

1 **Integrative proteogenomic analyses provide novel interpretations of type 1 diabetes risk loci**
2 **through circulating proteins**

3 Tianyuan Lu^{1,*}, Despoina Manousaki^{2,3,4}, Lei Sun^{1,5}, and Andrew D. Paterson^{5,6,7,*}

4
5 ¹Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto, Toronto,
6 ON, Canada

7 ²Department of Pediatrics, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

8 ³Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de
9 Montréal, Montreal, QC, Canada

10 ⁴Research Center of the Sainte-Justine University Hospital, Université de Montréal, Montreal,
11 QC, Canada

12 ⁵Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON,
13 Canada

14 ⁶Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto,
15 ON, Canada

16 ⁷Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada

17
18 *Correspondence to:

19 Tianyuan Lu (tianyuan.lu@mail.mcgill.ca) and Andrew D. Paterson
20 (andrew.paterson@sickkids.ca)

21 Address: Program in Genetics & Genomic Biology, The Hospital for Sick Children Research
22 Institute, PGCRL Rm 12.9835, 686 Bay Street, Toronto, ON M5G 0A4, Canada

23

24

25 **Abstract (150 words)**

26 Type 1 diabetes (T1D) requires new preventive measures and interventions. Circulating proteins
27 are promising biomarkers and drug targets. Leveraging genome-wide association studies
28 (GWASs) of T1D (18,942 cases and 501,638 controls) and circulating protein abundances
29 (10,708 individuals), the associations between 1,565 circulating proteins and T1D risk were
30 assessed through Mendelian randomization, followed by multiple sensitivity and colocalization
31 analyses, examinations of horizontal pleiotropy, and replications. Genetically increased
32 circulating abundances of *CTSH*, *IL27RA*, *SIRPG*, and *PGM1* were associated with an increased
33 risk of T1D, consistently replicated in other cohorts. Bulk tissue and single-cell gene expression
34 profiles revealed strong enrichment of *CTSH*, *IL27RA*, and *SIRPG* in immune system-related
35 tissues, and *PGM1* in muscle and liver tissues. Among immune cells, *CTSH* was enriched in B
36 cells and myeloid cells, while *SIRPG* was enriched in T cells and natural killer cells. These
37 proteins warrant exploration as T1D biomarkers or drug targets in relevant tissues.

38

39

40 Introduction

41 Type 1 diabetes is an autoimmune disease characterized by the destruction of pancreatic β
42 cells¹⁻⁶, which are responsible for producing insulin. Although traditionally considered a disease
43 of children and adolescents, type 1 diabetes can be diagnosed at any age and can affect a
44 significant proportion of the global population³⁻⁵. Despite continuing efforts to develop risk
45 predictors⁷⁻⁹, few effective preventive measures for type 1 diabetes have been implemented in
46 public health practice. Following diagnosis, preserving residual β cell function and delaying type
47 1 diabetes-associated autoimmunity present challenges due to the multifactorial and
48 heterogeneous nature of the disease^{3-5, 10}. Consequently, there is an urgent need for new
49 biomarkers and drug targets for type 1 diabetes.

50
51 Circulating molecules participate in various biological processes and play essential roles
52 encompassing immune responses, signaling cascades, and regulatory mechanisms¹¹⁻¹⁴. These
53 molecules may be promising biomarkers or drug targets because their abundances are
54 measurable and possibly modifiable. Autoantibodies to insulin, glutamic acid decarboxylase,
55 islet antigen-2, zinc transporter 8, and other circulating proteins have emerged as markers for
56 characterizing type 1 diabetes^{10, 15, 16}. Yet, establishing the causal roles of novel proteins is
57 difficult. The feasibility of conducting randomized controlled trials for these proteins remains
58 limited. Meanwhile, observational studies can encounter several pitfalls, including uncontrolled
59 confounding factors as well as reverse causation.

60
61 Mendelian randomization (MR) is an instrumental variable framework that can effectively
62 mitigate biases arising from confounding and reverse causation^{17, 18}. MR employs genetic
63 variants as instruments for a risk factor (i.e. a circulating protein), and evaluates the potential
64 causal effect of the risk factor on a disease outcome. MR relies on three core instrumental
65 variable assumptions^{17, 18}. First, a genetic instrument should strongly predict the risk factor,
66 known as the relevance assumption. Second, the genetic instrument should not be associated
67 with confounders of the risk factor-disease outcome relationship, known as the independence
68 assumption. Third, the genetic instrument should not act on the disease outcome through
69 alternative pathways other than the instrumented risk factor, known as the no horizontal
70 pleiotropy assumption. Recent large-scale proteo-genomic studies have identified ideal genetic
71 instruments for circulating protein abundances, which are lead variants in cis-protein
72 quantitative trait loci (cis-pQTLs)¹¹⁻¹³. These genetic variants are unlikely to be associated with
73 confounders that influence the risk factor-disease outcome association, because of the
74 randomization at conception. Moreover, their proximity to protein-coding genes suggests a
75 direct influence on protein abundances, thereby reducing the risk of horizontal pleiotropy.
76 Previous studies have implemented MR to investigate potential causal effects of circulating
77 protein abundances on the risks of several complex diseases¹⁹⁻²².

78
79 While cis-pQTL-facilitated MR can pinpoint target proteins, it is important to note that
80 circulating proteins originate from various sources¹¹⁻¹³. These include but are not limited to
81 endocrine cell secretion, cellular turnover and apoptosis, immune and inflammatory response,
82 and diet and nutrition. To further understand the underpinning disease mechanisms and open

83 new avenues for diagnostic and therapeutic advancements, it is crucial to identify candidate
84 tissues and cell types where the target proteins are primarily produced.

85
86 In this study, we conducted integrative proteogenomic analyses to systematically identify
87 potential biomarkers and drug targets for type 1 diabetes. We first capitalized on genetic
88 associations from large-scale genome-wide association studies (GWASs) to conduct MR, in
89 order to assess the associations between circulating protein abundances and type 1 diabetes
90 risk. We then prioritized target proteins through multiple sensitivity and colocalization analyses,
91 examinations of horizontal pleiotropy, and replications. Furthermore, we identified candidate
92 tissues and cell types through enrichment analyses, utilizing both bulk tissue and single-cell
93 gene expression profiles. Our findings underscore circulating proteins that exhibit a potential
94 causal effect on the risk of type 1 diabetes.

95 96 **Results**

97 ***Target protein prioritization through Mendelian randomization***

98 An overview of this study is presented in **Figure 1**. After identification of cis-genetic instruments
99 and data harmonization, associations between circulating abundances of 1,565 proteins and
100 type 1 diabetes risk were assessed using MR. MR analyses of 135 (8.6%) proteins utilized LD
101 proxies of cis-genetic instruments (**Methods**). Details of genetic instruments are provided in
102 **Supplementary Table S1**.

103
104 A total of 12 associations between circulating protein abundances and type 1 diabetes risk
105 reached the Bonferroni-corrected significance threshold (p -value $< 3.2 \times 10^{-5}$; **Figure 2** and
106 **Supplementary Figure S1**), excluding proteins whose coding genes map to the MHC region.
107 These significant associations had a minimal F-statistic of 46.6, indicating a low risk of weak
108 instrument bias. Full summary statistics of MR analyses are provided in **Supplementary Table**
109 **S2**.

110
111 Of these 12 proteins, circulating abundances of CTSH, ANXA2, and CCL25 were instrumented
112 using three cis-genetic instruments. Results of sensitivity analyses using weighted median,
113 penalized weighted median, weighted mode, and MR-Egger methods were highly consistent
114 with those obtained using the inverse variance weighted method (**Supplementary Table S3**).
115 MR-Egger intercepts largely overlapped with the null, suggesting a low risk of directional
116 horizontal pleiotropy (**Supplementary Table S3**).

117 118 ***Colocalization evidence, horizontal pleiotropy assessment, and replication***

119 Colocalization analyses and horizontal pleiotropy assessment were performed to verify MR
120 assumptions for these 12 proteins (**Methods**). Strong (PP.H4 $> 80\%$) or suggestive (PP.H4 $> 50\%$)
121 evidence of colocalization between circulating protein abundance and type 1 diabetes risk was
122 observed for CTSH, RHOC, IL27RA, ANXA2, SIRPG, CCL25, and PGM1. Conversely, colocalization
123 evidence was limited for EB13-IL27 complex, IL15RA, ERBB3, WARS, and ALDH2 (**Figures 2** and **3**,
124 **Supplementary Figure S2**, and **Supplementary Table S4**).

125

126 Among these target proteins supported by colocalization evidence, the genetic instruments of
127 both RHOC and ANXA2 were predicted to have stronger functional connections to neighboring
128 genes (i.e. *ICE2* and *ST7L*, respectively) other than their respective coding genes, as indicated by
129 V2G scores (**Methods** and **Supplementary Table S5**). In addition, the genetic instruments of
130 RHOC, ANXA2, SIRPG, and CCL25 have been associated with the expression or splicing of other
131 neighboring genes, which introduces an elevated risk of horizontal pleiotropy. In contrast, the
132 genetic instruments of *CTSH*, *IL27RA*, and *PGM1* demonstrated the strongest functional
133 connection to their respective coding genes, were not associated with the expression, splicing,
134 or translation of other neighboring genes, and had not been associated with other known risk
135 factors of type 1 diabetes in the Open Target database, thereby mitigating the risk of horizontal
136 pleiotropy (**Figures 2** and **Supplementary Tables S5** and **S6**).

137
138 Seven significant associations supported by colocalization evidence were re-evaluated when cis-
139 genetic instruments could be identified in the deCODE study or the UKB-PPP study, or when the
140 associations could be assessed based on the type 1 diabetes GWAS meta-analysis by Robertson
141 et al (**Methods**, **Figure 2**, and **Supplementary Table S7**). Six of the seven associations were
142 replicated using these additional resources with a consistent effect direction and a similar
143 magnitude of effect as obtained in the primary analyses (**Supplementary Table S8**). However,
144 based on the cis-genetic instrument identified in the UKB-PPP study, a one standard deviation
145 increase in genetically predicted circulating abundance of *CCL25* was not associated with the
146 risk of type 1 diabetes (odds ratio, OR = 1.01; 95% CI: 0.96-1.05; p-value = 0.82; **Supplementary**
147 **Table S8**).

148
149 Following these assessments, we prioritized *CTSH*, *IL27RA*, *SIRPG*, and *PGM1* as target proteins
150 (**Figures 2** and **3**), while cautioning a moderate risk of horizontal pleiotropy affecting the genetic
151 instrument of circulating *SIRPG* abundance. Specifically, genetically predicted circulating
152 abundances of *CTSH*, *IL27RA*, *SIRPG*, and *PGM1* were associated with increased odds of
153 developing type 1 diabetes, with ORs of 1.17 (*CTSH*; 95%: 1.10-1.24; p-value = 9.3×10^{-7} ; PP.H4 =
154 99.6%), 1.13 (*IL27RA*; 95%: 1.07-1.19; p-value = 2.3×10^{-5} ; PP.H4 = 92.7%), 1.37 (*SIRPG*; 95%:
155 1.26-1.49; p-value = 4.3×10^{-13} ; PP.H4 = 86.6%), and 1.66 (*PGM1*; 95%: 1.40-1.96; p-value =
156 3.9×10^{-9} ; PP.H4 = 71.3%) per one standard deviation increase, respectively.

157 158 ***Tissue and immune cell type enrichment of gene expression***

159 For each of the target protein-coding genes, enrichment of gene expression in 54 tissue sites
160 profiled by the GTEx Consortium was assessed to identify potential candidate tissues (**Methods**
161 and **Supplementary Table S9**). As a result, the expression of *CTSH*, *IL27RA*, and *SIRPG* was
162 enriched in the whole blood with a tissue-specific enrichment z-score > 10 (**Figure 4**).
163 Furthermore, *CTSH* expression was enriched in Epstein-Barr virus-transformed lymphocytes and
164 *SIRPG* expression was enriched in the spleen, while *IL27RA* expression was enriched in both of
165 these tissues (**Figure 4**). In contrast, the expression of *PGM1* exhibited enrichment in skeletal
166 muscle, heart (left ventricle), and liver (**Figure 4**).

167
168 Whole blood-specific cis-eQTL of *CTSH* and liver-specific cis-eQTL of *PGM1* demonstrated strong
169 evidence of colocalization with the genetic associations with type 1 diabetes risk, while tissue-

170 specific cis-eQTLs of *IL27RA* and *SIRPG* did not show evidence of colocalization (**Supplementary**
171 **Figure S3** and **Supplementary Table S10**). Importantly, the cis-genetic instruments of circulating
172 abundances of *IL27RA* and *SIRPG* were not strongly associated with their mRNA abundances in
173 these candidate tissues (**Supplementary Figure S4**). Meanwhile, there was strong evidence of
174 colocalization between genetic associations with multiple isoforms of *CTSH* and *SIRPG* in whole
175 blood and the genetic associations with the risk of type 1 diabetes (**Supplementary Table S11**).
176 These cis-sQTLs also overlapped with cis-pQTLs of *CTSH* and *SIRPG* (**Supplementary Figure S5**),
177 respectively.

178
179 Given the enrichment of *CTSH*, *IL27RA*, and *SIRPG* expression in immune system-related tissues,
180 we further examined cell type-specific gene expression based on single-cell transcriptomic
181 profiling of 329,762 immune cells, consisting of 45 curated cell types (**Methods** and **Figure 5A**).
182 Among these immune cells, it was evident that *CTSH* expression was enriched in B cells,
183 excluding pro-B cells and pre-B cells, as well as in myeloid cells (**Figure 5B**, **Supplementary**
184 **Figures 6A**, **Supplementary Figures 7-9**, and **Supplementary Table S12**). On the other hand,
185 *CTSH* expression was depleted in T cells, albeit with modest expression observed in effector
186 memory CD4⁺ T cells (Teffector/EM_CD4) and tissue-resident memory T-helper 1 and T-helper
187 17 cells (Trm_Th1/Th17). Meanwhile, the expression level of *IL27RA* was moderate and
188 relatively consistent across most cell types (**Figure 5C**, **Supplementary Figures 6B**,
189 **Supplementary Figures 7-9**, and **Supplementary Table S12**). In contrast, *SIRPG* expression was
190 enriched in T cells and natural killer cells, and depleted in B cells and myeloid cells (**Figure 5D**,
191 **Supplementary Figures 6C**, **Supplementary Figures 7-9**, and **Supplementary Table S12**).

192
193 ***Mendelian disorders and incident disease outcomes associated with target proteins***
194 Among the four target proteins, *PGM1* was implicated in congenital disorder of glycosylation
195 type 1t (CDG1T), an autosomal recessive disorder caused by *PGM1* deficiency due to
196 pathogenic homozygous or compound heterozygous mutations affecting the *PGM1* gene
197 (OMIM#614921, **Supplementary Table S13**). The other target proteins did not have known
198 implications in Mendelian disorders.

199
200 In the UK Biobank, observational associations between circulating protein abundances and
201 incident disease outcomes were only available for *CTSH*. Over 16 years of follow-up, a one
202 standard deviation increase in circulating *CTSH* abundance was associated with a 1.14-fold
203 increased hazard of mortality (95% CI: 1.10-1.17; p-value = 1.7×10^{-17}), and interestingly, a 1.16-
204 fold increased hazard of type 2 diabetes based on physician-made diagnosis (95% CI: 1.11-1.21;
205 p-value = 4.0×10^{-12} ; **Supplementary Figure S10** and **Supplementary Table S14**). In addition, a
206 one standard deviation increase in circulating *CTSH* abundance was associated with elevated
207 risks of systemic lupus erythematosus (hazard ratio, HR = 1.37; 95% CI: 1.17-1.60; p-value =
208 1.1×10^{-4}), rheumatoid arthritis (HR = 1.18; 95% CI: 1.09-1.27; p-value = 2.7×10^{-5}), and chronic
209 obstructive pulmonary disease (HR = 1.09; 95% CI: 1.05-1.14; p-value = 6.3×10^{-5} ;
210 **Supplementary Figure S10** and **Supplementary Table S14**).

211
212 **Discussion**

213 Type 1 diabetes impacts millions of individuals worldwide, causing acute and chronic
214 complications that profoundly deteriorate the quality of life and increase mortality rates¹⁻⁶.
215 Managing type 1 diabetes typically requires insulin injections for glycemic control, resulting in a
216 significant socioeconomic burden^{23, 24}. There is an urgent need for innovative strategies to
217 prevent, intervene early, and manage the disease. In this study, we conducted MR-guided
218 target discovery to systematically examine circulating proteins that may play a crucial role in
219 the etiology of type 1 diabetes. We also identified candidate tissues and cell types enriched of
220 target protein-coding gene expression. Our study presents a curated selection of candidate
221 proteins with the potential as biomarkers or drug targets.

222
223 Our integrative proteogenomic analyses prioritized four target proteins, CTSH, IL27RA, SIRPG,
224 and PGM1. Increased circulating abundances of these proteins were predicted to increase the
225 risk of type 1 diabetes. Specifically, CTSH (cathepsin H) is a lysosomal cysteine protease involved
226 in the degradation of lysosomal proteins^{25, 26}. CTSH in pancreatic islets may affect β cell
227 survival and insulin secretion by modulating apoptotic signaling pathways and transcription
228 factors^{27, 28}. The genomic locus within the CTSH gene has previously been associated with the
229 risk of type 1 diabetes^{29, 30}. In this study, we observed that *CTSH* expression was enriched in B
230 cells and myeloid cells, implying a potential role of CTSH in antigen presentation and antibody-
231 mediated immunity. Furthermore, although the genetic risk may be conferred by gene
232 expression, which may be mediated by DNA methylation³¹, colocalization between the genetic
233 associations with multiple isoforms of *CTSH* in the whole blood and the risk of type 1 diabetes
234 suggests that alternative splicing of *CTSH* may contribute to the disease pathogenesis. In the UK
235 Biobank, increased circulating CTSH abundance was linked to higher mortality and risks of
236 common autoimmune diseases. Importantly, while the observational association analyses did
237 not encompass type 1 diabetes as an outcome due to the limited number of cases, increased
238 circulating CTSH abundance was associated with an increased risk of type 2 diabetes, defined
239 based on physician-made diagnosis. This suggests the possibility of misdiagnosing type 1
240 diabetes as type 2 diabetes^{3, 32, 33} within the adult population of the UK Biobank. Increased
241 *CTSH* expression has also been associated with early-onset type 1 diabetes and rapid decline of
242 β cell function in other cohort studies^{27, 34, 35}. Taken together, our findings strongly encourage
243 functional follow-up studies to explicate the role of CTSH in type 1 diabetes and to evaluate its
244 potential as a biomarker or drug target.

245
246 IL27RA (alpha subunit of the interleukin 27 receptor) binds to IL27, a heterodimeric cytokine
247 composed of IL27p28 and EB13 subunits^{36, 37}. IL27 has both pro-inflammatory functions by
248 mediating T-helper 1 cell differentiation and increasing interferon γ production³⁶⁻³⁸, and anti-
249 inflammatory functions by inhibiting pro-inflammatory cytokines in T cells and promoting the
250 production of anti-inflammatory cytokines³⁹⁻⁴¹. However, due to the lack of colocalization
251 evidence and potential horizontal pleiotropic effects, we were unable to determine the effect
252 of IL27. On the other hand, the association between circulating abundance of IL27RA and the
253 risk of type 1 diabetes was substantiated by multiple lines of evidence. The results of our
254 enrichment analyses align with the involvement of IL27RA in cell-mediated and antibody-
255 mediated immunity by mediating IL27 signaling in various immune cells³⁶. While the functions
256 of IL27RA and IL27RA-mediated IL27 signaling in type 1 diabetes remain to be explored, we

257 posit that IL27 and IL27RA may regulate both innate and adaptive immune responses that
258 attack the pancreatic β cells.

259
260 SIRPG (signal-regulatory protein γ) is a receptor protein involved in the negative regulation of
261 receptor tyrosine kinase-coupled signaling processes⁴². It has been suggested that SIRPG
262 signaling may play an immunoregulatory role in maintaining peripheral immune tolerance and
263 preventing autoimmunity⁴³. In line with existing studies, our analyses demonstrated that *SIRPG*
264 expression was enriched in T cells and natural killer cells, where blocking of the SIRPG-CD47
265 interaction has been found to inhibit superantigen-induced T cell proliferation^{42, 44, 45}. These
266 findings imply the potential significance of investigating SIRPG as a T cell-specific target for type
267 1 diabetes, although it should be noted that the genetic instrument for circulating SIRPG
268 abundance was subject to a moderate risk of horizontal pleiotropy due to its associations with
269 the expression, splicing, or translation of neighboring genes encoding other signal-regulatory
270 proteins, SIRPB1, SIRPB2, and SIRPD.

271
272 PGM1 (phosphoglucomutase 1) is an enzyme that catalyzes the reversible conversion between
273 glucose 1-phosphate and glucose 6-phosphate⁴⁶, which are important intermediates in glucose
274 metabolism. The Mendelian disorders of PGM1 deficiency can result in congenital disorder of
275 glycosylation⁴⁷. Given the crucial functions of insulin in the uptake of glucose into cells and the
276 regulation of glycogen synthesis and breakdown⁴⁸, we hypothesize that PGM1 may play a role
277 in type 1 diabetes by affecting the balance between glycogen storage and glucose utilization,
278 particularly in muscle and liver tissues. Notably, previous GWASs have suggested that the
279 PGM1-increasing allele of the genetic instrument, which increases the risk of type 1 diabetes,
280 may have a marginal protective effect against type 2 diabetes^{49, 50}, although this association
281 was not genome-wide significant. Elucidating the precise involvement of PGM1 in diabetes
282 mellitus necessitates further efforts.

283
284 Our study has several strengths. First, we harnessed large-scale GWASs to increase the power
285 of MR and colocalization analyses. Specifically, we obtained genetic instruments for circulating
286 protein abundances from a large-scale proteo-genomic study with the highest coverage of the
287 circulating proteome to date, and conducted target discovery utilizing the largest meta-analysis
288 of type 1 diabetes GWASs. Notably, the association between circulating SIRPG abundance and
289 the risk of type 1 diabetes was identified in a previous MR study but was not supported by
290 colocalization evidence¹⁰, which is likely attributable to the smaller sample sizes of GWASs⁵¹.
291 This advantage is also evident when compared to GTEx cis-eQTL-based analyses, which had
292 much smaller sample sizes and failed to demonstrate colocalization evidence in most of the
293 candidate tissues. Second, after cis-pQTL-facilitated MR analyses, we subjected the genetic
294 instruments to rigorous scrutiny, ensuring that the no horizontal pleiotropy assumption was not
295 violated for prioritized proteins. Third, we bolstered the validity of our findings by replicating
296 our results using additional resources. Importantly, analyses using the genetic instruments
297 identified in the UKB-PPP study failed to replicate the association between circulating CCL25
298 abundance and type 1 diabetes risk, highlighting the potential influence of protein detection
299 platform and study population. These analyses collectively mitigated the risk of false positive
300 results. Although deprioritized target proteins, such as IL15RA^{52, 53} and ERBB3^{54, 55}, have been

301 previously linked to the risk of type 1 diabetes in different contexts, substantiating whether
302 these associations truly denote causal effects requires future efforts. Fourth, based upon MR-
303 prioritized target proteins, we further identified candidate tissues and cell types where the
304 target protein-coding gene expression was enriched. This characterization yielded valuable
305 insights into the biological relevance, disease mechanisms, as well as the therapeutic potential
306 of these target proteins.

307
308 Our study has important limitations. First of all, our findings have not been experimentally
309 validated, which should be the focus of follow-up studies. Second, our analyses were restricted
310 to populations predominantly of European ancestry. Given the substantial variability in the
311 prevalence and strong heterogeneity of type 1 diabetes across different populations in different
312 countries³⁻⁵, it is important to exercise caution when generalizing our findings to populations of
313 non-European ancestries. Third, it should be noted that all cis-pQTLs used in this study were
314 identified in middle-aged and older adults, whereas the type 1 diabetes GWASs included
315 patients who were more likely to develop the disease at a younger age. However, we posit that
316 the cis-genetic regulation of circulating protein abundances is likely consistent across age
317 distributions. Nevertheless, we strongly advocate for similar analyses to be conducted across
318 populations of diverse ancestries and demographic characteristics. Fourth, although existing
319 protein detection platforms have enabled the measurement of circulating abundances for
320 nearly 5,000 proteins, the possibility remains that potential target proteins lack valid genetic
321 instruments. Genetics-guided target discovery based on proteo-genomic studies featuring
322 increased sample sizes and enhanced coverage of the circulating proteome should be pursued
323 in the future. Last but not least, due to the strong variability and highly intricate LD structure of
324 the MHC region, we did not prioritize MHC gene-coded proteins, despite significant associations
325 identified through MR. Considering the well-established role of the MHC region in the
326 pathogenesis and progression of type 1 diabetes, future efforts should be dedicated to
327 elucidating the functional impacts of these proteins.

328
329 In conclusion, through integrative proteogenomic analyses, we identified significant
330 associations between circulating protein abundances and the risk of type 1 diabetes, which
331 further suggested possible causal effects of *CTSH*, *IL27RA*, *SIRPG*, and *PGM1*. The roles of *CTSH*,
332 *IL27RA*, and *SIRPG* in the immune system are underscored, with enrichment of *CTSH* expression
333 in B cells and myeloid cells, and *SIRPG* expression in T cells and natural killer cells. In contrast,
334 *PGM1* may influence the risk of type 1 diabetes through its impact on glucose metabolism,
335 particularly in muscle and liver tissues. Exploration of these target proteins as biomarkers or
336 viable candidates for drug targeting strategies while considering the candidate tissues and cell
337 types should be warranted in the context of type 1 diabetes.

338
339

340 **Online Methods**

341 ***Genome-wide association study of circulating protein abundances***

342 Genetic associations with circulating protein abundances were assessed in the Fenland study
343 based on 10,708 unrelated European ancestry individuals¹². Details of this study have been
344 described previously^{12, 56}. Abundances of 4,775 proteins and protein complexes from plasma
345 samples were measured using the SomaLogic SomaScan v4 assay, which includes 4,979 distinct
346 SOMAmer reagents. GWAS was conducted for each SOMAmer protein target, referred to as
347 “protein” hereinafter. The circulating abundances underwent rank-based inverse normal
348 transformation after regressing out the effects of age, sex, test site, and the first ten genetic
349 principal components¹². Conditional and joint (COJO) analyses⁵⁷ were performed to identify
350 conditionally independent lead variants with a p-value $< 1.0 \times 10^{-11}$, which represented the
351 Bonferroni-corrected genome-wide significance threshold. Cis-pQTL variants were defined for
352 each protein as conditionally independent lead variants located within 500 kb away from the
353 gene body of the protein-coding gene.

354

355 ***Genome-wide association study of type 1 diabetes***

356 Genetic associations with type 1 diabetes risk were assessed in a meta-analysis of GWASs by
357 Chiou et al. including up to 18,942 patients and 501,638 controls predominantly of European
358 ancestry from nine cohorts⁵⁸. Details of the participating cohorts and the meta-analysis have
359 been described previously⁵⁸. There was no known overlap between participants of the Fenland
360 study and participants of this meta-analysis.

361

362 ***Mendelian randomization and sensitivity analyses***

363 Two-sample MR was performed based on GWAS summary statistics to test associations
364 between the genetically predicted circulating abundance of each protein and type 1 diabetes
365 risk. Cis-pQTL variants identified in the Fenland study were used as genetic instruments. Trans-
366 genetic variants distal to the protein-coding genes likely act on other genes, thus to mitigate
367 the risk of horizontal pleiotropy, they were not used. If a cis-pQTL variant was unavailable in the
368 type 1 diabetes GWAS summary statistics, we attempted to identify a proxy as the genetic
369 instrument using the LDlink R package⁵⁹. The proxy should be in high linkage disequilibrium (LD;
370 $r^2 > 0.8$) with the cis-pQTL variant based on the LD reference panel consisting of non-Finnish
371 European ancestry populations in the 1000 Genomes Project phase 3⁶⁰. GWAS summary
372 statistics for genetic instruments were harmonized with forward strand alleles inferred using
373 allele frequency information. Palindromic variants with high minor allele frequency (MAF > 0.42)
374 were discarded to avoid allele mismatches.

375

376 Wald ratio estimates were derived for proteins with only one cis-genetic instrument, while
377 inverse variance weighted estimates were derived for proteins with two or more cis-genetic
378 instruments^{61, 62}. Associations with a p-value $< 3.2 \times 10^{-5}$ were considered significant,
379 representing the Bonferroni-corrected significance threshold to account for 1,565 tests. This
380 significance threshold may be overly conservative due to possible correlation and functional
381 relevance between proteins, but should effectively control the false positive rate.

382

383 For significant associations where the protein abundances were instrumented using three or
384 more cis-genetic instruments, we conducted sensitivity analyses using the weighted median,
385 penalized weighted median, weighted mode, and MR-Egger methods⁶¹⁻⁶⁴. An association was
386 considered robust to invalid instruments if these different methods yielded estimates with a
387 consistent effect direction and magnitude. A significant MR-Egger intercept (p-value < 0.05)
388 would indicate existence of directional horizontal pleiotropy⁶⁴. Furthermore, we calculated the
389 F-statistic for each test, where an F-statistic < 10 would indicate a risk of weak instrument bias
390⁶⁵. MR analyses were conducted using the TwoSampleMR R package version 0.5.6⁶⁶.

391

392 ***Colocalization analyses***

393 While most genetic instruments are typically not associated with confounders of the exposure-
394 outcome relationship, MR may be confounded by LD, where two genetic variants separately
395 influence the exposure and the outcome through different mechanisms but are correlated with
396 each other through LD⁶⁷. Colocalization analyses have been widely used to assess whether the
397 exposure and the outcome share the same causal genetic variants, in order to guard against
398 such confounding effects⁶⁷.

399

400 For significant associations, we performed colocalization analyses using PWCoCo¹⁶, leveraging
401 GWAS summary statistics of all variants located within 500 kb away from the cis-genetic
402 instruments, and an LD reference panel constructed using 5,000 randomly selected unrelated
403 European ancestry individuals from the UK Biobank⁶⁸. PWCoCo builds upon the classical
404 algorithm coloc⁶⁹, but allows for multiple causal variants in the same genomic region through
405 an implementation of COJO analyses for the exposure and the outcome separately, and
406 pairwise colocalization analyses of conditionally independent signals^{16, 57}. We used default
407 priors of PWCoCo, i.e. p_1 (prior probability of the exposure having a causal variant) = p_2 (prior
408 probability of the outcome having a causal variant) = 1.0×10^{-4} , and p_{12} (prior probability of the
409 exposure and the outcome sharing the same causal variant) = 1.0×10^{-5} . A colocalization
410 probability (PP.H4) > 80% was considered strong evidence of colocalization, while a PP.H4 > 50%
411 was considered suggestive evidence of colocalization. We excluded proteins whose coding
412 genes map to the major histocompatibility complex (MHC) region due to the strong variability
413 and highly intricate LD structure.

414

415 ***Annotation of genetic instruments and phenome-wide association study***

416 To further evaluate potential horizontal pleiotropic effects, we obtained variant-to-gene (V2G)
417 annotations⁷⁰ and phenome-wide associations from publicly available GWASs in Open Targets
418^{71, 72} (retrieved July 1, 2023) for each genetic instrument. Specifically, the V2G scores were
419 derived from a machine learning model trained to distinguish true causal genes from
420 neighboring genes in the same genomic region⁷⁰. Therefore, the V2G scores can be a
421 quantitative measure of the functional connection between a variant and a gene.

422

423 We considered a genetic instrument to be subject to a high risk of horizontal pleiotropy if the
424 gene with the highest V2G score paired with this variant was not the target protein-coding gene.
425 Furthermore, a genetic variant was considered to be subject to a moderate risk of horizontal
426 pleiotropy if it had been associated with the expression, splicing, or translation of one or more

427 genes in proximity other than the target protein-coding gene. Additionally, variants
428 demonstrating associations with other established risk factors of type 1 diabetes were also
429 considered to have a moderate risk of horizontal pleiotropy.

430

431 ***Replication of findings***

432 Significant associations that were supported by strong or suggestive colocalization evidence
433 were replicated in two ways. First, we repeated MR analyses for these proteins using cis-genetic
434 instruments identified in the deCODE study¹³ and the UK Biobank Pharma Proteomics Project
435 (UKB-PPP) study⁷³. The deCODE study measured circulating plasma abundances of 4,907
436 SOMAmer protein targets in 35,559 individuals from Iceland, employing the same SomaLogic
437 SomaScan v4 assay as in the Fenland study¹³. However, the genetic architecture underlying
438 circulating protein abundances and LD structures in the Icelandic population may differ from
439 those in other European ancestry populations due to extensive genetic drift⁷⁴. On the other
440 hand, the UKB-PPP study measured plasma circulating abundances of 2,923 protein analytes
441 using the Olink Explore 1536 platform⁷³. The discovery of pQTLs were conducted based on
442 35,571 European ancestry individuals. This cohort overlapped with participants in the meta-
443 analysis of type 1 diabetes GWASs by Chiou et al⁵⁸.

444

445 Second, we repeated MR analyses for these proteins leveraging a different meta-analysis of
446 type 1 diabetes GWASs by Robertson et al.⁷⁵, using cis-genetic instruments identified in the
447 Fenland study. This meta-analysis comprised up to 16,159 patients, 25,386 controls, and 6,143
448 trio families with an affected offspring and both parents, including 7,117 participants of non-
449 European (African, East Asian, or admixed) ancestries⁷⁵. Although participants in this meta-
450 analysis partially overlapped with those in the primary analysis by Chiou et al.⁵⁸, genotyping in
451 this study by Robertson et al. was conducted using the Illumina ImmunoChip, which provided
452 dense coverage in 188 immune-relevant genomic regions, but sparse coverage in other regions
453⁷⁵.

454

455 ***Quantification of tissue-specific gene expression***

456 We investigated the tissue specificity of prioritized proteins leveraging gene expression profiles
457 from the Genotype-Tissue Expression (GTEx) project version 8⁷⁶. Following previous studies⁷⁷,
458⁷⁸, we quantified the enrichment of gene expression in each of the 54 non-diseased tissue sites
459 across approximately 1,000 individuals. Specifically, we first retained genes that were detected
460 in at least 20% of the samples with at least 5 read counts. We performed per-tissue trimmed
461 mean of M-values (TMM) normalization using the edgeR R package⁷⁹. Then, we calculated the
462 median TMM value across individuals for each gene in each tissue. Subsequently, within each
463 tissue, we standardized the across-individual median TMM values, using the median and the
464 median absolute deviation across genes. Finally, for each gene, the tissue-specific enrichment z-
465 scores were calculated by standardizing the within-tissue standardized across-individual median
466 TMM values, using the median and the median absolute deviation across tissues. Tissues with
467 an enrichment z-score > 10 were considered to be enriched of expression of the corresponding
468 gene.

469

470 ***Genetic effects on tissue-specific gene expression***

471 Next, in each tissue that demonstrated enrichment of gene expression, we tested colocalization
472 of the genetic associations with gene expression and splicing patterns and the risk of type 1
473 diabetes for each target protein-coding gene. Genetic associations with mRNA abundances and
474 isoform abundances were obtained respectively from the cis-expression and cis-splicing
475 quantitative trait loci (eQTL and sQTL) analyses conducted by the GTEx Consortium⁷⁶.
476 Colocalization analyses were performed using PWCoCo with default priors as described above¹⁶.
477

478 ***Single-cell gene expression profiling***

479 Since tissues that play a significant role in the immune system were implicated, we further
480 investigated whether target protein-coding genes were enriched in specific immune cell types.
481 We obtained single-cell gene expression profiles from a cross-tissue analysis that included high-
482 quality 329,762 immune cells from 12 donors⁸⁰. Details of this study, including sample
483 collection, single-cell RNA sequencing and paired VDJ sequencing for T cell and B cell receptors,
484 and data processing, have been described previously⁸⁰. In this study, manual curation was
485 conducted after automated annotation, using existing cell type-specific gene expression
486 signatures to identify 45 cell types. These cell types were classified as: B cell compartment, T
487 cell compartment (predominantly T cells and natural killer cells), and myeloid compartment
488 (predominantly macrophages, monocytes, and dendritic cells)⁸⁰. We evaluated the normalized
489 gene expression level of each target protein-coding gene in each cell and compared the
490 distribution of gene expression levels between different cell types.
491

492 ***Clinically relevant variants affecting target protein-coding genes***

493 To assess whether the prioritized proteins may be associated with other human diseases, we
494 queried the ClinVar database (June 9th, 2023)⁸¹ to identify Mendelian disorders that are caused
495 by variants affecting the target protein-coding genes. Mendelian disorder-causing variants must
496 be pathogenic or likely pathogenic variants with at least one submitter providing assertion
497 criteria, which should not have conflicting interpretations.
498

499 ***Observational associations with incident disease outcomes in the UK Biobank***

500 Finally, we obtained observational association test statistics from a recent study to evaluate
501 whether measured circulating protein abundances could predict future disease outcomes in the
502 UK Biobank over 16 years of follow-up⁸². The associations between each protein and mortality
503 and 23 incident morbidities were estimated using Cox proportional hazards models, based on
504 49,234 individuals predominantly of European ancestry, adjusted for the fixed effects of age
505 and sex, or for age only in the case of sex-specific diseases⁸².
506

507 ***Acknowledgements***

508 T.L. has been supported by a Schmidt AI in Science Postdoctoral Fellowship. The funder has no
509 role in study design; collection, management, analysis and interpretation of data; or the
510 decision to submit for publication.

511

512 ***Conflicts of interest***

513 T.L. was employed by 5 Prime Sciences Inc. until Sept 2023. The research presented in this
514 paper was conducted independently, and 5 Prime Sciences Inc. was not involved in the design,
515 execution, analysis, or interpretation of the study. T.L. declares no ongoing conflicts of interest.
516 The other authors declare no conflicts of interest.

517

518 ***Data availability***

519 All results generated in this study are included in Supplementary Tables.

520

521 **References**

- 522 1. Gepts W. Pathologic anatomy of the pancreas in juvenile diabetes mellitus. *Diabetes*.
523 Oct 1965;14(10):619-33. doi:10.2337/diab.14.10.619
- 524 2. Eisenbarth GS. Type I diabetes mellitus. A chronic autoimmune disease. *N Engl J Med*.
525 May 22 1986;314(21):1360-8. doi:10.1056/NEJM198605223142106
- 526 3. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *The Lancet*.
527 2014;383(9911):69-82.
- 528 4. DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *The Lancet*.
529 2018;391(10138):2449-2462.
- 530 5. Katsarou A, Gudbjörnsdottir S, Rawshani A, et al. Type 1 diabetes mellitus. *Nature*
531 *reviews Disease primers*. 2017;3(1):1-17.
- 532 6. Daneman D. Type 1 diabetes. *The Lancet*. 2006;367(9513):847-858.
- 533 7. Oram RA, Patel K, Hill A, et al. A Type 1 Diabetes Genetic Risk Score Can Aid
534 Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. *Diabetes Care*. Mar
535 2016;39(3):337-44. doi:10.2337/dc15-1111
- 536 8. Redondo MJ, Geyer S, Steck AK, et al. A Type 1 Diabetes Genetic Risk Score Predicts
537 Progression of Islet Autoimmunity and Development of Type 1 Diabetes in Individuals at Risk.
538 *Diabetes Care*. Sep 2018;41(9):1887-1894. doi:10.2337/dc18-0087
- 539 9. Sharp SA, Rich SS, Wood AR, et al. Development and standardization of an improved
540 type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diabetes*
541 *care*. 2019;42(2):200-207.
- 542 10. Yazdanpanah N, Yazdanpanah M, Wang Y, et al. Clinically Relevant Circulating Protein
543 Biomarkers for Type 1 Diabetes: Evidence From a Two-Sample Mendelian Randomization Study.
544 *Diabetes Care*. Jan 1 2022;45(1):169-177. doi:10.2337/dc21-1049
- 545 11. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome.
546 *Nature*. Jun 2018;558(7708):73-79. doi:10.1038/s41586-018-0175-2
- 547 12. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic
548 convergence of human diseases. *Science*. Oct 14 2021:eabj1541. doi:10.1126/science.abj1541
- 549 13. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma
550 proteome with genetics and disease. *Nat Genet*. Dec 2021;53(12):1712-1721.
551 doi:10.1038/s41588-021-00978-w
- 552 14. Chen Y, Lu T, Pettersson-Kymmer U, et al. Genomic atlas of the plasma metabolome
553 prioritizes metabolites implicated in human diseases. *Nat Genet*. Jan 2023;55(1):44-53.
554 doi:10.1038/s41588-022-01270-1
- 555 15. Long AE, Gillespie KM, Rokni S, Bingley PJ, Williams AJ. Rising incidence of type 1
556 diabetes is associated with altered immunophenotype at diagnosis. *Diabetes*. 2012;61(3):683-
557 686.
- 558 16. Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping
559 the influence of the plasma proteome on complex diseases. *Nature genetics*. 2020;52(10):1122-
560 1131.
- 561 17. Skrivankova VW, Richmond RC, Woolf BA, et al. Strengthening the reporting of
562 observational studies in epidemiology using mendelian randomisation (STROBE-MR):
563 explanation and elaboration. *bmj*. 2021;375

- 564 18. Skrivankova VW, Richmond RC, Woolf BA, et al. Strengthening the Reporting of
565 Observational Studies in Epidemiology using Mendelian Randomization: the STROBE-MR
566 Statement. *JAMA*. 2021;326(16):1614-1621.
- 567 19. Yao C, Chen G, Song C, et al. Genome-wide mapping of plasma protein QTLs identifies
568 putatively causal genes and pathways for cardiovascular disease. *Nat Commun*. Aug 15
569 2018;9(1):3268. doi:10.1038/s41467-018-05512-x
- 570 20. Chong M, Sjaarda J, Pigeyre M, et al. Novel Drug Targets for Ischemic Stroke Identified
571 Through Mendelian Randomization Analysis of the Blood Proteome. *Circulation*. Sep 9
572 2019;140(10):819-830. doi:10.1161/CIRCULATIONAHA.119.040180
- 573 21. Lu T, Forgetta V, Greenwood CMT, Zhou S, Richards JB. Circulating Proteins Influencing
574 Psychiatric Disease: A Mendelian Randomization Study. *Biol Psychiatry*. Jan 1 2023;93(1):82-91.
575 doi:10.1016/j.biopsych.2022.08.015
- 576 22. Yoshiji S, Butler-Laporte G, Lu T, et al. Proteome-wide Mendelian randomization
577 implicates nephronectin as an actionable mediator of the effect of obesity on COVID-19 severity.
578 *Nature Metabolism*. 2023;5(2):248-264.
- 579 23. Beck RW, Bergenstal RM, Laffel LM, Pickup JC. Advances in technology for management
580 of type 1 diabetes. *Lancet*. Oct 5 2019;394(10205):1265-1273. doi:10.1016/S0140-
581 6736(19)31142-0
- 582 24. Lind M, Svensson AM, Kosiborod M, et al. Glycemic control and excess mortality in type
583 1 diabetes. *N Engl J Med*. Nov 20 2014;371(21):1972-82. doi:10.1056/NEJMoa1408214
- 584 25. Conus S, Simon HU. Cathepsins: key modulators of cell death and inflammatory
585 responses. *Biochem Pharmacol*. Dec 1 2008;76(11):1374-82. doi:10.1016/j.bcp.2008.07.041
- 586 26. Reiser J, Adair B, Reinheckel T. Specialized roles for cysteine cathepsins in health and
587 disease. *J Clin Invest*. Oct 2010;120(10):3421-31. doi:10.1172/JCI42918
- 588 27. Floyel T, Brorsson C, Nielsen LB, et al. CTSH regulates beta-cell function and disease
589 progression in newly diagnosed type 1 diabetes patients. *Proc Natl Acad Sci U S A*. Jul 15
590 2014;111(28):10305-10. doi:10.1073/pnas.1402571111
- 591 28. D'Angelo ME, Bird PI, Peters C, Reinheckel T, Trapani JA, Sutton VR. Cathepsin H is an
592 additional convertase of pro-granzyme B. *Journal of Biological Chemistry*. 2010;285(27):20514-
593 20519.
- 594 29. Cooper JD, Smyth DJ, Smiles AM, et al. Meta-analysis of genome-wide association study
595 data identifies additional type 1 diabetes risk loci. *Nat Genet*. Dec 2008;40(12):1399-401.
596 doi:10.1038/ng.249
- 597 30. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-
598 analysis find that over 40 loci affect risk of type 1 diabetes. *Nature genetics*. 2009;41(6):703-707.
- 599 31. Ye J, Richardson TG, Mcardle WL, et al. Identification of loci where DNA methylation
600 potentially mediates genetic risk of type 1 diabetes. *J Autoimmun*. Sep 2018;93:66-75.
601 doi:10.1016/j.jaut.2018.06.005
- 602 32. Usher-Smith JA, Thompson MJ, Sharp SJ, Walter FM. Factors associated with the
603 presence of diabetic ketoacidosis at diagnosis of diabetes in children and young adults: a
604 systematic review. *Bmj*. 2011;343
- 605 33. Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and
606 phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically

- 607 stratified survival analysis from UK Biobank. *Lancet Diabetes Endocrinol.* Feb 2018;6(2):122-129.
608 doi:10.1016/S2213-8587(17)30362-5
- 609 34. Mortensen HB, Swift PG, Holl RW, et al. Multinational study in children and adolescents
610 with newly diagnosed type 1 diabetes: association of age, ketoacidosis, HLA status, and
611 autoantibodies on residual beta-cell function and glycemic control 12 months after diagnosis.
612 *Pediatr Diabetes.* Jun 2010;11(4):218-26. doi:10.1111/j.1399-5448.2009.00566.x
- 613 35. Inshaw JR, Cutler AJ, Crouch DJ, Wicker LS, Todd JA. Genetic variants predisposing most
614 strongly to type 1 diabetes diagnosed under age 7 years lie near candidate genes that function
615 in the immune system and in pancreatic β -cells. *Diabetes Care.* 2020;43(1):169-177.
- 616 36. Yoshida H, Hunter CA. The immunobiology of interleukin-27. *Annu Rev Immunol.*
617 2015;33:417-43. doi:10.1146/annurev-immunol-032414-112134
- 618 37. Pflanz S, Timans JC, Cheung J, et al. IL-27, a heterodimeric cytokine composed of EB13
619 and p28 protein, induces proliferation of naive CD4+ T cells. *Immunity.* Jun 2002;16(6):779-90.
620 doi:10.1016/s1074-7613(02)00324-2
- 621 38. Owaki T, Asakawa M, Fukai F, Mizuguchi J, Yoshimoto T. IL-27 induces Th1
622 differentiation via p38 MAPK/T-bet- and intercellular adhesion molecule-1/LFA-1/ERK1/2-
623 dependent pathways. *J Immunol.* Dec 1 2006;177(11):7579-87.
624 doi:10.4049/jimmunol.177.11.7579
- 625 39. Artis D, Villarino A, Silverman M, et al. The IL-27 receptor (WSX-1) is an inhibitor of
626 innate and adaptive elements of type 2 immunity. *The Journal of Immunology.*
627 2004;173(9):5626-5634.
- 628 40. Stumhofer JS, Laurence A, Wilson EH, et al. Interleukin 27 negatively regulates the
629 development of interleukin 17-producing T helper cells during chronic inflammation of the
630 central nervous system. *Nat Immunol.* Sep 2006;7(9):937-45. doi:10.1038/ni1376
- 631 41. Stumhofer JS, Tait ED, Iijima WJQ, et al. A role for IL-27p28 as an antagonist of gp130-
632 mediated signaling. *Nature immunology.* 2010;11(12):1119-1126.
- 633 42. Brooke G, Holbrook JD, Brown MH, Barclay AN. Human lymphocytes interact directly
634 with CD47 through a novel member of the signal regulatory protein (SIRP) family. *The Journal of*
635 *Immunology.* 2004;173(4):2562-2570.
- 636 43. Sharp RC, Brown ME, Shapiro MR, Posgai AL, Brusko TM. The Immunoregulatory Role of
637 the Signal Regulatory Protein Family and CD47 Signaling Pathway in Type 1 Diabetes. *Front*
638 *Immunol.* 2021;12:739048. doi:10.3389/fimmu.2021.739048
- 639 44. Hatherley D, Graham SC, Turner J, Harlos K, Stuart DI, Barclay AN. Paired receptor
640 specificity explained by structures of signal regulatory proteins alone and complexed with CD47.
641 *Molecular cell.* 2008;31(2):266-277.
- 642 45. Dehmani S, Nerriere-Daguin V, Neel M, et al. SIRPgamma-CD47 Interaction Positively
643 Regulates the Activation of Human T Cells in Situation of Chronic Stimulation. *Front Immunol.*
644 2021;12:732530. doi:10.3389/fimmu.2021.732530
- 645 46. Quick CB, Fisher RA, Harris H. A kinetic study of the isozymes determined by the three
646 human phosphoglucomutase loci PGM1, PGM2, and PGM3. *Eur J Biochem.* Mar 1
647 1974;42(2):511-7. doi:10.1111/j.1432-1033.1974.tb03366.x
- 648 47. Tegtmeyer LC, Rust S, van Scherpenzeel M, et al. Multiple phenotypes in
649 phosphoglucomutase 1 deficiency. *N Engl J Med.* Feb 6 2014;370(6):533-42.
650 doi:10.1056/NEJMoa1206605

- 651 48. Wilcox G. Insulin and insulin resistance. *Clinical biochemist reviews*. 2005;26(2):19.
- 652 49. Inshaw JRJ, Sidore C, Cucca F, et al. Analysis of overlapping genetic association in type 1
653 and type 2 diabetes. *Diabetologia*. Jun 2021;64(6):1342-1347. doi:10.1007/s00125-021-05428-0
- 654 50. Xue A, Wu Y, Zhu Z, et al. Genome-wide association analyses identify 143 risk variants
655 and putative regulatory mechanisms for type 2 diabetes. *Nat Commun*. Jul 27 2018;9(1):2941.
656 doi:10.1038/s41467-018-04951-w
- 657 51. Forgetta V, Manousaki D, Istomine R, et al. Rare Genetic Variants of Large Effect
658 Influence Risk of Type 1 Diabetes. *Diabetes*. Apr 2020;69(4):784-795. doi:10.2337/db19-0831
- 659 52. Bobbala D, Mayhue M, Menendez A, Ilangumaran S, Ramanathan S. Trans-presentation
660 of interleukin-15 by interleukin-15 receptor alpha is dispensable for the pathogenesis of
661 autoimmune type 1 diabetes. *Cellular & molecular immunology*. 2017;14(7):590-596.
- 662 53. Kuczyński S, Winiarska H, Abramczyk M, Szczawińska K, Wierusz-Wysocka B, Dworacka
663 M. IL-15 is elevated in serum patients with type 1 diabetes mellitus. *Diabetes research and
664 clinical practice*. 2005;69(3):231-236.
- 665 54. Kaur S, Mirza AH, Brorsson CA, et al. The genetic and regulatory architecture of ERBB3-
666 type 1 diabetes susceptibility locus. *Mol Cell Endocrinol*. Jan 5 2016;419:83-91.
667 doi:10.1016/j.mce.2015.10.002
- 668 55. Wang H, Jin Y, Reddy MV, et al. Genetically dependent ERBB3 expression modulates
669 antigen presenting cell function and type 1 diabetes risk. *PLoS One*. Jul 26 2010;5(7):e11789.
670 doi:10.1371/journal.pone.0011789
- 671 56. O'Connor L, Brage S, Griffin SJ, Wareham NJ, Forouhi NG. The cross-sectional association
672 between snacking behaviour and measures of adiposity: the Fenland Study, UK. *British journal
673 of nutrition*. 2015;114(8):1286-1293.
- 674 57. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS
675 summary statistics identifies additional variants influencing complex traits. *Nat Genet*. Mar 18
676 2012;44(4):369-75, S1-3. doi:10.1038/ng.2213
- 677 58. Chiou J, Geusz RJ, Okino M-L, et al. Interpreting type 1 diabetes risk with genetics and
678 single-cell epigenomics. *Nature*. 2021;594(7863):398-402.
- 679 59. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-
680 specific haplotype structure and linking correlated alleles of possible functional variants.
681 *Bioinformatics*. Nov 1 2015;31(21):3555-7. doi:10.1093/bioinformatics/btv402
- 682 60. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic
683 variation. *Nature*. Oct 1 2015;526(7571):68-74. doi:10.1038/nature15393
- 684 61. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for
685 Mendelian randomization. *Stat Methods Med Res*. Oct 2017;26(5):2333-2355.
686 doi:10.1177/0962280215597579
- 687 62. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A
688 framework for the investigation of pleiotropy in two-sample summary data Mendelian
689 randomization. *Stat Med*. May 20 2017;36(11):1783-1802. doi:10.1002/sim.7221
- 690 63. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian
691 randomization with some invalid instruments using a weighted median estimator. *Genetic
692 epidemiology*. 2016;40(4):304-314.

- 693 64. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments:
694 effect estimation and bias detection through Egger regression. *International journal of*
695 *epidemiology*. 2015;44(2):512-525.
- 696 65. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in
697 Mendelian randomization studies. *Int J Epidemiol*. Jun 2011;40(3):755-64.
698 doi:10.1093/ije/dyr036
- 699 66. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal
700 inference across the human phenome. *Elife*. May 30 2018;7doi:10.7554/eLife.34408
- 701 67. Zuber V, Grinberg NF, Gill D, et al. Combining evidence from Mendelian randomization
702 and colocalization: Review and comparison of approaches. *Am J Hum Genet*. May 5
703 2022;109(5):767-782. doi:10.1016/j.ajhg.2022.04.001
- 704 68. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping
705 and genomic data. *Nature*. Oct 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
- 706 69. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between
707 pairs of genetic association studies using summary statistics. *PLoS Genet*. May
708 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383
- 709 70. Mountjoy E, Schmidt EM, Carmona M, et al. An open approach to systematically
710 prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nature*
711 *genetics*. 2021;53(11):1527-1533.
- 712 71. Ghousaini M, Mountjoy E, Carmona M, et al. Open Targets Genetics: systematic
713 identification of trait-associated genes using large-scale genetics and functional genomics.
714 *Nucleic Acids Res*. Jan 8 2021;49(D1):D1311-D1320. doi:10.1093/nar/gkaa840
- 715 72. Ochoa D, Hercules A, Carmona M, et al. Open Targets Platform: supporting systematic
716 drug-target identification and prioritisation. *Nucleic Acids Res*. Jan 8 2021;49(D1):D1302-D1310.
717 doi:10.1093/nar/gkaa1027
- 718 73. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and
719 health in the UK Biobank. *Nature*. Oct 2023;622(7982):329-338. doi:10.1038/s41586-023-
720 06592-6
- 721 74. Ebenesersdottir SS, Sandoval-Velasco M, Gunnarsdottir ED, et al. Ancient genomes from
722 Iceland reveal the making of a human population. *Science*. Jun 1 2018;360(6392):1028-1032.
723 doi:10.1126/science.aar2625
- 724 75. Robertson CC, Inshaw JRJ, Onengut-Gumuscu S, et al. Fine-mapping, trans-ancestral and
725 genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *Nat*
726 *Genet*. Jul 2021;53(7):962-971. doi:10.1038/s41588-021-00880-5
- 727 76. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human
728 tissues. *Science*. Sep 11 2020;369(6509):1318-1330. doi:10.1126/science.aaz1776
- 729 77. Verweij N, Haas ME, Nielsen JB, et al. Germline Mutations in CIDEA and Protection
730 against Liver Disease. *N Engl J Med*. Jul 28 2022;387(4):332-344. doi:10.1056/NEJMoa2117872
- 731 78. Akbari P, Gilani A, Sosina O, et al. Sequencing of 640,000 exomes identifies GPR75
732 variants associated with protection from obesity. *Science*. Jul 2
733 2021;373(6550)doi:10.1126/science.abf8683
- 734 79. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
735 expression analysis of digital gene expression data. *Bioinformatics*. Jan 1 2010;26(1):139-40.
736 doi:10.1093/bioinformatics/btp616

- 737 80. Dominguez Conde C, Xu C, Jarvis LB, et al. Cross-tissue immune cell analysis reveals
738 tissue-specific features in humans. *Science*. May 13 2022;376(6594):eabl5197.
739 doi:10.1126/science.abl5197
- 740 81. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among
741 sequence variation and human phenotype. *Nucleic acids research*. 2014;42(D1):D980-D985.
- 742 82. Gadd DA, Hillary RF, Kuncheva Z, et al. Blood protein levels predict leading incident
743 diseases and mortality in UK Biobank. *medRxiv*. 2023:2023.05. 01.23288879.
- 744

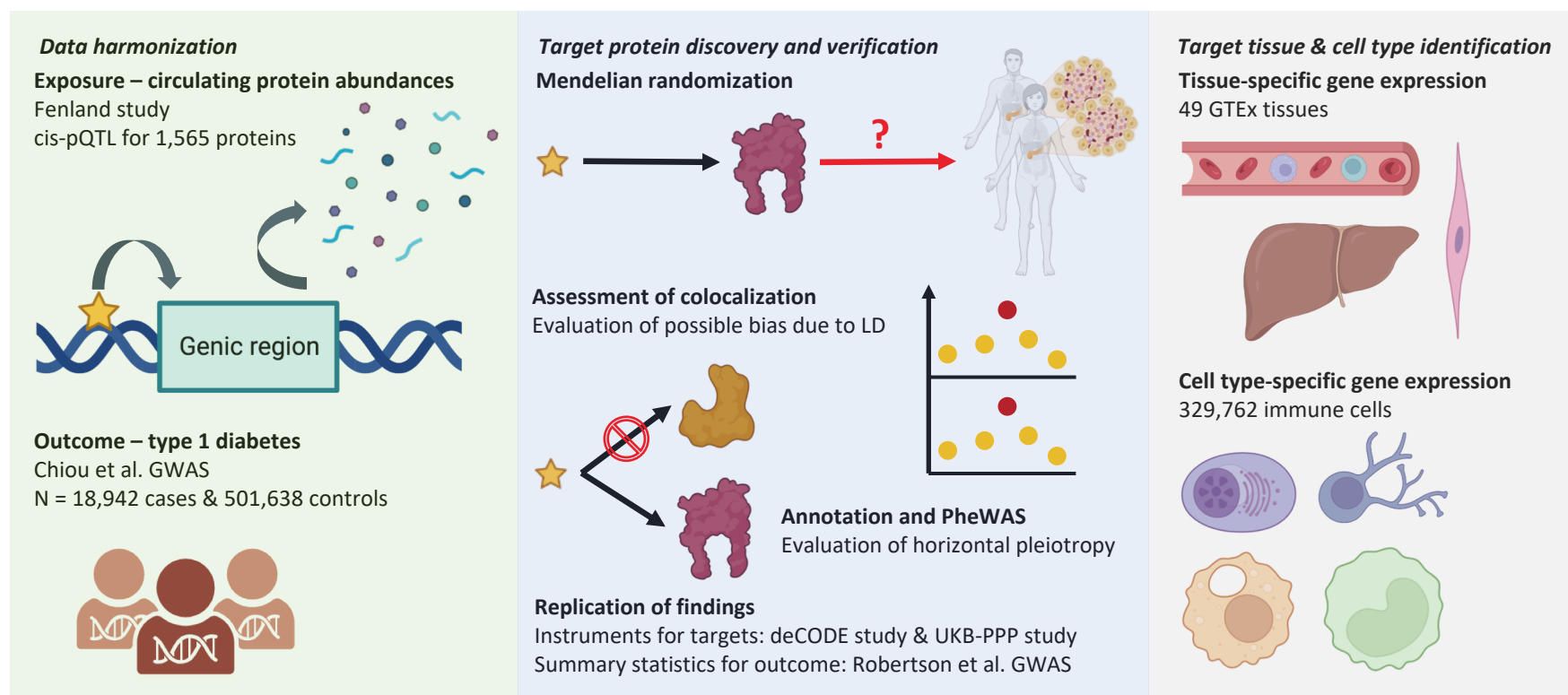


Figure 1. Overview of study. Mendelian randomization (MR) was conducted leveraging genome-wide association studies (GWASs) of circulating protein abundances in the Fenland study as well as a meta-analysis of type 1 diabetes GWASs. Colocalization analyses and evaluation of horizontal pleiotropy through annotation and phenome-wide association study (PheWAS) were conducted to verify MR assumptions. Significant associations were replicated using other proteomic studies and another meta-analysis of type 1 diabetes GWASs. Gene expression enrichment analyses were conducted to identify potential candidate tissues and cell types.

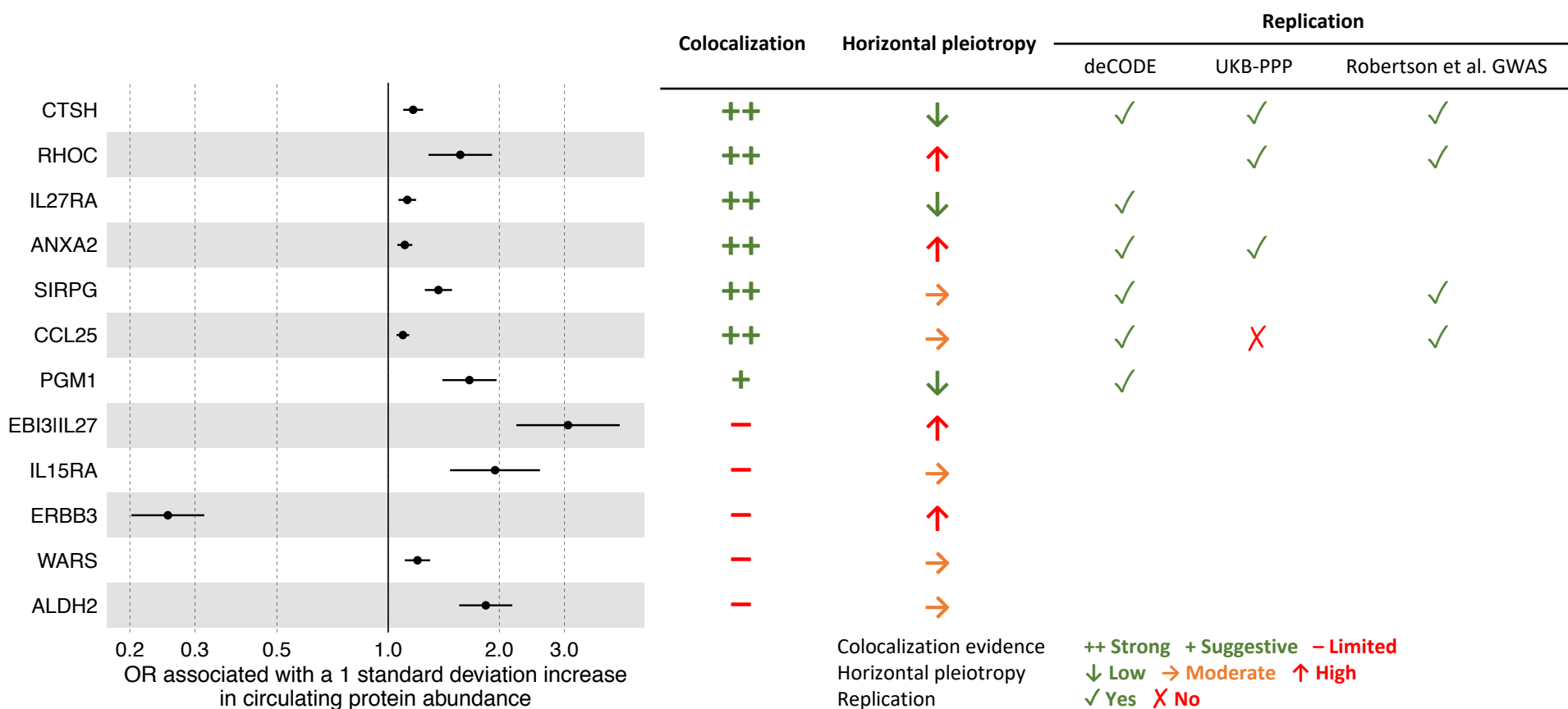


Figure 2. Target protein prioritization. Associations between circulating protein abundances and type 1 diabetes risk that withstood Bonferroni correction of multiple testing are illustrated. Target proteins are ordered by posterior probability of colocalization (Supplementary Table S4). A posterior probability of colocalization > 80% was considered strong evidence of colocalization, while a posterior probability of colocalization > 50% was considered suggestive evidence of colocalization. Risk of horizontal pleiotropy was assessed using V2G scores for quantifying functional connections between genetic instruments and target protein-coding genes, as well as phenome-wide associations for exploring potential pleiotropic pathways (Methods). Associations supported by strong or suggestive colocalization evidence were replicated using additional resources (Methods). Blank space indicates that no genetic instrument or proxy was identified to replicate the association. Target proteins were prioritized based on strong or suggestive colocalization evidence, the absence of a high risk of horizontal pleiotropy, and the consistent replication of associations with the risk of type 1 diabetes.

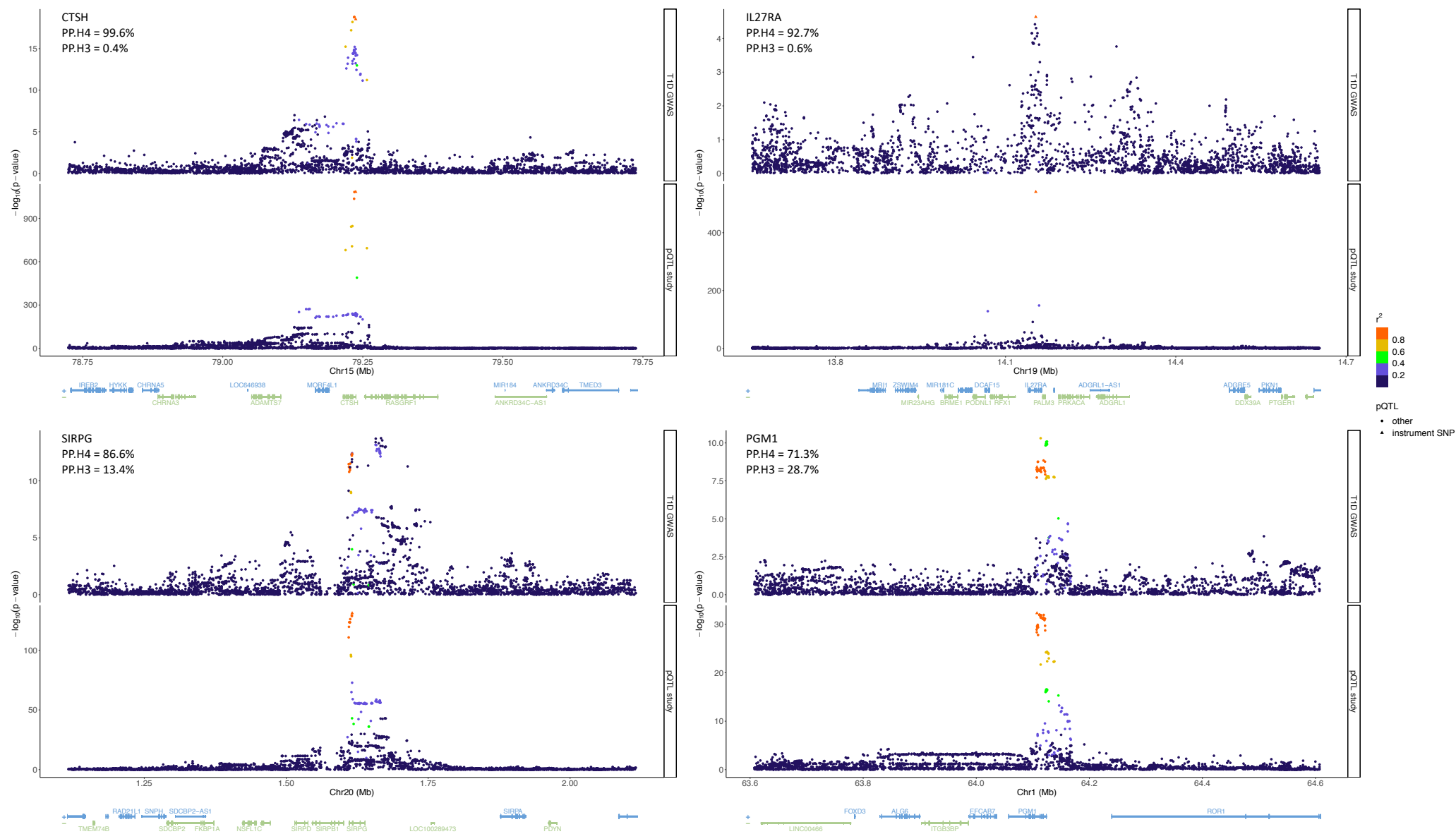


Figure 3. Colocalization of genetic associations with circulating abundances of prioritized target proteins and the risk of type 1 diabetes. The lead cis-genetic instruments are indicated. Genetic variants located in a ± 500 kb window centered around each genetic instrument are plotted with their significance in respective studies, and colored by the magnitude of correlation (linkage disequilibrium, LD r^2) with the corresponding instrument. For each target protein, the posterior probability of colocalization (PP.H4) and the posterior probability of co-existence of two distinct causal variants (PP.H3) are indicated. The UCSC known gene tracks are presented, with gene models colored by their respective strands.

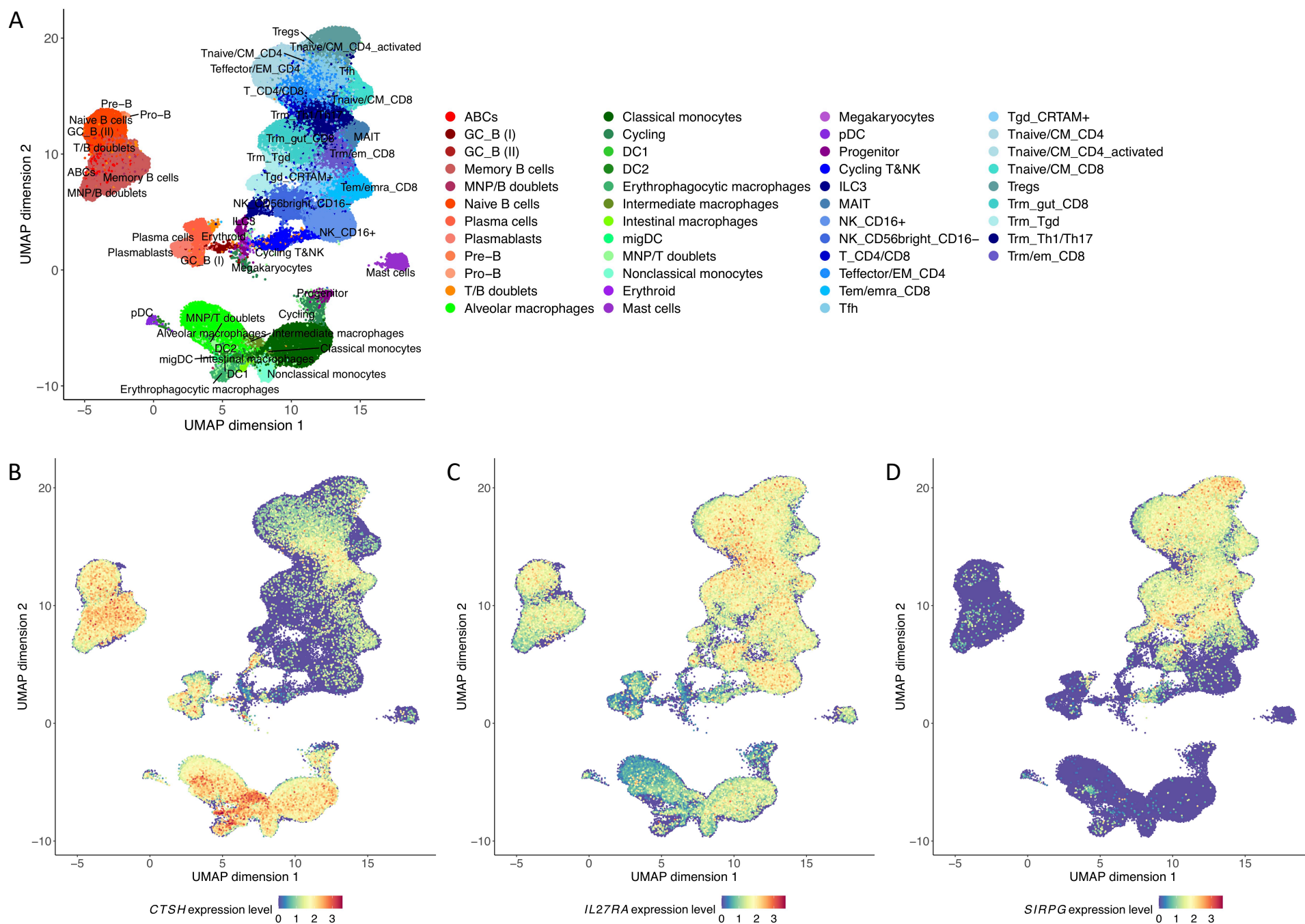


Figure 5. Single-cell gene expression profiles of *CTSH*, *IL27RA*, and *SIRPG* in immune cells. (A) Visualization of 329,762 immune cells based on Uniform Manifold Approximation and Projection (UMAP) of their transcriptomes. Cells are colored by manually curated cell types. Red colors: B cell compartment; Green colors: myeloid compartment; Purple colors: miscellaneous cell types; Blue colors: T cell compartment. Descriptions of cell types are available in Supplementary Table S11. Normalized gene expression levels of (B) *CTSH*, (C) *IL27RA*, and (D) *SIRPG* are visualized. UMAP coordinates, cell type annotations, and normalized gene expression levels were obtained from the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) under the accession ID SCP1845.