Universiteit Leiden
The Netherlands

Clustered Mendelian randomization analyses identify distinct and opposing pathways in the association between genetically influenced insulin-like growth factor-1 and type 2 diabetes mellitus

Wang, W.Y.; Tesfay, E.B.; Klinken, J.B. van; Dijk, K.W. van; Bartke, A.; Heemst, D. van; Noordam, R.

Diabetes

Clustered Mendelian randomization analyses identify distinct and opposing pathways in the association between genetically influenced insulin-like growth factor-1 and type 2 diabetes mellitus

Wenyi Wang [iD] ,[1]* Ephrem Baraki Tesfay,[2] Jan Bert van Klinken [iD] ,[1,3,4]
Ko Willems van Dijk,[1,5,6] Andrzej Bartke,[7] Diana van Heemst[2] and
Raymond Noordam [iD] [2]

[1]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, [2]Department of Internal Medicine, Section of Gerontology and Geriatrics; Leiden University Medical Center, Leiden, The Netherlands, [3]Laboratory Genetic Metabolic Diseases, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands, [4]Core Facility Metabolomics, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands, [5]Department of Internal Medicine, Division Endocrinology, Leiden University Medical Center, Leiden, The Netherlands, [6]Leiden Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands and [7]Department of Internal Medicine, Southern Illinois University School of Medicine, Springfield, IL, USA

*Corresponding author. Department of Human Genetics, Leiden University Medical Center, LUMC Main Building, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. E-mail: w.wang1@lumc.nl

Abstract

Background: There is inconsistent evidence for the causal role of serum insulin-like growth factor-1 (IGF-1) concentration in the pathogenesis of human age-related diseases such as type 2 diabetes (T2D). Here, we investigated the association between IGF-1 and T2D using (clustered) Mendelian randomization (MR) analyses in the UK Biobank.
Methods: We conducted Cox proportional hazard analyses in 451 232 European-ancestry individuals of the UK Biobank (55.3% women, mean age at recruitment 56.6 years), among which 13 247 individuals developed type 2 diabetes during up to 12 years of follow-up. In addition, we conducted two-sample MR analyses based on independent single nucleotide polymorphisms (SNPs) associated with IGF-1. Given the heterogeneity between the MR effect estimates of individual instruments (*P*-value for Q statistic = 4.03e−145), we also conducted clustered MR analyses. Biological pathway analyses of the identified clusters were performed by over-representation analyses.
Results: In the Cox proportional hazard models, with IGF-1 concentrations stratified in quintiles, we observed that participants in the lowest quintile had the highest relative risk of type 2 diabetes [hazard ratio (HR): 1.31; 95% CI: 1.23–1.39). In contrast, in the two-sample MR

analyses, higher genetically influenced IGF-1 was associated with a higher risk of type 2 diabetes. Based on the heterogeneous distribution of MR effect estimates of individual instruments, six clusters of genetically determined IGF-1 associated either with a lower or a higher risk of type 2 diabetes were identified. The main clusters in which a higher IGF-1 was associated with a lower risk of type 2 diabetes consisted of instruments mapping to genes in the growth hormone signalling pathway, whereas the main clusters in which a higher IGF-1 was associated with a higher risk of type 2 diabetes consisted of instruments mapping to genes in pathways related to amino acid metabolism and genomic integrity.

Conclusions: The IGF-1-associated SNPs used as genetic instruments in MR analyses showed a heterogeneous distribution of MR effect estimates on the risk of type 2 diabetes. This was likely explained by differences in the underlying molecular pathways that increase IGF-1 concentration and differentially mediate the effects of IGF-1 on type 2 diabetes.

Key words: Clustered Mendelian randomization analysis, cohort studies, Mendelian randomization analysis, type 2 diabetes, insulin-like growth factor-1

Key messages

- High genetically-influenced insulin-like growth factor-1 (IGF-1) was associated with higher risk of type 2 diabetes mellitus, but individual instruments show a heterogeneous distribution of effect estimates in Mendelian randomization (MR).
- Six clusters associated either with a lower or a higher risk of type 2 diabetes were identified, which mapped to distinct molecular pathways.
- Our results indicate that the association between IGF-1 and the risk of developing type 2 diabetes is context-dependent.
- Cause-specific, but not total, IGF-1 concentrations result in a more accurate risk assessment of an individual to develop type 2 diabetes mellitus.

Introduction

Insulin-like growth factor-1 (IGF-1) is a pleiotropic hormone that plays a major role in cellular growth, proliferation and survival.[1] The secretion of IGF-1, predominately by the liver, is promoted by growth hormone (GH); conversely, IGF-1 in the circulation feeds back centrally to the hypothalamus to inhibit GH secretion.[2] The availability of free IGF-1 in the blood is regulated by its association with distinct insulin-like growth factor-binding proteins (IGFBPs) which can increase IGF-1 half-life or block its binding to IGF-1 receptors.[3] IGF-1 plays a very important role in the aging process, and it was found that the level of IGF-1 increases during puberty and decreases gradually with increasing age during adulthood.[4,5] IGF-1 was also found to be involved in the pathophysiology of various diseases, including cancer, neurodegenerative disease, cardiovascular disease and type 2 diabetes mellitus.[6] Several (prospective) cohort studies have found that lower levels of

IGF-1 were associated with an increased risk of impaired glucose tolerance, increased insulin resistance and hence the development of diabetes mellitus.[7,8]

In contrast to these multivariable-adjusted association analyses, a recent study demonstrated that a higher genetically influenced IGF-1 concentration was associated with a higher risk of developing type 2 diabetes, using Mendelian randomization (MR) analyses.[9] MR is an approach to determine whether the association between risk factors and outcome is causal, by using genetic variants as instrumental variables,[10] and therefore does not meet the requirements for triangulation of observations done in epidemiology.[11] However in some circumstances, there is clear heterogeneity in the causal effects of the individual single nucleotide polymorphisms (SNPs) that are used as instrumental variables, which may indicate either pleiotropy or differences in biological pathways contributing to high levels of the exposure.[12] Clustered MR was recently

developed to provide a means to address the heterogeneity in causal effects by clustering variants that show similar individual causal estimates on the outcome.[12] Previously, such context-dependent MR analyses have been proposed to provide more biological perspective in causal associations.[13,14]

We hypothesized that heterogeneity in causal effects of individual variants could be a reflection of different biological mechanisms involved in the association between IGF-1 and type 2 diabetes. For example, variation influenced by processes causing insufficient GH signalling may have an impact on T2D different from variation influenced by processes causing increased GH resistance. Therefore in this study, we aimed to investigate the association between IGF-1 and incident type 2 diabetes followed by clustered MR analyses[15] in the large UK Biobank population, and explored the possible biological pathways involved in the clustered causal associations.

Methods

Study setting and study population
The UK Biobank is a very large prospective cohort study with over 500 000 participants aged 40–69 years at recruitment across the entire UK.[16] Participants were recruited between 2006 and 2010 in 22 assessment centres across the UK. Baseline examinations in all participants included physical measures, collection of blood, urine and saliva, a self-completed touch-screen questionnaire and a brief computer-assisted interview to investigate sociodemographics, family history, environmental factors, lifestyle, psychosocial factors etc.

The present project was accepted under project number 22474. We restricted the analyses to the UK Biobank participants who reported to be of European ancestry including British, Irish and any other European background, who had information available on serum IGF-1 concentration and who were in the full release imputed genomics datasets.

Biochemical analyses
Biological samples were collected to measure biochemical markers including IGF-1 at baseline (2006–10), comprising ~480 000 participant samples. Serum levels of IGF-1 were analysed using chemiluminescent Immunoassay (DiaSorin Liaison XL) with a one-step sandwich. Coefficients of variation derived from the internal quality control samples of the low, medium and high IGF-1 concentrations ranged from 6.03% to 6.18%. More information on assay performance of the UK Biobank Biomarker Project can be found online [https://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1227].

Outcome definitions
Prevalent and incident diagnosis of type 2 diabetes mellitus was identified in UK Biobank as the date of first appearance of non-insulin-dependent diabetes mellitus (data-field 130708 in the UK Biobank Database). This variable has been composed through a standard algorithm combining the data derived from hospital admissions (through linkage with the medical records from the National Health Service), general practitioners and death records, and through self-report. Based on the date of first appearance and the data of enrolment, we defined whether a case was prevalent (before enrolment) or incident (after enrolment).

Statistical analyses
The cross-sectional association between IGF-1 and prevalent type 2 diabetes was assessed by means of logistic regression modelling adjusting for age at recruitment, sex and body mass index (BMI).

Participants without diabetes mellitus at baseline were followed until the occurrence of type 2 diabetes, mortality or loss of follow-up, whichever occurred first. The association between IGF-1 levels (as continuous and stratified variables) and incidence of type 2 diabetes in the UK Biobank cohort was assessed using Cox proportional hazard models. Participants were categorized into five groups based on quintiles of IGF-1 concentration. Quintile 1 (lowest 20%) and quintile 5 (highest 20%) were used separately as reference groups to calculate the hazard ratio (HR), respectively. Potential confounders included sex, age at recruitment and baseline BMI based on height and weight measured at the assessment centres. The analyses were conducted in R using the survival package (version 3.2–7).[17] The Kaplan–Meier curve was plotted to visualize the difference of survival probability between IGF-1 quintiles and whether the proportionality assumption holds.

Genotyping and genetic imputations. Genome-wide genotype data for all 500 000 UK Biobank participants, generated using Affymetrix UK BiLEVE Axiom array (initial 50 000 participants) and the Affymetrix UK Biobank Axiom Array (remaining 450 000 participants), genotyped around 850 000 variants. All genetic data were quality-controlled centrally by UK Biobank resources. In addition, UK Biobank resources performed centralized imputations on approximately 96 million genotypes using the UK10K haplotype,[18] 1000 Genomes Phase 3[19] and Haplotype Reference Consortium (HRC) reference panels.[20] Autosomal SNPs were pre-phased using SHAPEIT3 and imputed using IMPUTE4. Related individuals were identified by estimating kinship coefficients for all pairs of samples using only markers weakly informative of ancestral background. More information on the genotyping

processes and genetic imputation can be found online [https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=263].

Genome-wide association analyses. Genome-wide association studies (GWAS) on continuous IGF-1 concentrations were performed on all European individuals to provide a list of independent lead SNPs to be used as instrumental variables for exposure (IGF-1) in the MR analyses. Analyses were performed using linear mixed models implemented in the program BOLT_LMM (version 2.3.2).[21] We adjusted the analyses for age, sex and the first 10 principal components, and corrected for the Kinship matrix to correct for familial relationships in the UK Biobank population. Analyses were done on the autosomal chromosomes only. SNPs with a minor allele frequency <0.01, as well as SNPs with an imputation quality <0.3, were excluded. *P*-values of SNPs smaller than $5e^{-8}$ were extracted and stored for the MR analyses. Visualization of the results was performed using the R-based packages ggplot2[22] and EasyStrata[23] [www.genepi-regensburg.de/easystrata].

Outcome data source. For this analysis, the instruments for outcome (type 2 diabetes) were acquired from the DIAGRAM Consortium, which is a publicly available summary-statistics GWAS meta-analysis of European ancestry comprising 32 studies and 898 130 individuals (74 124 type 2 diabetes cases and 824 006 controls).[24]

The summary statistics of outcome for glycated haemoglobin (HbA1c), fasting glucose, C-reactive protein (CRP) level and BMI were retrieved from an online source [https://gwas.mrcieu.ac.uk/] under the GWAS IDs ieu-b-104 (HbaA1c, $N = 46\,368$), ieu-b-114 (fasting glucose, $N = 133\,010$), ieu-b-35 (CRP, $N = 204\,402$) and ieu-b-40 (BMI, $N = 681\,275$).

Mendelian randomization analyses. Two-sample MR was performed with summary-based statistics of GWAS using the R-based statistical package Two Sample MR [http://github.com/MRCIEU/TwoSampleMR][25] This statistical package also connects to a large library of exposures from published GWAS to use as instrumental variables, which is aligned with the online GWAS catalogue.

For the present study, we performed clumping process (window size = 10 000 kb, $R2 < 0.001$) with the European samples from the 1000 Genomes Project, which were used to estimate linkage disequilibrium (LD) between SNPs. Among the pairs of SNPs with $R2$ above the specified threshold ($R2 = 0.001$), only the SNPs with the lowest *P*-value were retained to provide a list of independent lead SNPs from the MR analyses. Otherwise, the statistical power of MR analyses would be overestimated (e.g. underestimated standard errors of the summary estimates of the MR analyses). SNPs present in UK Biobank, but absent from the LD reference panel, were removed. On the basis of the significant independent lead SNPs (*P*-value <5e−8), we assessed their possible association with type 2 diabetes.

Methods for MR analyses of summary-level data based on a two-sample design have been described in detail previously.[26,27] Using inverse-variance-weighted (IVW) analyses, we combined the effects of the individual genetic instruments to obtain a genetically-determined association between exposure and outcome under the assumption of the absence of horizontal pleiotropy. However, given the large number of genetic instruments included in the present analyses, there is a high probability that at least some SNPs show pleiotropic effects. To test whether possible pleiotropic effects could bias the overall effect estimates (horizontal pleiotropy), we performed the sensitivity analyses MR Egger regression[28] and weighted median estimator (WME) analyses.[29] MR Egger does not force the regression line through the intercept and is, therefore, able to test for the presence of directional pleiotropy, and WME estimator assumes at least 50% of the instruments included in the MR analyses were valid.

Clustered Mendelian randomization analyses. Clustered MR analyses were conducted to identify groups of genetic variants that have similar causal estimates of the exposure on the outcome, which has been described in detail previously.[15] Briefly, this method performs likelihood-based clustering on a sample of ratio estimates and standard errors of genetic variants. An MR-Clust mixture model is firstly built for ratio estimates and the mixture proportion for cluster k, after which these two parameters are estimated by using an expectation-maximization (EM) algorithm and optimized by using maximum likelihood estimates (MLE). This MR-Clust mixture model automatically accounts for the possibilities that some or all of the ratio estimates may be drawn under two distribution—a null distribution (the ratio estimates are centred around zero) and a junk distribution (highly dispersed ratio estimates considered as outliers in the sample). The presence of null and junk clusters minimizes false-positive findings from the model, with no need to interpret. The maximization step to optimize the parameter values resembles the inverse-variance weighted estimates of the causal effect of the exposure on the outcome. Last, the numbers of substantive clusters are estimated based on the values derived from MLEs for each value of k possible substantive clusters, by minimizing the Bayesian information criterion (BIC) which helps to avoid ove-rparameterization. Therefore, if the causal estimates of each genetic variant on the outcome were similar (e.g. their ratio estimates were similar in direction, magnitude and precision), it was

divided into different clusters in which the included genetic variants were more homogeneous concerning the causal estimates. The inclusion probability of SNPs in each cluster was higher than 0.8.

In order to identify distinct causal effects of genetic variants derived from GWAS of continuous IGF-1 levels on type 2 diabetes, we made use of the R-based MR-Clust package [https://github.com/cnfoley/mrclust]. MR-Clust performs likelihood-based clustering on Wald ratio estimates and accompanying standard errors. Genetic instruments within a cluster share similar causal estimates (e.g. Wald ratio estimates are similar in direction, magnitude and precision) of the causal effect of the exposure on the outcome. MR analyses were repeated on all clusters to investigate the causal effects of each cluster on type 2 diabetes. The stability of the clusters was measured by repeating the same analyses using 1000 randomly generated subsets containing 200 SNPs and with the full set of 332 SNPs associated with IGF-1 levels. In addition, cross-validation analyses were performed by randomly sub-dividing the UKB participants into two subsets and repeating the same clustered MR analyses within each subset of the UKB.

Pathway analyses. We performed pathway analysis on all the six IGF-1 clusters, to provide insights into biological pathways which could explain the heterogeneity between causal estimates of genetic variants in different clusters on type 2 diabetes. Gene candidates of each locus were determined with the SNP2Gene function of FUMA (the Functional Mapping and Annotation of Genome-Wide Association Studies platform),[30] where we used a 20-kb positional map and included genes whose expression was associated with the locus in GTEx v8. Gene candidates of a locus which were also linked to a gene that had a clear biological relation to IGF-1 were excluded from the analysis. Specifically, we defined the list of prioritized genes as IGF1, IGF2, INS, IGFALS, GHR, GHSR, GH1, GH2, AKT3, FOXO3, which we based on their overlap with the IGF1 gene in the pathway databases that we used in the subsequent analysis.

Pathway analysis was performed using the pathway definitions of the KEGG (a bioinformatics resource for deciphering the genome) and Reactome database, which were downloaded from ConsensusPathDB[31] on 28 October 2021. Gene sets with more than 1000 genes or less than two were excluded. Since genes with a shared function sometimes colocalize on the genome, we developed a novel method for performing over-representation analysis that corrects for the inflation induced by the colocalization effect. In detail, we defined gene clusters of similar function and close proximity on the genome by merging all genes that were within a 100-kb distance on the genome (GENCODE v38) and that co-occurred in at least one pathway. This operation reduced the total number of genes that were covered in the KEGG and Reactome pathways from 11 855 to 9123 gene clusters and is completely independent from the set of gene candidates obtained from the GWAS results. Subsequently the pathway gene sets and the gene candidate set were mapped on the gene clusters, and the over-representation was done in terms of the gene clusters where loci that mapped multiple times to genes within the same cluster were counted only once. Over-representation analysis was done using Fisher's exact test and results were corrected for multiple testing with the Benjamini–Hochberg procedure where pathways with the same definition were counted as one. Only pathways with false-discovery rate (FDR) >0.1 were shown in the present study.

In addition, as follow-up analyses, we conducted two-sample MR analyses to associate IGF-1 and several metabolic traits related to type 2 diabetes including HbA1c, fasting glucose, CRP and BMI.

Results

Characteristics of the study population

In total, 451 232 European participants without diabetes mellitus at baseline were included in our study, of whom 13 247 developed type 2 diabetes in up to 12 years of follow-up (Table 1). Of the participants not developing type 2 diabetes, 55.7% were women, the mean age at recruitment was 56.5 (SD = 8.1) years, the average BMI was 27.1 (SD = 4.5) kg/m^2 and the mean IGF-1 levels were 21.5 (SD = 5.6) nmol/L. Of the participants developing type 2 diabetes during follow-up, women accounted for 41.9%, the mean age at recruitment was 59.3 (SD = 7.2) years, the average BMI was 31.7 (SD = 5.6) kg/m^2 and the mean IGF-1 level was 19.9 (SD = 6.5) nmol/L.

Table 1 Characteristics of the UK Biobank study population for prospective analyses

	Controls	Cases	Total
N	437 985	13 247	451 232
Age at recruitment, in years	56.5 (8.1)	59.3 (7.2)	56.6 (8.0)
Time to diagnosis, in years	–	5.3 (2.5)	–
% of women	55.7	41.9	55.3
BMI, in kg/m^2	27.1 (4.5)	31.7 (5.6)	27.2 (4.6)
IGF-1 levels, in nmol/L	21.5 (5.6)	19.9 (6.5)	21.5 (5.6)

Data presented as means with standard deviation (SD) or as stated otherwise. Information on BMI was missing for 7599 controls and 417 cases. Information on age at diagnosis was missing for 1228 cases. Information on IGF-1 levels is missing for 29 889 controls and 990 cases.

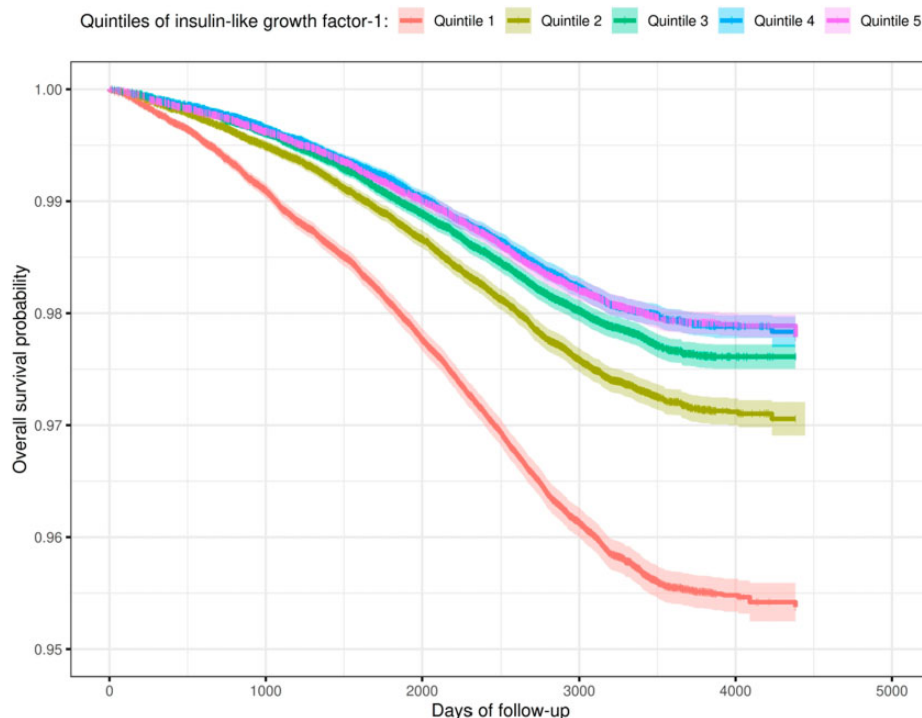IGF-1: insulin-like growth factor-1; BMI: body mass index.

Figure 1 Kaplan–Meier survival curves displaying the time to develop incident type 2 diabetes by quintiles of insulin-like growth factor-1 concentrations. The x-axis presents days of follow-up and the y-axis presents overall survival probability. Censoring is indicated by vertical marks

Table 2 Hazard ratio of incident type 2 diabetes according to quintiles of insulin-like growth factor-1 concentration[a]

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Ranges of IGF-1 (nmol/L)	1.44–16.73	16.73–19.93	19.93–22.63	22.63–25.77	25.77–126.77
Hazard ratio (95% CI)	Reference	0.77 (0.73–0.81)	0.70 (0.66–0.74)	0.68 (0.64–0.72)	0.76 (0.72–0.81)
Hazard ratio (95% CI)	1.31 (1.23–1.39)	1.01 (0.94–1.07)	0.92 (0.86–0.98)	0.89 (0.83–0.95)	Reference

[a]This table presents the hazard ratio (HR) and 95% confidence interval (CI) of incident type 2 diabetes by insulin-like growth factor-1 concentration in quintiles. The second row shows hazard ratios of incident type 2 diabetes using quintile 1 as reference group. The third row shows hazard ratios of incident type 2 diabetes using quintile 5 as reference group.

IGF-1, insulin-like growth factor-1.

Association between IGF-1 and prevalent and incident type 2 diabetes

The cross-sectional analyses showed that a higher level of IGF-1 was significantly associated with a lower odds of prevalent type 2 diabetes (odds ratio: 0.98; 95% CI: 0.97–0.98, per nmol/L increase of IGF-1).

Multivariable-adjusted Cox proportional hazard model analyses were performed to evaluate the association between IGF-1 levels (as quintile and continuous variables) and incident type 2 diabetes. The Kaplan–Meier curve (Figure 1) illustrated that participants in the lower IGF-1 quintiles (quintiles 1 and 2) had proportionally lower risk of type 2 diabetes compared with participants in the higher quintiles (quintiles 3, 4 and 5), which is consistent with the result of Cox proportional hazard model when IGF-1 is assessed as a continuous variable. The results showed that higher levels of IGF-1 are associated with lower risk of incident type 2 diabetes, with HR of 0.98 per nmol/L IGF-1 and 95% confidence interval (CI) of 0.97–0.98.

More specifically, individuals in quintile 2 had a lower risk of type 2 diabetes (HR: 0.77; 95% CI: 0.73–0.81) than individuals in quintile 1 (Table 2), as did those in quintile 3 (HR: 0.70; 95% CI: 0.66–0.74), quintile 4 (HR: 0.68; 95% CI: 0.64–0.72) and quintile 5 (HR: 0.76; 95% CI: 0.72–0.81). Taking participants in quintile 5 as the reference group, a higher risk was observed in quintile 1 (HR: 1.31; 95% CI: 1.23–1.39), and those in quintiles 3 (HR: 0.92; 95% CI: 0.86–0.98) and 4 (HR: 0.89; 95% CI: 0.83–0.95) had a lower risk of developing type 2 diabetes.

Mendelian randomization analyses

There were 95 877 significant SNPs ($P < 5e-8$) identified in the GWAS of continuous IGF-1 (Supplementary Figure S1,

available as Supplementary data at *IJE* online), of which 387 independent lead SNPs were derived. After quality control performed by the TwoSample MR package (e.g. excluding palindromic SNPs and SNPs with intermediate allele frequencies), 323 SNPs were used in the MR analyses (Supplementary Table S1, available as Supplementary data at *IJE* online). The MR estimates assessing the effect of IGF-1 on type 2 diabetes showed that a 1-nmol/L increase in IGF-1 was associated with a 1% higher risk of type 2 diabetes in the IVW analyses (OR: 1.01; 95% CI: 1.00–1.02). Similar results were obtained from MR Egger (OR: 1.02; 95% CI: 1.00–1.05) and WME analyses (OR: 1.01; 95% CI: 1.00–1.02).

Clustered Mendelian randomization analyses

We observed large heterogeneity in the individual MR effect estimates (Figure 2) as was also evidenced by the Q statistic ($P = 4.03e^{-145}$). The individual MR effect estimates were clustered into a total of six clusters with an inclusion probability of SNPs higher than 0.8 (Figure 3; Supplementary Table S2, available as Supplementary data at *IJE* online). MR estimates from different methods assessing the effect of six clusters on type 2 diabetes are presented in Figure 4, Table 3 and Supplementary Figure S2, available as Supplementary data at *IJE* online. In cluster 1 (IVW: OR: 1.54, 95% CI: 1.43–1.65), cluster 2 (IVW: OR: 1.03; 95% CI: 1.02–1.04) and cluster 5 (IVW: OR: 1.20; 95% CI: 1.18–1.22), higher levels of IGF-1 were associated with a higher risk of type 2 diabetes. On the

other hand, cluster 3 (IVW: OR: 0.92; 95% CI: 0.91–0.94), cluster 4 (IVW: OR: 0.62; 95% CI: 0.58–0.67) and cluster 6 (IVW: OR: 0.80; 95% CI: 0.76–0.85) showed that higher levels of IGF-1 were associated with lower risk of type 2 diabetes. The results from sensitivity analyses by using weighted median estimator and MR-Egger did not materially differ from the result of the IWV method. The results from the stability test of the MR-Clust demonstrated that IGF-1 clusters were consistently constructed across different random subsets of IGF-1 SNPs and across different runs with the same set of 332 IGF-1 SNPs (Supplementary Figures S3 and S4, available as Supplementary data at *IJE* online). The cross-validation analyses also provided evidence that at least three clusters for IGF-1 were reproducible (Supplementary Figures S5 and S6, available as Supplementary data at *IJE* online).

Pathway analyses

Gene candidates for each SNP and gene used for the following pathway analyses can be found in Supplementary Table S3 (available as Supplementary data at *IJE* online). For the pathway analyses, we found that cluster 2 and cluster 3 were mapped to specific pathways. Cluster 2 mapped to 11 pathways, including metabolism of xenobiotics by cytochrome P450 and chemical carcinogenesis (Supplementary Figure S7 and Table S4, available as Supplementary data at *IJE* online). Cluster 3 mapped to six pathways, including growth hormone synthesis, secretion and action, prolactin receptor signalling and growth hormone receptor signalling
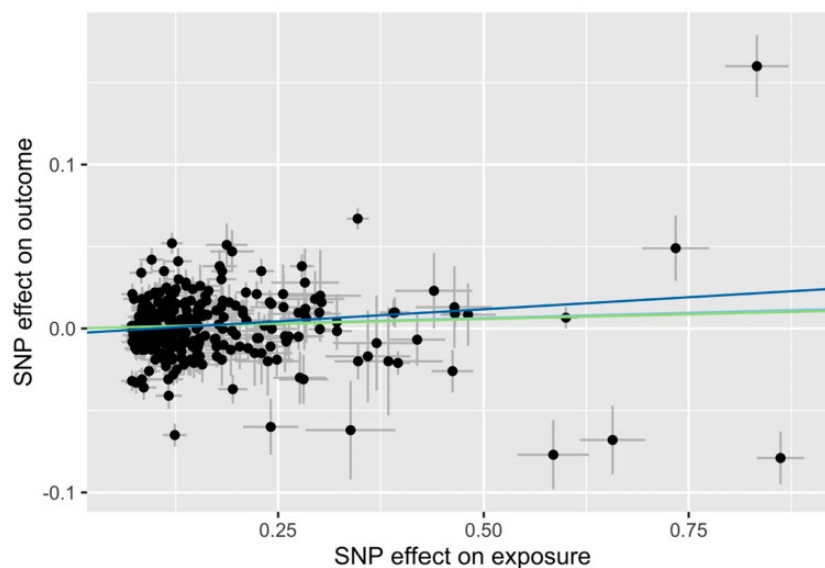

Figure 2 Scatter plots of the MR effect estimates of continuous IGF-1 on type 2 diabetes derived from different MR tests. The x-axis is the genetic association between SNPs and IGF-1 and the y-axis is the genetic association between SNPs and type 2 diabetes. Analyses were conducted using the inverse-variance-weighted, weighted median and MR Egger methods. The slope of each line presents the estimated MR effect for each method. MR, Mendelian randomization; IGF-1, insulin-like growth factor-1; SNPs, single nucleotide polymorphisms
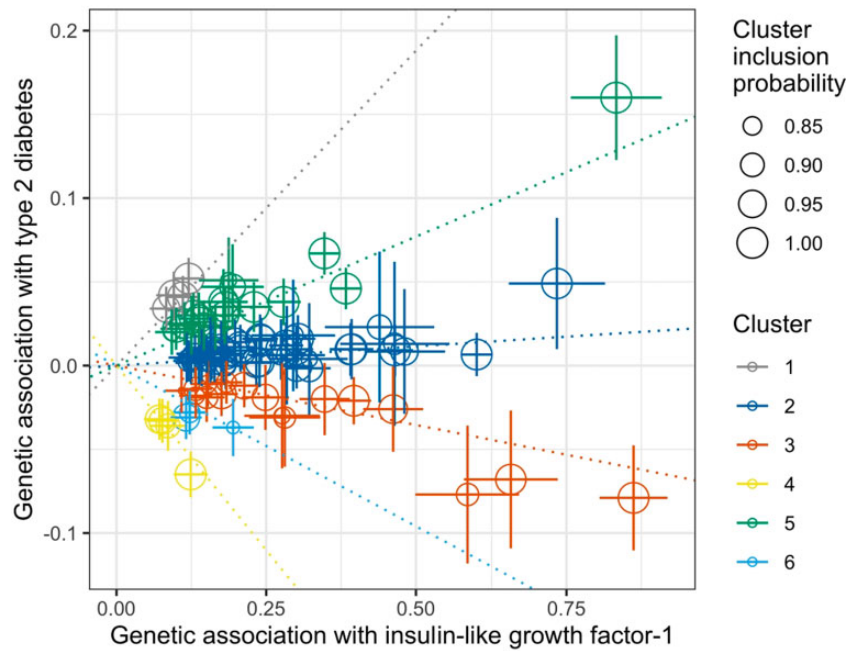
Figure 3 Scatter plots of the MR effect estimates of continuous IGF-1 on type 2 diabetes for clustered MR analyses. The x-axis is the genetic association between SNPs and IGF-1 and the y-axis is the genetic association between SNPs and type 2 diabetes. MR, Mendelian randomization; IGF-1, insulin-like growth factor-1; SNPs, single nucleotide polymorphisms
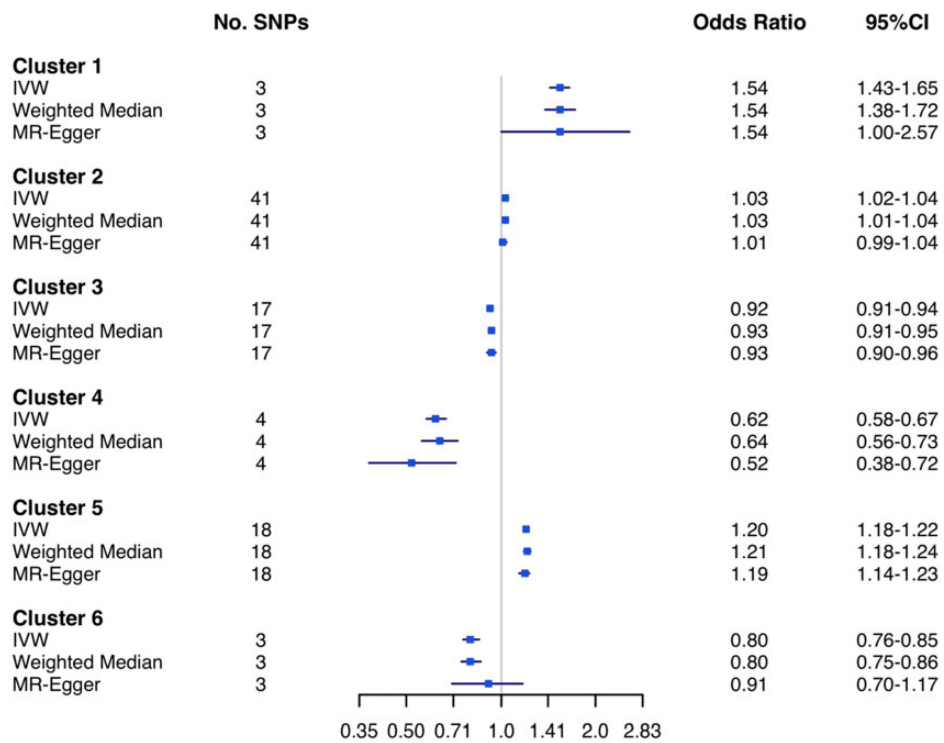

Figure 4 Forest plot of the MR estimates and MR sensitivity analyses for the six clusters of genetically predicted IGF-1 levels on type 2 diabetes. Odds ratios measured the associations between IGF-1 clusters and type 2 diabetes. No. SNPs, number of SNPs used in the separate MR analyses; 95% CI, 95% confidence interval; MR, Mendelian randomization; IGF-1, insulin-like growth factor-1; SNPs, single nucleotide polymorphisms

Table 3 Mendelian randomization effect estimates derived from different methods for six clusters on type 2 diabetes[a]

Clusters	Number of SNPs	Inverse-variance-weighted	Weighted median	MR Egger
		OR (95% CI)	OR (95% CI)	OR (95% CI)
Cluster 1	3	1.54 (1.43–1.65)	1.54 (1.38–1.72)	1.60 (1.00–2.57)
Cluster 2	41	1.03 (1.02–1.04)	1.02 (1.01–1.04)	1.01 (0.99–1.04)
Cluster 3	17	0.92 (0.91–0.94)	0.93 (0.91–0.95)	0.93 (0.90–0.96)
Cluster 4	4	0.62 (0.58–0.67)	0.64 (0.56–0.73)	0.52 (0.38–0.72)
Cluster 5	18	1.20 (1.18–1.22)	1.21 (1.18–1.24)	1.19 (1.14–1.23)
Cluster 6	3	0.80 (0.76–0.85)	0.80 (0.75–0.86)	0.91 (0.70–1.17)

[a]This table presents the odds ratio (OR) and 95% confidence interval (CI) of associations between six clusters and type 2 diabetes.
SNPs, single nucleotide polymorphisms.

(Supplementary Figure S8 and Supplementary Table S5, available as Supplementary data at *IJE* online). Clusters 1, 4, 5 and 6 did not map to any specific pathway.

Follow-up analyses of the clusters in relation to other traits

The results of follow-up analyses using two-sample MR are shown in Supplementary Table S6 (available as Supplementary data at *IJE* online). Using IVW, high genetically influenced IGF-1 levels through cluster 2, 3 and 6 were associated with higher HbA1c. In addition, high genetically-influenced IGF-1 levels through cluster 5 and cluster 6 were associated with lower CRP; and high genetically influenced IGF-1 through cluster 5 was also associated with lower fasting glucose. In general, results were (directionally) consistent in the weighted median and MR-Egger sensitivity analyses.

Discussion

In this study, we investigated the association of IGF-1 and type 2 diabetes using prospective multivariable-adjusted survival analyses followed by MR and clustered MR analyses in the UK Biobank. The results from the MR analyses showed that a genetically influenced higher level of IGF-1 was associated with a higher risk of type 2 diabetes, which was in contrast to the results from the prospective analyses showing that a higher concentration of IGF-1 was associated with a lower risk of type 2 diabetes. Since the underlying individual genetic instruments showed a heterogeneous distribution of MR effect estimates, clustered MR identified six clusters of genetic instruments for IGF-1 with different associations with type 2 diabetes, which mapped to distinct molecular pathways. Collectively, our results indicate that the association between IGF-1 and the risk of developing type 2 diabetes is context-dependent.

Findings from other prospective studies regarding the association between IGF-1 and type 2 diabetes have been inconsistent. Our results from the prospective analyses were in line with another cohort study showing that high levels of IGF-1 were associated with a lower risk of type 2 diabetes mellitus during 4.5 years of follow-up.[7] However, some nested case-cohort studies suggested there was no association between total IGF-1 levels and the risk of type 2 diabetes.[32,33] In addition, a cohort study found that the association between free IGF-1 and type 2 diabetes was dependent on the level of insulin in women.[34] Interestingly a recent study, that used a data-driven approach to metabolically subtype the UK biobank participants, found that the subgroup associated with increased prevalence and incidence of type 2 diabetes exhibited a decreased level of IGF-1.[35] The inconsistency of these findings reinforces the notion that the association between IGF-1 and type 2 diabetes is context-dependent, which is in line with the distinct biological mechanisms identified by the clustered MR analyses in our study.

Many studies found a J- or U- shaped association between IGF-1 and type 2 diabetes[7,36] or insulin resistance.[37] For example, one study showed that individuals with both low or high IGF-1 levels were at increased risk of developing diabetes, in a prospective cohort study.[36] Similarly, a U-shaped association between IGF-1 and measures of insulin resistance was found in a cross-sectional study in Danish adults.[37] Partly in line with these results, we observed a J-shaped relationship between IGF-1 and type 2 diabetes, with particularly low levels of IGF-1 being associated with an increased risk of developing type 2 diabetes.

The result of MR analyses of continuous IGF-1 and type 2 diabetes in our study was supported by a recent publication showing that genetically determined higher levels of IGF-1 were associated with increased risk of developing type 2 diabetes.[9] However, we observed large between-SNP heterogeneity, and the MR effect estimates were not proportional to each other. We identified six main clusters of IGF-1 genetic instruments with distinct effects on type 2 diabetes, by using clustered MR. After mapping the genetic instruments to genes and over-representation analysis

using the KEGG and Reactome databases, several clusters were mapped to specific pathways.

To gain insight into the biological basis of the identified IGF-1 genetic clusters in relation to type 2 diabetes, we performed pathway analyses and identified two clusters associated with specific pathways. Cluster 2 (in which higher genetically influenced IGF-1 was associated with a higher risk of type 2 diabetes) mapped to multiple pathways, including metabolism of xenobiotics by cytochrome P450 and chemical carcinogenesis. In addition, MR follow-up analyses using two-sample MR showed cluster 2 was associated with higher HbA1c, which supports our findings on type 2 diabetes. However, the exact interpretation of this cluster is complicated and requires additional research. It is tempting to speculate that since the cytochrome P450 system is likely involved in the breakdown of HbA1c, it influences the risk of type 2 diabetes by affecting the residence time of HbA1c in blood. In support of this notion, a direct effect of IGF-1 on the cytochrome P450 system has been demonstrated before.[38]

Cluster 3 (in which a higher level of genetically influenced IGF-1 was associated with a lower risk of type 2 diabetes) was mapped to pathways related to GH synthesis, secretion and action and to GH receptor signalling. However, counterintuitively, cluster 3 was associated with higher levels of Hb1Ac, despite being associated with a lower risk of type 2 diabetes. Although this finding seems difficult to interpret, this could mean that as IGF-1 plays an important role in growth and development, higher IGF-1 may have resulted in a larger pancreas, which confers an advantage in a non-diabetic situation (as used in the Hb1Ac analyses) by enabling a stronger response to rising glucose levels, resulting in lower HbA1C. However, under conditions of chronically elevated insulin demand, such strong pancreatic response may in the end cause an accelerated exhaustion of insulin production and a rapid development of a more insulin-dependent-like phenotype of type 2 diabetes. Previously, IGF-1 has been shown to interact with beta cells,[39] and to have direct effects on pancreas development and beta cell compensation to insulin resistance.[40] In support of the biological interpretation of the genes mapped to cluster 3, dysregulation of GH receptor signalling and the GH-IGF-1 axis can lead to multiple diseases such as type 2 diabetes. Mice with liver IGF-1 deficiency had a 4-fold increase in GH levels. Upon treatment with a GH antagonist, these mice had decreased blood glucose and insulin levels and increased peripheral insulin sensitivity compared with mice with untreated liver IGF-1 deficiency. These data indicate that the GH/IGF-1 axis plays a balancing role in insulin sensitivity and thus type 2 diabetes.[41] In addition, an epidemiological cross-sectional study showed that IGF-1 was associated with type 2

diabetes risk, but this association varied depending on the insulin levels: in individuals with low levels of insulin, IGF-1 decreased type 2 diabetes risk and in individuals with high levels of insulin IGF-1, increased type 2 diabetes risk.[34]

The main strength of our study is the extremely large sample size, allowing stratification of the genetic instruments with sample statistical power. A limitation of the present study is that the IGF-1 level used was total IGF-1 concentration and not free IGF-1 (e.g. relative to the concentration of IGF-binding proteins). Furthermore, the study was performed in a study of European-ancestry participants. Translation of the results to participants of non-European ancestry should be done with caution. Although the SNP-outcome dataset (DIAGRAM) used in two-sample MR analyses had an overlap of 49.3% with the SNP-exposure dataset (UKB only), we expect that the effect of this overlap is limited given the very large sample size, since it has been shown that one-sample MR performs well in large samples.[42] In addition, non-fasting blood sampling has likely resulted in somewhat increased variation in IGF-1 concentration and consequently a slight reduction in statistical power.[43] However, given that this variation is most likely independent from genetic make-up, no bias in the results was expected from this. Although sensitivity analyses in a random subset of the UK Biobank and clustering using random subsets of SNP-exposure relationships gave similar results, results warrant replication in an independent sample using SNP-exposure associations derived from a fully independent study sample. Last, as our multivariable-adjusted Coxproportional hazard models indicated a possible non-linear association between IGF-1 levels and type 2 diabetes, this was not taken into account in the MR analyses which were done on continuous levels. Whether the different clusters could possibly contribute to a non-linear association is currently not known. Furthermore, the more recently introduced methods for non-linear MR were not applicable here, given the overlap in the SNP-exposure and SNP-outcome datasets (both UK Biobank); the extent of bias introduced by one-sample non-linear MR (which was shown to be limited for the more classical two-sample MR methodologies[42]) is currently unknown.

Conclusion

In conclusion, from clustering analyses we found that genetically determined IGF-1 was associated with both a higher and a lower risk of development of type 2 diabetes, which is likely mediated by distinct biological mechanisms. Therefore, the total concentration of IGF-1 does not provide insight into the risk of developing type 2 diabetes.

Ethics approval

The data used in this work are publicly available. The UK Biobank obtained relevant participant consent and ethical approval. The analyses in UK Biobank were done under project number 22474. The UK Biobank study was approved by the North-West Multi-centre Research Ethics Committee (MREC). Access to information to invite participants was approved by the Patient Information Advisory Group (PIAG) from England and Wales. All participants in the UK Biobank study provided written informed consent.

Data availability

Data that support the findings of this study have been deposited in the UK Biobank under project number 22474. Data are available on request after approval of a research proposal by UK Biobank resources. Summary data for IGF-1 from GWAS analyses is available on the Figshare website [https://figshare.com/s/9ff69bbc84a67559989f].

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

W.W., E.B.T., J.B.vK., K.W.vD., A.B., D.vH., R.N.: substantial contributions to concept, design, acquisition of data, data analyses and data interpretation; drafting article and critical comment on the initial versions of the manuscript; final approval of the manuscript before submission.

Funding

Conflict of interest

None declared.

References

1. Rajpathak SN, Gunter MJ, Wylie-Rosett J *et al.* The role of insulin-like growth factor-I and its binding proteins in glucose homeostasis and type 2 diabetes. *Diabetes Metab Res Rev* 2009; **25**:3–12.
2. Nicholls AR, Holt RI. Growth hormone and insulin-like growth factor-1. *Front Horm Res* 2016;**47**:101–14.
3. Aguirre GA, De Ita JR, de la Garza RG, Castilla-Cortazar I. Insulin-like growth factor-1 deficiency and metabolic syndrome. *J Transl Med* 2016;**14**:3.
4. Poudel SB, Dixit M, Neginskaya M *et al.* Effects of GH/IGF on the aging mitochondria. *Cells* 2020;**9**:1384.
5. Sasako T, Ueki K. [Insulin/IGF-1 signaling and aging.] *Nihon Rinsho* 2016;**74**:1435–40.
6. Rosenzweig SA. The continuing evolution of insulin-like growth factor signaling. *F1000Res* 2020;**9**:F1000 Faculty Rev-205.
7. Sandhu MS, Heald AH, Gibson JM, Cruickshank JK, Dunger DB, Wareham NJ. Circulating concentrations of insulin-like growth factor-I and development of glucose intolerance: a prospective observational study. *Lancet* 2002;**359**:1740–45.
8. Rajpathak SN, McGinn AP, Strickler HD *et al.* Insulin-like growth factor-(IGF)-axis, inflammation, and glucose intolerance among older adults. *Growth Horm IGF Res* 2008;**18**:166–73.
9. Larsson SC, Michaëlsson K, Burgess S. IGF-1 and cardiometabolic diseases: a Mendelian randomisation study. *Diabetologia* 2020;**63**:1775–82.
10. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *JAMA* 2017;**318**:1925–26.
11. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;**45**:1866–86.
12. Foley CN, Mason AM, Kirk PDW, Burgess S. MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates. *Bioinformatics* 2021;**37**:531–41.
13. Noordam R, Oudt CH, Bos MM, Smit RAJ, van Heemst D. High-sensitivity C-reactive protein, low-grade systemic inflammation and type 2 diabetes mellitus: a two-sample Mendelian randomization study. *Nutr Metab Cardiovasc Dis* 2018;**28**:795–802.
14. Verkouter I, de Mutsert R, Smit RAJ *et al.* The contribution of tissue-grouped BMI-associated gene sets to cardiometabolic disease risk: a Mendelian randomization study. *Int J Epidemiol* 2020;**49**:1246–56.
15. Foley CN, Mason AM, Kirk PDW, Burgess S. MR-Clust: Clustering of genetic variants in Mendelian randomization with similar causal estimates. *Bioinformatics* 2021;**37**:531–41.
16. 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**:e1001779.
17. Therneau T (2022). *A Package for Survival Analysis in R*. R package version 3.3-1, https://CRAN.R-project.org/package=survival.
18. Walter K, Min JL, Huang J *et al.*; UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* 2015;**526**:82–90.
19. Auton A, Brooks LD, Durbin RM *et al.*; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;**526**:68–74.
20. McCarthy S, Das S, Kretzschmar W *et al.*; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;**48**:1279–83.
21. Loh PR, Tucker G, Bulik-Sullivan BK *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;**47**:284–90.
22. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer, 2016.
23. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* 2015;**31**:259–61.
24. 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**:1505–13.
25. Hemani G, Zheng J, Elsworth B *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;**7**:e34408.

26. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG; EPIC-InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol* 2015;**30**:543–52.

27. Noordam R, Smit RA, Postmus I, Trompet S, van Heemst D. Assessment of causality between serum gamma-glutamyltransferase and type 2 diabetes mellitus using publicly available data: a Mendelian randomization study. *Int J Epidemiol* 2016;**45**:1953–60.

28. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–25.

29. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;**40**:304–14.

30. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;**8**:1826.

31. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic Acids Res* 2009;**37**:D623–28.

32. Drogan D, Schulze MB, Boeing H, Pischon T. Insulin-like growth factor 1 and insulin-like growth factor–binding protein 3 in relation to the risk of type 2 diabetes mellitus: results from the EPIC-Potsdam Study. *Am J Epidemiol* 2016;**183**:553–60.

33. Similä ME, Kontto JP, Virtamo J *et al*. Insulin-like growth factor I, binding proteins-1 and -3, risk of type 2 diabetes and macronutrient intakes in men. *Br J Nutr* 2019;**121**:938–44.

34. Rajpathak SN, He M, Sun Q *et al*. Insulin-like growth factor axis and risk of type 2 diabetes in women. *Diabetes* 2012;**61**:2248–54.

35. Mulugeta A, Hyppönen E, Ala-Korpela M, Mäkinen V-P. Metabolic subgroups and cardiometabolic multimorbidity in the UK Biobank. *medRxiv* 2021.02.01: 21250893, preprint: not peer reviewed.

36. Schneider HJ, Friedrich N, Klotsche J *et al*. Prediction of incident diabetes mellitus by baseline IGF1 levels. *Eur J Endocrinol* 2011;**164**:223–29.

37. Friedrich N, Thuesen B, Jørgensen T *et al*. The association between IGF-I and insulin resistance: a general population study in Danish adults. *Diabetes Care* 2012;**35**:768–73.

38. Zhang B, Shozu M, Okada M *et al*. Insulin-like growth factor I enhances the expression of aromatase P450 by inhibiting autophagy. *Endocrinology* 2010;**151**:4949–58.

39. van Haeften TW, Twickler TB. Insulin-like growth factors and pancreas beta cells. *Eur J Clin Invest* 2004;**34**:249–55.

40. Kido Y, Nakae J, Hribal ML, Xuan S, Efstratiadis A, Accili D. Effects of mutations in the insulin-like growth factor signaling system on embryonic pancreas development and beta-cell compensation to insulin resistance. *J Biol Chem* 2002;**277**:36740–47.

41. Yakar S, Setser J, Zhao H *et al*. Inhibition of growth hormone action improves insulin sensitivity in liver IGF-1-deficient mice. *J Clin Invest* 2004;**113**:96–105.

42. Minelli C, Del Greco MF, van der Plaat DA, Bowden J, Sheehan NA, Thompson J. The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int J Epidemiol* 2021;**50**:1651–59.

43. Levine ME, Suarez JA, Brandhorst S *et al*. Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. *Cell Metab* 2014;**19**:407–17.