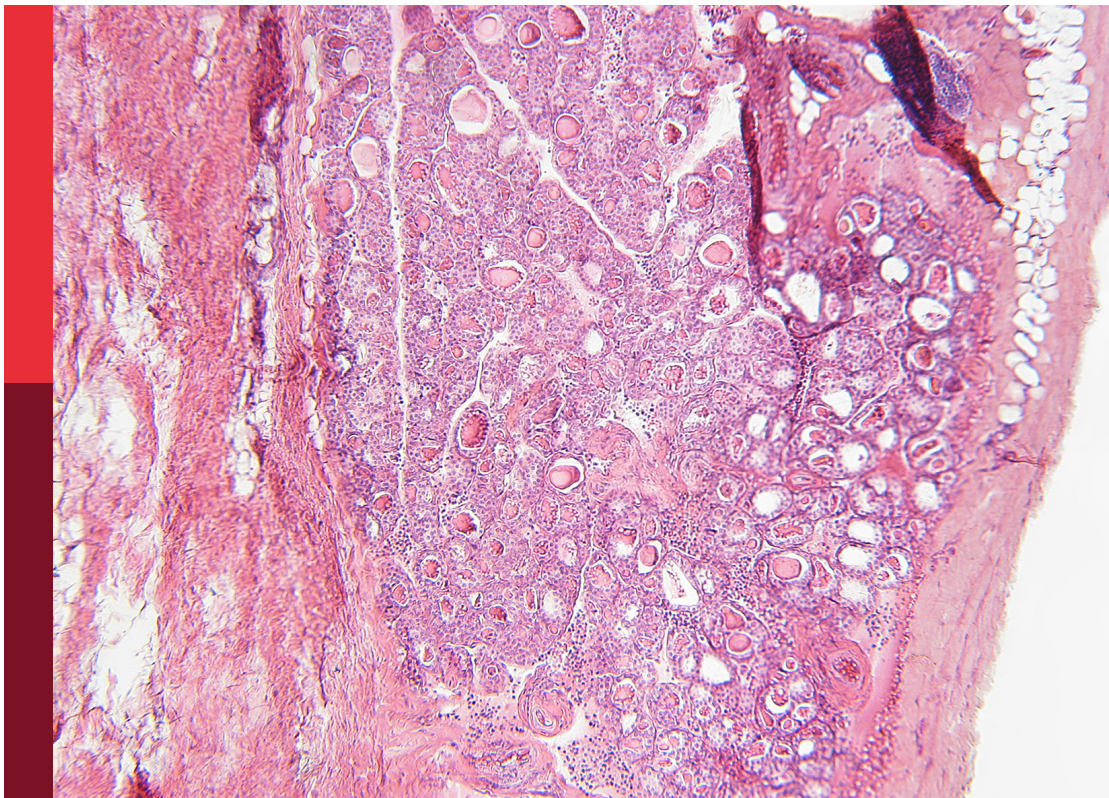
frontiers | Research Topics

Genetic variants and metabolic diseases, volume II

Edited by
Tarunveer Singh Ahluwalia and Kavita Jadhav

Coordinated by
Marta Szydlowska and Sarina Kajani

Published in
Frontiers in Endocrinology



ISSN 1664-8714
ISBN 978-2-8325-4467-9
DOI 10.3389/978-2-8325-4467-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Genetic variants and metabolic diseases, volume II

Topic editors

Tarunveer Singh Ahluwalia — Steno Diabetes Center Copenhagen (SDCC), Denmark

Kavita Jadhav — Epic-Bio, United States

Topic coordinators

Marta Szydlowska — Metabolism Bioscience, Early Research and Development, Cardiovascular, Renal and Metabolism, AstraZeneca, United States

Sarina Kajani — AstraZeneca, United States

Citation

Ahluwalia, T. S., Jadhav, K., Szydlowska, M., Kajani, S., eds. (2024). *Genetic variants and metabolic diseases, volume II*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-4467-9

Table of contents

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fendo.2023.1101934

OPEN ACCESS

EDITED BY
Kavita Jadhav,
AstraZeneca, United States

REVIEWED BY
Mihaela Stefan Lifshitz,
Albert Einstein College of Medicine, United States
Mohsen Ghanbari,
Erasmus Medical Center, Netherlands

*CORRESPONDENCE
Izortze Santin
izortze.santin@ehu.eus
Ainara Castellanos-Rubio
ainara.castellanos@ehu.eus

†These authors have equally contributed to this work

SPECIALTY SECTION
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 18 November 2022
ACCEPTED 24 January 2023
PUBLISHED 07 February 2023

CITATION
González-Moro I, Rojas-Márquez H, Sebastian-delaCruz M, Mentxaka-Salgado J, Olazagoitia-Garmendia A, Mendoza LM, Lluch A, Fantuzzi F, Lambert C, Ares Blanco J, Marselli L, Marchetti P, Cnop M, Delgado E, Fernández-Real JM, Ortega FJ, Castellanos-Rubio A and Santin I (2023) A long non-coding RNA that harbors a SNP associated with type 2 diabetes regulates the expression of *TGM2* gene in pancreatic beta cells.
Front. Endocrinol. 14:1101934.
doi: 10.3389/fendo.2023.1101934

A long non-coding RNA that harbors a SNP associated with type 2 diabetes regulates the expression of *TGM2* gene in pancreatic beta cells

Itziar González-Moro[1,2†], Henar Rojas-Márquez[2,3†], Maialen Sebastian-delaCruz[2,3], Jon Mentxaka-Salgado[1,2], Ane Olazagoitia-Garmendia[1,2,3], Luis Manuel Mendoza[1,2], Aina Lluch[4,5], Federica Fantuzzi[6], Carmen Lambert[7,8], Jessica Ares Blanco[7,9,10], Lorella Marselli[11], Piero Marchetti[11], Miriam Cnop[6,12], Elías Delgado[7,9,10,13], José Manuel Fernández-Real[4,5,14], Francisco José Ortega[4,5], Ainara Castellanos-Rubio[2,3,15,16*] and Izortze Santin[1,2,15*]

[1]Department of Biochemistry and Molecular Biology, University of the Basque Country UPV/EHU, Leioa, Spain, [2]Biocruces Bizkaia Health Research Institute, Barakaldo, Spain, [3]Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country, Leioa, Spain, [4]Institut d'Investigació Biomèdica de Girona, Girona, Spain, [5]CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn), Instituto de Salud Carlos III, Madrid, Spain, [6]ULB Center for Diabetes Research, Université Libre de Bruxelles, Brussels, Belgium, [7]Health Research Institute of the Principality of Asturias (ISPA), Oviedo, Spain, [8]University of Barcelona, Barcelona, Spain, [9]Endocrinology and Nutrition Department, Central University Hospital of Asturias (HUCA), Oviedo, Spain, [10]Department of Medicine, University of Oviedo, Oviedo, Spain, [11]Department of Clinical and Experimental Medicine, Cisanello University Hospital, Pisa, Italy, [12]Division of Endocrinology, Erasmus Hospital, Université Libre de Bruxelles, Brussels, Belgium, [13]Spanish Biomedical Research Network in Rare Diseases (CIBERER), Madrid, Spain, [14]Department of Medical Sciences, School of Medicine, University of Girona, Oviedo, Spain, [15]Diabetes and Associated Metabolic Diseases Networking Biomedical Research Centre, Madrid, Spain, [16]Ikerbasque - Basque Foundation for Science, Bilbao, Spain

Introduction: Most of the disease-associated single nucleotide polymorphisms (SNPs) lie in non- coding regions of the human genome. Many of these variants have been predicted to impact the expression and function of long non-coding RNAs (lncRNA), but the contribution of these molecules to the development of complex diseases remains to be clarified.

Methods: Here, we performed a genetic association study between a SNP located in a lncRNA known as LncTGM2 and the risk of developing type 2 diabetes (T2D), and analyzed its implication in disease pathogenesis at pancreatic beta cell level. Genetic association study was performed on human samples linking the rs2076380 polymorphism with T2D and glycemic traits. The pancreatic beta cell line EndoC-bH1 was employed for functional studies based on LncTGM2 silencing and overexpression experiments. Human pancreatic islets were used for eQTL analysis.

Results: We have identified a genetic association between LncTGM2 and T2D risk. Functional characterization of the LncTGM2 revealed its implication in the transcriptional regulation of TGM2, coding for a transglutaminase. The T2Dassociated risk allele in LncTGM2 disrupts the secondary structure of this lncRNA, affecting its stability and the expression of TGM2 in pancreatic beta cells. Diminished LncTGM2 in human beta cells impairs glucose-stimulated insulin release.

Conclusions: These findings provide novel information on the molecular mechanisms by which T2D-associated SNPs in lncRNAs may contribute to disease, paving the way for the development of new therapies based on the modulation of lncRNAs.

KEYWORDS

long non-coding RNA, type 2 diabetes, single nucleotide pholymorphism (SNP), pancreatic beta cell, transglutaminase 2

1 Introduction

Type 2 diabetes (T2D) is a complex metabolic disease that develops in genetically susceptible individuals (1). Indeed, the trigger of T2D development is presumed to be a combination of lifestyle and environmental factors working together with the genetic background (2). Genome-wide association studies (GWAS) have identified several genomic regions associated with the risk of T2D (3). Although these studies have provided a better understanding of T2D genetics, most of the genetic variants identified so far fall into non-coding regions of the genome. The molecular mechanism by which these variants increase risk of T2D remains to be clarified.

Transglutaminase 2 (TGM2) is a calcium-dependent multifunctional enzyme that can act as GTPase or transamidase, and that participates in several cellular processes, including apoptosis, cell adhesion or insulin release, among others (4). Disruption of *TGM2* in mice has been associated with increased glucose levels, and reduced insulin release in response to glucose (5). In addition, missense mutations in *TGM2* have been associated with early onset T2D and maturity onset diabetes of the young (MODY) (6).

A recent study identified a lncRNA (*LOC107987281* or *LncTMG2*) located within the first intron of the *TGM2* gene. The same study revealed that the expression of the lncRNA was tightly correlated with the expression of the *TGM2* coding gene in several cell lines and tumor tissues, suggesting its role as a cis acting transcriptional regulatory lncRNA (7).

LncRNAs are non-coding RNA molecules of more than 200 nucleotides in lenght that participate in several cellular and biological processes, including transcriptional regulation (8). Most of the complex disease-associated variants are located in non-coding regions of the human genome, and more specifically, in lncRNAs. The presence of disease-associated single nucleotide polymorphism (SNPs) in exonic regions of lncRNAs usually disrupt their secondary structure, affecting their capacity to interact with other macromolecules, and eventually altering their function (9). Although the function of most lncRNAs has not been annotated yet, there is already accumulating evidence of their implication in the development of several diseases, including metabolic disorders (10–12).

In the present work, we have described a genetic association between a SNP located in the coding sequence of *LncTGM2* and T2D and related traits. In addition, we have characterized the relation between *LncTGM2* and *TGM2* in pancreatic beta cells and unveiled the mechanisms by which *LncTGM2* might induce beta cell dysfunction in T2D.

2 Materials and methods

2.1 Association study

Cohort 1 consisted of 725 individuals (47 ± 11 years, 54% men) recruited in the northwest of Spain, including general population, and obesity and diabetes outpatient clinics in which the percentage of obese individuals was 72% and the percentage of type 2 diabetic individuals was 11% (13). Cohort 2 included 616 Caucasian subjects selected for a study of non-classic cardiovascular risk factors performed in the northwest of Spain (Asturias) (14). Participants (52 ± 12 years, 45% men, 26% obesity, 11% T2D) were randomly identified from a census and invited to participate.

Clinical characterization of human cohorts included a standardized questionnaire, physical examination and the performance of routine laboratory tests. Height and weight were measured by trained personnel using calibrated scales and a wall-mounted stadiometer, respectively, and with the participant in light clothing and without shoes. Body mass index (BMI) was calculated by dividing weight in kilograms by the square of the height in meters (kg/m^2). Obesity was set at BMI≥30 kg/m^2. The waist of the subjects was measured with a soft tape midway between the lowest rib and the iliac crest, hip circumference was measured at the widest part of the gluteal region, and waist-to-hip ratio was then calculated. Together with clinically relevant information and subsidiary data, the number of cigarettes/day (if any) and the use of

hormonal contraceptives were recorded. In those participants that agree (>75%), oral glucose tolerance test (OGTT) was performed to measure glucose tolerance. Blood samples from all the participants were collected, and after 15 minutes, tubes were centrifuged at 4,000 r.p.m. at room temperature. The serum and peripheral blood leukocytes were separated and immediately frozen at –80°C. Genomic DNA was extracted from blood samples following standard purification methods (QIAamp DNA Blood Mini Kit, Qiagen, Hilden, Germany) and DNA quantity and purity was determined using a spectrophotometer (GeneQuant, GE Health Care, Piscataway, USA). The targeted single nucleotide polymorphism (SNP) rs2076380 was genotyped by means of a predesigned rhAmp™ allelic discrimination assay (Hs.GT.rs2076380.A.1; Thermo Fisher Scientific, Massachusetts, USA) and the rhAmp Genotyping Master Mix (IDT, Coralville, USA), using a LightCycler 480 RT-qPCR System sequence detector (Roche Diagnostics, Barcelona, Spain). Replicates and positive and negative controls were included in all reactions.

2.2 Cell cultures and human cDNA samples

The EndoC-βH1 human pancreatic cell line (Univercell Biosolutions, Paris, France) was cultured in plates coated with Matrigel-fibronectin (100 mg/ml and 2 mg/ml, respectively; Sigma-Aldrich, Burlington, USA) in Opti-β1 medium (Univercell Biosolutions). DMEM containing 5.6 mmol/l glucose, 2% vol/vol Fetal Bovine Serum, 50 µmol/l 2-mercaptoethanol (Bio-Rad, Hercules, USA), 10 mmol/l nicotinamide (Calbiochem, Darmstadt, Germany), 5.5 µg/ml transferrin and 6.7 ng/ml selenite (Sigma-Aldrich) was used for transfection.

EndoC-βH1 cell line was Mycoplasma free as determined by the MycoAlert Mycoplasma Detection kit (Lonza). For the prevention of Mycoplasma contamination, Plasmocin Prophylactic (Invivogen, Toulouse, France) was added to the culture medium on a regular basis.

cDNA samples from human pancreatic islets were obtained from Cisanello University Hospital, Pisa, Italy. All the islets were isolated and cultured using the same experimental conditions and following established isolation procedures (15). Characteristics of islet preparations are described in Table S1. The Ethical Committee of Cisanello University Hospital approved experiments using human islets.

2.3 Silencing experiments

LncTGM2 silencing in the EndoC-βH1 cell line was performed by transfecting 30 nmol/l of a siRNA targeting *LncTGM2* (CD.Ri.214258.13.13, IDT) using Lipofectamine RNAimax reagent (Thermo Fisher Scientific) following the manufacturer's instructions.

2.4 Plasmid construction and transfection

For overexpressing plasmids, *LncTGM2* was purchased as a gBlock (IDT) and cloned into a modified pCMV6 vector using KpnI and FseI restriction enzymes (New England Biolabs, Ipswich, USA). Plasmids were transfected using Lipofectamine 2000 Transfection Reagent (Invitrogen, Carlsbad, USA) following the manufacturer's instructions.

2.5 Cell treatments

EndoC-βH1 cells were exposed to Actinomycin D (Sigma-Aldrich) at a final concentration of 5 µg/ml for 2, 4 or 6h. Palmitate treatment was performed by adding BSA-palmitic acid (0.5 mmol/l; 1:1) to DMEM/F-12, complemented with 0.25% vol/vol FBS, 50 µmol/l 2-mercaptoethanol (Bio-Rad), 10 mmol/l nicotinamide (Calbiochem), 5.5 µg/ml transferrin, 6.7 ng/ml selenite (Sigma-Aldrich), 100 units/ml penicillin and 100 µg/ml streptomycin (Lonza) for 4 or 8h.

2.6 Cellular fractionation

For *LncTGM2* RNA quantification in subcellular fractions of EndoC-βH1 cells, nuclei were isolated using C1 lysis buffer (1.28 mol/l sucrose, 40 mmol/l Tris -HCl pH 7.5, 20 mmol/l $MgCl_2$, 4% vol/vol Triton X-100). *LncTGM2*, *MEG3* (nuclear control) and *RPLP0* (cytoplasmic control) expression levels were measured by RT-qPCR and compared to the total amount of those RNAs in the whole cell lysate.

2.7 RNA isolation and RT-qPCR

RNA extraction was performed using the NucleoSpin RNA Kit (Macherey Nagel, Düren Germany) and expression values were determined by RT-qPCR using iTaq Universal SYBR Green Supermix (Bio-Rad) using specific primers for each target RNA (Table S2). All RT-qPCR measurements were performed in duplicates and expression levels were analyzed using the $2^{-\Delta\Delta Ct}$ method. A commercially available RNA panel set (Human total RNA master panel II, Clontech, Saint-Germain-en-Laye, France) was used to assess *LncTGM2* and *TGM2* expression levels in different human tissues.

2.8 Western blot analysis

EndoC-βH1 cells were washed with cold PBS and lysed in Laemmli buffer (62 mmol/l Tris-HCl, 100 mmol/l dithiothreitol (DTT), 10% vol/vol glycerol, 2% wt/vol SDS, 0.2 mg/ml bromophenol blue, 5% vol/vol 2-mercaptoethanol). Proteins in the lysate were separated by SDS-PAGE. After electrophoresis, proteins were transferred to nitrocellulose membranes using a Transblot-Turbo Transfer System (Bio-Rad) and blocked in 5% wt/vol non-fatty milk diluted in TBST (20 mmol/l Tris, 150 mmol/l NaCl and 0.1% vol/vol Tween 20) at room temperature for 1h. The membranes were incubated overnight at 4°C with a primary antibody specific for TGM2 (15100-1-AP, Proteintech Group, Rosemont, USA) diluted 1:1000 in 5% wt/vol BSA or anti-α-tubulin (Cat #T9026, Sigma-Aldrich) diluted 1:5000 in 5% wt/vol BSA. Immunoreactive bands

were revealed using the Clarity Max Western ECL Substrate (Bio-Rad) after incubation with a horseradish peroxidase-conjugated anti-rabbit (1:1000 dilution in 5% wt/vol non-fatty milk) or anti-mouse (1:5000 dilution in 5% wt/vol non-fatty milk) secondary antibody for 1h at room temperature. The immunoreactive bands were detected using a Bio-Rad Molecular Imager ChemiDoc XRS and quantified using ImageLab software (Bio-Rad).

2.9 *TGM2* promoter reporter assay

TGM2 promoter sequence was cloned into an empty pBV-Luc plasmid (Addgene, Watertown, USA) using KpnI and EcoRI restriction enzymes. EndoC-βH1 cells were transfected with a control vector (ovCTRL) or a vector overexpressing *LncTGM2* (ov*LncTGM2*), and co-transfected with the *TGM2* promoter reporter vector plus a pRL-CMV plasmid (used as an internal control) using Lipofectamine 2000 Transfection Reagent (Invitrogen). Dual-Luciferase Reporter Assay System (Promega, Madison, USA), was used to measure bioluminescence following the manufacturer's protocol.

2.10 *In silico* secondary structure prediction

Secondary structure of *LncTGM2* harboring the different alleles of rs2076380, rs7275079 and rs2067027 SNPs was predicted using the RNAsnp Web Server tool (16).

2.11 RNA mobility shift assay

LncTGM2 harboring rs2076380-A or rs2076380-G alleles were *in vitro* transcribed using T7 RNA Polymerase kit (TaKaRa, Kusatsu, Japan). RNAs were run in a native TBE 2% wt/vol agarose gel and migration profile was analyzed in a ChemiDoc XRS apparatus (Bio-Rad).

2.12 Insulin release

For insulin release experiments, *LncTGM2*-silenced EndoC-βH1 were left in Opti-β2 (Univercell Biosolutions) starving medium for 24h. After glucose starvation, cells were incubated in KREBS medium (Univercell Biosolutions) for 1h, and consecutively exposed to 0 or 20 mmol/l glucose for 40 minutes. Supernatant and lysate were harvested and insulin release and content measured by a commercial human insulin ELISA kit (Mercodia, Uppsala, USA) according to the manufacturer's instructions.

2.13 Statistics

The association between the rs2076380 single variation in the *TGM2* gene, clinical parameters and the risk of T2D was assessed using SPSS Statistics (IBM). Departures from Hardy-Weinberg equilibrium were tested in all groups using a chi-square goodness of fit test with one degree of freedom. The risk of developing T2D under exposure to rs2076380 *TGM2* genotypes was evaluated using logistic regression to estimate Odd Ratios (OR), considering a dominant model in which G-allele carriers (i.e., AG-heterozygotes plus GG-homozygotes) were the reference group. To compare groups with respect to continuous variables, one-way ANOVA for multiple comparisons was used. Other statistical tests and plots were performed using GraphPad Prism 8 software (Dotmatics). Significance-level was set at p-value <0.05. Results for *in vitro* functional studies are represented as means ± standard error of mean (S.E.M.).

3 Results

3.1 An exonic SNP in *LncTGM2* is associated with T2D risk

In order to determine the potential association of *LncTGM2* with T2D clinical parameters, we performed an association study by genotyping a SNP located in the exonic region of *LncTGM2* (rs2076380; chr20:38,165,027-38,165,227, hg38). This SNP can be considered as a tagSNP since it is in high linkage disequilibrium (LD>0.8) with other SNPs in the region (Figure S1). The *LncTGM2* SNP rs2076380 was tested in association with measures of T2D and other metabolic and clinical parameters in two independent cohorts (Table S3). In cohort 1, the frequency of AA-individuals for the *LncTGM2* SNP was 8.6%, similarly to the observed frequency in Cohort 2 (8.3%). These frequencies are in line with the observed frequency of the minor allele (A) in Caucasian populations (1000 Genomes Europe; A allele frequency = 0.32) (17) and Spanish control individuals (Medical Genome Project healthy controls from Spanish population; A allele frequency = 0.225) (18).

As observed in Figure 1, the percentage of known type 2 diabetic individuals was increased in individuals harboring the rs2076380-AA genotype in both cohorts (Cohort 1: OR=1.13 [0.999-1.27], Pearson's Chi-square p=0.006, two-sided Fisher's exact test p=0.013); and Cohort 2: OR=1.08 [0.996-1.18], Pearson's Chi-square p=0.018, two-sided Fisher's exact test p=0.026). For both cohorts, regression analyses depicted the impact of the polymorphism in *LncTGM2* on T2D incidence (ANOVA p-value of 0.026 in Cohort 1, and p=0.013 in Cohort 2) after correcting for sex and age. A codominant genetic model that included age, weight and sex effects was fitted to estimate the ORs between the exposure to the AA, AG and GG genotypes, the later as the reference group. The similar ORs for AG and GG genotypes obtained for the codominant model suggested the possibility of fitting a recessive model for AA-genotype carriers. This model allowed us to determine the OR between carriers of the AA genotype in relation to the G-allele porters. In this case, the residual deviance of the genotype, once age, weight and sex were added to the model, reached a p-value <0.05, indicating that the genotype effect was significant.

In addition, we observed that in Cohort 1, fasting glucose (p=0.005) and insulin levels (p=0.006) were increased compared to G-allele carriers (Table S3). However, in Cohort 2, association with fasting glucose only reached statistical significance in female participants (Table S3).

10.3389/fendo.2023.1101934

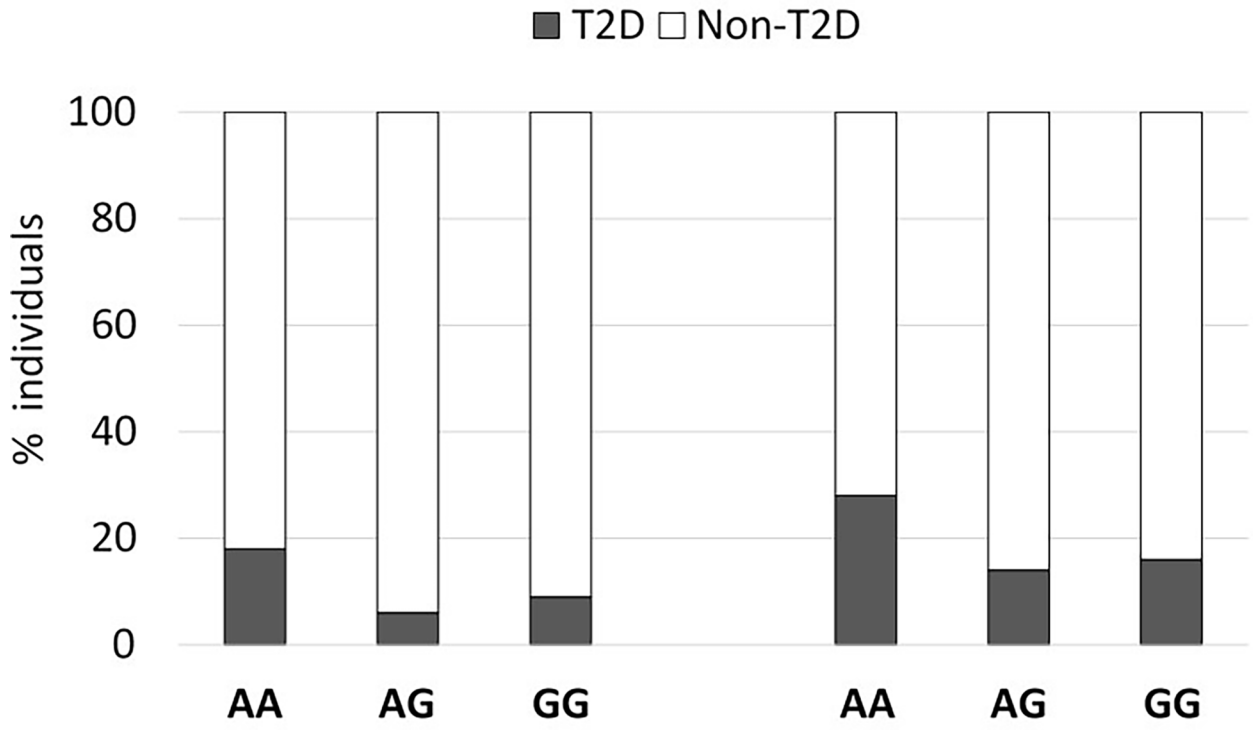


FIGURE 1
An exonic SNP in *LncTGM2* is associated with T2D risk. The graph shows the percent of T2D in two independent cohorts segregated according to their rs2076380 genotype.

3.2 *LncTGM2* expression is correlated with *TGM2* expression in several human tissues and regulated by lipotoxicity in pancreatic beta cells

Previous studies have correlated *LncTGM2* and *TGM2* expression in tumor tissues and some human cell lines, including lymphoblast (K562), promyeloblast (HL60) and monocyte (THP-1) cell lines (7). In order to clarify whether *LncTGM2* and *TGM2* expression was also correlated in healthy human tissues and in pancreatic beta cells, we first evaluated the expression of both genes in EndoC-βH1 cells and a set of human tissues. The highest expression of both, *LncTGM2* and *TGM2*, was found in lung, placenta and heart, and the expression in the EndoC-βH1 cell line was similar to that of intestine and liver (Figure S2). Spearman's correlation analysis showed a significant correlation between *LncTGM2* and *TGM2* expression across the tissues analyzed (R=0.87 (0.59-0.9); p<0.0001). Interestingly, a correlation was also seen in EndoC-βH1 cells using siRNA-driven inhibition of *LncTGM2*. As shown in Figure 2A, a 70% decrease of *LncTGM2* expression reduced *TGM2* mRNA expression by 20%, suggesting a potential implication of *LncTGM2* in the transcriptional regulation of *TGM2*.

In order to simulate the pathophysiological conditions of T2D in pancreatic beta cells, we next exposed EndoC-βH1 cells to palmitate (PA) as an *in vitro* model of lipotoxicity (19). As shown in Figure 2B, 4 and 8h PA exposure decreased both *LncTGM2* and *TGM2* expression in EndoC-βH1 cells, suggesting that in the presence of a lipotoxic insult the expression of both genes is reduced.

3.3 *LncTMG2* regulates the transcriptional activity of *TGM2*

Knowledge of the subcellular localization of lncRNAs is crucial to understand and characterize their function. In contrast to protein-coding mRNAs, lncRNA themselves should be located in their site of action, and thus, their location within the cell is crucial for their function. While nuclear lncRNAs are usually implicated in the regulation of transcriptional activity, cytoplasmic lncRNAs can participate for example, in the regulation of mRNA stability or in protein translation (20). Having this in mind, we next decided to analyze the subcellular localization of *LncTGM2* in EndoC-βH1 cells. As shown in Figure 2C, *LncTGM2* was detected in both nuclear and cytoplasmic fractions, but its expression level was significantly higher in the nuclear compartment, suggesting its potential implication in transcriptional regulation. Since expression of *LncTGM2* and *TGM2* was significantly correlated in pancreatic beta cells, we performed a promoter reporter assay to clarify whether *LncTGM2* was directly regulating the promoter activation of *TGM2* gene. To this aim, we constructed an expression vector coding for a luciferase under the control of the promoter of *TGM2*. The luciferase vector was then co-transfected in EndoC-βH1 cells with an empty overexpression plasmid (ovCTRL) or with the overexpression plasmid of *LncTGM2* (ov*LncTGM2*) and the activation of *TGM2* promoter was determined by measuring bioluminiscence. As shown in Figure 2D, the activation of the *TGM2* promoter was 1.5-fold higher in *LncTGM2*-overexpressing cells than in control cells, pointing out a role of *LncTGM2* in the activation of *TGM2* promoter, and consequently in the transcriptional activation of *TGM2*.

10.3389/fendo.2023.1101934

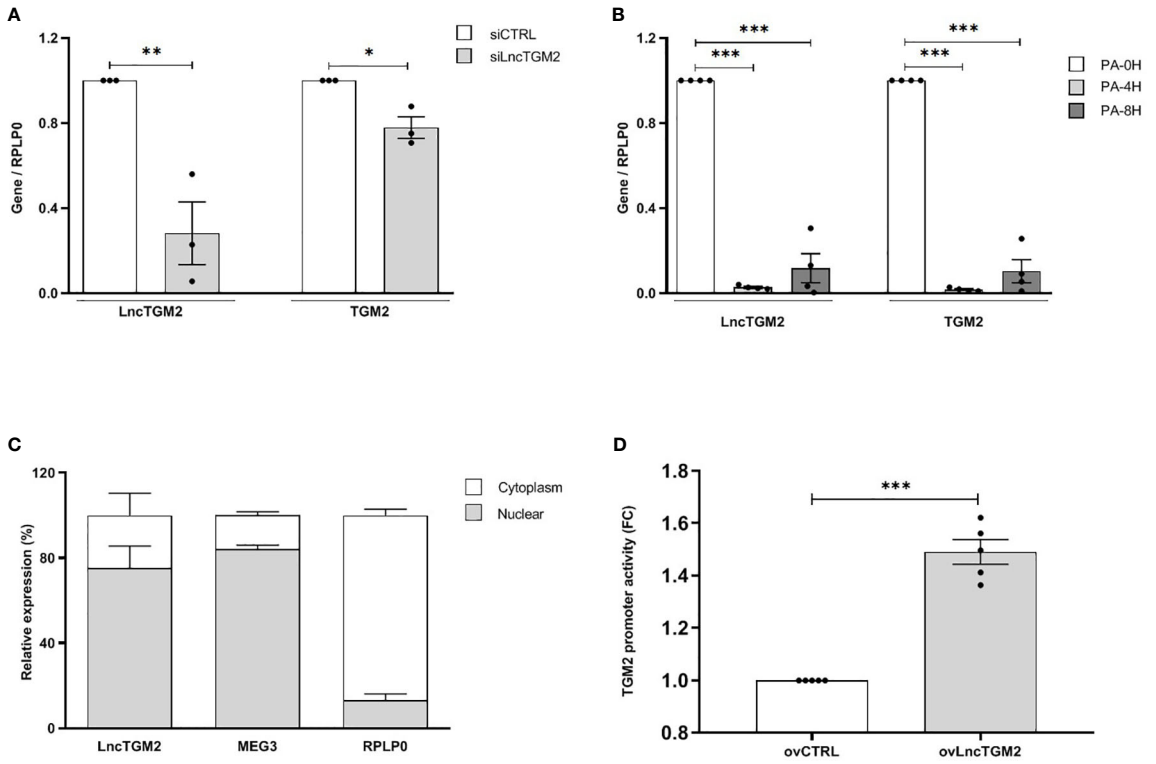


FIGURE 2
LncTGM2 co-expresses with *TGM2* and regulates its transcriptional activity in pancreatic beta cells. **(A)** *LncTGM2* was silenced in the EndoC-βH1 cell line using a siRNA, and **(B)** EndoC-βH1 cells were exposed to palmitate (0.5 mM) for 4 or 8h. *LncTGM2* and *TGM2* expression was assessed by RT-qPCR and normalized by the reference gene *RPLP0*. The results are means ± S.E.M. of 3-4 independent experiments; *p < 0.05, **p < 0.01, and ***p < 0.001 by Student's t-test. **(C)** RT-qPCR analysis of *LncTGM2*, *MEG3* (as nuclear marker) and *RPLP0* (as cytoplasmic marker) in nuclear and cytoplasmic fractions in EndoC-βH1 cells. **(D)** HEK-293 cells were transfected with a control vector (ovCTRL) or a vector overexpressing *LncTGM2* (ov*LncTGM2*), and co-transfected with a *TGM2* promoter luciferase reporter construct plus a pRL-CMV plasmid (used as internal control). After 48h of recovery, bioluminescence was measured.

3.4 The T2D-associated risk allele in *LncTGM2* disrupts its secondary structure impacting on its stability, and correlates with decreased expression of *TGM2* in beta cells

Disease-associated SNPs located within lncRNAs can affect their function through the disruption of their secondary structure (21, 22). As previously shown (Figure S1), the T2D-associated rs2076380 SNP is in high LD with other two SNPs located in the exonic region of *LncTGM2* (rs7275079 and rs2067027). To assess whether these SNPs alter the secondary structure of *LncTGM2*, we performed an *in silico* prediction analysis using the RNAsnp webserver from the Center for non-coding RNA in Technology and Health (23). Interestingly, rs2076380 was predicted to significantly alter the secondary structure of *LncTGM2* (p=0.0803), while the software did not predict any significant change in the structure of the lncRNA when the different alleles of rs7275079 or rs2067027 SNPs were present (p>0.2) (data not shown). As shown in Figure 3A, the predicted secondary structures of *LncTGM2* carrying the T2D protective (rs2076380-G) or risk allele (rs2076380-A) were significantly different. Consistent with the prediction, *in vitro*–transcribed forms of T2D protective and risk allele–harboring *LncTGM2* revealed different motilities on a native agarose gel (Figure 3B), suggesting a different conformation of the lncRNA in the presence of one or other allele in rs2076380.

Taking into account that the secondary structure of a lncRNA is crucial for its interaction with other macromolecules, and thus, for its function (9), we next decided to determine whether the genotype of the T2D-associated SNP in *LncTGM2* affected *TGM2* expression in human pancreatic islets. To this aim we genotyped rs2076380 SNP and measured *TGM2* expression in 16 cDNA samples from human islets, and performed an eQTL analysis. As shown in Figure S3, there was a trend for higher expression of *TGM2* in islets harboring the protective rs2076380-GG genotype compared to islets harboring the risk allele in heterozygosis (rs2076380-AG) or homozygosis (rs2076380-AA), although the differences did not reach statistical significance, probably due to the limited number of islets.

Next, to characterize the potential effect of each allele in rs2076380 SNP on the expression of both, *LncTGM2* and *TGM2*, we constructed two *LncTGM2* overexpression plasmids, one harboring the T2D risk allele (ov*LncTGM2*-A), and the other harboring the T2D protective allele (ov*LncTGM2*-G). Interestingly, allele-specific upregulation of *LncTGM2* in beta cells revealed that the expression level reached by transfecting ov*LncTGM2*-G plasmid was higher than the expression level obtained with ov*LncTGM2*-A plasmid (Figure 3C), suggesting that the T2D risk allele might be affecting the stability of *LncTGM2* RNA molecule.

10.3389/fendo.2023.1101934

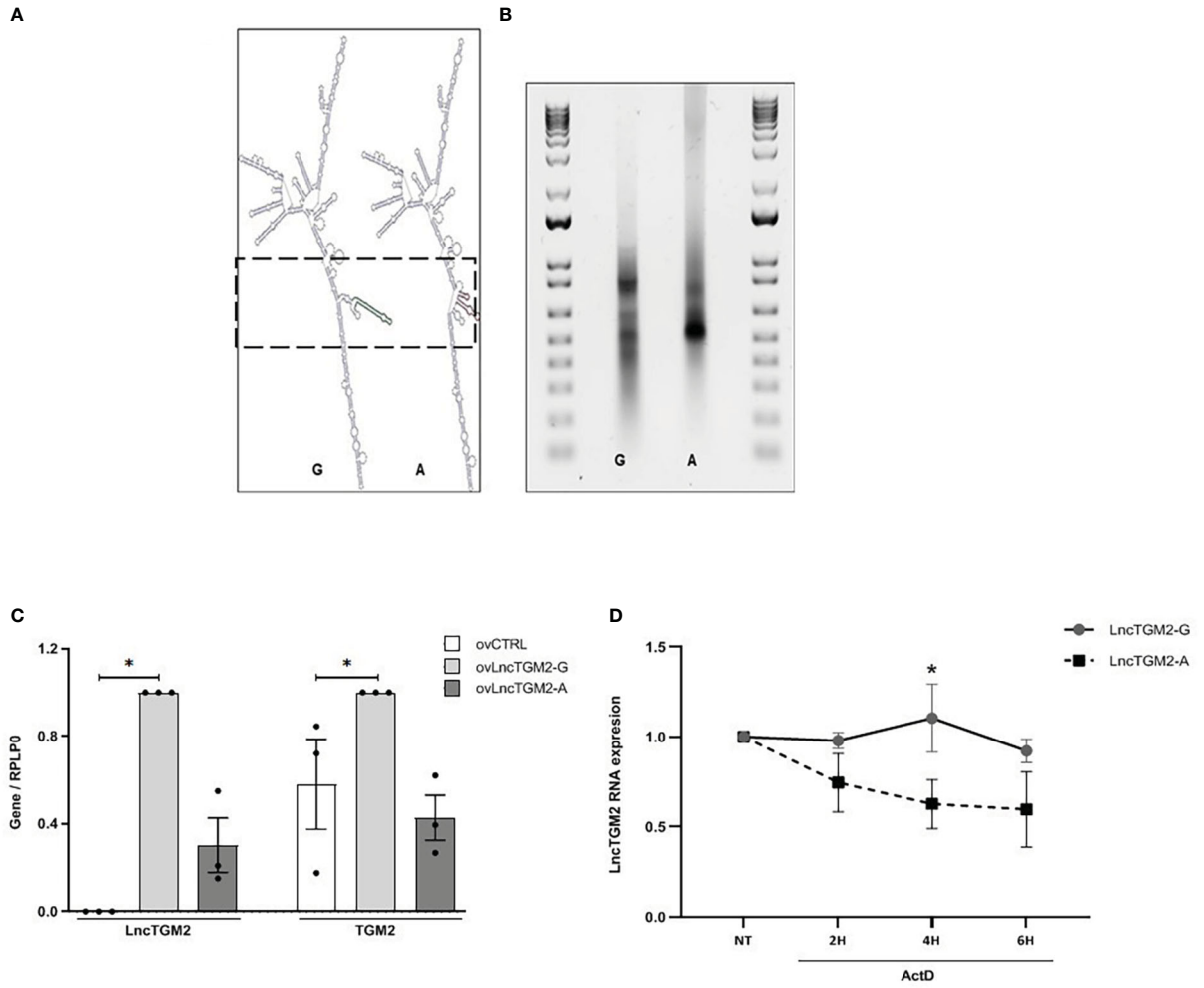


FIGURE 3
The T2D-associated risk allele in *LncTGM2* disrupts its secondary structure impacting on its stability, and correlates with decreased expression of *TGM2* in beta cells. (A) *In silico* prediction of the secondary structure of *LncTGM2* harboring each allele for rs2076380; T2D protective allele (G) or T2D risk allele **(A)**. **(B)** Electrophoretic mobility profiles of *in vitro*-transcribed *LncTGM2* molecule harboring the T2D protective allele (rs2076380-G) or the risk allele (rs2076380-A). **(C)** EndoC-βH1 cells were transfected with overexpression plasmids of *LncTGM2* harboring the protective (ov*LncTGM2*-G) or risk allele (ov*LncTGM2*-A) for T2D, and mRNA levels of *LncTGM2* and *TGM2* were determined by RT-qPCR and normalized to *RPLP0*. The results are means ± S.E.M. of 3 independent experiments; *p < 0.05 by Student's t-test. **(D)** EndoC-βH1 cells were transfected with *LncTGM2* overexpression plasmids harboring the protective (ov*LncTGM2*-G) or risk allele (ov*LncTGM2*-A) for T2D. EndoC-βH1 cells were exposed to Actinomycin D (ActD) (5 µg/ml) for 2, 4 or 6h and *LncTGM2* mRNA level was determined by RT-qPCR. The results are means ± S.E.M. of 3 independent experiments. *p < 0.05 ov*LncTGM2*-G vs. ov*LncTGM2*-A at the same time-point.

In order to directly test whether the T2D-associated polymorphism affected *LncTGM2* stability, we next performed an allele-specific overexpression of *LncTGM2* and exposed the EndoCβ-H1 cells to Actinomycin D, a drug that inhibits transcription. As shown in Figure 3D, *LncTGM2* harboring the protective allele (rs2076380-G) was more stable than the lncRNA harboring the risk allele (rs2076380-A) at all time-points, although the differences only reached statistical significance at 4h of Actinomycin D treatment ($p<0.05$). These results confirmed that the *LncTGM2* risk allele in the T2D-associated rs2076380 SNP reduced the stability of the lncRNA.

To clarify whether the decreased stability of *LncTGM2*-A affected its capacity to regulate *TGM2* expression, we next analyzed the expression of *TGM2* in EndoC-βH1 cells overexpressing *LncTGM2*-A or *LncTGM2*-G. As observed in Figure 3C, only the upregulation of the lncRNA harboring the protective allele (ov*LncTGM2*-G) increased the expression of *TGM2* mRNA. These results were also confirmed at the protein level (Figure S4).

In summary, these results suggested that the *LncTGM2* harboring the T2D risk allele induced less *TGM2* expression due to its reduced stability.

3.5 *LncTGM2* downregulation affects glucose-stimulated insulin secretion

Previous studies have shown that TGM2 might be implicated in insulin release through different mechanisms, including cytoplasmic actin remodeling and regulation of the action of other proteins during granule movement (24). Taking into account that our present results suggest that the T2D risk allele in *LncTGM2* might induce a decrease

10.3389/fendo.2023.1101934

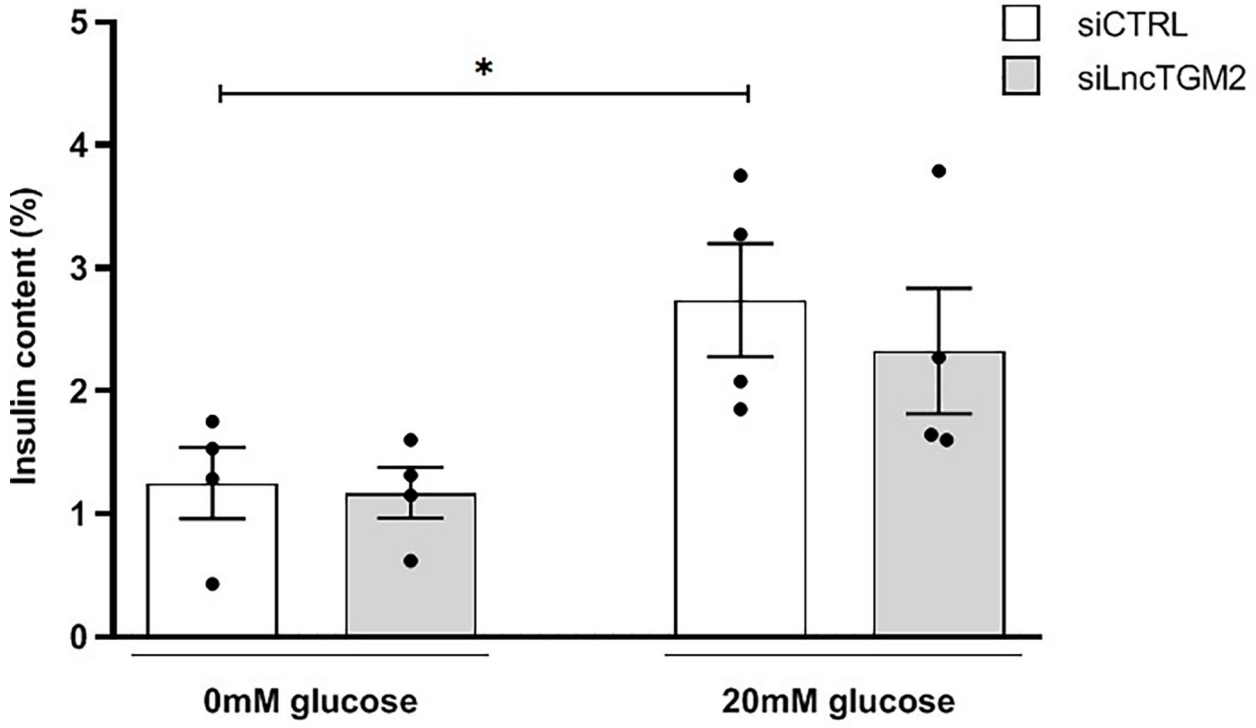


FIGURE 4
LncTGM2 downregulation affects glucose-stimulated insulin secretion. *LncTGM2* was silenced using a siRNA and EndoC-βH1 cells were exposed to 0 or 20 mmol/l glucose. Insulin release was determined by ELISA. The results are means ± S.E.M. of 4 independent experiments; *p<0.05 by Student's t-test.

in *TGM2* expression in pancreatic beta cells, we next decided to determine the potential contribution of *LncTGM2* in insulin release.

To this aim, we silenced *LncTGM2* with a specific siRNA in EndoC-βH1 cells and determined glucose-stimulated insulin release. As shown in Figure 4, high glucose stimulation in siCTRL-transfected EndoC-βH1 cells increased insulin secretion. In si*LncTGM2*-transfected beta cells, however, high glucose-induced insulin secretion (GSIS) was no longer statistically significant, suggesting that disruption of *LncTGM2* in pancreatic beta cell might affect GSIS through diminished expression of TGM2.

4 Discussion

In the current study, we identified a genetic association between *LncTGM2* and T2D and glycemic traits in two independent cohorts. Previous GWAS in larger Caucasian populations have not detected a genetic association between rs2076380 and T2D, however based on phenotype-wide association data (T2D knowledge portal), this polymorphism has been associated with T2D-related complications (e.g. microalbuminuria). Moreover, based on the T2D knowledge portal, the genomic region in which *LncTGM2* is located (also containing *TGM2*, *RPRD1B* and *KIAA1755* genes) has been associated with several metabolic and glycemic traits, including cardiovascular disease related parameters, cholesterol and type 2 diabetes. The main reason for the discordance between our findings and GWAS data may lie on the fact that our two cohorts are enriched for obese individuals (especially Cohort 1), and in our both cohorts, T2D incidence seem to be associated with obesity (data not shown). In this sense, several studies have described a link between TGM2 and obesity and associated glycemic traits. For example, a study found that loss of TGM2 sensitizes for diet-induced obesity-related inflammation and insulin resistance (25). Moreover, a network-based approach to assess the cellular processes associated with protein–protein interaction subnetworks of glycemic traits showed that TGM2 was associated with both, HOMA-β and HOMA-IR, suggesting a potential role of this protein in pancreatic beta cell function and insulin resistance (26). The same study concluded that HOMA-β-associated GWAS genes (which include *TGM2*) enriched pathways of fat metabolism, especially in adipose tissues, supporting the "lipotoxicity theory" of beta cell failure in T2D.

In line with this hypothesis, in the present study, we have observed a co-expression between *LncTGM2* and the coding gene *TGM2* in pancreatic beta cells under basal and lipotoxic conditions. Our data suggest that lipotoxicity, a typical feature of obesity-associated T2D, reduces *LncTGM2*, which in turn provokes a reduction of *TGM2* in pancreatic beta cells. Indeed, lipotoxicity (e.g. high fat diet) has been previously associated with *TGM2* expression reduction in other tissues, including liver (27).

Moreover, we propose a mechanism by which *LncTGM2* may affect glucose-stimulated insulin release through *TGM2* expression reduction in an allele-specific manner. The lncRNA *LncTGM2* lies within the first intron of the *TGM2* gene (9), which encodes a multifunctional enzyme that has been implicated in the pathogenesis of early onset T2D and MODY (6). Interestingly, early onset T2D and MODY-associated *TGM2* mutants have altered enzymatic activities, such as reduced transamidation and kinase activity that impact in glucose-stimulated insulin release (28).

Transcriptional regulation of *TGM2* is controlled by several transcription factors, including nuclear factor-kappa B, RA receptor/retinoid X receptor, liver X receptor and Sp1 (4). Here, we show for the first time that *LncTGM2* participates in the transcriptional regulation of *TGM2* in pancreatic beta cells. We observed that the T2D-associated risk allele in *LncTGM2* correlates with a reduction of *TGM2* expression in pancreatic beta cells. Moreover, our results suggest that a reduction in *TGM2* expression in human beta cells impair glucose-stimulated insulin release. These observations are in line with studies in rodents, in which reduced TGM2 activity has been linked to impaired glucose-stimulated insulin secretion (GSIS) (28), and also with data showing that naturally occurring mutations altering TGM2 enzymatic activities correlate with reduced insulin secretion (29). Interestingly, TGM2 has also been shown to interact with nuclear proteins (e.g. BAF and H3) immediately upon a glucose stimulus, suggesting that it may be involved not only in insulin secretion, but also in the regulation of glucose-induced gene transcription (30).

Although the molecular mechanisms by which *LncTGM2* participates in the regulation of *TGM2* transcription remain to be fully clarified, our results demonstrate that a T2D-associated polymorphism affects the secondary structure of the lncRNA, and, eventually, disrupts its function. Several other disease-associated SNPs that alter the secondary structure of lncRNAs affect the regulation of genes that participate in important pathways for disease pathogenesis, including type 1 diabetes and cardiovascular disease (31, 32). Here we demonstrate that the T2D risk allele in *LncTGM2* reduces its stability, affecting *TGM2* expression in pancreatic beta cells. Some studies have suggested that disease-associated SNPs in lncRNAs may affect RNA-turnover through disruption of the binding of proteins that regulate stability, and thus, affecting their biological function (33–35).

In conclusion, our results show that *LncTGM2* is associated with T2D and suggest that it might be implicated in disease pathogenesis through an allele-specific downregulation of *TGM2* in pancreatic beta cells. Our findings provide new information on the molecular mechanisms by which T2D-associated SNPs in lncRNAs cause disease and open the door to the development of novel diagnostic tools and therapeutic approaches based on lncRNA modulation.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding authors.

Author contributions

IS and AC-R conceived and designed the study. FO and AL performed the genetic association studies. ED, JF-R, CL and JA-B coordinated human samples, clinical information, written consents and intellectual content collection. LM and PM provided the human pancreatic islet material. HR-M, IG-M, JM-S, MS-C, AO-G, FF, LM-M and MC designed and performed the experimental procedures. HR-M and IG-M wrote the paper. IS and AC-R reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Ministerio de Ciencia, Innovación y Universidades (PID2019-104475GA-I00 to I.S, and PGC2018-097573-A-I00 to AC-R) and the European Foundation for the Study of Diabetes (EFSD) - EFSD/JDRF/Lilly Programme on Type 1 Diabetes Research to IS. FO (MS19/00109) is recipient of the Miguel Servet scheme, and AL (FI19/00045) was supported by the Instituto de Salud Carlos III (ISCIII); Ministerio de Ciencia, Innovación y Universidades, Gobierno de España (ES). HR-M (PRE2019-089350) is supported by predoctoral grant from the Ministerio de Ciencia, Innovacion y Universidades, Gobierno de España (ES) IG-M, MS-C, JM-S and AO-G were supported by Predoctoral Fellowship Grants from the UPV/EHU (Universidad del Pais Vasco/EuskalHerrikoUnibertsitatea) and the Basque Department of Education. MC is supported by the Fonds National de la RechercheScientifique (FNRS), the Francophone Foundation for Diabetes Research (sponsored by the French Diabetes Federation, Abbott, Eli Lilly, Merck Sharp & Dohme, and Novo Nordisk) and FF and MC by the EFSD/BoehringerIngelheim European Research Programme on Multi-System Challenges in Diabetes. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

The authors declare that this study received funding from the Francophone Foundation for Diabetes Research which is sponsored by Abbott, Eli Lilly, Merck Sharp & Dohme, and Novo Nordisk. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1101934/full#supplementary-material

References

1. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* (2014) 63:2158–71. doi: 10.2337/db13-0949

2. Laakso M. Biomarkers for type 2 diabetes. *Mol Metab* (2019) 27:S139–46. doi: 10.1016/j.molmet.2019.06.016

3. Aguiari G, Crudele F, Taccioli C, Minotti L, Corrà F, Keillor JW, et al. Dysregulation of transglutaminase type 2 through GATA3 defines aggressiveness and doxorubicin sensitivity in breast cancer. *Int J Biol Sci* (2022) 18:1–14. doi: 10.7150/ijbs.64167

4. Tatsukawa H, Furutani Y, Hitomi K, Kojima S. Transglutaminase 2 has opposing roles in the regulation of cellular functions as well as cell growth and death. *Cell Death Dis* (2016) 7:e2244–4. doi: 10.1038/cddis.2016.150

5. Bernassola F, Federici M, Corazzari M, Terrinoni A, Hribal ML, De Laurenzi V, et al. Role of transglutaminase 2 in glucose tolerance: knockout mice studies and a putative mutation in a MODY patient. *FASEB J* (2002) 16:1371–8. doi: 10.1096/fj.01-0689com

6. Porzio O, Massa O, Cunsolo V, Colombo C, Malaponti M, Bertuzzi F, et al. Missense mutations in theTGM2 gene encoding transglutaminase 2 are found in patients with early-onset type 2 diabetes. *Hum Mutat* (2007) 28:1150–0. doi: 10.1002/humu.9511

7. Minotti L, Baldassari F, Galasso M, Volinia S, Bergamini CM, Bianchi N. A long non-coding RNA inside the type 2 transglutaminase gene tightly correlates with the expression of its transcriptional variants. *Amino Acids* (2018) 50:421–38. doi: 10.1007/s00726-017-2528-9

8. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* (2009) 136:629–41. doi: 10.1016/j.cell.2009.02.006

9. Graf J, Kretz M. From structure to function: Route to understanding lncRNA mechanism. *BioEssays* (2020) 42:12. doi: 10.1002/bies.202000027

10. Wang S-H, Zhu X-L, Wang F, Chen S-X, Chen Z-T, Qiu Q, et al. LncRNA H19 governs mitophagy and restores mitochondrial respiration in the heart through Pink1/Parkin signaling during obesity. *Cell Death Dis* (2021) 12:557. doi: 10.1038/s41419-021-03821-6

11. Chen Y-T, Yang Q-Y, Hu Y, Liu X-D, de Avila JM, Zhu M-J, et al. Imprinted lncRNA Dio3os preprograms intergenerational brown fat development and obesity resistance. *Nat Commun* (2021) 12:6845. doi: 10.1038/s41467-021-27171-1

12. Zhao X-Y, Xiong X, Liu T, Mi L, Peng X, Rui C, et al. Long noncoding RNA licensing of obesity-linked hepatic lipogenesis and NAFLD pathogenesis. *Nat Commun* (2018) 9:2986. doi: 10.1038/s41467-018-05383-2

13. Moreno-Navarrete JM, Ortega FJ, Bassols J, Castro A, Ricart W, Fernández-Real JM. Association of circulating lactoferrin concentration and 2 nonsynonymous LTF gene polymorphisms with dyslipidemia in men depends on glucose-tolerance status. *Clin Chem* (2008) 54:301–9. doi: 10.1373/clinchem.2007.095943

14. Valdés S, Botas P, Delgado E, Álvarez F, Cadórniga FD. Population-based incidence of type 2 diabetes in northern Spain. *Diabetes Care* (2007) 30:2258–63. doi: 10.2337/dc06-2461

15. Marchetti P, Bugliani M, Lupi R, Marselli L, Massini M, Boggi U, et al. The endoplasmic reticulum in pancreatic beta cells of type 2 diabetes patients. *Diabetologia* (2007) 50:2486–94. doi: 10.1007/s00125-007-0816-8

16. Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. The RNAsnp web server: predicting SNP effects on local RNA secondary structure. *Nucleic Acids Res* (2013) 41:W475–9. doi: 10.1093/nar/gkt291

17. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* (2015) 526:68–74. doi: 10.1038/nature15393

18. García-Alonso L, Jiménez-Almazán J, Carbonell-Caballero J, Vela-Boza A, Santoyo-López J, Antiñolo G, et al. The role of the interactorme in the maintenance of deleterious variability in human populations. *Mol Syst Biol* (2014) 10:752. doi: 10.15252/msb.20145222

19. Ciregia F, Bugliani M, Ronci M, Giusti L, Boldrini C, Mazzoni MR, et al. Palmitate-induced lipotoxicity alters acetylation of multiple proteins in clonal β cells and human pancreatic islets. *Sci Rep* (2017) 7:13445. doi: 10.1038/s41598-017-13908-w

20. Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* (2015) 16:20. doi: 10.1186/s13059-015-0586-4

21. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol* (2018) 19:143–57. doi: 10.1038/nrm.2017.104

22. Castellanos-Rubio A, Fernandez-Jimenez N, Kratchmarov R, Luo X, Bhagat G, Green PHR, et al. A long noncoding RNA associated with susceptibility to celiac disease. *Science* (1979) 2016) 352:91–5. doi: 10.1126/science.aad0467

23. Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. RNAsnp: Efficient detection of local RNA secondary structure changes induced by SNPs. *Hum Mutat* (2013) 34:546–56. doi: 10.1002/humu.22273

24. Russo L, Marsella C, Nardo G, Massignan T, Alessio M, Piermarini E, et al. Transglutaminase 2 transamidation activity during first-phase insulin secretion: natural substrates in INS-1E. *Acta Diabetol* (2013) 50(1):61–72. doi: 10.1007/s00592-012-0381-6

25. Sághy T, Köröskényi K, Hegedűs K, Antal M, Bankó C, Bacsó Z, et al. Loss of transglutaminase 2 sensitizes for diet-induced obesity-related inflammation and insulin resistance due to enhanced macrophage c-src signaling. *Cell Death Dis* (2019) 10(6):439. doi: 10.1038/s41419-019-1677-z

26. Saxena A, Wahi N, Kumar A, Mathur SK. Functional interactomes of genes showing association with type-2 diabetes and its intermediate phenotypic traits point towards adipo-centric mechanisms in its pathophysiology. *Biomolecules* (2020) 10(4):601. doi: 10.3390/biom10040601

27. Liu H, Niu Q, Wang T, Dong H, Bian C. Lipotoxic hepatocytes promote nonalcoholic fatty liver disease progression by delivering microRNA-9-5p and activating macrophages. *Int J Biol Sci* (2021) 17(14):3745–59. doi: 10.7150/ijbs.57610

28. Sileno S, D'Oria V, Stucchi R, Alessio M, Petrini S, Bonetto V, et al. A possible role of transglutaminase 2 in the nucleus of INS-1E and of cells of human pancreatic islets. *J Proteomics* (2014) 96:314–27. doi: 10.1016/j.jprot.2013.11.011

29. Iismaa SE, Aplin M, Holman S, Yiu TW, Jackson K, Burchfield JG, et al. Glucose homeostasis in mice is transglutaminase 2 independent. *PloS One* (2013) 8:e63346. doi: 10.1371/journal.pone.0063346

30. Salter NW, Ande SR, Nguyen HK, Nyomba BLG, Mishra S. Functional characterization of naturally occurring transglutaminase 2 mutants implicated in early-onset type 2 diabetes. *J Mol Endocrinol* (2012) 48:203–16. doi: 10.1530/JME-11-0064

31. Gonzalez-Moro I, Olazagoitia-Garmendia A, Colli ML, Cobo-Vuilleumier N, Postler TS, Marselli L, et al. The T1D-associated lncRNA Lnc13 modulates human pancreatic β cell inflammation by allele-specific stabilization of STAT1 mRNA. *Proc Natl Acad Sci* (2020) 117:9022–31. doi: 10.1073/pnas.1914353117

32. Aguilo F, di Cecilia S, Walsh MJ. Long non-coding RNA ANRIL and polycomb in human cancers and cardiovascular disease. *Current Topics in Microbiology and Immunology* (2015) 394:29–39. doi: 10.1007/82_2015_455

33. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* (2021) 22:96–118. doi: 10.1038/s41580-020-00315-9

34. Sebastian-delaCruz M, Gonzalez-Moro I, Olazagoitia-Garmendia A, Castellanos-Rubio A, Santin I. The role of lncRNAs in gene expression regulation through mRNA stabilization. *Noncoding RNA* (2021) 7:3. doi: 10.3390/ncrna7010003

35. Aznaourova M, Schmerer N, Schmeck B, Schulte LN. Disease-causing mutations and rearrangements in long non-coding RNA gene loci. *Front Genet* (2020) 11:527484. doi: 10.3389/fgene.2020.527484

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 28 February 2023
DOI 10.3389/fendo.2023.1081741

OPEN ACCESS

EDITED BY
Anandwardhan Hardikar,
Western Sydney University, Australia

REVIEWED BY
Sameet Mehta,
Yale University, United States
Marija Petkovic,
Roskilde University, Denmark
Jalal Taneera,
University of Sharjah, United Arab Emirates

*CORRESPONDENCE
Tarunveer S. Ahluwalia
tarun.veer.singh.ahluwalia@regionh.dk

SPECIALTY SECTION
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 27 October 2022
ACCEPTED 07 February 2023
PUBLISHED 28 February 2023

CITATION
Uglebjerg N, Ahmadizar F, Aly DM, Cañadas-Garre M, Hill C, Naber A, Oddsson A, Singh SS, Smyth L, Trégouët D-A, Chaker L, Ghanbari M, Steinthorsdottir V, Ahlqvist E, Hadjadj S, Van Hoek M, Kavousi M, McKnight AJ, Sijbrands EJ, Stefansson K, Simons M, Rossing P and Ahluwalia TS (2023) Four missense genetic variants in *CUBN* are associated with higher levels of eGFR in non-diabetes but not in diabetes mellitus or its subtypes: A genetic association study in Europeans. *Front. Endocrinol.* 14:1081741. doi: 10.3389/fendo.2023.1081741

Four missense genetic variants in *CUBN* are associated with higher levels of eGFR in non-diabetes but not in diabetes mellitus or its subtypes: A genetic association study in Europeans

Nicoline Uglebjerg [1], Fariba Ahmadizar [2,3], Dina M. Aly [4], Marisa Cañadas-Garre [5,6,7], Claire Hill [5], Annemieke Naber [8], Asmundur Oddsson [9], Sunny S. Singh [8], Laura Smyth [5], David-Alexandre Trégouët [10], Layal Chaker [2,8], Mohsen Ghanbari [2], Valgerdur Steinthorsdottir [9], Emma Ahlqvist [4], Samy Hadjadj [11], Mandy Van Hoek [8], Maryam Kavousi [2], Amy Jayne McKnight [5], Eric J. Sijbrands [8], Kari Stefansson [9,12], Matias Simons [13], Peter Rossing [1,14] and Tarunveer S. Ahluwalia [1,15]*

[1]Complications Research, Steno Diabetes Center Copenhagen, Herlev, Denmark, [2]Department of Epidemiology, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands, [3]Department of Data Science & Biostatistics, Julius Global Health, University Medical Center Utrecht, Utrecht, Netherlands, [4]Department of Clinical Sciences, Lund University, Malmö, Sweden, [5]Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom, [6]GENYO Centre for Genomics and Oncological Research, Pfizer-University of Granada-Andalusian Regional Government, Granada, Spain, [7]Instituto de Investigación Biosanitaria de Granada (ibs.GRANADA), Granada, Spain, [8]Department of Internal Medicine, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands, [9]deCODE Genetics, Amgen, Inc., Reykjavik, Iceland, [10]University of Bordeaux, Institut National de la Santé et de la Recherche Médicale (INSERM), Bordeaux Population Health Research Center, Bordeaux, France, [11]Nantes Université, Centre Hospitalier Universitaire Nantes, Centre National de la Recherche Scientifique, INSERM, l'institut du thorax, Nantes, France, [12]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland, [13]Institute of Human Genetics, University Hospital Heidelberg, Heidelberg, Germany, [14]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark, [15]The Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen, Denmark

Aim: Rare genetic variants in the *CUBN* gene encoding the main albumin-transporter in the proximal tubule of the kidneys have previously been associated with microalbuminuria and higher urine albumin levels, also in diabetes. Sequencing studies in isolated proteinuria suggest that these variants might not affect kidney function, despite proteinuria. However, the relation of these *CUBN* missense variants to the estimated glomerular filtration rate (eGFR) is largely unexplored. We hereby broadly examine the associations between four *CUBN* missense variants and $eGFR_{creatinine}$ in Europeans with Type 1 (T1D) and Type 2 Diabetes (T2D). Furthermore, we sought to deepen our understanding of these variants in a range of single- and aggregate- variant analyses of other kidney-related traits in individuals with and without diabetes mellitus.

Methods: We carried out a genetic association-based linear regression analysis between four *CUBN* missense variants (*rs141640975*, *rs144360241*, *rs45551835*, *rs1801239*) and $eGFR_{creatinine}$ (ml/min/1.73 m^2, $CKD\text{-}EPI_{creatinine(2012)}$, natural log-

transformed) in populations with T1D (n ~ 3,588) or T2D (n ~ 31,155) from multiple European studies and in individuals without diabetes from UK Biobank (UKBB, n ~ 370,061) with replication in deCODE (n = 127,090). Summary results of the diabetes-group were meta-analyzed using the fixed-effect inverse-variance method.

Results: Albeit we did not observe associations between $eGFR_{creatinine}$ and *CUBN* in the diabetes-group, we found significant positive associations between the minor alleles of all four variants and $eGFR_{creatinine}$ in the UKBB individuals without diabetes with *rs141640975* being the strongest (Effect=0.02, $P_{eGFR\_creatinine}=2.2 \times 10^{-9}$). We replicated the findings for *rs141640975* in the Icelandic non-diabetes population (Effect=0.026, $P_{eGFR\_creatinine}=7.7 \times 10^{-4}$). For *rs141640975*, the $eGFR_{creatinine}$-association showed significant interaction with albuminuria levels (normo-, micro-, and macroalbuminuria; p = 0.03). An aggregated genetic risk score (GRS) was associated with higher urine albumin levels and $eGFR_{creatinine}$. The *rs141640975* variant was also associated with higher levels of $eGFR_{creatinine\text{-}cystatin\ C}$ (ml/min/1.73 m^2, $CKD\text{-}EPI_{2021}$, natural log-transformed) and lower circulating cystatin C levels.

Conclusions: The positive associations between the four *CUBN* missense variants and eGFR in a large population without diabetes suggests a pleiotropic role of *CUBN* as a novel eGFR-locus in addition to it being a known albuminuria-locus. Additional associations with diverse renal function measures (lower cystatin C and higher $eGFR_{creatinine\text{-}cystatin\ C}$ levels) and a *CUBN*-focused GRS further suggests an important role of *CUBN* in the future personalization of chronic kidney disease management in people without diabetes.

KEYWORDS

genetics, *CUBN*, cubilin, kidney function, eGFR, diabetes, non-diabetes, chronic kidney disease (CKD)

1 Introduction

Urine albumin or albuminuria is one of the most important biomarkers of kidney damage in individuals with or without diabetes. In healthy individuals, the glomerular filter in the kidneys retains most of the albumin, although a small amount can usually pass through to the tubular system (1). Reabsorption of albumin is facilitated by the kidney's proximal tubular cells (PTCs), ensuring that almost no albumin is excreted in urine under normal conditions (2, 3). Elevated excretion of albumin in the urine - initially coined as "microalbuminuria" - is one of the earliest signs of chronic kidney disease (CKD) and may be the kidney-related manifestation of general endothelial damage, where scarring of the glomerulus causes chronic leakiness through the filter of albumin and other proteins (4).

Over the past decades, the number of people with diabetes mellitus has more than doubled to a global prevalence of 537 million in 2021 (5), with serious consequences for the healthcare system and society. According to a recent European study (6), one in four hospitalized patients has diabetes. Up to 40% of individuals with diabetes develop diabetic kidney disease (DKD), which is associated with elevated cardiovascular morbidity and mortality and progresses to dependency on kidney replacement therapies such as dialysis and transplantation and is a leading cause of CKD (7).

In the recent years, studies have begun to unravel genetic aspects of albuminuria. Recently, we and others identified that genetic variants (single nucleotide variants (SNVs)) in the gene encoding for cubilin (*CUBN*) – the main albumin-transporter in PTCs (1, 8) – are associated with microalbuminuria and higher urine albumin levels in populations with and without diabetes (8–14). Four variants in the C-terminal end of cubilin have been of particular interest (*rs141640975 (c.5069C>T; p.Ala1690Val)*,

Abbreviations: ACEi, Angiotensin-converting enzyme inhibitors; AER, Albumin excretion rate (mg/24 hours); ALB, Urinary albumin level (mg/L); ARBs, Angiotensin receptor blockers; CKD, Chronic kidney disease; CKD-EPI, CKD Epidemiology Collaboration; *CUBN*, The gene encoding cubilin; DM, Diabetes mellitus; eGFR, Estimated glomerular filtration rate (ml/min/1.73 m^2); GRS, Genetic risk score; NDM, Non-diabetes; T1D, Type 1 diabetes; T2D, Type 2 diabetes; UACR, Urinary albumin-creatinine ratio (mg/mmol); UKBB, UK Biobank.

rs144360241 (c.6469A>G; p.Asn2157Asp), *rs45551835 (c.8741C>T; p.Ala2914Val)*, and *rs1801239 (c.8950A>G, p.Ile2984Val)*); these are functional (missense) variants that have been proposed to alter the function of cubilin, leading to a form of albuminuria that may reflect a lack of tubular reabsorption of albumin (i.e., tubular albuminuria) (8). *In silico* structural and damage prediction analyses of the variants indicate their potential to change secondary or even tertiary structure(s) in the cubilin protein and to have different degrees of damaging effects on protein function, disease, or both (8). Our recent study further suggests that the effect of some of these variants on urine albumin levels is 2-3 times higher in diabetes compared to non-diabetes (11).

However, the role of these *CUBN* variants in relation to estimated glomerular filtration rate (eGFR), a clinically used marker of kidney function, is largely unexplored, and most genetic studies have focused on the general population (8, 9, 11). Recent efforts to uncover the role of these variants specifically in diabetes – and to clearly separate the effect seen here from the effect in the non-diabetes-proportion of the general population – have been performed as relatively small secondary analyses without including *rs144360241* or diabetes subtypes (8). Thus far, only *rs45551835* has been connected to higher levels of eGFR in type 2 diabetes and *rs141640975* in non-diabetes (8). Therefore, we investigated the relationship between the four *CUBN* variants and eGFR in different contexts: First, we meta-analyzed studies of SNV-$eGFR_{creatinine}$ regressions in Europeans with type 1 (T1D) or type 2 diabetes mellitus (T2D). We then examined single- and aggregate-variant associations separately in diabetes and non-diabetes populations of a large, nationally representative cohort facilitating application of identical phenotype definitions, including the dependency of albuminuria-stage in SNV-$eGFR_{creatinine}$ associations, generation of a *CUBN*-specific genetic risk score (GRS), and identification of associations between individual SNVs and cystatin C-based measures of kidney function. Together, these analyses both seek to replicate previous associations in DM and NDM populations and to provide novel insights into the link between *CUBN* and eGFR.

2 Methods

2.1 Study design and cohorts

For the genetic association meta-analysis in diabetes mellitus (DM), we included data collected *via* three approaches (Figure 1): First, we acquired summary statistics from up to 15,200 individuals of European origin with either type 1 diabetes (T1D) or type 2 diabetes (T2D) subsetted from six cohorts: AfterEU (T1D) (15–18), Rotterdam (T2D) (19), DiaGene (T2D) (20), UK-ROI (T1D) (21), Genesis (T1D) (22) and ANDIS (T2D) (23). These studies (hereafter referred to as "DM cohorts") were invited to the study and given a harmonized analysis plan provided that any subset of the requested genetic variants was available. A description of each cohort can be found in the Supplemental text.

Second, we applied the same analysis plan to a subset of individuals with T2D (n ~ 14,860) from the UK Biobank (24) (henceforth referred to as "UKBB-T2D"). The approach we used to extract the T2D subset has been described previously (25, 26).

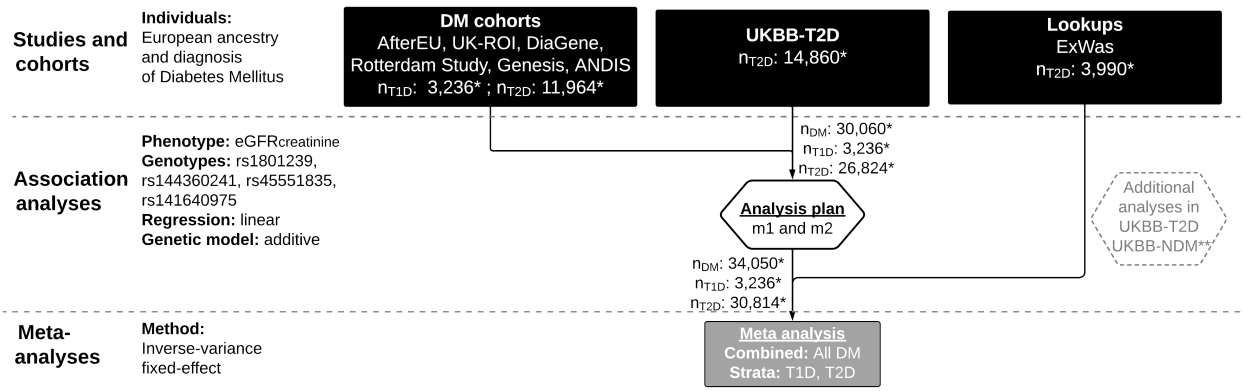


FIGURE 1
Flow chart of SNV-eGFRcreatinine meta-analyses in Diabetes. UKBB, UK Biobank; T2D, Type 2 diabetes; DM, diabetes mellitus; T1D, Type 1 diabetes; NDM, without diabetes mellitus; SNV, single nucleotide variant; eGFRcreatinine, Estimated glomerular filtration rate, natural log-transformed; PCs, Principal components of population structure; HbA1C, hemoglobin A1C; SBP, Systolic blood pressure; m1: model 1 (eGFRcreatinine ~ genotype + sex + age + 0-10 PCs); m2: model 2 (m1 + HbA1c + SBP + diabetes duration); * Sample sizes (n) reflect the maximal number of individuals (out of the total number of individuals in Table 1) available for rs45551835, model 1. ** See Supplementary Figure 1 for a flow chart of additional analyses. Figure made with LucidChart (lucid.app).

Third, we did a lookup in a subset of an exome-wide association study (henceforth referred to as "ExWas") that included 3,990 individuals with T2D from three Danish studies (Inter99, Vejle biobank and Addition-DK) described previously (11).

We also searched the Type 2 Diabetes Knowledge Portal [at time of search: www.type2diabetesgenetics.org, now: https://t2d.hugeamp.org/ (27)] for large-scale studies with publicly available summary statistics fulfilling the following criteria: Summary statistics should *a)* be readily available through the knowledge portal or a direct link to a study website; *b)* be available for diabetes-stratified and European-only populations; *c)* include at least one target genotype; *d)* be based on natural log-transformed eGFR values rather than non-transformed eGFR values; and *e)* be based on regression models with covariate adjustments comparable to those in the other cohorts in this study. However, as of 10 July 2020, no studies in the portal fulfilled our criteria, and no additional studies were included.

For additional analyses, we used *1)* a group of individuals without diabetes from UKBB (n ~ up to 370,000 individuals), henceforth referred to as "UKBB-NDM") and *2)* the UKBB-T2D group, which was also part of the meta-analysis (Supplementary Figure 1). 127,090 non-diabetes individuals from the Icelandic study deCODE participated as the replication cohort (Supplemental text).

This research work was conducted in accordance with the Helsinki Declaration. Ethical approval was previously obtained locally for individual studies. All participants gave written informed consent before participating.

2.2 Phenotype details

For the DM cohorts and UKBB (both NDM and T2D groups), we calculated the creatinine-based estimated glomerular filtration rate ($eGFR_{creatinine}$) with the Chronic Kidney Disease Epidemiology Collaboration creatinine equation ($CKD\text{-}EPI_{creatinine(2012)}$, ml/min/1.73 m^2 (28), natural log-transformed). We included it here as a continuous variable. Other measures of kidney function were also calculated for UKBB; see section 2.4.2.4.

2.3 Genotyping, imputation, quality control and variant selection

We obtained information on genotyping, imputation, and quality control of each cohort and summarized it in Supplementary Tables 1, 2.

Four variants were selected for further analysis: *rs141640975* (Chromosome (chr) 10, position (pos) 16992011 (genome-build GRCh37.p13)) with minor allele frequency (MAF) 0.002-0.009; *rs144360241* (chr 10, pos 16967417) with MAF 0.006-0.010; *rs45551835* (chr 10, pos 16932384) with MAF 0.016-0.021; and *rs1801239* (chr 10, pos 16919052) with MAF 0.097-0.114. For the deCODE study, the MAFs were in the same range except *rs144360241* (MAF: 0.002). The minor alleles of these variants (A, C, A, and C, respectively) were used as effect alleles.

We used LDlink version 5.1 (29) with the European (CEU + GBR) reference panel to confirm the independent relationship (Linkage Disequilibrium (LD) $r^2 < 0.1$) between these SNVs.

The SNVs were first used in single-variant analyses and were then combined into a genetic risk score (GRS; see description below).

2.4 Statistical methods

A flow chart of the meta-analyses is shown in Figure 1, and one of the additional analyses is shown in Supplementary Figure 1.

2.4.1 Study-level SNV-$eGFR_{creatinine}$ association analysis in diabetes and subsequent meta-analysis

In each DM cohort and UKBB-T2D, associations between $eGFR_{creatinine}$ and genetic variants were assessed assuming an additive genetic model. We used natural log-transformed $eGFR_{creatinine}$ in a linear regression model (model 1) adjusted for traditional clinical and genetic factors, i.e. age, gender, and study-specific covariates (i.e., 0-10 principal components of population structure to account for population stratification). To control for potential bias on kidney function in the diabetes population, another model was further adjusted for HbA_{1C}, systolic blood pressure (a proxy for medication with Angiotensin receptor blockers (ARBs) or Angiotensin-converting enzyme inhibitor (ACEi) frequently used in diabetes treatment) and diabetes duration (model 2). Some of the cohorts used summary statistics calculated prior to our query, so we allowed minor deviations in the included covariates (Supplementary Table 3). A list of software used for association analysis can be found in Supplementary Table 1. Each study dealt with missing data separately. Once all summary results were collected, we performed study-level quality control. Summary results were meta-analyzed using a fixed-effect inverse-variance method in the "Metagen" package in R (version 3.6.3). We report results in any diabetes mellitus subtype (denoted "combined") and in T1D and T2D subsets. Significant heterogeneity ($P_{het} < 0.05$) indicated variation across studies. Effect sizes (betas) are presented with 95% confidence intervals. We evaluate statistical significance at an FDR-corrected level of 0.05/4 = 0.0125 considering the number of tested SNVs.

2.4.2 Additional analyses in UKBB populations with diabetes and non-diabetes

To explore the interplay between *CUBN*-variants and kidney-related traits in more detail, we did a range of additional linear regressions in the UKBB NDM and T2D groups. Further, we also applied a combined genetic risk score (GRS). We based the analyses on model 1 and model 3. The latter was very similar to model 2, in that it included adjustment for model 1 and SBP but not HbA_{1c} and diabetes duration. The last two adjustments were absent from this model because they are less relevant in non-diabetes. We applied the same models in DM and NDM to provide consistency. Individuals were excluded if they had missing data for any variable.

2.4.2.1 SNV-eGFR$_{creatinine}$ association analysis in the UKBB population without diabetes and replication in the deCODE study

We examined SNV-eGFR$_{creatinine}$ associations in the UKBB NDM and T2D populations. It was advantageous to use the UKBB dataset here as it is a well-powered, phenotypically homogenous dataset (n ~ up to 370,000 individuals without diabetes). Since effects are based on natural log-transformed eGFR (trait) values, we also calculated the percental difference in mean, non-transformed eGFR per added effect allele for significant effects as follows: $\%\ difference = (e^{beta-1})*100\%$. . Again, we evaluated statistical significance at an FDR-corrected level of 0.0125.

SNV-eGFR$_{creatinine}$ associations identified in the UKBB NDM group were also examined in the Icelandic deCODE study (n_{NDM}=127,090) applying model 3.

2.4.2.2 Interaction with albuminuria

In order to examine whether the SNVs associated with eGFR$_{creatinine}$ in an albuminuria-dependent fashion, we assessed albuminuria-SNV interactions in SNV-eGFR$_{creatinine}$ regression models in individuals with T2D (n_{T2D} = 7,777) and without DM (n_{NDM} = 107,276) for whom continuous urine albumin levels were available (derived from the UKBB "microalbumin" field). The interaction term in the regression models included albuminuria groups as a factor defined from these albumin levels as follows: *i) normoalbuminuria:* =< 30 mg/L (n_{DM} = 5,566, n_{NDM} = 93,728), *ii) microalbuminuria:* 30-300 mg/L (incl. lower but not upper threshold, nN_{DM} = 1,954, n_{NDM} = 12,690), and *iii) macroalbuminuria:* >300 mg/L (incl. lower threshold, n_{DM} = 257, n_{NDM} = 858). We used regression models based on model 1 and 3 (i.e., *model 1*: ln(eGFR$_{creatinine}$) ~ SNV + albuminuria group + age + sex + SNV*albuminuria group and *model 3*: model 1 + SBP). A significant p-value (< 0.05) for the SNV*albuminuria interaction term was considered evidence for interaction. Interaction analysis was done whenever primary SNV-eGFR$_{creatinine}$ analyses were well-powered.

2.4.2.3 Genetic risk score association with microalbuminuria and eGFR$_{creatinine}$

We estimated an albuminuria genetic risk score (GRS) using the four albuminuria-associated *CUBN* missense SNVs. The GRS was generated for each study participant using the sum of individual SNV effect alleles in the UKBB dataset. We then examined the associations between GRS_{CUBN} and continuous urine microalbumin levels (mg/L) and eGFR$_{creatinine}$.

2.4.2.4 SNV vs. other kidney function-related traits in UKBB

We examined the associations between the study SNVs and 1) circulating serum Cystatin C levels (mg/L) and 2) the more recent eGFR$_{creatinine\text{-}cystatin\ C}$ equation (30) that uses both serum creatinine and cystatin C levels and applies to all ethnicities.

2.4.3 Power calculations

We used Quanto (version 1.2.4) (31) to calculate *post-hoc* power for main SNV-eGFR$_{creatinine}$ associations in DM and NDM groups.

For all power calculations in Quanto, we: *a)* chose a continuous design for independent individuals; *b)* assumed a gene-only hypothesis; *c)* assumed an additive inheritance mode; and *d)* set the two-sided type I error-rate to 0.05.

For the remaining options in Quanto, we typed in information specific to each variant and population (Supplementary Tables 13-14): For each variant, we used allele frequencies of the effect allele; for meta-analyses, this was done as a range of calculations spanning the frequencies reported by individual cohorts. We used effect sizes obtained through DM and NDM SNV-eGFR$_{creatinine}$ association analyses (main effect). Means and standard deviations of ln (eGFR$_{creatinine}$) were derived from UKBB subsets. Unless otherwise specified, total DM sample sizes were used.

3 Results

3.1 Clinical characteristics

Up to 34,743 individuals with diabetes mellitus (type 1 diabetes (T1D), n ~ 3,588, or type 2 diabetes (T2D), n ~ 31,155) and up to 370,061 without diabetes participated in the current study (Figure 1 and Supplementary Figure 1). Clinical characteristics of participating studies can be found in Table 1 and Supplementary Tables 4–7.

3.2 *CUBN* variants are not associated with eGFR$_{creatinine}$ in a diabetes meta-analysis

The effect of *rs144360241* on eGFR$_{creatinine}$ was studied in 32,904 individuals with diabetes. The variant was not available in UK-ROI (Supplementary Figures 2, 6). All eight studies contributed to the 34,050 individuals analyzed for *rs45551835* (Supplementary Figures 3 and 7). The *rs141640975* variant was available for 32,993 individuals and was unavailable in UK-ROI (Supplementary Figures 4, 8). The common variant, *rs1801239*, was available in all eight studies in 34,070 individuals (Supplementary Figures 2, 9).

After meta-analysis, none of the four *CUBN* variants were significantly positively associated with eGFR$_{creatinine}$ in the DM group, neither in the T1D or T2D subgroup [Table 2 (Model 1) and Table 3 (Model 2)]. However, the positive directionality of the effect for the T2D group was consistent with the directionality of effect for the combined group for all variants with non-zero effects. The T2D group carried the largest weight in the combined meta-analyses and UKBB carried the largest weight within the T2D group (Supplementary Figures 2–5). There was no evidence of heterogeneity across studies, except in model 2 for *rs45551835* and *rs1801239* (Table 3).

10.3389/fendo.2023.1081741

TABLE 1 Clinical characteristics of participating studies.

Study name	DM type	Individuals (N)	Males (N, %)	Age[##] [years]	BMI [kg/m^2]	eGFR$_{creatinine}$ [ml/min/1.73 m^2]	SBP [mmHg]	Diabetes duration [years]	Urinary albumin		
									AER [mg/24h]	UACR [mg/mmol]	ALB [mg/L]
AfterEU	T1D	854	492 (57.60)	43.67 (11.15)	24.23 (3.21)	89.48 (26.61)	139.22 (20.90)	28.02 (9.50)	29.00 (7.00 - 618.00)	NR	NR
UK-ROI	T1D	1,410	716 (50.80)	45.09 (11.35)	26.30 (4.40)	54.30 (30.00)	135.02 (20.80)	30.45 (9.70)	NA	NA	NA
GENESIS	T1D	1,324	700 (52.90)	41.37 (12.21)	22.21 (8.15)	80.87 (28.49)	129.41 (23.75)	24.91 (10.45)	9.00 (4.16-37.25)	NR	NR
DiaGene	T2D	1,886	1,011 (53.60)	65.24 (10.57)	30.47 (5.43)	78.33 (20.55)	141.83 (18.72)	10.09 (8.45)	NR	5.85 (30.45)	NR
Rotterdam	T2D	1,022	487 (47.70)	68.10 (9.70)	29.40 (4.80)	78.30 (16.40)	147.10 (21.70)	NA	NA	NA	NA
ANDIS	T2D	9,367	5,548 (59.22)	66.29 (13.29)	30.77 (5.70)	84.69 (30.92)	NA	8.07 (4.40)	NA	NA	NA
ExWas**	T2D	3,990	2,370 (59.30)	61.00 (8.50)	NA	79.00 (1.28)	NA	NA	NA	NA	NA
UKBB-T2D[#]	T2D	14,890	9,703 (65.10)	60.97 (6.28)	31.90 (5.70)	87.86 (15.73)	144.50 (18.20)	NA	NR	NR	16.00 (10.00-34.40)
UKBB-NDM[#]	NR*	370,061	166,976 (45.10)	56.73 (8.02)	27.10 (4.50)	90.81 (12.80)	139.90 (19.60)	NR	NR	NR	11.10 (8.30-18.10)

*Non-DM population. **The ExWas study comprises summary data from T2D individuals (discovery set). N, sample size; SD, standard deviation; BMI, Body-Mass Index; eGFR$_{creatinine}$, estimated glomerular filtration rate based on the CKD-EPI$_{2012}$ equation (non-transformed); SBP, Systolic blood pressure; AER, albumin excretion rate; IQR, Interquartile range; UACR, urinary albumin-creatinine ratio; UKBB, UK Biobank; ALB, continuous baseline urinary albumin level; T2D, Type 2 diabetes; DM, diabetes mellitus; NDM; non-DM; T1D, type 1 diabetes. NR, not relevant; NA, not available. [#]The UK Biobank urinary albumin measures are based on n=7,777 in T2D and n=370,061 in the NDM group. [##]The time point for age assessment is NA for Genesis. Age at recruitment was used in all other studies. Age, BMI, eGFR, and SBP have been deonted as mean (SD), while Urinary albumin measures have been denoted as median (IQR).

TABLE 2 meta-analysis of SNV-eGFR$_{creatinine}$ summary data in diabetes mellitus and its subtypes (model 1).

Genetic variant (EA)	Diabetes type	N	Effect (Beta [95% CI])	P$_{HET}$	P-value
rs144360241 (C)	T1D	2,177	-0.14 [-0.32; 0.05]	0.38	0.15
	T2D	30,727	0.01 [-0.01; 0.04]	0.62	0.40
	Combined DM	32,904	0.01 [-0.02; 0.04]	0.42	0.53
rs45551835 (A)	T1D	3,236	-0.02 [-0.13; 0.08]	0.34	0.69
	T2D	30,814	0.01 [0.00; 0.03]	0.08	0.09
	Combined DM	34,050	0.01 [0.00; 0.02]	0.15	0.10
rs141640975 (A)	T1D	2,177	0.16 [-0.11; 0.44]	0.75	0.25
	T2D	30,816	0.00 [-0.03; 0.03]	0.53	0.83
	Combined DM	32,993	0.01 [-0.02; 0.03]	0.60	0.73
rs1801239 (C)	T1D	3,236	-0.01 [-0.06; 0.03]	0.20	0.57
	T2D	30,834	0.00 [0.00; 0.01]	0.21	0.64
	Combined DM	34,070	0.00 [0.00; 0.01]	0.23	0.59

SNV, single nucleotide variant; eGFR$_{creatinine}$, log-transformed estimated glomerular filtration rate based on the CKD-EPI$_{2012}$ equation; EA, effect allele (i.e., minor allele); N, sample size; Beta, Beta coefficient; CI, confidence interval; P$_{Het}$, P-value for heterogeneity across studies. P$_{Het}$< 0.05 indicates variation; T1D: Type 1 diabetes; T2D: Type 2 diabetes; Combined DM: T1D and T2D combined.

10.3389/fendo.2023.1081741

TABLE 3 meta-analysis of SNV-eGFR$_{creatinine}$ summary data in diabetes mellitus and its subtypes (model 2).

Genetic variant (EA)	Population	N	Effect (Beta [95% CI])	P_{HET}	P-value
	T1D	1,916	-0.12 [-0.32; 0.08]	0.26	0.25
	T2D	15,745	0.01 [-0.02; 0.04]	0.37	0.66
rs144360241 (C)	Combined DM	17,661	0.00 [-0.02; 0.03]	0.32	0.78
	T1D	2,712	-0.05 [-0.16; 0.07]	0.25	0.43
	T2D	15,724	0.01 [0.00; 0.03]	0.03*	0.14
rs45551835 (A)	Combined DM	18,436	0.01 [0.00; 0.03]	0.05	0.18
	T1D	1,916	0.10 [-0.17; 0.38]	0.4	0.46
	T2D	15,746	0.00 [-0.04; 0.05]	0.58	0.88
rs141640975 (A)	Combined DM	17,662	0.01 [-0.04; 0.05]	0.67	0.8
	T1D	2,712	0.00 [-0.04; 0.05]	0.53	0.94
	T2D	15,741	0.00 [-0.01; 0.01]	0.03*	0.77
rs1801239 (C)	Combined DM	18,453	0.00 [-0.01; 0.01]	0.15	0.76

SNV, single nucleotide variant; eGFR$_{creatinine}$, log-transformed estimated glomerular filtration rate based on the CKD-EPI$_{2012}$ equation; EA, effect allele (i.e., minor allele); N, sample size; Beta, Beta coefficient; CI, confidence interval; P_{Het}, P-value for heterogeneity across studies. P_{Het}< 0.05 indicates variation; T1D: Type 1 diabetes; T2D: Type 2 diabetes; Combined DM: T1D and T2D combined.

3.3 *CUBN* variants are associated with higher eGFR$_{creatinine}$ in non-diabetes

In UKBB-NDM, we observed larger eGFR$_{creatinine}$-levels for minor alleles compared to major alleles for all four *CUBN* variants in both models, except for *rs1801239* in NDM, model 3 (Table 4 and Supplementary Table 8): The effect and standard deviation of *rs144360241* was, for model 1 (model 3), 0.008 ± 0.002 (0.007 ± 0.002), corresponding to a difference of +0.8% (+0.7%) in mean eGFR$_{creatinine}$ (ml/min/1.73 m^2) for each additional copy of the affect allele, *C*. For *rs45551835*, the effect was 0.005 ± 0.001 (0.004 ± 0.001), corresponding to a difference of +0.5% (+0.4%) in mean eGFR$_{creatinine}$ per copy of the *A*-allele. *rs141640975* had the largest effect size, 0.02 ± 0.003 (0.02 ± 0.003), corresponding to a +2.02% (+2.02%) difference in mean eGFR$_{creatinine}$ for each additional *A*-allele. The common variant, *rs1801239*, had the smallest effect size of 0.001 ± 0.0005, corresponding to a +0.1% difference in eGFR$_{creatinine}$ for each *C*-allele. We replicated the finding that *rs141640975* was significantly associated with higher eGFR$_{creatinine}$ in non-diabetes in an Icelandic study (deCODE, n = 127,090, effect = 0.026, SE = 0.007, $P_{eGFR\_creatinine} = 7.7 \times 10^{-4}$, model 3, Supplementary Table 8). None of the other SNVs were replicated (data not shown). Meta-analysis for the *rs141640975*-eGFR-association in the NDM studies (UKBB and deCODE) is depicted in Supplementary Figure 10.

In UKBB-T2D, none of the variants had statistically significant associations with eGFR$_{creatinine}$, although the effects of three of the variants (except *rs141640975*) were in the same direction as in NDM (Table 4 and Supplementary Table 8).

TABLE 4 Summary results for SNV-eGFR$_{creatinine}$ analyses in UKBB (model 1).

Genetic variant (EA)	EAF	Population **	N	Effect (Beta [SE])	P-value
	0.004	NDM ***	369,832	0.008 (0.002)	0.0008*
rs144360241 (C)	0.004	T2D ****	14,882	0.02 (0.02)	0.23
	0.014	NDM ***	369,028	0.005 (0.001)	0.0004*
rs45551835 (A)	0.014	T2D ****	14,860	0.01 (0.01)	0.13
	0.003	NDM ***	369,987	0.02 (0.003)	2.2×10^{-9}*
rs141640975 (A)	0.003	T2D ****	14,885	-0.01 (0.02)	0.71
	0.10	NDM ***	369,849	0.001 (0.0005)	0.006*
rs1801239 (C)	0.10	T2D ****	14,880	0.00 (0.00)	0.42

SNV, single-nucleotide variant; eGFR$_{creatinine}$, estimated glomerular filtration rate (natural log-transformed); EA, effect allele (i.e., minor allele); N, sample size; EAF, Effect allele frequency; Beta, Beta coefficient; SE, standard error; NDM, without Diabetes Mellitus; T2D, Type 2 diabetes. *Statistically significant (P< 0.05). ** For completeness, we also show the results for T2D, which were part of DM meta-analyses for model 1. *** out of total 370,061 individuals. **** out of total 14,892 individuals.

10.3389/fendo.2023.1081741

TABLE 5 Interaction with albuminuria in SNV-$eGFR_{creatinine}$ analyses in UKBB (model 1).

Genetic variant (EA)	Population	N	P-value of interaction term#
rs144360241 (C)	NDM **	107,202	0.67
rs45551835 (A)	NDM **	106,964	0.88
rs141640975 (A)	NDM **	107,255	0.03*
rs1801239 (C)	NDM **	107,216	0.49

SNV, single-nucleotide variant; $eGFR_{creatinine}$, estimated glomerular filtration rate (natural log-transformed); EA, effect allele (i.e., minor allele); N, sample size; NDM, without Diabetes Mellitus; *Statistically significant (P< 0.05). ** out of total 107,276 individuals with continuous urinary albumin levels. Albuminuria-SNV interaction was only tested when primary SNV-eGFRcreatinine associations were significant. # Interaction term is SNV*albuminuria groups (normo-, micro-, and macro albuminuria).

3.4 Associations of *rs141640975* with $eGFR_{creatinine}$ depend on albuminuria-status in non-diabetes

To examine whether the SNVs are associated with $eGFR_{creatinine}$ in an albuminuria-dependent fashion, we included albuminuria*SNV interactions in two regression models. For the first model, we observed significant interaction for *rs141640975* in UKBB-NDM ($P_{interaction}$ = 0.03, Table 5). This was also observed in the other model ($P_{interaction}$ = 0.04, Supplementary Table 9). An interaction plot showed that for the eGFR-SNV-association, the effect on eGFR was even higher for more elevated albuminuria-levels (Supplementary Figure 11).

3.5 A *CUBN*-based GRS for albuminuria is associated with $eGFR_{creatinine}$ in non-diabetes

We combined the four *CUBN* variants into a genetic risk score for albuminuria, verified its associations with continuous urine albumin levels and tested it against $eGFR_{creatinine}$ in UKBB-T2D and UKBB-NDM. The GRS was associated with higher levels of both traits, except for eGFR in T2D (Tables 6, 7).

TABLE 6 Summary results for GRS_{CUBN}-$eGFR_{creatinine}$ and -ALB analyses in UKBB (model 1).

Trait	Population	N	Effect (Beta [SE])	P-value
ALB	NDM **	106,814	0.05 (0.004)	2×10^{-16}*
	T2D ***	7,741	0.08 (0.02)	0.004*
$eGFR_{creatinine}$	NDM	368,521	0.002 (0.0004)	2×10^{-6}*
	T2D	14,837	0.004 (0.003)	0.2

GRS_{CUBN}, A genetic risk score based on a combination of the four CUBN genetic variants (minor alleles); N, sample size; Beta, Beta estimate; SE, standard error; ALB, continuous urinary albumin (mg/L, natural log-transformed); $eGFR_{creatinine}$, estimated glomerular filtration rate (natural log-transformed); NDM, without Diabetes Mellitus; T2D, Type 2 diabetes. *Statistically significant (P< 0.05). ** out of total 107,276 individuals with continuous urinary albumin levels. *** out of total 7,777 individuals with continuous urinary albumin levels.

TABLE 7 Summary results for GRS_{CUBN}-$eGFR_{creatinine}$ and -ALB analyses in UKBB (model 3).

Trait	Population	N	Effect (Beta [SE])	P-value
ALB	NDM **	99,180	0.05 (0.004)	2×10^{-16}*
	T2D ***	7,182	0.08 (0.02)	3×10^{-4}*
$eGFR_{creatinine}$	NDM	343,988	0.002 (0.0004)	2×10^{-5}*
	T2D	13,828	0.005 (0.003)	0.1

GRS_{CUBN}, A genetic risk score based on a combination of the four CUBN genetic variants (minor alleles); N, sample size; Beta, Beta estimate; SE, standard error; ALB, continuous urinary albumin (mg/L, natural log-transformed); $eGFR_{creatinine}$, estimated glomerular filtration rate (natural log-transformed); NDM, without Diabetes Mellitus; T2D, Type 2 diabetes. *Statistically significant (P< 0.05). ** out of total 107,276 individuals with continuous urinary albumin levels. *** out of total 7,777 individuals with continuous urinary albumin levels.

3.6 *rs141640975* is associated with additional markers of kidney function in non-diabetes

We examined the associations between the study SNVs and two additional markers of kidney function. The SNV *rs141640975* was associated with higher levels of $eGFR_{creatine\text{-}cystatin\ C}$ [a more recent ethnicity-independent GFR-estimator (28)] and lower levels of cystatin C, both observed in NDM (Supplementary Tables 10–12). The $eGFR_{creatinine\text{-}cystatin\ C}$ association of *rs144360241* was borderline significant in NDM.

3.7 Estimated power

3.7.1 Meta-analysis (diabetes mellitus)

Given the ranges of EAFs obtained from individual studies participating in meta-analyses, we reached a power level of 35-43% for *rs45551835*, 16-23% for *rs1444360241*, and 9-21% for *rs141640975* in the DM group (Supplementary Table 14). Effect sizes were assumed from the individual meta-analysis $eGFR_{creatinine}$-associations of each SNV. We did not calculate power for *rs1801239* as the effect in the DM meta-analysis was 0.0.

3.7.2 Association of SNVs with eGFR (UKBB population without diabetes)

In NDM, the power for main $eGFR_{creatinine}$ analyses was between 70-99% for the four variants (Supplementary Table 15).

4 Discussion

Recently, we demonstrated that individuals carrying the minor allele of the *CUBN* missense variant *rs141640975* had higher albuminuria-levels than non-carriers. The effect of this variant was stronger in individuals with diabetes (DM) compared to those without diabetes (NDM) (11). In continuation of these findings, Bedin et al. (8) performed secondary lookups for *CUBN*-variants in

the CKDGen eGFR GWAS study population, reporting that missense variants in *CUBN* may also be associated with higher levels of eGFR in the general population. Our current large-scale study aimed to examine the effect of minor alleles of three rare *CUBN* missense variants (*rs144360241 (c.6469A>G; p.Asn2157Asp)*, *rs45551835 (c.8741C>T; p.Ala2914Val)* and *rs141640975* (*c.5069C>T; p.Ala1690Val*)) and one common variant (*rs1801239 (c.8950A>G; p.Ile2984Val)*) on $eGFR_{creatinine}$ levels separately in people with and without diabetes (n_{DM} ~ 34,000 individuals, n_{NDM} ~ 370,000 individuals), including stratification for diabetes-type and supplemented by tests on circulating cystatin C levels, the recently updated eGFR-equation based on creatinine and cystatin C (30), and aggregate-variant tests. We were able to replicate the association between creatinine-based eGFR and *rs141640975* in NDM and report new insightful connections with the alternative measures of kidney function for all four SNVs.

Previously, a borderline association between *rs45551835* and higher eGFR-levels has been reported in a smaller type 2 diabetes (T2D) population from Denmark (8, 11), a finding which we could not replicate in our meta-analysis of up to 34,432 individuals with diabetes and its subtypes. Like the initial study (8), we could not establish a link between eGFR and the three other variants within the diabetes group. As for *rs45551835*, it was surprising to be unable to replicate the earlier findings as the current study has a larger sample size compared to earlier efforts. Our *post-hoc* power assessment indicated that insufficient power might be at play, even with a larger sample size for the diabetes group (8). We also speculated whether the apparent lack of association between *CUBN* and eGFR in our diabetes meta-analysis could be due to use of Angiotensin receptor blockers (ARBs) or Angiotensin-converting enzyme inhibitor (ACEi) medication which is frequently used in diabetes treatment. As part of our sensitivity analyses, we included models adjusted for systolic blood pressure (a proxy for such medication) and did not find evidence that this could explain why no association was found in the diabetes group. Another reason could be the allele frequency of the variants may differ between Danish and UK populations. We need further validation in well-powered populations to confirm the relationship between the *rs45551835* and eGFR in diabetes, especially in T2D. In case of a true lack of association, *CUBN* may be associated with higher levels of urine albumin (11) with no pleiotropic effect to eGFR in this population.

We proceeded to single- and aggregate-variant analyses in the UK Biobank (UKBB), shifting focus to non-diabetes populations. For all four *CUBN* variants, we report significantly higher $eGFR_{creatinine}$-levels in individuals without diabetes harboring more copies of the minor alleles compared to individuals with fewer or no copies of the minor alleles in the same group. For *rs141640975*, we observed the strongest association with $eGFR_{creatinine}$ ($P = 2.2 \times 10^{-9}$) with replication in the Icelandic study (deCODE, $P = 7.7 \times 10^{-4}$), confirming what has previously been observed for this SNV in NDM (8) – but also a significant interaction between the SNV and albuminuria stages ($P_{INT} < 0.05$). Taken together with the already known associations of the minor alleles with higher albuminuria (11), this not only demonstrates genetic pleiotropy of *CUBN* for albuminuria and eGFR in non-diabetes but also implies that these two associations are intertwined for this SNV, where the effect on eGFR is even higher for more elevated albuminuria-levels. Here, *CUBN* demonstrates a classic genetic pleiotropy phenomenon where a DNA variant influences multiple traits, usually in the same domain with concordant or sometimes discordant effects as observed earlier in complex disorders (32). Further validation of independent biological or related causal effects might be required in additional follow up studies.

This finding is unusual as there is no obvious clinical or pathophysiological explanation for such an albuminuria-eGFR pattern in the context of non-diabetes. It has been suggested that the tubular albuminuria observed in presence of C-terminal variants in *CUBN* has a benign or even slightly protective effect on kidney function in chronic kidney disease if glomerular albuminuria is also present (8, 33, 34). Another recent study on chronic isolated proteinuria suggests that different C-terminal *CUBN* variants uncouple proteinuria from glomerular filtration barrier through declined cubilin expression accompanied by aberrant amnionless (AMN) localization in renal tubules. AMN is part of the receptor complex (along with cubilin and megalin) necessary for tubular reabsorption of albumin. This is suggested to create a benign condition, not requiring any further proteinuria lowering treatment (35). In non-diabetes, where the population can be assumed to consist mostly of healthy individuals, a concept of such protectiveness is less relevant. However, it is possible that an undetected subpopulation with relevant comorbidities exists in the non-diabetes group.

Our *CUBN* aggregate-variant method – which was defined as a genetic risk score (GRS) combining the four variants – showed that a higher number of C-terminal *CUBN* risk alleles is associated with higher urine albumin and $eGFR_{creatinine}$ levels and confirms both the single-variant association with higher urine albumin levels reported previously in diabetes and non-diabetes (11, 14), and the consistency of the overall effects on urine albumin levels being greater in diabetes compared to non-diabetes (10, 11). Through GRS_{CUBN}, we also saw that a higher number of minor alleles across the four variants was associated with higher $eGFR_{creatinine}$-levels in the UKBB population without diabetes, which is in line with our single-variant findings and the previous findings for *rs45551835* (8). Using aggregate-variant methods is an optimal way to examine combined genetic effects and has been used extensively for polygenic traits (13, 36). Using GRS is highly relevant here as three of the four variants are rare and mostly present as heterozygous variants in our populations. This might substantiate with some additional power to detect effects and adds further certainty to the presence of a *CUBN*-eGFR relationship in non-diabetes. Nevertheless, we still do not find an association with eGFR in T2D, even when the variants are combined in a GRS.

Finally, we examined the association between the study SNVs and two alternative markers of kidney function. In non-diabetes, the minor alleles of *rs141640975* and *rs144360241* were associated with higher levels of $eGFR_{creatinine\text{-}cystatin\ C}$. This measure was estimated using a recent update to the equation, $CKD\text{-}EPI_{2021}$, which does not include ethnicity and is a more precise indicator of kidney function in comparison to the $CKD\text{-}EPI_{creatinine(2012)}$ equation which is based only on creatinine. Our results using the

conventional $eGFR_{creatinine}$ equation are concordant with our results from the updated equation in terms of directionality of effect and with our finding that *rs141640975* is associated with lower cystatin C levels, which is another indicator of kidney function. It should be noted, though, that considering Table 1 and Supplementary Tables 4–6, the 0.1% – 2.02% higher mean eGFR we report for each minor allele is modest and may reflect that individual harboring these genetic variants have normal kidney function rather than a better kidney function.

A strength of our study is the restriction to specifically diabetes- and non-diabetes-only subgroups so that effects from mixed diabetes-status are minimized. Heterogeneity is likely to be present in meta-analyses of a diverse set of cohorts originally used for different research purposes. Indeed, some of the cohorts included in our meta-analyses differ regarding available covariates and/or kidney disease status. However, we did not observe heterogeneity in our meta-analyses. In addition to this, we could minimize heterogeneity in the remainder of our analyses by using data from the UKBB, which is a nationally representative cohort facilitating application of identical phenotype definitions across subgroups. Another strength is the broad spectrum of additional analyses that we explored in the UKBB population to nuance our findings on the relationship between eGFR and *CUBN*. The judicious use of UKBB leveraging individual-level genotype information to investigate interaction-analyses based on albuminuria groupings is a great strength of the current study, especially for rare variants.

A major limitation is that we did not have sufficient statistical power for our meta-analyses in the diabetes group due to the limited availability of suitable datasets. Consequently, interpretations of T2D findings should not be overstated and we thus could not demonstrate, nor disprove, the presence of a *CUBN*-eGFR relationship in this population. Although we demonstrate that C-terminal missense variants in *CUBN* are associated with different measures of normal (or even higher) kidney function in non-diabetes, we emphasize that the current study is insufficient to establish causality. Finally, using multiple-testing-corrected significance thresholds might be too conservative when testing a very small number of variants from the same locus as it may remove true associations. In genome-wide studies, a conservative threshold of 5×10^{-8} is generally agreed upon for novel associations. There is less consensus on when and how to appropriately apply multiple testing correction in smaller-scale genetic studies dealing with a mixture of new and known associations. Nevertheless, we deemed that it would be fair to apply FDR-correction of the significance threshold to our primary analyses in DM and NDM.

In conclusion, the current study identifies the existence of pleiotropic genetic effects of *CUBN* on two facets of kidney function – albuminuria and eGFR – by reporting SNV-eGFR associations in a large study population without diabetes. The interaction between *rs141640975* and albuminuria-status on $eGFR_{creatinine}$ in this population and its associations with lower cystatin C and higher levels of $eGFR_{creatinine\text{-}cystatin\ C}$ expands our knowledge of these variants in relation to measures of kidney function. The demonstration of a *CUBN*-focused GRS in relation to albuminuria and $eGFR_{creatinine}$ further suggests an important role of *CUBN*-variants in the future personalization of chronic kidney disease management.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical approval has previously been obtained locally for each individual study. The patients/participants provided their written informed consent to participate in this study.

Author contributions

NU, MS, PR, and TA contributed to conception and design of the study. NU wrote the first draft of the manuscript.NU, MC-G, CH, AN, AO, SS, D-AT, EA, MH, AM, EJS, MS, PR, and TA contributed to manuscript revision. D-AT, VS, KS, EA, MH, AM, PR and TA acquired data. NU, FA, MC-G, CH, AN, SS, LS, D-AT, and TA performed statistical analysis. NU, MC-G, LS, D-AT, AM, MS, PR, and TA contributed to interpretation of data. TA and MS acquired funding and TA administered this project. PR and TA supervised the project. LC and MG had other roles. All authors contributed to the article and approved the submitted version.

Funding

NU, MS, and TA were supported by the Novo Nordisk Foundation, Steno Collaborative Grant (NNF18OC0052457). TA and PR were supported by Steno Diabetes Center Copenhagen, Herlev, Denmark. Funding of individual studies contributing to this project can be found in Supplementary Table 16.

Acknowledgments

This research has been conducted using data from UK Biobank database (www.ukbiobank.ac.uk), project ID numbers, 32683 and 71699. Acknowledgements of individual studies can be found in Supplementary Table 16.

Conflict of interest

Author MC-G was employed by Pfizer-University of Granada-Andalusian Regional Government. Authors AO, VS, and KS were employed by Amgen, Inc. PR reports personal fees from Bayer during the conduct of the study. He has received research support and personal fees from AstraZeneca and Novo Nordisk, and personal fees from Astellas Pharma, Boehringer Ingelheim, Eli

Lilly, Gilead Sciences, Mundipharma, Sanofi, and Vifor Pharma. All fees are given to Steno Diabetes Center Copenhagen.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1081741/full#supplementary-material

References

1. Amsellem S, Gburek J, Hamard G, Nielsen R, Willnow TE, Devuyst O, et al. Cubilin is essential for albumin reabsorption in the renal proximal tubule. *J Am Soc Nephrol* (2010) 21:1859–67. doi: 10.1681/ASN.2010050492

2. Yang J, Xu Y, Deng L, Zhou L, Qiu L, Zhang Y, et al. CUBN gene mutations may cause focal segmental glomerulosclerosis (FSGS) in children. *BMC Nephrol* (2022) 23:15. doi: 10.1186/s12882-021-02654-x

3. Christensen EI, Nielsen R, Birn H. From bowel to kidneys: The role of cubilin in physiology and disease. *Nephrol Dial Transplant* (2013) 28:274–81. doi: 10.1093/ndt/gfs565

4. Deckert T, Feldt-Rasmussen B, Borch-Johnsen K, Jensen T, Kofoed-Enevoldsen A. Albuminuria reflects widespread vascular damage. *Steno hypothesis Diabetologia* (1989) 32:219–26. doi: 10.1007/BF00285287

5. International Diabetes Federation. *IDF Diabetes Atlas, 10th edn*. Brussels, Belgium: International Diabetes Federation. (2021), 2021.

6. Kufeldt J, Kovarova M, Adolph M, Staiger H, Bamberg M, Haring HU, et al. Prevalence and distribution of diabetes mellitus in a maximum care hospital: Urgent need for HbA1c-screening. *Exp Clin Endocrinol Diabetes* (2018) 126:123–9. doi: 10.1055/s-0043-112653

7. Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: Challenges, progress, and possibilities. *Clin J Am Soc Nephrol* (2017) 12:2032–45. doi: 10.2215/CJN.11491116

8. Bedin M, Boyer O, Servais A, Li Y, Villoing-Gaude L, Tete MJ, et al. Human c-terminal CUBN variants associate with chronic proteinuria and normal renal function. *J Clin Invest* (2020) 130:335–44. doi: 10.1172/JCI129937

9. Boger CA, Chen MH, Tin A, Olden M, Kottgen A, de Boer IH, et al. CUBN is a gene locus for albuminuria. *J Am Soc Nephrol* (2011) 22:555–70. doi: 10.1681/ASN.2010060598

10. Casanova F, Tyrrell J, Beaumont RN, Ji Y, Jones SE, Hattersley AT, et al. A genome-wide association study implicates multiple mechanisms influencing raised urinary albumin-creatinine ratio. *Hum Mol Genet* (2019) 28:4197–207. doi: 10.1093/hmg/ddz243

11. Ahluwalia TS, Schulz CA, Waage J, Skaaby T, Sandholm N, van Zuydam N, et al. A novel rare CUBN variant and three additional genes identified in europeans with and without diabetes: Results from an exome-wide association study of albuminuria. *Diabetologia* (2019) 62:292–305. doi: 10.1007/s00125-018-4783-z

12. Haas ME, Aragam KG, Emdin CA, Bick AG, International Consortium for Blood P, Hemani G, et al. Genetic association of albuminuria with cardiometabolic disease and blood pressure. *Am J Hum Genet* (2018) 103:461–73. doi: 10.1016/j.ajhg.2018.08.004

13. Teumer A, Li Y, Ghasemi S, Prins BP, Wuttke M, Hermle T, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun* (2019) 10:4130. doi: 10.1038/s41467-019-11576-0

14. Teumer A, Tin A, Sorice R, Gorski M, Yeo NC, Chu AY, et al. Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes* (2016) 65:803–17. doi: 10.2337/db15-1313

15. Charmet R, Duffy S, Keshavarzi S, Gyorgy B, Marre M, Rossing P, et al. Novel risk genes identified in a genome-wide association study for coronary artery disease in patients with type 1 diabetes. *Cardiovasc Diabetol* (2018) 17:61. doi: 10.1186/s12933-018-0705-0

16. Dahlstrom EH, Saksi J, Forsblom C, Uglebjerg N, Mars N, Thorn LM, et al. The low-expression variant of FABP4 is associated with cardiovascular disease in type 1 diabetes. *Diabetes* (2021) 70:2391–401. doi: 10.2337/db21-0056

17. Winther SA, Ollgaard JC, Tofte N, Tarnow L, Wang Z, Ahluwalia TS, et al. Utility of plasma concentration of trimethylamine n-oxide in predicting cardiovascular and renal complications in individuals with type 1 diabetes. *Diabetes Care* (2019) 42:1512–20. doi: 10.2337/dc19-0048

18. Lajer M, Jorsal A, Tarnow L, Parving HH, Rossing P. Plasma growth differentiation factor-15 independently predicts all-cause and cardiovascular mortality as well as deterioration of kidney function in type 1 diabetic patients with nephropathy. *Diabetes Care* (2010) 33:1567–72. doi: 10.2337/dc09-2174

19. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam study. *Eur J Epidemiol* (2020) 35:483–517. doi: 10.1007/s10654-020-00640-5

20. van Herpt TTW, Lemmers RFH, van Hoek M, Langendonk JG, Erdtsieck RJ, Bravenboer B, et al. Introduction of the DiaGene study: Clinical characteristics, pathophysiology and determinants of vascular complications of type 2 diabetes. *Diabetol Metab Syndr* (2017) 9:47. doi: 10.1186/s13098-017-0245-x

21. McKnight AJ, Patterson CC, Pettigrew KA, Savage DA, Kilner J, Murphy M, et al. A GREM1 gene variant associates with diabetic nephropathy. *J Am Soc Nephrol* (2010) 21:773–81. doi: 10.1681/ASN.2009070773

22. Hadjadj S, Pean F, Gallois Y, Passa P, Aubert R, Weekers L, et al. Different patterns of insulin resistance in relatives of type 1 diabetic patients with retinopathy or nephropathy: the genesis France-Belgium study. *Diabetes Care* (2004) 27:2661–8. doi: 10.2337/diacare.27.11.2661

23. Mansour Aly D, Dwivedi OP, Prasad RB, Karajamaki A, Hjort R, Thangam M, et al. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat Genet* (2021) 53:1534–42. doi: 10.1038/s41588-021-00948-2

24. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature* (2018) 562:203–9. doi: 10.1038/s41586-018-0579-z

25. Eastwood SV, Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK biobank. *PloS One* (2016) 11:e0162388. doi: 10.1371/journal.pone.0162388

26. Noordam R, Lall K, Smit RAJ, Laisk T, Estonian Biobank Research T, Metspalu A, et al. Stratification of type 2 diabetes by age of diagnosis in the UK biobank reveals subgroup-specific genetic associations and causal risk profiles. *Diabetes* (2021) 70:1816–25. doi: 10.2337/db20-0602

27. Common Metabolic Diseases Knowledge Portal. Available at: https://hugeamp.org/. Accessed 10 Jul 2020. 2020.

28. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, et al. Investigators, Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med* (2012) 367:20–9.

29. Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* (2015) 31:3555–7. doi: 10.1093/bioinformatics/btv402

30. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New creatinine- and cystatin c-based equations to estimate GFR without race. *N Engl J Med* (2021) 385:1737–49. doi: 10.1056/NEJMoa2102953

31. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med* (2002) 21:35–50. doi: 10.1002/sim.973

32. Gratten J, Visscher PM. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med* (2016) 8:78. doi: 10.1186/s13073-016-0332-x

33. Beenken A, Barasch JM, Gharavi AG. Not all proteinuria is created equal. *J Clin Invest* (2020) 130:74–6. doi: 10.1172/JCI133250

34. Quinlan C. CUBN variants uncouple proteinuria from kidney function. *Nat Rev Nephrol* (2020) 16:135–6. doi: 10.1038/s41581-019-0242-4

35. Gan C, Zhou X, Chen D, Chi H, Qiu J, You H, et al. Novel pathogenic variants in CUBN uncouple proteinuria from renal function. *J Transl Med* (2022) 20:480. doi: 10.1186/s12967-022-03706-y

36. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* (2019) 51:957–72. doi: 10.1038/s41588-019-0407-x

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 16 March 2023
DOI 10.3389/fendo.2023.1115890

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Feng Jiang,
Fudan University, China
Lei Luo,
Beijing Youan Hospital, Capital Medical University, China

*CORRESPONDENCE
Jun-jie Zhang
zhangjunjielab@hotmail.com
Fei Xu
xufei8586@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 04 December 2022
ACCEPTED 02 March 2023
PUBLISHED 16 March 2023

CITATION
Zhang J-j, Shen Y, Chen X-y, Jiang M-l, Yuan F-h, Xie S-l, Zhang J and Xu F (2023) Integrative network-based analysis on multiple Gene Expression Omnibus datasets identifies novel immune molecular markers implicated in non-alcoholic steatohepatitis. *Front. Endocrinol.* 14:1115890. doi: 10.3389/fendo.2023.1115890

Integrative network-based analysis on multiple Gene Expression Omnibus datasets identifies novel immune molecular markers implicated in non-alcoholic steatohepatitis

Jun-jie Zhang[1*†], Yan Shen[2†], Xiao-yuan Chen[2], Man-lei Jiang[3], Feng-hua Yuan[1], Shui-lian Xie[1], Jie Zhang[3] and Fei Xu[3*]

[1]Center for Molecular Pathology, Department of Basic Medicine, Gannan Medical University, Ganzhou, China, [2]Department of Publication Health and Health Management, Gannan Medical University, Ganzhou, China, [3]Department of Hepatology, The Affiliated Fifth People's Hospital of Ganzhou, Gannan Medical University, Ganzhou, China

Introduction: Non-alcoholic steatohepatitis (NASH), an advanced subtype of non-alcoholic fatty liver disease (NAFLD), has becoming the most important aetiology for end-stage liver disease, such as cirrhosis and hepatocellular carcinoma. This study were designed to explore novel genes associated with NASH.

Methods: Here, five independent Gene Expression Omnibus (GEO) datasets were combined into a single cohort and analyzed using network biology approaches.

Results: 11 modules identified by weighted gene co-expression network analysis (WGCNA) showed significant association with the status of NASH. Further characterization of four gene modules of interest demonstrated that molecular pathology of NASH involves the upregulation of hub genes related to immune response, cholesterol and lipid metabolic process, extracellular matrix organization, and the downregulation of hub genes related to cellular amino acid catabolic, respectively. After DEGs enrichment analysis and module preservation analysis, the Turquoise module associated with immune response displayed a remarkably correlation with NASH status. Hub genes with high degree of connectivity in the module, including CD53, LCP1, LAPTM5, NCKAP1L, C3AR1, PLEK, FCER1G, HLA-DRA and SRGN were further verified in clinical samples and mouse model of NASH. Moreover, single-cell RNA-seq analysis showed that those key genes were expressed by distinct immune cells such as microphages, natural killer, dendritic, T and B cells. Finally, the potential transcription factors of Turquoise module were characterized, including NFKB1, STAT3, RFX5, ILF3, ELF1, SPI1, ETS1 and CEBPA, the expression of which increased with NASH progression.

10.3389/fendo.2023.1115890

Discussion: In conclusion, our integrative analysis will contribute to the understanding of NASH and may enable the development of potential biomarkers for NASH therapy.

KEYWORDS

non-alcoholic steatohepatitis, weighted gene co-expression network analysis, hub genes, immune response, transcription factors

Introduction

Non-alcoholic fatty liver disease (NAFLD) is likely to become the most common chronic liver disease, affecting about 25% in the adult population (1). It is characterized by excessive accumulation of hepatic triacylglycerol (TG) and encompasses a spectrum of liver pathologies ranging from isolated steatosis (non-alcoholic fatty liver, NAFL) to non-alcoholic steatohepatitis (NASH), a more severe form of fatty liver disease featured by lobular inflammatory infiltrates, hepatocyte ballooning and fibrosis (2). Up to 30% of the patients with NAFLD will process to NASH (3), which may eventually progress to cirrhosis, hepatocellular carcinoma (HCC) and liver failure (4). Moreover, NASH is considered the hepatic manifestation of metabolic syndrome, commonly alongside serious extrahepatic diseases, such as dyslipidemia, hypertension, obesity and type 2 diabetes mellitus (T2DM) (5, 6), and multiple pathogenic pathways are involved in NASH progression.

Previous studies have contributed greatly to our understanding of genetic and environmental risk factors in the pathogenesis of NAFLD. Genome-wide association studies (GWAS) have revealed genetic variants in several loci (*PNPLA3*, *TM6SF2*, *GCKR*, *MTARC1* and *HSD17B13*) that promote NAFLD risks in humans (7–11), which highlights the dysregulation of gene expression and/or function as an important players in the development and progression of NASH. Integrating multi-omics approaches including genomics, transcriptomics, proteomics and metabolomics have provided additional insights (12–15), which may not be elucidated by genomics analysis alone. In addition, previous bioinformatics analyses in cross-sectional studies have facilitated the exploration of potential biomarkers related to NAFLD/NASH (16–19). However, for complex disease trait, the comprehensive molecular characterization of NASH are still not entirely deciphered. As a consequence, no effective pharmacological therapies targeting NASH are presently available. Hence, further exploration into the molecular pathogenesis of NASH and diagnostic biomarkers are essential to build novel approaches for management of NASH.

Network biology approaches have proven effective for uncovering new perturbed pathways underlying molecular pathology (18, 20, 21). Contrary to traditional differential expression analysis methods based on gene expression profiling, network-based approaches investigate the correlation among changing genes from a systematic perspective. Weighted gene co-expression network analysis (WGCNA) has become a frequently used method for multigene analysis, which establishes gene sets (modules) from observed gene expression data using unsupervised hierarchical clustering. WGCNA is widely used for exploring the relationship between diverse gene sets and clinical features (22, 23), providing insights into functions of co-expression gene modules and detecting hub genes related to the clinical characteristics of various diseases (24, 25).

In the present work, we aimed to identify deregulated modules, hub genes and transcription factors (TFs) associated with NASH by integrating transcriptomic data with biological network analysis between normal liver tissues and NASH tissues. We obtained five liver transcriptome datasets from the Gene Expression Omnibus (GEO) database (26). We first generated MergeCohort by merging five pre-processed datasets. Based on the combining expression matrix, differentially expressed gene (DEG) analysis was performed to identify genes associated with NASH. After that, through integrative analyses of co-expression gene network, functional annotation, TF-target regulatory network and validation analysis, we detected several promising candidate biomarkers for NASH. Our integrative study provides a comprehensive view on the molecular processes of NASH and may discover potential therapeutic target for NASH treatment.

Methods

Data collection

We obtained the expressing profiles of mRNA of NASH and normal control from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/) (26). We searched the microarray and next-generation sequencing (NGS) studies with the keywords: “Fatty liver”, “Non-alcoholic”, “Gene expression”, “Homo sapiens”, “Microarray” and “RNA sequencing”. Datasets were selected based on the following criterial (1): Containing at least 10 total samples (2); Samples must Contain at least five patients in both NASH group and healthy control group (3); Raw data or gene expression profiles were available in GEO (4). Pathways related to lipid metabolism, inflammation and fibrosis were significantly (normalized enrichment score (NES) more than 1.0 and a false discovery rate (FDR) below 0.25) enriched between the two groups in the gene set enrichment analysis (GSEA) (Supplementary Tables S2, S3), which was carried out with the Java GSEA (version 3.0) (27) platform with the ‘Signal2Noise’ metric to create a ranked list and a ‘gene set’ permutation type. The flowchart was shown in Figure 1.

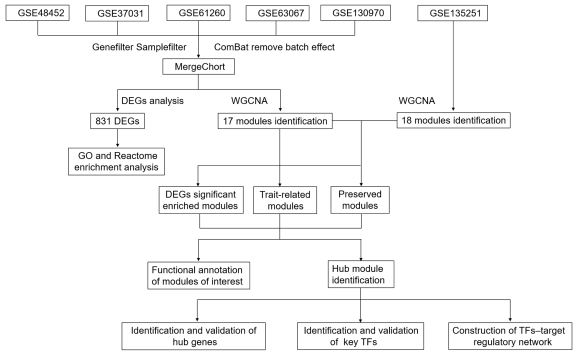


FIGURE 1
Flowchart.

Data processing

For each dataset, we download raw expression data and pre-processed using standard approaches. Specially, gene chip datasets were normalized by the robust multi-average (RMA) method with oligo/Bioconductor (28). For RNA-seq datasets, reads count information were generated by StringTie using a Python script (prepDE.py) and raw counts were normalized across samples following TMM method in edgeR package. After filtering low abundance expression genes and outlier samples, we applied the ComBat (version 3.20.0) method in the sva R package to remove the batch effects (29) from five datasets (GSE48452, GSE37031, GSE61260, GSE63067 and GSE130970) and combined these five datasets into a single cohort (MergeCohort), which contains 67 normal and 97 NASH tissue samples. Subsequently, the expression matrix of MergeCohort was used for differentially expressed genes (DEGs) identification between NASH and healthy control samples. It is worth noticing that we applied Wilcoxon's rank-sum test to assess the differential expression, the corrected threshold was p less than 0.05, and the absolute difference of means more than 0.3. Gene ontology (GO) and Reactome enrichment analyses were performed for DEGs using hypergeometric test, which is conducted by the python package gseapy (version 0.9.16; https://github.com/zqfang/gseapy), all gene sets of GO term and Reactome pathway were obtained from database source of Enrichr (30). Only GO terms or Reactome pathways were considered as significantly enriched by using the criterion with a corresponding p value less than 0.05.

Weight gene co-expression network construction, module detection and preservation analysis of the co-expression modules

5,000 transcripts with maximal variability across all patients (n = 164) based on the median absolute deviation in the MergeCohort were kept for WGCNA and tested by the WGCNA R package (22). In our work, the power threshold of 5 was selected to calculate biweight midcorrelations and weighted adjacency matrix, the soft thresholding parameter was defined using the scale-free topology fit model. We identified the gene modules based on the 'hybrid' method and parameters deepSplit = 4, mergeCutHeight = 0.15 and minModuleSize = 50. Modules are identified as branches in the dendrogram with Dynamic Tree Cut algorithm (22). Subsequently, we assessed the relevance of a module eigengene (ME) to the disease status using the Pearson correlation. An intramodular connectivity (K_{in}) was defined to measure for each gene on the base of its correlation with the remaining genes in a given module. Genes with highest K_{in} are identified as hub genes. Cytoscape version 3.8.2 was used for visualization. In order to understand the extent of module preservation in MergeCohort, a publicly available expression profiling of high throughput RNA sequencing dataset GSE135251 including 10 controls, 51 NAFL and 155 NASH was used, processed as described above. Module preservation analysis was carried out by using Module preservation function in WGCNA package introduced by Langfelder et al. (31) and described in detail in Oldham et al. (32). Moreover, to investigate the module similarity among different cohorts, we applied hypergeometric test to evaluate whether the genes from each MergeCohort module significantly overlapped with the genes from each of GSE135251 module. The overlap was regarded as significant when p value below 0.05.

Functional annotation of the modules

In order to determine the functional significance of the identified modules, we firstly performed GO and KEGG pathway enrichment analysis for the gene lists of each module of co-expression network on the basis of Enrichr (30) as described above. Moreover, we carried out disease enrichment analysis for the gene lists of each module by using DisGeNet (33). The statistical significance threshold level for all disease terms was p value less than 0.05 (Benjamini-Hochberg corrected for multiple comparisons) and we presented top 20 for each disease-associated module. Additionally, to obtain regulatory information of transcription factors (TFs) and target genes, Transcriptional Regulatory Relationships Unraveled by Sentence based Text mining (TRRUST) v2 database (https://www.grnpedia.org/trrust/) (34) were supplied for Enrichr (30), conducted by the python package gseapy (version 0.9.16; https://github.com/zqfang/gseapy). In addition, ChIP-X Enrichment Analysis 3 (ChEA3) database (https://maayanlab.cloud/chea3/) (35) was adopted to further validate the significantly enriched transcription factors over module genes. After obtaining TF–target regulatory relationships, a TF-target network, which contained TFs regulating Turquoise modules' genes, was reconstructed.

Single cell RNA-sequencing analyses

We investigated the expression patterns of top 25 hub genes in Turquoise module using scRNA-seq analyses of human liver tissues from public scRNA-seq data (GSE136103) (36). In our study, only four samples including two healthy liver tissue samples (GSM4041156 and GSM4041159) and two NAFLD liver tissue samples (GSM4041162 and GSM4041163) were analyzed with

Seurat package (version 3.1.5) (37). First, 2000 highly variable genes (n = 2,000) were identified using the R package *SCTransfom* (version 0.2.1). Subsequently, principal component analysis was performed, and the appropriate principal components (PCs) for dimensionality reduction were decided using the *JackStraw* function. Clusters were identified with the Seurat function *FindClusters* with the resolution set at 0.4. This method resulted in 18 clusters, which were visualized by Uniform Manifold Approximation and Projection (UMAP) analysis. Clusters were then annotated by using the expression of known genes. We annotated cell types based on cell markers and the R package SingleR (36, 38).

Results

Information of included GEO datasets

According to the previously established inclusion criteria, GSE48452, GSE37031, GSE61260, GSE63067 and GSE130970 were included in this study. There are 104 NASH patients and 70 controls in these five datasets. After outlier removal, 97 NASH patients and 67 controls were retained in the following analysis. The detail information of the five datasets was shown in Supplementary Table S1. In order to eliminate the bath effect from different platforms and batches, we used the combat function to eliminate the batch effect from five datasets. A total of 12579 genes were detected by merging different platforms. Before removing the batch effect, samples were clusters in batch according to the top two principal components (PCs) of the expression values before normalization (Figure S1A). In contrast, when the samples from five platforms were merged, the overall expression in the samples was uniformly distributed based on principal component analysis, suggesting that the batch effect caused by different platforms that had effect on the estimation of molecular biological differences was successfully corrected (Figure S1B). In addition, we used dataset GSE135251 as the validation dataset in this study.

Identification of DEGs in the NASH patients

Principle component analysis plot of the gene expression matrix of five combined dataset (MergeCohort) distinguished between NASH and control group is shown in Figure 2A. Total of 831 DEGs (Benjamin-Hochberg adjusted p value < 0.05, absolute difference of mean > 0.3) among control and NASH in MergeCohort were identified, consisting of 600 upregulated and 231 downregulated DEGs (Figure 2B; Supplementary Table S4).

Function and pathway enrichment analysis of DEGs

In the present study, we performed GO and Reactome pathway enrichment analysis to determine the potential functions of 831 DEGs in the pathogenesis of NASH. The biological process analysis (Figure 2C; Supplementary Table S5) revealed that in the NASH, these genes were associated with multiple immunity-related pathways, such as the cytokine-mediated signaling pathway, cellular response to cytokine stimulus and neutrophil activation involved in immune response. Several ECM-related pathways were also enriched such as extracellular matrix organization and extracellular structure

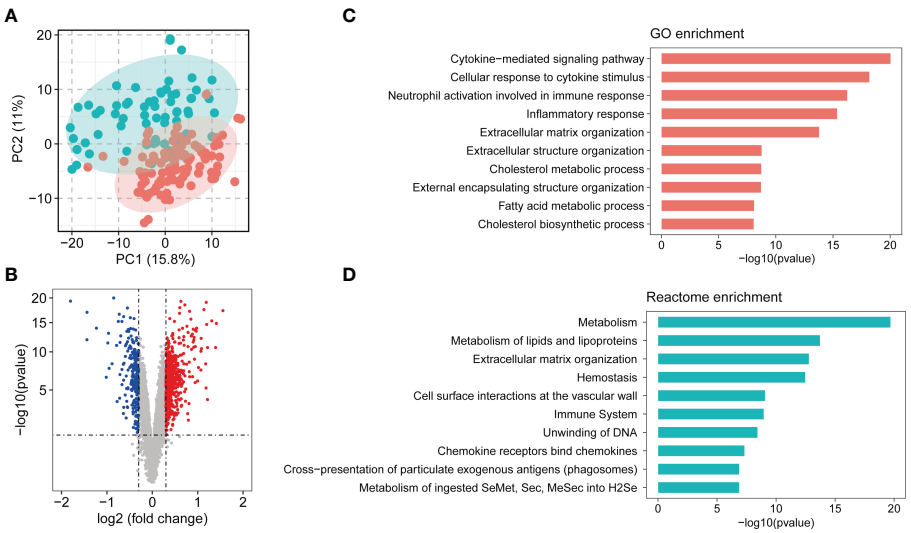


FIGURE 2
Overview of combining gene expression profiles in healthy controls and nonalcoholic steatohepatitis (NASH) patients. **(A)** Principle component plot of samples based on top 500 most variable gene expression from combining gene expression profiles (MergeCohort). NASH patients are marked in red; healthy controls are marked in green. **(B)** Volcano plot of differentially expressed genes (DEGs) between NASH patients and healthy controls. DEGs are listed in Supplemental Table S4. 600 genes upregulated and 200 genes downregulated are shown in red and blue, respectively. **(C)** Top 10 enriched biological functions of DEGs determined by Gene Ontology (GO) enrichment analysis. **(D)** Top 10 enriched Reactome pathways of DEGs determined by Reactome pathway enrichment analysis.

organization. Moreover, metabolic process, such as cholesterol metabolic process, fatty acid metabolic process, cholesterol biosynthetic process and other biological process (Supplementary Table S5) were also identified. Reactome pathway analysis was performed to investigate the pathway based on the DEGs (Supplementary Table S6). The top 10 pathways are shown in Figure 2D. Among them, metabolism, metabolism of lipids and lipoproteins, extracellular matrix organization, immune system, chemokine receptors bind chemokines were significantly enriched. Therefore, the outcomes above suggested that metabolism, ECM-related pathways and immunity-related pathways play an important role in development and procession of NASH.

WGCNA and identification of module associated with NASH disease status

To capture discrete groups of co-expression genes correlated with NASH status and to integrate the identified expression divergences into a higher system level context, a co-expression network analysis (WGCNA) was conducted based on the top 5000 median absolute deviation (MAD) genes from the MergeCohort. Keep to the scale-free topology criterion, β=5 was considered in this study (Figure 3A). According to dynamic tree cut, the hierarchical clustering dendrogram resulted in 17 different gene modules, as displayed in Figure 3B. 909 genes failed to fit within a distinct group and were assigned to the Grey module which was neglected in the present study. The size of modules ranged from 86 (Grey60 module) to 734 (Turquoise module) (Figure 3C). DEGs enrichment in each module was shown in Figure 3D, in which upregulated genes was mostly significantly enriched in Turquoise ($n = 233$, $p = 1.93 \times 10^{-44}$), and followed by Cyan ($n = 54$, $p = 1.24 \times 10^{-15}$), Grey60 ($n = 40$, $p = 2.05 \times 10^{-13}$), Tan ($n = 48$, $p = 1.59 \times 10^{-9}$) and Magenta ($n = 47$, $p = 2.77 \times 10^{-4}$), downregulated genes was significantly enriched in Black ($n = 107$, $p = 9.25 \times 10^{-86}$) and Brown module ($n = 68$, $p = 1.07 \times 10^{-24}$). To investigate which co-expression modules are associated with NASH status, we then correlated the expression of eigengenes (genes representing the expression profile of each module) with NASH status. The relationship between all the modules and the NASH status are displayed in a correlation heatmap, in which Y-axis corresponds to groups of genes (modules) and the X-axis represents the NASH status (Figure 3E). Of the 17 co-expression modules, 11 WGCNA modules to be correlated with NASH status at a Pearson correlation ($p < 1.47 \times 10^{-3}$), which is determined based on Bonferroni correction. Among them, nine modules (Cyan, Grey60, Turquoise, Magenta, Purple, Lightcyan, Tan, Midnightblue and Blue) were positively correlated with NASH disease status, two modules (Black and Brown) were negatively associated with NASH disease status (Figure 3E).

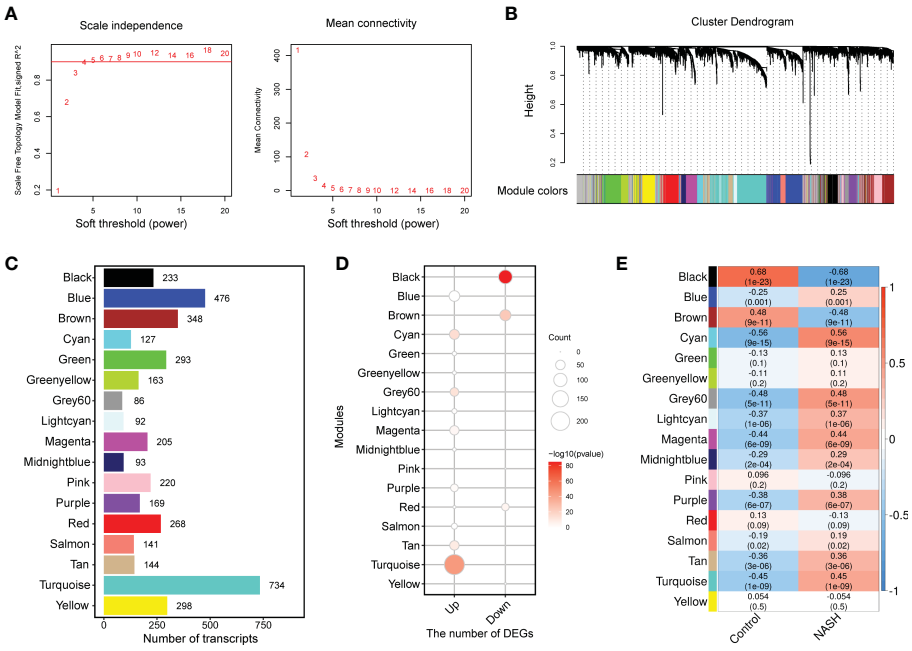


FIGURE 3
WGCNA network and module identification. **(A)** Soft-thresholding calculation of MergeCohort. The left panel displays the scale-free fit index versus soft-thresholding power. The right panel shows the mean connectivity versus soft-thresholding power. Power 5 was selected, for which the fit index curve flattens out upon reaching a high value (> 0.9). **(B)** The Cluster dendrogram of co-expression network modules from WGCNA depending on a dissimilarity measure (1-TOM). The leaves in the tree represent genes and the colors in the horizontal bar indicate co-expression module determined by the dynamic tree cut algorithm. **(C)** Number of genes in each module. **(D)** Enrichment of upregulated and downregulated DEGs in each module. **(E)** Heatmap showing the association between module eigengenes (rows) and NASH disease status (column). Associated *p* values were computed using the cor.test R function. The color scale in the heat map represents the magnitude of the Pearson correlation coefficients. Number in each cell contained corresponding correlation coefficient and *p* value (in brackets). WGCNA, weighted gene correlation network analysis; TOM, topological overlap matrix.

10.3389/fendo.2023.1115890

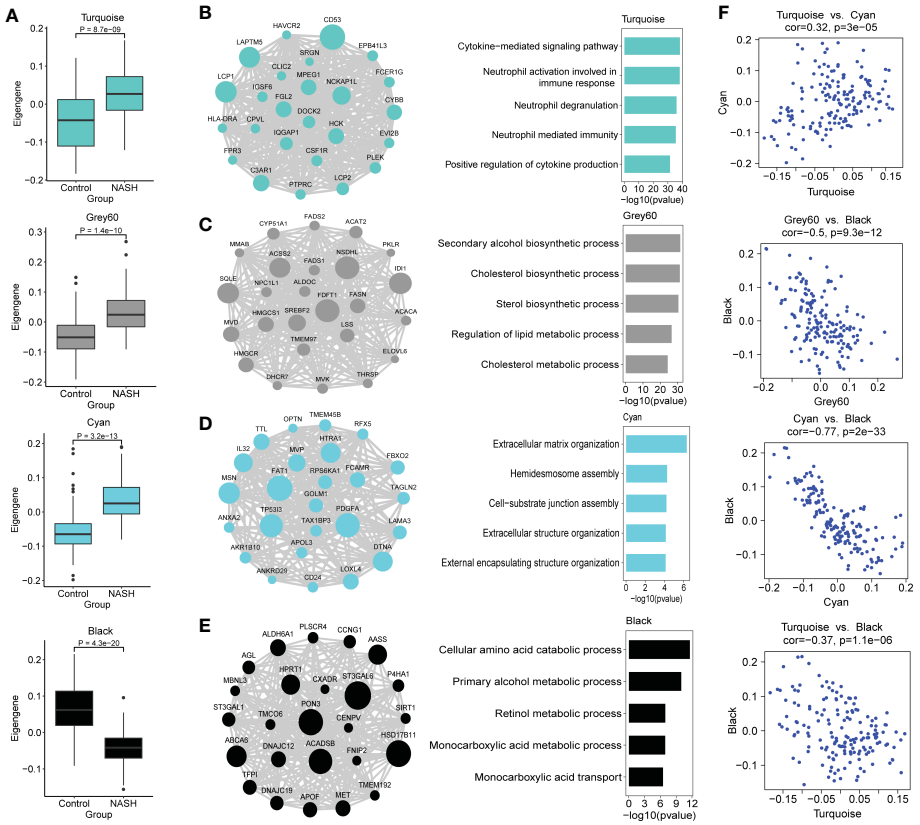


FIGURE 4
Functional characterization of co-expression modules of interest identified by WGCNA. **(A)** Box and Whisker plots representing the expression of module eigengenes Turquoise, Grey60, Cyan, Black between NASH (n = 97) and healthy control (n = 67) samples. Data are presented as median with first and third quartiles as the box edges. Differences between group were estimated by Student's t test. **(B–E)** The network of hub genes (module genes within the top 25 genes with the highest intromodular connectivity values (kWithin)) (left panel) and top GO terms (right panel) of the modules Turquoise **(B)**, Grey60 **(C)**, Cyan **(D)** and Black **(E)** are shown. In the network diagrams, node sizes correspond to kWithin in the module. For the bars plot, the bars in the GO enrichment results represent the -$\log_{10}$(pvalue). **(F)** Scatterplots of module eigengenes show positive correlation between Turquoise and Cyan, and negative correlation between Grey60, Cyan, Turquoise and Black, respectively.

Functional characterization of co-expression modules of interest

Because we were more concerned about the modules whose expression was different between NASH and control group, we compared the eigengenes from NASH samples to the expression of control in every module, and these results were used to further assess whether the modules were associated with NASH status. Modules Cyan, Grey60 and Turquoise exhibited an upregulation of the eigengenes in NASH, whereas module black showed lower expression in NASH (Figure 4A). In order to investigate whether the co-expression modules cover the information associated with validated networks, the existing data on protein-protein interactions from the STRING database was used to test the biological characteristics of the detected modules in this study. All the modules showed significant enrichment in interactions (p < 0.01), therefore indicating that the modules detected in the present work are biologically relevant (Supplementary Table S7). In addition, the NASH status positively correlated modules showed much higher average node degree (AND), particularly module Turquoise (AND = 22.4).

We then conducted GO and KEGG pathway enrichment of the NASH-associated modules to further investigate the gene functions by Enrichr. Top biological process and KEGG pathway in each module are shown in Table 1. Turquoise module was upregulated in NASH patients, contained hub genes related to immune response (*CD53*, *LAPTM5*, *LCP1*, *NCKAP1L*, *C3AR1* and *FGL2*) (Figure 4B), and enriched for GO categories to cytokine-mediated signaling pathway, neutrophil activation involved in immune response and neutrophil degranulation (Figure 4B). Grey60 module with hub genes such as *FDFT1*, *NSDHL*, *IDI1*, *SQLE*, *ACSS2*, *SREBF2*, *HMGCR*, *FASN*, *LSS*, *ACAT2*, *FADS1*, *FADS2* and *ELOVL6* was upregulated in NASH (Figure 4C), which were mainly participating in cholesterol and lipid metabolic process (Figure 4C). The majority of the GO terms enriched in module Cyan were primarily related to extracellular matrix organization and extracellular structure organization (Figure 4D), including hub genes related to fibrosis (*PDGFA*, *LOXL4*, *MSN*, *LAMA3* and *AKR1B10*) (Figure 4D). However, the majority of the GO terms enrich in Black module were related to cellular amino acid catabolic and primary alcohol metabolic process (*ACADSB*, *AASS* and *ALDH6A1*) (Figure 4E). The complete annotation for each module can be found in Supplementary Tables S8, S9.

10.3389/fendo.2023.1115890

TABLE 1 Top GO and pathway enrichment in each module.

Module	Category	Term	*P*-value	FDR
Black	GOTERM_BP	Cellular amino acid catabolic process	2.37×10^{-12}	3.95×10^{-09}
Blue	GOTERM_BP	Extracellular matrix organization	6.18×10^{-37}	1.57×10^{-33}
Brown	GOTERM_BP	Cellular amino acid catabolic process	5.27×10^{-09}	1.06×10^{-05}
Cyan	GOTERM_BP	Extracellular matrix organization	4.82×10^{-07}	5.88×10^{-04}
Grey60	GOTERM_BP	Secondary alcohol biosynthetic process	2.39×10^{-32}	1.54×10^{-29}
Lightcyan	GOTERM_BP	T cell activation	4.17×10^{-13}	3.44×10^{-10}
Magenta	GOTERM_BP	DNA metabolic process	2.69×10^{-45}	3.48×10^{-42}
Midnightblue	GOTERM_BP	IRE1-mediated unfolded protein response	7.75×10^{-16}	6.39×10^{-13}
Purple	GOTERM_BP	Regulation of glycogen metabolic process	2.31×10^{-06}	3.06×10^{-03}
Tan	GOTERM_BP	Neutrophil degranulation	8.86×10^{-16}	7.05×10^{-13}
Turquoise	GOTERM_BP	Cytokine-mediated signaling pathway	3.47×10^{-39}	8.55×10^{-36}
Black	KEGG_PATHWAY	Metabolism of xenobiotics by cytochrome P450	2.94×10^{-05}	3.85×10^{-03}
Blue	KEGG_PATHWAY	ECM-receptor interaction	3.54×10^{-19}	8.42×10^{-17}
Brown	KEGG_PATHWAY	Glycine, serine and threonine metabolism	2.24×10^{-08}	5.78×10^{-06}
Cyan	KEGG_PATHWAY	Mitophagy	9.22×10^{-04}	0.11
Grey60	KEGG_PATHWAY	Steroid biosynthesis	1.01×10^{-14}	8.99×10^{-13}
Lightcyan	KEGG_PATHWAY	Primary immunodeficiency	1.14×10^{-17}	1.39×10^{-15}
Magenta	KEGG_PATHWAY	DNA replication	5.62×10^{-27}	7.20×10^{-25}
Midnightblue	KEGG_PATHWAY	Protein processing in endoplasmic reticulum	3.05×10^{-21}	2.75×10^{-19}
Purple	KEGG_PATHWAY	Axon guidance	1.62×10^{-04}	3.11×10^{-02}
Tan	KEGG_PATHWAY	Cytokine-cytokine receptor interaction	4.47×10^{-12}	8.81×10^{-10}
Turquoise	KEGG_PATHWAY	Osteoclast differentiation	2.48×10^{-18}	6.45×10^{-16}

We next explored the relationship of eigengenes among the annotated modules. Upregulated immune Turquoise module was positively correlated with Cyan module related to fibrosis (r = 0.32, $p = 3.0 \times 10^{-5}$) (Figure 4F), suggesting that Turquoise module related to immune response that drives fibrosis in NASH, which confirmed the results of previous studies (20). Interestingly, Cyan, Grey60 and Turquoise modules was negatively correlated with Black module that is enriched in amino acid metabolic processes (Figure 4F). The high negatively correlation (r = -0.77, $p = 2.0 \times 10^{-33}$) between the upregulated fibrosis module Cyan and downregulated Black module that is enriched in metabolic processes (Figure 4F), which indicated that perturbations in amino acid metabolism are likely involved in NASH pathogenesis (39, 40).

Module preservation analysis indicates the presence of NASH-associated co-expression module function in immune response

To find out whether the identified modules were common in another dataset, we examined the module preservation statistics between the MergeCohort and one recently published large NASH datatset GSE135251 (13). In particular, we assumed co-expression modules of MergeCohort as reference dataset and the co-expression modules of GSE135251 as test dataset. We utilized the principle described in (22). The score of Zsummary more than 10 represents strongly preserved module, less than 2 denotes non-preserved module while the value between 2 and 10 implies moderately preserved module. We plotted the scatterplot of Zsummary scores against the sizes of MergeCohort modules (Figure 5A). All modules have a Zsummary statics greater than 2, suggesting that all modules were preserved in GSE135251. The lowest preservation is the Red module (Zsummary = 6.37). Particularly, MergeCohort module Turquoise (MergeCohort_Turquoise) exhibited Zsummary preservation score (Zsummary = 42.68) higher than 40. To provide a more intuitive picture of the preservation of each co-expression module identified, we evaluated module overlaps of MergeCohort and GSE135251 (Figure 5B), we found that MergeCohort_Turquoise show the most significantly overlapping with GSE135251 module Turquoise (GSE135251_Turquoise). Moreover, we discovered a highly positively correlation between the intromodular connectivity of 289 genes overlapped in MergeCohort_Turquoise and GSE135251_Turquoise (Spearman's

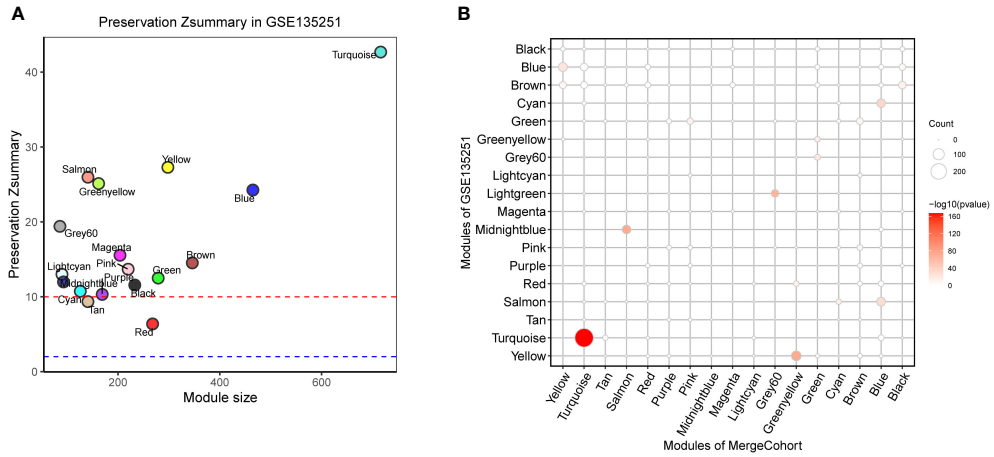


FIGURE 5
Module preservation of MergeCohort in GSE135251 dataset. **(A)** Preservation Zsummary statistics of MergeCohort in GSE135251 dataset. Each point represents a module. Point color reflects the module color as used in Figures 3B–E of MergeCohort. Points are also labeled by the name of the module. The dashed blue and red lines indicate the rough thresholds for week (Z = 2) and strong (Z = 10) evidence of module preservation. **(B)** Overlaps of MergeCohort and GSE135251 modules. Each axis is labelled by the corresponding module name. The size of each dot represents the number of overlapping genes in the intersection of corresponding MergeCohort and GSE135251 modules while the color implies -$\log_{10}$ of the hypergeometric enrichment *p* value.

correlation = 0.62, $p = 1.3 \times 10^{-9}$) (Figures 6A, B), which indicated those two modules have similar co-expression pattern.

To comprehensively evaluate the biological functions related to MergeCohort_Turquoise and GSE135251_Turquoise, we next calculated the statistical significance of enrichment of genes with the association in disease-related gene sets from the DisGeNET database (33) and KEGG pathway gene sets. We observed that genes in MergeCohort_Turquoise and GSE135251_Turquoise were significantly enriched by liver disease-related gene sets (liver cirrhosis) and multiple immune disease-related gene sets (autoimmune disease, immunosuppression and inflammatory bowel disease) (Figure 6C; Supplementary Tables S10, S11). Interestingly, these two modules were also significantly enriched in atherosclerosis and arteriosclerosis. Notably, we observed that genes in MergeCohort_Turquoise, which shows the highest module similarity with GSE135251_Turquoise (289 out of 734; hypergeometric test *p* value = 5.33×10^{-168}) (Figure 6A) are both significant enriched in phagosome, osteoclast differentiation, cell adhesion molecules, antigen processing and presentation, B cell receptor signaling pathway (Figure 6D). In addition, the MergeCohort_Turquoise was upregulated in NASH and is also the third most significant module, and showed the greater number

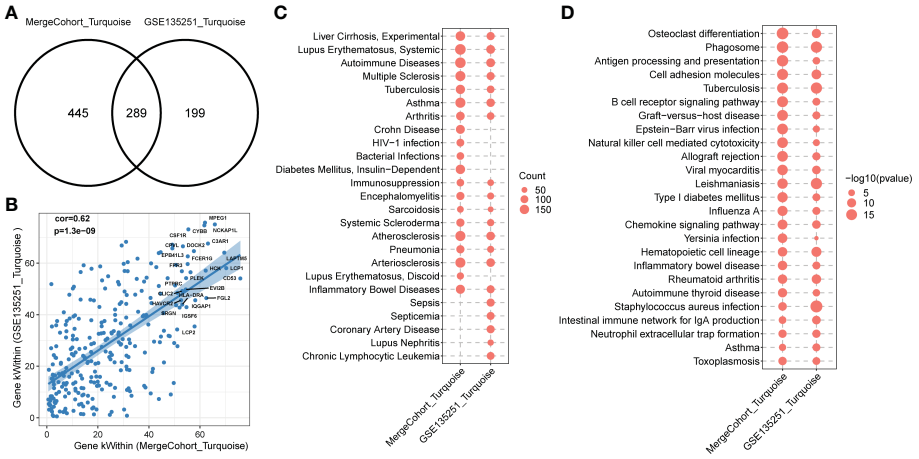


FIGURE 6
Functional enrichment of MergeCohort_Turquoise and GSE135251_Turquoise module. **(A)** Venn diagram displays number of genes overlapped between MergeCohort_Turquoise and GSE135251_Turquoise module. **(B)** Spearman's correlation between the kWithin of common genes (*n* = 289) overlapped between each module. Top 25 hub genes with the highest kWithin from MergeCohort_Turquoise module are shown. **(C)** Dot-plot heatmap shows top 20 significantly enriched disease by genes in each module. The size of each dot represent the gene counts enriched in each disease term. **(D)** Dot-plot heatmap shows top 20 significantly enriched KEGG pathways by genes in each module. The size of each dot represents the -$\log_{10}$ of *p* value for each KEGG pathway term.

of statistically differential expressed genes, with 233 of the 734 genes being upregulated (fold change > 1.2; $p < 0.05$) and none significantly downregulated (Figure 3D). Considering all these results, we will choose the co-expression Turquoise module from MergeCohort for further analysis.

Validation of hub genes in Turquoise module

Hub genes were upregulated in the liver from NASH patients. Focusing on the MergeCohort_Turquoise module, we firstly explored the top 25 hub genes including *CD53*, *LCP1*, *LAPTM5*, *NCKAP1L*, *C3AR1*, *PLEK*, *FCER1G*, *HLA-DRA* and *SRGN* that had a high intramodular connectivity (K.in). The expression level of those core genes were all upregulated in four cohorts (GSE130970, GSE48452, GSE61260 and GSE63067) involved in this study Figure 7A, suggesting that these hub genes may play fundamental role in NASH development. The PPI network of these 25 hub genes was showed in Figure 7B.

Hub genes were positively correlated with clinical characteristics. We further investigated the relationship between the changes in expression of these 25 hub genes and the histological phenotype in GSE130970 (Figure 7C). Our results demonstrated that each of the 25 key genes were positively correlated with the NAFLD activity score, and *FPR3* has the highest correlation (r = 0.53, $p = 1.49 \times 10^{-4}$). *LCP1* gene was the most associated gene with steatosis grade (r = 0.46, $p = 1.16 \times 10^{-3}$) and the lobular inflammation grade (r = 0.32, $p = 3.06 \times 10^{-2}$). Moreover, *FPR3* associated most with the cytological ballooning grade (r = 0.53, $p = 1.82 \times 10^{-4}$). *SRGN* was the most relevant gene with the fibrosis stage (r = 0.35, $p = 1.84 \times 10^{-2}$). Additionally, *C3AR1* showed significant correlation with all the clinical parameters, especially higher correlation with the cytological ballooning grade (r = 0.51, $p = 2.94 \times 10^{-4}$).

Hub genes were upregulated in the liver from the choline deficient L-amino acid defined high fat diet (CDAHFD) model of NASH in mouse. Furthermore, to explore the significance of the hub genes in mouse, we mined public available microarray data (GSE120977) (41) to validate the mRNA levels of the abovementioned genes, except *Hla-dra*, *Clic2* and *Fpr3* gene which was lacking in the dataset. Intriguingly, several of the hub genes displayed either a significant or a trending higher expression in mouse individuals fed with CDAHFD diets at 12 weeks compared with the controls. For instance, 14 genes, namely *Cd53*,

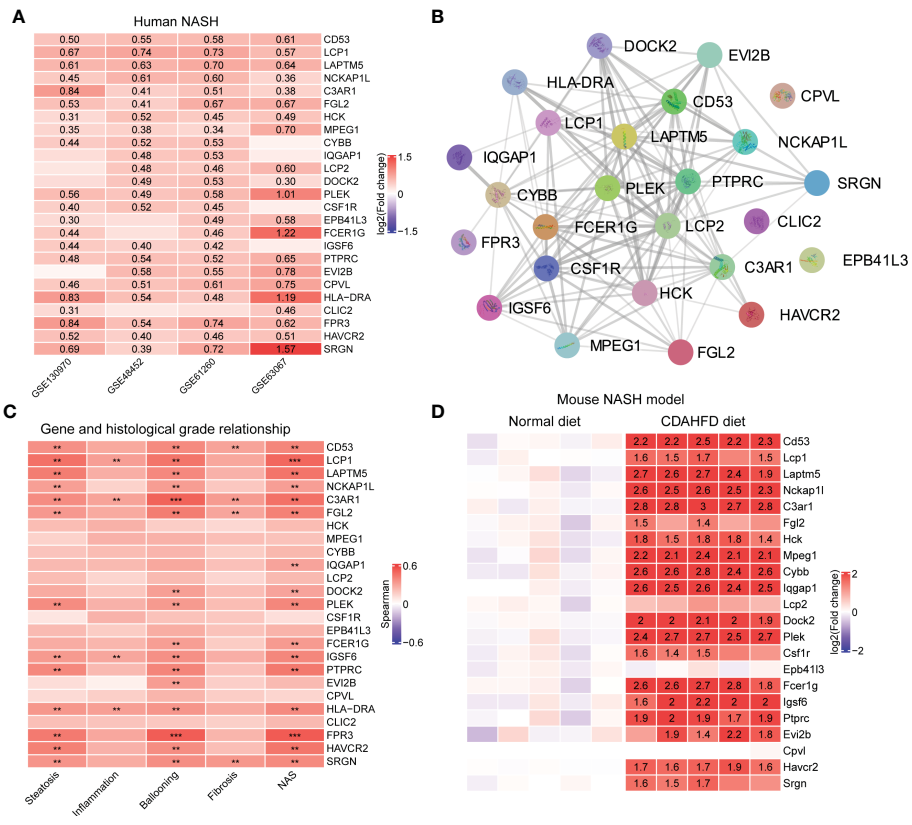


FIGURE 7
Validation of hub genes in MergeCohort_Turquoise module. **(A)** Heatmap shows the expression patterns of top 25 hub genes in human liver tissues according to four datasets (GSE130970, GSE48452, GSE61260 and GSE63067). The numbers in heatmap represent $\log_2$ value of fold change between NASH patients and healthy controls. **(B)** The protein-protein interactions among top 25 hub genes were retrieved by the STRING database. **(C)** Heatmap shows the Person correlation coefficients of top 25 hub genes and clinical parameters of NAFLD according to GSE130970 dataset. *p* values are overlaid on the heatmap (**$p < 0.01$ and ***$p < 0.001$). **(D)** Heatmap shows the expression patterns of top 25 hub genes in mouse liver tissue according to GSE120977 dataset. The numbers in heatmap represent $\log_2$ value of fold change between the CDAHFD and chow diet control group. CDAHFD, choline deficient L-amino acid defined high fat diet.

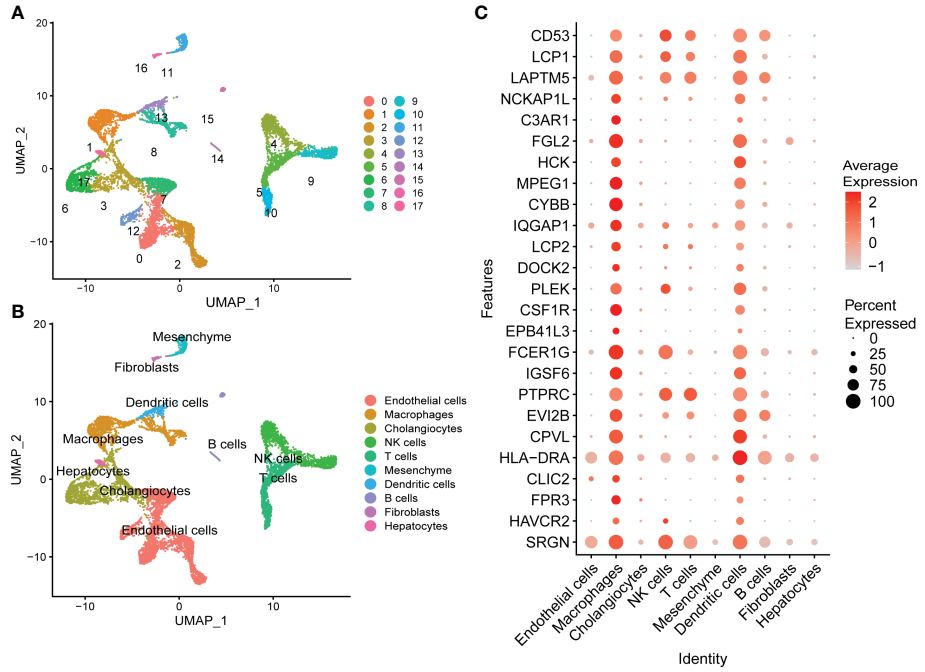


FIGURE 8
Assessment of the expression patterns of hub genes in MergeCohort_Turquoise module in different types of cells using publicly available healthy and cirrhotic scRNA-seq from dataset GSE136103. **(A)** UMAP visualization of different cell clusters from healthy (n = 2) and cirrhotic (n = 2) human livers. **(B)** UMAP visualization of cell types from healthy (n = 2) and cirrhotic (n = 2) human livers. Cells were annotated as endothelial cells, macrophages, cholangiocytes, NK cells, T cells, mesenchyme, dendritic cells, B cells, fibroblasts, and hepatocytes based on the expression of lineage markers. **(C)** Dot plot shows the expression patterns of top 25 hub genes in different types of liver cells. Size of the dot indicates proportion of the cell population that expresses each gene. Color represents level of expression. UMAP, uniform manifold approximation and projection.

Laptm5, *Nckap1l*, *C3ar1*, *Hck*, *Mpeg1*, *Cybb*, *Iqgap1*, *Dock2*, *Plek*, *Fcer1g*, *Igsf6*, *Ptprc* and *Havcr2*, which were strongly upregulated in mouse fed with CDAHFD chow (Figure 7D), supporting the notion that these hub genes were also activated during progression of mouse NASH model.

Identification of cell clusters contributions to the NASH-associated Turquoise module integrating single-cell RNA-seq analysis

To investigate how potential hub genes identified in MergeCohort_Turquoise module change within specific cell populations during NASH progression, we carried out an integrated scRNA-seq analysis using publicly available scRNA-seq data from healthy and cirrhotic liver samples. Clustering revealed 17 populations of cells comprising 10 distinct cell types (Figures 8A, B; Supplementary Figure S2). We identified Endothelial cells, macrophages, cholangiocytes, NK cells, T cells, mesenchyme, dendritic cells, B cells, fibroblasts, and hepatocytes within the scRNA-seq data based on the expression of lineage specific markers as annotated with integration of discoveries from human liver cell atlas and the annotation analysis with SingleR. The expression patterns of the top 25 genes in the MergeCohort_Turquoise module were analyzed by scRNA-seq analyses of liver tissues. Those key genes in MergeCohort_Turquoise module including *CD53*, *LCP1*, *LAPTM5*, *PTPRC* and *SRGN* expressed by distinct immune cells such as microphages, NK cells, T cells, dendritic cells and B cells, and most of them, namely *FGL2*, *HCK*, *MPEG1*, *CYBB*, *CSF1R*, *IGSF6*, *CPVL* and *HLA-DRA* were mainly expressed by macrophages, dendritic cells (Figure 8C; Supplementary Figure S3), which indicated that the macrophages and dendritic cells play an important role in the pathogenesis of NASH.

Identification of TFs that regulate the Turquoise modules

The results of the analysis above showed that hub genes in MergeCohort_Turquoise module were enriched in immunity. Because co-expressed genes tend to be co-regulated by the common transcription factors (TFs), we further conducted TFs enrichment analysis (hypergeometric test) using the genes from the MergeCohort_Turquoise and GSE135251_Turquoise modules to obtain key regulatory genes, based on TRRUST database (34). Our results indicated that *NFKB1*, *SPI1*, *RELA*, *CIITA*, *HIVEP2*, *SP1*, *RFXANK*, *RFXAP*, *RFX5*, *IRF1* are the top 10 most significantly enriched TFs in MergeCohort_Turquoise module (Figure 9A). Moreover, we adopted ChEA3 database (35) to validate the significantly enriched transcription factors over MergeCohort_Turquoise module genes. As a result, ChEA3 analysis identified 27 of the 33 significant TFs for

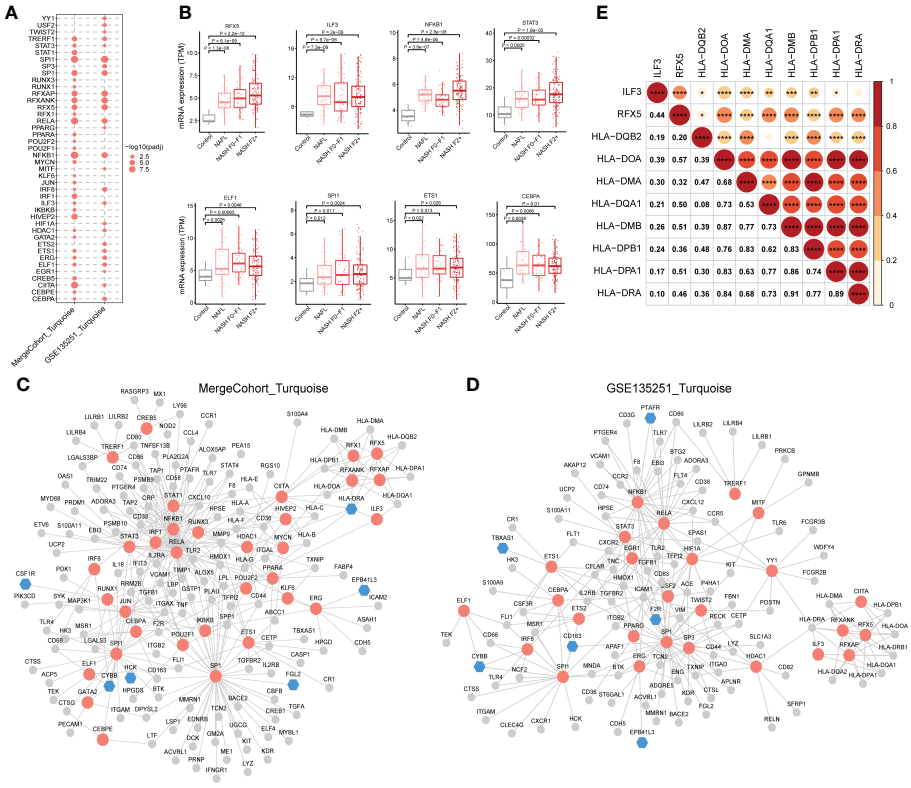


FIGURE 9
Regulatory relationship between enriched transcription factors and their target genes in NASH-associated module. **(A)** Dot-plot heatmap shows enriched transcription factors in MergeCohort_Turquoise and GSE135251_Turquoise module. The size of each dot represents the -$\log_{10}$ of adjusted *p* value for each transcription factor. **(B)** Boxplots shows mRNA hepatic expression of the enriched transcription factors including *RFX5*, *ILF3*, *NFKB1*, *STAT3*, *ELF1*, *SPI1*, *ETS1* and *CEBPA* according to GSE135251 dataset. The *p* value was calculated by Student's t test. **(C, D)** The regulatory networks between enriched transcription factors and associated target genes in MergeCohort_Turquoise **(C)** and GSE135251_Turquoise module **(D)**, respectively. Red color represents transcription factors, blue color represents target hub genes, grey color represents other target genes. **(E)** Pearson correlations for mRNA hepatic expression of transcription factors (*RFX5* and *ILF3*) and associated target genes (*HLA-DQB2*, *HLA-DOA*, *HLA-DMA*, *HLA-DQA1*, *HLA-DMB*, *HLA-DPB1*, *HLA-DPA1* and *HLA-DRA*) in GSE135251 dataset. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ and ****$p < 0.0001$.

MergeCohort_Turquoise module genes with TRRUST database, the other six TFs were part of their targets (Table S12). We also found that *NFKB1*, *SPI1*, *RELA*, *CIITA*, *SP1*, *RFXANK*, *RFXAP*, *RFX5*, *TRERF1*, *ELF1*, *STAT3*, *ERG*, *ETS1*, *ILF3*, *CEBPA*, *HDAC1* and *IRF8* are significantly enriched TFs in both MergeCohort_Turquoise and GSE135251_Turquoise module (Figure 9A). Furthermore, we observed significantly increased of hepatic expression of *RFX5*, *ILF3*, *NFKB1*, *STAT3*, *ELF1*, *SPI1*, *ETS1* and *CEBPA* in NAFL and NASH compared to the control group ($p < 0.05$) (Figure 9B).

Next, the regulatory networks were constructed for the enriched TFs and associated target genes in each of the modules (Figures 9C, D). We observed that *RFX5* and *ILF3*, an important transcriptional factor mainly expressed in the liver, upregulated from mild to advanced NASH, regulates the expression of genes involved in antigen processing and presentation of exogenous peptide antigen *via* MHC class II, including *HLA-DQB2*, *HLA-DOA*, *HLA-DMA*, *HLA-DQA1*, *HLA-DMB*, *HLA-DPB1*, *HLA-DPA1* and *HLA-DRA*. Notably, the gene expression of *RFX5* and *ILF3* positively correlated with MHCII gene expression (Figure 9E). We found 41 genes are regulated by the *NFKB1* transcription factor. As known, *NFKB1* regulates the expression of genes associated with cytokine-mediated signaling pathway (e.g., *TNF*, *CXCL10*, *MMP9* and *TGFB1*) and immune response (e.g., *CD74*, *CD58*, *CD80* and *CD86*) (Figure 9C). Moreover, *STAT3* regulates the expression of gene in Wound healing involved in inflammatory response, including *HMOX1*, *TIMP1*, *TGFB1* and *F2R*. Interestingly, *SPI1* regulated gene involved in immune effector process (e.g., *CTSG*, *CD68*, *IFIT3* and *IL18*) including hub genes (*CYBB* and *HCK*) in MergeCohort_Turquoise module. *SP1* regulated gene involved in cell activation (e.g., *TIMP1*, *LTF*, *FGL2* and *LYZ*).

For further analysis the expression of the hub genes and key TFs *in vitro* models of NASH, we retrieved public available RNA-seq data (the RNA-seq data of L02 hepatocytes (PRJNA726826) and murine primary hepatocytes (PRJNA726846) treated with palmitic acid and oleic acid (PAOA) for 0h, 12h and 24h, respectively (42)), we found hub genes (*CD53* and *SRGN*) and key TFs (*NFKB1*, *ELF1* and *EST1*) displayed higher expression in L02 hepatocytes treated with PAOA (Figure S4A). Moreover, we observed that hub genes (*Lcp1* and *Fcer1g*) and key TFs (*Ilf3*, *stat3* and *Est1*) showed increased expression in murine primary hepatocytes with PAOA treatment (Figure S4B). Together, these TFs and target genes identified in our study provide a promising list for investigators

or companies interested in conducting preclinical study into the mechanisms of and treatments for NASH both *in vitro* and *in vivo*.

Discussion

The global epidemic of NASH is a serious public health problem, the pathogenesis of NASH still remains unclear. Moreover, although liver biopsy currently remains the reference standard for diagnosis of NASH, it is an intrusive operation with risks and many shortcomings. Thus, identifying novel non-invasive biomarkers in NASH is of paramount importance in the prevention and therapy of this disease.

Thanks to the rapid development of high-throughput sequencing technology and gene chip technology, more and more researchers are actively pursuing molecular markers using data mining and analysis of sequencing data or gene chips to the diagnosis and treatment of disease (19, 43, 44). In our study, we analyzed gene expression profiles of NASH patients and normal controls from five independent GEO data sets. The batch of various platforms or batches is removed. DEGs were identified between normal liver tissues and NASH tissues, based on 831 DEGs between Normal-NASH group, we performed GO and Reactome pathway analysis to explore underlying mechanism of NASH. The results showed that enriched pathways were involved in metabolism pathways, inflammatory response and immune response, extracellular matrix organization (Figures 2C, D), conforming their association with NASH development and progression.

Subsequently, we constructed a co-expression network and identified 17 different modules by WGCNA, among which 11 modules were significantly associated with the status of NASH. DEG numbers showed a significant enrichment in seven important modules (Figure 3D). The results of this study indicated that the identified modules are biologically rational, majority of which are enriched for specific GO terms and KEGG pathways, sharing some commonality with the existing literature. For example, module Black and Brown, are markedly negative correlated with NASH status. Both the Black and Brown were most significantly enriched in cellular amino acid catabolic process. Recent studies showed that deregulation in amino acid metabolism seem to be involved in the appearance of NASH (39, 45). In addition, previous research has demonstrated that lipid metabolism significantly altered during NASH progression (46). Our data found Grey60 module that was significantly upregulated in NASH, enriched in the lipid metabolism pathways, encompassing hub genes related to cholesterol metabolism (*FDFT1*, *NSDHL*, *IDI1*, *SQLE*, *MVD*, *HMGCS1*, *HMGCR* and *LSS*) as well as fatty acid metabolism (*FASN*, *ELOVL6*, *FADS1*, *FADS2*, *ACACA*, *ELOVL6*, *PKLR* and *THRSP*) (Figure 4C). Similarly, previous biological network analysis identified cholesterol synthesis genes in human NAFLD (e.g., *FDFT1*, *NSDHL*, *IDI1*, *SQLE*, *MVD*, *HMGCS1* and *HMGCR*) and fatty acid metabolism genes (e.g., *Fasn*, *Thrsp* and *Pklr*) in NAFLD mouse model that were also reported to be deregulated by (47) and (18), respectively. Thus, despite the differences in study design, the three studies coverage on a number of key biological findings.

Inflammation is an important factor driving NASH progression. Our current systematic transcriptomic analysis also highlighted the importance of the Turquoise module in modulating NASH occurrence and development. This study found that the immune-related pathways were mostly enriched in the Turquoise module, which contained the highest number of differentially deregulated genes (Figure 3D). Moreover, we demonstrated the highest preservation of the Turquoise module between the MergeCohort and validation dataset GSE135251 (Figure 5A). The top hub genes overexpression in NASH samples and linking immune-related pathways belonged to *CD53*, *LCP1*, *LAPTM5*, *NCKAP1L*, *C3AR1*, *FGL2*, *PLEK*, *HLA-DRA*, *FPR3* and *SRGN*, which also showed positive correlation with histological grade (Figure 7C). Further validation by mouse NASH model, the expression of *CD53*, *LCP1*, *LAPTM5*, *NCKAP1L*, *C3AR1*, *FGL2*, *PLEK* and *SRGN* were significantly upregulated (Figure 7D). The role of *CD53*, *C3AR1*, *NCKAP1L* and *FGL2* genes in regulation of immune responses has recently been proposed in previous studies. CD53 is a member of the tetraspanin membrane protein family that may be involved in transmembrane signal transduction (48). *CD53* has been reported to associate with liver inflammation and insulin sensitivity (49). LAPTM5 is a transmembrane protein which is preferentially expressed in immune cells, and it acts as a positive regulator of proinflammatory signaling pathways in macrophages (50). Previous study revealed that *LAPTM5* could interact with *CDC42*, and promote its degradation, then suppressed the activation of MAPK signaling pathway, hence ameliorated NASH in mouse (51). Besides, *LAPTM5* has been shown to be significantly upregulated in HCC tissues compared to normal liver tissues, and Pan et al. reported that *LAPTM5* could remarkably accelerate autophagic flux by promoting fusion of lysosomes with autophagosomes to drive lenvatinib resistance in HCC (52). Moreover, C3AR1 is a G protein-coupled receptor (GPCR) protein, which participates in the complement system and can stimulate the production of IL-1β and TGFβ (53). Interestingly, Han et al. found that *C3ar1* knockout mice showed drastically less severe fibrosing steatohepatitis, concomitantly with reduced hepatic stellate cells (HSCs) activation when compared with the wildtype littermates (54). In addition, the mRNA level of *LCP1* in liver tissue of NAFLD patients was strongly increased (300%) compare to the control group in a previous GWAS study (55), and Miller et al. used proteomic method to describe the proteome of NAFLD and observed that *LCP1* performed well in distinguishing the disease state from control group, NAFL from NASH and fibrosis grading (56). Notably, our study also found that the Turquoise module including hub gene *HLA-DRA*, displayed higher expression in NASH, which associated with NAFLD loci found by GWAS, and genetic variants of *HLA−DRA* has been recently reported to affect hepatitis development in a Korean population (57). Additionally, it has been shown that *SRGN*, *CD53*, *NCKAP1L*, *LCP1*, *EVI2B*, *MPEG1* and *TYROBP* may be potential pathological target gene for NAFLD and NASH, which is highly similar to our Turquoise module (58).

It should be noted that NASH is regarded as an inflammatory subtype of NAFLD with steatosis and evidence of hepatocyte injury and interactions between multiple immune cells. Increasing

evidence has demonstrated the high heterogeneity and plasticity of macrophage populations in human liver (59). For example, Ramachandran et al. adopted scRNA-seq approach to discover a disease-associated TREM2+/CD9+ macrophage population that was remarkably expanded in human cirrhotic livers. Therapeutic inhibition of $CCR2^{+}$ bone marrow-derived macrophages has been reported to alleviate inflammation and fibrosis in mouse NASH and fibrosis in human disease (36, 60). Similarly, our integrated scRNA-seq analysis revealed that the hub genes in the Turquoise module were mainly enriched in macrophage and dendritic cells, conforming the importance of which during NASH progression. For instance, our study found that expression of *FGL2* was elevated in macrophages and dendritic cells (Figure 8C). A recent study demonstrated that *Fgl2* expression in the livers of both humans and mice with NASH was significantly increased along with the accumulation of hepatic macrophages (61). Moreover, we found that the expression of *CSF1R* gene, a marker for pan-macrophages reported to be involved in hepatic fibrosis, was also considered as a potential marker for hepatocarcinogenesis (62). By analyzing the association between *LCP1* and immune cells, Zhang et al. found *LCP1* was significantly positively related to memory B cells as well as M1 macrophages (58). Our study also observed that hub gene *HLA-DRA* was higher expressed in both macrophages and dendritic cells (Figure 8C). Intriguingly, previous reports examining human NASH livers using single-cell RNA sequencing reported that M-Mac-1 included three genes, *HLA-DRA*, *HLA-DQA2* and *HLA-DQB2* (63), which was related to NAFLD loci (57, 64, 65). Further, recent study reported that cDC-related gene expression signatures in human livers were associated with NASH pathology (66). These findings emphasized the importance of further studies of the subpopulations of inflammatory macrophages and dendritic cells in NASH progression. However, more single-cell transcriptome data focusing on NASH progression among NASH patients are needed in future studies.

Several studies involving transcription factors have indicated therapeutic effects in NASH (67, 68), for example, transcription factors including *PPARs*, *LXR* and *FXR* are mainly known for their roles in altering lipid metabolism in NAFLD/NASH development. Agonists of *PPARs* and *FXR* have been investigated extensively in mouse models (69, 70), clinical trials presently are ongoing to test the effects of these drugs for potential NASH treatments. In addition, *PPARs*, *LXR* and *FXR* not only regulate lipid metabolism but also exert anti-inflammatory functions *via* direct and indirect mechanisms as shown by the suppression of several proinflammatory genes (71–74). Therefore, the detection of an immune-related transcription factor seems to be essential for the identification of novel therapeutic targets in NAFLD/NASH. In present study, we observed that the immune-related module enriched TFs including *NFKB1*, *STAT3*, *RFX5*, *ILF3*, *ELF1*, *SPI1*, *ETS1* and *CEBPA*, the expression of which enhanced with NASH progression (Figure 9B). Among the TFs, *NFKB1*, *STAT3*, *SPI1*, *ETS1*, *CEBPA* and *ELF1* have been reported to be linked to NAFLD/NASH by literature searching.

NF-κB is a protein complex that plays a central role in regulating the expression of cytokines and chemokines, and recent studies suggest that NF-κB is highly activated both in mice and patients with NASH (75, 76). *NFKB1* (p105/p50), a member of NF-κB family, emerging evidence suggests that NF-κB1-gene-coded proteins p105 and p50 have critical regulatory activities of inflammatory responses (77, 78). Previous study have showed that *Nfkb1*-deficient mice enhanced NASH progression to fibrosis by favouring NKT cell recruitment (79). In addition, Jurk et al. reported that loss of *Nfkb1* in mouse promoted ageing-related chronic liver disease, featured by steatosis, hepatitis, fibrosis and HCC (80), which point to the possible relevance of polymorphisms in human *NFKB1* gene as a risk factor for the progression of inflammatory disease (81).

STAT family members with inflammatory biological functions notably *STAT1* and *STAT3* have been linked to NAFLD and NASH. Grohmann and colleagues demonstrated that the oxidative hepatic environment in obesity restrained the *STAT1* and *STAT3* phosphatase *TCPTP*, which led to potentiate *STAT1* and *STAT3* signaling, and further increase the risk of developing NASH and HCC in the setting of nutritional excess (82). On the other hand, the suppression of *TCPTP*, coupled with heightened *STAT1* and *STAT3* signaling, were easily detectable events in the livers of patients with NASH (82). Moreover, a recently study revealed that dampening *IL6/STAT3* activity alleviated the I148M-mediated susceptibility to NAFLD, while boosting it in wild-type liver cultures enhanced the development of NAFLD (83). Additionally, downregulation of *STAT3* expression can activate autophagy and inhibit the inflammatory response of NASH (84, 85). Interestingly, other transcription factor such as *SPI1*, *ETS1* and *CEBPA* have been described to be a promising target for NASH prevention and treatment. Liu et al. applied proteomics strategy to identify *SPI1* as critical TF, *SPI1* expression was positively related to resistance indicator HOMA-IR and the inflammatory marker TNFA in human liver biopsies, and inhibition of *SPI1* ameliorated metabolic dysfunction and NASH (86). It has been proven that *Ets1* acted as a positive regulator of TGF-β1 signaling, which accelerated the development of NASH in mice (87). Notably, Vujkovic et al. recently presented a GWAS study and identified 77 genome-wide loci significantly associated with NAFLD (diagnosed using elevated ALT as a proxy for NAFLD), of interest is that for nine SNPs, the cATL risk allele was associated with lower BMI including *CEBPA* (65).

There are few studies of *RFX5*, *ELF1* and *ILF3* that have been reported at present in the field of NAFLD and NASH. *RFX5*, a classical transcription regulator of MHCII gene expression in the immune system. It has been previously shown that *RFX5* displayed higher transcriptional activity in both human NASH and mouse model of NASH (68). Interestingly, *RFX5* mRNA has previously been shown overexpressed in HCC compared with non-tumor tissue, which promoted HCC progression *via* transcriptionally activating *KDM4A*, *TPP1* and *YWHAQ* (88–90). Moreover, our results also showed that *RFX5* are the prominent regulators of expression of HLA class II genes in the immune-related module. Interestingly, *RFX5* was recently reported to enhance surface expression of *HLA-DR* molecules, which promoted tissue macrophages-dependent expansion of antigen-specific T cells in rheumatoid arthritis (91). In addition, *ELF1* regulated hub gene *CYBB* in MergeCohort_Turquoise module, the mechanism of TAZ-

induced *Cybb* leading to liver tumor formation in NASH has been well defined (92).

ILF3, also known as NF90/NF110, encodes a double-stranded RNA (dsRNA)-binding protein which can regulate gene expression and stabilize mRNA (93, 94). Recent studies have reported insights into the possible physiological roles of *ILF3* in dyslipidemia, the cardiovascular system, neurodegenerative disorder as well as in tumorigenesis and progression of different cancers. Zhang et al. demonstrated that *ILF3* together with another eight transcription regulators control late-onset Alzheimer's disease (LOAD) risk genes *HLA-DRB1* and *HLA-DQA1* expression in human microglial cells (95). Moreover, there is evidence that *ILF3* could have an important role in inflammatory pathophysiology *in vivo*, Nazitto et al. identified *ILF3* as negative regulator of innate immune response and dendritic cell (DC) maturation, and found that knockdown of *ILF3* led to significantly elevated expression of genes (*CD86*, *CD80* and *HLA-DR*) associated with DC maturation in the primary human monocyte-derived DCs during stimulation with viral mimetics or classic innate agonists (96). In addition, previous studies have revealed the essential roles of deregulated lncRNA *ILF3* divergent transcript (*ILF3-AS1*) in HCC, Bo et al. found that *ILF3-AS1* expression was significantly increased in HCC tissues and also associated with prognosis of HCC patients, and knockdown of ILF3-AS1 expression suppressed HCC cell proliferation, migration and invasion (97). Yan et al. also observed that *ILF3-AS1* silencing inhibited the hepatocellular carcinoma tumor growth (98). However, the regulation roles of *RFX5* and *ILF3* on *HLA-DR* molecules in the progression of NASH have also not been well defined. Therefore, our results provide a very meaningful direction for future research.

In summary, unlike previous studies with limitation of a few human NASH transcriptome data or focusing on individual genes influencing NASH progression, our network-driven strategy generated a comprehensive and unbiased view of the modules, hub genes and critical transcriptional factors associated with NASH. In particular, the Turquoise module and regulators involving immune-related pathways especially transcription factor *RFX5* coordinating antigen processing and presenting function in NASH progression deserve further attention. The main limitation of present study is that all conclusions are based on transcriptomic data from human and lack verification from relevant experiments *in vitro*/*in vivo* disease models. Nevertheless, it provides useful and novel molecular candidates in dysregulated pathways for NASH prognosis and therapeutic targets.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

Conception and design: J-JZ and FX. Acquisition and analysis of data: J-JZ, YS and X-YC. Investigation: J-JZ, YS, X-YC, M-LJ, F-HY and JZ. Software: J-JZ. Validation: YS, X-YC, M-LJ, S-LX and JZ. Visualization: J-JZ. Writing–original draft: J-JZ. Writing–review & editing: J-JZ, X-YC, F-HY and FX. Funding: J-JZ. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (82200653), the doctoral startup fund of Gannan Medical University (QD202112).

Acknowledgments

We would like to thank all the researchers who have shared their data in GEO and SRA databases.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1115890/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Principal component analysis (PCA) of gene expression data set with the first two components. **(A)** PCA plot without batch effect elimination. **(B)** PCA plot with batch effect elimination with ComBat algorithm. PC1, first principal component; PC2, second principal component.

SUPPLEMENTARY FIGURE 2
Integrated scRNA-seq analysis. **(A)** Significant principal components (PCs) were determined via the JackStraw function in Seurat R-packages. PCs 1-17 were used for graph-based clustering (resolution = 0.4) to identify distinct clusters. **(B)** UMAP visualization of scRNA-seq data from four healthy (n = 2) and cirrhotic (n = 2) human livers annotated by liver sample. **(C)** UMAP visualization of cirrhotic and healthy control groups annotated by liver disease status. UMAP, uniform manifold approximation and projection.

SUPPLEMENTARY FIGURE 3
UMAP plots show the expression and distribution of top 25 hub genes in MergeCohort_Turquoise module for each cell type. The expression

levels of those hub genes are expressed by the color transition from red to grey.

SUPPLEMENTARY FIGURE 4
Assessment of the expression patterns of hub genes and key TFs in MergeCohort_Turquoise module in vitro models of NASH using publicly available RNA-seq data of L02 hepatocytes (PRJNA726826) and murine primary hepatocytes (PRJNA726846) treated with palmitic acid and oleic acid (PAOA) for 0h, 12h and 24h, respectively. Heatmap shows the expression patterns of hub genes and key TFs in in L02 hepatocytes **(A)** and mouse primary hepatocytes **(B)** with PAOA treatment for 12 h and 24 h (1 technical replicate of 3 biological replicates for each group).

SUPPLEMENTARY TABLE 1
Characteristics of six liver transcriptome datasets from GEO comparing NASH patients with healthy controls (HC).

SUPPLEMENTARY TABLE 2
Significant enriched pathway of GSEA in Control-NASH group of five GEO datasets.

SUPPLEMENTARY TABLE 3
Significant enriched pathway of GSEA within the intersection of more than 4 list in Control-NASH group of five GEO datasets.

SUPPLEMENTARY TABLE 4
DEGs identified in MergeCohort between HC and NASH.

SUPPLEMENTARY TABLE 5
GO analysis of DEGs between HC and NASH.

SUPPLEMENTARY TABLE 6
Reactome pathways analysis of DEGs between HC and NASH.

SUPPLEMENTARY TABLE 7
Characteristics of gene modules obtained by WGCNA.

SUPPLEMENTARY TABLE 8
GO analysis of genes in each module.

SUPPLEMENTARY TABLE 9
KEGG pathways analysis of genes in each module.

SUPPLEMENTARY TABLE 10
DisGeNET enrichment analysis of genes in MergeCohort_Turquoise and GSE135251_Turquoise.

SUPPLEMENTARY TABLE 11
KEGG pathway analysis of genes in MergeCohort_Turquoise and GSE135251_Turquoise.

SUPPLEMENTARY TABLE 12
Transcription factor enrichment analysis by ChEA3 of MergeCohort_Turquoise module genes.

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-Analytic assessment of prevalence, incidence, and outcomes. *Hepatology* (2016) 64:73–84. doi: 10.1002/hep.28431

2. Brunt EM. Pathology of nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* (2010) 7:195–203. doi: 10.1038/nrgastro.2010.21

3. Williams CD, Stengel J, Asike MI, Torres DM, Shaw J, Contreras M, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: A prospective study. *Gastroenterology* (2011) 140:124–31. doi: 10.1053/j.gastro.2010.09.038

4. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: Current concepts and future challenges. *Nat Rev Gastroenterol Hepatol* (2019) 16:411–28. doi: 10.1038/s41575-019-0145-7

5. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American association for the study of liver diseases. *Hepatology* (2018) 67:328–57. doi: 10.1002/hep.29367

6. Sanyal AJ. Past, present and future perspectives in nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* (2019) 16:377–86. doi: 10.1038/s41575-019-0144-8

7. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* (2008) 40:1461–5. doi: 10.1038/ng.257

8. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PloS Genet* (2011) 7:e1001324. doi: 10.1371/journal.pgen.1001324

9. Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjærg-Hansen A, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* (2014) 46:352–6. doi: 10.1038/ng.2901

10. Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. *N Engl J Med* (2018) 378:1096–106. doi: 10.1056/NEJMoa1712191

11. Emdin CA, Haas ME, Khera AV, Aragam K, Chaffin M, Klarin D, et al. A missense variant in mitochondrial amidoxime reducing component 1 gene and protection against liver disease. *PloS Genet* (2020) 16:e1008629. doi: 10.1371/journal.pgen.1008629

12. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort☆. *J Hepatol* (2020) 73:505–15. doi: 10.1016/j.jhep.2020.04.003

13. Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M, et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med* (2020) 12:eaba4448. doi: 10.1126/scitranslmed.aba4448

14. Sveinbjornsson G, Ulfarsson MO, Thorolfsdottir RB, Jonsson BA, Einarsson E, Gunnlaugsson G, et al. Multiomics study of nonalcoholic fatty liver disease. *Nat Genet* (2022) 54:1652–63. doi: 10.1038/s41588-022-01199-5

15. Zhang X-J, She Z-G, Wang J, Sun D, Shen L-J, Xiang H, et al. Multiple omics study identifies an interspecies conserved driver for nonalcoholic steatohepatitis. *Sci Transl Med* (2021) 13:eabg8117. doi: 10.1126/scitranslmed.abg8117

16. Jia X, Zhai T. Integrated analysis of multiple microarray studies to identify novel gene signatures in non-alcoholic fatty liver disease. *Front Endocrinol* (2019) 10:599. doi: 10.3389/fendo.2019.00599

17. Wu C, Zhou Y, Wang M, Dai G, Liu X, Lai L, et al. Bioinformatics analysis explores potential hub genes in nonalcoholic fatty liver disease. *Front Genet* (2021) 12:772487. doi: 10.3389/fgene.2021.772487

18. Yang H, Arif M, Yuan M, Li X, Shong K, Türkez H, et al. A network-based approach reveals the dysregulated transcriptional regulation in non-alcoholic fatty liver disease. *iScience* (2021) 24:103222. doi: 10.1016/j.isci.2021.103222

19. Gao R, Wang J, He X, Wang T, Zhou L, Ren Z, et al. Comprehensive analysis of endoplasmic reticulum-related and secretome gene expression profiles in the progression of non-alcoholic fatty liver disease. *Front Endocrinol* (2022) 13:967016. doi: 10.3389/fendo.2022.967016

20. Esmaili S, Langfelder P, Belgard TG, Vitale D, Azardaryany MK, Alipour Talesh G, et al. Core liver homeostatic Co-expression networks are preserved but respond to perturbations in an organism- and disease-specific manner. *Cell Syst* (2021) 12:432–45.e7. doi: 10.1016/j.cels.2021.04.004

21. Misselbeck K, Parolo S, Lorenzini F, Savoca V, Leonardelli L, Bora P, et al. A network-based approach to identify deregulated pathways and drug effects in metabolic syndrome. *Nat Commun* (2019) 10:5215. doi: 10.1038/s41467-019-13208-z

22. Langfelder P, Horvath S. WGCNA: An r package for weighted correlation network analysis. *BMC Bioinform* (2008) 9:559. doi: 10.1186/1471-2105-9-559

23. Zhang B, Horvath S. A general framework for weighted gene Co-expression network analysis. *Stat Appl Genet Mol Biol* (2005) 4(Article17). doi: 10.2202/1544-6115.1128

24. Saris CG, Horvath S, van Vught PW, van Es MA, Blauw HM, Fuller TF, et al. Weighted gene Co-expression network analysis of the peripheral blood from amyotrophic lateral sclerosis patients. *BMC Genom* (2009) 10:405. doi: 10.1186/1471-2164-10-405

25. Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H. Gene Co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat Commun* (2014) 5:3231. doi: 10.1038/ncomms4231

26. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets–update. *Nucleic Acids Res* (2013) 41:D991–5. doi: 10.1093/nar/gks1193

27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

28. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* (2003) 4:249–64. doi: 10.1093/biostatistics/4.2.249

29. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034

30. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* (2016) 44:W90–7. doi: 10.1093/nar/gkw377

31. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PloS Comput Biol* (2011) 7:e1001057. doi: 10.1371/journal.pcbi.1001057

32. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA* (2006) 103:17973–8. doi: 10.1073/pnas.0605938103

33. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. Disgenet: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* (2016) 45:D833–9. doi: 10.1093/nar/gkw943

34. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* (2018) 46:D380–d6. doi: 10.1093/nar/gkx1013

35. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: Transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res* (2019) 47:W212–w24. doi: 10.1093/nar/gkz446

36. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* (2019) 575:512–8. doi: 10.1038/s41586-019-1631-3

37. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM3rd, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177:1888–902.e21. doi: 10.1016/j.cell.2019.05.031

38. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* (2019) 20:163–72. doi: 10.1038/s41590-018-0276-y

39. Rom O, Liu Y, Liu Z, Zhao Y, Wu J, Ghrayeb A, et al. Glycine-based treatment ameliorates nafld by modulating fatty acid oxidation, glutathione synthesis, and the gut microbiome. *Sci Transl Med* (2020) 12:eaaz2841. doi: 10.1126/scitranslmed.aaz2841

40. Leung H, Long X, Ni Y, Qian L, Nychas E, Siliceo SL, et al. Risk assessment with gut microbiome and metabolite markers in nafld development. *Sci Transl Med* (2022) 14:eabk0855. doi: 10.1126/scitranslmed.abk0855

41. Min-DeBartolo J, Schlerman F, Akare S, Wang J, McMahon J, Zhan Y, et al. Thrombospondin-I is a critical modulator in non-alcoholic steatohepatitis (NASH). *PloS One* (2019) 14:e0226854. doi: 10.1371/journal.pone.0226854

42. Wang L, Zhang X, Lin ZB, Yang PJ, Xu H, Duan JL, et al. Tripartite motif 16 ameliorates nonalcoholic steatohepatitis by promoting the degradation of phospho-TAK1. *Cell Metab* (2021) 33:1372–88.e7. doi: 10.1016/j.cmet.2021.05.019

43. Xie X, Zhang Y, Yu J, Jiang F, Wu C. Significance of m^6A regulatory factor in gene expression and immune function of osteoarthritis. *Front Physiol* (2022) 13:918270. doi: 10.3389/fphys.2022.918270

44. Yu J, Xie X, Zhang Y, Jiang F, Wu C. Construction and analysis of a joint diagnosis model of random forest and artificial neural network for obesity. *Front Med* (2022) 9:906001. doi: 10.3389/fmed.2022.906001

45. Hoyles L, Fernández-Real J-M, Federici M, Serino M, Abbott J, Charpentier J, et al. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. *Nat Med* (2018) 24:1070–80. doi: 10.1038/s41591-018-0061-3

46. Loomba R, Quehenberger O, Armando A, Dennis EA. Polyunsaturated fatty acid metabolites as novel lipidomic biomarkers for noninvasive diagnosis of nonalcoholic steatohepatitis. *J Lipid Res* (2015) 56:185–92. doi: 10.1194/jlr.P055640

47. Chella Krishnan K, Kurt Z, Barrere-Cain R, Sabir S, Das A, Floyd R, et al. Integration of multi-omics data from mouse diversity panel highlights mitochondrial dysfunction in non-alcoholic fatty liver disease. *Cell Syst* (2018) 6:103–15.e7. doi: 10.1016/j.cels.2017.12.006

48. Yeung L, Anderson JML, Wee JL, Demaria MC, Finsterbusch M, Liu YS, et al. Leukocyte tetraspanin CD53 restrains α_3 integrin mobilization and facilitates cytoskeletal remodeling and transmigration in mice. *J Immunol* (2020) 205:521–32. doi: 10.4049/jimmunol.1901054

49. Ehses JA, Lacraz G, Giroix MH, Schmidlin F, Coulaud J, Kassis N, et al. IL-1 antagonism reduces hyperglycemia and tissue inflammation in the type 2 diabetic GK rat. *Proc Natl Acad Sci USA* (2009) 106:13998–4003. doi: 10.1073/pnas.0810087106

50. Glowacka WK, Alberts P, Ouchida R, Wang JY, Rotin D. LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. *J Biol Chem* (2012) 287:27691–702. doi: 10.1074/jbc.M112.355917

51. Jiang L, Zhao J, Yang Q, Li M, Liu H, Xiao X, et al. Lysosomal-associated protein transmembrane 5 ameliorates non-alcoholic steatohepatitis through degradating CDC42. *Res Square* (2022). doi: 10.21203/rs.3.rs-2065929/v1

52. Pan J, Zhang M, Dong L, Ji S, Zhang J, Zhang S, et al. Genome-scale CRISPR screen identifies LAPTM5 driving lenvatinib resistance in hepatocellular carcinoma. *Autophagy* (2022) 7:1–15. doi: 10.1080/15548627.2022.2117893

53. Li L, Yin Q, Tang X, Bai L, Zhang J, Gou S, et al. C3a receptor antagonist ameliorates inflammatory and fibrotic signals in type 2 diabetic nephropathy by suppressing the activation of TGF-β/smad3 and IKBα pathway. *PloS One* (2014) 9:e113639. doi: 10.1371/journal.pone.0113639

54. Han J, Zhang X, Lau JK-C, Fu K, Lau HC, Xu W, et al. Bone marrow-derived macrophage contributes to fibrosing steatohepatitis through activating hepatic stellate cells. *J Pathol* (2019) 248:488–500. doi: 10.1002/path.5275

55. Adams LA, White SW, Marsh JA, Lye SJ, Connor KL, Maganga R, et al. Association between liver-specific gene polymorphisms and their expression levels with nonalcoholic fatty liver disease. *Hepatology* (2013) 57:590–600. doi: 10.1002/hep.26184

56. Miller MH, Walsh SV, Atrih A, Huang JT, Ferguson MA, Dillon JF. Serum proteome of nonalcoholic fatty liver disease: A multimodal approach to discovery of biomarkers of nonalcoholic steatohepatitis. *J Gastroenterol Hepatol* (2014) 29:1839–47. doi: 10.1111/jgh.12614

57. Hong M, Jung J, Jin H-S, Hwang D. Genetic polymorphism of HLA-DRA and alcohol consumption affect hepatitis development in the Korean population. *Genes Genomics* (2022) 44:1109–16. doi: 10.1007/s13258-022-01286-1

58. Zhang X, Li J, Liu T, Zhao M, Liang B, Chen H, et al. Identification of key biomarkers and immune infiltration in liver tissue after bariatric surgery. *Dis Markers* (2022) 2022:4369329. doi: 10.1155/2022/4369329

59. MacParland SA, Liu JC, Ma X-Z, Innes BT, Bartczak AM, Gage BK, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun* (2018) 9:4383. doi: 10.1038/s41467-018-06318-7

60. Xiong X, Kuang H, Ansari S, Liu T, Gong J, Wang S, et al. Landscape of intercellular crosstalk in healthy and Nash liver revealed by single-cell secretome gene analysis. *Mol Cell* (2019) 75:644–60.e5. doi: 10.1016/j.molcel.2019.07.028

61. Hu J, Wang H, Li X, Liu Y, Mi Y, Kong H, et al. Fibrinogen-like protein 2 aggravates nonalcoholic steatohepatitis *Via* interaction with TLR4, eliciting inflammation in macrophages and inducing hepatic lipid metabolism disorder. *Theranostics* (2020) 10:9702–20. doi: 10.7150/thno.44297

62. Iio E, Ocho M, Togayachi A, Nojima M, Kuno A, Ikehara Y, et al. A novel glycobiomarker, wisteria floribunda agglutinin macrophage colony-stimulating factor receptor, for predicting carcinogenesis of liver cirrhosis. *Int J Cancer* (2016) 138:1462–71. doi: 10.1002/ijc.29880

63. Fred RG, Steen Pedersen J, Thompson JJ, Lee J, Timshel PN, Stender S, et al. Single-cell transcriptome and cell type-specific molecular pathways of human non-alcoholic steatohepatitis. *Sci Rep* (2022) 12:13484. doi: 10.1038/s41598-022-16754-7

64. Doganay L, Katrinli S, Colak Y, Senates E, Zemheri E, Ozturk O, et al. HLA DQB1 alleles are related with nonalcoholic fatty liver disease. *Mol Biol Rep* (2014) 41:7937–43. doi: 10.1007/s11033-014-3688-2

65. Vujkovic M, Ramdas S, Lorenz KM, Guo X, Darlay R, Cordell HJ, et al. A multiancestry genome-wide association study of unexplained chronic ALT elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. *Nat Genet* (2022) 54:761–71. doi: 10.1038/s41588-022-01078-z

66. Deczkowska A, David E, Ramadori P, Pfister D, Safran M, Li B, et al. $XCR1^+$ type 1 conventional dendritic cells drive liver pathology in non-alcoholic steatohepatitis. *Nat Med* (2021) 27:1043–54. doi: 10.1038/s41591-021-01344-3

67. Steensels S, Qiao J, Ersoy BA. Transcriptional regulation in non-alcoholic fatty liver disease. *Metabolites* (2020) 10:283. doi: 10.3390/metabo10070283

68. Loft A, Alfaro AJ, Schmidt SF, Pedersen FB, Terkelsen MK, Puglia M, et al. Liver-Fibrosis-Activated transcriptional networks govern hepatocyte reprogramming and intra-hepatic communication. *Cell Metab* (2021) 33:1685–700.e9. doi: 10.1016/j.cmet.2021.06.005

69. Lefere S, Puengel T, Hundertmark J, Penners C, Frank AK, Guillot A, et al. Differential effects of selective- and pan-PPAR agonists on experimental steatohepatitis and hepatic macrophages☆. *J Hepatol* (2020) 73:757–70. doi: 10.1016/j.jhep.2020.04.025

70. Radun R, Trauner M. Role of FXR in bile acid and metabolic homeostasis in NASH: Pathogenetic concepts and therapeutic opportunities. *Semin Liver Dis* (2021) 41:461–75. doi: 10.1055/s-0041-1731707

71. Cariello M, Piccinin E, Moschetta A. Transcriptional regulation of metabolic pathways *Via* lipid-sensing nuclear receptors PPARs, FXR, and LXR in NASH. *Cell Mol Gastroenterol Hepatol* (2021) 11:1519–39. doi: 10.1016/j.jcmgh.2021.01.012

72. Gordon S. Alternative activation of macrophages. *Nat Rev Immunol* (2003) 3:23–35. doi: 10.1038/nri978

73. Joseph SB, Castrillo A, Laffitte BA, Mangelsdorf DJ, Tontonoz P. Reciprocal regulation of inflammation and lipid metabolism by liver X receptors. *Nat Med* (2003) 9:213–9. doi: 10.1038/nm820

74. Wang YD, Chen WD, Wang M, Yu D, Forman BM, Huang W. Farnesoid X receptor antagonizes nuclear factor kappaB in hepatic inflammatory response. *Hepatology* (2008) 48:1632–43. doi: 10.1002/hep.22519

75. Mussbacher M, Salzmann M, Brostjan C, Hoesel B, Schoergenhofer C, Datler H, et al. Cell type-specific roles of NF-κB linking inflammation and thrombosis. *Front Immunol* (2019) 10:85. doi: 10.3389/fimmu.2019.00085

76. Severa M, Islam SA, Waggoner SN, Jiang Z, Kim ND, Ryan G, et al. The transcriptional repressor BLIMP1 curbs host defenses by suppressing expression of the chemokine CCL8. *J Immunol* (2014) 192:2291–304. doi: 10.4049/jimmunol.1301799

77. Beinke S, Ley SC. Functions of NF-kappaB1 and NF-KappaB2 in immune cell biology. *Biochem J* (2004) 382:393–409. doi: 10.1042/bj20040544

78. Panzer U, Steinmetz OM, Turner JE, Meyer-Schwesinger C, von Ruffer C, Meyer TN, et al. Resolution of renal inflammation: A new role for NF-kappaB1 (p50) in inflammatory kidney diseases. *Am J Physiol Renal Physiol* (2009) 297:F429–39. doi: 10.1152/ajprenal.90435.2008

79. Locatelli I, Sutti S, Vacchiano M, Bozzola C, Albano E. NF-κB1 deficiency stimulates the progression of non-alcoholic steatohepatitis (NASH) in mice by promoting NKT-Cell-Mediated responses. *Clin Sci* (2013) 124:279–87. doi: 10.1042/cs20120289

80. Jurk D, Wilson C, Passos JF, Oakley F, Correia-Melo C, Greaves L, et al. Chronic inflammation induces telomere dysfunction and accelerates ageing in mice. *Nat Commun* (2014) 2:4172. doi: 10.1038/ncomms5172

81. Cheng CW, Su JL, Lin CW, Su CW, Shih CH, Yang SF, et al. Effects of NFKB1 and NFKBIA gene polymorphisms on hepatocellular carcinoma susceptibility and clinicopathological features. *PloS One* (2013) 8:e56130. doi: 10.1371/journal.pone.0056130

82. Grohmann M, Wiede F, Dodd GT, Gurzov EN, Ooi GJ, Butt T, et al. Obesity drives STAT-1-Dependent NASH and STAT-3-Dependent HCC. *Cell* (2018) 175:1289–306.e20. doi: 10.1016/j.cell.2018.09.053

83. Park J, Zhao Y, Zhang F, Zhang S, Kwong AC, Zhang Y, et al. IL-6/STAT3 axis dictates the PNPLA3-mediated susceptibility to non-alcoholic fatty liver disease. *J Hepatol* (2022). doi: 10.1016/j.jhep.2022.08.022

84. Li YL, Li XQ, Wang YD, Shen C, Zhao CY. Metformin alleviates inflammatory response in non-alcoholic steatohepatitis by restraining signal transducer and activator of transcription 3-mediated autophagy inhibition *In vitro* and *In vivo*. *Biochem Biophys Res Commun* (2019) 513:64–72. doi: 10.1016/j.bbrc.2019.03.077

85. Mohammed S, Nicklas EH, Thadathil N, Selvarani R, Royce GH, Kinter M, et al. Role of necroptosis in chronic hepatic inflammation and fibrosis in a mouse model of increased oxidative stress. *Free Radic Biol Med* (2021) 164:315–28. doi: 10.1016/j.freeradbiomed.2020.12.449

86. Liu Q, Yu J, Wang L, Tang Y, Zhou Q, Ji S, et al. Inhibition of PU.1 ameliorates metabolic dysfunction and non-alcoholic steatohepatitis. *J Hepatol* (2020) 73:361–70. doi: 10.1016/j.jhep.2020.02.025

87. Liu D, Wang K, Li K, Xu R, Chang X, Zhu Y, et al. Ets-1 deficiency alleviates nonalcoholic steatohepatitis *Via* weakening TGF-β1 signaling-mediated hepatocyte apoptosis. *Cell Death Dis* (2019) 10:458. doi: 10.1038/s41419-019-1672-4

88. Zhao Y, Xie X, Liao W, Zhang H, Cao H, Fei R, et al. The transcription factor RFX5 is a transcriptional activator of the TPP1 gene in hepatocellular carcinoma. *Oncol Rep* (2017) 37:289–96. doi: 10.3892/or.2016.5240

89. Chen DB, Zhao YJ, Wang XY, Liao WJ, Chen P, Deng KJ, et al. Regulatory factor X5 promotes hepatocellular carcinoma progression by transactivating tyrosine 3-Monooxygenase/Tryptophan 5-monooxygenase activation protein theta and suppressing apoptosis. *Chin Med J* (2019) 132:1572–81. doi: 10.1097/cm9.0000000000000296

90. Chen DB, Xie XW, Zhao YJ, Wang XY, Liao WJ, Chen P, et al. RFX5 promotes the progression of hepatocellular carcinoma through transcriptional activation of Kdm4a. *Sci Rep* (2020) 10:14538. doi: 10.1038/s41598-020-71403-1

91. Hu Z, Zhao TV, Huang T, Ohtsuki S, Jin K, Goronzy IN, et al. The transcription factor RFX5 coordinates antigen-presenting function and resistance to nutrient stress in synovial macrophages. *Nat Metab* (2022) 4:759–74. doi: 10.1038/s42255-022-00585-x

92. Wang X, Zeldin S, Shi H, Zhu C, Saito Y, Corey KE, et al. TAZ-induced cybb contributes to liver tumor formation in non-alcoholic steatohepatitis. *J Hepatol* (2022) 76:910–20. doi: 10.1016/j.jhep.2021.11.031

93. Shi L, Godfrey WR, Lin J, Zhao G, Kao PN. NF90 regulates inducible IL-2 gene expression in T cells. *J Exp Med* (2007) 204:971–7. doi: 10.1084/jem.20052078

94. Jayachandran U, Grey H, Cook AG. Nuclear factor 90 uses an ADAR2-like binding mode to recognize specific bases in dsRNA. *Nucleic Acids Res* (2016) 44:1924–36. doi: 10.1093/nar/gkv1508

95. Zhang X, Zou M, Wu Y, Jiang D, Wu T, Zhao Y, et al. Regulation of the late onset alzheimer's disease associated *HLA-DQA1/DRB1* expression. *Am J Alzheimers Dis Other Demen* (2022) 37:15333175221085066. doi: 10.1177/15333175221085066

96. Nazitto R, Amon LM, Mast FD, Aitchison JD, Aderem A, Johnson JS, et al. ILF3 is a negative transcriptional regulator of innate immune responses and myeloid dendritic cell maturation. *J Immunol* (2021) 206:2949–65. doi: 10.4049/jimmunol.2001235

97. Bo C, Li N, He L, Zhang S, An Y. Long non-coding RNA ILF3-AS1 facilitates hepatocellular carcinoma progression by stabilizing ILF3 mRNA in an m^6A-dependent manner. *Hum Cell* (2021) 34:1843–54. doi: 10.1007/s13577-021-00608-x

98. Yan G, Chang Z, Wang C, Gong Z, Xin H, Liu Z. LncRNA ILF3-AS1 promotes cell migration, invasion and emt process in hepatocellular carcinoma *Via* the miR-628-5p/MEIS2 axis to activate the notch pathway. *Dig Liver Dis* (2022) 54:125–35. doi: 10.1016/j.dld.2021.04.036

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 04 April 2023
DOI 10.3389/fendo.2023.1127336

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Erfu Xie,
Nanjing Medical University, China
Abeer Ramadan,
National Research Centre, Egypt

*CORRESPONDENCE
Fengqiong Han
29113126@qq.com
Rongrong He
rongronghe@jnu.edu.cn
Runmin Guo
1314ivu@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 19 December 2022
ACCEPTED 20 March 2023
PUBLISHED 04 April 2023

CITATION
Zeng Q, Zou D, Liu N, Wei Y, Yang J, Wu W, Han F, He R and Guo R (2023) Association of miR-196a2 and miR-27a polymorphisms with gestational diabetes mellitus susceptibility in a Chinese population.
Front. Endocrinol. 14:1127336.
doi: 10.3389/fendo.2023.1127336

Association of miR-196a2 and miR-27a polymorphisms with gestational diabetes mellitus susceptibility in a Chinese population

Qiaoli Zeng[1,2,3,4,5†], Dehua Zou[2,3†], Na Liu[6†], Yue Wei[7,8], Jing Yang[9], Weibiao Wu[10], Fengqiong Han[9*], Rongrong He[2,3*] and Runmin Guo[1,4,5,11*]

[1]Department of Internal Medicine, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [2]State Key Laboratory of Quality Research in Chinese Medicine, School of Pharmacy, Macau University of Science and Technology, Taipa, Macau, China, [3]Guangdong Engineering Research Center of Chinese Medicine & Disease Susceptibility, Jinan University, Guangzhou, Guangdong, China, [4]Key Laboratory of Research in Maternal and Child Medicine and Birth Defects, Guangdong Medical University, Foshan, Guangdong, China, [5]Matenal and Child Research Institute, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [6]Department of Pediatrics, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [7]Department of Ultrasound, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [8]Department of Ultrasound, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China, [9]Department of Obstetric, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [10]Medical Genetics Laboratory, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [11]Department of endocrinology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China

Introduction: MiR-196a2 and miR-27a play a key role in the regulation of the insulin signaling pathway. Previous studies have indicated that miR-27a rs895819 and miR-196a2 rs11614913 have a strong association with type 2 diabetes (T2DM), but very few studies have investigated their role in gestational diabetes mellitus (GDM).

Methods: A total of 500 GDM patients and 502 control subjects were enrolled in this study. Using the SNPscan™ genotyping assay, rs11614913 and rs895819 were genotyped. In the data treatment process, the independent sample t test, logistic regression and chi-square test were used to evaluate the differences in genotype, allele, and haplotype distributions and their associations with GDM risk. One-way ANOVA was conducted to determine the differences in genotype and blood glucose level.

Results: There were obvious differences in prepregnancy body mass index (pre-BMI), age, systolic blood pressure (SBP), diastolic blood pressure (DBP) and parity between GDM and healthy subjects ($P < 0.05$). After adjusting for the above factors, the miR-27a rs895819 C allele was still associated with an increased risk of GDM (C vs. T: OR=1.245; 95% CI: 1.011-1.533; $P = 0.039$) and the TT-CC

genotype of rs11614913-rs895819 was related to an increased GDM risk (OR=3.989; 95% CI: 1.309-12.16; P = 0.015). In addition, the haplotype T-C had a positive interaction with GDM (OR=1.376; 95% CI: 1.075-1.790; P=0.018), especially in the 18.5 ≤ pre-BMI < 24 group (OR=1.403; 95% CI: 1.026-1.921; P=0.034). Moreover, the blood glucose level of the rs895819 CC genotype was significantly higher than that of the TT and TC genotypes (P < 0.05). The TT-CC genotype of rs11614913-rs895819 showed that the blood glucose level was significantly higher than that of the other genotypes.

Discussion: Our findings suggest that miR-27a rs895819 is associated with increased GDM susceptibility and higher blood glucose levels.

KEYWORDS

gestational diabetes mellitus (GDM), miR-196a2, miR-27a, rs11614913, rs895819, case-control study

1 Introduction

Gestational diabetes (GDM) is a common disease in pregnancy that is determined by the first diagnosis of hyperglycemia (1). GDM is harmful to the health of pregnant women and fetuses to a certain extent. For pregnant women, it may increase the incidence of complications, such as pregnancy hypertension, cardiovascular disease and glucose metabolism inhibition (2). For the fetus, there is a risk of premature birth and neonatal hypoglycemia (3). Therefore, to prevent and treat the occurrence of GDM, it is necessary to explore its pathogenesis and risk factors. The pathogenesis of GDM may include impaired insulin secretion and insulin resistance (2). Dietary, environmental and genetic factors contribute to GDM development (4), among which single nucleotide polymorphisms (SNPs) are an important genetic variation factor (5).

MicroRNAs (miRNAs) play a key regulatory role in the metabolic signaling pathway during pregnancy (6), which may influence islet β-cell differentiation and islet development (7). A growing number of studies have shown that SNPs in miRNAs have an impact on their maturation, expression and function. The dysregulation of miRNA expression is associated with cancer, diabetes and cardiovascular disease development (8, 9). It has been reported that miR-196a2 and miR-27a are involved in the regulation of the insulin signaling pathway and have a strong correlation with diabetes mellitus (DM) (10–20). There have been many previous studies on the association of miR-196a2 rs11614913 and miR-27a rs895819 polymorphisms with type 2 diabetes (T2DM) (21–29), but very few studies have investigated the association of these miRNAs with GDM (20).

The Oral Glucose Tolerance Test (OGTT) is currently regarded as the gold standard for diagnosing GDM (30), yet it is a cumbersome process, requiring fasting and multiple blood draws, and is associated with nausea and vomiting, resulting in reduced patient compliance. Additionally, the OGTT is performed between 24-28 weeks of gestation, providing a limited timeframe to implement interventions to improve pregnancy outcomes. Therefore, it is essential to find ways to increase patient compliance and facilitate early detection. In recent years, SNPs have been explored as potential molecular biomarkers for GDM screening (31). While the correlation between miR-27a rs895819, miR-196a2 rs11614913 and gestational diabetes has been less studied, the identification of sensitive and specific biomarkers through the detection of these SNPs may offer potential for GDM risk prediction and intervention strategies.

Therefore, this study evaluated the association between the single SNPs rs11614913 and rs895819, SNP-SNP and GDM risk and further explored the correlation between genotype and blood glucose level. We conducted a Chinese case-control study to assess whether miR-196a2 rs11614913 and miR-27a rs895819 are associated with GDM risk. Further meta-analysis was performed to estimate the relationships between rs11614913 and rs895819 and DM.

2 Materials and methods

2.1 Study subjects

This study protocol was approved by the Ethics Committee of Shunde Women and Children's Hospital of Guangdong Medical University, and subjects for this study were selected through the following criteria: (i) voluntary informed consent; (ii) never diagnosed with diabetes; (iii) Han ethnicity; (iv) age not less than 18 years; (v) no pregnancy complications; and (vi) no glucose-lowering medication. A total of 1002 pregnant Chinese Han women were recruited, including 500 in the GDM group and 502 in the control group. Based on the GDM diagnostic criteria of the International Association of Diabetes and Pregnancy Study Groups (IADPSG), during 24-28 weeks of pregnancy, pregnant women took 75 g glucose for the glucose tolerance test (OGTT), and subjects with at least one glucose level measurement equal to or

above the threshold value (fasting blood glucose level, FBP ≥ 5.1 mmol/L, 1 hour blood glucose level, 1 h-PG ≥ 10.0 mmol/L or 2 hour blood glucose level, 2 h-PG ≥ 8.5 mmol/L) were diagnosed with GDM, while subjects with normoglycemic levels were deemed healthy controls. This study was performed based on the principles of the Declaration of Helsinki.

2.2 Data collection

General clinical information, such as age, ethnicity, height, systolic blood pressure (SBP), diastolic blood pressure (DBP), prepregnancy weight, and parity (primipara or multipara) were gathered. The prepregnancy body mass index (pre-BMI, Kg/m2) was calculated as prepregnancy weight (Kg) divided by the square of the height (m^2). According to BMI, the obesity criteria of Chinese people were divided into the following groups: obesity (≥28 Kg/m^2), overweight (24 Kg/m^2 ≤ BMI <28 Kg/m^2), normal (18.5 Kg/m^2 ≤ BMI <24 Kg/m^2), and underweight (<18.5Kg/m^2).

2.3 SNP genotyping

The QIAamp DNA blood kit (Qiagen, Germany) was used to extract genomic DNA. Genotypes of individual SNPs were detected using the SNPscan method, and the raw data were collected on an ABI3730XL sequencer and analyzed with GeneMapper 4.1 software (Applied Biosystems, USA) (Genesky Technologies Inc., Shanghai, China). The accuracy of genotyping results was ensured by further quality control.

2.4 Statistical analyses

All statistical analyses were performed using SPSS 20.0 software (SPSS, Chicago, IL, USA). Independent sample t test was used for comparison of continuous variables (mean ± standard deviation); discontinuous variables, including Hardy-Weinberg equilibrium (HWE) in the control group, were compared using chi-square tests. After adjusting for potential confounders (including age, pre-BMI, blood pressure, and parity), the association of SNP, SNP-SNP and risk of GDM was assessed by dominance ratio (OR) and 95% confidence interval (CI) using binary logistic regression analysis. One-way ANOVA was used to analyze the correlation between SNP, SNP-SNP and blood glucose levels. The least significant difference (LSD) method was used for multiple comparisons. Bilateral $P < 0.05$ was statistically significant.

2.5 Bioinformatics analyses

Utilized the UCSC database (http://genome.ucsc.edu/) to locate SNPs with a minimum allele frequency of one percent or higher. The effect of variation on RNA folding and the stability of mRNA secondary structure was analyzed using RNAfold Web Servers (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). In addition, the miRWalk (http://mirwalk.umm.uni-heidelberg.de/) online tool was used to predict the GDM-related genes that were coregulated by miR-196a2 and miR-27a.

2.6 Meta-analysis

Different combinations of the terms rs11614913, rs895819, gestational diabetes mellitus, GDM, type 2 diabetes mellitus, T2DM and type 1 diabetes mellitus, T1DM were used to comprehensively search the literature through the PubMed, Chinese National Knowledge Infrastructure and Google Scholar databases with no limitations. The inclusion criteria were case-control or cohort studies that assessed the association of rs11614913 and rs895819 with GDM/T2DM/T1DM with sufficient raw data. Studies that did not meet the diagnostic criteria and studies with data that were not in HWE were excluded. Two authors supervised each other to extract the basic data in the article. The overall and subgroup meta-analysis of five genetic models used the fixed or random effects model according to the level of heterogeneity (32). Publication bias was determined using Egger's and Begg's tests. All meta-analyses were performed using STATA v.16.0 software (Stata Corporation, TX, USA).

3 Results

3.1 General clinical characteristics of the subjects

This case-control study included 500 GDM and 502 healthy controls for whom the genotypes of miR-27a rs895819 and miR-196a2 rs11614913 were detected. Clinical baseline information is listed in Table 1. The mean age, pre-BMI, SBP, DBP, and blood glucose levels were significantly higher in the GDM group than in the control group ($P < 0.05$). Moreover, the parity of the GDM group was significantly different from that of the control group ($P < 0.05$).

3.2 The association of rs11614913 and rs895819 with GDM risk

3.2.1 Overall analysis results

Table 2 shows the results of Hardy-Weinberg equilibrium (HWE) analysis and minor allele frequencies (MAF) for the 2 SNPs in the control group. The results were consistent with HWE ($P > 0.05$). The (unadjusted and adjusted) OR and 95% CI of the correlation between genotype and GDM were estimated in five models (codominant homozygous, codominant heterozygous, dominant, recessive and allele models) for each polymorphism. Before adjustment, the results showed the rs895819 dominant model (CC+TC vs. TT: OR=1.293; 95% CI: 1.008-1.658; $P = 0.043$) and the rs895819 allele model (C vs. T: OR=1.257; 95% CI: 1.032-1.532; $P = 0.023$) associated with increased GDM risk. After adjusting for age, pre-BMI, SBP, DBP, and parity, the results of the

10.3389/fendo.2023.1127336

TABLE 1 Basic and stratified characteristic of participants of the study.

Variables	Cases (%)	Controls (%)	t/x2	*P*
Age, year (mean ± SD)	31.01 ± 4.32	28.66 ± 4.37	-8.56	**<0.001**
			49.2	**<0.001**
<30	192 (38.4)	304 (60.6)		
≥30	308 (61.6)	198 (39.4)		
pre-BMI, kg/m2	21.51 ± 3.10	20.53 ± 2.58	-5.42	**<0.001**
			27.8	**<0.001**
<18.5	67 (13.4)	95 (18.9)		
18.5 ≤ BMI < 24	336 (67.2)	365 (72.7)		
≥24	97 (19.4)	42 (8.3)		
SBP, mmHg	116.69 ± 10.96	114.33 ± 10.18	-3.53	**<0.001**
DBP, mmHg	69.77 ± 7.80	68.23 ± 7.26	-3.23	**0.001**
FBP, mmol/L	4.82 ± 0.64	4.50 ± 0.31	-9.75	**<0.001**
1h-PG, mmol/L	10.17 ± 1.60	7.66 ± 1.27	-26.22	**<0.001**
2h-PG, mmol/L	8.91 ± 1.60	6.69 ± 0.99	-25.85	**<0.001**
Parity (n)			8.88	**0.003**
Primipara	210 (42)	258 (51.4)		
Multipara	290 (58)	244 (48.6)		

pre-BMI, pre-gestational body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; FBP, fasting blood glucose level; 1h-PG, 1 hour blood glucose level; 2h-PG, 2 hour blood glucose level.

rs895819 allele model (C vs. T: OR=1.245; 95% CI: 1.011-1.533; P = 0.039) remained significantly associated with increased GDM risk (Table 3). However, no significant correlation with GDM risk was found for rs11614913 (Table 3).

3.2.2 Stratified analysis results

Subsequently, the association of the 2 SNPs in 5 models with GDM susceptibility was tested using stratified analysis by age or pre-BMI. Notably, for the rs895819 dominant model (CC+TC vs. TT: OR=1.515; 95% CI: 1.053-2.179; P = 0.025), rs895819 codominant heterozygote model (TC vs. TT: OR=1.514; 95% CI: 1.036-2.214; P = 0.032) and rs895819 allele model (C vs. T: OR=1.353; 95% CI: 1.015-1.802; P = 0.039) the results showed a significantly increased GDM risk in subjects younger than 30 years of age. In the 18.5 ≤ pre-BMI <24 group, the results of the rs895819 dominant model (CC+TC vs. TT: OR=1.434; 95% CI: 1.064-1.933; P = 0.018), rs895819 codominant heterozygote model (TC vs. TT: OR=1.402; 95% CI: 1.024-1.918; P = 0.035) and rs895819 allele model (C vs. T: OR=1.335; 95% CI: 1.052-1.693; P = 0.017) showed that rs895819 was significantly related to increased GDM risk; however, after correction, no significant difference was found. In addition, no significant correlation with GDM risk was found for rs11614913 (Supplementary Table 1-4).

3.3 The association between rs11614913-rs895819 and GDM risk

We further investigated the effect of rs11614913-rs895819 interactions. The model included three genotypes and alleles of miRNA polymorphisms. The results after adjusting for age, pre-BMI, SBP, DBP, and parity showed that the TT-CC genotype of miR-196a2 rs11614913 and miR-27a rs895819 was associated with increased GDM risk (OR=3.989; 95% CI: 1.309-12.16; P = 0.015). In addition, the haplotype T-C was significantly associated with increased GDM risk (OR=1.376; 95% CI: 1.075-1.790; P = 0.018) (Table 4), especially in the group with 18.5 ≤ pre-BMI < 24 (OR=1.403; 95% CI: 1.026-1.921; P = 0.034) (Supplementary Tables 5, 6).

TABLE 2 SNPs information and HWE test in the controls.

SNP	Min/Maj	Chr. position	MAF	HWE(P)
rs11614913	C/T	chr12:53991815	0.462	0.411
rs895819	C/T	chr19:13836478	0.247	0.996

HWE, Hardy–Weinberg equilibrium; Min, minor allele; Maj, major allele; MAF, frequency of minor allele.

10.3389/fendo.2023.1127336

TABLE 3 The associations between SNPs and GDM risk in overall subjects.

SNP	Genetic Models	Cases (freq) (n=500)	Controls (freq) (n=502)	Crude OR (95 % CI)	Crude *P*	Adjusted OR (95 % CI)	Adjusted *P*
rs895819	Codominant model						
	TT	252 (0.504)	285 (0.567)	1(ref)		1(ref)	
	TC	204 (0.408)	186 (0.37)	1.240 (0.955-1.611)	0.106	1.233 (0.936-1.622)	0.136
	CC	44 (0.088)	31 (0.061)	1.605 (0.984-2.620)	0.058	1.576 (0.939-2.644)	0.085
	Aelle model						
	T	708 (0.708)	756 (0.752)	1(ref)		1(ref)	
	C	292 (0.292)	248 (0.247)	1.257 (1.032-1.532)	**0.023**	1.245 (1.011-1.533)	**0.039**
	Dominant Model						
	TT	252 (0.504)	285 (0.567)	1(ref)		1(ref)	
	CC+TC	248 (0.496)	217 (0.433)	1.293 (1.008-1.658)	**0.043**	1.281 (0.985-1.665)	0.064
	Recessive Model						
	TC+TT	456 (0.912)	471 (0.939)	1(ref)		1(ref)	
	CC	44 (0.088)	31 (0.061)	1.466 (0.910-2.363)	0.116	1.440 (0.870-2.384)	0.156
rs11614913	Codominant model						
	TT	142 (0.284)	148 (0.294)	1(ref)		1(ref)	
	TC	254 (0.508)	245 (0.488)	1.081 (0.809-1.443)	0.6	1.003 (0.738-1.362)	0.985
	CC	104 (0.208)	109 (0.217)	0.994 (0.698-1.417)	0.975	0.930 (0.639-1.353)	0.704
	Aelle model						
	T	538 (0.538)	541 (0.538)	1(ref)		1(ref)	
	C	462 (0.462)	463 (0.462)	1.003 (0.842-1.196)	0.97	0.968 (0.804-1.165)	0.728
	Dominant Model						
	TT	142 (0.284)	148 (0.294)	1(ref)		1(ref)	
	CC+TC	358 (0.716)	354 (0.706)	1.054 (0.802-1.385)	0.706	0.981 (0.734-1.309)	0.894
	Recessive Model						
	TC+TT	396 (0.792)	393 (0.783)	1(ref)		1(ref)	
	CC	104 (0.208)	109 (0.217)	0.947 (0.700-1.282)	0.724	0.928 (0.673-1.279)	0.648

Adjusted P value calculated by logistic regression with adjustment for age, pre-BMI, SBP,DBP and parity.

3.4 Association between genotype and blood glucose level

The results of the OGTT experiment showed that the 2-hour blood glucose level of the CC genotype of rs895819 was significantly higher than those of the TT and TC genotypes ($P < 0.05$)(Table 5). For the interaction genotype of rs11614913-rs895819, the fasting blood glucose level of the CC-TC genotype was higher than that of TC-TT ($P < 0.05$), and the 1-hour and 2-hour blood glucose levels of the TT-CC genotype were significantly higher than those of the TT-TC, TC-TC and CC-TT genotypes ($P < 0.05$) (Table 6).

3.5 Effects of variants on miRNA secondary structure

The locations of the SNP mutation sites in the studied miRNAs are shown in Figure 1. The analysis of the effect of SNPs on the local miRNA structure showed that the centroid secondary structure in dot-bracket notation with a minimum free energy (MFE) of the T and C rs895819 alleles are -34.3 kcal/mol and -30.4 kcal/mol, respectively, and the size of the miRNA hairpin loop increases when the T allele is replaced by the C allele. (Figures 2A, B). Thermodynamically, the lower the MFE,

10.3389/fendo.2023.1127336

TABLE 4 The associations between combined genotype/aelle and GDM risk.

Genotype combination		Cases (freq)	Controls (freq)	Crude OR(95 % CI)	Crude *P*	Adjusted OR(95 % CI)	Adjusted *P*
rs11614913	rs895819	(n=500)	(n=502)				
TT	TT	65 (0.13)	85(0.169)	1(ref)		1(ref)	
	TC	63(0.126)	58(0.116)	1.420 (0.878-2.298)	0.153	1.414 (0.852-2.347)	0.18
	CC	14(0.028)	5(0.010)	3.661 (1.255-10.69)	**0.018**	3.989 (1.309-12.16)	**0.015**
TC	TT	136(0.272)	139(0.277)	1.279 (0.857-1.909)	0.227	1.165 (0.762-1.781)	0.482
	TC	97(0.194)	85(0.169)	1.492 (0.966-2.305)	0.071	1.455 (0.917-2.307)	0.111
	CC	21(0.042)	21(0.041)	1.307 (0.659-2.596)	0.443	1.268 (0.613-2.622)	0.523
CC	TT	51(0.102)	61(0.121)	1.093 (0.668-1.789)	0.723	1.140 (0.675-1.925)	0.625
	TC	44(0.088)	43(0.085)	1.338 (0.787-2.273)	0.281	1.159 (0.663-2.029)	0.604
	CC	9(0.018)	5(0.010)	2.354 (0.753-7.359)	0.141	1.654 (0.488-5.604)	0.419
Aelle combination		Cases (freq)	Controls (freq)	Crude OR(95 % CI)	Crude *P*	Adjusted OR(95 % CI)	Adjusted *P*
rs11614913	**rs895819**	(2n=1000)	(2n=1004)				
T	T	329(0.329)	367(0.366)	1(ref)		1(ref)	
	C	209(0.209)	174(0.173)	1.340 (1.043-1.721)	**0.022**	1.376 (1.057-1.790)	**0.018**
C	T	379(0.379)	389(0.387)	1.087 (0.885-1.335)	0.427	1.078 (0.868-1.339)	0.495
	C	83(0.083)	74(0.074)	1.251 (0.884-1.770)	0.205	1.116 (0.773-1.610)	0.559

Adjusted P value calculated by logistic regression with adjustment for age, pre-BMI, SBP, DBP and parity.

the more stable the miRNA structure. Thus, these variations may affect the processing of pre-miRNAs.The centroid secondary structure in dot-bracket notation with an MFE of the C and T rs11614913 alleles are −49.9 kcal/mol and -44.3 kcal/mol, respectively (Figures 2C, D). This suggests that the local miRNA structure of the C allele may be more stable than that of the T allele. Figure 3 shows the base pair probabilities of wild-type and mutant-type, suggesting the difference between wild-type and mutant-type.

3.6 Meta-analysis results

The final analysis included 12 studies (including our study): 6 studies related to rs11614913 and GDM/T2DM/T1DM (1/4/1) and 7 studies related to rs895819 and GDM/T2DM (2/5). Table 7 shows the characteristics of the studies. In the overall analysis, no significant associations were found between rs11614913 and rs895819 and DM. In the subgroup meta-analysis, the results of the rs895819 dominant model (CC+TC vs. TT: OR=0.699; 95% CI:

TABLE 5 Relationship between polymorphisms genotype and blood glucose levels.

SNP	Genotype	FBG (mmol/L)	1 h-PG (mmol/L)	2 h-PG (mmol/L)
rs11614913	TT	4.650 ± 0.402	8.919 ± 1.734	7.822 ± 1.552
	TC	4.661 ± 0.426	9.036 ± 1.893	7.927 ± 1.718
	CC	4.714 ± 0.840	9.031 ± 2.189	7.833 ± 2.023
	F	0.912	0.348	0.385
	P	>0.05	>0.05	>0.05
rs895819	TT	4.641 ± 0.440	8.947 ± 1.890	7.764 ± 1.687[a]
	TC	4.704 ± 0.666	9.004 ± 1.998	7.917 ± 1.788[b]
	CC	4.690 ± 0.375	9.387 ± 1.630	8.483 ± 1.791
	F	1.505	1.581	5.31
	P	>0.05	>0.05	**<0.05**

[a]LSD was used to compare the blood glucose levels of three rs895819 genotypes: the difference of 2-hour blood glucose between CC and TT genotypes was statistically significant, P = 0.001.
[b]LSD was used to compare the blood glucose levels of three rs895819 genotypes: the difference of 2-hour blood glucose between CC and TC genotypes was statistically significant, P = 0.014.

10.3389/fendo.2023.1127336

TABLE 6 Relationship between polymorphisms combined genotype and blood glucose levels.

Genotype combination		FBG (mmol/L)	1 h-PG (mmol/L)	2 h-PG (mmol/L)
rs11614913	rs895819			
TT	TT	4.632 ± 0.408	8.733 ± 1.773[b]	7.583 ± 1.538[cd]
	TC	4.678 ± 0.405	8.972 ± 1.654[b]	7.924 ± 1.480[c]
	CC	4.618 ± 0.338	9.992 ± 1.598	8.985 ± 1.586
TC	TT	4.638 ± 0.442[a]	9.092 ± 1.851	7.896 ± 1.715[c]
	TC	4.687 ± 0.421	8.951 ± 1.998[b]	7.928 ± 1.732[c]
	CC	4.698 ± 0.335	9.046 ± 1.706	8.135 ± 1.699
CC	TT	4.660 ± 0.482	8.871 ± 2.106[b]	7.675 ± 1.788[cd]
	TC	4.772 ± 1.170	9.157 ± 2.398	7.886 ± 2.245[c]
	CC	4.770 ± 0.522	9.522 ± 1.249	8.781 ± 2.200
	F	0.697	1.276	2.159
	P	>0.05	>0.05	<0.05

[a] LSD was used to compare the blood glucose levels of nine genotype combinations: the difference of FBG between CC-TC and TC-TT genotype combination was statistically significant, P < 0.05.
[b] LSD was used to compare the blood glucose levels of nine genotype combinations: the difference of 1-hour blood glucose between TT-CC and other genotype combination were statistically significant, P < 0.05.
[c] LSD was used to compare the blood glucose levels of nine genotype combinations: the difference of 2-hour blood glucose between TT-CC and other genotype combination were statistically significant, P < 0.05.
[d] LSD was used to compare the blood glucose levels of nine genotype combinations: the difference of 2-hour blood glucose between CC-CC and other genotype combination were statistically significant, P < 0.05.

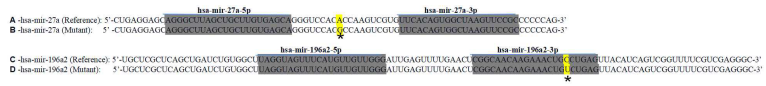


FIGURE 1
RNA precursor sequence and mutation sites (marked with asterisk). A hsa-miR-27a (reference), B hsa-miR-27a (mutant), C hsa-miR-196a2 (reference), D hsa-miR-196a2 (mutant).

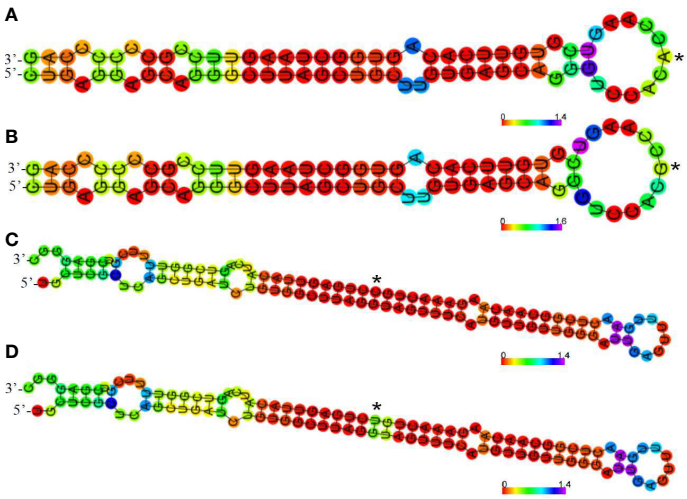


FIGURE 2
Centroid secondary structure of pre-mir-27a and pre-mir-196a2. The size of the miRNA hairpin loop increases when the rs895819 T allele is replaced by the C allele. **(A)** hsa-miR-27a (reference), **(B)** hsa-miR-27a (mutant), **(C)** hsa-miR-196a2 (reference), **(D)** hsa-miR-196a2 (mutant).

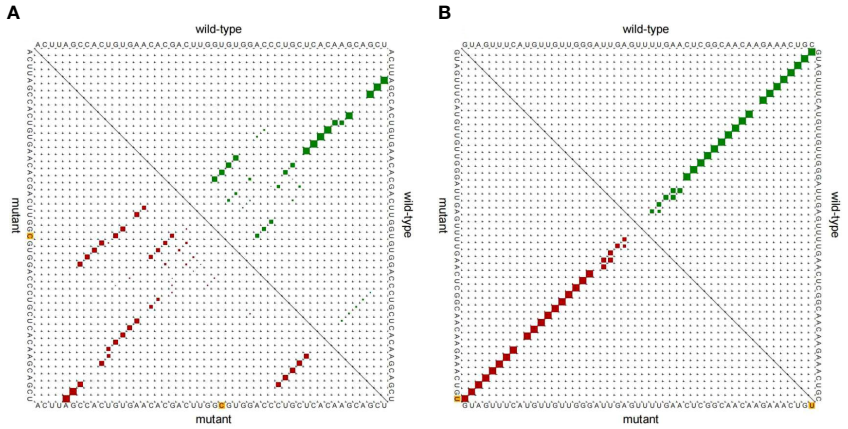


FIGURE 3
Base pair probability of local region. **(A)** miR-27a rs895819 T>C, **(B)** miR-196a2 rs11614913 C>T.

TABLE 7 Characteristics of each study included in the meta-analysis.

SNP							Allele distribution				Genotype distribution						
							Cases (n)		Controls (n)		Cases (n)			Controls (n)			
rs11614913	Athor	Year	Ethnicity	Type	Cases (n)	Controls (n)	C	T	C	T	CC	CT	TT	CC	CT	TT	HWE
	Zeng et al.(Our study)	2023	Asian	GDM	500	502	462	538	463	541	104	254	142	109	245	148	>0.05
	MIR et al.	2022	Caucasian	T2DM	100	100	145	55	165	35	51	43	6	70	25	5	>0.05
	Khan et al.	2021	Caucasian	T2DM	338	236	346	330	333	139	84	178	76	130	73	33	>0.05
	Huang et al.	2021	Asian	T2DM	497	782	413	581	691	873	81	251	165	138	415	229	>0.05
	Ibrahim et al.	2019	Caucasian	T1DM	150	150	175	125	206	94	59	57	34	71	64	15	>0.05
	Buraczynska et al.	2014	Caucasian	T2DM	920	834	1224	616	1001	667	414	396	110	292	417	125	>0.05
rs895819	Athor	Year	Ethnicity	Type	Cases (n)	Controls (n)	T	C	T	C	TT	TC	CC	TT	TC	CC	HWE
	Zeng et al. (Our study)	2023	Asian	GDM	500	502	708	292	756	248	252	204	44	285	186	31	>0.05
	Choi et al.	2022	Asian	T2DM	238	247	317	159	277	217	106	105	27	84	109	54	>0.05
	Ghaedi et al.	2016	Caucasian	T2DM	204	209	301	107	280	138	108	85	11	97	86	26	>0.05
	Wang et al.	2015	Asian	T2DM	995	967	1469	521	1415	519	554	361	80	526	363	78	>0.05
	Li et al.	2015	Asian	T2DM	738	610	1064	412	900	320	371	322	45	330	240	40	>0.05
	Wang et al.	2014	Asian	GDM	837	848	1293	381	1257	439	482	329	26	469	319	60	>0.05
	Ciccacci et al.	2013	Caucasian	T2DM	148	147	247	49	219	75	101	45	2	83	53	11	>0.05

n number, T1DM type 1 diabetes mellitus, T2DM type 2 diabetes mellitus, GDM gestational diabetes mellitus, HWE Hardy–Weinberg equilibrium.

0.518-0.943; $P = 0.019$), rs895819 recessive model (CC vs. TC+TT: OR=0.365; 95% CI: 0.190-0.701; $P = 0.002$), rs895819 codominant homozygous model (CC vs. TT: OR=0.305; 95% CI: 0.156-0.595; $P < 0.001$) and rs895819 allele model (C vs. T: OR=0.667; 95% CI: 0.524-0.848; $P = 0.001$) showed that the tested models were associated with decreased T2DM risk in a Caucasian population (Figure 4). No significant difference was found in other groups (data not shown).

4 Discussion

MiRNAs affect gene expression through posttranscriptional regulation and are involved in many important physiological processes (21). Polymorphisms of miRNAs may affect their maturation, expression and function, which may lead to human disease susceptibility (22). It has been found that miRNA polymorphisms are associated with a variety of cancers, T2DM,

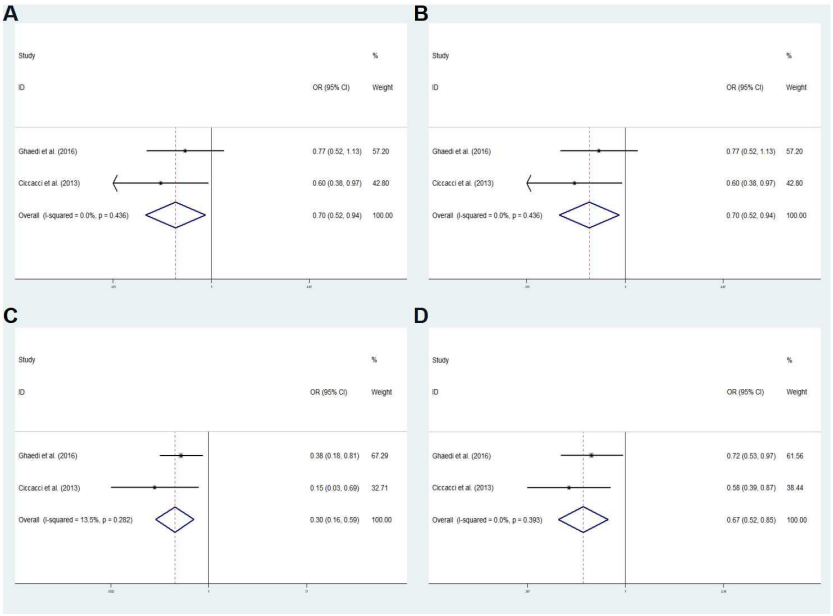


FIGURE 4
Subgroup meta-analysis for the association between miR-27a rs895819 and T2DM susceptibility in a Caucasian population in fixed effects model. **(A)** Dominant model, CC+TC vs. TT. **(B)** Recessive model, CC vs. TC+TT. **(C)** co-dominant homozygous model, CC vs. TT. **(D)** Allele model C vs. T. OR odds ratio, CI, confidence interval; I2: measurement to quantify the degree of heterogeneity in meta-analyses.

GDM and cardiovascular diseases (23–25). In this study, we evaluated the associations between miR-27a rs895819 and miR-196a2 rs11614913 and GDM susceptibility in a Chinese population.

The results showed that the miR-27a rs895819 C allele was associated with increased GDM risk, and a previous study indicated that the miR-27a CC genotype was associated with increased T2DM risk in an overweight Chinese population (18). Zhu et al. showed that miR-27a rs895819 variant genotypes were also associated with an increased risk of T2MD in both the age ≥ 60 years (GG genotype) and male subgroup (AG genotype and dominant model) (26). MiRNA SNPs may contribute to the development of GDM by changing the expression of target genes. The results of this study showed that the rs895819 mutant C allele increased the production of mature miR-27a and suppressed the expression of its target genes compared to the wild T allele (27). Furthermore, human Drosha selectively cleaves RNA hairpins with larger terminal loops (28). Thus, as long as the size of the miRNA loop is altered by mutation or deletion, the maturation process of Drosha is affected. The miR-27a secondary structure analysis found that the size of the miRNA hairpin loop increases when the T allele is replaced by the C allele. This enlargement has been shown to accelerate the maturation of miR-27a, resulting in the upregulation of miR-27a (28). Interestingly, the expression of the miR-27a rs895819 AG and GG genotypes was significantly higher than that of the AA genotype (29). The *peroxisome proliferator-activated receptorγ*(*PPARγ*) is the target gene of miR-27a, and miR-27a rs895819 variants may further negatively regulate the expression of the *PPARγ* gene, it may directly down-regulates the adiponectin (33). Adiponectin deficiency is strongly associated with insulin resistance in pregnancy, Khosrowbeygi et al. showed that the adiponectin level of the GDM group was significantly lower than that of healthy pregnant women (34). Thus miR-27a rs895819 variants is considered to be associated with insulin resistance and diabetes.

However, the results of studies in Caucasians were in contrast to these results. Ciccacci et al. showed that the miR-27a rs895819 G allele played a protective role against T2DM in an Italian study (17), and Ghaedi et al. also found that the miR-27a rs895819 C allele played a protective role against T2DM in an Iranian cohort (16). Our meta-analysis verified the above results. These conflicting results with those of our study may be related to ethnic differences in the study populations. In a recent study of a Korean population, it was found that the G allele in a recessive model and the GG genotype of miR-27a rs895819 were significantly associated with decreased T2DM risk, but the sample sizes of T2DM and healthy controls were only 238 and 247, respectively (15). The results of only one study on GDM showed that the miR-27a rs895819 C allele decreased GDM risk in a Chinese population (20), which is contrary to our results. After summarizing the data of our study and the above study, we conducted a meta-analysis and found no correlation between rs895819 and GDM. Therefore, more research on GDM is especially important.

Moreover, miR-196a2 may regulate the insulin signaling pathway, and miR-196a2 variants are involved in T2DM development (10, 35). Rs11614913 is located in the 3p arm of miR-196a2 (36), which may affect the maturation of pre-miRNAs and target gene binding (37). Previous studies showed that the miR−196a2 rs11614913 T allele and CT genotype were associated with an increased T2DM risk in the Saudi Arabian population. Huang et al. showed that the rs11614913 C allele was significantly associated with decreased T2DM susceptibility in the smoking subgroup (13), but a study of a Pakistani population found an increased association between the miR−196a2 rs11614913 C allele and T2DM risk. Our

results did not find a significant correlation between miR-196a2 rs11614913 and GDM in the Chinese population. The meta-analysis results did not find a significant association between miR-196a2 rs11614913 and DM risk. These contradicting findings may be related to the different sample sizes of studies and ethnic differences.

The combined SNP genotype analysis indicated that the TT-CC genotype of miR-196a2 rs11614913 and miR-27a rs895819 was associated with an increased risk of GDM susceptibility. This detection of the interaction of rs11614913 and rs895819 in GDM was defined as an epistatic influence, which generally explains the absence or underestimation of heritability when only a single SNP is included in a disease susceptibility study (38). According to the results of our research, miR-196a2 rs11614913 probably has no impact on GDM. However, miR-196a2 rs11614913 and miR-27a rs895819 may jointly affect the development of GDM. Remarkably, the combined genotype and haplotype methods have high potential for application in association research (39). In the haplotype results, the allele combination T-C haplotype of miR-196a2 rs11614913 and miR-27a rs895819 was significantly associated with increased GDM risk, especially in the group with 18.5 ≤ pre-BMI <24. The results of the miRWalk database analysis showed that both miR-196a2 and miR-27a can target the Adiponectin gene, which is related to GDM. MiR-196a2 rs11614913 and miR-27a rs895819 variants may negatively regulate Adiponectin gene expression and increase susceptibility to GDM. Therefore, further functional verification is necessary.

The results of correlation analysis between genotype and blood glucose level showed that the 2-h blood glucose level of the miR-27a rs895819 CC genotype was significantly higher than that of the TT and TC genotypes. The 1-h and 2-h blood glucose levels of the TT-CC genotypes of rs11614913 and rs895819 were significantly higher than those of other combinations. Previous studies have shown that miR-27a in cluster C was positively correlated with fasting blood glucose level, which may play a key role in early hyperglycemia and contribute to the development of diabetes (40).

5 Conclusions

In general, our research is the first to confirm that miR-27a rs895819 may contribute to GDM susceptibility in pregnant Chinese women. However, one of the limitations of this study is the limited sample size. In addition, multicenter and further functional studies are needed to gain more insight into the association between rs895819 and GDM. Importantly, future research should verify some selected targets through luciferase analysis and evaluate the regulatory effect of these miRNA mutations on target gene expression.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The study was agreed by the Ethics Committee of Shunde Women and Children's Hospital of Guangdong Medical University (Maternity and Child Healthcare Hospital of Shunde Foshan). The patients/participants provided their written informed consent to participate in this study.

Author contributions

QZ, DZ and NL contributed equally to this study. QZ, JY, WW and FH collected clinical data and samples, QZ, DZ and NL did data analyzes, QZ, DZ, YW and RG wrote the manuscript. FH, RH and RG supervised the whole research. All authors contributed to the article and approved the submitted version.

Funding

Support from the National Natural Science Foundation of China (81873649); Doctoral scientific research Initiate funding project of Shunde Women and Children's Hospital of Guangdong Medical University (Maternity and Child Healthcare Hospital of Shunde Foshan) (2020BSQD007); Youth Talent Project of Shunde Women and Children's Hospital of Guangdong Medical University (Maternity and Child Healthcare Hospital of Shunde Foshan) (2023QNRC023).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1127336/full#supplementary-material

References

1. Buchanan TA, Xiang AH, Page KA. Gestational diabetes mellitus: risks and management during and after pregnancy. *Nat Rev Endocrinol* (2012) 8(11):639–49. doi: 10.1038/nrendo.2012.96

2. Szmuilowicz ED, Josefson JL, Metzger BE. Gestational diabetes mellitus. *Endocrinol Metab Clin North Am* (2019) 48(3):479–93. doi: 10.1016/j.ecl.2019.05.001

3. Wu Y, Liu B, Sun Y, Du Y, Santillan MK, Santillan DA, et al. Association of maternal prepregnancy diabetes and gestational diabetes mellitus with congenital anomalies of the newborn. *Diabetes Care* (2020) 43(12):2983–90. doi: 10.2337/dc20-0261

4. Modzelewski R, Stefanowicz-Rutkowska MM, Matuszewski W, Bandurska-Stankiewicz EM. Gestational diabetes mellitus-recent literature review. *J Clin Med* (2022) 11(19):5736. doi: 10.3390/jcm11195736

5. Chen X, Wang W, Li R, Yu J, Gao L. Association between polymorphisms in microRNAs and susceptibility to diabetes mellitus: A meta-analysis. *Med (Baltimore)* (2019) 98(44):e17519. doi: 10.1097/MD.0000000000017519

6. Cai M, Kolluru GK, Ahmed A. Small molecule, big prospects: MicroRNA in pregnancy and its complications. *J Pregnancy* (2017) 2017:6972732. doi: 10.1155/2017/6972732

7. Poy MN, Spranger M, Stoffel M. microRNAs and the regulation of glucose and lipid metabolism. *Diabetes Obes Metab* (2007) 9 Suppl 2:67–73. doi: 10.1111/j.1463-1326.2007.00775.x

8. Jiménez-Lucena R, Rangel-Zúñiga OA, Alcalá-Díaz JF, López-Moreno J, Roncero-Ramos I, Molina-Abril H, et al. Circulating miRNAs as predictive biomarkers of type 2 diabetes mellitus development in coronary heart disease patients from the CORDIOPREV study. *Mol Ther Nucleic Acids* (2018) 12:146–57. doi: 10.1016/j.omtn.2018.05.002

9. Méndez-Mancilla A, Lima-Rogel V, Toro-Ortíz JC, Escalante-Padrón F, Monsiváis-Urenda AE, Noyola DE, et al. Differential expression profiles of circulating microRNAs in newborns associated to maternal pregestational overweight and obesity. *Pediatr Obes* (2018) 13(3):168–74. doi: 10.1111/ijpo.12247

10. Ibrahim AA, Ramadan A, Wahby AA, Hassan M, Soliman HM, Abdel Hamid TA. Micro-RNA 196a2 expression and miR-196a2 (rs11614913) polymorphism in T1DM: a pilot study. *J Pediatr Endocrinol Metab* (2019) 32(10):1171–9. doi: 10.1515/jpem-2019-0226

11. Mir MM, Mir R, Alghamdi MAA, Wani JI, Elfaki I, Sabah ZU, et al. Potential impact of *GCK*, *MIR-196A-2* and *MIR-423* gene abnormalities on the development and progression of type 2 diabetes mellitus in asir and tabuk regions of Saudi Arabia. *Mol Med Rep* (2022) 25(5):162. doi: 10.3892/mmr.2022.12675

12. Khan MS, Rahman B, Haq TU, Jalil F, Khan BM, Maodaa SN, et al. Deciphering the variants located in the *MIR196A2*, *MIR146A*, and *MIR423* with type-2 diabetes mellitus in Pakistani population. *Genes (Basel)* (2021) 12(5):664. doi: 10.3390/genes12050664

13. Huang Q, Chen H, Xu F, Liu C, Wang Y, Tang W, et al. Relationship of microRNA locus with type 2 diabetes mellitus: a case-control study. *Endocr Connect* (2021) 10(11):1393–402. doi: 10.1530/EC-21-0261

14. Buraczynska M, Zukowski P, Wacinski P, Ksiazek K, Zaluska W. Polymorphism in microRNA-196a2 contributes to the risk of cardiovascular disease in type 2 diabetes patients. *J Diabetes Complications* (2014) 28(5):617–20. doi: 10.1016/j.jdiacomp.2014.05.006

15. Choi Y, Hong SH. Genetic association between miR-27a and miR-449b polymorphisms and susceptibility to diabetes mellitus. *BioMed Rep* (2022) 16(5):37. doi: 10.3892/br.2022.1520

16. Ghaedi H, Tabasinezhad M, Alipoor B, Shokri F, Movafagh A, Mirfakhraie R, et al. The pre-mir-27a variant rs895819 may contribute to type 2 diabetes mellitus susceptibility in an Iranian cohort. *J Endocrinol Invest* (2016) 39(10):1187–93. doi: 10.1007/s40618-016-0499-4

17. Ciccacci C, Di Fusco D, Cacciotti L, Morganti R, D'Amato C, Greco C, et al. MicroRNA genetic variations: association with type 2 diabetes. *Acta Diabetol* (2013) 50 (6):867–72. doi: 10.1007/s00592-013-0469-7

18. Wang TT, Chen YJ, Sun LL, Zhang SJ, Zhou ZY, Qiao H. Affection of single-nucleotide polymorphisms in miR-27a, miR-124a, and miR-146a on susceptibility to type 2 diabetes mellitus in Chinese han people. *Chin Med J (Engl)* (2015) 128(4):533–9. doi: 10.4103/0366-6999.151112

19. Li Y, Zhang Y, Li X, Shi L, Tao W, Shi L, et al. Association study of polymorphisms in miRNAs with T2DM in Chinese population. *Int J Med Sci* (2015) 12(11):875–80. doi: 10.7150/ijms.12954

20. Wang X, Nie M, Li W, Ping F, Ma L, Gao J, et al. Associations of single nucleotide polymorphism in has-miR-27a and has-miR-124a gene with gestational diabetes mellitus. *Basic Clin Med* (2014) 34:1553–7.

21. Cirillo F, Catellani C, Lazzeroni P, Sartori C, Street ME. The role of MicroRNAs in influencing body growth and development. *Horm Res Paediatr* (2020) 93(1):7–15. doi: 10.1159/000504669

22. Lv H, Pei J, Liu H, Wang H, Liu J. A polymorphism site in the pre-miR-34a coding region reduces miR-34a expression and promotes osteosarcoma cell proliferation and migration. *Mol Med Rep* (2014) 10(6):2912–6. doi: 10.3892/mmr.2014.2582

23. Wang J, Li J, Qiu H, Zeng L, Zheng H, Rong X, et al. Association between miRNA-196a2 rs11614913 T>C polymorphism and Kawasaki disease susceptibility in southern Chinese children. *J Clin Lab Anal* (2019) 33(7):e22925. doi: 10.1002/jcla.22925

24. Hu Z, Chen J, Tian T, Zhou X, Gu H, Xu L, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J Clin Invest* (2008) 118(7):2600–8. doi: 10.1172/JCI34934

25. Zhu L, Chu H, Gu D, Ma L, Shi D, Zhong D, et al. A functional polymorphism in miRNA-196a2 is associated with colorectal cancer risk in a Chinese population. *DNA Cell Biol* (2012) 31(3):350–4. doi: 10.1089/dna.2011.1348

26. Zhu Z, Zhang Y, Bai R, Yang R, Shan Z, Ma C, et al. Association of genetic polymorphisms in MicroRNAs with type 2 diabetes mellitus in a Chinese population. *Front Endocrinol (Lausanne)* (2021) 11:587561. doi: 10.3389/fendo.2020.587561

27. Sun Q, Gu H, Zeng Y, Xia Y, Wang Y, Jing Y, et al. Hsa-mir-27a genetic variant contributes to gastric cancer susceptibility through affecting miR-27a and target gene expression. *Cancer Sci* (2010) 101(10):2241–7. doi: 10.1111/j.1349-7006.2010.01667.x

28. Zeng Y, Yi R, Cullen BR. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme drosha. *EMBO J* (2005) 24(1):138–48. doi: 10.1038/sj.emboj.7600491

29. Song B, Yan G, Hao H, Yang B. rs11671784 G/A and rs895819 A/G polymorphisms inversely affect gastric cancer susceptibility and miR-27a expression in a Chinese population. *Med Sci Monit* (2014) 20:2318–26. doi: 10.12659/MSM.892499

30. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* (2010) 33(3):676–82. doi: 10.2337/dc09-1848

31. Dias S, Pheiffer C, Abrahams Y, Rheeder P, Adam S. Molecular biomarkers for gestational diabetes mellitus. *Int J Mol Sci* (2018) 19(10):2926. doi: 10.3390/ijms19102926

32. Zeng Q, Zou D, Wei Y, Ouyang Y, Lao Z, Guo R. Association of vitamin d receptor gene rs739837 polymorphism with type 2 diabetes and gestational diabetes mellitus susceptibility: a systematic review and meta-analysis. *Eur J Med Res* (2022) 27 (1):65. doi: 10.1186/s40001-022-00688-x

33. Kersten S, Desvergne B, Wahli W. Roles of PPARs in health and disease. *Nature* (2000) 405(6785):421–4. doi: 10.1038/35013000

34. Khosrowbeygi A, Rezvanfar MR, Ahmadvand H. Tumor necrosis factor- α, adiponectin and their ratio in gestational diabetes mellitus. *Caspian J Intern Med* (2018) 9(1):71–9. doi: 10.22088/cjim.9.1.71

35. Zhuang GQ, Wang YX. A tiny RNA molecule with a big impact on type 2 diabetes mellitus susceptibility. *BioMed Environ Sci* (2017) 30(11):855–61. doi: 10.3967/bes2017.116

36. Landi D, Gemignani F, Barale R, Landi S. A catalog of polymorphisms falling in microRNA-binding regions of cancer genes. *DNA Cell Biol* (2008) 27(1):35–43. doi: 10.1089/dna.2007.0650

37. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* (2007) 129(7):1401–14. doi: 10.1016/j.cell.2007.04.040

38. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* (2010) 11(6):446–50. doi: 10.1038/nrg2809

39. Ridolfi E, Fenoglio C, Cantoni C, Calvi A, De Riz M, Pietroboni A, et al. Expression and genetic analysis of MicroRNAs involved in multiple sclerosis. *Int J Mol Sci* (2013) 14(3):4375–84. doi: 10.3390/ijms14034375

40. Karolina DS, Tavintharan S, Armugam A, Sepramaniam S, Pek SL, Wong MT, et al. Circulating miRNA profiles in patients with metabolic syndrome. *J Clin Endocrinol Metab* (2012) 97(12):E2271–6. doi: 10.1210/jc.2012-1996

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 23 May 2023
DOI 10.3389/fendo.2023.1154271

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Polyxeni Mantzouratou,
National and Kapodistrian University of Athens Medical School, Greece
Mario García Urena,
University of Copenhagen, Denmark

*CORRESPONDENCE
Bin Wang
colin_iverson@163.com
Xiaoying Zhang
zhangxy6689996@163.com

†These authors have contributed equally to this work

RECEIVED 30 January 2023
ACCEPTED 04 May 2023
PUBLISHED 23 May 2023

CITATION
Wang M, Mei K, Chao C, Di D, Qian Y, Wang B and Zhang X (2023) Rheumatoid arthritis increases the risk of heart failure-current evidence from genome-wide association studies.
Front. Endocrinol. 14:1154271.
doi: 10.3389/fendo.2023.1154271

Rheumatoid arthritis increases the risk of heart failure-current evidence from genome-wide association studies

Min Wang†, Kun Mei†, Ce Chao, Dongmei Di, Yongxiang Qian, Bin Wang* and Xiaoying Zhang*

Department of Cardiothoracic Surgery, The Third Affiliated Hospital of Soochow University, Changzhou, Jiangsu, China

Background: Numerous studies have demonstrated that rheumatoid arthritis (RA) is related to increased incidence of heart failure (HF), but the underlying association remains unclear. In this study, the potential association of RA and HF was clarified using Mendelian randomization analysis.

Methods: Genetic tools for RA, HF, autoimmune disease (AD), and NT-proBNP were acquired from genome-wide studies without population overlap. The inverse variance weighting method was employed for MR analysis. Meanwhile, the results were verified in terms of reliability by using a series of analyses and assessments.

Results: According to MR analysis, its genetic susceptibility to RA may lead to increased risk of heart failure (OR=1.02226, 95%CI [1.005495-1.039304], *P*=0.009067), but RA was not associated with NT-proBNP. In addition, RA was a type of AD, and the genetic susceptibility of AD had a close relation to increased risk of heart failure (OR=1.045157, 95%CI [1.010249-1.081272], *P*=0.010825), while AD was not associated with NT-proBNP. In addition, the MR Steiger test revealed that RA was causal for HF and not the opposite (P = 0.000).

Conclusion: The causal role of RA in HF was explored to recognize the underlying mechanisms of RA and facilitate comprehensive HF evaluation and treatment of RA.

KEYWORDS

rheumatoid arthritis, autoimmune disease, heart failure, NT-proBNP, Mendelian randomization analysis, genome-wide association study

Introduction

Rheumatoid arthritis (RA) is an autoimmune disease with a worldwide lifetime prevalence of 1% (1), and more common in women, which accounts for 75% of all RA cases (2). RA is typically indicated by the presence of autoantibodies, including anti-cyclic citrullinated peptide and rheumatoid factor, years before the disease can be detected (3), and the most common clinical manifestations caused by these autoantibodies are distal joint pain and joint deformity caused by involvement of synovial joints. Current therapies for RA include antirheumatic drugs (DMARDs), anti-tumor necrosis factor-alpha inhibitors (e.g., adalimumab, etanercept, and infliximab) and non-tumor necrosis factor inhibitors (e.g., ababtreotide, rituximab, toximab) (4). If untreated or poorly controlled, it may lead to interrupted physical function and increased mortality owing to increased cardiovascular risk.

Despite progress in the treatment of RA, which achieves disease activity control in most patients, the life expectancy of RA patients remains low due to the complications of cardiovascular diseases (5, 6). It was found that RA patients had a risk of heart failure 1.87% higher than that of the general population (7), and it was not associated with cardiovascular risk factors (8). The incidence of sudden cardiac death of RA patients is twice that of normal controls, and it is secondary to non-ischemic heart disease, ischemic heart disease and arrhythmia (9). Meanwhile, it is shown that the prevalence of non-ischemic heart disease (heart failure) in RA patients is significantly higher than that of ischemic heart disease (10). N-Terminal Pro-Brain Natriuretic Peptide (NT-proBNP) is now established for the diagnosis of heart failure, but new evidence also points to the role of NT-proBNP in diagnosing myocardial ischemia in asymptomatic patients for primary prevention. NT-proBNP has been shown to be elevated in RA, and this elevation is not significantly related to cardiac function (11). Whether RA can directly affect the change of NT-proBNP, the causal relationship remains unknown. It is noteworthy that these observational studies have different sample size and the results are indeed dependent on confounding factors, and the specific mechanism has yet to be clarified.

Confirmation of causality is challenging due to complex confounders of RA and HF risk. The causal relationship of exposure and outcomes without bias was assessed, and the instrumental variables (IVs) were genetic variation in MR analysis (12). In virtue of the unique advantages of IVs, MR analysis is independent from conventional confounding factors, allowing causal inference (13, 14). Genome-wide association studies (GWAS) provide reliable IVs. In this study, MR analysis was performed on two samples to clarify the potential causality of HF risk and genetic susceptibility to RA and AD without interference from side effects of drug or common risk factors, which is critical for prevention and treatment of RA and even AD.

Methods

Study design and data sources

A two-sample MR approach and classical MR analysis were involved in this study. The data related to RA were acquired from a meta-analysis of GWAS, which included 14,361 cases and 42,923 controls. GWAS data for AD (42,202 cases and 17,6590 controls) were acquired online (https://www.finngen.fi/en). For the outcome dataset, single nucleotide polymorphisms (SNPs) for HF were acquired from a meta-analysis of GWAS (47,309 cases and 930,014 controls). The data for NT-proBNP were acquired from GWAS (21,758 samples). Table 1 summarizes demographic profiles involved. The details of the GWAS are provided in Supplementary Table 1.

We performed a two-sample MR study to assess the causality of CVD risk and genetic susceptibility to RA. Herein, SNPs served as IVs (15). An overview of the research design is presented in Figure 1. The entire process satisfied the three main hypotheses of classical MR analysis: 1. exposure is directly affected IVs; 2. IVs had no correlation with confounders; 3. IVs directly impact outcome risk via exposure, instead of other pathways. Additionally, ethical approval was available for all original studies, along with informed consent. Herein, we followed the latest (STROBE-MR) guidelines (16).

Ethical approval

A MR study by using GWAS summary statistics was employed in this study, and ethical approval had been obtained for each

TABLE 1 Instrumental variable assessment and data source.

Traits	Data sources	Sample size (cases/controls)	Ancestry	R^2(%) for RA/AD (Total)	F for RA/AD (Total)
Exposures					
RA	PMID:24390342	14,361/42,923	European		
AD	FinnGen	42,202/17,6590	European		
Outcomes					
HF	PMID:31919418	47,309/930,014	European	0.76/0.008	1442/44
NT-proBNP	PMID:33067605	21,758	European	0.76/0.008	1453/38

$F=R^2(N-K-1)/[K(1-R^2)]$, $R^2 = 2\times(1-EAF)\times EAF\times(b/SD)^2$, among which $SD=SE\times N^{1/2}$, where N refers to the sample size of GWAS, b refers to an effect estimated on adipokines, SE refers to the SD of b, and EAF refers to an effect allele frequency.

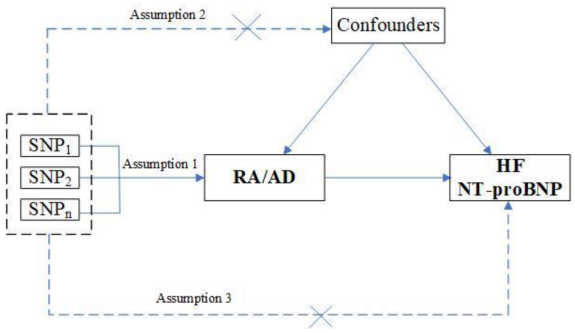


FIGURE 1
Study design flowchart of the Mendelian randomization study. The Mendelian randomization method is based on three hypotheses: 1. the instrumental variables is closely related to exposure; 2. instrumental variables is independent of any confounding factor; 3. instrumental variables affects the results only through exposure but not through other ways.

GWAS. The summary statistics were obtained online (https://www.ebi.ac.uk). All data are accessible and no restriction was set.

Selection of IVs

Genetic variants that are closely related to RA ($P < 5 \times 10^{-8}$) were regarded as instrumental variables. We made sure to include only SNPs that were independent (r2<0.001 in 10.000kb) performing LD-clumping with a European reference panel from 1000G (17). Meanwhile, secondary phenotypes were searched for each SNP in order to exclude potential pleiotropic effects. We did not find SNPs associated with confounders (hypertension, diabetes, obesity, and smoking) in PhenoScanner V2. Specifically, SNPs corresponding to the outcome-related phenotypes ($P < 5 \times 10^{-8}$) were excluded, while other SNPs were kept. After that, variance (R^2) and F-statistics were employed to evaluate the strength of instrumental variables so that weak-tool bias can be avoided (18). Herein, the formula is as follows: $F=R^2(NK-1)/[K(1-R^2)]$, where N denotes the sample number of the chosen GWAS, K denotes the number of SNPs involved, and R^2 denotes the explained variance (cumulative) of the chosen SNPs during exposure. F>10 indicates a strong correlation of exposure and instrumental variables, and the MR analysis results are independent on weak-tool bias.

MR-analysis

All statistical analyses were conducted using R software (version 4.2.0, R Foundation for Statistical Computing), the MR analysis was performed using the "TwoSampleMR" package (version 0.5.6). For each set of IVs, we harmonized exposure and outcome data to ensure the effect sizes for each GWAS were aligned to the same alleles. Similarly, different exposures (e.g., AD) and outcomes (NT-proBNP) were adjusted in a similar way. The inverse variance weighting (IVW) method was dominant in the MR analysis (15). Meanwhile, MR-PRESSO, MR-RAPS, maximum-likelihood, MR-Egger, and median weighting were employed to clarify the causality (18). Different hypotheses about the effectiveness of IVs were made by using each method. Estimation of median weighting is executed if half of IVs are invalid. MR-Egger was used because it corrects for horizontal pleiotropy, despite lower statistical capability. Specifically, the MR-RAPS was responsible for horizontal multiplicity correction by contour scores adjusted, resulting in reduced deviation due to horizontal multiplicity. And the MR-PRESO method could automatically identify and remove outliers (IVW linear regression) to correct the MR estimation (19). The directionality that exposure causes outcome was verified using the MR Steiger test, $P < 0.05$ was regarded as statistically significant. These methods were used to comprehensively investigate causality.

Multivariable Mendelian randomization analysis

Multivariable MR (MVMR) analysis was implemented for significant exposure-outcome pairs identified by univariate MR analysis. Specifically, four confounders, Diabetes (IEU GWAS ID: "ukb-b-10753"), Obesity (IEU GWAS ID: "finn-b-E4_OBESITY"), Hypertension (IEU GWAS ID: "finn-b-I9_HYPTENS") and Smoking (IEU GWAS ID: "ieu-b-142"), were included for MVMR analysis. After combining the GWAS summary level datasets of exposure and the four confounders, it should be ensured that each IV is strongly correlated (P < 5e−8) with at least one or more of the exposure or the three confounders. Then, the SNPs within a window size of 10,000 kb were pruned under the threshold of $r^2 < 0.001$ to mitigate LD. Finally, after excluding palindromic SNPs, outcome-related SNPs (P<0.05), and SNPs not present in outcome GWAS summary data, we used the IVW method to assess causal effects after adjusting for confounders.

Pleiotropy and heterogeneity analyses

As primary analysis we applied the Causal Analysis Using Summary Effect Estimates (CAUSE) approach, which has been demonstrated to outperform other established methods to detect causal relationships in the presence of pleiotropy, CAUSE avoids more false positives induced by correlated horizontal pleiotropy than other methods (20). In this case, CAUSE analysis was conducted to determine whether the relationship between RA and HF was causal (causal model) or induced by correlated horizontal pleiotropy (shared model). When P<0.05 it means that the causal model is preferred over the shared one, indicated that the causal relationship between RA and HF is real and not a false positive due to the correlated horizontal pleiotropy. A series of methods were used for sensitivity analysis in this study. First, the heterogeneity of different SNP estimates was evaluated by the Cochran's Q test. If $P > 0.05$, no heterogeneity was indicated. Although the random-effects model could be used, the fixed-effect IVW method was dominant. Second, the horizontal pleiotropy of IVs was investigated by using the MR-Egger intercept method (21). Average of the horizontal pleiotropic effect was estimated based on the intercept across SNPs

in the MR-Egger test, and the IVW estimate might be biased if $P < 0.05$. Third, a single SNP could generate the results was verified by using the leave-one-out sensitivity test. Leave-one-out method shown how the IVW causal effect when remove each variant from the analysis. This allows to detect heterogeneity since if the IVW changes drastically, that means that a variant is contributing way more than the others. Importantly, this is not always a sign of pleiotropy, but always a sign of heterogeneity in the data being analyzed. Fourth, the presence of pleiotropy was directly detected by generating funnel and forest plots. "Two-Sample MR", "MR-PRESSO", "CAUSE" and "mr.raps" packages in R software were used for statistical analysis.

Results

Causality of genetic susceptibility to RA and AD on the risk for HF

As shown in Table 2, results obtained by the IVW method indicated that RA was related to increased risk of HF. As observed, the prevalence of HF in RA cases was 1.014-fold that of the control group (95% CI [1.0009-1.0281], OR=1.014, P=0.036) (Supplementary Figure 1), and increase of the OR of AD by one unit leads to increased HF risk (95% CI [1.010-1.081], OR=1.045, P=0.011) (Supplementary Figure 3). MR analysis of RA and HF indicated that the results of the Weighted median analyses were highly consistent with those obtained by the IVW method. In the strict CAUSE, the causal model was shown to be a better fit than the sharing model (95% CI [2.461-2.823], OR=2.642, p = 1.8e-30), indicating a causal association between RA and HF. More supporting statistics were listed in Supplementary Table 2. MR analysis of AD and HF showed that the results of the MR Egger analyses were highly consistent with those obtained by the IVW method. The causal assumption of RA or AD and HF was verified via the MR Steiger test, and the result showed RA or AD influence on HF was the correct causal direction (P = 0.000). The details of the MR Steiger test are provided in Supplementary Table 3.

Causality of the risk for NT-proBNP and genetic susceptibility to RA and AD

As shown in Table 3, the prevalence of NT-proBNP (β=-0.0114, SE =0.0150, P=0.4467) in the RA group was not significantly different from that of the control group (Supplementary Figure 2). The results listed were consistent with those obtained by the IVW method. Meanwhile, no significant association was observed between AD and NT-proBNP risk (β=0.0722, SE =0.0265, P=0.7851) (Supplementary Figure 4). It was also confirmed by analyses listed in the table.

Results of multivariable Mendelian randomization analysis

As shown in Table 4, We performed an MVMR analysis to assess the causal effect of RA on HF after adjusting for four confounding factors (diabetes, obesity, hypertension and smoking). MVMR analysis identified that all of these four confounders were taken into account, the causal relationship between RA and HF was not obvious (OR = 1.022968, 95% CI [0.9994881-1.047000], P = 0.055266). indicating that no significant direct causal effect was detected for RA on HF risk, while jointly modeling diabetes, obesity, hypertension and smoking.

Analysis of horizontal pleiotropy and heterogeneity

As shown in Table 5, a series of methods were employed for MR analysis regarding the correlation of RA, AD and HF to determine the presence of significant horizontal pleiotropy and heterogeneity in the present study. First, the P-value was > 0.05 in the heterogeneity test, demonstrating that SNPs had negligible heterogeneity (Table 5). The fixed-effect IVW method was dominant in this MR analysis. The "leave-one-out" sensitivity analysis demonstrated that IVs involved in the present study had

TABLE 2 MR estimates of RA and AD on the risk for HF.

Disease	Methods	SNPs(n)	OR	95%CI	*P*-value
RA	MR Egger	112	1.006100	0.983752-1.028956	0.596722
	Weighted median	112	1.006559	0.983870-1.029771	0.574114
	IVW	112	1.014421	1.000941-1.028084	0.035929
	Simple mode	112	1.060100	1.012280-1.110178	0.014707
	Weighted mode	112	1.009812	0.988696-1.031380	0.367098
AD	MR Egger	39	1.074839	1.014617-1.138636	0.01899
	Weighted median	39	1.032231	0.992336-1.07373	0.114699
	IVW	39	1.045157	1.010249-1.081272	0.010825
	Simple mode	39	0.982959	0.904143-1.068645	0.689153
	Weighted mode	39	1.037358	1.000383-1.075699	0.054899

TABLE 3 MR estimates of RA and AD on the risk for NT-proBNP.

Disease	Methods	SNPs(n)	β	SE	*P*-value
RA	IVW	114	-0.011421	0.015009	0.446705
	Weighted median	114	0.014745	0.026135	0.572620
	MR Egger	114	-0.014908	0.027667	0.591073
	Weighted mode	114	0.009568	0.026682	0.720572
	Simple mode	114	0.0176783	0.052438	0.736644
AD	IVW	52	0.007217	0.026466	0.785090
	Weighted median	52	0.037362	0.037515	0.276857
	MR Egger	52	0.041910	0.045491	0.361601
	Weighted mode	52	0.0373620	0.036093	0.305781
	Simple mode	52	-0.041060	0.589407	0.589407

TABLE 4 MVMR analysis for assessing the causal effect of RA on HF.

Exposure	SNPs	OR	95% CI	P-value	F-statistic
RA	38	1.022968	0.9994881-1.047000	5.526636e-02	37.79063
Diabetes mellitus	37	1.657085	0.7339741-3.741182	2.241354e-01	3.369213
Obesity	2	1.062153	0.9828610-1.147841	1.276904e-01	10.22393
Hypertension	28	1.190533	1.1231576-1.261950	4.423231e-09	15.66378
Smoking	17	1.133324	1.0477772-1.225855	1.774866e-03	25.55431

negligible impact on such results (Supplementary Figures 5-8), and the funnel plot illustrates an asymmetric distribution of single IVs (Supplementary Figure 9), suggesting that the causality was not likely to be affected by potential bias. The MR Steiger test indicated that there was no reverse causality (Supplementary Table 3).

Discussion

In the present study, MR analysis was first performed to investigate the potential causal relationship of HF risk and the susceptibility to RA. RA is the most common autoimmune disease. The causal relationship of HF risk and AD was thus evaluated by MR analysis. The results showed that the genetic susceptibility to RA and AD was correlated with an increase in HF risk. The MR Steiger test further showed that there was no evidence of reverse causality in our study. The limited evidence from MR analysis supported the potential causal relationship between RA and AD and HF risk.

HF is a cardiovascular syndrome associated with RA and also contributes to the incidence and death of RA (22). In the population-based RA cohort, the incidence of HF was about twice

TABLE 5 Heterogeneity and pleiotropy test of RA and AD from HF and NT-proBNP GWAS.

Exposure	Outcomes	Pleiotropy test			Heterogeneity test					
		MR-Egger			MR-Egger			Inverse-variance weighted		
		Intercept	SE	*P*	Q	Q_df	Q_*pval*	Q	Q_df	Q_*pval*
RA										
	HF	0.001384	0.001547	0.372925	109.7300	110	0.489326	110.530	111	0.494726
	NT-proBNP	0.000523	0.003477	0.880793	117.3168	112	0.346731	117.3404	113	0.370953
AD										
	HF	-0.0039	0.0033	0.2475	60.5797	37	0.0086	62.8402	38	0.0068
	NT-proBNP	-0.0046	0.0049	0.3532	45.4325	47	0.5376	46.3118	48	0.5422

the incidence in the general population (22, 23). As a complex clinical syndrome, HF involves a variety of potential risk factors and causes, among which hypertension and ischemic heart disease are most common (24). Clinically, HF is classified based on the left ventricular ejection fraction (LVEF): 1. Reduced LVEF is defined as ≤40%, i.e. those with a significant reduction in LV systolic function. This is designated as HFrEF. 2. Patients with a LVEF between 41% and 49% have mildly reduced LV systolic function, i.e. HFmrEF. 3. Those with symptoms and signs of HF, with evidence of structural and/or functional cardiac abnormalities and/or raised natriuretic peptides (NPs), and with an LVEF ≥50%, have HFpEF (25). Along with aggravated population aging, the prevalence of HFpEF has been rising in recent years. A recent retrospective study found that 64% of the RA patients are combined with HFpEF (26). HFpEF is more common among RA patients compared to the general HF population without RA (27). A follow-up survey using cardiac ultrasonography showed that the development of subclinical changes in the diastolic function among RA patients was more rapid within 5 years compared to the general population (28). Mantel et al. compared the incidence of 10,000 Swedish patients with ischemic and non-ischemic heart failure. They reported a rapid increase in the HF risk following the onset of FA and a close connection with high disease activity (10). RA patients were related to a higher incidence of HF and IHD throughout the course of observation, and RA was more significantly correlated with the high HF risk (29). Recent advances in the treatment of RA have decreased the incidence of cardiovascular diseases in RA patients, but these patients are still at a higher risk for IHD. Besides, the HF risk increases as the duration and severity of RA increase (10, 30). Nicola et al. proved that compared to the non-RA population, the risk of congestive heart failure was significantly increased in the RA population, with an odds ratio of 1.87 during the 30-year follow-up (22). Similarly, according to Wolfe and Michaud, HF was common among RA patients (22). Michael J Ahlers et al. performed a retrospective case-control study of 9,889 RA patients and 9,889 controls without autoimmune diseases, who were matched for age, gender, and race. It was found that the HF risk was increased by 21% in RA patients and such an increase was irrelevant to the conventional cardiovascular risk factors (26). This estimate agrees with the increased HF risk associated with RA at the Swedish and Danish National Patient Registry (10, 29). Nevertheless, the above reported increase in the HF risk was smaller than that reported by Nicola in the presence of RA, which was 87% (22). Recently, some scholars reported that among RA patients diagnosed in Denmark from 1978 to 2008, RA was associated with an increase in HF-related admissions (31). The above evidence has indicated that RA does increase the risk of HF. Four main factors have been identified as contributors of a higher HF risk in RA patients (32): 1. Conventional cardiovascular risk factors, including smoking, dyslipidemia, hypertension, obesity and diabetes, which usually exist concurrently with the risk factors for RA; 2. The use of glucocorticoids and non-steroidal anti-inflammatory drugs will increase the HF risk; 3. The presence of anti-citrulline peptide antibodies and rheumatoid factors in RA patients was an independent risk factor for HF; 4. An increase in the RA disease activity alongside a continuous cardiovascular impact of systemic inflammation is another primary risk factor for HF.

However, the increased prevalence of hypertension and IHD in RA patients may not fully explain the higher HF risk in RA patients (24). A previous study showed that a significant increase in the mortality of HF among RA patients might be related to coronary artery disease (CAD) (22). Other research showed that RA is a typical chronic inflammatory disease and related to an increase in the HF risk. The latter, however, is uncorrelated with the conventional cardiovascular risk factors (including CAD) (10, 29, 33). The HF phenotype in RA patients is different from that in non-RA patients. The former usually presents with diastolic dysfunction, hypotension and high ejection fraction. Thus, RA and non-RA patients may vary in the mechanism of myocardial injury (27, 34). The newly diagnosed RA patients were associated with a significant increase in the incidence of HF events five years before the diagnosis, although few of them presented with typical features of cardiovascular risks, including hypertension and hypercholesterolemia. These facts suggest that CVD is not only a late complication of RA (35). RA-related inflammation may be a critical factor for the progression to HF. The HF risk may be even increased in an absence of IHD risk if the patients have RA-related inflammation. It has been reported that the risk of non-ischemic heart failure is increased at an early stage and closely connected with the severity of RA (10). In another study, the SLE/RA inpatients were analyzed, and the prevalence of HF in the population was 16.4%. Besides, the likelihood of HF in RA patients was significantly lower than that in SLE (36). The above results proved from another perspective that RA-related HF is not caused by shared risk factors alone, since SLE and HF also share some common risk factors. PARK E et al. found that an increase in HF risk in RA patients might not be explained by IHD alone. Non-ischemic HF is related to the severity of RA, implying that RA-related factors and autoimmune process are related to the risk of the HF phenotype above (37).

In the present study, the causal relationship between RA and NT-proBNP was analyzed, but the result was negative. Recently, Baniaamam et al. conducted a prospective study of 51 RA patients, where echocardiography and baseline tests were performed on those with moderate to high disease activity, along with an assessment after six months of treatment with anti-tumor necrosis factor. Although the NT-proBNP level was decreased by 23% after six months of treatment, no adverse effect on the cardiac function was observed (38). The above results suggest that the RA-related impact on cardiac function is not manifested as changes in NT-proBNP. However, controversy continues over the predictive performance of HF-related biomarkers, such as B-type natriuretic peptide (BNP) or NT-proBNP, for cardiac injury. Some authors believe that these factors are sensitive, non-invasive predictors for subclinical CVD and are all-cause mortality predictors independent of conventional risk factors for CV (39). Evidence has shown that an increased NT-proBNP level in RA patients is related to

inflammatory markers (40). However, some researchers did not prove the relationship between the NTproBNP level and left ventricular function in RA patients (41, 42), which also agreed with ours findings.

The relationship between RA and NT-proBNP is complex. In this study, there was no causal relationship between RA and serum NT-proBNP level. In the study of Armstrong et al. although researchers observed an increase in the median NT-proBNP level in the RA group, the increase in NT-proBNP level was significantly correlated with DAS28 and age, and had no direct correlation with RA itself (43). In addition, NT-proBNP may play an indispensable role in regulating the immune system and endocrine system (44–46), including the aging process of individuals, etc (47). These findings all reveal that NT-proBNP levels increase with age, so we speculate that the increased NT-proBNP levels in RA patients may be related to accelerated aging, rather than causally related to the disease itself. However, studies have shown that accelerated aging only explains 16% of the increase in BNP in RA patients (48). Therefore, the increase of BNP in RA patients is largely due to other unknown causes.

DMARDs and TNF-α inhibitors are usually prescribed as standard treatments for RA (42). TNF-α inhibitors are effective for controlling the activity and progression of RA. However, their risks in increasing incidence and deaths of cardiovascular diseases remain disputable, particularly RA patients already with a higher risk for cardiovascular complications (49). One study indicated that a higher dose of TNF-α inhibitors may cause HF deterioration and shortened life span (50). According to a randomized placebo-controlled clinical trial, TNF-α inhibitors did not have a considerable efficacy when used to treat symptomatic HF patients (51). Danish scholars performed a follow-up of RA patients that lasted for over 20 years, and it was found that the biological treatments for RA did not change the risks of IHD and HF (29). According to another study, the dose of glucocorticoids and TNF inhibitors was adjusted in the multivariate regression analysis, and it was found that the increased risk of HF in RA patients was independent of these drugs (31).

Inflammation is considered as a critical mechanism for the development of HF, especially HFpEF (52). Both ESR and CRP were correlated with increased risk of HF in RA patients (10). Evidence from the Mayo Clinic suggests that a higher level of inflammatory markers is related to a higher risk of HF (53). It has been found that an increase in the inflammatory activity related to the pathogenesis of RA may have myocardial effects, leading to HF shortly after RA diagnosis. In sepsis, TNF-α and other cytokines were related to the reduction in myocardial contractility after *in vitro* exposure for ≥10 min (54). Cardiomyocytes may also respond to inflammatory stimuli and express chemokines, cytokines, and cell adhesion molecules, leading to leukocyte recruitment and reduced cardiomyocyte contractility (55). Inflammation can also induce endothelial dysfunction, myocardial hypertrophy and fibrosis, which further results in HF (56). The incidence of HFpEF is also higher in other diseases related to chronic inflammation, such as obesity, diabetes and chronic kidney disease. It is implied that an increase in circulating proinflammatory cytokines in RA patients may be a critical factor in the pathogenesis of HF (57). Interestingly, those with the highest level of C-reactive protein (CRP) are also faced with the highest risk for HF, which highlights the role of inflammation in the pathogenesis. After stratified based on HF subtypes, the CRP level was higher in HFpEF than in HFrEF, indicating that inflammation might be a more important risk factor for HFpEF in RA (26).

RA is a chronic autoimmune inflammatory disease. Our study proved that RA was related to a higher risk for HF. To verify the results, MR analysis was performed, and a potential causal relationship of HF risk and the genetic susceptibility to AD was indicated. This finding coincided with our expectations. Another recent study showed that as an autoimmune disease, SLE was related to a higher risk of venous thromboembolism, ischemic cerebral infarction, and HF (58). Some researchers also performed MR analysis for this purpose, and it was found that RA was correlated with a higher risk of angina, hypertension, arrhythmia, and coronary heart disease (59). Others reported a correlation between MS and the risk of CAD, myocardial infarction, HF, and cerebral stroke (60). All the results above are consistent with our findings.

The clinical diagnosis and treatment of AD and HF should be carefully evaluated, considering the causal relationship of HF risk and the genetic susceptibility for RA and AD. In fact, rheumatologists have become increasingly aware of the relationship between CVD and RA. In the European Society of Cardiology guideline, RA is considered as an independent cardiovascular risk factor (61). The European League Against Rheumatism (EULAR) has published official advice for monitoring CV risk in RA patients (62). It is suggested that the CVD risk score should be multiplied by 1.5 in RA patients. Such a correction may improve the estimate of the cardiac risk in these patients. Therefore, earlier preventive tests and medication treatment are recommended if necessary.

Advantages and limitation

A recent report involved MR analysis of the genetic susceptibility for cardiovascular risks. So far, the causal relationship between CVD risk and SLE and other autoimmune diseases has been analyzed, but few studies have been devoted to the potential relationship of HF risk and RA through MR analysis. We first performed MR analysis on RA and even AD and HF risk to identify any causal relationship. Secondly, large-scale GWAS was employed to collect more comprehensive genetic data in RA and HF, thereby avoiding the influence of conventional confounding factors and eliminating the potential of reverse causality. Lastly, consistent results were obtained through several repeat analyses, and an absence of biases was verified by the heterogeneity and pleiotropy analyses.

However, our study had some limitations. Firstly, pleiotropy was analyzed using multiple methods, but potential multiplicity

might still exist. Secondly, we reported a lower OR value, compared with other studies, and more studies are needed to further document the clinical significance of this OR value. Thirdly, the F- statistics of obesity in MVMR analysis is lower than 10, which may cause a certain bias in the statistical results of MVMR, and the interpretation of the results should be very cautious.

Summary

In conclusion, our study found the first evidence supporting the potential causal relationship of HF risk and RA and AD, which facilitates further investigation into the pathogenesis of RA and AD and comprehensive assessment of the RA-related HF and the associated treatments. Further studies are required to reduce the incidence and mortality of RA-related HF.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

MW and KM designed the study and drafted the article. CC and BW conducted data acquisition. DD, YQ and XZ performed data analysis and manuscript revision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Funding from Young Talent Development plan of Changzhou Health Commission (CZQM2020034, CZQM2020004). Young talents Science and technology project of Changzhou Health Commission (QN201913). The National Natural Science Fund (81701584).

Acknowledgments

We thank all the participants and researchers for their participation in this MR study. The IEU Open GWAS project and European Bioinformatics Institute GWAS Catalog provide summary data for the analyses.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1154271/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Mendelian randomization analysis of RA and the risk of HF.

SUPPLEMENTARY FIGURE 2
Mendelian randomization analysis of RA and the risk of NT-proBNP.

SUPPLEMENTARY FIGURE 3
Mendelian randomization analysis of AD and the risk of HF.

SUPPLEMENTARY FIGURE 4
Mendelian randomization analysis of AD and the risk of NT-proBNP.

SUPPLEMENTARY FIGURE 5
The MR "leave-one-out" sensitivity analysis of RA on HF.

SUPPLEMENTARY FIGURE 6
The MR "leave-one-out" sensitivity analysis of RA on NT-proBNP.

SUPPLEMENTARY FIGURE 7
The MR "leave-one-out" sensitivity analysis of AD on HF.

SUPPLEMENTARY FIGURE 8
The MR "leave-one-out" sensitivity analysis of AD on NT-proBNP.

SUPPLEMENTARY FIGURE 9
Funnel plots of RA/AD with HF/BNP. The X-axis represents odds ratio (OR), and the Y-axis represents standard error (SE). (A)RA to HF. (B) AD to HF. (C) RA to NT-proBNP. (D)AD to NT-proBNP.

SUPPLEMENTARY TABLE 1
SNPs used to analyze the causal relationship between RA and HF, RA and NT-proBNP, AD and HF, AD and NT-proBNP.

References

1. Wasserman A. Rheumatoid arthritis: common questions about diagnosis and management. *Am Fam Physician* (2018) 97(7):455–62.

2. Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. *Autoimmun Rev* (2003) 2(3):119–25. doi: 10.1016/S1568-9972(03)00006-5

3. Nielen MM, Van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MH, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum* (2004) 50(2):380–6. doi: 10.1002/art.20018

4. Singh JA, Furst DE, Bharat A, Curtis JR, Kavanaugh AF, Kremer JM, et al. 2012 update of the 2008 American college of rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res (Hoboken)* (2012) 64(5):625–39. doi: 10.1002/acr.21641

5. Bandyopadhyay D, Banerjee U, Hajra A, Chakraborty S, Amgai B, Ghosh RK, et al. Trends of cardiac complications in patients with rheumatoid arthritis: analysis of the united states national inpatient sample; 2005-2014. *Curr Probl Cardiol* (2021) 46 (3):100455. doi: 10.1016/j.cpcardiol.2019.100455

6. Semb AG, Ikdahl E, Wibetoe G, Crowson C, Rollefstad S. Atherosclerotic cardiovascular disease prevention in rheumatoid arthritis. *Nat Rev Rheumatol* (2020) 16(7):361–79. doi: 10.1038/s41584-020-0428-y

7. Avina-Zubieta JA, Thomas J, Sadatsafavi M, Lehman AJ, Lacaille D. Risk of incident cardiovascular events in patients with rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis* (2012) 71(9):1524–9. doi: 10.1136/annrheumdis-2011-200726

8. Masoud S, Lim PB, Kitas GD, Panoulas V. Sudden cardiac death in patients with rheumatoid arthritis. *World J Cardiol* (2017) 9(7):562–73. doi: 10.4330/wjc.v9.i7.562

9. Blyszczuk P, Szekanecz Z. Pathogenesis of ischaemic and non-ischaemic heart diseases in rheumatoid arthritis. *RMD Open* (2020) 6(1):e001032. doi: 10.1136/rmdopen-2019-001032

10. Mantel A, Holmqvist M, Andersson DC, Lund LH, Askling J. Association between rheumatoid arthritis and risk of ischemic and nonischemic heart failure. *J Am Coll Cardiol* (2017) 69(10):1275–85. doi: 10.1016/j.jacc.2016.12.033

11. George J, Mackle G, Manoharan A, Khan F, Struthers AD. High BNP levels in rheumatoid arthritis are related to inflammation but not to left ventricular abnormalities: a prospective case-control study. *Int J Cardiol* (2014) 172(1):e116–8. doi: 10.1016/j.ijcard.2013.12.119

12. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* (2003) 32(1):1–22. 10.1093/ije/dyg070

13. Nattel S. Canadian Journal of cardiology January 2013: genetics and more. *Can J Cardiol* (2013) 29(1):1–2. doi: 10.1016/j.cjca.2012.11.015

14. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent developments in mendelian randomization studies. *Curr Epidemiol Rep* (2017) 4 (4):330–45. doi: 10.1007/s40471-017-0128-6

15. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* (2008) 27(8):1133–63. doi: 10.1002/sim.3034

16. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: the STROBE-MR statement. *JAMA* (2021) 326 (16):1614–21. doi: 10.1001/jama.2021.18236

17. Pistis G, Porcu E, Vrieze SI, Sidore C, Steri M, Danjou F, et al. Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *Eur J Hum Genet* (2015) 23(7):975–83. doi: 10.1038/ejhg.2014.216

18. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-egger regression: the role of the I2 statistic. *Int J Epidemiol* (2016) 45 (6):1961–74. doi: 10.1093/ije/dyw220

19. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* (2018) 50(5):693–8. doi: 10.1038/s41588-018-0099-7

20. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet* (2020) 52(7):740–7. doi: 10.1038/s41588-020-0631-4

21. 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(2):512–25. doi: 10.1093/ije/dyv080

22. Nicola PJ, Maradit-Kremers H, Roger VL, Jacobsen SJ, Crowson CS, Ballman KV, et al. The risk of congestive heart failure in rheumatoid arthritis: a population-based study over 46 years. *Arthritis Rheum* (2005) 52(2):412–20. doi: 10.1002/art.20855

23. Wolfe F, Michaud K. Heart failure in rheumatoid arthritis: rates, predictors, and the effect of anti-tumor necrosis factor therapy. *Am J Med* (2004) 116(5):305–11. doi: 10.1016/j.amjmed.2003.09.039

24. Crowson CS, Nicola PJ, Kremers HM, O'Fallon WM, Therneau TM, Jacobsen SJ, et al. How much of the increased incidence of heart failure in rheumatoid arthritis is attributable to traditional cardiovascular risk factors and ischemic heart disease? *Arthritis Rheum* (2005) 52(10):3039–44. doi: 10.1002/art.21349

25. Task Force M, Mcdonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the European society of cardiology (ESC). with the special contribution of the heart failure association (HFA) of the ESC. *Eur J Heart Fail* (2022) 24(1):4–131. doi: 10.1002/ejhf.2333

26. Ahlers MJ, Lowery BD, Farber-Eger E, Wang TJ, Bradham W, Ormseth MJ, et al. Heart failure risk associated with rheumatoid arthritis-related chronic inflammation. *J Am Heart Assoc* (2020) 9(10):e014661. doi: 10.1161/JAHA.119.014661

27. Davis JM3rd, Roger VL, Crowson CS, Kremers HM, Therneau TM, Gabriel SE. The presentation and outcome of heart failure in patients with rheumatoid arthritis differs from that in the general population. *Arthritis Rheum* (2008) 58(9):2603–11. doi: 10.1002/art.23798

28. Davis JM3rd, Lin G, Oh JK, Crowson CS, Achenbach SJ, Therneau TM, et al. Five-year changes in cardiac structure and function in patients with rheumatoid arthritis compared with the general population. *Int J Cardiol* (2017) 240:379–85. doi: 10.1016/j.ijcard.2017.03.108

29. Logstrup BB, Ellingsen T, Pedersen AB, Kjaersgaard A, Bøtker HE, Maeng M. Development of heart failure in patients with rheumatoid arthritis: a Danish population-based study. *Eur J Clin Invest* (2018) 48(5):e12915. doi: 10.1111/eci.12915

30. Kao AH, Krishnaswami S, Cunningham A, Edmundowicz D, Morel PA, Kuller LH, et al. Subclinical coronary artery calcification and relationship to disease duration in women with rheumatoid arthritis. *J Rheumatol* (2008) 35(1):61–9.

31. Khalid U, Egeberg A, Ahlehoff O, Lane D, Gislason GH, Lip GYH, et al. Incident heart failure in patients with rheumatoid arthritis: a nationwide cohort study. *J Am Heart Assoc* (2018) 7(2):e007227. doi: 10.1161/JAHA.117.007227

32. Nair S, Singh Kahlon S, Sikandar R, Peddemul A, Tejovath S, Hassan D, et al. Tumor necrosis factor-alpha inhibitors and cardiovascular risk in rheumatoid arthritis: a systematic review. *Cureus* (2022) 14(6):e26430. doi: 10.7759/cureus.26430

33. Schattner A. Patients with new-onset rheumatoid arthritis had increased risk for ischemic and nonischemic heart failure. *Ann Intern Med* (2017) 167(2):JC8. doi: 10.7326/ACPJC-2017-167-2-008

34. Giles JT, Fert-Bober J, Park JK, Bingham CO, Andrade F, Fox-Talbot K, et al. Myocardial citrullination in rheumatoid arthritis: a correlative histopathologic study. *Arthritis Res Ther* (2012) 14(1):R39. doi: 10.1186/ar3752

35. Nikiphorou E, De Lusignan S, Mallen CD, Khavandi K, Bedarida G, Buckley CD, et al. Cardiovascular risk factors and outcomes in early rheumatoid arthritis: a population-based study. *Heart* (2020) 106(20):1566–72. doi: 10.1136/heartjnl-2019-316193

36. Chang CM, Lin JR, Fu TC. Associations between sarcopenia, heart failure and myocardial infarction in patients with systemic lupus erythematosus and rheumatoid arthritis. *Front Med (Lausanne)* (2022) 9:882911. doi: 10.3389/fmed.2022.882911

37. Park E, Griffin J, Bathon JM. Myocardial dysfunction and heart failure in rheumatoid arthritis. *Arthritis Rheumatol* (2022) 74(2):184–99. doi: 10.1002/art.41979

38. Baniaamam M, Handoko ML, Agca R, Heslinga SC, Konings TC, van Halm VP, et al. The effect of anti-TNF therapy on cardiac function in rheumatoid arthritis: an observational study. *J Clin Med* (2020) 9(10):3145. doi: 10.3390/jcm9103145

39. Solus J, Chung CP, Oeser A, Avalos I, Gebretsadik T, Shintani A, et al. Amino-terminal fragment of the prohormone brain-type natriuretic peptide in rheumatoid arthritis. *Arthritis Rheum* (2008) 58(9):2662–9. doi: 10.1002/art.23796

40. Provan SA, Semb AG, Hisdal J, Stranden E, Agewall S, Dagfinrud H, et al. Remission is the goal for cardiovascular risk management in patients with rheumatoid arthritis: a cross-sectional comparative study. *Ann Rheum Dis* (2011) 70(5):812–7. doi: 10.1136/ard.2010.141523

41. Lazurova I, Tomas L. Cardiac impairment in rheumatoid arthritis and influence of anti-TNFalpha treatment. *Clin Rev Allergy Immunol* (2017) 52(3):323–32. doi: 10.1007/s12016-016-8566-3

42. Tomas L, Lazurova I, Oetterova M, Pundová L, Petrášová D, Studenčan M. Left ventricular morphology and function in patients with rheumatoid arthritis. *Wien Klin Wochenschr* (2013) 125(9-10):233–8. doi: 10.1007/s00508-013-0349-8

43. Armstrong DJ, Gardiner PV, O'kane MJ. Rheumatoid arthritis patients with active disease and no history of cardiac pathology have higher brain natriuretic peptide (BNP) levels than patients with inactive disease or healthy control subjects. *Ulster Med J* (2010) 79(2):82–4.

44. Casserly BP, Sears EH, Gartman EJ. The role of natriuretic peptides in inflammation and immunity. *Recent Pat Inflammation Allergy Drug Discovery* (2010) 4(2):90–104. doi: 10.2174/187221310791163125

45. Shaw SM, Fildes JE, Puchalka CM, Basith M, Yonan N, Williams SG, et al. BNP directly immunoregulates the innate immune system of cardiac transplant recipients *in vitro*. *Transpl Immunol* (2009) 20(3):199–202. doi: 10.1016/j.trim.2008.08.010

46. Omland T, Hagve TA. Natriuretic peptides: physiologic and analytic considerations. *Heart Fail Clin* (2009) 5(4):471–87. doi: 10.1016/j.hfc.2009.04.005

47. Paganelli R, Di Iorio A, Cherubini A, Lauretani F, Mussi C, Volpato S, et al. Frailty of older age: the role of the endocrine–immune interaction. *Curr Pharm Des* (2006) 12(24):3147–59. doi: 10.2174/138161206777947533

48. Crowson CS, Liang KP, Therneau TM, Kremers HM, Gabriel SE. Could accelerated aging explain the excess mortality in patients with seropositive rheumatoid arthritis? *Arthritis Rheum* (2010) 62(2):378–82. doi: 10.1002/art.27194

49. Feldman AM, Combes A, Wagner D, Kadakomi T, Kubota T, Li YY, et al. The role of tumor necrosis factor in the pathophysiology of heart failure. *J Am Coll Cardiol* (2000) 35(3):537–44. doi: 10.1016/S0735-1097(99)00600-2

50. Kotyla PJ. Bimodal function of anti-TNF treatment: shall we be concerned about anti-TNF treatment in patients with rheumatoid arthritis and heart failure? *Int J Mol Sci* (2018) 19(6):1739. doi: 10.3390/ijms19061739

51. Sarzi-Puttini P, Atzeni F, Shoenfeld Y, Ferraccioli G. TNF-alpha, rheumatoid arthritis, and heart failure: a rheumatological dilemma. *Autoimmun Rev* (2005) 4 (3):153–61. doi: 10.1016/j.autrev.2004.09.004

52. Paulus WJ, Tschope C. A novel paradigm for heart failure with preserved ejection fraction: comorbidities drive myocardial dysfunction and remodeling through coronary microvascular endothelial inflammation. *J Am Coll Cardiol* (2013) 62(4):263–71. doi: 10.1016/j.jacc.2013.02.092

53. Maradit-Kremers H, Nicola PJ, Crowson CS, Ballman KV, Jacobsen SJ, Roger VL, et al. Raised erythrocyte sedimentation rate signals heart failure in patients with rheumatoid arthritis. *Ann Rheum Dis* (2007) 66(1):76–80. doi: 10.1136/ard.2006.053710

54. Krishnagopalan S, Kumar A, Parrillo JE, Kumar A. Myocardial dysfunction in the patient with sepsis. *Curr Opin Crit Care* (2002) 8(5):376–88. doi: 10.1097/00075198-200210000-00003

55. Marchant DJ, Boyd JH, Lin DC, Granville DJ, Garmaroudi FS, McManus BM, et al. Inflammation in myocardial diseases. *Circ Res* (2012) 110(1):126–44. doi: 10.1161/CIRCRESAHA.111.243170

56. Lim SL, Lam CS, Segers VF, Brutsaert DL, De Keulenaer GW. Cardiac endothelium-myocyte interaction: clinical opportunities for new heart failure therapies regardless of ejection fraction. *Eur Heart J* (2015) 36(31):2050–60. doi: 10.1093/eurheartj/ehv132

57. Patel RB, Shah SJ. Drug targets for heart failure with preserved ejection fraction: a mechanistic approach and review of contemporary clinical trials. *Annu Rev Pharmacol Toxicol* (2019) 59:41–63. doi: 10.1146/annurev-pharmtox-010818-021136

58. Gao N, Kong M, Li X, Wei D, Zhu X, Hong Z, et al. Systemic lupus erythematosus and cardiovascular disease: a mendelian randomization study. *Front Immunol* (2022) 13:908831. doi: 10.3389/fimmu.2022.908831

59. Qiu S, Li M, Jin S, Lu H, Hu Y. Rheumatoid arthritis and cardio-cerebrovascular disease: a mendelian randomization study. *Front Genet* (2021) 12:745224. doi: 10.3389/fgene.2021.745224

60. Yang F, Hu T, He K, Ying J, Cui H. Multiple sclerosis and the risk of cardiovascular diseases: a mendelian randomization study. *Front Immunol* (2022) 13:861885. doi: 10.3389/fimmu.2022.861885

61. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: the sixth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European association for cardiovascular prevention & rehabilitation (EACPR). *Eur Heart J* (2016) 37(29):2315–81. doi: 10.1093/eurheartj/ehw106

62. Agca R, Heslinga SC, Rollefstad S, Heslinga M, McInnes IB, Peters MJ, et al. EULAR recommendations for cardiovascular disease risk management in patients with rheumatoid arthritis and other forms of inflammatory joint disorders: 2015/2016 update. *Ann Rheum Dis* (2017) 76(1):17–28. doi: 10.1136/annrheumdis-2016-209775

frontiers | Frontiers in Endocrinology

TYPE Review
PUBLISHED 30 May 2023
DOI 10.3389/fendo.2023.1163001

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Inês Cebola,
Imperial College London, United Kingdom
Valeriya Lyssenko,
University of Bergen, Norway

*CORRESPONDENCE
Niina Sandholm
niina.sandholm@helsinki.fi
Per-Henrik Groop
per-henrik.groop@helsinki.fi

RECEIVED 10 February 2023
ACCEPTED 10 May 2023
PUBLISHED 30 May 2023

CITATION
Sandholm N, Dahlström EH and Groop P-H (2023) Genetic and epigenetic background of diabetic kidney disease. *Front. Endocrinol.* 14:1163001. doi: 10.3389/fendo.2023.1163001

Genetic and epigenetic background of diabetic kidney disease

Niina Sandholm[1,2,3]*, Emma H. Dahlström[1,2,3] and Per-Henrik Groop[1,2,3,4]*

[1]Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland, [2]Department of Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland, [3]Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [4]Department of Diabetes, Central Clinical School, Monash University, Melbourne, VIC, Australia

Diabetic kidney disease (DKD) is a severe diabetic complication that affects up to half of the individuals with diabetes. Elevated blood glucose levels are a key underlying cause of DKD, but DKD is a complex multifactorial disease, which takes years to develop. Family studies have shown that inherited factors also contribute to the risk of the disease. During the last decade, genome-wide association studies (GWASs) have emerged as a powerful tool to identify genetic risk factors for DKD. In recent years, the GWASs have acquired larger number of participants, leading to increased statistical power to detect more genetic risk factors. In addition, whole-exome and whole-genome sequencing studies are emerging, aiming to identify rare genetic risk factors for DKD, as well as epigenome-wide association studies, investigating DNA methylation in relation to DKD. This article aims to review the identified genetic and epigenetic risk factors for DKD.

KEYWORDS

diabetic kidney disease, kidney failure, GWAS, genome sequencing, exome sequencing, epigenetics, epigenome-wide association study, EWAS

1 Introduction

A total of 537 million people worldwide have diabetes (1), characterized by elevated blood glucose. Despite treatment, which aims to normalize the blood glucose concentrations, diabetes can lead to micro- and macrovascular organ damage through various molecular pathways, including increased reactive oxygen species, which further affect the downstream pathways such as the polyol pathway flux, advanced glycation end-product formation and activation, protein kinase C activation, and the hexosamine pathway flux (2). These microvascular complications include diabetic kidney disease (DKD), sight-threatening proliferative diabetic retinopathy, and diabetic neuropathy. The complications reduce the quality of life, increase mortality, and account for the majority of the health care costs for diabetes (3, 4). Together, 30%–50% of individuals with diabetes develop DKD (5–7).

Individuals with type 1 diabetes (T1D) develop diabetes early in life and, thus, have a particularly high lifetime risk of developing complications. In up to 20% of individuals with T1D, DKD leads to kidney failure requiring dialysis or kidney transplantation (8). Because of the improvements in the management and treatment of both diabetes and its complications (9), the 25-year cumulative incidence of DKD has halved in those diagnosed in the 1980s compared to those diagnosed in the 1970s. However, there was no further improvement in the later cohorts, and 36% of individuals with severe DKD still progressed to kidney failure within 15 years (6). DKD also substantially increases the risk of CVD, and as many as 40% of individuals with T1D and DKD develop CVD by the age of 40 (10).

DKD is characterized by urinary albumin excretion and gradually decreasing renal function, measured or estimated as glomerular filtration rate (eGFR). Urinary albumin excretion can be classified as normal or mildly increased, moderately, or severely increased albuminuria; the two latter ones are also called micro- and macroalbuminuria. The classical view has been that albuminuria represents an earlier sign of DKD, followed by reduced eGFR and eventually kidney failure, but a substantial proportion of individuals with DKD may present with reduced kidney function even without albuminuria (11). On the tissue level, DKD is characterized by glomerular and tubular basement membrane thickening, mesangial expansion, glomerulosclerosis, podocyte effacement, and, ultimately, nephron loss (12). It is of note, however, that kidney biopsies are rarely taken for diagnostic purposes. Therefore, any chronic kidney disease (CKD) in an individual with diabetes is *a priori* considered as DKD, irrespective of the underlying pathophysiology (11). Lack of a biopsy proof is less of a problem in T1D because most of the individuals with T1D and DKD have histologically true diabetic nephropathy.

DKD is a complex multifactorial disease in which both genetic and environmental risk factors contribute to the development and progression of the disease. However, the exact molecular mechanisms leading to DKD remain poorly understood. Apart from albuminuria and eGFR, no other biomarkers are yet in clinical use for monitoring disease progression or identification of individuals at risk, and only a few treatment options exist for the prevention of DKD, especially in individuals with T1D. To address these issues, genetic studies aim to identify the underlying molecular mechanisms leading to DKD. Here, we review the genetic factors that have been identified for DKD, mainly based on genome-wide association studies (GWASs) performed within the latest decade and summarize the main findings from epigenetic studies—being the potential dynamic link between genes and the environment—investigating the DNA methylation changes associated with DKD.

2 Heritability of DKD

Three decades ago, family studies reported clustering of DKD in siblings with T1D, suggesting an inherited component of the disease (13–17). More recently, a genome-wide estimation of the narrow-sense DKD heritability—the proportion of phenotypic variance explained by additive genetic factors—based on unrelated individuals with T1D reported 24%–42% heritability of DKD, depending on the phenotype definition. The heritability estimates were as high as 59% when adjusted for sex, diabetes duration and age at diabetes diagnosis, and with a tendency to higher heritability estimates for the more severe definitions (18). Similar analyses in individuals with T2D suggested only 8%–25% heritability for DKD, potentially reflecting more heterogeneous mechanisms leading to DKD in T2D in addition to a more important contribution of environmental factors (19, 20). Indeed, a sub-analysis of individuals with T2D from the Action to Control Cardiovascular Risk in Diabetes trial suggested that the gene–treatment interaction explains a large part of the phenotypic variance in microalbuminuria. Nevertheless, the heritability estimates for albuminuria and eGFR both in T1D and T2D range between 7% and 75% (19, 21–25).

3 Common genetic variants associated with DKD

3.1 Early genetic studies for DKD

The early genetic studies on DKD utilized various microsatellite markers and single-nucleotide polymorphisms (SNPs) for family-based linkage studies to identify chromosomal regions co-segregating with DKD. One of the strongest linkage peaks with a logarithm of odds (LOD) score of 3.1 was obtained in a candidate gene study of the *AGTR1* on chromosome 3q (26), and many genome-wide linkage scans reported a suggestive linkage peak on the extended 3q21-q29 region (27–31). Subsequent fine-mapping efforts of candidate genes on the 3q region, comparing the allele frequencies of tens or hundreds of SNPs in unrelated DKD cases and controls, suggested, e.g., *ADIPOQ* (32) and *NCK1* (33) to be involved in DKD. A linkage analysis in Turkish families with T2D and DKD identified a strong linkage peak on chr18q22.3–23 (LOD score = 6.1) (34), subsequently fine-mapped to a polymorphism in the *CNDP1* gene associated with both DKD and serum carnosinase concentrations (35).

In addition to the positional candidates, biological candidate gene studies were performed on the basis of information and hypotheses of the underlying biology. However, the results were mostly inconclusive, with limited statistical evidence due to the small sample number, lenient statistical threshold, and lack of external replication (36). The findings with the strongest statistical evidence include variants on the promoter region of the *EPO* gene encoding for erythropoietin [rs1617640, p-value = 2.7×10^{-11} (37)], as well as in the *SLC19A3* gene encoding for a high-affinity thiamine (vitamin B) transporter [rs12694743, $p = 2.30 \times 10^{-8}$ (38)], both associated with a combined phenotype of kidney failure and diabetic retinopathy.

3.2 Genome-wide association studies on DKD

To overcome the limitations of the candidate gene studies, the first GWASs covering hundreds of thousands of SNPs were pursued

nearly two decades ago, identifying genetic risk factors for both T2D (39–41) and T1D (42). The GWASs have since identified thousands of genetic loci affecting common complex diseases, supporting the multifactorial genetic background and the common disease/common variant (CDCV) hypothesis that suggests that common genetic factors significantly contribute to the risk of common diseases and traits (43). Because of the burden of multiple testing of hundreds of thousands, or even millions of genetic variants, only associations reaching the stringent threshold of a p-value $< 5 \times 10^{-8}$ are considered genome-wide significant. The GWASs on DKD have to date identified 41 loci genome-wide significantly associated with various case-control definitions of DKD, as detailed in Table 1.

3.2.1 Genome-wide association studies on DKD in type 1 diabetes

One of the first GWASs on DKD included 1,705 individuals with T1D from the Genetics of Kidneys in Diabetes (GoKinD) collection and suggested multiple putative susceptibility loci, including a variant in the *FRMD3* gene suggestively associated with DKD (p-value = 5.0×10^{-7}) (54) and replicated by some of the subsequent studies (54, 55). Re-analysis of the data, including imputed variants, suggested additional loci, including *SORBS1* (56); variants in the same gene were also supported by a later GWAS including 1,462 additional individuals with T1D, but the association was attenuated in the replication (57).

The first GWAS meta-analysis on DKD combining data across multiple studies was undertaken by the Genetics of Nephropathy, an International Effort consortium. The GWAS meta-analysis discovery stage included 6,691 participants of European ancestry and with T1D from the GoKinD US, the Finnish Diabetic Nephropathy (FinnDiane) Study, and from the All Ireland-Warren 3-Genetics of Kidneys in Diabetes UK and Republic of Ireland (UK-ROI) Collection. The combined meta-analysis with 11,847 participants with T1D resulted in two loci, an intronic variant rs7583877 in *AFF3*, and an intergenic rs12437854 between in the *RGMA* and *MCTP2* genes associated with kidney failure in T1D with a p-value $< 5 \times 10^{-8}$. Furthermore, the authors reported a suggestive association for rs7588550 in the *ERBB4* gene associated with DKD (p-value = 2.1×10^{-7}). *In vitro* analyses on a renal epithelial cell line suggested that *AFF3* influences the transforming growth factor–β1 (TGF-β1)–induced fibrotic responses (44).

Of note, nearly 90% of the GWAS findings are located on non-coding regions and are enriched for gene regulatory regions, rather than changing the protein amino acid sequence and structure (58, 59). The associated genetic variant does not necessarily affect the gene expression of the underlying or the closest gene, and, thus, a common challenge in GWAS is to identify the target gene of the non-coding regulatory variants. With large expression quantitative trait locus (eQTL) databases that are now available, one can link the genotypes to gene expression levels. On the basis of eQTL data from whole blood in the eQTLGen.org database, the rs7583877 variant in the *AFF3* gene is indeed associated with *AFF3* gene expression (p-value = 2.9×10^{-19}) (60).

In the same consortium, an analysis stratified by gender identified a variant between the *SP3* and *CDCA7* genes associated with kidney failure in women (rs4972593, p-value = 3.9×10^{-8}) (45). Multiple estrogen-responsive elements were predicted near rs4972593, and the *SP3* gene showed higher expression in kidney glomeruli in women (45). Furthermore, the Sp3 transcription factor directly interacts with the estrogen receptor-α (61) and regulates kidney-related genes such as *TGFBI*, *CD2AP*, and *VEGFA*, supporting its role in kidney failure in women with T1D.

The largest GWAS on DKD in T1D to date was performed by the Diabetic Nephropathy Collaborative Research Initiative (DNCRI) consortium, including up to 19,406 individuals with T1D and of European ancestry from 17 cohorts. The analysis comprised 10 different case-control definitions for DKD, based on either albuminuria, eGFR, or both. Altogether, 16 loci reached a p-value $< 5 \times 10^{-8}$, with the strongest association for a common missense mutation rs55703767 (Asp326Tyr) in the collagen type IV alpha 3 chain (*COL4A3*) gene, associated with a 21% lower risk of DKD (p-value = 5.3×10^{-12}) (49). The gene encodes a major structural component of the glomerular basement membrane (GBM). In kidney biopsies of the Renin Angiotensin System Study (RASS) study participants with T1D and normal AER, the carriers of the protective variant had thinner GBM (49). The variant effect was dependent on glycemia, as the association at rs55703767 was observed only among individuals with $HbA_{1c} \geq 7.5\%$ in the HbA_{1c}-stratified sub-analysis of 4,321 FinnDiane participants with longitudinal HbA_{1c} measurements. Similarly, in the Diabetes Control and Complications Trial (DCCT), followed by the Epidemiology of Diabetes Interventions and Complications (DCCT-EDIC) study, the rs55703767 effect on DKD was stronger among those recruited in the secondary cohort and randomized to conventional treatment and therefore had higher HbA_{1c}. Thus, the *COL4A3* rs55703767 association with DKD seems specific to diabetes and amplified by poor glucose control (49). The lead loci in the DNCRI meta-analysis also included other collagen-related findings: association with microalbuminuria for the rs116772905 variant in the *DDR1* gene encoding the epithelial discoidin domain-containing receptor 1, which binds collagens including type IV collagen; and gene aggregate analysis found variants in the *COL20A1* gene associated with severe CKD.

3.2.2 Genome-wide association studies on DKD in type 2 diabetes

One of the first GWASs on DKD among individuals with T2D and the first transethnic meta-analysis of DKD included 4,909 individuals with T2D from the Family Investigation of Nephropathy and Diabetes (FIND) consortium in the discovery cohort and, altogether, 13,736 individuals in the final meta-analysis (including 6,229 non-diabetic controls). The analysis identified rs12523822 near the *SCAF8* and *CNKSR3* genes associated with a 43% lower risk of DKD in American Indians (p-value = 5.7×10^{-9}) and with directionally consistent results across the ethnic groups (46). *CNKSR3* is a direct mineralocorticoid receptor target gene highly expressed in the renal cortical collecting ducts. The gene is involved in the transepithelial sodium transport and is upregulated in response to physiologic aldosterone concentrations (62). Clinically, renin-angiotensin-aldosterone system blockade is the

10.3389/fendo.2023.1163001

TABLE 1 Variants genome-wide significantly (p-value < 5×10^{-8}) associated with DKD.

SNP	Reported gene	Diabetes population	Phenotype	N cases vs. controls	*P*-value	EA	NEA	OR	Refs
rs7583877	*AFF3*	T1D	ESKD	1,786 vs. 8,718	1.2×10^{-8}	C	T	1.29	(44)
rs12437854	*RGMA/MCTP2*	T1D	ESKD	1,786 vs. 8,718	2.0×10^{-9}	G	T	1.8	(44)
rs4972593	*SP3/CDCA7*	Women with T1D	ESKD	688 vs. 2,009	3.9×10^{-8}	A	T	1.81	(45)
rs12523822	*SCAF8/CNKSR3*	T1D + T2D [a]	DKD	5,226 vs. 8,510	1.3×10^{-8}	G	C	0.73	(46)
rs56094641	*FTO*	T2D	DKD	4,022 vs. 6,980	7.7×10^{-10}	G	A	1.23	(47)
rs9942471	*GABRR1*	T2D	Microalbuminuria	1,989 vs. 2,238	4.5×10^{-8}	A	C	1.25	(19)
rs72858591	*RND3/RBM43*	T2D cases vs. non-diabetic controls	ESKD	3,432 vs. 6,977	4.5×10^{-8}	C	T	1.42	(48)
rs58627064	*SLITRK3*	T2D cases vs. non-diabetic controls	ESKD	3,432 vs. 6,977	6.8×10^{-10}	T	G	1.62	(48)
rs142563193	*ENPP7*	T2D cases vs. non-diabetic controls	ESKD	3,432 vs. 6,977	1.2×10^{-8}	A	G	0.74	(48)
rs142671759	*ENPP7*	T2D cases vs. non-diabetic controls	ESKD	3,432 vs. 6,977	5.5×10^{-9}	C	T	2.26	(48)
rs4807299	*GNG7*	T2D cases vs. non-diabetic controls	ESKD	3,432 vs. 6,977	3.2×10^{-8}	A	C	1.67	(48)
rs9622363	*APOL1*	T2D cases vs. non-diabetic controls	ESKD	3,432 vs. 6,977	1.4×10^{-10}	A	G	0.77	(48)
rs75029938	*GRAMD3*	T2D excluding APOL1 carriers [b]	ESKD	2,768 vs. 6,059	2.0×10^{-9}	T	C	1.89	(48)
rs17577888	*MGAT4C*	T2D excluding APOL1 carriers [b]	ESKD	2,768 vs. 6,059	3.9×10^{-8}	T	G	0.67	(48)
rs55703767	*COL4A3*	T1D	DKD	4,948 vs. 12,076	5.3×10^{-12}	T	G	0.79	(49)
rs12615970	*COLEC11*	T1D	CKD	4,266 vs. 14,838	9.4×10^{-9}	G	A	0.76	(49)
rs142823282	*TAMM41*	T1D	Microalbuminuria	2,477 vs. 12,113	1.1×10^{-11}	G	A	6.75	(49)
rs145681168	*HAND2-AS1*	T1D	Microalbuminuria	2,477 vs. 12,113	5.4×10^{-9}	G	A	5.53	(49)
rs118124843	*DDR1*	T1D	Microalbuminuria	2,477 vs. 12,113	3.4×10^{-8}	T	C	3.78	(49)
rs77273076	*MBLAC1*	T1D	Microalbuminuria	2,477 vs. 12,113	1.0×10^{-8}	T	C	9.12	(49)
rs551191707	*PRNCR1*	T1D	ESKD vs. macroalbuminuria	2,187 vs. 2,725	4.4×10^{-8}	CA	C	1.7	(49)
rs144434404	*BMP7*	T1D	Microalbuminuria	2,477 vs. 12,113	4.7×10^{-9}	T	C	6.75	(49)
rs115061173	*LINC01266*	T1D	ESKD	2,187 vs. 12,101	4.1×10^{-8}	A	T	9.39	(49)
rs116216059	*STAC*	T1D	ESKD	2,187 vs. 17,216	1.4×10^{-8}	A	C	8.76	(49)
rs191449639	*MUC7*	T1D	DKD	4,948 vs. 12,076	1.3×10^{-8}	A	T	32.5	(49)
rs149641852	*SNCAIP*	T1D	CKD extreme	2,235 vs. 14,993	1.4×10^{-8}	T	G	9.03	(49)
rs183937294	*PLEKHA7*	T1D	Microalbuminuria	2,477 vs. 12,113	1.7×10^{-8}	G	T	17.3	(49)
rs61983410	*STXBP6*	T1D	Microalbuminuria	2,477 vs. 12,113	3.1×10^{-8}-	T	C	0.79	(49)
rs113554206	*PAPLN*	T1D	Macroalbuminuria	2,751 vs. 12,124	8.5×10^{-9}	A	G	4.62	(49)
rs185299109	*LINC00470/METTL4*	T1D	CKD	4,266 vs. 14,838	1.3×10^{-8}	T	C	20.7	(49)
rs72763500	*NID1*	T2D	DKD	11,327 vs. 7,513	2.6×10^{-8}	C	T	0.79	(50)

(Continued)

TABLE 1 Continued

SNP	Reported gene	Diabetes population	Phenotype	N cases vs. controls	*P*-value	EA	NEA	OR	Refs
rs12917707	*UMOD*	T2D	DKD	11,327 vs. 7,513	4.5×10^{-8}	T	G	0.86	(20, 50)
rs538044833 [c]	*CCSER1*	T1D	CKD	727 vs. 3,962	2.8×10^{-8}	C	T	3.0	(51)
rs72831309	*TENM2*	T1D + T2D	CKD + DKD	4,122 vs. 13,972	9.8×10^{-9}	A	G	2.08	(52)
rs55703767	*COL4A3*	T1D + T2D	DKD	6,705 vs. 15,430	3.6×10^{-11}	T	G	0.86	(52)
rs141560952	*DIS3L2*	Any diabetes vs. healthy controls	CKD	1,194 vs. 9,568	3.6×10^{-9}	AGGG	A	192.6	(53)
rs425827	*KRT6B*	Any diabetes vs. healthy controls	CKD	1,194 vs. 9,568	2.7×10^{-9}	A	T	5.31	(53)
rs73038008	*PLD1*	T1D or T2D	DKD [d]	1,973 vs. 5,734	1.7×10^{-8}	C	T	2.55	(20)
rs77924615	*PDILT/UMOD*	T1D or T2D	DKD [d]	1,973 vs. 5,734	7.8×10^{-9}	A	G	0.75	(20)
rs75733846	*WSCD2*	T2D	ESRD [e]	121 vs. 4,197	3.7×10^{-8}	T	C	7.16	(20)
rs559427701	*SETDB2*	T2D	ESRD [e]	121 vs. 4,197	4.0×10^{-9}	A	C	11.36	(20)
rs62202699	*LOC105372639*	T2D	Microalbuminuria [f]	702 vs. 2,210	4.3×10^{-9}	T	C	2.97	(20)

SNP: Variant rs-identifier. EA: Effect allele. NEA: non-effect allele. OR: odds ratio. Refs: If multiple references are given, then the data in other columns for the same locus are taken from the first listed reference.
[a]Not all controls had diabetes.
[b]Controls did not have diabetes.
[c]Identified as underlying a linkage peak for DKD.
[d]CKD/DKD in self-reported, primary care, hospital, or death records.
[e]Dialysis or a rise of serum creatinine to 3.3 mg/dl (292 µmol/L).
[f]UACR ≥3.4 mg/mmol.

main therapy for individuals with DKD and many other kidney diseases (63, 64). It is of note that the Finerenone in Reducing Kidney Failure and Disease Progression in Diabetic Kidney Disease (FIDELIO-DKD) trial with the non-steroidal mineralocorticoid-receptor-antagonist finerenone on top of standard of care showed cardio- and renoprotection in albuminuric individuals with T2D (65).

As end-stage kidney disease (ESKD) is disproportionately affecting African Americans (AAs), a subsequent FIND study GWAS focused on AAs and was extended to 3,432 T2D-ESKD cases and 6,977 non-diabetic non-nephropathy controls ($N = 10,409$), followed by a discrimination analysis in 2,756 T2D non-nephropathy controls to exclude T2D-associated variants. Six independent variants located in or near *RND3/RBM43*, *SLITRK3*, *ENPP7*, *GNG7*, *EFNB2*, and *APOL1* were associated with T2D-ESKD (p-value $< 5 \times 10^{-8}$), whereby variants in *EFNB2*, *GNG7*, and *APOL1* were also associated with all-cause ESKD (48). *EFNB2* encodes Ephrin-B2 and is expressed in the developing nephron and contributes to the glomerular microvascular assembly (66). The *APOL1* missense mutations rs73885319 (Ser342Gly), rs60910145 (Ile384Met), and rs71785313 (Asn388 and Tyr389 deletion), also known as the *APOL1* G1 and G2 haplotypes, are only found in individuals with African ancestry and are a major contributor to non-diabetic ESKD in AAs (48, 67, 68). To enrich for T2D-associated ESKD, an analysis excluding the *APOL1* ESKD-risk allele carriers identified additional variants in the *GRAMD3* (rs75029938, p-value $= 2.0 \times 10^{-9}$) and *MGAT4C* (rs17577888, p-value $= 3.9 \times 10^{-8}$) genes (48).

A GWAS in 7,614 Japanese individuals with T2D found the rs56094641 in the *FTO* gene to be associated with DKD (p-value $= 7.6 \times 10^{-10}$) (47). *FTO* is one of the strongest genetic loci for obesity and adiposity (69), and rs56094641 is in linkage disequilibrium (LD) with the obesity signal such that the DKD risk-associated allele is also associated with obesity. Indeed, other Mendelian randomization studies utilizing genetic information suggest that obesity is a causal risk factor for DKD (52, 70). However, the association between rs56094641 and DKD was not affected by adjustment for body mass index (BMI), suggesting that the locus affects DKD through another mechanism than an increase in BMI (47). Indeed, the *FTO* locus has been highlighted as a pleiotropic one, associated with multiple biomarkers and traits such as sweet vs. salty taste preference through modifying the regulatory properties of enhancers targeting the *IRX3* and *IRX5* gene expression in various tissues (71, 72).

The SUrrogate markers for Micro- and Macrovascular hard endpoints for Innovative diabetes Tools (SUMMIT) Consortium GWAS meta-analysis of DKD in T2D included 5,717 individuals of European ancestry and with T2D at the discovery stage. After joint analysis with additional European individuals, rs9942471 upstream *GABRR1*, encoding the rho1 subunit of the GABA type a receptor, was associated with microalbuminuria (p-value $= 4.5 \times 10^{-8}$), although the association did not replicate in Asian individuals or in individuals with T1D (19). The variant is in LD with the lead eQTL association signal for *GABRR1* expression in multiple tissues (19). Extended to individuals with T1D and other ethnicities, the joint meta-analysis involved up to 40,340 subjects with diabetes.

However, meta-analysis with individuals with T1D (18) revealed no loci for dichotomous DKD phenotypes. Nevertheless, variants in the *UMOD* and *PRKAG2* loci, previously associated with eGFR and CKD in the general population (73, 74), were associated with eGFR also in individuals with diabetes (Table 2) (19).

3.2.3 Genome-wide association studies on DKD in combined diabetes populations

Meta-analysis of the DNCRI [T1D (49)] and SUMMIT consortia [both T1D (18) and T2D (19)], excluding the overlap between the consortia, and harmonized for the 10 phenotype

TABLE 2 Variants associated with eGFR in diabetes.

SNP	Reported gene	Diabetes population	Phenotype	N total	*P*-value	EA	NEA	Beta	Refs
rs12917707 [a,b]	*UMOD*	T1D + T2D	log eGFR per allele	11,522	2.5×10^{-8}	T	G	0.0266	(19, 50, 75)
rs11864909 [a]	*UMOD*	T1D + T2D	ml/min/1.73 m^2	23,708	2.3×10^{-12}	T	C	2.11	(19)
rs1974990	*SSB*	T1D + T2D	ml/min/1.73 m^2	13,158	4.8×10^{-8}	G	T	4.07	(19)
rs10224002 [a]	*PRKAG2*	T1D + T2D	ml/min/1.73 m^2	22,165	2.7×10^{-8}	A	G	2.01	(19, 50)
rs267738 [a]	*CERS2*	Any	log eGFR per allele	176,573	2.7×10^{-8}	T	G	−0.0065	(76)
rs4665972 [a]	*SNX17*	Any	log eGFR per allele	170,721	3.3×10^{-9}	T	C	0.0057	(76)
rs10206899 [a]	*ALMS1P*	Any	log eGFR per allele	143,419	1.6×10^{-8}	T	C	−0.0068	(76)
rs1047891 [a]	*CPS1*	Any	log eGFR per allele	170,741	5.6×10^{-12}	A	C	−0.007	(76)
rs4663171	*SH3BP4*	Any	log eGFR per allele	170,901	8.8×10^{-9}	A	T	−0.0072	(76)
rs28817415 [a]	*SHROOM3*	Any	log eGFR per allele	176,910	9.9×10^{-26}	T	C	−0.0091	(76)
rs10857147 [a]	*FGF5*	Any	log eGFR per allele	170,848	2.4×10^{-10}	A	T	−0.0061	(76)
rs434215 [a,b]	*TPPP*	Any	log eGFR per allele	119,397	3.5×10^{-19}	A	G	−0.0119	(76)
rs3812036 [a]	*SLC34A1*	Any	log eGFR per allele	170,458	2.1×10^{-12}	T	C	−0.0073	(76)
rs34246779 [a]	*HMGN4*	Any	log eGFR per allele	172,626	1.1×10^{-8}	A	G	−0.0091	(76)
rs3101824 [a,b]	*SLC22A2*	Any	log eGFR per allele	176,569	3.6×10^{-23}	T	C	−0.0143	(76)
rs11761603 [a]	*UNCX*	Any	log eGFR per allele	168,668	4.8×10^{-15}	T	C	0.0075	(76)
rs6464165 [a]	*PRKAG2*	Any	log eGFR per allele	136,252	4.0×10^{-21}	T	C	0.0107	(76)
rs9314272 [a]	*STC1*	Any	log eGFR per allele	177,021	9.4×10^{-10}	A	G	−0.0054	(76)
rs7033278 [a]	*PIP5K1B* [c]	Any	log eGFR per allele	176,480	1.3×10^{-10}	T	C	0.0062	(76)
rs80282103 [a]	*LARP4B*	Any	log eGFR per allele	176,591	6.8×10^{-11}	A	T	0.0109	(76)
rs55917128	*LOXL4*	Any	log eGFR per allele	176,998	4.6×10^{-8}	T	C	−0.0048	(76)
rs963837 [a,b]	*DCDC5*	Any	log eGFR per allele	170,722	2.4×10^{-34}	T	C	−0.0108	(76)
rs2004649 [a]	*MAP3K11*	Any	log eGFR per allele	176,918	6.1×10^{-10}	A	G	−0.0055	(76)
rs10899482 [a]	*GAB2*	Any	log eGFR per allele	177,039	1.2×10^{-8}	A	C	−0.0058	(76)
rs2461700 [a]	*GATM*	Any	log eGFR per allele	177,144	1.5×10^{-15}	T	C	0.008	(76)
rs17631603 [a]	*WDR72*	Any	log eGFR per allele	177,042	6.6×10^{-15}	A	G	0.0068	(76)
rs11636251 [a]	*NRG4*	Any	log eGFR per allele	171,081	1.9×10^{-14}	T	C	−0.0069	(76)
rs77924615 [a,b]	*UMOD/PDILT*	Any	log eGFR per allele	170,741	1.9×10^{-106}	A	G	0.0234	(76)
rs9895661 [a]	*BCAS3*	Any	log eGFR per allele	176,461	6.7×10^{-10}	T	C	0.0066	(76)
rs8096658 [a]	*NFATC1*	Any	log eGFR per allele	167,173	1.6×10^{-12}	C	G	0.0067	(76)
rs6015028 [a]	*PCK1*	Any	log eGFR per allele	176,558	1.4×10^{-9}	A	T	−0.0071	(76)
rs1882961 [a,b]	*NRIP1*	Any	log eGFR per allele	176,630	3.6×10^{-14}	T	C	−0.0073	(76)
rs9607518 [a]	*MAFF*	Any	log eGFR per allele	170,649	2.2×10^{-8}	T	C	−0.0049	(76)

SNP: Variant rs-identifier. EA: Effect allele. NEA: non-effect allele. Beta: effect size beta estimate. Refs: If multiple references are given, then the data in other columns for the same locus are taken from the first listed reference.
[a]Associated with eGFR also in the general population.
[b]Significant effect size difference between individuals with and without diabetes.
[c]DNA methylation of CpGs in the gene region associated with DKD (77, 78).

definitions of DKD for available cohorts, included nearly 27,000 individuals with diabetes (52). The meta-analysis identified a novel intronic variant, rs72831309 in the *TENM2* gene, to be associated with a lower risk of the combined CKD-DKD phenotype (p-value = 9.8×10^{-9}). *TENM2* gene expression in kidney tubules correlated positively with eGFR (p-value = 1.6×10^{-8}) and negatively with tubulointerstitial fibrosis (p-value = 2.0×10^{-9}). In addition, the gene-level analysis identified 10 genes significantly associated with DKD (*COL20A1*, *DCLK1*, *EIF4E*, *PTPRN–RESP18*, *GPR158*, *INIP–SNX30*, *LSM14A*, and *MFF*; p-value $<2.7 \times 10^{-6}$). Transcriptome-wide association study integrating GWAS with human glomerular and tubular gene expression data demonstrated a higher tubular *AKIRIN2* gene expression associated with DKD (p-value = 1.1×10^{-6}). Expression of multiple lead genes correlated with renal phenotypes, e.g., tubular *DCLK1* expression correlated with fibrosis (p-value = 7.4×10^{-16}) and *SNX30* expression with eGFR (p-value = 5.8×10^{-14}), and negatively with fibrosis (p-value < 2.0×10^{-16}) (52).

In addition to the disease-specific cohorts, large population-based biobanks allow analyses of an increasing number of samples and phenotypes. A GWAS on DKD in the UK Biobank included 13,123 unrelated individuals with diabetes and of European origin. Of note, the heritability estimate for DKD, defined based on ICD-10 codes (E11.2, T2D with kidney complications, or any CKD code assigned after diabetes) or a measurement of albuminuria or eGFR, was only 0.027 with a standard deviation (SD) of 0.03; heritability estimate for eGFR in T2D was higher, 0.1 with an SD of 0.01. GWAS on DKD and eGFR identified variants in the *UMOD* and *PRKAG2* loci (50). Meta-analysis with the SUMMIT T2D study further identified a novel variant, rs72763500, associated with the combined DKD definition. The variant is associated with alternative gene splicing of the *NID1* gene (50), encoding for nidogen-1, a sulfated glycoprotein involved in the development of GBM, where it binds to laminin and type IV collagen (79). Another study in the UK Biobank, although focused on heritability estimates for diabetic micro- and macrovascular complications, additionally found a variant rs73038008 near *PLD1* associated with DKD (self-reported or medical records); as well as variants in *WSCD2* and *SETDB2* associated with ESKD and in *LOC105372639* associated with microalbuminuria (20).

3.2.4 Genome-wide association studies on albuminuria and eGFR in diabetes

In addition to the dichotomous case-control definitions of DKD, GWASs have also explored albuminuria and eGFR as continuous traits in individuals with diabetes (Figure 1). Only few studies have identified variants with genome-wide significance for albuminuria (Table 3) or eGFR (Table 2), and most of these loci were identified in diabetes-specific sub-analyses of larger general population studies.

A GWAS including 1,925 Finnish individuals with T1D identified rs10011025 in the *GLRA3* associated with albuminuria (p-value = 1.5×10^{-9}) (25). The association did not replicate in 3,771 other European individuals with T1D (p-value = 0.04, opposite direction) (25); however, the association was subsequently replicated in 1,259 additional Finnish individuals with T1D (81). The association was pronounced in individuals with HbA_{1c} > 7%. The *GLRA3* gene encodes the α3 subunit of glycine receptors. In pancreatic α-cells, glycine receptors stimulate glucagon release in response to glycine, thus counterbalancing the effects of insulin (83). Interestingly, the association with albuminuria was only evident among individuals with a 24-h urine collection. Because exercise can acutely increase albuminuria due to excess hemodynamic pressure (84), the authors hypothesized that the variant might affect renal sensitivity to hemodynamic pressure (81). Of note, in the eQTLGen database, the rs10011025 variant is associated with the expression of the *HPGD* gene, encoding for the 15-hydroxyprostaglandin dehydrogenase that catalyzes the prostaglandin catabolic pathway; prostaglandins are locally acting vasodilators and regulate renal hemodynamics in the kidneys (85).

Another GWAS on albuminuria included 54,450 individuals from the general population, confirming the previously identified *CUBN* locus (86) for albuminuria. In the sub-analysis of 5,825 individuals with diabetes, variants in the *HS6ST1* (rs13427836, p-value = 6.3×10^{-7}) and *RAB38/CTSC* loci (rs649529, p-value = 5.8×10^{-7}) were suggestively associated with albuminuria in subjects with, but not without diabetes (87). *RAB38* expression was found higher in the tubules of individuals with DKD compared to healthy controls, and *Rab38* knockout resulted in higher urinary albumin concentrations in diabetic rat models (87). A larger study including

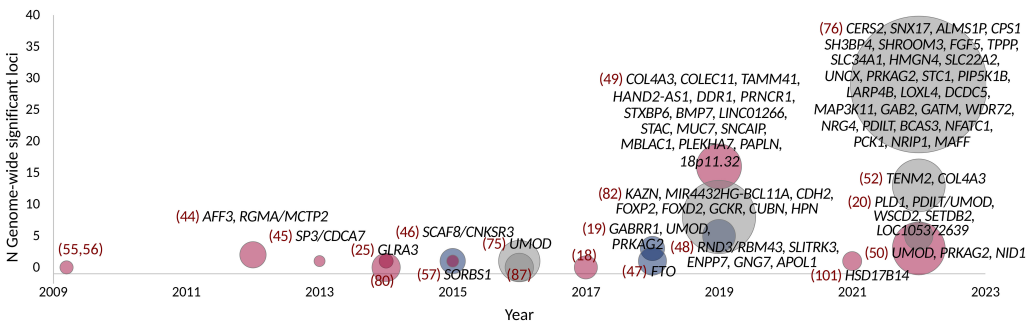


FIGURE 1
GWAS on DKD, albuminuria, and eGFR in diabetes. Point size indicates the number of samples. Studies with individuals with T1D are colored red, T2D with blue, and combined T1D + T2D, any type of diabetes or unspecified type of diabetes with gray. Gene names indicate loci reaching genome-wide significance (p-value < 5×10^{-8}).

TABLE 3 Variants associated with albuminuria in diabetes.

SNP	Reported gene	Diabetes population	Phenotype	N	*P*-value	EA	NEA	Beta	Refs
rs10011025	*GLRA3*	T1D	$\log_{10}$ AER	1,925	1.5×10^{-9}	G	A	0.21	(25, 81)
rs59825600	*KAZN*	Any	sd of log(UACR)	40,668	3.6×10^{-8}	A	G	−0.075	(82)
rs6688849 [a]	*FOXD2*	Any	sd of log(UACR)	51,215	4.1×10^{-9}	A	G	−0.049	(82)
rs780093 [a]	*GCKR*	Any	sd of log(UACR)	51,515	1.5×10^{-13}	T	C	0.049	(82)
rs6706313	*MIR4432HG-BCL11A*	Any	sd of log(UACR)	51,162	2.8×10^{-8}	A	G	−0.041	(82)
rs17137004	*FOXP2*	Any	sd of log(UACR)	51,294	2.7×10^{-8}	A	G	−0.036	(82)
rs74375025 [a]	*CUBN*	Any	sd of log(UACR)	50,641	1.1×10^{-24}	A	G	0.106	(82)
rs4258701	*CDH2*	Any	sd of log(UACR)	51,328	1.1×10^{-8}	T	C	0.039	(82)
rs149131600 [a]	*HPN*	Any	sd of log(UACR)	46,939	3.5×10^{-8}	T	C	0.050	(82)

SNP: Variant rs-identifier. EA: Effect allele. NEA: non-effect allele. Beta: effect size beta estimate. Refs: If multiple references are given, then the data in other columns for the same locus are taken from the first listed reference. AER, albumin excretion rate. UACR, urinary albumin-to-creatinine ratio.
[a]Significant also in the general population, but with larger effect in diabetes.

564,257 individuals, of which 51,541 individuals with diabetes, identified eight loci associated with albuminuria in diabetes; all had larger effect among individuals with diabetes, and four (*KAZN*, *MIR4432HG-BCL11A*, *FOXP2*, and *CDH2*) were only found in the secondary analysis limited to diabetes (82).

Finally, a GWAS including 178,691 individuals with diabetes from the CKD Genetics (CKDGen) consortium and large biobank studies identified 29 genome-wide significant loci for eGFR, including 27 novel loci for eGFR in diabetes; among these, variants near *SH3BP4* and *LOXL4* were not associated with eGFR in the 1,296,113 individuals without diabetes (76).

3.3 Overlap between genetic factors for DKD and general population kidney traits

In the general population, nearly 900 genetic loci have been identified for eGFR in meta-analyses, including over 1.5 million individuals (88). Diabetes is one of the key risk factors for CKD, and 31% of the CKD-associated disability-adjusted life years can be attributed to diabetes (89). Other main risk factors for CKD include hypertension, obesity, and high age, all commonly seen among individuals with T2D in particular. In individuals with T1D, the majority of DKD is due to diabetic nephropathy. On the contrary, the renal lesions in kidney biopsies of DKD in T2D are heterogeneous, and a substantial proportion of the biopsies do not show the typical characteristics of diabetic nephropathy (90). However, kidney biopsies are rarely taken, and DKD is defined as any CKD in an individual with diabetes (91). Therefore, the question arises, how much of the genetic background of DKD is shared with the CKD and eGFR in the general population?

The DKD loci identified in individuals with T1D in the DNCRI consortium did not replicate in the general population GWAS for eGFR (49); conversely, the loci associated with eGFR in the general population (92) were not associated with DKD in T1D apart from the *UMOD* locus (49). On the contrary, some of the first findings for DKD in T2D included the *UMOD* and *PRKAG2* loci known from the general population (19), as well as the *APOL1* variant responsible for the majority of kidney failures in AAs (48). The CKDGen GWAS on eGFR including 133,413 individuals, of which 16,477 with diabetes, found that the effect size of the eGFR loci identified in the full population were highly correlated between individuals with and without diabetes (correlation coefficient of 0.80) (75). A more recent study on eGFR from the CKDGen consortium, including nearly 1.5 million participants of which 178,691 with T2D, systematically sought for differences in effect size between individuals with and without diabetes. They identified seven eGFR loci with significant difference in individuals with and without diabetes, as well as four loci with suggestive difference; in all but one, the effect was more pronounced or exclusively seen among individuals with diabetes (76). Similarly, in a GWAS for eGFR decline studied as a longitudinal trait in the general population, the effect sizes of the nine identified variants were on average two-fold higher in individuals with diabetes (93). Finally, the effect of the rs10795433 variant in the *CUBN* locus—the major locus for albuminuria—was larger among individuals with diabetes compared to those without diabetes (87). In addition, a rare *CUBN* variant rs141640975 had three times stronger effect in individuals with T2D compared with those without (94). Furthermore, rs141640975 was associated with higher eGFR but only in the non-diabetes population, suggesting pleiotropic effects on both kidney function measures (95).

In the DNCRI-SUMMIT GWAS meta-analysis for DKD, the similarity of DKD with kidney traits in the general population (of note, including individuals with diabetes) was assessed on a genome-wide scale instead of single-variant level, using the LD score regression approach. The albuminuria-based DKD definition, including microalbuminuria, was genetically correlated with microalbuminuria in the general population, both in the pooled analysis, and separately for individuals with T1D or T2D; of note, the correlation was over two-fold stronger in individuals with T2D. In addition, the eGFR-based CKD definition was also correlated with eGFR and CKD in individuals with T2D, but not in T1D despite more than three times more individuals with T1D (52). The analysis suggests that DKD in T2D has a larger proportion of shared genetic background with the general population, e.g., due to other

co-existing risk factors such as aging, overweight, hypertension, and other glomerular diseases, while less overlap is observed between the general population kidney traits and DKD in T1D representing a purer form of diabetic nephropathy. The LD score regression with cardiometabolic and other traits further suggested that a proportion of the genetic background of DKD is shared with genetic risk factors, e.g., for aging (mother's age at death), obesity, and smoking (52). However, the confidence intervals remain large, and further studies are needed to estimate the proportion of risk attributable to each risk factor.

Some interesting discrepancies also exist between DKD and the general population: For example, the missense variant rs55703767 in *COL4A3* is one of the strongest findings for DKD in T1D, but the effect is modified by glycemia, and the variant does not seem to affect kidney traits in the general population. On the contrary, variants in the flanking *COL4A4* (collagen type IV alpha 4 chain) gene were associated with albuminuria in the general population (rs57858280, p-value = 9×10^{-11}) (82); according to the GTEx portal, the variant may affect the *COL4A4* splicing (https://gtexportal.org/). Rare mutations in both *COL4A3* and *COL4A4* cause Alport syndrome, a monogenic disease of basement membranes that frequently leads to ESKD, as well as thin basement membrane nephropathy and focal segmental glomerulosclerosis (96).

3.4 Overlap between genetic factors for DKD and diabetes

Some studies have suggested a correlation between the genetic risk factors predisposing to insulin resistance or T2D and DKD (18, 19, 52). Of note, these studies found no correlation between genetic risk factors predisposing to T1D and DKD. T2D was modestly causally associated with DKD in a Mendelian randomization study of individuals with either T1D or T2D (p-value = 0.02), but only obesity related traits remained significantly associated with DKD when using methods accounting for pleiotropic effects (52). However, among the lead variants for DKD, albuminuria, or eGFR in diabetes, only the albuminuria-associated *FTO* locus [rs56094641 (47)] has been associated with T2D. In addition, the albuminuria-associated rs780093 (82) in the highly polygenic *GCKR* locus, as well as the eGFR-associated rs4665972 (in *SNX17*, but in LD with variants mapped to *GCKR*), rs11864909 (*UMOD*), rs10206899 (*ALMS1P*), rs10899482 (*GAB2*), and rs9607518 (*MAFF*), are in LD with variants associated with T2D (https://ldlink.nci.nih.gov/?tab=ldtrait ; search for any "diabetes" in GWAS Catalog for variants in LD ($R^2 \geq 0.8$ in European population), 21 March 2023), providing some evidence of genetic overlap between T2D and eGFR in diabetes.

4 From common to rare genetic variants for DKD

While common variants have a large effect on complex traits at the population level (43), the low frequency and rare variants can have a high impact on the individual level (97). In particular, protein-altering variants (PAVs), i.e., exon variants that change the protein amino acid sequence, can directly impact protein function. For example, 71% of severe *LDLR* mutation carriers had hypercholesterolemia in the UK Biobank WES data (98). To identify chromosomal regions harboring rare variants for DKD, a linkage study based on GWAS data of 6,019 FinnDiane study participants included 177 small pedigrees such as sib-ships, parent-offspring pairs, and more distant relations, with, altogether, 452 individuals, all with T1D. Eight chromosomal regions reached a significant LOD score > 3.3 (51). Many of these regions harbor genes in which mutations cause rare syndromes with kidney complications, such as *ARHGAP24* associated with focal segmental glomerulosclerosis (99) and *FRAS1* associated with the familial Fraser syndrome (100). Overlap with loci causing rare kidney syndromes supports the role of rare variants in the development of DKD. Interestingly, one suggestive linkage peak was observed in the *NID1* locus, recently associated with DKD in T2D (50). While a rare rs538044833 variant in the *CCSER1* locus was externally replicated (p-value = 2.8×10^{-8}), the resolution remains low even in the GWAS-based linkage studies, hindering further fine-mapping and interpretation of the results.

In addition, on the basis of GWAS data, enriched for rare PAVs with the ExomeChip array, a gene aggregate meta-analysis including 4,196 individuals with T1D found PAVs in the hydroxysteroid 17-β dehydrogenase 14 (*HSD17B14*) gene exome-wide significantly (p-value < 5×10^{-7}) associated with the disease progression from DKD to kidney failure. The gene and protein expression were attenuated in human diabetic proximal tubules and in mouse kidney injury models (101).

The GWAS genotyping chips cover only a portion of the PAVs, and genotype imputation quality largely depends on the variant minor allele count in the reference sample and can be limited for rare variants (102, 103). A whole-exome sequencing (WES) on DKD, including 997 individuals with T1D, did not find any variants or genes reaching robust exome-wide significance (18) but found suggestive evidence of association, e.g., for PAVs in the *THADA* gene, previously associated with T2D (104). A WES of 593 DKD cases and 2,066 healthy controls of European and African ancestry, with subsequent discriminatory analyses and replication in up to 11,487 multi-ancestry participants from the Trans-Omics for Precision Medicine study, identified an in-frame insertion rs141560952 in the *DIS3L2* gene (p-value = 3.6×10^{-9}), and a *KRT6B* splice-site variant rs425827 associated with DKD (p-value = 2.7×10^{-9}). Both variants were associated with DKD also when compared with diabetes controls without DKD, but with lower statistical significance (p-value = 1.4×10^{-4} and 2.8×10^{-4}). Furthermore, gene aggregate analyses identified *ERAP2* (p-value = 4.03×10^{-8}) and *NPEPPS* (p-value = 1.51×10^{-7}); both are expressed in the kidney and implicated in the renin-angiotensin-aldosterone system–modulated immune response (53). However, the discriminatory analyses suggest that the *ERAP2* and *NPEPPS* may be primarily associated with diabetes *per se*, subsequently leading to DKD (53).

While WES mainly covers the protein-coding sequence, a whole-genome sequencing (WGS) study of 76 Finnish sibling pairs with T1D but discordant for DKD found significant

enrichment of variants in DKD in gene promoter and enhancer regions, as well as for specific transcription factor binding sites (105), but larger studies are required to pinpoint the most relevant regulatory regions. Gene aggregate analysis of PAVs suggested protein kinase C isoforms (*PRKCE* and *PRKCI*) and protein tyrosine kinase 2 (*PTK2*) involved in DKD (105); of note, a recent GWAS on albuminuria in the general population highlighted variants in the *PRKCI* and demonstrated that a podocyte-specific deletion of aPKClambda/iota in mice results in severe proteinuria (82). A recent multi-ethnic WGS in 23,732 individuals identified three novel rare intronic variants for eGFR in the general population (106), and larger WGS for DKD are needed to identify the rare variants contributing to DKD.

5 Epigenetic factors for DKD

Studies focusing on epigenetic modifications have emerged in an increasing number during the last years. Epigenetic modifications can be described as chemical modifications of the DNA (or RNA) that can induce changes in gene expression without changing the underlying sequence. In contrast to an individual's genetic variation, which is constant across tissues and throughout lifetime, epigenetic modifications are dynamic and modifiable. Thus, epigenetic changes may vary between tissues, cell types, and developmental stages and can even be affected by environmental factors. Furthermore, in disease states, the methylation patterns can change either as a cause or a consequence of the disease (107). In this way, epigenetic factors provide a link between the genome and the environment and can potentially reflect an individual's risk of developing a disease more accurately at a given time. Although epigenetic changes are dynamic, there is evidence that epigenetic modifications, such as DNA methylation, persist in blood years after acute illness or metabolic changes in the body (108, 109). Consequently, epigenetic factors have been suggested as an underlying mechanism for metabolic memory (110, 111). Metabolic memory in diabetes refers to the sustained harmful effect of hyperglycaemia on diabetic complications, initially observed in the DCCT-EDIC study, even after improved glycaemic control (112, 113). In line with this observation, subsequent work in DCCT-EDIC has identified several epigenetic changes associated with metabolic memory (110, 111). A combination of DNA methylation levels at several HbA_{1c}-associated sites explained as much as 71 to 97% of the association between HbA_{1c} and diabetic complications in the DCCT (114), further reinforcing the connection between epigenetic changes and metabolic memory.

DNA methylation is the most frequently studied epigenetic modification and occurs at cytosine bases of cytosine–phosphate–guanine dinucleotide sites (CpGs) in the DNA sequence. In addition to DNA methylation, additional epigenetic modifications exist, such as histone modifications (acetylation and methylation), and their role in DKD has also been explored. For example, dysregulation of histone H3 lysine 27 trimethylation (H3K27me3) in TGF-β1–induced gene expression has been associated with DKD (115). Histone modifications associated with DKD are reviewed, e.g., in (116), and are out of the scope of this review, where we focus on DNA methylation changes.

5.1 Various study settings for DNA methylation

Although whole-genome bisulfite sequencing for the analysis of the methylome has been done for DKD, sample sizes have been small (117). Studies assessing DNA methylation patterns across the genome, known as epigenome-wide association studies (EWASs) or methylome-wide association studies (MWASs), have primarily relied on Illumina's BeadChip platforms, which have evolved from the Illumina 27K array with only ~27,000 sites to the Illumina 450K with ~450,000 and the EPIC array containing methylation levels at ~850,000 sites. However, this number of CpGs only accounts for a small amount of all the CpGs in the genome, totalling up to ~30 million (118). The EWASs have applied various significance thresholds, but a p-value below 9×10^{-8} has been suggested as a threshold for robust significance, adequately controlling for the false positive rate for the EPIC array (94). The genome-wide significance threshold recommended for Illumina's 450K BeadChip is p-value $< 2.4 \times 10^{-7}$ or p-value $< 3.6 \times 10^{-8}$ (119), although the false discovery rate (FDR) has been widely used (Table 4). Contrary to the GWAS, which initially yielded few significant loci with increasing number of findings with larger studies, in EWAS, the use of varying thresholds, combined with unaddressed inflated test statistics especially in the early EWAS (131), has led to a quite varying number of identified methylation loci in the studies performed so far.

Most EWASs performed on DKD have examined DNA methylation in blood. Still, other tissues have been used, such as kidney samples micro-dissected into kidney tubules (125) and even saliva (121). The epigenetic changes observed in the kidney tissue likely reflect the local changes more accurately. Indeed, EWAS on fibrosis in kidney tissue samples identified 65 differentially methylated CpGs that were enriched on kidney regulatory regions (125). Another promising target tissue for studying kidney disease would be the urine, which can be collected non-invasively and easily from larger datasets. Urine, however, contains few nucleated cells and extracting a sufficient amount of DNA from urine has turned out to be a challenge (132).

5.2 Over 150 CpGs associated with DKD and related traits

To date, methylation levels at over 150 CpG sites across the genome have been associated with DKD, eGFR, or albuminuria (p-value $< 9 \times 10^{-8}$), in studies including both T1D and T2D (Figure 2; Table 4; Supplementary Table 1), with the majority assessing DNA methylation in blood. The first DKD-EWAS identified DNA methylation levels at 19 CpGs associated with DKD in T1D (FDR < 0.05) using Illumina's 27K array (120), highlighting one CpG located upstream of the *UNC13B* gene. An intronic SNP (rs2281999) in the same *UNC13B* gene was

10.3389/fendo.2023.1163001

TABLE 4 EWASs on kidney disease and related traits in individuals with diabetes.

Study	Ethnicity	Tissue	Phenotype	Cases	Controls	N Total	CpGs (array)	p-threshold	N significant CpGs
Bell, 2010 (120)	White European	Blood	DKD	96 (T1D: 100%)	96 (T1D: 100%)	192	27,578 (27K)	P_{FDR}< 0.05	19 (P_{FDR}< 0.05); none with P_{FDR} <10^{-8}
Sapienza, 2010 (121)	African American/ Hispanic	Saliva	DKD	24 (T2D: 87%, T1D: 13%)	24 (T2D: 100%)	48	27,578 (27K)	Diffscore** > 20 or < −20	2,870, of which 30 remained significant after FDR adjustment (P_{FDR} < 0.05)
Smyth, 2014 (122)	White European	Blood	CKD/DKD	255 (T1D: 44%)	152 (T1D: 74%)	407	485,577 (450K)	P_{FDR} < 10^{-8}	52 CpGs (P_{FDR} < 10^{-8}) in 23 genes
Swan, 2015 (123)	White European	Blood	DKD	196 (T1D: 100%)	246 (T1D: 100%)	442	450* (27k, 450K)	P_{FDR} < 10^{-8}	54 (P_{FDR} < 10^{-8})
Qiu, 2018 (124)	American PIMA Indians	Blood	eGFR; ESKD; eGFR slope	80 (T2D: 100%)	101 (T2D: 100%)	181	397,063 (450K)	P_{FDR} < 0.05	eGFR and ESKD: none (P_{FDR} < 0.05); 77 (eGFR slope, P_{FDR} < 0.05)
Gluck, 2019 (125)	Mixed	Kidney tubules	degree of kidney fibrosis	91 (22 with DKD)	0	91	321,473 (450 K)	P_{FDR} < 0.05	Degree of fibrosis: 203 (P_{FDR} < 0.05) of which 65 replicated (p < 0.05)
Sheng, 2020 (126)	Mixed	Blood	eGFR, eGFR slope, albuminuria	473 (all with diabetes)	0	473	866,836 (EPIC)	$P < 5 \times 10^{-5}$ (discovery), p < 6.4×10^{-8} (Bonferroni)	Albuminuria: 73 ($P < 5 \times 10^{-5}$), eGFR: 99 ($P < 5 \times 10^{-5}$); 1 (6.4×10^{-8}), eGFR slope: 111 ($P < 5 \times 10^{-5}$); 3 (6.4×10^{-8})
Smyth, 2020 (127)	White European	Blood	DKD	150 (T1D: 100%)	100 (T1D: 100%)	677	482,421 (450K)	P_{FDR} < 10^{-8}, Δβ > 0.2	22
Kim, 2021 (128)	East Asian	Blood	DKD	87 (T2D: 100%)	80 (T2D: 100%)	167	749 315 (EPIC)	P_{FDR} < 9.0×10^{-8}	3 (P_{FDR} < 9.0×10^{-8})
Smyth, 2021 (129)	White European	Blood	ESKD (4 analysis models)	107 (T1D: 100%)	253 (T1D: 100%)	360	862,927 (EPIC)	P_{FDR} < 10^{-8}, FC ± 2	36 (P_{FDR} < 10^{-8}, FC ± 2 across all four models)
Lecamwasam, 2021 (130)	Mixed	Blood	late CKD (eGFR<45) vs. early (eGFR≥45)	38 (T1D: 8%, T2D: 87%)	83 (T1D: 20%, T2D: 80%)	119	764 333 (EPIC)	P_{FDR} < 0.05	1 (P_{FDR} < 0.05)
Smyth, 2022 (77)	White European	blood	DKD (3 analysis models)	651 (T1D:100%)	653 (T1D: 100%)	1304	763 064 (EPIC)	P_{FDR} <9×10^{-8}	32 (P_{FDR} < 9×10^{-8})

27K, Illumina Infinium HumanMethylation 27K; 450K, Illumina Infinium HumanMethylation 450K; EPIC, Illumina Infinium HumanMethylation EPIC v1.
*Only CpGs within mitochondrial genes were surveyed.
**Diffscore = $10\text{sgn}(\text{b-value}_{\text{ESRD}} - \text{b-value}_{\text{diabetes no nephropathy}})$ log10p.

identified for DKD in T1D in a prior genetic association study including genetic variants in 127 candidate genes (133). More recent methylation arrays, with higher coverage have enabled identification of additional CpGs. Using the 450K array, Smyth et al. identified 53 CpGs within 23 genes with differential methylation in participants with CKD, of which approximately half had T1D. Of the 23 genes, six were in genes that are biological candidates for kidney disease: *CUX1*, *ELMO1*, *FKBP5*, *INHBA-AS1*, *PTPRN2*, and *PRKAG2* (122). Of these, genetic variants within the *PRKAG2*, encoding a protein kinase involved in cellular energy metabolism, have also been associated with eGFR in GWAS on kidney disease, both in individuals with and without diabetes (19, 73, 74). Following this study, several EWAS have been performed (Table 4), focusing mainly on DKD (77, 123, 127) and ESKD (129) in T1D but also on DKD in T2D (128) or eGFR in individuals with diabetes of unspecified/mixed type (124, 126, 130), yielding a plethora of sites that are differentially methylated, shown in Figure 2 (CpGs with *p*-value < 9×10^{-8}). The most recent and largest study, including 1,304 individuals with T1D, identified 32 sites with altered methylation in DKD (77), of which 23 were specific to the EPIC array. Methylation levels at seven CpGs were epigenome-wide significantly and differentially methylated after accounting for differences in multiple clinical risk factors (HbA_{1c}, HDL cholesterol, triglycerides, BMI, smoking, and duration of diabetes), in addition to age, sex, and six cell-type proportions. These seven included two intergenic CpGs on chromosome 19 and four CpGs located within genes *PTBP3*, *NME7*, *SLC1A5*, and *SLC27A3* and one CpG within a long non-coding RNA (*LINC01800*).

Methylation levels at only a few CpGs have been associated with DKD in multiple studies (Figure 2). This can partly be explained by the higher coverage of Illumina's EPIC array, with many CpGs on that array not present on previous arrays and, therefore, not testable. Consequently, one-third of the differentially methylated CpGs identified for DKD or eGFR in studies using the EPIC array (77, 126, 129) were novel and not available on previous arrays (Supplementary Table 1). However, methylation loci that have been repeatedly associated with DKD, include CpG within genes *C5orf66*, *FKBP5* (77, 122), and *PIP5K1C* (77, 129). In addition, higher methylation at the intergenic CpG cg17944885, located on chromosome 19 within a zinc finger gene cluster, has been repeatedly associated not only with DKD and eGFR in diabetes (77, 126, 130) but also with CKD and eGFR in the general population (78, 134, 135), as well as eGFR in other more specific cohorts, such as men with human immunodeficiency virus (HIV) (136). Moreover, CpGs within the *IRF2* (cg05165263) and *SLC27A3* (cg21961721) gene, both with higher methylation levels in DKD in T1D (77), have also been associated with eGFR (p-value = 5×10^{-5} and 8×10^{-5}) in the general population (135), although not among the reported top loci.

Although most of the DNA methylation association studies performed on DKD have covered the whole genome, targeted approaches have been undertaken as well. Swan et al. evaluated DNA methylation levels associated with DKD for CpGs located within genes influencing mitochondrial function in 442 individuals with long term T1D (123). Although methylation levels at several CpG sites reached the threshold for epigenome-wide significance (p-value < 9×10^{-8}), none of the differentially methylated CpG sites has emerged in subsequent EWASs.

A few CpGs identified as differentially methylated in DKD to date (Figure 2) also appear in EWAS on traits that are considered risk factors for DKD. Lower methylation of cg19693031 located in the 3′-untranslated region of the *TXNIP* gene has been recurrently observed in the context of diabetes and glycemia, such as persistently higher HbA_{1c} both in T2D and T1D (109, 126, 137). *TXNIP* encodes for the thioredoxin-interacting protein, which by binding to thioredoxin induces oxidative stress and apoptosis. Although it is mainly considered a glycemia-related methylation locus, it not only shows repeated associations with albuminuria and DKD (77, 109), explaining alone up to 45% of the HbA_{1c} association with DKD (114), but also associates with DKD and triglycerides independently of HbA_{1c} (109). Intriguingly, methylation levels at cg19693031 are also under genetic influence by SNPs located within the *SLC2A1* gene encoding for the glucose transporter 1 (GLUT1) (109). A recent EWAS on DKD performed a systematic trait enrichment analysis and found significant overlap with EWAS findings for traits and diseases such as aging, smoking, systolic and diastolic blood pressure, eGFR, and HbA_{1c} (77). Our lookup of the significant CpGs identified for DKD, eGFR, fibrosis, and albuminuria to date (Figure 2; 160 CpGs as listed in Supplementary Table 1) in the EWAS catalogue (associations with p-value < 9×10^{-8}; http://www.ewascatalog.org , accessed 31 January, 2023) found an overlap with DKD risk factors including dyslipidemia (CpGs within *SLC1A5*, *TXNIP*, and *CPT1A*), HbA_{1c} (*TXNIP*), blood pressure (CpGs within *SLC1A5*, *TXNIP*, *CPT1A*, and *PTBP3*) and obesity (CpGs within *SLC1A5*, *TXNIP*, *CPT1A*, and *FKBP5*; Supplementary Table 2). For example, in the *CPT1A* gene, methylation at cg17058475 was associated with DKD in T1D (77) and has been robustly associated with the

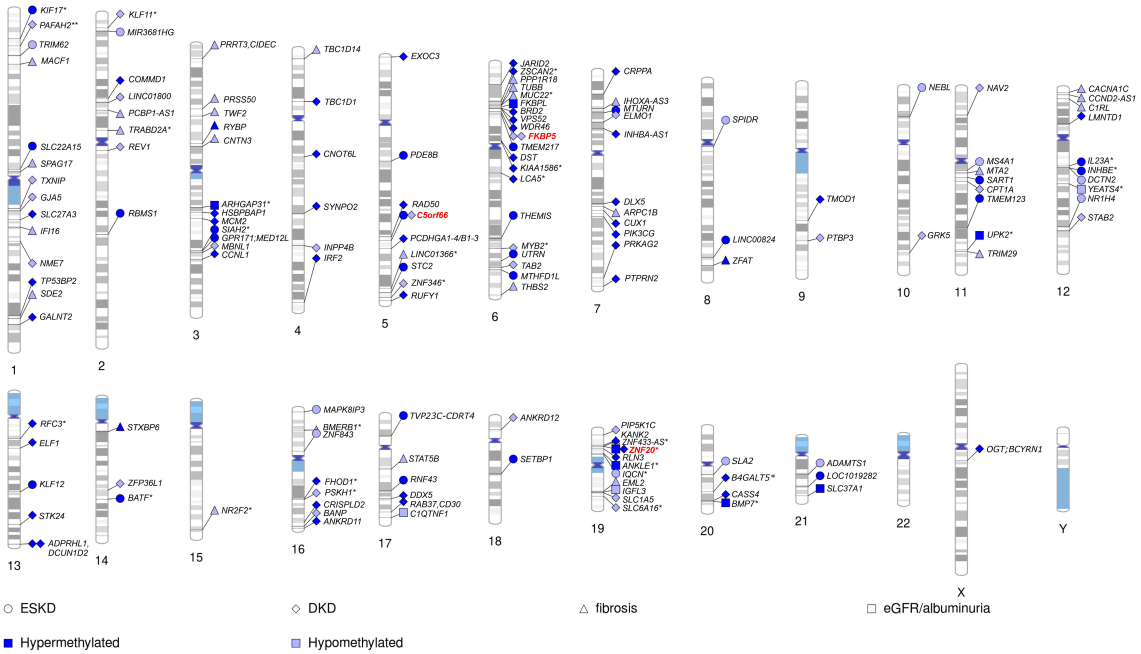


FIGURE 2
Chromosomal ideogram including CpGs methylation associated with kidney disease (DKD and ESKD), fibrosis, eGFR, or albuminuria in diabetes. For intergenic CpGs (*), the nearest gene is given. Hypermethylated CpGs (in kidney disease vs. controls) are denoted by a dark blue colour and hypomethylated by a light blue colour. CpGs appearing among top loci in multiple studies on kidney disease in diabetes denoted by a red color.

triglycerides (118) in the general population. *CPT1A* encodes a key enzyme in the fatty acid metabolism, namely, the hepatic isoform of carnitine palmitoyl transferase 1 (138), controlling the fatty acid flux in the liver. In addition, the genetic variants in the gene were also associated with triglycerides and HDL cholesterol in a recent GWAS (139).

5.3 DNA methylation for prediction of DKD

Several studies have provided evidence suggesting that changes in DNA methylation patterns could be used to predict DKD or its progression. Using 91 kidney tissue samples, Gluck et al. found that information from 471 differentially methylated CpGs in the kidneys helped them to predict kidney disease progression (125). However, the utility of kidney tissue–specific DNA methylation patterns as potential biomarkers remain limited, as individuals with DKD do not routinely undergo kidney biopsy. As an alternative, a study with methylation data from 831 individuals constructed methylation risk scores for 607 phenotypes based on electronic health records and suggested that blood methylation was particularly good in identifying individuals with pre-existing kidney failure and related traits (140). An EWAS in 181 American Indians with diabetes identified methylation levels at 77 CpG sites associated with eGFR decline over a 6-year period (124). Methylation at two CpGs (cg25799291 and cg22253401 in *FSTL5*) improved prediction of eGFR decline even when baseline eGFR and Albumin-to-creatinine ratio (ACR) were included in the model (124). In addition, in T1D, methylation levels at baseline can be used to predict progression of DKD. In total, 20 of the 32 differentially methylated CpGs in DKD in T1D predicted future progression to kidney failure in 397 individuals with DKD, 13 even after accounting for eight clinical risk factors (77). Furthermore, methylation at the two intergenic CpGs located within the zinc finger gene cluster on chromosome 19 predicted kidney failure, independent of baseline eGFR.

5.4 Epigenetic changes—the cause or the consequence?

Because of the dynamic nature of epigenetic changes, the methylation changes observed at CpGs in DKD can be either a cause or a consequence of the disease. To separate the causal methylation changes from the consequential, EWASs have also attempted Mendelian randomization, which uses genetic information to infer causality (77, 126, 128). Although these analyses have been partly hampered by the lack of genetic variants influencing CpG methylation, some causal associations have been observed. For example, Mendelian randomization suggested that higher methylation levels at cg23527387 located within the *REV1* gene reduces the risk of DKD in T1D (77). On the other hand, no evidence for causality was found for cg19693031 (*TXNIP*) or cg17944885 (between *ZNF788P* and *ZNF625-ZNF20*), suggesting that methylation changes observed at these sites are consequential to kidney disease or its other manifestations, e.g., hyperglycemia. Kim et al. used Mendelian randomization in the opposite direction, i.e., to assess the causal effect of metabolic phenotypes on CpG methylation changes identified in their EWAS on T2D (n = 8) and DKD (n = 3). These analyses revealed that fasting glucose resulted in 2% hypomethylation of cg00574958 located in the *CPT1A* gene, whereas HbA1c or BMI did not causally affect the cg00574958 methylation. Genetically determined eGFR, however, was associated with 7% hypomethylation of cg19693031 within *TXNIP* (p-value = 0.045), as well as hypomethylation of all the CpGs identified for DKD in T2D, including three CpGs within genes: *COMMD1*, *TMOD1*, and *FHOD1*.

6 Discussion

During the last 5 years, both GWAS and EWAS have identified an expanding number of genetic loci for DKD. Nearly 80 genetic loci have reached genome-wide statistical significance for DKD, albuminuria, or eGFR in diabetes to date. Much of this increase is not only due to larger meta-analyses of existing diabetes cohorts but also due to CKD studies in the general population including a substantial number of individuals with diabetes, as well as general population biobank studies. Even larger meta-analyses combining multiple biobank studies are likely to result in more genetic loci contributing to DKD. One of the major challenges of such studies will be how to best ascertain cases with DKD, either based on ICD codes that do not capture DKD well, self-reported DKD, or single measurements of albuminuria or eGFR, both of which vary over time. General population biobanks may also be affected by selection bias including healthier than average individuals (141), leading to a limited number of individuals with severe DKD or ESKD or with long-lasting diabetes: As DKD takes decades to develop (6), ideal study controls would only include individuals with diabetes without DKD despite a long diabetes duration.

The number of identified genetic loci now also allows comparison of the findings and the genetic overlap between general population CKD and DKD in T1D and T2D. The general population loci for eGFR seem to affect eGFR also in individuals with diabetes, especially those with T2D (76). For some variants, the effect size is markedly higher in the individuals with diabetes than in those without (e.g., UMOD, rs77924615, $beta_{DM} = -0.019$, $beta_{noDM} = -0.011$, $P_{diff} = 1.3 \times 10^{-27}$; *TPPP*, rs4663171, $beta_{DM} = -0.011$, $beta_{noDM} = -0.004$; $P_{diff} = 2.5 \times 10^{-9}$), potentially reflecting the elevated risk and accumulated risk factors for kidney complications among individuals with diabetes. On the other hand, genetic risk factors for DKD in T1D seem to differ from the general population (52). These support the notions from the clinical and epidemiological studies suggesting that individuals with T2D can have either DKD, non-DKD, or both, whereby individuals with T1D mainly develop diabetic nephropathy with a different pathophysiology from the general CKD (11, 90). Therefore, future genetic studies on DKD will need to balance between maximizing the number of samples (any diabetes, or even the general population with focus on diabetes) but with a more heterogeneous phenotype, and a cleaner DKD phenotype in T1D with diabetic nephropathy as a more likely underlying cause, but with a more limited number of samples.

GWASs on DKD have been performed in various populations beyond the European ancestry (46–48), and some of the identified variants are population-specific, e.g., the *APOL1* variants associated with all-cause and diabetic ESKD in AAs (48, 67, 68). For many complex diseases, such as T2D, extension to further populations, as well as larger multi-ancestry GWAS meta-analyses have yielded novel genetic susceptibility loci by increasing the total sample size and capturing additional variants with ancestry-correlated heterogeneity in the allelic effect sizes (104, 142). Multi-ancestry GWASs also provide improved fine-mapping resolution of the detected association signals, i.e., can provide a smaller number of variants in the credible set including the underlying causal variant among the many associated ones (142). Therefore, such multi-ancestry studies are likely to reveal novel loci with improved fine-mapping for DKD as well. On the contrary, homogenous study populations may be particularly important in sequencing studies aiming to identify rare genetic risk factors for DKD.

Although there are known differences in the methylation pattern of a number of CpGs between different ethnicities (143), there is a lack of ethnic diversity in EWAS, which are based mainly on individuals of European ancestry (144, 145). A recent multi-ancestry EWAS on kidney function (135) revealed several population-specific methylation patterns for eGFR in the general population with little overlap between African and European populations. These discrepancies, however, could be due to both genetic and environmental differences between the different ethnic groups. The expansion of EWAS datasets in DKD to include multi-ancestry populations is still lacking.

The GWASs have also enabled creation of polygenic risk scores (PRSs) that may be used for risk stratification and identification of affected traits and phenotypes. In general population, PRS on eGFR was associated with incident CKD and kidney failure in the Atherosclerosis Risk in Communities study with 8.6% of the individuals having diabetes (146). In diabetes, smaller studies have shown that genetic risk scores for DKD improved the prediction of DKD in Han Chinese with T2D (147). In the ADjuVANt Chemotherapy in the Elderly (ADVANCE) trial with individuals with T2D, a multi-phenotype PRS, based on variants from the general population GWAS, predicted micro- and macrovascular complications and suggested that the PRS can identify high-risk individuals, who would benefit from intensified diabetes treatment (148); similarly, a general population PRS for coronary artery disease (CAD) was associated with CAD also among individuals with T1D (149). However, no large-scale PRS for DKD have yet been published, and larger GWASs on DKD are needed to create diabetes-specific PRS for DKD and to assess their utility compared to general population PRS.

To date, several CpG sites with altered methylation levels in DKD have been identified across the genome. Understanding the underlying mechanism behind these changes would be critical, i.e., are the observed changes driven by kidney disease or some other manifestation that emerges as the disease progress, and whether the changes are causal for the development or progression of DKD. In addition, methylation levels are also influenced by the genetics. Insights to the complex network behind the findings might therefore require integrating DNA methylation results with results from multiple other sources such as GWAS as well as transcriptomic and proteomic data. Some efforts in that direction have already been made. Indeed, a recent study demonstrated that DNA methylation explains a larger fraction of kidney disease heritability than gene expression by integrating GWAS data with methylomic and transcriptomic data obtained from 446 kidney tissue samples (88).

DNA methylation markers have proven useful for the prediction of DKD progression. Current studies, however, have focused on the later stages of kidney disease, when AER is severely increased or when kidney failure has occurred. EWASs at earlier stages of DKD, when AER is only moderately increased, could potentially identify additional CpGs and perhaps even more importantly, enable the prediction of early changes using DNA methylation. Although DNA methylation scores have not yet been as extensively implemented in risk prediction as the PRSs, methylation scores show a great promise as they incorporate information from both the genes and the environment. In a recent study, methylation scores improved the prediction of a range of clinical diagnoses and traits, including kidney disease, outperforming the predictive ability of polygenetic risk scores (140). However, the dynamic nature of methylation as well as its tissue-specificity introduces limitations regarding causality, time span of effect, and target tissue. By incorporating genetic information, causality can be addressed, and future studies may also be facilitated by emerging single-cell sequencing technologies that enable more targeted analyses, such as exploring the causal effects of DNA methylation at the single-cell level in the kidneys.

Author contributions

NS and ED revised the literature and wrote the manuscript. P-HG critically revised the manuscript for the scientific content. All authors agree to be accountable for the content of the work. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from Folkhälsan Research Foundation, Wilhelm and Else Stockmann Foundation, “Liv och Hälsa” Society, Sigrid Jusélius Foundation, Helsinki University Central Hospital Research Funds (TYH2023403), and Academy of Finland (316664).

Conflict of interest

P-HG has received investigator-initiated research grants from Eli Lilly and Roche; is an advisory board member for AbbVie, Astellas, AstraZeneca, Bayer, Boehringer Ingelheim, Cebix, Eli Lilly, Janssen, Medscape, Merck Sharp & Dohme, Mundipharma, Nestlé, Novartis, Novo Nordisk, and Sanofi; and has received lecture fees from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Elo Water,

Genzyme, Merck Sharp & Dohme, Medscape, Novartis, Novo Nordisk, PeerVoice, Sanofi, and Sciarc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1163001/full#supplementary-material

References

1. International Diabetes Federation. *IDF diabetes atlas* (2021). Available at: https://diabetesatlas.org/ (Accessed August 26, 2022).

2. Brownlee M. Biochemistry and molecular cell biology of diabetic complications. *Nature* (2001) 414:813–20. doi: 10.1038/414813a

3. American Diabetes Association. Economic costs of diabetes in the U.S. in 2007. *Diabetes Care* (2008) 31:596–615. doi: 10.2337/dc08-9017

4. Lithovius R, Harjutsalo V, Forsblom C, Groop PHFinnDiane Study Group. Cumulative cost of prescription medication in outpatients with type 1 diabetes in Finland. *Diabetologia* (2011) 54:496–503. doi: 10.1007/s00125-010-1999-y

5. Costacou T, Orchard TJ. Cumulative kidney complication risk by 50 years of type 1 diabetes: the effects of sex, age, and calendar year at onset. *Diabetes Care* (2017) 41:426–33. doi: 10.2337/dc17-1118

6. Jansson Sigfrids F, Groop P-H, Harjutsalo V. Incidence rate patterns, cumulative incidence, and time trends for moderate and severe albuminuria in individuals diagnosed with type 1 diabetes aged 0-14 years: a population-based retrospective cohort study. *Lancet Diabetes Endocrinol* (2022) 10:489–98. doi: 10.1016/S2213-8587(22)00099-7

7. Thomas MC, Weekes AJ, Broadley OJ, Cooper ME, Mathew TH. The burden of chronic kidney disease in Australian patients with type 2 diabetes (the NEFRON study). *Med J Aust* (2006) 185:140–4. doi: 10.5694/j.1326-5377.2006.tb00499.x

8. Harjutsalo V, Maric C, Forsblom C, Thorn L, Waden J, Groop PH, et al. Sex-related differences in the long-term risk of microvascular complications by age at onset of type 1 diabetes. *Diabetologia* (2011) 54:1992–9. doi: 10.1007/s00125-011-2144-2

9. Schernthaner G, Schernthaner GH. Diabetic nephropathy: new approaches for improving glycemic control and reducing risk. *JNephrol* (2013) 26:975–85. doi: 10.5301/jn.5000281

10. Tuomilehto J, Borch-Johnsen K, Molarius A, Forsen T, Rastenyte D, Sarti C, et al. Incidence of cardiovascular disease in type 1 (insulin-dependent) diabetic subjects with and without diabetic nephropathy in Finland. *Diabetologia* (1998) 41:784–90. doi: 10.1007/s001250050988

11. Anders H-J, Huber TB, Isermann B, Schiffer M. CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. *Nat Rev Nephrol* (2018) 14:361–77. doi: 10.1038/s41581-018-0001-y

12. Najafian B, Mauer M. Morphologic features of declining renal function in type 1 diabetes. *Semin Nephrol* (2012) 32:415–22. doi: 10.1016/j.semnephrol.2012.07.003

13. Quinn M, Angelico MC, Warram JH, Krolewski AS. Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia* (1996) 39:940–5. doi: 10.1007/BF00403913

14. Seaquist ER, Goetz FC, Rich S, Barbosa J. Familial clustering of diabetic kidney disease. *NEnglJMed* (1989) 320:1161–5. doi: 10.1056/NEJM198905043201801

15. Earle K, Walker J, Hill C, Viberti G. Familial clustering of cardiovascular disease in patients with insulin-dependent diabetes and nephropathy. *NEnglJMed* (1992) 326:673–7. doi: 10.1056/NEJM199203053261005

16. Harjutsalo V, Katoh S, Sarti C, Tajima N, Tuomilehto J. Population-based assessment of familial clustering of diabetic nephropathy in type 1 diabetes. *Diabetes* (2004) 53:2449–54. doi: 10.2337/diabetes.53.9.2449

17. Borch-Johnsen K, Norgaard K, Hommel E, Mathiesen ER, Jensen JS, Deckert T, et al. Is diabetic nephropathy an inherited complication. *Kidney Int* (1992) 41:719–22. doi: 10.1038/ki.1992.112

18. Sandholm N, Van Zuydam N, Ahlqvist E, Juliusdottir T, Deshmukh HA, Rayner NW, et al. The genetic landscape of renal complications in type 1 diabetes. *J Am Soc Nephrol* (2017) 28:557–74. doi: 10.1681/ASN.2016020231

19. van Zuydam NR, Ahlqvist E, Sandholm N, Deshmukh H, Rayner NW, Abdalla M, et al. A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. *Diabetes* (2018) 67:1414–27. doi: 10.2337/db17-0914

20. Kim J, Jensen A, Ko S, Raghavan S, Phillips LS, Hung A, et al. Systematic heritability and heritability enrichment analysis for diabetes complications in UK biobank and ACCORD studies. *Diabetes* (2022) 71:1137–48. doi: 10.2337/db21-0839

21. Fogarty DG, Rich SS, Hanna L, Warram JH, Krolewski AS. Urinary albumin excretion in families with type 2 diabetes is heritable and genetically correlated to blood pressure. *Kidney Int* (2000) 57:250–7. doi: 10.1046/j.1523-1755.2000.00833.x

22. Forsblom CM, Kanninen T, Lehtovirta M, Saloranta C, Groop LC. Heritability of albumin excretion rate in families of patients with type II diabetes. *Diabetologia* (1999) 42:1359–66. doi: 10.1007/s001250051450

23. Krolewski AS, Poznik GD, Placha G, Canani L, Dunn J, Walker W, et al. A genome-wide linkage scan for genes controlling variation in urinary albumin excretion in type II diabetes. *Kidney Int* (2006) 69:129–36. doi: 10.1038/sj.ki.5000023

24. Langefeld CD, Beck SR, Bowden DW, Rich SS, Wagenknecht LE, Freedman BI. Heritability of GFR and albuminuria in caucasians with type 2 diabetes mellitus. *Am J Kidney Dis* (2004) 43:796–800. doi: 10.1053/j.ajkd.2003.12.043

25. Sandholm N, Forsblom C, Makinen VP, McKnight AJ, Osterholm AM, He B, et al. Genome-wide association study of urinary albumin excretion rate in patients with type 1 diabetes. *Diabetologia* (2014) 57:1143–53. doi: 10.1007/s00125-014-3202-3

26. Moczulski DK, Rogus JJ, Antonellis A, Warram JH, Krolewski AS. Major susceptibility locus for nephropathy in type 1 diabetes on chromosome 3q: results of novel discordant sib-pair analysis. *Diabetes* (1998) 47:1164. doi: 10.2337/diabetes.47.7.1164

27. Osterholm AM, He B, Pitkaniemi J, Albinsson L, Berg T, Sarti C, et al. Genome-wide scan for type 1 diabetic nephropathy in the Finnish population reveals suggestive linkage to a single locus on chromosome 3q. *Kidney Int* (2007) 71:140–5. doi: 10.1038/sj.ki.5001933

28. Rogus JJ, Poznik GD, Pezzolesi MG, Smiles AM, Dunn J, Walker W, et al. High-density single nucleotide polymorphism genome-wide linkage scan for susceptibility genes for diabetic nephropathy in type 1 diabetes. *Diabetes* (2008) 57:2519–26. doi: 10.2337/db07-1086

29. Wessman M, Forsblom C, Kaunisto MA, Soderlund J, Ilonen J, Sallinen R, et al. Novel susceptibility locus at 22q11 for diabetic nephropathy in type 1 diabetes. *PloS One* (2011) 6:e24053. doi: 10.1371/journal.pone.0024053

30. Imperatore G, Hanson RL, Pettitt DJ, Kobes S, Bennett PH, Knowler WC. Sib-Pair linkage analysis for susceptibility genes for microvascular complications among pima indians with type 2 diabetes. *Pima Diabetes Genes Group Diabetes* (1998) 47:821–30. doi: 10.2337/diabetes.47.5.821

31. Bowden DW, Colicigno CJ, Langefeld CD, Sale MM, Williams A, Anderson PJ, et al. A genome scan for diabetic nephropathy in African americans. *Kidney Int* (2004) 66:1517–26. doi: 10.1111/j.1523-1755.2004.00915.x

32. Vionnet N, Tregouet D, Kazeem G, Gut I, Groop PH, Tarnow L, et al. Analysis of 14 candidate genes for diabetic nephropathy on chromosome 3q in European populations: strongest evidence for association with a variant in the promoter region of the adiponectin gene. *Diabetes* (2006) 55:3166–74. doi: 10.2337/db06-0271

33. He B, Österholm AM, Hoverfält A, Forsblom C, Hjörleifsdóttir EE, Nilsson AS, et al. Association of genetic variants at 3q22 with nephropathy in patients with type 1 diabetes mellitus. *Am J Hum Genet* (2009) 84:5–13. doi: 10.1016/j.ajhg.2008.11.012

34. Vardarli I, Baier LJ, Hanson RL, Akkoyun I, Fischer C, Rohmeiss P, et al. Gene for susceptibility to diabetic nephropathy in type 2 diabetes maps to 18q22.3-23. *Kidney Int* (2002) 62:2176–83. doi: 10.1046/j.1523-1755.2002.00663.x

35. Janssen B, Hohenadel D, Brinkkoetter P, Peters V, Rind N, Fischer C, et al. Carnosine as a protective factor in diabetic nephropathy: association with a leucine repeat of the carnosinase gene CNDP1. *Diabetes* (2005) 54:2320–7. doi: 10.2337/diabetes.54.8.2320

36. Mooyaart A, Valk EJJ, van Es L, Bruijn J, de Heer E, Freedman B, et al. Genetic associations in diabetic nephropathy: a meta-analysis. *Diabetologia* (2011) 54:544–53. doi: 10.1007/s00125-010-1996-1

37. Tong Z, Yang Z, Patel S, Chen H, Gibbs D, Yang X, et al. Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. *ProcNatlAcadSciUSA* (2008) 105:6998–7003. doi: 10.1073/pnas.0800454105

38. Porta M, Toppila I, Sandholm N, Hosseini SM, Forsblom C, Hietala K, et al. Variation in SLC19A3 and protection from microvascular damage in type 1 diabetes. *Diabetes* (2016) 65:1022–30. doi: 10.2337/db15-1247

39. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* (2007) 445:881–5. doi: 10.1038/nature05616

40. Diabetes Genetics Initiative of Broad Institute of Harvard, MIT, Lund University and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* (2007) 316:1331–6. doi: 10.1126/science.1142358

41. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science* (2007) 316:1341–5. doi: 10.1126/science.1142382

42. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *NatGenet* (2007) 39:857–64. doi: 10.1038/ng2068

43. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* (2001) 17:502–10. doi: 10.1016/s0168-9525(01)02410-6

44. Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, et al. New susceptibility loci associated with kidney disease in type 1 diabetes. *PloS Genet* (2012) 8:e1002921. doi: 10.1371/journal.pgen.1002921

45. Sandholm N, McKnight AJ, Salem RM, Brennan EP, Forsblom C, Harjutsalo V, et al. Chromosome 2q31.1 associates with ESRD in women with type 1 diabetes. *JAmSocNephrol* (2013) 24:1537–43. doi: 10.1681/ASN.2012111122

46. Iyengar SK, Sedor JR, Freedman BI, Kao WH, Kretzler M, Keller BJ, et al. Genome-wide association and trans-ethnic meta-analysis for advanced diabetic kidney disease: family investigation of nephropathy and diabetes (FIND). *PloS Genet* (2015) 11:e1005352. doi: 10.1371/journal.pgen.1005352

47. Taira M, Imamura M, Takahashi A, Kamatani Y, Yamauchi T, Araki SI, et al. A variant within the FTO confers susceptibility to diabetic nephropathy in Japanese patients with type 2 diabetes. *PloS One* (2018) 13:e0208654. doi: 10.1371/journal.pone.0208654

48. Guan M, Keaton JM, Dimitrov L, Hicks PJ, Xu J, Palmer ND, et al. Genome-wide association study identifies novel loci for type 2 diabetes-attributed end-stage kidney disease in African americans. *Hum Genomics* (2019) 13:21. doi: 10.1186/s40246-019-0205-7

49. Salem RM, Todd JN, Sandholm N, Cole JB, Chen W-M, Andrews D, et al. Genome-wide association study of diabetic kidney disease highlights biology involved in glomerular basement membrane collagen. *J Am Soc Nephrol* (2019) 30:2000–16. doi: 10.1681/ASN.2019030218

50. Khattab A, Torkamani A. Nidogen-1 could play a role in diabetic kidney disease development in type 2 diabetes: a genome-wide association meta-analysis. *Hum Genomics* (2022) 16:47. doi: 10.1186/s40246-022-00422-y

51. Haukka J, Sandholm N, Valo E, Forsblom C, Harjutsalo V, Cole JB, et al. Novel linkage peaks discovered for diabetic nephropathy in individuals with type 1 diabetes. *Diabetes* (2021) 70:986–95. doi: 10.2337/db20-0158

52. Sandholm N, Cole JB, Nair V, Sheng X, Liu H, Ahlqvist E, et al. Genome-wide meta-analysis and omics integration identifies novel genes associated with diabetic kidney disease. *Diabetologia* (2022) 65:1495–509. doi: 10.1007/s00125-022-05735-0

53. Pan Y, Sun X, Mi X, Huang Z, Hsu Y, Hixson JE, et al. Whole-exome sequencing study identifies four novel gene loci associated with diabetic kidney disease. *Hum Mol Genet* (2022), 32:1048-60 ddac290. doi: 10.1093/hmg/ddac290

54. Freedman BI, Langefeld CD, Lu L, Divers J, Comeau ME, Kopp JB, et al. Differential effects of MYH9 and APOL1 risk variants on FRMD3 association with diabetic ESRD in African americans. *PloS Genet* (2011) 7:e1002150. doi: 10.1371/journal.pgen.1002150

55. Pezzolesi MG, Jeong J, Smiles AM, Skupien J, Mychaleckyj JC, Rich SS, et al. Family-based association analysis confirms the role of the chromosome 9q21.32 locus in the susceptibility of diabetic nephropathy. *PloS One* (2013) 8:e60301. doi: 10.1371/journal.pone.0060301

56. Pezzolesi MG, Skupien J, Mychaleckyj JC, Warram JH, Krolewski AS. Insights to the genetics of diabetic nephropathy through a genome-wide association study of the GoKinD collection. *SeminNephrol* (2010) 30:126–40. doi: 10.1016/j.semnephrol.2010.01.004

57. Germain M, Pezzolesi MG, Sandholm N, McKnight AJ, Susztak K, Lajer M, et al. SORBS1 gene, a new candidate for diabetic nephropathy: results from a multi-stage genome-wide association study in patients with type 1 diabetes. *Diabetologia* (2015) 58:543–8. doi: 10.1007/s00125-014-3459-6

58. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U.S.A.* (2009) 106:9362–7. doi: 10.1073/pnas.0903103106

59. Project Consortium ENCODE, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489:57–74. doi: 10.1038/nature11247

60. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Large-Scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* (2021) 53:1300–10. doi: 10.1038/s41588-021-00913-z

61. Stoner M, Wang F, Wormke M, Nguyen T, Samudio I, Vyhlidal C, et al. Inhibition of vascular endothelial growth factor expression in HEC1A endometrial cancer cells through interactions of estrogen receptor alpha and Sp3 proteins. *JBiolChem* (2000) 275:22769–79. doi: 10.1074/jbc.M002188200

62. Ziera T, Irlbacher H, Fromm A, Latouche C, Krug SM, Fromm M, et al. Cnksr3 is a direct mineralocorticoid receptor target gene and plays a key role in the regulation of the epithelial sodium channel. *FASEB J* (2009) 23:3936–46. doi: 10.1096/fj.09-134759

63. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD. The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. *Collab Study Group NEnglJMed* (1993) 329:1456–62. doi: 10.1056/NEJM199311113292004

64. Brenner BM, Cooper ME, de Zeeuw D, Keane WF, Mitch WE, Parving HH, et al. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *NEnglJMed* (2001) 345:861–9. doi: 10.1056/NEJMoa011161

65. Bakris GL, Agarwal R, Anker SD, Pitt B, Ruilope LM, Rossing P, et al. Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. *N Engl J Med* (2020) 383:2219–29. doi: 10.1056/NEJMoa2025845

66. Takahashi T, Takahashi K, Gerety S, Wang H, Anderson DJ, Daniel TO. Temporally compartmentalized expression of ephrin-B2 during renal glomerular development. *J Am Soc Nephrol* (2001) 12:2673–82. doi: 10.1681/ASN.V12122673

67. Tzur S, Rosset S, Shemer R, Yudkovsky G, Selig S, Tarekegn A, et al. Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene. *Hum Genet* (2010) 128:345–50. doi: 10.1007/s00439-010-0861-0

68. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, et al. Association of trypanolytic ApoL1 variants with kidney disease in African americans. *Science* (2010) 329:841–5. doi: 10.1126/science.1193032

69. Loos RJF, Yeo GSH. The bigger picture of FTO: the first GWAS-identified obesity gene. *Nat Rev Endocrinol* (2014) 10:51–61. doi: 10.1038/nrendo.2013.227

70. Todd JN, Dahlstrom EH, Salem RM, Sandholm N, Forsblom C, Group FS, et al. Genetic evidence for a causal role of obesity in diabetic kidney disease. *Diabetes* (2015) 64:4238–46. doi: 10.2337/db15-0254

71. Sobreira DR, Joslin AC, Zhang Q, Williamson I, Hansen GT, Farris KM, et al. Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of IRX3 and IRX5. *Science* (2021) 372:1085–91. doi: 10.1126/science.abf1008

72. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* (2021) 53:1415–24. doi: 10.1038/s41588-021-00931-x

73. Köttgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *NatGenet* (2009) 41:712–7. doi: 10.1038/ng.377

74. Böger CA, Gorski M, Li M, Hoffmann MM, Huang C, Yang Q, et al. Association of eGFR-related loci identified by GWAS with incident CKD and ESRD. *PloS Genet* (2011) 7:e1002292. doi: 10.1371/journal.pgen.1002292

75. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *NatCommun* (2016) 7:10023. doi: 10.1038/ncomms10023

76. Winkler TW, Rasheed H, Teumer A, Gorski M, Rowan BX, Stanzick KJ, et al. Differential and shared genetic effects on kidney function between diabetic and non-diabetic individuals. *Commun Biol* (2022) 5:580. doi: 10.1038/s42003-022-03448-z

77. Smyth LJ, Dahlström EH, Syreeni A, Kerr K, Kilner J, Doyle R, et al. Epigenome-wide meta-analysis identifies DNA methylation biomarkers associated with diabetic kidney disease. *Nat Commun* (2022) 13:7891. doi: 10.1038/s41467-022-34963-6

78. Schlosser P, Tin A, Matias-Garcia PR, Thio CHL, Joehanes R, Liu H, et al. Meta-analyses identify DNA methylation associated with kidney function and damage. *Nat Commun* (2021) 12:7174. doi: 10.1038/s41467-021-27234-3

79. Dziadek M. Role of laminin-nidogen complexes in basement membrane formation during embryonic development. *Experientia* (1995) 51:901–13. doi: 10.1007/BF01921740

80. Sambo F, Malovini A, Sandholm N, Stavarachi M, Forsblom C, Makinen VP, et al. Novel genetic susceptibility loci for diabetic end-stage renal disease identified through robust naive bayes classification. *Diabetologia* (2014) 57:1611–22. doi: 10.1007/s00125-014-3256-2

81. Sandholm N, Haukka JK, Toppila I, Valo E, Harjutsalo V, Forsblom C, et al. Confirmation of GLRA3 as a susceptibility locus for albuminuria in Finnish patients with type 1 diabetes. *Sci Rep* (2018) 8:12408. doi: 10.1038/s41598-018-29211-1

82. Teumer A, Li Y, Ghasemi S, Prins BP, Wuttke M, Hermle T, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun* (2019) 10:4130. doi: 10.1038/s41467-019-11576-0

83. Li C, Liu C, Nissim I, Chen J, Chen P, Doliba N, et al. Regulation of glucagon secretion in normal and diabetic human islets by gamma-hydroxybutyrate and glycine. *JBiolChem* (2013) 288:3938–51. doi: 10.1074/jbc.M112.385682

84. Climie RE, Srikanth V, Keith LJ, Davies JE, Sharman JE. Exercise excess pressure and exercise-induced albuminuria in patients with type 2 diabetes mellitus. *AmJPhysiolHeart CircPhysiol* (2015) 308:H1136–42. doi: 10.1152/ajpheart.00739.2014

85. Kim G-H. Renal effects of prostaglandins and cyclooxygenase-2 inhibitors. *Electrolyte Blood Press* (2008) 6:35–41. doi: 10.5049/EBP.2008.6.1.35

86. Böger CA, Chen MH, Tin A, Olden M, Kottgen A, de Boer IH, et al. CUBN is a gene locus for albuminuria. *JAmSocNephrol* (2011) 22:555–70. doi: 10.1681/ASN.2010060598

87. Teumer A, Tin A, Sorice R, Gorski M, Yeo NC, Chu AY, et al. Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes* (2016) 65:803–17. doi: 10.2337/db15-1313

88. Liu H, Doke T, Guo D, Sheng X, Ma Z, Park J, et al. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nat Genet* (2022) 54:950–62. doi: 10.1038/s41588-022-01097-w

89. Chronic Kidney Disease Collaboration GBD. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet* (2020) 395:709–33. doi: 10.1016/S0140-6736(20)30045-3

90. Di Vincenzo A, Bettini S, Russo L, Mazzocut S, Mauer M, Fioretto P. Renal structure in type 2 diabetes: facts and misconceptions. *J Nephrol* (2020) 33:901–7. doi: 10.1007/s40620-020-00797-y

91. de Boer IH, Khunti K, Sadusky T, Tuttle KR, Neumiller JJ, Rhee CM, et al. Diabetes management in chronic kidney disease: a consensus report by the American diabetes association (ADA) and kidney disease: improving global outcomes (KDIGO). *Diabetes Care* (2022) 45:3075–90. doi: 10.2337/dci22-0027

92. Gorski M, van der Most PJ, Teumer A, Chu AY, Li M, Mijatovic V, et al. Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci Rep* (2017) 7:45040. doi: 10.1038/srep45040

93. Gorski M, Rasheed H, Teumer A, Thomas LF, Graham SE, Sveinbjornsson G, et al. Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies. *Kidney Int* (2022) 102:624–39. doi: 10.1016/j.kint.2022.05.021

94. Ahluwalia TS, Schulz CA, Waage J, Skaaby T, Sandholm N, van Zuydam N, et al. A novel rare CUBN variant and three additional genes identified in europeans with and without diabetes: results from an exome-wide association study of albuminuria. *Diabetologia* (2019) 62:292–305. doi: 10.1007/s00125-018-4783-z

95. Uglebjerg N, Ahmadizar F, Aly DM, Cañadas-Garre M, Hill C, Naber A, et al. Four missense genetic variants in CUBN are associated with higher levels of eGFR in non-diabetes but not in diabetes mellitus or its subtypes: a genetic association study in europeans. *Front Endocrinol (Lausanne)* (2023) 14:1081741. doi: 10.3389/fendo.2023.1081741

96. Voskarides K, Damianou L, Neocleous V, Zouvani I, Christodoulidou S, Hadjiconstantinou V, et al. COL4A3/COL4A4 mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy. *J Am Soc Nephrol* (2007) 18:3004–16. doi: 10.1681/ASN.2007040444

97. Jukarainen S, Kiiskinen T, Kuitunen S, Havulinna AS, Karjalainen J, Cordioli M, et al. Genetic risk factors have a substantial impact on healthy life years. *Nat Med* (2022) 28:1893–901. doi: 10.1038/s41591-022-01957-2

98. Jurgens SJ, Choi SH, Morrill VN, Chaffin M, Pirruccello JP, Halford JL, et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK biobank. *Nat Genet* (2022) 54:240–50. doi: 10.1038/s41588-021-01011-w

99. Akilesh S, Suleiman H, Yu H, Stander MC, Lavin P, Gbadegesin R, et al. Arhgap24 inactivates Rac1 in mouse podocytes, and a mutant form is associated with familial focal segmental glomerulosclerosis. *J Clin Invest* (2011) 121:4127–37. doi: 10.1172/JCI46458

100. Pitera JE, Scambler PJ, Woolf AS. Fras1, a basement membrane-associated protein mutated in Fraser syndrome, mediates both the initiation of the mammalian kidney and the integrity of renal glomeruli. *Hum Mol Genet* (2008) 17:3953–64. doi: 10.1093/hmg/ddn297

101. Mychaleckyj J, Valo E, Ichimura T, Ahluwalia T, Dina C, Miller R, et al. Association of coding variants in hydroxysteroid 17-beta dehydrogenase 14 (HSD17B14) with reduced progression to end stage kidney disease in type 1 diabetes. *J Am Soc Nephrol* (2021), 32:2634-51 ASN.2020101457. doi: 10.1681/ASN.2020101457

102. Si Y, Vanderwerff B, Zöllner S. Why are rare variants hard to impute? Coalescent models reveal theoretical limits in existing algorithms. *Genetics* (2021) 217:iyab011. doi: 10.1093/genetics/iyab011

103. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* (2019) 570:71–6. doi: 10.1038/s41586-019-1231-2

104. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet* (2020) 52:680–91. doi: 10.1038/s41588-020-0637-y

105. Guo J, Rackham OJL, Sandholm N, He B, Osterholm AM, Valo E, et al. Whole-genome sequencing of Finnish type 1 diabetic siblings discordant for kidney disease reveals DNA variants associated with diabetic nephropathy. *JAmSocNephrol* (2020) 31:309–23. doi: 10.1681/ASN.2019030289

106. Lin BM, Grinde KE, Brody JA, Breeze CE, Raffield LM, Mychaleckyj JC, et al. Whole genome sequence analyses of eGFR in 23,732 people representing multiple ancestries in the NHLBI trans-omics for precision medicine (TOPMed) consortium. *eBioMedicine* (2021) 63:103157. doi: 10.1016/j.ebiom.2020.103157

107. Ding H, Zhang L, Yang Q, Zhang X, Li X. Epigenetics in kidney diseases. *Adv Clin Chem* (2021) 104:233–97. doi: 10.1016/bs.acc.2020.09.005

108. Balnis J, Madrid A, Hogan KJ, Drake LA, Adhikari A, Vancavage R, et al. Persistent blood DNA methylation changes one year after SARS-CoV-2 infection. *Clin Epigenet* (2022) 14:94. doi: 10.1186/s13148-022-01313-8

109. Miller RG, Mychaleckyj JC, Onengut-Gumuscu S, Orchard TJ, Costacou T. TXNIP DNA methylation is associated with glycemic control over 28 years in type 1 diabetes: findings from the Pittsburgh epidemiology of diabetes complications (EDC) study. *BMJ Open Diabetes Res Care* (2023) 11:e003068. doi: 10.1136/bmjdrc-2022-003068

110. Miao F, Chen Z, Genuth S, Paterson A, Zhang L, Wu X, et al. Evaluating the role of epigenetic histone modifications in the metabolic memory of type 1 diabetes. *Diabetes* (2014) 63:1748–62. doi: 10.2337/db13-1251

111. Chen Z, Miao F, Paterson AD, Lachin JM, Zhang L, Schones DE, et al. Epigenomic profiling reveals an association between persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort. *Proc Natl Acad Sci U.S.A.* (2016) 113:E3002–3011. doi: 10.1073/pnas.1603712113

112. Writing Team for the DCCT/EDIC Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the epidemiology of diabetes interventions and complications (EDIC) study. *JAMA* (2003) 290:2159–67. doi: 10.1001/jama.290.16.2159

113. DCCT/EDIC Research Group, de Boer IH, Sun W, Cleary PA, Lachin JM, Molitch ME, et al. Intensive diabetes therapy and glomerular filtration rate in type 1 diabetes. *NEnglJMed* (2011) 365:2366–76. doi: 10.1056/NEJMoa1111732

114. Chen Z, Miao F, Braffett BH, Lachin JM, Zhang L, Wu X, et al. DNA Methylation mediates development of HbA1c-associated complications in type 1 diabetes. *Nat Metab* (2020) 2:744–62. doi: 10.1038/s42255-020-0231-8

115. Jia Y, Reddy MA, Das S, Oh HJ, Abdollahi M, Yuan H, et al. Dysregulation of histone H3 lysine 27 trimethylation in transforming growth factor-β1-induced gene expression in mesangial cells and diabetic kidney. *J Biol Chem* (2019) 294:12695–707. doi: 10.1074/jbc.RA119.007575

116. Keating ST, van Diepen JA, Riksen NP, El-Osta A. Epigenetics in diabetic nephropathy, immunity and metabolism. *Diabetologia* (2018) 61:6–20. doi: 10.1007/s00125-017-4490-1

117. Park J, Guan Y, Sheng X, Gluck C, Seasock MJ, Hakimi AA, et al. Functional methylome analysis of human diabetic kidney disease. *JCI Insight* (2019) 4:e128886. doi: 10.1172/jci.insight.128886

118. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics* (2011) 98:288–95. doi: 10.1016/j.ygeno.2011.07.007

119. Saffari A, Silver MJ, Zavattari P, Moi L, Columbano A, Meaburn EL, et al. Estimation of a significance threshold for epigenome-wide association studies. *Genet Epidemiol* (2018) 42:20–33. doi: 10.1002/gepi.22086

120. Bell CG, Teschendorff AE, Rakyan VK, Maxwell AP, Beck S, Savage DA. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med Genomics* (2010) 3:33. doi: 10.1186/1755-8794-3-33

121. Sapienza C, Lee J, Powell J, Erinle O, Yafai F, Reichert J, et al. DNA Methylation profiling identifies epigenetic differences between diabetes patients with ESRD and diabetes patients without nephropathy. *Epigenetics* (2011) 6:20–8. doi: 10.4161/epi.6.1.13362

122. Smyth LJ, McKay GJ, Maxwell AP, McKnight AJ. DNA Hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics* (2014) 9:366–76. doi: 10.4161/epi.27161

123. Swan EJ, Maxwell AP, McKnight AJ. Distinct methylation patterns in genes that affect mitochondrial function are associated with kidney disease in blood-derived DNA from individuals with type 1 diabetes. *DiabetMed* (2015) 32:1110–5. doi: 10.1111/dme.12775

124. Qiu C, Hanson RL, Fufaa G, Kobes S, Gluck C, Huang J, et al. Cytosine methylation predicts renal function decline in American indians. *Kidney Int* (2018) 93:1417–31. doi: 10.1016/j.kint.2018.01.036

125. Gluck C, Qiu C, Han SY, Palmer M, Park J, Ko Y-A, et al. Kidney cytosine methylation changes improve renal function decline estimation in patients with diabetic kidney disease. *Nat Commun* (2019) 10:2461. doi: 10.1038/s41467-019-10378-8

126. Sheng X, Qiu C, Liu H, Gluck C, Hsu JY, He J, et al. Systematic integrated analysis of genetic and epigenetic variation in diabetic kidney disease. *PNAS* (2020) 117:29013–24. doi: 10.1073/pnas.2005905117

127. Smyth LJ, Patterson CC, Swan EJ, Maxwell AP, McKnight AJ. DNA Methylation associated with diabetic kidney disease in blood-derived DNA. *Front Cell Dev Biol* (2020) 8:561907. doi: 10.3389/fcell.2020.561907

128. Kim H, Bae JH, Park KS, Sung J, Kwak SH. DNA Methylation changes associated with type 2 diabetes and diabetic kidney disease in an East Asian population. *J Clin Endocrinol Metab* (2021) 106:e3837–51. doi: 10.1210/clinem/dgab488

129. Smyth LJ, Kilner J, Nair V, Liu H, Brennan E, Kerr K, et al. Assessment of differentially methylated loci in individuals with end-stage kidney disease attributed to diabetic kidney disease: an exploratory study. *Clin Epigenet* (2021) 13:99. doi: 10.1186/s13148-021-01081-x

130. Lecamwasam A, Novakovic B, Meyer B, Ekinci EI, Dwyer KM, Saffery R. DNA Methylation profiling identifies epigenetic differences between early versus late stages of

diabetic chronic kidney disease. *Nephrol Dial Transplant* (2021) 36:2027–38. doi: 10.1093/ndt/gfaa226

131. Guintivano J, Shabalin AA, Chan RF, Rubinow DR, Sullivan PF, Meltzer-Brody S, et al. Test-statistic inflation in methylome-wide association studies. *Epigenetics* (2020) 15:1163–6. doi: 10.1080/15592294.2020.1758382

132. Lecamwasam A, Sexton-Oates A, Carmody J, Ekinci EI, Dwyer KM, Saffery R. DNA Methylation profiling of genomic DNA isolated from urine in diabetic chronic kidney disease: a pilot study. *PloS One* (2018) 13:e0190280. doi: 10.1371/journal.pone.0190280

133. Tregouet DA, Groop PH, McGinn S, Forsblom C, Hadjadj S, Marre M, et al. G/T substitution in intron 1 of the UNC13B gene is associated with increased risk of nephropathy in patients with type 1 diabetes. *Diabetes* (2008) 57:2843–50. doi: 10.2337/db08-0073

134. Chu AY, Tin A, Schlosser P, Ko Y-A, Qiu C, Yao C, et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun* (2017) 8:1286. doi: 10.1038/s41467-017-01297-7

135. Breeze CE, Batorsky A, Lee MK, Szeto MD, Xu X, McCartney DL, et al. Epigenome-wide association study of kidney function identifies trans-ethnic and ethnic-specific loci. *Genome Med* (2021) 13:74. doi: 10.1186/s13073-021-00877-z

136. Chen J, Huang Y, Hui Q, Mathur R, Gwinn M, So-Armah K, et al. Epigenetic associations with estimated glomerular filtration rate among men with human immunodeficiency virus infection. *Clin Infect Dis* (2020) 70:667–73. doi: 10.1093/cid/ciz240

137. Soriano-Tárraga C, Jiménez-Conde J, Giralt-Steinhauer E, Mola-Caminal M, Vivanco-Hidalgo RM, Ois A, et al. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum Mol Genet* (2016) 25:609–19. doi: 10.1093/hmg/ddv493

138. Aslibekyan S, Claas SA. Methylation in CPT1A, lipoproteins, and epigenetics. In: Patel V, Preedy V, editors. *Handbook of nutrition, diet, and epigenetics*. (Cham, Switzerland: Springer International Publishing (2017). p. 1–17. doi: 10.1007/978-3-319-31143-2_108-1

139. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* (2021) 600:675–9. doi: 10.1038/s41586-021-04064-3

140. Thompson M, Hill BL, Rakocz N, Chiang JN, Geschwind D, Sankararaman S, et al. Methylation risk scores are associated with a collection of phenotypes within electronic health record systems. *NPJ Genom Med* (2022) 7:50. doi: 10.1038/s41525-022-00320-1

141. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol* (2017) 186:1026–34. doi: 10.1093/aje/kwx246

142. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* (2022) 54:560–72. doi: 10.1038/s41588-022-01058-3

143. Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. *Genome Biol* (2012) 13:R8. doi: 10.1186/gb-2012-13-2-r8

144. Breeze CE, Wong JYY, Beck S, Berndt SI, Franceschini N. Diversity in EWAS: current state, challenges, and solutions. *Genome Med* (2022) 14:71. doi: 10.1186/s13073-022-01065-3

145. Breeze CE, Beck S, Berndt SI, Franceschini N. The missing diversity in human epigenomic studies. *Nat Genet* (2022) 54:737–9. doi: 10.1038/s41588-022-01081-4

146. Yu Z, Jin J, Tin A, Köttgen A, Yu B, Chen J, et al. Polygenic risk scores for kidney function and their associations with circulating proteome, and incident kidney diseases. *J Am Soc Nephrol* (2021) 32:3161–73. doi: 10.1681/ASN.2020111599

147. Liao L-N, Li T-C, Li C-I, Liu C-S, Lin W-Y, Lin C-H, et al. Genetic risk score for risk prediction of diabetic nephropathy in han Chinese type 2 diabetes patients. *Sci Rep* (2019) 9:19897. doi: 10.1038/s41598-019-56400-3

148. Tremblay J, Haloui M, Attaoua R, Tahir R, Hishmih C, Harvey F, et al. Polygenic risk scores predict diabetes complications and their response to intensive blood pressure and glucose control. *Diabetologia* (2021) 64:2012–25. doi: 10.1007/s00125-021-05491-7

149. Lithovius R, Antikainen AA, Mutter S, Valo E, Forsblom C, Harjutsalo V, et al. Genetic risk score enhances coronary artery disease risk prediction in individuals with type 1 diabetes. *Diabetes Care* (2022), 45(734–41):dc210974. doi: 10.2337/dc21-0974

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 31 May 2023
DOI 10.3389/fendo.2023.1159714

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Gopi Sundaramoorthy,
Madras Diabetes Research Foundation, India
Marcia Hiriart,
Universidad Nacional Autonoma de Mexico, Mexico

*CORRESPONDENCE
Jinzhi Huang
huangjzgd@163.com
Yue Wei
weiyue138@163.com
Runmin Guo
runmin.guo@gdmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 06 February 2023
ACCEPTED 12 May 2023
PUBLISHED 31 May 2023

CITATION
Zeng Q, Tan B, Han F, Huang X, Huang J, Wei Y and Guo R (2023) Association of *solute carrier family 30 A8 zinc transporter gene* variations with gestational diabetes mellitus risk in a Chinese population. *Front. Endocrinol.* 14:1159714. doi: 10.3389/fendo.2023.1159714

Association of *solute carrier family 30 A8 zinc transporter gene* variations with gestational diabetes mellitus risk in a Chinese population

Qiaoli Zeng[1,2,3†], Bing Tan[1,2,4†], Fengqiong Han[5†], Xiujuan Huang[6], Jinzhi Huang[7*], Yue Wei[8*] and Runmin Guo[1,2,3,9*]

[1]Department of Internal Medicine, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [2]Key Laboratory of Research in Maternal and Child Medicine and Birth Defects, Guangdong Medical University, Foshan, Guangdong, China, [3]Maternal and Child Research Institute, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [4]Department of Endocrinology, Second Affiliated Hospital of Guangdong Medical University, Zhanjiang, China, [5]Department of Obstetric, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [6]Department of Children's Health, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [7]Department of Gynaecology, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [8]Department of Ultrasound, Shunde Women and Children's Hospital (Maternity and Child Healthcare Hospital of Shunde Foshan), Guangdong Medical University, Foshan, Guangdong, China, [9]Department of Endocrinology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China

Background: The *solute carrier family 30 A8 zinc transporter* (*SLC30A8*) plays a crucial role in insulin secretion. This study aimed to investigate the impact of *SLC30A8* gene polymorphisms on gestational diabetes mellitus (GDM).

Methods: The research objective was to select 500 patients with GDM and 502 control subjects. Rs13266634 and rs2466293 were genotyped using the SNPscan™ genotyping assay. Statistical tests, such as the chi-square test, t-test, logistic regression, ANOVA, and meta-analysis, were conducted to determine the differences in genotypes, alleles, and their associations with GDM risk.

Results: Statistically significant differences were observed in age, pregestational BMI, SBP, DBP, and parity between individuals with GDM and healthy subjects ($P < 0.05$). After adjusting for these factors, rs2466293 remained significantly associated with an increased risk of GDM in overall subjects (GG+AG vs. AA: OR = 1.310; 95% CI: 1.005-1.707; $P = 0.046$, GG vs. AA: OR = 1.523; 95% CI: 1.010-2.298; $P = 0.045$ and G vs. A: OR = 1.249; 95% CI: 1.029-1.516; $P = 0.024$). Rs13266634 was still found to be significantly associated with a decreased risk of GDM in individuals aged ≥ 30 years (TT vs. CT+CC: OR = 0.615; 95% CI: 0.392-0.966; $P = 0.035$, TT vs. CC: OR = 0.503; 95% CI: 0.294-0.861; $P = 0.012$ and T vs. C: OR =0.723; 95% CI: 0.557-0.937; $P = 0.014$). Additionally, the haplotype CG was found to be associated with a higher risk of GDM ($P < 0.05$). Furthermore,

pregnant women with the CC or CT genotype of rs13266634 exhibited significantly higher mean blood glucose levels than those with the TT genotype ($P < 0.05$). Our findings were further validated by the results of a meta-analysis.

Conclusion: The *SLC30A8* rs2466293 polymorphism was found to be associated with an increased risk of GDM, while rs13266634 was associated with a decreased risk of GDM in individuals aged ≥ 30 years. These findings provide a theoretical basis for GDM testing.

KEYWORDS

gestational diabetes mellitus, solute carrier family 30 A8 zinc transporter, SNP, rs13266634, rs2466293, case-control study

1 Introduction

Gestational diabetes mellitus (GDM) is a global concern, and its incidence has increased by over 30% in numerous countries during the past few years (1, 2). GDM is characterized by β-cell dysfunction, insulin resistance, and abnormal glucose utilization (3, 4), but its pathogenesis is not yet clear. Increasing evidence indicates that environmental and genetic factors are implicated in the development of GDM. Single nucleotide polymorphisms (SNPs) are a common type of genetic variation, and polymorphisms in different genes may be associated with GDM (5).

The *solute carrier family 30 A8 zinc transporter* (*SLC30A8*) gene encodes ZnT8, which is primarily expressed in pancreatic β-cells and is in charge of delivering zinc from the cytoplasm into insulin vesicles (6). *SLC30A8* is involved in the secretion of insulin (7). The zinc stabilizes the insulin hexamer in secretory insulin vesicles, making it resistant to degradation (8). Insulin packaged into secretory vesicles can be released immediately upon glucose stimulation (7). The rs13266634 polymorphism is a missense C to T variant in exon 9 of the *SLC30A8* gene, and the amino acid changes from arginine (R) to tryptophan (W) at position 325 (8). Thus, rs13266634 has been thought to be related to diabetes risk, as it affects the expression of *SLC30A8*, and negative regulation of ZnT8 is considered to disrupt the stability of insulin molecules (9). The polymorphism rs2466293 is in the 3′-UTR of the *SLC30A8* gene, and rs2466293 may impact *SLC30A8* post-transcriptional regulation by binding to miRNA (10). MiRNAs are closely related to gene level regulation; hence, rs2466293 in the seed sites of miRNA targets can create or disrupt miRNA-binding sites that further influence disease susceptibility (11). In this context, this study researched the influence of rs13266634 and rs2466293 polymorphisms on GDM risk.

2 Materials and methods

2.1 Study subjects

From 1 August 1 2021 to 31 January 31 2022, a total of 1,002 unrelated Chinese Han pregnant women (500 GDM cases and 502 controls) were recruited for our study at the obstetric clinic of Shunde Maternal and Child Health Hospital, Guangdong Medical University. All individuals underwent a routine 75-gram oral glucose tolerance test (OGTT) during 24-28 weeks of gestation. A control group consisting of pregnant women at 24 to 28 weeks of gestation was selected over the same period. The inclusion criteria were as follows: voluntarily provided written informed consent, not previously diagnosed with diabetes, Han nationality, aged ≥ 18 years, no pregnancy complications, and not taking hypoglycemic medicines. Participants who did not meet the above criteria were excluded.

2.2 Data collection

Information including age, height, pregestational weight, parity (primipara or multipara), blood pressure, race, pregnancy condition, and other clinical information were obtained at 24-28 gestational weeks. Pregestational body mass index (pre-BMI, Kg/m^2) was calculated as pregestational weight (Kg) divided by height squared (m^2). The Chinese standards for obesity were as follows: underweight (< 18.5 Kg/m^2), normal (18.5-24 Kg/m^2), overweight (24-28 Kg/m^2), and obese (≥ 28 kg/m^2).

2.3 SNP genotyping

A total of 2 mL of EDTA-treated blood was immediately stored in the freezer. Genomic DNA was extracted and purified from blood cells by a QIAamp DNA Blood Kit (Qiagen, Germany). Genotypes of candidate SNPs were determined using the SNPscan™ genotyping assay (Genesky Technologies Inc., Shanghai, China). Pre-experiments were conducted before formal experiments. In order to check the genotyping data accuracy, 6% of the samples were randomly selected for duplicate analysis using Sanger sequencing.

2.4 Statistical analyses

Continuous variables following normal distribution were reported as means ± SD, and the independent sample t-test was used to determine the differences between the relevant parameters of the two groups. In cases where the assumption of normality was violated, non-parametric tests were employed. Qualitative data were analyzed using the chi-square ($\chi 2$) test. The Hardy-Weinberg equilibrium (HWE) test, assessed through the goodness-of-fit $\chi 2$, was used to ensure that the control group was representative of the population. The risk of GDM was evaluated using six genetic models, namely, codominant homozygous, codominant heterozygous, dominant, recessive, overdominant, and allele models, through the $\chi 2$ test and logistic regression analysis. Crude and adjusted odds ratios (ORs) and their corresponding 95% confidence intervals (CIs) were presented, with adjustments made for covariates such as age, pre-BMI, etc. Stratified analysis was performed to further examine the potential influence of age and pre-BMI on the results. The frequency distribution of haplotypes was calculated using Haploview 4.2 software. The association between SNPs and blood glucose levels was investigated using one-way ANOVA. For multiple comparisons, the least significant difference (LSD) method was used. Statistical analyses were performed using SPSS 20.0 (SPSS Inc., Chicago, IL, USA), and a *P*-value < 0.05 was considered statistically significant.

3 Results

3.1 General clinical characteristics of the subjects

The study included 500 GDM cases and 502 non-diabetic controls for the evaluation of the *SLC30A8* genotype. Table 1 presents the clinical baseline information and stratified features. The mean age, pre-BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), fasting plasma glucose (FPG), 1 h-PG, and 2 h-PG were significantly higher in the GDM group than in the

TABLE 1 Basic and stratified characteristic of participants of the study.

Variables	Cases (%) (n = 500)	Controls (%) (n = 502)	t/$\chi 2$	*P*
Age, year (mean ± SD)	31±4	29±4	-8.56	< **0.001**
pre-BMI, kg/m^2	21.51±3.10	20.53±2.58	-5.42	< **0.001**
SBP, mmHg	117±11	114±10	-3.53	< **0.001**
DBP, mmHg	70±8	68±7	-3.23	**0.001**
FPG, mmol/L	4.82±0.64	4.50±0.31	-9.75	< **0.001**
1h-PG, mmol/L	10.17±1.60	7.66±1.27	-26.22	< **0.001**
2h-PG, mmol/L	8.91±1.60	6.69±0.99	-25.85	< **0.001**
Parity (n)			8.88	**0.003**
Primipara	210 (42)	258 (51.4)		
Multipara	290 (58)	244 (48.6)		
Variables	Cases (%) (n = 500)	Controls (%) (n = 502)	**$\chi 2$**	***P***
Age, year			49.2	< **0.001**
< 30	192 (38.4)	304 (60.6)		
≥ 30	308 (61.6)	198 (39.4)		
pre-BMI, kg/m^2			27.8	< **0.001**
< 18.5	67 (13.4)	95 (18.9)		
18.5 ≤ BMI < 24	336 (67.2)	365 (72.7)		
≥ 24	97 (19.4)	42 (8.3)		

pre-BMI pre-gestational body mass index, SBP systolic blood pressure, DBP diastolic blood pressure, FPG fasting plasma glucose, bold values indicate the P ≤ 0.05.

control group ($P < 0.05$). Furthermore, there was a significant difference in parity between the GDM and control groups ($P < 0.05$).

3.2 The association between polymorphisms and GDM risk

3.2.1 Overall analysis results

Table 2 presents the minor allele frequency (MAF) and the results of the HWE analysis for two SNPs in the control group. The results were in conformity with HWE ($P > 0.05$). Table 3 shows the ORs with corresponding 95% CIs and associated *P* values estimated for the relationship between genotypes and GDM in the six models (codominant homozygous, codominant heterozygous, dominant, recessive, overdominant, and allele models) for each polymorphism. *SLC30A8* rs2466293 was found to be significantly associated with an increased risk of GDM in the dominant model (GG+AG vs. AA: OR = 1.288; 95% CI: 1.003-1.655; $P = 0.047$), codominant homozygous model (GG vs. AA: OR = 1.499; 95% CI: 1.014-2.217; $P = 0.043$), and allele model (G vs. A: OR = 1.237; 95% CI: 1.029-1.487; $P = 0.023$). Further evaluation was performed using a logistic regression method to adjust for age, pre-BMI, SBP, DBP, and parity. The results indicated a strong association between *SLC30A8* rs2466293 and an increased risk of GDM in the dominant model (GG+AG vs. AA: OR = 1.310; 95% CI: 1.005-1.707; $P = 0.046$), codominant homozygous model (GG vs. AA: OR = 1.523; 95% CI: 1.010-2.298; $P = 0.045$), and allele model (G vs. A: OR = 1.249; 95% CI: 1.029-1.516; $P = 0.024$). However, no significant association was found in rs13266634.

3.2.2 Stratified analysis results

Subsequently, the associations between two SNPs and susceptibility to GDM in six models were tested using stratified analysis for age or pre-BMI. Notably, protective roles were detected in subjects aged ≥ 30 years for rs13266634 under the dominant model (TT+CT vs. CC: OR = 0.648; 95% CI: 0.431-0.975; $P = 0.037$), codominant homozygous (TT vs. CC: OR = 0.517; 95% CI: 0.307-0.872; $P = 0.013$) and allele model (T vs. C: OR = 0.728; 95% CI: 0.565-0.938; $P = 0.014$). After adjustments, rs13266634 was significantly associated with lower GDM risk under the recessive model (TT vs. CT+CC: OR = 0.615; 95% CI: 0.392-0.966; $P = 0.035$), codominant homozygous model (TT vs. CC: OR = 0.503; 95% CI: 0.294-0.861; $P = 0.012$) and allele model (T vs. C: OR =0.723; 95% CI: 0.557-0.937; $P = 0.014$) (Table 4). Moreover, these associations were more evident in subjects aged ≥ 30 years for rs2466293 under the dominant model (GG+AG vs. AA: OR = 1.445; 95% CI: 1.007-2.073; $P = 0.045$) and allele model (G vs. A: OR = 1.337; 95% CI: 1.024-1.747; $P = 0.033$). After these abovementioned factors were adjusted, rs2466293 was significantly related to higher GDM odds under the dominant model (GG+AG vs. AA: OR = 1.579; 95% CI: 1.086-2.295; $P = 0.017$), codominant heterozygous (AG vs. AA: OR = 1.519; 95% CI: 1.020-2.263; $P = 0.040$), and allele model (G vs. A: OR = 1.399; 95% CI: 1.064-1.839; $P = 0.016$) (Table 4). However, no significant associations were found in subjects aged < 30 years (Supplementary Table 1). Nevertheless, the results indicated no significant relationship between rs13266634 or rs2466293 and GDM susceptibility in subjects in the pre-BMI stratified analysis.

3.3 Haplotype and linkage disequilibrium analyses

The study found that two SNPs, rs13266634 and rs2466293, were in strong linkage disequilibrium ($D' > 0.99$) with each other (Figure 1). The CG haplotype consisting of these SNPs was significantly associated with higher GDM risk (OR = 1.231; 95% CI: 1.024-1.48; $P = 0.026$). In addition, the age-stratified analysis revealed that haplotype CG was associated with higher GDM risk in subjects aged ≥ 30 years (OR = 1.328; 95% CI: 1.016-1.734; $P = 0.037$), while haplotype TA was associated with lower GDM risk in subjects aged ≥ 30 years (OR = 0.722; 95% CI: 0.560-0.931; $P = 0.011$). However, no significant associations were found with age < 30 years (Table 5).

3.4 The association between polymorphism genotype and blood glucose levels

The fasting glucose and 1-h PG levels of pregnant women with different genotypes were analyzed by age stratification (Table 6). The results showed that the glucose indexes of the rs13266634 CC genotype were higher than those of the TT genotype in subjects aged ≥ 30 years (all $P < 0.05$), and the 1-h PG level of the CC genotype was significantly higher than the CT genotype.

3.5 Meta–analysis results

Relevant references were searched for based on the PubMed and Google Scholar databases to evaluate the relationship between *SLC30A8* rs13266634 or rs2466293 and GDM. Eight eligible studies were included in the rs13266634 and GDM analysis, and two studies were related to *SLC30A8* rs2466293 and GDM. In total,

TABLE 2 SNPs information and HWE test in the controls.

SNP	Min/Maj	Chr. position	MAF	HWE (*P*)
rs13266634	T/C	chr8:117172544	0.473	0.894
rs2466293	G/A	chr8:117173699	0.325	0.627

Min minor allele, Maj major allele, MAF frequency of minor allele, HWE Hardy–Weinberg equilibrium.

10.3389/fendo.2023.1159714

TABLE 3 The associations between SNPs in SLA30C8 gene and GDM risk in overall subjects.

Model	Cases (%) (n = 500)	Controls (%) (n = 502)	Crude OR (95 % CI)	Crude *P*	Adjusted OR (95 % CI)	Adjusted *P*
rs13266634						
Codominant model						
CC	161 (32.2)	142 (28.3)	1(ref)		1(ref)	
CT	240 (48.0)	245 (48.8)	0.864 (0.648-1.152)	0.319	0.824 (0.609-1.116)	0.212
TT	99 (19.8)	115 (22.9)	0.759 (0.535-1.078)	0.124	0.754 (0.520-1.092)	0.135
Aelle model						
C	562 (56.2)	529 (52.7)	1(ref)		1(ref)	
T	438 (43.8)	475 (47.3)	0.868 (0.728-1.035)	0.115	0.862 (0.716-1.037)	0.115
Dominant Model						
CC	161 (32.2)	142 (28.3)	1(ref)		1(ref)	
TT+CT	339 (67.8)	360 (71.7)	0.831 (0.634-1.008)	0.178	0.802 (0.604-1.006)	0.129
Recessive Model						
CT+CC	401 (80.2)	387 (77.1)	1(ref)		1(ref)	
TT	99 (19.8)	115 (22.9)	0.831 (0.614-1.125)	0.23	0.848 (0.616-1.169)	0.314
Overdominant model						
TT+CC	260 (52.0)	257 (51.2)	1(ref)		1(ref)	
CT	240 (48.0)	245 (48.8)	0.968 (0.756-1.241)	0.799	0.926 (0.713-1.203)	0.565
rs2466293						
Codominant model						
AA	201 (40.2)	233 (46.4)	1(ref)		1(ref)	
AG	224 (44.8)	211 (42.0)	1.231 (0.943-1.606)	0.127	1.251 (0.944-1.658)	0.119
GG	75 (15.0)	58 (11.6)	1.499 (1.014-2.217)	**0.043**	1.523 (1.010-2.298)	**0.045**
Aelle model						
A	626 (62.6)	677 (67.4)	1(ref)		1(ref)	
G	374 (37.4)	327 (32.6)	1.237 (1.029-1.487)	**0.023**	1.249 (1.029-1.516)	**0.024**
Dominant Model						
AA	201 (40.2)	233 (46.4)	1(ref)		1(ref)	
GG+AG	299 (59.8)	269 (53.6)	1.288 (1.003-1.655)	**0.047**	1.310 (1.005-1.707)	**0.046**
Recessive Model						
AG+AA	425 (85.0)	444 (88.4)	1(ref)		1(ref)	
GG	75 (15.0)	58 (11.6)	1.351 (0.935-1.951)	0.109	1.360 (0.925-1.999)	0.118
Overdominant model						
GG+AA	276 (55.2)	291 (58.0)	1(ref)		1(ref)	
AG	224 (44.8)	211 (42.0)	1.119 (0.872-1.437)	0.377	1.131 (0.86-1.472)	0.36

Adjusted *P* value calculated by logistic regression with adjustment for age, pre-BMI, SBP, DBP and parity, bold values indicate the P ≤ 0.05.

the fixed-effects model was used for analysis. Rs13266634 was shown to be significantly associated with a decreased risk of GDM in the following models: dominant model (TT+CT vs. CC: OR = 0.751; 95% CI: 0.674-0.838; $P < 0.001$), recessive model (TT vs. CT+CC: OR = 0.736; 95% CI: 0.629-0.861; $P < 0.001$), overdominant model (CT vs. TT+CC: OR = 0.878; 95% CI: 0.789-0.977; $P < 0.001$), codominant homozygous model (TT vs. CC: OR = 0.643; 95% CI: 0.542-0.763; $P < 0.001$), codominant heterozygous model (CT vs. CC: OR = 0.789; 95% CI: 0.703-0.885; $P < 0.001$), and allele model (T vs. C: OR = 0.795; 95% CI: 0.734-0.860; $P < 0.001$) (Figure 2). In addition, *SLC30A8* rs2466293 was associated with increased GDM risk in the dominant model

10.3389/fendo.2023.1159714

TABLE 4 The associations between SNPs in SLA30C8 gene and GDM risk in subjects aged ≥ 30 years.

Model	Cases (%) (n = 308)	Controls (%) (n = 198)	Crude OR (95 % CI)	Crude *P*	Adjusted OR (95 % CI)	Adjusted *P*
rs13266634						
Codominant model						
CC	98 (31.8)	46 (23.2)	1(ref)		1(ref)	
CT	156 (50.7)	103 (52.0)	0.711 (0.463-1.093)	0.119	0.736 (0.475-1.141)	0.171
TT	54 (17.5)	49 (24.8)	0.517 (0.307-0.872)	**0.013**	0.503 (0.294-0.861)	**0.012**
Aelle model						
C	352 (57.1)	195 (49.2)	1(ref)		1(ref)	
T	264 (42.9)	201 (50.8)	0.728 (0.565-0.938)	**0.014**	0.723 (0.557-0.937)	**0.014**
Dominant Model						
CC	98 (31.8)	46 (23.2)	1(ref)		1(ref)	
TT+CT	210 (68.2)	152 (76.8)	0.648 (0.431-0.975)	**0.037**	0.661 (0.436-1.003)	0.052
Recessive Model						
CT+CC	254 (82.5)	149 (75.2)	1(ref)		1(ref)	
TT	54 (17.5)	49 (24.8)	0.646 (0.418-1.000)	0.05	0.615 (0.392-0.966)	**0.035**
Overdominant model						
TT+CC	152 (49.4)	95 (48.0)	1(ref)		1(ref)	
CT	156 (50.6)	103 (52.0)	0.947 (0.662-1.353)	0.736	0.988 (0.686-1.425)	0.95
rs2466293						
Codominant model						
AA	120 (39.0)	95 (48.0)	1(ref)		1(ref)	
AG	141 (45.8)	81 (40.9)	1.378 (0.939-2.022)	0.101	1.519 (1.020-2.263)	**0.04**
GG	47 (15.2)	22 (11.1)	1.691 (0.953-3.001)	0.071	1.784 (0.994-3.203)	0.053
Aelle model						
A	381 (61.9)	271 (68.4)	1(ref)		1(ref)	
G	235 (38.1)	125 (31.6)	1.337 (1.024-1.747)	**0.033**	1.399 (1.064-1.839)	**0.016**
Dominant Model						
AA	120 (39.0)	95 (48.0)	1(ref)		1(ref)	
GG+AG	188 (61.0)	103 (52.0)	1.445 (1.007-2.073)	**0.045**	1.579 (1.086-2.295)	**0.017**
Recessive Model						
AG+AA	261 (84.7)	176 (88.9)	1(ref)		1(ref)	
GG	47 (15.3)	22 (11.1)	1.441 (0.839-2.475)	0.184	1.447 (0.834-2.508)	0.189
Overdominant model						
GG+AA	167 (54.2)	117 (59.1)	1(ref)		1(ref)	
AG	141 (45.8)	81 (40.9)	1.220 (0.850-1.750)	0.281	1.323 (0.910-1.923)	0.143

Adjusted *P* value calculated by logistic regression with adjustment for age, pre-BMI and SBP. bold values indicate the P ≤ 0.05.

(GG+AG vs. AA: OR = 1.184; 95% CI: 1.013-1.383; *P* = 0.034), recessive model (GG vs. AG + AA : OR = 1.408; 95% CI: 1.135-1.747; *P* = 0.002), codominant homozygous model (GG vs. AA : OR = 1.474; 95% CI: 1.167-1.861; *P* = 0.001), and allele model (G vs. A: OR = 1.195; 95% CI: 1.069-1.336; *P* = 0.002), and no significant association was found in other genetic models (Figure 3). There was no obvious evidence of publication bias in the genetic models, and these results are consistent with Egger's tests (all *P* > 0.05).

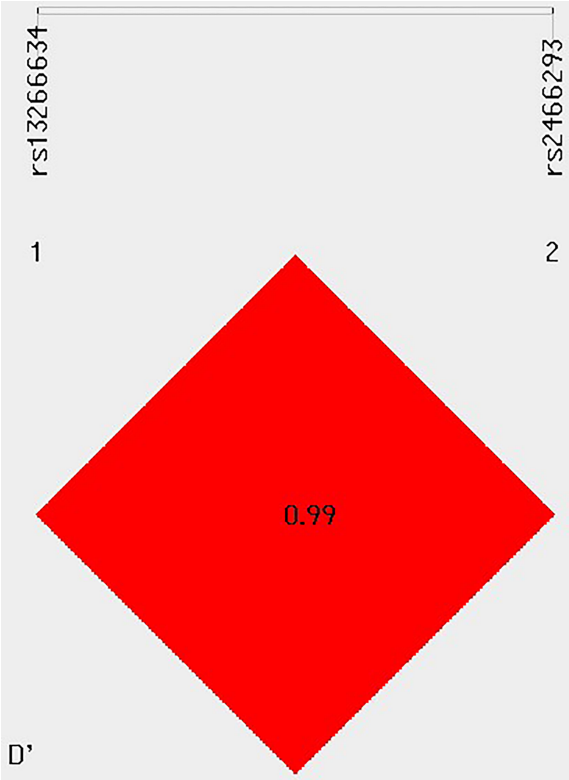


FIGURE 1
Linkage disequilibrium (LD) between multiple loci of the *SLC30A8* gene (rs13266634 C/T and rs2466293 A/G).

4 Discussion

The role of genetic factors in the GDM process has been verified by previous findings (12). Rs13266634 is a non-synonymous SNP in *SLC30A8*, and a protective role for the rs13266634 T allele, which reduces GDM risk, has been proposed in a Swedish population (8). In contrast, six studies from Brazil, the United States, Denmark, the Republic of Korea, and China failed to replicate the results (13–18). Therefore, further verification is necessary. Moreover, rs2466293 is a polymorphism in miRNA-binding sites (miR-binding SNP). Recent findings indicated that rs2466293 impacted the development of GDM (19), but more extensive research is needed for verification. This research paper conducted a case-control study to estimate the association of *SLC30A8* rs13266634 or rs2466293 with GDM among six different genetic models in a Chinese population.

In the research process, we explored the relationship between *SLC30A8* gene multiformity and GDM risk. In the overall analysis, the findings indicated that *SLC30A8* rs13266634 showed no association with GDM risk, but *SLC30A8* rs2466293 was shown to be significantly related to increased GDM risk under the dominant (GG+AG), codominant homozygous (GG), and allele (G) genetic models that were unadjusted and adjusted for age, pre-BMI, SBP, DBP, and parity. In this study, women with GDM were older than healthy controls. It has been pointed out that the prevalence of GDM increases with age, and the incidence is higher among women over 30 years of age (20). Therefore, further studies used a cutoff point of 30 years of age and analyzed the association between polymorphic variants and GDM after stratification by age. Interestingly, after adjusting for age, pre-pregnancy BMI, and SBP, our findings indicated that the SNP rs13266634 in *SLC30A8* was found to have a protective effect against GDM risk in subjects aged ≥30 years under the recessive and homozygous dominant genetic models, while *SLC30A8* rs2466293 was significantly associated with increased GDM risk in patients aged ≥30 years under the dominant and heterozygous dominant genetic models. These results are in accordance with some scholarly studies (8, 13, 19). Furthermore, the CG haplotype, comprised of SNPs rs13266634 and rs2466293, was significantly associated with an increased risk of GDM in the overall analysis. In a further analysis stratified by age, the CG haplotype was also

TABLE 5 Haplotype analysis of the rs132666342 and rs2466293 SNPs of the SLA30C8 gene for the GDM and controls.

Haplotype	Cases (%)	Controls (%)	χ2	*P*	OR (95 % CI)
CA	189 (18.9)	202 (20.1)	0.474	0.49	0.925 (0.741-1.154)
TA	437 (43.7)	475 (47.3)	2.633	0.104	0.864 (0.724-1.03)
CG	373 (37.3)	327 (32.5)	4.932	**0.026**	1.231 (1.024-1.48)
Haplotype	**Cases (%)**	**Controls (%)**	**χ2**	***P***	**OR (95 % CI)**
Age (years) <30					
CA	71 (18.4)	132 (21.7)	1.5	0.22	0.817 (0.592-1.128)
CG	139 (36.1)	202 (33.2)	0.922	0.336	1.14 (0.872-1.49)
TA	174 (45.3)	274 (45.0)	0.005	0.939	1.01 (0.781-1.305)
Age (years) ≥ 30					
CA	118 (19.1)	70 (17.6)	0.348	0.554	1.103 (0.795-1.53)
TA	263 (42.6)	201 (50.7)	6.311	**0.011**	0.722 (0.56-0.931)
CG	234 (37.9)	125 (31.5)	4.342	**0.037**	1.328 (1.016-1.734)

10.3389/fendo.2023.1159714

TABLE 6 Association between SNPs polymorphisms genotype and blood glucose levels.

Genotype	FPG (mmol/L)	1h-PG (mmol/L)	2h-PG (mmol/L)
rs13266634			
Age (years) < 30			
CC	4.66±0.614	8.49±1.923	7.30±1.925
CT	4.57±0.570	8.38±1.81	7.31±1.594
TT	4.66±1.077	8.55±2.277	7.68±1.925
F	0.903	0.28	2.123
P	> 0.05	> 0.05	> 0.05
Age (years) ≥ 30			
CC	4.79±0.546	9.87±1.717	8.62±1.818
CT	4.74±0.556	9.50±1.884[b]	8.25±1.724
TT	4.64±0.482[a]	9.10±1.679[a]	8.12±1.733[a]
F	2.225	5.262	2.766
P	**< 0.05**	**< 0.05**	**< 0.05**
rs2466293			
Age (years) < 30			
AA	4.67±0.887	8.37±2.033	7.48±1.733
AG	4.58±0.606	8.44±1.886	7.27±1.577
GG	4.61±0.493	8.78±1.921	7.49±1.443
F	0.752	1.009	0.945
P	> 0.05	> 0.05	> 0.05
Age (years) ≥ 30			
AA	4.76±0.532	9.40±2.014	8.31±1.917
AG	4.68±0.577	9.62±1.682	8.33±1.693
GG	4.81±0.435	9.56±1.57	8.40±1.478
F	1.739	0.781	0.07
P	> 0.05	> 0.05	> 0.05

[a]LSD was used to compare the blood glucose levels of three rs13266634 genotypes: the difference of blood glucose between CC and TT genotypes was statistically significant, all P < 0.05. [b] LSD was used to compare the blood glucose levels of three rs13266634 genotypes: the difference of 1-h blood glucose between CC and CT genotypes was statistically significant, P < 0.05. P < 0.05, bold values indicate the P < 0.05.

associated with an increased risk of GDM in individuals aged ≥ 30 years, while the TA haplotype was associated with a reduced risk of GDM in the same age group. These results suggest that the T allele of rs13266634 in *SLC30A8* can be considered a protective factor for GDM, while the G allele of rs2466293 may be a risk factor for GDM.

Wang et al. found that the C allele of rs2466293 increased susceptibility to GDM in the Chinese population (19), which was consistent with our research findings. In addition, our study found that the TT homozygous genotype of rs13266634 and the T allele decreased the risk of developing GDM in subjects aged ≥ 30 years. Similarly, previous research has demonstrated that the T allele of rs13266634 protects against the risk of GDM in the Swedish population (8), which was consistent with our findings. Moreover, in populations of Filipinos, Swedes, Koreans, and Chinese individuals, there was evidence of an association between the C allele of *SLC30A8* rs13266634 and a higher risk of GDM (8, 13, 14, 18). However, other studies have not found any association between rs13266634 and the risk of GDM in populations of Danes and Europeans (15–17). Inconsistencies in these results may be related to differences in ethnicity, environment, or limited study sample sizes. Therefore, a comprehensive meta-analysis was carried out with a larger number of different populations (ethnicities) to identify the relationship of SLC30A8 SNPs with GDM risk. Rs13266634 was demonstrated to have a protective effect in every genetic model ($P < 0.05$) in eight eligible studies (including our study), and significant findings of rs13266634 could also be observed in both the Caucasian and Asian subgroups. SLC30A8 rs2466293 was found to be significantly related to higher GDM risk in the relevant models (codominant homozygous and allele models) ($P < 0.05$) based on two Chinese population studies.

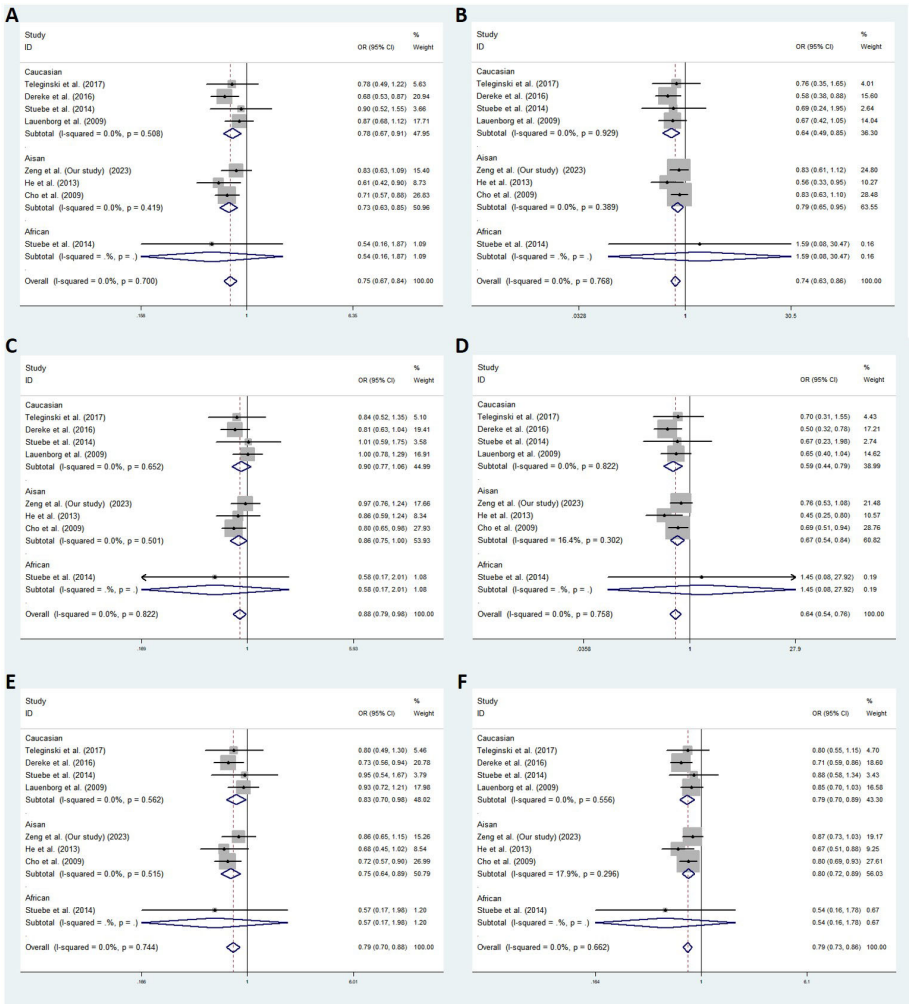


FIGURE 2
Meta-analysis with a fixed effects model for the association between *SLC30A8* rs13266634 and GDM susceptibility. **(A)** dominant model, TT+CT vs. CC **(B)** recessive model, TT vs. CT+CC **(C)** overdominant model, CT vs. TT +CC **(D)** codominant homozygous model,TT vs.CC **(E)** codominant heterozygous model, CT vs.CC **(F)** allele model, T vs. **(C)** OR: odds ratio, CI: confidence interval, I-squared: measure to quantify the degree of heterogeneity in meta-analyses.

GDM and T2DM are considered to have similar pathogenesis. In a study of diabetic mice, *SLC30A8* gene expression levels were inhibited in the pancreas of animals with this pathology, indicating that it is related to diabetes (9). Studies have shown that the *SLC30A8* rs13266634 C allele is associated with glucose regulation in GWASs (21, 22). In addition, studies based on fluorescence and radiation have proposed a hypothesis that the rs13266664-T allele reduces *SLC30A8* activity, which changes insulin synthesis and reduces GDM susceptibility based on this mechanism (23–25). In addition, genetic variation in the 3'UTR, a miRNA target gene, can affect the interaction between miRNA and target mRNA. We queried the rs2466293 polymorphism located using the "MirSNP" database (http://bioinfo.life.hust.edu.cn/miRNASNP/). According to the results, it can be inferred that rs2466293 creates eight and destroys three putative miRNA target sites, which may impact the expression of *SLC30A8* and lead to a higher risk of GDM. However, functional research is necessary to further confirm its mechanism.

According to the abovementioned research, this study obtained a conclusion that the age and pre-BMI of the GDM group were significantly higher than those of the control group, and logistic regression analysis indicated that the increase in age and pre-BMI were important risk factors for GDM. SBP, DBP, and parity in the GDM group were significantly higher than those in the other group. It can be inferred that patients with GDM were prone to pregnancy-induced hypertension syndrome. Moreover, a previous study found that the *SLC30A8* rs13266634 C allele was correlated with higher fasting glucose levels among women with gestational high BMI (26). Our study also showed that the *SLC30A8* rs13266634 C allele had an influence on higher fasting glucose, 1-h, and 2-h glucose levels among pregnant women over the age of 30 years, which was similar to the results of previous studies. The *SLC30A8* rs13266634 C allele may affect the normal secretion of insulin. Wang et al. found a significant relationship between the C allele of rs2466293 with higher plasma glucose (19), but no differences were found in our study. Therefore, further relevant research is necessary.

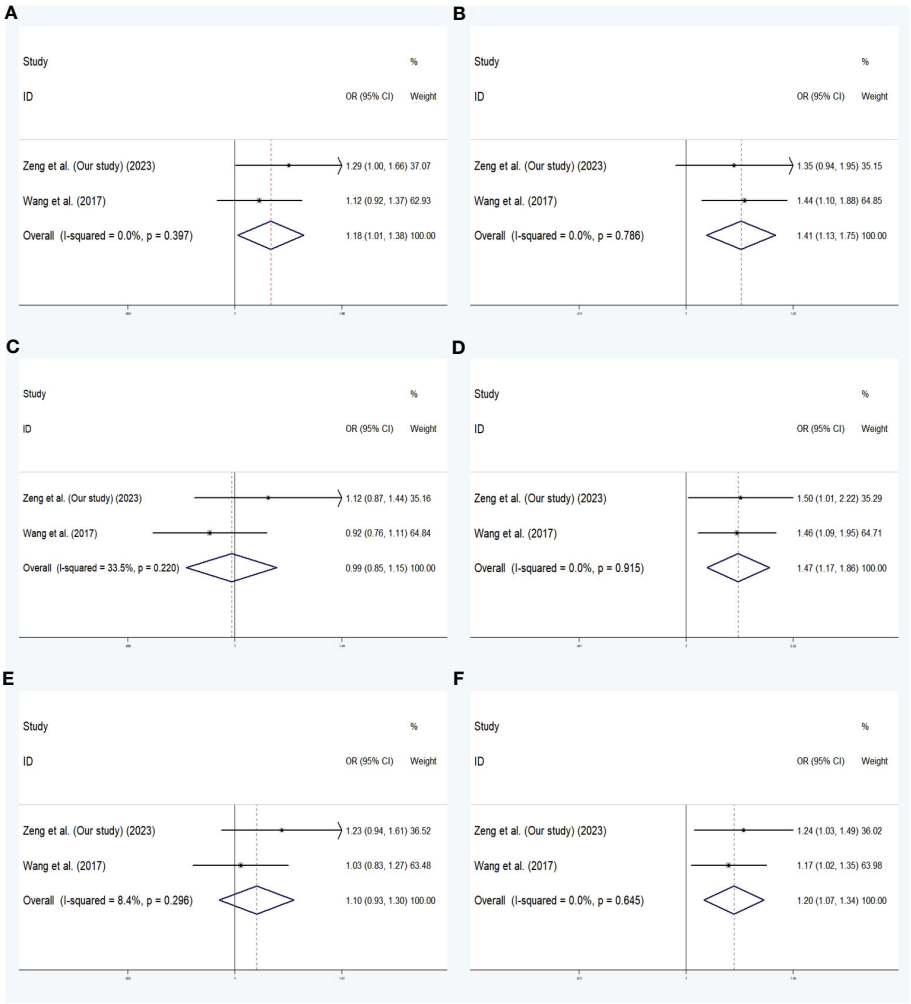


FIGURE 3
Meta-analysis with a fixed effects model for the association between *SLC30A8* rs2466293 and GDM susceptibility. **(A)** dominant model,GG+AG vs. AA **(B)** recessive model,GG vs. AG + AA **(C)** overdominant model, AG vs. GG +AA **(D)** codominant homozygous model, GG vs. AA **(E)** codominant heterozygous model, AG vs.AA **(F)** allele model, G vs. **(A)** OR: odds ratio, CI: confidence interval, I-squared: measure to quantify the degree of heterogeneity in meta-analyses.

There are still several limitations in this study. First, due to the modest sample size of the GDM and control groups, future studies need to validate our observations in a larger cohort. Second, the data used in this study were insufficient, such as the lack of fasting insulin data, to accurately measure and evaluate pancreatic islet β-cell function. Finally, the study subjects were limited to Chinese individuals, and additional research is necessary to confirm our findings in diverse populations.

5 Conclusions

In conclusion, in subjects aged ≥ 30 years, *SLC30A8* rs13266634 exhibited a protective relationship against GDM susceptibility, while the results indicated associations of rs2466293 with the risk of GDM. The haplotype CG was also associated with a higher risk of GDM, and the haplotype TA was associated with a lower risk of GDM in subjects aged ≥ 30 years. In general, our findings provide more clues for studying the precise mechanism of the development of GDM.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: PRJEB61053, ERZ16808993.

Ethics statement

The study was agreed by the Ethics Committee of Shunde Maternal and Child Health Hospital of Guangdong Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

QZ, BT and FH contributed equally to this study. QZ, FH and BT collected clinical data and samples. QZ, XH and JH performed data analyses. QZ, YW and RG wrote the manuscript. JH and YW supervised the whole research. All authors contributed to the article and approved the submitted version.

Funding

Support from the National Natural Science Foundation of China (81873649); Doctoral scientific research Initiate funding project of Shunde Women and Children's Hospital of Guangdong Medical University (Maternity and Child Healthcare Hospital of Shunde Foshan) (2020BSQD007); Medical Research Project of Foshan Health Bureau (20210289); Guangdong Medical University Research Foundation (GDMUM2020008); Youth Talent Project of Shunde Women and Children's Hospital of Guangdong Medical University (Maternity and Child Healthcare Hospital of Shunde Foshan) (2023QNRC023).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1159714/full#supplementary-material

References

1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective [published correction appears in diabetes care. *Diabetes Care* (2007) 30 Suppl 2:S141–6. doi: 10.2337/dc07-s206

2. Zhu Y, Zhang C. Prevalence of gestational diabetes and risk of progression to type 2 diabetes: a global perspective. *Curr Diabetes Rep* (2016) 16(1):7. doi: 10.1007/s11892-015-0699-x

3. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus–present and future perspectives. *Nat Rev Endocrinol* (2011) 8 (4):228–36. doi: 10.1038/nrendo.2011.183

4. Martin AO, Simpson JL, Ober C, Freinkel N. Frequency of diabetes mellitus in mothers of probands with gestational diabetes: possible maternal influence on the predisposition to gestational diabetes. *Am J Obstet Gynecol* (1985) 151(4):471–5. doi: 10.1016/0002-9378(85)90272-8

5. Huopio H, Cederberg H, Vangipurapu J, Hakkarainen H, Pääkkönen M, Kuulasmaa T, et al. Association of risk variants for type 2 diabetes and hyperglycemia with gestational diabetes. *Eur J Endocrinol* (2013) 169(3):291–7. doi: 10.1530/EJE-13-0286

6. Chimienti F, Devergnas S, Pattou F, Schuit F, Garcia-Cuenca R, Vanderwalle B, et al. *In vivo* expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion. *J Cell Sci* (2006) 119(Pt 20):4199–206. doi: 10.1242/jcs.03164

7. Maruthur NM, Clark JM, Fu M, Linda Kao WH, Shuldiner AR. Effect of zinc supplementation on insulin secretion: interaction between zinc and *SLC30A8* genotype in old order Amish. *Diabetologia* (2015) 58(2):295–303. doi: 10.1007/s00125-014-3419-1

8. Dereke J, Palmqvist S, Nilsson C, Landin-Olsson M, Hillman M. The prevalence and predictive value of the *SLC30A8* R325W polymorphism and zinc transporter 8 autoantibodies in the development of GDM and postpartum type 1 diabetes. *Endocrine* (2016) 53(3):740–6. doi: 10.1007/s12020-016-0932-7

9. Seman NA, Mohamud WN, Östenson CG, Brismar K, Gu HF. Increased DNA methylation of the *SLC30A8* gene promoter is associated with type 2 diabetes in a Malay population. *Clin Epigenet* (2015) 7(1):30. doi: 10.1186/s13148-015-0049-5

10. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* (2004) 116(2):281–97. doi: 10.1016/s0092-8674(04)00045-5

11. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research [published correction appears in nat rev cancer. *Nat Rev Cancer* (2010) 10(6):389–402. doi: 10.1038/nrc2867

12. Tsai PJ, Roberson E, Dye T. Gestational diabetes and macrosomia by race/ethnicity in Hawaii. *BMC Res Notes* (2013) 6:395. doi: 10.1186/1756-0500-6-395

13. He M, Ban B, Bian D, Li P, Zhang M, Man D, et al. Association of rs13266634 polymorphism in zinc transporter solute carrier family 30-member 8 gene with gesta-tional diabetes mellitus. *Chin J Diabetes Mellit* (2013) 5(1):38–43.

14. Benny P, Ahn HJ, Burlingame J, Lee MJ, Miller C, Chen J, et al. Genetic risk factors associated with gestational diabetes in a multi-ethnic population. *PloS One* (2021) 16(12):e0261137. doi: 10.1371/journal.pone.0261137

15. Teleginski A, Welter M, Frigeri HR, Réa RR, Souza EM, Alberton D, et al. Leptin (rs7799039) and solute carrier family 30 zinc transporter (rs13266634) polymorphisms in Euro-Brazilian pregnant women with gestational diabetes. *Genet Mol Res* (2017) 16 (1). doi: 10.4238/gmr16019515

16. Stuebe AM, Wise A, Nguyen T, Herring A, North KE, Siega-Riz AM. Maternal genotype and gestational diabetes. *Am J Perinatol* (2014) 31(1):69–76. doi: 10.1055/s-0033-1334451

17. Lauenborg J, Grarup N, Damm P, Borch-Johnsen K, Jørgensen T, Pederson O, et al. Common type 2 diabetes risk gene variants associate with gestational diabetes. *J Clin Endocrinol Metab* (2009) 94(1):145–50. doi: 10.1210/jc.2008-1336

18. Cho YM, Kim TH, Lim S, Choi SH, Shin HD, Lee HK, et al. Type 2 diabetes-associated genetic variants discovered in the recent genome-wide association studies are related to gestational diabetes mellitus in the Korean population. *Diabetologia* (2009) 52(2):253–61. doi: 10.1007/s00125-008-1196-4

19. Wang X, Li W, Ma L, Ping F, Liu J, Wu X, et al. Investigation of miRNA-binding site variants and risk of gestational diabetes mellitus in Chinese pregnant women. *Acta Diabetol* (2017) 54(3):309–16. doi: 10.1007/s00592-017-0969-y

20. Murgia C, Berria R, Minerba L, Sulis S, Murenu M, Portoghese E, et al. Risk assessment does not explain high prevalence of gestational diabetes mellitus in a large group of sardinian women. *Reprod Biol Endocrinol* (2008) 6:26. doi: 10.1186/1477-7827-6-26

21. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* (2007) 445 (7130):881–5. doi: 10.1038/nature05616

22. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science* (2007) 316(5829):1341–5. doi: 10.1126/science.1142382

23. Carvalho S, Molina-López J, Parsons D, Corpe C, Maret W, Hogstrand C. Differential cytolocation and functional assays of the two major human *SLC30A8* (ZnT8) isoforms. *J Trace Elem Med Biol* (2017) 44:116–24. doi: 10.1016/j.jtemb.2017.06.001

24. Kim I, Kang ES, Yim YS, Ko SJ, Jeong SH, Rim JH, et al. A low-risk ZnT-8 allele (W325) for post-transplantation diabetes mellitus is protective against cyclosporin a-induced impairment of insulin secretion. *Pharmacogenom J* (2011) 11(3):191–8. doi: 10.1038/tpj.2010.22

25. Nicolson TJ, Bellomo EA, Wijesekara N, Loder MK, Baldwin JM, Gyulkhandanyan AV, et al. Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes* (2009) 58(9):2070–83. doi: 10.2337/db09-0551

26. Wang T, Liu H, Wang L, Huang T, Li W, Zheng Y, et al. Zinc-associated variant in *SLC30A8* gene interacts with gestational weight gain on postpartum glycemic changes: a longitudinal study in women with prior gestational diabetes mellitus. *Diabetes* (2016) 65(12):3786–93. doi: 10.2337/db16-0730

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 07 June 2023
DOI 10.3389/fendo.2023.1140499

OPEN ACCESS

EDITED BY
Richard Ivell,
University of Nottingham, United Kingdom

REVIEWED BY
Ye Bai,
Chongqing Medical University, China
Jingya Wang,
University of Birmingham, United Kingdom

*CORRESPONDENCE
Jian Zhao
jzhao.epi@gmail.com
Jun Zhang
junjimzhang@sina.com

†These authors have contributed equally to this work

‡These authors share senior authorship

RECEIVED 09 January 2023
ACCEPTED 25 May 2023
PUBLISHED 07 June 2023

CITATION
Liu D, Gan Y, Zhang Y, Cui L, Tao T, Zhang J and Zhao J (2023) Fetal genome predicted birth weight and polycystic ovary syndrome in later life: a Mendelian randomization study.
Front. Endocrinol. 14:1140499.
doi: 10.3389/fendo.2023.1140499

Fetal genome predicted birth weight and polycystic ovary syndrome in later life: a Mendelian randomization study

Dong Liu[1†], Yuexin Gan[1†], Yue Zhang[2], Linlin Cui[3], Tao Tao[4], Jun Zhang[1,5*‡] and Jian Zhao[1,5,6*‡]

[1]Ministry of Education and Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [3]Center for Reproductive Medicine, The Second Hospital, Cheeloo College of Medicine, National Research Center for Assisted Reproductive Technology and Reproductive Genetics, Shandong University, Jinan, Shandong, China, [4]Department of Endocrinology and Metabolism, Renji Hospital, School of Medicine, Shanghai JiaoTong University, Shanghai, China, [5]Department of Maternal and Child Health, School of Public Health, Shanghai Jiao Tong University, Shanghai, China, [6]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom

Associations between lower birth weight and higher polycystic ovary syndrome (PCOS) risk have been reported in previous observational studies, however, the causal relationship is still unknown. Based on decomposed fetal and maternal genetic effects on birth weight (n = 406,063), we conducted a two-sample Mendelian randomization (MR) analysis to assess potential causal relationships between fetal genome predicted birth weight and PCOS risk using a large-scale genome-wide association study (GWAS) including 4,138 PCOS cases and 20,129 controls. To further eliminate the maternally transmitted or non-transmitted effects on fetal growth, we performed a secondary MR analysis by utilizing genetic instruments after excluding maternally transmitted or non-transmitted variants, which were identified in another birth weight GWAS (n = 63,365 parent-offspring trios from Icelandic birth register). Linkage disequilibrium score regression (LDSR) analysis was conducted to estimate the genetic correlation. We found little evidence to support a causal effect of fetal genome determined birth weight on the risk of developing PCOS (primary MR analysis, OR: 0.86, 95% CI: 0.52 to 1.43; secondary MR analysis, OR: 0.86, 95% CI: 0.54 to 1.39). In addition, a marginally significant genetic correlation (r_g = -0.14, se = 0.07) between birth weight and PCOS was revealed *via* LDSR analysis. Our findings indicated that observed associations between birth weight and future PCOS risk are more likely to be attributable to genetic pleiotropy driven by the fetal genome rather than a causal mechanism.

KEYWORDS

Mendelian randomization, birth weight, polycystic ovary syndrome, fetal genome, genetic pleiotropy

Introduction

Polycystic ovary syndrome (PCOS), affecting 6% – 9% of women of reproductive age, is the most common endocrine condition (1). Based on previous studies, insulin resistance, obesity, and androgen excess may contribute together and play crucial roles in PCOS development (2, 3). In addition, an increasing body of evidence suggests a strong genetic component in its aetiology (4, 5). However, the aetiology of PCOS remains largely unknown, and no efficient therapeutic treatments or prevention measures for PCOS are available. According to the Developmental Origins of Health and Disease (DOHaD) hypothesis, early life abnormal growth and development were associated with the risk of developing various chronic diseases in later life (6, 7). Birth weight, a common indicator reflecting intrauterine fetal growth, has been widely studied on its long-term impact on adulthood health outcomes (8–11). Interestingly, observational associations between birth weight and PCOS risk in later life have been reported in a recent meta-analysis and multiple cohort studies (12–17). However, these associations were not well replicated in other independent large-scale cohort studies (18–20). Given that observational studies are commonly prone to residual confounding or reverse causation (21), the causal relationship between birth weight and the risk of developing PCOS remains unknown.

Mendelian randomization (MR), which is a causal inference technique using genetic variants randomly allocated during conception as instrumental variables, is less prone to residual confounding or reverse causation bias (22). In a previous study, little evidence was found to support a causal effect of birth weight on PCOS risk by using MR (P = 0.22) (23). However, this study used offspring genetic variants associated with birth weight as instrumental variables without adjusting for maternal genotypes, which were correlated with fetal genotypes ($r \approx 0.5$) (24, 25) (Supplemental Figure 1). Thus, their effect estimates of birth weight on PCOS risk might be biased by the maternal genetic effects. In addition, recent studies suggested that composite or complex traits can be explained by multiple components or distinct biological pathways (26–28). Like other complex traits, variation in birth weight can also be explained by different components, such as fetal genetically regulated components and maternal adverse intrauterine environment components (29–31). Dissecting these components of birth weight is essential to understand the underlying biological mechanism. Recently, several studies investigated possible mechanisms between birth weight and cardiometabolic risk by using different components of birth weight. Based on structural equation model (SEM) and weighted linear model (WLM) methods, Warrington et al. and Moen et al. recently separated genetic effects on birth weight into maternal and fetal components to investigate the causal mechanisms between birth weight and future cardiometabolic risk (29, 32). Their findings suggested that associations between birth weight and adulthood cardiometabolic outcomes were attributable to fetal genetic effects rather than intrauterine programming (29, 32). Moreover, from a genomic perspective, Juliusdottir et al. discriminated the effects of transmitted and non-transmitted alleles on birth weight by using a long-range phasing (LRP) method based on the Icelandic fetal growth samples to investigate inheritance patterns affecting birth weight (30). This study indicated that associations between birth weight and most cardiometabolic risk factors were driven by the fetal genome (30), whereas it is still unclear whether birth weight affects PCOS in the same manner.

Recently, two large-scale genome-wide association study (GWAS) meta-analyses on PCOS released their summary statistics (4, 33), which provided opportunities for assessing potential causal relationships between birth weight and PCOS risk. Thus, in this study, we aimed to investigate whether there is a causal effect of fetal genome determined birth weight on PCOS risk using two-sample MR analysis. Considering other potential mechanisms that might underpin the association between birth weight and PCOS, such as genetic pleiotropy, we also assessed the genetic correlation between birth weight and PCOS risk by conducting linkage disequilibrium score regression (LDSR) analysis which is mainly used to identify shared genetic variation between two traits across the whole genome (34, 35).

Materials and methods

Data sources and study populations

A schematic overview of the study design is presented in Figure 1 and detailed data sources information can be found in Supplemental Table 1. We used two sets of birth weight summary statistics obtained from GWASs conducted by the Early Growth Genetics (EGG) consortium (http://egg-consortium.org) and the Icelandic birth register to construct two sets of instrumental variables (IVs) for the primary and secondary MR analysis, respectively. GWAS of birth weight conducted by the EGG Consortium included 406,063 individuals of European ancestry (29), where maternal and fetal genetic effects on birth weight were separated by using SEM. In the primary MR analysis, we used the summary statistics of fetal genetic effects on the offspring's birth weight after adjusting for correlated maternal genotypes. Of note, the original birth weight GWAS categorized 305 genome-wide significant ($P < 5\times10^{-8}$) single nucleotide polymorphisms (SNPs) identified into 5 groups based on the effects of maternal and/or fetal genotypes on offspring birth weight: 1) fetal effect only, 2) maternal effect only, 3) fetal and maternal effects with the same direction, 4) fetal and maternal effects with the opposite directions, and 5) unclassified (29). Among these variants, 28 SNPs were identified as having fetal genetic effects on birth weight (SEM classification: "fetal only" or "fetal and maternal").

The outcome data were obtained from a large-scale GWAS meta-analysis of PCOS, including 4,138 cases and 20,129 controls of European ancestry from six cohorts (Rotterdam, Oxford, EGCUT, deCODE, Chicago, and Boston) (4). To further validate the results of the MR analysis, we used summary statistics of PCOS GWAS meta-analysis in the FinnGen and Estonian Biobank (EstBB) as replication data, which included 3,609 cases and 229,788 controls (33).

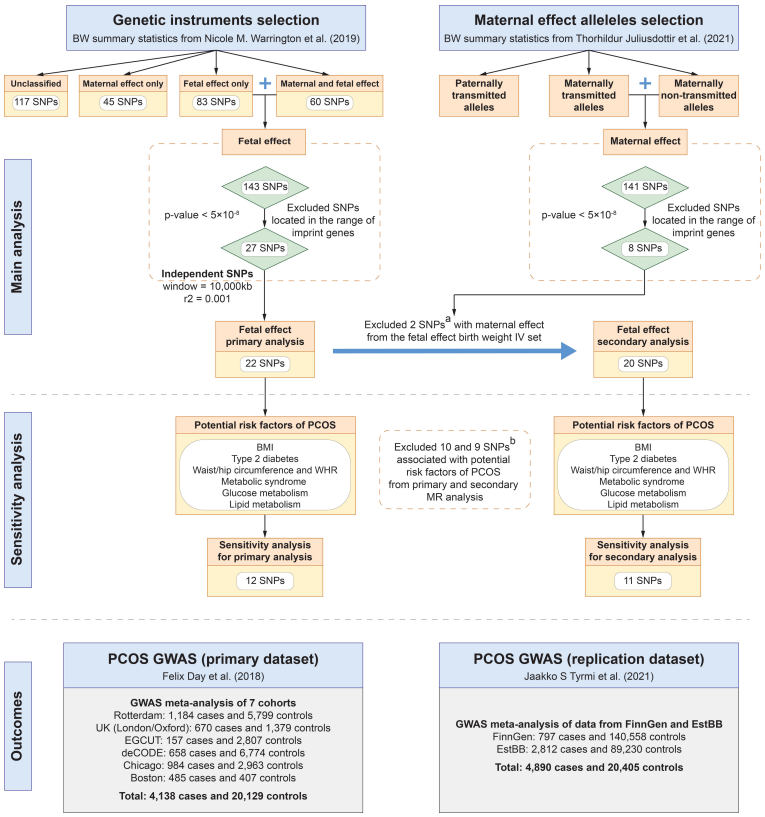


FIGURE 1
Study design of MR analyses. **(A)** rs560887 and rs10872678 were identified as maternally transmitted and non-transmitted alleles respectively in the birth weight GWAS by Juliusdottir et al. (30). **(B)** SNPs were genome-wide significantly associated with potential confounders of PCOS, including BMI, type 2 diabetes, waist/hip circumference, waist-to-hip ratio, metabolic syndrome, glucose metabolism, and lipid metabolism. BMI, body mass index; BW, birth weight; EstBB, Estonian Biobank; GWAS, genome-wide association study; PCOS, polycystic ovary syndrome; SNP, single nucleotide polymorphism; WHR, waist-to-hip ratio.

Genetic instruments selection

The genetic instruments selection procedure was conducted in the following steps. First, in the primary MR analysis, statistically significant ($P < 5\times10^{-8}$) genetic variants were selected from summary statistics of birth weight GWAS conducted by the EGG Consortium (29). To ensure that genetic variants are independent, a stringent linkage disequilibrium (LD) threshold ($r^2 < 0.001$ and window size = 10,000 kb) was used for LD clumping, with the European subsample of 1,000 Genome Project data as reference panel (36). Moreover, we excluded genetic instruments located in the range of imprinted genes to minimize the heterogeneous effect of variants on phenotypes in the population. Considering potential violations of the MR core assumptions, that is, maternal genetic effects confounded fetal genetic variants which were used as IVs and the outcome (i.e., PCOS), we identified and excluded SNPs that exerted maternal genetic effects on birth weight from the set of IVs. To further eliminate maternal genetic effects on birth weight from IVs used in the primary analysis, we identified and excluded maternally transmitted and non-transmitted alleles based on a GWAS meta-analysis on birth weight by Juliusdottir et al. from 63,365 parent-offspring trios (30), to construct IVs for the secondary MR analysis. The allele-specific effects of maternally transmitted or non-transmitted on birth weight were used to represent the maternal and fetal genetic effects, respectively (37). Finally, five maternally transmitted and three maternally non-transmitted SNPs that reached a genome-wide significant level on birth weight were identified from the GWAS by Juliusdottir et al. (30)

Furthermore, we extracted SNP-POCS associations for each genetic instrument from two independent PCOS GWASs conducted by Day et al. and Tyrmi et al., respectively (4, 33). If a certain instrument was not available in the summary data, a proxy SNP in high LD in the European population was identified using LDlink (https://ldlink.nci.nih.gov/?tab=ldproxy). After that, data harmonization was performed to combine SNP-birth weight and SNP-PCOS associations using the "harmonise_data" function in the TwoSample MR package (36), in which ambiguous or palindromic SNPs were excluded.

As a result, we retained a total of 22 SNPs as genetic instruments in the primary MR analysis from birth weight GWAS conducted by Warrington et al. (29) and 20 SNPs after excluding two maternally transmitted or non-transmitted SNPs (i.e., rs560887 and rs10872678 which were identified in the GWAS by Juliusdottir et al. (30)) in the secondary MR analysis (Table 1). To minimize the risk of violating the IV assumptions, we identified SNPs associated

TABLE 1 Characteristics of instrumental variables for birth weight used in the primary MR analysis.

SNP	CHR	Position	Gene	EA	OA	EAF	Beta	SE	P	F*
rs80278614	1	119412317	*TBX15*	A	G	0.05	0.05	0.009	4.03×10^{-8}	30.1
rs2551347	2	23912401	*KLHL29*	T	C	0.75	0.03	0.005	2.20×10^{-9}	35.8
rs17034876	2	46484310	*EPAS1*	T	C	0.70	0.04	0.005	5.47×10^{-17}	70.2
rs560887 [a,b]	2	169763148	*G6PC2*	C	T	0.70	-0.02	0.004	2.78×10^{-8}	30.9
rs11708067 [b]	3	123065778	*ADCY5*	G	A	0.25	0.06	0.005	6.26×10^{-32}	138.3
rs1482852 [b]	3	156798294	*LOC339894*	A	G	0.60	0.05	0.004	7.56×10^{-39}	170.0
rs4144829 [b]	4	17903654	*LCORL*	C	T	0.26	0.03	0.005	1.12×10^{-11}	46.1
rs35261542 [b]	6	20675792	*CDKAL1*	C	A	0.74	0.05	0.005	3.23×10^{-26}	112.2
rs10872678 [a]	6	152039964	*ESR1*	T	C	0.72	0.03	0.005	8.23×10^{-10}	37.7
rs138715366	7	44246271	*YKT6/GCK*	C	T	0.99	0.24	0.022	1.43×10^{-25}	109.3
rs112139215	7	73034559	*MLXIPL*	A	C	0.07	0.06	0.008	1.20×10^{-11}	46.0
rs13266210	8	41533514	*ANK1*	A	G	0.78	0.03	0.005	3.05×10^{-9}	35.2
rs28457693	9	98217348	*PTCH1*	G	A	0.11	0.04	0.007	1.70×10^{-9}	36.3
rs1112718 [b]	10	94479107	*HHEX/IDE*	G	A	0.41	0.04	0.004	1.51×10^{-17}	72.7
rs7076938 [b]	10	115789375	*ADRB1*	T	C	0.73	0.03	0.005	2.91×10^{-10}	39.7
rs4444073	11	10331664	*ADM*	A	C	0.51	0.02	0.004	2.20×10^{-8}	31.3
rs7968682 [b]	12	66371880	*HMGA2*	G	T	0.49	0.04	0.004	4.87×10^{-20}	84.0
rs75844534	15	38667117	*SPRED1*	A	C	0.12	0.04	0.006	1.54×10^{-8}	32.0
rs7402983 [b]	15	99193276	*IGF1R*	A	C	0.41	0.03	0.004	4.61×10^{-10}	38.8
rs222857	17	7164563	*CLDN7*	T	C	0.57	0.03	0.004	5.77×10^{-10}	38.4
rs11698914	20	31327144	*COMMD7*	C	G	0.23	0.03	0.005	2.75×10^{-9}	35.3
rs1012167 [b]	20	39159119	*MAFB*	C	T	0.41	0.02	0.004	1.86×10^{-8}	31.6

* The selected instruments explain 0.3% of the variation in birth weight in the primary MR analysis. The F statistic of individual SNPs ranged from 30.1 to 170.0 with an average F statistic of 58.2.
[a]. Maternally transmitted or non-transmitted alleles were excluded from the secondary MR analysis.
[b]. SNPs were genome-wide significantly associated with potential confounders of PCOS, including BMI, type 2 diabetes, waist/hip circumference, waist-to-hip ratio, metabolic syndrome, glucose metabolism, and lipid metabolism.
BMI, body mass index; CHR: chromosome; EA: effect allele; EAF: effect allele frequency; OA: other allele; P, N, and F indicate p-value, sample size, and F statistic, respectively; SE: standard error; SNP: single-nucleotide polymorphism.

with risk factors for PCOS, including body mass index (BMI), type 2 diabetes, waist/hip circumference, waist-to-hip ratio, metabolic syndrome, glucose metabolism, and lipid metabolism, by searching the GWAS Catalog database (https://www.ebi.ac.uk/gwas/) and the PhenoScanner database (version 2; http://phenoscanner.medschl.cam.ac.uk/). After excluding associated SNPs, 12 and 11 SNPs were retained as genetic instruments in each set of IVs, respectively (Figure 1, Supplemental Tables 2, 3).

Primary MR analysis

Main analysis

The multiplicative random-effects inverse-variance weighted (IVW) method was used as the main analysis (38, 39). Wald ratio estimate for each SNP was calculated by dividing the per allele effect on PCOS by the per allele change in the standard deviation (SD) of birth weight, followed by meta-analyzing the estimates *via* the multiplicative random-effects IVW method, which eventually yielded the IVW estimates. The IVW estimates can be interpreted as the odds ratio (OR) of PCOS risk for one SD change in birth weight.

Sensitivity analyses

Assessment of the IV assumptions

To test the MR relevance assumption (i.e., whether the selected IVs have strong associations with birth weight), the F statistic was calculated for each genetic instrument in our study (40). Furthermore, to ensure that the exclusion restriction assumption holds, Cochran's Q statistic in the IVW analysis (38, 39) was used to assess the heterogeneity of the causal estimates between genetic variants (41). The intercept term of MR-Egger regression was used to test for directional pleiotropy. In addition, we conducted the leave-one-out (LOO) (42) and the Mendelian Randomization

Pleiotropy RESidual Sum and Outlier (MR-PRESSO) (43) analyses to detect strong influential SNPs or outliers.

Robust MR methods

Given that the IVW method provides a biased estimate in the presence of unbalanced horizontal pleiotropy (i.e., directional pleiotropy), we carried out sensitivity analyses by using several pleiotropic-robust methods, including MR-Egger (44), weighted median (45), weighted mode (46), and MR-PRESSO (43) methods, to enhance the robustness of causal inference. When the assumption of the Instrument Strength Independent on Direct Effect (InSIDE) holds, the MR-Egger regression will generate consistent estimates even in the presence of directional pleiotropy (47). The assumption of InSIDE allows for the pleiotropy effects of IVs but requires that the SNP-exposure effects are independent of the pleiotropic effects of SNPs on the outcome, which is a weaker assumption than the IVW assumption. However, the MR-Egger estimate is less precise than the IVW estimate, particularly when the SNP-exposure effect estimates of each genetic variant are relatively homogeneous. Furthermore, we conducted the weighted median analysis which provides reliable estimates when up to 50% of the weight comes from valid IVs. We also carried out the weighted mode analysis which assumes that the most common effect estimate is a consistent estimate of the true effect and allows the majority of variants to be invalid (46). Finally, MR-PRESSO analysis was conducted to estimate the causal effect after correcting for horizontal pleiotropy by removing outliers (43).

Secondary and replication MR analysis

A secondary MR analysis was conducted using the fetal genetic associations extracted from the birth weight GWAS by Warrington et al. (29), after excluding maternally transmitted or non-transmitted alleles that were identified in the GWAS by Juliusdottir et al. (30) In addition, to validate the causal estimates in the primary MR analysis, a replication MR analysis was performed using data from an independent PCOS GWAS meta-analysis in the FinnGen and EstBB (33). To increase the statistical power and precision of causal estimates, a fixed-effect meta-analysis was conducted to pool the IVW estimates from the primary/secondary and replication analyses.

LDSR analysis

LDSR analysis was conducted to assess the genetic correlation between offspring birth weight and PCOS risk by using the fetal genetic associations with birth weight after adjusting for maternal genotypes. First, we conducted LDSR analysis based on summary statistics from birth weight GWAS conducted by Warrington et al. (29) and PCOS GWAS conducted by Day et al. (4) For replication, LDSR analysis was performed based on the summary statistic from another independent PCOS GWAS conducted by Tyrmi et al. (33) The heritability of a single trait or the genetic correlation between two traits can be estimated using LDSR analysis based on the LD structure of a reference panel. Unlike MR, LDSR analysis assesses the genetic correlation between two traits by using genetic variants from the whole genome rather than the causal effect between two traits.

All statistical analyses were conducted using the R packages "TwoSampleMR", "MRPRESSO" and "meta" in R software, version 4.0.0 (R Foundation for Statistical Computing, Vienna, Austria). LDSR analysis was performed using the LDSC software, version 4.0.0 (https://github.com/bulik/ldsc) (34, 35).

Results

Main analysis

The main analysis by IVW suggested little evidence to support a causal relationship between fetal genome determined birth weight and PCOS risk. Causal effect estimates of fetal genome determined birth weight on PCOS risk in the primary MR analysis equated to an OR of PCOS of 0.86 (95% CI: 0.52 to 1.43) for one SD increase in birth weight (Figure 2). Replication analysis using another independent data source from PCOS GWAS meta-analysis generated a consistent causal association of fetal genome determined birth weight with offspring PCOS risk (OR: 0.87, 95% CI: 0.60 to 1.24). Further, consistent estimates (regarding both effect directions and magnitudes) were obtained after meta-analyzing the IVW estimates (OR: 0.87, 95% CI: 0.65 to 1.16) in the primary and replication analyses (Figure 2). After excluding the maternally transmitted and non-transmitted effects, the MR analysis results suggested a null causal effect of fetal genetically predicted birth weight on PCOS risk in both secondary and replication MR analyses (secondary IVW OR: 0.86, 95% CI: 0.54 to 1.39; replication IVW OR: 0.86, 95% CI: 0.58 to 1.26) (Figure 3). A similar pooled IVW estimate was observed (OR: 0.86, 95% CI: 0.64 to 1.16).

Sensitivity analyses

Assessment of the IV assumptions

Genetic instruments for fetal genome determined birth weight, including 22 SNPs, ranged from 30.1 to 170.0 with an average F statistic of 58.2, indicating the absence of weak instruments (Table 1). No evidence for heterogeneity between SNP specific causal effect estimates was found for the primary IVs set (*P* for Cochran Q heterogeneity test = 0.10, replication: P = 0.21) and the secondary IVs set (P = 0.26, replication: P = 0.15), respectively. The proximity of the intercept to the origin in the scatter plot (Supplemental Figure 2) and no significant difference of the intercept from zero in MR-Egger regression suggested little evidence for directional pleiotropy (primary IVs set: P = 0.85, replication: P = 0.08; secondary IVs set: P = 0.70, replication: P = 0.07) (Supplemental Table 4). Meanwhile, the LOO analysis did not

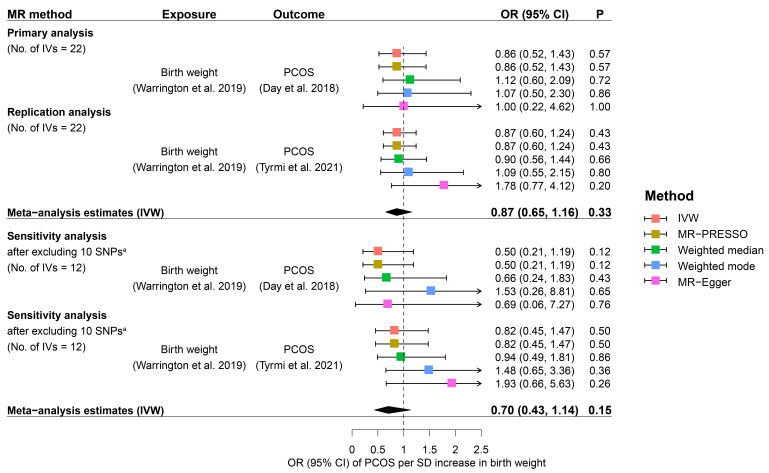


FIGURE 2
Causal effects of fetal genome determined birth weight on future PCOS risk estimated in the primary MR analysis. Squares represent ORs of PCOS per SD increase in birth weight. Error bars represent 95% confidence intervals. A. 10 SNPs that were genome-wide significantly associated with potential confounders of PCOS, including BMI, type 2 diabetes, waist/hip circumference, waist-to-hip ratio, metabolic syndrome, glucose metabolism, and lipid metabolism, were excluded from the MR analysis. BMI, body mass index; CI, confidence interval; IVs, instrumental variables; IVW, inverse variance weighted; MR, Mendelian randomization; MR-PRESSO, Mendelian Randomization Pleiotropy RESidual Sum and Outlier; OR, odds ratio; P, p-value; PCOS, polycystic ovary syndrome; SD, standard deviation; SNP, single nucleotide polymorphism.

detect influential genetic instruments for birth weight in the MR analysis (Supplemental Figure 3). The MR-PRESSO global test did not detect any outliers (primary IVs set: P = 0.13, replication: P = 0.23; secondary IVs set: P = 0.30, replication: P = 0.15) (Supplemental Table 5). For sensitivity analyses by using IVs after excluding potential confounder-related SNPs, we found little evidence for heterogeneity between the causal effect estimates for each SNP, directional pleiotropy, and any outliers.

Results from robust MR methods

The results from robust MR methods are presented in Figures 2 and 3 which were broadly consistent with the IVW analysis results. For the primary IVs set consisting of 22 SNPs, non-significant causal effects of fetal genome determined birth weight on PCOS risk were observed by using MR-PRESSO, weighted median, weighted mode, and MR-Egger methods, respectively (Figure 2). Consistent

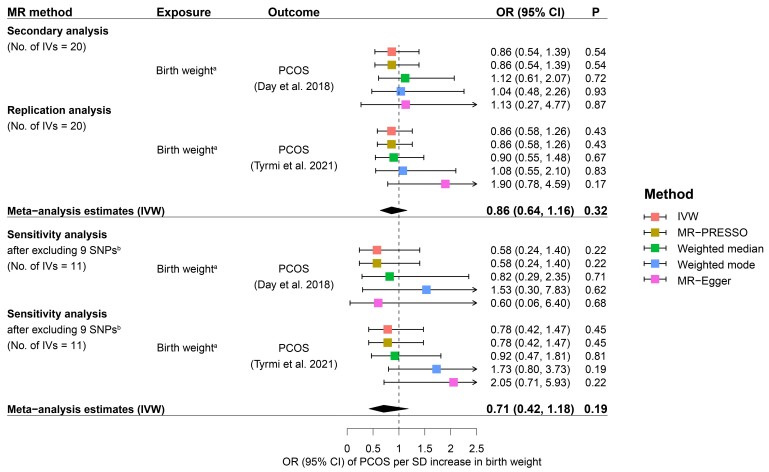


FIGURE 3
Causal effects of fetal genome determined birth weight on future PCOS risk estimated in the secondary MR analysis. Squares represent ORs of PCOS per SD increase in birth weight. Error bars represent 95% confidence intervals. **(A)** rs560887 and rs10872678were identified as maternally transmitted and non-transmitted alleles respectively in the birth weight GWAS by Juliusdottir et al. (30). **(B)** 9 SNPs that were genome-wide significantly associated with potential confounders of PCOS, including BMI, type 2 diabetes, waist/hip circumference, waist-to-hip ratio, metabolic syndrome, glucose metabolism, and lipid metabolism, were excluded from the MR analysis. BMI, body mass index; CI, confidence interval; IVs, instrumental variables; IVW, inverse variance weighted; MR, Mendelian randomization; MR-PRESSO, Mendelian Randomization Pleiotropy RESidual Sum and Outlier; OR, odds ratio; P, p-value; PCOS, polycystic ovary syndrome; SD, standard deviation; SNP, single nucleotide polymorphism.

causal effect estimates derived from robust MR analysis were revealed for the secondary IVs set consisting of 20 SNPs (Figure 3). The results of sensitivity analysis after excluding potential risk factor related SNPs were consistent with the results of the primary and secondary analyses (Figures 2, 3).

LDSR analyses

There was a marginally significant genetic correlation (r_g = -0.14, se = 0.07, P = 0.05) between birth weight and PCOS on a genome-wide scale. Although the result of replication LDSR analysis showed a non-significant genetic correlation between birth weight and PCOS (r_g = -0.16, se = 0.12, P = 0.18), the effect directions and magnitudes were consistent with one another.

Discussion

In this study, we used MR to test the potential causal relationship between fetal genome predicted birth weight and PCOS risk. From a genomic perspective, it is important to discriminate between the maternally transmitted alleles or intrauterine environment effects and the fetal own genetic effects on birth weight. To confirm our results, we further tested whether there was a causal effect of the fetal genome predicted birth weight on offspring PCOS risk, after excluding maternal transmitted and non-transmitted (i.e., maternal intrauterine environment effects) alleles. Our findings provided little evidence for a causal effect of the fetal genome-determined birth weight on offspring developing PCOS in later life. These findings were consistent with previous observational studies that there was no difference in birth weight between women with PCOS and controls (18–20, 23, 48).

Notably, controversial findings were observed in other studies (12, 49), and the LDSR analysis results of the present study suggested a marginally significant genetic correlation between the two traits. Meanwhile, the potential pleiotropic effects underpinning the link between birth weight and PCOS were reported. A recent study found that two genetic variants (i.e., rs2910164 C > G and rs182052 G > A) in genes *MIR146A* and *ADIPOQ*, both of which were related to PCOS, were associated with birth weight (50). Although a causal effect of birth weight on PCOS risk was not observed in the present MR analysis, genetically pleiotropic effects of variants that contribute to the associations between birth weight and PCOS cannot be ruled out. Our study suggested that the association between birth weight and PCOS is likely to be driven by genetic pleiotropy of variants on the fetal genome.

It is noteworthy that observational studies and animal experiments demonstrated that prenatal exposure to androgens possibly in combination with a genetic predisposition may affect birth weight and subsequent PCOS (51–55). In the present study, the potential confounding of maternal genetic effects was minimized by using fetal genetic variants associated with birth weight as IVs and further excluding maternal transmitted or non-transmitted genetic variants.

Strengths and limitations

There are several strengths in our study. We benefited from large sample sizes and study design yielding more reliable results. First, to our best knowledge, we used the summary statistics from the largest published birth weight GWAS with adjusting for maternal genetic effects (n = 406,063 European ancestral individuals) to select genetic variants as IVs, and the outcome data were also extracted from the latest or largest GWAS meta-analyses on PCOS (4, 29). Second, as mentioned above, Warrington et al. separated maternal and fetal genetic effects on birth weight by using SEM. We used the fetal genetic effects on their own birth weight as the IV-exposure associations, after adjusting for maternal genetic effects, which could provide insights into the underlying biological or pathogenic mechanisms between fetal growth and PCOS development in later life. Third, we also constructed IVs for birth weight by filtering out maternal transmitted and non-transmitted variants using summary statistics from a study in which the study design is different from the study conducted by Warrington et al. to minimize the confounding bias due to maternal genetic effects. Fourth, we performed a series of sensitivity analyses with multiple sets of IVs and robust MR methods to strengthen the robustness of causal inference. The sensitivity analysis results were consistent with the results of the main analysis.

Several limitations deserve discussion. First, similar to Chen et al. in their description of the methodology, the allele-specific effects on offspring birth weight/fetal growth by maternally non-transmitted, paternally transmitted, and maternally transmitted alleles were used to represent maternal genetic effect, fetal genetic effect, and combination of both, respectively (37). As suggested in the study conducted by Chen et al. (37), we filtered out maternal non-transmitted and transmitted alleles that indicated maternal genetic effects. However, the allele-specific effects on offspring birth weight/fetal growth by maternally transmitted alleles were composed of maternal and fetal effects. In the original study, genetic dissection of maternal and fetal genetic effects was not performed by modeling maternal and fetal effects using linear combinations of these three haplotype effects, that is maternal genetic effect, fetal genetic effect, and a combination of both. Therefore, more large-scale studies are needed to dissect maternal and fetal genetic effects on birth weight using linear combinations of these three haplotype effects in the future. Second, in our study, there exited moderate sample overlap between data on birth weight (in GWAS by EGG Consortium (29)) and PCOS (in GWAS conducted by Day et al. (4)) Up to 2,867 women in the 1958 British Birth Cohort (56) and the Rotterdam Study (57) in the Netherlands National Trial Register (www.trialregister.nl) were included in both GWASs (4, 29). Sample overlap in two-sample MR analysis would bias causal effects estimation (i.e., inflate the false positive rate) (58), whereas in our study null causal effects of birth weight on PCOS risk were revealed in both primary and replication analyses, thus the potential bias due to sample overlap would not alter the conclusion of our findings. Third, PCOS, as a common and complex genetic disease with multiple etiologies, is caused by genes and environmental factors. In the current study, we

focus on explaining the genetic correlation between birth weight and PCOS risk. Postnatal environmental effects need to be further tested for with genotypes of father-offspring pairs in the future since paternal genotypes might be associated with offspring PCOS risk after adjusting for offspring genotypes in the presence of postnatal environmental effects. Fourth, previous studies suggested that low birth weight was associated with PCOS development (12), however, other findings supported that women born with extra high birth weight increased the risk of PCOS (15). These inconsistent findings might suggest a non-linear causal effect of birth weight on PCOS risk. The present study was limited by its two-sample MR design and GWAS summary statistics used to assess the potential non-linear effect. It is warranted to be investigated through one-sample MR analysis when individual-level data are available. In addition, for the replication analysis of LDSR, a genetic correlation between birth weight with PCOS did not reach statistical significance. Considering that populations, in which the original GWAS meta-analysis for the replication analysis was conducted, were mainly from the FinnGen and Estonian Biobank (33) that were not fully consistent with populations where the birth weight GWAS was conducted, population stratification might arise. Finally, LD scores estimated from European samples of 1000 Genomes reference data may not represent LD scores well for heterogeneous meta-analyses of GWAS, these may lead to the reduced accuracy of results from LDSR analysis (59). However, both results of genetic correlation based on two different data sets showed an inverse genetic correlation. Therefore, we believe that an inverse genetic correlation between birth weight and PCOS is plausible. To avoid a chance finding, genomic restricted maximum likelihood analysis with individual-level genotype data is needed to further validate our results in the future.

Conclusions

In conclusion, our findings provided little evidence for a causal effect of fetal genome predicted birth weight on developing PCOS in later life. However, we found evidence for genetic pleiotropy between birth weight and the future PCOS risk, which has the potential to explain the relationship observed in previous observational studies. In this study, although birth weight within the normal range (i.e., 2,500 to 4,000 grams) may not be causally associated with the risk of PCOS in later life, the potential non-linear causal associations between low/high birth weight and PCOS development need to be further investigated. Further, strong evidence for the genetic pleiotropy between fetal-genome predicted birthweight and later life PCOS risk not only suggests a shared genetic basis but provides novel insight into the common intervention and treatment targets for these two phenotypes.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

DL, YG, JuZ, and JiZ initiated the study. DL, YG, and JiZ undertook statistical analyses and drafted the manuscript. All authors contributed to the interpretation of analysis results and critical revision of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China (grant number 2021YJRC02).

Acknowledgments

Data on birth weight has been contributed by the EGG Consortium using the UK Biobank Resource and has been downloaded from www.egg-consortium.org. We thank all the participants and researchers involved in the FinnGen, Estonian Biobank, UK Biobank, and Icelandic birth register study, and are grateful to Dr. Nicole M. Warrington et al., Dr. Thorhildur Juliusdottir et al., Dr. Felix Day et al., and Dr. Jaakko S Tyrmi et al. for sharing valuable resources.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1140499/full#supplementary-material

References

1. Wolf WM, Wattick RA, Kinkade ON, Olfert MD. Geographical prevalence of polycystic ovary syndrome as determined by region and Race/Ethnicity. *Int J Environ Res Pu* (2018) 15(11):2589. doi: 10.3390/ijerph15112589

2. Escobar-Morreale HF. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. *Nat Rev Endocrinology*. (2018) 14(5):270–84. doi: 10.1038/nrendo.2018.24

3. Carvalho LML, Dos Reis FM, Candido AL, Nunes FFC, Ferreira CN, Gomes KB. Polycystic ovary syndrome as a systemic disease with multiple molecular pathways: a narrative review. *Endocrine Regulations*. (2018) 52(4):208–21. doi: 10.2478/enr-2018-0026

4. Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-Scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PloS Genet* (2018) 14(12):e1007813. doi: 10.1371/journal.pgen.1007813

5. Xita N, Georgiou I, Tsatsoulis A. The genetic basis of polycystic ovary syndrome. *Eur J Endocrinol* (2002) 147(6):717–25. doi: 10.1530/eje.0.1470717

6. Mandy M, Nyirenda M. Developmental origins of health and disease: the relevance to developing nations. *Int Health* (2018) 10(2):66–70. doi: 10.1093/inthealth/ihy006

7. Barker DJ. The origins of the developmental origins theory. *J Intern Med* (2007) 261(5):412–7. doi: 10.1111/j.1365-2796.2007.01809.x

8. Wang Y-X, Li Y, Rich-Edwards JW, Florio AA, Shan Z, Wang S, et al. Associations of birth weight and later life lifestyle factors with risk of cardiovascular disease in the USA: a prospective cohort study. *E ClinicalMedicine* (2022) 51:101570. doi: 10.1016/j.eclinm.2022.101570

9. Zeng P, Zhou X. Causal association between birth weight and adult diseases: evidence from a mendelian randomization analysis. *Front Genet* (2019) 10:618. doi: 10.3389/fgene.2019.00618

10. Horikoshi M, Beaumont RN, Day FR, Warrington NM, Kooijman MN, Fernandez-Tajes J, et al. Genome-wide associations for birth weight and correlations with adult disease. *Nature* (2016) 538(7624):248–52. doi: 10.1038/nature19806

11. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of *In utero* and early-life conditions on adult health and disease. *N Engl J Med* (2008) 359(1):61–73. doi: 10.1056/NEJMra0708473

12. Sadrzadeh S, Hui EVH, Schoonmade LJ, Painter RC, Lambalk CB. Birthweight and PCOS: systematic review and meta-analysis. *Hum Reprod Open* (2017) 2017(2): hox010. doi: 10.1093/hropen/hox010

13. Koivuaho E, Laru J, Ojaniemi M, Puukka K, Kettunen J, Tapanainen JS, et al. Age at adiposity rebound in childhood is associated with PCOS diagnosis and obesity in adulthood-longitudinal analysis of BMI data from birth to age 46 in cases of PCOS. *Int J Obes (Lond).* (2019) 43(7):1370–9. doi: 10.1038/s41366-019-0318-z

14. Stracquadanio M, Ciotta L. Low birth-weight is a PCOS risk factor for southern-Italian women. *Gynecol Endocrinol* (2017) 33(5):373–7. doi: 10.1080/09513590.2017.1283487

15. Mumm H, Kamper-Jorgensen M, Nybo Andersen AM, Glintborg D, Andersen M. Birth weight and polycystic ovary syndrome in adult life: a register-based study on 523,757 Danish women born 1973-1991. *Fertil Steril.* (2013) 99(3):777–82. doi: 10.1016/j.fertnstert.2012.11.004

16. Davies MJ, March WA, Willson KJ, Giles LC, Moore VM. Birthweight and thinness at birth independently predict symptoms of polycystic ovary syndrome in adulthood. *Hum Reprod* (2012) 27(5):1475–80. doi: 10.1093/humrep/des027

17. Cresswell JL, Barker DJP, Osmond C, Egger P, Phillips DIW, Fraser RB. Fetal growth, length of gestation, and polycystic ovaries in adult life. *Lancet* (1997) 350 (9085):1131–5. doi: 10.1016/S0140-6736(97)06062-5

18. Aarestrup J, Pedersen DC, Thomas PE, Glintborg D, Holm JC, Bjerregaard LG, et al. Birthweight, childhood body mass index, height and growth, and risk of polycystic ovary syndrome. *Obes Facts.* (2021) 14(3):283–90. doi: 10.1159/000515294

19. Legro RS, Roller RL, Dodson WC, Stetter CM, Kunselman AR, Dunaif A. Associations of birthweight and gestational age with reproductive and metabolic phenotypes in women with polycystic ovarian syndrome and their first-degree relatives. *J Clin Endocrinol Metab* (2010) 95(2):789–99. doi: 10.1210/jc.2009-1849

20. Laitinen J, Taponen S, Martikainen H, Pouta A, Millwood I, Hartikainen AL, et al. Body size from birth to adulthood as a predictor of self-reported polycystic ovary syndrome symptoms. *Int J Obes* (2003) 27(6):710–5. doi: 10.1038/sj.ijo.0802301

21. Smith GD. Mendelian randomization for strengthening causal inference in observational Studies:Application to gene × environment interactions. *Perspect Psychol Sci* (2010) 5(5):527–45. doi: 10.1177/1745691610383505

22. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* (2008) 123(1):15–33. doi: 10.1007/s00439-007-0448-6

23. Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Commun* (2015) 6(1):8464. doi: 10.1038/ncomms9464

24. Evans DM, Moen G-H, Hwang L-D, Lawlor DA, Warrington NM. Elucidating the role of maternal environmental exposures on offspring health and disease using two-sample mendelian randomization. *Int J Epidemiol.* (2019) 48(3):861–75. doi: 10.1093/ije/dyz019

25. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GÅ, et al. Avoiding dynastic, assortative mating, and population stratification biases in mendelian randomization through within-family analyses. *Nat Commun* (2020) 11 (1):3519. doi: 10.1038/s41467-020-17117-4

26. Foley CN, Mason AM, Kirk PDW, Burgess S. MR-clust: clustering of genetic variants in mendelian randomization with similar causal estimates. *Bioinformatics* (2021) 37(4):531–41. doi: 10.1093/bioinformatics/btaa778

27. Walter S, Kubzansky LD, Koenen KC, Liang L, Tchetgen Tchetgen EJ, Cornelis MC, et al. Revisiting mendelian randomization studies of the effect of body mass index on depression. *Am J Med Genet B* (2015) 168b(2):108–15. doi: 10.1002/ajmg.b.32286

28. Sulc J, Sonrel A, Mounier N, Auwerx C, Marouli E, Darrous L, et al. Composite trait mendelian randomization reveals distinct metabolic and lifestyle consequences of differences in body shape. *Commun Biol* (2021) 4(1):1064. doi: 10.1038/s42003-021-02550-y

29. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, Laurin C, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet* (2019) 51(5):804–14. doi: 10.1038/s41588-019-0403-1

30. Juliusdottir T, Steinthorsdottir V, Stefansdottir L, Sveinbjornsson G, Ivarsdottir EV, Thorolfsdottir RB, et al. Distinction between the effects of parental and fetal genomes on fetal growth. *Nat Genet* (2021) 53(8):1135–42. doi: 10.1038/s41588-021-00896-x

31. Hattersley AT, Beards F, Ballantyne E, Appleton M, Harvey R, Ellard S. Mutations in the glucokinase gene of the fetus result in reduced birth weight. *Nat Genet* (1998) 19(3):268–70. doi: 10.1038/953

32. Moen G-H, Brumpton B, Willer C, Åsvold BO, Birkeland KI, Wang G, et al. Mendelian randomization study of maternal influences on birthweight and future cardiometabolic risk in the HUNT cohort. *Nat Commun* (2020) 11(1):5404. doi: 10.1038/s41467-020-19257-z

33. Tyrmi JS, Arffman RK, Pujol-Gualdo N, Kurra V, Morin-Papunen L, Sliz E, et al. Leveraging northern European population history: novel low-frequency variants for polycystic ovary syndrome. *Hum Reprod* (2021) 37(2):3552–65. doi: 10.1101/2021.05.20.21257510

34. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* (2015) 47 (11):1236–41. doi: 10.1038/ng.3406

35. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* (2015) 47(3):291–5. doi: 10.1038/ng.3211

36. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife* (2018) 7:e34408. doi: 10.7554/eLife.34408

37. Chen J, Bacelis J, Sole-Navais P, Srivastava A, Juodakis J, Rouse A, et al. Dissecting maternal and fetal genetic effects underlying the associations between maternal phenotypes, birth outcomes, and adult phenotypes: a mendelian-randomization and haplotype-based genetic score analysis in 10,734 mother-infant pairs. *PloS Med* (2020) 17(8):e1003305. doi: 10.1371/journal.pmed.1003305

38. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* (2013) 37 (28):658–65. doi: 10.1002/gepi.21758

39. Lawlor DA. Commentary: two-sample mendelian randomization: opportunities and challenges. *Int J Epidemiol.* (2016) 45(3):908–15. doi: 10.1093/ije/dyw127

40. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol.* (2011) 40(3):755–64. doi: 10.1093/ije/dyr036

41. Bowden J, Holmes MV. Meta-analysis and mendelian randomization: a review. *Res Synth Methods* (2019) 10(4):486–96. doi: 10.1002/jrsm.1346

42. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. *Epidemiology* (2017) 28:30–42. doi: 10.1097/EDE.0000000000000559

43. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* (2018) 50(5):693–8. doi: 10.1038/s41588-018-0099-7

44. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-egger regression: the role of the I2 statistic. *Int J Epidemiol.* (2016) 45(6):1961–74. doi: 10.1093/ije/dyw220

45. 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(4):304–14. doi: 10.1002/gepi.21965

46. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data mendelian randomization *via* the zero modal pleiotropy assumption. *Int J Epidemiol.* (2017) 46(6):1985–98. doi: 10.1093/ije/dyx102

47. 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(2):512–25. doi: 10.1093/ije/dyv080

48. Paschou SA, Ioannidis D, Vassilatou E, Mizamtsidi M, Panagou M, Lilis D, et al. Birth weight and polycystic ovary syndrome in adult life: is there a causal link? *PloS One* (2015) 10(3):e0122050. doi: 10.1371/journal.pone.0122050

49. de Zegher F, Reinehr T, Malpique R, Darendeliler F, López-Bermejo A, Ibáñez L. Reduced prenatal weight gain and/or augmented postnatal weight gain precedes polycystic ovary syndrome in adolescent girls. *Obesity* (2017) 25(9):1486–9. doi: 10.1002/oby.21935

50. Silva LR, Melo AS, Salomão KB, Mazin SC, Tone LG, Cardoso VC, et al. MIR146A and ADIPOQ genetic variants are associated with birth weight in relation to gestational age: a cohort study. *J Assist Reprod Gen* (2022) 39(8):1873–86. doi: 10.1007/s10815-022-02532-x

51. Crespi EJ, Steckler TL, Mohankumar PS, Padmanabhan V. Prenatal exposure to excess testosterone modifies the developmental trajectory of the insulin-like growth factor system in female sheep. *J Physiol* (2006) 572(Pt 1):119–30. doi: 10.1113/jphysiol.2005.103929

52. Sathishkumar K, Elkins R, Chinnathambi V, Gao H, Hankins GD, Yallampalli C. Prenatal testosterone-induced fetal growth restriction is associated with down-regulation of rat placental amino acid transport. *Reprod Biol Endocrinol* (2011) 9:110. doi: 10.1186/1477-7827-9-110

53. Manikkam M, Crespi EJ, Doop DD, Herkimer C, Lee JS, Yu S, et al. Fetal programming: prenatal testosterone excess leads to fetal growth retardation and postnatal catch-up growth in sheep. *Endocrinology* (2004) 145(2):790–8. doi: 10.1210/en.2003-0478

54. Dumesic DA, Hoyos LR, Chazenbalk GD, Naik R, Padmanabhan V, Abbott DH. Mechanisms of intergenerational transmission of polycystic ovary syndrome. *Reproduction* (2020) 159(1):R1–R13. doi: 10.1530/REP-19-0197

55. Beckett EM, Astapova O, Steckler TL, Veiga-Lopez A, Padmanabhan V. Developmental programing: impact of testosterone on placental differentiation. *Reproduction* (2014) 148(2):199–209. doi: 10.1530/REP-14-0055

56. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National child development study). *Int J Epidemiol.* (2005) 35(1):34–41. doi: 10.1093/ije/dyi183

57. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, et al. The Rotterdam study: 2014 objectives and design update. *Eur J Epidemiol* (2013) 28(11):889–926. doi: 10.1007/s10654-013-9866-z

58. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample mendelian randomization. *Genet Epidemiol* (2016) 40(7):597–608. doi: 10.1002/gepi.21998

59. Ni G, Moser G, Wray NR, Lee SH. Estimation of genetic correlation *via* linkage disequilibrium score regression and genomic restricted maximum likelihood. *Am J Hum Genet* (2018) 102(6):1185–94. doi: 10.1016/j.ajhg.2018.03.021

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 16 June 2023
DOI 10.3389/fendo.2023.1177268

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Amalia Sertedaki,
National and Kapodistrian University of Athens, Greece
Magdalena Szopa,
Jagiellonian University, Poland

*CORRESPONDENCE
Venkatesan Radha
radharv@yahoo.co.in;
drradha@mdrf.in

RECEIVED 01 March 2023
ACCEPTED 31 May 2023
PUBLISHED 16 June 2023

CITATION
Kavitha B, Ranganathan S, Gopi S, Vetrivel U, Hemavathy N, Mohan V and Radha V (2023) Molecular characterization and re-interpretation of *HNF1A* variants identified in Indian MODY subjects towards precision medicine.
Front. Endocrinol. 14:1177268.
doi: 10.3389/fendo.2023.1177268

Molecular characterization and re-interpretation of *HNF1A* variants identified in Indian MODY subjects towards precision medicine

Babu Kavitha[1], Sampathkumar Ranganathan[2], Sundaramoorthy Gopi[1], Umashankar Vetrivel[3,4], Nagarajan Hemavathy[3], Viswanathan Mohan[5] and Venkatesan Radha[1*]

[1]Department of Molecular Genetics, Madras Diabetes Research Foundation, Indian Council of Medical Research (ICMR) Centre for Advanced Research on Diabetes, Affiliated to University of Madras, Chennai, India, [2]Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India, [3]Department of Bioinformatics, Vision Research Foundation, Chennai, India, [4]Department of Virology Biotechnology, Indian Council of Medical Research (ICMR)-National Institute of Traditional Medicine, Belagavi, India, [5]Department of Diabetology, Madras Diabetes Research Foundation, Chennai and Dr. Mohan's Diabetes Specialties Centre, International Diabetes Federation (IDF) Centre of Education, Chennai, India

Background: *HNF1A* is an essential component of the transcription factor network that controls pancreatic β-cell differentiation, maintenance, and glucose stimulated insulin secretion (GSIS). A continuum of protein malfunction is caused by variations in the *HNF1A* gene, from severe loss-of-function (LOF) variants that cause the highly penetrant Maturity Onset Diabetes of the Young (MODY) to milder LOF variants that are far less penetrant but impart a population-wide risk of type 2 diabetes that is up to five times higher. Before classifying and reporting the discovered variations as relevant in clinical diagnosis, a critical review is required. Functional investigations offer substantial support for classifying a variant as pathogenic, or otherwise as advised by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) ACMG/AMP criteria for variant interpretation.

Objective: To determine the molecular basis for the variations in the *HNF1A* gene found in patients with monogenic diabetes in India.

Methods: We performed functional protein analyses such as transactivation, protein expression, DNA binding, nuclear localization, and glucose stimulated insulin secretion (GSIS) assay, along with structural prediction analysis for 14 *HNF1A* variants found in 20 patients with monogenic diabetes.

Results: Of the 14 variants, 4 (28.6%) were interpreted as pathogenic, 6 (42.8%) as likely pathogenic, 3 (21.4%) as variants of uncertain significance, and 1 (7.14%) as benign. Patients harboring the pathogenic/likely pathogenic variants were able to

successfully switch from insulin to sulfonylureas (SU) making these variants clinically actionable.

Conclusion: Our findings are the first to show the need of using additive scores during molecular characterization for accurate pathogenicity evaluations of *HNF1A* variants in precision medicine.

KEYWORDS

Maturity Onset Diabetes of Young (MODY) subtype-3, acmg-amp guidelines, re-interpretation, pathogenic variants, functional characterization, structural analysis, ACMG-AMP guidelines

1 Introduction

The hepatocyte nuclear factor 1A (***HNF1A***)gene (MIM # 142410) encodes a crucial member of an auto-regulatory transcription circuit in mature and developing pancreas. Heterozygous mutations in *HNF1A* result in the most common form of MODY namely subtype *HNF1A*-MODY. Autosomal dominant inheritance, early onset, and progressive β-cell deterioration resulting in severe hyperglycemia define this type of monogenic diabetes (1–3). This kind of MODY has the highest prevalence and is more common than other subtypes, and it is more common in Europe, North America, and Asia (4–7).

Individuals with *HNF1A* MODY are likely to develop extra pancreatic symptoms such as glycosuria which will appear even before the onset of diabetes due to a low renal glucose threshold (8). This is mainly because *HNF1A* is expressed in tissues such as the kidney, liver, and small intestine, in addition to β-cells. The risk of micro- and macro-vascular problems in *HNF1A*-MODY is comparable to that of T1D and T2DM (9) and hence strict glucose management is required for these individuals. Patients harboring pathogenic variants in *HNF1A* gene are sensitive to low doses of sulfonylureas (10).

The *HNF1A* protein consists of three functional domains namely a dimerization domain (1 – 33 aa), a bipartite DNA-binding domain (homeo domain 100 –184 aa; POU domain 198 –281 aa), and a transactivation domain (282 –631 aa) (11, 12). It binds to DNA as a homodimer or with the structurally related transcription factor *HNF1B* as heterodimers (13, 14). To date, about 564 MODY-causing variants have been identified in the *HNF1A* gene (15, 16). These variations include missense, nonsense, frameshift, in-frame deletions/insertions/duplications, splice site, promoter region, and whole/partial gene deletions. Analyses of these variants have demonstrated that some of them render the protein unstable and poorly expressed (17, 18). Some of the variants affect either the DNA binding or transactivation ability of *HNF1A*. However, patients with the latter type of variants do not exhibit more severe phenotypes (19–21). Finally, a subgroup of variants exert a dominant-negative effect over the normal protein.

It is important that these candidate variants are subjected to rigorous evaluation of pathogenicity to avoid false annotation of causality, which would be an impediment to the translation of genomic research findings to clinical practice and precision medicine. False assignment of pathogenicity can also have severe consequences for patients, resulting in incorrect prognostic and therapeutic advice. Therefore, a comprehensive map is needed, linking mutation status, effect on protein function, and clinical effect that is genotype-function-phenotype. The recent American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) (ACMG-AMP) guidelines classification is based on five tier score system namely pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB) and benign (B) (22). Our previous studies have shown that *HNF1A* -MODY is the most prevalent subtype in India (3) and we identified several variants which were of uncertain significance, Assessing the pathogenicity of these rare protein-coding genetic variants in *HNF1A* is very important in our patient cohort before assigning causality to these variants, as this may lead to change of treatment.

Functional investigation constitutes one of the strongest pieces of evidence for classifying a variant as pathogenic or benign (23). Each variant needs to be assessed by genomic, bioinformatic, structural, and functional lines of evidence for classifying them as pathogenic or benign. Hence, we hypothesized that functional evaluation would enhance the interpretation of the pathogenicity of *HNF1A* variants identified in individuals from families of Indian MODY subjects.

2 Materials and methods

2.1 Subjects

We investigated 14 *HNF1A* variants found in 20 unrelated individuals (11 females and 9 males) from 20 non-consanguineous Indian families. Patients were selected for MODY genetic screening based on the following criteria: a family history of diabetes in multiple generations; an early age at onset of diabetes (< 35 years); lack of obesity, ketosis, and beta cell autoimmunity with detectable endogenous insulin reserve as measured by C peptide which is one of the best biomarkers; and diabetes controllable without insulin for at least 2 years. The study was carried out in compliance with the

Helsinki Declaration (2000); all study participants (or their guardians) provided written, informed consent, and the study was approved by the Madras Diabetes Research Foundation's local institutional ethics committee.

2.2 Genomic analyses

Genomic DNA was isolated from whole blood using the standard protocol. Direct sequencing was carried out on an ABI 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA) using the Big Dye terminator V3.1 chemistry, and the sequences were compared with the public databases. Published primer sequences were used to amplify the DNA for *HNF1A* gene. In addition to the sequencing of patients, we also sequenced 100 normal glucose-tolerant subjects (fasting value <100 mg/dL and 2 hours value <140 mg/dL) to check for the presence or absence of variants in them.

2.3 ACMG classification

All *HNF1A* variations were assessed using the ACMG guidelines, which classify variants as pathogenic (class 5), likely pathogenic (class 4), uncertain significance (class 3), likely benign (class 2), or benign (class 1). Criteria used for the classification of variants are listed in Supplementary Table 1. Public databases such as PubMed, the Human Gene Mutation Database, ClinVar, and LOVD were used and the genome aggregation database (GnomAD) was referred to for population frequency. Bioinformatic prediction tools such as SIFT, PolyPhen2, Mutation Taster, PROVEAN, CADD Score, i mutant 2.0, and Grantham scores were used to assess the pathogenicity (Supplementary Table 2).

2.4 Functional analysis

Human *HNF1A* cDNA (NCBI Entrez Gene BC104910.1) (NM_000545.5) in pcDNA 3.1 His/C vector (Invitrogen Inc, Carlsbad, CA, USA), was used as a template for constructing individual *HNF1A* variants using the QuikChange Lightning Site-directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA), and all constructs were verified by Sanger sequencing. Transiently transfected HeLa and INS1 cells with WT, empty vector (pcDNA3.1), or variant *HNF1A* cDNA were used in functional studies, investigating *HNF1A* (i) transcriptional activity using a rat albumin (in HeLa cells) and *HNF4A P2* (in INS1 cells) promoter-linked luciferase reporter assay system; (ii) DNA binding ability was analyzed using Episeeker DNA-protein binding assay kit (Abcam, ab117139) and a biotinylated oligonucleotide (Sigma Aldrich, St. Luis, MO, US) containing the *HNF1A* binding site in the rat albumin promoter; (iii) protein expression in whole cell lysates by immunoblotting;(iv) nuclear localization by indirect immunocytochemistry; and (v) the glucose-stimulated insulin secretion (GSIS) capacity of the variant *HNF1A* in INS1 β-cells were measured using insulin ELISA kit (Mercodia, Sweden). A detailed methodology is described in the Supplementary Material.

2.5 Structural analysis

The human *HNF1A* protein sequence (P20823) was downloaded from the UniProt database. The Consurf server was used to obtain amino acid conservation scores within the orthologous protein family by comparing 150 homologous sequences. For the structure-based stability prediction, the available crystal structure of *HNF1A* in complex with DNA, PDB ID-1IC8 was remodeled with missing residues and was refined using Modeller10v. The refined Wild type (WT) *HNF1A* was considered for stability analysis of *HNF1A* and also the impact of mutants in the *HNF1A*-DNA complex. The structure of mutants was modeled with a WT-*HNF1A* template using Modeller10v, and the refined WT and MT *HNF1A* were subjected to molecular dynamics simulation studies using Gromacs2020 (10.1080/07391102.2021.1965030). Subsequently, PCA and FEL analyses were carried out to determine the near-native conformation, wherein the *HNF1A*-DNA interactions were analyzed using DNAproDB. A detailed methodology is given in the Supplementary Material.

2.6 Statistical analysis

The results of functional analyses of individual variants are presented as mean (in %) ± standard deviation (SD) and normalized to WT *HNF1A* activity (set as 100%), unless otherwise specified. Experiments were carried out on at least 3 independent occasions unless otherwise specified in the figure legends. Statistical differences between individual variants and WT function were analyzed using GraphPad Prism software (version 8.1.1, GraphPad Software, Inc. San Diego, CA, USA) and raw data (i.e., firefly/renilla ratios) and an unpaired 2-tailed t-test based on n=3. A p-value < 0.05 was considered statistically significant.

3 Results

3.1 Clinical and biochemical characteristics of the subjects with *HNF1A* variants

A total of 14 missense *HNF1A* variants identified in 20 clinical MODY patients were included in this study. All the patients were heterozygous for the variants. In three families, we were able to observe the segregation of variants in affected family members, but for other patients, family samples were not available. Pedigrees of the available families are shown in Supplementary Figure 1. All were negative for β-cell autoantibodies such as GAD and ZnT8 antibodies. The mean ± SD of biochemical parameters were as follows: age at onset of diabetes, 21 ± 6.5 years; Body Mass Index (BMI) - 23 ± 4 kg/m2; duration of diabetes, 9.9 ± 6.7 years; Fasting plasma glucose - 181 ± 64 mg/dL; post prandial plasma glucose - 277 ± 97 mg/dL; glycated hemoglobin (HbA1C)- 9.2 ± 2.4%; fasting C-peptide was 0.9 ± 0.4 pmol/L; stimulated C- peptide was 1.5 ± 0.6 pmol/L; total cholesterol - 169 ± 41 mg/dL; triglycerides - 137 ± 82 mg/dL; High Density Lipoprotein (HDL)- cholesterol - 39 ± 8.5 mg/

dL and Low Density Lipoprotein (LDL)- cholesterol - 94 ± 36 mg/dL. Prior to functional genetic investigations, 11 patients were on insulin treatment; one patient was on insulin + metformin; four patients were on insulin + SU; one patient was on metformin alone and three patients were on SU treatment alone before the genetic investigation. Clinical and biochemical parameters are summarized in Table 1.

Among the 14 variants, four variants (p.Lys120Asn, p.Gln125His,p.Ala367Val,p.Asp602Asn) were novel and not reported in the literature, three variants were previously reported by us (3, 24), and the remaining seven variants were reported in other studies (20, 25–29). Of the 14 variants included in this study, six variants reside in DNA binding domain (91-281 a.a), specifically four variants were mapped to POU_S domain (91-181 a.a), one variant was mapped to POU_H domain (203-279 a.a) and one variant reside in the interface between the POUs and POU_H domains of *HNF1A* protein. The other, eight variants were mapped to the transactivation domain (282- 631 a.a) of *HNF1A* protein (Supplementary Figure 2).

3.2 Functional evaluation

3.2.1 Altered transcriptional activity of *HNF1A* variants

In HeLa cells compared to the WT *HNF1A* activity (set as 100%), the measured levels of transcriptional activity (TA) for five (p.Asn127*,p.Val134Ile,p.Arg200Trp and p.Gly292Fs*25)of the 14 variants were significantly lower (<40%) (Figure 1A, Table 2). Three variants (p.Lys120Asn,p.Pro379Ser, and p.Leu611Pro) had TA activity <50%, while two variants (p.Gln125His and p.Thr354Met) had TA activity of 53 and 62% respectively and reduction observed in all these variants were significant. Two variants p.Ala367Val (61%) and p.Asp602Asn (51%) showed a mildly reduced TA. Two other variants (p.Ala301Thr and p.Glu619Lys) demonstrated TA levels comparable to WT *HNF1A* levels (Figure 1A, Table 2). TA was consistently higher for all these variants when using *HNF4A-P2* promoter in INS-1 cells (activity range 32%–137%) (Figure 1B, Table 2) versus rat albumin promoter in HeLa cells. This is most likely due to interference of endogenous *HNF1A* in INS-1 cells (2- to 4-fold higher basal promoter activity).

3.2.2 Effect of variants on DNA- binding activity of *HNF1A* to target DNA sequence

Three variants (p.Asn127*, p.Arg200Trp and p.Arg272His) localized in the DBD and one variant (p.Gly292Fs*25) in TAD demonstrated severely reduced (<40%) activity. All other variants showed normal binding activity comparable to WT (Figure 1C, Table 2).

3.2.3 Effect of variants on *HNF1A* protein expression

Two variants (p.Gly292Fs*25 and p.Ala301Thr) showed significantly reduced protein expression level (<60%); while four variants (p.Gln125His,p.Asn127*,p.Arg200Trp and p.Asp602Asn), demonstrated reduced expression level (61-75%) and were also significant (Figure 1D, Table 2).

3.2.3 Effect of variants on nuclear localization of *HNF1A* protein

All the 14 *HNF1A* variants were assessed for their ability to translocate to the nucleus of the cell in order to regulate their target gene expression. Only four variants showed reduced (~57-67%) nuclear translocation as assessed by indirect immunocytochemistry (Figure 1E, Table 2). Other variants showed normal nuclear translocation.

3.2.4 Effect of variants on insulin secretion

All 14 variants were also assessed for insulin secretion using GSIS. Under basal conditions (2.8mM glucose), these variants produced insulin in the range of 3-15µg/L of insulin and under stimulated conditions using 16.7mM glucose they produced 1-45µg/L of insulin. When they were treated with 100µM glibenclamide (GBC), the stimulated insulin secretion was enhanced ranging from 8-48µg/L in all the 14 variants tested (Figure 1F, Table 2).

3.3 Structural evaluation

Structural analysis was performed for variants found in DNA binding domain. These variants were mapped onto the crystal structure of *HNF1A* protein (PDB ID: 1IC8). Thereby, all the missense variants, namely p.Lys120Asn, p.Gln125His, p.Val134Ile, p.Arg200Trp, and p.Arg272His, were subjected to the following predictions such as sequence and structural-based stability prediction followed by molecular dynamics (MD).

Sequence-based stability study revealed that the *HNF1A* structure is destabilized by the variants p.Lys120Asn, p.Gln125His, p.Arg200Trp, and p.Arg272His, but not by the variant p.Val134Ile. The crystal structure of *HNF1A* in association with DNA (PDB ID-1IC8), was further modified with missing residues and refined using Modeller10v for the structure-based stability prediction (Figure 2A). According to structure-based prediction, the *HNF1A* variants **p.Lys120Asn, p.Arg200Trp, and p.Arg272His** were shown to have a larger destabilizing impact and more molecular flexibility than the other variants. Among these variants, the p.Arg200Trp variant has a higher destabilizing impact. Variants p.Gln125His and p.Val134Ile had the least destabilizing impact (Figures 2B–K). Since the three variants p.Lys120Asn, p.Arg200Trp and p.Arg272His, showed higher destabilizing effects they were chosen for the MD study.

3.3.1 Molecular dynamics stability analysis of the wild and mutant complexes

The WT-*HNF1A* template was used to simulate the structures of the mutants p.Lys120Asn, p.Arg200Trp, and p.Arg272His. The revised WT and MT *HNF1A* were then submitted to MD simulation investigations using Gromacs2020. When the complexes' MD trajectories were compared to the WT, the variant p.Arg272His showed higher divergence than the variants p.Lys120Asn and p.Arg200Trp in the initial period of simulation. However, variant p.Lys120Asn showed more deviations than

TABLE 1 Clinical and biochemical workup of subjects with *HNF1A* gene variants.

S. No	Patient ID	Gender	Variant	Age at onset (Years)	Duration of Diabetes (Years)	BMI (Kg/m2)	Fasting plasma glucose (mg/dl)	Post prandial plasma glucose (mg/dl)	HbA1C (%)	Fasting C-peptide (pmol/l)	Stimulated-C-peptide (pmol/l)	Total cholesterol (mg/dl)	Triglycerides (mg/dl)	HDL (mg/dl)	LDL (mg/dl)
1	M-026	F	p.Lys120Asn	14	3.7	19.1	188	315	7.1	0.7	1.1	127	61	33	82
2	M-027	M	p.Gln125His	26	6.3	24	134	248	6.9	1	2.2	150	167	32	85
3	M-028	F	p.Asn127Del	14.9	18.1	19.1	277	414	9.5	0.6	0.8	177	134	47	101
4	M-124	M	p.Val134Ile	26.7	6.3	21.9	194	390	9.8	0.5	0.8	136	174	27	94
5	M-125	M	p.Arg200Trp	22.8	16.1	17.9	161	280	8.3	0.5	1.2	152	84	47	88
6	M-126	F	p.Arg200Trp	11	1	23.2	114	171	–	0.9	–	-	-	-	-
7	M-129	F	p.Arg272His	26	8	26.9	106	204	6.4	1.2	2	250	71	45	49
8	M-130	F	p.Arg272His	23	5	23	125	220	6.9	1	2.3	191	209	28	121
9	M-131	F	p.Gly292fs*25	19.1	13	17.3	204	197	10.8	1.1	2	211	176	44	132
10	M-035	F	p.Gly292fs*25	11	4	18.6	127	225	8.7	0.9	1.5	153	114	59	98
11	M-132	M	p.Ala301Thr	28	19	-	114	155	7.3	–	–	193	136	47	125
12	M-133	M	p.Thr354Met	24.8	5	16.2	159	243	6.9	0.7	1.3	125	77	39	71
13	M-138	F	p.Ala367Val	11.6	5	24.1	219	291	11	1	1.6	138	65	43	82
14	M-134	M	p.Pro379Ser	26	6.8	24	268	310	11.4	–	–	270	150	31	209
15	M-135	F	p.Pro379Ser	23	3	26.3	250	310	11.2	2.16	–	145	95	41	85
16	M-036	M	p.Pro379Ser	24	10	27.6	305	521	15.4	0.2	0.3	187	439	37	40
17	M-136	F	p.Pro379Ser	14	–	21.2	289	431	12.7	0.56	1.31	145	95	41	85
18	M-139	F	p.Asp602Asn	14	5	20	159	280	9	2	2.6	195	110	40	70
19	M-137	M	p.Leu611Pro	28.8	18.2	31.6	108	147	6	1.1	3	154	95	30	105
20	M-040	M	p.Glu619Lys	32	27	26.3	134	191	9.5	0.7	1.4	117	160	25	60

10.3389/fendo.2023.1177268

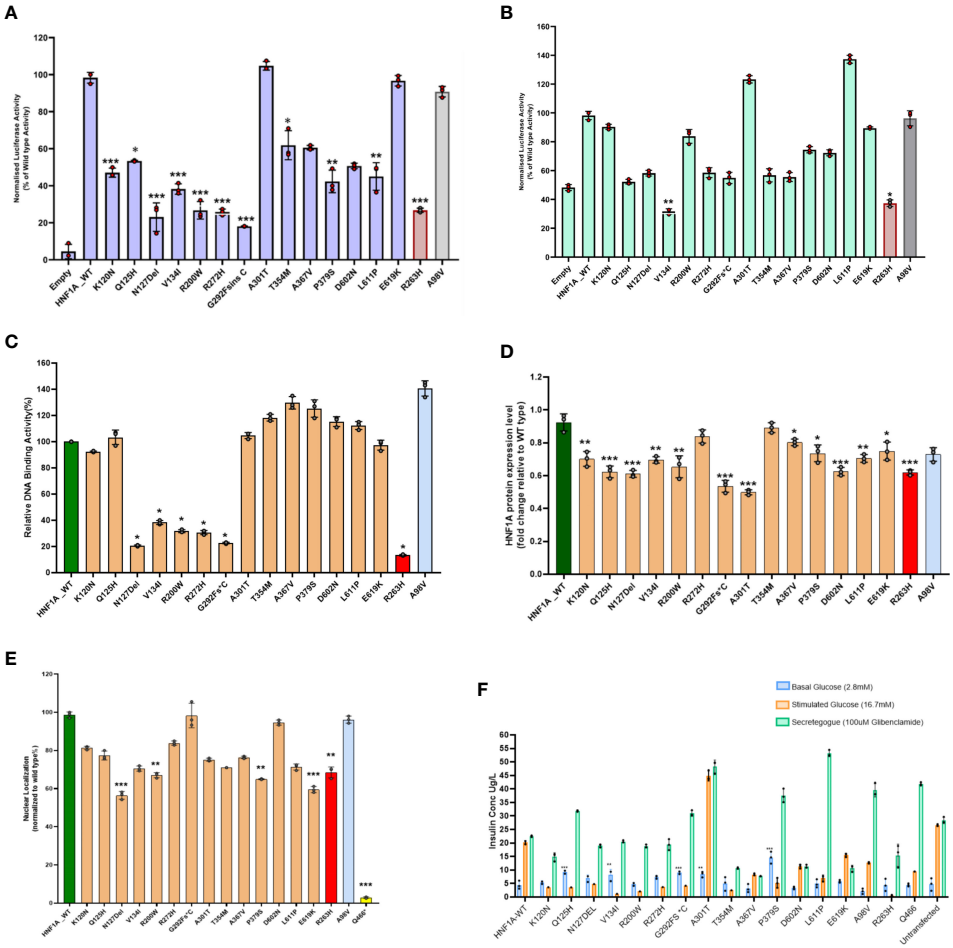


FIGURE 1
Summary of the data obtained from functional studies. **(A)** Transcriptional activity of the *HNF1A* protein variants in HeLa cells; **(B)** Transcriptional activity of the *HNF1A* protein variants in Ins1 cells using HNF4A P2 promoter; **(C)** Assessment of the DNA binding ability of the *HNF1A* protein variants; **(D)** Protein expression levels of the *HNF1A* protein variants in HeLa cells; **(E)** Nuclear localization of the *HNF1A* protein variants in HeLa cells; **(F)** Variant effect on Glucose Stimulated Insulin Secretion. Red bar indicates MODY 3 control variant; Grey bar indicates type 2 diabetes risk variant; Yellow bar indicates variant with poor nuclear translocation effect in *HNF1A* gene. Each bar represents the mean of three independent experiments (n=3) ± SD. P-values were obtained by un-paired student t-test. **** indicates p value <0.001; ** indicates p value <0.01; * indicates p value <0.05.*

p.Arg272His during the last 20 ns of the root mean square deviation (RMSD) plot, a numerical measurement representing the difference between WT and variant protein structures (Figure 2L). The root mean square fluctuation (RMSF) plot, is a calculation of individual residue flexibility, or how much a particular residue moves (fluctuates) during a simulation (Figure 2M), and this showed that residues that interact with DNA were found to have larger deviations in all of the complexes; in particular, residues 179 and 180 of the p.Arg272His variant showed higher deviations of 0.9 nm and 192-193 of the p.Arg272His variant showed higher fluctuations of about 1 nm among the complexes. When compared to WT, the variants p.Lys120Asn and p.Arg272His lost their contact with DNA at the residue level, and their total interactions with DNA also decreased (Figures 2N, O). However, the variant p.Arg200Trp had an increased frequency of interactions with DNA and a greater accessible surface area of all buried solvents (Figures 2N, O). Particularly, the variant residue Trp200 interacts with the minor groove of DNA. From these results, it was revealed that variants p.Lys120Asn and p.Arg272His had lost their interaction with DNA resulting in structural defects.

3.4 Reinterpretation of *HNF1A* variants based on molecular characterization

Pathogenic *HNF1A* variants causing *HNF1A*-MODY are often characterized by significantly decreased TA, poor DNA binding, impaired nuclear targeting, and/or lower protein expression levels in the range of ~20-35% when compared to WT (100%) (19, 21, 30–33). In this study, the cut-off considerations were set at a slightly different level compared to the previous study by Althari et al. (31). Being a more distilled cohort of clinically proven MODY patients, the cut-off of TA<40% was used for pathogenic variants, and TA activity between 40-60% was used for likely pathogenic variants. In addition to this, DNA binding activity, GSIS, and clinical course were considered for ascribing pathogenic and likely pathogenic variants. Therefore, over and above the ACMG/AMP guidelines, the functional and clinical work such as the response to SU have been considered together to re-interpret the variants.

Variants p.Gly292Fs*25 and p.Asn127* were interpreted as pathogenic variants since they have low TA activity along with the

TABLE 2 Summary of the functional studies of the *HNF1A* variants identified in Indian MODY subjects.

	S.No	Amino acid change at protein level	Nucleotide change at c.DNA level	Functional Study								Structure Prediction		
				Transactivation Assay (% WT)		DNA Binding Activity (% WT)	Protein Expression (% WT)	Nuclear Localisation (% WT)	GSIS (Insulin Levels)			Sequence Based Prediction	Structure Based prediction	Molecular Dynamics
				HeLa	Ins 1				Basal	Stimulated	On adding 100μM GBC			
DNA Binding Domain	1	p.K120N	c.360G>C	47	90	92	76	81	5	4	15	Destabilization effect	Higher Destabilization effect	Defect
	2	p.Q125H	c.375G>C	53	52	103	67	77	9	4	32	Destabilization effect	Least Destabilization effect	–
	3	p.N127del	c.377_379delACA	23	58	21	66	57	7	5	19	–	–	–
	4	p.V134I	c.400G>A	38	32	38	75	71	8	1	21	No defect	Least Destabilization effect	–
	5	p.R200W	c.598C>T	27	84	32	71	67	5	2	19	Destabilization effect	Higher Destabilization effect	No defect
	6	p.R272H	c.815G>A	26	59	31	91	84	7	4	19	Destabilization effect	Higher Destabilization effect	Defect
Transactivation Domain	7	p.G292fs*25	c.872-873dupC	18	55	23	58	98	9	4	31	–	–	–
	8	p.A301T	c.901G>A	105	123	105	54	75	8	45	48	–	–	–
	9	p.T354M	c.1061C>T	62	57	118	97	71	5	2	11	–	–	–
	10	p.A367V	c.1100C>T	61	56	130	87	76	3	8	8	–	–	–
	11	p.P379S	c.1135C>T	42	75	125	80	65	15	5	37	–	–	–
	12	p.D602N	c.1804G>A	51	72	115	68	95	3	11	11	–	–	–
	13	p.L611P	c.1832T>C	45	137	112	76	71	5	7	25	–	–	–
	14	p.E619K	c.1855G>A	97	90	97	81	60	6	16	11	–	–	–
	15	p.Arg263His	c.788G>A	27	37	13	67	69	4	1	15	–	–	–
	16	p.Ala98Val	c.293C>T	91	96	141	76	96	2	13	26	–	–	–
	17	p. Gln466*	c.1396 C>T	_	_	_	_	7	_	_	_	–	–	–

Shaded in grey are used as control for the functional assay.

10.3389/fendo.2023.1177268

10.3389/fendo.2023.1177268

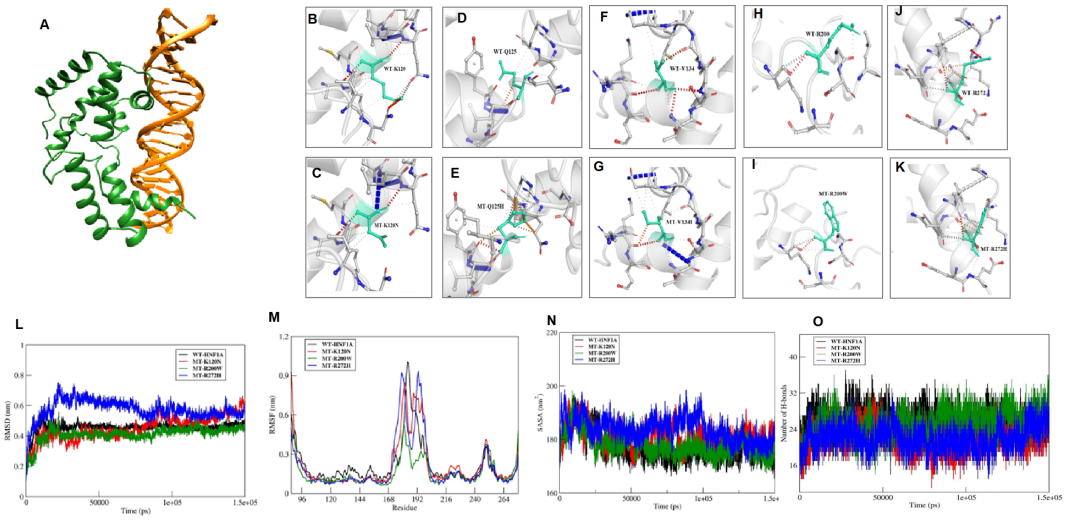


FIGURE 2
In Sillico structural prediction of wild type and variant *HNF1A* protein. **(A)** Remodelled and Refined Wild-type (WT) *HNF1A*-DNA complex; **B-K**) Prediction of Interatomic Interactions of the Wild and mutant forms of *HNF1A* variants, where the Wild-type and mutant residues are coloured in light-green and are also represented as sticks alongside the surrounding residues which are involved in any type of interactions; **(L-O)** Molecular dynamics simulation analysis of the wild and MT forms of *HNF1A* complexes **(L)** RMSD plot **(M)** RMSF plot **(N)** Solvent accessible surface area plot; **(O)** Number of inter hydrogen bonds maintained throughout the MD production run within *HNF1A* and DNA.

reduced DNA binding activity and defect in insulin secretion.p.Arg272His was reinterpreted as a pathogenic variant from their initial interpretation. Seven variants (p.Lys120Asn, p.Gln125His, p.Val134Ile, p.Arg200Trp, p.Thr354Met, p.Pro379Ser, and p.Leu611Pro) were reclassified as likely pathogenic variants from VUS. Three variants (p.Ala367Val, p.Asp602Asn, and p.Glu619Lys) remained VUS after reinterpretation whereas variant p.Ala301Thr was reinterpreted as benign from VUS (Figure 3, Table 3).

3.5 Clinical follow-up of the patients with *HNF1A* variants

Variants designated as pathogenic/likely pathogenic based on functional assessment were investigated for clinical actionability by collecting the follow-up details of the patients over a period of time. The patient (M-026) with variant p.Lys120Asn has been switched from insulin to two doses of SU (glimepiride) along with metformin per day. The patient M-027 with the mutation p.Gln125His (likely pathogenic variant) developed diabetes at the age of 25.7 years and had diabetes for 7 years. Before genetic testing, the patient was treated with insulin and oral hypoglycemic agents (OHA). As a result of genetic studies, the patient was transferred from insulin to two doses of gliclazide per day. His HbA1C levels dropped from 9.6% to 6.4% after his therapy was changed.

Patient M-028, who carries the pathogenic variant p.Asn127*, is diagnosed with diabetes at the age of 14.9 years, with a duration of 15.6 years (Figure 4). The patient was on OHA for around two years before being started on insulin. She is currently on insulin and SU therapy since her β cell reserve was low (CPF-0.6 and CPS-0.9) and she started to develop microvascular and macrovascular complications. Patient M-124 harboring the variant p.Val134Ile

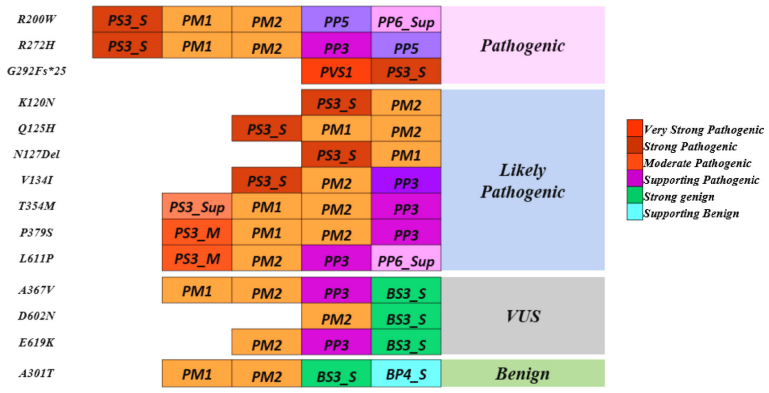


FIGURE 3
Reinterpretation of *HNF1A* variants using molecular characterization.

TABLE 3 Summary of re-interpretation of *HNF1A* gene variants and their clinical actionability, identified in Indian MODY patients based on molecular characterization.

	S.No	Amino acid change at protein level	Nucleotide change at c.DNAlevel	Variant Interpretation_ ACMG guidelines 2015		Functional Study								Structure Prediction			Reinterpation Based on functional evidence		Clinical Actionability
						Transactivation Assay (% WT)		DNABinding Activity (% WT)	Protein Expression (% WT)	Nuclear Localisation (% WT)	GSIS (Insulin Levels)			Sequence Based Predic-tion	Structure Based predic-tion	Molecular Dynamics			
				Evidence	Classification	HeLa	Ins 1				Basal	Stimulated	On adding 100μM GBC				Evidence	Classification	
DNA Binding Domain	1	p.K120N	c.360G>C	PM1, PM2	VUS	47	90	92	76	81	5	4	15	Destabilization effect	Higher Destabilization effect	Defect	PS3_Moderate, PP3_Strong	LP	Actionable
	2	p.Q125H	c.375G>C	PM1, PM2	VUS	53	52	103	67	77	9	4	32	Destabilization effect	Least Destabilization effect	–	PS3_Moderate, PP3 and PP6	LP	Actionable
	3	p.N127del	c.377_379delACA	PM1, PM2	VUS	23	58	21	66	57	7	5	19	–	–	–	PS3_Strong	P	Actionable
	4	p.V134I	c.400G>A	PM1, PM2,PP3	VUS	38	32	38	75	71	8	1	21	No defect	Least Destabilization effect	–	PS3_Strong	LP	Actionable
	5	p.R200W	c.598C>T	PM1, PM2,PP5	VUS	27	84	32	71	67	5	2	19	Destabilization effect	Higher Destabilization effect	No defect	PS3_Strong	P	Actionable
	6	p.R272H	c.815G>A	PM1, PM2, PP3, PP5	LP	26	59	31	91	84	7	4	19	Destabilization effect	Higher Destabilization effect	Defect	PS3_Strong	P	Actionable
Transactivation Domain	7	p.G292fs*25	c.872-873dupC	PVS1	LP	18	55	23	58	98	9	4	31	–	–	–	PS3_Strong	P	Actionable
	8	p.A301T	c.901G>A	PM1, PM2	VUS	105	123	105	54	75	8	45	48	–	–	–	BS3_Strong, BP4_Strong	B	_
	9	p.T354M	c.1061C>T	PM1, PM2,PP3	VUS	62	57	118	97	71	5	2	11	–	–	–	PS3_Supporting	LP	Actionable
	10	p.A367V	c.1100C>T	PM1, PM2	VUS	61	56	130	87	76	3	8	8	–	–	–	BS3_Strong, BP4_Strong	VUS	Unresolved
	11	p.P379S	c.1135C>T	PM1, PM2, PM5,PP3	LP	42	75	125	80	65	15	5	37	–	–	–	PS3_Moderate	LP	Actionable
	12	p.D602N	c.1804G>A	PM1, PM2	VUS	51	72	115	68	95	3	11	11	–	–	–	BS3_Strong	VUS	Unresolved
	13	p.L611P	c.1832T>C	PM1, PM2,PP3	VUS	45	137	112	76	71	5	7	25	–	–	–	PS3_Moderate	LP	Actionable
	14	p.E619K	c.1855G>A	PM1, PM2,PP3	VUS	97	90	97	81	60	6	16	11	–	–	–	BS3_Strong	VUS	Unresolved

P, Pathogenic; LP, Likely Pathogenic; B, Benign; VUS, Variant of Uncertain significance.

10.3389/fendo.2023.1177268

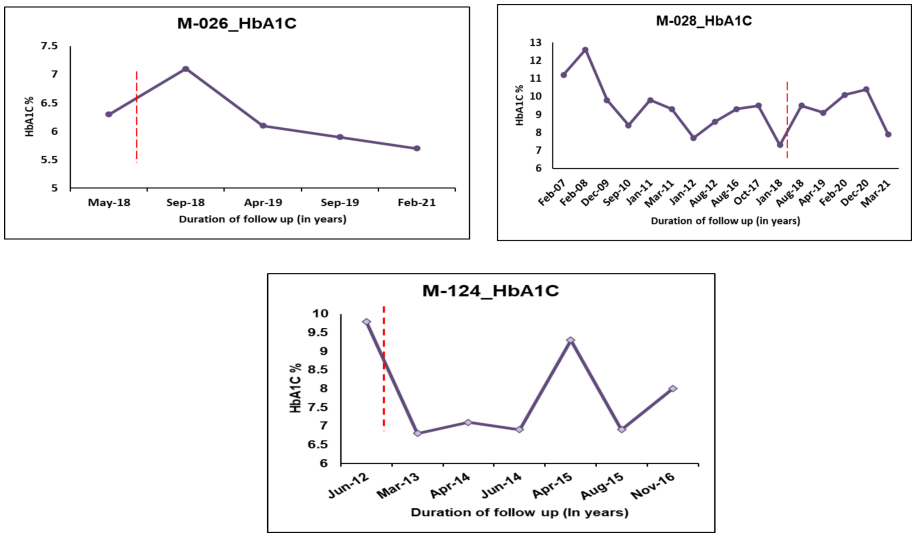


FIGURE 4
HbA1C Trajectories of few *HNF1A* MODY patients after change in treatment based on re-interpretation of the variants. Dotted lines indicate change of treatment.

(Likely pathogenic variant) was diagnosed with diabetes at the age of 26.7 years with diabetes duration of 4 years. Based on functional evidence, patient M-124 with variant p.Val134Ile was transitioned from insulin to a single dose of glipizide per day.

Patient M-126 with the pathogenic variant p.Arg200Trp was switched from insulin to SU. It was advised to continue with SU for patient M-125 who had the same variant. Statins were given for patient M-125 in order to maintain a normal lipid profile. Previous studies have shown two other amino acid changes at the same codon such as p.Arg200Gly and p.Arg200Gln in multiple SU-sensitive *HNF1A*-MODY families (34, 35). The functional effects of these two variants, p.Arg200Gly and p.Arg200Gln, were however not mentioned. All of the patients, including the one from this study, who have the variation in this codon respond to SU. This suggests that the variation is pathogenic and clinically actionable. Patients with pathogenic variant (p.Arg200Trp, p.Arg272His and p.Gly292Fs*25) and likely pathogenic variant (p.Thr354Met and p.Leu611Pro) were also shifted from insulin to SU therapy.

4 Discussion

The comprehension of disease mechanisms is improved by well-established functional investigations on variants, which also offer proof for the pathogenicity of the variants. Studies have demonstrated that functional studies help to clarify the interpretation of *HNF1A*-MODY variants, particularly in the absence of familial segregation or phenotypic data (32).

In this study, we have performed molecular characterization of 14 *HNF1A* variants identified in 20 unrelated individuals from 20 non-consanguineous families among Indian MODY subjects, where the majority of variants have not been reported. Normal transactivation activity of *HNF1A* protein, which depends on the capacity to bind target promoters (DNA) and on an adequate quantity of cellular (nuclear) protein, is necessary for normal *HNF1A* transcription factor function.

Because not all functional tests represent the underlying process and not all variants have the same effects on function (36), we aimed at improving the understanding and interpretation of these findings. Therefore, multiple assays were employed to fully examine the effects of a variant in order to come to a conclusion. These variants were examined utilizing *in vitro* functional pipelines, such as luciferase assays for transactivation, which measure the transcriptional activity of *HNF1A* variants, as well as assays of DNA binding activity, protein expression, and subcellular localization to determine the impact of the variants on the protein function. Additionally, a GSIS assay to examine the impact of these variants on insulin secretion was performed. A distinctive feature of this work is the *in silico* structural analyses to determine if it might identify the variants with functional defects. Since the crystal structure of *HNF1A* is available only for the DNA binding domain, structural investigations were carried out for the missense variants identified only in that region.

A multi-pronged approach using the ACMG guidelines, the functional and structural analyses have been considered together to re-classify these variants. In this work, we focused on the scoring systems and the criteria for re-interpreting the variants. PS3 was assigned when data from well-established *in vitro* functional studies supported a detrimental effect on the gene or gene product; PP3 was assigned when multiple lines of computational evidence and structural prediction supported a detrimental effect on the gene or gene product (conservation, evolutionary, etc.); and BS3 was assigned when well-established *in vitro* functional studies showed no detrimental effect on protein function. In addition, multiple levels of strength, such as strong, moderate, and supporting levels based on functional and structural data were applied to the scoring

approaches employed in this study. Of the 14 variants considered in this study, 1 variant p.Arg272His was interpreted as likely pathogenic, and 11 variants were interpreted as VUS initially based on the ACMG/AMP guidelines. (Figure 3, Table 3).

According to previous studies on the effects of pathogenic *HNF1A*-MODY variants, pathogenic and MODY causal variants impair *HNF1A* activity, DNA binding, and localization (40% compared to WT *HNF1A*) (21, 32), whereas type 2 diabetes risk variants have an impact on *HNF1A* function ranging from 40% -60% compared to WT (30, 31, 33).

Based on the aforementioned cut-offs, many degrees of strength were assigned to each scoring criterion. PS3_Strong scoring criteria were assigned to variants that showed <40% activity than WT activity in at least two functional assays; PS3_Moderate was assigned to variants that showed activity between 40 and 60%; and PS3_Supporting was assigned to variants that showed activity less than 65%. PP3_Strong criterion was assigned when the variant showed defects in all the *in silico* structural prediction analysis. The variant meeting the BS3_Strong criterion had no negative effect on protein function in any of the functional experiments.

The *p.Arg272His* previously interpreted as likely pathogenic was re-interpreted as *pathogenic* based on the evidence *PS3_Strong, PM1, PM2, PP5, and PP3_Strong*. One variant *p.Arg200Trp* interpreted as VUS was re-interpreted as *pathogenic* based on the evidence *PS3_Strong, PM1, PM2, PP3_Supporting, and PP5*. Variant *p.Gly292Fs*25* was interpreted as *pathogenic* based on the evidence *PVS1 and PS3_Strong* and variant *p.Asn127** was interpreted as *likely pathogenic* based on the evidence *PS3_Strong, PM1*. Variants *p.Lys120Asn* and *p.Gln125His* interpreted as VUS was re-interpreted into *likely pathogenic* based on the evidence *PS3_Moderate, PM2, PP3_Strong, and PS3_Moderate, PM2, PP3_Supporting, PP6* respectively. Variant *p.Val134Ile* was re-interpreted into *likely pathogenic* based on evidence *PS3_Strong* and *PM2*. Variant *p.Thr354Met* was re-interpreted as likely pathogenic based on *PS3_Supporting, PM1, PM2, and PP3*. Variant *p.Pro379Ser* was re-interpreted as *likely pathogenic* based on the evidence *PS3_Moderate, PM1, PM2, and PP3*. Variant *p.Leu611Pro* was re-interpreted as *likely pathogenic* based on the evidence *PS3_Moderate, PM2, PP3, and PP6_Supporting*. Variant *p.Ala367Val* remains *VUS* based on the evidence *PM1, PM2, PP3, and BS3_Strong*. Variants *p.Asp602Asn* and *p.Glu619Lys* remain *VUS* based on the evidence *PM2, BS3_Strong and PM2, PP3, and BS3_Strong* respectively. Variant *p.Ala301Thr* was re-interpreted as *benign* based on the evidence *PM1, PM2, BS3_Strong, and BP4_Strong* (Table 3). It is crucial to remember that functional evidence does not always associate a variant to disease outcome; in order to determine clinical actionability, the functional data must be assessed in combination with clinical data (30). It is important to be aware of the fact that both functional and longitudinal clinical follow up are important to establish the clinical actionability of the variants.

Clinical actionability is generally defined as clinically prescribed interventions that are effective for preventing or delaying clinical disease, lowering clinical burden, or improving clinical outcomes in an adult who has not previously received a diagnosis and are specific to the genetic disorder under consideration (37). Based on our results, 4 out of 14 (28.6%) variants were interpreted as pathogenic, 6 variants (42.8%) as likely pathogenic, 3 variants (21.4%) as variants of uncertain significance, and 1 variant (7.14%) as a benign variant. Patients with the ten P/LP variants were able to successfully switch from insulin to SU and sustain good glycemic control, thus making these variants clinically actionable (Table 3).

We performed 3D structural analysis to check whether *in-silico* analysis corroborated with functional investigations in identifying the pathogenic variants and also to have a structural understanding of the variant *HNF1A* proteins. Our *in-silico* analysis showed that variants p.Gln125His, p.Val134Ile have lesser structural defects while variants p.Lys120Asn and p.Arg272His have severe structural defects, and the variant p.Arg200Trp has moderate structural defects. In the case of the p.Val134Ile variant, we found differences between the functional and structural data. Although *in-silico* structural analysis showed that it has a lesser destabilizing effect despite being predicted to be a highly conserved structural residue, our functional data showed that variant p.Val134Ile has a defect in DNA binding thus down-regulating the target genes resulting in reduced insulin secretion (Table 2). Moreover, the patient follow-up also showed that the patient (M-124) responded well to treatment change to SU, making this variant a clinically actionable one (Figure 4).

Our study has a few limitations. Since we could not obtain family samples for many patients, we were unable to conduct family co-segregation studies. In some patients, we did not have adequate clinical data.

In summary, this paper exemplifies the importance of performing molecular characterization after genetic testing, since the understanding of the functional basis of genotypes helps in understanding the phenotype which could lead to changes in clinical treatment for monogenic disorders like MODY. Our findings are the first to show the need of using additive scores during molecular characterization for accurate pathogenicity evaluations of *HNF1A* variants in precision medicine. Furthermore, it is also one of the first to introduce structural understanding to functional implications. The study has led to the delineation of the VUS into pathogenic and disease-causing MODY variants, from non-pathogenic variants. Patients with most pathogenic *HNF1A* variants benefit from OHA treatment; hence, this would assist clinicians in determining the best course of action for patients. While the combination of functional and structural-based approaches may lead to increased certainty in variant–phenotype correlation in a research setting, a functional understanding of the variants helps in precision diagnosis and treatment in a monogenic disorder such as MODY.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional ethics committee, MDRF. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

VR and BK designed and implemented the functional study. BK analyzed the data and wrote the manuscript. SR designed and performed the structural analysis. UV and NH analyzed the structural data. SG performed segregation analysis. VM collected the clinical data and analyzed the manuscript. VR analyzed all data and corrected the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Indian Council of Medical Research (ICMR), India, through the project Functional Studies on Variants of Pancreatic β-cell genes (*HNF1A, HNF4A, ABCC8, and KCNJ11*) in monogenic diabetes – an experimental approach with clinical translational potential; grant no: No. 5/4/5-2/Diab/2020-NCD-III awarded to VR.

Acknowledgments

The authors thank the patients and their parents for giving the blood samples for the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1177268/full#supplementary-material

References

1. Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* (1996) 384(6608):455–8. doi: 10.1038/384455a0

2. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract Endocrinol Metab* (2008) 4(4):200–13. doi: 10.1038/ncpendmet0778

3. Radha V, Ek J, Anuradha S, Hansen T, Pedersen O, Mohan V. Identification of novel variants in the hepatocyte nuclear factor-1alpha gene in south Indian patients with maturity onset diabetes of young. *J Clin Endocrinol Metab* (2009) 94(6):1959–65. doi: 10.1210/jc.2008-2371

4. Kavvoura FK, Owen KR. Maturity onset diabetes of the young: clinical characteristics, diagnosis and management. *Pediatr Endocrinol Rev* (2012) 10(2):234–42.

5. Radha V, Mohan V. Genetic basis of monogenic diabetes. *Curr Sci* (2017) 113:1277–86. doi: 10.18520/cs/v113/i07/1277-1286

6. Broome DT, Pantalone KM, Kashyap SR, Philipson LH. Approach to the patient with MODY-monogenic diabetes. *J Clin Endocrinol Metab* (2021) 106(1):237–50. doi: 10.1210/clinem/dgaa710

7. Hattersley AT, Greeley SAW, Polak M, Rubio-Cabezas O, Njølstad PR, Mlynarski W, et al. ISPAD clinical practice consensus guidelines 2018: the diagnosis and management of monogenic diabetes in children and adolescents. *Pediatr Diabetes* (2018) 19:47–63. doi: 10.1111/pedi.12772

8. Pontoglio M, Prié D, Cheret C, Doyen A, Leroy C, Froguel P, et al. *HNF1A*lpha controls renal glucose reabsorption in mouse and man. *EMBO Rep* (2000) 1(4):359–65. doi: 10.1093/embo-reports/kvd071

9. Steele AM, Shields BM, Shepherd M, Ellard S, Hattersley AT, Pearson ER. Increased all-cause and cardiovascular mortality in monogenic diabetes as a result of mutations in the *HNF1A* gene. *Diabetes Med* (2010) 27(2):157–61. doi: 10.1111/j.1464-5491.2009.02913.x

10. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT. Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* (2003) 362(9392):1275–81. doi: 10.1016/S0140-6736(03)14571-0

11. Baumhueter S, Mendel DB, Conley PB, Kuo CJ, Turk C, Graves MK, et al. HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. *Genes Dev* (1990) 4(3):372–9. doi: 10.1101/gad.4.3.372

12. Tronche F, Yaniv M. HNF1, a homeoprotein member of the hepatic transcription regulatory network. *BioEssays* (1992) 14(9):579–87. doi: 10.1002/bies.950140902

13. Mendel DB, Crabtree GR. HNF-1, a member of a novel class of dimerizing homeodomain proteins. *J Biol Chem* (1991) 266(2):677–80. doi: 10.1016/S0021-9258(17)35222-5

14. Galán M, García-Herrero CM, Azriel S, Gargallo M, Durán M, Gorgojo , JJ, et al. Differential effects of HNF-1α mutations associated with familial young-onset diabetes on target gene regulation. *Mol Med* (2011) 17(3-4):256–65. doi: 10.2119/molmed.2010.00097

15. Cooper DN, Krawczak M. Human gene mutation database. *Hum Genet* (2021) 98(5):629. doi: 10.1007/s004390050272

16. Ellard S, Colclough K. Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (*HNF1A*) and 4 alpha (HNF4A) in maturity-onset diabetes of the young. *Hum Mutat* (2006) 27(9):854–69. doi: 10.1002/humu.20357

17. Vaxillaire M, Abderrahmani A, Boutin P, Bailleul B, Froguel P, Yaniv M, et al. Anatomy of a homeoprotein revealed by the analysis of human MODY3 mutations. *J Biol Chem* (1999) 274(50):35639–46. doi: 10.1074/jbc.274.50.35639

18. Valkovicova T, Skopkova M, Stanik J, Gasperikova D. Novel insights into genetics and clinics of the *HNF1A*-MODY. *Endocr Regul* (2019) 53(2):110–34. doi: 10.2478/enr-2019-0013

19. Bjørkhaug L, Sagen JV, Thorsby P, Søvik O, Molven A, Njølstad PR. Hepatocyte nuclear factor-1 alpha gene mutations and diabetes in Norway. *J Clin Endocrinol Metab* (2003) 88(2):920–31. doi: 10.1210/jc.2002-020945

20. Bellanné-Chantelot C, Carette C, Riveline JP, Valéro R, Gautier JF, Larger E, et al. The type and the position of *HNF1A* mutation modulate age at diagnosis of diabetes in patients with maturity-onset diabetes of the young (MODY)-3. *Diabetes* (2008) 57(2):503–8. doi: 10.2337/db07-0859

21. Balamurugan K, Bjørkhaug L, Mahajan S, Kanthimathi S, Njølstad PR, Srinivasan N, et al. Structure-function studies of *HNF1A* (MODY3) gene mutations in south Indian patients with monogenic diabetes. *Clin Genet* (2016) 90(6):486–95. doi: 10.1111/cge.12757

22. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med* (2015) 17(5):405–24. doi: 10.1038/gim.2015.30

23. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant interpretation: functional assays to the rescue. *Am J Hum Genet* (2017) 101(3):315–25. doi: 10.1016/j.ajhg.2017.07.014

24. Mohan V, Radha V, Nguyen TT, Radha V, Nguyen TT, Stawiski EW, et al. Comprehensive genomic analysis identifies pathogenic variants in maturity-onset diabetes of the young (MODY) patients in south India. *BMC Med Genet* (2018) 19 (1):22. doi: 10.1186/s12881-018-0528-6

25. Thomas H, Badenberg B, Bulman M, Lemm I, Lausen J, Kind L, et al. Evidence for haploinsufficiency of the human *HNF1A*lpha gene revealed by functional characterization of MODY3-associated mutations. *Biol Chem* (2002) 383(11):1691–700. doi: 10.1515/BC.2002.190

26. Chèvre JC, Hani EH, Boutin P, Vaxillaire M, Blanché H, Vionnet N, et al. Mutation screening in 18 Caucasian families suggest the existence of other MODY genes. *Diabetologia* (1998) 41(9):1017–23. doi: 10.1007/s001250051025

27. Kaisaki PJ, Menzel S, Lindner T, Oda N, Rjasanowski I, Sahm J, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in MODY and early-onset NIDDM: evidence for a mutational hotspot in exon 4. *Diabetes* (1997) 46(3):528–35. doi: 10.2337/diab.46.3.528

28. Dusátková P, Průhová S, Sumník Z, Kolousková S, Obermannová B, Cinek O, et al. *HNF1A* mutation presenting with fetal macrosomia and hypoglycemia in childhood prior to onset of overt diabetes. *J Pediatr Endocrinol Metab* (2011) 24(5-6):377–9. doi: 10.1515/jpem.2011.083

29. Elbein SC, Teng K, Yount P, Scroggin E. Linkage and molecular scanning analyses of MODY3/hepatocyte nuclear factor-1 alpha gene in typical familial type 2 diabetes: evidence for novel mutations in exons 8 and 10. *J Clin Endocrinol Metab* (1998) 83(6):2059–65. doi: 10.1210/jcem.83.6.4874

30. Najmi LA, Aukrust I, Flannick J, Molnes J, Burtt N, Molven A, et al. Functional investigations of *HNF1A* identify rare variants as risk factors for type 2 diabetes in the general population. *Diabetes* (2017) 66(2):335–46. doi: 10.2337/db16-0460

31. Althari S, Najmi LA, Bennett AJ, Aukrust I, Rundle JK, Colclough K, et al. Unsupervised clustering of missense variants in *HNF1A* using multidimensional functional data aids clinical interpretation. *Am J Hum Genet* (2020) 107(4):670–82. doi: 10.1016/j.ajhg.2020.08.016

32. Malikova J, Kaci A, Dusatkova P, Aukrust I, Torsvik J, Vesela K, et al. Functional analyses of *HNF1A*-MODY variants refine the interpretation of identified sequence variants. *J Clin Endocrinol Metab* (2020) 105(4):dgaa051. doi: 10.1210/clinem/dgaa051

33. SIGMA Type 2 Diabetes Consortium, Estrada K, Aukrust I, Burtt NP, Mercader JM, García-Ortiz H, et al. Association of a low-frequency variant in *HNF1A* with type 2 diabetes in a Latino population. *JAMA. Diabetes Consortium.* (2014) 311(22):2305–14. doi: 10.1001/jama.2014.6511

34. Brnich SE, Rivera-Muñoz EA, Berg JS. Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. *Hum Mutat* (2018) 39(11):1531–41. doi: 10.1002/humu.23609

35. Zubkova N, Burumkulova F, Plechanova M, Burtt NP, Mercader JM, García-Ortiz H, et al. High frequency of pathogenic and rare sequence variants in diabetes-related genes among Russian patients with diabetes in pregnancy. *Acta Diabetol* (2019) 56(4):413–20. doi: 10.1007/s00592-018-01282-6

36. Pruhova S, Ek J, Lebl J, Sumnik Z, Saudek F, Andel M, et al. Genetic epidemiology of MODY in the Czech republic: new mutations in the MODY genes HNF-4alpha, GCK and HNF-1alpha. *Diabetologia* (2003) 46(2):291–5. doi: 10.1007/s00125-002-1010-7

37. Hunter JE, Irving SA, Biesecker LG, Buchanan A, Jensen B, Lee K, et al. A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. *Genet Med* (2016) 18(12):1258–68. doi: 10.1038/gim.2016.40

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 27 July 2023
DOI 10.3389/fendo.2023.1133260

OPEN ACCESS

EDITED BY
Kavita Jadhav,
Epic-Bio, United States

REVIEWED BY
Yuan Shi,
Children's Hospital of Chongqing Medical University, China
Bo Li,
Jilin University, China

*CORRESPONDENCE
Hui Wu
wuhui@jlu.edu.cn

RECEIVED 03 January 2023
ACCEPTED 06 June 2023
PUBLISHED 27 July 2023

CITATION
Qu Y, Chen L, Guo S, Liu Y and Wu H (2023) Genetic liability to multiple factors and uterine leiomyoma risk: a Mendelian randomization study. *Front. Endocrinol.* 14:1133260. doi: 10.3389/fendo.2023.1133260

Genetic liability to multiple factors and uterine leiomyoma risk: a Mendelian randomization study

Yangming Qu[1], Lanlan Chen[2], Shijie Guo[1], Ying Liu[1] and Hui Wu[1]*

[1]Department of Neonatology, the First Hospital of Jilin University, Changchun, Jilin, China, [2]Department of Hepatobiliary and Pancreatic Surgery, the First Hospital of Jilin University, Changchun, Jilin, China

Background and objective: Uterine leiomyoma is the most common benign tumor in females of reproductive age. However, its causes have never been fully understood. The objective of our study was to analyze the causal association between various factors and uterine leiomyoma using Mendelian randomization (MR).

Methods: Genetic variables associated with risk factors were obtained from genome-wide association studies. Summary-level statistical data for uterine leiomyoma were obtained from FinnGen and the UK Biobank (UKB) consortium. We used inverse variance weighted, MR-Egger, and weighted median methods in univariate analysis. Multivariable MR analysis was used to identify independent risk factors. A fixed-effect model meta-analysis was used to combine the results of the FinnGen and UKB data.

Results: In the FinnGen data, higher genetically predicted age at natural menopause, systolic blood pressure (SBP), diastolic blood pressure (DBP), and fasting insulin were associated with an increased risk of uterine leiomyoma, while higher age at menarche was associated with a reduced risk of uterine leiomyoma. Multivariable MR analysis of SBP and DBP showed that higher DBP might be an independent risk factor of uterine leiomyoma. In the UKB data, the results for age at natural menopause, SBP, DBP, and age at menarche were replicated. The result of the meta-analysis suggested that uterine leiomyoma could also be affected by polycystic ovary syndrome (PCOS), endometriosis, and 2-hour glucose level.

Conclusion: Our MR study confirmed that earlier menstrual age, hypertension, obesity, and elevated 2-hour glucose post-challenge were risk factors for uterine leiomyoma, and the causal relationship between smoking and uterine leiomyoma was ruled out. In addition, later age of menopause and endometriosis were found to increase the risk of uterine leiomyoma, while PCOS was found to decrease the risk.

KEYWORDS

uterine leiomyoma, risk factor, Mendelian randomization, influence, study

1 Introduction

Uterine leiomyomas (fibroids) are the most common benign tumor in females of reproductive age. Due to differences in race, diagnostic criteria, and study designs, epidemiological reports of uterine leiomyoma incidence vary widely, ranging from 4.5 to 68.6% (1). Since many uterine leiomyomas are asymptomatic and the histological incidence is more than twice as high as the clinical incidence, its true incidence may be underestimated in most studies (2). Johnson et al. (3) reported that the incidence of uterine leiomyoma increased with age, with a cumulative incidence of over 70% by menopause. In the United States, uterine leiomyoma accounts for 29% of gynecological hospitalizations among women aged 15-54 and 40-60% of all hysterectomies (4, 5).

Uterine leiomyoma has been shown to seriously impair woman's quality of life and may lead to endometrial cancer (6). It can cause extensive or prolonged menstrual bleeding (leading to anemia, fatigue, and dysmenorrhea), abdominal swelling, painful intercourse, bladder or bowel dysfunction (leading to urinary incontinence or retention, pain, or constipation), and reproductive problems (such as impaired fertility, pregnancy complications, and miscarriage) (7, 8). If left untreated, it can even lead to death (9).

At present, the main treatment of uterine leiomyoma is hysterectomy, which is expensive and affects fertility. Nearly a quarter of women who have tried non-surgical treatment for fibroids choose to have surgery within a year (10). Many women opt for minimally invasive treatments to preserve their uterus, such as myomectomy, uterine artery embolization, and endometrial ablation. However, relapse is common after treatment (10). Therefore, it is necessary to clarify the risk factors of uterine leiomyoma for early prevention.

Several risk factors such as early age at menarche, early age at first birth, obesity, and hypertension have been established as increasing the risk of uterine leiomyoma (5, 11). However, due to the large number of undetected patients and the large bias of epidemiological data and risk factor evidence, its etiology is still far from being fully understood. In addition, some conflicting conclusions make it difficult to discover the true cause of uterine leiomyoma. For example, earlier studies have shown that smoking has a protective effect on fibroids (12, 13), while subsequent studies have shown that smoking increases the risk of uterine leiomyoma (14). One study showed that, among black women, those who self-reported PCOS had a 65% increased risk of fibroids compared with those who did not self-report PCOS (15). However, another study has shown that patients with PCOS had a lower risk of fibroids than women with normal ovaries (16). Additionally, some studies have found an inverse correlation between diabetes and uterine leiomyoma (13, 15, 17), and other researchers hypothesize that insulin stimulates fibroid growth (18, 19).

Therefore, whether there is a causal relationship between uterine leiomyoma and these factors still needs further analysis. Mendelian randomization (MR) is an emerging method of epidemiological causal inference, which uses genetic variation to determine the causal relationships between risk factors and outcomes. It relies on the natural random assortment of genetic variation during meiosis to distribute genetic variation randomly in a population, reducing bias caused by confounding or reverse causation (20). In our study, MR was used to explore the causal relationship between 20 risk factors and uterine leiomyomas. To our knowledge, this is the first MR study to examine the risk factors for uterine leiomyoma.

2 Methods

2.1 Summary statistics for risk factors

The summary statistics of anthropometric traits were from the GIANT (Genetic Investigation of Anthropometric Traits) consortium. For body mass index (BMI), the genome-wide association study (GWAS) included 234,069 Europeans and used sex, age, age squared, and principal components as covariates (21). For waist circumference, hip circumference, and waist-to-hip ratio, the GWAS included 210,088 Europeans and adjusted for age, age square, and study-specific covariates if necessary (22).

The summary statistics of DBP and SBP were obtained from the International Consortium for Blood Pressure, with 757,601 participants of European ancestry, and sex, age, and age squared were adjusted (23).

The summary statistics of serum 25-hydroxyvitamin D concentrations were from the SUNLIGHT consortium with 79,366 participants of European ancestry (24). The lead genetic variants of plasma vitamin C were derived from a GWAS meta-analysis of 52,018 Europeans from the Fenland study, the European Prospective Investigation into Cancer and Nutrition (EPIC)-InterAct study, the EPIC-Norfolk study, and the EPIC-CVD study (25).

In terms of smoking and drinking, the GWAS was conducted by the Sequencing Consortium of Alcohol and Nicotine use, which included 249,752 European participants for smoking and 335,394 European participants for drinking (26). Smoking was defined as the average number of cigarettes smoked per day, while drinking was the average number of alcoholic drinks consumed per week (including all types of alcohol). Age, sex, age-by-sex interaction, and the top 10 genetic principal components were used as covariates.

Three reproductive traits were involved in our study. The summary statistics of age at menarche (AAM) were from the largest meta-analysis of the ReproGen consortium, 23andMe, and the UK Biobank cumulatively including 329,345 women of European ancestry (27). The summary statistics of age at natural menopause (ANM) were from the ReproGen consortium with 69,360 European women (28). The summary statistics of age at first birth (AFB) were from a GWAS with 69,360 European individuals (29).

The summary statistics of PCOS were from a large-scale genome-wide meta-analysis with 10,074 PCOS cases and 103,164 controls of European ancestry (30). Cases were diagnosed with PCOS based on National Institutes of Health (NIH) or Rotterdam Criteria or by self-report, and age and BMI were used as covariables. The summary data for endometriosis included 17,045 endometriosis cases and 191,858 controls, 93% of whom were European (31).

The summary-level genetic data of T2D (Type 2 diabetes) were from the Diabetes Genetics Replication and Meta analysis consortium (32). A total of 74,124 T2D cases and 824,006 controls of European ancestry from 32 GWAS (with and without adjustment for BMI) were included.

The GWAS summary statistics of glycemic traits, including fasting glucose, fasting insulin, glycated hemoglobin (HbA1c), and 2-hour glucose post-challenge in an oral glucose tolerance test, were obtained from MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium) (33). The GWAS included 281,416 participants, 70% of whom were of European ancestry. Our study used only European summary statistics and adjusted for covariates specific to the study.

2.2 GWAS summary statistics of uterine leiomyoma from the FinnGen and UKB consortia data sets

The GWAS summary statistics of uterine leiomyoma were obtained from FinnGen (https://r4.finngen.fi/) and the UKB. In the FinnGen data, the GWAS included 18,060 cases and 105,519 controls of European ancestry. The GWAS from the UKB included 4,351 cases and 332,848 controls and was conducted by the Neale Lab (http://www.nealelab.is/uk-biobank). Uterine leiomyoma is defined as ICD (International Classification of Diseases) 10: D25. When assessing causality, the FinnGen GWAS was used as the discovery set and the UKB GWAS as the validation set, considering that the FinnGen data had a higher proportion of cases. The main design of this study is shown in Supplementary Figure 1.

2.3 Ethics and consent statement

Specific ethical and consent statements for each GWAS in this study can be found in the original GWAS publications. The FinnGen Biobank GWAS was approved by the FinnGen Steering Committee. The Neale Lab received approval to conduct the GWAS from the Ethics Advisory Committee of the UKB. All of these data are de-identified, freely downloadable, and can be used without restriction.

2.4 Statistical analysis

We used a two-sample Mendelian randomization analysis to explore the potential causal relationship between 20 risk factors and uterine leiomyoma. Single nucleotide polymorphism (SNPs) with genome-wide significance ($P < 5 \times 10^{-8}$) and minor allele frequency >0.01 were included. Then, these SNPs were clumped based on the linkage disequilibrium $r^2 < 0.01$. The power of each SNP was assessed using F statistics (34) ($F = beta^2/se^2$), and the general F statistics of each exposure were also calculated. SNPs with weak statistical power were deleted (F statistics< 10).

We used inverse variance weighted (IVW) analysis as the primary statistical method. Although this method assumes that there is no heterogeneity between genetic variants (potentially due to pleiotropy), it has the strongest power to detect associations (35). In addition, we used two sensitivity analyses methods, including the MR-Egger (36) and weighted median (37) methods, as supplements to IVW. MR-Egger intercept and MR-PRESSO (38) methods were used to detect horizontal pleiotropy, and Cochran's Q statistic was used to assess the heterogeneity. When there were outliers, the MR-PRESSO-corrected results would be reported in the main results. If heterogeneity still exists, the median based estimation was used as primary analysis. A false discovery rate (FDR) was used to adjust for multiple testing. In multivariable MR (MVMR) analysis, the IVW model was also the main method and the MR-Egger method was the complementary method.

A fixed-effects model meta-analysis was used to combine the results of the training set and verification set. All statistical analyses were performed using R software 4.1.2 (https://www.r-project.org/). The IVW, MR-Egger and weighted median methods were performed using the R packages "Two Sample MR" and "Mendelian Randomization". The MVMR was performed using the R packages "Mendelian Randomization" and "MVMR". P<0.05 was used as significance threshold. The mRnd was used to calculate the statistical power (39) for MR (https://cnsgenomics.shinyapps.io/mRnd/).

3 Results

3.1 Summary characteristics of risk factors

The number of SNPs ranged from 6 to 821, explaining 0.15% to 7.01% of the variance. The F statistics of each SNP and exposure were greater than 10, indicating that all instrumental variables had sufficient validity (Table 1).

3.2 Discovery results of uterine leiomyoma in the FinnGen consortium data set

In the FinnGen data set, a higher genetically predicted age at natural menopause (OR=1.0864 per standard deviation of age at natural menopause increase, 95%CI=1.0429-1.1317,$P=6.97\times10^{-5}$), SBP (OR=1.0073 per standard deviation of SBP increase, 95% CI=1.0026-1.0120,$P=2.30\times10^{-3}$), DBP (OR=1.0118 per standard deviation of DBP increase, 95%CI=1.0040-1.0197,$P=3.09\times10^{-3}$), and fasting insulin (OR=1.7342 per standard deviation of fasting insulin increase, 95%CI=1.1455-2.6253, $P=9.25\times10^{-3}$) were associated with an increased risk of uterine leiomyoma, while a genetically predicted higher age at menarche (OR=0.8435 per standard deviation of age of menarche increase, 95%CI=0.7999-0.8894, $P=3.27\times10^{-10}$) was associated with a reduced risk of uterine leiomyoma (FDR<0.05). T2D, endometriosis, and BMI showed a positive association with uterine leiomyoma risk (FDR> 0.05 and IVW $P< 0.05$). Both SBP and DBP are indicators of blood pressure, and some SNPs may be associated with both SBP and DBP. Therefore, we used multivariable MR to adjust the results of SBP and DBP. Multivariable MR analysis of SBP and DBP showed that a higher DBP might be an independent risk factor of uterine leiomyoma (adjusted OR=1.0309, 95%CI=1.0091-1.0533, $P=5.34\times10^{-3}$), while SBP was not significant (adjusted OR=0.9913, 95%CI=0.9787-1.0041, $P=0.184$).

10.3389/fendo.2023.1133260

TABLE 1 Summary characteristics of risk factors.

Exposure	Data source	NSNP	Unit	Sample	R^2 (%)	F	PMID
BMI	GIANT consortium	92	SD	234,069	2.39	62.27	25673413
Serum 25-hydroxyvitamin D concentrations	SUNLIGHT consortium	6	SD	79,366	0.8	106.67	31100827
Drinking	Sequencing Consortium of Alcohol and Nicotine use	39	SD	335,394	0.84	72.84	30643251
PCOS	A large-scale genome-wide meta-analysis	14	logOR	113,238	0.50	40.64	30566500
Endometriosis	GWAS	14	logOR	208,903	0.26	38.89	28537267
Smoking	Sequencing Consortium of Alcohol and Nicotine use	28	SD	249,752	1.04	93.73	30643251
Age at menarche	ReproGen consortium	321	SD	329,345	6.29	68.80	28436984
2-hour glucose	MAGIC	14	SD	281,416	0.31	62.50	34059833
Fasting glucose	MAGIC	69	SD	281,416	2.74	114.87	34059833
Fasting insulin	MAGIC	36	SD	281,416	0.70	55.10	34059833
HbA1c	MAGIC	78	SD	281,416	2.84	105.43	34059833
Age at natural menopause	ReproGen consortium	48	SD	69,360	4.37	65.99	26414677
Age at first birth	GWAS	10	SD	251,151	0.15	37.73	27798627
T2D	Diabetes Genetics Replication and Meta-analysis consortium	248	logOR	898,130	1.78	65.61	30297969
SBP	International Consortium for Blood Pressure	776	SD	757,601	6.49	67.69	30224653
DBP	International Consortium for Blood Pressure	821	SD	757,601	7.01	69.49	30224653
Circulating vitamin C concentration	A GWAS meta-analysis	10	SD	52,018	1.72	91.02	33203707
Waist-to-hip ratio	GIANT consortium	34	SD	210,088	0.76	47.31	25673412
Waist circumference	GIANT consortium	45	SD	210,088	1.26	59.56	25673412
Hip circumference	GIANT consortium	55	SD	210,088	1.42	55.01	25673412

The results of heterogeneity, pleiotropy, weighted median, and MR-Egger are shown in Table 2. There was heterogeneity in age at natural menopause, SBP, DBP, fasting insulin, age of menarche, T2D, and endometriosis, and they all showed MR-PRESSO-corrected results if outliers were detected. No horizontal pleiotropy was found. For some robust MR estimators, such as endometriosis, age at menarche, age at natural menopause, fasting insulin, T2D, and SBP, their IVW results were supported by MR-Egger or weighted median. However, the IVW results of waist-to-hip ratio and DBP were not supported by MR-Egger or weighted median, which may be related to the presence of horizontal pleiotropy and the detected outliers. The statistical power for the FinnGen outcome ranged from 96% to 100%.

3.3 Validation results of uterine leiomyoma in the UKB consortium data set

In the validation set, the MR results of SBP, DBP, age at menarche, and age at natural menopause were consistent with the training set. A higher age at natural menopause, SBP, and DBP were associated with an increased risk of uterine leiomyoma, while a higher age of menarche was associated with a reduced risk of uterine leiomyoma (Figure 1). No horizontal pleiotropy was found for these risk factors. After removing outliers, the odds of uterine leiomyoma increased per 1-SD increase in SBP (OR=1.0002, 95%CI=1.0001-1.0003, $P=1.01\times10^{-4}$), age at natural menopause (OR=1.0013, 95%CI=1.0006-1.0020, $P=2.08\times10^{-4}$), and DBP (OR=1.0002, 95%CI=1.0001-1.0004, $P=8.95\times10^{-3}$). Moreover, 1-SD increase in age at menarche (OR=0.9980, 95%CI=0.9969-0.9990, $P=1.73\times10^{-4}$) was associated with a reduced risk of uterine leiomyoma.

In addition, the results of the validation set showed that genetic liability to PCOS (OR=0.9974, 95%CI=0.9954-0.9994, $P=1.09\times10^{-2}$), 2-hour glucose level (OR=1.0032, 95%CI=1.0001-1.0064, $P=4.57\times10^{-2}$), and endometriosis (OR=1.0038, 95%CI=1.0001-1.0075, $P=4.34\times10^{-2}$) were also influential factors for uterine fibroids.

It is worth noting that the statistical power of the UKB results was not sufficient (<50%). The reason may be that UKB data set has fewer cases than the FinnGen data set, resulting in lower statistical power.

3.4 Combined result of uterine leiomyoma from meta-analysis

The results of the meta-analysis further confirmed the previous findings that a higher age at natural menopause (OR=1.0013, 95% CI=1.0006-1.0020, $P=2.94\times10^{-4}$), SBP (OR=1.0002, 95%CI=1.0001-

10.3389/fendo.2023.1133260

TABLE 2 Two-sample Mendelian randomization estimates of MR-Egger and weighted median methods.

	NSNP	MR-Egger				Weighted median				$P_{heterogenelty}$	$P_{pleiotropy}$
		OR	95% LCI	95% UCI	*p*	OR	95% LCI	95% UCI	*p*		
FinnGen											
BMI	91	1.2241	0.9059	1.6539	0.191	1.2042	1.0064	1.4409	0.049	1.114	0.589
Waist-to-hip ratio	33	1.5922	0.6209	4.0826	0.340	1.2690	0.9468	1.7007	0.111	0.005	0.503
Waist circumference	45	0.9948	0.5621	1.7607	0.986	1.1309	0.9001	1.4209	0.291	0.307	0.689
Hip circumference	54	1.3795	0.8947	2.1270	0.151	1.1408	0.9363	1.3900	0.191	0.076	0.210
Serum 25-hydroxyvitamin D concentrations	6	1.6075	0.9683	2.6684	0.140	1.3320	0.9813	1.8081	0.066	0.420	0.301
Circulating vitamin C concentration	10	1.0356	0.8117	1.3213	0.785	0.9769	0.8169	1.1683	0.798	0.647	0.585
Drinking	39	0.3618	0.1484	0.8818	0.031	0.4423	0.2678	0.7308	0.001	0.003	0.094
Smoking	28	0.9393	0.8054	1.0956	0.433	0.9348	0.8267	1.0570	0.282	0.937	0.757
PCOS	13	0.6201	0.3372	1.1406	0.153	0.8898	0.8089	0.9788	0.016	<0.001	0.261
Endometriosis	14	5.0109	0.8617	29.141	0.098	1.1982	1.0337	1.3889	0.016	<0.001	0.189
Age at menarche	303	0.8870	0.7682	1.0243	0.103	0.8584	0.7904	0.9324	<0.001	<0.001	0.461
Age at natural menopause	43	1.2002	1.0752	1.3397	0.002	1.1337	1.0906	1.1784	<0.001	<0.001	0.064
Age at first birth	10	0.7951	0.3458	1.8282	0.604	1.0355	0.9139	1.1732	0.584	0.886	0.577
2-hour glucose	13	0.6588	0.3459	1.2547	0.230	0.9582	0.8047	1.1409	0.631	<0.001	0.138
Fasting glucose	65	1.1888	0.7707	1.8337	0.437	1.1895	0.9260	1.5281	0.174	<0.001	0.347
Fasting insulin	36	1.6367	0.4093	6.5449	0.491	1.7667	1.0794	2.8915	0.024	<0.001	0.932
HbA1c	74	1.2990	0.7970	2.1172	0.297	1.0847	0.7648	1.5386	0.648	0.079	0.339
T2D	230	0.9826	0.9149	1.0553	0.631	1.0539	1.0009	1.1096	0.046	<0.001	0.054
SBP	741	1.0139	1.0016	1.0263	0.027	1.0051	0.9989	1.0113	0.105	<0.001	0.255
DBP	776	1.0061	0.9866	1.0260	0.544	1.0060	0.9953	1.0167	0.275	<0.001	0.533
UKB											
BMI	91	0.9990	0.9931	1.0049	0.736	1.0024	0.9988	1.0059	0.197	0.633	0.501
Waist-to-hip ratio	33	0.9964	0.9754	1.0178	0.741	1.0047	0.9982	1.0113	0.153	0.010	0.505
Waist circumference	44	1.0014	0.9896	1.0133	0.817	1.0008	0.9958	1.0058	0.760	0.605	0.910
Hip circumference	54	1.0029	0.9946	1.0113	0.498	1.0030	0.9988	1.0072	0.152	0.483	0.651
Serum 25-hydroxyvitamin D concentrations	6	0.9973	0.9875	1.0072	0.621	0.9972	0.9910	1.0035	0.384	0.925	0.966
Circulating vitamin C concentration	10	0.9975	0.9921	1.0029	0.383	0.9989	0.9953	1.0024	0.531	0.126	0.316
Smoking	28	0.9990	0.9957	1.0024	0.580	1.0000	0.9972	1.0030	0.972	0.497	0.935
Drinking	38	0.9991	0.9879	1.0104	0.877	1.0010	0.9923	1.0099	0.812	0.037	0.858
PCOS	12	0.9901	0.9810	0.9995	0.066	0.9979	0.9957	1.0000	0.052	0.047	0.153
Endometriosis	14	1.0142	0.9938	1.0350	0.200	1.0041	1.0013	1.0071	0.005	<0.001	0.333
Age at menarche	306	0.9974	0.9958	0.9991	0.075	0.9983	0.9966	0.9999	0.049	0.031	0.686
Age at natural menopause	45	1.0016	1.0000	1.0033	0.058	1.0011	1.0004	1.0019	0.003	<0.001	0.677
Age at first birth	10	0.9987	0.9812	1.0166	0.891	1.0023	0.9996	1.0051	0.096	0.583	0.788

(Continued)

TABLE 2 Continued

	NSNP	MR-Egger				Weighted median				$P_{heterogenelty}$	$P_{pleiotropy}$
		OR	95% LCI	95% UCI	*p*	OR	95% LCI	95% UCI	*p*		
2-hour glucose	14	0.9994	0.9907	1.0082	0.897	1.0014	0.9981	1.0047	0.416	0.046	0.375
Fasting glucose	67	1.0011	0.9936	1.0086	0.781	1.0029	0.9975	1.0083	0.299	0.065	0.542
Fasting insulin	35	0.9907	0.9671	1.0148	0.450	1.0002	0.9905	1.0100	0.967	0.097	0.458
HbA1c	75	1.0013	0.9923	1.0103	0.782	0.9983	0.9913	1.0054	0.645	0.136	0.830
T2D	236	1.0003	0.9990	1.0017	0.639	1.0003	0.9991	1.0015	0.665	0.005	0.937
SBP	757	1.0001	0.9999	1.0003	0.330	1.0000	0.9999	1.0001	0.448	0.038	0.968
DBP	794	1.0000	0.9996	1.0004	0.931	1.0002	1.0001	1.0004	0.045	<0.001	0.279

1.0003, $P=1.01\times10^{-4}$), and DBP (OR=1.0002, 95%CI=1.0001-1.0003,$P=8.95\times10^{-3}$) are risk factors for uterine fibroids, and a higher age at menarche (OR=0.9979, 95%CI=0.9969-0.9990, $P=1.02\times10^{-4}$) is a protective factor for uterine leiomyoma (Figure 2).

In addition, the results of meta-analysis suggested that uterine leiomyoma may also be affected by PCOS (OR=0.9974, 95% CI=0.9954-0.9994, $P=1.09\times10^{-2}$), endometriosis (OR=1.0038, 95% CI=1.0002-1.0076, $P=4.33\times10^{-2}$), and 2-hour glucose levels (OR=1.0032, 95%CI=1.0001-1.0064, $P=4.57\times10^{-2}$). In fact, PCOS and endometriosis were also significant in the FinnGen results, although they failed to pass FDR correction. Therefore, the results of analysis of the FinnGen and UKB data sets and the meta-analysis of PCOS and endometriosis were consistent. The difference between the results of the FinnGen and UKB data sets for 2-hour glucose level may be related to the different number of SNPs in the instrumental variables.

4 Discussion

Our MR study found that a genetically predicted higher age at natural menopause, SBP, DBP, endometriosis, and elevated 2-hour glucose level were risk factors for uterine leiomyoma, and a higher age at menarche and PCOS were protective factors for uterine leiomyoma.

An earlier age at menarche is thought to be associated with an increased risk of uterine leiomyoma (40, 41), and this finding is further supported by our study. Women with an earlier age at menarche had higher levels of estradiol and estrone and lower levels of sex hormone-binding globulin in their hormonal milieu than women with a later age at menarche (42, 43). Fibroids were found to have more estrogen receptors, lower estradiol metabolism, and a stronger transcriptional response to estrogen than myometrium (44). Therefore, higher estrogen and progesterone levels may increase the risk of fibroids. Animal models have also confirmed that hormonal stimulation can increase tumor proliferation and decrease apoptosis (45). There are few studies on the effect of menopausal age on uterine leiomyoma risk. Our results suggest that a later age at menopause is associated with an increased risk of uterine leiomyoma. A prospective study of female teachers also found a reduced risk of fibroids in postmenopausal women compared to premenopausal women (13). The National Institute of Environmental Health Sciences (NIEHS) Fibroid Growth Study found that the growth rate of uterine leiomyoma in white women was related to age (46). Rapid growth of uterine leiomyoma after the age of 30, especially in premenopause, is consistent with age-related changes in estrogen and progesterone (47). Therefore, the effect of menopausal age on fibroids may also be related to hormone levels. In addition, mitotic activity in the myometrium is greatest during the luteal phase of the menstrual cycle, and prolonged exposure to the menstrual cycle may increase the risk of uterine leiomyoma (48). This also suggests that earlier menstruation age and later menopause age can increase the risk of uterine leiomyoma.

Several studies have found a significant positive association between hypertension and uterine leiomyoma, but most of these studies are retrospective studies, cross-sectional studies, or prospective studies with small sample sizes and have not successfully established a causal association (49–51). Our results suggest that higher SBP and DBP are causally associated with an increased risk of uterine leiomyoma. During the onset of hypertension, angiotensin is hydrolyzed to angiotensin I, which is then converted to angiotensin II by the angiotensin converting enzyme (ACE) (52). Angiotensin II has been reported to significantly increase the number of uterine leiomyoma cells in a dose-dependent manner (53). Hsieh et al. found that mutations in angiotensin-converting enzyme activation genes were significantly associated with leiomyoma susceptibility (54). A recent study reported a 31.8% reduction in clinically diagnosed uterine leiomyoma in hypertensive adult women who had previously used angiotensin-converting enzyme inhibitors (ACEis) compared with those who had not used ACEis (55). Therefore, hypertension may cause uterine leiomyoma through production of angiotensin II. In addition, hypertension can induce fibroid proliferation and fibrogenesis by inducing smooth muscle cell injury through mechanical shear stress, which may also lead to uterine leiomyoma (56).

The relationship between diabetes mellitus and uterine leiomyoma has been controversial for many years. On the one hand, some observational studies have found a lower incidence of fibroids in diabetic patients and hypothesized that diabetes may

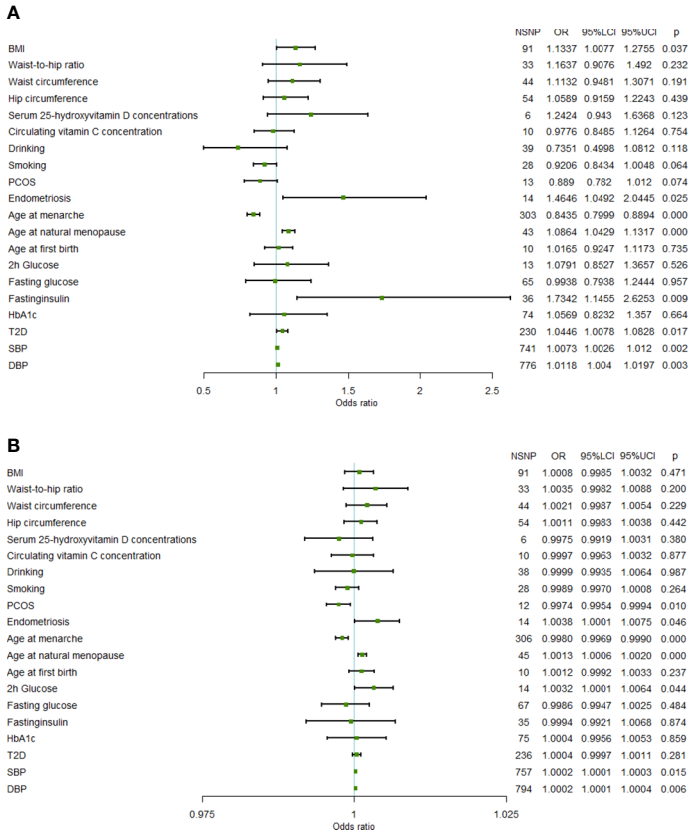


FIGURE 1
Forest plot of Mendelian randomization results. **(A)** Mendelian randomization results in the FinnGen data set. **(B)** Mendelian randomization results in the UKB data set (95%LCI, lower limit of 95% CI; 95%UCI, upper limit of 95% CI; BMI, body mass index; PCOS, polycystic ovary syndrome; 2h glucose, 2-hour glucose after oral glucose tolerance test; HbA1c, glycated hemoglobin; T2D, Type 2 diabetes; SBP, systolic blood pressure; DBP, diastolic blood pressure; NSNP, number of single nucleotide polymorphisms).

inhibit tumor development by causing vascular dysfunction (57). On the other hand, diabetes is often accompanied by obesity and hypertension, which may increase the risk of uterine leiomyoma. Since it is not possible to assess the relationship between diabetes and fibroids in untreated diabetic populations, observational estimates may capture both the effect of disease and treatment effects on fibroids. A study that indirectly examined the relationship between diabetes and treatment found that the protective effects of diabetes was only present in diabetic patients receiving the drug (15). Elevated 2-hour glucose post-challenge, an indicator of

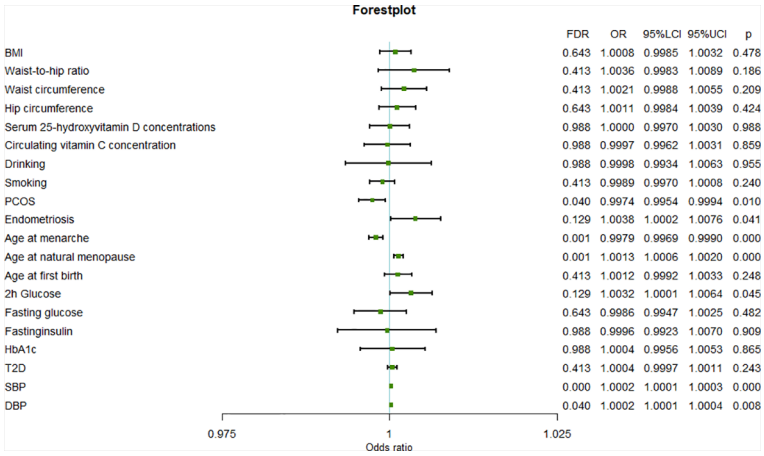


FIGURE 2
Forest plot of the results from meta-analysis.

diabetes, was found in our study to be associated with an increased risk of uterine leiomyoma. However, fasting glucose, another indicator of diabetes, was not causally associated with uterine leiomyoma in either the FinnGen or the UKB data sets. In addition, elevated fasting insulin levels and type 2 diabetes were found to be associated with an increased risk of uterine leiomyoma in the FinnGen data, but this association was not found in the UKB data. Due to the low power of the UKB results, we were unable to determine the relationship between diabetes and uterine leiomyoma. To be sure, the risk of uterine leiomyoma should be considered when 2-hour glucose post-challenge is elevated.

Endometriosis polycystic ovary syndrome (PCOS) and uterine leiomyoma are common non-cancerous gynecological diseases in women. Our study found that endometriosis was associated with an increased risk of uterine leiomyoma, while PCOS was associated with a reduced risk of uterine leiomyoma. Although there is no direct evidence that endometriosis is a influence factor of uterine leiomyomas, Uimari et al. found that 20% of patients with symptomatic fibroids had endometriosis, and 26% of patients with symptomatic endometriosis had fibroids (58). Hemmings et al. (59) also reported that patients with endometriosis were more likely to develop uterine leiomyoma than patients without endometriosis. Physiologically, endometriosis tissues have been shown to express aromatase and produce estrogen independently of the ovary (60), which is a major cause of uterine leiomyoma. Therefore, aromatase inhibitors for endometriosis may reduce the risk of uterine leiomyoma. There are few studies on the relationship between PCOS and uterine leiomyoma, and existing studies are controversial (15, 16). Our study found that PCOS was associated with a reduced risk of uterine leiomyoma. Our results are consistent with a large later study using data from PPCOS I (National Institute of Child Health and Human Development Cooperative Reproductive Medicine Network Pregnancy in Polycystic Ovary Syndrome I), PPCOS II (Pregnancy in Polycystic Ovary Syndrome II), and AMIGOS (Assessing Multiple Intrauterine Gestations from Ovarian Stimulation) and found PCOS patients had a reduced risk of uterine leiomyoma compared to unexplained infertility patients (61). The results of this large study using ultrasound diagnosis were more reliable than those of a small sample using self-reported data. PCOS patients are anovulatory and have limited exposure of myometrium to progesterone, which has been shown to stimulate leiomyoma growth through a group of key genes that regulate apoptosis and proliferation, and may be the cause of this association (62–64).

Obesity has been consistently recognized as a risk factor for uterine leiomyoma. Given the insufficient power of the UKB database and the fact that a recent Mendelian randomization study (65) using different instrumental variables found that a higher BMI slightly increased the risk of uterine fibroids in the UKB data set, we concluded that obesity is unquestionably a risk factor for uterine leiomyomas.

Several studies have reported a negative association between age at first birth and uterine leiomyoma risk (40, 41, 66), while our study found no such association. Pregnancy may lead to decreased estrogen receptor levels in myometrium (67). Postpartum reduction of collagen content and smooth muscle cytoplasm can eliminate or shrink uterine leiomyoma (68). In addition, the vascular distribution of uterine leiomyoma is different from that of the myometrium, and delivery ischemia and uterine remodeling can give priority to the elimination of uterine leiomyoma (69, 70). But the relationship between age at first birth and uterine leiomyoma may be non-linear. Donnaet al. (71) reported that the effect of age at first birth on uterine leiomyoma was not linear, and mid-reproductive (25-29 years) delivery appeared to be most protective against fibroids development. Larger fibroids are more common in women over the age of 40. If women give birth at a young age, the disease may not develop. But there may be no benefit if the first pregnancy is too late, as some tumors may have grown too large. Our study was unable to determine whether there is a non-linear relationship between age and uterine leiomyoma risk, so more research is needed.

There are some advantages to our study: (1) This was a Mendelian randomized study that could find causal associations. (2) Our study found that some previously unknown factors, such as uterine leiomyoma and menopausal age, were associated with uterine leiomyoma. In addition, the influence of previously controversial factors such as PCOS, smoking, and diabetes on uterine leiomyoma were identified. And (3), participants in all GWAS studies were of predominantly European ancestry, with less racial bias. Discovery data sets, validation sets, and meta-analysis were used to increase the reliability of the results. Our research also has some shortcomings: (1) The influence of pleiotropy in the MR design, including horizontal pleiotropic and vertical pleiotropic; we used two sensitivity analysis methods to detect pluripotency, including MR-Egger intercept and MR-PRESSO, in the hope of minimizing bias. (2) The power of the results verified in the UKB data set was lower, resulting in several factors that were found to be significant in the FinnGen data set but not in the UKB. (3) The fact that the GWAS studies were mainly Europeans may have influenced the extrapolation of the results. In addition, the non-linear relationships could not be detected in this study. And (4), the genetic instruments were variants identified through GWAS analyses with p-values < $5x10^{-8}$. As a result, the estimates of these genetic effects tend to be upwardly biased due to a phenomenon known as the “winner’s curse”.

In conclusion, our MR study confirmed that earlier menstrual age, hypertension, obesity, and elevated 2-hour glucose post-challenge were risk factors for uterine leiomyoma, and ruled out the causal relationship between smoking and uterine leiomyoma. In addition, a later age of menopause and endometriosis were found to increase the risk of uterine leiomyoma, while PCOS was found to decrease the risk.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ReproGen: (http://www.reprogen.org/); GIANT: (http://portals.broadinstitute.org/

collaboration/giant/index.php/GIANT_consortium_data_files); MAGIC: (https://magicinvestigators.org/downloads/); GLGC: (http://lipidgenetics.org/#data-downloads-title); UKB: (http://www.nealelab.is/uk-biobank); FinnGen: (https://r4.finngen.fi/).

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. All the data used were from the GWAS studies, specific ethical and consent statements for each GWAS in this study can be found in the original GWAS publications.

Author contributions

HW and YQ conceptualized and designed the study, drafted the initial manuscript, and reviewed and revised the manuscript. YQ and LC collected the data and carried out the initial analyses. SG and YL reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1133260/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Main design of this study.

References

1. Millien C, Manzi A, Katz AM, Gilbert H, Smith Fawzi MC, Farmer PE, et al. Assessing burden, risk factors, and perceived impact of uterine fibroids on women's lives in rural Haiti: implications for advancing a health equity agenda, a mixed methods study. *Int J Equity Health* (2021) 20:1. doi: 10.1186/s12939-020-01327-9

2. Bulun SE. Uterine fibroids. *N Engl J Med* (2013) 369:1344–55. doi: 10.1056/NEJMra1209993

3. Johnson G, MacLehose RF, Baird DD, Laughlin-Tommaso SK, Hartmann KE. Uterine leiomyomata and fecundability in the right from the start study. *Hum Reprod* (2012) 27:2991–7. doi: 10.1093/humrep/des263

4. Whiteman MK, Kuklina E, Jamieson DJ, Hillis SD, Marchbanks PA. Inpatient hospitalization for gynecologic disorders in the united states. *Am J Obstet Gynecol* (2010) 202:541. doi: 10.1016/j.ajog.2009.12.013

5. Pavone D, Clemenza S, Sorbi F, Fambrini M, Petraglia F. Epidemiology and risk factors of uterine fibroids. *Best Pract Res Clin Obstet Gynaecol* (2018) 46:3–11. doi: 10.1016/j.bpobgyn.2017.09.004

6. Kho PF, Mortlock S, Rogers PAW, Nyholt DR, Montgomery GW, Spurdle AB, et al. Genetic analyses of gynecological disease identify genetic relationships between uterine fibroids and endometrial cancer, and a novel endometrial cancer genetic risk region at the WNT4 1p36. *12 locus. Hum Genet* (2021) 140:1353–65. doi: 10.1007/s00439-021-02312-0

7. Stewart EA, Cookson CL, Gandolfo RA, Schulze-Rath R. Epidemiology of uterine fibroids: a systematic review. *Bjog* (2017) 124:1501–12. doi: 10.1111/1471-0528.14640

8. Stewart EA, Laughlin-Tommaso SK, Catherino WH, Lalitkumar S, Gupta D, Vollenhoven B. Uterine fibroids. *Nat Rev Dis Primers* (2016) 2:16043. doi: 10.1038/nrdp.2016.43

9. Anneveldt KJ, van 't Oever HJ, Nijholt IM, Dijkstra JR, Hehenkamp WJ, Veersema S, et al. Systematic review of reproductive outcomes after high intensity focused ultrasound treatment of uterine fibroids. *Eur J Radiol* (2021) 141:109801. doi: 10.1016/j.ejrad.2021.109801

10. Donnez J, Dolmans MM. Uterine fibroid management: from the present to the future. *Hum Reprod Update* (2016) 22:665–86. doi: 10.1093/humupd/dmw023

11. Flake GP, Andersen J, Dixon D. Etiology and pathogenesis of uterine leiomyomas: a review. *Environ Health Perspect* (2003) 111:1037–54. doi: 10.1289/ehp.5787

12. Parazzini F, Negri E, La Vecchia C, Rabaiotti M, Luchini L, Villa A, et al. Uterine myomas and smoking. *Results an Ital study. J Reprod Med* (1996) 41:316–20.

13. Templeman C, Marshall SF, Clarke CA, DeLellis Henderson K, Largent J, Neuhausen S, et al. Risk factors for surgically removed fibroids in a large cohort of teachers. *Fertil Steril* (2009) 92:1436–46. doi: 10.1016/j.fertnstert.2008.08.074

14. Dragomir AD, Schroeder JC, Connolly A, Kupper LL, Hill MC, Olshan AF, et al. Potential risk factors associated with subtypes of uterine leiomyomata. *Reprod Sci* (2010) 17:1029–35. doi: 10.1177/1933719110376979

15. Wise LA, Palmer JR, Stewart EA, Rosenberg L. Polycystic ovary syndrome and risk of uterine leiomyomata. *Fertil Steril* (2007) 87:1108–15. doi: 10.1016/j.fertnstert.2006.11.012

16. Abdel-Gadir A, Oyawoye OO, Chander BP. Coexistence of polycystic ovaries and uterine fibroids and their combined effect on the uterine artery blood flow in relation to age and parity. *J Reprod Med* (2009) 54:347–52.

17. Baird DD, Travlos G, Wilson R, Dunson DB, Hill MC, D'Aloisio AA, et al. Uterine leiomyomata in relation to insulin-like growth factor-I, insulin, and diabetes. *Epidemiology* (2009) 20:604–10. doi: 10.1097/EDE.0b013e31819d8d3f

18. Poretsky L, Kalin MF. The gonadotropic function of insulin. *Endocr Rev* (1987) 8:132–41. doi: 10.1210/edrv-8-2-132

19. Okolo S. Incidence, aetiology and epidemiology of uterine fibroids. *Best Pract Res Clin Obstet Gynaecol* (2008) 22:571–88. doi: 10.1016/j.bpobgyn.2008.04.002

20. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *Jama* (2017) 318:1925–6. doi: 10.1001/jama.2017.17219

21. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* (2015) 518:197–206. doi: 10.1038/nature14177

22. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* (2015) 518:187–96. doi: 10.1038/nature14132

23. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* (2018) 50:1412–25. doi: 10.1038/s41588-018-0205-x

24. Jiang X, O'Reilly PF, Aschard H, Hsu YH, Richards JB, Dupuis J, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin d levels. *Nat Commun* (2018) 9:260. doi: 10.1038/s41467-017-02662-2

25. Zheng JS, Luan J, Sofianopoulou E, Imamura F, Stewart ID, Day FR, et al. And type 2 diabetes: genome-wide association study and mendelian randomization analysis in European populations. *Diabetes Care* (2021) 44:98–106. doi: 10.2337/dc20-1328

26. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* (2019) 51:237–44. doi: 10.1038/s41588-018-0307-5

27. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* (2017) 49:834–41. doi: 10.1038/ng.3841

28. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-Scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* (2015) 47:1294–303. doi: 10.1038/ng.3412

29. Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet* (2016) 48:1462–72. doi: 10.1038/ng.3698

30. Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-Scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PloS Genet* (2018) 14:e1007813. doi: 10.1371/journal.pgen.1007813

31. Sapkota Y, Steinthorsdottir V, Morris AP, Fassbender A, Rahmioglu N, De Vivo I, et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat Commun* (2017) 8:15539. doi: 10.1038/ncomms15539

32. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, 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. doi: 10.1038/s41588-018-0241-6

33. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* (2021) 53:840–60. doi: 10.1038/s41588-021-00852-9

34. Chen L, Yang H, Li H, He C, Yang L, Lv G. Insights into modifiable risk factors of cholelithiasis: a mendelian randomization study. *Hepatology* (2022) 75:785–96. doi: 10.1002/hep.32183

35. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife* (2018) 7:e34408. doi: 10.7554/eLife.34408

36. 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. doi: 10.1093/ije/dyv080

37. 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. doi: 10.1002/gepi.21965

38. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* (2018) 50:693–8. doi: 10.1038/s41588-018-0099-7

39. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in mendelian randomization studies. *Int J Epidemiol* (2013) 42:1497–501. doi: 10.1093/ije/dyt179

40. Wise LA, Palmer JR, Harlow BL, Spiegelman D, Stewart EA, Adams-Campbell LL, et al. Reproductive factors, hormonal contraception, and risk of uterine leiomyomata in African-American women: a prospective study. *Am J Epidemiol* (2004) 159:113–23. doi: 10.1093/aje/kwh016

41. Terry KL, De Vivo I, Hankinson SE, Missmer SA. Reproductive characteristics and risk of uterine leiomyomata. *Fertil Steril* (2010) 94:2703–7. doi: 10.1016/j.fertnstert.2010.04.065

42. Abetew DF, Enquobahrie DA, Dishi M, Rudra CB, Miller RS, Williams MA. Age at menarche, menstrual characteristics, and risk of preeclampsia. *ISRN Obstet Gynecol* (2011) 2011:472083. doi: 10.5402/2011/472083

43. Emaus A, Espetvedt S, Veierød MB, Ballard-Barbash R, Furberg AS, Ellison PT, et al. 17-beta-estradiol in relation to age at menarche and adult obesity in premenopausal women. *Hum Reprod* (2008) 23:919–27. doi: 10.1093/humrep/dem432

44. Velez Edwards DR, Baird DD, Hartmann KE. Association of age at menarche with increasing number of fibroids in a cohort of women who underwent standardized ultrasound assessment. *Am J Epidemiol* (2013) 178:426–33. doi: 10.1093/aje/kws585

45. Hunter DS, Hodges LC, Eagon PK, Vonier PM, Fuchs-Young R, Bergerson JS, et al. Influence of exogenous estrogen receptor ligands on uterine leiomyoma: evidence from an *in vitro*/*in vivo* animal model for uterine fibroids. *Environ Health Perspect* (2000) 108 Suppl 5:829–34. doi: 10.1289/ehp.00108s5829

46. Peddada SD, Laughlin SK, Miner K, Guyon JP, Haneke K, Vahdat HL, et al. Growth of uterine leiomyomata among premenopausal black and white women. *Proc Natl Acad Sci USA* (2008) 105:19887–92. doi: 10.1073/pnas.0808188105

47. Baird DD, Kesner JS, Dunson DB. Luteinizing hormone in premenopausal women may stimulate uterine leiomyomata development. *J Soc Gynecol Investig* (2006) 13:130–5. doi: 10.1016/j.jsgi.2005.12.001

48. Wise LA, Laughlin-Tommaso SK. Epidemiology of uterine fibroids: from menarche to menopause. *Clin Obstet Gynecol* (2016) 59:2–24. doi: 10.1097/GRF.0000000000000164

49. Silver MA, Raghuvir R, Fedirko B, Elser D. Systemic hypertension among women with uterine leiomyomata: potential final common pathways of target end-organ remodeling. *J Clin Hypertens (Greenwich)* (2005) 7:664–8. doi: 10.1111/j.1524-6175.2005.04384.x

50. Takeda T, Sakata M, Isobe A, Miyake A, Nishimoto F, Ota Y, et al. Relationship between metabolic syndrome and uterine leiomyomas: a case-control study. *Gynecol Obstet Invest* (2008) 66:14–7. doi: 10.1159/000114250

51. Haan YC, Diemer FS, van der Woude L, Van Montfrans GA, Oehlers GP, Brewster LM. The risk of hypertension and cardiovascular disease in women with uterine fibroids. *J Clin Hypertens (Greenwich)* (2018) 20:718–26. doi: 10.1111/jch.13253

52. Narkiewicz K. Diagnosis and management of hypertension in obesity. *Obes Rev* (2006) 7:155–62. doi: 10.1111/j.1467-789X.2006.00226.x

53. Isobe A, Takeda T, Sakata M, Miyake A, Yamamoto T, Minekawa R, et al. Dual repressive effect of angiotensin II-type 1 receptor blocker telmisartan on angiotensin II-induced and estradiol-induced uterine leiomyoma cell proliferation. *Hum Reprod* (2008) 23:440–6. doi: 10.1093/humrep/dem247

54. Hsieh YY, Lee CC, Chang CC, Wang YK, Yeh LS, Lin CS. Angiotensin I-converting enzyme insertion-related genotypes and allele are associated with higher susceptibility of endometriosis and leiomyoma. *Mol Reprod Dev* (2007) 74:808–14. doi: 10.1002/mrd.20474

55. Fischer NM, Nieuwenhuis TO, Singh B, Yenokyan G, Segars JH. Angiotensin-converting enzyme inhibitors reduce uterine fibroid incidence in hypertensive women. *J Clin Endocrinol Metab* (2021) 106:e650–9. doi: 10.1210/clinem/dgaa718

56. Humphrey JD. Mechanisms of arterial remodeling in hypertension: coupled roles of wall shear and intramural stress. *Hypertension* (2008) 52:195–200. doi: 10.1161/HYPERTENSIONAHA.107.103440

57. Walocha JA, Litwin JA, Miodoński AJ. Vascular system of intramural leiomyomata revealed by corrosion casting and scanning electron microscopy. *Hum Reprod* (2003) 18:1088–93. doi: 10.1093/humrep/deg213

58. Uimari O, Järvelä I, Ryynänen M. Do symptomatic endometriosis and uterine fibroids appear together? *J Hum Reprod Sci* (2011) 4:34–8. doi: 10.4103/0974-1208.82358

59. Hemmings R, Rivard M, Olive DL, Poliquin-Fleury J, Gagné D, Hugo P, et al. Evaluation of risk factors associated with endometriosis. *Fertil Steril* (2004) 81:1513–21. doi: 10.1016/j.fertnstert.2003.10.038

60. Bulun SE, Lin Z, Imir G, Amin S, Demura M, Yilmaz B, et al. Regulation of aromatase expression in estrogen-responsive breast and uterine disease: from bench to treatment. *Pharmacol Rev* (2005) 57:359–83. doi: 10.1124/pr.57.3.6

61. Huang H, Kuang H, Sun F, Diamond MP, Legro RS, Coutifaris C, et al. Lower prevalence of non-cavity-distorting uterine fibroids in patients with polycystic ovary syndrome than in those with unexplained infertility. *Fertil Steril* (2019) 111:1011–1019.e1. doi: 10.1016/j.fertnstert.2019.01.020

62. Moravek MB, Yin P, Ono M, J.S.t. Coon, Dyson MT, Navarro A, et al. Ovarian steroids, stem cells and uterine leiomyoma: therapeutic implications. *Hum Reprod Update* (2015) 21:1–12. doi: 10.1093/humupd/dmu048

63. Blake RE. Leiomyomata uteri: hormonal and molecular determinants of growth. *J Natl Med Assoc* (2007) 99:1170–84.

64. Rein MS, Barbieri RL, Friedman AJ. Progesterone: a critical role in the pathogenesis of uterine myomas. *Am J Obstet Gynecol* (1995) 172:14–8. doi: 10.1016/0002-9378(95)90077-2

65. Venkatesh SS, Ferreira T, Benonisdottir S, Rahmioglu N, Becker CM, Granne I, et al. Obesity and risk of female reproductive conditions: a mendelian randomisation study. *PloS Med* (2022) 19:e1003679. doi: 10.1371/journal.pmed.1003679

66. Marshall LM, Spiegelman D, Goldman MB, Manson JE, Colditz GA, Barbieri RL, et al. A prospective study of reproductive factors and oral contraceptive use in relation to the risk of uterine leiomyomata. *Fertil Steril* (1998) 70:432–9. doi: 10.1016/S0015-0282(98)00208-8

67. Kawaguchi K, Fujii S, Konishi I, Iwai T, Nanbu Y, Nonogaki H, et al. Immunohistochemical analysis of oestrogen receptors, progesterone receptors and ki-67 in leiomyoma and myometrium during the menstrual cycle and pregnancy. *Virchows Arch A Pathol Anat Histopathol* (1991) 419:309–15. doi: 10.1007/BF01606522

68. Schwartz SM, Marshall LM, Baird DD. Epidemiologic contributions to understanding the etiology of uterine leiomyomata. *Environ Health Perspect* (2000) 108 Suppl 5:821–7. doi: 10.1289/ehp.00108s5821

69. Burbank F. Childbirth and myoma treatment by uterine artery occlusion: do they share a common biology? *J Am Assoc Gynecol Laparosc* (2004) 11:138–52. doi: 10.1016/S1074-3804(05)60189-2

70. Laughlin SK, Schroeder JC, Baird DD. New directions in the epidemiology of uterine fibroids. *Semin Reprod Med* (2010) 28:204–17. doi: 10.1055/s-0030-1251477

71. Baird DD, Dunson DB. Why is parity protective for uterine fibroids? *Epidemiology* (2003) 14:247–50. doi: 10.1097/01.EDE.0000054360.61254.27

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 04 September 2023
DOI 10.3389/fendo.2023.1200183

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Shahid Banday,
University of Massachusetts Medical School, United States
Jelena Djordjevic,
University of Belgrade, Serbia

*CORRESPONDENCE
Mohamed Ahdi
m.ahdi@amsterdamumc.nl

†These authors share first authorship

RECEIVED 04 April 2023
ACCEPTED 30 June 2023
PUBLISHED 04 September 2023

CITATION
Ahdi M, Gerards MC, Smits PHM, Meesters EW, Brandjes DPM, Nieuwdorp M and Gerdes VEA (2023) Genetic glucocorticoid receptor variants differ between ethnic groups but do not explain variation in age of diabetes onset, metabolic and inflammation parameters in patients with type 2 diabetes.
Front. Endocrinol. 14:1200183.
doi: 10.3389/fendo.2023.1200183

Genetic glucocorticoid receptor variants differ between ethnic groups but do not explain variation in age of diabetes onset, metabolic and inflammation parameters in patients with type 2 diabetes

Mohamed Ahdi[1*†], Maaike C. Gerards[1†], Paul H.M. Smits[2], Eelco W. Meesters[3], Dees P. M. Brandjes[1], Max Nieuwdorp[1] and Victor E. A. Gerdes[1,3]

[1]Department of Vascular Medicine, Amsterdam University Medical Centers (UMCs), Amsterdam, Netherlands, [2]Department of Molecular Biology, Atalmedial, Amsterdam, Netherlands, [3]Department of Internal Medicine, Spaarne Hospital, Hoofddorp, Netherlands

Aims: The effect of excess glucocorticoid receptor (GR) stimulation through glucocorticoid medication or cortisol on glucose metabolism is well established. There are genetic GR variants that result in increased or decreased GR stimulation. We aimed to determine the prevalence of genetic GR variants in different ethnic groups in a cohort of patients with type 2 diabetes, and we aimed to determine their association with age of diabetes onset and metabolic and inflammation parameters.

Methods: A cross-sectional analysis was performed in a multiethnic cohort (n = 602) of patients with established type 2 diabetes. Polymorphisms in the GR gene that have previously been associated with altered glucocorticoid sensitivity (*TthIII*I, ER22/23EK N363S, *Bcl*I and 9β) were determined and combined into 6 haplotypes. Associations with age of diabetes onset, HbA1c, hs-CRP and lipid values were evaluated in multivariate regression models.

Results: The prevalence of the SNPs of N363S and *Bcl*I was higher in Dutch than in non-Dutch patients. We observed a lower prevalence of the SNP 9β in Dutch, South(East) Asian and Black African patients versus Turkish and Moroccan patients. We did not detect an association between SNPs and diabetes age of onset or metabolic parameters. We only found a trend for lower age of onset and higher HbA1c in patients with 1 or 2 copies of haplotype 3 (*TthIII*I + 9β).

10.3389/fendo.2023.1200183

Conclusions: The prevalence of genetic GR variants differs between patients of different ethnic origins. We did not find a clear association between genetic GR variants and age of diabetes onset or metabolic and inflammation parameters. This indicates that the clinical relevance of GR variants in patients with established type 2 diabetes is limited.

KEYWORDS

glucocorticoid receptor, diabetes, glucocorticoid medication, cortisol, glucose metabolism, ethnicity, inflammation (markers)

Introduction

The onset and course of type 2 diabetes mellitus is determined by a combination of environmental and genetic risk factors. The effect of excess glucocorticoid receptor (GR) stimulation through glucocorticoid medication or cortisol on the incidence of type 2 diabetes is well established (1, 2). It is unknown whether genetic GR variants that are associated with increased GR stimulation also contribute to a diabetogenic phenotype. Ethnic origin is one of the factors associated with the incidence and course of type 2 diabetes and metabolic syndrome (3). Ethnicity as a determinant for disease consists of shared origin and genetics but also shared social and environmental background (4). If the prevalence of genetic GR variants differs in populations from different ethnicities and geographical regions, differences in functioning of this receptor could partly explain differences in onset and outcome of type 2 diabetes among ethnic groups.

The GR is expressed in almost every cell in the body (5). Binding of the GR by cortisol or glucocorticoid medication results in transrepression and transactivation of certain genes. Transrepression contributes to suppression of inflammation, and transactivation contributes to regulation of energy metabolism. Excess transactivation has effects that are comparable to metabolic derangements in type 2 diabetes (6). In clinical practice, we frequently encounter the effects of excess transactivation due to supraphysiological GR stimulation. Examples are acute disturbance of glucose metabolism due to high-dose glucocorticoid therapy and increased incidence of type 2 diabetes in Cushing syndrome (1, 2). Glucocorticoid signalling can also affect lipid metabolism, resulting in higher levels of triglycerides and total cholesterol (7).

The gene that encodes the glucocorticoid receptor (NR3C1) consists of 157,582 base pairs and is located on chromosome 5 (8). Single nucleotide polymorphisms (SNPs) can induce changes in the configuration and sensitivity of the GR, which may impact the binding and regulation of gene expression with glucocorticoids, and may subsequently affect inflammatory suppression and glucose metabolism (9). There are functional GR variants (SNPs) that can potentially change the transactivation and/or transrepression capacity of the GR gene. A schematic overview of the GR gene including the locations of these SNPs within the gene has been published before (10, 11).

GR variants B*cl*I (rs41423247) and N363S (rs6195) are associated with increased transactivation (sensitivity to glucocorticoids), whereas an SNP at ER22/23EK (rs6189) is associated with diminished transactivation. On the other hand, the GR variant 9β (rs6198) is associated with lower transrepression (10, 12). A fifth SNP, *TthIII*I (rs10052957) does not affect glucocorticoid sensitivity on itself but can result in glucocorticoid resistance in the presence of ER22/23EK (10). Form a clinical perspective, it has been observed that both BclI and N363S variations are linked to abdominal obesity, although there have been conflicting findings regarding N363S. Furthermore, N363S has been associated with higher levels of LDL-cholesterol and an increased risk of cardiovascular disease, while the ER22/23EK polymorphism has been associated with a reduced risk of dementia but an increased risk of major depression. Additionally, the 9β variant has been linked to increased inflammatory markers, rheumatoid arthritis, post-traumatic stress disorder, and cardiovascular disease (10, 13).The prevalence of the SNPs varies in previous studies, with a minor allele frequency (MAF) of 1.9-3% for ER22/23EK and 29.6-38.6% for B*cl*I (11, 14–17).

In this study, we aim to determine the association between genetic GR variants and the incidence and course of type 2 diabetes and metabolic syndrome in patients with established type 2 diabetes from different ethnic groups. We hypothesize that GR SNPs resulting in increased transactivation are associated with a lower age of diabetes onset and impaired glycemic control, and SNPs resulting in diminished transrepression are associated with a higher level of inflammation in patients with established type 2 diabetes. The second aim is to evaluate whether the prevalence of genetic GR variants differs between ethnic groups.

Methods

We performed a cross-sectional analysis in a multi-ethnic cohort of patients with type 2 diabetes who were treated in

secondary care (18). The participants consisted of consecutive individuals who visited the outpatient clinic of MC Slotervaart in Amsterdam for their annual comprehensive diabetes assessment between May 2009 and December 2010. Only those who provided written informed consent, and for whom DNA material was available for analysis were included in this study. For all participants, data on ethnic origin, diabetes onset, glucose- and lipid-lowering treatment, established complications, vital, anthropometric and laboratory parameters were registered. The study protocol was approved by the institutional review board.

The diagnosis of type 2 diabetes was based on the general practitioner (GP) referral letter in combination with clinical and biochemical characteristics determined at our clinic. GAD antibodies and C-peptide were determined in case of doubt regarding the type of diabetes. Age of diabetes onset was retrieved from the GP referral letter and checked with the patient. In case of discrepancy between referral letter and patient history, the age of onset as told by the patient was considered true. Ethnicity was determined according to the country of birth of either the patient or his or her parent and by last name analysis (4). The following ethnic groups were considered: native Dutch, Turkish, Moroccan, Southeast Asians (comprising 57% Hindustani and 25% Indonesians), and Black Africans (with 78% being Surinamese Creoles). Additional information can be found in the caption of Table S1.

Laboratory assays

Blood samples were obtained by standard phlebotomy after a 10-hour overnight fast.

Depending on the patients' informed consent form, an additional 10 ml EDTA-anticoagulated whole blood sample was collected. Following immediate centrifugation (15 minutes, 3000rpm, 1860g at 15°C), the isolated "buffy-coat" was carefully separated using a Pasteur pipette tube and stored in 0.5 ml vials at -70°C until assayed.

Total genomic DNA was isolated from the frozen 'buffy-coat', using the total nucleic acid (TNA) protocol on the MagNAPure LC (Roche Diagnostics). PCR primers (forwards and reverse) as well as MGB probes were designed using Primer Express Software v3.0.1 of Life Technologies. Five NR3C1 SNPs were determined by real-time polymerase chain reaction (RT-PCR): *TthIII*I (rs10052957: guanine > adenine), ER22/23EK (rs6189: guanine > adenine and rs6190: guanine > adenine), N363S (rs6195: adenine > guanine), B*cl*I (rs41423247: cytosine > guanine) and 9β (rs6198: adenine > guanine) followed by the allelic discrimination protocol on an ABI 7500 real-time PCR thermocycler (Thermo Fisher) as described previously (19). These SNPs combine into 6 haplotypes, as previously shown (11). For each haplotype, 3 genotype combinations were distinguished as carrying 0, 1, or 2 copies of the haplotype allele. To show that the designed assays were able to detect the indicated SNPs, all assays were validated before using well-characterized DNA kindly provided by P. Noordijk (Leiden University Medical Centre) from each individual SNP.

The routine analysis of these samples for HbA1c was performed using a Menarini (Adams™ HA-8160, Arkray Inc, Kyoto, Japan) automated HPLC analyser. Serum total- and HDL-cholesterol and triglycerides were determined using standard laboratory procedures within 4 hours after sampling with an automated analyser (Synchron® LX20, Beckman Coulter Inc, Fullerton CA, USA). LDL-cholesterol was calculated using the Friedewald formula (20). High-sensitivity C-reactive protein (hs-CRP) was determined with a near infrared particle immunoassay rate methodology (Beckman Brea, CA).

Statistical analysis

We estimated beta-coefficients for the change in age of diabetes onset, glycemic control, inflammation and lipid parameters for each SNP in linear regression models. Age, sex, diabetes duration, BMI, glucose and lipid-lowering medication and ethnicity were assessed as confounders if applicable. Potential confounders were selected if we presumed a theoretical relationship with SNP status and outcome, in combination with a statistical association (21). Outcome variables that had a non-normal distribution were transformed to approximate normality. To ease interpretation, we presented back-transformed values of those variables in the outcome tables.

We performed an *a priori* power estimation on the difference in age of diabetes onset in the absence or presence of different polymorphisms. The power generally increases as a SNP is more prevalent (and as the effect on glucocorticoid sensitivity is stronger) (22). For the least prevalent SNP - ER22/23EK (94% wild type) - univariate regression analysis with a 5% significance level will have 85% power to detect the difference of 4 years (standard deviation ± 9) in age of onset between patients with a glucocorticoid-resistant genotype and patients with a glucocorticoid-sensitive genotype when the total sample size is 624 patients.

Results

Patients and genotyping

From a total of 983 patients with type 2 diabetes, 602 patients had available DNA samples and were included. There were no significant differences in demographics, clinical variables, and complications between patients with and patients without available DNA data (Table S1). Overall, patients had a reasonably well-regulated diabetes with an average HbA1c level of 7.3% (54 mmol/mol) and the average diabetes duration was 11.9 ± 8.5 years. Fourty-four percent of the participants were of non-native Dutch origin, comprising individuals from Turkish (n = 45), Moroccan (n = 101), Southeast Asian (n = 79), and black African (n = 40)

10.3389/fendo.2023.1200183

backgrounds. Non-Dutch patients were as compared with Dutch patients, less frequently males (46 versus 59%, p < 0.001), were younger (mean 57.4 versus 65.8 years), their average age of diabetes onset was 8.9 years earlier, and they had a poorer level of glycemic control (mean 7.6 versus 7.0% [60 vs. 53 mmol/mol]. The characteristics of the study population are shown in Table S1.

SNPs of the glucocorticoid receptor gene

In 557 patients (93%), at least 1 SNP could be determined. As shown in Table 1, the minor allele frequency varied from 1.6% (ER22/23EK) to 31.6% (*Bcl*I). The prevalence of SNPs was not associated with sex or age. We found a higher prevalence of the N363S SNP and a lower prevalence of the *Bcl*I CC genotype in Dutch than in non-Dutch patients (p < 0.01). Additionally, we observed a difference in the prevalence of 9β in Dutch, Southeast Asian and Black African patients versus Turkish and Moroccan patients, but this difference was not significant (p = 0.094).

Association SNPs of the GR gene with age of diabetes onset and parameters of metabolic syndrome

In the overall study population, diabetes was diagnosed at the age of 50.4 years. We could not detect a clear influence of the SNPs of the GR gene and age of onset, except for patients who were heterozygous for the 9β SNP (50.9 versus 49.2, p adj 0.02). Patients with the 9β SNP showed a trend toward higher HbA1c and CRP (Table 2). Patients with at least 1 copy of the N363S polymorphism had a lower LDL cholesterol. We did not observe any effect of the *TthIII*I, ER22/23EK and *Bcl*I polymorphisms on glycemic control, inflammation or lipid parameters.

Association of haplotypes of the GR gene with age of diabetes onset and parameters of metabolic syndrome

Haplotype 1 (which does not contain any SNP, wild type) had a minor allele frequency of 47.9%. The minor allele frequency of other haplotypes varied from 1.6% (haplotype 6) to 20.1% (haplotype 2). Patients who had 1 or 2 copies of haplotype 3 showed a trend toward a lower age of diabetes onset and a higher HbA1c, which is in accordance with the results of the individual haplotypes (Table 3). Patients who had at least 1 copy of haplotype 5 showed a trend toward lower LDL cholesterol. No associations were found for the other haplotypes.

Discussion

We studied the association between genetic variants of the GR and metabolic and inflammation parameters in a multiethnic cohort of patients with established type 2 diabetes in secondary care. We observed a different prevalence of genetic variants between patients of different ethnic origins. We did not find a clear association between genetic variants and age of diabetes onset, glycemic control, lipid parameters or inflammation, and we found only a trend for lower age of onset for patients with haplotype 3. This suggests that the clinical relevance of these genetic variants for the onset of diabetes and the course of established diabetes seems to be minor.

TABLE 1 Prevalence of SNPs of the glucocorticoid receptor gene by ethnic origin.

Genotype		Dutch *n (%)*	Turkish *n (%)*	Moroccan *n (%)*	SE Asian *n (%)*	Black African *n (%)*	P*
TTH111I (rs10052957)	CC	162 (49.7)	24 (54.5)	41 (42.3)	52 (67.5)	18 (46.2)	
	CT	139 (42.6)	18 (40.9)	46 (47.4)	24 (31.2)	19 (48.7)	
	TT	25 (7.7)	2 (4.5)	10 (10.3)	1 (1.3)	2 (5.1)	0.211
ER22/23EK (rs6189/rs6190)	GG/GG	319 (95.8)	43 (95.6)	98 (100)	77 (98.7)	38 (95)	
	GA/GA	14 (4.2)	2 (4.4)	0 (0)	1 (1.3)	2 (5)	0.211
N363S (rs6195)	AA	298 (90.3)	44 (97.8)	96 (97.0)	76 (97.4)	39 (100)	
	AG	30 (9.1)	1 (2.2)	3 (3.0)	2 (2.6)	0 (0)	
	GG	2 (0.6)	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
BCLI (rs41423247)	CC	123 (38.9)	23 (57.5)	46 (48.9)	42 (54.5)	25 (67.6)	
	CG	163 (51.6)	16 (40)	35 (37.2)	31 (40.3)	9 (24.3)	
	GG	30 (9.5)	1 (2.5)	13 (13.8)	4 (5.2)	3 (8.1)	0.002
9β (rs6198)	AA	232 (71.4)	25 (56.8)	56 (58.9)	64 (83.1)	34 (87.2)	
	AG	93 (28.6)	19 (43.2)	39 (41.1)	13 (16.9)	5 (12.8)	0.094

*Statistical significance of differences between ethnic groups is tested through a chi square test for trend.

10.3389/fendo.2023.1200183

TABLE 2 Association of SNPs of the glucocorticoid receptor gene with clinical characteristics.

Genotype		N	MAF *(%)*	Age DM onset *(years)*	HbA1c *(%)*	hs-CRP *(mmol/l)*	Total chol *(mmol/l)*	Triglyc *(mmol/l)*	LDL chol *(mmol/l)*
TTH111I (rs10052957)	**CC**	297		50.5 (11.4)	7.2 (1.2)	4.4 (6.6)	4.2 (1.0)	1.7 (1.0)	2.3 (0.8)
	T vs. CC	286	28.0	50.3 (11.6)	7.4 (1.3)	4.6 (6.8)	4.3 (1.0)	1.9 (2.3)	2.4 (0.8)
adjusted	Beta			-0.52 (0.91)	0.15 (0.10)	0.08 (0.55)	0.03 (0.08)	0.16 (0.15)	-0.02 (0.06)
	P			0.57	0.12	0.88	0.68	0.26	0.75
ER22/23EK (rs6189/rs6190)	**GG/GG**	575		50.5 (11.5)	7.3 (1.3)	4.5 (6.7)	4.2 (1.0)	1.8 (1.8)	2.4 (0.8)
	A vs GG/GG	19	1.6	51.4 (10.7)	7.2 (0.6)	3.6 (3.2)	4.0 (0.8)	1.6 (0.9)	2.2 (0.8)
adjusted	Beta			-0.12 (2.54)	-0.25 (0.27)	-0.62 (1.53)	-0.22 (0.23)	-0.24 (0.41)	-0.21 (0.18)
	P			0.96	0.36	0.69	0.33	0.57	0.24
N363S (rs6195)	**AA**	553		50.3 (11.5)	7.3 (1.3)	4.5 (6.7)	4.2 (1.0)	1.7 (1.0)	2.4 (0.8)
	G vs AA	38	3.4	52.7 (11.3)	7.2 (1.0)	3.9 (6.8)	4.2 (0.6)	1.9 (1.0)	2.1 (0.5)
adjusted	Beta			0.29 (1.86)	-0.01 (0.20)	-0.81 (1.11)	-0.11 (0.16)	0.17 (0.17)	-0.24 (0.13)
	P			0.87	0.95	0.46	0.51	0.33	0.06
BCLI (rs41423247)	**CC**	259		50.0 (11.3)	7.3 (1.2)	4.1 (5.7)	4.2 (0.9)	1.7 (0.9)	2.4 (0.8)
	G vs. CC	305	31.6	50.8 (11.7)	7.2 (1.3)	4.7 (7.1)	4.2 (1.1)	1.9 (2.3)	2.3 (0.8)
adjusted	Beta			-0.51 (0.94)	0.03 (0.10)	0.55 (0.54)	-0.05 (0.08)	0.18 (0.15)	-0.09 (0.06)
	P			0.59	0.75	0.31	0.54	0.23	0.18
9β (rs6198)	**AA**	411		50.9 (11.4)	7.2 (1.2)	4.4 (6.5)	4.2 (1.0)	1.8 (2.0)	2.3 (0.8)
	G vs AA	169	14.6	49.2 (11.4)	7.4 (1.4)	4.9 (7.2)	4.3 (0.9)	1.8 (1.0)	2.4 (0.8)
adjusted	Beta			-2.25 (0.99	0.16 (0.11)	0.41(0.61)	0.01 (0.09)	0.02 (0.16)	-0.03 (0.07)
	P			0.02	0.14	0.50	0.90	0.90	0.64

Data are presented as mean (sd). Adjustments in the multivariate linear regression model: Age of onset was adjusted for sex and ethnicity; HbA1c was adjusted for sex, ethnicity, diabetes duration, insulin use and metformin use; hsCRP was adjusted for age and sex; lipid spectrum was adjusted for sex, age, use of lipid lowering medication and metformin.

The development of type 2 diabetes is a combination of genetic and environmental risk factors. Whereas mutations underlying monogenic diabetes have direct clinical consequences, genetic variants in multifactorial forms of diabetes have a much weaker association (23). In patients with established diabetes, such as in our study population, HbA1c and lipid parameters are affected by medication and BMI. Despite adjusting for these confounding factors, we did not find an association. Additionally, for the time of diabetes onset – a parameter that is unbiased by glucose-lowering treatment - we did not find an association with genetic variants of the glucocorticoid receptor.

The SNPs N363S and ER22/23EK, which were previously associated with increased and decreased transactivation, respectively, did not affect the age of diabetes onset. Interestingly, the N363S SNP, which we hypothesized to result in diabetes onset at a younger age, showed a trend towards later diabetes onset. Despite the increased prevalence of N363S in Dutch patients compared to patients of Turkish and Moroccan origin, Dutch patients were diagnosed with diabetes at a later age. In patients with at least one copy of SNP 9β, we observed a trend for a higher level of hs-CRP, which is in line with our hypothesis.

Although specific effects on transrepression and transactivation have been established *in vitro* for all analysed SNPs, clinical studies have shown contradictory results. For example, the ER22/23EK SNP reduced GC-induced transactivation *in vitro*, and supportive evidence was found by increased insulin sensitivity and lower fasting insulin concentration in a Dutch cohort (24). However, ER22/23EK was associated with higher HbA1c levels in a cohort of patients older than 85 years old (17). Minor allele frequency was not different between these cohorts, arguing against an age difference as an explanation for the contradictory findings. Glucose metabolism is a highly regulated process in which multiple genetic and environmental factors are intertwined with an eventual effect of glucocorticoid sensitivity (25). The absence of an association in our study suggests that there is no clinically relevant effect of GR variants on glucose metabolism and that the previous contradictory findings may have arisen by chance.

10.3389/fendo.2023.1200183

TABLE 3 Association of haplotypes of the GR receptor gene with clinical characteristics.

Haplotype	Copies	N	Age DM onset (years)	HbA1c (%)	Hs-CRP (mmol/l)	Total chol. (mmol/l)	Triglycer. (mmol/l)	LDL chol. (mmol/l)
1	0	147	49.5 (11.7)	7.3 (1.3)	4.7 (7.0)	4.2 (1.0)	1.8 (1.1)	2.3 (0.7)
wild type	1	288	51.0 (11.6)	7.3 (1.2)	4.5 (6.6)	4.2 (1.0)	1.7 (1.0)	2.4 (0.8)
	2	124	50.3 (11.2)	7.3 (1.2)	4.0 (6.0)	4.2 (0.9)	1.6 (1.0)	2.3 (0.8)
adjusted	Beta		1.53 (0.68	-0.08 (0.07)	-0.13 (0.40)	0.01 (0.06)	-0.11 (0.06)	0.05 (0.05)
	P		0.03	0.27	0.74	0.88	0.09	0.24
2	0	360	50.6 (11.4)	7.3 (1.2)	4.3 (6.3)	4.2 (1.0)	1.7 (1.0)	2.3 (0.8)
BCLI	1	178	50.2 (11.8)	7.3 (1.3)	4.7 (7.3)	4.2 (1.0)	1.7 (1.0)	2.3 (0.8)
	2	24	50.5 (11.3)	7.3 (1.1)	4.0 (3.6)	4.3 (1.0)	1.9 (1.1)	2.4 (0.9)
adjusted	Beta		-0.80 (0.81)	0.02 (0.08)	0.14 (0.48)	0.00 (0.07)	0.04 (0.08)	-0.01 (0.06)
	P		0.32	0.78	0.76	0.97	0.58	0.84
3	0	431	50.9 (11.4)	7.3 (1.2)	4.4 (6.5)	4.2 (1.0)	1.8 (2.0)	2.3 (0.8)
TthIII **+ 9β**	1	134	49.3 (11.3)	7.4 (1.4)	4.7 (7.3)	4.3 (1.0)	1.8 (1.1)	2.4 (0.8)
	2	12	44.5 (14.2)	7.8 (1.4)	5.1 (6.2)	3.8 (0.6)	1.7 (0.7)	2.1 (0.4)
adjusted	Beta		-2.33 (0.93)	0.14 (0.10)	0.14 (0.56)	-0.01 (0.08)	0.04 (0.15)	-0.03 (0.07)
	P		0.01	0.14	0.80	0.92	0.79	0.64
4	0	442	50.1 (11.5)	7.3 (1.2)	4.3 (6.3)	4.2 (1.0)	1.7 (1.0)	2.4 (0.8)
TthIII **+** ***BCLI***	1	120	52.2 (11.6)	7.2 (1.3)	5.0 (7.3)	4.2 (1.0)	1.8 (1.2)	2.3 (0.7)
	2	120	45.0 (10.8)	7.4 (0.9)	2.0 (1.3)	4.7 (0.8)	1.9 (1.4)	2.7 (0.7)
adjusted	Beta		0.45 (1.05)	-0.01(0.11)	0.16 (0.61)	-0.02 (0.09)	0.04 (0.10)	-0.05 (0.07)
	P		0.67	0.93	0.79	0.83	0.67	0.49
5	0	554	50.3 (11.5)	7.3 (1.3)	4.5 (6.7)	4.2 (1.0)	1.7 (1.0)	2.4 (0.8)
N363S	1	38	52.9 (11.2)	7.3 (1.0)	3.8 (6.9)	4.1 (0.6)	1.9 (1.0)	2.1 (0.5)
	2	38	48.5 (17.7)	6.2 (0.3)	6.7 (6.7)	4.8 (0.9)	1.6 (0.8)	2.8 (1.0)
adjusted	Beta		0.00 (1.72)	-0.03(0.18)	-0.65 (1.03)	-0.07 (0.15)	0.14 (0.16)	-0.18 (0.12)
	P		1.00	0.87	0.53	0.65	0.38	0.13
6	0	577	50.4 (11.5)	7.3 (1.3)	4.5 (6.7)	4.2 (1.0)	1.8 (1.8)	2.4 (0.8)
TthIII **+ ER22/ 23EK + 9β**	1	19	51.4 (10.7)	7.2 (0.6)	3.6 (3.2)	4.0 (0.8)	1.6 (0.9)	2.2 (0.8)
	2	0	–	–	–	–	–	–
adjusted	Beta		-0.09 (2.54)	-0.25 (0.27)	-0.62(1.53)	-0.22 (0.23)	-0.23 (0.41)	-0.21 (0.18)
	P		0.97	0.36	0.69	0.33	0.57	0.24

Data are presented as mean (sd). Adjustments in the multivariate linear regression model: Age of onset was adjusted for sex and ethnicity; HbA1c was adjusted for sex, ethnicity, diabetes duration, insulin use and metformin use; hsCRP was adjusted for age and sex; lipid spectrum was adjusted for sex, age, use of lipid lowering medication and metformin.

Our study has both strengths and weaknesses. A strength of our study is the extensive data with both detailed information on treatment as well as laboratory parameters and therefore the ability to correct for possible confounders. By including all consecutive patients in our clinic, we established a cohort that is representative for the secondary care diabetes population in an urban area. However, the heterogeneity of our study population regarding age, diabetes duration and origin might also have blunted the effect of genetic variants on metabolic parameters. A weakness of our study arises from the cross-sectional nature of the cohort. The age of diabetes onset is determined retrospectively, and we cannot exclude the possibility of recall or information bias on this outcome parameter. Although a diagnostic delay in type 2 diabetes is frequently observed, in previous studies, the duration of delay was not affected by ethnicity of the patient (26, 27). Furthermore, we do not have data on the socioeconomic position of patients, which could be an uncontrolled confounder between ethnicity and diabetes outcome parameters.

In conclusion, we observed that the prevalence of SNPs of the glucocorticoid receptor was different between ethnic groups. We

found a modest association between the 9β SNP of the GR and the level of systemic inflammation in patients with established and well-regulated type 2 diabetes. However, genetic variants of the GR did not explain the variation in age of diabetes onset and level of glycemic control; therefore, its clinical relevance for patients with established type 2 diabetes is limited.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers are as follows: ER22/23EK (rs6189): VCV000155925.10 - ClinVar - NCBI (nih.gov); 9β (rs6198): VCV000351314.5 - ClinVar - NCBI (nih.gov); N363S (rs6195 has merged into rs56149945): VCV000016150.9 - ClinVar - NCBI (nih.gov); TTH111I (rs10052957): rs10052957 RefSNP Report - dbSNP - NCBI (nih.gov); BCLI (rs41423247): rs41423247 RefSNP Report - dbSNP - NCBI (nih.gov).

Ethics statement

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5).

Author contributions

MA and MG (shared first authors) designed the study protocol, wrote the manuscript, collected data, and performed statistical analyses. PS performed the molecular biological procedures contributed to the discussion, and reviewed/edited the manuscript. VG and DB designed the study protocol contributed to the discussion, and reviewed/edited the manuscript. EM and MN contributed to the discussion and reviewed/edited the manuscript. MA and MG had full access to all data in the study and take responsibility for the integrity of data and the accuracy of data analysis. All authors contributed to the article and approved the submitted version.

Funding

The authors declare that this study received funding from Novo Nordisk BV. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Acknowledgments

We are grateful to the employers of the Molecular Biology and Clinical Chemistry departments of the former 'MC Slotervaart' Hospital, Amsterdam, The Netherlands for their time and efforts in facilitating and performing all laboratory analyses and procedures.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1200183/full#supplementary-material

References

1. Gulliford MC, Charlton J, Latinovic R. Risk of diabetes associated with prescribed glucocorticoids in a large population. *Diabetes Care* (2006) 29(12):2728–9. doi: 10.2337/dc06-1499

2. Pivonello R, De Leo M, Vitale P, Cozzolino A, Simeoli C, De Martino MC, et al. Pathophysiology of diabetes mellitus in cushing's syndrome. *Neuroendocrinology* (2010) 92 Suppl 1):77–81. doi: 10.1159/000314319

3. Lanting LC, Joung IM, Mackenbach JP, Lamberts SW, Bootsma AH. Ethnic differences in mortality, end-stage complications, and quality of care among diabetic patients: a review. *Diabetes Care* (2005) 28(9):2280–8. doi: 10.2337/diacare.28.9.2280

4. Senior PA, Bhopal R. Ethnicity as a variable in epidemiological research. *BMJ* (1994) 309(6950):327–30. doi: 10.1136/bmj.309.6950.327

5. Pujols L, Mullol J, Roca-Ferrer J, Torrego A, Xaubet A, Cidlowski JA, et al. Expression of glucocorticoid receptor alpha- and beta-isoforms in human cells and tissues. *Am J Physiol Cell Physiol* (2002) 283(4):C1324–31. doi: 10.1152/ajpcell.00363.2001

6. van Raalte DH, Ouwens DM, Diamant M. Novel insights into glucocorticoid-mediated diabetogenic effects: towards expansion of therapeutic options? *Eur J Clin Invest* (2009) 39(2):81–93. doi: 10.1111/j.1365-2362.2008.02067.x

7. Arnaldi G, Scandali VM, Trementino L, Cardinaletti M, Appolloni G, Boscaro M. Pathophysiology of dyslipidemia in cushing's syndrome. *Neuroendocrinology* (2010) 92 Suppl 1):86–90. doi: 10.1159/000314213

8. *NR3C1 nuclear receptor subfamily 3 group c member 1 [Homo sapiens (human)]* (2023). Available at: https://www.ncbi.nlm.nih.gov/gene/2908#bibliography.

9. Rosmond R. The glucocorticoid receptor gene and its association to metabolic syndrome. *Obes Res* (2002) 10(10):1078–86. doi: 10.1038/oby.2002.146

10. Manenschijn L, van den Akker EL, Lamberts SW, van Rossum EF. Clinical features associated with glucocorticoid receptor polymorphisms. an overview. *Ann N Y Acad Sci* (2009) 1179:179–98. doi: 10.1111/j.1749-6632.2009.05013.x

11. van Raalte DH, van Leeuwen N, Simonis-Bik AM, Nijpels G, van Haeften TW, Schafer SA, et al. Glucocorticoid receptor gene polymorphisms are associated with reduced first-phase glucose-stimulated insulin secretion and disposition index in women, but not in men. *Diabetes Med* (2012) 29(8):e211–6. doi: 10.1111/j.1464-5491.2012.03690.x

12. van Rossum EF, Koper JW, van den Beld AW, Uitterlinden AG, Arp P, Ester W, et al. Identification of the BclI polymorphism in the glucocorticoid receptor gene: association with sensitivity to glucocorticoids *in vivo* and body mass index. *Clin Endocrinol (Oxf).* (2003) 59(5):585–92. doi: 10.1046/j.1365-2265.2003.01888.x

13. Castro-Vale I, Duraes C, van Rossum EFC, Staufenbiel SM, Severo M, Lemos MC, et al. The glucocorticoid receptor gene (NR3C1) 9beta SNP is associated with posttraumatic stress disorder. *Healthcare (Basel)* (2021) 9(2):173. doi: 10.3390/healthcare9020173

14. Ross IL, Levitt NS, van der Merwe L, Schatz DA, Johannsson G, Dandara C, et al. Investigation of glucocorticoid receptor polymorphisms in relation to metabolic parameters in addison's disease. *Eur J Endocrinol* (2013) 168(3):403–12. doi: 10.1530/EJE-12-0808

15. Srivastava N, Prakash J, Lakhan R, Agarwal CG, Pant DC, Mittal B. Influence of bcl-1 gene polymorphism of glucocorticoid receptor gene (NR3C1, rs41423247) on blood pressure, glucose in northern indians. *Indian J Clin Biochem* (2011) 26(2):125–30. doi: 10.1007/s12291-010-0099-6

16. Syed AA, Halpin CG, Irving JA, Unwin NC, White M, Bhopal RS, et al. A common intron 2 polymorphism of the glucocorticoid receptor gene is associated with insulin resistance in men. *Clin Endocrinol (Oxf).* (2008) 68(6):879–84. doi: 10.1111/j.1365-2265.2008.03175.x

17. Kuningas M, Mooijaart SP, Slagboom PE, Westendorp RG, van Heemst D. Genetic variants in the glucocorticoid receptor gene (NR3C1) and cardiovascular disease risk. the Leiden 85-plus study. *Biogerontology* (2006) 7(4):231–8. doi: 10.1007/s10522-006-9021-2

18. Ahdi M, Gerdes VE, Graaff R, Kuipers S, Smit AJ, Meesters EW. Skin autofluorescence and complications of diabetes: does ethnic background or skin color matter? *Diabetes Technol Ther* (2015) 17(2):88–95. doi: 10.1089/dia.2013.0374

19. Bosch TM, Bakker R, Schellens JH, Cats A, Smits PH, Beijnen JH. Rapid detection of the DPYD IVS14+1G>A mutation for screening patients to prevent fluorouracil-related toxicity. *Mol Diagn Ther* (2007) 11(2):105–8. doi: 10.1007/BF03256229

20. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* (1972) 18(6):499–502. doi: 10.1093/clinchem/18.6.499

21. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* (2002) 359(9302):248–52. doi: 10.1016/S0140-6736(02)07451-2

22. B-Rao C. Sample size considerations in genetic polymorphism studies. *Hum Hered* (2001) 52(4):191–200. doi: 10.1159/000053376

23. McCarthy MI. Genomics, type 2 diabetes, and obesity. *N Engl J Med* (2010) 363 (24):2339–50. doi: 10.1056/NEJMra0906948

24. van Rossum EF, Koper JW, Huizenga NA, Uitterlinden AG, Janssen JA, Brinkmann AO, et al. A polymorphism in the glucocorticoid receptor gene, which decreases sensitivity to glucocorticoids *in vivo*, is associated with low insulin and cholesterol levels. *Diabetes* (2002) 51(10):3128–34. doi: 10.2337/diabetes.51.10.3128

25. Kussmann M, Morine MJ, Hager J, Sonderegger B, Kaput J. Perspective: a systems approach to diabetes research. *Front Genet* (2013) 4:205. doi: 10.3389/fgene.2013.00205

26. Samuels TA, Cohen D, Brancati FL, Coresh J, Kao WH. Delayed diagnosis of incident type 2 diabetes mellitus in the ARIC study. *Am J Manag Care* (2006) 12 (12):717–24.

27. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the united states, 1988-2012. *JAMA* (2015) 314(10):1021–9. doi: 10.1001/jama.2015.10029

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 06 September 2023
DOI 10.3389/fendo.2023.1186252

OPEN ACCESS

EDITED BY
Jeff M. P. Holly,
University of Bristol, United Kingdom

REVIEWED BY
Eric Bonnet,
Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), France
Vanessa Vermeirssen,
Ghent University, Belgium

*CORRESPONDENCE
Sean Bankier
Sean.Bankier@uib.no

RECEIVED 15 March 2023
ACCEPTED 14 July 2023
PUBLISHED 06 September 2023

CITATION
Bankier S, Wang L, Crawford A, Morgan RA, Ruusalepp A, Andrew R, Björkegren JLM, Walker BR and Michoel T (2023) Plasma cortisol-linked gene networks in hepatic and adipose tissues implicate corticosteroid-binding globulin in modulating tissue glucocorticoid action and cardiovascular risk.
Front. Endocrinol. 14:1186252.
doi: 10.3389/fendo.2023.1186252

Plasma cortisol-linked gene networks in hepatic and adipose tissues implicate corticosteroid-binding globulin in modulating tissue glucocorticoid action and cardiovascular risk

Sean Bankier[1,2,3]*, Lingfei Wang[3], Andrew Crawford[1,4], Ruth A. Morgan[1,5], Arno Ruusalepp[6,7,8], Ruth Andrew[1], Johan L. M. Björkegren[8,9,10], Brian R. Walker[1,11] and Tom Michoel[2,3]

[1]University/BHF Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom, [2]Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway, [3]Division of Genetics and Genomics, The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, [4]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, [5]SRUC, The Roslin Institute, Edinburgh, United Kingdom, [6]Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia, [7]Department of Cardiology, Institute of Clinical Medicine, Tartu University, Tartu, Estonia, [8]Clinical Gene Networks AB, Stockholm, Sweden, [9]Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden, [10]Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [11]Clinical and Translational Research Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

Genome-wide association meta-analysis (GWAMA) by the Cortisol Network (CORNET) consortium identified genetic variants spanning the *SERPINA6/SERPINA1* locus on chromosome 14 associated with morning plasma cortisol, cardiovascular disease (CVD), and *SERPINA6* mRNA expression encoding corticosteroid-binding globulin (CBG) in the liver. These and other findings indicate that higher plasma cortisol levels are causally associated with CVD; however, the mechanisms by which variations in CBG lead to CVD are undetermined. Using genomic and transcriptomic data from The Stockholm Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET) study, we identified plasma cortisol-linked single-nucleotide polymorphisms (SNPs) that are trans-associated with genes from seven different vascular and metabolic tissues, finding the highest representation of trans-genes in the liver, subcutaneous fat, and visceral abdominal fat, [false discovery rate (FDR) = 15%]. We identified a subset of cortisol-associated trans-genes that are putatively regulated by the glucocorticoid receptor (GR), the primary transcription factor activated by cortisol. Using causal inference, we identified GR-regulated trans-genes that are responsible for the regulation of tissue-specific gene networks. Cis-expression Quantitative Trait Loci (eQTLs) were used as genetic instruments for identification of pairwise causal relationships from which gene networks could be reconstructed. Gene networks were identified in the liver, subcutaneous fat, and visceral abdominal fat, including a high confidence gene network specific to subcutaneous adipose (FDR = 10%)

under the regulation of the interferon regulatory transcription factor, *IRF2*. These data identify a plausible pathway through which variation in the liver CBG production perturbs cortisol-regulated gene networks in peripheral tissues and thereby promote CVD.

KEYWORDS

cortisol, corticosteroid-binding globulin, gene networks, systems genetics, causal inference

1 Introduction

The steroid cortisol is the major glucocorticoid hormone involved in mediating the human stress response, with effects on metabolism, cardiovascular homeostasis, and inflammation (1). Excessive cortisol production occurs in Cushing's syndrome either in response to chronic activation of the hypothalamic-pituitary-adrenal (HPA) axis by increased adrenocorticotropic hormone (ACTH) secretion or through autonomous production of cortisol in an adrenocortical tumor (2). The incidence of Cushing's syndrome is low, with the number of cases estimated to be between 0.7 and 2.4 cases per million (3). It results in insulin resistance, obesity and hypertension with increased risk of cardiovascular disease (CVD). Similarly, higher plasma cortisol within the population, in the absence of overt Cushing's syndrome, is associated with risk factors for CVD such as hypertension (4) and type II diabetes (1, 5).

Interindividual variation in plasma cortisol levels has a genetic basis with heritability estimated between 30% and 60% (6). The Cortisol Network (CORNET) consortium conducted a genome-wide association meta-analysis (GWAMA) with the intention of uncovering genetic influences on the HPA axis function (7). This was followed in 2021 with an updated GWAMA of 25,314 individuals across 17 population-based cohorts of European ancestries (8), expanded from 12,597 individuals in the original GWAMA. In an additive genetic model, the new CORNET GWAMA identified 73 genome-wide significant single-nucleotide polymorphisms (SNPs) associated with variation for plasma cortisol at a single locus on chromosome 14. These SNPs were used in a two-sample Mendelian randomization analysis showing that higher cortisol is causative for CVD (8).

The locus on chromosome 14 spans the genes *SERPINA6* and *SERPINA1* that both play roles in the regulation of corticosteroid-binding globulin (CBG), a plasma protein produced in the liver that is responsible for binding 80%–90% of cortisol in the blood (9, 10). *SERPINA6* encodes CBG (11), and *SERPINA1* encodes α1-antitrypsin, an inhibitor of neutrophil elastase, a serine protease that can cleave the reactive center loop of CBG resulting in a 9–10-fold reduction in binding affinity to cortisol (12, 13).

The CORNET GWAMA showed that 21 cortisol-associated SNPs were also cis-expression Quantitative Trait Loci (eQTLs) for *SERPINA6* in the liver and demonstrated that the genetic variation associated with plasma cortisol is driven by *SERPINA6* rather than *SERPINA1* (8). However, although variation in CBG production could explain changes in total plasma cortisol, it is the free fraction of cortisol that is considered to equilibrate with target tissue concentrations and signal through intracellular glucocorticoid receptors (GR) (14, 15). While CBG deficiency may be associated with symptoms (16–18), variations in CBG have not been shown conclusively to influence the tissue response to cortisol in humans.

To test the hypothesis that cortisol-associated genetic variants in the *SERPINA6/SERPINA1* locus influence cortisol delivery to, and hence action in, extrahepatic tissues, we investigated transcriptome-wide associations between cortisol-associated SNPs and gene transcripts across seven different vascular and metabolic tissues from the Stockholm Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET) study (19). As well as conducting a multi-tissue eQTL analysis using STARNET transcriptomics and plasma cortisol-associated SNPs, we identified tissue-specific trans-eQTL-associated genes under the regulation of GR. Moreover, we used a causal inference framework, with cis-eQTLs as genetic instruments, for the reconstruction of causal gene networks within STARNET tissues.

These results provide evidence that genetic variations in CBG production in liver influence extra-hepatic cortisol signaling and provide plausible pathways leading to CVD.

2 Materials and methods

2.1 Data

STARNET is a cohort-based study of 600 individuals undergoing coronary artery bypass grafting (CABG) for coronary artery disease (CAD) and was used as the primary discovery cohort in this study. These individuals underwent blood genotyping preoperatively for 951,117 genomic markers, and during surgery, seven different tissue samples were obtained and underwent RNA-sequencing (RNA-seq): liver, skeletal muscle, atherosclerotic aortic root, internal mammary artery, visceral abdominal fat, subcutaneous fat, and whole blood. STARNET data are available through a database of Genotypes and Phenotypes (dbGaP) application (accession no. phs001203.v2.p1). A detailed description of data processing can be found in the Supplemental Material of this article (section S1.1).

The Stockholm Atherosclerosis Gene Expression (STAGE) study (n = 114) (20) and the Metabolic Syndrome in Man (METSIM) study (n = 982) (21) were used in the replication of causal gene networks identified using STARNET. Gene expression data for the METSIM and STAGE studies are available publicly at Gene expression omnibus (GEO) (accession no. GSE70353 and GSE40231, respectively). Microarray data for the liver, subcutaneous fat, and visceral abdominal fat were used from the STAGE study, and gene expression data from subcutaneous fat were measured in the METSIM study using RNA-seq.

2.2 Multi-tissue trans-eQTL discovery

A list of SNPs associated with plasma cortisol was obtained from the summary statistics of the 2021 GWAMA conducted by the CORNET consortium (available at https://datashare.ed.ac.uk/handle/10283/3836) (8). We filtered this list to obtain SNPs that were found to be associated with plasma cortisol at a level of genome-wide significance ($p< 5 \times 10^{-8}$) that were taken forward 68 and tested against all genes across STARNET tissues.

The secondary linkage test (P2) is a likelihood ratio test in the Findr package (22) (version 1.0.8) that was used to identify associations between a given SNP (E) and a gene (B) using categorical regression. P2 proposes a null hypothesis where E and B are independent and alternative hypotheses where E is causal for B (E →B). Maximum likelihood estimators are then used to obtain a log likelihood ratio (LLR) between the alternative and null hypotheses. The LLR is then converted to the posterior probability of the alternative hypothesis $\mathcal{H}_{alt}^{(P2)}$ being true with empirical estimation of the local false discovery rate (FDR) as a value from 0 to 1 (Equation 1).

$$P(E \rightarrow B) = P(\mathcal{H}_{alt}^{(P2)} | LLR^{(P2)}) . \quad (1)$$

2.3 Identification of glucocorticoid-regulated trans-genes

Multiple datasets were used to identify genes that had prior evidence of putative regulation by GR (23–27). These datasets have been filtered to include targets for NR3C1, the gene that encodes GR.

Trans-genes were categorized according to evidence of GR regulation from datasets shown in Supplementary Table S1. Genes were scored against these criteria: 1) appearing in a transcription factor database (ENCODE, TRANSFAC, CHEA); 2) identified as a GR target from chromatin immunoprecipitation sequencing (ChIP-seq) experiment in adipocytes from Yu et al. (23); 3) differentially expressed in response to dexamethasone treatment in adipocytes from Yu et al. (23); and 4) murine homolog of human gene differentially expressed in response to dexamethasone treatment using adrenalectomized mice (FC >1; p-value<0.05) (24). Genes were then ranked according to how well they met the criteria for GR regulation (+1 for each item matched from criteria 1–4).

2.4 Causal gene network reconstruction

Pairwise causal inference was used for the reconstruction of cortisol-responsive transcriptional networks across STARNET tissues using cis-eQTL genotypes as genetic instruments with gene expression data from STARNET, as implemented by the Findr software (22). A detailed description of these methods can be found in the Supplementary Material of this article (Section S1.2).

2.5 Transcription factor target enrichment

Lists of known transcription factor targets for both *NR3C1* and *IRF2* were obtained from ENCODE and TRANSFAC datasets, respectively. These datasets were used to test for an enrichment of known transcription factor targets within novel gene sets derived from gene network targets. This was performed using Fisher's exact test from the Python module Scipy Stats (28) and involved the creation of a 2 × 2 contingency table based on a tissue-specific background consisting of all genes available in the corresponding tissue.

2.6 Gene network replication

Correlations between gene network targets were calculated using gene expression data from STARNET, STAGE, and METSIM. Gene expression matrices were filtered to only include the target genes under investigation. Correlation matrices of corresponding Pearson correlation coefficients as absolute values were constructed in Python.

A background gene set was constructed from the overlapping genes between the STARNET gene expression set that was used for network discovery and the corresponding gene expression set that was being used for replication. The previously described correlation analysis was then repeated using a random set of genes (the same size as the target set) selected from the background gene set. The Kruskal–Wallis test was implemented in Python using Scipy Stats (28) to test if the targeted and randomly sampled correlations follow the same distribution. Both the targeted and random correlations were then plotted as a boxplot using the Python plotting package Seaborn (29).

2.7 Gene expression clustering

Hierarchical clustering was performed on correlation values between network targets using the discovery (STARNET) gene expression data and hierarchical clustering from Scipy Stats (28) in Python. The leaves list that resulted from the clustering of the discovery dataset was then extracted and applied to the correlations between target genes from the corresponding replication dataset. Both sets of clustered correlation values were then plotted as opposing correlation heatmaps with Seaborn (29).

3 Results

3.1 Cortisol-associated trans-genes

SNPs associated with plasma cortisol at the *SERPINA6/SERPINA1* locus have previously been linked as expression single-nucleotide polymorphisms (eSNPs) for *SERPINA6* in the liver (8). Using genotype and tissue-specific RNA-seq data from the STARNET cohort, we explored the hepatic and extrahepatic consequences of genetic variation for plasma cortisol using 73 cortisol-associated SNPs at genome-wide significance ($p < 5 \times 10^{-8}$) identified from the CORNET GWAMA (8). We identified 704 eQTL associations in cis and trans between plasma cortisol-associated SNPs and genes measured across all STARNET tissues, composed of 262 unique genes and 72 SNPs at a 15% FDR threshold (Supplementary Tables S2, S3).

The tissues with the greatest number of trans-genes were the liver, subcutaneous fat, and visceral abdominal fat, with a combined total of 157 trans-genes and 422 total SNP–gene associations (FDR = 15%) (Figure 1A). The vast majority of trans-eQTL associations were specific to a single tissue. A single trans-gene, the glycosyltransferase-encoding gene *OGT*, was identified in both the liver and visceral abdominal fat. However, as this was the only cross-tissue trans-gene identified, suggesting that the transcriptional impact of genetic variation at the *SERPINA6/SERPINA1* locus is highly tissue-specific. The CORNET GWAMA describes four blocks of SNPs in linkage disequilibrium (LD), which represent the cortisol-associated variation at the *SERPINA6/SERPINA1* locus (8). We observed that LD blocks 2 and 4 represent the majority of the variation across all tissues in the trans-gene sets (Figures 1B, C).

3.2 GR-regulated trans-genes associated with plasma cortisol

As the GR is the primary mechanism by which cortisol influences transcription, we sought to identify a subset of cortisol-associated trans-genes that were also regulated by the GR. The cortisol-associated trans-genes identified in this study were compared to sets of known GR targets identified from different sources as described in Supplementary Table S1. This included large projects such as ENCODE, TRANSFAC, and CHEA that predict transcription factor-binding targets from high-throughput transcription factor-binding assays. We also included predicted GR targets from perturbation-based experiments in specific tissues. ChIP-seq and microarray analysis has been used to identify 274 glucocorticoid-regulated genes in 3TS-L1 adipocytes, a murine-derived cell line (23). In addition, RNA-seq data in subcutaneous fat from adrenalectomized mice treated with dexamethasone, a GR agonist, have been used to identify genes that are differentially expressed (24).

The greatest number of unique cortisol-associated trans-genes was identified in the liver (n = 43), subcutaneous fat (n = 54), and visceral abdominal fat (n = 59) at a 15% FDR threshold. The involvement of these tissues in glucocorticoid signaling and physiological effects has been well documented in the literature (31–34); therefore, the identification of GR-regulated trans-genes was restricted to these tissues. Comparisons of genes identified as glucocorticoid-regulated in 3T3-L1 adipocytes were only made with subcutaneous and visceral adipose trans-genes. Likewise, as the murine RNA-seq experiments were restricted to subcutaneous adipose, only subcutaneous adipose trans-genes were compared to these differentially expressed genes.

In the liver trans-gene set, 19/43 genes were identified that were present in either the ENCODE, TRANSFAC, or CHEA datasets (FDR = 15%) (Figure 1D; Supplementary Table S4). This includes *SERPINA6* that is cis-associated with genetic variation for plasma cortisol, as described previously (8). One gene, *CPEB2*, was identified in more than one dataset and was present in both ENCODE and CHEA. *CPEB2* (posterior probability = 0.89) is a regulator of translation, splice variants of which have been linked to cancer metastasis (35).

Visceral adipose tissue had the largest number of cortisol-associated trans-genes. Here, 21/59 of these genes had some evidence of being targets of GR (Figure 1E; Supplementary Table S5). There were five genes that had been identified as GR targets from both high-throughput transcription factor-binding assays and adipose-specific experiments. These include *CD163* and *LUC7L3*. *CD163* is a hemoglobin scavenger protein that is expressed in macrophages and involved in the clearance of hemoglobin/haptoglobin complexes that may play a role in the protection from oxidative damage. It also plays a role in activating macrophages as part of the inflammatory response (36). *LUC7L3*, also known as CROP, encodes a protein that is involved in alternative splicing and is associated with human heart failure (37). It has also been shown to play a role in the inhibition of hepatitis B replication (38).

Of the cortisol-associated trans-genes identified in subcutaneous adipose (FDR = 15%), 28/54 genes were either present in a transcription factor dataset or identified from the adipose-specific perturbation datasets (Figure 1F; Supplementary Table S6). There were 13 genes that had been identified as GR targets from both high-throughput transcription factor-binding assays and adipose-specific experiments. These include *RNF13* that encodes IRE1α-interacting protein that plays an important role in the endoplasmic reticulum (ER) stress response through regulation of IRE1α, a critical sensor of unfolded proteins (39). Also *IRF2*, encoding the transcription factor Interferon Regulatory Factor 2 that plays an important role as a repressor of *IRF1* that in turn is involved in the interferon-mediated immune response (40). Furthermore, *IRF1* has previously been identified as a marker for glucocorticoid sensitivity in peripheral blood (41).

3.3 Reconstruction of cortisol-associated gene networks

Having identified cortisol-associated trans-genes that are regulated by GR, causal estimates were obtained for pairwise relationships between GR-regulated trans-genes and all other genes within the given tissue. This was carried out for all GR-

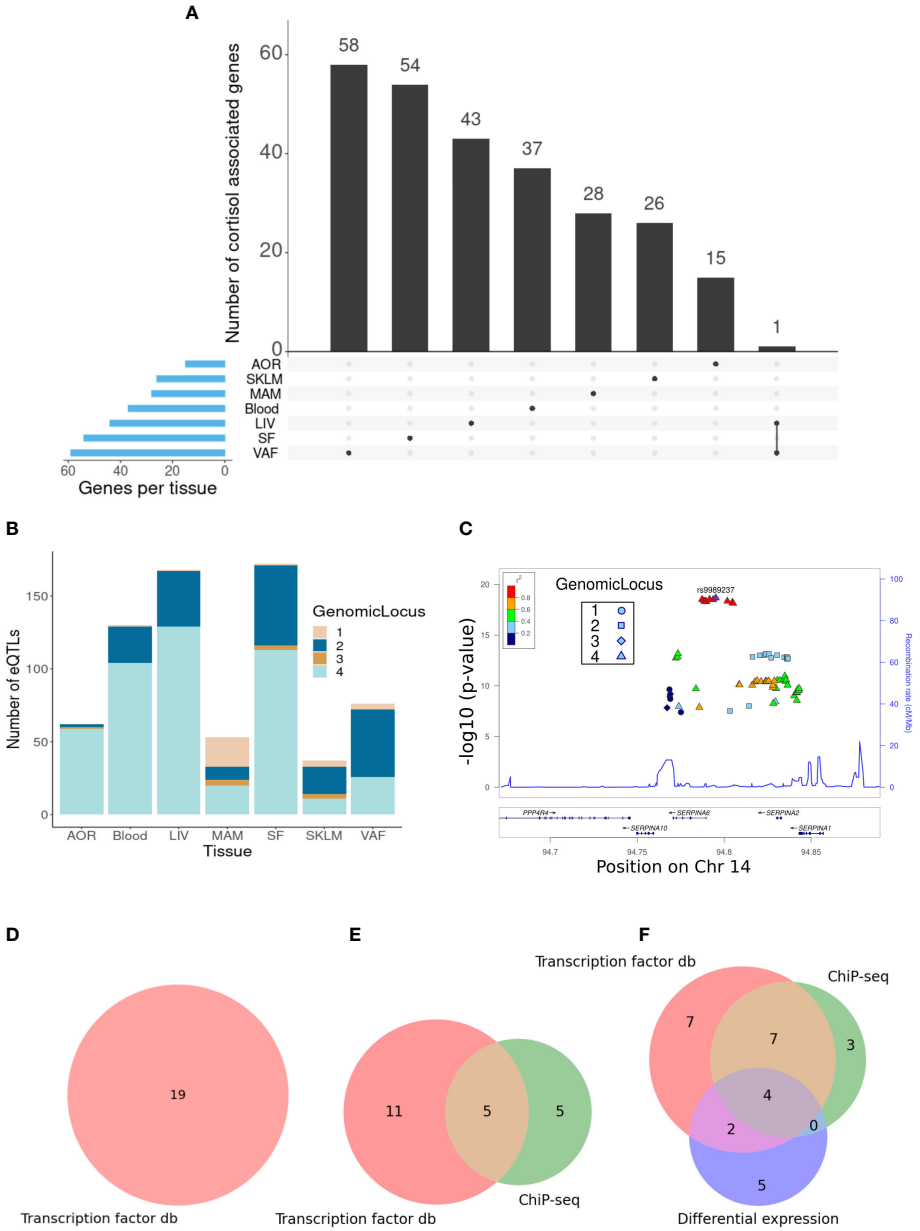


FIGURE 1
Identification of cortisol-associated trans-genes across STARNET tissues (FDR = 15%). **(A)** Upset plot showing the distribution of trans-genes across STARNET tissues, including genes shared by multiple tissues. Tissues include the atherosclerotic aortic root (AOR), skeletal muscle (SKLM), internal mammary artery (MAM), blood (Blood), liver (LIV), subcutaneous fat (SF), and visceral abdominal fat (VAF). **(B)** Distribution of trans-eQTLs across tissues and colored by genomic locus (LD block) of associated SNP. **(C)** LocusZoom plot (30) showing the location of cortisol-associated SNPs within defined LD blocks. **(D)** Venn diagrams where groupings represent different sources used to identify GR-linked trans-genes in the liver, **(E)** visceral abdominal fat, and **(F)** subcutaneous fat. These sources include transcription factor databases (db), ChIP-seq from perturbation-based experiments (23), and differential expression of dexamethasone-treated mice (24).

regulated trans-genes in the liver, subcutaneous fat, and visceral abdominal fat with a valid cis-eQTL instrument (12, 19, and 7 genes, respectively) (Supplementary Table S7). A 10% global FDR threshold was then imposed for each gene set (Table 1). Primary networks were obtained by filtering to include only GR trans-genes with a minimum of four target genes at the global FDR threshold.

In the liver, we identified a single gene network driven by *CPEB2*, which was found to be trans-associated with the cortisol-associated SNP rs4905194 (Figure 2A). This network contained 48 causal interactions driven by *CPEB2* at a 10% FDR threshold (Figure 2D; Supplementary Table S9). It is notable that *CPEB2* appears as the only network regulator in the liver considering it was also the cortisol-associated trans-gene with the strongest links to GR regulation from the liver trans-gene set. A detailed description of the *CPEB2* network and all other networks identified can be found in the Supplementary Information (Section S2.1).

In subcutaneous fat, two major subnetworks were identified under the regulation of the genes *RNF13* and *IRF2*. This includes a

10.3389/fendo.2023.1186252

TABLE 1 Number of network targets following FDR filtering.

Tissue	FDR threshold	Total targets	Network regulator	Regulator targets
Liver	15%	197	*CPEB2*	190
	10%	48	*CPEB2*	44
Subcutaneous fat	15%	1,701	*RNF13*	416
			IRF2	247
			PBX2	883
	10%	486	*RNF13*	215
			IRF2	128
			PBX2	138
Visceral abdominal fat	15%	396	*CD163*	378
			LUC7L3	15
	10%	17	*CD163*	4
			LUC7L3	11

Total targets include all pairwise interactions at the given threshold, and network regulators correspond to trans-genes with at least four network targets at the given FDR threshold. Inclusive of network regulators present at both 10% and 15% thresholds.

total of 343 causal relationships across both subnetworks, including two genes shared by both subnetworks. *RNF13* was found to be trans-associated with the cortisol-associated SNP rs11622665 (Figure 2B) and represents the largest subcutaneous fat subnetwork with 215 gene targets at a 10% FDR threshold (Figure 2E; Supplementary Table S10).

The transcription factor *IRF2*, which was associated with the cortisol-linked SNP rs8022616 (Figure 2C), was found to putatively regulate a network of 128 genes (FDR = 10%) (Figure 2F). Some notable targets of *IRF2* include *LDB2* (posterior probability = 0.94) and *LIPA* (posterior probability = 0.91). GWAS suggests functions for *LIPA* related to CAD and ischemic cardiomyopathy (42), while *LDB2* has been demonstrated to be involved in the development of atherosclerosis (43). Additionally, cortisol has been shown to induce a 5-fold reduction in *LDB2* expression in adipocytes (44).

Predicted *IRF2* transcription factor targets have been previously described as part of the TRANSFAC dataset. We examined the overlap between predicted *IRF2* targets in TRANSFAC, and gene targets within the *IRF2* causal networks were identified in subcutaneous fat. A true network of *IRF2* targets would be expected to show an enrichment of predicted *IRF2*. Using Fisher's exact test on data from subcutaneous fat, at a 10% FDR threshold, the *IRF2* network had 128 target genes, 35 of which were also predicted *IRF2* targets ($p = 0.08$); at a 15% FDR threshold, 104/247 causal targets were also predicted targets of *IRF2* in TRANSFAC ($p = 0.005$). Decreasing the global FDR beyond this threshold increased the number of TRANSFAC targets within the pool of causal targets, however at a lower enrichment ($p = 0.046$) (Supplementary Table S12).

In addition to examining the prevalence of *IRF2* targets within the *IRF2* causal network, we investigated the overlap between network genes that are also regulated by GR. We observed an enrichment of ENCODE GR targets at 15% and 20% FDR thresholds ($p < 0.05$) including 68 and 138 GR targets, respectively. No GR enrichment was observed in either CHEA or TRANSFAC datasets for *IRF2* networks.

3.4 Co-expression of cortisol network targets in independent datasets

Causal gene networks represent coordinated changes in gene expression in response to changes in the expression of network regulators. Therefore, it is possible to examine if these changes in gene expression are present in independent datasets using gene expression data alone. We used RNA-seq and microarray data from the METSIM and STAGE datasets, respectively, to compare patterns in gene expression within causal networks predicted from STARNET. As METSIM only contains gene expression data for subcutaneous fat, analysis was restricted to the causal networks identified in STARNET subcutaneous fat.

Absolute correlation coefficients between the targets of the previously described network regulators were calculated, and their distributions were compared to distributions of random sets of genes selected from the replication gene expression data, the same size as the corresponding target gene set. The difference between targeted and random distributions was formalized using the Kruskal–Wallis test for each subnetwork (Table 2).

In the liver, correlations between network targets of the single subnetwork under the regulation of *CPEB2* were observed in STARNET and STAGE. Hierarchical clustering within the STARNET liver also revealed clustering of correlated genes that were retained when the clustered gene order was then applied to the STAGE liver (Figure 3A). Correlations between the 44 *CPEB2* target genes in the STAGE liver were stronger than their random counterparts ($p = 8.2 \times 10^{-32}$), with this shift also being observed in the STARNET liver ($p = 2.32 \times 10^{-197}$) (Figure 3D).

10.3389/fendo.2023.1186252

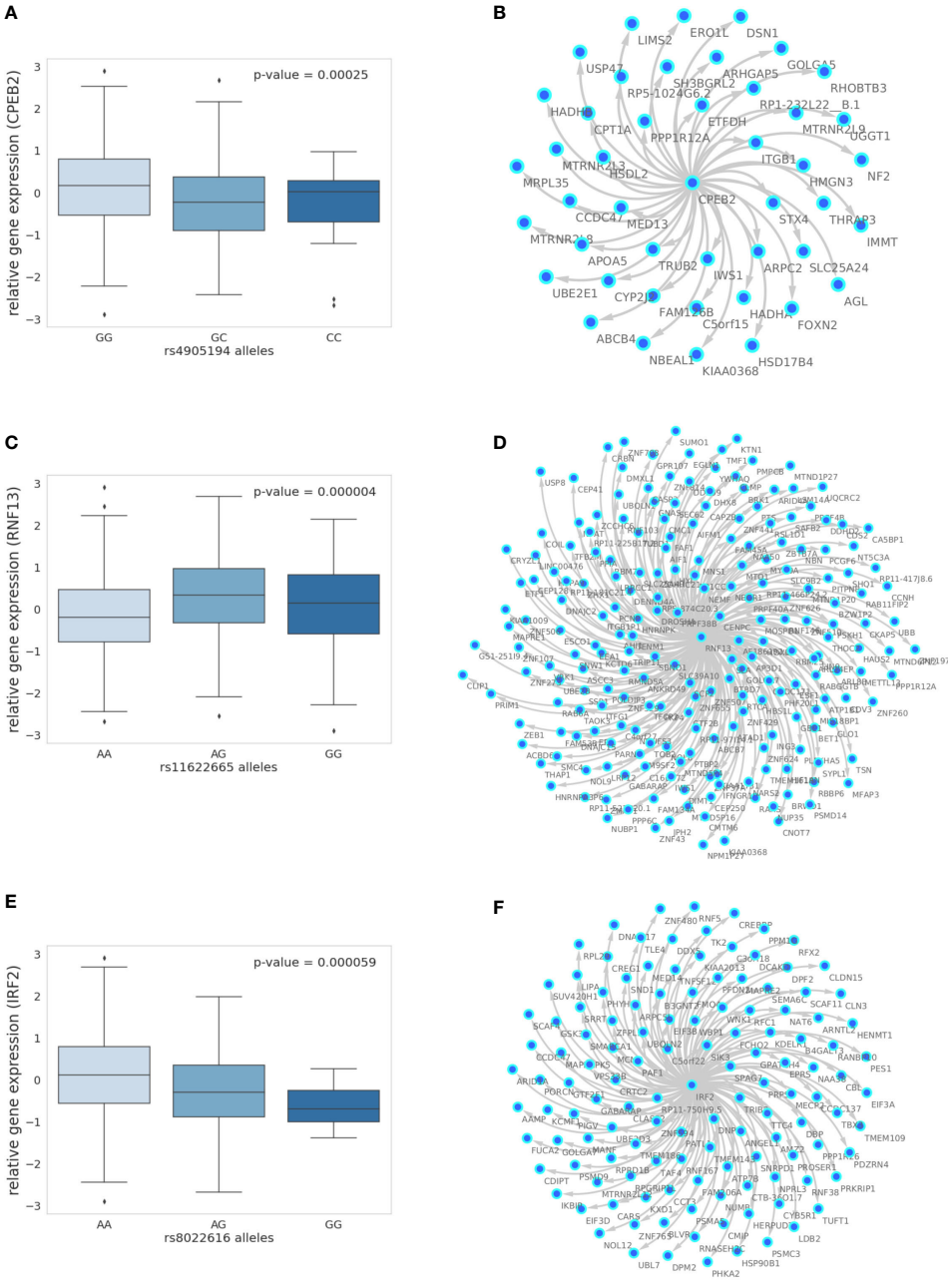


FIGURE 2
The 10% FDR gene networks in STARNET across different tissues. **(A)** Gene expression boxplot in the liver showing trans-association with cortisol-linked SNP rs4905194 and *CPEB2*, **(B)** in subcutaneous fat between rs11622665 and *RNF13* and **(C)** rs8022616 and *IRF2* (p-value obtained from Kruskal–Wallis test statistic). Box shows quarterlies of the dataset, with whiskers indicating the upper and lower variability of the distribution. **(D)** Causal gene network reconstructed from pairwise interactions from GR-regulated trans-genes against all other genes in the corresponding tissue for *CPEB2*, **(E)** *RNF13*, and **(F)** *IRF2*. Edges represent Bayesian posterior probabilities of pairwise interaction between genes (nodes) exceeding 10% global FDR. Arrow indicates direction of regulation, and interactions were only retained where parent node had at least four targets.

In subcutaneous fat, correlations were observed between the network targets of *RNF13* and *IRF2*, and hierarchical clustering patterns from STARNET were applied to the replication datasets of STAGE and METSIM (Figures 3B, C). For *RNF13*, similar patterns of co-expression were observed in the STAGE subcutaneous fat following clustering; however, this was not the case in the METSIM dataset (Figure 3B). Despite this, *RNF13* targets appeared more highly correlated than their randomly selected counterparts in STARNET ($p< 1.0 \times 10^{-300}$), STAGE ($p< 1.0 \times 10^{-300}$) and to a lesser extent in METSIM ($p = 2.3 \times 10^{-7}$) (Figure 3E).

In subcutaneous fat, patterns of co-expression between *IRF2* targets were conserved most prominently in METSIM; however, co-expression was less strongly correlated compared with *RNF13* targets (Figure 3C). *IRF2* subcutaneous fat subnetwork targets

TABLE 2 Correlations between network targets within replication datasets.

Replication dataset	Tissue	Network regulator	p-value	No. target genes
METSIM	Subcutaneous fat	*IRF2*	$< 1.0\times10^{-300}$	128
		RNF13	2.3×10^{-7}	215
STAGE	Liver	*CPEB2*	8.2×10^{-32}	44
	Subcutaneous fat	*IRF2*	8.3×10^{-86}	128
		RNF13	$< 1.0\times10^{-300}$	215
	Visceral abdominal fat	*CD163*	2.6×10^{-3}	4
		LUC7L3	4.4×10^{-1}	11

The Kruskal–Wallis test calculated for the distribution of correlations between network targets compared to correlations within random gene sets of the same size.

were more strongly correlated than their random counterparts in STARNET ($p< 1.0 \times 10^{-300}$), STAGE ($p = 8.35 \times 10^{-86}$), and METSIM ($p< 1.0 \times 10^{-300}$) (Figure 3F).

4 Discussion

In this study, we have characterized the impact that genetic variation for plasma cortisol has upon tissue-specific gene expression. We showed that cortisol-linked genetic variants at the *SERPINA6/SERPINA1* locus mediate changes in gene expression in trans across multiple tissues, in addition to the cis-associations in the liver that have been described previously (8). We have scrutinized these trans-associations to identify a subset of genes that are regulated by glucocorticoids and in turn regulate downstream transcriptional networks, thus providing a deeper understanding of the transcriptional landscape driven by cortisol-linked genetic variation that may underpin the progression to CVD.

CBG, as encoded by *SERPINA6*, is responsible for binding cortisol in the blood. It has remained uncertain whether variation in CBG impacts the availability of cortisol within tissues, since any resulting change in free cortisol concentrations would be expected to be adjusted by negative feedback of the HPA axis (45). However, deleterious mutations in CBG are associated with dysfunction in animals and humans, suggesting an impact of CBG on cortisol signaling (45). Our major finding that downstream transcriptomic changes in extrahepatic tissues are associated with genetic variation at the *SERPINA6* locus lends strong support to the hypothesis that CBG influences tissue delivery of cortisol and modulates glucocorticoid-induced changes in gene expression.

For the STARNET study, whole-blood samples were taken preoperatively and all other tissues including the liver were taken during the CABG surgery. In addition to any rise in cortisol due to anxiety and disturbed sleep in anticipation of surgery, the human stress response to surgery has been well characterized and results in stimulation of the HPA axis leading to high levels of cortisol in the blood both during and post-surgery (46). Surgery is also associated with a very rapid fall in CBG production. Therefore, it is uncertain if cortisol-associated gene expression patterns observed in STARNET would also be observed in an unstressed healthy population. It may be that CBG influences the dynamic range of alterations in free plasma cortisol during stress rather than affecting the delivery of cortisol to tissues in unstressed conditions. However, considering that co-expression of the network targets was reproducible within independent samples from the METSIM study, obtained under nonsurgical conditions, this suggests that the cortisol-associated networks we inferred from STARNET do operate also in unstressed conditions.

The tissues with the greatest number of trans-genes identified were the liver and both subcutaneous and visceral abdominal fat, all tissues known to play a role in glucocorticoid biology. In the liver, glucocorticoids have extensive effects on glucose and fatty acid metabolism (31, 32), while in adipose tissue, glucocorticoids regulate lipogenesis and lipid turnover (33, 34). Skeletal muscle is also a major target of glucocorticoids, where they modulate protein and glucose metabolism (47). A lack of available data for identifying tissue-specific GR targets in other tissues means that potential GR targets may have been missed in tissues outside of the liver and adipose.

We identified a subset of GR-responsive genes in the liver, subcutaneous fat, and visceral adipose fat. However, we did not observe a statistical enrichment of GR-regulated genes in any of these trans-gene sets. This does not negate the identification of GR targets that are associated with plasma cortisol, but it may imply that there are some effects of cortisol-linked genetic variation that are mediated by mechanisms other than directly by GR either through secondary regulation by GR-regulated genes or through the alternative mineralocorticoid receptor. Indeed, some of the genes with higher levels of evidence for GR regulation also demonstrated regulation of transcription networks, e.g., *CPEB2*, *IRF2*, and *RNF13*. This supports our strategy of setting a relatively lenient FDR threshold and then filtering to identify cortisol-associated trans-genes with prior evidence of GR regulation.

It should be noted that different FDR thresholds were used for the trans-gene discovery and for the network reconstruction. Initially, we selected a more lenient threshold of 15% for the identification of trans-genes, considering that trans-eQTLs tend to exhibit weaker associations compared to their cis counterparts (47). We then decided to restrict our list of trans-associations by implementing a biological rather than a statistical threshold,

10.3389/fendo.2023.1186252

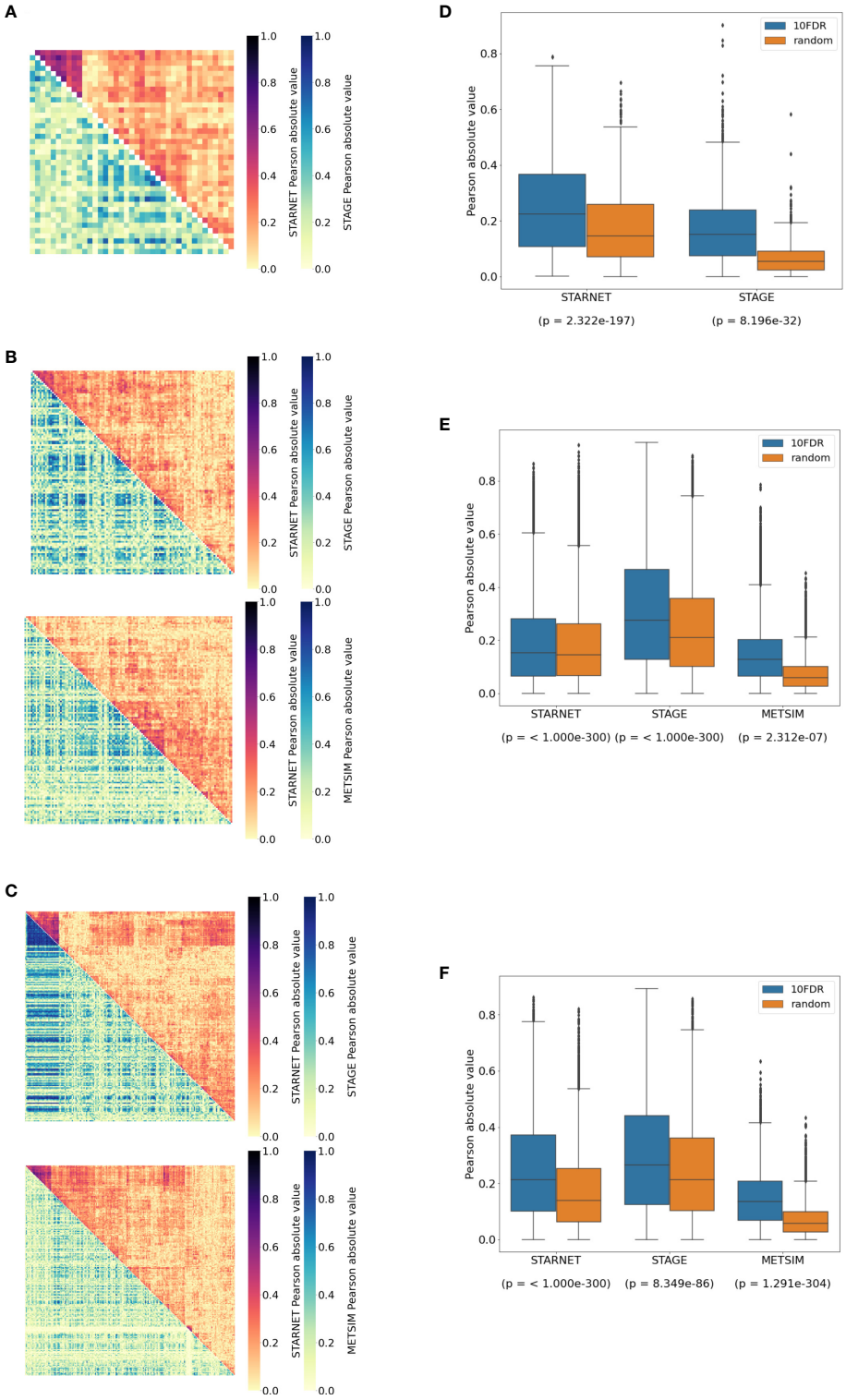


FIGURE 3
Replication of cortisol-associated gene networks in independent datasets. **(A)** Correlation heatmap showing pairwise Pearson correlations between *CPEB2*, **(B)** *IRF2*, and **(C)** *RNF13* network targets. Hierarchical clustering of genes in STARNET (discovery) was applied to the same genes within replication datasets. **(D)** Correlations between network targets in discovery vs. replication datasets for *CPEB2*, **(E)** *IRF2*, and **(F)** *RNF13* networks. The Kruskal–Wallis test calculated for the distribution of correlations between network targets compared to correlations within random gene sets of the same size.

limiting the number of trans-genes to those with evidence of GR regulation. However, given that there was no biological threshold implemented with network reconstruction, a more stringent FDR threshold was appropriate. The 10% FDR in this context implies that 1 in 10 edges of a given network is a potential false positive. However, given the strength of the replication within independent datasets, this suggests that these networks are considerably robust.

We identified causal gene networks in the liver, subcutaneous fat, and visceral abdominal fat where cortisol-associated trans-genes act as regulators of subnetworks within overarching tissue-specific networks. Pairwise causal relationships were established between network regulators and downstream targets using cis-eQTLs as genetic instruments. This approach has the benefit of generating directed relationships between a regulator and target while accounting for any unobserved confounding. However, a drawback of this approach is that we are limited by only being able to examine GR-regulated trans-genes with valid cis-eQTLs. This means that there could be valid cortisol-responsive networks regulated by GR trans-genes that we were unable to predict due to lack of a corresponding instrument.

IRF2 stands out as a network regulator of particular interest. There is strong evidence of GR regulation, where *IRF2* has been identified as a GR target from published dexamethasone-treated adipocyte ChIP-seq experiments (23) and as a putative GR target within ENCODE. It is robustly associated with its corresponding cis-eQTL instrument, and there is an enrichment of *IRF2* targets within our predicted *IRF2*-regulated causal network. Additionally, we show evidence of regulation by glucocorticoids within the targets of *IRF2*, potentially suggesting evidence of a feed-forward loop motif (48). Interestingly, the genotype for rs8022616, the cortisol-associated SNP linked to *IRF2* expression in subcutaneous fat, is associated with a decrease in *IRF2* expression. Previous evidence suggests that interferon signaling is inhibited by glucocorticoids (49, 50).

Although we have determined the direction of causality between the regulator and target genes, we do not know if the expression of the target gene is upregulated or downregulated in response to modulation of the regulator. This could be investigated through functional experiments within a relevant cell line, whereby the differential gene expression of target genes is measured in response to perturbation of the network regulator. To take this one step further, the results of a cell line experiment could be used to determine the dynamics of the putative cortisol networks using systems biology approaches for modelling gene expression (51).

In conclusion, we have linked genetic variation for plasma cortisol to changes in gene expression across the genome, beyond that which has been previously described at the *SERPINA6/SERPINA1* locus (8) and extending to adipose tissue as well as the liver. Furthermore, we have shown that a subset of these trans-genes is driven by the GR and in turn drives transcriptional networks across different tissues. These networks have been found to be robust and their network targets appear co-expressed within independent gene expression datasets of the same tissue. Further study of these networks and their downstream targets could be used to enhance our mechanistic understanding of the pathways linking cortisol with complex diseases as described in observational studies.

Data availability statement

All code used in the analyses presented in this study are available at the following repository: https://github.com/sbankier/cortisol_networks/tree/main. Data from the Stockholm Tartu Atherosclerosis Reverse Networks Engineering Task study (STARNET) are available through a database of Genotypes and Phenotypes (dbGaP) application (accession no. phs001203.v2.p1). Gene expression data from The Stockholm Atherosclerosis Gene Expression study (STAGE) and the Metabolic Syndrome in Man study (METSIM) are available publicly at GEO (accession no. GSE70353 and GSE40231, respectively). The summary statistics from the CORNET GWAMA are available at Edinburgh DataShare: https://datashare.ed.ac.uk/handle/10283/3836.

Ethics statement

The studies involving human participants were reviewed and approved by ethical approvals: Tartu, Dnr 154/7 and 188/M-12, Mount Sinai, IRB-20-03781. The patients/participants provided their written informed consent to participate in this study.

Author contributions

SB, TM, and BW contributed to the conception and design of this research. SB conducted all formal analyses and visualizations and wrote the article, supervised by TM, BW, and RA. LW and TM developed and supported the use of and interpretation of outputs from the software Findr. AC contributed to data analysis and interpretation for the CORNET consortium. RM conducted the experiments and contributed to data analysis of dexamethasone-treated mice. AR and JB provided access to and contributed to interpretation of data from the STARNET cohort. All authors reviewed the article and approved the submitted version.

Funding

This work has benefited from UK research and Innovation (UKRI) funding through a Medical Research Council (MRC) PhD studentship (project reference 1938124). Funding has also been provided by the Wellcome Trust (project number 107049/Z/15/Z) and the Norwegian Research Council (NFR) (project number 312045).

Conflict of interest

Authors JB and AR was employed by the company Clinical Gene Networks AB.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1186252/full#supplementary-material

References

1. Walker BR. Glucocorticoids and cardiovascular disease. *Eur J Endocrinol* (2007) 157(5):545–59. doi: 10.1530/EJE-07-0455

2. Raff H, Carroll Ty. Cushing's syndrome: from physiological principles to diagnosis and clinical care. *J Physiol* (2015) 593(3):493–506. doi: 10.1113/jphysiol.2014.282871

3. Newell-Price J, Bertagna X, Grossman AB, Nieman LK. Cushing's syndrome. *Lancet* (2006) 367(9522):1605–17. doi: 10.1016/S0140-6736(06)68699-6

4. Whitworth JA, Brown MA, Kelly JJ, Williamson PM. Mechanisms of cortisol-induced hypertension in humans. *Steroids Aldosterone Hypertension* (1995) 60(1):76–80. doi: 10.1016/0039-128X(94)00033-9

5. Chiodini I, Adda G, Scillitani A, Coletti F, Morelli V, Lembo SD, et al. Cortisol Secretion in Patients With Type 2 Diabetes: Relationship with chronic complications. *Diabetes Care* (2007) 30.1:83–8. doi: 10.2337/dc06-1267

6. Bartels M, Geus, de Eco JC, Kirschbaum C, Sluyter F, Boomsma DI. Heritability of daytime cortisol levels in children. *Behav Genet* (2003) 33.4:421–433. doi: 10.1023/A:1025321609994

7. Bolton JL, Hayward C, Direk N, Lewis JG, Hammond GL, Hill LA, et al. Genome wide association identifies common variants at the *SERPINA6/SERPINA1* locus influencing plasma cortisol and corticosteroid binding globulin. *PloS Genet* (2014) 10.7:e1004474. doi: 10.1371/journal.pgen.1004474

8. Crawford AA, Bankier S, Altmaier E, Barnes CLK, Clark DW, Ermel R, et al. Variation in the SERPINA6/SERPINA1 locus alters morning plasma cortisol, hepatic corticosteroid binding globulin expression, gene expression in peripheral tissues, and risk of cardiovascular disease. *J Hum Genet* (2021), 1–12. doi: 10.1038/s10038-020-00895-6

9. Hammond GL. Molecular properties of corticosteroid binding globulin and the sex-steroid binding proteins*. *Endocrine Rev* (1990) 11.1:65–79. doi: 10.1210/edrv-11-1-65

10. Perogamvros I, Ray DW, Trainer PJ. Regulation of cortisol bioavailability—effects on hormone measurement and action. *Nat Rev Endocrinol* (2012) 8.12:717–27. doi: 10.1038/nrendo.2012.134

11. Pemberton PA, Stein PE, Pepys MB, Potter JM, Carrell RW. Hormone binding globulins undergo serpin conformational change in inflammation. *Nature* (1988) 336:6196. doi: 10.1038/336257a0

12. Chan WL, Carrell RW, Zhou A, Read RJ. How changes in affinity of corticosteroid-binding globulin modulate free cortisol concentration. *J Clin Endocrinol Metab* (2013) 98.8:3315–22. doi: 10.1210/jc.2012-4280

13. Nenke MA, Holmes M, Rankin W, Lewis JG, Torpy DJ. Corticosteroid binding globulin cleavage is paradoxically reduced in alpha-1 antitrypsin deficiency: Implications for cortisol homeostasis. *Clinica Chimica Acta* (2016) 452:27–31. doi: 10.1016/j.cca.2015.10.028

14. Lewis JG, Bagley CJ, Elder PA, Bachmann AW, Torpy DJ. Plasma free cortisol fraction reflects levels of functioning corticosteroid-binding globulin. *Clinica Chimica Acta* (2005) 359.1:189–94. doi: 10.1016/j.cccn.2005.03.044

15. Oakley RH, Cidlowski JA. The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. *J Allergy Clin Immunol* (2013) 132.5:1033–44. doi: 10.1016/j.jaci.2013.09.007

16. Torpy DJ, Bachmann AW, Grice JE, Fitzgerald SP, Phillips PJ, Whitworth JA, et al. Familial corticosteroid-binding globulin deficiency due to a novel null mutation: association with fatigue and relative hypotension. *J Clin Endocrinol Metab* (2001) 86.8:3692–700. doi: 10.1210/jcem.86.8.7724

17. Buss C, Schuelter U, Hesse J, Moser D, Phillips DI, Hellhammer D, et al. Haploinsufficiency of the *SERPINA6* gene is associated with severe muscle fatigue: A *de novo* mutation in corticosteroid-binding globulin deficiency. *J Neural Transm* (2007) 114.5:563–9. doi: 10.1007/s00702-006-0620-5

18. Simard M, Hill LA, Lewis JG, Hammond GL. Naturally occurring mutations of human corticosteroid-binding globulin. *J Clin Endocrinol Metab* (2015) 100.1:E129–39. doi: 10.1210/jc.2014-3130

19. Franzén O, Ermel R, Cohain A, Akers NK, Narzo AD, Talukdar HA, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* (2016) 353.6301:827–30. doi: 10.1126/science.aad6970

20. Talukdar HA, Foroughi Asl H, Jain RK, Ermel R, Ruusalepp A, Franzén O, et al. Cross-tissue regulatory gene networks in coronary artery disease. *Cell Syst* (2016) 2.3:196–208. doi: 10.1016/j.cels.2016.02.002

21. Laakso M, Kuusisto J, Stančáková A, Kuulasmaa T, Pajukanta P, Lusis AJ, et al. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res* (2017) 58.3:481–93. doi: 10.1194/jlr.O072629

22. Wang L, Michoel T. Efficient and accurate causal inference with hidden confounders from genome-transcriptome variation data. *PloS Comput Biol* (2017) 13.8:e1005703. doi: 10.1371/journal.pcbi.1005703

23. Yu C-Y, Mayba O, Lee JV, Tran J, Harris C, Speed TP, et al. Genome-wide analysis of glucocorticoid receptor binding regions in adipocytes reveal gene network involved in triglyceride homeostasis. *PloS One* (2010) 5.12:e15188. doi: 10.1371/journal.pone.0015188

24. Bell RMB, Villalobos E, Nixon M, Miguelez-Crespo A, Murphy L, Fawkes A, et al. Carbonyl reductase 1 amplifies glucocorticoid action in adipose tissue and impairs glucose tolerance in lean mice. *Mol Metab* (2021) 48:101225. doi: 10.1016/j.molmet.2021.101225

25. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature* (2012) 489.7414:101–108. doi: 10.1038/nature11233

26. Matys V, Fricke E, Geffers R, Gößling E, Haubrock M, Hehl R, et al. TRANSFAC ®: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* (2003) 31.1:374–8. doi: 10.1093/nar/gkg108

27. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, and Ma'ayan A ChEA: transcription factor regulation inferred from integrating genome wide ChIP-X experiments. *Bioinformatics* (2010) 26.19:2438–44. doi: 10.1093/bioinformatics/btq466

28. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* (2020) 17.3:261–72. doi: 10.1038/s41592-019-0686-2

29. Waskom ML. seaborn: statistical data visualization. *J Open Source Softw* (2021) 6.60:3021. doi: 10.21105/joss.03021

30. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* (2010) 26.18:2336–7. doi: 10.1093/bioinformatics/btq419

31. Rahimi L, Rajpal A, Ismail-Beigi F. Glucocorticoid-induced fatty liver disease. *Diab Metab Syndr Obesity: Targets Ther* (2020) 13:1133–45. doi: 10.2147/DMSO.S247379

32. Præstholm SM, Correia CM, Grøntved L. Multifaceted control of GR signaling and its impact on hepatic transcriptional networks and metabolism. *English. Front Endocrinol* (2020) 11. doi: 10.3389/fendo.2020.572981

33. Pavlatou MG, Vickers KC, Varma S, Malek R, Sampson M, Remaley AT, et al. Circulating cortisol-associated signature of glucocorticoid-related gene expression in subcutaneous fat of obese subjects. *Obesity* (2013) 21.5:960–7. doi: 10.1002/oby.20073

34. Lee RA, Harris CA, Wang J-C. Glucocorticoid receptor and adipocyte biology. *Nucl receptor Res* (2018) 5. doi: 10.32527/2018/101373

35. DeLigio JT, Lin G, Chalfant CE, Park MA. Splice variants of cytosolic polyadenylation element–binding protein 2 (*CPEB2*) differentially regulate pathways linked to cancer metastasis. *J Biol Chem* (2017) 292.43:17909–18. doi: 10.1074/jbc.M117.810127

36. Etzerodt A, Moestrup S. *CD163* and inflammation: biological, diagnostic, and therapeutic aspects. *Antioxid Redox Signaling* (2013) 18.17:2352–63. doi: 10.1089/ars.2012.4834

37. Gao G, Xie A, Huang S.-C, Zhou A, Zhang J, Herman AM, et al. The role of RBM25/*LUC7L3* in abnormal cardiac sodium channel splicing regulation in human heart failure. *Circulation* (2011) 124.10:1124–31. doi: 10.1161/CIRCULATIONAHA.111.044495

38. Li Y, Ito M, Sun S, Chida T, Nakashima K, Suzuki T, et al. *LUC7L3*/CROP inhibits replication of hepatitis B virus via suppressing enhancer II/basal core promoter activity. *Sci Rep* (2016) 6.1:36741. doi: 10.1038/srep36741

39. Arshad M, Ye Z, Gu X, Wong CK, Liu Y, Li D, et al. *RNF13*, a RING Finger Protein, Mediates Endoplasmic Reticulum Stress induced Apoptosis through the Inositol-requiring Enzyme (IRE1α)/c-Jun NH2-terminal Kinase Pathway. *J Biol Chem* (2013) 288.12:8726–36. doi: 10.1074/jbc.M112.368829

40. Harada H, Kitagawa M, Tanaka N, Yamamoto H, Harada K, Ishihara M, et al. Anti-oncogenic and oncogenic potentials of interferon regulatory factors-1 and -2. *Science* (1993) 259.5097:971–4. doi: 10.1126/science.8438157

41. Chapin WJ, Lenkala D, Mai Y, Mao Y, White SR, Huang RS, et al. Peripheral blood IRF1 expression as a marker for glucocorticoid sensitivity. *Pharmacogenetics Genomics* (2015) 25.3:126–33. doi: 10.1097/FPC.0000000000000116

42. Zhang H, Reilly MP. *LIPA* variants in genome-wide association studies of coronary artery diseases. *Arteriosclerosis Thrombosis Vasc Biol* (2017) 37.6:1015–7. doi: 10.1161/ATVBAHA.117.309344

43. Shang M-M, Talukdar Husain A, Hofmann Jennifer J, Niaudet Colin Asl Hassan Foroughi, Jain Rajeev K, et al. Lim domain binding 2. *Arteriosclerosis Thrombosis Vasc Biol* (2014) 34.9:2068–77. doi: 10.1161/ATVBAHA.113.302709

44. Bujalska IJ, Quinkler M, Tomlinson JW, Montague CT, Smith DM, and Stewart PM Expression profiling of 11beta-hydroxysteroid dehydrogenase type-1 and glucocorticoid target genes in subcutaneous and omental human preadipocytes. *J Mol Endocrinol* (2006) 37.2:327–40. doi: 10.1677/jme.1.02048

45. Lightman SL, Birnie MT, Conway-Campbell BL. Dynamics of ACTH and cortisol secretion and implications for disease. *Endocrine Rev* (2020) 41.3:470–90. doi: 10.1210/endrev/bnaa002

46. Finnerty CC, Mabvuure NT, Ali A, Kozar RA, Herndon DN. The surgically induced stress response. *JPEN. J parenteral enteral Nutr* (2013) 37.50:21S–9S. doi: 10.1177/0148607113496117

47. Pierce BL, Tong L, Chen LS, Rahaman R, Argos M, Jasmine F, et al. Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: A genome-wide analysis among 1,800 South Asians. *PloS Genet* (2014) 10.12:e1004818. doi: 10.1371/journal.pgen.1004818

48. Mangan S, Alon U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci* (2003) 100.21:11980–5. doi: 10.1073/pnas.2133841100

49. Hu X, Li W-P, Meng C, Ivashkiv LB. Inhibition of IFN-γ Signaling by glucocorticoids. *J Immunol* (2003) 170.9:4833–9. doi: 10.4049/jimmunol.170.9.4833

50. Flammer JR, Dobrovolna J, Kennedy MA, Chinenov Y, Glass CK, Ivashkiv LB, et al. The type I interferon signaling pathway is a target for glucocorticoid inhibition. *Mol Cell Biol* (2010) 30.19:4564–74. doi: 10.1128/MCB.00146-10

51. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science* (2002) 297.5584:1183–6. doi: 10.1126/science.1070919

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 04 October 2023
DOI 10.3389/fendo.2023.1185956

OPEN ACCESS

EDITED BY
Tarunveer Singh Ahluwalia,
Steno Diabetes Center Copenhagen (SDCC), Denmark

REVIEWED BY
Anton Terasmaa,
National Institute of Chemical Physics and Biophysics, Estonia
Tajudeen Yahaya,
Federal University, Birnin Kebbi, Nigeria
Edith Hofer,
Medical University of Graz, Austria
Sulev Kõks,
Murdoch University, Australia

*CORRESPONDENCE
Jehad Abubaker
jehad.abubakr@dasmaninstitute.org
Thangavel Alphonse Thanaraj
alphonse.thangavel@dasmaninstitute.org

†These authors have contributed equally to this work

RECEIVED 14 March 2023
ACCEPTED 11 September 2023
PUBLISHED 04 October 2023

CITATION
Hammad MM, Abu-Farha M, Hebbar P, Anoop E, Chandy B, Melhem M, Channanath A, Al-Mulla F, Thanaraj TA and Abubaker J (2023) The miR-668 binding site variant rs1046322 on *WFS1* is associated with obesity in Southeast Asians. *Front. Endocrinol.* 14:1185956. doi: 10.3389/fendo.2023.1185956

The miR-668 binding site variant rs1046322 on *WFS1* is associated with obesity in Southeast Asians

Maha M. Hammad[1,2†], Mohamed Abu-Farha[1†], Prashantha Hebbar[3†], Emil Anoop[4], Betty Chandy[4], Motasem Melhem[4], Arshad Channanath[3], Fahd Al-Mulla[3], Thangavel Alphonse Thanaraj[3*] and Jehad Abubaker[1*]

[1]Department of Biochemistry and Molecular Biology, Dasman Diabetes Institute, Kuwait, Kuwait, [2]Department of Pharmacology and Toxicology, Faculty of Medicine, Kuwait University, Kuwait, Kuwait, [3]Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait, Kuwait, [4]Special Service Facility Department, Dasman Diabetes Institute, Kuwait, Kuwait

The Wolfram syndrome 1 gene (*WFS1)* is the main causative locus for Wolfram syndrome, an inherited condition characterized by childhood-onset diabetes mellitus, optic atrophy, and deafness. Global genome-wide association studies have listed at least 19 *WFS1* variants that are associated with type 2 diabetes (T2D) and metabolic traits. It has been suggested that miRNA binding sites on *WFS1* play a critical role in the regulation of the wolframin protein, and loss of WFS1 function may lead to the pathogenesis of diabetes. In the Hungarian population, it was observed that a 3' UTR variant from *WFS1*, namely rs1046322, influenced the affinity of miR-668 to *WFS1* mRNA, and showed a strong association with T2D. In this study, we genotyped a large cohort of 2067 individuals of different ethnicities residing in Kuwait for the *WFS1* rs1046322 polymorphism. The cohort included 362 Southeast Asians (SEA), 1045 Arabs, and 660 South Asians (SA). Upon performing genetic association tests, we observed significant associations between the rs1046322 SNP and obesity traits in the SEA population, but not in the Arab or SA populations. The associated traits in SEA cohort were body mass index, BMI (β=1.562, P-value=0.0035, P_{emp}=*0.0072*), waist circumference, WC (β=3.163, P-value=0.0197, P_{emp}=*0.0388*) and triglyceride, TGL (β=0.224, P-value=0.0340). The association with BMI remained statistically significant even after multiple testing correction. Among the SEA individuals, carriers of the effect allele at the SNP had significantly higher BMI [mean of 27.63 (3.6) Kg/m^2], WC [mean of 89.9 (8.1) cm], and TGL levels [mean of 1.672 (0.8) mmol/l] than non-carriers of the effect allele. Our findings suggest a role for *WFS1* in obesity, which is a risk factor for diabetes. The study also emphasizes the significant role the ethnic background may play in determining the effect of genetic variants on susceptibility to metabolic diseases.

KEYWORDS

WFS1, ethnicity, obesity, triglycerides, polymorphism, waist circumference

1 Introduction

Obesity has now become a global epidemic with an alarmingly increasing rate of incidence (1). The World Health Organization (WHO) reports that the worldwide prevalence of obesity nearly tripled between 1975 and 2016 and is expected to double in the next 25 years (2, 3). Different ethnicities exhibit different rates of obesity. This is evident in the different rates of obesity in ethnicities such as those of Southeast Asians, Arabs, and South Asians. The Southeast Asian countries have some of the lowest rates of overweight and obesity globally ranging from 2.2 to 15.5%. A recent review of all national surveys for some of the major South Asian countries (including Afghanistan, Bangladesh, India and Sri Lanka) reported that the prevalence of being overweight or obese in adults ranged from 22.4 to 52.4% (4). As for Arabian countries, WHO reported that the prevalence of obesity has a wide range between 4 and 55% (5).

Obesity is a complex multifactorial disorder with several risk factors that contribute to its development. People with obesity not only suffer from a poor quality of life, but are also at risk of developing serious complications, such as diabetes, cardiovascular diseases, sleep disorders, or hypertension (1). Therefore, it is important to understand the underlying multifactorial causes of obesity. Although environmental factors and lifestyle practices are the main causes of obesity, genetic susceptibility also plays a very significant role. Several reports confirm the association of certain genes with obesity, fat distribution, energy expenditure, and appetite regulation. Studies have shown that 40–70% of the variation in body mass index (BMI) among individuals can be attributed to genetic factors (6, 7). Furthermore, at least 200 genetic variants have been reported to be associated with obesity in several populations; however, such studies have focused on Caucasians from Europe (8, 9). Ethnicity can play an important role in determining the genetic susceptibility of an individual to obesity (10, 11).

Wolfram syndrome 1 gene (*WFS1*) was identified in the year of 1998 on chromosome 4p16 as a novel gene that causes a rare autosomal recessive neurodegenerative disorder, namely Wolfram syndrome (WFS) (12) *alias* DIDMOAD syndrome (diabetes insipidus, diabetes mellitus, optic atrophy, and deafness). It is clinically characterized by the juvenile-onset of diabetes mellitus as the main symptom in the early stage of the disease, and by bilateral progressive optic atrophy in later stages (13–15). The *WFS1* gene encodes wolframin, a protein present in the membrane of the endoplasmic reticulum (ER), and is mainly detected in certain brain regions as well as in pancreatic β-cells and the heart (16, 17). Several *WFS1* polymorphic variants have been associated with the risk of developing diabetes mellitus (18, 19). Among these variants, two microRNA-single-nucleotide polymorphisms (miR-SNPs), rs1046322 and rs9457, were strongly associated with both type 1 and type 2 diabetes, and this association remained statistically significant after applying multiple corrections (19). Considering that obesity is a risk factor for diabetes associated with insulin resistance, we aimed to evaluate, in the present study, the association of these two *WFS1* variants with obesity in individuals of different ethnicities.

2 Materials and methods

2.1 Study population

This study included a cohort of 2067 participants residing in Kuwait. Upon enrolment, we recorded the following information: age, sex, baseline characteristics (height, weight, waist circumference (WC)), and underlying diagnosed disorders (such as diabetes). The study protocol was reviewed and approved by the Ethical Review Committee of Dasman Diabetes Institute and was conducted in accordance with the guidelines of the Declaration of Helsinki and the US Federal Policy for the Protection of Human Subjects. All participants signed an informed consent form before participating in the study. The ethnicity of each subject was defined via self-reporting and was confirmed through detailed questioning on parental lineage up to three generations.

2.2 Sample processing

We collected blood samples in accordance with established institutional guidelines. After confirming that the participant was under an overnight fast, we collected blood samples in the morning, between 8 and 11 am. We performed DNA extraction using a Gentra Puregene kit (Qiagen, Valencia, CA, USA) and assessed quantification using Quant-iT PicoGreen dsDNA Assay Kits (Life Technologies, Grand Island, NY, USA) and an Epoch Microplate Spectrophotometer (BioTek Instruments). We checked absorbance values at 260–280 nm for adherence to an optical density range of 1.8–2.1.

2.3 Anthropometric measurements and blood biochemistry

The BMI of each participant was calculated as the ratio of their weight (Kg) to height (m) squared. We assessed lipid profiles, including triglyceride (TGL), low density lipoprotein (LDL), high density lipoprotein (HDL), and total cholesterol (TC) levels, using a Siemens Dimension RXL integrated chemistry analyzer (Diamond Diagnostics, Holliston, MA, USA).

2.4 Genotyping

We performed candidate SNP genotyping using the TaqMan Genotyping Assay on an ABI 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). We set the polymerase chain reaction (PCR) sample with 10 ng of DNA, 5 × FIREPol Master Mix (Solis BioDyne, Estonia), and 1 µl of 20 × TaqMan SNP Genotyping Assay. We set thermal cycling conditions at 60°C for 1 min and 95°C for 15 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. We used Sanger sequencing to validate certain selected cases of homozygous and heterozygous genotypes using a BigDye Terminator v3.1 Cycle Sequencing Kit on a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

2.5 Quality assessment of the rs1046322 and rs9457 SNPs

We assessed the quality and statistical association of the rs1046322 and rs9457 SNPs using the PLINK genome association analysis toolset (version 1.9). Next, for quality assessments, we determined minor allele frequency (MAF) and consistency with the Hardy–Weinberg equilibrium for the *WFS1* variants.

2.6 Statistical analysis

Data are presented as mean± standard deviation (SD). We determined statistical significance using Student's t-test for quantitative variables and Fisher's exact test for categorical variables, and P values ≤ 0.05 were considered significant. We assessed allele-based associations between the rs1046322 variant and the quantitative traits (BMI, TGL, and WC) using genetic models based on additive mode of inheritance (GG versus GA versus AA) adjusted for the confounders of age, sex and diabetes status. We assessed changes in the mean of phenotype measurement using regression coefficient (Beta), where a positive regression coefficient indicated that the minor allele increases the risk effect. Multiple comparisons were corrected by generating empirical P values (P_{emp}) using the max(T) permutation procedure available in PLINK, based on 10,000 permutations. A threshold of < 0.05 was set for both the *P* value and P_{emp} value to assess the statistical significance of the association signal. Any quantitative trait value lesser than Q1–1.5 × the interquartile range (IQR) or higher than Q3 + 1.5 × IQR was considered to be an outlier and was excluded from the statistical analyses. Statistical analyses were performed using PLINK, version 1.9, and R software, version 4.0.2.

3 Results

3.1 Characteristics of the study participants and genotyping data

The average rate of successful genotyping of the two SNPs rs1046322 and rs9457 in each of the three subpopulations, namely Arab, South Asians, and Southeast Asians was > 99%, and the SNP was within the Hardy–Weinberg equilibrium. Of the two SNPs, the rs9457 did not show significant associations with any of the examined obesity traits in any of the three ethnic cohorts, though an association is seen with HDL in the Arab cohort with a P value of 0.0334 *albeit* with an insignificant empirical P (P_{emp}) value (Supplementary Table S1; the allele and genotype frequencies for the variant are listed in the notes to the table). Thus, rs9457 was not included in further analyses.

The frequencies of the minor allele (A) at the *WFS1* rs1046322 SNP in the Arab, South Asian, and Southeast Asian populations were 17.27%, 11.67%, and 6.77%, respectively. Of the 362 genotyped samples from the Southeast Asian population, 314 (86.7%) were homozygous for G, only 1 (0.3%) was homozygous for A, and 47 (13%) were heterozygous (GA). Of the 1045 genotyped samples from the Arab population, 715 (68.4%) were homozygous for G, 31 (3%) were homozygous for A, and 299 (28.6%) were heterozygous (GA). Of the 660 genotyped samples from the South Asian population, 515 (78%) were homozygous for G, 9 (1.4%) were homozygous for A, and 136 (20.6%) were heterozygous (GA).

The mean (SD) age of participants in the Southeast Asian cohort was 41.1 (9.4) years, in the Arab cohort was 47.6 (11.6) years and in the South Asian cohort was 42.9 (9.7) years. Thus, the participants in each of the three ethnic cohorts are uniformly largely middle-aged. Table 1 presents a genotype-wide distribution (GG versus (GA+AA)) at rs1046322 of the characteristics of the three ethnic cohorts. After examining the genotype-specific differences in the phenotypic traits, we observed statistically significant differences in the Southeast Asian population (Table 1). The phenotypic traits that showed statistically significant genotypic differences in the Southeast Asian population were as follows: (i) Mean (SD) of BMI was significantly higher in the participants that harbored the A allele as compared to non-carriers [27.63 (3.6) Kg/m^2 vs. 26.02 (3.6) Kg/m^2; P = 0.004] (Figure 1A and Table 1); (ii) WC of the SNP carriers was significantly higher than that of the non-carriers [89.9 (8.1) cm vs. 85.9 (9.6) cm; P = 0.006] (Figure 1B and Table 1); and (iii) Participants that harbored the A allele had higher TGL as compared to non-carriers [1.672 (0.8) mmol/l vs. 1.43 (0.7) mmol/l; P = 0.037] (Figure 1C and Table 1).

3.2 Association between *WFS1* rs1046322 and obesity-related markers

The association tests for the variant and obesity markers showed that obesity traits were significantly associated with the *WFS1* rs1046322 variant only in the SEA subpopulation, and not in the Arab or South Asian populations (Table 2). The traits showing statistically significant differences included BMI (β:1.562, P = 0.0035), WC (β:3.163, P = 0.0197), and TGL (β:0.224, P = 0.034). Further, the associations with BMI and WC also exhibited significant empirical P_{emp}values of 0.0072 and 0.0388, respectively establishing the BMI and WC as strong contenders for association with the SNP.

3.3 P value threshold after correction for multiple testing

Having found association of the SNP with triglyceride at a P value of 0.034, we investigated the associations of the SNP with the cholesterol traits of HDL, LDL and total cholesterol. We found that the associations for these cholesterol traits with the SNP were insignificant (Supplementary Table S2). As regards multiple testing for considering the cholesterol traits along with triglycerides and BMI and WC, it is to be noted that not all the cholesterol traits are independent of each other. In our earlier

10.3389/fendo.2023.1185956

TABLE 1 Overview of the Southeast Asian, Arab, and South Asian populations as per genotype distribution of the *WFS1* rs1046322 variant.

	Southeast Asians (SEA) MAF = 6.77%			Arabs (Arab) MAF = 17.27%			South Asians (SA) MAF = 11.67%		
Trait	GG	GA+AA	P Value	GG	GA+AA	P Value	GG	GA+AA	P Value
Distribution*	314 (86.7)	47 + 1 (13 + 0.3)		715 (68.4)	299 + 31 (28.6 + 3)		515 (78.0)	136 + 9 (20.6 + 1.4)	
Sex* (M:F)	109: 205 (34.7: 65.3)	16: 32 (33.3: 66.7)		393: 322 (54.97: 45.03)	157: 173 (47.58: 52.42)		376: 139 (73: 27)	101: 44 (69.66: 30.34)	
Age$ (years)	40.59 (9.351)	44.6 (8.997)		48.2 (11.24)	46.35 (12.19)		42.83 (9.935)	43.15 (8.861)	
Diabetes* (NO : YES)	265: 49 (84.4: 15.6)	37: 11 (77.1: 22.9)	0.213	437: 278 (61.1: 38.9)	203: 127 (61.5: 38.5)	0.946	352: 163 (68.3: 31.7)	100: 45 (69: 31)	0.9197
Obesity* (NO : YES)	261: 53 (83.1: 16.9)	37: 11 (77.1: 22.9)	0.312	317: 396 (44.5: 55.5)	149: 181 (45.2: 54.8)	0.841	405: 109 (78.8: 21.2)	105: 40 (72.4: 27.6)	0.116
BMI$ (Kg/m^2)	26.02 (3.6)	27.63 (3.6)	**0.004**	31.26 (5.5)	30.87 (5.469)	0.288	26.73 (3.841)	27.34 (3.939)	0.099
Waist Circumference$ (cm)	85.9 (9.6)	89.9 (8.1)	**0.006**	101.3 (12.06)	100.6 (11.97)	0.407	91.8 (9.04)	92.98 (9.2)	0.188
TGL$ (mmol/l)	1.43 (0.7)	1.672 (0.8)	**0.037**	1.396 (0.6)	1.389 (0.7)	0.868	1.419 (0.64)	1.4 (0.63)	0.756
TC$ (mmol/l)	5.41 (0.98)	5.61 (1.1)	0.208	4.998 (1.007)	5.017 (0.9873)	0.788	5.189 (0.983)	5.172 (0.917)	0.849
LDL$ (mmol/l)	3.39 (0.9)	3.59 (0.9)	0.133	3.16 (0.9)	3.18 (0.88)	0.807	3.42 (0.9)	3.39 (0.8)	0.7301
HDL$ (mmol/l)	1.29 (0.3)	1.209 (0.3)	0.114	1.141 (0.3)	1.138 (0.312)	0.877	1.082 (0.242)	1.062 (0.262)	0.404

*Number (%), $Mean (SD), MAF, minor allele frequency; SD, standard deviation; IQR, interquartile range; BMI, body mass index; TGL, triglyceride; TC, total cholesterol; LDL, low-density lipoprotein cholesterol; HDL, high-density lipoprotein cholesterol.
Statistically significant data are bolded.

work (20), we queried independent variables among the four lipid traits by way of performing Pearson correlation analysis between the traits followed by matSpD analysis (21) (http://neurogenetics.qimrberghofer.edu.au/matSpD/), and found that the estimated effective number of independent traits among the lipid traits as 3. Upon including the WC and BMI as independent traits, we have a total of 5 independent traits tested for associations with the SNP. The P value threshold after multiple testing correction for significant associations turns out to be 0.01 (=0.05/5). Thus, after multiple testing correction, only the BMI association with the rs1046322 SNP (at a P value of 0.0035) remains significant.

4 Discussion

In this study, the *WFS1* rs1046322 and rs9457 were assessed for association with obesity in the ethnic populations of Arabs, South Asians and Southeast Asians. The rs9457 SNP did not exhibit any

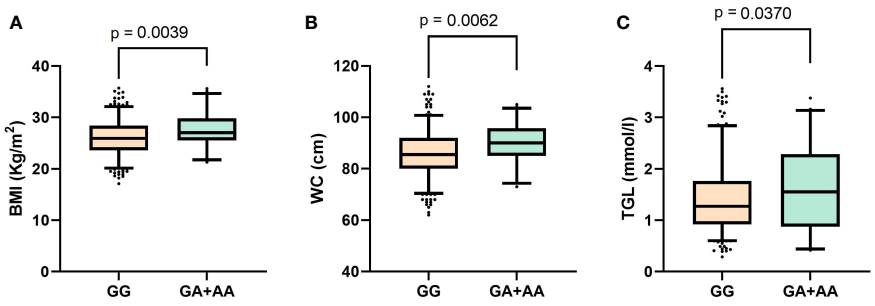


FIGURE 1
Boxplots displaying data distribution for the phenotype traits **(A)** Body mass index, **(B)** Waist circumference, **(C)** Triglyceride levels in Southeast Asian individuals with genotypes (GA+AA) containing the effect allele or homozygous (GG) genotypes for reference alleles of the *WFS1* rs1046322 variant.

significant associations with obesity traits. However, the results demonstrated that *WFS1* rs1046322 is significantly associated with BMI and WC in the Southeast Asian population, but not in the Arab or South Asian populations. The SNP is also associated with TGL levels only in the Southeast Asian population, and not the other two subpopulations. Furthermore, carriers of the SNP had significantly higher BMI, WC, and TGL levels, as compared to non-carriers in the Southeast Asian subpopulation alone.

The 1000 Genome Project reported a MAF of 8% for the *WFS1* rs1046322 variant. However, this SNP varies in its frequency across different populations, ranging from 7.2% in Chileans to 50% in Siberians. Although MAFs in the subpopulations of the present study fall within the lower range, we observed some differences among the three groups, with the Southeast Asian population showing the lowest MAF (6.8%), followed by the South Asian (11.7%) and Arab (17.3%) populations.

Upper body, or truncal, obesity is strongly associated with obesity-related complications, such as diabetes and cardiovascular diseases. Considering that WC is a measure of truncal obesity, it is interesting to note the effect sizes of the associations between the variant and WC and BMI (Table 2). The observed effect sizes in the SEA population indicate that the effect allele in *WFS1* rs1046322 resulted in an increase by 3.163 cm in WC and by 1.562 Kg/m^2 in BMI. Interestingly, a recent *in vitro* study by Ivask et al. (22) examined *WFS1* heterozygous mouse model for response to high fat diet (HFD) in terms of body weight and metabolic characteristics. The authors found that the impaired body weight gain found in *WFS1* mutant mice is prevented by HFD. They further observed that in *WFS1* heterozygous mutant mice, HFD impaired the normalized insulin secretion and the expression of endoplasmic reticulum (ER) stress genes in isolated pancreatic islets. HFD increased the expression of *Ire1α* and *Chop* in pancreas and decreased the expression of *Ire1α* and *Atf4* in liver from these mutant mice. The authors concluded that quantitative *WFS1* gene deficiency predisposes carriers of single functional *WFS1* copy to diabetes and metabolic syndrome and makes them susceptible to environmental factors such as HFD.

There is a lack of literature reports on associations between *WFS1* and obesity. However, previous studies have linked several *WFS1* SNPs with type 2 diabetes and biomarkers related to diabetes across various ethnicities, including the United Kingdom population, Swedish population and Ashkenazi population (23–26). One of the recently published large studies that confirmed the association between *WFS1* and type 2 diabetes included 81,412 type 2 diabetes patients and 370,832 healthy individuals of diverse ancestries (27). Furthermore, the DESIR (Data from Epidemiological Study on the Insulin Resistance Syndrome) prospective study demonstrated in French cohorts that allelic variations at three SNPs in the *WFS1* gene were associated with incident type 2 diabetes (28).

Considering the established role of wolframin as an ER stress regulator that negatively regulates ER stress signaling, discovering a link between the gene and obesity does not come as a surprise. Additional studies are required to further investigate the mechanism for this regulation. Given that wolframin plays a key role in mediating the ER export of vesicular cargo proteins, it could be speculated that it regulates the processing and release of different gut hormones or melanocortin hormones in the brain, similar to its role in regulating proinsulin cleavage and insulin secretion (29). In addition, a recent study demonstrated that WFS1 regulates anti-inflammatory responses in pancreatic β-cells. Specifically, the study reported that the pancreatic islets of WFS1 whole-body knockout mice display M1-macrophage infiltration and hypervascularization (30).

WFS1 rs1046322 is a 3' UTR variant and is a putative miRNA (miR-668) binding site polymorphism. This variant was previously shown to influence the affinity of miR-668 to *WFS1* mRNA (31). Though there exist no previous studies investigating the role of miR-668 in obesity, it has been recently shown that miR-668-3p can suppress mediators of inflammation and oxidative stress (32). Therefore, it would be interesting to examine the effect of this

TABLE 2 Association tests for the *WFS1* rs1046322 variant (A as the effect allele) with the phenotypic traits of BMI, WC and TGL using genetic models based on additive mode of inheritance (GG versus GA versus AA).

	Southeast Asians				Arabs				South Asians			
Trait	Sample Size	Effect Size (β value) [95% CI]	P Value	P_{emp} Value	Sample Size	Effect Size (β value) [95% CI]	P Value	P_{emp} Value	Sample Size	Effect Size (β value) [95% CI]	P Value	P_{emp} Value
BMI	351	1.562 [0.52, 2.60]	**0.0035**	**0.0072**	1017	0.224 [0.016, 0.432]	0.1905	0.333	640	0.467 [-0.19, 1.12]	0.1609	0.2868
WC	354	3.163 [0.52, 5.81]	**0.0197**	**0.0388**	869	-0.583 [-1.99, 0.82]	0.4162	0.6534	636	0.911 [-0.60, 2.42]	0.2380	0.4229
TGL	345	0.224 [0.02, 0.43]	**0.0340**	0.0725	857	-0.006 [-0.08, 0.07]	0.8847	0.9877	619	-0.014 [-0.12, 0.095]	0.7996	0.9595

[95% CI], 95% confidence intervals; P_{emp}, empirical P values; BMI, body mass index; WC, waist circumference; TGL, triglyceride.
The models were corrected for the confounders of age, sex and diabetes status.
Statistically significant data are bolded.

miRNA in obesity and its expression level in participants who suffer from obesity or metabolic syndrome.

In conclusion, the findings of the present study suggest an ethnic-specific role for *WFS1* in obesity. While the current study included a large cohort with three different ethnic populations, further studies would benefit by examining the observed associations in more diverse ethnic populations. The study also highlights the importance of including ethnic groups that are under-represented in current global genetic studies of genotype–phenotype associations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The study protocol was reviewed and approved by the Ethical Review Committee of Dasman Diabetes Institute. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MH, MA-F, TT and JA contributed to conception and design of the study. PH and AC organized the database and performed the statistical analysis. MH and PH wrote the first draft of the manuscript. MA-F and TT wrote sections of the manuscript. EA, BC and MM performed the assays. TAT, JA and FA-M revised and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Kuwait Foundation for the Advancement of Sciences (KFAS; research project RA HM-2018-039).

Acknowledgments

We would like to thank the National Dasman Diabetes Biobank Core Facility at DDI for their contribution in sample processing. We are also indebted to Kuwait Foundation for the Advancement of Sciences (KFAS) for providing financial support for this research project (RA HM-2018-039). The funding agency was not involved in data collection, analysis, or interpretation; trial design; patient recruitment; or any aspect pertinent to the study. Lastly, we would like to thank Ms. Lubaina Koti for editing the manuscript for language, structure, and accuracy.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1185956/full#supplementary-material

SUPPLEMENTARY TABLE 1
Association tests for the *WFS1* rs9457 variant (G as the effect allele) with phenotypic traits using genetic models based on additive mode of inheritance (CC versus GC versus GG). The model was adjusted for the confounders of age, sex and diabetes status.

SUPPLEMENTARY TABLE 2
Association tests for the *WFS1* rs1046322 variant (A as the effect allele) with phenotypic traits (including cholesterol traits) using genetic models based on additive mode of inheritance (GG versus GA versus AA). The model was adjusted for the confounders of age, sex, and diabetes status.

References

1. Hruby A, Hu FB. The epidemiology of obesity: A big picture. *PharmacoEconomics* (2015) 33(7):673–89. doi: 10.1007/s40273-014-0243-x

2. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* (2014) 384(9945):766–81. doi: 10.1016/S0140-6736(14)60460-8

3. Kelly T, Yang W, Chen CS, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes* (2008) 32(9):1431–7. doi: 10.1038/ijo.2008.102

4. Awasthi A, Panduranga AB, Deshpande A. Prevalence of overweight/obesity in South Asia: A narrative review. *Clin Epidemiol Global Health* (2023) 22:101316. doi: 10.1016/j.cegh.2023.101316

5. Badran M, Laher I. Obesity in arabic-speaking countries. *J Obes* (2011) 2011:686430. doi: 10.1155/2011/686430

6. Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. *Nat Rev Genet* (2022) 23(2):120–33. doi: 10.1038/s41576-021-00414-z

7. Loos RJ. Recent progress in the genetics of common obesity. *Br J Clin Pharmacol* (2009) 68(6):811–29. doi: 10.1111/j.1365-2125.2009.03523.x

8. El-Sayed Moustafa JS, Froguel P. From obesity genetics to the future of personalized obesity therapy. *Nat Rev Endocrinol* (2013) 9(7):402–13. doi: 10.1038/nrendo.2013.57

9. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* (2010) 42(11):937–48. doi: 10.1038/ng.686

10. Albuquerque D, Nóbrega C, Manco L, Padez C. The contribution of genetics and environment to obesity. *Br Med Bull* (2017) 123(1):159–73. doi: 10.1093/bmb/ldx022

11. Byrd AS, Toth AT, Stanford FC. Racial disparities in obesity treatment. *Curr Obes Rep* (2018) 7(2):130–8. doi: 10.1007/s13679-018-0301-3

12. Inoue H, Tanizawa Y, Wasson J, Behn P, Kalidas K, Bernal-Mizrachi E, et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *NatGenet* (1998) 20(2):143–8. doi: 10.1038/2441

13. Fraser FC, Gunn T. Diabetes mellitus, diabetes insipidus, and optic atrophy. An autosomal recessive syndrome? *J Med Genet* (1977) 14(3):190–3. doi: 10.1136/jmg.14.3.190

14. Richardson JE, Hamilton W. Diabetes insipidus, diabetes mellitus, optic atrophy, and deafness. 3 cases of (DIDMOAD) syndrome. *Arch Dis Childhood* (1977) 52 (10):796. doi: 10.1136/adc.52.10.796

15. Philbrook C, Fritz E, Weiher H. Expressional and functional studies of Wolframin, the gene function deficient in Wolfram syndrome, in mice and patient cells. *Exp Gerontol* (2005) 40(8):671–8. doi: 10.1016/j.exger.2005.06.008

16. Rigoli L, Bramanti P, Di Bella C, De Luca F. Genetic and clinical aspects of Wolfram syndrome 1, a severe neurodegenerative disease. *Pediatr Res* (2018) 83 (5):921–9. doi: 10.1038/pr.2018.17

17. Fonseca SG, Ishigaki S, Oslowski CM, Lu S, Lipson KL, Ghosh R, et al. Wolfram syndrome 1 gene negatively regulates ER stress signaling in rodent and human cells. *J Clin Invest* (2010) 120(3):744–55. doi: 10.1172/JCI39678

18. Fawcett KA, Wheeler E, Morris AP, Ricketts SL, Hallmans G, Rolandsson O, et al. Detailed investigation of the role of common and low-frequency WFS1 variants in type 2 diabetes risk. *Diabetes* (2010) 59(3):741–6. doi: 10.2337/db09-0920

19. Elek Z, Nemeth N, Nagy G, Nemeth H, Somogyi A, Hosszufalusi N, et al. MicroRNA binding site polymorphisms in the WFS1 gene are risk factors of diabetes mellitus. *PloS One* (2015) 10(10):e0139519. doi: 10.1371/journal.pone.0139519

20. Hebbar P, Nizam R, Melhem M, Alkayal F, Elkum N, John SE, et al. Genome-wide association study identifies novel recessive genetic variants for high TGs in an Arab population. *J Lipid Res* (2018) 59(10):1951–66. doi: 10.1194/jlr.P080218

21. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* (2005) 95(3):221–7. doi: 10.1038/sj.hdy.6800717

22. Ivask M, Volke V, Raasmaja A, Koks S. High-fat diet associated sensitization to metabolic stress in Wfs1 heterozygous mice. *Mol Genet Metab* (2021) 134(1-2):203–11. doi: 10.1016/j.ymgme.2021.07.002

23. Minton JA, Hattersley AT, Owen K, McCarthy MI, Walker M, Latif F, et al. Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. *Diabetes* (2002) 51(4):1287–90. doi: 10.2337/diabetes.51.4.1287

24. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* (2007) 39(8):951–3. doi: 10.1038/ng2067

25. Franks PW, Rolandsson O, Debenham SL, Fawcett KA, Payne F, Dina C, et al. Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. *Diabetologia* (2008) 51(3):458–63. doi: 10.1007/s00125-007-0887-6

26. Wasson J, Permutt MA. Candidate gene studies reveal that the WFS1 gene joins the expanding list of novel type 2 diabetes genes. *Diabetologia* (2008) 51(3):391–3. doi: 10.1007/s00125-007-0920-9

27. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet* (2018) 50(4):559–71. doi: 10.1038/s41588-018-0084-1

28. Cheurfa N, Brenner GM, Reis AF, Dubois-Laforgue D, Roussel R, Tichet J, et al. Decreased insulin secretion and increased risk of type 2 diabetes associated with allelic variations of the WFS1 gene: the Data from Epidemiological Study on the Insulin Resistance Syndrome (DESIR) prospective study. *Diabetologia* (2011) 54(3):554–62. doi: 10.1007/s00125-010-1989-0

29. Wang L, Liu H, Zhang X, Song E, Wang Y, Xu T, et al. WFS1 functions in ER export of vesicular cargo proteins in pancreatic beta-cells. *Nat Commun* (2021) 12 (1):6996. doi: 10.1038/s41467-021-27344-y

30. Morikawa S, Blacher L, Onwumere C, Urano F. Loss of function of WFS1 causes ER stress-mediated inflammation in pancreatic beta-cells. *Front Endocrinol (Lausanne)* (2022) 13:849204. doi: 10.3389/fendo.2022.849204

31. Kovacs-Nagy R, Elek Z, Szekely A, Nanasi T, Sasvari-Szekely M, Ronai Z. Association of aggression with a novel microRNA binding site polymorphism in the wolframin gene. *Am J Med Genet Part B: Neuropsychiatr Genet* (2013) 162(4):404–12. doi: 10.1002/ajmg.b.32157

32. Gao Z, Gao Q, Lv X. MicroRNA-668-3p protects against oxygen-glucose deprivation in a rat H9c2 cardiomyocyte model of ischemia-reperfusion injury by targeting the stromal cell-derived factor-1 (SDF-1)/CXCR4 signaling pathway. *Med Sci Monitor Int Med J Exp Clin Res* (2020) 26:e919601. doi: 10.12659/MSM.919601

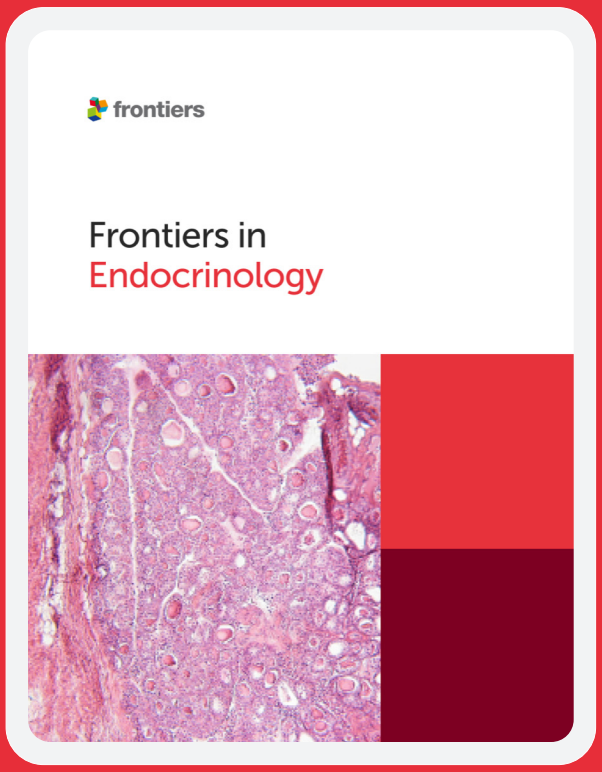
Frontiers in Endocrinology

Explores the endocrine system to find new therapies for key health issues

The second most-cited endocrinology and metabolism journal, which advances our understanding of the endocrine system. It uncovers new therapies for prevalent health issues such as obesity, diabetes, reproduction, and aging.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact