Research article

The impact of growth differentiation factor 15 (GDF-15) on coronary artery disease, breast and colorectal cancer risk: A Mendelian randomization study

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

[1]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, People's Republic of China.

[2]City University of New York, Graduate School of Public Health and Health Policy, New York, United States of America.

*Corresponding author:

Dr Shiu Lun Au Yeung,

School of Public Health,

Li Ka Shing Faculty of Medicine,

The University of Hong Kong,

7 Sassoon Road, Pokfulam,

Hong Kong SAR, China

Email: ayslryan@hku.hk

Phone: 852-3917 6740

Fax: 852-3520 1945

Shan Luo: luoshan@hku.hk

CM Schooling: cms1@hku.hk

Abstract: 329

Word count: 3,764

Tables: 2

Figure: 1

Electronic supplementary material: 1

Running title: GDF-15 on CAD and cancer using MR

Tweet: This Mendelian randomization study does not provide convincing evidence that growth differentiation factor 15 relates to risk of coronary artery disease, breast or colorectal cancer.

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 randomization study to examine the role of GDF-15 in risk of coronary artery disease (CAD), breast and colorectal cancer. Secondary analyses included examination of the association of GDF-15 with type 2 diabetes, glycaemic traits, blood pressure, lipids and body mass index.

Methods: We obtained single nucleotide polymorphisms (SNPs) strongly (p value < 5 x 10^{-8}) predicting GDF-15 from a genome wide association study (GWAS) (n=5,440) and applied them to genetic studies of CAD (CARDIoGRAMplusC4D 1000 Genomes-based GWAS (n=184,305)), type 2 diabetes (DIAbetes Genetics Replication And Meta-analysis (n=898,130)), glycaemic traits (the Meta-Analyses of Glucose and Insulin-related traits Consortium (HbA1c: n=123,665; fasting glucose: n= 46,186)), blood pressure, breast cancer and colorectal cancer (UK Biobank (n<=401,447)), lipids (Global Lipids Genetic Consortium (n<=92,820)) and adiposity (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 (BCAC).

Results: Using 5 SNPs, increased GDF-15 was associated with lower CAD (odds ratio (OR) 0.93 per standard deviation (SD), 95% confidence interval (CI) 0.87 to 0.99) and breast cancer (OR 0.89 per SD, 95% CI 0.82 to 0.96), with similar results from lead SNP analysis. However, the associations with CAD (OR 0.99 per SD increase, 95% CI 0.93 to 1.04) and breast cancer (OR 0.97 per SD increase, 95% CI 0.94 to 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, breast or colorectal cancer. Whether the observed inverse association of metformin use with cancer risk is via other unexplored mechanistic pathways warrants further investigation.

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-glycemic effects of metformin on cardiovascular disease and cancer

What is the key question?

- To investigate the causal role of GDF-15 in risk of coronary artery disease (CAD), colorectal cancer and breast cancer using Mendelian randomization

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, breast or colorectal cancer. Whether the inverse association of metformin use with cancer is an artefact of biases or acts via other unexplored mechanistic pathways warrants further investigation, which would help validate whether this medication could be repurposed for cancer prevention.

Introduction

Metformin is the first line medication for diabetes, given its superiority and safety compared to other diabetes medications (e.g. sulfonylureas),[1] which suggests metformin reduces glycaemia via potentially less hazardous pathways than other medications.[1] Some randomized controlled trials 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 sulphonylurea 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 HbA1c 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 randomization studies are increasingly used to ascertain causes of disease since they utilize 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 2 single nucleotide polymorphisms (SNPs) (rs62122429 and rs62122430) relevant to GDF-15 with CAD in CARDIoGRAM and found rs62122429 associated with CAD whilst rs62122430 was not.[4] This finding suggests GDF-15 may have a role in CAD development although a recent Mendelian randomization 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 metfomin's potential anti-cancer property.[9] Here, we conducted Mendelian randomization to evaluate the impact of GDF-15on CAD and cancer (breast and colorectal) risk. For completeness, we also assessed its relation with type 2 diabetes and glycaemic traits (HbA1c and fasting glucose), given GDF-15 is strongly related to metformin use, and other established CAD risk factors (systolic and diastolic blood pressure, high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL) cholesterol, and body mass index (BMI)).

Method

Study design

This is a 2 sample Mendelian randomization study using summary statistics from genome wide association studies (GWAS). It has 3 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. Thirdly, 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 standard deviations (SD). The genetic predictors of GDF-15 were extracted from a GWAS consisting of 4 cohorts (n=5,440) 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 Aging 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, systolic blood pressure, antihypertensive medication use, diabetes mellitus, and 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] ESM table 1 shows the 9 correlated SNPs, identified in the GWAS, which explained 21.5% of the variance of GDF-15. Six (6) 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, 5 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, HbA1c and glucose, systolic and diastolic blood pressure (mmHg), LDL cholesterol (SD), and HDL cholesterol (SD), and BMI (SD). The studies used are described below.

Genetic associations with coronary artery disease

Genetic associations with CAD were obtained from CARDIoGRAMplusC4D 1000 Genomes-based GWAS, a meta-analysis of GWAS of CAD studies (cases=60,801, non-cases=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 by-pass 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≤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 (4,562 cases, 382,756 non-cases) and breast cancer (12,898 cases,

388,549 non-cases) in the UK Biobank were defined using UK Biobank 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 4 principal

components.

Genetic associations with type 2 diabetes

Genetic association 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 anti-diabetic mediation, and self-report.

Imputation was performed based on the HRC reference panel for all studies expect 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 HbA1c and glucose

Genetic associations with HbA1c and glucose were obtained from the Meta-Analysis of Glucose

and Insulin-related traits consortium (MAGIC), a meta-analysis of GWAS of HbA1c 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 HbA1c 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 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 (Utah residents of Northern and Western European Ancestry) 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 hours 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 body mass index

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.) 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 systolic and diastolic blood pressure

Genetic associations with systolic and diastolic blood pressure (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.[22] Seated blood pressure was measured using the Omron HRM-7015IT digital

BP monitor. Genetic associations with blood pressure (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.[23,24] 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. [25,26] 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 Breast Cancer Association Consortium (BCAC)

As a validation, we examined the association of GDF-15 with CAD using the UK Biobank, and with breast cancer using 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, non-cases=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.[27] These GWAS adjusted for up to 10 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. *Inverse variance weighting 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.[28]

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.[29] 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.[29]

5. Mendelian randomization 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).[30] Assuming 15% of the variance explained, this study is powered to find
relatively small effects, such as an odds ratio 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).[25,26]

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 figures 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 (odds ratio (OR) 0.93 per SD increase, 95% confidence interval (CI) 0.87 to 0.99). However, the association was not evident in the UK Biobank (OR 0.99 per SD increase, 95% CI 0.93 to 1.04). GDF-15 was unrelated to type 2 diabetes risk (OR 1.02 per SD, 95%CI 0.98 to 1.06), HbA1c (0.084 per SD, 95%CI -0.046 to 0.215) or glucose (0.004 per SD, 95%CI -0.023 to 0.031). Higher GDF-15 was associated with lower HDL cholesterol (-0.05 per SD increase, 95% CI -0.09 to -0.02), but not with LDL cholesterol, BMI, systolic blood pressure (SBP) or diastolic blood pressure (DBP). Higher GDF-

15 was associated with lower breast cancer risk (OR 0.89 per SD increase, 95% CI 0.82 to 0.96) but the association was non-significant for colorectal cancer risk (OR 0.91 per SD increase, 95% CI 0.80 to 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 to 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 IV exclusion-restriction assumption, except for DBP (p value: 0.046), BMI (p value <0.001) and HbA1c (p value: 0.009), as shown in ESM table 6.

ESM table 7 shows the Mendelian randomization analysis restricted to instruments from the same gene region. 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 randomization 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 GDF-15 unlikely increases the risk of CAD, breast cancer, and colorectal cancer.[6,7] Although our main analysis suggested a potential inverse association of GDF-15 wih CAD and breast cancer risk, these findings were not consistently observed in the validation studies. Our study does therefore not support that GDF-15, a suggested 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.[31-33] 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 ischemic injury.[34,35] 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.[36] Mechanistically, GDF-15 could reduce CAD risk via anti-inflammatory and/or anti-hypertrophic effects although this is inconsistent with our findings.[37] 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,[38] 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.[39] Our study also suggests GDF-15 is

unrelated to type 2 diabetes and glycaemic traits, which is inconsistent with previous

observational studies.[40] 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.[41]

Overall, our study does not support that increased GDF-15 levels 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),[42] and sex hormones,[43] which are potential causes of CAD,[44-46]

as well as some cancers.[47] Future studies should further systematically explore other potential

pathways by which metformin may have an impact such as via detailed metabolomic analyses

using Mendelian randomization, which may be useful in identifying new interventions to treat

these diseases. Similar approaches have already been used for statins,[48] which are more

efficacious than other lipid lowering drugs in reducing CAD.[49]

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 randomization which is less susceptible to confounding,

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 value >0.08)).[22,25] 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 red blood cells.[50,51] 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 IV exclusion-restriction assumptions for the main outcomes (CAD, breast and colorectal cancer), there was some evidence of potential violation for DBP, BMI and HbA1c. 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 randomization 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 blood pressure 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.[27] However, these differences also highlight the importance of validation, whenever possible, for Mendelian randomization. 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 randomization design as we do not have access to genetic summary statistics for GDF-15. Lastly, 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.

Data availability

All data used were publicly available, the data sources can be found under Acknowledgements and the references.

Funding

Duality of interest

The authors declared no conflict of interest, financial or otherwise.

Contribution statement

SLAY and CMS designed the study, wrote the analysis plan and interpreted the results. SLAY undertook analyses with feedback from SL and CMS. 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. Sulfonylureas as second line drugs in type 2 diabetes and the risk of cardiovascular and hypoglycaemic events: population based cohort study. *BMJ.* 2018;362:k2693.
2. Griffin SJ, Leaver JK, Irving GJ. Impact of metformin on cardiovascular disease: a meta-analysis of randomised trials among people with type 2 diabetes. *Diabetologia.* 2017;60(9):1620-1629.
3. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care.* 2012;35(12):2665-2673.
4. Gerstein HC, Pare G, Hess S, et al. Growth Differentiation Factor 15 as a Novel Biomarker for Metformin. *Diabetes Care.* 2017;40(2):280-283.
5. Natali A, Nesti L, Venturi E, et al. 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.* 2018.
6. Taddei S, Virdis A. Growth differentiation factor-15 and cardiovascular dysfunction and disease: malefactor or innocent bystander? *Eur Heart J.* 2010;31(10):1168-1171.
7. Li C, Wang X, Casal I, et al. Growth differentiation factor 15 is a promising diagnostic and prognostic biomarker in colorectal cancer. *J Cell Mol Med.* 2016;20(8):1420-1426.
8. Holmes MV, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol.* 2017;14(10):577-590.
9. Cheung CL, Tan KCB, Au PCM, Li GHY, Cheung BMY. Evaluation of GDF15 as a therapeutic target of cardiometabolic diseases in human: A Mendelian randomization study. *EBioMedicine.* 2019.
10. Jiang J, Thalamuthu A, Ho JE, et al. A Meta-Analysis of Genome-Wide Association Studies of Growth Differentiation Factor-15 Concentration in Blood. *Front Genet.* 2018;9:97.
11. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med.* 2016;35(11):1880-1906.

12. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47(10):1121-1130.
13. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet.* 2018.
14. Sudlow C, Gallacher J, Allen N, et al. 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.
15. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol.* 2013;31(12):1102-1110.
16. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;50(11):1505-1513.
17. Wheeler E, Leong A, Liu CT, et al. 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.* 2017;14(9):e1002383.
18. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet.* 2010;42(2):105-116.
19. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45(11):1274-1283.
20. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet.* 2018;27(20):3641-3649.
21. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518(7538):197-206.
22. The Neale Lab. UK Biobank GWAS. http://www.nealelab.is/uk-biobank/. Accessed.
23. Rosner B. *Fundamentals of biostatistics.* 4th ed. Belmont, Calif.: Duxbury Press; 1995.
24. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *International Journal of Epidemiology.* 2011;40(3):740-752.
25. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 2018;7.
26. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol.* 2017;46(6):1734-1739.
27. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92-94.
28. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med.* 2017;36(11):1783-1802.
29. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol.* 2017;32(5):377-389.
30. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol.* 2014;43(3):922-929.
31. Andersson C, Enserro D, Sullivan L, et al. Relations of circulating GDF-15, soluble ST2, and troponin-I concentrations with vascular function in the community: The Framingham Heart Study. *Atherosclerosis.* 2016;248:245-251.
32. Wallentin L, Hijazi Z, Andersson U, et al. 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.* 2014;130(21):1847-1858.

33. Pavo N, Wurm R, Neuhold S, et al. GDF-15 Is Associated with Cancer Incidence in Patients with Type 2 Diabetes. *Clin Chem.* 2016;62(12):1612-1620.

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

35. Kempf T, Eden M, Strelau J, et al. The transforming growth factor-beta superfamily member growth-differentiation factor-15 protects the heart from ischemia/reperfusion injury. *Circ Res.* 2006;98(3):351-360.

36. Daniels LB, Clopton P, Laughlin GA, Maisel AS, Barrett-Connor E. Growth-differentiation factor-15 is a robust, independent predictor of 11-year mortality risk in community-dwelling older adults: the Rancho Bernardo Study. *Circulation.* 2011;123(19):2101-2110.

37. Ago T, Sadoshima J. GDF15, a cardioprotective TGF-beta superfamily protein. *Circ Res.* 2006;98(3):294-297.

38. Mullican SE, Lin-Schmidt X, Chin CN, et al. GFRAL is the receptor for GDF15 and the ligand promotes weight loss in mice and nonhuman primates. *Nat Med.* 2017;23(10):1150-1157.

39. Bauskin AR, Brown DA, Kuffner T, et al. Role of macrophage inhibitory cytokine-1 in tumorigenesis and diagnosis of cancer. *Cancer Res.* 2006;66(10):4983-4986.

40. Bao X, Borne Y, Muhammad IF, et al. Growth differentiation factor 15 is positively associated with incidence of diabetes mellitus: the Malmo Diet and Cancer-Cardiovascular Cohort. *Diabetologia.* 2018.

41. Hirst JA, Farmer AJ, Ali R, Roberts NW, Stevens RJ. Quantifying the effect of metformin treatment and dose on glycemic control. *Diabetes Care.* 2012;35(2):446-454.

42. Nagi DK, Yudkin JS. 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.* 1993;16(4):621-629.

43. Campagnoli C, Berrino F, Venturelli E, et al. Metformin decreases circulating androgen and estrogen levels in nondiabetic women with breast cancer. *Clin Breast Cancer.* 2013;13(6):433-438.

44. Song C, Burgess S, Eicher JD, O'Donnell CJ, Johnson AD. Causal Effect of Plasminogen Activator Inhibitor Type 1 on Coronary Heart Disease. *J Am Heart Assoc.* 2017;6(6).

45. Zhao JV, Schooling CM. Coagulation Factors and the Risk of Ischemic Heart Disease: A Mendelian Randomization Study. *Circ Genom Precis Med.* 2018;11(1):e001956.

46. Schooling CM, Luo S, Au Yeung SL, et al. Genetic predictors of testosterone and their associations with cardiovascular disease and risk factors: A Mendelian randomization investigation. *Int J Cardiol.* 2018;267:171-176.

47. Chlebowski RT, Hendrix SL, Langer RD, et al. Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. *JAMA.* 2003;289(24):3243-3253.

48. Wurtz P, Wang Q, Soininen P, et al. Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J Am Coll Cardiol.* 2016;67(10):1200-1210.

49. Schooling CM, Au Yeung SL, Leung GM. Why do statins reduce cardiovascular disease more than other lipid modulating therapies? *Eur J Clin Invest.* 2014;44(11):1135-1140.

50. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell.* 2016;167(5):1415-1429 e1419.

51. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res.* 2018;122(3):433-443.

Figure legend: Figure 1: Selection process of single nucleotide polymorphisms (SNPs) for growth differentiation factor -15 (GDF-15) and the data sources used

Table 1: The impact of growth differentiation factor 15 (GDF-15) on coronary artery disease (CAD) risk, type 2 diabetes and glycaemic traits, CAD risk factors, breast and colorectal cancer risk using Mendelian randomization

| [a]Outcome | [b]Data source | Inverse variance weighting with fixed effects | |
| | | Per standard deviation increase GDF-15 | |
		Odds ratio/ Beta	95% CI
CAD	CARDIoGRAM	0.93	0.87 to 0.99
CAD	UK Biobank	0.99	0.93 to 1.04
Type 2 diabetes	DIAGRAM	1.02	0.98 to 1.06
HbA1c (mmol/mmol)	MAGIC	0.084	-0.046 to 0.215
Fasting glucose (mmol/L)	MAGIC	0.004	-0.023 to 0.031
SBP (mmHg)	UK Biobank	-0.09	-0.35 to 0.18
DBP (mmHg)	UK Biobank	0.10	-0.05 to 0.25
LDL-C (SD)	Global Lipids Genetic Consortium	-0.01	-0.05 to 0.03
HDL-C (SD)	Global Lipids Genetic Consortium	-0.05	-0.09 to -0.02
BMI (SD)	GIANT	0.002	-0.009 to 0.01
Breast cancer	UK Biobank	0.89	0.82 to 0.96
Breast cancer	BCAC	0.97	0.94 to 1.01
Colorectal cancer	UK Biobank	0.91	0.80 to 1.04

[a] BMI: Body mass index; CI: confidence interval; DBP: Diastolic blood pressure; HDL-C: High density lipoprotein cholesterol; LDL-C: Low density lipoprotein cholesterol; SBP: Systolic blood pressure

[b] BCAC: Breast cancer association consortium; CARDIoGRAM: CARDIoGRAMplusC4D 1000 Genomes-based GWAS; DIAGRAM: DIAbetes Genetics Replication And Meta-analysis; GIANT: Genetic Investigation of ANthropometric Traits; MAGIC: the Meta-Analyses of Glucose and Insulin-related traits Consortium;

Table 2: The impact of growth differentiation factor 15 (per standard deviation (SD)) on coronary artery disease (CAD), type 2 diabetes and glycaemic traits, CAD risk factors, breast and colorectal risk using Mendelian randomization with the lead SNP (rs888663) as the instrument

[a]Outcome	[b]Data source	Wald estimate Odds ratio/Beta	95% Confidence Interval
CAD	CARDIoGRAM	0.87	0.81 to 0.95
CAD	UK Biobank	0.93	0.87 to 1.00
Type 2 diabetes	DIAGRAM	1.00	0.96 to 1.06
HbA1c (mmol/mmol)	MAGIC	0.0036	-0.167 to 0.174
Fasting glucose (mmol/L)	MAGIC	0.006	-0.031 to 0.042
SBP (mmHg)	UK Biobank	-0.34	-0.69 to 0.006
DBP (mmHg)	UK Biobank	0.005	-0.19 to 0.20
LDL-C (SD)	Global Lipids Genetic Consortium	-0.03	-0.08 to 0.02
HDL-C (SD)	Global Lipids Genetic Consortium	-0.03	-0.07 to 0.02
BMI (SD)	Genetic Investigation of ANthropometric Traits	-0.010	-0.024 to 0.004
Breast cancer	UK Biobank	0.80	0.72 to 0.89
Breast cancer	BCAC	0.99	0.94 to 1.04
Colorectal cancer	UK Biobank	0.88	0.74 to 1.05

[a]BMI: Body mass index; DBP: Diastolic blood pressure; HDL-C: High density lipoprotein cholesterol; LDL-C: Low density lipoprotein cholesterol; SBP: Systolic blood pressure

[b] BCAC: Breast cancer association consortium; CARDIoGRAM: CARDIoGRAMplusC4D 1000 Genomes-based GWAS; DIAGRAM: DIAbetes Genetics Replication And Meta-analysis; MAGIC: the Meta-Analyses of Glucose and Insulin-related traits Consortium