Metabolism 178 (2026) 156552

Contents lists available at ScienceDirect

Metabolism

journal homepage: www.journals.elsevier.com/metabolism

A biological-systems-based analysis using proteomic and metabolic network inference reveals mechanistic insights into hepatic steatosis

Natalie N. Atabaki [a,b,c,1,*], Daniel E. Coral [a], Hugo Pomares-Millan [a,d], Kieran Smith [c], Harry H. Behjat [e], Robert W. Koivula [c], Andrea Tura [f], Hamish Miller [c,g], Katherine E. Pinnick [c], Leandro Z. Agudelo [h,i], Kristine H. Allin [b], Andrew A. Brown [j], Elizaveta Chabanova [k], Piotr J. Chmura [l], Ulrik P. Jacobsen [l], Adem Y. Dawed [j], Petra J.M. Elders [m], Juan J. Fernandez-Tajes [a], Ian M. Forgie [j], Mark Haid [n], Tue H. Hansen [b,o], Angus G. Jones [p], Tarja Kokkola [q], Sebastian Kalamajski [a], Anubha Mahajan [r], Timothy J. McDonald [s], Donna McEvoy [t], Mirthe Muilwijk [u,v], Konstantinos D. Tsirigos [w], Jagadish Vangipurapu [q,x], Sabine van Oort [m], Henrik Vestergaard [b,y], Jerzy Adamski [z,aa,ab], Joline W. Beulens [u], Søren Brunak [l], Emmanouil T. Dermitzakis [ac], Giuseppe N. Giordano [a], Ramneek Gupta [ad,ae], Torben Hansen [b], Leen M. 't Hart [u,v,af,ag], Andrew T. Hattersley [p], Leanne Hodson [c], Markku Laakso [q], Ruth J.F. Loos [b,ah], Jordi Merino [b], Mattias Ohlsson [ai], Oluf Pedersen [aj], Martin Ridderstråle [ak], Hartmut Ruetten [al], Femke Rutters [u,v], Jochen M. Schwenk [am], Jeremy Tomlinson [c], Mark Walker [an], Hanieh Yaghootkar [ao,ap], Fredrik Karpe [c], Mark I. McCarthy [aq], Elizabeth Louise Thomas [ar], Jimmy D. Bell [ar], Andrea Mari [f], Imre Pavo [as], Ewan R. Pearson [j], Ana Viñuela [j], Paul W. Franks [a,at,1,*]

[a] Department of Clinical Science, Genetic and Molecular Epidemiology, Lund University Diabetes Centre, Helsingborg, Sweden
[b] Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[c] The National Institute for Health and Care Research (NIHR) Oxford Biomedical Research Centre (BRC), OCDEM, Radcliffe Department of Medicine, University of Oxford, Oxford, UK
[d] Department of Epidemiology, Geisel School of Medicine at Dartmouth, Dartmouth-Hitchcock Medical Center, Lebanon, USA
[e] Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Lund, Sweden
[f] Institute of Neuroscience, National Research Council, Padova, Italy
[g] Barts Liver Centre, Queen Mary University London and Barts Health NHS Trust, London, UK
[h] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, USA
[i] Broad Institute of MIT and Harvard, Cambridge, USA
[j] Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, Scotland, UK
[k] Department of Diagnostic Radiology, Copenhagen University Hospital Herlev Gentofte, Herlev, Denmark
[l] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark
[m] Department of General Practice, Amsterdam UMC- location Vumc, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands
[n] Metabolomics and Proteomics Core, German Research Center for Environmental Health, Helmholtz Zentrum München, Neuherberg, Germany
[o] Department of Medicine, Zealand University Hospital, Køge, Denmark
[p] Department of Clinical and Biomedical Sciences, University of Exeter College of Medicine & Health, Exeter, UK
[q] Internal Medicine, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland
[r] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[s] Blood Sciences, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
[t] Diabetes Research Network, Royal Victoria Infirmary, Newcastle upon Tyne, UK
[u] Amsterdam UMC, Epidemiology and Data Science, Amsterdam, the Netherlands
[v] Amsterdam Public Health, Health Behaviors & Chronic Diseases, Amsterdam, the Netherlands
[w] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark
[x] A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland

* Corresponding authors at: Department of Clinical Sciences, Lund University, Svartbrödragränden 3-5, Almahuset, Helsingborg Hospital, 251 87, Helsingborg, Sweden.
E-mail addresses: natalie.atabaki@med.lu.se (N.N. Atabaki), paul.franks@med.lu.se (P.W. Franks).

https://doi.org/10.1016/j.metabol.2026.156552
Received 30 June 2025; Accepted 28 January 2026
Available online 6 February 2026

[y] Steno Diabetes Center, Copenhagen, Denmark
[z] Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[aa] Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[ab] Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
[ac] Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
[ad] Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark
[ae] Disease Intelligence Pte Ltd, Singapore, Singapore
[af] Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands
[ag] Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
[ah] The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[ai] Center for Environmental and Climate Science, Lund University, Lund, Sweden
[aj] Center for Clinical Metabolic Research, Herlev and Gentofte University Hospital, Copenhagen, Denmark
[ak] Department of Clinical Science, Lund University, Malmö, Sweden
[al] Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim, Germany
[am] Science for Life Laboratory, Department of Protein Science, KTH Royal Institute of Technology, Stockholm, Sweden
[an] Translational and Clinical Research Institute, Faculty of Medical Sciences, University of Newcastle, Newcastle upon Tyne, UK
[ao] School of Electronics and Computer Science, Faculty of Engineering and Physical Science, University of Southampton, Southampton, UK
[ap] School of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK
[aq] Wellcome Centre for Human Genetics, University of Oxford, UK
[ar] Research Centre for Optimal Health, School of Life Sciences, University of Westminster, London, UK
[as] Eli Lilly Regional Operations Ges.m.b.H., Vienna, Austria
[at] Precision Healthcare University Research Institute, Queen Mary University of London, London, UK

ARTICLE INFO

Keywords:
Basal insulin secretion
Hepatic steatosis
MASLD
Type 2 diabetes
Proteomics
Bayesian networks
Mendelian randomization

ABSTRACT

Objective: To delineate organ-specific and systemic drivers of metabolic dysfunction-associated steatotic liver disease (MASLD), we applied integrative causal inference across clinical, imaging, and proteomic domains in individuals with and without type 2 diabetes (T2D).
Methods: Bayesian network analyses and complementary two-sample Mendelian randomization were used to quantify causal pathways linking adipose distribution, glycemia, and insulin dynamics with liver fat in the IMI-DIRECT prospective cohort study. Data included frequently sampled metabolic challenge tests, MRI-derived abdominal and hepatic fat content, serological biomarkers, and Olink plasma proteomics from 331 adults with new-onset T2D and 964 adults without diabetes, with harmonized protocols enabling replication.
Results: High basal insulin secretion rate (BasalISR), estimated via C-peptide deconvolution, emerged as the primary potential causal driver of liver fat accumulation in both cohorts. BasalISR, a clearance-independent measure of β-cell insulin output distinct from peripheral insulin levels, was independently linked to hepatic steatosis. Visceral adipose tissue exhibited bidirectional associations with liver fat, suggesting a self-reinforcing metabolic loop. Of 446 analyzed proteins, 34 mapped to these metabolic networks (27 in the non-diabetes network, 18 in the T2D network, and 11 shared). Key proteins directly associated with liver fat included GUSB, ALDH1A1, LPL, IGFBP1/2, CTSD, HMOX1, FGF21, AGRP, and ACE2. Sex-stratified analyses identified GUSB in females and LEP in males as the strongest protein predictors of liver fat.
Conclusions: BasalISR may better capture early β-cell-driven disturbances contributing to MASLD. These findings outline a multifactorial, sex- and disease stage–specific proteo-metabolic architecture of hepatic steatosis and identify potential biomarkers or therapeutic targets.

1. Introduction

About a third of the global adult population is estimated to have metabolic dysfunction-associated steatotic liver disease (MASLD), which is predicted to increase markedly in the coming decades [1]. In people with type 2 diabetes (T2D), the prevalence of MASLD is about 67% and rising [2]. Although the epidemiological association between intrahepatic lipid accumulation and T2D is well described, its underlying metabolic and proteomic features remain poorly understood, hindering the prevention and treatment of MASLD [3].

Hyperinsulinemia, especially excessive early-phase insulin secretion, is a common prelude to T2D and MASLD [4,5], and the hypersecretion of insulin often seen in people with T2D is exacerbated when MASLD co-occurs [6]. It is important to distinguish between insulin secretion, the rate at which insulin is released by pancreatic β-cells into the portal vein, and circulating plasma insulin levels, which reflect both secretion and hepatic insulin clearance. This is because approximately 50–80% of endogenously secreted insulin is extracted by the liver during the first pass through the portal circulation, meaning that plasma insulin concentrations can substantially underestimate the actual rate of insulin production [7,8].

Insulin is produced in the pancreas through the enzymatic cleavage of proinsulin into insulin and C-peptide, which are co-secreted in equimolar amounts. Unlike insulin, C-peptide bypasses hepatic extraction, has a longer half-life, and is primarily cleared by the kidneys [9,10]. As a result, C-peptide serves as a physiologically stable proxy for quantifying true insulin secretion, independent of variable hepatic clearance. Mathematical deconvolution of plasma C-peptide concentrations, using established kinetic models, allows accurate estimation of β-cell secretory rates, including basal insulin secretion [11,12]. This distinction is especially important in the context of MASLD, where impaired insulin clearance may confound interpretations of plasma insulin levels alone.

When adipocyte lipid storage capacity is exceeded, ectopic lipid accumulation occurs, with the liver being a major sink for excess triacylglycerol. An overabundance of metabolic substrates (glucose and fatty acids) drives liver-specific insulin resistance and corresponding de novo hepatic glucose and lipid production. This degenerative cycle leads to systemic elevation in glucose and lipids—major risk factors for T2D [13–16].

Understanding the plasma protein network and its relationship with glucose and insulin dynamics and adipose distribution may elucidate how these metabolic features interact. Recent proteomic studies have explored specific features of MASLD [3,17–19]. Nevertheless, a

[1] Joint corresponding authors.

comprehensive system-wide model is lacking, hindering understanding of the causal drivers and consequences of liver fat accumulation and its effects on glucose and insulin homeostasis. Addressing this gap may inform diagnostic, preventive, and therapeutic strategies for MASLD.

To this end, we applied Bayesian network analysis, a probabilistic graphical modeling approach that allows inference of potential causal relationships between interdependent variables, and Mendelian randomization (MR), a statistical genetics method that uses nuclear DNA variants as instrumental variables to strengthen causal inference. These approaches have greater capacity for causal inference than conventional regression or correlation analyses. Thus, these approaches can help quantify directionality and minimize residual confounding. Together, they offer a complementary framework for investigating complex metabolic networks.

Here, we applied Bayesian network and bidirectional two-sample MR analyses to the IMI-DIRECT T2D and non-T2D datasets; these are derived from comprehensively phenotyped prospective cohort studies with measurements of glucose and insulin dynamics (using frequently sampled metabolic challenge tests), MRI-derived abdominal and liver fat content, serological biomarkers, and Olink plasma proteomics. Our aim was to determine potential causal networks linking plasma proteins, glucose and insulin dynamics across their physiological circulation—-from production to clearance—in relation to hepatic lipid accumulation.

2. Materials and methods

2.1. IMI-DIRECT cohorts and measures

To avoid imputation biases, analyses focused on complete-case data from IMI-DIRECT, a multicenter prospective cohort study involving European-ancestry adults diagnosed with T2D ($n = 331$) and those without T2D ($n = 964$). The latter group included participants with normal or prediabetic glycemia (ascertained by glycated hemoglobin A1C (HbA1c), fasting glucose, or 2-h glucose). All participants provided written informed consent at enrollment, and the study protocol was approved by the regional research ethics committees of each clinical study center [20,21]. Examinations and biosample collections in both cohorts, including for serological biomarkers and Olink plasma proteomics, were conducted after a 10-h overnight fast.

Frequently sampled mixed-meal tolerance tests (MMTTs) and oral glucose tolerance tests (OGTTs) were carried out in the diabetes and non-diabetes cohorts, respectively. Insulin secretion estimates were derived from C-peptide deconvolution using the method described by Van Cauter et al. [11], expressed as pmol/min per square meter of estimated body surface area. The insulin secretion rate reflects the continuous output of insulin by pancreatic β-cells and offers a more direct and mechanistically informative measure of β-cell function than circulating insulin concentrations alone, which are also influenced by hepatic insulin clearance. Visceral and subcutaneous fat were assessed using T1-weighted MRI scans, while liver and pancreas fat content were measured using a T2*-based multi-echo technique from MRI scans [22].

Biospecimens from IMI-DIRECT study centers were manually randomized using a mix-shake-distribute procedure and placed into 96-well plates. Proteins in EDTA plasma samples were quantified at SciLifeLab in Stockholm using the Cardiometabolic, Cardiovascular II, Cardiovascular III, Development, and Metabolism panels (Table S1) provided by Olink Proteomics AB (Uppsala, Sweden), following the guidelines for Proximity Extension Assays (PEA) [23]. Further details on the study design and core characteristics are provided elsewhere [20,21,68].

2.2. Bayesian network analysis

We employed Bayesian networks to create directed acyclic graphs (DAGs) that visually depict potential causal connections and interactions between clinical and protein features, using established methods. In these graphs, each variable is represented as a node, and the arcs connecting them suggest possible cause-and-effect relationships. Bayesian network inference is probabilistic and model-dependent; accordingly, several sensitivity analyses were performed across learning algorithms, thresholds, and resampling strategies to assess robustness.

The network skeleton was constructed using the Semi-Interleaved HITON-PC (SI-HITON-PC) algorithm. SI-HITON-PC is a constraint-based algorithm that performs conditional correlation–based independence tests (p-value ≤ 0.01) [24]. This process selects protein features and determines the presence of arcs between clinical and proteomic nodes. The algorithm identifies the Markov blanket for each node, which is the minimal set of nodes that makes the node conditionally independent of other nodes in the network. Directional arcs were then assigned using a hill-climbing score-based algorithm, which orients the arcs based on causal Bayesian theory while avoiding cycles in the DAG and optimizing goodness of fit. To assess robustness, we also compared SI-HITON-PC with the max–min hill-climbing (MMHC) algorithm, another hybrid method combining constraint- and score-based learning. Results were highly consistent, and SI-HITON-PC was retained given its benchmarking in high-dimensional biomedical data [24].

These steps were carried out through a 10-fold cross-validation process, aggregating predictive and posterior correlation scores from multiple runs of the model. To create a robust structure that resists network perturbations, we used model averaging, resulting in an averaged network based on the frequency of observed potential arcs across all networks. Bootstrap aggregation (boot.strength, $R = 500$) was also performed, confirming the stability of key arcs beyond cross-validation. The strength of each arc was determined by how frequently it appeared in the model; the probability of the assigned direction was also calculated. The purpose of these analyses was to determine the probability of true-positive discovery and causal effect directionality. To ensure the networks functioned appropriately, we inverse-normalized all numeric variables to fit a Gaussian Bayesian network with a multivariate normal distribution.

Variables from IMI-DIRECT were adjusted for the clinical center where participants were studied (known as the 'center effect'), age, and sex (when appropriate) in a linear model, including each variable per model. For proteins, in addition to biological adjustments, plate adjustments were made. Residuals were extracted from these models and rank-normalized to fulfill the Gaussian Bayesian network assumption. All the analyses, including the preprocessing, protein selection, and network construction performed separately in the T2D and non-diabetes cohorts.

2.3. Mendelian randomization analysis

Following the construction of the Bayesian networks, we performed bidirectional two-sample MR analysis [25] using GWAS summary statistics on connections between proteins and clinical variables that were strongly associated but had low directional probability (suggesting bidirectional causal associations).

For associations between proteins and clinical variables, we utilized the SMR (Summary-based Mendelian randomization) and HEIDI (Heterogeneity in Dependent Instruments) methodologies [26]. We applied the top pQTL genetic loci as the instrumental *cis*-pQTLs within a ± 1 Mb window of the gene encoding the target protein. HEIDI analysis was conducted to assess the heterogeneity of each lead SNP estimate relative to those in linkage disequilibrium (LD), helping to distinguish biological pleiotropic effects (p-value_Heidi > 0.01) from those driven solely by LD.

For reciprocal associations between clinical variables and proteins, we employed the inverse variance weighted (IVW) method as the primary MR analysis, along with MR-Egger. We considered MR findings to be statistically significant if the causal association among IVW and MR Egger was directionally concordant and the causal association for IVW passed a false discovery rate (FDR) corrected threshold (p-value $< 7e^{-4}$)

for multiple testing (i.e., 0.05/(34 × 2)). To detect and account for directional pleiotropy, the MR-Egger intercept test was conducted (*p*-value > 0.05).

We harmonized GWAS summary statistics to the 1000 Genomes Phase 3 European reference panel (GRCh37/hg19), excluding multiallelic variants, indels, rare variants (MAF < 1%), and palindromic SNPs with ambiguous strand orientation (MAF > 40%). During harmonization, strand mismatches were corrected, and allele frequency concordance within ±0.10 was required. Summary statistics were aligned on effect and non-effect alleles between exposure and outcome datasets, resulting in harmonized data with consistent allele orientation and frequency alignment. The LD estimates for HEIDI analyses were derived from the same reference panel. For IVW MR analyses, instruments were selected based on genome-wide significant associations (p-value < 5×10^{-8}; F-statistics > 28), whereas for HEIDI tests, all LD SNPs had F-statistics > 10.

Statistical analyses were conducted using *R* software version 4.2.1. The Bayesian networks were constructed using the *bnlearn* package [27], and the MR analysis was performed using *TwoSampleMR* [28]. Bayesian networks were visualized using *Cytoscape* software version 3.10.3 [29]. The scripts used can be found at https://github.com/Natalie-Atabaki/BN.

3. Results

Within the IMI-DIRECT cohorts, 331 participants with T2D and 964 without T2D had the required clinical ($n = 19$) and proteomic ($n = 446$) variables for a complete-case analysis. To assess potential selection bias introduced by excluding participants with missing data, baseline demographic, anthropometric, and metabolic characteristics of the complete-case and full IMI-DIRECT datasets for both cohorts were compared (Tables S2–S3). Standardized mean differences (SMDs) across all clinical and biochemical variables were negligible (|SMD| < 0.1), indicating minimal bias from case exclusion. These clinical variables were initially selected based on their association with hepatic lipid accumulation within the IMI-DIRECT cohorts [30] or from existing literature [13–16,31]; these included glycemic, β-cell function, adipose volume, and hepatic lipid content variables. Some clinical biomarkers (e.g., bilirubin and amylin) were not measured in IMI-DIRECT and could therefore not be evaluated; this reflects cohort design rather than selective reporting.

3.1. Proteomic feature selection in IMI-DIRECT cohorts

Of 446 analyzed Olink plasma proteins, 34 were identified as key components in the metabolic network, including CTRC, PON3, KITLG, APOM, CD4, LDLR, MFGE8, LPL, MATN2, FAM3C, FGF21, TGM2,

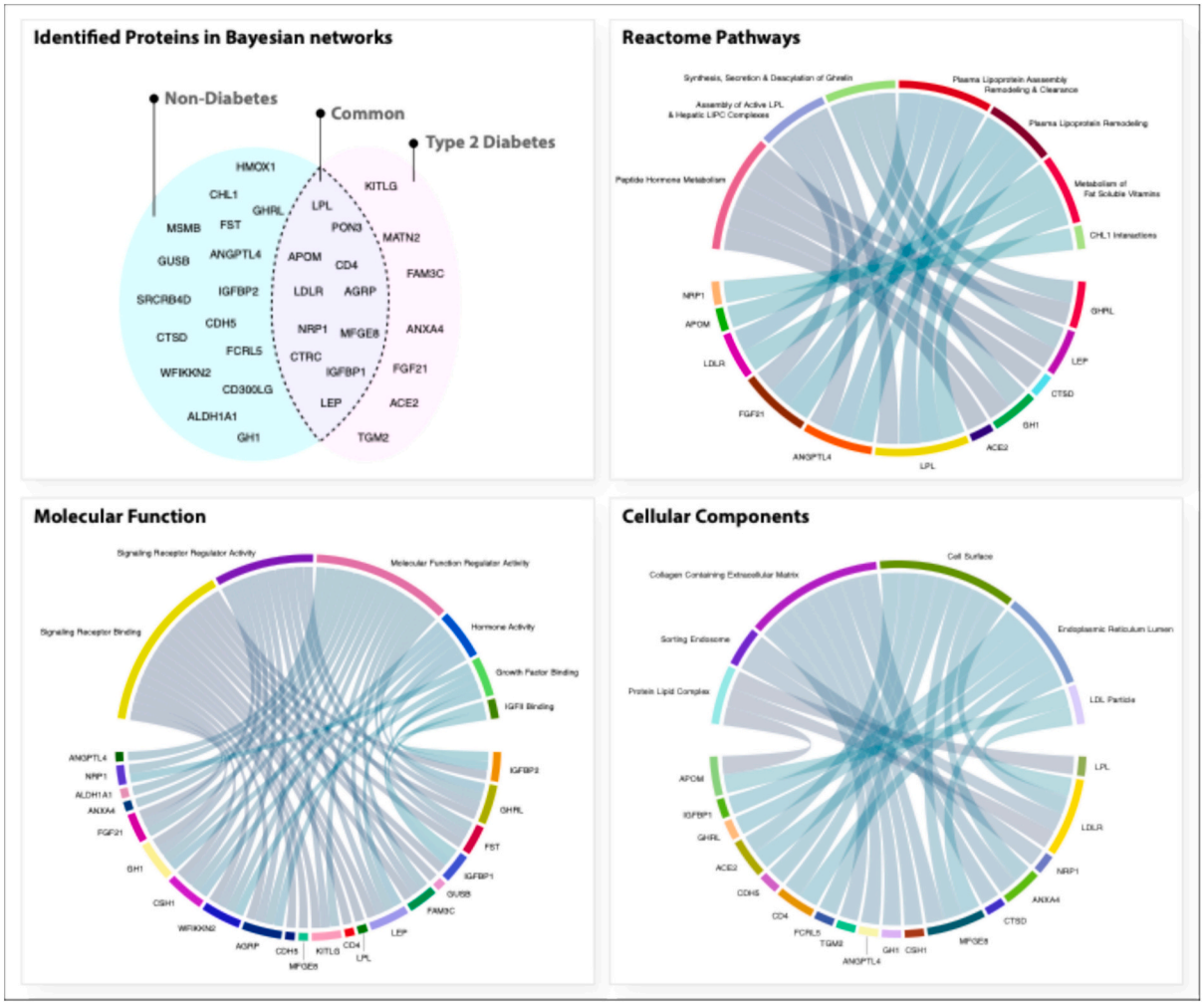

Fig. 1. Proteins identified in Bayesian networks and their functional enrichment across metabolic pathways.
(Top left) Venn diagram showing the distribution of 34 proteins identified in the Bayesian networks across non-diabetes and type 2 diabetes cohorts; (Top right) Reactome pathway enrichment highlighting associated biological processes; (Bottom left) molecular function enrichment analysis; (Bottom right) cellular component enrichment. The background gene set for the hypergeometric test included 18,731 protein-coding genes.

IGFBP1, LEP, AGRP, ANXA4, ACE2, HMOX1, CHL1, MSMB, NRP1, FST, GUSB, GHRL, ANGPTL4, IGFBP2, CDH5, CTSD, SRCRB4D, FCRL5, WFIKKN2, CD300LG, ALDH1A1, and GH1. The selection of these proteins was part of the network-building process, which was conducted separately for the complete-case data of the T2D and non-diabetes cohorts using the SI-HITON-PC algorithm with a conditional independence test threshold of p-value ≤ 0.01. This process identified 27 proteins in the non-diabetes network, and 18 in the T2D network, 11 of them being shared across both cohorts. Functional enrichment analysis revealed that these proteins were associated with several cellular

Table 1
Conditional regression models for each node in the averaged Bayesian network (non-diabetes cohort, $n = 964$).
Each row lists the dependent variable (Outcome) and its set of parent variables (Exposures), along with coefficient estimates and standard errors (SEs). All variables were inverse-normal transformed. The conditional regression coefficients represent local, network-conditioned effects rather than marginal associations; arc strength and direction probabilities therefore provide the primary evidence for robustness of the inferred relationships.

Outcome|Exposures
HbA1c|GlucoseSens
Coefficients: (Intercept) = −1.183e-06, GlucoseSens = −1.225e-01, SE = 0.993
Insulin|TwoInsulin
Coefficients: (Intercept) = 1.589e-07, TwoInsulin =6.582e-0, SE = 0.753
TwoInsulin|Glucose + TwoGlucose + GlucoseSens
Coefficients: (Intercept) = −4.684e-07, Glucose = 1.054e-01, TwoGlucose = 6.223e-01, GlucoseSens = 3.149e-01, SE = 0.724
HDL|BasalISR
Coefficients: (Intercept) = 5.153e-07, BasalISR = −4.039e-01, SE = 0.915
TG|TwoGlucose + HDL + BasalISR
Coefficients: (Intercept) = 4.122e-07, TwoGlucose = 1.213e-01, HDL = −3.399e-01, BasalISR = 3.007e-01, SE = 0.814
BasalISR|Insulin + TwoGlucose + TwoInsulin + OGIS
Coefficients: (Intercept) = −9.746e-08, Insulin = 6.091e-01, TwoGlucose = −1.274e-01, TwoInsulin = 1.389e-01, OGIS = −2.682e-01, SE = 0.483
OGIS|Glucose + TwoGlucose + GlucoseSens
Coefficients: (Intercept) = 2.072e-06, Glucose = −5.395e-01, TwoGlucose = −4.335e-01, GlucoseSens = −3.140e-01, SE = 0.594
TotGLP1min0|Glucose + BasalISR + VAT
Coefficients: (Intercept) = 3.660e-07, Glucose = −9.979e-02, BasalISR = 3.708e-01, VAT = 1.885e-01, SE = 0.880
LiverFat|BasalISR
Coefficients: (Intercept) = 8.713e-08, BasalISR = 5.498e-01, SE = 0.835
VAT|Glucose + BasalISR + OGIS
Coefficients: (Intercept) = 5.781e-07, Glucose = −1.576e-01, BasalISR = 4.264e-01, OGIS = −3.127e-01, SE = 0.788
SAT|OGIS + VAT
Coefficients: (Intercept) = −1.306e-17, OGIS = −2.228e-01, VAT = 4.467e-01, SE = 0.805
Clins|Glucose + TwoGlucose + TwoInsulin + OGIS
Coefficients: (Intercept) = −8.650e-07, Glucose = 2.134e-01, TwoGlucose = 4.762e-01, TwoInsulin = −3.891e-01, OGIS = 7.694e-01, SE = 0.593
Clinsb|Insulin + TwoGlucose + BasalISR + OGIS + LiverFat + Clins
Coefficients: (Intercept) = 2.41e-07, Insulin = −1.154e+00, TwoGlucose = −1.134e-01, BasalISR = 6.804e-01, OGIS = −9.197e-02, LiverFat = 6.161e-02, Clins = 3.470e-01, SE = 0.438
HOMA_IR|Glucose + Insulin + BasalISR + Clinsb
Coefficients: (Intercept) = −7.378e-07, Glucose = 1.515e-01, Insulin = 1.004, BasalISR = −5.359e-02, Clinsb = 3.691e-02, SE = 0.075
PON3|TwoInsulin + VAT + IGFBP2
Coefficients: (Intercept) = 0.224, TwoInsulin = −0.172, VAT = −0.211, IGFBP2 = 0.277, SE = 0.830
APOM|Clins
Coefficients: (Intercept) = 0.245, Clins = 0.191, SE = 0.869
LDLR|TG + PON3 + IGFBP2
Coefficients: (Intercept) = −0.039, TG = 0.689, PON3 = 0.2114343, IGFBP2 = −0.103, SE = 0.679
AGRP|Glucose + TG + CD4
Coefficients: (Intercept) = −0.037, Glucose = 0.059, TG = 0.268, CD4 = 0.461, SE = 0.797
LPL|VAT + LEP
Coefficients: (Intercept) = 0.124, VAT = −0.433, LEP = 0.188, SE = 0.870
LEP|BasalISR + VAT + SAT + Clins
Coefficients: (Intercept) = −0.275, BasalISR = 0.236, VAT = 0.155, SAT = 0.313, Clins = −0.089, SE = 0.673
IGFBP1|BasalISR
Coefficients: (Intercept) = −0.002, BasalISR = −0.601, SE = 0.855
CTRC|CDH5
Coefficients: (Intercept) = 0.297, CDH5 = 0.132, SE = 0.924
IGFBP2|IGFBP1
Coefficients: (Intercept) = 0.135, IGFBP1 = 0.607, SE = 0.671
CDH5|Clinsb
Coefficients: (Intercept) = −1.305, Clinsb = −0.305, SE = 1.724
Glucagonmin0|TotGLP1min0
Coefficients: (Intercept) = 3.262e-08, TotGLP1min0 = 3.678e-01, SE = 0.930
CTSD|LiverFat + LDLR
Coefficients: (Intercept) = −0.977, LiverFat = 0.414, LDLR = 0.514, SE = 1.401
PancFat|VAT
Coefficients: (Intercept) = 3.799e-09, VAT = 3.666e-01, SE = 0.930

AGRP: agouti-related peptide; ALDH1A1: aldehyde dehydrogenase 1 family member A1; APOM: apolipoprotein M; BasalISR: basal insulin secretion rate at the beginning of the oral glucose tolerance test (OGTT); CD4: cluster of differentiation 4; CDH5: cadherin-5; Clins: mean insulin clearance during OGTT-calculated as mean insulin secretion/mean insulin concentration; Clinsb: basal insulin clearance; CTRC: chymotrypsin C; CTSD: cathepsin D; FGF21: fibroblast growth factor 21; Glucagonmin0: fasting glucagon; Glucose: fasting plasma glucose; GlucoseSens: glucose sensitivity; GUSB: β-glucuronidase; HbA1c: glycated hemoglobin A1C; HDL: high-density lipoprotein cholesterol; IGFBP1/2: insulin-like growth factor binding proteins 1 and 2; Insulin: fasting plasma insulin; KITLG: KIT ligand; LDLR: low-density lipoprotein receptor; LEP: leptin; LiverFat: hepatic fat content; LPL: lipoprotein lipase; MFGE8: milk fat globule-epidermal growth factor 8; OGIS: oral glucose insulin sensitivity index according to the method of Mari et al. [12]; PancFat: pancreas fat; PON3: paraoxonase 3; SAT: subcutaneous adipose tissue; TG: triglycerides; TotGLP1min0: fasting total GLP-1; TwoGlucose/TwoInsulin: 2-h post-load values from OGTT; VAT: visceral adipose tissue.

components, including protein-lipid macromolecular complexes, low-density lipoprotein (LDL) particles, and the endoplasmic reticulum lumen. Their molecular functions were enriched in signaling receptor binding, hormone activity, and growth factor interactions. Pathway analysis showed an overrepresentation of biological processes related to lipoprotein metabolism, ghrelin processing, and fat-soluble vitamin metabolism, highlighting their potential roles in metabolic regulation across health and disease states (Fig. 1) [32–35].

3.2. Metabolic Bayesian network in IMI-DIRECT non-diabetes cohort

To construct the metabolic network, we performed structural and parameter learning using clinical and proteomic features as nodes within the non-diabetic cohort ($n = 964$). To evaluate the network's fit to the data, we compared the Bayesian information criterion (BIC) scores of the learned network with the BIC density of 1000 randomly generated (null) networks (Fig. S1). To ensure robustness and parsimony, we restricted the network to arcs with both strength and directional probabilities ≥ 0.8. Table 1 outlines the conditional density of each node within the restricted network, along with parameter estimates for its upstream (parent) variables, indicating the magnitude of their relationships within the network.

The constructed network (Fig. 2) identified basal insulin secretion rate (BasalISR) by pancreatic β-cells as the main direct causal determinant of hepatic lipid content. Insulin clearance, calculated from basal values as insulin secretion/insulin concentration (Clinsb), was identified as a downstream effect of liver fat. BasalISR was the parental node for both liver fat and visceral adipose tissue (VAT), leading to fat accumulation in ectopic and visceral abdominal areas. Given VAT's strong correlation with liver fat, we hypothesized that BasalISR directly and indirectly (via VAT accumulation) drives liver fat accumulation. Table 1 also provides the conditional regression coefficients (β) and standard errors (SE) for key node relationships (strength and directional probabilities $\geq$ 0.8). For example, BasalISR was positively associated with liver fat ($\beta = 0.55$, SE = 0.84) and inversely ($\beta = -0.60$, SE = 0.86) with IGFBP1 (insulin-like Growth Factor Binding Protein), while liver fat was

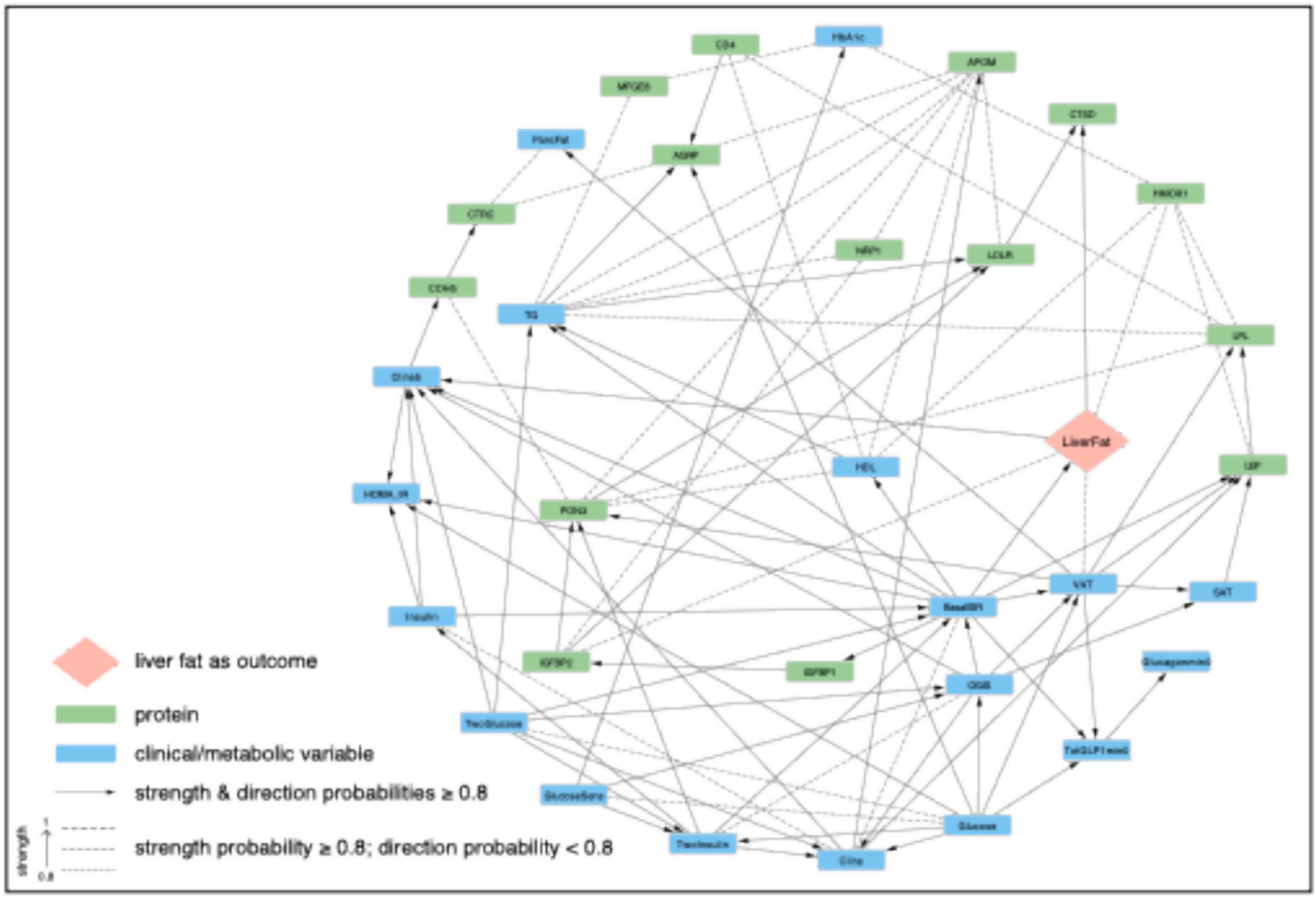

Fig. 2. Bayesian network of metabolic and proteomic interactions in the IMI-DIRECT non-diabetes cohort ($n = 964$).
The graph displays directed relationships among clinical and proteomic variables. Nodes are color-coded: blue (clinical/metabolic), green (proteins), peach (liver fat as outcome). Solid arrows represent directed associations with high confidence (strength and direction probability ≥ 0.8), while dashed arrows indicate less confident directionality.
AGRP: agouti-related peptide; ALDH1A1: aldehyde dehydrogenase 1 family member A1; APOM: apolipoprotein M; BasalISR: basal insulin secretion rate at the beginning of the oral glucose tolerance test (OGTT); CD4: cluster of differentiation 4; CDH5: cadherin-5; Clins: mean insulin clearance during OGTT-calculated as mean insulin secretion/mean insulin concentration; Clinsb: basal insulin clearance; CTRC: chymotrypsin C; CTSD: cathepsin D; FGF21: fibroblast growth factor 21; Glucagonmin0: fasting glucagon; Glucose: fasting plasma glucose; GlucoseSens: glucose sensitivity; GUSB: β-glucuronidase; HbA1c: glycated hemoglobin A1C; HDL: high-density lipoprotein cholesterol; IGFBP1/2: insulin-like growth factor binding proteins 1 and 2; Insulin: fasting plasma insulin; KITLG: KIT ligand; LDLR: low-density lipoprotein receptor; LEP: leptin; LiverFat: hepatic fat content; LPL: lipoprotein lipase; MFGE8: milk fat globule-epidermal growth factor 8; OGIS: oral glucose insulin sensitivity index according to the method of Mari et al. [12]; PancFat: pancreas fat; PON3: paraoxonase 3; SAT: subcutaneous adipose tissue; TG: triglycerides; TotGLP1min0: fasting total GLP-1; TwoGlucose/TwoInsulin: 2-h post-load values from OGTT; VAT: visceral adipose tissue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

positively (β = 0.41, SE = 1.40) associated with CTSD (cathepsin D). These estimates reflect local conditional effects within the network rather than independent effect sizes. Arc strength and direction probabilities (Tables S4–S6) therefore provide the primary evidence for robust relationships, with regression estimates serving as supportive indicators of effect magnitude.

The derived network revealed a direct inverse effect of basal hyperinsulinemia on IGFBP1, which then positively affected IGFBP2. Insulin suppresses IGFBP1 production in the liver, indicating that individuals with basal insulin hypersecretion have lower IGFBP1 levels owing to this suppressive feedback loop. The network also suggests a direct causal effect of liver fat on CTSD levels, an enzyme involved in protein degradation. CTSD has also been associated with lipid metabolism, inflammation, and hepatic fibrosis in nonalcoholic fatty liver disease (NAFLD) [36]. Higher HMOX1 (heme Oxygenase 1) levels were strongly associated with increased liver fat, which likely reflects its role in metabolic stress responses. This could be either a reaction to liver fat-induced oxidative stress or to the products of heme degradation (e.g., iron) that exacerbate lipid peroxidation, contributing to fat storage and inflammation in the liver [37]. Plasma proteins CTSD, IGFBP2, HMOX1, GUSB (β-glucuronidase), LPL (lipoprotein Lipase), ALDH1A1 (aldehyde dehydrogenase 1 family member A1), and AGRP (agouti-related protein) were directly linked to liver fat at different strength and directional probability levels (Tables S4–S6 and Fig. S2).

3.3. Metabolic Bayesian network in IMI-DIRECT diabetes cohort

To identify causal networks specific to individuals with T2D, we conducted Bayesian network analyses within the T2D cohort of the IMI-DIRECT study ($n = 331$) in a similar manner as before (Table 2, Figs. 3, and S3). In alignment with the non-diabetes network, BasalISR emerged as a putative upstream driver influencing liver fat accumulation (β = 0.50, SE = 0.87) with the FGF21 (fibroblast growth factor 21) protein as a downstream target of liver fat (β = 0.37, SE = 0.83). FGF21, predominantly secreted by the liver, is a key modulator of hepatic glucose and lipid metabolism. Elevated circulating FGF21 levels have been reported in patients with non-alcoholic fatty liver disease [38], possibly reflecting a compensatory response to hepatic lipid accumulation. This response may promote lipolysis, enhance fatty acid oxidation, and improve insulin sensitivity. However, in the context of T2D, the effectiveness of this compensatory mechanism may be diminished due to preexisting metabolic dysregulation.

ACE2 (angiotensin-converting enzyme 2) was also directly linked to liver fat, albeit with a weaker association that excluded it from the parsimonious network (Fig. S4). As a component of the renin-angiotensin system (RAS), ACE2 mitigates the proinflammatory and profibrotic effects of angiotensin II, and its elevation in fatty liver disease may reflect a protective, compensatory adaptation to lipid overload and oxidative stress [39]. Notably, increased hepatic ACE2 expression has been implicated in facilitating SARS-CoV-2 entry, potentially

Table 2
Conditional regression models for each node in the averaged Bayesian network (type 2 diabetes cohort, $n = 331$).
Each row lists the dependent variable (Outcome) and its set of parent variables (Exposures), along with coefficient estimates and standard errors (SEs). All variables were inverse-normal transformed. The conditional regression coefficients represent local, network-conditioned effects rather than marginal associations; arc strength and direction probabilities therefore provide the primary evidence for robustness of the inferred relationships.

Outcome|Exposures
Insulin|TwoInsulin + Clinsb
Coefficients: (Intercept) = −1.183e-06, GlucoseSens = −1.225e-01, SE = 0.993
TwoGlucose|TwoInsulin + GlucoseSens + OGIS + Clins
Coefficients: (Intercept) = −1.475e-06, TwoInsulin = 2.882e-01, GlucoseSens = −3.685e-01, OGIS = −6.688e-01, Clins = 3.660e-01, SE = 0.560
BasalISR|Insulin + Clinsb
Coefficients: (Intercept) = −9.682e-18, Insulin = 1.286, Clinsb = 5.610e-01, SE = 0.458
GlucoseSens|BasalISR
Coefficients: (Intercept) = 8.583e-18, BasalISR = 2.483e-01, SE = 0.969
OGIS|Glucose + TwoInsulin + GlucoseSens + Clins
Coefficients: (Intercept) = −8.499e-08, Glucose = −6.819e-01, TwoInsulin = −4.463e-01, GlucoseSens = −1.165e-01, Clins = 1.061e-01, SE = 0.380
Glucagonmin0|HDL + HOMA_IR
Coefficients: (Intercept) = 1.098e-17, HDL = −1.729e-01, HOMA_IR = 2.110e-01, SE = 0.957
LiverFat|BasalISR
Coefficients: (Intercept) = 4.452e-08, BasalISR = 4.991e-01, SE = 0.867
SAT|VAT
Coefficients: (Intercept) = 5.551e-18, VAT = 5.138e-01, SE = 0.858
HOMA_IR|Glucose + Insulin
Coefficients: (Intercept) = 3.413e-08, Glucose = 2.738e-01, Insulin = 8.944e-01, SE = 0.143
CTRC|CD4
Coefficients: (Intercept) = −0.214, CD4 = 0.194, SE = 0.931
APOM|HDL
Coefficients: (Intercept) = −0.581, HDL = 0.308, SE = 0.891
LDLR|TG
Coefficients: (Intercept) = 0.001, TG = 0.637, SE = 0.790
MFGE8|TG + MATN2
Coefficients: (Intercept) = −0.126, TG = 0.517, MATN2 = 0.385, SE = 0.855
LPL|PON3
Coefficients: (Intercept) = −0.056, PON3 = 0.411, SE = 0.876
FGF21|LiverFat
Coefficients: (Intercept) = 0.289, LiverFat = 0.371, SE = 0.831
TGM2|PancFat
Coefficients: (Intercept) = −0.428, PancFat = −0.199, SE = 1.122

APOM: apolipoprotein M; BasalISR: basal insulin secretion rate at the beginning of mixed-meal tolerance test (MMTT); CD4: cluster of differentiation 4; Clins: mean insulin clearance during MMTT-calculated as mean insulin secretion/mean insulin concentration; Clinsb: basal insulin clearance; CTRC: chymotrypsin C; FGF21: fibroblast growth factor 21; Glucagonmin0: fasting glucagon; Glucose: fasting plasma glucose; GlucoseSens: glucose sensitivity; HDL: high-density lipoprotein cholesterol; HOMA_IR: homeostatic model assessment of insulin resistance; Insulin: fasting plasma insulin; LDLR: low-density lipoprotein receptor; LiverFat: hepatic fat content; LPL: lipoprotein lipase; MATN2: matrilin-2; MFGE8: milk fat globule-EGF factor 8; OGIS: oral glucose insulin sensitivity index according to the method of Mari et al. [12]; PancFat: pancreas fat; PON3: paraoxonase 3; SAT: subcutaneous adipose tissue; TG: triglycerides; TGM2: transglutaminase 2; TwoGlucose: 2-h post-load glucose (MMTT); TwoInsulin: 2-h post-load insulin (MMTT); VAT: visceral adipose tissue.

contributing to liver injury and adverse outcomes in individuals with diabetes [40]. Additionally, both VAT and HbA1c were strongly associated with liver fat, although no definitive causal direction could be established (shown as dashed lines in Fig. 3). Detailed information on all arcs is reported in Tables S7–S9 and Fig. S4. As before, the β estimates represent local conditional effects within the network, while arc strength and direction probabilities provide the primary evidence for robustness.

3.4. Complementary directional inference with Mendelian randomization

For some of the arcs, although the strength of the connection was high, the direction of the connection could not be determined with high probability (dashed lines in Figs. 2 and 3). This suggests either a bidirectional association or insufficient data to confidently determine causal direction. To investigate this, we performed two-sample MR analyses for connections from clinical variables to the proteins and SMR–HEIDI to study the association from proteins to the clinical traits, using the latest summary statistics from publicly available databases (Table S10). The results are summarized in Table 3 and detailed in Table S11.

We identified potential positive causal associations from high-density lipoprotein (HDL) to HMOX1, APOM (apolipoprotein M), and KITLG (KIT ligand) proteins. HDL is well-known for its anti-inflammatory and antioxidant properties. The observed association with HMOX1, a stress-response enzyme, may reflect HDL's role in mitigating oxidative stress through upregulation of HMOX1. APOM, predominantly associated with HDL particles, contributes to lipid metabolism and endothelial function, so a causal link from HDL aligns with its established biological role as a stabilizing carrier of the APOM protein [41]. The association with KITLG, a cytokine regulating hematopoiesis and cell survival, may reflect HDL-mediated modulation of inflammatory and hematopoietic signaling pathways [42].

Triglycerides (TG) also showed positive causal effects on LDLR (low-density lipoprotein receptor), AGRP, and MFGE8 (Milk fat globule–EGF factor 8 protein) protein levels. Elevated TG levels can perturb lipid homeostasis, promoting upregulation of LDLR to enhance lipid clearance from circulation [43]. The link to AGRP, a neuropeptide that promotes appetite and energy intake, may reflect a compensatory feedback in which increased TG levels signal energy surplus, thereby stimulating AGRP expression as part of energy balance regulation. TG-driven upregulation of MFGE8 may indicate an adaptive anti-inflammatory response, as MFGE8 facilitates apoptotic cell clearance and supports tissue repair during lipid-induced stress [44]. In contrast, TG exhibited a negative association with NRP1 (neuropilin-1), possibly reflecting suppression of NRP1 expression by TG-induced endothelial dysfunction, thereby impairing angiogenic and lipid-handling capacity

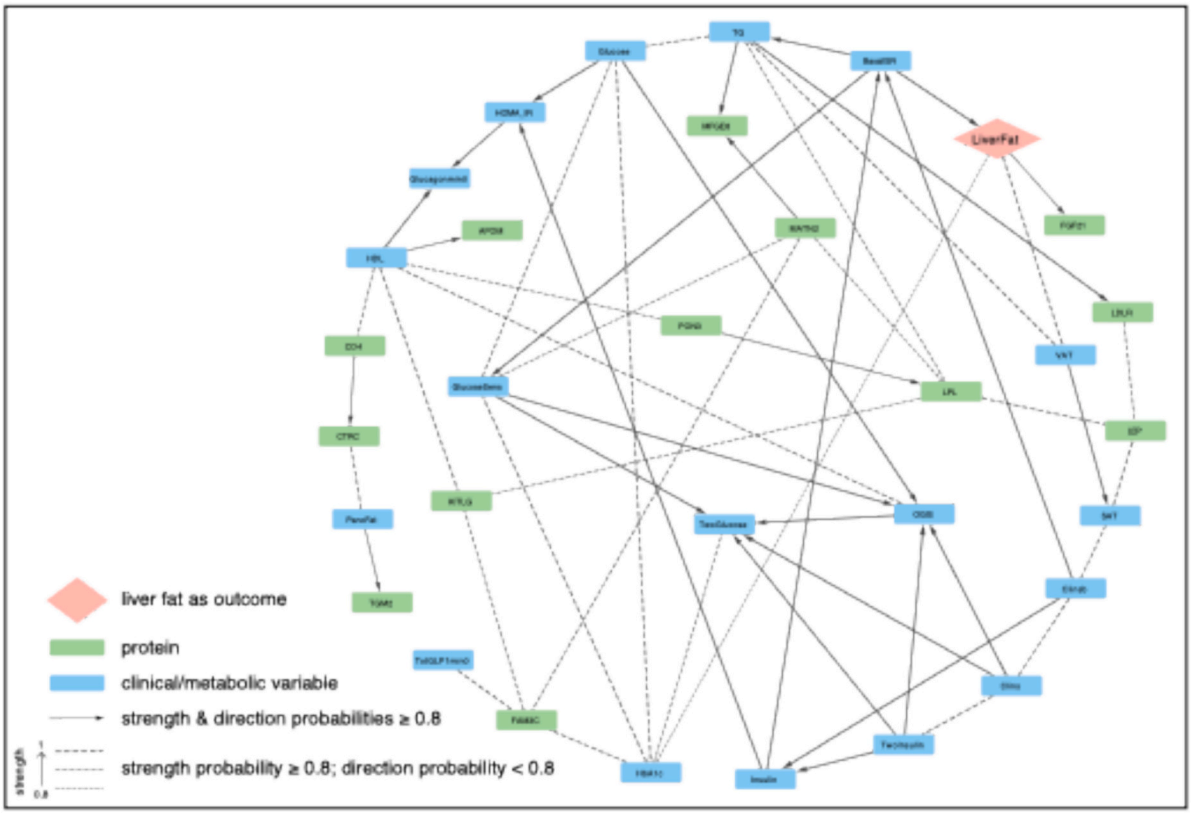

Fig. 3. Bayesian network of metabolic and proteomic interactions in the IMI-DIRECT type 2 diabetes cohort (n = 331).
The graph displays directed relationships among clinical and proteomic variables. Nodes are color-coded: blue (clinical/metabolic), green (proteins), peach (liver fat as outcome). Solid arrows represent directed associations with high confidence (strength and direction probability ≥ 0.8), while dashed arrows indicate less confident directionality.
APOM: apolipoprotein M; BasalISR: basal insulin secretion rate at the beginning of the mixed-meal tolerance tests (MMTT); CD4: cluster of differentiation 4; Clins: mean insulin clearance during MMT-calculated as mean insulin secretion/mean insulin concentration; Clinsb: basal insulin clearance; CTRC: chymotrypsin C; FGF21: fibroblast growth factor 21; Glucagonmin0: fasting glucagon; Glucose: fasting plasma glucose; GlucoseSens: glucose sensitivity; HDL: high-density lipoprotein cholesterol; HOMA_IR: homeostatic model assessment of insulin resistance; Insulin: fasting plasma insulin; LDLR: low-density lipoprotein receptor; LiverFat: hepatic fat content; LPL: lipoprotein lipase; MATN2: matrilin-2; MFGE8: milk fat globule-EGF factor 8; OGIS: oral glucose insulin sensitivity index according to the method of Mari et al. [12]; PancFat: pancreas fat; PON3: paraoxonase 3; SAT: subcutaneous adipose tissue; TG: triglycerides; TGM2: transglutaminase 2; TwoGlucose/TwoInsulin: 2-h post-load values from MMTT; VAT: visceral adipose tissue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Mendelian randomization (MR) analyses to determine directionality of associations between clinical variables and proteins.
This table presents the results of bidirectional MR and SMR–HEIDI analyses used to infer the direction of causal relationships between clinical traits and proteins for which Bayesian networks indicated strong associations but uncertain directionality (Figs. 2 and 3). Highlighted rows represent statistically significant findings after correction, with effect estimates, standard errors, and *p*-values shown for both forward (clinical → protein) and reverse (protein → clinical) directions. Color shading indicates the direction with stronger evidence: peach for positive β estimates, and blue for negative β estimates.

Link	BETA_IVW	SE_IVW	PVAL_IVW	NSNP_IVW	PVAL_B0Egger	BETA_SMR	SE_SMR	PVAL_SMR	PVAL_HEIDI	NSNP_HEIDI
AGRP - Glucose	0.032	0.057	5.76E-01	107	3.50E-01	0.041	0.037	2.73E-01	6.63E-01	6
PON3 - TwoGlucose	0.384	0.308	2.13E-01	3	7.49E-01	−0.070	0.023	2.16E-03	3.40E-01	20
HMOX1 - HDL	0.260	0.018	4.03E-46	1905	3.46E-01	0.004	0.018	8.30E-01	1.21E-01	20
APOM - HDL	0.465	0.018	1.01E-143	1912	4.70E-01	−0.007	0.004	9.38E-02	4.50E-03	20
APOM - TG	−0.001	0.022	9.68E-01	1691	1.37E-01	0.003	0.004	3.84E-01	4.47E-09	20
LDLR - TG	0.940	0.018	0.00E+00	1696	2.43E-01	−0.012	0.029	6.75E-01	NA	NA
AGRP - TG	0.600	0.015	0.00E+00	1699	8.82E-01	0.071	0.024	2.92E-03	1.55E-03	20
LPL - TG	−0.396	0.022	9.50E-70	1701	2.67E-09	−0.056	0.003	4.15E-98	4.21E-102	20
IGFBP1 - BasalISR	NA	NA	NA	NA	NA	−0.376	0.881	6.69E-01	7.35E-01	17
HMOX1 - LiverFat	0.198	0.024	4.22E-17	15	4.26E-01	0.050	0.105	6.33E-01	3.72E-01	20
CTSD - LiverFat	0.269	0.041	7.14E-11	15	9.50E-01	−0.010	0.026	7.11E-01	3.42E-01	20
PON3 - VAT	0.084	0.203	6.80E-01	5	3.76E-02	−0.021	0.023	3.73E-01	6.05E-02	20
LPL - VAT	0.151	0.130	2.43E-01	5	3.37E-02	−0.017	0.018	3.44E-01	5.19E-01	20
LEP - VAT	0.158	0.213	4.58E-01	5	6.56E-02	NA	NA	NA	NA	NA
LEP- SAT	0.614	0.084	2.28E-13	1	NA	NA	NA	NA	NA	NA
HMOX1 - HbA1c	−0.756	0.105	6.88E-13	128	8.26E-01	−0.003	0.019	8.68E-01	1.61E-01	20
MFGE8 - HbA1c	0.159	0.104	1.27E-01	128	8.74E-01	0.000	0.005	1.00E+00	6.67E-01	20
MFGE8 - TG	0.712	0.017	0.00E+00	1686	6.76E-01	−0.008	0.004	6.22E-02	4.26E-01	20
PON3 - HDL	0.478	0.016	2.07E-189	1909	1.80E-06	−0.029	0.005	4.79E-10	2.41E-01	20
CD4 - HDL	−0.268	0.017	3.65E-58	1898	5.60E-05	−0.002	0.004	6.89E-01	4.37E-01	20
NRP1 - TG	−0.338	0.021	5.79E-57	1687	2.69E-01	−0.006	0.004	1.91E-01	8.67E-01	20
CTRC - PancFat	−0.716	0.141	3.79E-07	10	9.59E-01	0.038	0.025	1.28E-01	7.55E-01	20
IGFBP2 - LiverFat	−0.009	0.044	8.43E-01	15	2.87E-01	0.107	0.129	4.11E-01	NA	NA
FAM3C - HbA1c	−0.152	0.063	1.61E-02	128	4.11E-05	−0.019	0.013	1.45E-01	1.64E-01	20
KITLG - HDL	0.428	0.023	8.36E-80	1916	5.38E-01	−0.024	0.024	3.22E-01	2.65E-01	7
FGF21 - LiverFat	0.115	0.068	9.40E-02	15	4.20E-02	0.024	0.064	7.07E-01	2.34E-01	20
TGM2 - PancreasFat	0.013	0.049	7.89E-01	10	1.59E-01	−0.023	0.052	6.61E-01	8.14E-01	20
MATN2 - GlucoseSense	0.013	0.030	6.68E-01	25	2.21E-01	0.013	0.026	6.09E-01	2.72E-01	20

AGRP: agouti-related peptide; APOM: apolipoprotein M; BasalISR: basal insulin secretion rate at the beginning of frequently sampled mixed-meal tolerance tests (MMTT) or oral glucose tolerance tests (OGTT); BETA_IVW / SE_IVW / PVAL_IVW / NSNP_IVW: causal effect estimate, standard error, *p*-value, and number of SNPs from inverse variance weighted MR; BETA_SMR / SE_SMR / PVAL_SMR / PVAL_HEIDI / NSNP_HEIDI: effect estimate, standard error, p-value, HEIDI test p-value, and number of SNPs from summary-based MR; PVAL_B0Egger: p-value for MR-Egger intercept (pleiotropy test); CD4: cluster of differentiation 4; CDH5: cadherin 5; CTRC: chymotrypsin C; CTSD: cathepsin D; FAM3C: family with sequence similarity 3 member C; FGF21: fibroblast growth factor 21; Glucose: fasting plasma glucose; GlucoseSense: glucose sensitivity; HbA1c: hemoglobin A1c; HDL: high-density lipoprotein cholesterol; HMOX1: heme oxygenase 1; HOMA-IR: homeostatic model assessment for insulin resistance; IGFBP1: insulin-like growth factor-binding protein 1; IGFBP2: insulin-like growth factor-binding protein 2; KITLG: KIT ligand; LDLR: low-density lipoprotein receptor; LEP: leptin; LiverFat: hepatic fat content; LPL: lipoprotein lipase; MATN2: matrilin-2; MFGE8: milk fat globule-EGF factor 8; NRP1: neuropilin 1; PancFat: pancreas fat; PON3: paraoxonase 3; SAT: subcutaneous adipose tissue; TG: triglycerides; TGM2: transglutaminase 2; TwoInsulin: 2-h post-load insulin (OGTT/MMTT); VAT: visceral adipose tissue.
*Insulin fold change (IFC) was used as a proxy for TwoGlucose.
*xinsdG30: insulin secretion over glucose increment at 30 min (ΔIns/ΔG30) was used as a proxy for GlucoseSense.

in hepatocytes [45]. Liver fat demonstrated positive causal effects on HMOX1 and CTSD, consistent with the Bayesian network (Fig. 2).

HbA1c and pancreas fat showed negative causal effects on HMOX1 and CTRC (chymotrypsin C), respectively. Chronic hyperglycemia, reflected by elevated HbA1c, may suppress HMOX1 expression through mechanisms involving oxidative and altered transcriptional regulation. Similarly, fat infiltration of the pancreas can impair acinar cell function—the primary source of CTRC—thereby reducing enzyme synthesis or secretion. PON3 (paraoxonase 3) was the only protein exhibiting a negative causal effect on clinical variables, specifically on insulin fold change (proxy for 2-h post-load insulin) and HDL, consistent with its antioxidative and lipid-regulating role.

We further performed bidirectional two-sample MR analyses between clinical variables using IVW (Table S12) and between proteins using SMR–HEIDI (Table S13), focusing on associations identified with high confidence in the Bayesian networks. Several associations were confirmed through MR, including positive effects of fasting insulin and glucose on HOMA-IR (homeostatic model assessment of insulin resistance), a negative effect of glucose sensitivity (xInsdG30, insulin secretion over glucose increment at 30 min, used as a proxy) on HbA1c, and a bidirectional negative association with 2-h glucose. VAT and pancreatic fat also displayed a bidirectional relationship, consistent with shared lipid storage and mobilization pathways.

The complementary MR analyses helped assign directionality to thirteen associations identified through Bayesian network analysis that demonstrated strong associations but where directionality could not be confidently ascertained. In addition, among clinical variables with higher directional probability (≥0.8), five more associations were validated through MR (Table S14).

3.5. *Bayesian networks and MASLD posterior probabilistic inference in IMI-DIRECT*

To evaluate the potential direct and indirect contributions of each variable to liver fat accumulation, we conducted posterior probabilistic inference of MASLD (defined as liver fat > 5% [46–49]) using the constructed Bayesian networks. Focusing on liver fat as the outcome of interest, we estimated posterior probabilities by conditioning not only on its direct parent nodes but also individually on each variable in the network. This comprehensive conditioning strategy enabled the identification of both direct and indirect influences on the likelihood of MASLD. To facilitate this analysis, all continuous variables were discretized into low, moderate, and high levels using Hartemink's method [50].

In the non-diabetes network, the baseline probability of MASLD was 34%, which rose to 71% after conditioning on high BasalISR (Fig. 4).

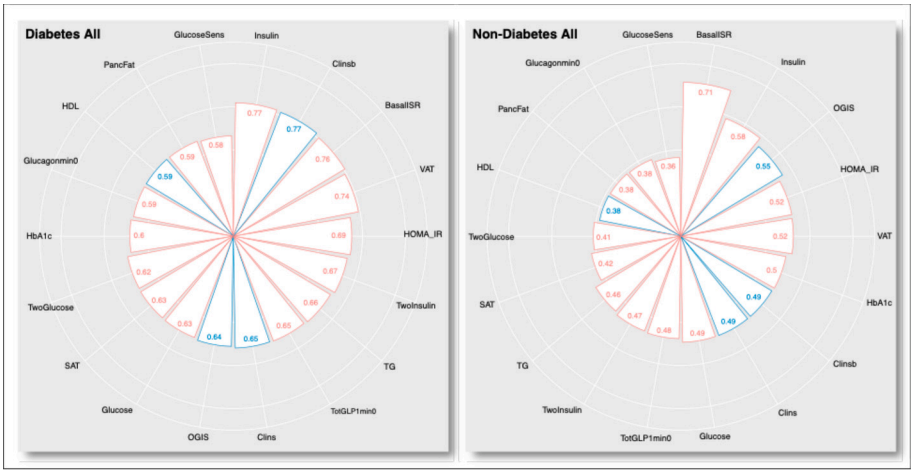

Fig. 4. Posterior probabilities of MASLD based on clinical predictors across IMI-DIRECT diabetes and non-diabetes cohorts.
Radar plots display the posterior probability of MASLD (as liver fat > 5%) after conditioning on individual clinical variables in (left panel) individuals with T2D and (right panel) those without diabetes. Bars represent the conditional probability of MASLD given high (red) or low (blue) levels of each variable.
BasalISR: basal insulin secretion rate at the beginning of the frequently sampled mixed-meal tolerance tests (MMTT) or oral glucose tolerance tests (OGTT); Glucose: fasting plasma glucose; GlucoseSens: glucose sensitivity; HbA1c: glycated hemoglobin A1C; HDL: high-density lipoprotein cholesterol; Insulin: fasting plasma insulin; LiverFat: hepatic fat content; MASLD: metabolic dysfunction-associated steatotic liver disease; PancFat: pancreas fat; SAT: subcutaneous adipose tissue; TG: triglycerides; TotGLP1min0: fasting total GLP-1; TwoGlucose/TwoInsulin: 2-h post-load values from OGTT/MMTT; T2D: type 2 diabetes; VAT: visceral adipose tissue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Proteomic data further revealed distinct modulators of MASLD risk (Fig. 5). In this group, the greatest increases (~30%) in MASLD probability were observed after conditioning on high levels of GUSB and ALDH1A1. GUSB is involved in lysosomal degradation and has been implicated in inflammation and obesity-related metabolic dysfunction [3,51]; its upregulation may signal hepatic stress or immune activation. NAFLD-associated genetic variants have also been reported to correlate with plasma GUSB levels [3]. ALDH1A1 has been linked to hepatic lipid accumulation through its regulation of PPAR signaling [52,53]. Additionally, low levels of IGFBP1 and LPL were associated with increased MASLD risk (~15%). IGFBP1 is negatively regulated by insulin and is typically suppressed in insulin-resistant states, serving as a surrogate marker of hepatic insulin action [54]. LPL hydrolyzes circulating TGs, and its reduced activity can impair lipid clearance, promoting ectopic fat deposition in the liver [55].

In the diabetes network, the unconditioned MASLD probability was

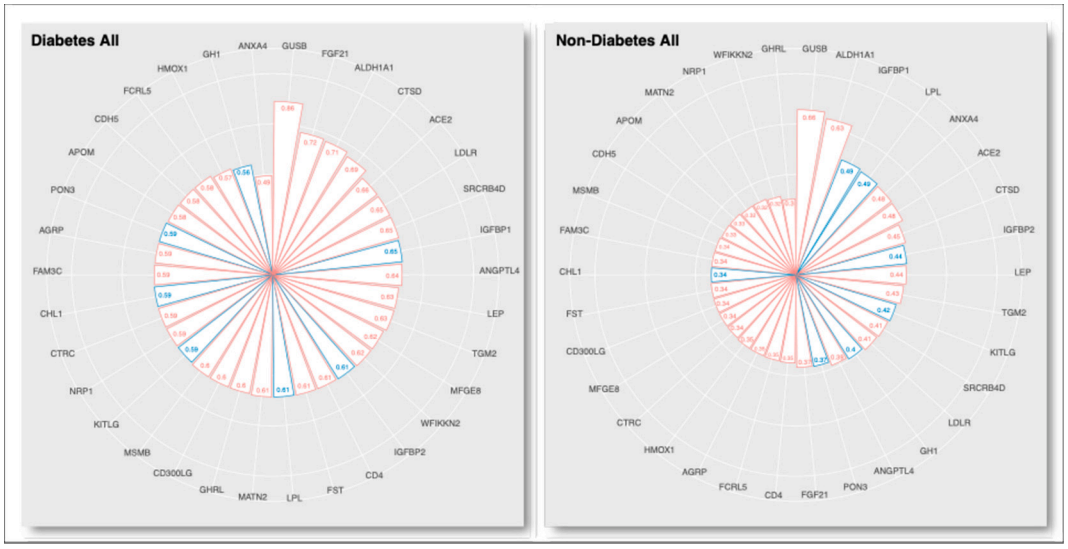

Fig. 5. Posterior probabilities of MASLD based on proteomic predictors in IMI-DIRECT diabetes and non-diabetes cohorts.
Radar plots display the conditional probability of MASLD (as liver fat > 5%) after conditioning on individual proteomic variables in (left panel) T2D and (right panel) non-diabetes cohorts. Bars indicate high (red) or low (blue) levels of the corresponding protein.
ACE2: angiotensin-converting enzyme 2; AGRP: agouti-related peptide; ALDH1A1: aldehyde dehydrogenase 1 family member A1; ANXA4: annexin A4; APOM: apolipoprotein M; CD300LG: CD300 molecule-like family member G; CD4: cluster of differentiation 4; CDH5: cadherin-5; CHL1: cell adhesion molecule L1-like; CTRC: chymotrypsin C; CTSD: cathepsin D; FAM3C: family with sequence similarity 3 member C; FCRL5: Fc receptor-like 5; FGF21: fibroblast growth factor 21; FST: follistatin; GH1: growth hormone 1; GHRL: ghrelin; GUSB: β-glucuronidase; HMOX1: heme oxygenase 1; IGFBP1: insulin-like growth factor-binding protein 1; IGFBP2: insulin-like growth factor-binding protein 2; KITLG: KIT ligand; LEP: leptin; LDLR: low-density lipoprotein receptor; LPL: lipoprotein lipase; MATN2: matrilin-2; MFGE8: milk fat globule-EGF factor 8; MSMB: microseminoprotein β; NRP1: neuropilin-1; PON3: paraoxonase 3; SRCRB4D: scavenger receptor cysteine-rich domain-containing protein B4D; TGM2: transglutaminase 2; WFIKKN2: WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain-containing protein 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

58%, with similar increases (~20%) observed after conditioning on high fasting insulin, BasalISR, or VAT, and low Clinsb, reinforcing the roles of hyperinsulinemia, insulin resistance, and visceral adiposity in hepatic steatosis. In this network, GUSB again demonstrated the strongest association (~30% increase), further supporting its potential role in disease progression. High levels of FGF21, ALDH1A1, and CTSD also elevated MASLD probability (~15%). Across both the non-diabetes and diabetes networks, with baseline MASLD probabilities of 34% and 58%, respectively, the most informative combination of markers predicting the highest MASLD risk (over 90%) included elevated BasalISR, VAT, GUSB, and ALDH1A1, alongside low levels of LPL and IGFBP1. This pattern highlights a coordinated disruption of pathways related to insulin resistance, inflammation, and lipid metabolism.

Sex-stratified analyses revealed subgroup-specific patterns (Figs. S5 and S6). Among non-diabetic males (n = 793), ALDH1A1 and LEP (leptin) showed the strongest effects, while GUSB was most influential in non-diabetic females (n = 171). In the T2D cohort (n = 193 males, n = 138 females), LEP remained the most influential proteomic marker in males, while GUSB remained dominant in females. Among clinical variables, BasalISR and VAT were the strongest predictors of MASLD in females, underscoring the pivotal role of insulin secretion and visceral adiposity in hepatic lipid accumulation. In males, BasalISR was most predictive in the non-diabetic state, whereas subcutaneous adipose tissue (SAT) had a greater impact in diabetes, highlighting sex- and disease stage–specific differences in adipose distribution and metabolic regulation.

As a sensitivity analysis, we varied the Hartemink discretization from 3 to 5 levels and observed similar ranking of the proteins and clinical traits. Furthermore, to assess sensitivity to prior strength, we varied the BDeu equivalent sample size (ESS = 5 and 10), representing moderate and stronger prior weights, respectively. The results were stable across these settings.

4. Discussion

This analysis focused on determining causal mechanisms underlying liver fat accumulation in adults with and without T2D using comprehensive clinical, imaging, and proteomic phenotyping in the IMI-DIRECT cohorts. We describe complex interactions between insulin dynamics, adipose distribution, and specific circulating proteins in MASLD risk. Several key findings replicated across cohorts, with some effects being specific to sex and disease state.

Our results suggest that BasalISR may be a primary, upstream causal driver of liver fat accumulation, particularly in individuals without diabetes. This highlights the importance of directly assessing β-cell insulin output, rather than relying solely on plasma insulin concentrations. Circulating insulin levels reflect the net result of insulin secretion and hepatic clearance, the latter of which is often impaired in MASLD. In contrast, BasalISR, estimated via C-peptide deconvolution, provides an accurate, physiologically grounded measure of insulin production, independent of clearance variability. This distinction is clinically meaningful, as it allows us to identify individuals with excessive insulin output who may not exhibit elevated plasma insulin due to compensatory hepatic clearance.

This aligns with prior mechanistic evidence indicating that insulin influences hepatic lipid metabolism via de novo lipogenesis, more through hormonal regulation (insulin levels) than because of substrate availability (glucose levels) [56]. While many studies have highlighted associations between high insulin levels and metabolic disease [5], few have provided clear evidence for the causal role of elevated insulin secretion itself. There are also longitudinal and MR studies examining insulin resistance and fatty liver [57,58]. However, there is an absence of published research explicitly focused on determining the causal relationship of basal insulin hypersecretion and liver fat accumulation, which the current work addresses. Moreover, excess adiposity has been shown to increase basal insulin secretion even in the absence of insulin resistance, and dietary interventions that reduce adiposity can lower insulin secretion without necessarily improving insulin sensitivity—further supporting the modifiability of this pathway [59].

Our results concur with prior research showing that even in people with normoglycemia, small changes in insulin levels can affect lipid accumulation and adipogenesis independent of glucose homeostasis [60,61]. The directionality of relationships between insulin secretion, insulin resistance, and liver fat is complex [62], with evidence supporting both insulin resistance driving hyperinsulinemia [63] and vice versa [56,64]. In our Bayesian networks, an arc from BasalISR to HOMA-IR and from OGIS (oral glucose insulin sensitivity index) to BasalISR was observed, reflecting the complexity of these effects. The Bayesian networks also uncovered a bidirectional relationship between liver fat and VAT, suggesting a self-perpetuating cycle of fat redistribution and metabolic stress. In the T2D cohort, while BasalISR remained important, VAT exerted a similar causal influence on liver fat, likely reflecting the shift in disease drivers as insulin secretion capacity diminishes. Crosstalk between VAT and the liver, mediated by free fatty acids entering the portal circulation, likely underpins this relationship [13]. Additionally, a negative association between Clinsb and liver fat was observed, suggesting hepatic steatosis may impair insulin catabolism, exacerbating systemic hyperinsulinemia. Sex-specific differences in adipose distribution emerged: VAT was the strongest predictor of MASLD in females, while SAT had a stronger influence in males with diabetes. These findings highlight the need for sex- and disease stage–specific interventions that consider differences in fat compartment function and metabolic regulation.

Proteomic analyses revealed 34 proteins associated with liver fat in the Bayesian networks, with 11 replicating across both diabetes and non-diabetes cohorts. Proteins such as GUSB, ALDH1A1, LPL, IGFBP1/2, CTSD, HMOX1, FGF21, AGRP, and LEP played central roles in modulating MASLD risk. GUSB, a lysosomal enzyme linked to inflammation and cellular stress, showed the strongest association with MASLD across cohorts and was particularly important in females. The greater influence of GUSB in females is consistent with prior findings suggesting sex-specific regulation, potentially mediated by estrogen and body composition. GUSB activity has been reported to be higher in overweight females (but not males), possibly reflecting hormonal fluctuations and systemic low-grade inflammation during reproductive transitions [65]. ALDH1A1, a key regulator of retinoid metabolism and lipid homeostasis, had strong effects in males and was consistently associated with MASLD risk across metabolic states. Prior evidence indicates higher ALDH1A1 expression and activity in male adipose tissue, contributing to increased visceral adiposity and ectopic fat deposition [66]. LPL and IGFBP1, inversely associated with MASLD, reflect disturbances in lipid clearance and hepatic insulin sensitivity, respectively. Their depletion likely contributes to ectopic fat accumulation and systemic insulin resistance. CTSD and HMOX1, both enzymes involved in oxidative and lysosomal stress responses, were upregulated in relation to liver fat, suggesting a compensatory reaction to lipotoxicity. FGF21, particularly in T2D, appeared as a stress-induced hepatokine, reinforcing its role as a compensatory factor in advanced metabolic dysfunction. Sex-stratified Bayesian inference further revealed that LEP was more influential in men. LEP is central to energy homeostasis and hepatic fat oxidation; its dysregulation may impair hepatic lipid handling, particularly in males [67]. Elevated LEP levels in fatty liver may also indicate hepatic LEP resistance, where LEP fails to stimulate lipid turnover. These findings suggest that hormonal and immunometabolic mechanisms interact with fat distribution to shape liver fat regulation across sex and disease stage.

By estimating posterior probabilities conditioned on single variables and combinations of key markers, we were able to quantify how specific features modify MASLD risk. The greatest increases in MASLD probability were observed following conditioning on combinations of high BasalISR, VAT, GUSB, ALDH1A1, and low IGFBP1 and LPL, demonstrating the additive influence of insulin dynamics and proteomic factors. These probabilistic models highlight the clinical utility of

combining hormonal and proteomic data to stratify populations by MASLD risk. Functional enrichment analysis further supported these findings by highlighting pathways central to MASLD pathophysiology. Proteins clustered in lipoprotein metabolism and LDL particle pathways (impaired lipid clearance), ghrelin processing (appetite regulation), and hormone- or growth factor-related signaling (FGF21, IGFBPs, LEP), suggesting that disrupted lipid handling, altered energy balance, and compensatory endocrine responses collectively contribute to hepatic fat accumulation.

Where causal effect directionality was unclear in the Bayesian networks, we applied bidirectional two-sample MR. This analysis validated thirteen associations and helped assign directionality: HDL → HMOX1, APOM, KITLG: reflecting HDL's antioxidant and signaling roles. TG → LDLR, AGRP and MFGE8, indicating adaptive responses to lipid excess, including enhanced lipid clearance, energy balance regulation, and anti-inflammatory signaling. Conversely, TG showed a negative association with NRP1, suggesting TG-induced suppression of lipid transport and angiogenic pathways. Liver fat → HMOX1, CTSD: supporting a model where hepatic lipid accumulation triggers oxidative and lysosomal stress responses. HbA1c and pancreas fat → lower HMOX1 and CTRC: indicating that chronic hyperglycemia and fat infiltration may suppress protective or digestive pathways. PON3 was the only protein showing negative causal effects on insulin fold change and HDL, consistent with its antioxidant role in limiting lipid oxidation and moderating insulin-driven lipid turnover. These findings enhance the interpretability of our network results and help delineate causal vs. consequential biological processes in MASLD. While many proteomic associations overlap with prior studies, the key contribution of our work is their integration into a directed causal framework, which highlights potential causal pathways and goes beyond simple associations reported elsewhere.

The integration of multiple phenotypic layers (including clinical, imaging, metabolic, and proteomic data) within the causal modeling framework deployed here provides robust inference of directionality and mechanistic insight into complex metabolic interactions. Despite these strengths, we should also highlight several methodological limitations. A proportion of participants were excluded due to missing data, which reduced the sample size and may have introduced selection bias. However, comparison of baseline and complete-case datasets showed negligible differences ($|SMD| < 0.1$), suggesting minimal impact on representativeness. MR assumptions may not hold for all proteins due to pleiotropy or weak instruments, and Bayesian inference depends on the completeness and quality of included variables and assumptions. Even though the estimations of arc strength and direction are conservative, these are probabilistic inferences, and residual confounding or unmeasured common causes cannot be fully excluded. Accordingly, inferred directionality throughout this study reflects probabilistic, model-based relationships supported by convergent evidence from Bayesian network inference and Mendelian randomization, rather than definitive causal effects, and should be interpreted as hypothesis-generating. It should also be emphasized that the model-derived posterior probabilities represent comparative likelihoods within the inferred network, rather than absolute clinical risks, and should therefore be interpreted in a relative, model-based context. These estimates depend on discretization, and although variable rankings were stable across different discretization levels and prior strengths, the numerical probabilities should be interpreted cautiously and not as absolute risk estimates. The use of inverse-normal transformation improved model stability by reducing skewness and outliers, but may attenuate non-linear relationships and influence arc directionality. While SI-HITON-PC addresses collinearity by restricting features to Markov blankets, residual correlations among proteins could still affect arc selection and directionality, introducing further uncertainty.

In terms of generalizability, restricting the analysis to individuals of European ancestry may have increased internal validity but limits generalizability to other ancestries; this is important given that differences in MASLD phenotypes have been observed across ancestries [46]. Replication also remains a key limitation: the differences between the two cohorts (i.e., one is of people without and the other with diabetes at enrollment) mean that some or all the findings that did not replicate may simply reflect disease-state-specific effects. Disentangling this is challenging, as no comparable cohorts are available for replication, but targeted testing of key pathways such as BasalISR → liver fat in other datasets would provide valuable confirmation. To assess the stability of this pathway, we re-estimated the network structure using Tabu search and observed a consistent BasalISR → liver fat link across learners, with comparable arc strengths and directional probabilities in both cohorts. Nonetheless, some variability across algorithms is expected and should be considered when interpreting the results. Finally, the observed direction of BasalISR preceding liver fat accumulation should be interpreted with caution, as hepatic insulin extraction could also contribute to higher circulating insulin levels. However, insulin clearance was included as a separate variable in the network, partly accounting for hepatic insulin handling and helping to mitigate reverse causation effects.

In summary, this study provides a comprehensive systems-level analysis of MASLD pathogenesis. We identify basal insulin hypersecretion as a putative upstream driver, distinct from hyperinsulinemia, in the development of MASLD, independent of glycemic status. BasalISR, as measured by C-peptide deconvolution, offers a direct and clearance-independent index of β-cell output, making it a promising candidate for early risk prediction. Nevertheless, given the cross-sectional design and model-based inference, these findings remain probabilistic with confirmation required through longitudinal and interventional studies. Our analyses also show that insulin dynamics, visceral fat, and circulating proteins such as GUSB and ALDH1A1 interact in sex- and disease-specific ways to influence risk of fatty liver disease. These insights might support precision medicine approaches to MASLD prevention, including targeted metabolic modulation in high-risk individuals and the potential use of plasma proteomics for risk stratification. This work is intended as a systems-level, discovery analysis of deeply phenotyped human cohorts, integrating clinical, imaging, proteomic, and genetic evidence to delineate plausible mechanistic pathways rather than to provide definitive clinical validation. Future research should examine the longitudinal impact of modifying these pathways and extend analyses to more diverse populations.

Ethics statements

Approval for the study protocol was obtained from each of the regional research ethics review boards separately (Lund, Sweden: 20130312105459927, Copenhagen, Denmark: H-1-2012-166 and H-1-2012-100, Amsterdam, Netherlands: NL40099.029.12, Newcastle, Dundee and Exeter, UK: 12/NE/0132). All participants provided written informed consent at enrolment. The research conformed to the ethical principles for medical research involving human participants outlined in the Declaration of Helsinki.

CRediT authorship contribution statement

Natalie N. Atabaki: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Daniel E. Coral:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Hugo Pomares-Millan:** Writing – review & editing, Methodology, Formal analysis. **Kieran Smith:** Writing – review & editing, Methodology. **Harry H. Behjat:** Writing – review & editing, Visualization. **Robert W. Koivula:** Writing – review & editing, Project administration, Investigation, Data curation. **Andrea Tura:** Writing – review & editing, Methodology, Data curation. **Hamish Miller:** Writing – review & editing. **Katherine E. Pinnick:** Writing – review & editing. **Leandro Z. Agudelo:** Writing – review & editing. **Kristine H. Allin:** Writing – review & editing, Investigation. **Andrew A. Brown:** Writing – review &

editing, Data curation. **Elizaveta Chabanova:** Writing – review & editing. **Piotr J. Chmura:** Writing – review & editing, Resources. **Ulrik P. Jacobsen:** Writing – review & editing, Resources. **Adem Y. Dawed:** Writing – review & editing, Methodology, Data curation. **Petra J.M. Elders:** Writing – review & editing, Investigation. **Juan J. Fernandez-Tajes:** Writing – review & editing, Data curation. **Ian M. Forgie:** Writing – review & editing, Project administration. **Mark Haid:** Writing – review & editing, Data curation. **Tue H. Hansen:** Writing – review & editing, Project administration. **Angus G. Jones:** Writing – review & editing, Methodology, Data curation. **Tarja Kokkola:** Writing – review & editing, Data curation. **Sebastian Kalamajski:** Writing – review & editing. **Anubha Mahajan:** Writing – review & editing, Data curation. **Timothy J. McDonald:** Writing – review & editing, Investigation, Data curation. **Donna McEvoy:** Writing – review & editing. **Mirthe Muilwijk:** Writing – review & editing, Data curation. **Konstantinos D. Tsirigos:** Writing – review & editing, Resources. **Jagadish Vangipurapu:** Writing – review & editing. **Sabine van Oort:** Writing – review & editing. **Henrik Vestergaard:** Writing – review & editing, Investigation. **Jerzy Adamski:** Writing – review & editing, Project administration, Funding acquisition. **Joline W. Beulens:** Writing – review & editing, Project administration, Investigation. **Søren Brunak:** Writing – review & editing, Resources, Project administration, Investigation, Funding acquisition. **Emmanouil T. Dermitzakis:** Writing – review & editing, Data curation. **Giuseppe N. Giordano:** Writing – review & editing, Project administration, Data curation. **Ramneek Gupta:** Writing – review & editing, Investigation, Data curation. **Torben Hansen:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Leen M. 't Hart:** Writing – review & editing, Data curation. **Andrew T. Hattersley:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Leanne Hodson:** Writing – review & editing. **Markku Laakso:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Ruth J.F. Loos:** Writing – review & editing. **Jordi Merino:** Writing – review & editing. **Mattias Ohlsson:** Writing – review & editing, Methodology. **Oluf Pedersen:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Martin Ridderstråle:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Hartmut Ruetten:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Femke Rutters:** Writing – review & editing, Investigation, Data curation. **Jochen M. Schwenk:** Writing – review & editing, Project administration, Methodology, Data curation. **Jeremy Tomlinson:** Writing – review & editing. **Mark Walker:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Hanieh Yaghootkar:** Writing – review & editing. **Fredrik Karpe:** Writing – review & editing. **Mark I. McCarthy:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Elizabeth Louise Thomas:** Writing – review & editing, Methodology, Data curation. **Jimmy D. Bell:** Writing – review & editing, Project administration, Methodology, Data curation. **Andrea Mari:** Writing – review & editing, Methodology, Data curation. **Imre Pavo:** Writing – review & editing, Investigation, Funding acquisition, Conceptualization. **Ewan R. Pearson:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Ana Viñuela:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation. **Paul W. Franks:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, NNA used ChatGPT in order to improve the English and flow. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Declaration of competing interest

KHA is an employee of Novo Nordisk. AD works for Novo Nordisk Research Centre Oxford. SB reports ownerships in Hobe Therapeutics, Novo Nordisk, and Eli Lilly & Co. MR owns stock in Novo Nordisk A/S. JMS received speaker travel support from Olink. MMcC is an employee of Genentech and a holder of Roche stock. Within the past five years, PWF has received consulting honoraria from Novo Nordisk A/S, Qatar Foundation, and Zoe Ltd. PWF was an employee of the Novo Nordisk Foundation(2021-–2024). PWF has received investigator-initiated grants (paid to institution) from numerous pharmaceutical companies as part of the Innovative Medicines Initiative of the European Union.

Acknowledgments

The work leading to this publication has received support from the Innovative Medicines Initiative Joint. Undertaking under grant agreement n°115317 (DIRECT, https://directdiabetes.org/), resources of which are composed of financial contribution from the European Union Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution. We thank all the participants and study center staff in IMI DIRECT for their contribution to the study. NNA was funded by the Swedish Research Council (VR - Avtals-ID: 2021-06714_3) for her postdoctoral fellowship (2022–2025) and acknowledges support from the Henning och Johan Throne-Holsts Foundation. DEC is supported by an EFSD/Lilly research grant. PJC, UPJ and SB acknowledge grants from the Novo Nordisk Foundation (NNF14CC0001, NNF17OC0027594, NNF24SA0098829). RJFL is supported by grants from the National Institutes of Health (R01DK107786; R01DK110113; R01HG010297; R01HL142302; R01DK075787; R01HL156991; U01HG011723; R01DK123019; R01HL151152; R01HL158884), the Danish National Research Fund (DNRF161) and the Novo Nordisk Foundation (NNF20OC0059313). JM acknowledges awards from EFSD/Novo Nordisk Foundation Future Leaders Award (no. 0094134) and the European Union (HORIZON-EIC-2023-PATHFINDERCHALLENGES-01-101161509). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them.NNA, RJFL and JM are supported by the Novo Nordisk Foundation grant NNF23SA0084103. HY is supported by Diabetes UK (grant 23/0006598). Researchers affiliated with Lund University Center have been supported by grants from the Swedish Research Council (#2019-01348), Strategic Research Area Exodiab (grant agreement number 2009-1039), Swedish Foundation for Strategic Research (LUDC-IRC, 15-0067), and the European Commission (ERC-CoG_NASCENT - 681742).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.metabol.2026.156552.

References

[1] Younossi ZM, Golabi P, Paik JM, Henry A, Van Dongen C, Henry L. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. Hepatology 2023;77(4):1335–47.

[2] En Li Cho E, Ang CZ, Quek J, Fu CE, Lim LKE, Heng ZEQ, et al. Global prevalence of non-alcoholic fatty liver disease in type 2 diabetes mellitus: an updated systematic review and meta-analysis. Gut 2023;72(11):2138–48.

[3] 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(11):1652–63.

[4] Hyogo H, Yamagishi S, Maeda S, Kimura Y, Ishitobi T, Chayama K. Increased insulinogenic index is an independent determinant of nonalcoholic fatty liver disease activity score in patients with normal glucose tolerance. Dig Liver Dis 2012;44(11):935–9.

[5] Zhang AMY, Wellberg EA, Kopp JL, Johnson JD. Hyperinsulinemia in obesity, inflammation, and cancer. Diabetes Metab J 2021;45(4):622.
[6] Chai SY, Pan XY, Song KX, Huang YY, Li F, Cheng XY, et al. Differential patterns of insulin secretion and sensitivity in patients with type 2 diabetes mellitus and nonalcoholic fatty liver disease versus patients with type 2 diabetes mellitus alone. Lipids Health Dis 2014;13:7.
[7] Meier JJ, Veldhuis JD, Butler PC. Pulsatile insulin secretion dictates systemic insulin delivery by regulating hepatic insulin extraction in humans. Diabetes 2005;54(6):1649–56.
[8] Polonsky KS, Given BD, Van Cauter E. Twenty-four-hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. J Clin Invest 1988;81(2):442–8.
[9] Jones AG, Hattersley AT. The clinical utility of C-peptide measurement in the care of patients with diabetes. Diabet Med 2013;30(7):803–17.
[10] Wahren J, Ekberg K, Johansson J, Henriksson M, Pramanik A, Johansson BL, et al. Role of C-peptide in human physiology. Am J Physiol Endocrinol Metab 2000;278(5):E759–68.
[11] Van Cauter E, Mestrez F, Sturis J, Polonsky KS. Estimation of insulin secretion rates from C-peptide levels. Comparison of individual and standard kinetic parameters for C-peptide clearance. Diabetes 1992;41(3):368–77.
[12] Mari A, Pacini G, Murphy E, Ludvik B, Nolan JJ. A model-based method for assessing insulin sensitivity from the oral glucose tolerance test. Diabetes Care 2001;24(3):539–48.
[13] Scoditti E, Sabatini S, Carli F, Gastaldelli A. Hepatic glucose metabolism in the steatotic liver. Nat Rev Gastroenterol Hepatol 2024;21(5):319–34.
[14] Utzschneider KM, Kahn SE. Review: the role of insulin resistance in nonalcoholic fatty liver disease. J Clin Endocrinol Metab 2006;91(12):4753–61.
[15] Liu Z, Zhang Y, Graham S, Wang X, Cai D, Huang M, et al. Causal relationships between NAFLD, T2D and obesity have implications for disease subphenotyping. J Hepatol 2020;73(2):263–76.
[16] Taylor R. Type 2 diabetes: etiology and reversibility. Diabetes Care 2013;36(4):1047–55.
[17] Boel F, Akimov V, Teuchler M, Terkelsen MK, Wernberg CW, Larsen FT, et al. Deep proteome profiling of metabolic dysfunction-associated steatotic liver disease. Commun Med (Lond) 2025;5(1):56.
[18] Govaere O, Hasoon M, Alexander L, Cockell S, Tiniakos D, Ekstedt M, et al. A proteo-transcriptomic map of non-alcoholic fatty liver disease signatures. Nat Metab 2023;5(4):572–8.
[19] Loesch DP, Garg M, Matelska D, Vitsios D, Jiang X, Ritchie SC, et al. Identification of plasma proteomic markers underlying polygenic risk of type 2 diabetes and related comorbidities. Nat Commun 2025;16(1):2124.
[20] Koivula RW, Heggie A, Barnett A, Cederberg H, Hansen TH, Koopman AD, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: rationale and design of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia 2014;57(6):1132–42.
[21] Koivula RW, Forgie IM, Kurbasic A, Vinuela A, Heggie A, Giordano GN, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: descriptive characteristics of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia 2019;62(9):1601–15.
[22] Thomas EL, Fitzpatrick JA, Malik SJ, Taylor-Robinson SD, Bell JD. Whole body fat: content and distribution. Prog Nucl Magn Reson Spectrosc 2013;73:56–80.
[23] Assarsson E, Lundberg M, Holmquist G, Bjorkesten J, Thorsen SB, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PloS One 2014;9(4):e95192.
[24] Scutari M, Howell P, Balding DJ, Mackay I. Multiple quantitative trait analysis using bayesian networks. Genetics 2014;198(1):129–37.
[25] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;23(R1):R89–98.
[26] Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016;48(5):481–7.
[27] Scutari MD, J-B.. Bayesian networks with examples in R. New York: Chapman and Hall/CRC; 2021.
[28] 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.
[29] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–504.
[30] Atabaki-Pasdar N, Ohlsson M, Vinuela A, Frau F, Pomares-Millan H, Haid M, et al. Predicting and elucidating the etiology of fatty liver disease: a machine learning modeling and validation study in the IMI DIRECT cohorts. PLoS Med 2020;17(6):e1003149.
[31] Leite NC, Villela-Nogueira CA, Cardoso CR, Salles GF. Non-alcoholic fatty liver disease and diabetes: from physiopathological interplay to diagnosis and treatment. World J Gastroenterol 2014;20(26):8377–92.
[32] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8(1):1826.
[33] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25(1):25–9.
[34] Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics 2023;224(1).
[35] Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The reactome pathway knowledgebase 2024. Nucleic Acids Res 2024;52(D1):D672–8.
[36] Ding L, Goossens GH, Oligschlaeger Y, Houben T, Blaak EE, Shiri-Sverdlov R. Plasma cathepsin D activity is negatively associated with hepatic insulin sensitivity in overweight and obese humans. Diabetologia 2020;63(2):374–84.
[37] Malaguarnera L, Madeddu R, Palio E, Arena N, Malaguarnera M. Heme oxygenase-1 levels and oxidative stress-related parameters in non-alcoholic fatty liver disease patients. J Hepatol 2005;42(4):585–91.
[38] Tucker B, Li H, Long X, Rye KA, Ong KL. Fibroblast growth factor 21 in non-alcoholic fatty liver disease. Metabolism 2019;101:153994.
[39] Soldo J, Heni M, Konigsrainer A, Haring HU, Birkenfeld AL, Peter A. Increased hepatic ACE2 expression in NAFL and diabetes-a risk for COVID-19 patients? Diabetes Care 2020;43(10):e134–6.
[40] Cano L, Desquilles L, Ghukasyan G, Angenard G, Landreau C, Corlu A, et al. SARS-CoV-2 receptor ACE2 is upregulated by fatty acids in human MASH. JHEP Rep 2024;6(1):100936.
[41] Borup A, Christensen PM, Nielsen LB, Christoffersen C. Apolipoprotein M in lipid metabolism and cardiometabolic diseases. Curr Opin Lipidol 2015;26(1):48–55.
[42] Sethumadhavan A, Mani M. Kit activates interleukin-4 receptor and effector signal transducer and activator of transcription 6 independent of its cognate ligand in mouse mast cells. Immunology 2020;159(4):441–9.
[43] Palmisano BT, Yu S, Neuman JC, Zhu L, Luu T, Stafford JM. Low-density lipoprotein receptor is required for cholesteryl ester transfer protein to regulate triglyceride metabolism in both male and female mice. Physiol Rep 2021;9(4):e14732.
[44] Rout M, Malone-Perez MW, Park G, Lerner M, Kimble Frazer J, Apple B, et al. Contribution of circulating Mfge8 to human T2DM and cardiovascular disease. Gene 2024;927:148712.
[45] Wilson AM, Shao Z, Grenier V, Mawambo G, Daudelin JF, Dejda A, et al. Neuropilin-1 expression in adipose tissue macrophages protects against obesity and metabolic syndrome. Sci Immunol 2018;3(21).
[46] Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology 2023;77(5):1797–835.
[47] Azizi N, Naghibi H, Shakiba M, Morsali M, Zarei D, Abbastabar H, et al. Evaluation of MRI proton density fat fraction in hepatic steatosis: a systematic review and meta-analysis. Eur Radiol 2025;35(4):1794–807.
[48] Park S, Kwon JH, Kim SY, Kang JH, Chung JI, Jang JK, et al. Cutoff values for diagnosing hepatic steatosis using contemporary MRI-proton density fat fraction measuring methods. Korean J Radiol 2022;23(12):1260–8.
[49] Reeder SB, Sirlin CB. Quantification of liver fat with magnetic resonance imaging. Magn Reson Imaging Clin N Am 2010;18(3):337–57 [ix].
[50] Hartemink AJ. Principled computational methods for the validation discovery of genetic regulatory networks. Massachusetts Institute of Technology; 2001.
[51] Basinska A, Florianczyk B. Beta-glucuronidase in physiology and disease. Ann Univ Mariae Curie Sklodowska Med 2003;58(2):386–9.
[52] Kiefer FW, Vernochet C, O'Brien P, Spoerl S, Brown JD, Nallamshetty S, et al. Retinaldehyde dehydrogenase 1 regulates a thermogenic program in white adipose tissue. Nat Med 2012;18(6):918–25.
[53] Paik J, Kim A, Fogassy K, Snyder JM, Brabb T, Dill-McFarland KA, et al. Weight loss and metabolic effects of an ALDH1A1-specific inhibitor, FSI-TN42, in a diet induced mouse model of obesity. Int J Obes (Lond) 2025;49(3):507–15.
[54] Stanley TL, Fourman LT, Zheng I, McClure CM, Feldpausch MN, Torriani M, et al. Relationship of IGF-1 and IGF-binding proteins to disease severity and glycemia in nonalcoholic fatty liver disease. J Clin Endocrinol Metab 2021;106(2):e520–33.
[55] Shang R, Rodrigues B. Lipoprotein lipase as a target for obesity/diabetes related cardiovascular disease. J Pharm Pharm Sci 2024;27:13199.
[56] Johnson JD. On the causal relationships between hyperinsulinaemia, insulin resistance, obesity and dysglycaemia in type 2 diabetes. Diabetologia 2021;64(10):2138–46.
[57] Tang A, Coster ACF, Tonks KT, Heilbronn LK, Pocock N, Purtell L, et al. Longitudinal changes in insulin resistance in normal weight, overweight and obese individuals. J Clin Med 2019;8(5).
[58] Li R, Wang T, Luo H, Fan Y, Guan Y, Tian Y. Causal effects and mediating pathways of metabolic dysfunction-associated fatty liver disease on novel subtypes of adult-onset diabetes: a two-step Mendelian randomization study. Nutr Metab Cardiovasc Dis 2025;35(6):103976.
[59] van Vliet S, Koh HE, Patterson BW, Yoshino M, LaForest R, Gropler RJ, et al. Obesity is associated with increased basal and postprandial beta-cell insulin secretion even in the absence of insulin resistance. Diabetes 2020;69(10):2112–9.
[60] Dankner R, Chetrit A, Shanik MH, Raz I, Roth J. Basal state hyperinsulinemia in healthy normoglycemic adults heralds dysglycemia after more than two decades of follow up. Diabetes Metab Res Rev 2012;28(7):618–24.
[61] Templeman NM, Skovso S, Page MM, Lim GE, Johnson JD. A causal role for hyperinsulinemia in obesity. J Endocrinol 2017;232(3):R173–83.
[62] Czech MP. Insulin action and resistance in obesity and type 2 diabetes. Nat Med 2017;23(7):804–14.
[63] Boucher J, Kleinridders A, Kahn CR. Insulin receptor signaling in normal and insulin-resistant states. Cold Spring Harb Perspect Biol 2014;6(1).
[64] Janssen J. Hyperinsulinemia and its pivotal role in aging, obesity, type 2 diabetes, cardiovascular disease and cancer. Int J Mol Sci 2021;22(15).
[65] Funk JL, Wertheim BC, Frye JB, Blew RM, Nicholas JS, Chen Z, et al. Association of ss-glucuronidase activity with menopausal status, ethnicity, adiposity, and inflammation in women. Menopause 2023;30(2):186–92.

[66] Ziouzenkova O, Orasanu G, Sharlach M, Akiyama TE, Berger JP, Viereck J, et al. Retinaldehyde represses adipogenesis and diet-induced obesity. Nat Med 2007;13(6):695–702.
[67] Polyzos SA, Kountouras J, Mantzoros CS. Leptin in nonalcoholic fatty liver disease: a narrative review. Metabolism 2015;64(1):60–78.
[68] Brown AA, Fernandez-Tajes JJ, Hong Mg, et al. Genetic analysis of blood molecular phenotypes reveals common properties in the regulatory networks affecting complex traits. Nat Commun 2023;14:5062. https://doi.org/10.1038/s41467-023-40569-3.