

1 Genetic regulation of the plasma proteome and its link to 2 cardiometabolic disease in Greenlandic Inuit

3

4 Authors

5 Sara E. Stinson^{1*}, Renzo F. Balboa^{2*}, Mette K. Andersen¹, Frederik F. Stæger², Shixu He², Anne
6 Cathrine Baun Thuesen¹, Long Lin², Peter Bjerregaard^{3,4}, Christina V.L. Larsen^{3,4}, Niels
7 Grarup¹, Marit E. Jørgensen^{3,4,5}, Ida Moltke^{2,#}, Anders Albrechtsen^{2,#} and Torben Hansen^{1,#}

8

9 *These authors contributed equally

10 #Corresponding authors

11

12 Affiliations

13 ¹Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical
14 Sciences, University of Copenhagen, Copenhagen, Denmark

15 ²Department of Biology, University of Copenhagen, Copenhagen, Denmark

16 ³National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark

17 ⁴Greenland Center for Health Research, University of Greenland, Nuuk, Greenland

18 ⁵Steno Diabetes Center, Nuuk, Greenland

19

20 Corresponding authors

21 Addresses: Ole Maaløes Vej 5, DK-2200, Copenhagen N, Denmark (I. Moltke and A.
22 Albrechtsen); Blegdamsvej 3B, DK-2200, Copenhagen N, Denmark (T. Hansen).

23 Emails: ida@bio.ku.dk (I. Moltke), aalbrechtsen@bio.ku.dk (A. Albrechtsen),
24 torben.hansen@sund.ku.dk (T. Hansen)

25

26 Keywords

27 Proteomics, Genomics, Diverse populations, Greenland Inuit, Cardiovascular disease, Type 2
28 diabetes, pQTL analysis

29 **Summary**

30 **Background**

31 Circulating proteins play essential roles in numerous complex diseases, yet our understanding of
32 the genetic influences on inflammation and cardiovascular-related proteins in non-European
33 isolated populations remains sparse. Here, we aimed to characterise the genetic architecture of
34 plasma protein biomarkers in the Greenlandic population.

36 **Methods**

37 Using combined data from Greenlandic population cohorts (n=3,707 individuals), including
38 genotypes and plasma proteomics (177 proteins) from Olink Target 96 Inflammation and
39 Cardiovascular II panels, we performed a protein quantitative trait loci (pQTL) study using a
40 linear mixed model, accounting for relatedness and population structure.

42 **Findings**

43 Mapping of 177 plasma proteins in 3,707 adult Greenlandic individuals (mean age 47.9; 54.5%
44 female) reveal 251 primary pQTLs, 235 additive (92 *cis* and 143 *trans*) and 16 recessive (1 *cis*
45 and 15 *trans*), 48 secondary pQTLs, and 70 novel pQTLs (28%). We demonstrate a higher
46 proportion of variance in protein levels explained in Greenlanders compared to Europeans from
47 the UK Biobank (e.g. IL-27, IgGFCRII-b, IL-16, and Gal-9). We describe changes in expression
48 of inflammation and cardiovascular-related proteins associated with known high impact Arctic-
49 specific variants, including in *CPT1A*, *TBC1D4*, *HNF1A*, *LDLR*, and *PCSK9*.

51 **Interpretation**

52 These findings highlight the importance of genome-wide plasma proteomic analyses in
53 Greenlanders, and diverse populations in general, with implications for biomarker and
54 therapeutic target development.

56 **Funding**

57 Novo Nordisk Foundation, The Independent Research Fund Denmark, and Karen Elise Jensen
58 Foundation.

59 **Research in context**

60 **Evidence before this study**

61 Recent affinity-based proteomic studies have been performed in large European biobank-scale
62 cohorts such as the UK Biobank and deCODE. Several smaller-scale studies have also been
63 performed in isolated European populations, e.g. MANOLIS and Pomak (Hellenic), Orkney
64 (Scotland), and Vis (Croatia). Studies in non-European populations are also beginning to emerge,
65 including in the China Kadoorie Biobank. Studies performed in diverse populations can identify
66 population-specific variants in genes implicated in regulating the expression of proteins, which
67 may be causally linked to inflammation and cardiovascular disease. In particular, small and
68 historically isolated populations, such as the Greenlandic population, are more likely to harbour
69 common variants with larger effect sizes that may contribute to health and disease.

70

71 **Added value of this study**

72 This study reports 251 primary protein quantitative trait loci (pQTLs) associated with the
73 abundance of 177 plasma proteins, 28% of which have not been previously reported. We
74 identified 48 additional pQTLs in a secondary conditional analysis. We identified novel pQTLs
75 that were common in Greenland, but rare globally (e.g. *ST1A1*, *DCN*). We found common
76 pQTLs which explained a substantial proportion of variance (>30%) in protein abundance (e.g.
77 *IL-27*, *IgGFCRIIb*, *IL-16*, *Gal-9*) when compared to Europeans. We examined differences in
78 protein abundance in carriers of Arctic-specific variants (e.g. *CPT1A*, *TBC1D4*, *HNF1A*, *LDLR*,
79 *PCSK9*) which are implicated in lipid metabolism and cardiometabolic disease, revealing
80 underlying biological mechanisms.

81

82 **Implications of all the available evidence**

83 Given that both genetics and the environment affect protein levels causally linked to disease, it is
84 crucial to perform genome-wide association studies in smaller populations of diverse genetic
85 ancestry to ensure equity in genetic discovery. Investigating the effect of previously identified
86 Arctic-specific variants on protein expression revealed links to therapeutic targets for metabolic
87 disease, which may have implications for the health care system in Greenland and beyond,
88 including access to treatment.

89 **Introduction**

90 Circulating proteins measured in the blood represent intermediate phenotypes for disease states
91 and are strongly linked to both causative genetic variants and disease outcomes¹. Thus,
92 differences in protein expression make ideal candidates to study, in order to understand the
93 molecular mechanisms of complex disease.

94 Proteomics data has only recently become accessible to large-scale genome-wide association
95 studies (GWAS), largely due to analytical limitations specific to the composition of the plasma
96 proteome². Advances in high-throughput proteomics methods (e.g. Olink, SomaScan, and mass-
97 spectrometry approaches) now enable the measurement of plasma levels of hundreds to
98 thousands of proteins concurrently^{3,4}. Unlike mass-spectrometry-based proteomics, the Olink
99 Proximity Extension Assay captures less abundant proteins, allowing for the quantification of
100 cytokines and chemokines relevant to the pathophysiology of cardiometabolic disease⁵.

101 Identification of genetic variants associated with altered plasma protein levels (protein
102 quantitative trait loci, pQTLs) can enhance our understanding of the role of proteins in health and
103 disease⁶. Discovery of genetic determinants of protein levels help elucidate the causal role of a
104 given protein in disease risk, which may allow for improved risk stratification to enhance early
105 diagnosis and identification of potential drug targets for treatment or prevention of such
106 diseases⁷.

107 To date, the vast majority of studies have restricted analyses of the genetic architecture of the
108 plasma proteome to European ancestries⁸⁻¹³, this includes large-scale biobanks such as the UK
109 Biobank Pharma Proteomics Project (UKB-PPP, Olink)¹⁴ and deCODE Genetics (SomaScan).
110 For example, the discovery analysis in UKB-PPP was limited to European individuals, with
111 replication in a limited number of non-European individuals¹⁴. Although separate efforts have
112 been made to analyse pQTLs in non-EUR, including East Asian^{15,16} and Middle Eastern¹⁷
113 ancestries, diverse populations are in comparison, highly underrepresented.

114 Two studies have been performed in isolated European populations, including MANOLIS,
115 Pomark, and ORCADES in combination with whole-genome sequencing^{18,19}. Studying isolated
116 populations has the power to enhance the discovery of disease-associated loci due to the
117 enrichment of certain rare variants with larger effects^{20,21}.

118 The Greenlandic population, comprising ~57,000 individuals, mainly has Inuit ancestry, but due
119 to recent admixture it also has some European ancestry²². The population is a small and
120 historically isolated founder population, which has experienced an increase in the prevalence of
121 type 2 diabetes (T2D) over the last 25 years²³. Previous studies of the population have revealed
122 that complex diseases, including T2D, often manifest differently and have Arctic-specific causes,
123 when compared to other global populations. How lipid-related and cardiometabolic disease-
124 causing variants are associated with differences in protein expression within the Greenlandic
125 population, and its comparison with other populations, have yet to be studied. An investigation
126 into these relationships will allow for a better understanding of the mechanism of disease within
127 this population, and other diverse populations in general.

128 Here, we examined the genetic architecture of the targeted plasma proteome in 3,707
129 Greenlanders. The aims of this study were to: (i) perform GWAS of 177 plasma proteins
130 measured by the Olink Target 96 Inflammation and Cardiovascular II panels; (ii) compare the
131 genetic architecture of the targeted proteome between Greenlanders and European populations
132 through comparisons with the UKB-PPP, and (iii) assess the impact of known high impact
133 Arctic-specific variants on circulating protein levels.

134

135 **Methods**

136 **Ethics statement**

137 The study has received ethical approval from the Science Ethics Committee in Greenland
138 (project 505-42, project 505-95, project 2011–13 [ref. no. 2011–056978], project 2017-5582,
139 project 2015–22 [ref. no. 2015-16426], and project 2021-09 [16669546]). The study was
140 conducted in accordance with the Declaration of Helsinki, second revision. All participants gave
141 written informed consent.

142

143 **Study populations**

144 A total of 3,707 participants with available proteomics, genomics, and phenotype data, were
145 included in this study; the participants were pooled from three population-survey cohorts: the
146 Greenland Population Study (B99) (n = 15; 1998–2001)²⁴, the Inuit Health in Transition (IHIT)
147 cohort (n = 2,130; 2005–2010)²⁵, and the B2018 cohort (n = 1,587; 2017-2019)²⁶.

148

149 **Plasma proteomics**

150 Relative levels of 184 plasma proteins were measured in 3,707 participants using the Olink
151 Target 96 Inflammation and Cardiovascular II panels across two batches. Seven proteins were
152 common to both panels, yielding a total of 177 unique proteins that were measured. The Olink
153 batches were bridged and normalised using 16 controls using the *OlinkAnalyze* R package
154 (<https://cran.r-project.org/web/packages/OlinkAnalyze/index.html>). Normalised protein
155 expression values (NPX) on a \log_2 scale were used for downstream analyses. Protein values
156 below the lower limit of detection were included in analyses, while those failing Olink's internal
157 quality control warnings were excluded.

158

159 **Genotyping and imputation**

160 All samples were genotyped using the Multi-Ethnic Genotyping Array (MEGA chip, Illumina),
161 composed of ~2M variants across the human genome. This chip was used due to the inclusion of
162 variants selected from Indigenous Americans and East Asians, maximising variant capture in
163 Inuit. Genotypes were jointly called using GenCall within GenomeStudio (Illumina). Duplicate
164 samples, individuals with >0.05 missing genotypes, and variants with >0.01 missing genotypes
165 were excluded from analyses. After quality control, our variant set comprised 1.6M variants used
166 for the present study.

167 The same reference panel described in Senftleber et al.²⁷ was utilised in this study. Briefly, whole
168 genome sequencing data from 447 Greenlanders were genotyped using GATK, and phased using
169 SHAPEIT2²⁸, using available duo/trio information. These data were then combined with 1000
170 Genomes²⁹ data and utilised as a reference for imputing common (MAF >0.5%) variants using
171 IMPUTE2³⁰. The imputed data was used to estimate a genetic relationship matrix (GRM) and
172 admixture proportions for inclusion in genome-wide association analyses, described below.

173 **Genome-wide association analyses**

174 For association analyses, a linear mixed model (LMM) was used, accounting for relatedness and
175 population structure (GEMMA³¹ version 0.98.5). The GRM was estimated from SNPs with <1%
176 missingness, a MAF >5%, info-score >0.95. PCAngsd³² was used to calculate per-site inbreeding
177 coefficient (site-F) accounting for population structure. Sites with site-F > |0.05| and likelihood-

178 ratio test P-value $<1 \times 10^{-6}$ were also excluded from the GRM estimation. Association tests
179 assuming additive and recessive models were performed using a score test. Effect sizes and
180 standard errors were estimated using a restricted maximum likelihood approach. Normalised
181 protein expression (NPX) values were quantile transformed to a normal distribution using a rank-
182 based inverse normal transformation performed separately for each sex. *P*-values are from
183 analyses of transformed data and effect size estimates are from the same analyses and thus
184 reported in standard deviations as β_{SD} . In all analyses, sex, age, and cohort were included as
185 covariates.

186

187 **Definition and refinement of significant loci**

188 Primary pQTLs were first defined by distance-based clumping of ± 10 Mb, reporting one
189 representative SNP-protein association with the lowest P-value. For the proteins that are present
190 in both the Cardiovascular II and Inflammation panels, we took pQTLs with the lowest P-value
191 in either panel to avoid duplication. We reported the total number of identified significant
192 associations using the conventional genome-wide threshold of 5×10^{-8} .

193

194 Some SNP-protein pairs were genome-wide significant when tested in both additive and
195 recessive models. To avoid double counting and to retain independence, if the same SNP-protein
196 pair was genome-wide significant in both additive and recessive models, we reported the lowest
197 p-value from the two, defined as ‘mostly additive’ if $P_{add} < P_{rec}$ and ‘mostly recessive’ if $P_{rec} <$
198 P_{add} . SNPs within 1Mb significant for the same phenotype were also merged to the SNP with the
199 lowest p-value, even if SNPs were detected with different models.

200

201 To test whether a SNP exhibited a ‘truly’ additive/recessive mode of inheritance, we tested
202 whether the full model with full flexibility to the mode of inheritance (i.e. the alternative mode of
203 inheritance is added as a covariate in the LMM) was significantly better than when using either
204 an additive or recessive model, respectively. We subsequently defined the following categories
205 for this test using an alpha threshold of 0.05: a) recessive/reject additive - where the full model is
206 significantly better than the additive model, and the full model is not significantly better than the
207 recessive model; b) additive/reject recessive - where the full model is not significantly better
208 than the additive model, and the full model is significantly better than the recessive model; c) co-

209 dominant/reject both - where the full model is significantly better than both the additive and
210 recessive models, respectively.

211
212 To perform *cis/trans* classifications, we extracted the start/end information for all proteins by
213 mapping UniProt IDs or gene names to Ensembl BioMart. Using these annotations, pQTLs were
214 subsequently classified into four categories; for each protein, a *cis*-pQTL was defined when a
215 lead SNP resided within 100 Kb upstream or downstream (± 100 Kb) of either the start or end of
216 the gene encoding the associated protein, taking whichever was closer to the SNP; a *semi-cis*-
217 pQTL was defined when a SNP resided within ± 100 kb-5Mb of the gene encoding the associated
218 protein; a *semi-trans*-pQTL was defined when a SNP resided within the same chromosome but
219 >5 Mb away; and a *trans*-pQTL was defined when a SNP was located in a different chromosome
220 with respect to the corresponding protein-coding gene.

221
222 **Secondary associations and conditional analyses**
223 To test for secondary signals, we performed conditional analyses where primary SNPs were used
224 as covariates as a factor within both additive and recessive models. SNPs that were genome-wide
225 significant when conditioning with any primary SNP within 10 Mb, were classified as a
226 secondary association.

227 Similar to how we reported primary associations, secondary pQTL were classified as the lowest
228 p-value from either the additive or recessive models, defined as ‘mostly additive’ if $P_{\text{add}} < P_{\text{rec}}$ and
229 ‘mostly recessive’ if $P_{\text{rec}} < P_{\text{add}}$. These secondary associations were then reported with their
230 primary SNP.

231
232 **Proportion of variance explained by pQTLs**
233 To ensure a fairer comparison when investigating effect sizes and proportion of variance
234 explained (PVE) by pQTLs, we performed the same GWAS analysis described above using a
235 random subset of 3,707 individuals from the UKB-PPP publication¹⁴. Of the 177 unique proteins
236 assayed within this study, 176 were matched using UniProtIDs (data for VSIG2 was not
237 available in the dataset described by Sun et al.¹⁴).

238 The following equation was used to estimate PVEs by pQTLs from GWAS summary statistics
239 for each protein:

$$\hat{\beta}^2 = \frac{\beta^2}{\beta^2 + \text{SE}^2 \cdot N}$$

240 Where β represents the effect size, SE represents the standard error of the effect size, and N
241 represents the number of individuals assayed.

242 Comparisons were then made between our Greenlandic dataset (n=3,707), the UKB-PPP subset
243 (n=3,707), and the UKB-PPP summary dataset (n=54,219; described in Supplementary Table 10
244 in Sun et al.¹⁴).

245

246 **Identification of novel pQTLs**

247 To assess the novelty of the identified pQTLs in this study, we compared our results at genome-
248 wide significance level to the results of 34 previously published pQTL studies, including the
249 most recent UKB-PPP publication¹⁴. For all studies, we retained the pQTLs at the reported
250 significance levels. We defined the novelty of locus-protein pairs if no variants residing within \pm
251 10Mb of the primary pQTLs in this study had been reported in previous plasma/serum pQTL
252 studies for the corresponding primary protein as in Thareja et al.¹⁷; described in Supplementary
253 Table 4. Comparisons were done at protein level by matching the reported gene name from each
254 study. Gene names were mapped based on Uniprot IDs, through Uniprot ID mapping in case of
255 missing values.

256 In some cases, locus-protein pairs have been previously reported, yet the putative variant within
257 this study could be novel. To identify such variants, we plotted SNP-protein effects from UK
258 Biobank and the Greenlandic study together in Supplementary Figure 1 and for potentially novel
259 pairs in Supplementary Figure 2.

260

261 **Analysis of known high-impact cardiometabolic-related variants**

262 Summary data describing the associations between known high-impact Arctic-specific variants
263 in *CPT1A* (rs80356779), *TBC1D4* (rs61736969), *HNF1A* (rs2135845768), *LDLR* (rs730882082),

264 and *PCSK9* (rs4609471) on the abundance of 177 Olink proteins measured in this study were
265 examined. False discovery rate (FDR) was calculated to account for multiple testing.

266

267 **Role of funding source**

268 None of the funding agencies had any role in the study design or collection, analysis, or
269 interpretation of the data.

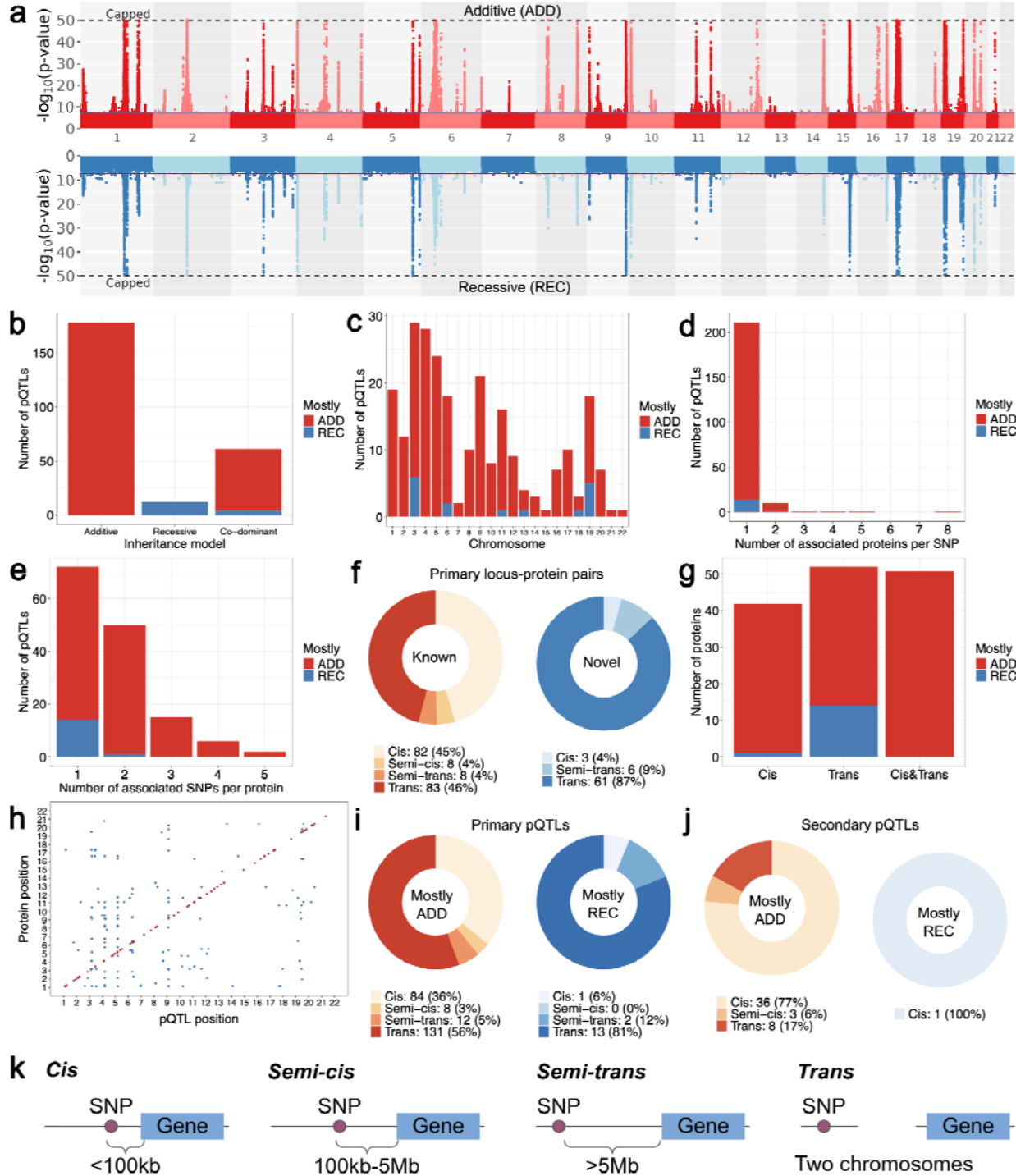
270

271 **Results**

272 **Genetic architecture of circulating plasma proteome**

273 We analysed 177 proteins measured using the Olink Target Inflammation and Cardiovascular II
274 panels (Supplementary Table 1) in three Greenlandic cohorts, totaling 3,707 participants from 19
275 different localities in Greenland (Supplementary Table 2). To identify pQTLs we used both an
276 additive and a recessive model in a linear mixed model framework, which resulted in low
277 genomic inflation for both the additive (mean $\lambda = 0.985 \pm 0.0246$) and the recessive ($\lambda = 0.989 \pm$
278 0.0250) model (Supplementary Figure 3). We identified a total of 251 genome-wide significant
279 ($P < 5 \times 10^{-8}$) primary associations between 225 genomic regions and 133 proteins (Figure 1,
280 Supplementary Table 3), where primary associations means that all SNPs associated with the
281 same protein are more than 10Mb apart. Of these primary pQTLs, 235 were classified as mostly
282 additive and 16 as mostly recessive, however, for around one in four we can reject both the
283 additive and the recessive model in favour of the full model (figure 1b). As expected, we found
284 that the majority of primary SNPs were associated with a single protein, yet a large portion of
285 proteins (56.4%) were associated with more than one SNP. In 34 previously published pQTL
286 studies, we found that 181 of our identified locus-protein associations have been previously
287 reported in a peer-reviewed article, and that these are in almost equal proportions in *cis* and *trans*
288 (Figure 1f). Conversely, the vast majority of the remaining 70 (28%) locus-protein pairs were
289 found to be novel (± 10 Mb from the nearest variant previously reported to be associated to the
290 same protein in a peer-reviewed article), in which the vast majority of these associations were in
291 *trans* or semi-*trans* (96%; Figure 1f; Figure 1k, Supplementary Table 4). Interestingly, two of the
292 three novel associations reported in *cis* that did not overlap with any of the examined published
293 studies have lead SNPs that are very rare in European populations, yet were common in the Inuit
294 ancestry component of the Greenlanders (chr16:28596376:T:C for ST1A1 - MAF=0.064 in Inuit

295 and MAF=0.001 in Europeans; chr12:91158410:T:C for DCN - MAF=0.230 in Inuit and
296 MAF<0.001 in Europeans; Supplementary Tables 3 & 4). Of the primary pQTLs, 93 (37%) were
297 local-acting (85 *cis*, 8 *semi-cis*; Figure 1k; Methods) and 158 (63%) were distant-acting (14
298 *semi-trans* and 144 *trans*; Figure 1k; Methods). Of the 177 proteins assayed within this study, 51
299 unique proteins were identified as having both *cis* and *trans* pQTLs (defined as $<\pm 5\text{Mb}$ of the
300 gene encoding the associated protein and $>\pm 5\text{Mb}$ of the gene encoding the associated protein or
301 on a different chromosome, respectively). In addition to the 251 primary associations,
302 conditional analyses further revealed 48 secondary pQTLs, of which, 47 were additive (36 *cis*, 3
303 *semi-cis*; 8 *trans*) and one recessive (in *cis*) (Supplementary Table 5).
304



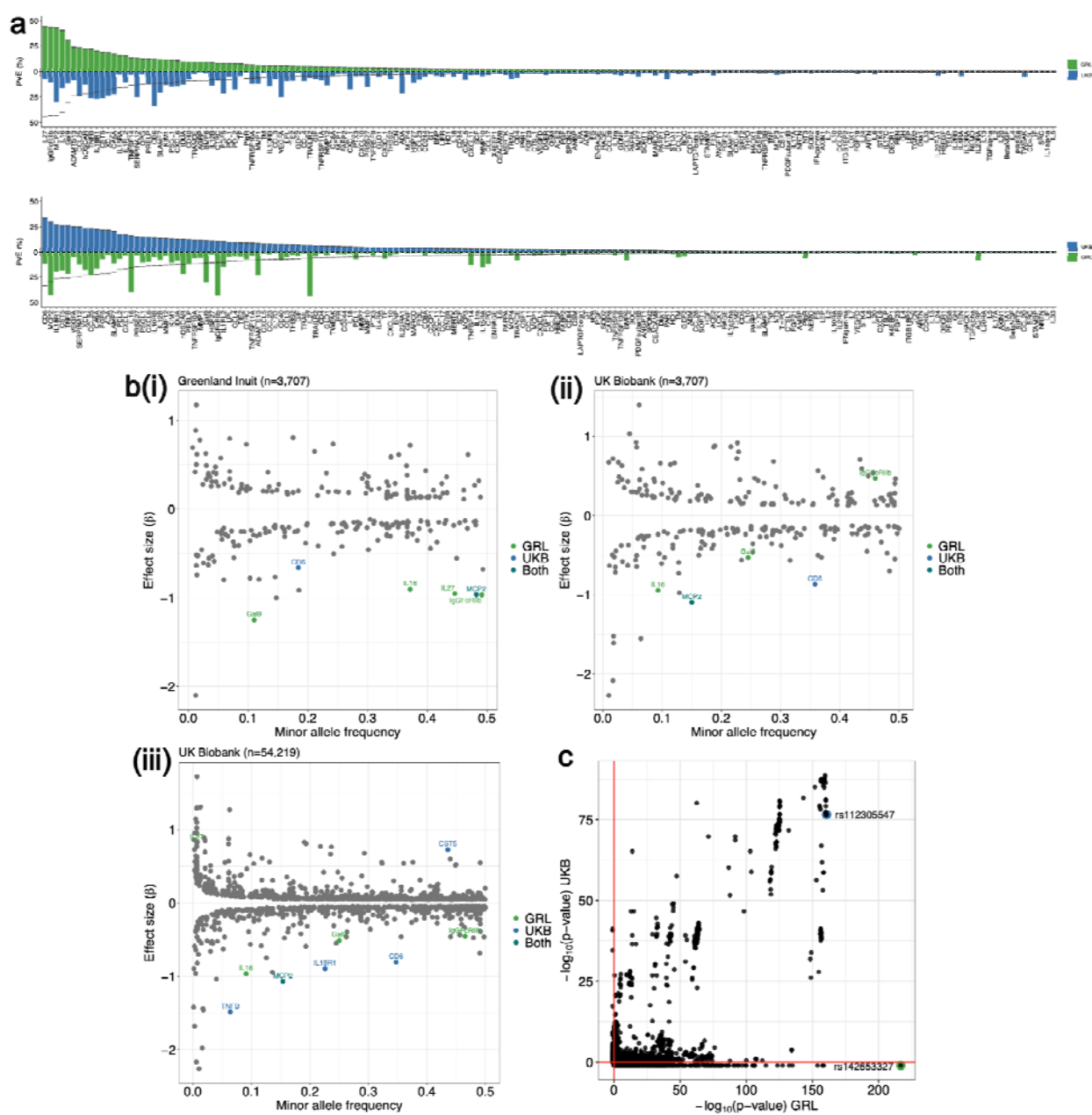
305
306
307

308 **Figure 1. Genetic architecture of the targeted plasma proteome in the Greenlandic population.** **a** Manhattan
309 plot of p-values combined for all proteins under additive (red) and recessive (blue) models obtained using GEMMA.
310 A genome-wide significance threshold of 5×10^{-8} is indicated by horizontal blue and red lines. P-values are capped at
311 1×10^{-50} , as indicated by dotted black lines. **b** Primary pQTLs classified according to inheritance model where
312 ‘additive’ and ‘recessive’ indicate pQTLs that are not rejected in favour of the full model while for ‘co-dominant’
313 both the additive and recessive models are rejected. Bar colours indicate the model in which the lowest p-value was
314 reported, where red indicates a lower p-value when using an additive model (‘mostly additive’) and blue, a lower p-
315 value when using a recessive model (‘mostly recessive’). **c** Number of primary pQTLs detected per chromosome. **d**
316 Number of associated proteins per variant. **e** Number of primary pQTLs per protein. **f** Number of known and novel
317 locus-protein associations in *cis*, *semi-cis*, *semi-trans* and *trans* (defined in k). Known indicates associations within
318 ± 10 Mb of any significant pQTL reported in previous studies (Methods). Novel indicates associations $>\pm 10$ Mb of
319 any significant pQTL reported in previous studies (Methods). **g** Number of proteins with pQTLs in *cis*, *trans* or
320 both. *Trans* pQTLs here are defined as >5 Mb from the start/end positions of the primary gene relevant for each
321 Olink protein (a combination of *trans* and *semi-trans* as defined in k, for simplicity). **h** Dot plot indicating pQTL
322 positions across the genome. *Trans* pQTLs are defined as >5 Mb from the start/end positions of the primary gene
323 relevant for each Olink protein, (a combination of *trans* and *semi-trans* as defined in k, for simplicity). Red - *cis*
324 QTLs; Blue - *trans* pQTLs. **i** Number of primary pQTLs defined as *cis*, *semi-cis*, *semi-trans* and *trans* (defined in k).
325 **j** Number of secondary pQTLs defined as *cis*, *semi-cis*, *semi-trans* and *trans* (defined in k). **k** Definitions of *cis*,
326 *semi-cis*, *semi-trans* and *trans* pQTLs. *Cis* pQTLs are defined as SNP-protein pairs which are within 100kb of the
327 start/end of the gene encoding the associated protein. *Semi-cis* pQTLs are defined as SNP-protein pairs which are
328 between 100-kb-5Mb of the start/end of the gene encoding the associated protein. *Semi-trans* pQTLs are defined as
329 SNP-protein pairs which are >5 Mb of the start/end of the encoding gene, but on the same chromosome. *Trans*
330 pQTLs are defined as SNP-protein pairs which are on different chromosomes from the gene encoding the associated
331 protein.

332 333 **Common, high-impact pQTLs in Greenlanders**

334 We next explored PVE for pQTLs detected in Greenlandic individuals and performed
335 comparisons with Olink data from the UK Biobank¹⁴. The UK Biobank individuals tended to
336 have a larger proportion of significant pQTLs with a lower PVE, and exhibited a lower number
337 of high-impact pQTLs compared to the Greenlandic individuals (Figure 2a-b, Supplementary
338 Table 6). In the Greenlandic individuals, several proteins, including IL-27, IgGFCRII-b, MCP-2
339 and IL-16 and Gal-9 exhibited PVEs $>30\%$ (Figure 2a). pQTLs related to the same proteins in
340 UK Biobank exhibited PVEs $<20\%$ and were generally less common in allele frequency when
341 compared to Greenlanders (Figure 2bi-iii). Two of these variants, rs1801274 associated with
342 IgGFCRII-b, rs142653327 associated with Gal-9, are missense variants. Specifically, the

343 association of IgGfCII-b rs1801274 G-allele has been shown to confer susceptibility to
344 autoimmune diseases, including Kawasaki disease and ulcerative colitis³³. Furthermore,
345 rs142653327 is a novel pQTL variant for Gal-9, which is not present in Europe (Figure 2c).
346 However, a secondary association for Gal-9 found in Greenland was also significant within UK
347 Biobank (rs112305547; Figure 2c). Overall, these results suggest that Greenlanders exhibit more
348 common and high-impact variants when compared to UK Biobank individuals, and that some of
349 these variants may carry a functional impact.
350



351

352

353 **Figure 2. Proportion of variance explained by pQTLs in Greenlanders vs Europeans from UK Biobank. a**
354 Proportion of variance explained (PVE) for Greenlanders (GRL; green; n=3,707) and the UK Biobank cohort¹⁴
355 (UKB; blue; n=3,707) for 176 proteins common between both studies, ordered by highest PVEs in Greenlanders
356 (top) and highest PVEs in UKB (bottom). Dashed lines indicate the equivalent PVE for each protein as reported in
357 Greenlanders (top) and UKB (bottom), respectively. **b** PVE vs minor allele frequency (MAF) for genome-wide
358 significant pQTLs detected in (i) Greenlandic individuals using GEMMA (n=3,707); (ii) a random subset of UKB
359 individuals using GEMMA (n=3,707); and (iii) using summary statistics from Sun et al. (2024) (n=54,219). pQTLs
360 with PVE >30% for either Greenlandic individuals, UKB individuals or both in (i) and (ii) are highlighted in green,
361 blue and turquoise, respectively. PVE >30% for Greenlandic individuals, and PVE >15% for UKB individuals or
362 meeting both criteria in (iii) are highlighted in green, blue and turquoise, respectively. **c** Comparisons of $-\log_{10} P$
363 values for variant-protein associations described by the Greenlanders dataset (n=3,707), and a subset of individuals
364 from the UKB-PPP (n=3,707). Variant-protein associations tested using an additive model between Gal-9 and SNPs
365 within 5Mb of rs142653327 are plotted. Red lines indicate $-\log_{10} P$ values of 0 for both datasets. Points at -1 indicate
366 associations that were unable to be overlapped with the orthogonal dataset. Green circle - indicates the lead QTL in
367 the Greenland cohort (rs142653327); Blue circle - indicates the lead secondary association reported in the Greenland
368 cohort (rs112305547).

369

370 **Associations with known high-impact Arctic-specific variants**

371 Previous studies have identified causal, high-impact variants related to population-specific
372 effects for lipid metabolism, T2D, and CVD within Greenlanders, including *CPT1A*³⁴,
373 *TBC1D4*³⁵, *HNFI1A*³⁶, *LDLR*³⁷, and *PCSK9*²⁷. As such, we investigated the protein expression for
374 these known variants to gain a better understanding of potential mechanisms leading to these trait
375 differences (Figure 3).

376

377 A known variant in *CPT1A* (rs80356779)³⁴, previously found to modulate fatty acid composition,
378 with large effect sizes on weight and height, herein was associated with lower plasma levels of
379 55 proteins at a significance level of FDR<0.05, including eight proteins significant at
380 FDR<0.001. Stem cell factor, also termed kit ligand (SCF/KITLG) and follistatin (FS)
381 represented the most significant proteins associated with this variant.

382

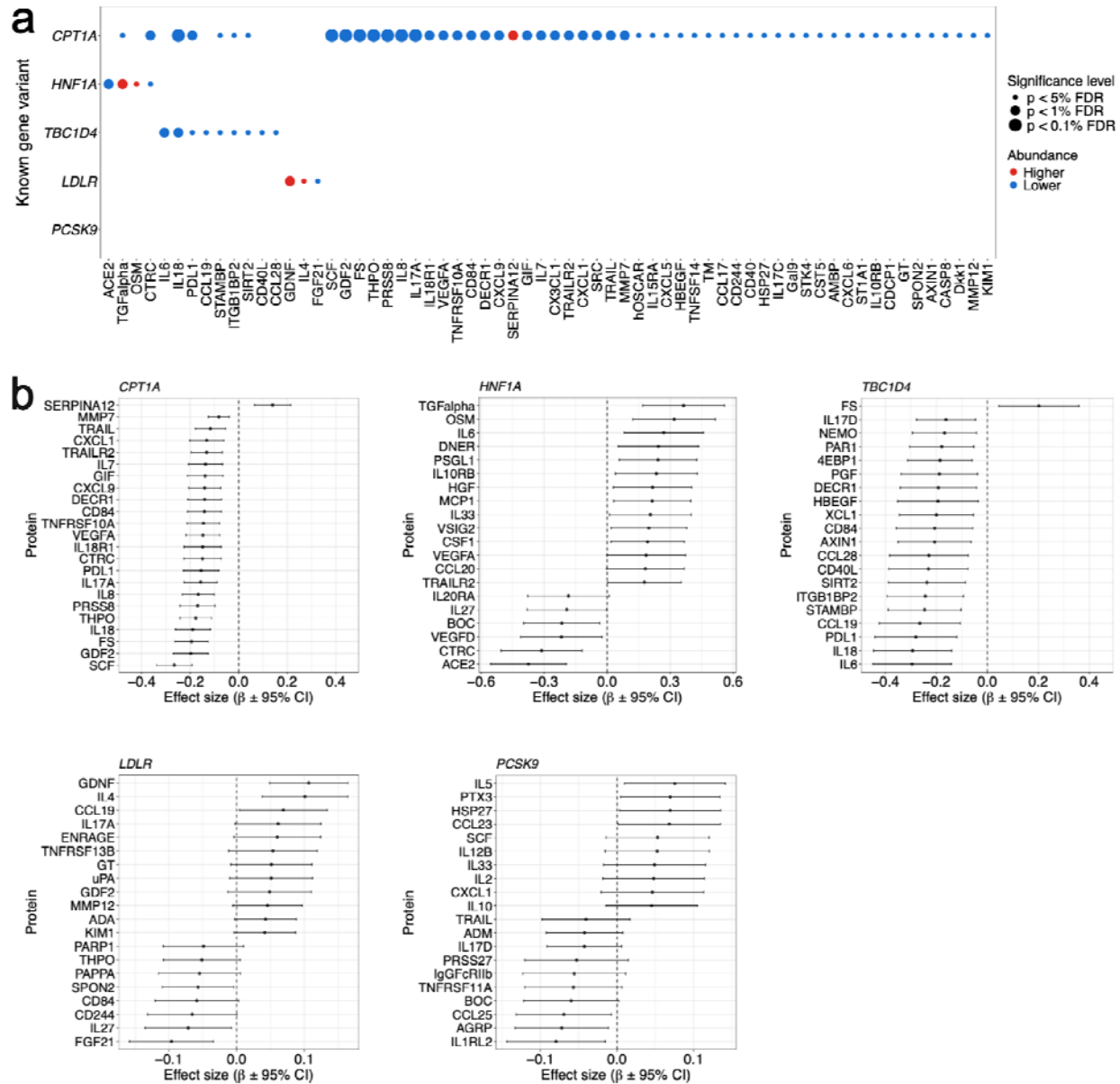
383 The known high-impact variant in *TBC1D4* (rs61736969)³⁵, which causes insulin resistance in
384 the skeletal muscle and explains a significant proportion of the variance in diabetes cases in

385 Greenland, was associated with lower levels of nine plasma proteins after a 5% FDR correction
386 (i.e. IL-6, IL-18, CCL-19, CCL-28, CD40L, ITGB1BP2, PD-L1, STAMBP and SIRT2).

387
388 A variant in the maturity onset diabetes of the young (MODY) gene, *HNF1A* (rs2135845768)³⁶,
389 which explains up to 2.5% of variance in diabetes in Greenland (together with *TBC1D4* explains
390 up to 18% of diabetes cases), was associated with lower plasma levels of two proteins
391 (angiotensin converting enzyme 2 [ACE2] and chymotrypsin-C [CTR]) and higher levels of
392 two proteins (oncostatin-M [OSM] and protransforming growth factor alpha [TGF-alpha]), after
393 a 5% FDR correction.

394
395 A known variant in *LDLR* (rs730882082)³⁷ which has a large impact on familial
396 hypercholesterolemia (FH) in Greenland, was associated with lower plasma levels of one protein
397 (fibroblast growth factor 21 [FGF-21]), and higher levels of two proteins (glial cell line-derived
398 neurotrophic factor [GDNF] and interleukin-4 [IL-4]), after a 5% FDR correction.

399
400 The *PCSK9* loss-of-function variant (p.Arg46Leu; rs11591147) originally identified in European
401 GWAS has led to PCSK9-inhibitors as a treatment for dyslipidemia³⁸. We examined whether any
402 proteins were associated with the recently identified *PCSK9* (rs4609471)²⁷ variant in Greenland,
403 but found no significant differences, after correction for multiple testing using FDR.



404

405

406 **Figure 3. Known high-impact cardiometabolic-related Greenlandic variants.** **a** Significant associations between

407 known high-impact cardiometabolic-related Greenlandic variants - *CPT1A* (rs80356779), *HNF1A* (rs2135845768),

408 *TBC1D4* (rs61736969), *LDLR* (rs730882082) and *PCSK9* (rs4609471) - and inflammatory and cardiovascular-

409 related plasma proteins (FDR<0.05). Bubble sizes describe significance levels after multiple testing; small bubbles

410 indicate pQTLs with FDR<0.05 and large bubbles FDR<0.01. Red - higher abundance; blue - lower abundance. **b**

411 Forest plots for known high-impact variants within the same five genes and the top 20 proteins with the lowest FDR.

412 The centre estimate and error bars on the x-axis describe the effect size ($\beta \pm 95\% \text{ CI}$).

413 Discussion

414 Here, we present a pQTL study in an isolated non-European population with 177 circulating
415 inflammation- and cardiovascular-related proteins, measured using the Olink Proximity
416 Extension Assay in 3,707 Greenlanders. We identified 251 genome-wide significant primary
417 locus-protein pairs (93 *cis*, 158 *trans*), involving 225 different loci for 133 unique proteins. The
418 251 primary locus-protein pairs included many of the same locus-protein pairs found in prior
419 studies performed in Europeans, with 181 (~72.1%) reported previously in 34 peer-reviewed
420 studies, but also revealed 70 locus-protein pair associations (3 *cis*, 6 *semi-trans*, 61 *trans*) that
421 have not been reported previously. We identify 16 pQTLs which are purely recessive, 15 out of
422 16 are in *semi-trans/trans* (e.g. PRSS27, FABP2 and TF). This highlights the importance of
423 conducting analyses in diverse populations, including alternative modes of inheritance.

424
425 Cross-ancestry differences have been highlighted by a previous mass-spectrometry based
426 proteomics study conducted in a population of 2,958 Han Chinese participants, wherein proteins
427 which are positively associated with BMI in European populations, showed opposite direction of
428 effects in East Asians.¹⁵ An additional study found that up to 89% of pQTLs identified in
429 European studies can be replicated in an Arab population of 2,935 individuals, when using the
430 SOMAscan platform.¹⁷ So far, the pQTL studies performed in isolated populations are of
431 European ancestry. An earlier study used a unique approach of whole-genome sequencing and
432 three Olink panels (CVDI, CVDII, metabolism) across three isolated European cohorts
433 (MANOLIS, Pomak, and ORCADES), identifying 5 rare variant pQTLs.¹⁸

434
435 To examine whether high-impact pQTLs were more common in the Greenlandic population, we
436 performed an additional GWAS for a subset of UK Biobank individuals, matched by the sample
437 size of our analysis for Greenlanders (n=3,707 in both populations) and using both additive and
438 recessive models. The three pQTLs with the highest PVE in Greenlanders (IL-27, IgGFCRII-b,
439 MCP-2) explained 44.3%, 43.7%, and 43.2%, compared to the top three pQTLs in the sample-
440 size matched UK Biobank (CD6, MCP-2, IL-18R1) GWAS, which explained 33.9%, 30.0%, and
441 26.9%.

442

443 Certain genetic variants have undergone selection as an adaptation to the extreme conditions of
444 the Arctic and the intake of specialised diets rich in proteins and polyunsaturated fatty acids
445 (PUFAs). An example of this is the Arctic-specific variant in *CPT1A* (rs80356779), which is a
446 known modulator of fatty acid composition (including decreased levels of circulating omega-6
447 and increased omega-3 PUFAs), which may affect the regulation of growth hormones, with large
448 effects on height and weight³⁴. Herein, we demonstrated that this variant has pleiotropic effects
449 on plasma protein levels, with a general pattern of reduced inflammatory and cardiovascular-
450 related protein biomarkers, which may be a result of the reduction in circulating PUFAs,
451 primarily omega-6. *CPT1A* carriers exhibited significantly lower plasma levels of SCF, a protein
452 that has been shown to play a central role in inflammation (cell survival, proliferation, and
453 hematopoiesis) and is a suggested therapeutic target for inflammatory diseases³⁹. *CPT1A* carriers
454 also had significantly lower plasma levels of FS, a hepatokine which is elevated in patients with
455 T2D and is associated with adipose tissue insulin resistance⁴⁰.

456
457 The common loss-of-function variant in *TBC1D4*³⁵ (rs61736969) responsible for muscle insulin
458 resistance and explains up to 15.5% of variance in diabetes in Greenland, and in the present
459 study was associated with lower expression of nine plasma proteins, most of which are
460 interleukins (IL-6, IL-18), chemokines (CCL-19, CCL-28), other immune mediators (CD40L,
461 PD-L1), integrins (ITGB1BP2), metalloproteases (STAMBP), and transferases (SIRT2). IL-6 is
462 elevated with obesity, but on the other hand is also secreted as a myokine from muscles during
463 exercise, exerting anti-inflammatory and anti-diabetogenic effects. Physical activity has been
464 shown to reduce postprandial hyperglycemia in *TBC1D4* carriers⁴¹, and IL-6 levels may be
465 increased as a result of exercise. Moreover, lower expression of ITGB1BP2 (integrin beta-1-
466 binding protein 2) was also observed in *TBC1D4* carriers, a protein that plays an essential role in
467 the maturation and organisation of muscle cells⁴², which could be driven by the impaired insulin
468 resistance in skeletal muscle. It remains unclear whether *TBC1D4* carriers are at an increased
469 risk of cardiovascular disease⁴³. In the present study, we see elevations in several cardiovascular-
470 disease related biomarkers, including ITGB1BP2⁴², STAMBP (linked to congenital
471 malformations of the heart)⁴⁴ and SIRT2 (functions in inhibiting pathological cardiac
472 hypertrophy)⁴⁵.

473

474 The known Inuit-specific variant in the *MODY* gene *HNF1A*³⁶ (rs2135845768), explaining up to
475 2.5% of variance in diabetes in Greenland, was associated with altered expression of four
476 (*ACE2*, *CTRC*, *OSM*, *TGF-alpha*) proteins in the present study. Plasma abundance of *ACE2* was
477 significantly lower in *HNF1A* carriers. *HNF1A* is known to induce *ACE2* expression in the
478 pancreatic islet cells. *ACE2* plays an important role in maintaining glycemia and β -cell function,
479 and has been suggested as a potential therapeutic target for treatment of diabetes⁴⁶.

480

481 A known Arctic-specific variant in *LDLR*³⁷ (rs730882082) which poses a high risk of FH in
482 nearly 30% of Greenlanders, associated with altered expression of three plasma proteins (*FGF-*
483 *21*, *GDNF*, and *IL-4*) in the present study. Specifically, *FGF-21* levels were lower in carriers of
484 the *LDLR* variant, a protein responsible for fatty acid oxidation and regulation of systemic
485 glucose homeostasis and insulin sensitivity. *FGF-21* analogs are being developed for their
486 beneficial effects on modulate lipid and lipoprotein levels, and therefore could potentially serve
487 as treatment for patients with FH⁴⁷. Clinical trials for *FGF-21* analogs are also ongoing for
488 metabolic-dysfunction-associated steatotic liver disease.

489

490 **Limitations**

491 Our study has several limitations. Firstly, our pQTL analysis is restricted to the 177 proteins
492 measured by two Olink panels. An untargeted approach would allow for identification of more
493 and potentially novel pQTLs, yet targeted approaches allow for the study of biologically relevant
494 protein biomarkers. Inclusion of orthogonal methods such as mass spectrometry-based
495 proteomics, would resolve the possibility of epitope effects, yet this kind of validation is difficult
496 to perform at scale. Tissue- and cell-type specific pQTL studies will be valuable in future studies
497 to understand differences in genetic signals across tissues. We did not examine ratios between
498 protein levels which may identify novel associations, when multiple proteins are associated with
499 the same variant⁴⁸. Furthermore, the raw abundance of proteins does not take into consideration
500 protein-protein interactions, mapping the interactome can lead to additional biological insights⁴⁹.

501 **Strengths**

502 Identification of pQTLs aid in the understanding of mechanisms of complex disease and can
503 identify promising targets for pharmacological interventions. The majority of all pQTL studies
504 performed to date have been conducted in European populations, limiting the generalizability of

505 findings across populations. Although an attempt was made to identify pQTLs in non-European
506 ancestry in the UKB-PPP (953 with South Asian and 1,513 with African ancestry), the sample
507 size is smaller than the present study⁵⁰. Here, we explore pQTLs in a historically isolated
508 Greenlandic population, identify associations related to alternative modes of inheritance, and
509 highlight the need to study diverse populations, through comparisons to large datasets generated
510 in the UK Biobank, primarily of European ancestry. The challenge in genetic research in
511 indigenous populations is that it raises complex psychological, ethical, social, and political
512 issues. These issues have been raised in a series of user surveys on the use of genetics in
513 metabolic research. Generally, the communities fully support the research with a specific request
514 to investigate the significance of traditional and modern Inuit lifestyle, in order to understand
515 how advising individuals regarding diet and lifestyle, can be linked in a culturally appropriate
516 way, to better inform health effects of genetic profiles.

517

518 **Conclusions**

519 The pQTLs identified and analysed in this study provide a resource to unravel the genetic
520 architecture of the targeted plasma proteome in the Greenlandic population. There was a larger
521 proportional variance explained in Greenlanders when compared to European ancestries, which
522 has relevant implications for disease outcomes. Variants with large population impact on lipid
523 metabolism (*CPT1A*) and risk of cardiometabolic disease in Greenland (*TBC1D4*, *HNFI1A*, and
524 *LDLR*), show distinct patterns with inflammatory cytokines and cardiovascular disease
525 biomarkers, pointing towards potential therapeutic targets.

526

527

528 **Contributors**

- 529 ● Conceptualization: SES, RFB, MKA, NG, MEJ, IM, AA, TH
- 530 ● Data curation: SES, MKA, ACBT, NG, MEJ, IM, AA, TH
- 531 ● Formal analysis: SES, RFB, FFS, SH, LL
- 532 ● Funding acquisition: PB, CVLL, NG, MEJ, IM, AA, TH
- 533 ● Investigation: SES, RFB, FFS, PB, CVLL, NG, MEJ, IM, AA, TH
- 534 ● Methodology: SES, RFB, FFS, NG, MEJ, IM, AA, TH
- 535 ● Project administration: SES, RFB, MKA, PB, CVLL, NG, MEJ, IM, AA, TH
- 536 ● Resources: PB, CVLL, NG, MEJ, IM, AA, TH
- 537 ● Software: AA
- 538 ● Supervision: NG, IM, AA, TH
- 539 ● User surveys: MEJ
- 540 ● Visualisation: SES, RFB, FFS, IM, AA, TH
- 541 ● Writing - original draft: SES, RFB, IM, AA, TH
- 542 ● Writing - review & editing: MKA, FFS, SH, ACBT, LL, PB, CVLL, NG, MEJ

543

544 **Data sharing statement**

545 Summary data from the study may be available by reasonable request to the authors.

546

547 **Declaration of interests**

548 The population surveys of Greenland were funded by the Department of Health, Greenland, The
549 Novo Nordisk Foundation (NNF17OC0028136 & NNF17SH0027192), The Independent
550 Research Fund Denmark, and Karen Elise Jensen Foundation. The Novo Nordisk Foundation
551 Center for Basic Metabolic Research is an independent research centre at the University of
552 Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation
553 (NNF18CC0034900). SES received funding from The Novo Nordisk Foundation, Copenhagen
554 Bioscience PhD Program (grant number: NNF18CC0033668). RFB and IM are supported by a
555 Villum Young Investigator grant (VIL19114), awarded to IM. RFB, SH, LL and AA are funded
556 by the Novo Nordisk Foundation (NNF20OC0061343). AA is supported by the Independent
557 Research Fund Denmark (8021-00360B). None of the funding agencies had any role in study

558 design or collection, analysis, or interpretation of data. NG is currently employed at Novo
559 Nordisk. All other authors have no further conflicts of interest to declare.

560

561 **Acknowledgements**

562 We would like to acknowledge all the participants and the staff of the Greenland Population
563 Health Surveys (Befolkningsundersøgelsen i Grønland). We would also like to thank Annemette
564 Forman and Tina Hvidtfeldt Lorentzen from the Novo Nordisk Foundation Center for Basic
565 Metabolic Research, University of Copenhagen for their technical support.

566

567 **Appendix A. Supplementary Data**

568

569 **Appendix B. Supplementary Figures**

570

571 References

- 572 1 Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function.
573 *Nature* 2016; **537**: 347–55.
- 574 2 Deutsch EW, Omenn GS, Sun Z, *et al.* Advances and Utility of the Human Plasma
575 Proteome. *J Proteome Res* 2021; **20**: 5241–63.
- 576 3 Katz DH, Robbins JM, Deng S, *et al.* Proteomic profiling platforms head to head:
577 Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods.
578 *Sci Adv* 2022; **8**: eabm5164.
- 579 4 Meissner F, Geddes-McAlister J, Mann M, Bantscheff M. The emerging role of mass
580 spectrometry-based proteomics in drug discovery. *Nat Rev Drug Discov* 2022; **21**: 637–54.
- 581 5 Smith JG, Gerszten RE. Emerging Affinity-Based Proteomic Technologies for Large-Scale
582 Plasma Profiling in Cardiovascular Disease. *Circulation* 2017; **135**: 1651–64.
- 583 6 Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large
584 population-based studies. *Nat Rev Genet* 2020; **22**: 19–37.
- 585 7 Aguet F, Alasoo K, Li YI, *et al.* Molecular quantitative trait loci. *Nature Reviews Methods
586 Primers* 2023; **3**: 1–22.
- 587 8 Sun BB, Maranville JC, Peters JE, *et al.* Genomic atlas of the human plasma proteome.
588 *Nature* 2018; **558**: 73–9.
- 589 9 Pietzner M, Wheeler E, Carrasco-Zanini J, *et al.* Mapping the proteo-genomic convergence
590 of human diseases. *Science* 2021; **374**: eabj1541.
- 591 10 Folkersen L, Gustafsson S, Wang Q, *et al.* Genomic and drug target evaluation of 90
592 cardiovascular proteins in 30,931 individuals. *Nature Metabolism* 2020; **2**: 1135–48.
- 593 11 Emilsson V, Ilkov M, Lamb JR, *et al.* Co-regulatory networks of human serum proteins link
594 genetics to disease. *Science* 2018; **361**: 769–73.
- 595 12 Enroth S, Johansson A, Enroth SB, Gyllensten U. Strong effects of genetic and lifestyle
596 factors on biomarker variation and use of personalized cutoffs. *Nat Commun* 2014; **5**: 4684.
- 597 13 Suhre K, Arnold M, Bhagwat AM, *et al.* Connecting genetic risk to disease end points
598 through the human blood plasma proteome. *Nat Commun* 2017; **8**: 14357.
- 599 14 Sun BB, Chiou J, Traylor M, *et al.* Plasma proteomic associations with genetics and health
600 in the UK Biobank. *Nature* 2023; **622**: 329–38.
- 601 15 Xu F, Yu EY-W, Cai X, *et al.* Genome-wide genotype-serum proteome mapping provides
602 insights into the cross-ancestry differences in cardiometabolic disease susceptibility. *Nat
603 Commun* 2023; **14**: 1–12.

- 604 16 Said S, Pozarickij A, Lin K, *et al.* Ancestry diversity in the genetic determinants of the
605 human plasma proteome and associated new drug targets. *medRxiv.* 2023; :
606 2023.11.13.23298365.
- 607 17 Thareja G, Belkadi A, Arnold M, *et al.* Differences and commonalities in the genetic
608 architecture of protein quantitative trait loci in European and Arab populations. *Hum Mol*
609 *Genet* 2023; **32**: 907–16.
- 610 18 Gilly A, Klaric L, Park Y-C, *et al.* Gene-based whole genome sequencing meta-analysis of
611 250 circulating proteins in three isolated European populations. *Mol Metab* 2022; **61**:
612 101509.
- 613 19 Png G, Barysenka A, Repetto L, *et al.* Mapping the serum proteome to neurological
614 diseases using whole genome sequencing. *Nat Commun* 2021; **12**: 1–12.
- 615 20 Wojcik GL, Graff M, Nishimura KK, *et al.* Genetic analyses of diverse populations
616 improves discovery for complex traits. *Nature* 2019; **570**: 514–8.
- 617 21 Easteal S, Arkell RM, Balboa RF, *et al.* Equitable Expanded Carrier Screening Needs
618 Indigenous Clinical and Population Genomic Data. *Am J Hum Genet* 2020; **107**: 175–82.
- 619 22 Moltke I, Fumagalli M, Korneliussen TS, *et al.* Uncovering the genetic history of the
620 present-day Greenlandic population. *Am J Hum Genet* 2015; **96**: 54–69.
- 621 23 Jørgensen ME, Bjerregaard P, Borch-Johnsen K. Diabetes and impaired glucose tolerance
622 among the inuit population of Greenland. *Diabetes Care* 2002; **25**: 1766–71.
- 623 24 Bjerregaard P, Curtis T, Borch-Johnsen K, *et al.* Inuit health in Greenland: a population
624 survey of life style and disease in Greenland and among Inuit living in Denmark. *Int J*
625 *Circumpolar Health* 2003; **62 Suppl 1**: 3–79.
- 626 25 Jørgensen ME, Borch-Johnsen K, Stolk R, Bjerregaard P. Fat distribution and glucose
627 intolerance among Greenland Inuit. *Diabetes Care* 2013; **36**: 2988–94.
- 628 26 Bjerregaard P, Larsen CVL, Olesen I, *et al.* The Greenland population health survey 2018 –
629 methods of a prospective study of risk factors for lifestyle related diseases and social
630 determinants of health amongst Inuit. *Int J Circumpolar Health* 2022; published online June
631 16. DOI:10.1080/22423982.2022.2090067.
- 632 27 Senftleber NK, Andersen MK, Jørsboe E, *et al.* GWAS of lipids in Greenlanders finds
633 association signals shared with Europeans and reveals an independent PCSK9 association
634 signal. *Eur J Hum Genet* 2023; published online Oct 30. DOI:10.1038/s41431-023-01485-8.
- 635 28 Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and
636 population genetic studies. *Nat Methods* 2013; **10**: 5–6.
- 637 29 A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.

- 638 30 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method
639 for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
- 640 31 Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies.
641 *Nat Genet* 2012; **44**: 821–4.
- 642 32 Meisner J, Albrechtsen A. Inferring Population Structure and Admixture Proportions in
643 Low-Depth NGS Data. *Genetics* 2018; **210**: 719–31.
- 644 33 Zhang C 'e, Wang W, Zhang H 'e, Wei L, Guo S. Association of FCGR2A rs1801274
645 polymorphism with susceptibility to autoimmune diseases: A meta-analysis. *Oncotarget*
646 2016; **7**: 39436–43.
- 647 34 Fumagalli M, Moltke I, Grarup N, *et al.* Greenlandic Inuit show genetic signatures of diet
648 and climate adaptation. *Science* 2015; **349**: 1343–7.
- 649 35 Moltke I, Grarup N, Jørgensen ME, *et al.* A common Greenlandic TBC1D4 variant confers
650 muscle insulin resistance and type 2 diabetes. *Nature* 2014; **512**: 190–3.
- 651 36 Thuesen ACB, Stæger FF, Kaci A, *et al.* A novel splice-affecting HNF1A variant with large
652 population impact on diabetes in Greenland. *Lancet Reg Health Eur* 2023; **24**: 100529.
- 653 37 Jørsboe E, Andersen MK, Skotte L, *et al.* An LDLR missense variant poses high risk of
654 familial hypercholesterolemia in 30% of Greenlanders and offers potential of early
655 cardiovascular disease intervention. *HGG Adv* 2022; **3**: 100118.
- 656 38 Shapiro MD, Tavori H, Fazio S. PCSK9: From Basic Science Discoveries to Clinical Trials.
657 *Circ Res* 2018; **122**: 1420–38.
- 658 39 Stem cell factor and its receptor c-Kit as targets for inflammatory diseases. *Eur J*
659 *Pharmacol* 2006; **533**: 327–40.
- 660 40 Wu C, Borné Y, Gao R, *et al.* Elevated circulating follistatin associates with an increased
661 risk of type 2 diabetes. *Nat Commun* 2021; **12**: 1–10.
- 662 41 Schnurr TM, Jørsboe E, Chadt A, *et al.* Physical activity attenuates postprandial
663 hyperglycaemia in homozygous TBC1D4 loss-of-function mutation carriers. *Diabetologia*
664 2021; **64**: 1795–804.
- 665 42 Brancaccio M, Fratta L, Notte A, *et al.* Melusin, a muscle-specific integrin beta1-interacting
666 protein, is required to prevent cardiac failure in response to chronic pressure overload. *Nat*
667 *Med* 2003; **9**: 68–75.
- 668 43 Jørgensen ME, Tvermosegaard M, Rønn PF, *et al.* Do Greenlandic Carriers of the TBC1D4
669 p.Arg684Ter Variant Have Increased Risk of Cardiovascular Disease? *Diabetes* 2018; **67**:
670 439 – P.
- 671 44 Spielmann N, Miller G, Oprea TI, *et al.* Extensive identification of genes involved in

- 672 congenital and structural heart disorders and cardiomyopathy. *Nature Cardiovascular*
673 *Research* 2022; **1**: 157–73.
- 674 45 Wu B, You S, Qian H, *et al.* The role of SIRT2 in vascular-related and heart-related
675 diseases: A review. *J Cell Mol Med* 2021; **25**: 6470–8.
- 676 46 Pedersen KB, Chhabra KH, Nguyen VK, Xia H, Lazartigues E. The transcription factor
677 HNF1 α induces expression of angiotensin-converting enzyme 2 (ACE2) in pancreatic islets
678 from evolutionarily conserved promoter motifs. *Biochim Biophys Acta* 2013; **1829**: 1225–
679 35.
- 680 47 Besseling J, Ong KL, Huijgen R, *et al.* The impact of LDLR function on fibroblast growth
681 factor 21 levels. *Atherosclerosis* 2015; **241**: 322–5.
- 682 48 Suhre K. Genetic associations with ratios between protein levels detect new pQTLs and
683 reveal protein-protein interactions. bioRxiv. 2023; published online July 20.
684 DOI:10.1101/2023.07.19.549734.
- 685 49 Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a
686 multiomics approach. *Nat Rev Cardiol* 2021; **18**: 313–30.
- 687 50 Eldjarn GH, Ferkingstad E, Lund SH, *et al.* Large-scale plasma proteomics comparisons
688 through genetics and disease associations. *Nature* 2023; **622**: 348–58.
- 689