

1 **Linking the plasma proteome to genetics in individuals from continental Africa**
2 **provides insights into type 2 diabetes pathogenesis**

3 Opeyemi Soremekun^{1,2,3}, Young-Chan Park¹, Mauro Tutino¹, Allan Kalungi³, N. William
4 Rayner¹, Moffat Nyirenda³, Segun Fatumo^{*1,3,4}, Eleftheria Zeggini^{*1,5}

5 ¹Institute of Translational Genomics, Helmholtz Munich, 85764 Neuherberg, Germany

6 ²Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences,
7 University of KwaZulu-Natal, Westville Campus, Durban, South Africa.

8 ³Medical Research Council, Uganda Virus Research Institute and London School of Hygiene
9 and Tropical Medicine (MRC/UVRI & LSHTM), Entebbe, Uganda

10 ⁴Precision Healthcare University Research Institute Queen Mary University of London, United
11 Kingdom

12 ⁵TUM School of Medicine, Technical University Munich and Klinikum Rechts der Isar, 81675
13 Munich, Germany

14

15 *Joint Corresponding Authors

16

17

18 **Individuals of African ancestry remain largely underrepresented in genetic and**
19 **proteomic studies. Here, we measure the levels of 2,873 proteins using the Olink**
20 **proximity extension assay in plasma samples from 163 individuals with type 2 diabetes**
21 **(T2D) or prediabetes and 362 normoglycemic controls from the Ugandan population for**
22 **the first time. We identify 88 differentially expressed proteins between the two groups**
23 **and 208 proteins associated with cardiometabolic traits. We link genome-wide data to**
24 **protein expression levels and construct the first protein quantitative trait locus (pQTL)**
25 **map in this population. We identify 399 independent associations with 346 (86.7%) cis-**
26 **pQTLs and 53 (13.3%) trans-pQTLs. 16.7% of the cis-pQTLs and all of the trans-pQTLs**
27 **have not been previously reported in African-ancestry individuals. Of these, 37 pQTLs**
28 **have not been previously reported in any population. We find evidence for colocalization**
29 **between a pQTL for SIRPA and T2D genetic risk. Mendelian randomization analysis**
30 **identified 20 proteins causally associated with T2D. Our findings reveal proteins causally**
31 **implicated in the pathogenesis of T2D, which may be leveraged for personalized medicine**

32 **NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**
tailored to African-ancestry individuals.

33 **Main**

34 Type 2 diabetes (T2D) is becoming a major public health concern in Africa, congruent with
35 the complex interplay of genetic, environmental and socio-economic factors ¹⁻³. The shift
36 from traditional to urbanized and sedentary lifestyles, accompanied by dietary changes, has
37 contributed to this increasing prevalence⁴⁻⁶. According to the International Diabetes Federation
38 (IDF), it is predicted that, globally, people with T2D will rise by 51% reaching 700.2 million
39 by 2045 from 463 million in 2019 ⁷. A significant increase of 143% is anticipated in Africa,
40 with numbers expected to rise from 19.4 million in 2019 to 47.1 million in 2045 ⁷. Hemoglobin
41 A1c (HbA1c), also known as glycated hemoglobin ⁸, provides an estimate of the blood sugar
42 level over a period of two to three months by measuring the percentage of hemoglobin with
43 attached glucose^{9,10}. An HbA1c level of 6.5% or higher on two separate tests typically indicates
44 diabetes. Levels between 5.7% and 6.4% suggest prediabetes, and values below 5.7% are
45 considered normal ¹¹.

46

47 Proteins drive many biological functions, can be used as biomarkers of disease onset and
48 progression, and are the primary targets of drug therapies. With advancement in technologies,
49 high-throughput quantification of circulating proteins on an epidemiological scale is now
50 possible. Combining proteomic and genomic data for blood-based pQTLs has led to the
51 identification of hundreds of associations between genetic variants and protein levels ¹²⁻¹⁶. A
52 fraction of African-ancestry individuals in diaspora have been studied in proteomics studies to
53 date ^{15,17}, with continental Africans largely underrepresented.

54 To address this, we have measured 2873 proteins using the Olink PEA Explore assay in plasma
55 samples of 163 individuals with prediabetes or type 2 diabetes (cases) (defined as HbA1c >
56 5.7%) as well as 362 normoglycemic controls (defined as HbA1c < 5.7%) (Table 1). We have
57 performed differential protein expression analysis between the two groups, and have
58 undertaken proteomic genetic association analysis to identify sequence variants influencing
59 protein levels. We have subsequently examined the role of the identified pQTLs in type 2
60 diabetes using colocalization and Mendelian randomization analysis.

61

62

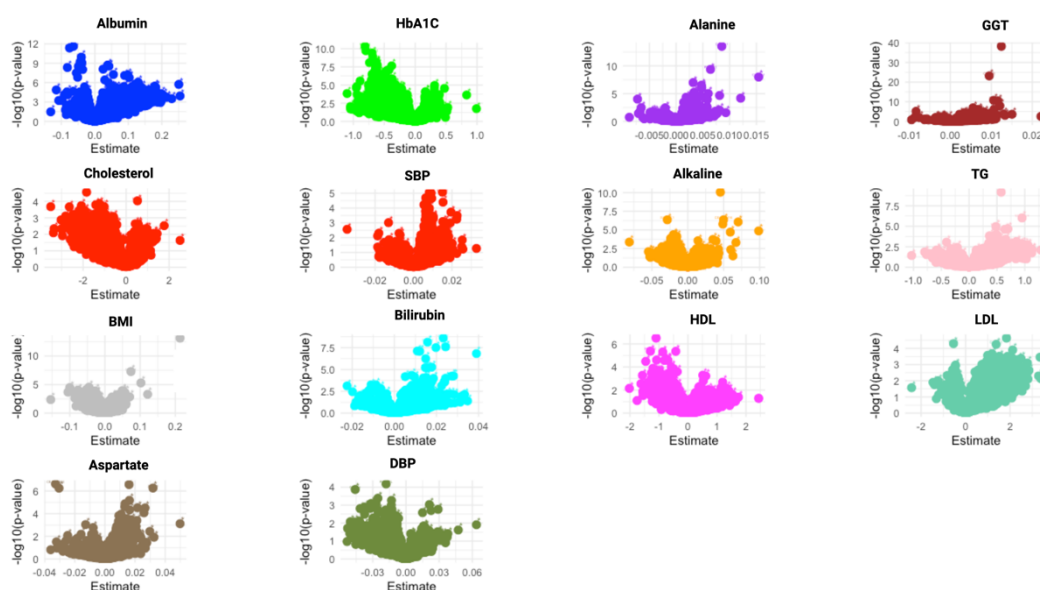
63

64

65 **Table 1:** Clinical Characteristics of study participants

	Cases	Controls
Number of participants (n, %)	163(31.05%)	362(68.95%)
Age (year) (mean ± SD)	49.82 ± 18.5	50.39 ± 17.76
Male (n, %)	47(28.83%)	139(38.40%)
Female (n, %)	116(71.17%)	223(61.60%)
BMI (kg/m ²)	23.4 ± 4.68	22.06 ± 4.35
HbA1c (%)	6.46 ± 1.24	5.13 ± 0.48

66 First, we studied the association between protein levels and cardiometabolic traits measured in
67 the Ugandan cohort (supplementary 1). A total of 208 proteins were associated with HbA1c,
68 42 with HDL, and 46 with LDL at a false discovery rate (FDR) of 5% (Figure 1). Some of the
69 associations, such as ERCC1 found to be associated with HbA1c ($P_{adj} = 6.77 \times 10^{-7}$) and HDL
70 ($P_{adj} = 1.91 \times 10^{-2}$), have been shown to affect glucose intolerance in a progeroid-deficient
71 animal model causing an autoinflammatory response that leads to fat loss and insulin resistance
72 ¹⁸. PTPN9 was also found to be associated with HbA1c (2.96×10^{-6}); this protein belongs to a
73 family of protein tyrosine phosphatases (PTN) that are known to be involved in cellular insulin
74 resistance associated with T2D ^{19,20}. Due to the role PTPN9 plays in insulin resistance, it is
75 currently being investigated as a potential diabetes drug target ^{21,22}.

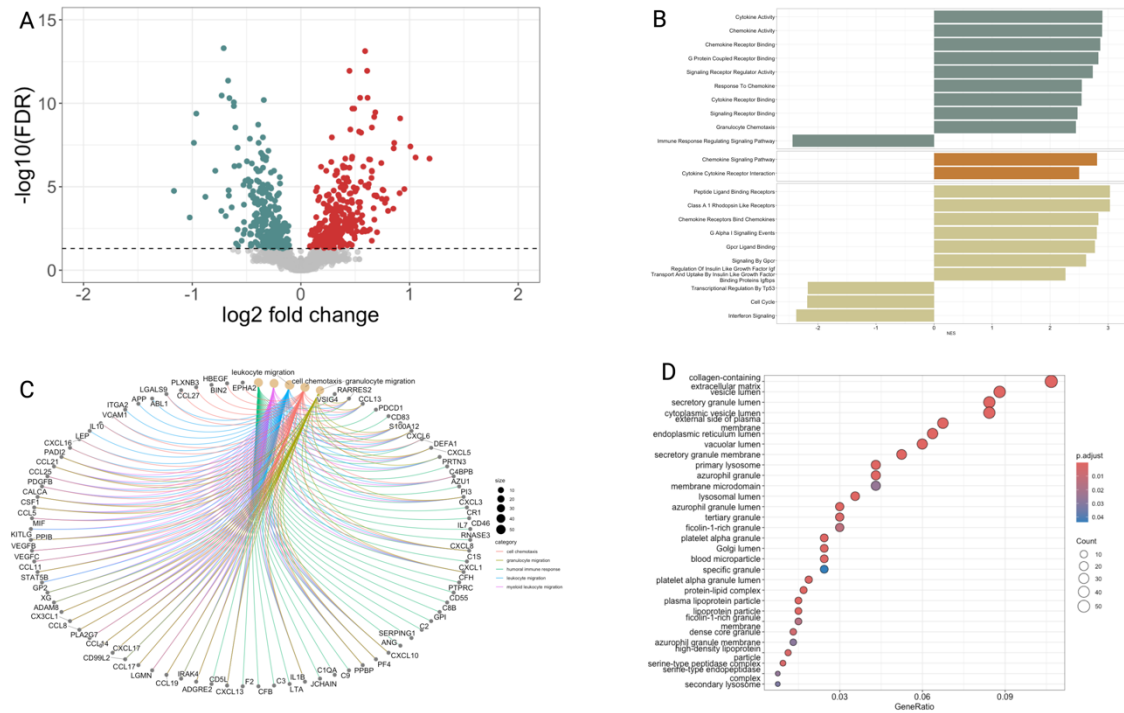


76
77 **Figure 1:** Association of protein levels with clinical traits. The y axis represents the FDR-
78 adjusted $-\log_{10}(p\text{-value})$ of the association and the x axis of each plot represents the effect size

79 estimated with linear regression. [Body mass index (BMI), Systolic blood pressure (SBP),
80 Diastolic blood pressure (DBP), High density lipoproteins (HDL), Gamma glutamate
81 transferase (GGT), Triglycerides (TG), Low density lipoproteins (LDL)].
82

83 Next, we sought to identify differentially expressed protein (DEP) levels between cases and
84 controls. Using a set of 2873 unique proteins, DEPs were defined based on a 2-fold change
85 ($|\log_2\text{FC}| > 0.5$) in expression levels at a false discovery rate (FDR) of 5%. This led to the
86 identification of 88 DEPs. Among these, 57 were found to be significantly upregulated, with
87 \log_2 fold changes ranging from 0.50 to 1.18, while 31 proteins were downregulated with \log_2
88 fold changes between -0.51 to -1.17 (Figure 2A), (supplementary 2). EDIL3 ((EGF-like repeats
89 and discoïdin I-like domains 3), associated with processes such as cell adhesion, migration and
90 vascular development, showed the most significant upregulation with an adjusted p-value of
91 1.2×10^{-13} . EDIL3 has been shown to be differentially expressed in adipose tissue of insulin-
92 resistant and insulin sensitive individuals^{23,24}, and to be involved in angiogenesis²⁵⁻²⁷.
93 Impaired angiogenesis has been implicated in the progression of diabetic retinopathy and
94 nephropathy^{28,29}. Studies also indicate that EDIL3 is involved in inflammatory responses,
95 which are known to contribute to insulin resistance, a hallmark of T2D³⁰. LPCAT2
96 (Lysophosphatidylcholine Acyltransferase 2) was identified as the most significantly
97 downregulated protein with $P_{\text{adj}} = 4.41 \times 10^{-14}$. LPCAT2 plays a critical role in lipid
98 metabolism, specifically in the remodeling of phospholipids in cell membranes^{31,32}.
99 Downregulation of LPCAT2 may lead to changes in membrane lipids that contribute to
100 impaired insulin signaling. The DEPs were primarily enriched in GO terms such as chemokine
101 receptor binding, and chemokine and cytokine activity (Figure 2B, supplementary 3). For the
102 first time, EDIL3 and LPCAT2 are identified to be differentially expressed between diabetes
103 cases and controls.

104



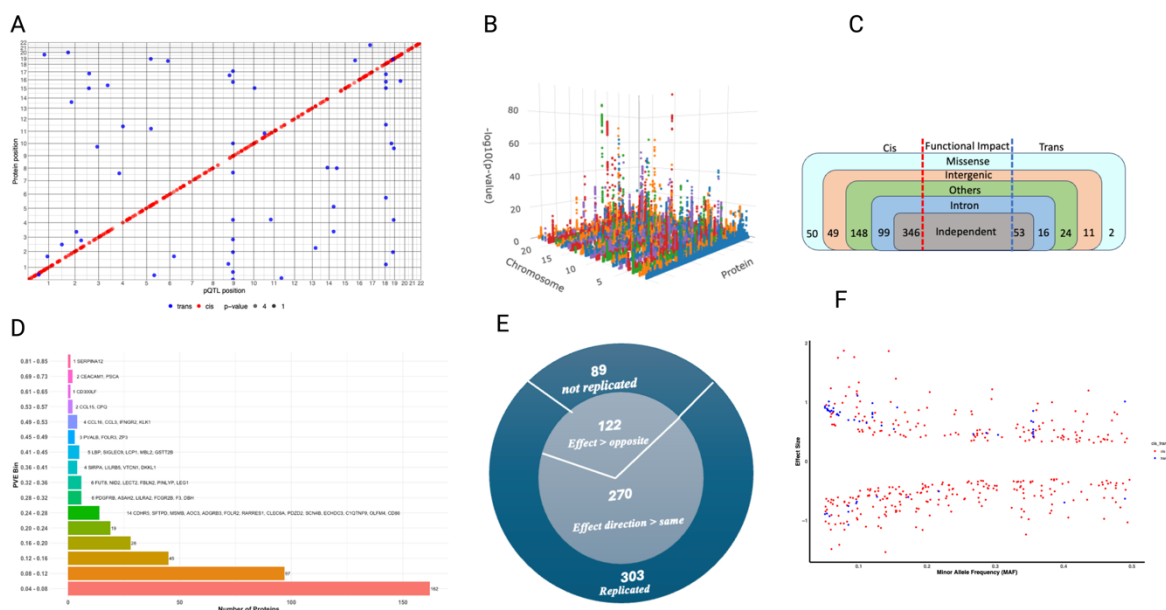
105

106 **Figure 2:** Volcano plot showing the differentially expressed proteins, with significantly
 107 overexpressed proteins annotated in red and downregulated proteins in teal using linear model
 108 implemented in Limma (A). Gene sets significantly associated with DEPs at 5% FDR (B).
 109 Biological process of significantly associated DEPs at 5% FDR(C). Cellular component of
 110 significantly associated DEPs at 5% FDR using a hypergeometric test (D)

111

112

113 We then undertook pQTL analysis with up to 15.8 million imputed variants for 2873 proteins,
 114 following quality control. Using approximate conditional and joint stepwise model selection,
 115 we identified 399 independent associations following multiple testing correction at P value
 116 thresholds of $P < 1.46 \times 10^{-6}$ and $P < 2.2 \times 10^{-10}$ for cis and trans pQTLs, respectively
 117 (supplementary 4). We identified 346 (86.7%) cis-pQTLs, and 53 (13.3%) trans-pQTLs. Seven
 118 proteins had both cis- and trans-pQTLs. We also identified 4 trans-pQTLs located within the
 119 following pleiotropic loci: *PRSS27*, *CDH1*, *EPCAM*, *LPO*, *LEGI*, *TSPAN8*, *MUC2*, *SELE*,
 120 *IL7R*, *ALPI*, *KLK1*, *CDH17*, *CDH5*, *PTPRM*, and *SERPINI2*.



121

122 **Figure 3:** Scatterplot of pQTL variant location against the location of the gene encoding the
 123 target protein. Each dot represents an independent variant. Cis-pQTLs are colored in red, while
 124 trans-pQTLs are colored in blue (A). 3D Manhattan plot of identified pQTLs. (x axis = proteins,
 125 y axis = chromosome location, and z axis = $-\log_{10}$ p-values of each association) (B). Summary
 126 of identified pQTLs showing their functional consequences (C). Proportion of variance
 127 explained (PVE) by the conditionally independent pQTLs categorized into bins (D).
 128 Replication of conditionally independent pQTLs in the UKBB-PPP African-ancestry data
 129 highlighting replication via p values and direction of effect (E). Assessment of the relationship
 130 between the pQTLs effect sizes and minor allele frequency (MAF) (F).

131

132 To determine the novelty of the pQTLs identified in the Ugandan population, we compared
 133 against pQTLs of 47 genome-wide pQTL studies (supplementary 5). We identified 6
 134 independent cis-pQTLs and 31 independent trans-pQTLs that have not been previously
 135 reported in any population (supplementary 6). We compared our pQTL findings against the
 136 African-ancestry data of the UKBB-PPP and found that 16.7% (58 out of 346) of the
 137 discovered cis-pQTLs and all trans-pQTLs have not been previously reported (supplementary
 138 7). We tested the conditionally independent Uganda population pQTLs for replication in the
 139 UKBB-PPP. Of the 399 pQTLs, we were able to test 392 in the UKBB-PPP data. Of these, 303
 140 replicated at $P \leq 1.2 \times 10^{-4}$ (Bonferroni-corrected threshold) and 270 of these also had the same
 141 effect estimate direction (supplementary 8).

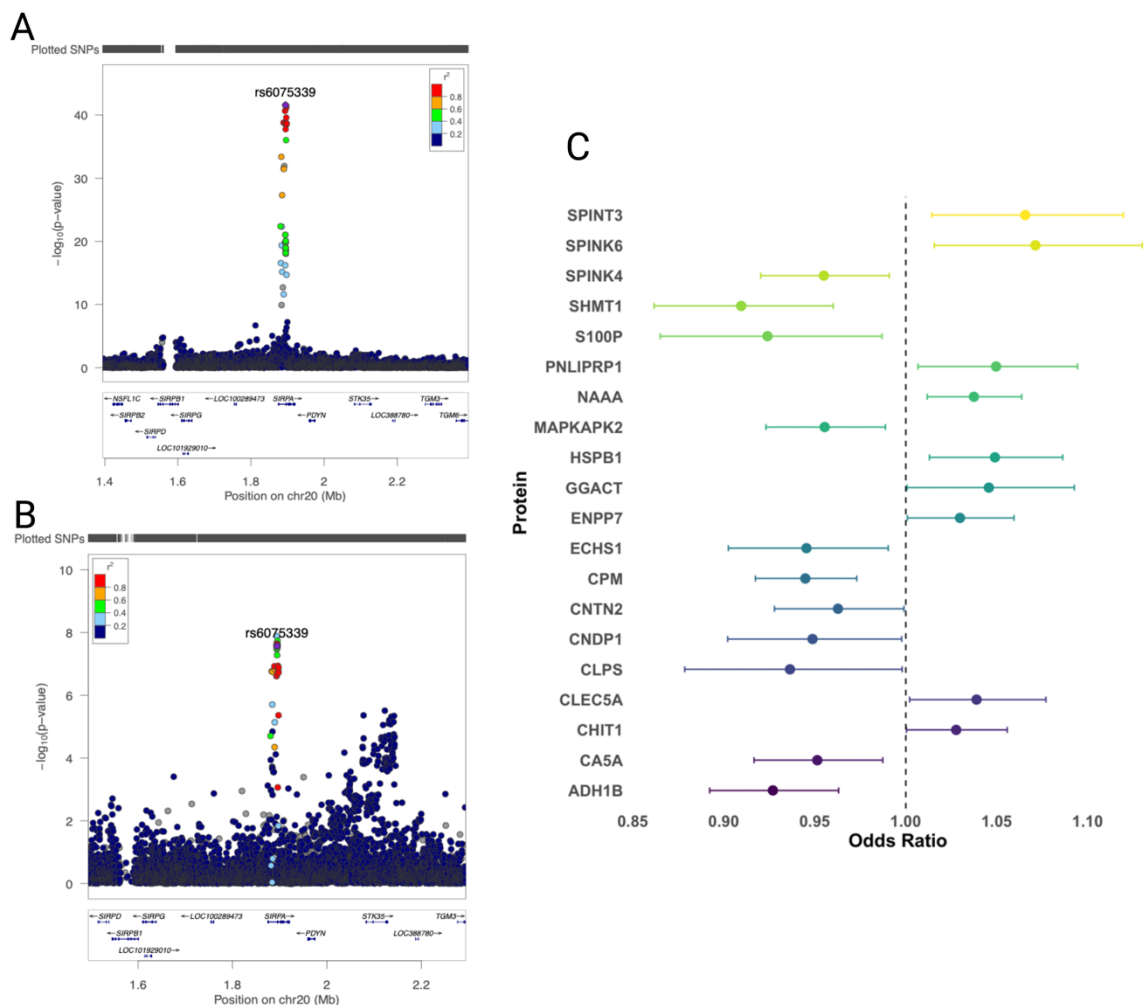
142

143 Next, we performed colocalization analysis to determine shared risk variants between the
 144 pQTLs and T2D using the largest multi-ancestry genome wide association study (GWAS)
 145 meta-analysis to date³³. We found evidence for a shared T2D risk variant (posterior probability

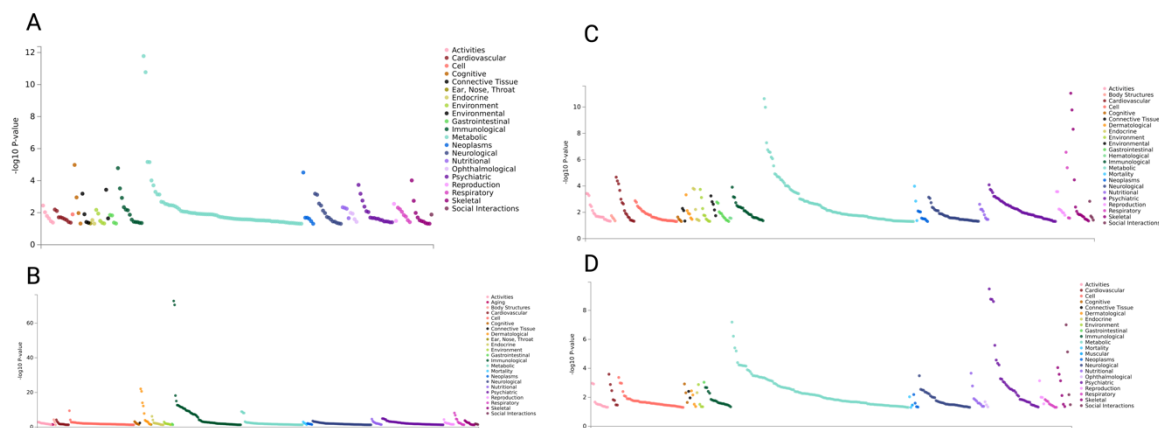
146 4 (PP4) = 95.5%) with pQTLs for the signal regulatory protein alpha (SIRPA) protein (Figure
147 4A,B). Integrating pQTLs with T2D GWAS revealed protein-disease colocalization that may
148 suggest the involvement of SIRPA in the pathogenesis of T2D. SIRPA is a member of the SIRP
149 family and is expressed on the surface of most T cells and some B cells. Genetic studies have
150 implicated SIRP signaling in diabetes pathogenesis. For example, a SNP in human *SIRPγ*,
151 encoding a SIRP family receptor that also binds CD47, was found to be associated with type 1
152 diabetes (T1D)³⁴. Wong *et al.*, also identified SIRPA as a causal gene that mediates diabetes
153 susceptibility in a nonobese diabetic mouse model³⁵.

154
155 To examine the causal relationship between the identified cis-pQTLs and T2D, we undertook
156 a Mendelian randomization analysis. We found 20 proteins to be causally associated with T2D.
157 Higher levels of SPINT3, SPINK6, PNLIPRP1, NAAA, HSPB1, GGACT, ENPP7, CLEC5A,
158 and CHIT1 were associated with increased risk of T2D, while SPINK4, SHMT1, S100P,
159 MAPKAPK2, ECHS1, CPM, CNTN2, CNDP1, CLPS, CA5A, and ADH1B showed a
160 protective effect on T2D risk (Figure 4C, supplementary 9). Several of these proteins such as
161 PNLIPRP1, NAAA, HSPB1, ENPP7, CLEC5A, CHIT1, SHMT1, S100P, MAPKAPK2,
162 ECHS1, CPM, CNDP1, CLPS, CA5A, and ADH1B have been shown to be associated with
163 diabetes through pathways involving lipid metabolism, inflammation, protease activity, and
164 hypermethylation etc³⁶⁻³⁹. PheWAS analysis also showed that some of these proteins, such as
165 SPINT3, SHMT1, ADH1B, and CNTN2, are associated with metabolic traits (Figure 5).

166
167 Lastly, we assembled a list of 1804 postulated effector genes for T2D from 9 GWAS studies.
168 If a gene coding for any of the proteins associated with identified pQTL in our study was found
169 in the curated list, we defined such gene/protein as reported, if not, we classified them as
170 previously unresolved. We identified 320 proteins previously unresolved as potentially linked
171 to effector genes for T2D based on these GWAS signals (supplementary 10).



172
 173 **Figure 4:** LocusZoom plots of the colocalizing SIRPA pQTL and T2D risk variant. The top
 174 panel shows the T2D GWAS p-values, while the bottom panel shows pQTL p-values for the
 175 same region (A,B). Mendelian randomization forest plot for proteins causally associated with
 176 T2D (C).



177

178 **Figure 5:** PheWAS plots for SPINT3 (A), CNTN2 (B), SHMT1 (3), and ADH1B (D).

179 Our work takes a first step toward addressing the underrepresentation of continental African
180 individuals in genetics and proteomics studies. As a result, we were able to delineate the
181 molecular landscape of 2873 unique proteins in a context that might be pivotal to understanding
182 drivers of T2D pathophysiology, identified 58 African ancestry-specific cis-pQTLs that have
183 not been reported previously, and identified 20 proteins that are causally associated with T2D.
184 The generalizability of these findings may be limited for the continent, since the population is
185 drawn from a single demographic group within Africa. There is, hence, a need to include more
186 ancestrally diverse populations in future studies.

187

188 Here, we have used the Olink targeted proteomic assay, which has some limitations, for
189 example only a subset of the full proteome is studied, and the affinity of aptamers may be
190 affected by missense variants. While HbA1c is a highly standardized and accurate test with
191 lower intraindividual variability compared to fasting glucose, in African ancestry individuals
192 using HbA1c as a blood sugar level indicator may not provide the full spectrum of the metabolic
193 conditions associated with T2D due to the prevalence of hemoglobinopathies such as glucose-
194 6-phosphate dehydrogenase (G6PD) deficiency. In individuals with G6PD deficiency, there is
195 increased susceptibility to hemolysis, which may lead to reduced HbA1c level potentially
196 leading to missed T2D diagnosis ^{40,41}.

197

198 In conclusion, the novel associations and causally-associated proteins identified offer
199 promising avenues for the development of targeted therapies and personalized treatment
200 strategies for T2D, contributing to improved management and prevention of this global health
201 challenge. Our findings demonstrate the utility and discovery opportunities afforded by the
202 inclusion of African ancestry individuals in large-scale proteomic studies.

203 **Methods**

204 **Study population**

205 The participants were selected from the Uganda Genome Resource (UGR), which is a subset
206 of the General Population Cohort (GPC). As previously described ^{42,43}, the GPC is a
207 population-based cohort made up of over 22,000 people from 25 nearby communities in the
208 remote Southwest Ugandan sub-county of Kyamulibwa sub-county, which is a part of Kalungu
209 district. We selected 528 samples from the UGR based on their age, sex and HbA1c. Following

210 hemolysis of anticoagulated whole blood, the concentrations of total haemoglobin and HbA1c
211 were measured using turbidimetric inhibition immunoassay-quant Haemoglobin Alc Gen.⁴².
212 In addition to the genotype QC described in⁴², we used Hardy Weinberg p-value (pHWE) <
213 1×10^{-6} .

214 **Association with clinical characteristics**

215 We used linear regression to determine the association between the protein levels and systolic
216 Systolic Blood Pressure (SBP), Diastolic Blood Pressure, alanine, albumin, alkaline, aspartate,
217 bilirubin, cholesterol, gamma-glutamyl transferase, High-Density Lipoprotein, Low-Density
218 Lipoprotein, triglycerides, and Hemoglobin A1c. All *p*-values were FDR corrected.

219 **Differential Expressed Proteins and Functional enrichment**

220 We determined differential expressed proteins (DEP) between the cases and control using
221 limma⁴⁴, and we used Benjamini–Hochberg False Discovery Rate (FDR) for multiple testing
222⁴⁵. DEP are defined as proteins with an FDR < 5% and a fold change greater than 0.5 ($|\log_2FC|$
223 > 0.5). To better understand the functional impact of the proteins, we used enrichr tools from
224 clusterProfiler⁴⁶ and then used Pearson's correlation analysis to investigate the association
225 between the expression levels of the DEPs. STRING database⁴⁷ was used to generate protein-
226 protein interaction network for the upregulated and downregulated protein categories.

227

228 **Proteomics Quality Control**

229 Olink's proximity extension assay (PEA) technology⁴⁸ was used to measure the plasma level
230 of 2978 proteins in 528 samples across eight Olink panels. The levels of protein expression
231 were measured logarithmically as Normalized Protein eXpression (NPX) units. We adjusted
232 all phenotypes using a linear regression for age, sex, plate number, sample collection season,
233 followed by an inverse-normal transformation of the residuals. During the quality control (QC)
234 process, we excluded one sample because the PCR plate well was empty, additional 2 samples
235 were further excluded due to missingness greater than 40%. Two samples SC1 and SC2 were
236 also excluded, these samples were initially included by O-link for internal control process. For
237 assay QC, 40 assays were excluded as they did not have Normalized Protein eXpression (NPX)
238 values. Additionally, we excluded 31 assays that had fraction of assay warning greater than
239 15%. No assay was excluded because of limit of detection (LOD). In all, 525 samples and 2873
240 assays remained after QC and were subsequently used for further analysis.

241

242 **Single-point association**

243 Covariates such as sex, age, plate, and mean protein expression per sample were regressed
244 using R's LM function. The residuals were then translated into z-scores and used
245 for association analysis. We used the single-point-analysis-pipeline version 0.0.2 (dev branch)
246 [<https://github.com/hmgu-itg/single-point-analysis-pipeline/tree/dev>] to perform the
247 association analysis. GCTA version 1.93.2 beta was used to conduct a mixed linear model
248 association (MLMA) analysis, genetic relationship matrix (GRM) function within the GCTA
249 software was used to estimate the genetic relationships among the individuals. We then used
250 GCTA-COJO, which is designed for approximate conditional and joint stepwise model
251 selection to identify independent associated variants at each locus.

252 **Significance threshold**

253 Cis significant threshold was determined by first multiplying the Bayes Factors (BF) by 2875,
254 and values over 1 were capped at 1. The BF was estimated using eigenMT⁴⁹. eigenMT
255 calculates M_{eff} as the number of ranked eigenvalues from the adjusted genotype correlation
256 matrix needed to account for 99% of the detected genotype variability. Subsequently, the
257 corrected p values were adjusted for multiple testing by applying the False Discovery Rate
258 (FDR) method. Q-values were then calculated using the qvalue package, allowing for the
259 identification of a subset of significant associations based on a q-value threshold of <0.05 .
260 Finally, the cis threshold for significance in pQTL analysis was determined by averaging the
261 smallest non-significant p-value and the largest significant p-value. This method resulted in a
262 cis P -value threshold of $1.462E^{-06}$. Trans threshold was calculated based on the effective
263 number of variants (N_{eff}) and number of protein traits (M_{eff}). The N_{eff} was derived by performing
264 LD-pruning with the following parameters indep 500 5 0.2 in Plink 1.9⁵⁰. This resulted in
265 an N_{eff} of 452593 unique variants. The M_{eff} was calculated using the M_{eff} function and Gao
266 method in poolr R package⁵¹. The trans P -value threshold is $2.227E^{-10}$. Variants within 1
267 megabase (Mb) upstream or downstream of the encoding genes are referred to as cis-pQTLs
268 while trans-pQTLs are those found beyond 1mb relative to the encoding gene. Ensembl's
269 Variant Effect Predictor (VEP) was used to determine the functional impact of the variants.

270 **Comparison of pQTLs to prior published data**

271
272
273 To determine the novelty of our pQTL, we used an in house-built database of previously
274 identified signals of 46 genome-wide pQTL studies including the UKB-PPP¹⁵. We evaluated
275

276 novelty by identifying novel loci and novel variants. Novel loci were defined as those with no
277 published variants within $\pm 1\text{Mb}$ of our variants. For variants at known loci, we checked their
278 rsIDs against those previously reported. Variants with no prior matches were further
279 conditioned in the context of other known variants at that locus. These were classified as novel
280 if the significance of their association p-value persisted even after adjusting for other known
281 variants.

282

283 **Colocalisation analysis**

284 We perform Bayesian-based colocalisation analysis using the Coloc.fast function
285 (<https://github.com/tobyjohnson/gtx>) between our pQTL signals and multi-ancestry T2D
286 GWAS summary statistics³³ from DIAGRAM database. To assume shared genetics, we used
287 default priors and posterior probability of $(\text{PP.H4}) \geq 0.8$ ⁵².

288

289 **Mendelian Randomization**

290 We used the TwoSampleMR package⁵³ to undertake an MR analysis between the proteins and
291 African ancestry T2D to further determine causality and the direction of effect. Wald ratio test
292 was conducted for proteins with a single cis-pQTL variant and IVW was used as the main
293 analysis in instances with more than 1 variants.

294 **Identification of Effector Genes**

295 To find putative effector genes for T2D, we compiled effector genes associated with T2D
296 GWAS. This dataset was curated from nine papers published in the Type 2 Diabetes
297 Knowledge Portal (T2DKP), resulting in a collection of 1,804 distinct effector genes. For
298 classification purposes, proteins that were documented in our curated list were labelled
299 "reported." Those not found on the list were classified as "unresolved."

300

301 **Ethics**

302 The study was approved by the Uganda Virus Research Institute Research and Ethics
303 Committee (UVRI REC #GC/127/907) and the Uganda National Council for Science and
304 Technology (UNCST HS2527ES).

305

306 **Code Availability**

307 Analyses were performed using publicly available software

308 **Acknowledgement**

309 We acknowledge the support of the Core Facility-Metabolomics and Proteomics at Helmholtz
310 Munich. We thank the core facility for help with sample preparation and protein measurement.
311 We thank all participants who contributed to the Uganda Genome Resource. UGR/GPC was
312 supported by the UK Medical Research Council (MRC) and the UK Department for
313 International Development (DFID) under the MRC/DFID Concordat agreement, through core
314 funding to the MRC/UVRI and LSHTM Uganda Research Unit. The 2023 Award Fellowship
315 support of the Alexander Von Humboldt Stiftung Foundation to OS is acknowledged. SF was
316 supported by the Wellcome Trust grant number 220740/Z/20/Z.

317

318

319

320 Reference

- 321 1. Tremblay, J. & Hamet, P. Environmental and genetic contributions to diabetes.
322 *Metabolism* **100**, 153952 (2019).
- 323 2. Tekola-Ayele, F., Adeyemo, A. A. & Rotimi, C. N. Genetic Epidemiology of Type 2
324 Diabetes and Cardiovascular Diseases in Africa. *Prog. Cardiovasc. Dis.* **56**, 251–260
325 (2013).
- 326 3. Motala, A. A., Mbanya, J. C., Ramaiya, K., Pirie, F. J. & Ekoru, K. Type 2 diabetes
327 mellitus in sub-Saharan Africa: challenges and opportunities. *Nat. Rev. Endocrinol.* **18**,
328 219–229 (2022).
- 329 4. Issaka, A. *et al.* Association between urban-rural location and prevalence of type 2
330 diabetes and impaired fasting glucose in West Africa: a cross-sectional population-
331 based epidemiological study. *BMJ Open* **13**, e063318 (2023).
- 332 5. Doherty, M. L., Owusu-Dabo, E., Kantanka, O. S., Brawer, R. O. & Plumb, J. D. Type
333 2 diabetes in a rapidly urbanizing region of Ghana, West Africa: a qualitative study of
334 dietary preferences, knowledge and practices. *BMC Public Health* **14**, 1069 (2014).
- 335 6. Sidahmed, S., Geyer, S. & Beller, J. Socioeconomic inequalities in diabetes
336 prevalence: the case of South Africa between 2003 and 2016. *BMC Public Health* **23**,
337 324 (2023).
- 338 7. Saeedi, P. *et al.* Global and regional diabetes prevalence estimates for 2019 and
339 projections for 2030 and 2045: Results from the International Diabetes Federation
340 Diabetes Atlas, 9(th) edition. *Diabetes Res. Clin. Pract.* **157**, 107843 (2019).
- 341 8. Yazdanpanah, S. *et al.* Evaluation of glycated albumin (GA) and GA/HbA1c ratio for

- 342 diagnosis of diabetes and glycemic control: A comprehensive review. *Crit. Rev. Clin.*
343 *Lab. Sci.* **54**, 219–232 (2017).
- 344 9. Weykamp, C. HbA1c: a review of analytical and clinical aspects. *Ann. Lab. Med.* **33**,
345 393–400 (2013).
- 346 10. Day, A. HbA1c and diagnosis of diabetes. The test has finally come of age. *Annals of*
347 *clinical biochemistry* vol. 49 7–8 (2012).
- 348 11. Cohen, M. P. & Hud, E. Measurement of plasma glycoalbumin levels with a
349 monoclonal antibody based ELISA. *J. Immunol. Methods* **122**, 279–283 (1989).
- 350 12. Png, G. *et al.* Identifying causal serum protein–cardiometabolic trait relationships
351 using whole genome sequencing. *Hum. Mol. Genet.* **32**, 1266–1275 (2023).
- 352 13. Dhindsa, R. S. *et al.* Rare variant associations with plasma protein levels in the UK
353 Biobank. *Nature* **622**, 339–347 (2023).
- 354 14. Zhao, J. H. *et al.* Genetics of circulating inflammatory proteins identifies drivers of
355 immune-mediated disease risk and therapeutic targets. *Nat. Immunol.* **24**, 1540–1551
356 (2023).
- 357 15. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK
358 Biobank. *Nature* **622**, 329–338 (2023).
- 359 16. Gilly, A. *et al.* Genome-wide meta-analysis of 92 cardiometabolic protein serum
360 levels. *Mol. Metab.* **78**, 101810 (2023).
- 361 17. Zhang, J. *et al.* Plasma proteome analyses in individuals of European and African
362 ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat.*
363 *Genet.* **54**, 593–602 (2022).
- 364 18. Karakasilioti, I. *et al.* DNA damage triggers a chronic autoinflammatory response,
365 leading to fat depletion in NER progeria. *Cell Metab.* **18**, 403–415 (2013).
- 366 19. Kennedy, B. P. & Ramachandran, C. Protein tyrosine phosphatase-1B in diabetes.
367 *Biochem. Pharmacol.* **60**, 877–883 (2000).
- 368 20. Hendriks, W. J. A. J. & Pulido, R. Protein tyrosine phosphatase variants in human
369 hereditary disorders and disease susceptibilities. *Biochim. Biophys. Acta - Mol. Basis*
370 *Dis.* **1832**, 1673–1696 (2013).
- 371 21. Yoon, S.-Y., Ahn, D., Hwang, J. Y., Kang, M. J. & Chung, S. J. Linoleic acid exerts
372 antidiabetic effects by inhibiting protein tyrosine phosphatases associated with insulin
373 resistance. *J. Funct. Foods* **83**, 104532 (2021).
- 374 22. Chung, S. J. Dual Inhibition of PTPN6 and PTPN9. 1–13 (2024).
- 375 23. Yu, Y. *et al.* Bioinformatics analysis of candidate genes and potential therapeutic

- 376 drugs targeting adipose tissue in obesity. *Adipocyte* **11**, 1–10 (2022).
- 377 24. Elbein, S. C. *et al.* Global Gene Expression Profiles of Subcutaneous Adipose and
378 Muscle From Glucose-Tolerant, Insulin-Sensitive, and Insulin-Resistant Individuals
379 Matched for BMI. *Diabetes* **60**, 1019–1029 (2011).
- 380 25. Tabasum, S. *et al.* EDIL3 as an Angiogenic Target of Immune Exclusion Following
381 Checkpoint Blockade. *Cancer Immunol. Res.* **11**, 1493–1507 (2023).
- 382 26. Gasca, J. *et al.* EDIL3 promotes epithelial–mesenchymal transition and paclitaxel
383 resistance through its interaction with integrin $\alpha V\beta 3$ in cancer cells. *Cell Death*
384 *Discov.* **6**, 86 (2020).
- 385 27. Shen, W. *et al.* EDIL3 knockdown inhibits retinal angiogenesis through the induction
386 of cell cycle arrest in vitro . *Mol Med Rep* **16**, 4054–4060 (2017).
- 387 28. Yu, C.-G. *et al.* Endothelial Progenitor Cells in Diabetic Microvascular Complications:
388 Friends or Foes? *Stem Cells Int.* **2016**, 1803989 (2016).
- 389 29. Tahergorabi, Z. & Khazaei, M. Imbalance of angiogenesis in diabetic complications:
390 the mechanisms. *Int. J. Prev. Med.* **3**, 827–838 (2012).
- 391 30. Kumar, M. A. *et al.* Extracellular vesicles as tools and targets in therapy for diseases.
392 *Signal Transduct. Target. Ther.* **9**, 27 (2024).
- 393 31. Abate, W., Alrammah, H., Kiernan, M., Tonks, A. J. & Jackson, S. K.
394 Lysophosphatidylcholine acyltransferase 2 (LPCAT2) co-localises with TLR4 and
395 regulates macrophage inflammatory gene expression in response to LPS. *Sci. Rep.* **10**,
396 10355 (2020).
- 397 32. Wang, B. & Tontonoz, P. Phospholipid Remodeling in Physiology and Disease. *Annu.*
398 *Rev. Physiol.* **81**, 165–188 (2019).
- 399 33. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology.
400 *Nature* **627**, 347–357 (2024).
- 401 34. Barrett, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40
402 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
- 403 35. Wong, A. S. L. *et al.* Polymorphism in the Innate Immune Receptor SIRP α Controls
404 CD47 Binding and Autoimmunity in the Nonobese Diabetic Mouse. *J. Immunol.* **193**,
405 4833–4844 (2014).
- 406 36. Fanelli, F. *et al.* Profiling plasma N-Acylethanolamine levels and their ratios as a
407 biomarker of obesity and dysmetabolism. *Mol. Metab.* **14**, 82–94 (2018).
- 408 37. Rowles, J. E. *et al.* Are Heat Shock Proteins an Important Link between Type 2
409 Diabetes and Alzheimer Disease? *Int. J. Mol. Sci.* **21**, (2020).

- 410 38. Chen, S.-T. *et al.* CLEC5A is critical for dengue-virus-induced lethal disease. *Nature*
411 **453**, 672–676 (2008).
- 412 39. Maurin, L. *et al.* PNLIPRP1 hypermethylation in exocrine pancreas links type 2
413 diabetes and cholesterol metabolism. *Diabetes* (2024) doi:10.2337/db24-0215.
- 414 40. Breeyear, J. H. *et al.* Adaptive selection at G6PD and disparities in diabetes
415 complications. *Nat. Med.* (2024) doi:10.1038/s41591-024-03089-1.
- 416 41. Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type
417 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-
418 wide meta-analysis. *PLOS Med.* **14**, e1002383 (2017).
- 419 42. Gurdasani, D. *et al.* Uganda Genome Resource Enables Insights into Population
420 History and Genomic Discovery in Africa. *Cell* **179**, 984-1002.e36 (2019).
- 421 43. Fatumo, S. *et al.* Uganda Genome Resource: A rich research database for genomic
422 studies of communicable and non-communicable diseases in Africa. *Cell genomics* **2**,
423 None (2022).
- 424 44. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-
425 sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- 426 45. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
427 Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
- 428 46. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing
429 biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
- 430 47. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association
431 networks and functional enrichment analyses for any sequenced genome of interest.
432 *Nucleic Acids Res.* **51**, D638–D646 (2023).
- 433 48. Petrera, A. *et al.* Multiplatform Approach for Plasma Proteomics: Complementarity of
434 Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein
435 Profiling. *J. Proteome Res.* **20**, 751–762 (2021).
- 436 49. Davis, J. R. *et al.* An Efficient Multiple-Testing Adjustment for eQTL Studies that
437 Accounts for Linkage Disequilibrium between Variants. *Am. J. Hum. Genet.* **98**, 216–
438 224 (2016).
- 439 50. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and
440 richer datasets. *Gigascience* **4**, 7 (2015).
- 441 51. Cinar, O. & Viechtbauer, W. The poolr Package for Combining Independent and
442 Dependent p Values. *J. Stat. Softw.* **101**, 1–42 (2022).
- 443 52. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic

- 444 Association Studies Using Summary Statistics. *PLOS Genet.* **10**, e1004383 (2014).
- 445 53. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across
446 the human phenome. *Elife* **7**, (2018).
- 447
- 448
- 449