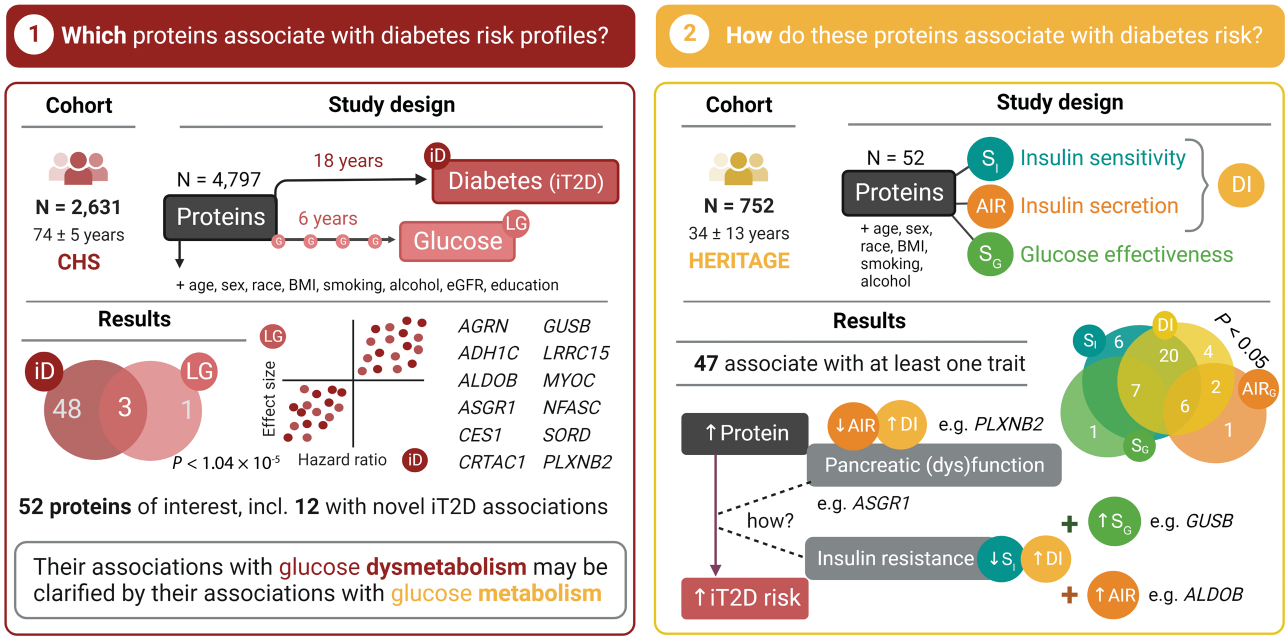
diabetes
American Diabetes Association

Plasma Proteomic Risk Markers of Incident Type 2 Diabetes Reflect Physiologically Distinct Components of Glucose-Insulin Homeostasis

Héléne T. Cronjé, Michael Y. Mi, Thomas R. Austin, Mary L. Biggs, David S. Siscovick, Rozenn N. Lemaitre, Bruce M. Psaty, Russell P. Tracy, Luc Djoussé, Jorge R. Kizer, Joachim H. Ix, Prashant Rao, Jeremy M. Robbins, Jacob L. Barber, Mark A. Sarzynski, Clary B. Clish, Claude Bouchard, Kenneth J. Mukamal, Robert E. Gerszten, and Majken K. Jensen

Diabetes 2023;72(5):666–673 | https://doi.org/10.2337/db22-0628



AIR, acute insulin response; eGFR, estimated glomerular filtration rate; iD, incident diabetes; incl, including; iT2D, incident type 2 diabetes; LG, longitudinal glucose; ±, SD.

Plasma Proteomic Risk Markers of Incident Type 2 Diabetes Reflect Physiologically Distinct Components of Glucose-Insulin Homeostasis

Héléne T. Cronjé,[1] Michael Y. Mi,[2] Thomas R. Austin,[3] Mary L. Biggs,[4] David S. Siscovick,[5] Rozenn N. Lemaitre,[6] Bruce M. Psaty,[6,7] Russell P. Tracy,[8] Luc Djoussé,[9] Jorge R. Kizer,[10,11] Joachim H. Ix,[12] Prashant Rao,[2] Jeremy M. Robbins,[2] Jacob L. Barber,[13] Mark A. Sarzynski,[13] Clary B. Clish,[14] Claude Bouchard,[15] Kenneth J. Mukamal,[16] Robert E. Gerszten,[2] and Majken K. Jensen[1,17]

Diabetes 2023;72:666–673 | https://doi.org/10.2337/db22-0628

High-throughput proteomics allows researchers to simultaneously explore the roles of thousands of biomarkers in the pathophysiology of diabetes. We conducted proteomic association studies of incident type 2 diabetes and physiologic responses to an intravenous glucose tolerance test (IVGTT) to identify novel protein contributors to glucose homeostasis and diabetes risk. We tested 4,776 SomaScan proteins measured in relation to 18-year incident diabetes risk in participants from the Cardiovascular Health Study (*N* = 2,631) and IVGTT-derived measures in participants from the HERITAGE Family Study (*N* = 752). We characterize 51 proteins that were associated with longitudinal diabetes risk, using their respective 39, 9, and 8 concurrent associations with insulin sensitivity index (S_I), acute insulin response to glucose (AIR_G), and glucose effectiveness (S_G). Twelve of the 51 diabetes associations appear to be novel, including β-glucuronidase, which was associated with increased

ARTICLE HIGHLIGHTS

- Plasma proteins are associated with the risk of incident diabetes in older adults independent of various demographic, lifestyle, and biochemical risk factors.
- These same proteins are associated with subtle differences in measures of glucose homeostasis earlier in life.
- Proteins that are associated with lower insulin sensitivity in individuals without diabetes tend to be associated with appropriate compensatory mechanisms, such as a stronger acute insulin response or higher glucose effectiveness.
- Proteins that are associated with future diabetes risk, but not with insulin insensitivity, tend to be associated with lower glucose effectiveness and/or impaired acute insulin response.

[1]Department of Public Health, Section of Epidemiology, University of Copenhagen, Copenhagen, Denmark
[2]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA
[3]Department of Epidemiology, University of Washington, Seattle, WA
[4]Department of Biostatistics, University of Washington, Seattle, WA
[5]New York Academy of Medicine, New York, NY
[6]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA
[7]Department of Epidemiology, Health Systems and Population Health, University of Washington, Seattle, WA
[8]Department of Pathology Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, VT
[9]Division of Aging, Brigham and Women's Hospital and Harvard Medical School, Boston, MA
[10]Cardiology Section San Francisco Veterans Affairs Health Care System, San Francisco, CA
[11]Department of Medicine, Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA
[12]Division of Nephrology-Hypertension, University of California, San Diego, La Jolla, CA
[13]Department of Exercise Science, University of South Carolina, Columbia, SC
[14]Broad Institute of MIT and Harvard, Cambridge, MA
[15]Pennington Biomedical Research Center, Baton Rouge, LA
[16]Division of General Medicine, Beth Israel Deaconess Medical Center, Boston, MA
[17]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

Corresponding authors: Héléne T. Cronjé, toinet.cronje@sund.ku.dk, and Michael Y. Mi, mmi@bidmc.harvard.edu

Received 18 July 2022 and accepted 2 February 2023

This article contains supplementary material online at https://doi.org/10.2337/figshare.22009409.

H.T.C. and M.Y.M. contributed equally to this work. K.J.M., R.E.G., and M.K.J. contributed equally to this work.

diabetes risk and lower S_G, suggesting an alternative pathway to insulin for glucose disposal; and plexin-B2, which also was associated with increased diabetes risk, but with lower AIR_G, and not with S_I, indicating a mechanism related instead to pancreatic dysfunction. Other novel protein associations included alcohol dehydrogenase-1C, fructose-bisphosphate aldolase-B, sorbitol dehydrogenase with elevated type 2 diabetes risk, and a leucine-rich repeat containing protein-15 and myocilin with decreased risk.

The heterogeneity of type 2 diabetes pathogenesis continues to complicate efforts at risk assessment and therapeutic targeting (1,2). High-throughput proteomics provides an avenue to disentangle this complexity through the simultaneous measurement of thousands of proteins using a single plasma specimen (3). Although progress has been made (3–6), prospective proteomics analyses in diverse population-based cohorts remain few and limited. Furthermore, the integration of population-derived diabetes risk with sophisticated physiologic studies may better define how proteins relate to diabetic subphenotypes.

We present a large-scale proteomics association study of incident type 2 diabetes and longitudinal glucose trajectories, the latter of which, to our knowledge, have not been previously evaluated. We further explore prospective protein associations in a complementary analysis of multiple glucose homeostasis traits using a frequently sampled intravenous glucose tolerance test (IVGTT). The combination of these data sets characterizes protein biomarkers of type 2 diabetes risk across physiologically distinct components of glucose homeostasis before the onset of overt disease.

RESEARCH DESIGN AND METHODS

Study Cohorts

The Cardiovascular Health Study (CHS) is a population-based prospective study initiated in 1989 to investigate cardiovascular disease risk factors in adults aged 65 years or older (7). We included 3,188 CHS participants who had frozen plasma samples from the 1992–1993 visit for proteomics assessment, which was used as the baseline for these analyses (8). Diabetes status was ascertained through June 2010 (i.e., 18-year follow-up) using plasma glucose data, medication inventories, and insurance claims (9). Plasma glucose level was measured at the 1992–1993, 1994–1995, 1996–1997, and 1998–1999 visits, and participants were asked to fast before collection except in 1994–1995 (9). We complemented CHS prospective analyses with a cross-sectional investigation of baseline data from the Health, Risk Factors, Exercise Training and Genetics (HERITAGE) Family Study (hereafter, HERITAGE). HERITAGE is a 20-week, single-arm exercise intervention study investigating genetic and cardiometabolic contributors to endurance exercise response in 763 participants recruited within family units (10). Participants underwent a fasted IVGTT comprising of 20 g/m^2 body surface area of glucose administered over 3 min. Sixteen blood samples were collected between 15 min before and 180 min after glucose administration (11) and were used to generate measures of glycemic control using MINMOD software (12).

Plasma Proteomics

Participants in CHS and HERITAGE had 4,796 unique proteins assayed by 4,979 aptamers on the SomaScan 4.0 platform (SomaLogic, Boulder, CO) (13). We excluded 45 nonfasting CHS samples and 17 that failed quality control. Data from both cohorts were normalized using hybridization controls and plate calibrators, and then $\log_2$ transformed and standardized to *z* scores. In CHS, adaptive normalization was additionally applied (8).

Statistical Analysis

We used Cox proportional hazard regression models to test associations of aptamers with incident type 2 diabetes. For this analysis, we additionally excluded 32 participants with missing data and 463 with prevalent diabetes. Age, sex, race, clinic site, and estimated glomerular filtration rate were included as covariates in our base model. Additionally, we adjusted our main model for BMI, smoking status, alcohol consumption status, and education. In exploratory analyses, we tested for interaction effects of sex on statistically significant base-model associations.

As a secondary outcome, we used linear mixed-effect models to assess associations of aptamers with 6-year longitudinal fasting glucose levels. We excluded nonfasting measurements (fasting time <8 h) and adjusted glucose levels for those taking glucose-lowering medications based on their published glucose lowering effects (+30 mg/dL for oral hypoglycemics and +60 mg/dL for insulin or dual therapy [14]). A Bonferroni-adjusted statistical significance threshold of $P < 2.1 \times 10^{-5}$ ($P < 0.05$ adjusted for 4,979 statistical tests) was applied for proteome-wide associations, and a false discovery rate (FDR) of <0.05 was applied for interaction effects.

Proteins that significantly associated with CHS outcomes were further tested for their associations with IVGTT-derived measures in HERITAGE using linear mixed-effect models. The base model included age, sex, and race as fixed effects and family grouping as random effects, and the main model additionally included BMI, smoking status, and alcohol consumption status. Participants who had missing model covariates (n = 11) were excluded. Per the model, we excluded those who had missing trait data. Statistical significance was defined as $P < 0.05$. We performed all analysis using R, version 4.1.3 (15).

Aptamer Specificity

To support aptamer-protein specificity, we incorporated published mass spectrometry–based verification data, *cis* protein quantitative trait loci (pQTLs), and internal data on correlations between SomaScan and Olink proteomic data in HERITAGE (16).

Mendelian Randomization

We tested for causal relationships between proteins and diabetes or glycemic traits using Mendelian randomization (MR)

Table 1—Plasma proteins associated with incident type 2 diabetes in CHS after full covariate adjustment

UniProt identifier	Gene symbol	Target protein name	HR (95% CI)	*P* value	Also associated with		Prior evidence†
					Glucose*	IVGTT‡	
Q9BRK3	*MXRA8*	Matrix-remodeling-associated protein 8	0.65 (0.57–0.73)	1.8×10^{-12}	Yes	S_I, AIR_G, DI, S_G	Yes
Q92729	*PTPRU*	Receptor-type tyrosine-protein phosphatase U	1.32 (1.21–1.43)	2.4×10^{-11}	No	S_I, DI, S_G	Yes
P08236	*GUSB*	β-glucuronidase	1.46 (1.30–1.63)	2.5×10^{-11}	Yes	S_I, DI, S_G	No
P42785	*PRCP*	Lysosomal Pro-X carboxypeptidase	1.46 (1.31–1.64)	5.4×10^{-11}	Yes	S_I, DI	Yes
Q03154	*ACY1*	Aminoacylase-1	1.36 (1.24–1.48)	5.6×10^{-11}	No	S_I, AIR_G, DI	Yes
Q13790	*APOF*	Apolipoprotein F	0.68 (0.61–0.77)	7.2×10^{-11}	Yes	S_I, AIR_G, DI	Yes
P18065	*IGFBP2*	Insulin-like growth factor–binding protein 2	0.66 (0.58–0.75)	2.8×10^{-10}	Yes	S_I, AIR_G, DI	Yes
Q15848	*ADIPOQ*	Adiponectin	0.68 (0.61–0.77)	4.0×10^{-10}	Yes	S_I, DI, S_G	Yes
P52758	*HRSP12*	Ribonuclease UK114	1.34 (1.21–1.47)	3.1×10^{-9}	No	S_I, DI	Yes
P82980	*RBP5*	Retinol-binding protein 5	1.30 (1.19–1.42)	1.1×10^{-8}	No	S_I, DI	Yes
P14384	*CPM*	Carboxypeptidase M	1.32 (1.20–1.46)	1.9×10^{-8}	Yes	S_I, DI	Yes
P55103	*INHBC*	Inhibin β C chain	1.40 (1.25–1.58)	2.3×10^{-8}	Yes	S_I, AIR_G, DI	Yes
P07327	*ADH1A*	Alcohol dehydrogenase 1A	1.28 (1.17–1.40)	3.0×10^{-8}	No	S_I, AIR_G, DI	Yes
P04278	*SHBG*	Sex hormone–binding globulin	0.72 (0.64–0.81)	3.5×10^{-8}	No	S_I, AIR_G, DI	Yes
P10912	*GHR*	Growth hormone receptor	1.40 (1.24–1.58)	6.5×10^{-8}	Yes	S_I, AIR_G, DI	Yes
Q9BU40	*CHRDL1*	Chordin-like protein 1	0.69 (0.61–0.79)	7.1×10^{-8}	Yes	S_I, AIR_G	Yes
P07306	*ASGR1*	Asialoglycoprotein receptor 1	1.29 (1.17–1.41)	7.4×10^{-8}	No	S_I, DI	No
P09467	*FBP1*	Fructose-1,6-bisphosphatase 1	1.28 (1.17–1.39)	8.9×10^{-8}	No	S_I, AIR_G, DI	Yes
Q86TH1	*ADAMTSL2*	ADAMTS-like protein 2	1.30 (1.18–1.43)	1.0×10^{-7}	No	S_I, AIR_G, DI	Yes
Q8WXD2	*SCG3*	Secretogranin-3	0.74 (0.66–0.82)	1.1×10^{-7}	No	S_I, AIR_G, DI	Yes
O94933	*SLITRK3*	SLIT and NTRK-like protein 3	0.75 (0.67–0.83)	1.2×10^{-7}	No	S_I, AIR_G, DI	Yes
Q03167	*TGFBR3*	Transforming growth factor β receptor type 3	0.74 (0.66–0.83)	1.3×10^{-7}	No	S_I, AIR_G, DI	Yes
P23141	*CES1*	Liver carboxylesterase 1	1.29 (1.17–1.43)	2.1×10^{-7}	No	S_I, DI	No
P00326	*ADH1C*	Alcohol dehydrogenase 1C	1.28 (1.16–1.40)	3.1×10^{-7}	No	S_I, DI	No
Q6UXZ4	*UNC5D*	Netrin receptor UNC5D	0.72 (0.63–0.82)	3.1×10^{-7}	No	S_I, AIR_G, DI	Yes
Q9BZR6	*RTN4R*	Reticulon-4 receptor	1.34 (1.20–1.50)	3.9×10^{-7}	Yes	S_I, DI	Yes
Q8N142	*ADSSL1*	Adenylosuccinate synthetase isozyme 1	1.26 (1.15–1.37)	3.9×10^{-7}	No	S_I, DI	Yes
P05062	*ALDOB*	Fructose-bisphosphate aldolase B	1.27 (1.16–1.40)	4.3×10^{-7}	No	S_I, AIR_G, DI	No
O15335	*CHAD*	Chondroadherin	0.76 (0.68–0.84)	4.6×10^{-7}	No	S_I, DI	Yes
O15031	*PLXNB2*	Plexin-B2	1.36 (1.21–1.54)	6.1×10^{-7}	No	S_I, AIR_G	No
Q99727	*TIMP4*	Metalloproteinase inhibitor 4	0.74 (0.66–0.84)	6.2×10^{-7}	No	AIR_G, DI	Yes
P04424	*ASL*	Argininosuccinate lyase	1.25 (1.15–1.37)	8.7×10^{-7}	No	S_I, AIR_G, DI	Yes
O94856	*NFASC*	Neurofascin	1.22 (1.13–1.33)	9.4×10^{-7}	No	S_G	No
P08319	*ADH4*	Alcohol dehydrogenase 4	1.26 (1.15–1.38)	9.9×10^{-7}	No	S_I, AIR_G, DI	Yes
Q99519	*NEU1*	Sialidase-1	1.18 (1.11–1.27)	1.1×10^{-6}	No	S_I, DI	Yes
Q9NQ79	*CRTAC1*	Cartilage acidic protein 1	0.77 (0.69–0.86)	1.1×10^{-6}	No	S_I, DI, S_G	No
Q00796	*SORD*	Sorbitol dehydrogenase	1.29 (1.16–1.42)	1.2×10^{-6}	No	S_I, AIR_G, DI	No

Continued on p. 669

Table 1—Continued

UniProt identifier	Gene symbol	Target protein name	HR (95% CI)	*P* value	Also associated with		Prior evidence†
					Glucose*	IVGTT‡	
O95954	*FTCD*	Formimidoyltransferase-cyclodeaminase	1.28 (1.16–1.41)	1.4×10^{-6}	No	S_I, DI	Yes
Q6NW40	*RGMB*	RGM domain family member B	0.73 (0.64–0.83)	1.5×10^{-6}	No	S_I, AIR_G, DI	Yes
Q7Z3B1	*NEGR1*§	Neuronal growth regulator 1§	0.74 (0.65–0.84)	1.7×10^{-6}	No	S_I, DI, S_G	Yes
Q96EE4	*CCDC126*	Coiled-coil domain-containing protein 126	0.76 (0.68–0.85)	1.7×10^{-6}	No	S_I, AIR_G, DI, S_G	Yes
O00468	*AGRN*	Agrin	1.27 (1.15–1.41)	2.4×10^{-6}	No	S_I, DI	No
Q969E1	*LEAP2*	Liver-expressed antimicrobial peptide 2	1.29 (1.16–1.44)	2.4×10^{-6}	No	No	Yes
Q99972	*MYOC*	Myocilin	0.77 (0.69–0.86)	2.9×10^{-6}	No	No	No
Q13478	*IL18R1*	Interleukin-18 receptor 1	1.21 (1.12–1.32)	3.6×10^{-6}	No	S_I, DI	Yes
Q13332	*PTPRS*	Receptor-type tyrosine-protein phosphatase S	0.78 (0.70–0.87)	3.7×10^{-6}	No	S_I, AIR_G, DI	Yes
Q01581	*HMGCS1*	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	1.27 (1.15–1.40)	3.9×10^{-6}	No	S_I	Yes
Q9Y617	*PSAT1*	Phosphoserine aminotransferase	1.25 (1.14–1.37)	4.0×10^{-6}	No	S_I, AIR_G, DI	Yes
P35442	*THBS2*	Thrombospondin-2	1.26 (1.14–1.39)	4.5×10^{-6}	No	AIR_G, DI	Yes
Q8TF66	*LRRC15*	Leucine-rich repeat-containing protein 15	0.77 (0.69–0.86)	4.8×10^{-6}	No	S_I, DI	No
Q96GG9	*DCUN1D1*	DCN1-like protein 1	1.22 (1.12–1.34)	9.4×10^{-6}	No	S_I	Yes
Q9POT7	*TMEM9*	Transmembrane protein 9	0.99 (0.88–1.10)	0.40	Yes	No	No

The model was adjusted for age, sex, race, clinic site, estimated glomerular filtration rate, BMI, smoking status, alcohol consumption status, and education. AIR_G, acute insulin response to glucose; DI, Disposition index; HR, hazard ratio; S_I: insulin sensitivity; S_G: glucose response. HR presents change in risk per SD increase in log2 protein unit at baseline. *Also associated with longitudinal glucose in the CHS cohort. ‡Also associated with at glycemic traits in the HERITAGE cohort. †Published proteomic associations with incident type 2 diabetes. §Targeted by two distinct aptamers; summary statistics for both are reported in the Supplementary Material.

with the *TwoSampleMR* R package. For genetic instruments of aptamers, we included published *cis* pQTLs in people of European ancestry (17), filtered for $r^2 > 0.1$ within a 10-Mb window based on the 1000 Genomes Project European reference using the ld_clump function from the *ieugwasr* R package. We included genetic instruments from the most comprehensive genome-wide association studies on diabetes (18) and glycemic traits (19), restricted to individuals of European ancestry to increase the validity of linkage disequilibrium assumptions. We reported Wald ratios for aptamers with single-variant instruments, inverse variance–weighted estimates for instruments containing multiple genetic variants, and MR-Egger regression estimates as sensitivity analysis, where applicable. Statistical significance was defined as $P < 0.05$.

Data and Resource Availability

CHS data can be acquired through dbGaP and the National Heart, Lung, and Blood Institute BioLINCC (https://chs-nhlbi.org/CHS_DistribPolicy). HERITAGE data can be accessed from the corresponding authors upon reasonable request.

RESULTS

CHS participants were aged 64 to98 years, 63% were female, and 14% were Black (Supplementary Table 1). During follow-up, 410 participants (15.6%) developed type 2 diabetes (median time to diagnosis, 7.0 [interquartile range = 3.9–11.6] years). The HERITAGE cohort was younger (16 to 66 years) and included more male (45%) and Black (36%) participants (Supplementary Table 1).

Prospective Proteomic Association Analysis

In the main model, 51 proteins were associated with incident type 2 diabetes (Table 1). All but two of the associated aptamers had published *cis* pQTLs; 12 have been validated by mass spectrometry, and 7 had strong ($P > 0.70$) correlations with Olink measurements. In our base model, 59 additional associations were statistically significant ($P < 2.1 \times 10^{-5}$). Of the 110 protein associations observed in these models (Supplementary Table 2), 29 are reported for the first time, to our knowledge, and 81 are supported by prior proteomic investigations. None of the proteomic associations were modified by sex ($P_{FDR} \geq 0.05$, Supplementary Table 3).

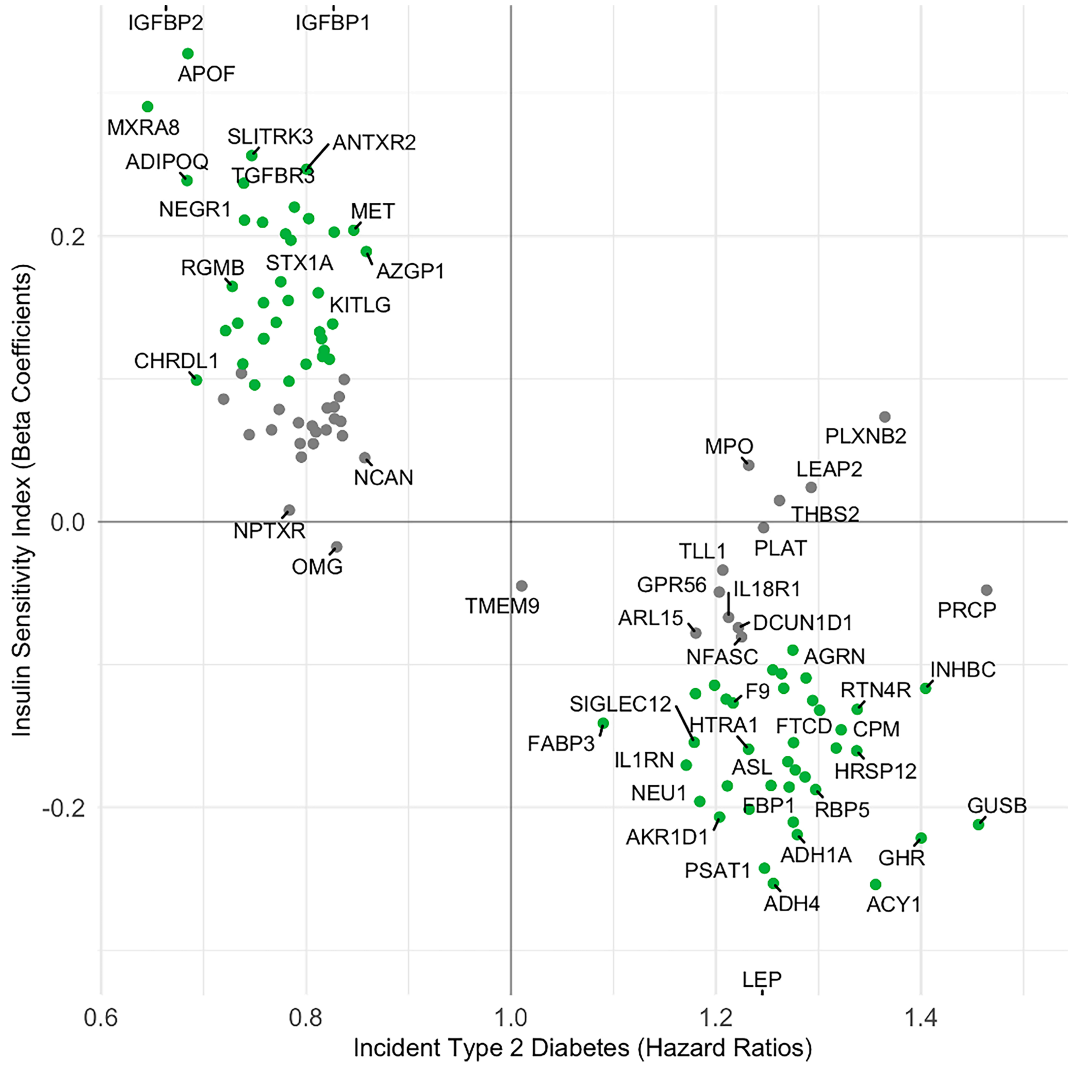


Figure 1— Comparison of association plots of incident type 2 diabetes associations in the CHS cohort and S_I in the HERITAGE cohort. Proteins are labeled using their Entrez gene symbol. All proteins were associated with incident type 2 diabetes at $P < 1.04 \times 10^{-5}$ upon adjustment for age, sex, race, clinic site, and estimated glomerular filtration rate. Green indicates a protein associated with S_I at $P < 0.05$ upon adjustment for age, sex, race, family grouping, BMI, smoking status, and alcohol consumption status. Grey indicates a statistically nonsignificant protein–S_I association. Because of their strong effect sizes, IGFBP1, IGFBP2, and LEP are plotted off the axes in some of the panels to ensure that other proteins can be better visualized.

Protein associations with longitudinal glucose levels largely reflected those observed in the incident diabetes analyses (Supplementary Fig. 1). Of the four significant associations observed in our main longitudinal glucose model and 13 in our base model, only transmembrane protein 9 and heart-type fatty acid–binding protein were not associated with incident type 2 diabetes (Supplementary Table 4).

Integration of IVGTT-Based Associations

We carried forward 112 proteins from the base CHS models and tested their associations with the following four IVGTT traits in HERITAGE (Supplementary Table 6): insulin sensitivity index (S_I), acute insulin response to glucose (AIR_G), disposition index (DI), and glucose effectiveness (S_G). Supplementary Fig. 2 shows the correlation matrix of these traits with available clinical glycemic measures in HERITAGE.

Of the 52 significantly associated proteins from the CHS main models, S_I, AIR_G, D_I, and S_G were respectively associated with 39, 9, 39, and 8 proteins (Table 1, Supplementary Table 5), and 47 (90%) were associated with at least one IVGTT trait. Proteins associated with lower insulin sensitivity were consistently associated with higher diabetes risk, and vice versa (Fig. 1). Within these protein–S_I association clusters, we observed further branching of concurrent protein associations along two insulin-resistance adaptations: AIR_G, representing the acute secretion of insulin to a glucose challenge (Fig. 2*A*), or S_G, an estimate of circulating glucose's inherent impact on suppressing its production and enhancing its disposal (Fig. 3*A*). In addition, we observed diabetes risk–associated proteins that did not concurrently associate with S_I but instead associated solely with AIR_G and S_G (Figs. 2*B* and 3*B*).

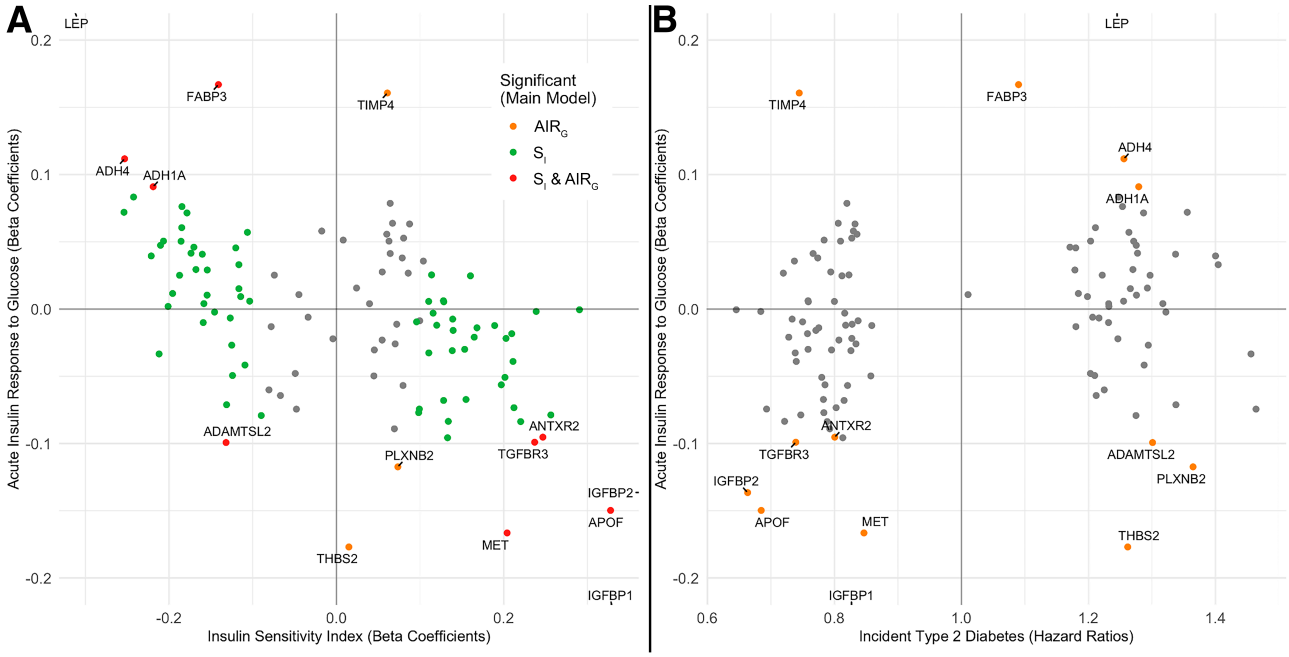


Figure 2—Comparison of association plots of AIR_G. *A*: Protein associations with AIR_G versus with S_I in the HERITAGE cohort. *B*: Protein associations with AIR_G in the HERITAGE cohort versus with incident type 2 diabetes associations in the CHS cohort. Proteins are labeled using their Entrez gene symbol. All proteins were associated with incident type 2 diabetes at $P < 1.04 \times 10^{-5}$ upon adjustment for age, sex, race, clinic site, and estimated glomerular filtration rate. Orange and green indicate a protein associated with AIR_G or S_I, respectively, at $P < 0.05$ upon adjustment for age, sex, race, family grouping, BMI, smoking status, and alcohol consumption status. Red and grey respectively indicate that both or neither association(s) were statistically significant. Because their strong effect sizes, IGFBP1, IGFBP2, and LEP are plotted off the axes in some of the panels to ensure that other proteins can be better visualized.

We identified genetic instruments for 90 aptamers (n = 89 proteins) with significant associations in the CHS base models. We report supportive MR evidence for the relationship of 13 proteins with diabetes, including 7 with concordant direction of effect observed in CHS, and for 21 proteins with fasting glucose, insulin, HbA_{1c}, and/or 2-h glucose levels after an oral glucose tolerance test (Supplementary Tables 6 and 7).

DISCUSSION

We report 112 proteomic associations (including 30 apparently novel and 7 MR-supported associations) with incident type 2 diabetes and/or longitudinal change in glucose in a population-based cohort of U.S. adults aged 65 years or older. When further investigated in an independent younger and healthier cohort, 90% of these proteins shared associations with at least one cross-sectional measure of glucose homeostasis after administration of an intravenous glucose load. That these proteins can demonstrate strong relationships with glucose metabolism and dysmetabolism across two mixed-ancestry cohorts spanning the entire adult age spectrum speaks to the robustness of their involvement. Additionally, by aligning proteins associated with diabetes risk along discrete axes of their role in glucose homeostasis, we identified patterns of association that may offer mechanistic insights.

The first such pattern relates to insulin resistance. We observed a general concordance between protein–diabetes risk relationships and S_I; in other words, proteins reflecting higher insulin sensitivity were also markers of decreased diabetes risk and vice versa. As the initial compensatory mechanism for insulin resistance (lower S_I), insulin production increases (higher AIR_G; i.e., more insulin in the first 10 min of the IVGTT). This is reflected in our data by the contrasting associations some proteins have with S_I and AIR_G, including well-established diabetes risk markers like leptin, as well as proteins previously reported to be associated with diabetes risk without mechanistic context (e.g., apolipoprotein-F, alcohol dehydrogenase-4, alcohol dehydrogenase-1A). Proteins for which we report a novel diabetes association and that cluster within this pattern include fructose-bisphosphate aldolase-B (ALDOB), sorbitol dehydrogenase, and cartilage acidic protein-1. For ALDOB, we report supportive MR evidence of a concordant increased risk for diabetes. Ultimately, some of the proteins in this group may be downstream markers of insulin resistance, including fructose-1,6-bisphosphatase-1 and argininosuccinate lyase, that are transcriptionally regulated by insulin (20,21) but also are associated with increased diabetes risk.

A second pattern of associations relates to pancreatic function. These proteins were associated with diabetes risk and insulin secretion (AIR_G) but not with S_I. We validated the association of metalloproteinase inhibitor-4 with lower

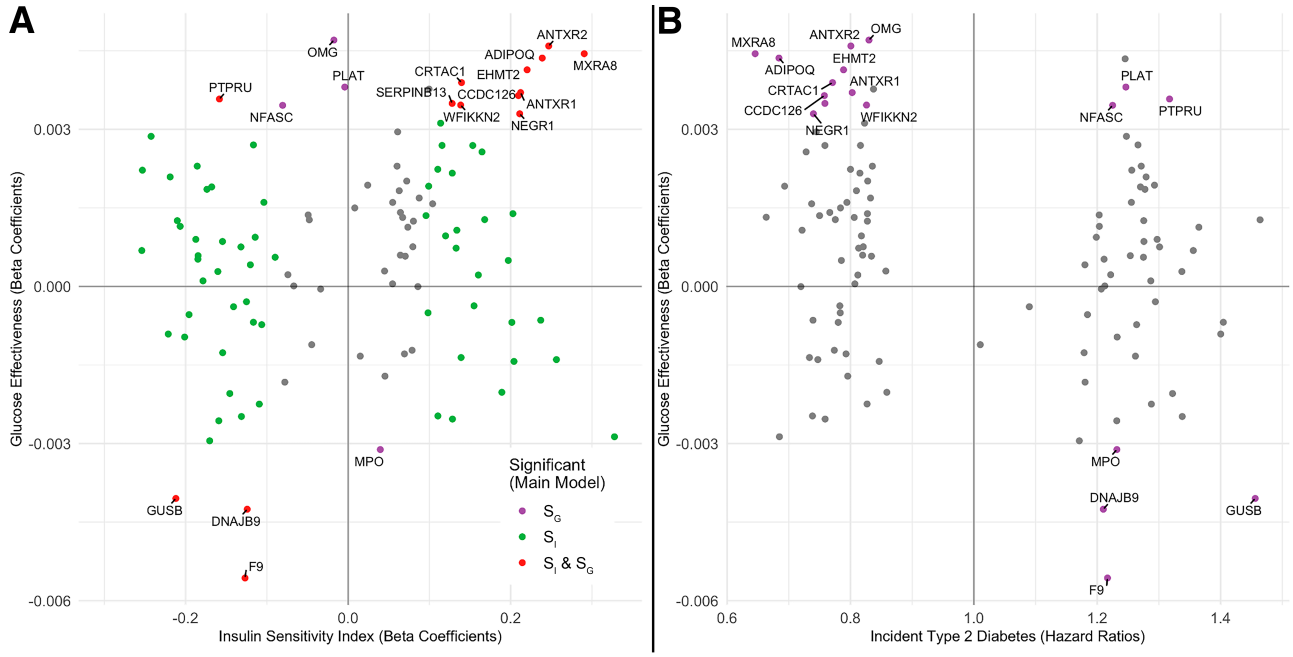


Figure 3—Comparison of association plots of S_G. *A*: Protein associations with S_G versus with S_I in the HERITAGE cohort. *B*: Protein associations with S_G in the HERITAGE cohort versus with incident type 2 diabetes associations in the CHS cohort. Proteins are labeled using their Entrez gene symbol. All proteins were associated with incident type 2 diabetes at $P < 1.04 \times 10^{-5}$ upon adjustment for age, sex, race, clinic site, and estimated glomerular filtration rate. Purple and green indicate a protein associated with S_G or S_I, respectively, at $P < 0.05$ upon adjustment for age, sex, race, family grouping, BMI, smoking status, and alcohol consumption status. Red and grey respectively indicate that both or neither association(s) were statistically significant. Because of their strong effect sizes, IGFBP1, IGFBP2, and LEP are plotted off the axes in some of the panels to ensure that other proteins can be better visualized.

diabetes risk and report it as a possible marker of pancreatic health through its association with insulin response in the absence of an insulin sensitivity–related association. We highlight thrombospondin-2, a matricellular protein associated with impaired wound healing and severity of liver fibrosis in those with diabetes (22,23), as a possible marker of subtle β-cell dysfunction preceding diabetes onset through its association with increased diabetes risk and lower AIR_G, but not with S_I. We observed a similar pattern of association for Plexin-B2, which has previously only been associated with prevalent diabetes. The discordant observational and MR-based associations between Plexin-B2 and diabetes risk supports prior evidence that plexin-B2 likely causally relates to diabetes risk and is causally affected by it (5).

The third cluster relates to insulin-independent glucometabolic pathways. β-Glucuronidase (GUSB) was most strongly associated with incident diabetes within this cluster. GUSB catalyzes the hydrolysis of glucuronic acids and breakdown of mucopolysaccharides, which are elevated in individuals with prevalent diabetes, and is hypothesized to metabolize excess glucose through the glucuronic acid cycle (24). Alongside the positive GUSB–S_G association, GUSB is associated inversely with S_I, suggesting that before the onset of overt diabetes, GUSB may be upregulated to compensate for hyperglycemia as the body becomes more insulin resistant. Although this hypothetical mechanism is not causal for diabetes, it may still have a role akin to sodium-glucose cotransporter-2 inhibitors, which lower glucose levels independent of insulin (25).

Our study has some limitations. Although the SomaLogic assay covers the proteome broadly, it still preselects a panel of proteins, and many proteins remain unexplored. Although we provide data that support aptamer specificity, aptamer performance varies, and the data may inaccurately reflect true protein abundance. Although we validated some of our associations with diabetes risk using published literature, we also report numerous novel associations that require replication in independent prospective cohorts. Targets of potential causal relevance might be particularly useful for further investigation in perturbational studies.

In conclusion, our findings suggest that proteins may capture subtle changes in glucometabolic health years before onset of clinical disease and provide support for further investigation of candidate proteins related to diabetes pathogenesis.

Acknowledgments. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The authors thank Drs. D.C. Rao (Professor Emeritus, Washington University School of Medicine, St. Louis, MO), James S. Skinner (Professor Emeritus, Indiana University, Bloomington, IN), Tuomo Rankinen (Pennington Biomedical Research Center, Baton Rouge, LA), Jacques Gagnon (Marine Biotechnology Research Centre, Rimouski, Quebec, Canada), and the late Drs. Arthur S. Leon and Jack H. Wilmore for contributions to the planning, data collection, and conduct of the HERITAGE project.

Funding. The Cardiovascular Health Study was supported by the National Heart, Lung, and Blood Institute (NHLBI) (contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and 75N92021D00006; and grants U01HL080295, U01HL130114, and R01HL144483), with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided by the National Institute on Aging (grant R01AG023629). The HERITAGE Family Study was initially funded by the NHLBI through the following grants: HL45670 (author C.B., principal investigator [PI]) from 1992 to 2010, HL47317 (D.C. Rao, PI) from 1992 to 2010, HL47321 (J.H. Wilmore, PI) from 1992 to 2003, HL47323 (A.S. Leon, PI) from 1992 to 2003, and HL47327 (J.S. Skinner, PI) from 1992 to 2003. Current funding for HERITAGE is provided to M.A.S. (R01HL146462) and to R.E.G. and M.A.S. (R01NR019628). H.T.C. and M.K.J. are supported by grants from the Novo Nordic Foundation Challenge Program: Harnessing the Power of Big Data to Address the Societal Challenge of Aging (NNF17OC0027812). M.Y.M. is supported by the National Institutes of Health (NIH) (grant 5T32HL007208). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Duality of Interest. B.M.P. serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. L.D. receives research support from Novartis Pharmaceuticals. J.R.K. reports stock ownership in Abbott, Bristol-Myers Squibb, Johnson & Johnson, Medtronic, Merck, and Pfizer. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. H.T.C., M.Y.M., T.R.A., M.L.B., D.S.S., R.N.L., B.M.P., R.P.T., M.A.S., C.B.C., C.B., K.J.M., R.E.G., and M.K.J. designed this study. B.M.P., R.P.T., L.D., J.H.I., M.A.S., C.B.C., C.B., R.E.G., and K.J.M. acquired the data. M.Y.M. and T.R.A. analyzed the data. H.T.C., M.Y.M., T.R.A., B.M.P., L.D., J.R.K., P.R., J.M.R., J.L.B., M.A.S., C.B.C., C.B. K.J.M., R.E.G., and M.K.J. interpreted the data. H.T.C. and M.Y.M. wrote the manuscript. All authors critically reviewed the manuscript for intellectual content and approved it for publication. M.K.J. and R.E.G. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the American Heart Association Scientific Sessions 2022, Chicago, IL, 5–7 November 2022.

References

1. Lin X, Xu Y, Pan X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. Sci Rep 2020;10:14790

2. Mansour Aly D, Dwivedi OP, Prasad RB, et al.; Regeneron Genetics Center. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. Nat Genet 2021;53:1534–1542

3. Williams SA, Kivimaki M, Langenberg C, et al. Plasma protein patterns as comprehensive indicators of health. Nat Med 2019;25:1851–1857

4. Elhadad MA, Jonasson C, Huth C, et al. Deciphering the plasma proteome of type 2 diabetes. Diabetes 2020;69:2766–2778

5. Gudmundsdottir V, Zaghlool SB, Emilsson V, et al. Circulating protein signatures and causal candidates for type 2 diabetes. Diabetes 2020;69:1843–1853

6. Ngo D, Benson MD, Long JZ, et al. Proteomic profiling reveals biomarkers and pathways in type 2 diabetes risk. JCI Insight 2021;6:e144392

7. Fried LP, Borhani NO, Enright P, et al. The cardiovascular health study: design and rationale. Ann Epidemiol 1991;1:263–276

8. Austin TR, McHugh CP, Brody JA, et al. Proteomics and population biology in the Cardiovascular Health Study (CHS): design of a study with mentored access and active data sharing. Eur J Epidemiol 2022;37:755–765

9. Kizer JR, Arnold AM, Benkeser D, et al. Total and high-molecular-weight adiponectin and risk of incident diabetes in older people. Diabetes Care 2012;35:415–423

10. Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, Gagnon J. The HERITAGE family study. Aims, design, and measurement protocol. Med Sci Sports Exerc 1995;27:721–729

11. Boulé NG, Weisnagel SJ, Lakka TA, et al.; HERITAGE Family Study. Effects of exercise training on glucose homeostasis: the HERITAGE Family Study. Diabetes Care 2005;28:108–114

12. Boston RC, Stefanovski D, Moate PJ, Sumner AE, Watanabe RM, Bergman RN. MINMOD Millennium: a computer program to calculate glucose effectiveness and insulin sensitivity from the frequently sampled intravenous glucose tolerance test. Diabetes Technol Ther 2003;5:1003–1015

13. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. Nature Precedings. 2010:1-.

14. Tsapas A, Avgerinos I, Karagiannis T, et al. Comparative effectiveness of glucose-lowering drugs for type 2 diabetes: a systematic review and network meta-analysis. Ann Intern Med 2020;173:278–286

15. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing, 2021

16. Katz DH, Robbins JM, Deng S, et al. Proteomic profiling platforms head to head: leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. Sci Adv 2022;8:eabm5164

17. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet 2021;53:1712–1721

18. Mahajan A, Spracklen CN, Zhang W, et al.; FinnGen; eMERGE Consortium. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 2022;54:560–572

19. Chen J, Spracklen CN, Marenne G, et al.; Lifelines Cohort Study; Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC). The trans-ancestral genomic architecture of glycemic traits. Nat Genet 2021;53:840–860

20. el-Maghrabi MR, Lange AJ, Kümmel L, Pilkis SJ. The rat fructose-1,6-bisphosphatase gene. Structure and regulation of expression. J Biol Chem 1991;266:2115–2120

21. Husson A, Renouf S, Fairand A, Buquet C, Benamar M, Vaillant R. Expression of argininosuccinate lyase mRNA in foetal hepatocytes. Regulation by glucocorticoids and insulin. Eur J Biochem 1990;192:677–681

22. Lee CH, Seto WK, Lui DT, et al. Circulating thrombospondin-2 as a novel fibrosis biomarker of nonalcoholic fatty liver disease in type 2 diabetes. Diabetes Care 2021;44:2089–2097

23. Kunkemoeller B, Bancroft T, Xing H, et al. Elevated thrombospondin 2 contributes to delayed wound healing in diabetes. Diabetes 2019;68:2016–2023

24. Miller BF, Keyes FP, Curreri PW. Increase of serum β-glucuronidase activity in human diabetes mellitus. JAMA 1966;195:189–192

25. Braunwald E. Gliflozins in the management of cardiovascular disease. N Engl J Med 2022;386:2024–2034