ARTICLE

The impact of GDF-15, a biomarker for metformin, on the risk of coronary artery disease, breast and colorectal cancer, and type 2 diabetes and metabolic traits: a Mendelian randomisation study

Shiu Lun Au Yeung[1] · Shan Luo[1] · C. Mary Schooling[1,2]

Abstract

Aims/hypothesis Growth differentiation factor 15 (GDF-15), a suggested biomarker for metformin use, may explain the potential cardioprotective and anti-cancer properties of metformin. We conducted a Mendelian randomisation study to examine the role of GDF-15 in risk of coronary artery disease (CAD) and breast and colorectal cancer. Secondary analyses included examination of the association of GDF-15 with type 2 diabetes, glycaemic traits, BP, lipids and BMI.

Methods We obtained SNPs strongly (p value $<5 \times 10^{-8}$) predicting GDF-15 from a genome-wide association study (GWAS) ($n = 5440$) and applied them to genetic studies of CAD (CARDIoGRAMplusC4D 1000 Genomes-based GWAS [$n = 184,305$]), type 2 diabetes (DIAGRAM [DIAbetes Genetics Replication And Meta-analysis; $n = 898,130$]), glycaemic traits (MAGIC [the Meta-Analyses of Glucose and Insulin-related traits Consortium; HbA$_{1c}$: $n = 123,665$; fasting glucose: $n = 46,186$]), BP, breast cancer and colorectal cancer (UK Biobank [$n \leq 401,447$]), lipids (GLGC [Global Lipids Genetic Consortium; $n \leq 92,820$]) and adiposity (GIANT [Genetic Investigation of ANthropometric Traits Consortium; $n = 681,275$]). Causal estimates were obtained using inverse variance weighting, taking into account correlations between SNPs. Sensitivity analyses included focusing on the lead SNP (rs888663) and validation for CAD in the UK Biobank and for breast cancer in the Breast Cancer Association Consortium.

Results Using 5 SNPs, increased GDF-15 was associated with lower CAD (OR 0.93 per SD increase, 95% CI 0.87, 0.99) and breast cancer (OR 0.89 per SD increase, 95% CI 0.82, 0.96), with similar results from lead SNP analysis. However, the associations with CAD (OR 0.99 per SD increase, 95% CI 0.93, 1.04) and breast cancer (OR 0.97 per SD increase, 95% CI 0.94, 1.01) in the validation studies were not as apparent. GDF-15 was not associated with type 2 diabetes, glycaemic traits, CAD risk factors or colorectal cancer.

Conclusions/interpretation There is no convincing evidence that GDF-15 reduces risk of CAD or breast or colorectal cancer. Whether the observed inverse association of metformin use with cancer risk is via other unexplored mechanistic pathways warrants further investigation.

Keywords Cancer · Coronary artery disease · Growth differentiation factor 15 · Mendelian randomisation · metformin

✉ Shiu Lun Au Yeung
ayslryan@hku.hk

[1] School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong SAR, China

[2] Graduate School of Public Health and Health Policy, City University of New York, New York, NY, USA

Abbreviations

BCAC	Breast Cancer Association Consortium
CAD	Coronary artery disease
CVD	Cardiovascular disease
DBP	Diastolic BP
DIAGRAM	DIAbetes Genetics Replication And Meta-analysis
GDF-15	Growth differentiation factor 15
GIANT	Genetic Investigation of Anthropometric Traits
GLGC	Global Lipids Genetics Consortium

Research in context

What is already known about this subject?

- Metformin use is inversely associated with cardiovascular disease and cancer, although these observations are susceptible to confounding by indication and to misallocation of exposure time

- Growth differentiation factor 15 (GDF-15) is a biomarker of metformin use and may explain the non-glycaemic effects of metformin on cardiovascular disease and cancer risk

What is the key question?

- Does genetically predicted GDF-15 influence the risk of coronary artery disease (CAD), colorectal cancer, breast cancer, type 2 diabetes or glycaemic traits?

What are the new findings?

- Although increased GDF-15 appeared to decrease the risk of CAD and breast cancer in the main analyses, these associations were not observed in the validation study

- GDF-15 was not associated with type 2 diabetes, glycaemic traits, CAD risk factors or colorectal cancer

How might this impact on clinical practice in the foreseeable future?

- There is no convincing evidence that GDF-15 reduces risk of CAD or breast or colorectal cancer. Whether the inverse association of metformin use with cancer is an artefact of biases or a result of other unexplored mechanistic pathways warrants further investigation, which would help validate whether this medication could be repurposed for cancer prevention

GWAS	Genome-wide association study
HRC	Haplotype Reference Consortium
InSIDE	Instrument strength independent of direct effects
IVW	Inverse variance weighting
LD	Linkage disequilibrium
MAGIC	Meta-Analysis of Glucose and Insulin-related traits Consortium
NSPHS	Northern Sweden Population Health Study
ORIGIN	Outcome Reduction with Initial Glargine Intervention
PheWAS	Phenome-wide association study
PIVUS	Prospective Investigation of the Vasculature in Uppsala Seniors
SAIGE	Scalable and Accurate Implementation of GEneralized mixed model
SBP	Systolic BP
Sydney MAS	Sydney Memory and Ageing Study

Introduction

Metformin is the first line medication for diabetes, given its superiority and safety compared with other diabetes medications (e.g. sulfonylureas) [1], which suggests metformin reduces blood glucose levels via potentially less hazardous pathways than other medications [1]. Some RCTs have also suggested a potential beneficial effect of metformin on cardiovascular disease (CVD) although the evidence remains inconclusive [2]. In addition, the relation of metformin use with cancer remains controversial because most studies suggesting a protective effect are observational and hence are open to confounding by indication and to misallocation of exposure time [3]. Assessing the impact of potential downstream factors of metformin, such as growth differentiation factor 15 (GDF-15), may provide additional insight to elucidate the health impacts of metformin. GDF-15 (also known as macrophage inhibitory cytokine-1 [MIC-1]) was found to be a strong biomarker of metformin use, but not of sulfonylurea use, in a cross-sectional analysis of the Outcome Reduction with Initial Glargine Intervention (ORIGIN) trial. The association of metformin use with GDF-15 remained after adjusting for glycaemic traits, such as HbA_{1c} and glucose, which suggests GDF-15 could be a potential non-glycaemic effect of metformin use, possibly explaining the effect of metformin on other health outcomes, including coronary artery disease (CAD) and cancer [4]. Although GDF-15 is also associated with other cardiovascular risk factors such as plasma creatinine, diuretic use and smoking, metformin use appears to be the main contributor to GDF-15 [5]. Observationally, GDF-15 is positively related to CVD and cancer [6, 7]. However, this association could be driven by reverse causation, confounding and selection bias (i.e. selective survival before recruitment). Elucidating the role of GDF-15 in CVD and cancer may help clarify whether metformin could be repurposed to further mitigate CVD and to reduce cancer risk.

Mendelian randomisation studies are increasingly used to ascertain causes of disease since they utilise genetic variants of exposure randomly assigned at conception, and hence are less vulnerable to confounding than observational studies [8]. A previous study examined the relation of two SNPs (rs62122429 and rs62122430) relevant to GDF-15 with CAD in CARDIoGRAM and found rs62122429 associated with CAD while rs62122430 was not [4]. This finding suggests GDF-15 may have a role in CAD development, although a recent Mendelian randomisation study did not provide strong evidence for the role of GDF-15 in cardiometabolic diseases. Nevertheless, this study only used 3 SNPs as instruments and did not formally assess its role in cancer risk, which is directly relevant to the evaluation of metformin's potential anti-cancer property [9]. Here, we conducted Mendelian randomisation to evaluate the impact of GDF-15 on CAD and cancer (breast and colorectal) risk. For completeness, we also assessed its relation with type 2 diabetes and glycaemic traits (HbA$_{1c}$ and fasting glucose), given GDF-15 is strongly related to metformin use, and other established CAD risk factors (systolic [SBP] and diastolic [DBP] BP, HDL-cholesterol, LDL-cholesterol and BMI).

Methods

Study design This is a two sample Mendelian randomisation study using summary statistics from genome-wide association studies (GWAS). It has three main assumptions [8]. First, the genetic instruments, i.e. SNPs, should be associated with GDF-15. Second, the genetic instruments should be independent of confounders. Third, the exclusion-restriction assumption that the genetic instruments should only be linked with the outcome via GDF-15.

Exposure: genetic predictors of GDF-15 The exposure in this study was genetically predicted GDF-15, in SDs. The genetic predictors of GDF-15 were extracted from a GWAS consisting of four cohorts ($n = 5440$) of people of European descent: the Framingham Offspring Cohort, the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) Study, the Northern Sweden Population Health Study (NSPHS), and the Sydney Memory and Ageing Study (Sydney MAS) [10]. The mean age was 62 years and 53% of the participants were women. Serum/plasma GDF-15 was measured using immunoassays. Imputation was based on the HapMap2 reference panel (Framingham; PIVUS and Sydney MAS) and 1000 Genomes Project Phase 3 reference panel (NSPHS). We used estimates adjusted for age, sex, SBP, antihypertensive medication use, diabetes mellitus, smoking status and year of data collection (NSPHS only). Population stratification was accounted for using different methods in the studies, such as principal components and kinship matrix adjustment [10]. Electronic supplementary material (ESM) Table 1 shows the nine correlated SNPs,

identified in the GWAS, which explained 21.5% of the variance of GDF-15. Six SNPs were in the *PGPEP1* gene region whilst 3 SNPs were in the *GDF15* gene region. We excluded 3 SNPs (rs3746181, rs1363120 and rs1054564) in high linkage disequilibrium (LD) ($r^2 \geq 0.8$) and with higher p values than the other SNPs based on LD in European samples from the 1000 Genomes project. We also excluded a SNP (rs16982345) which did not reach genome-wide significance (p value $>5 \times 10^{-8}$). As such, five correlated SNPs were used as genetic instruments [11].

Outcomes The primary outcomes were CAD, breast cancer and colorectal cancer. The secondary outcomes were type 2 diabetes, HbA$_{1c}$ (mmol/mol) and glucose (mmol/l), SBP and DBP (mmHg), LDL-cholesterol (per SD change), HDL-cholesterol (per SD change) and BMI (per SD change). The studies used are described below.

Genetic associations with CAD Genetic associations with CAD were obtained from CARDIoGRAMplusC4D 1000 Genomes-based GWAS, a meta-analysis of GWAS of CAD studies (cases = 60,801, controls = 123,504) of people of mainly European descent (77%) with imputation using the 1000 Genomes Phase 1 v3 reference panel [12]. The definition of CAD varied across studies, such as diagnosis of myocardial infarction, acute coronary syndrome, chronic stable angina, or coronary stenosis >50%. Diagnoses of CAD also varied across studies, such as clinical diagnosis, procedures (coronary angiography results or bypass surgery), use of medications or symptoms that indicate angina, or self-report of a doctor diagnosis, as described elsewhere [12]. CARDIoGRAMplusC4D 1000 Genomes-based GWAS was adjusted for study-specific covariates and genomic control.

Genetic associations with colorectal and breast cancer Genetic associations with colorectal and breast cancer risk were obtained from the UK Biobank summary statistics generated by Scalable and Accurate Implementation of GEneralized mixed model (SAIGE) and restricted to white British of European ancestry ($n \leq 401,447$) [13]. The mean age for all the UK Biobank participants at recruitment was 57 years [14]. In brief, SAIGE fits a null logistic mixed model to estimate the variance component and the model parameters, followed by ascertainment of the gene-phenotype association with the application of saddlepoint approximation to the score test statistics [13, 14]. SAIGE is particularly suitable for studies with extreme imbalance in case–control ratios and is able to control for sample relatedness [13]. Imputation was based on the Haplotype Reference Consortium (HRC) imputation reference panel. Colorectal cancer (4562 cases, 382,756 controls) and breast cancer (12,898 cases, 388,549 controls) in the UK Biobank were defined using UK Biobank phenome-wide association study (PheWAS) codes, which is an aggregate of relevant International Classification of Diseases codes (153 for

colorectal cancer; 174 for breast cancer) [15]. SAIGE was adjusted for sex, birth year and the first four principal components.

Genetic associations with type 2 diabetes Genetic associations with type 2 diabetes were obtained from DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, a meta-analysis of 32 GWAS of type 2 diabetes studies (cases = 74,124, controls = 824,006) of European ancestry [16]. Type 2 diabetes was defined in various ways, such as previous diagnosis, the use of glucose-lowering medication, and self-report. Imputation was performed based on the HRC reference panel for all studies except deCODE which was imputed based on the Icelandic panel. The GWAS was adjusted for study-specific covariates, principal components, and corrected for genomic control. In this study, we used summary statistics without adjustment for BMI to avoid the possibility of collider bias [16].

Genetic associations with HbA$_{1c}$ and glucose Genetic associations with HbA$_{1c}$ and glucose were obtained from the Meta-Analysis of Glucose and Insulin-related traits consortium (MAGIC), a meta-analysis of GWAS of HbA$_{1c}$ in 159,940 adults without diabetes, imputed using Phase 2 of the International HapMap Project reference panel [17]. In this study, we only used genetic associations from 123,665 participants of European ancestry. Genetic associations with HbA$_{1c}$ were adjusted for age, sex, study-specific covariates and genomic control. Genetic associations with glucose were generated from a meta-analysis of GWAS of fasting glucose in up to 46,186 participants of European descent without diabetes, imputed using the HapMap CEU (Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection) reference panel [18]. Adjustment was made for age, sex, study-specific covariates and genomic control.

Genetic associations with HDL- and LDL-cholesterol Genetic associations with lipids were obtained from the Global Lipids Genetics Consortium (GLGC), a meta-analysis of GWAS of different study designs, which included 188,577 participants of European ancestry [19]. Genotyping was done using two arrays: the GWAS array (n = 94,595) and the Metabochip array (n = 93,982). Imputation was based on the CEU HapMap release 22 reference panel. However, the SNPs used in this study were only present in one of the panels (n ≤ 92,820). Blood lipids were typically sampled after more than 8 h of fasting. Participants were excluded if they took lipid lowering medications in the majority of studies. LDL-cholesterol was directly measured in 24% of the participants and was estimated using the Friedewald formula in the remaining 76% of the participants. HDL-cholesterol was directly measured. The GWAS was adjusted for age, sex and genomic control.

Genetic associations with BMI Genetic associations with BMI were obtained from the Genetic Investigation of ANthropometric Traits (GIANT) Consortium, which is a meta-analysis of GWAS of different study designs (n = 681,275) [20]. Imputation was based on HapMap CEU release 22 reference panel (GWAS by Locke et al) [20] and the HRC imputation reference panel (UK Biobank). BMI was calculated based on height and weight (either measured or self-reported). The GWAS was adjusted for age and other study-specific covariates (e.g. principal components), and corrected for genomic control.

Genetic associations with SBP and DBP Genetic associations with SBP and DBP (automated reading) were obtained from summary statistics generated from the UK Biobank, restricted to ~337,000 unrelated individuals of white British ancestry, with imputation using the HRC imputation reference panel [21]. Seated BP was measured using the Omron HRM-7015IT digital BP monitor (Omron Healthcare, Kyoto, Japan). Genetic associations with BP (mmHg) were obtained using multivariable linear regression, adjusted for age, age squared, sex, the interactions of age and age squared with sex, and the first 20 principal components.

Statistical analysis

We calculated the mean F statistics of the instruments used in this study to assess potential weak instruments bias [22, 23]. To take into account correlations between the SNPs when calculating the estimates using inverse variance weighting (IVW), we obtained the correlation matrix based on European 1000 Genomes data obtained from MR-Base. We then used IVW with the correlation matrix in our main analysis to correct for correlation between SNPs [24, 25]. We used fixed effects IVW since variation in the Wald estimates (gene-outcome association divided by gene-exposure association) for each SNP are likely due to sampling error only, given they were primarily from two gene regions (*PGPEP1* and *GDF15*). To assess the robustness of results, we also included sensitivity analyses.

1. Validation of association with CAD using the UK Biobank, and breast cancer using the Breast Cancer Association Consortium As a validation, we examined the association of GDF-15 with CAD using the UK Biobank, and with breast cancer using the Breast Cancer Association Consortium (BCAC) given the differences in study design, sample selection and analytic strategy between the disease specific GWAS and the UK Biobank. We used SAIGE UK Biobank summary statistics. The PheWAS code for CAD is 411 (cases = 31,355, controls = 377,103) [13]. Genetic associations with breast cancer were obtained from a large GWAS of breast cancer, including BCAC, DRIVE and iCOGS (cases = 122,977, controls = 105,974), with imputation based on the 1000 Genomes Project Phase 3 panels [26]. These GWAS adjusted for up to ten principal components and other study-specific covariates

(e.g. study and country). For breast cancer we only included associations from women of European descent.

2. IVW with multiplicative random effects We repeated the analyses using IVW with multiplicative random effects. This model gives the same point estimate as the fixed effect model with the variance of the estimate increased when heterogeneity between SNP-specific Wald estimates is high (a sign of potentially invalid SNPs). The multiplicative random effects model is less sensitive to biases introduced by weaker SNPs than additive random effects [27].

3. Lead SNP analysis We repeated the analyses only using the lead SNP (rs888663, p value with GDF-15: 2.64×10^{-35}) to further rule out the possibility of bias due to inclusion of invalid SNPs.

4. MR-Egger intercept test As another means of identifying potential pleiotropy, we used MR-Egger intercept test, which can be implemented with correlated variants and does not rely on the instrument strength independent of direct effects (InSIDE) assumption [28]. However, we did not perform MR-Egger since the InSIDE assumption is unlikely to be satisfied with correlated instruments and hence estimates from MR-Egger are likely biased [28].

5. Mendelian randomisation restricted to instruments from the same gene region We also repeated the analyses using SNPs from each gene region separately. Consistent estimates using SNPs from different gene regions would provide more convincing evidence concerning the causal role of GDF-15.

Power calculation In the GDF-15 GWAS the variance explained by all 9 SNPs was 21.5% [10]; how much of that variance is explained by the 5 SNPs used here is unclear. We provide power calculations assuming the SNPs explained 21% of the variance and more conservatively 15% of the variance (ESM Table 2) [29]. Assuming 15% of the variance was explained by the SNPs, this study is powered to find relatively small effects, such as an OR of 1.04 per SD change in GDF-15 for CAD using CARDIoGRAMplusC4D 1000 Genomes-based GWAS, and 0.024 per SD change for lipids.

All analyses were performed using R Version 3.3.3 (R Development Core Team, Vienna, Austria) with the R packages (TwoSampleMR) and (MendelianRandomization) [24, 25].

Ethics approval This study only used publicly available data and hence no ethics approval was required.

Results

One SNP (rs1227731) from *GDF15* and 4 SNPs (rs888663, rs749451, rs3195944 and rs17725099) from *PGPEP1* were

used in the main analysis, as shown in ESM Table 1, and corresponding correlation matrix in ESM Table 3. ESM Table 4 and ESM Figs. 1–5 show the associations of the genetic variants with GDF-15 and with the outcomes. Figure 1 shows the selection process for the GDF-15 SNPs and the GWAS used for the outcomes. The mean F statistic for the instruments was 108, suggesting weak instrument bias is unlikely.

Table 1 shows the relation of GDF-15 with CAD and its risk factors using IVW with fixed effects. Higher GDF-15 was associated with lower CAD risk (OR 0.93 per SD increase, 95% CI 0.87, 0.99). However, the association was not evident in the UK Biobank (OR 0.99 per SD increase, 95% CI 0.93, 1.04). GDF-15 was unrelated to type 2 diabetes risk (OR 1.02 per SD increase, 95% CI 0.98, 1.06), HbA$_{1c}$ (0.084 per SD increase, 95% CI −0.046, 0.215) or glucose (0.004 per SD increase, 95% CI −0.023, 0.031). Higher GDF-15 was associated with lower HDL-cholesterol (−0.05 per SD increase, 95% CI −0.09, −0.02), but not with LDL-cholesterol, BMI, SBP or DBP. Higher GDF-15 was associated with lower breast cancer risk (OR 0.89 per SD increase, 95% CI 0.82, 0.96) but the association was non-significant for colorectal cancer risk (OR 0.91 per SD increase, 95% CI 0.80, 1.04) albeit directionally similar. The estimate for breast cancer risk using BCAC was smaller than the UK Biobank estimate and was non-significant (OR 0.97 per SD increase, 95% CI 0.94, 1.01).

ESM Table 5 shows the relation of GDF-15 with the outcomes using IVW with multiplicative random effects, which gave similar results as with fixed effects although with wider confidence intervals.

Table 2 shows the relation of GDF-15 with the outcomes using the lead SNP (rs888663) only. When compared with the main results, the estimates were most consistent for CAD and breast cancer, with directionally similar estimates for colorectal cancer. The associations with other outcomes were non-significant.

The MR-Egger intercept test did not provide strong evidence of potential violation of the exclusion-restriction assumption, except for DBP ($p = 0.046$), BMI ($p < 0.001$) and HbA$_{1c}$ ($p = 0.009$), as shown in ESM Table 6.

ESM Table 7 shows the Mendelian randomisation analysis restricted to instruments from the same gene region. The majority of estimates were directionally consistent, with the association of GDF-15 with lower HDL-cholesterol most evident.

Discussion

This is one of the first Mendelian randomisation studies examining the role of GDF-15 in CAD, breast and colorectal cancer, as well as CAD risk factors, type 2 diabetes and glycaemic traits. Although GDF-15 might have potential utility in prediction, as shown in previous observational studies, our study suggests it is unlikely that GDF-15 increases the risk
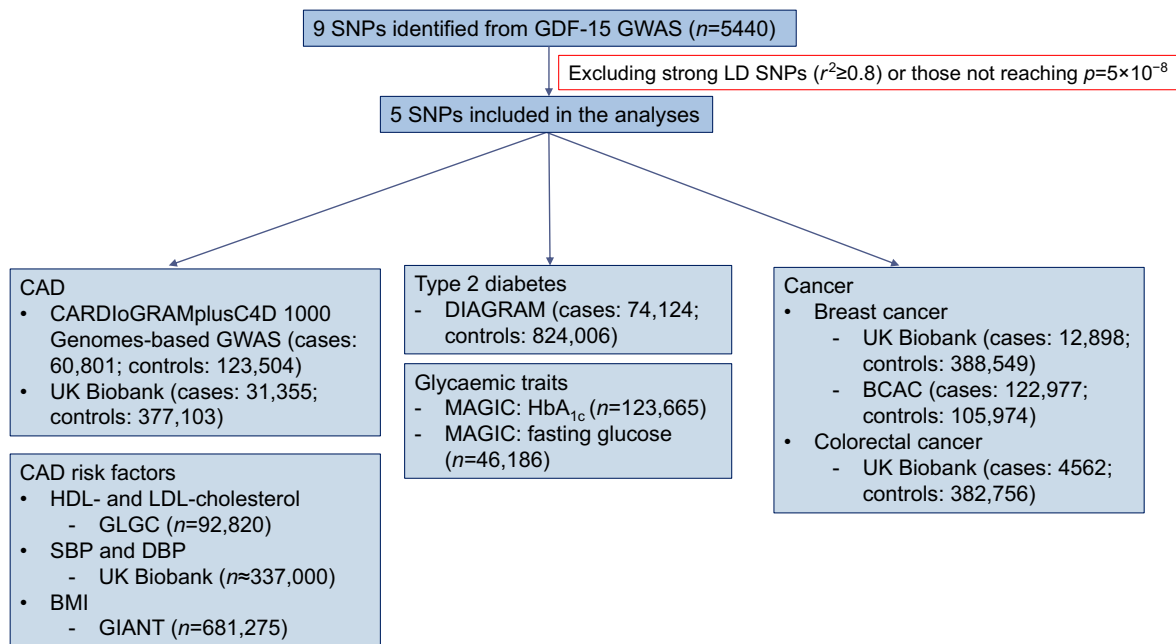
Fig. 1 Selection process of SNPs for GDF-15 and the data sources used. SNPs associated with GDF-15 were selected from a GWAS and used to analyse genetic associations with the outcomes shown, using data from various studies, as detailed in the flow chart. *BCAC*, Breast Cancer Association Consortium; *BMI*; Body mass index; *DBP*, Diastolic blood pressure; *DIAGRAM*, DIAbetes Genetics Replication and meta-analysis; *GIANT*, Genetic Investigation of ANthropometric traits; *GLGC*, Global Lipids Genetic Consortium; *GWAS*, Genome wide association studies; *HDL*, High density lipoprotein; *LD*, Linkage disequilibrium; *LDL*, Low density lipoprotein; *MAGIC*, Meta-Analyses of Glucose and Insulin-related traits Consortium; *SBP*, Systolic blood pressure

of CAD, breast cancer and colorectal cancer [6, 7]. Although our main analysis suggested a potential inverse association of

GDF-15 with CAD and breast cancer risk, these findings were not consistently observed in the validation studies. Our study, therefore, does not support that GDF-15, a suggested

Table 1 The impact of GDF-15 on CAD risk, type 2 diabetes and glycaemic traits, CAD risk factors, breast and colorectal cancer risk using Mendelian randomisation

Outcome	Data source	IVW with fixed effects	
		Per SD increase GDF-15	
		OR/β	95% CI
CAD	CARDIoGRAM	0.93	0.87, 0.99
CAD	UK Biobank	0.99	0.93, 1.04
Type 2 diabetes	DIAGRAM	1.02	0.98, 1.06
HbA$_{1c}$ (mmol/mol)	MAGIC	0.084	−0.046, 0.215
Fasting glucose (mmol/l)	MAGIC	0.004	−0.023, 0.031
SBP (mmHg)	UK Biobank	−0.09	−0.35, 0.18
DBP (mmHg)	UK Biobank	0.10	−0.05, 0.25
LDL-C (per SD change)	GLGC	−0.01	−0.05, 0.03
HDL-C (per SD change)	GLGC	−0.05	−0.09, −0.02
BMI (per SD change)	GIANT	0.002	−0.009, 0.01
Breast cancer	UK Biobank	0.89	0.82, 0.96
Breast cancer	BCAC	0.97	0.94, 1.01
Colorectal cancer	UK Biobank	0.91	0.80, 1.04

CARDIoGRAM, CARDIoGRAMplusC4D 1000 Genomes-based GWAS; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol

Table 2 The impact of GDF-15 (per SD change) on CAD, type 2 diabetes and glycaemic traits, CAD risk factors, breast and colorectal cancer risk using Mendelian randomisation with the lead SNP (rs888663) as the instrument

Outcome	Data source	Wald estimate	
		OR/β	95% CI
CAD	CARDIoGRAM	0.87	0.81, 0.95
CAD	UK Biobank	0.93	0.87, 1.00
Type 2 diabetes	DIAGRAM	1.00	0.96, 1.06
HbA$_{1c}$ (mmol/mol)	MAGIC	0.0036	−0.167, 0.174
Fasting glucose (mmol/l)	MAGIC	0.006	−0.031, 0.042
SBP (mmHg)	UK Biobank	−0.34	−0.69, 0.006
DBP (mmHg)	UK Biobank	0.005	−0.19, 0.20
LDL-C (per SD change)	GLGC	−0.03	−0.08, 0.02
HDL-C (per SD change)	GLGC	−0.03	−0.07, 0.02
BMI (per SD change)	GIANT	−0.010	−0.024, 0.004
Breast cancer	UK Biobank	0.80	0.72, 0.89
Breast cancer	BCAC	0.99	0.94, 1.04
Colorectal cancer	UK Biobank	0.88	0.74, 1.05

CARDIoGRAM, CARDIoGRAMplusC4D 1000 Genomes-based GWAS; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol

biomarker for metformin use, is causally associated with CAD, breast or colorectal cancer.

Previous observational studies mainly found higher GDF-15 associated with higher CVD risk and cancer risk [30–32]. However, increased GDF-15 could be a symptom of disease, i.e. reverse causation, where the disease increases GDF-15. Some in vitro studies suggest GDF-15 is overexpressed in cancer and myocardial GDF-15 mRNA is induced rapidly on ischaemic injury [33, 34]. Differences in findings between our study and previous studies could have occurred for several reasons. GDF-15 could simply be an early marker of underlying disease, a reflection of confounding by ill-health [5], or a consequence of selection or survivor bias in studies of older people [35]. Mechanistically, GDF-15 could reduce CAD risk via anti-inflammatory and/or anti-hypertrophic effects although this is inconsistent with our findings [36]. Animal studies suggest increases in GDF-15 may result in weight loss via regulation of appetite, although these findings were not always consistent and may depend on other conditions such as interactions with other receptors [37], which may explain the lack of association with BMI in our study. GDF-15 might have different effects on cancer depending on the cancer stage which needs to be clarified and the underlying mechanisms elucidated in future studies [38]. Our study also suggests GDF-15 is unrelated to type 2 diabetes and glycaemic traits, which is inconsistent with previous observational studies [39]. However, these studies could be confounded by other metabolic markers, as well as by ill-health status. Given GDF-15 is a strong correlate of metformin use, it is also unlikely that increases in GDF-15 lead to type 2 diabetes since this is inconsistent with the effect of metformin on glycaemic traits [40].

Overall, our study does not support that increased GDF-15 level is a potential explanation for the observed inverse association between metformin use and CAD or some cancers. Alternative pathways are in need of further investigation. For example, metformin decreases plasminogen-activator inhibitor type 1 (PAI-1) [41], and sex hormones [42], which are potential causes of CAD [43–45], as well as some cancers [46]. Future studies should further systematically explore other potential pathways by which metformin may have an impact such as via detailed metabolomic analyses using Mendelian randomisation, which may be useful in identifying new interventions to treat these diseases. Similar approaches have already been used for statins [47], which are more efficacious than other lipid lowering drugs in reducing CAD [48].

This is one of the first studies to evaluate the impact of GDF-15, a biomarker for metformin, on CAD and cancer risk using Mendelian randomisation, which is less susceptible to confounding than observational studies, nevertheless limitations exist. First, the validity of the study depends on the chosen genetic instruments. The instruments strongly predicted GDF-15 and were unlikely confounded (with no evidence of association with Townsend deprivation index, age completed full time education or ever smoking in the UK Biobank [$p > 0.08$]) [21, 24]. We cannot rule out the possibility of violation of the exclusion-restriction assumption. For example, rs888663 is associated with CAD whilst rs1227731 and rs3195944 are associated with erythrocytes [49, 50]. However, it is unclear whether these pleiotropic effects are downstream effects of GDF-15 (which does not violate the exclusion-restriction assumption) or effects independent of GDF-15 (which violates the assumption). We were unable to use weighted median or MR-Egger estimates because of the use of correlated SNPs. Although the MR-Egger intercept test suggested no violation of the exclusion-restriction assumptions for the main outcomes (CAD, breast and colorectal cancer), there was some evidence of potential violation for DBP, BMI and HbA$_{1c}$. As such, our study needs confirmation when more genetic instruments for GDF-15 from different gene regions have been identified in a larger GWAS. Second, we were unable to directly evaluate the impact of metformin on CAD and cancer risk, which would require individual genetic data (e.g. UK Biobank) [14] and definitive targets of metformin (e.g. AMP-activated protein kinase). Such Mendelian randomisation studies would be able to emulate more fully the impact of metformin. Third, some participants in the UK Biobank were taking antihypertensives, so the observed BP measurements for these participants were underestimated, leading to misclassification. Fourth, the identified genetic instruments may not be the true genetic signals for GDF-15 (i.e. false positives), leading to potential weak instrument bias. However, this is unlikely given the mean F statistic for the instruments used was much greater than 10. Fifth, our study was unable to identify if exogenous (e.g. metformin) and endogenous GDF-15 (naturally occurring) have different effects, which could be answered using other designs, such as regression discontinuity applied to comprehensive electronic health records. Sixth, the main analyses showed a potential inverse association of GDF-15 with CAD and breast cancer but these findings were not consistently observed in the validation studies. Such differences may be driven by methodological differences generating potential biases, such as healthy volunteers in the UK Biobank or the lack of age adjustment in some of the studies within BCAC since some of them had few or no controls [26]. However, these differences also highlight the importance of validation, whenever possible, for Mendelian randomisation. Seventh, although it is possible that GDF-15 is influenced by previous health status, i.e. reverse causation, we were unable to examine this possibility using a bi-directional Mendelian randomisation design as we do not have access to genetic summary statistics for GDF-15. Finally, we were unable to assess differences by age or sex, because sufficiently large age- or sex-specific studies are not available. We also had to assume a linear relation of GDF-15 with the outcomes.

There is no convincing evidence that GDF-15 may reduce risk of CAD, breast or colorectal cancer. Whether the inverse

association of metformin use with cancer risk is an artefact of biases or acts via other unexplored mechanistic pathways warrants future investigation, to help validate whether this medication can be repurposed for cancer prevention.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement SLAY and CMS designed the study, and wrote the analysis plan. SLAY undertook analyses with feedback from SL and CMS, and all authors interpreted the results in the study. SLAY wrote the first draft of the manuscript with critical feedback and revisions from SL and CMS. All authors gave final approval of the version to be published. SLAY is the guarantor of this work.

References

1. Douros A, Dell'Aniello S, Yu OHY, Filion KB, Azoulay L, Suissa S (2018) Sulfonylureas as second line drugs in type 2 diabetes and the risk of cardiovascular and hypoglycaemic events: population based cohort study. BMJ 362:k2693. https://doi.org/10.1136/bmj.k2693

2. Griffin SJ, Leaver JK, Irving GJ (2017) Impact of metformin on cardiovascular disease: a meta-analysis of randomised trials among people with type 2 diabetes. Diabetologia 60(9):1620–1629. https://doi.org/10.1007/s00125-017-4337-9

3. Suissa S, Azoulay L (2012) Metformin and the risk of cancer: time-related biases in observational studies. Diabetes Care 35(12):2665–2673. https://doi.org/10.2337/dc12-0788

4. Gerstein HC, Pare G, Hess S et al (2017) Growth differentiation factor 15 as a novel biomarker for metformin. Diabetes Care 40(2):280–283. https://doi.org/10.2337/dc16-1682

5. Natali A, Nesti L, Venturi E et al (2018) Metformin is the key factor in elevated plasma growth differentiation factor-15 levels in type 2 diabetes: a nested, case-control study. Diabetes Obes Metab 21(2):412–416. https://doi.org/10.1111/dom.13519

6. Taddei S, Virdis A (2010) Growth differentiation factor-15 and cardiovascular dysfunction and disease: malefactor or innocent bystander? Eur Heart J 31(10):1168–1171. https://doi.org/10.1093/eurheartj/ehq077

7. Li C, Wang X, Casal I et al (2016) Growth differentiation factor 15 is a promising diagnostic and prognostic biomarker in colorectal cancer. J Cell Mol Med 20(8):1420–1426. https://doi.org/10.1111/jcmm.12830

8. Holmes MV, Ala-Korpela M, Davey Smith G (2017) Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol 14(10):577–590. https://doi.org/10.1038/nrcardio.2017.78

9. Cheung CL, Tan KCB, Au PCM, Li GHY, Cheung BMY (2019) Evaluation of GDF15 as a therapeutic target of cardiometabolic diseases in human: a Mendelian randomization study. EBioMedicine. 41:85–90. https://doi.org/10.1016/j.ebiom.2019.02.021

10. Jiang J, Thalamuthu A, Ho JE et al (2018) A meta-analysis of genome-wide association studies of growth differentiation factor-15 concentration in blood. Front Genet 9:97. https://doi.org/10.3389/fgene.2018.00097

11. Burgess S, Dudbridge F, Thompson SG (2016) Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med 35(11):1880–1906. https://doi.org/10.1002/sim.6835

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(10):1121–1130. https://doi.org/10.1038/ng.3396

13. Zhou W, Nielsen JB, Fritsche LG et al (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 50(9):1335–1341. https://doi.org/10.1038/s41588-018-0184-y

14. Sudlow C, Gallacher J, Allen N et al (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12(3):e1001779. https://doi.org/10.1371/journal.pmed.1001779

15. Denny JC, Bastarache L, Ritchie MD et al (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 31(12):1102–1110. https://doi.org/10.1038/nbt.2749

16. Mahajan A, Taliun D, Thurner M et al (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50(11):1505–1513. https://doi.org/10.1038/s41588-018-0241-6

17. Wheeler E, Leong A, Liu CT et al (2017) Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med 14(9):e1002383. https://doi.org/10.1371/journal.pmed.1002383

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

19. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM et al (2013) Discovery and refinement of loci associated with lipid

levels. Nat Genet 45(11):1274–1283. https://doi.org/10.1038/ng.2797

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

21. The Neale Lab UK Biobank GWAS. Available from http://www.nealelab.is/uk-biobank/ [Accessed: 22 Oct, 2018]

22. Rosner B (1995) Fundamentals of biostatistics, 4th edn. Duxbury Press, Belmont, CA

23. Pierce BL, Ahsan H, Vanderweele TJ (2011) Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. Int J Epidemiol 40(3):740–752. https://doi.org/10.1093/ije/dyq151

24. Hemani G, Zheng J, Elsworth B, et al. (2018) The MR-Base platform supports systematic causal inference across the human phenome. Elife 7. https://doi.org/10.7554/eLife.34408

25. Yavorska OO, Burgess S (2017) MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol 46(6):1734–1739. https://doi.org/10.1093/ije/dyx034

26. Michailidou K, Lindstrom S, Dennis J et al (2017) Association analysis identifies 65 new breast cancer risk loci. Nature 551(7678):92–94. https://doi.org/10.1038/nature24284

27. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J (2017) A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Stat Med 36(11):1783–1802. https://doi.org/10.1002/sim.7221

28. Burgess S, Thompson SG (2017) Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol 32(5):377–389. https://doi.org/10.1007/s10654-017-0255-x

29. Burgess S (2014) Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol 43(3):922–929. https://doi.org/10.1093/ije/dyu005

30. Andersson C, Enserro D, Sullivan L et al (2016) Relations of circulating GDF-15, soluble ST2, and troponin-I concentrations with vascular function in the community: The Framingham Heart Study. Atherosclerosis 248:245–251. https://doi.org/10.1016/j.atherosclerosis.2016.02.013

31. Wallentin L, Hijazi Z, Andersson U et al (2014) Growth differentiation factor 15, a marker of oxidative stress and inflammation, for risk assessment in patients with atrial fibrillation: insights from the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial. Circulation 130(21):1847–1858. https://doi.org/10.1161/CIRCULATIONAHA.114.011204

32. Pavo N, Wurm R, Neuhold S et al (2016) GDF-15 is associated with cancer incidence in patients with type 2 diabetes. Clin Chem 62(12):1612–1620. https://doi.org/10.1373/clinchem.2016.257212

33. Liu T, Bauskin AR, Zaunders J et al (2003) Macrophage inhibitory cytokine 1 reduces cell adhesion and induces apoptosis in prostate cancer cells. Cancer Res 63(16):5034–5040

34. Kempf T, Eden M, Strelau J et al (2006) The transforming growth factor-beta superfamily member growth-differentiation factor-15 protects the heart from ischemia/reperfusion injury. Circ Res 98(3):351–360. https://doi.org/10.1161/01.RES.0000202805.73038.48

35. Daniels LB, Clopton P, Laughlin GA, Maisel AS, Barrett-Connor E (2011) Growth-differentiation factor-15 is a robust, independent predictor of 11-year mortality risk in community-dwelling older adults: the Rancho Bernardo Study. Circulation 123(19):2101–2110. https://doi.org/10.1161/CIRCULATIONAHA.110.979740

36. Ago T, Sadoshima J (2006) GDF15, a cardioprotective TGF-beta superfamily protein. Circ Res 98(3):294–297. https://doi.org/10.1161/01.RES.0000207919.83894.9d

37. Mullican SE, Lin-Schmidt X, Chin CN et al (2017) GFRAL is the receptor for GDF15 and the ligand promotes weight loss in mice and nonhuman primates. Nat Med 23(10):1150–1157. https://doi.org/10.1038/nm.4392

38. Bauskin AR, Brown DA, Kuffner T et al (2006) Role of macrophage inhibitory cytokine-1 in tumorigenesis and diagnosis of cancer. Cancer Res 66(10):4983–4986. https://doi.org/10.1158/0008-5472.CAN-05-4067

39. Bao X, Borne Y, Muhammad IF et al (2019) Growth differentiation factor 15 is positively associated with incidence of diabetes mellitus: the Malmo Diet and Cancer-Cardiovascular Cohort. Diabetologia. 62(1):78–86. https://doi.org/10.1007/s00125-018-4751-7

40. Hirst JA, Farmer AJ, Ali R, Roberts NW, Stevens RJ (2012) Quantifying the effect of metformin treatment and dose on glycemic control. Diabetes Care 35(2):446–454. https://doi.org/10.2337/dc11-1465

41. Nagi DK, Yudkin JS (1993) Effects of metformin on insulin resistance, risk factors for cardiovascular disease, and plasminogen activator inhibitor in NIDDM subjects. A study of two ethnic groups. Diabetes Care 16(4):621–629. https://doi.org/10.2337/diacare.16.4.621

42. Campagnoli C, Berrino F, Venturelli E et al (2013) Metformin decreases circulating androgen and estrogen levels in nondiabetic women with breast cancer. Clin Breast Cancer 13(6):433–438. https://doi.org/10.1016/j.clbc.2013.08.012

43. Song C, Burgess S, Eicher JD, O'Donnell CJ, Johnson AD (2017) Causal effect of plasminogen activator inhibitor type 1 on coronary heart disease. J Am Heart Assoc 6(6). https://doi.org/10.1161/JAHA.116.004918

44. Zhao JV, Schooling CM (2018) Coagulation factors and the risk of ischemic heart disease: a Mendelian randomization study. Circ Genom Precis Med 11(1):e001956. https://doi.org/10.1161/CIRCGEN.117.001956

45. Schooling CM, Luo S, Au Yeung SL et al (2018) Genetic predictors of testosterone and their associations with cardiovascular disease and risk factors: A Mendelian randomization investigation. Int J Cardiol 267:171–176. https://doi.org/10.1016/j.ijcard.2018.05.051

46. Chlebowski RT, Hendrix SL, Langer RD et al (2003) Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. JAMA 289(24):3243–3253. https://doi.org/10.1001/jama.289.24.3243

47. Wurtz P, Wang Q, Soininen P et al (2016) Metabolomic profiling of statin use and genetic inhibition of HMG-CoA reductase. J Am Coll Cardiol 67(10):1200–1210. https://doi.org/10.1016/j.jacc.2015.12.060

48. Schooling CM, Au Yeung SL, Leung GM (2014) Why do statins reduce cardiovascular disease more than other lipid modulating therapies? Eur J Clin Investig 44(11):1135–1140. https://doi.org/10.1111/eci.12342

49. Astle WJ, Elding H, Jiang T et al (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. Cell 167(5):1415–1429 e1419. https://doi.org/10.1016/j.cell.2016.10.042

50. van der Harst P, Verweij N (2018) Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ Res 122(3):433–443. https://doi.org/10.1161/CIRCRESAHA.117.312086