



# Clinically Relevant Circulating Protein Biomarkers for Type 1 Diabetes: Evidence From a Two-Sample Mendelian Randomization Study

Diabetes Care 2022;45:169–177 | <https://doi.org/10.2337/dc21-1049>

Nahid Yazdanpanah,<sup>1</sup>  
Mojgan Yazdanpanah,<sup>1</sup> Ye Wang,<sup>2</sup>  
Vincenzo Forgetta,<sup>2</sup> Michael Pollak,<sup>2,3,4</sup>  
Constantin Polychronakos,<sup>5,6,7</sup>  
J. Brent Richards,<sup>2,3,6,8,9</sup> and  
Despoina Manousaki<sup>1,10</sup>

## OBJECTIVE

To identify circulating proteins influencing type 1 diabetes susceptibility using Mendelian randomization (MR).

## RESEARCH DESIGN AND METHODS

We used a large-scale two-sample MR study, using *cis* genetic determinants (protein quantitative trait loci [pQTL]) of up to 1,611 circulating proteins from five large genome-wide association studies, to screen for causal associations of these proteins with type 1 diabetes risk in 9,684 case subjects with type 1 diabetes and 15,743 control subjects. Further, pleiotropy-robust MR methods were used in sensitivity analyses using both *cis* and *trans*-pQTL.

## RESULTS

We found that a genetically predicted SD increase in signal regulatory protein gamma (SIRPG) level was associated with increased risk of type 1 diabetes risk (MR odds ratio [OR] 1.66 [95% 1.36–2.03];  $P = 7.1 \times 10^{-7}$ ). The risk of type 1 diabetes increased almost twofold per genetically predicted standard deviation (SD) increase in interleukin-27 Epstein-Barr virus-induced 3 (IL27-EBI3) protein levels (MR OR 1.97 [95% CI 1.48–2.62];  $P = 3.7 \times 10^{-6}$ ). However, an SD increase in chymotrypsinogen B1 (CTRB1) was associated with decreased risk of type 1 diabetes (MR OR 0.84 [95% CI 0.77–0.90];  $P = 6.1 \times 10^{-6}$ ). Sensitivity analyses using MR methods testing for pleiotropy while including *trans*-pQTL showed similar results. While the MR-Egger suggested no pleiotropic effect ( $P$  value MR-Egger intercept = 0.31), there was evidence of pleiotropy in MR-PRESSO ( $P$  value global test = 0.006).

## CONCLUSIONS

We identified three novel circulating protein biomarkers associated with type 1 diabetes risk using an MR approach. These biomarkers are promising targets for development of drugs and/or of screening tools for early prediction of type 1 diabetes.

Type 1 diabetes is a chronic autoimmune disorder causing destruction of the insulin-producing pancreatic  $\beta$ -cell. With a constantly rising prevalence, it represents a significant public health burden (1). While therapies do not exist to prevent type 1

<sup>1</sup>Research Center of the Sainte-Justine University Hospital, University of Montreal, Montreal, Quebec, Canada

<sup>2</sup>Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

<sup>3</sup>Department of Medicine, McGill University, Montreal, Quebec, Canada

<sup>4</sup>Department of Oncology, McGill University, Montreal, Quebec, Canada

<sup>5</sup>Department of Pediatrics, McGill University, Montreal, Quebec, Canada

<sup>6</sup>Department of Human Genetics, McGill University, Montreal, Quebec, Canada

<sup>7</sup>Centre of Excellence in Translational Immunology, Montreal, Quebec, Canada

<sup>8</sup>Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada

<sup>9</sup>Department of Twin Research, King's College London, London, U.K.

<sup>10</sup>Departments of Pediatrics, Biochemistry and Molecular Medicine, University of Montreal, Montreal, Quebec, Canada

Corresponding author: Despoina Manousaki, [despina.manousaki@umontreal.ca](mailto:despina.manousaki@umontreal.ca)

Received 15 May 2021 and accepted 18 October 2021

This article contains supplementary material online at <https://doi.org/10.2337/figshare.16826689>.

© 2021 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/journals/pages/license>.

diabetes to date, ongoing research aims to identify interventions to preserve  $\beta$ -cell function and delay disease onset before autoantibodies emerge. Strategies to identify individuals at risk for type 1 diabetes as candidates for these interventions involve genotyping for human leukocyte antigen alleles or polygenic risk scores (2). In search of potential predictive biomarkers and drug targets, several circulating molecules have previously been shown to be dysregulated in type 1 diabetes, mostly metabolites and immunophenotypes (3,4). Yet, very few biomarkers exist to predict early development of type 1 diabetes.

Recent advances in large-scale proteomics have enabled the simultaneous measurement of thousands of circulating proteins, and, when combined with evidence from human genetics, such molecules are a potential source for identification of new screening tools and, eventually, drug targets (4). Indeed, in the recent years, *de novo* drug development or repositioning of currently available drugs targeting candidate proteins provided opportunities to deliver new drugs to prevent type 1 diabetes (5).

Observational evidence in search of predictive circulating proteins for risk of type 1 diabetes has been limited due to small sample sizes (6) and failure to identify clearly discriminating proteins (7). Also, the serum proteome can be highly sensitive to sample processing. Moreover, proteomics biomarker research is prone to potential bias due to differences in the sample groups, since the samples often come from highly selective cross-sectional cohorts and frequently lack validation cohorts. Thus, it is challenging to comprehensively characterize and compare the serum/plasma proteome in case-control studies (8). Most proteomics studies for type 1 diabetes looked at differences in the protein profiles of patients with already diagnosed type 1 diabetes (9), but since our goal is prevention, we need to understand if the protein biomarkers change prior to disease onset. Further, case-control studies sampling the proteome after disease onset are subject to reverse causation, in which the disease itself leads to changes in biomarkers, in this case due to the severe metabolic instability introduced at the onset of type 1 diabetes (10). Last, it

is cost-prohibitive to apply throughput proteomic profiling to large cohorts. As a result of the above limitations, the existing epidemiological associations between serum proteins and type 1 diabetes have not generally yielded predictive biomarkers or successful drug targets. Better evidence is needed to understand the role of serum proteins in type 1 diabetes.

Mendelian randomization (MR) is an established genetic epidemiology method that uses human genetics to ascertain whether a given biomarker, such as a serum protein, is implicated in disease etiology (11). MR considers genetic variants strongly associated with an exposure as instrumental variables for that exposure. Since these variants are naturally randomized at conception, they are not generally influenced by environmental confounders. Also, a disease state does not typically change germline genetic variation. As such, MR limits both confounding and reverse causation, which compromise causal inference in observational studies (11).

Two-sample MR (12) allows for scanning of hundreds of biomarkers for causal association with a given outcome, if genome-wide association studies (GWAS) have been done for both the biomarkers and the outcome in separate populations (4). This can happen if the biomarker and the disease have shared causal variants (single-nucleotide polymorphisms [SNPs]) at the same locus or if their causal SNPs are jointly associated due to linkage disequilibrium (LD). To assess the presence of LD, a statistical method called colocalization analysis can be applied (13).

In this study, we applied MR analyses to identify circulating proteins outside the MHC altering the risk of type 1 diabetes. We first identified *cis*-acting SNPs for circulating proteins (meaning SNPs within the genes that encode the proteins [14]) in five large-scale protein GWAS (15–19). Using those variants (*cis*-protein quantitative trait loci [*cis*-pQTL]) as instruments for up to 1,611 proteins, we estimated the effect of genetically altered levels of these proteins on type 1 diabetes risk among 9,684 case subjects with type 1 diabetes and 15,743 control subjects from a large European GWAS (20).

## RESEARCH DESIGN AND METHODS

### Study Exposures

#### *cis*-pQTL GWAS

MR instruments for circulating proteins (*cis*-pQTL) were obtained from five proteomic GWAS (15–19). Circulating proteins in the GWAS by Sun et al. (18), Emilsson et al. (15), Suhre et al. (17), and Yao et al. (19) were measured using the SomaLogic platform, while in the GWAS by Folkersen et al. (16), the O-link platform was used (Supplementary Table 1).

### Study Outcomes

#### Type 1 Diabetes GWAS

To assess the association of *cis*-pQTL with type 1 diabetes risk, we retrieved the effects of these *cis*-pQTL from a large type 1 diabetes GWAS meta-analysis on 9,684 case subjects with type 1 diabetes and 15,743 control subjects from 12 European cohorts by Forgetta et al. (20).

### Statistical Analyses

#### Two-Sample MR

We performed two-sample MR analyses implemented in the “TwoSampleMR” R package (version 0.5.5) (21), using the Wald ratio to estimate the effect of each circulating protein on type 1 diabetes risk. We looked up the lead SNPs (*cis*-pQTL) for association with protein levels in the five proteomic GWAS (15–19). We subsequently queried the lead *cis*-pQTL for association with the outcome in the GWAS by Forgetta et al. (20). To compute the Wald ratios, SNP-exposure effects were used against SNP-outcome effects to compute a single MR estimate reflecting the effect of each protein linked to a *cis*-pQTL on type 1 diabetes risk. Bonferroni correction was used to control for the total number of distinct proteins tested in our MR experiments. The results were presented as MR odds ratio (OR) and 95% CIs for risk of type 1 diabetes per genetically predicted 1-SD change in circulating protein level.

For the *cis*-pQTL of proteins significantly associated with type 1 diabetes in our MR study after Bonferroni correction, we computed the proportion of the variance explained of the respective protein ( $R^2$ ) (Supplementary Materials). Moreover, an inverse-variance weighted approach

and various pleiotropy-robust MR methods (MR-Egger, weighted-median, weighted-mode and MR-PRESSO) were performed in multi-instrument MR analyses, including *trans*-pQTL for a number of candidate proteins (Supplementary Materials).

### MR Assumptions

MR relies upon three main assumptions. First, it requires a robust association between SNPs and the levels of an exposure. This is unlikely to be a concern since all *cis*-pQTL have been associated with their respective protein's level at a genome-wide significant level ( $P \leq 5 \times 10^{-8}$ ) and were defined as the SNPs with the lowest  $P$  value within 1 Mb of the transcription start site of the gene encoding the measured protein. For *cis*-pQTL that were not present in the type 1 diabetes GWAS, SNPs in high LD (defined by an LD  $R^2 \geq 0.8$  in the 1000 Genomes phase 3 European panel) were identified as proxies using SNiPA (<https://snipa.helmholtz-muenchen.de/snipa3/>). As another metric of strength of association of the genetic instrument to the exposure, we calculated the  $F$  statistic of the *cis*-pQTL, which, as a rule of thumb, should be  $>10$  to consider a SNP as a strong MR instrument.

According to the second MR assumption, the genetic instruments (*cis*-pQTL) should not be associated with confounders of the relationship between the exposure and the outcome. A potential source of such confounding is ancestry, which is decreased by ensuring that all case and control subjects in the GWAS are of European ancestry using Eigenstrat (22).

The third MR assumption, known as the exclusion restriction assumption, requires that the genetic variant does not influence the outcome, except through its effect upon the exposure (horizontal pleiotropy). Potential bias due to horizontal pleiotropy is greatly reduced in our study since the instruments in our main MR studies are *cis*-acting SNPs, meaning within 300 kb of the genes that encode the proteins (14). *cis*-pQTL are considered to have a higher biological prior for a direct and definite influence on the protein compared with *trans*-pQTL. *trans*-pQTL are SNPs mapping to regions remote from the gene encoding the

protein of interest, and thus, they are less likely to impact the levels of this protein independently of the levels of the proteins encoded by their respective gene (horizontal pleiotropy). For this reason, we opted to include *trans*-pQTL in MR sensitivity analyses only for the candidate proteins prioritized by our main MR analyses.

To further eliminate potential bias due to horizontal pleiotropy, we used PhenoScanner v2 (23) to assess any reported associations of the *cis*-pQTL of the MR-prioritized proteins with potential confounders (at  $P \leq 1 \times 10^{-8}$ ), and with other circulating protein levels (i.e., *trans*-pQTL). For *cis*-pQTL of our MR-prioritized proteins measured on SomaLogic, we assessed the possibility of potential aptamer-binding effects, in which the presence of protein-altering variants (PAVs) may affect protein measurements (Supplementary Materials). Since PAVs can influence the binding of a protein to an antibody in ELISAs used for protein measurements, this information is important for future validation studies of the candidate proteins.

### Expression QTL Assessment

We evaluated the *cis*-pQTL of the MR-prioritized proteins for effects on gene expression or for evidence of being expression quantitative trait loci (eQTL) using Genotype-Tissue Expression (GTEx) (24) (Supplementary Materials and Supplementary Fig. 1).

### Colocalization Analysis

To assess potential confounding by LD, we checked whether the loci harboring the *cis*-pQTL of the MR-prioritized proteins are also associated with type 1 diabetes using colocalization as implemented in the coloc R package (13). Visualization of colocalization results was performed using the LocusCompareR R package. To estimate the posterior probability of each genomic locus containing a single variant affecting both the protein and the type 1 diabetes, we analyzed all SNPs with minor allele frequency  $>0.01$  within 1 Mb of the *cis*-pQTLs. The analyses were undertaken for each protein with MR evidence for association with type 1 diabetes using summary-level results from Sun et al. (18). Note that a limitation of colocalization is the assumption

of a single shared common causal SNP; however, in reality, genetic loci may contain several causal SNPs (Supplementary Materials).

### Sample Size and Power Analysis

Statistical power in this MR study is a function of the variance of the protein levels explained by the protein-increasing allele of the *cis*-pQTL and the sample size of the type 1 diabetes cohort. The method used has been described previously (25).

### ELISA Validation

In order to be able to test candidate proteins as screening tools for individuals at high risk for type 1 diabetes, we proceeded to validation of commercially available assays for our MR-prioritized proteins. Known quantities of each protein were measured to test accuracy and validity (Supplementary Materials).

### Data and Resource Availability

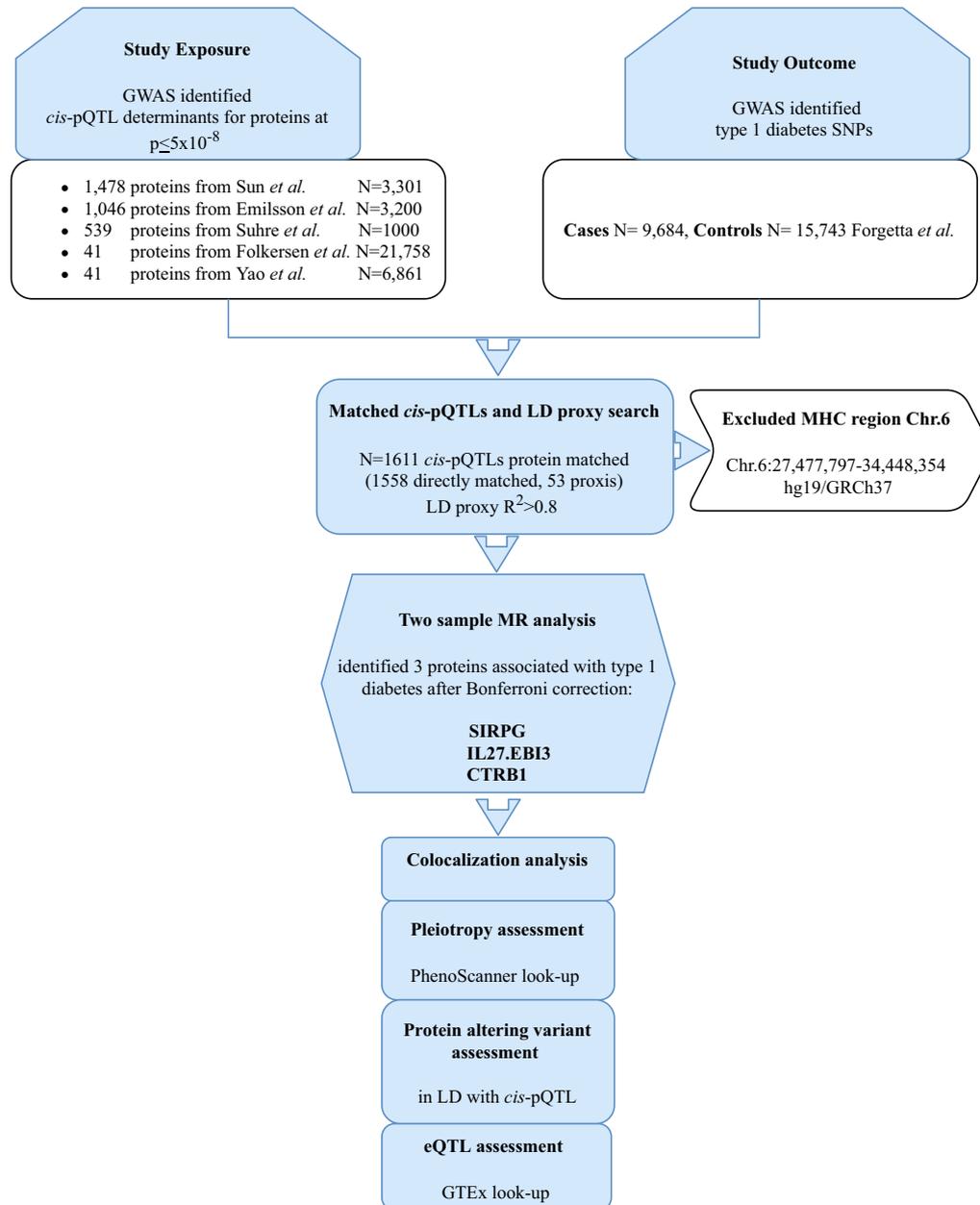
Data from proteomics studies are available from the referenced peer-reviewed studies or their corresponding authors, as applicable. Summary statistics for the type 1 diabetes GWAS are publicly available for download from the GWAS catalog. The statistical code needed to reproduce the results in the article is available upon request.

## RESULTS

### MR Using *cis*-pQTLs and Pleiotropy Assessment

We obtained *cis*-pQTL from five proteomic GWAS (Sun et al. [18],  $N = 3,301$ ; Emilsson et al. [15],  $N = 3,200$ ; Suhre et al. [17],  $N = 1,000$ ; Folkersen et al. [16],  $N = 21,758$ ; and Yao et al. [19],  $N = 6,861$ ) and used them as instruments to evaluate the causal role of circulating proteins on type 1 diabetes. After excluding proteins in the MHC region and LD clumping ( $r^2 < 0.3$ ), 2,028 unique *cis*-pQTL were associated with protein levels (at  $P \leq 5 \times 10^{-8}$ ) (Fig. 1). After querying the *cis*-pQTL in the type 1 diabetes GWAS, 1,611 distinct *cis*-pQTL ( $N = 1,558$  directly matched and  $N = 53$  proxies) were retrieved (Supplementary Table 2).

Using these 1,611 *cis*-pQTLs as genetic instruments for their respective proteins, our MR analyses suggested, after Bonferroni correction ( $P = 0.05/1,611$  or  $3 \times$



**Figure 1**—Flowchart displays the selection process to identify *cis*-pQTL as instruments for our MR study.

$10^{-5}$ ), associations with risk of type 1 diabetes for three circulating proteins: signal regulatory protein gamma (SIRPG), interleukin-27 Epstein-Barr virus-induced 3 (IL27.EBI3), and chymotrypsinogen B1 (CTRB1) (Table 1). Notably, using *cis*-pQTL effect estimates on protein levels from Emilsson *et al.* (15), increased genetically predicted SIRPG levels were associated with increased risk of type 1 diabetes (MR OR 1.66 [95% CI 1.36–2.03];  $P = 7.1 \times 10^{-7}$ ) per SD increase in SIRPG. We cross-validated this finding using estimates for SIRPG from Sun *et al.* (18) (Table 1). Moreover,

using the data from Emilsson *et al.* (15), the risk of type 1 diabetes was increased almost twofold per genetically predicted SD increase in IL27.EBI3 levels (MR OR 1.97 [95% CI 1.48–2.62];  $P = 3.7 \times 10^{-6}$ ). However, using *cis*-pQTL effect estimates from Emilsson *et al.* (15), a genetically predicted SD increase in CTRB1 was associated with decreased risk of type 1 diabetes (MR OR 0.84 [95% CI 0.77–0.90];  $P = 6.1 \times 10^{-6}$ ). The effect for CTRB1 was cross-validated using estimates from Sun *et al.* (Table 1). All *cis*-pQTL for the three candidate proteins had an *F* statistic

$>10$ , implying that they were strong MR instruments.

We next assessed whether the *cis*-pQTL associated with SIRPG levels in Emilsson *et al.* (15) and Sun *et al.* (18) (rs2281808 and rs6043409) were associated with other phenotypes in PhenoScanner (23). We found that rs6043409 was strongly associated with circulating serum SIRPG levels ( $P = 8.13 \times 10^{-24}$ ) and was not associated with any other traits or proteins (Supplementary Tables 3 and 4), which reduces the probability that our MR estimate for SIRPG was driven by horizontal pleiotropy. Querying

**Table 1—MR results for circulating proteins obtaining a significant MR *P* value (after Bonferroni correction) for association with type 1 diabetes**

Protein	Chr.	Position	rs number <i>cis</i> -pQTL	EAF	EA	MR $\beta$ (SE)	MR OR	95% CI	<i>P</i> value (unadjusted)	<i>R</i> <sup>2</sup>	<i>F</i> statistics	Source
SIRPG	20	1610551	rs2281808	0.66	C	0.51 (0.10)	1.66	1.36 – 2.03	$7.1 \times 10^{-7}$	0.03	83.9	Emilsson et al. (15)
SIRPG	20	1610551	rs6043409	0.66	G	0.47 (0.10)	1.60	1.32 – 1.92	$1.1 \times 10^{-6}$	0.03	106.67	Sun et al. (18)
IL27.EBI3.2829.19.2	16	28503533	rs181209	0.33	T	0.68 (0.15)	1.97	1.48 – 2.62	$3.7 \times 10^{-6}$	0.01	48.21	Sun et al. (18)
CTRB1	16	75221319	rs8051363	0.74	G	−0.22 (0.05)	0.80	0.73 – 0.88	$6.1 \times 10^{-6}$	0.10	374.6	Emilsson et al. (15)
CTRB1	16	75221319	rs8051363	0.70	G	−0.18 (0.04)	0.84	0.77 – 0.90	$6.1 \times 10^{-6}$	0.17	696.88	Sun et al. (18)

Source indicates the protein GWAS providing the estimate of the effect of the *cis*-pQTL on protein level. Results express changes in type 1 diabetes risk per 1 SD increase in protein level. EA, effect allele; EAF, effect allele frequency;  $\beta$ , effect on type 1 diabetes.

the *cis*-pQTL for IL27.EBI3 (rs181209) in PhenoScanner, we did not find evidence of pleiotropic associations, other than its association with circulating IL27.EBI3 ( $P = 8.71 \times 10^{-12}$ ) (Supplementary Table 5). The *cis*-pQTL for CTRB1 (rs8051363) was strongly associated with circulating CTRB1 ( $P = 1.58 \times 10^{-157}$ ), and two other proteins, namely inactive pancreatic lipase-related protein 1 and retinoic acid receptor responder protein 1, suggesting potential pleiotropic effects (Supplementary Table 6). Given the known involvement of these proteins in multiple physiological processes, this specific MR analysis may suffer from horizontal pleiotropy.

Further, we performed multi-instrument MR analysis for the three candidate proteins incorporating *trans*-pQTL whenever available. Of the three proteins, we found available *trans*-pQTL in Sun et al. (18) only for CTRB1 ( $n = 11$  SNPs). While the inverse-variance weighted estimate and weighted-median were significant, the MR-Egger and weighted-mode methods provided nonsignificant results. Although the MR-Egger intercept suggested no pleiotropic effect ( $P$  value intercept = 0.31), the  $P$  value of the MR-PRESSO global test was 0.006, implying the presence of horizontal pleiotropy (Supplementary Table 9A and B).

### Colocalization Analysis

To assess confounding due to LD, we tested the probability that the genetic determinants of SIRPG levels were shared with those of type 1 diabetes using colocalization. We found that the posterior probability that SIRPG levels and type 1 diabetes shared a single causal signal was low (posterior

probability 4 or  $H_4 = 0.025$ ), while the two traits were linked through two independent SNPs in the 1-Mb locus around the rs6043409 *cis*-pQTL ( $H_3 = 0.98$ ) (Fig. 2A). Colocalization analysis showed similar results between type 1 diabetes and either IL27.EBI3 ( $H_3 = 0.96$ ) (Fig. 2B) or CTRB1 level ( $H_3 = 1.00$ ) (Fig. 2C). For all three proteins, these results suggest that there is not a single shared causal signal but rather two independent SNPs in the same locus, implying a possible bias due to LD.

### PAV Assessment

We assessed if the *cis*-pQTLs for the MR-prioritized proteins were PAVs or in LD ( $R^2 > 0.8$ ) with PAVs. The rs6043409 (SIRPG) is a missense variant and therefore could be subject to potential binding effects. rs2281808, the *cis*-pQTL for SIRPG in the GWAS by Emilsson et al. (15), is an intronic variant in LD with the missense SNP rs60443409 ( $R^2 = 0.94$ ). rs8051363 (CTRB1) is an intronic variant that is not in LD with missense variants. rs181209 (IL27.EBI3) is an intronic variant in almost perfect LD ( $R^2 \geq 0.97$ ) with rs181206 and rs56354901, which are splicing region variants and could also be subject to potential binding effects (Supplementary Table 10).

### eQTL Assessment

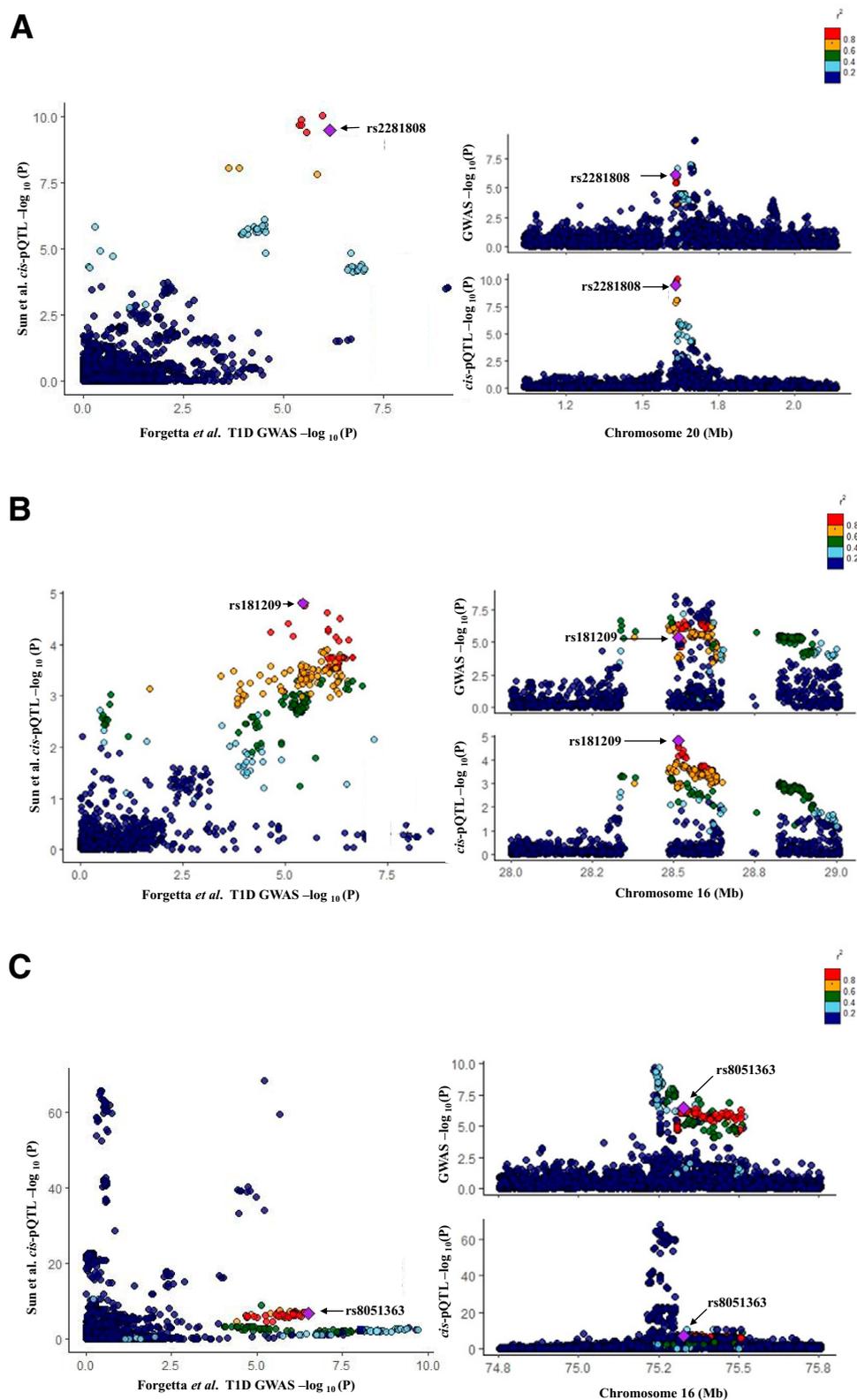
Based on our GTEx search (24), rs6043409 (*cis*-pQTL for SIRPG) affects *SIRPG* expression in the following tissues: whole blood ( $P = 1.7 \times 10^{-39}$ ) and spleen ( $P = 4.0 \times 10^{-12}$ ). Both rs6043409 and rs2281808 in *SIRPG* are in high LD ( $R^2 = 0.94$ ). We found that rs2281808 was an eQTL for *SIRPG* in the following tissues: whole blood ( $P = 5.8 \times$

$10^{-33}$ ) and spleen ( $P = 5.8 \times 10^{-9}$ ). With regards to *IL27.EBI3*, we found that the *cis*-pQTL rs181209 was an eQTL for *IL27* in liver ( $P = 1.8 \times 10^{-33}$ ), while rs8051363 was an eQTL for *CTRB1* in pancreas ( $P = 2.7 \times 10^{-23}$ ) (Supplementary Fig. 1).

Main MR analysis showed that per each SD increase in circulating CTRB1 level instrumented by its *cis*-pQTL, there was a decrease in the risk of type 1 diabetes. While the G allele of the rs8051363 pQTL increased circulating CTRB1 level, when the same SNP is acting as an eQTL its G allele decreases *CTRB1* expression in pancreas. This opposite direction of the effect of rs8051363 when it is acting as pQTL versus eQTL for CTRB1 argues in favor of a possible PAV effect. Nevertheless, our assessment for PAV for the CTRB1 rs8051363 was negative (Supplementary Tables 9 and 10). Contrarily to CTRB1, the T allele of the rs181209 pQTL increases circulating IL27.EBI3 level, while it also increases *IL27.EBI3* expression in liver. Also, carrying the G allele of the rs6043409 pQTL leads to higher circulating SIRPG level and increased *SIRPG* expression in pancreas.

### MR Power Calculation

We computed the power of our MR studies at an  $\alpha = 0.05/1,611 = 3 \times 10^{-5}$  (Bonferroni correction). Using estimates from an observational study (8), our power to detect an association between a specific protein level and type 1 diabetes was estimated at 100%, since the median variance explained per protein by *cis*-pQTL is 4.6% (17,18), and the portion of case subjects with type 1 diabetes is 38% in our consortium of



**Figure 2**—A: Colocalization analysis of the *cis*-pQTL for SIRPG level and type 1 diabetes risk: the *cis*-pQTL rs6043409 is shown as a purple diamond. The lead SNP is shown with an arrow. The  $P$  values of the SNPs in the SIRPG locus were extracted from the GWAS by Sun et al. (18) and from the type 1 diabetes GWAS by Forgetta et al. (20). B: Colocalization analysis of the *cis*-pQTL for IL27.EBI3 level and type 1 diabetes risk: the *cis*-pQTL rs181209 is shown as a purple diamond. The lead SNP is shown with an arrow. The  $P$  values of the SNPs in the IL27.EBI3 locus were extracted from the GWAS by Sun et al. (18) and from the type 1 diabetes GWAS by Forgetta et al. (20). C: Colocalization analysis of the *cis*-pQTL for CTRB1 level and type 1 diabetes risk: the *cis*-pQTL rs8051363 is shown as a purple diamond. The lead SNP is shown with an arrow. The  $P$  values of the SNPs in the CTRB1 locus were extracted from the GWAS by Sun et al. (18) and from the type 1 diabetes GWAS by Forgetta et al. (20). The dots in the scatter plots are colored according to their LD to the colocalization lead variant (*cis*-pQTL). LD is calculated based on the 1000 Genomes phase 3 European references.

25,427 individuals (20) (Supplementary Table 7).

### Results of the ELISA Validation

We validated available commercial assays for each of the three candidate proteins (SIRPG, IL27.EBI3, and CTRB1). Interassay percentage coefficients of variation were  $<10$ , and recovery rates were close to 100%. These results demonstrate that the assays were well-run and validate them for measurement of our MR-prioritized proteins (Supplementary Table 8).

### CONCLUSIONS

We used two-sample MR studies to evaluate causal effects of 1,611 plasma proteins from five large proteomics GWAS (15–19) on type 1 diabetes in 9,684 case subjects and 15,743 control subjects (20). We found convincing evidence that genetically altered levels of SIRPG, IL27.EBI3, and CTRB1 protein levels are causally associated with type 1 diabetes susceptibility. Notably, per SD increase in their circulating level, SIRPG was associated with an  $\sim 60\%$  increase in risk of type 1 diabetes, IL-27 doubled this risk, whereas CTRB1 had a protective effect with an  $\sim 20\%$  decrease in type 1 diabetes risk. Our *in silico* follow-up analysis provided evidence that the genes annotated to the *cis*-pQTL for SIRPG, IL-27, and CTRB1 are mainly expressed in immune cells, pancreas and liver. Finally, we successfully validated immune assays for measurement of each of the three candidate proteins.

In this study, we screened for association with type 1 diabetes only proteins outside the MHC for two reasons: first, HLA (26) haplotypes including genes in the MHC are already used as biomarkers for type 1 diabetes screening, and this genetic association (one of the strongest among complex traits) is mostly mediated by PAVs; second, the complex patterns of LD in the MHC region render the assessment of bias due to LD particularly difficult.

The development of serum biomarkers for type 1 diabetes was initiated with the identification of autoantibodies against islet cell antigens (27). Proteomics technologies such as liquid chromatography-mass spectrometry provided additional insight into the pathogenesis of type 1 diabetes and

enhanced identification of biomarkers. The advent of high-throughput and sensitive “-omics” technologies have allowed large scale profiling of protein levels in patients with type 1 diabetes (28) to discover biomarkers. Once biomarkers are identified, widely used methods such as targeted mass spectrometry (29) or ELISAs could be implemented to measure particular proteins of interest. However, care must be taken to avoid the impact of sequence variation (PAV) in the endogenous proteins. A strength of the SomaLogic aptamer methodology is the simultaneous measurement of many proteins. This is a key value for biomarker discovery, but more targeted methods may be adequate for validation and potential clinical use. Also, the impact of sequence variants is likely harder to predict in aptamer-based methods such as the SomaLogic.

Our analyses have prioritized SIRPG as a functionally relevant biomarker for type 1 diabetes. This protein is a member of the SIRP family and is expressed on the surface of most T cells and some B cells. Decreased SIRP $\gamma$  expression on CD8<sup>+</sup> T cells has been associated with a lowered activation threshold and enhanced cytotoxic potential, suggesting SIRP $\gamma$  could function as a negative regulator of activation (30). Dysregulated *SIRPG* expression has been demonstrated in other autoimmune conditions, such as systemic lupus erythematosus (31). Binding of SIRP $\gamma$  with its ligand CD47 has been shown to facilitate cell adhesion, while engagement of SIRP $\gamma$  on T cells by CD47 on antigen-presenting cells may enhance antigen-specific T-cell proliferation (31). Further, polymorphisms in *SIRPG* gene may interfere with transcription factors important in T-cell development (30).

The two *cis*-pQTL of SIRPG have been associated with type 1 diabetes in GWAS: rs2281808 (C > T; intronic) and rs6043409 (G > A; A263 V) (31,32). The T allele of rs2281808 is a risk allele for type 1 diabetes (33). There is evidence that the rs2281808 T variant is associated with a reduction in SIRP $\gamma$  expression on human T cells, leading to a hyperactivated state with lower activation threshold in healthy donors (31), suggesting that disruption in SIRP $\gamma$  levels may lead to immune dysregulation

in individuals with autoimmune diseases. These results are in the opposite direction of those of our MR study, which indicated that higher levels of SIRPG are linked to an increased risk of type 1 diabetes. Longitudinal case-control studies are required to clarify the role of SIRPG in type 1 diabetes.

IL-27 is another protein with prior evidence of association with type 1 diabetes. It is a heterodimer composed of two noncovalently associated subunits, including EBI3 and IL27p28 (34). IL-27 has been shown to have both pro- and anti-inflammatory activities in several autoimmune diseases (34). It was initially found to enhance the function of type 1 helper T cells and suppress type 2 helper T cells and regulatory T cells and to have an anti-inflammatory role (35). Evidence suggests that IL-27 can be secreted from adipocytes in response to inflammatory stress (36), and its levels increase in patients with type 2 diabetes after Roux-en-Y gastric bypass (37). Notably, IL-27 has been proposed as a drug target in type 1 diabetes (26). *In vitro* and animal studies suggest that IL-27 is involved in the pathogenesis of type 1 diabetes (38). In mice, IL-27 is critical for type 1 diabetes development and influences differentiation of CD4 and CD8 T cells in pancreatic islets (39).

*IL27* (encoding the p28 subunit) is also a known type 1 diabetes locus (33). The missense variant (rs181206) in *IL27* is in strong LD with the intronic variant rs181209 in our study. rs181206 is associated with elevated STAT1 and interferon regulatory factor 1 transcript levels in human CD4, as well as increased IL-27 function (40). In a recent study, circulating IL-27 level was suggested as a potential biomarker of early cardiac dysfunction in women with type 1 diabetes and heart disease (41). Collectively, these studies suggest a potential role of *IL27* variants in type 1 diabetes pathogenesis and effects of IL-27 on risk of type 1 diabetes in the same direction as the one observed in our MR analysis.

*CTRB1* encodes a member of the serine protease family, a component of digestive enzymes secreted by the pancreas, and is located adjacent to a related chymotrypsinogen gene, *CTRB2* (42). In some human populations, an alternate haplotype is present, which inverts haplotype exon 1 and flanking sequence in *CTRB1* and *CTRB2* (42).

Shared variants at the *CTRB1* locus have shown opposite effects on type 2 and type 1 diabetes risk and have allelic effects on islet regulatory activity (42). Interestingly, we observed the rs8051363 has an opposite effect on *CTRB1* circulating levels compared with its effect on *CTRB1* expression, implying a possible PAV effect. Nevertheless, our assessment for PAV for the *CTRB1* rs8051363 was negative. Primary cleavage specificity and catalytic efficiency of *CTRB1* and *CTRB2* are different, with *CTRB2* being more active on most substrates (43). This finding is in line with previous evidence on risk variants affecting *CTRB1/2* expression in pancreas and pancreatic islets and glucagon-like peptide 1–mediated insulin secretion (42). In another study, the association of the lead SNP rs8051363 with *CTRB1/2* expression was similar but smaller, supporting a causal role in chronic pancreatitis (44).

The *CTRB1/2* locus is associated with a reduced response to treatment with dipeptidyl peptidase 4 inhibitors in patients with type 2 diabetes. Carriers of the G allele at rs7202877 showed increased chymotrypsin mRNA expression and activity in the pancreas and feces, respectively. Accordingly, a change in the isoform expression ratio likely affects chymotrypsin activity in the pancreas. The rs8051363 in *CTRB1* was linked with both amylase and lipase serum levels, used to diagnose pancreatitis, and was previously shown to be associated with alcoholic chronic pancreatitis (44). Our results are consistent with a previous study showing weak evidence for colocalization between eQTL and GWAS signals for *CTRB1/2* locus (rs7202877; posterior probability 35.2%) in both type 1 and type 2 diabetes (45). Also, a previous report correlated risk variants with *CTRB1/2* expression in pancreas and pancreatic islets and glucagon-like peptide 1–mediated insulin secretion (42). In this direction, it has been positioned that *CTRB1/2* locus affects diabetes susceptibility via the incretin pathway and is an overlapping locus between type 1 and type 2 diabetes (46). The above evidence supports the protective effect of circulating *CTRB1* on risk of type 1 diabetes, which is in agreement with our MR results.

Our follow-up analyses showed that the three candidate proteins were expressed in immune cells, pancreas, and liver, which highlights the complex

multicellular interactions between pancreatic  $\beta$ -cells and immune cells during the preclinical stage of type 1 diabetes. As such, serum proteomic signatures in individuals who will develop type 1 diabetes may reflect ongoing processes in the aforementioned tissues. Despite type 1 and type 2 diabetes being considered genetically distinct, there is evidence supporting an overlap in their genetic etiology (47). As such, genetic susceptibility to altered protein pathways provides a significant resource not only for novel therapies, but also for drug repurposing in diabetes.

This study used the largest protein GWAS consortia to date in terms of number of tested proteins and sample sizes (15–19), in addition to the largest type 1 diabetes GWAS (20). The main strength of our study is that it allowed for assessment of the causal role of biomarkers using MR, an approach that is robust to unmeasured error due to confounding and reverse causation. Also, SIRPG and *CTRB1* proteins were measured in both GWAS by Emilsson et al. (15) and Sun et al. (18) using an aptamer-based technique; since both GWAS demonstrated the presence of genome-wide significant *cis*-pQTL for both proteins, this minimizes the possibility of measurement error. Our main MR analysis limits the risk of horizontal pleiotropy by using *cis*-pQTL as instruments since these SNPs may directly influence the transcription and/or translation of these genes (48). Presence of horizontal pleiotropy including *trans*-pQTL in the MR study for *CTRB1* confirmed the major advantage of selecting *cis* instruments with high biological plausibility. Hence, by selecting *cis* SNPs, there is reduced risk of pleiotropy, which strengthens our MR results, indicating that specific proteins directly alter type 1 diabetes risk (14). The large sample sizes of both the GWAS ensure adequate statistical power to show clinically relevant associations between protein levels and risk of type 1 diabetes. The Bonferroni-adjusted *P* value might be overly conservative here, given the nonindependence of circulating proteins; however, it minimizes type 1 error.

There are a few considerable limitations in our study. Firstly, our approach tested effects of circulating protein levels and could not capture potential tissue-specific effects. Secondly, we did

not validate the identified protein biomarkers as predictors of development of type 1 diabetes in independent longitudinal case-control cohorts, since this exceeded the scope of the present work. However, we validated commercially available assays for the three candidate proteins, which ensures the feasibility of such validation studies. Further, we confirmed the robustness of our findings by cross-validating the results using instruments from the GWAS from either Emilsson et al. (15) or Sun et al. (18). Thirdly, although the *cis*-pQTL for the three candidate proteins were significantly associated with type 1 diabetes, based on the colocalization results, the three SNPs are likely in LD with causal variants for type 1 diabetes in these loci and are not causal themselves. Nonetheless, the fact that all three protein loci harbored SNPs likely to be causal for type 1 diabetes is reassuring. Finally, the findings of our MR studies are based on GWAS of European ancestry and cannot be generalized to other ancestries. This underscores the importance of large-scale biobanks including individuals of non-European ancestry.

In conclusion, our MR study provided evidence that genetically altered levels of three circulating proteins are associated with type 1 diabetes risk. The identified proteins are promising for development of early screening tools for type 1 diabetes, while they may be appealing drug targets. Future studies are needed to validate these candidate proteins as predictive biomarkers in longitudinal case-control type 1 diabetes cohorts. However, such studies are often strongly influenced by confounding and reverse causation.

**Funding.** D.M. received a Pediatric Endocrine Society Clinical Scholar Award for this research.

The funding body had no involvement in the study design, data collection, analysis and interpretation of results, or writing of this manuscript.

**Duality of Interest.** J.B.R. has served as an advisor to GlaxoSmithKline and Deerfield Capital; his institution has received investigator-initiated grant funding from Eli Lilly and Company, GlaxoSmithKline, and Biogen for projects unrelated to this research; and he is the founder of 5 Prime Sciences. No other potential conflicts of interest relevant to this article were reported.

**Author Contributions.** N.Y., M.Y., and D.M. conducted the analysis and interpretation of data and produced the first draft of the manuscript. V.F., C.P., J.B.R., and D.M. designed the study. Y.W. and M.P. undertook the ELISA validation studies. All authors reviewed and approved the final version. D.M. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## References

- Vehik K, Dabelea D. The changing epidemiology of type 1 diabetes: why is it going through the roof? *Diabetes Metab Res Rev* 2011;27:3–13
- Sharp SA, Weedon MN, Hagopian WA, Oram RA. Clinical and research uses of genetic risk scores in type 1 diabetes. *Curr Opin Genet Dev* 2018;50:96–102
- Long AE, Gillespie KM, Rokni S, Bingley PJ, Williams AJ. Rising incidence of type 1 diabetes is associated with altered immunophenotype at diagnosis. *Diabetes* 2012;61:683–686
- Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* 2020;52:1122–1131
- Chaimowitz NS, Ebenezer SJ, Hanson IC, Anderson M, Forbes LR. STAT1 gain of function, type 1 diabetes, and reversal with JAK inhibition. *N Engl J Med* 2020;383:1494–1496
- Moulder R, Bhosale SD, Erkkilä T, et al. Serum proteomes distinguish children developing type 1 diabetes in a cohort with HLA-conferred susceptibility. *Diabetes* 2015;64:2265–2278
- McGuire JN, Eising S, Wägner AM, Pociot F. Screening newborns for candidate biomarkers of type 1 diabetes. *Arch Physiol Biochem* 2010;116:227–232
- Zhi W, Purohit S, Carey C, Wang M, She JX. Proteomic technologies for the discovery of type 1 diabetes biomarkers. *J Diabetes Sci Technol* 2010;4:993–1002
- do Nascimento de Oliveira V, Lima-Neto ABM, van Tilburg MF, et al. Proteomic analysis to identify candidate biomarkers associated with type 1 diabetes. *Diabetes Metab Syndr Obes* 2018;11:289–301
- Bonifacio E. Predicting type 1 diabetes using biomarkers. *Diabetes Care* 2015;38:989–996
- Smith GD, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22
- Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37:658–665
- Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10:e1004383
- 1000 Genomes Project Consortium; Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65
- Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* 2018;361:769–773
- Folkersen L, Gustafsson S, Wang Q, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab* 2020;2:1135–1148
- Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017;8:14357
- Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature* 2018;558:73–79
- Yao C, Chen G, Song C, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat Commun* 2018;9:3268
- Forgetta V, Manousaki D, Istomine R, et al.; DCCT/EDIC Research Group. Rare genetic variants of large effect influence risk of type 1 diabetes. *Diabetes* 2020;69:784–795
- Yavorska OO, Burgess S. Mendelian-Randomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;46:1734–1739
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–909
- Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* 2019;35:4851–4853
- Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank* 2015;13:307–308
- Brion MJ, Shakhbuzov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013;42:1497–1501
- Zamani F, Almasi S, Kazemi T, Jahanban Esfahlan R, Aliparasti MR. New approaches to the immunotherapy of type 1 diabetes mellitus using interleukin-27. *Adv Pharm Bull* 2015;5(Suppl. 1):599–603
- Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O, Akerblom HK. Environmental triggers and determinants of type 1 diabetes. *Diabetes* 2005;54(Suppl. 2):S125–S136
- Yi L, Swensen AC, Qian WJ. Serum biomarkers for diagnosis and prediction of type 1 diabetes. *Transl Res* 2018;201:13–25
- Frohnert BJ, Webb-Robertson BJ, Bramer LM, et al. Predictive modeling of type 1 diabetes stages using disparate data sources. *Diabetes* 2020;69:238–248
- Sinha S, Renavikar PS, Crawford MP, et al. Altered expression of SIRPγ on the T-cells of relapsing remitting multiple sclerosis and type 1 diabetes patients could potentiate effector responses from T-cells. *PLoS One* 2020;15:e0238070
- Sinha S, Borchering N, Renavikar PS, et al. An autoimmune disease risk SNP, rs2281808, in SIRPγ is associated with reduced expression of SIRPγ and heightened effector state in human CD8 T-cells. *Sci Rep* 2018;8:15440
- Brooke G, Holbrook JD, Brown MH, Barclay AN. Human lymphocytes interact directly with CD47 through a novel member of the signal regulatory protein (SIRP) family. *J Immunol* 2004;173:2562–2570
- Barrett JC, Clayton DG, Concannon P, et al.; Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009;41:703–707
- Meka RR, Venkatesha SH, Dudics S, Acharya B, Moudgil KD. IL-27-induced modulation of autoimmunity and its therapeutic potential. *Autoimmun Rev* 2015;14:1131–1141
- Li MS, Liu Z, Liu JQ, Zhu X, Liu Z, Bai XF. The Yin and Yang aspects of IL-27 in induction of cancer-specific T-cell responses and immunotherapy. *Immunotherapy* 2015;7:191–200
- Nam H, Ferguson BS, Stephens JM, Morrison RF. Modulation of IL-27 in adipocytes during inflammatory stress. *Obesity (Silver Spring)* 2016;24:157–166
- Katsogiannis P, Kamble PG, Pereira MJ, et al. Changes in circulating cytokines and adipokines after RYGB in patients with and without type 2 diabetes. *Obesity (Silver Spring)* 2021;29:535–542
- Ciecko AE, Foda B, Barr JY, et al. Interleukin-27 is essential for type 1 diabetes development and Sjögren syndrome-like inflammation. *Cell Rep* 2019;29:3073–3086.e5
- Fujimoto H, Hirase T, Miyazaki Y, et al. IL-27 inhibits hyperglycemia and pancreatic islet inflammation induced by streptozotocin in mice. *Am J Pathol* 2011;179:2327–2336
- Kasela S, Kisand K, Tserel L, et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet* 2017;13:e1006643
- Łukawska-Tatarczuk M, Franek E, Czupryniak L, et al. Sirtuin 1, visfatin and IL-27 serum levels of type 1 diabetic females in relation to cardiovascular parameters and autoimmune thyroid disease. *Biomolecules* 2021;11:1110
- ‘t Hart LM, Fritsche A, Nijpels G, et al. The *CTRB1/2* locus affects diabetes susceptibility and treatment via the incretin pathway. *Diabetes* 2013;62:3275–3281
- Szabó A, Sahin-Tóth M. Determinants of chymotrypsin C cleavage specificity in the calcium-binding loop of human cationic trypsinogen. *FEBS J* 2012;279:4283–4292
- Rosendahl J, Kirsten H, Hegyi E, et al.; PanEuropean Working group on ACP. Genome-wide association study identifies inversion in the *CTRB1-CTRB2* locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut* 2018;67:1855–1863
- Nyaga DM, Vickers MH, Jefferies C, Fadason T, O’Sullivan JM. Untangling the genetic link between type 1 and type 2 diabetes using functional genomics. *Sci Rep* 2021;11:13871
- Inshaw JRJ, Sidore C, Cucca F, et al. Analysis of overlapping genetic association in type 1 and type 2 diabetes. *Diabetologia* 2021;64:1342–1347
- Stolerman ES, Florez JC. Genomics of type 2 diabetes mellitus: implications for the clinician. *Nat Rev Endocrinol* 2009;5:429–436
- Montgomery SB, Dermitzakis ET. From expression QTLs to personalized transcriptomics. *Nat Rev Genet* 2011;12:277–282