diagnoses for incident CHD events in our data have been obtained from validated national registries with comprehensive coverage [33, 34].

Our study also has some limitations. First, we restricted our analysis to lead SNPs with genome-wide significance. Consequently, the variants only explain a small proportion of trait variability. Identifying new loci for these traits could highly increase the effect sizes between the GRSs and CHD and increase the power in MR analysis. Another way to create more powerful GRSs is to combine multiple genetic variants from a single gene region. Even if the variants are correlated, this method can provide more precise causal estimates [35]. Second, limited data of glycaemic traits restricted us to perform powerful multivariable analysis with measured risk factor data. Thus, individual-level analysis with detailed phenotypic and genetic information is required to further elucidate the causal processes between insulin resistance and CHD. Third, even though our statistical modelling approach and sensitivity analyses minimise the potential for pleiotropic effects on our set of well-known CHD risk traits, it is possible that the tested SNPs have large effects on other factors and causal pathways leading to CHD. As further GWAS results for potential risk factors are published, these can be added to the future analyses of causality and thus minimise the possibility of pleiotropy.

In summary, our findings from complementary summary- and individual-level data analyses support the causal role of insulin and insulin resistance in the pathogenesis of CHD. Efficient prevention and treatment of insulin resistance is essential to prevent future CHD events.

References

1. Baigent C, Keech A, Kearney PM et al (2005) Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 366:1267–1278

2. Shepherd J, Cobbe SM, Ford I et al (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. N Engl J Med 333:1301–1307

3. Heart Protection Study Collaborative Group (2002) MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet 360:7–22

4. Barter PJ, Caulfield M, Eriksson M et al (2007) Effects of torcetrapib in patients at high risk for coronary events. N Engl J Med 357:2109–2122

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

6. Holmes MV, Asselbergs FW, Palmer TM et al (2014) Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J 36:539–550

7. Willer CJ, Schmidt EM, Sengupta S et al (2013) Discovery and refinement of loci associated with lipid levels. Nat Genet 45:1274–1283

8. Ehret GB, Munroe PB, Rice KM et al (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478:103–109

9. Shungin D, Winkler TW, Croteau-Chonka DC et al (2015) New genetic loci link adipose and insulin biology to body fat distribution. Nature 518:187–196

10. Locke AE, Kahali B, Berndt SI et al (2015) Genetic studies of body mass index yield new insights for obesity biology. Nature 518:197–206

11. Scott RA, Lagou V, Welch RP et al (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 44:991–1005

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

13. Burgess S, Thompson SG (2015) Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol 181:251–260

14. Burgess S, Dudbridge F, Thompson SG (2015) Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". Am J Epidemiol 181:290–291

15. Johnson T (2012) Efficient calculation for multi-SNP genetic risk scores. Available from https://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf, accessed 6 June 2016

16. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529

17. Vaara S, Nieminen MS, Lokki ML et al (2012) Cohort profile: the Corogene study. Int J Epidemiol 41:1265–1271

18. Aromaa A, Koskinen S (eds) (2004) Health and functional capacity in Finland. Baseline results of the Health 2000 health examination survey. National Public Health Institute, Helsinki

19. Bluher M (2010) The distinction of metabolically 'healthy' from 'unhealthy' obese individuals. Curr Opin Lipidol 21:38–43

20. Grundy SM, Brewer HB Jr, Cleeman JI et al (2004) Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation 109:433–438

21. Pyorala K (1979) Relationship of glucose tolerance and plasma insulin to the incidence of coronary heart disease: results from two population studies in Finland. Diabetes Care 2:131–141

22. Pyorala K, Savolainen E, Kaukola S, Haapakoski J (1985) Plasma insulin as coronary heart disease risk factor: relationship to other risk factors and predictive value during 9 1/2-year follow-up of the Helsinki Policemen Study population. Acta Medica Scand Suppl 701:38–52

23. Pyorala M, Miettinen H, Laakso M, Pyorala K (1998) Hyperinsulinemia predicts coronary heart disease risk in healthy middle-aged men: the 22-year follow-up results of the Helsinki Policemen Study. Circulation 98:398–404

24. Pyorala M, Miettinen H, Laakso M, Pyorala K (1998) Hyperinsulinemia and the risk of stroke in healthy middle-aged men: the 22-year follow-up results of the Helsinki Policemen Study. Stroke 29:1860–1866

25. Pyorala M, Miettinen H, Laakso M, Pyorala K (2000) Plasma insulin and all-cause, cardiovascular, and noncardiovascular mortality: the 22-year follow-up results of the Helsinki Policemen Study. Diabetes Care 23:1097–1102

26. Rutter MK, Meigs JB, Sullivan LM, D'Agostino RB Sr, Wilson PW (2005) Insulin resistance, the metabolic syndrome, and incident cardiovascular events in the Framingham Offspring Study. Diabetes 54:3252–3257

27. Jeppesen J, Hansen TW, Rasmussen S, Ibsen H, Torp-Pedersen C, Madsbad S (2007) Insulin resistance, the metabolic syndrome, and risk of incident cardiovascular disease: a population-based study. J Am Coll Cardiol 49:2112–2119

28. Bonora E, Kiechl S, Willeit J et al (2007) Insulin resistance as estimated by homeostasis model assessment predicts incident symptomatic cardiovascular disease in caucasian subjects from the general population: the Bruneck study. Diabetes Care 30:318–324

29. Ruige JB, Assendelft WJ, Dekker JM, Kostense PJ, Heine RJ, Bouter LM (1998) Insulin and risk of cardiovascular disease: a meta-analysis. Circulation 97:996–1001

30. Lorenzo C, Haffner SM, Stancakova A, Laakso M (2010) Relation of direct and surrogate measures of insulin resistance to cardiovascular risk factors in nondiabetic finnish offspring of type 2 diabetic individuals. J Clin Endocrinol Metab 95:5082–5090

31. Holmes MV, Lange LA, Palmer T et al (2014) Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. Am J Hum Genet 94:198–208

32. Yaghootkar H, Scott RA, White CC et al (2014) Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease and type 2 diabetes. Diabetes 63:4369–4377

33. Pajunen P, Koukkunen H, Ketonen M et al (2005) The validity of the Finnish Hospital Discharge Register and causes of death register data on coronary heart disease. Eur J Cardiovasc Prev Rehabil 12:132–137

34. Sund R (2012) Quality of the Finnish Hospital Discharge Register: a systematic review. Scandinavian J Public Health 40:505–515

35. Burgess S, Dudbridge F, Thompson SG (2015) Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med 35:1880–1906