REVIEW

Causal factors underlying diabetes risk informed by Mendelian randomisation analysis: evidence, opportunities and challenges

Shuai Yuan[1] · Jordi Merino[2,3,4,5] · Susanna C. Larsson[1,6]

Abstract

Diabetes and its complications cause a heavy disease burden globally. Identifying exposures, risk factors and molecular processes causally associated with the development of diabetes can provide important evidence bases for disease prevention and spur novel therapeutic strategies. Mendelian randomisation (MR), an epidemiological approach that uses genetic instruments to infer causal associations between an exposure and an outcome, can be leveraged to complement evidence from observational and clinical studies. This narrative review aims to summarise the evidence on potential causal risk factors for diabetes by integrating published MR studies on type 1 and 2 diabetes, and to reflect on future perspectives of MR studies on diabetes. Despite the genetic influence on type 1 diabetes, few MR studies have been conducted to identify causal exposures or molecular processes leading to increased disease risk. In type 2 diabetes, MR analyses support causal associations of somatic, mental and lifestyle factors with development of the disease. These studies have also identified biomarkers, some of them derived from the gut microbiota, and molecular processes leading to increased disease risk. These studies provide valuable data to better understand disease pathophysiology and explore potential therapeutic targets. Because genetic association studies have mostly been restricted to participants of European descent, multi-ancestry cohorts are needed to examine the role of different types of physical activity, dietary components, metabolites, protein biomarkers and gut microbiome in diabetes development.

Keywords Causality · Diabetes · Mendelian randomisation · Review · Risk factor

Abbreviations

BCAA   Branched-chain amino acid
BCKD   Branched-chain α-ketoacid dehydrogenase
MR   Mendelian randomisation

✉ Susanna C. Larsson
    susanna.larsson@ki.se

1   Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

2   Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

3   Programs in Metabolism and Medical and Population Genetics, Eli and Edythe L. Broad Institute of MIT and Harvard, Cambridge, MA, USA

4   Department of Medicine, Harvard Medical School, Boston, MA, USA

5   Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

6   Unit of Medical Epidemiology, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

Introduction

Diabetes is a leading health issue that causes severe disease and has a huge economic burden worldwide [1, 2]. Many epidemiological studies have assessed the causes of diabetes to provide an evidence base for disease prevention. For example, in type 2 diabetes, an exposure-wide umbrella review including 142 factors identified a wide range of biomarkers, medical conditions and dietary, lifestyle, environmental and psychosocial factors that were associated with the risk of disease [3]. The picture is somewhat different for type 1 diabetes owing to the strong genetic contribution and less influence of external factors. In addition to genetic factors, only a few environmental factors, including birthweight and childhood obesity, have been linked to type 1 diabetes [4]. While results from observational studies have provided initial evidence of potential exposures associated with diabetes, residual confounding and reverse causation limit our understanding of the complex set of factors underlying the development of

Standard terms and key concepts in MR studies

Collider bias	A bias caused by genetic associations obtained from genome-wide association analysis with adjustment for certain covariates. This can also be caused by stratifying the population based on a collider or studying the association between a risk factor for the disease and disease progression
Confounding and reverse causality	These two biases can usually be minimised but may be introduced by using invalid genetic instrumental variables, such as involving genetic variants with pleiotropic associations with confounders
Horizontal pleiotropy	Instrumental variables influence the outcome not only through the exposure but also through alternative pathways. Horizontal pleiotropy can be balanced or imbalanced. Imbalanced horizontal pleiotropy generates bias in MR estimates. For example, if the genetic variants for smoking are associated with a risk factor for type 2 diabetes, such as physical activity, in an imbalanced way, the MR association between smoking and type 2 diabetes will be biased by horizontal pleiotropy. Whether the association is biased by positive (leading to an exaggerated estimate) or negative (leading to an attenuated estimate) horizontal pleiotropy can be assessed by comparing the associations in the inverse variance weighted and MR-Egger regression methods
Instrumental variables	Genetic variants (i.e. SNPs) that are strongly associated with the exposure and not associated with confounders and that affect the outcome merely via the exposure, such as cis-SNPs that encode a protein
Linkage disequilibrium	The non-random assortment of genetic variants that can be understood as the correlation of genetic variants. Linkage disequilibrium can be used as a criterion for genetic instrument selection and for identifying proxy genetic variants
MR	An epidemiological method based on observational genetic data for causal inference
MR sensitivity analyses	MR statistical analysis methods supplementing the main analysis (usually the inverse variance method). Includes a wide range of methods, such as the frequently used weighted median, weighted mode, MR-Egger, MR-PRESSO and contamination mixture methods, and newly developed methods based on summary-level data, such as the MR-Cause method. These analyses can examine the robustness of MR results and provide indications about outliers, heterogeneity and pleiotropy
Multiple instruments	The genetic score of more than one genetic instrumental variables
Multivariable MR analysis	An analysis that includes at least two traits proxied by instrumental variables and provides MR estimates after mutual adjustment for the included traits. This method helps with minimisation of pleiotropy and mediation estimation
Non-linear MR	An MR design for examining the non-linear association between the exposure and the outcome. The design requires individual-level data
One-sample MR	Genetic associations for the exposure and the outcome from one dataset
Population structure bias	This bias can be generated by using genetic association data from populations of different ancestries, in which ancestry is correlated with both phenotype and genotype. For example, the results are likely to be biased if MR analysis used genetic instruments from a European population and outcome data based on a non-European population and vice versa. This bias can be minimised by using data from populations of the same ancestry and adjusting for top-ranked genetic principal components

Standard terms and key concepts in MR studies: continued

Two-sample MR	Genetic associations for the exposure and the outcome from two independent datasets. This design can incorporate summary-level data from multiple sources and thus increase the power. However, the results may be influenced by population features of different studies
Vertical pleiotropy	Instrumental variables associated with traits in the same pathway from the exposure to the outcome. In detail, the exposure proxied by genetic instruments may influence a downstream factor and therefore influence the risk of the outcome. This type of pleiotropy does not bias MR estimates. For example, the association between genetically predicted alcohol consumption and type 2 diabetes risk is mediated by blood pressure, which does not violate the MR assumptions
Weak instrument bias	A bias caused by weak genetic instrumental variables that explain a small phenotypic variance, coupled with a small sample size in relation to the outcome. The strength of the instrumental variable can be assessed by the F statistic. The power of the analysis can be estimated

diabetes. Thus, whether the factors observed in previous observational studies are causally associated with the risk of diabetes remains unconfirmed. A clear appraisal of the causal risk factors for diabetes is of great importance for disease prevention.

Mendelian randomisation (MR) is an epidemiological method that can strengthen causal inference by using genetic variants as instrumental variables [5]. An instrumental variable is a variable that satisfies three main conditions: (1) it is associated with the exposure (relevance assumption); (2) it does not share a common cause with the outcome (independence assumption); and (3) it is related to the outcome only through the exposure (exclusion restriction assumption) (Fig. 1). The text box summarises the common terms used in MR studies and their key concepts and limitations. As genetic variants are randomly assorted at conception and thus are generally unassociated with environmental and self-adopted factors, MR is believed to be less affected by measured and unmeasured confounding factors. This narrative review aims to summarise the evidence on potential causal risk factors for diabetes by integrating published MR studies on type 1 and 2 diabetes, and to reflect on future perspectives of MR studies on diabetes.

Causal exposures and risk factors for type 1 diabetes

Because there is a strong genetic component in type 1 diabetes, MR studies of type 1 diabetes are limited and only a few potentially modifiable risk factors have been identified (Table 1). Low birthweight [6], childhood obesity [6, 7] and a higher abundance of the *Bifidobacterium* genus [8] have been associated with an increased risk of type 1 diabetes.

MR studies have found no associations of adult body size [6], features of the liver or pancreas [9] and serum 25-hydroxyvitamin D levels [10] with type 1 diabetes.

A protein-wide MR study examined the associations of 1611 circulating protein biomarkers with the risk of type 1 diabetes and identified associations for signal regulatory protein gamma, IL-27 Epstein–Barr virus-induced 3 and chymotrypsinogen B1 [11]. These findings linking certain viral infections, particularly by enteroviruses (e.g. coxsackievirus), with the risk of type 1 diabetes are consistent with recent observational studies [12], thus providing an avenue to better understand and prevent this disease.

Causal exposures and risk factors for type 2 diabetes

Most MR studies on glycaemic outcomes have focused on type 2 diabetes. Our previous exposure-wide MR study examined the associations of 97 exposures with risk of type 2 diabetes using data from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (74,124 cases and 824,006 controls). In total, 34 factors that were possibly causally associated with the risk of type 2 diabetes were identified [13]. In Table 2 we summarise and update the associations of a wide range of exposures with type 2 diabetes from MR studies on diabetes.

Somatic and psychological health status The results of MR studies of somatic and psychological health status in relation to type 2 diabetes are summarised in Table 2. Contradictory associations were reported for LDL-cholesterol and type 2 diabetes, with an inverse association observed in a European
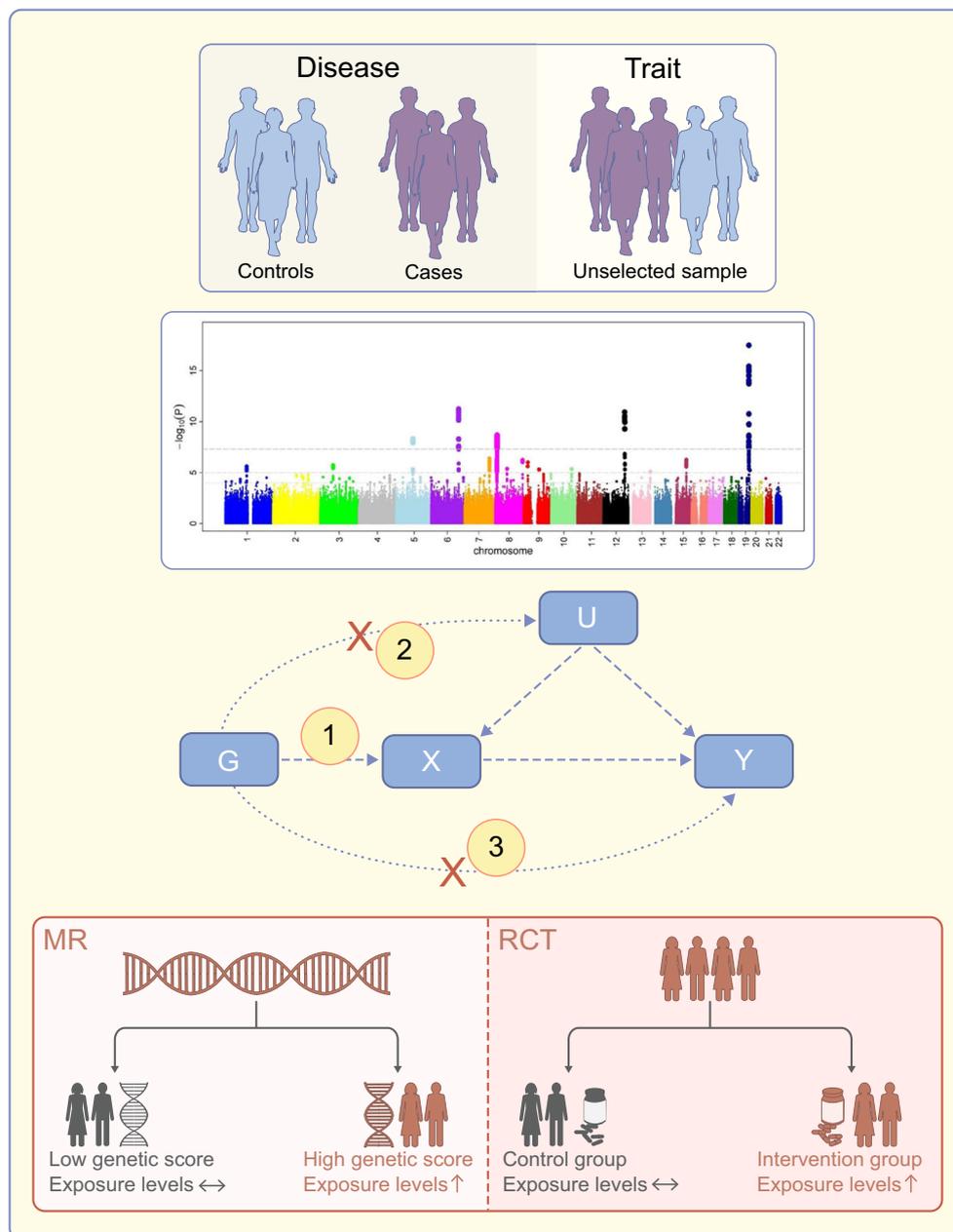
Fig. 1 Study design and assumptions of MR analysis. The process of MR analysis is shown from top to bottom. In detail, MR analysis is based on genome-wide association analyses of the exposure and outcome. Genetic instruments for the exposure are independent SNPs that are strongly associated with the exposure of interest in a genome-wide association analysis in an unselected sample, such as a general population. Likewise, summary-level data on the outcome are obtained from a genome-wide association analysis of a binary phenotype that defines the population into cases and controls. The directed acyclic graph represents the study design and assumptions of MR analysis; G indicates the genetic instruments, X indicates the exposure of interest, Y indicates the outcome of interest, and U indicates the confounders. There are three important assumptions in MR analysis. Assumption 1 indicates that the genetic variants used as the instrumental variable should be robustly associated with the exposure. Assumption 2 indicates that the instrumental variable should not be associated with any confounders. Assumption 3 indicates that the instrumental variable used should affect the risk of the outcome only through the risk factor, not through alternative pathways. Regarding causal inference, the MR design resembles that of an RCT; specifically, the random allocation of genetic variants in MR mimics the randomisation process of RCTs, which minimises confounding effects. Source: Manhattan plot reproduced from Ikram et al [75], available under a CC BY 2.5 licence (https://creativecommons.org/licenses/by/2.5/). This figure is available as a downloadable slide

population and a positive association in an African population [13–15]. A recent study further identified that the diabetogenic effect of low levels of LDL-cholesterol might be mediated by increased BMI [16]. Lower levels of bilirubin (a marker of liver function) [17], testosterone [18] and thyrotropin [19] were associated with an increased risk of type 2 diabetes in

Table 1 MR studies on the causes of type 1 diabetes

Exposure	First author, year of publication	SNPs	Cases/controls	IVW estimate, OR (95% CI)	Robustness in sensitivity analyses[a]
Birthweight	Richardson, 2022 [6]	161	15,573/158,408	0.66 (0.47, 0.92) per SD increase	Yes
Childhood obesity	Censin, 2017 [7][b]	23	5913/8828	1.32 (1.06, 1.64) per SD increase	Yes
Childhood body size	Richardson, 2022 [6]	280	15,573/158,408	2.05 (1.20, 3.50) per change in body size category (thinner, average, plumper)	Yes
Adult body size	Richardson, 2022 [6]	515	15,573/158,408	1.60 (1.05, 2.45) per change in body size category (thinner, average, plumper)	No
Liver fat	Martin, 2022 [9][b]	10	9358/15,705	1.06 (0.90, 1.27) per SD increase	Yes
Liver volume	Martin, 2022 [9][b]	11	9358/15,705	0.92 (0.67, 12.7) per SD increase	Yes
Pancreas fat	Martin, 2022 [9][b]	9	9358/15,705	1.26 (0.82, 1.93) per SD increase	No
Pancreas volume	Martin, 2022 [9][b]	16	9358/15,705	1.55 (0.85, 2.85) per SD increase	No
Serum 25OHD levels	Manousaki, 2021 [10][b]	69	9358/15,705	1.09 (0.86, 1.40) per SD decrease	Yes
Circulating protein biomarkers	Yazdanpanah, 2022 [11][b]	1 cis-pQTL for each	9358/15,705	1.66 (1.36, 2.03) per SD increase in SIRPG; 1.97 (1.48, 2.62) per SD increase in IL27.EBI3; 0.80 (0.73, 0.88) per SD increase in CTRB1	Yes
Gut microbiota	Xu, 2022 [8][b]	5	6683/12,173 (discovery); 3041/449,223 (replication)	1.61 (1.34, 1.92) for higher abundance of the *Bifidobacterium* genus	Yes
Telomere length	Haycock, 2017 [64][b]	6	7514/9045	0.71 (0.51, 0.98) per SD increase	Not reported

[a] Robustness in sensitivity analyses was defined as the association remaining stable in any of the following three sensitivity analyses: weighted median, MR-Egger and MR-PRESSO. We defined the sensitivity analyses as robust for studies on metabolites and proteins as the associations remained stable in the analyses based on different data sources

[b] Study was based on summary-level data

25OHD, 25-hydroxyvitamin D; CTRB1, chymotrypsinogen B1; IL27.EBI3, IL-27 Epstein–Barr virus-induced 3; IVW, inverse variance weighted; pQTL, protein quantitative trait loci; SIRPG, signal regulatory protein gamma

Table 2 MR studies on the causes of type 2 diabetes

Exposure	First author, year of publication	SNPs	Cases/controls	IVW estimate, OR (95% CI)	Robustness in sensitivity analyses[a]
Somatic health status					
Systolic blood pressure	Aikens, 2017 [65][b]	28	37,293/125,686	1.02 (1.01, 1.03) per 1 mmHg increase	Yes
HDL-cholesterol	Fall, 2015 [14][b]	140	34,840/114,981 (EUR)	0.86 (0.79, 0.94) per SD increase	Yes
	Soremekun, 2022 [15][b]	18	23,305/30,140 (AA)	0.92 (0.84, 0.99) per SD increase	Yes
Uric acid levels	Sluijs, 2015 [66]	24	41,508/NA	1.01 (0.87, 1.16) per 1 mg/dl (59.5 μmol/l) increase	Not reported
Rheumatoid arthritis	Zhang, 2022 [67][b]	70	74,124/824,006	1.04 (1.02, 1.05) per log-odds	Yes
Liver function	De Silva, 2019 [68]	3 for AST	64,094/607,012	1.25 (1.14, 2.38) per SD increase in AST	Yes
		4 for ALT		1.45 (1.10, 1.92) per SD increase in ALT	
		13 for ALP		0.91 (0.86, 0.97) per SD increase in ALP	
		22 for GGT		0.92 (0.80, 1.06) per SD increase in GGT	
Liver volume	Martin, 2022 [9][b]	11	55,005/400,308	1.27 (1.08, 1.49) per SD increase	Yes
Pancreas volume	Martin, 2022 [9]	17	55,005/400,308	0.76 (0.62, 0.96) per SD increase	Yes
Polycystic ovary syndrome	Zhu, 2021 [69][b]	14 for EUR 13 for EA	74,124/824,006 (EUR) 77,418/356,122 (EA)	0.97 (0.92, 1.01) per log-OR in EUR 0.98 (0.96, 1.01) per log-OR in EA	Yes
Gallstone disease	Wang, 2019 [70]	8	2233/4767 (EA)	1.17 (0.90, 1.52) per SD increase in log-OR	Not reported
Circulating bilirubin levels	Abbasi, 2015 [17]	1	210/3171	0.75 (0.62, 0.92) per SD increase in log-transformed levels	Not reported
Age at menarche	Cao, 2020 [71][b]	122	26,676/132,532	0.83 (0.78, 0.88) per 1 year increase	Not reported
Testosterone levels	Ruth, 2020 [23][b]	254 (women) 231 (men)	17,790/243,645 (women) 34,990/150,760 (men)	1.37 (1.22, 1.53) per SD increase in women 0.86 (0.76, 0.98) per SD increase in men	Yes
SHBG levels	Yuan, 2022 [24]	450	103,290/1,007,191	0.49 (0.43, 0.57) per nmol/l increase	Yes
Sleep and mental health					
Sleep duration	Yuan, 2020 [13][b]	73	74,124/824,006	0.83 (0.62, 1.12) per h/night	Yes
Short sleep (<7 h/night)	Yuan, 2020 [13][b]	25	74,124/824,006	1.14 (0.92, 1.41) per log-OR	Yes
Long sleep (>9 h/night)	Yuan, 2020 [13][b]	8	74,124/824,006	0.79 (0.47, 1.34) per log-OR	Yes
Insomnia	Yuan, 2020 [13][b]	208	74,124/824,006	1.17 (1.11, 1.23) per log-OR	Yes
Depression	Tang, 2020 [72][b]	96	74,124/824,006	1.26 (1.10, 1.43) per log-OR	Yes
Adiposity-related factors					
Birthweight	Huang, 2019 [31]	7	28,806/52,691 (EUR) 12,349/27,317 (EA)	1.96 (1.42, 2.71) per SD decrease in EUR 1.39 (1.18, 1.62) per SD decrease in EA	Yes
Childhood obesity	Geng, 2018 [25][b]	15	34,840/114,981	1.83 (1.46, 2.30) per SD increase	Yes
Adulthood overall obesity (BMI)	Wainberg, 2019 [26]	57	13,982/273,412	1.31 (1.11, 1.53) per kg/m² increase for BMI <25 kg/m² 1.36 (1.28, 1.45) per kg/m² increase for BMI 25–30 kg/m² 1.25 (1.20, 1.31) per kg/m² increase for BMI >30 kg/m²	Not reported
Adulthood overall obesity (BMI)	Dale, 2017 [27]	97	34,840/114,981	1.98 (1.41, 2.78) per SD increase	Yes
Adulthood central obesity (WHR adjusted for BMI)	Dale, 2017 [27]	49	34,840/114,981	1.82 (1.38, 2.24) per SD increase	Yes

Table 2 (continued)

Exposure	First author, year of publication	SNPs	Cases/controls	IVW estimate, OR (95% CI)	Robustness in sensitivity analyses[a]
Liver fat	Martin, 2022 [9][b]	10	55,005/400,308	1.27 (1.08, 1.49) per SD increase	Yes
Pancreas fat	Martin, 2022 [9][b]	9	55,005/400,308	1.02 (0.76, 1.37) per SD increase	Yes
Whole-body fat mass	Larsson, 2020 [29][b]	581	26,676/132,532	1.96 (1.54, 2.50) per kg/m² increase	
Visceral fat mass	Karlsson, 2019 [28]	102	6473/101,623 (men)	2.50 (1.98, 3.14) per 1 kg increase for men	Yes
			3044/106,371 (women)	7.34 (4.48, 12.0) per 1 kg increase for women	
Plasma adiponectin	Nielsen, 2021 [30]	10	62,829/596,487	1.26 (1.01, 1.57) per log-transformed decrease	No
Lifestyle factors					
Smoking initiation	Yuan, 2019 [33][b]	377	74,124/824,006	1.28 (1.20, 1.37) per SD increase in log-OR	Yes
Lifetime smoking index	Yuan, 2020 [13][b]	126	74,124/824,006	1.61 (1.36, 1.91) per SD increase	Yes
Alcohol consumption	Yuan, 2020 [13][b]	83	74,124/824,006	1.08 (0.80, 1.45) per SD increase in log drinks/week	No
Coffee consumption	Yuan, 2020 [13][b]	12	74,124/824,006	1.59 (1.09, 2.32) per 50% increase	Yes
Milk consumption	Vissers 2019 [73]	1	9686/12,134	0.99 (0.93, 1.05) per 15 g/day increase	No sensitivity analysis
Vigorous physical activity	Meisinger, 2020 [32][b]	8	74,124/824,006	0.83 (0.56, 1.23) per SD increase in accelerations	Yes
Nutritional factors					
Serum 25OHD levels	Yuan, 2019 [37][b]	7	74,124/824,006	0.94 (0.88, 0.99) per SD increase	Yes
Serum vitamin K levels	Zwakenberg, 2019 [40]	4	69,547/551,336	0.93 (0.89, 0.97) per ln nmol/l increase	Yes
Serum iron levels	Wang, 2021 [41][b]	3	74,124/824,006	1.07 (1.02, 1.12) per SD increase	Yes
Serum fatty acid levels	Yuan, 2020 [42][b]	1–4	74,124/824,006	Eight out of ten fatty acids identified: 0.86 (0.81, 0.91) per SD increase in palmitoleic acid	Influenced by *FADS1/2* genes
Amino acids	Lotta, 2016 [50]	5 (isoleucine)	47,877/267,694	1.44 (1.26, 1.65) per SD increase in isoleucine	Yes
		1 (leucine)		1.85 (1.41, 2.42) per SD increase in leucine	
		1 (valine)		1.54 (1.28, 1.84) per SD increase in valine	
Amino acids	Yuan, 2020 [13][b]	1–2	74,124/824,006	0.51 (0.45, 0.58) per SD increase in alanine	No sensitivity analysis
				1.15 (1.03, 1.29) per SD increase in phenylalanine	
				0.86 (0.76, 0.98) per SD increase in tyrosine	
Inflammatory and protein biomarkers					
IGF-1 levels	Larsson, 2020 [46][b]	416	74,124/824,006	1.14 (1.05, 1.24) per SD increase	Yes
C-reactive protein	Noordam, 2018 [74][b]	15	12,171/56,862	1.15 (0.93, 1.42) per ln mg/l increase	No
IL-1 receptor antagonist	Yuan, 2020 [13][b]	2	74,124/824,006	1.13 (1.01, 1.27) per 0.22 SD increase	No sensitivity analysis
IL-6 receptor	Swerdlow, 2012 [49]	1	12,859/86,807	0.97 (0.94, 1.00) per minor allele	No sensitivity analysis
Circulating protein biomarkers	Ghanbari, 2022 [58][b]	1 for each	74,124/824,006	20 of 1089 proteins identified	Yes

[a] Robustness in sensitivity analyses was defined as the association remaining stable in any of the following three sensitivity analyses: weighted median, MR-Egger and MR-PRESSO. We defined the sensitivity analyses as robust for studies on metabolites and proteins as the associations remained stable in the analyses based on different data sources. No sensitivity analysis means that the sensitivity analysis could not be performed because of a limited number of genetic variants

[b] Study based on summary-level data

25OHD, 25-hydroxyvitamin D; AA, African American; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; EA, East Asian; EUR, European; GGT, gamma-glutamyl transferase; NA, not available; SHBG, sex hormone-binding globulin

some MR studies, but not all [13, 20–22]. Sex-specific associations were observed for testosterone [23, 24], with an increased risk of type 2 diabetes in women but a decreased risk in men with higher testosterone levels [23]. Insomnia, but no other sleep-related traits, was associated with type 2 diabetes [13].

Adiposity-related factors Similar to the large body of evidence from prospective observational studies, childhood obesity, adulthood overall obesity and central obesity, excessive liver fat and whole-body and visceral fat mass were all associated with an increased risk of type 2 diabetes [9, 25–29]. Plasma levels of adiponectin, an adipocyte-secreted hormone, are decreased in individuals with obesity, which was associated with an increased risk of type 2 diabetes [30]. However, this association was inconsistent in MR sensitivity analyses [30], suggesting that the association may be biased by pleiotropy (e.g. from fat mass). Several MR studies have found that lower birthweight, independent of adult body weight, is associated with a higher risk of type 2 diabetes [31], which may suggest a role of the uterine environment and fetal development in the development of type 2 diabetes.

Lifestyle and nutritional factors MR studies have strengthened the causal role of cigarette smoking in type 2 diabetes and failed to convincingly confirm the effects of physical activity and alcohol and coffee consumption on type 2 diabetes risk [13, 32, 33]. Although alcohol consumption instrumented by 83 SNPs was not associated with type 2 diabetes, the main SNP that associates with higher alcohol consumption and alcohol abuse in European populations (i.e. rs1229984 in the *ADH1B* gene) was significantly associated with an increased risk of disease [13]. A robust inverse association between coffee consumption and type 2 diabetes risk has been reported in many observational studies [34]. However, genetically predicted higher coffee consumption was not associated with a decreased risk of type 2 diabetes in MR studies [13, 35]. Pleiotropic effects of the SNPs used may cause this lack of association (e.g. from fat mass or other hot beverages or caffeine-containing drinks) and the inverse relationship between genetically proxied coffee consumption and plasma caffeine levels (i.e. the genetic variants with the strongest association with higher coffee consumption are associated with lower plasma caffeine levels) [36].

An MR study found an inverse association between circulating 25-hydroxyvitamin D levels and type 2 diabetes risk [37], and this association might be driven by the vitamin D synthesis pathway [37–39]. Lower levels of vitamin K1 (phylloquinone) [40] and higher levels of iron [41] were associated with an increased risk of type 2 diabetes. Eight out of ten plasma fatty acids were found to be associated with type 2

diabetes; however, the associations, with the exception of palmitoleic acid, were driven by SNPs in the *FADS1/2* genes [42]. Thus, whether these associations were biased by this pleiotropic gene, which encodes a key enzyme in fatty acid metabolism, remains unknown [43].

Despite the popularity of MR studies for investigating dietary and lifestyle exposures in diabetes and cardiometabolic diseases, there are unique challenges in such studies of these time-varying, compositional and intercorrelated exposures [44]. For example, MR analyses of nutritional exposures based on genetic instruments for a single measure of diet collected in midlife bear an underlying assumption that, on average, the dietary assessment tool is representative of long-term habitual intake. Furthermore, like many behavioural exposures, nutrition is intercorrelated with numerous other lifestyle and environmental factors. Recent studies have documented that confounding and reverse causation affecting traditional epidemiological studies may also impact genetic associations [45]. A recent study has shown that half of the genetic variants associated with diet are the consequence of increased BMI and that it is possible to use genetics to correct for confounding and reverse causation to strengthen genetic correlations and causal inference [45].

IGF-1 and inflammatory biomarkers Genetically predicted elevated levels of IGF-1, a peptide hormone similar in molecular structure to insulin, were positively associated with the risk of type 2 diabetes [46]. Given the heterogeneous effects of IGF-1-associated SNPs on type 2 diabetes, a recent MR analysis examined several clusters of IGF-1-associated SNPs in relation to type 2 diabetes and specified that this overall positive association might be explained by pathways related to amino acid metabolism and genomic integrity [47]. However, the main cluster of IGF-1-associated SNPs that were associated with a decreased risk of type 2 diabetes mapped to the growth hormone signalling pathway [47], possibly mediated by pleiotropic effects from fat mass, as growth hormone secretion is decreased in obesity [48].

As for inflammatory biomarkers, the IL-1 and IL-6 pathways may be involved in the development of type 2 diabetes [13, 49], even though the evidence is weak. One additional minor allele of the *IL6R* SNP rs7529229 (corresponding to the effect of taking tocilizumab 4–8 mg/kg every 4 weeks) was suggestively associated with a reduced risk of type 2 diabetes (OR 0.97, 95% CI 0.94, 1.00), which implied a possible role of IL-6 receptor blockade in type 2 diabetes prevention.

Circulating metabolites and proteins One of the first demonstrations of the use of MR to study circulating metabolites was in relation to the previously reported epidemiological association between plasma levels of branched-chain amino acids

(BCAAs) and the risk of type 2 diabetes [50]. In an MR analysis using genetic variation at the *PPM1K* locus (which encodes a mitochondrial phosphatase that activates branched-chain α-ketoacid dehydrogenase [BCKD]), an increase in leucine, isoleucine and valine levels was associated with an increased odds of type 2 diabetes [50]. However, given that BCKD has a range of substrates besides leucine, isoleucine and valine, untangling which of these substrates causes type 2 diabetes is challenging. A separate MR analysis of BCAAs showed that higher BCAA levels have no causal effects on insulin resistance but, rather, genetically raised insulin resistance drives higher circulating fasting BCAA levels [51]. A metabolome-wide MR approach confirmed evidence of the strong reverse causal effect, indicating that the genetic predisposition to type 2 diabetes may trigger early changes in valine and leucine [52]. Other products of amino acid catabolism, such as 2-aminoadipic acid (2-AAA) or α-hydroxybutyrate, are strongly associated with incident type 2 diabetes in observational studies [53], but MR studies have failed to demonstrate evidence of causality [54]. There are many reasons for the discrepancies between observational studies and MR studies, but the fact that observational studies have been conducted in a mixture of individuals with normoglycaemia and impaired glucose tolerance could explain these differences. A study in the Framingham cohort restricted to individuals with strict normoglycaemia at baseline (fasting glucose <5.6 mmol/l) provided evidence of a subset of 19 metabolites associated with the risk of diabetes among apparently healthy individuals [55]. Pathway enrichment analyses and MR showed that metabolites in the nitrogen metabolism pathway are causally related to the development of diabetes [55].

Integration of genomic and small molecule data across platforms enables the discovery of regulators of human metabolism and translation into clinical insights. A recent genome-wide meta-analysis of 174 metabolite levels across six cohorts, including up to 86,507 participants, identified ~500 genetic loci influencing metabolite levels [56]. Among many relevant findings for dysglycaemia, the study provided evidence that a missense p.Asp470Asn (rs17681684) variant in the *GLP2R* gene, which encodes the receptor for glucagon-like peptide 2, was associated with a 4% higher type 2 diabetes risk. Findings from a metabolome-wide MR analysis further identified new metabolites that potentially play a causal role in type 2 diabetes, including betaine, glutamic acid, lysine, alanine and mannose [52].

High-throughput detection and quantification of serum proteins in a large human population can provide insight into the molecular processes underlying diabetes risk. A protein-wide MR study examined the associations of 164 proteins with genome-wide association summary statistics available from the independent INTERVAL study and identified 16 proteins as potentially having a causal effect on the development of type 2 diabetes [57]. A recent protein-wide MR study examined the associations of 1089 circulating protein biomarkers with the risk of type 2 diabetes [58]. The analyses identified 20 proteins that might be causally associated with type 2 diabetes. These findings may provide evidence to support therapeutic development in type 2 diabetes.

MR studies on circulating metabolites and proteins usually employ a *cis*-variant located in an encoding gene region as the instrumental variable, which satisfies three key assumptions of MR. However, these MR associations can still be influenced by the genome-wide associations analyses on metabolites and proteins as well as corresponding profiling process (possible bias caused by batch effects) [59] and different high-throughput platforms [60]. Of note, using *cis*-variants as instrumental variables may not always completely rule out horizontal pleiotropy, especially when one gene regulates several metabolites and proteins that are not in a common pathway. In this case, multivariable MR analysis or removing the pleiotropic SNPs may help reduce this bias.

Gut microbiota and related metabolites With increasing evidence suggesting that the human gut microbiome plays a role in immune function and metabolic disease, there is a need to discriminate between microbiome features that are causal for disease and those that are a consequence of disease or its treatment. A study including genome-wide genetic data, gut metagenomic sequencing and measurements of faecal short-chain fatty acids showed that a host genetic-driven increase in gut production of butyrate was associated with improved insulin response following an oral glucose test. In contrast, abnormalities in the production or absorption of propionate were causally related to an increased risk of type 2 diabetes [61]. Another two-sample MR study identified seven genera of gut microbiota nominally associated with type 2 diabetes [62]. For gut microbiota-related metabolites, a separate study found that genetically predicted higher trimethylamine *N*-oxide and carnitine levels were not associated with higher odds of type 2 diabetes. However, the study found possible associations of high choline and low betaine levels with an increased risk of type 2 diabetes [63]. Of note, although many genome-wide association analyses of the gut microbiome have been carried out, high-quality MR studies on the gut microbiome in relation to diabetes are limited [8]. This may raise doubt over the applicability of host genetic variants as an instrumental variable to mimic the function of the gut microbiome.

Assessment of included MR studies on diabetes

The overall quality of the MR studies included was satisfactory, with careful genetic instrument selection criteria, comparatively large sample sizes and different approaches to

testing the robustness of the findings. As for the examination of the assumptions of MR, assumption 1 was usually found to be satisfied by using genetic variants associated with the exposure of interest at the genome-wide significance level. However, there was no unified threshold for linkage disequilibrium of SNPs. Using a high or low threshold of linkage disequilibrium could lead to an inflated rate of type 1 and 2 errors, respectively. As MR analysis can minimise confounding, the associations are less likely to be biased by confounding but cannot be completely immune to this bias, especially when genetic instruments have large pleiotropy effects. Except for studies using individual-level data, whether genetic instruments were primarily associated with other phenotypes or were associated with confounders was rarely examined in these MR studies. The most common bias in MR analysis is horizontal pleiotropy caused by violation of assumption 3, the exclusion restriction assumption, which means that genetic variants affect the outcome through alternative pathways, not only through the exposure of interest. The associations with type 1 and 2 diabetes summarised in this review were robust in sensitivity analyses, and most studies used MR-Egger or MR pleiotropy residual sum and outlier (MR-PRESSO) to detect potential horizontal pleiotropy. Of note, even though statistical methods can detect and minimise the influence of horizontal pleiotropy, instrumental variable selection is a crucial process for reducing the bias. Using genetic variants in genes with well-understood biological functions as instrumental variables usually satisfies the assumptions of MR analysis and thus generates precise and correct associations. However, it is difficult to identify specific genetic variants for certain exposures, especially for health behaviours and complex phenotypes. Therefore, a thoughtful examination of pleiotropy should be conducted in analyses using multiple genetic instruments. Evidence from observational studies and clinical trials should be used in interpreting MR findings. Robust MR findings, in turn, should be examined in clinical trials. In addition, it is tricky to interpret MR results, especially for binary exposures. Given that the exposure in MR analysis is not an exact phenotype but is proxied by the effects of genetic variants on a certain trait, this genetically proxied exposure usually mimics a lifetime chronic effect, which hinders the exploration of time-specific associations.

Future perspectives

- The null findings in previous MR studies may have been caused by inadequate power, particularly for weak associations of exposures proxied by a few SNPs that explained a small phenotypic variance. For exposures with robust associations in traditional observational studies, the neutral associations in MR studies deserve to be re-examined in well-powered studies with robust genetic

instruments for the exposures and large sample sizes for the diabetes outcomes.

- Most previous MR studies were based on summary-level data, which do not allow the exploration of potential non-linear associations (e.g. J- or U-shaped); rather, it can only be assumed that the association is linear without a threshold effect. MR analysis using individual-level data from large-scale biobanks and studies is needed to examine the non-linearity of the associations.
- More effort should be put into MR studies on non-heritable exposures or exposures without genetic association information. For example, MR analyses of the association of diet and physical activity with diabetes risk are warranted.
- Most MR studies have been based on data from European populations. With more and more data available from other populations, such as Asian and African populations, future MR studies are encouraged to include data from multi-ancestry cohorts.
- Even though the associations between protein biomarkers and diabetes risk were examined in a few MR studies [11, 58], more independent verification is needed to confirm these findings. In addition, the intermediate roles of blood proteins and metabolites in the pathways from environmental exposure to diabetes should be investigated to provide evidence for treatment and intervention.
- Even though many statistical approaches, such as the weighted median, MR-Egger, MR-PRESSO, MR-Cluster and contamination mixture methods, have been developed to detect pleiotropy and verify the association with different assumptions, more efforts are needed to generate new statistical approaches to handle pleiotropy and other limitations.

Conclusion

This review has integrated data from published MR studies on type 1 and 2 diabetes to highlight the many possible causal risk factors for dysglycaemia. While few studies have been conducted for type 1 diabetes, most MR analyses support that social, demographic, metabolic and lifestyle factors are causally associated with the development of type 2 diabetes. More MR studies in multi-ancestry cohorts are needed to examine the role of diet in the development of diabetes. MR investigations based on data on metabolites, protein biomarkers and the gut microbiome may help to illustrate the pathological molecular basis of diabetes.

References

1. Ferguson D, Finck BN (2021) Emerging therapeutic approaches for the treatment of NAFLD and type 2 diabetes mellitus. Nat Rev Endocrinol 17(8):484–495. https://doi.org/10.1038/s41574-021-00507-z

2. Sun H, Saeedi P, Karuranga S et al (2022) IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 183:109119. https://doi.org/10.1016/j.diabres.2021.109119

3. Bellou V, Belbasis L, Tzoulaki I, Evangelou E (2018) Risk factors for type 2 diabetes mellitus: an exposure-wide umbrella review of meta-analyses. PLoS One 13(3):e0194127. https://doi.org/10.1371/journal.pone.0194127

4. Rewers M, Ludvigsson J (2016) Environmental risk factors for type 1 diabetes. Lancet 387(10035):2340–2348. https://doi.org/10.1016/s0140-6736(16)30507-4

5. Burgess S, Thompson SG (2015) Mendelian randomization: methods for using genetic variants in causal estimation. Chapman and Hall/CRC, London, UK

6. Richardson TG, Crouch DJM, Power GM et al (2022) Childhood body size directly increases type 1 diabetes risk based on a lifecourse Mendelian randomization approach. Nat Commun 13(1):2337. https://doi.org/10.1038/s41467-022-29932-y

7. Censin JC, Nowak C, Cooper N, Bergsten P, Todd JA, Fall T (2017) Childhood adiposity and risk of type 1 diabetes: a Mendelian randomization study. PLoS Med 14(8):e1002362. https://doi.org/10.1371/journal.pmed.1002362

8. Xu Q, Ni JJ, Han BX et al (2021) Causal relationship between gut microbiota and autoimmune diseases: a two-sample Mendelian randomization study. Front Immunol 12:746998. https://doi.org/10.3389/fimmu.2021.746998

9. Martin S, Sorokin EP, Thomas EL et al (2022) Estimating the effect of liver and pancreas volume and fat content on risk of diabetes: a Mendelian randomization study. Diabetes Care 45(2):460–468. https://doi.org/10.2337/dc21-1262

10. Manousaki D, Harroud A, Mitchell RE et al (2021) Vitamin D levels and risk of type 1 diabetes: a Mendelian randomization study. PLoS Med 18(2):e1003536. https://doi.org/10.1371/journal.pmed.1003536

11. Yazdanpanah N, Yazdanpanah M, Wang Y et al (2022) Clinically relevant circulating protein biomarkers for type 1 diabetes: evidence from a two-sample Mendelian randomization study. Diabetes Care 45(1):169–177. https://doi.org/10.2337/dc21-1049

12. Op de Beeck A, Eizirik DL (2016) Viral infections in type 1 diabetes mellitus–why the β cells? Nat Rev Endocrinol 12(5):263–273. https://doi.org/10.1038/nrendo.2016.30

13. Yuan S, Larsson SC (2020) An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. Diabetologia 63(11):2359–2371. https://doi.org/10.1007/s00125-020-05253-x

14. Fall T, Xie W, Poon W et al (2015) Using genetic variants to assess the relationship between circulating lipids and type 2 diabetes. Diabetes 64(7):2676–2684. https://doi.org/10.2337/db14-1710

15. Soremekun O, Karhunen V, He Y et al (2022) Lipid traits and type 2 diabetes risk in African ancestry individuals: a Mendelian randomization study. EBioMedicine 78:103953. https://doi.org/10.1016/j.ebiom.2022.103953

16. Wu P, Moon JY, Daghlas I et al (2022) Obesity partially mediates the diabetogenic effect of lowering LDL cholesterol. Diabetes Care 45(1):232–240. https://doi.org/10.2337/dc21-1284

17. Abbasi A, Deetman PE, Corpeleijn E et al (2015) Bilirubin as a potential causal factor in type 2 diabetes risk: a Mendelian randomization study. Diabetes 64(4):1459–1469. https://doi.org/10.2337/db14-0228

18. Syed AAS, He L, Shi Y (2020) The potential effect of aberrant testosterone levels on common diseases: a Mendelian randomization study. Genes (Basel) 11(7):721. https://doi.org/10.3390/genes11070721

19. Kuś A, Marouli E, Del Greco MF et al (2021) Variation in normal range thyroid function affects serum cholesterol levels, blood pressure, and type 2 diabetes risk: a Mendelian randomization study. Thyroid 31(5):721–731. https://doi.org/10.1089/thy.2020.0393

20. Mohammadi-Shemirani P, Chong M, Pigeyre M, Morton RW, Gerstein HC, Paré G (2020) Effects of lifelong testosterone exposure on health and disease using Mendelian randomization. Elife 9:e58914. https://doi.org/10.7554/eLife.58914

21. Hou L, Li H, Si S et al (2021) Exploring the causal pathway from bilirubin to CVD and diabetes in the UK biobank cohort study: observational findings and Mendelian randomization studies. Atherosclerosis 320:112–121. https://doi.org/10.1016/j.atherosclerosis.2020.12.005

22. Bos MM, Smit RAJ, Trompet S, van Heemst D, Noordam R (2017) Thyroid signaling, insulin resistance, and 2 diabetes mellitus: a Mendelian randomization study. J Clin Endocrinol Metab 102(6):1960–1970. https://doi.org/10.1210/jc.2016-2816

23. Ruth KS, Day FR, Tyrrell J et al (2020) Using human genetics to understand the disease impacts of testosterone in men and women. Nat Med 26(2):252–258. https://doi.org/10.1038/s41591-020-0751-5

24. Yuan S, Wang L, Sun J et al (2022) Genetically predicted sex hormone levels and health outcomes: phenome-wide Mendelian randomization investigation. Int J Epidemiol. https://doi.org/10.1093/ije/dyac036

25. Geng T, Smith CE, Li C, Huang T (2018) Childhood BMI and adult type 2 diabetes, coronary artery diseases, chronic kidney disease, and cardiometabolic traits: a Mendelian randomization analysis. Diabetes Care 41(5):1089–1096. https://doi.org/10.2337/dc17-2141

26. Wainberg M, Mahajan A, Kundaje A et al (2019) Homogeneity in the association of body mass index with type 2 diabetes across the UK Biobank: a Mendelian randomization study. PLoS Med 16(12):e1002982. https://doi.org/10.1371/journal.pmed.1002982

27. Dale CE, Fatemifar G, Palmer TM et al (2017) Causal associations of adiposity and body fat distribution with coronary heart disease, stroke subtypes, and type 2 diabetes mellitus: a Mendelian

randomization analysis. Circulation 135(24):2373–2388. https://doi.org/10.1161/circulationaha.116.026560

28. Karlsson T, Rask-Andersen M, Pan G et al (2019) Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. Nat Med 25(9):1390–1395. https://doi.org/10.1038/s41591-019-0563-7

29. Larsson SC, Burgess S (2020) Fat mass and fat-free mass in relation to cardiometabolic diseases: a two-sample Mendelian randomization study. J Intern Med 288(2):260–262. https://doi.org/10.1111/joim.13078

30. Nielsen MB, Çolak Y, Benn M, Nordestgaard BG (2021) Low plasma adiponectin in risk of type 2 diabetes: observational analysis and one- and two-sample Mendelian randomization analyses in 756,219 individuals. Diabetes 70(11):2694–2705. https://doi.org/10.2337/db21-0131

31. Huang T, Wang T, Zheng Y et al (2019) Association of birth weight with type 2 diabetes and glycemic traits: a Mendelian randomization study. JAMA Netw Open 2(9):e1910915. https://doi.org/10.1001/jamanetworkopen.2019.10915

32. Meisinger C, Linseisen J, Leitzmann M, Baurecht H, Baumeister SE (2020) Association of physical activity and sedentary behavior with type 2 diabetes and glycemic traits: a two-sample Mendelian randomization study. BMJ Open Diabetes Res Care 8(2):e001896. https://doi.org/10.1136/bmjdrc-2020-001896

33. Yuan S, Larsson SC (2019) A causal relationship between cigarette smoking and type 2 diabetes mellitus: a Mendelian randomization study. Sci Rep 9(1):19342. https://doi.org/10.1038/s41598-019-56014-9

34. Carlström M, Larsson SC (2018) Coffee consumption and reduced risk of developing type 2 diabetes: a systematic review with meta-analysis. Nutr Rev 76(6):395–417. https://doi.org/10.1093/nutrit/nuy014

35. Wang X, Jia J, Huang T (2022) Coffee types and type 2 diabetes mellitus: large-scale cross-phenotype association study and Mendelian randomization analysis. Front Endocrinol (Lausanne) 13:818831. https://doi.org/10.3389/fendo.2022.818831

36. Cornelis MC, Munafo MR (2018) Mendelian randomization studies of coffee and caffeine consumption. Nutrients 10(10):1343. https://doi.org/10.3390/nu10101343

37. Yuan S, Jiang X, Michaelsson K, Larsson SC (2019) Genetic prediction of serum 25-hydroxyvitamin D, calcium, and parathyroid hormone levels in relation to development of type 2 diabetes: a Mendelian randomization study. Diabetes Care 42(12):2197–2203. https://doi.org/10.2337/dc19-1247

38. Lu L, Bennett DA, Millwood IY et al (2018) Association of vitamin D with risk of type 2 diabetes: a Mendelian randomisation study in European and Chinese adults. PLoS Med 15(5):e1002566. https://doi.org/10.1371/journal.pmed.1002566

39. Afzal S, Brøndum-Jacobsen P, Bojesen SE, Nordestgaard BG (2014) Vitamin D concentration, obesity, and risk of diabetes: a Mendelian randomisation study. Lancet Diabetes Endocrinol 2(4):298–306. https://doi.org/10.1016/s2213-8587(13)70200-6

40. Zwakenberg SR, Remmelzwaal S, Beulens JWJ et al (2019) Circulating phylloquinone concentrations and risk of type 2 diabetes: a Mendelian randomization study. Diabetes 68(1):220–225. https://doi.org/10.2337/db18-0543

41. Wang X, Fang X, Zheng W et al (2021) Genetic support of a causal relationship between iron status and type 2 diabetes: a Mendelian randomization study. J Clin Endocrinol Metab 106(11):e4641–e4651. https://doi.org/10.1210/clinem/dgab454

42. Yuan S, Larsson SC (2020) Association of genetic variants related to plasma fatty acids with type 2 diabetes mellitus and glycaemic traits: a Mendelian randomisation study. Diabetologia 63(1):116–123. https://doi.org/10.1007/s00125-019-05019-0

43. Jäger S, Cuadrat R, Hoffmann P, Wittenbecher C, Schulze MB (2020) Desaturase activity and the risk of type 2 diabetes and coronary artery disease: a Mendelian randomization study. Nutrients 12(8):2261. https://doi.org/10.3390/nu12082261

44. Merino J, Tobias DK (2022) The unique challenges of studying the genetics of diet and nutrition. Nat Med 28(2):221–222. https://doi.org/10.1038/s41591-021-01626-w

45. Pirastu N, McDonnell C, Grzeszkowiak EJ et al (2022) Using genetic variation to disentangle the complex relationship between food intake and health outcomes. PLoS Genet 18(6):e1010162. https://doi.org/10.1371/journal.pgen.1010162

46. Larsson SC, Michaëlsson K, Burgess S (2020) IGF-1 and cardiometabolic diseases: a Mendelian randomisation study. Diabetologia 63(9):1775–1782. https://doi.org/10.1007/s00125-020-05190-9

47. Wang W, Tesfay EB, van Klinken JB et al (2022) 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. Int J Epidemiol. https://doi.org/10.1093/ije/dyac119

48. Frystyk J, Brick DJ, Gerweck AV, Utz AL, Miller KK (2009) Bioactive insulin-like growth factor-I in obesity. J Clin Endocrinol Metab 94(8):3093–3097. https://doi.org/10.1210/jc.2009-0614

49. Swerdlow DI, Holmes MV, Kuchenbaecker KB et al (2012) The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet 379(9822):1214–1224. https://doi.org/10.1016/s0140-6736(12)60110-x

50. Lotta LA, Scott RA, Sharp SJ et al (2016) Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis. PLoS Med 13(11):e1002179. https://doi.org/10.1371/journal.pmed.1002179

51. Mahendran Y, Jonsson A, Have CT et al (2017) Genetic evidence of a causal effect of insulin resistance on branched-chain amino acid levels. Diabetologia 60(5):873–878. https://doi.org/10.1007/s00125-017-4222-6

52. Porcu E, Gilardi F, Darrous L et al (2021) Triangulating evidence from longitudinal and Mendelian randomization studies of metabolomic biomarkers for type 2 diabetes. Sci Rep 11(1):6197. https://doi.org/10.1038/s41598-021-85684-7

53. Wang TJ, Ngo D, Psychogios N et al (2013) 2-Aminoadipic acid is a biomarker for diabetes risk. J Clin Invest 123(10):4309–4317. https://doi.org/10.1172/jci64801

54. Shi M, Wang C, Mei H et al (2022) Genetic architecture of plasma alpha-aminoadipic acid reveals a relationship with high-density lipoprotein cholesterol. J Am Heart Assoc 11(11):e024388. https://doi.org/10.1161/jaha.121.024388

55. Merino J, Leong A, Liu CT et al (2018) Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. Diabetologia 61(6):1315–1324. https://doi.org/10.1007/s00125-018-4599-x

56. Lotta LA, Pietzner M, Stewart ID et al (2021) A cross-platform approach identifies genetic regulators of human metabolism and health. Nat Genet 53(1):54–64. https://doi.org/10.1038/s41588-020-00751-5

57. Gudmundsdottir V, Zaghlool SB, Emilsson V et al (2020) Circulating protein signatures and causal candidates for type 2 diabetes. Diabetes 69(8):1843–1853. https://doi.org/10.2337/db19-1070

58. Ghanbari F, Yazdanpanah N, Yazdanpanah M, Richards JB, Manousaki D (2022) Connecting genomics and proteomics to identify protein biomarkers for adult and youth-onset type 2 diabetes: a two-sample Mendelian randomization study. Diabetes 71(6):1324–1337. https://doi.org/10.2337/db21-1046

59. Leek JT, Scharpf RB, Bravo HC et al (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 11(10):733–739. https://doi.org/10.1038/nrg2825

60. Pietzner M, Wheeler E, Carrasco-Zanini J et al (2021) Synergistic insights into human health from aptamer- and antibody-based

proteomic profiling. Nat Commun 12(1):6822. https://doi.org/10.1038/s41467-021-27164-0

61. Sanna S, van Zuydam NR, Mahajan A et al (2019) Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet 51(4):600–605. https://doi.org/10.1038/s41588-019-0350-x

62. Yang Q, Lin SL, Kwok MK, Leung GM, Schooling CM (2018) The roles of 27 genera of human gut microbiota in ischemic heart disease, type 2 diabetes mellitus, and their risk factors: a Mendelian randomization study. Am J Epidemiol 187(9):1916–1922. https://doi.org/10.1093/aje/kwy096

63. Jia J, Dou P, Gao M et al (2019) Assessment of causal direction between gut microbiota-dependent metabolites and cardiometabolic health: a bidirectional Mendelian randomization analysis. Diabetes 68(9):1747–1755. https://doi.org/10.2337/db19-0153

64. Haycock PC, Burgess S, Nounu A et al (2017) Association between telomere length and risk of cancer and non-neoplastic diseases: a Mendelian randomization study. JAMA Oncol 3(5):636–651. https://doi.org/10.1001/jamaoncol.2016.5945

65. Aikens RC, Zhao W, Saleheen D et al (2017) Systolic blood pressure and risk of type 2 diabetes: a Mendelian randomization study. Diabetes 66(2):543–550. https://doi.org/10.2337/db16-0868

66. Sluijs I, Holmes MV, van der Schouw YT et al (2015) A Mendelian randomization study of circulating uric acid and type 2 diabetes. Diabetes 64(8):3028–3036. https://doi.org/10.2337/db14-0742

67. Zhang K, Jia Y, Wang R et al (2022) Rheumatoid arthritis and the risk of major cardiometabolic diseases: a Mendelian randomization study. Scand J Rheumatol 1–7. https://doi.org/10.1080/03009742.2022.2070988

68. De Silva NMG, Borges MC, Hingorani AD et al (2019) Liver function and risk of type 2 diabetes: bidirectional Mendelian randomization study. Diabetes 68(8):1681–1691. https://doi.org/10.2337/db18-1048

69. Zhu T, Cui J, Goodarzi MO (2021) Polycystic ovary syndrome and risk of type 2 diabetes, coronary heart disease, and stroke. Diabetes 70(2):627–637. https://doi.org/10.2337/db20-0800

70. Wang F, Wang J, Li Y et al (2019) Gallstone disease and type 2 diabetes risk: a Mendelian Randomization study. Hepatology 70(2):610–620. https://doi.org/10.1002/hep.30403

71. Cao M, Cui B (2020) Negative effects of age at menarche on risk of cardiometabolic diseases in adulthood: a Mendelian randomization study. J Clin Endocrinol Metab 105(2):dgz071. https://doi.org/10.1210/clinem/dgz071

72. Tang B, Yuan S, Xiong Y, He Q, Larsson SC (2020) Major depressive disorder and cardiometabolic diseases: a bidirectional Mendelian randomisation study. Diabetologia 63(7):1305–1311. https://doi.org/10.1007/s00125-020-05131-6

73. Vissers LET, Sluijs I, van der Schouw YT et al (2019) Dairy product intake and risk of type 2 diabetes in EPIC-InterAct: a Mendelian randomization study. Diabetes Care 42(4):568–575. https://doi.org/10.2337/dc18-2034

74. Noordam R, Oudt CH, Bos MM, Smit RAJ, van Heemst D (2018) High-sensitivity C-reactive protein, low-grade systemic inflammation and type 2 diabetes mellitus: a two-sample Mendelian randomization study. Nutr Metab Cardiovasc Dis 28(8):795–802. https://doi.org/10.1016/j.numecd.2018.03.008

75. Ikram MK, Xueling S, Jensen RA et al (2010) Four novel loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. PLoS Genet 6(10):e1001184. https://doi.org/10.1371/journal.pgen.1001184