

Integrative genetic and immune cell analysis of plasma proteins in healthy donors identifies novel associations involving primary immune deficiency genes

Barthelemy Caron¹, Etienne Patin², Maxime Rotival², Bruno Charbit³, Matthew L Albert⁴, Lluís Quintana-Murci^{2,5,#}, Darragh Duffy^{3,6,#,*}, Antonio Rausell^{1,7,#,*}, and the Milieu Intérieur Consortium[†]

¹ Université de Paris, INSERM UMR1163, Imagine Institute, Clinical Bioinformatics Laboratory, Paris, F-75006, France.

² Human Evolutionary Genetics Unit, Institut Pasteur, UMR2000, CNRS, Paris, F-75015, France.

³ Cytometry and Biomarkers UTechS, CRT, Institut Pasteur, Paris, F-75015, France.

⁴ Insitro, San Francisco, California, CA, 94080, USA.

⁵ Human Genomics and Evolution, Collège de France, Paris, F-75005, France

⁶ Translational Immunology Lab, Institut Pasteur, Paris, F-75015, France

⁷ Molecular genetics service, AP-HP, Necker Hospital for Sick Children, Paris, F-75015, France

(#) Equal contribution

(*) Correspondence: darragh.duffy@pasteur.fr , antonio.rausell@institutimagine.org

Keywords: pQTL, plasma proteins, immune cells, immune variability

Abstract

Blood plasma proteins play an important role in immune defense against pathogens, including cytokine signaling, the complement system and the acute-phase response. Recent large-scale studies have reported genetic (*i.e.* quantitative trait loci, pQTLs) and non-genetic factors, such as age and sex, as major determinants to inter-individual variability in immune response variation. However, the contribution of blood cell composition to plasma protein heterogeneity has not been fully characterized and may act as a confounding factor in association studies. Here, we evaluated plasma protein levels from 400 unrelated healthy individuals of western European ancestry, who were stratified by sex and two decades of life (20-29 and 60-69 years), from the Milieu Intérieur cohort. We quantified 297 proteins by Luminex in a clinically certified laboratory and their levels of variation were analysed together with 5.2M single-nucleotide polymorphisms. With respect to non-genetic variables, we included more than 700 lifestyle and biochemical factors, as well as counts of seven circulating immune cell populations measured by hemogram and

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
standardized flow cytometry. Collectively, we found 152 significant associations involving 49 proteins and 20

non-genetic variables. Consistent with previous studies, age and sex showed a global, pervasive impact on plasma protein heterogeneity, while body mass index and other health status variables were among the non-genetic factors with the highest number of associations. After controlling for these covariates, we identified 100 and 12 pQTLs acting in *cis* and *trans*, respectively, collectively associated with 87 plasma proteins and including 30 novel genetic associations. Genetic factors explained the largest fraction of the variability of plasma protein levels, as compared to non-genetic factors. In addition, blood cell fractions, including leukocytes, lymphocytes and three types of polymorphonuclear cells, had a larger contribution to inter-individual variability than age and sex, and appeared as confounders of specific genetic associations. Finally, we identified new genetic associations with plasma protein levels of eight monogenic Mendelian disease genes including three primary immunodeficiency genes (Ficolin-3, Interleukine-2 Receptor alpha and FAS). Our study identified novel genetic and non-genetic factors associated to plasma protein levels which may inform health status and disease management.

Introduction

Plasma proteins play important physiological roles in human health and disease. They participate in immune responses against pathogens (e.g. interferons, chemokines and complement factors^{1,2}), blood clotting³, hormone transport^{4,5}, and energy metabolism regulation⁶. Plasma protein levels reflect the balance of diverse biological processes including active cellular secretion⁷⁻⁹, tissue leakage^{10,11}, protein degradation¹² and protein excretion in urine¹³. Plasma proteins are widely used as markers of the physiological state of an individual and represent ~42% of all requested blood-based laboratory tests¹⁰. As of today, the US Food and Drug Agency (FDA) approved 235 plasma proteins as diagnostic, prognostic, risk predictive or treatment response biomarkers (<http://mrmassaydb.proteincentre.com/fdaassay/>¹⁴) for a broad range of diseases such as cancer¹⁵⁻¹⁷, pulmonary defects¹⁸, autoimmune¹⁹ and metabolic diseases²⁰. In addition to their association with clinical outcomes, natural heterogeneity of plasma protein levels among the general population has been widely reported but is not considered in clinical applications. Recent large-scale studies performed both in healthy and disease cohorts have identified both non-genetic (e.g. age and sex) and genetic factors (i.e. quantitative trait loci, pQTLs) that determine variable plasma protein levels²¹⁻²³. pQTLs are enriched in disease-susceptibility loci identified from GWAS studies^{23,24}, and could have protective or modifying effects, potentially in conjunction with pathogenic mutations leading to disease due to altered expression levels, e.g. loss of homeostasis, proteotoxic stress or insufficiency²⁵. Yet, the assessment of the genetic associations reported by previous studies did not characterize the specific cell types accounting for the observed variation in plasma proteins. Thus, it remains unclear whether a fraction of the plasma protein variability initially

associated with a pQTL could have been confounded by concomitant heterogeneity in their cellular sources. This may be especially relevant for plasma proteins displaying immune-related functions, since significant variability in immune cell fractions is observed across individuals driven by both genetic and non-genetic factors^{26,27}.

Here, we present an in-depth characterization of heterogeneity in plasma protein levels in healthy individuals from the Milieu Intérieur study²⁸, with a focus on immune-related proteins. The Milieu Intérieur consortium aims at characterizing the genetic and environmental factors underlying the observed variability of the immune response in a healthy population²⁸. This study was performed on 400 individuals equally distributed by sex and across two decades of life (aged 30-39, and 60-69). We evaluated the association of 229 plasma protein concentrations with a total of 254 non-genetic factors including lifestyle, environmental, physiological and blood biochemical variables as well as with 5,201,100 common single nucleotide polymorphisms (SNPs). To control for the natural variation in blood cell populations, we systematically accounted for the levels of seven major blood-cell fractions, including leukocytes, lymphocytes and three types of polymorphonuclear cells. We found that together with age and sex, blood cell fractions explain an important fraction of the inter-individual plasma protein variability. After controlling for such factors, we identified 112 pQTLs associated with 87 proteins, 24 of which are reported here for the first time. Among these, nine are associated in *cis* to monogenic Mendelian disease genes (MMDGs), including 3 primary immunodeficiency (PID) genes. Such genetic variants may have potential clinical value as susceptibility or protective factors for immune-related diseases.

Results

Variation of plasma protein levels in a well-defined healthy population

We quantified the concentrations of 297 plasma proteins in 400 healthy individuals from the Milieu Intérieur (MI) study²⁸ using CLIA certified assays. After quality control accounting for detection limits (**Methods**), 229 proteins were retained for downstream analyses, including 141 immune-related plasma proteins (*i.e.* proteins with previously identified immune functions or produced by immune cells, **Methods**). First, we evaluated the impact of host genetics and non-genetic factors on plasma protein levels. Genome-wide association tests for each of the 229 proteins were performed against a total of 5,201,100 common SNPs (minor allele frequency ≥ 0.05). Covariates that we systematically included in the analysis were age, sex, counts of 7 major blood-cell sub-populations (lymphocytes, leukocytes, neutrophils, basophils, eosinophils, monocytes and platelets)

and the first two principal components of a principal component analysis of the genetic data, representing genetic stratification in the sample (**Methods**). Additional non-genetic factors were selected among 254 lifestyle, environmental, physiological and blood biochemical variables, and added as confounders in a protein-specific manner (**Methods**). The relative contribution (marginal correlation, CAR score, **Methods**) of non-genetic, genetic and cell fraction components to the inter-individual variability of the 229 plasma proteins evaluated in this work is presented in **Figure 1** and **Supplementary Table 1-3 (Methods)**.

Consistent with previous studies^{21–23,29–32}, we found that age and sex had a widespread effect on plasma protein levels, each explaining on average 2.8% of the total observed variability (**Figure 1, Supplementary Table 1, Methods**). Similar figures were observed for the subset of immune-related proteins, *i.e.* 2.3% and 2.1% for age and sex respectively. For specific proteins, however, the observed contribution of age and sex was particularly large, in line with previous findings. For example, age explained 30.2% of growth differentiation factor 15 variability (GDF15)^{29,33}, while variability attributed to sex was 42.8% for Leptin³⁴, 39% for Stromelysin-1 (MMP3)^{29,35}, 36.2% for FSH and 32.1% for LH^{29,36}. Moreover, when accounting for potential covariation among the 229 proteins through a principal component analysis (PCA), both age and sex showed a strong association with the global heterogeneity of protein levels (univariate linear modeling against PC1 coordinates, p-values = 1.8e-07 and 1.4e-06, respectively; and PC2, p-values = 4.2e-24 and 3.5e-05, respectively; **Supplementary Figure 1, Methods**). While previous studies mostly assessed the global impact of age and sex on plasma proteins, their effects appear highly heterogeneous across proteins.

Blood cell fractions explain a substantial part of plasma protein level heterogeneity

To assess the potential effect of circulatory cell counts on plasma protein levels, we next quantified their relative contribution to the inter-individual variability of each protein (**Methods**). Taken together, blood-cell fractions explained on average 3.6% of the variability of the observed plasma protein levels. This contribution was comparatively higher than those of age and sex (two-sided Wilcoxon test p-value = 6.4e-11 and 2.3e-14, respectively; **Figure 1, Supplementary Table 1**). Furthermore, blood-cell fractions explained significantly more variability for immune-related proteins than for the rest of proteins evaluated (mean explained variability 4% and 2.9% respectively, one-sided Wilcoxon test, P = 4.8e-02). Platelet counts alone explained an average of 1.6% of the variability of immune related proteins, as compared to 0.78% for the rest of proteins (one-sided Wilcoxon test, p-value = 3.9e-04), with contributions as high as 16.3% for the Neutrophil Activating Peptide 2, and 13.3% for Thrombospondin-1. These results highlight the contribution of blood-cell fractions to the variability of plasma protein levels, and support their consideration as a potential confounding factor in the assessment of genetic associations.

Plasma lipids and body mass index are important covariates of specific plasma proteins

Two other classes of non-genetic factors were found to substantially associate with plasma levels of specific proteins. First, plasma lipids such as triglycerides, HDL, LDL and total cholesterol were associated with expression levels of 25 proteins, including various components of cholesterol particles as well as proteins involved in lipid transport (ApoA1, ApoB, ApoC1, ApoC3, ApoD, ApoE, FABP-adipocyte, SHBG), metabolism and homeostasis (Adiponectin, Carboxypeptidase B2, C3, CFH, C-peptid, Endoglin, FGF21, IGFBP2, Leptin, Leptin Receptor, PEDF, Prostatin, PSAT, RBP4, SAP, tPA). Second, anthropometric factors such as body mass index (BMI) and abdominal circumference were associated with plasma levels of 20 proteins, most of which also associate with plasma lipids (e.g. Adiponectin, ApoD, C-peptid, FABP-adipocyte, SHBG). Blood lipids and anthropometric factors accounted on average for 11% and 8.1% of the variability of the associated plasma proteins, respectively. Yet, the highest association was observed between HDL and the Apolipoprotein A-1 (marginal correlation of 44.6%)³⁷ (**Figure 1, Figure 2, Supplementary Table 2, Methods**). Globally, anthropometric factors and plasma lipid levels are known to be markers of physical shape and overall health. Interestingly, the association of complement factors (C3 and CFH) with anthropometric traits may reflect the low level inflammation induced by higher body mass, both of which associate with obesity, cardiovascular diseases and increased susceptibility to infections^{38–40}.

Contribution of host genetics on plasma protein levels

Genome wide association testing against the 229 plasma proteins identified 112 pQTLs, including 100 *cis*- and 12 *trans*-pQTLs, and collectively involving 87 proteins and 111 SNPs (FDR \leq 0.05; **Figure 3, Supplementary Figure 2, Supplementary Figure 3, and Supplementary Table 3, Methods**). Among the 87 proteins, 76 were only associated with *cis*-SNPs, 9 with only *trans*-SNPs and 2 with both. Sixty-two proteins were associated with one SNP, and 25 with two independent SNPs. Interestingly, three loci aggregated several pQTLs associated with different proteins. First, among the 12 *trans*-pQTLs identified, 4 (E-selectin-rs2519093; PECAM-1-rs2519093; Cadherin-1-rs635634; FASLG receptor-rs687621) were in moderate ($R^2 = 0.45$ rs2519093 and rs687621) to high ($R^2 = 0.99$, rs2519093 and rs635634) linkage disequilibrium (LD) and colocalized in a 18kb region of chromosome 9, previously described as the ABO locus, and known to be associated with the expression of many plasma proteins^{22,24,41,42}. Second, two SNPs on chromosome 1 in high LD ($R^2 \geq 0.99$) at the *CFHR4* locus associated, respectively, in *cis* (rs60642321) with *CFHR1* plasma levels, and in *trans* (rs115094736) with *TFR1* plasma levels. While the former had previously been reported in blood⁴³, the latter was, to the best of our knowledge, not reported before, neither

as a pQTL nor as an eQTL. In addition, no physical or regulatory interactions have been reported between *TFRC* and any other gene or protein in a 500kb window centered on the associated SNP (as reported in STRING-db⁴⁴, **Methods**). Last, two SNPs (rs584007 and rs3826688) in high LD ($R^2 \geq 0.99$), located on chromosome 19, associated in *cis* with the plasma levels of Apo E and Apo C1, respectively. Both SNPs are located within a known *ApoE* enhancer⁴⁵, and were previously described as *cis*-eQTLs of both genes (in blood or in other tissues), hinting at a potential co-regulation of the expression of both genes^{46,47}. Among the 12 *trans*-pQTLs identified therein, six were located within a maximum of one Mb from a known eQTL for the same gene (based on QTLbase⁴⁸). Last, only one gene located at the *trans*-pQTL locus was shown to interact with the associated gene (*SORT1* and *GRN*, respectively; physical or regulatory interactions reported in STRING-db, **Methods**). Indeed, *GRN* encoded protein, Progranulin, was shown to bind to *SORT1* encoded protein, Sortilin 1, based on co-immunoprecipitation experiments performed in various mice cell lines^{49–51} and in green monkey fibroblasts, and on co-expression experiments in human breast cancer cell lines^{52,53} (**Supplementary Table 3**). Individually considered, *cis*-pQTLs explained a mean of 12.6% of the associated protein levels (marginal correlation interquartile range from 4.3% to 14.6%). The highest variance explained (marginal correlation, CAR score, **Methods**) by *cis*-pQTLs were for the rs7041 polymorphism and Vitamin D-binding protein (65.6%), rs2856448 for the Tenascin-X protein (61.5%) and rs60642321 for the Complement factor H-related protein 1 (54.4%). Significant *trans*-pQTLs explained a mean of 12.9% of the variability of the associated protein levels (interquartile range from 8.1% to 14.6%), with a maximum contribution of 32.2% in the case of the rs115094736 SNP for the Transferrin receptor protein 1. Considering all *cis*-pQTLs and all *trans*-pQTLs, they explained on average 16.2% and 14.1% of the total variance, respectively. The per-protein global contribution of *cis*-QTLs to plasma level heterogeneity were lower for immune related proteins as compared to the rest of the evaluated proteins (13.8% and 19.5% on average, respectively, two-sided Wilcoxon test p-value = 0.10).

Characteristics of the pQTL summary statistics

We next evaluated the relationship between the pQTL minor allele frequency, its absolute effect size, the absolute variability of the associated protein, and the fraction of such variability explained by the pQTL. Consistent with previous studies^{21,27,54–56}, we observed a significant negative correlation between the minor allele frequency (MAF) and the absolute effect size of the SNPs (Spearman's rho = -0.30, p-value = 1e-03). In addition, we observed a mild positive correlation between the MAF and the variability explained by the pQTLs (Spearman's rho = 0.25, p-value = 7e-03), as well as a strong positive correlation between the absolute effect size and the explained variability (Spearman's rho = 0.57, p-value < 2.2e-16). A strong

correlation was also observed between the absolute effect size of the pQTLs and the inter-individual standard deviation of the associated proteins (Spearman's $\rho = 0.70$, $p\text{-value} < 2.2e-16$). These results are in line with previous studies^{21,27,54–56} observing that the more frequent a regulatory SNP is in the population, the more it contributes to the variability of the associated protein, but with lower effect size and lower inter-individual variance. Notwithstanding, a bias associating high MAF SNPs with low effect size SNPs due to higher statistical power cannot be excluded^{54,57}.

Accounting for blood cell fractions reveals new genetic associations

To evaluate the impact of considering blood cell fractions in the evaluation of genetic associations with plasma protein levels, we tested whether a linear model accounting for cell fractions better fits protein levels than a simpler model not considering them as covariates. Out of the 112 pQTLs reported in this work, the addition of cell fractions significantly improved the linear model in 42 of the cases (one-way ANOVA, F test $p\text{-values} \leq 0.05$, **Methods**). In consequence, we repeated a genome-wide pQTL assessment as previously described, while excluding blood-cell fractions from the covariates (**Methods, Supplementary Note 1**). We found genetic associations for 94 proteins, as compared to 87 proteins initially identified, with 84 proteins in common. Thus, three proteins were specific to the analysis accounting for blood cell fractions as covariates, while 10 proteins were specific to the analysis not considering blood cell fractions (**Supplementary Table 4, Methods**). Moreover, 6 of the 13 proteins showing different pQTL results between the two settings were in turn significantly associated with at least one of the seven circulatory cell fractions tested herein (CEACAM6, Hemopexin, Resistin, TARC, Thrombospondin 1 and TTR; **Supplementary Table 4, Methods**). Yet, the nominal $p\text{-values}$ of genetic associations from one analysis to the other remained significant. These results show that cell fractions are an important factor for the study of genetic and non-genetic associations with plasma protein variability across healthy individuals. However, a mediator role can't be directly inferred from the previous associations.

Replication of previously reported plasma proteins presenting *cis*-pQTLs

We then evaluated the extent to which our study replicated previously reported plasma protein associations with proximal genetic polymorphisms (*i.e.* *cis*-pQTLs) in three recent large-scale studies^{21–23} (**Supplementary Table 5**). When considering a nominal $p\text{-value}$ threshold of 0.01, we replicated 98.85% of the $n=87$ significant associations reported by the three studies, out of the $n=131$ plasma proteins common with our study. Using a more stringent significance threshold based on FDR correction (**Methods**), the replication rate would drop to 52.2%, 63.6% and 77.3% for Sun *et al.*, Suhre *et al.*, and Deming *et al.*

respectively (**Figure 4**). From a complementary perspective, 54 out of 78 (66.7%) and 5 out of 11 (45.5%) of the proteins with *cis*- and *trans*-pQTLs in our study, respectively, had been previously reported by at least one of 4 large scale studies^{21–23,32}. Conversely, we identified 24 novel genetically regulated plasma proteins, collectively associated to a total of 24 *cis*- and 6 *trans*-pQTLs (**Supplementary Table 3**).

Clinical relevance of plasma proteins and associated genetic factors

We characterized the potential medical interest of the pQTLs identified and their associated genes. Both *cis* and *trans* pQTLs reported in our study were significantly enriched in GWAS-based disease- or trait-associated SNPs, showing ~7 and ~7.8 times more GWAS hits, respectively, than expected (43% and 41.6% observed as compared to an expectation of 6.2% and 5%, odds ratio of 12 and 15.8 respectively, with a re-sampling p-value < 1e-04; **Methods, Figure 5, Supplementary Table 3**). Seventeen of the 87 proteins with pQTLs are FDA-approved biomarkers, including two plasma proteins not previously associated with pQTLs, *i.e.* Cancer Antigen 15-3 and Sex Hormone-Binding Globulin. In addition, among the 87 genes collectively associated to the 112 pQTLs identified, 29 genes are known monogenic Mendelian diseases genes (MMDGs), including six primary immunodeficiency genes (PIDs) (**Methods**). Notably, the plasma protein levels of eight out of these 29 MMDGs, including three PID genes - *i.e.*: *FCN3*, *IL2RA* and *FAS* - had not been previously reported to be associated with genetic polymorphisms (pQTLs) in reference repositories (**Figure 6, Supplementary Note 2, Methods**). The identification of pQTLs associated to such Mendelian disease genes may contribute to the genetic characterization of the observed incomplete penetrance or severity heterogeneity across patients suffering from primary immune deficiencies.

Discussion

In this work, we characterized non-genetic and genetic factors explaining the natural heterogeneity of 229 plasma protein levels observed in healthy individuals. We replicated previous findings^{21–23,29–32} describing that age and sex have a global impact on plasma proteins, while anthropometric variables, blood lipids and metabolic markers are also relevant factors for specific proteins. In addition, we characterized the contribution of seven major blood-cell fractions to inter-individual heterogeneity, and found that their contribution to plasma protein variability was higher than age and sex. Moreover, our results suggest that blood-cell fractions may act as important confounders of genetic associations with specific plasma protein levels. In addition to non-genetic factors, we identified 100 and 12 pQTLs acting in *cis* and *trans*, respectively, associated with 87 plasma proteins. However, the inclusion of cellular covariates in the assessment of genetic associations led to the identification of three novel proteins with pQTLs, while

abrogated the signal for 10 proteins which would have otherwise led to positive hits. This could potentially be explained by the fact that all 13 proteins are expressed by specific circulating immune cell populations^{58–67}. However, the interactions between these proteins, blood cell populations and genetic variants are less obvious to interpret, as both direct and indirect effects or co-occurring mechanisms could be involved.

Although our study replicated a large number of previously reported genetic associations with plasma proteins^{21–23,32}, it also identified 24 proteins whose plasma levels had not been previously associated with genetic factors. This may stem from the well-defined healthy nature of our study population, which may reduce potential confounding lifestyle or medical factors, or from a low measurement error. Interestingly, of the newly identified associations, eight include proteins encoded by MMDGs, 3 of which are known to cause primary immunodeficiencies, *i.e.*: Ficolin-3, Interleukin-2 Receptor alpha and FAS (**Figure 6, Supplementary Note 2**).

Primary immunodeficiencies are caused by rare variants leading either to loss- or gain-of-function consequences in the affected genes^{68,69}. However, such mutations are often not fully penetrant and the associated symptoms are heterogeneous between and within families. Among possible explanations for this heterogeneity, low to mild effect common variants, such as the pQTLs identified in this work, might act as modifier of the corresponding diseases, by increasing or decreasing the expression of the corresponding proteins, and consequently mitigating or aggravating the consequences of causal variants.

Thus, the common variant associated with Ficolin-3 plasma levels identified in this work, rs2504780 (AF = 10.7%, 1:27710876, T>A), is located 9.5kb upstream of FCN3, and associated with a diminution of Ficolin-3 levels (size effect = -3.79 µg/mL per alternative allele, **Figure 6A, Supplementary Note 2**) comparable with heterozygous FCN3 loss-of-function variants⁷⁰ causing immunodeficiency 41 with lymphoproliferation and autoimmunity (OMIM 606367, **Supplementary Note 2**). This variant could be a risk factor for Ficolin-3 deficiency and might play a role in the observed etiology of both complete Ficolin-3 deficiency or Ficolin-3 haploinsufficiency in the response to infection and autoimmunity. Future analysis of auto-antibodies in our cohort may allow us to directly test this hypothesis. Another example are the two independent *cis*-pQTLs associated to interleukin 2 receptor alpha (IL2RA) levels, *i.e.* rs12722497 (6:6095928; C>A, AF = 10%) and rs2104286 (6:6099045; T>C, AF = 24%), both located in the first intron of *IL2RA*. *IL2RA*, expressed at the membrane of regulatory T cells, allows the control of the proliferation of responder T cells⁷¹. Complete *IL2RA* deficiency causes immunodeficiency, lymphoproliferation and auto-immunity^{19,72–74}. The rs12722497 polymorphism is associated with an increased plasma levels of IL2RA (size effect = +699 pg/mL per alternative allele; **Figure 6B left**), while rs2104286 associated with a decreased expression of IL2RA (size effect = -200 pg/mL per alternative allele; **Figure 6B right**). These two variants might affect the normal

behavior of regulatory T cells and alter their ability to modulate the proliferation of responder T cells, similarly to heterozygous *IL2RA* loss-of-function variant carriers⁷¹. Last, we identified a common variant associated in *trans* with FAS plasma levels, rs687621 (9:136137065, A>G, AF = 36%) which increases the expression of FAS (size effect = +2.79 ng/mL per alternative allele; **Figure 6C**). This variant could contribute to a protective role against the haploinsufficient forms of autoimmune lymphoproliferative syndrome (ALPS, OMIM 601859) by increasing the expression of FAS in heterozygous loss-of-function variant carriers. However, the rs687621 polymorphism is located at the *ABO* locus, which is known to associate with the expression of many plasma proteins^{22,24,41,42}. Such an association hotspot could be explained by the glycosyltransferase activity of ABO proteins⁷⁵, which by transferring glycosyl residuals on target proteins may potentially alter its binding affinity of the associated antibody in immunoassays⁷⁶, thus constituting a technical artefact. In light of these potential caveats, the biological relevance of the FAS-associated *trans*-pQTL identified should be taken with caution, prior to replication. The common genetic associations identified here for plasma protein levels of PID genes could be further characterized through genetic fine-mapping and functional characterization.

Plasma protein levels can be considered as end-of-chain signal integrators, and their levels are influenced by several molecular mechanisms (e.g. mRNA transcription, Kozak sequence affinity and other translation initiation mechanisms, codon usage, translation rate, post-transcriptional modifications⁷⁷⁻⁸¹). A combination of targeted genome and transcriptome sequencing, ribosome occupancy assay and intracellular protein assays in the tissue of interest would allow the identification of the causal variants and the molecular mechanisms mediating the observed associations. Finally, the phenotypic consequences of pQTLs associated to plasma levels coded by PID genes should be further characterized both in healthy and PID patients, where protective or modifier roles could be further established.

Methods

The Milieu Intérieur cohort

The 400 donors in this study were a subset of the 1,000 healthy donors of the Milieu Intérieur cohort recruited at BioTrial (Rennes, France). The Milieu Intérieur cohort was approved by the Comité de Protection des Personnes – Ouest 6 (Committee for the protection of persons) on June 13th, 2012 and by French Agence nationale de sécurité du médicament (ANSM) on June 22nd, 2012. The study is sponsored by Institut Pasteur (Pasteur ID-RCB Number: 2012-A00238-35), and was conducted as a single centre interventional study without an investigational product. The original protocol was registered under ClinicalTrials.gov (study# NCT01699893). The samples and data used in this study were formally established as the Milieu Intérieur biocollection (NCT03905993), with approvals by the Comité de Protection des Personnes – Sud Méditerranée and the Commission nationale de l'informatique et des libertés (CNIL) on April 11, 2018. Donors included in this sub-study were stratified by sex and were between the ages of 30-39 (n=200) or 60-69 (n=200) years old. Participants were selected based on stringent inclusion and exclusion criteria, as detailed elsewhere²⁸. To minimize the influence of population substructure on genomic analyses, the study was restricted to individuals of self-reported Metropolitan French origin for three generations (*i.e.*, with parents and grand-parents born in continental France). Fasting whole blood samples were collected in EDTA tubes and plasma was separated following high-speed centrifugation and stored at -80°C until analysis. Standard blood testing and complete hemogram was performed on fresh aliquots, while protein immunoassays were performed on frozen aliquots.

Quantification of plasma protein levels in 400 healthy individuals.

The protein immunoassays and the blood tests were performed on samples taken the same day and analyzed at different times, on fresh or frozen aliquots. Blood chemical and major cell fractions were estimated through direct enumeration and standard blood panels. The concentrations of 297 plasma proteins of 400 individuals were quantified by Luminex multi-analyte immunoassays (Discovery Map v3.3 from Myriad RBM), as previously described⁸². Protein levels were analysed and compared with their respective lower limit of quantification (LLOQ). Among the 297 assayed proteins, 68 proteins were reported at a concentration lower than the LLOQ in at least 20% of the individuals and were filtered out. For the 229 proteins that were kept, reported concentrations lower than the LLOQ were replaced by NAs, to prevent incorrect protein-environment or protein-genotype associations due to undetected or undetectable proteins. Next, for each protein, plasma level distribution across individuals were tested for normality using the Shapiro-Wilk test on the raw and log-transformed values within the 2.5% and 97.5% percentiles. Shapiro-Wilk null hypothesis was

not rejected (p -value ≤ 0.001 , after multiple testing correction) for a total of 50 (22%) and 183 (80%) proteins, depending on whether raw or log-transformed values were used. These results suggested that the majority of raw protein plasma levels followed a log-normal distribution and were thus log-transformed for downstream analysis.

Filters and tests for non-genetic variables

For each individual from the Milieu Intérieur cohort, an extensive electronic clinical record file was filled, gathering 754 lifestyle, environmental and medical history variables as well as blood metabolites and enzymes levels from standard blood test and erythrocytes enumeration²⁸. First, variables with names describing repetitive measurements over several visits after the first visit were filtered out. Second, redundant columns informative about the sex of the individual were removed. Third, mono-factorial variables, character variables, variables with missing values in 20% or more of the individuals, or varying in less than 10 individuals, variables correlated with other variables with a Spearman correlation coefficient of 1, and variables providing redundant information about the same phenotype were filtered out, for a final number of 254 variables. Then, for each of the 254 non-genetic factors, a univariate linear regression analysis was performed against the log-transformed expression levels of each of the 229 plasma proteins evaluated. Age and sex were systematically included as covariates in all such regressions, consistent with their pervasive influence shown in previous studies, as well as with their association with many of the non-genetic factors evaluated^{28,83}. Univariate linear regression was performed between each pair of proteins and non-genetic factors. In addition, to reduce the sensitivity of the linear models to outliers, the ten lowest and ten highest values of each protein were removed from the regression analysis. Significance was declared at p -value ≤ 0.05 after Bonferroni multiple testing correction accounting for the number of tests ($n = 58,166$). A total of 152 significant associations collectively involving 49 proteins and 20 non-genetic factors were found. In addition, seven major blood-cell fractions, *i.e.* leukocytes, lymphocytes, monocytes, neutrophils, eosinophils, basophils and platelets, were assessed through hemogram on fresh aliquots, along the other blood chemicals and enzymes²⁸.

Genotyping and imputation

Each individual from the Milieu Intérieur cohort was genotyped by the HumanOmniExpress-24 BeadChip (Illumina), covering 719,665 SNPs. 245,766 rare functional variants were also genotyped on a HumanExome-12 BeadChip (Illumina). After quality control, both datasets were merged, for a total of 723,341 SNPs. Next, IMPUTE v.2 was used to perform genotype imputation, on 1Mb windows buffered by

an additional 1Mb. Before imputation, SNPs were phased using 500 conditioning haplotypes, 50 MCMC, 10 burn-in and 10 pruning iterations. SNPs and allelic states were aligned to the imputation reference panel from the 1000 Genome Project Phase 1 v3 (2010/11/23). SNPs with dissimilar alleles (even after flipping) or ambiguous C/G or A/T alleles were filtered out. Imputation yielded a total of set of 37,895,612 SNPs. Removing SNPs with information metric ≤ 0.8 , duplicated or monomorphic SNPs, and SNPs with missingness $> 5\%$ (SNPs with genotyping probability lower than 0.8 in an individual were considered as missing) reduced the set to 11,395,554 SNPs. Further removing non-SNP variants and filtering out variants with a MAF < 0.05 (with the `--snps-only` option of PLINK v1.9) in the 400 sampled individuals, resulted in a final set of 5,201,100 SNPs. First and second components of a Principal Component Analysis of the OmniExpress array, were performed with reference populations⁸⁴.

Genome-wide association testing of plasma proteins

To perform the pQTL mapping of plasma proteins, we chose to use a multivariate approach by incorporating, for each protein, the associated non-genetic variables as covariates, in addition to sex, age, the 7 major blood cell fractions (leukocytes, lymphocytes, monocytes, neutrophils, eosinophils, basophils and platelets) and the two first PCs of the genetic data. If a non-genetic variable was a redundant measure with the corresponding protein (e.g. CRP), it was not added as a covariable in the model. We used a first linear mixed model to correct the protein expression levels for their specific covariates and for kinship, using per chromosome genetic relationship matrix (GRM) computed using GenABEL v1.8 (leaving one chromosome out). The analysis was performed separately for *cis* and *trans* QTLs, and the false discovery rate (FDR) was computed independently for *cis*-acting and *trans*-acting SNPs⁸⁵. *Cis*-acting SNPs were defined as SNPs located at a maximum distance of 1MB from the transcription start or end site of the corresponding gene. For each protein and each kind of associations tested (*cis* or *trans*), the minimal raw p-value was reported. In addition, for each protein, 100 permutations were performed between all *cis* or *trans* SNPs, and the minimal p-value of each of these permutations was extracted. Next, proteins were ascendingly sorted based on their raw p-values. Then, the FDR was computed, for each protein, as the mean over the N=100 permutations of the number of times its raw p-value is lower than the nth permutation from all proteins, divided by the rank of the corresponding protein. Protein-SNP pairs were considered as significant when the corresponding FDR was equal to or lower than 0.05. To investigate the potential presence of secondary pQTLs, we performed the same analysis a second time, incorporating the genotype of the most significant SNP detected in the first round of analysis as an additional covariate. The FDR was computed independently on each analysis iteration, both considering the same number of proteins and permutations. The significance thresholds

corresponding to first-round and second-round pQTLs were respectively around $2.2e-05$ and $1.5e-05$ for *cis* and respectively around $9e-10$ and $2e-10$ for *trans* (significance thresholds were computed as the mean between the last significant and the first non-significant p-values). The genome wide analysis yielded a total of 100 *cis*-pQTLs and 12 *trans*-pQTLs.

Contribution of non-genetic and genetic components in the variability of plasma proteins

The relative contribution of the various environmental and genetic variables was assessed using the correlation-adjusted marginal correlation score (CAR score⁸⁶) from the care package in R. The CAR score is the shrunk estimator of the adjusted coefficient of determination (R^2) of each independent variable in a linear model, which considers the marginal correlation between variables. The CAR score is determined for each independent variable within a model, representing their independent contribution to the total variability of the dependent variable. The sum of the CAR score attributed to each variable in a model is equal to the model adjusted R^2 . For each protein, the relative contribution of its significantly associated non-genetic variable was assessed at once. In case only a single variable was significantly associated with a protein, we used the adjusted coefficient of determination (R^2) of the variable as its relative contribution in the variability of the corresponding protein levels. Then, the relative contribution of the identified pQTLs was assessed by computing their CAR score in protein-specific models incorporating age, sex, the protein-specific covariates and the corresponding genotypes. The effect size of pQTLs was computed following a two-stage model similar to the GWAS. A first linear model was used to regress out the associated covariates (and previously identified SNP in the case of SNPs identified by the conditional analysis) from the log-transformed and non-transformed plasma protein levels, and a second linear model was used to regress the residuals of the first against the tested SNP. The Beta was extracted from this model and used as the SNP effect size.

Global impact of age and sex on plasma protein levels

In order to characterize the global impact of age and sex on plasma protein heterogeneity, while accounting for the collinearity of several plasma proteins, we performed a principal component analysis (PCA) on the expression levels of the 229 proteins across the 400 individuals. When considered independently, only PC1 and PC2 explained more than 5% of the total variability (**Supplementary Figure 1**).

Assessment of gene-gene interactions

The interactions between *trans*-pQTL associated proteins and candidate proteins coded by genes located within a 500kb window centered around the associated SNP were assessed using STRING-db v11⁴⁴ at

<https://string-db.org/>. All protein-coding genes were queried at once through the “Multiple proteins” option and default settings (organism: “Homo sapiens”). Proteins were considered to interact when they were shown to be direct neighbors in protein-protein interaction networks, or when one of the protein was shown to be directly or indirectly involved in the regulation of the other.

Contribution of blood cell fractions in protein level predictions

To quantify the relevance of blood cell fractions in the prediction of plasma protein levels, we used a one-way ANOVA to compare, for each pQTL, the predictions coming from two models. A first linear model considered the genotype of the corresponding SNP, the previously defined protein-specific covariates, as well as age, sex, the two first PCs of the genetic data, and the 7 blood-cell fractions. A second linear model was evaluated by considering all variables used in the previous one, with the exception of the 7 blood-cell fractions. The models including pQTLs obtained from the conditional analysis additionally corrected for the SNP used for their identification. To assess the potential relevance of the different circulatory cell fractions in the different results obtained from the two pQTL mapping, the proteins were first corrected for age and sex through linear regression, and the resulting residuals regressed against each circulatory cell counts individually. Association p-values were corrected independently for each protein, and are reported in **Supplementary Table 4**.

Replication of previously identified plasma protein QTLs.

We compared our significant SNP-protein pairs with four studies analyzing the genetic basis of plasma protein levels. Sun et al. reported 1927 pQTLs, resulting from the analysis of 3622 plasma proteins in 3301 individuals; Suhre et al. reported 539 pQTLs from the analysis of 1124 plasma proteins in 1000 individuals; Deming et al. reported 56 pQTLs from the analysis of 146 plasma proteins in 818 individuals; and Zhong et al. reported 144 pQTLs from the analysis of 107 plasma proteins in 101 individuals. The replication of the reported pQTLs was performed at the protein-level rather than at the SNP-level, due to the poor overlap in terms of sentinel SNPs reported in previous studies^{21–23,32} and our imputed set of 5,201,100 SNPs. A protein previously reported as *cis*-regulated (*i.e.* reported as significantly associated with a SNP annotated as “*cis*” by the authors) was considered replicated when it was significantly associated with a SNP located closer than 1 Mb around the gene extremities. The same consideration was applied to *trans*-regulated proteins with all SNPs further than 1Mb from the corresponding gene.

Protein and gene annotations

Proteins were classified as immune-related when they were either i) annotated as adaptive immune proteins (**Supplementary Table 2** from ⁸⁷), or innate immune proteins (**Supplementary Table 1** from ⁸⁷, and **Supplementary Table 1** from ⁸⁸) or ii) produced in sufficient concentrations in substantial fractions of immune cells, as described in ⁸⁹. The list of primary immunodeficiency genes was obtained from **Supplementary Table 3** from ⁸⁷. Gene-disease annotations were obtained from OMIM (downloaded at <https://www.omim.org/downloads/> on the 2019/06/10). Entries were parsed following ⁹⁰. Mendelian disease genes were selected for their level of supporting evidence equal to 3 and for not having a “somatic” flag, and monogenic mendelian disease genes (MMDGs) were further selected for not being flagged as “complex”. The list of secreted proteins was downloaded from UniProt (<https://www.uniprot.org/>) on the 2020/02/03, using the keywords: locations:(location:"Secreted [SL-0243]" type:component) AND organism:"Homo sapiens (Human) [9606]". The list of FDA approved targeted proteins was downloaded on the 2020/02/05 from <http://mrmassaydb.proteincentre.com/fdaassay/>.

The protein classes were taken from Myriad RBM Discovery Map V3.3 table. Proteins were considered enriched or depleted in a specific class when the proportion of proteins in that class was larger than in 10,000 randomly sampled set of proteins of the same size, coming from the tested set of 229 proteins. Previous reports of pQTLs associated with monogenic Mendelian disease genes, with primary immunodeficiency genes or with genes coding for FDA approved biomarkers were identified through QTLbase (<http://mulinlab.org/qtlbase>)⁴⁸.

Disease loci enrichment

To assess the potential association with diseases or other traits of the *cis* and *trans* pQTLs reported in this work was assessed, we used hits from the NHGRI-EBI Genome Wide Association Studies (GWAS) Catalog, downloaded on the 2019/03/22 (file name `gwas_catalog_v1.0-associations_e95_r2019-03-22.tsv`). SNPs associated with a trait or a disease with a reported p-value $\leq 1e-08$ and mapped to autosomes were kept. SNPs associated with traits containing “blood”, “plasma” or “serum” and “protein” were removed. A set of pQTLs was declared as enriched when the proportion of pQTLs in a set that were GWAS SNPs or in linkage disequilibrium with GWAS SNPs ($r^2 \geq 0.8$) was larger in 95% of 10,000 randomly sampled set of SNPs of the same size, matched by MAF (bins of 5%). Randomly sampled SNPs were drawn from 122,757 and 384,897 SNPs, selected respectively from the 1,674,134 and 5,201,100 SNPs tested for *cis* and *trans* associations respectively (with the `--indep-pairwise 100 5 0.5` function of PLINK v1.9). When several traits or diseases were associated with one locus, the most significant one was selected.

Functional annotation of pQTLs

The *cis*-pQTL molecular consequences were assessed using VEP v97. Each *cis*-pQTL was annotated based on the canonical transcript of the gene coding for the regulated protein.

Author: The Milieu Intérieur Consortium†.

† The Milieu Intérieur Consortium¶ is composed of the following team leaders: Laurent Abel (Hôpital Necker), Andres Alcover, Hugues Aschard, Philippe Bouso, Nollaig Bourke (Trinity College Dublin), Petter Brodin (Karolinska Institutet), Pierre Bruhns, Nadine Cerf-Bensussan (INSERM UMR 1163 – Institut Imagine), Ana Cumano, Caroline Demangel, Christophe d'Enfert, Ludovic Deriano, Marie-Agnès Dillies, James Di Santo, Françoise Dromer, Gérard Eberl, Jost Enninga, Jacques Fellay (EPFL, Lausanne), Ivo Gomperts-Boneca, Milena Hasan, Magnus Fontes (Institut Roche), Gunilla Karlsson Hedestam (Karolinska Institutet), Serge Hercberg (Université Paris 13), Molly Ingersoll, Rose Anne Kenny (Trinity College Dublin), Olivier Lantz (Institut Curie), Frédérique Michel, Hugo Mouquet, Cliona O'Farrelly (Trinity College Dublin), Etienne Patin, Sandra Pellegrini, Stanislas Pol (Hôpital Cochin), Antonio Rausell (INSERM UMR 1163 – Institut Imagine), Frédéric Rieux-Laucat (INSERM UMR 1163 – Institut Imagine), Lars Rogge, Anavaj Sakuntabhai, Olivier Schwartz, Benno Schwikowski, Spencer Shorte, Frédéric Tangy, Antoine Toubert (Hôpital Saint-Louis), Mathilde Touvier (Université Paris 13), Marie-Noëlle Ungeheuer, Christophe Zimmer, Matthew L. Albert (Insitro)§, Darragh Duffy§, Lluís Quintana-Murci§,

¶ unless otherwise indicated, partners are located at Institut Pasteur, Paris

§ co-coordinators of the Milieu Intérieur Consortium

Additional information can be found at: www.milieuinterieur.fr

Sample Information

Samples came from the Milieu Intérieur Cohort, which was approved by the *Comité de Protection des Personnes – Ouest 6* (Committee for the protection of persons) on June 13th, 2012 and by French *Agence nationale de sécurité du médicament* (ANSM) on June 22nd, 2012. The study is sponsored by Institut Pasteur (Pasteur ID-RCB Number: 2012-A00238-35), and was conducted as a single centre interventional study without an investigational product. The original protocol was registered under [ClinicalTrials.gov](https://clinicaltrials.gov) (study NCT01699893). The samples and data used in this study were formally established as the Milieu Intérieur biocollection (NCT03905993), with approvals by the *Comité de Protection des Personnes – Sud Méditerranée* and the *Commission nationale de l'informatique et des libertés* (CNIL) on April 11, 2018.

Data Access

The SNP array data that support the findings of this study have been deposited in the European Genome-Phenome Archive (EGA) with the accession code EGAS00001002460. Further data access is provided for research use only after review and approval by the Milieur Intérieur data access committee. Requests can be sent to milieuinterieurdac@pasteur.fr

Acknowledgments

This work benefited from support of the French government's Invest in the Future programme. This programme is managed by the Agence Nationale de la Recherche, reference ANR-10-LABX-69-01. The Clinical Bioinformatics Laboratory of the Imagine Institute was partly supported by the French National Research Agency (ANR) "Investissements d'Avenir" Program (Grant ANR-10-IAHU-01).

References

1. Belardelli, F. Role of interferons and other cytokines in the regulation of the immune response. *APMIS* **103**, 161–179 (1995).
2. Ray, S., Patel, S. K., Kumar, V., Damahe, J. & Srivastava, S. Differential expression of serum/plasma proteins in various infectious diseases: Specific or nonspecific signatures. *Prot. Clin. Appl.* **8**, 53–72 (2014).
3. Davie, E. W. Introduction to the blood coagulation cascade and cloning of blood coagulation factors. *J Protein Chem* **5**, 247–253 (1986).
4. Partridge, W. M. Plasma protein-mediated transport of steroid and thyroid hormones. *American Journal of Physiology-Endocrinology and Metabolism* **252**, E157–E164 (1987).
5. Partridge, W. M. Targeted Delivery of Hormones to Tissues by Plasma Proteins. in *Comprehensive Physiology* (ed. Terjung, R.) cp070114 (John Wiley & Sons, Inc., 2011). doi:10.1002/cphy.cp070114.
6. Qaid, M. M. & Abdelrahman, M. M. Role of insulin and other related hormones in energy metabolism - A review. *Cogent Food & Agriculture* **2**, (2016).
7. Dimou, E. & Nickel, W. Unconventional mechanisms of eukaryotic protein secretion. *Current Biology* **28**, R406–R410 (2018).
8. Rabouille, C. Pathways of Unconventional Protein Secretion. *Trends in Cell Biology* **27**, 230–240 (2017).
9. Zhao, K.-W., Murray, E. J. B. & Murray, S. S. HK2 Proximal Tubule Epithelial Cells Synthesize and Secrete Plasma Proteins Predominantly Through the Apical Surface. *J. Cell. Biochem.* **118**, 924–933 (2017).
10. Geyer, P. E., Holdt, L. M., Teupser, D. & Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol* **13**, 942 (2017).
11. Anderson, N. L. The Clinical Plasma Proteome: A Survey of Clinical Assays for Proteins in Plasma and Serum. *Clinical Chemistry* **56**, 177–185 (2010).

12. Heutinck, K. M., ten Berge, I. J. M., Hack, C. E., Hamann, J. & Rowshani, A. T. Serine proteases of the human immune system in health and disease. *Molecular Immunology* **47**, 1943–1955 (2010).
13. Jia, L. *et al.* An Attempt to Understand Kidney's Protein Handling Function by Comparing Plasma and Urine Proteomes. *PLoS ONE* **4**, e5146 (2009).
14. Amur, S., LaVange, L., Zineh, I., Buckman-Garner, S. & Woodcock, J. Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization. *Clin. Pharmacol. Ther.* **98**, 34–46 (2015).
15. Leth-Larsen, R., Lund, R. R. & Ditzel, H. J. Plasma Membrane Proteomics and Its Application in Clinical Cancer Biomarker Discovery. *Mol Cell Proteomics* **9**, 1369–1382 (2010).
16. Enroth, S. *et al.* High throughput proteomics identifies a high-accuracy 11 plasma protein biomarker signature for ovarian cancer. *Commun Biol* **2**, 221 (2019).
17. Rho, J. & Lampe, P. High-Throughput Analysis of Plasma Hybrid Markers for Early Detection of Cancers. *Proteomes* **2**, 1–17 (2014).
18. Wu, A. C. *et al.* Current Status and Future Opportunities in Lung Precision Medicine Research with a Focus on Biomarkers. An American Thoracic Society/National Heart, Lung, and Blood Institute Research Statement. *Am J Respir Crit Care Med* **198**, e116–e136 (2018).
19. Goudy, K. *et al.* Human IL2RA null mutation mediates immunodeficiency with lymphoproliferation and autoimmunity. *Clinical Immunology* **146**, 248–261 (2013).
20. Ojha, A., Ojha, U., Mohammed, R., Chandrashekar, A. & Ojha, H. Current perspective on the role of insulin and glucagon in the pathogenesis and treatment of type 2 diabetes mellitus. *CPAA Volume* **11**, 57–65 (2019).
21. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
22. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* **8**, 14357 (2017).
23. Alzheimer's Disease Neuroimaging Initiative (ADNI) *et al.* Genetic studies of plasma analytes identify novel potential biomarkers for several complex traits. *Sci Rep* **6**, 18092 (2016).
24. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat Commun* **9**, 3268 (2018).
25. Moriya, H. Quantitative nature of overexpression experiments. *MBoC* **26**, 3932–3939 (2015).
26. Piasecka, B. *et al.* Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. *Proc Natl Acad Sci USA* **115**, E488–E497 (2018).
27. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19 (2016).
28. Thomas, S. *et al.* The Milieu Intérieur study — An integrative approach for study of human immunological variance. *Clinical Immunology* **157**, 277–293 (2015).
29. Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med* **25**, 1843–1850 (2019).
30. Bjelosevic, S. *et al.* Quantitative Age-specific Variability of Plasma Proteins in Healthy Neonates, Children and Adults. *Mol Cell Proteomics* **16**, 924–935 (2017).
31. Enroth, S., Johansson, Å., Enroth, S. B. & Gyllensten, U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat Commun* **5**, 4684 (2014).

32. Zhong, W. *et al.* Whole-genome sequence association analysis of blood proteins in a longitudinal wellness cohort. *Genome Med* **12**, 53 (2020).
33. Doerstling, S., Hedberg, P., Öhrvik, J., Leppert, J. & Henriksen, E. Growth differentiation factor 15 in a community-based sample: age-dependent reference limits and prognostic impact. *Uppsala Journal of Medical Sciences* **123**, 86–93 (2018).
34. Shi, H., Seeley, R. J. & Clegg, D. J. Sexual differences in the control of energy homeostasis. *Frontiers in Neuroendocrinology* **30**, 396–404 (2009).
35. Manicourt, D.-H., Fujimoto, N., Obata, K. & Thonar, E. J.-M. A. Serum levels of collagenase, stromelysin-1, and timp-1. *Arthritis & Rheumatism* **37**, 1774–1783 (1994).
36. Svechnikov, K. & Söder, O. Ontogeny of gonadal sex steroids. *Best Practice & Research Clinical Endocrinology & Metabolism* **22**, 95–106 (2008).
37. Kontush, A. *et al.* Structure of HDL: Particle Subclasses and Molecular Components. in *High Density Lipoproteins* (eds. von Eckardstein, A. & Kardassis, D.) vol. 224 3–51 (Springer International Publishing, 2015).
38. Zewinger, S. *et al.* Apolipoprotein C3 induces inflammation and organ damage by alternative inflammasome activation. *Nat Immunol* **21**, 30–41 (2020).
39. Huttunen, R. & Syrjänen, J. Obesity and the risk and outcome of infection. *Int J Obes* **37**, 333–340 (2013).
40. Falagas, M. E. & Kompoti, M. Obesity and infection. *The Lancet Infectious Diseases* **6**, 438–446 (2006).
41. de Vries, P. S. *et al.* Whole-genome sequencing study of serum peptide levels: the Atherosclerosis Risk in Communities study. *Human Molecular Genetics* **26**, 3442–3450 (2017).
42. Ruffieux, H. *et al.* A fully joint Bayesian quantitative trait locus mapping of human protein abundance in plasma. <http://biorxiv.org/lookup/doi/10.1101/524405> (2019) doi:10.1101/524405.
43. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
44. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**, D607–D613 (2019).
45. Shih, S.-J. *et al.* Duplicated Downstream Enhancers Control Expression of the Human Apolipoprotein E Gene in Macrophages and Adipose Tissue. *Journal of Biological Chemistry* **275**, 31567–31572 (2000).
46. LifeLines Cohort Study *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet* **50**, 493–497 (2018).
47. Gillies, C. E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *The American Journal of Human Genetics* **103**, 232–244 (2018).
48. Zheng, Z. *et al.* QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Research* **48**, D983–D991 (2020).
49. Gass, J. *et al.* Progranulin regulates neuronal outgrowth independent of Sortilin. *Mol Neurodegeneration* **7**, 33 (2012).
50. Prudencio, M. *et al.* Misregulation of human sortilin splicing leads to the generation of a nonfunctional progranulin receptor. *Proceedings of the National Academy of Sciences* **109**, 21510–21515 (2012).
51. Jian, J. *et al.* Progranulin directly binds to the CRD2 and CRD3 of TNFR extracellular domains. *FEBS Letters* **587**, 3428–3436 (2013).

52. Rhost, S. *et al.* Sortilin inhibition limits secretion-induced progranulin-dependent breast cancer progression and cancer stem cell expansion. *Breast Cancer Res* **20**, 137 (2018).
53. Lee, W. C. *et al.* Targeted manipulation of the sortilin–progranulin axis rescues progranulin haploinsufficiency. *Human Molecular Genetics* **23**, 1467–1478 (2014).
54. The UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
55. Khan, Z. *et al.* Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures. *Science* **342**, 1100–1104 (2013).
56. Popadin, K. Y. *et al.* Gene Age Predicts the Strength of Purifying Selection Acting on Gene Expression Variation in Humans. *The American Journal of Human Genetics* **95**, 660–674 (2014).
57. Josephs, E. B., Stinchcombe, J. R. & Wright, S. I. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytol* **214**, 21–33 (2017).
58. Nichols, W. L. *et al.* von Willebrand disease (VWD): evidence-based diagnosis and management guidelines, the National Heart, Lung, and Blood Institute (NHLBI) Expert Panel report (USA). *Haemophilia* **14**, 171–232 (2008).
59. Bonnefoy, A. *et al.* Thrombospondin-1 controls vascular platelet recruitment and thrombus adherence in mice by protecting (sub)endothelial VWF from cleavage by ADAMTS13. *Blood* **107**, 955–964 (2006).
60. Patel, L. *et al.* Resistin is expressed in human macrophages and directly regulated by PPAR γ activators. *Biochemical and Biophysical Research Communications* **300**, 472–476 (2003).
61. Buckley, A. R. Prolactin, a lymphocyte growth and survival factor. *Lupus* **10**, 684–690 (2001).
62. Montgomery, D. W. Prolactin production by immune cells. *Lupus* **10**, 665–675 (2001).
63. Katakura, T., Miyazaki, M., Kobayashi, M., Herndon, D. N. & Suzuki, F. CCL17 and IL-10 as Effectors That Enable Alternatively Activated Macrophages to Inhibit the Generation of Classically Activated Macrophages. *J Immunol* **172**, 1407–1413 (2004).
64. Achuthan, A. *et al.* Granulocyte macrophage colony-stimulating factor induces CCL17 production via IRF4 to mediate inflammation. *Journal of Clinical Investigation* **126**, 3453–3466 (2016).
65. Hosomi, S. *et al.* CEACAM1 on activated NK cells inhibits NKG2D-mediated cytolytic function and signaling: Innate immunity. *Eur. J. Immunol.* **43**, 2473–2483 (2013).
66. Sarantis, H. & Gray-Owen, S. D. Defining the Roles of Human Carcinoembryonic Antigen-Related Cellular Adhesion Molecules during Neutrophil Responses to *Neisseria gonorrhoeae*. *Infect. Immun.* **80**, 345–358 (2012).
67. Kolla, V. *et al.* Carcinoembryonic cell adhesion molecule 6 in human lung: regulated expression of a multifunctional type II cell protein. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **296**, L1019–L1030 (2009).
68. Fischer, A. & Rausell, A. What do primary immunodeficiencies tell us about the essentiality/redundancy of immune responses? *Semin Immunol* **36**, 13–16 (2018).
69. Casanova, J.-L. & Abel, L. Human genetics of infectious diseases: Unique insights into immunological redundancy. *Semin Immunol* **36**, 1–12 (2018).
70. Munthe-Fog, L. *et al.* Immunodeficiency Associated with *FCN3* Mutation and Ficolin-3 Deficiency. *N Engl J Med* **360**, 2637–2644 (2009).

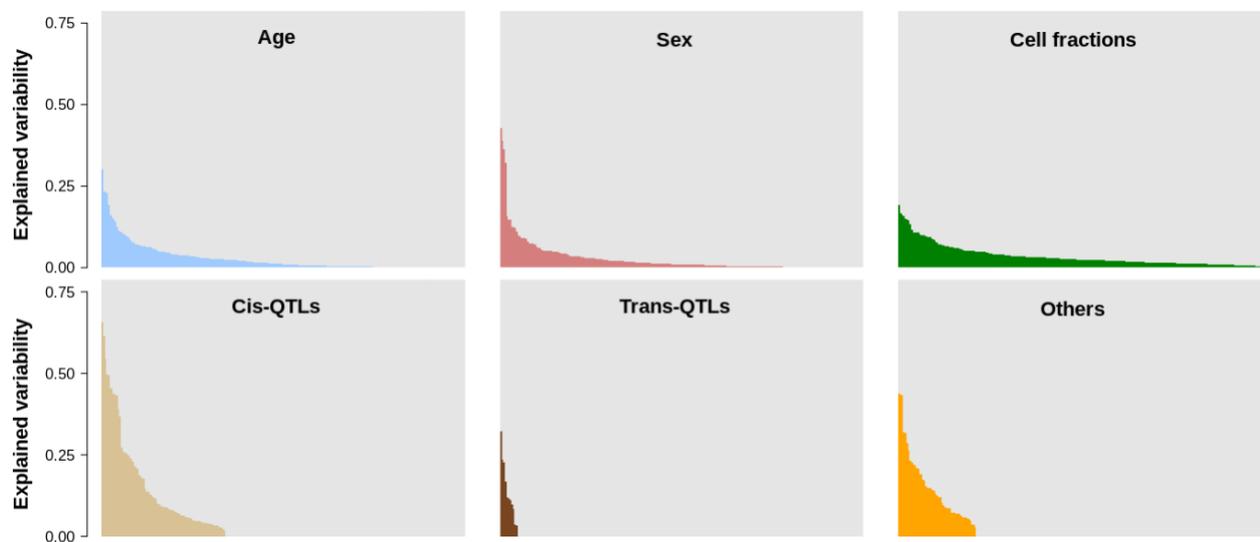
71. Roth, T. L. *et al.* Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **559**, 405–409 (2018).
72. Caudy, A. A., Reddy, S. T., Chatila, T., Atkinson, J. P. & Verbsky, J. W. CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked–like syndrome, and defective IL-10 expression from CD4 lymphocytes. *Journal of Allergy and Clinical Immunology* **119**, 482–487 (2007).
73. Bezrodnik, L., Caldirola, M. S., Seminario, A. G., Moreira, I. & Gaillard, M. I. Follicular bronchiolitis as phenotype associated with CD25 deficiency: CD25 deficiency. *Clin Exp Immunol* **175**, 227–234 (2014).
74. Sharfe, N., Dadi, H. K., Shahar, M. & Roifman, C. M. Human immune disorder arising from mutation of the chain of the interleukin-2 receptor. *Proceedings of the National Academy of Sciences* **94**, 3168–3171 (1997).
75. Qasim, A. N. & Reilly, M. P. Genetics of Atherosclerotic Cardiovascular Disease. in *Emery and Rimoin's Principles and Practice of Medical Genetics* 1–37 (Elsevier, 2013). doi:10.1016/B978-0-12-383834-6.00061-6.
76. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet* **22**, 19–37 (2021).
77. Acevedo, J. M., Hoermann, B., Schlimbach, T. & Teleman, A. A. Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Sci Rep* **8**, 4018 (2018).
78. Hanson, G. & Collier, J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* **19**, 20–30 (2018).
79. Collart, M. A. & Weiss, B. Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Research* **48**, 1043–1055 (2020).
80. Nedialkova, D. D. & Leidel, S. A. Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell* **161**, 1606–1618 (2015).
81. Nachtergaele, S. & He, C. The emerging biology of RNA post-transcriptional modifications. *RNA Biology* **14**, 156–163 (2017).
82. Duffy, D. *et al.* The ABCs of viral hepatitis that define biomarker signatures of acute viral hepatitis: DUFFY ET AL. *Hepatology* **59**, 1273–1282 (2014).
83. Michalski, M. *et al.* H-ficolin (ficolin-3) concentrations and FCN3 gene polymorphism in neonates. *Immunobiology* **217**, 730–737 (2012).
84. Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238–242 (2010).
85. Michalski, M. *et al.* Ficolin-3 activity towards the opportunistic pathogen, *Hafnia alvei*. *Immunobiology* **220**, 117–123 (2015).
86. Zuber, V. & Strimmer, K. High-Dimensional Regression and Variable Selection Using CAR Scores. *Statistical Applications in Genetics and Molecular Biology* **10**, (2011).
87. Fischer, A. & Rausell, A. Primary immunodeficiencies suggest redundancy within the human immune system. *Sci. Immunol.* **1**, eaah5861 (2016).
88. Deschamps, M. *et al.* Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. *The American Journal of Human Genetics* **98**, 5–21 (2016).
89. Rausell, A. *et al.* Common homozygosity for predicted loss-of-function variants reveals both redundant and advantageous effects of dispensable human genes. *Proc Natl Acad Sci USA* **117**, 13626–13636 (2020).

90. Caron, B., Luo, Y. & Rausell, A. NCBoost classifies pathogenic non-coding variants in Mendelian diseases through supervised learning on purifying selection signals in humans. *Genome Biol* **20**, 32 (2019).
91. Endo, Y., Matsushita, M. & Fujita, T. Role of ficolin in innate immunity and its molecular basis. *Immunobiology* **212**, 371–379 (2007).
92. Michalski, M. *et al.* Primary Ficolin-3 deficiency – Is it associated with increased susceptibility to infections? *Immunobiology* **220**, 711–713 (2015).
93. Schlapbach, L. J. *et al.* Congenital H-ficolin deficiency in premature infants with severe necrotising enterocolitis. *Gut* **60**, 1438–1439 (2011).
94. Hein, E. *et al.* Functional Analysis of Ficolin-3 Mediated Complement Activation. *PLoS ONE* **5**, e15443 (2010).
95. Barkai, L. J. *et al.* Decreased Ficolin-3-mediated Complement Lectin Pathway Activation and Alternative Pathway Amplification During Bacterial Infections in Patients With Type 2 Diabetes Mellitus. *Front. Immunol.* **10**, 509 (2019).
96. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414.e24 (2016).
97. Patsopoulos, N. A., the Bayer Pharma MS Genetics Working Group, the Steering Committees of Studies Evaluating IFN β -1b and a CCR1-Antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium & de Bakker, P. I. W. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol.* **70**, 897–912 (2011).
98. Risk Alleles for Multiple Sclerosis Identified by a Genomewide Study. *N Engl J Med* **357**, 851–862 (2007).
99. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118–1125 (2010).
100. Yamada, A., Arakaki, R., Saito, M., Kudo, Y. & Ishimaru, N. Dual Role of Fas/FasL-Mediated Signal in Peripheral Immune Tolerance. *Front. Immunol.* **8**, (2017).
101. Rieux-Laucat, F., Magérus-Chatinet, A. & Neven, B. The Autoimmune Lymphoproliferative Syndrome with Defective FAS or FAS-Ligand Functions. *J Clin Immunol* **38**, 558–568 (2018).
102. Magerus-Chatinet, A. *et al.* Onset of autoimmune lymphoproliferative syndrome (ALPS) in humans as a consequence of genetic defect accumulation. *J. Clin. Invest.* **121**, 106–112 (2011).
103. Kuehn, H. S. *et al.* FAS Haploinsufficiency Is a Common Disease Mechanism in the Human Autoimmune Lymphoproliferative Syndrome. *J.I.* **186**, 6035–6043 (2011).
104. Magerus-Chatinet, A. *et al.* FAS-L, IL-10, and double-negative CD4–CD8– TCR α/β + T cells are reliable markers of autoimmune lymphoproliferative syndrome (ALPS) associated with FAS loss of function. *Blood* **113**, 3027–3030 (2009).
105. Le Deist, F. *et al.* Clinical, immunological, and pathological consequences of Fas-deficient conditions. *The Lancet* **348**, 719–723 (1996).

Figures

Figure 1: Contribution of environmental and genetic factors to the variability of plasma protein levels. The variability explained by the associated genetic and non-genetic factors in the levels of 229 plasma protein levels, taken independently (A) or altogether (B). Each vertical bar represents the total variability of a protein, with the contribution of the considered (colored) or other and unknown (grey) factors summing up to 1. Age (blue), sex (brick-red), *cis* (light brown) and *trans* (dark brown) pQTLs correspond directly to the assessed relative importance, the cell fractions category (green) represents the cumulated relative importance of lymphocytes, leukocytes, neutrophils, eosinophils, basophils, monocytes and platelets, and the “Others” category (orange) represent the sum of the relative importance all other non-represented variables. In A, the grey area represents 1 minus the sum of all non-considered and unknown factors. In B, the “Uncharacterized” category was computed as 1 minus the sum of all other variables or groups of variables.

A



B

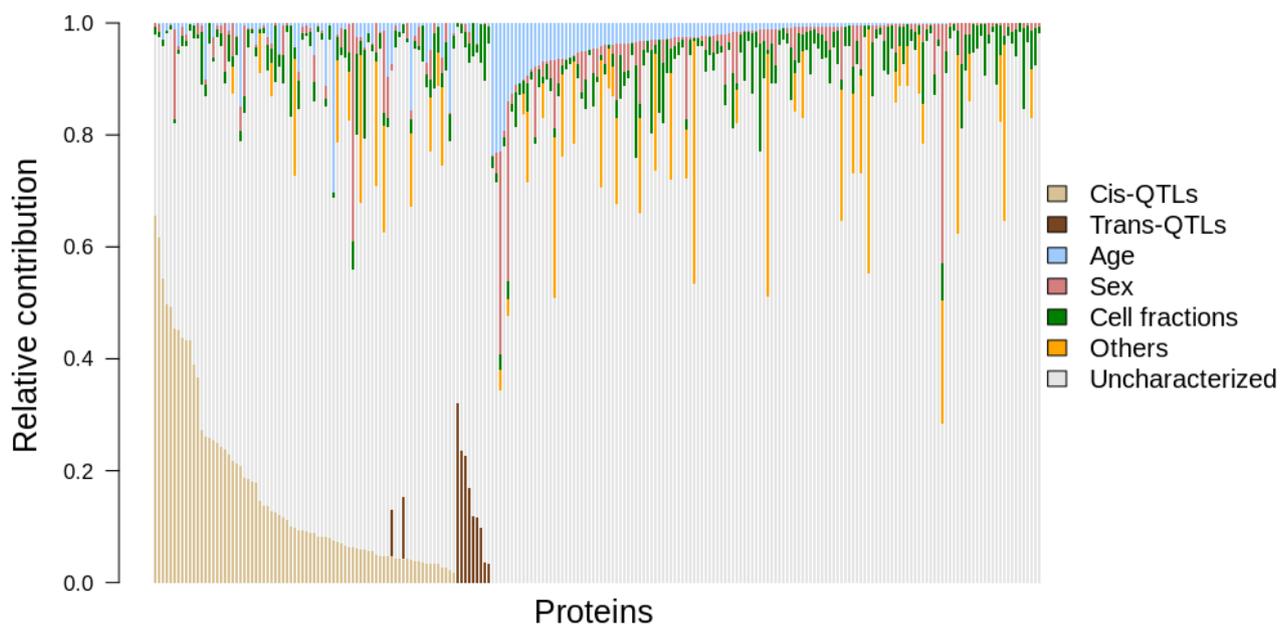


Figure 2: Relative contribution of selected factors in the variability of plasma protein levels

Relative contribution (CAR score⁸⁶) of the 20 significantly associated factors to the variability of plasma protein levels. Variables are sorted depending on the number of significant associations with proteins. The number of proteins significantly associated with each variable is reported on the x axis. The diameter of each dot represents the median CAR score of the corresponding factor in the variation of the associated plasma proteins.

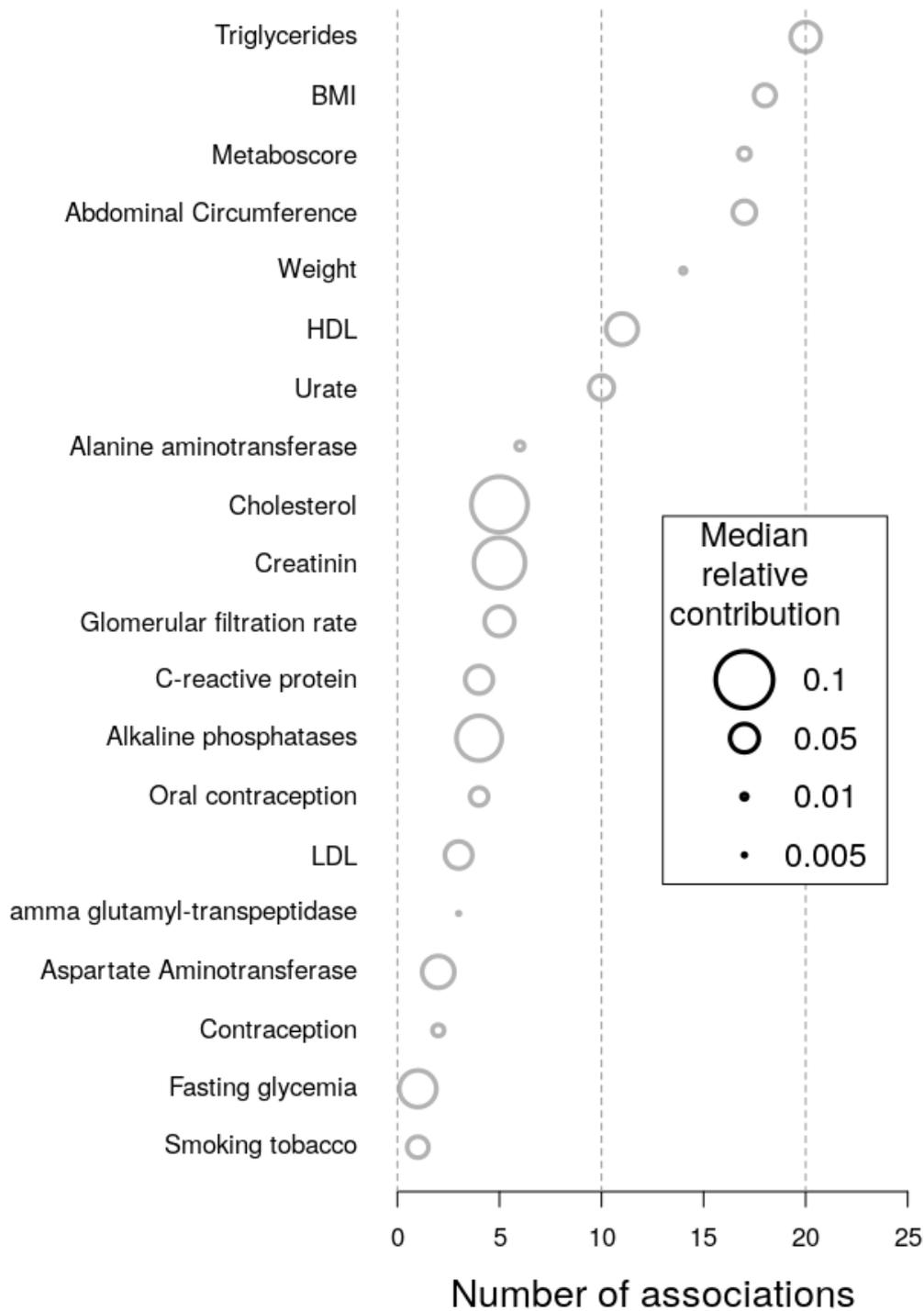


Figure 3: Plasma protein QTLs localization and co-regulation

The genomic positions of the *cis* (blue) and *trans* (orange) pQTLs identified in this work (x axis) and the genomic location of the gene coding for the associated protein (y axis). The point size is proportional to the uncorrected association p-value as reflected in the legend.

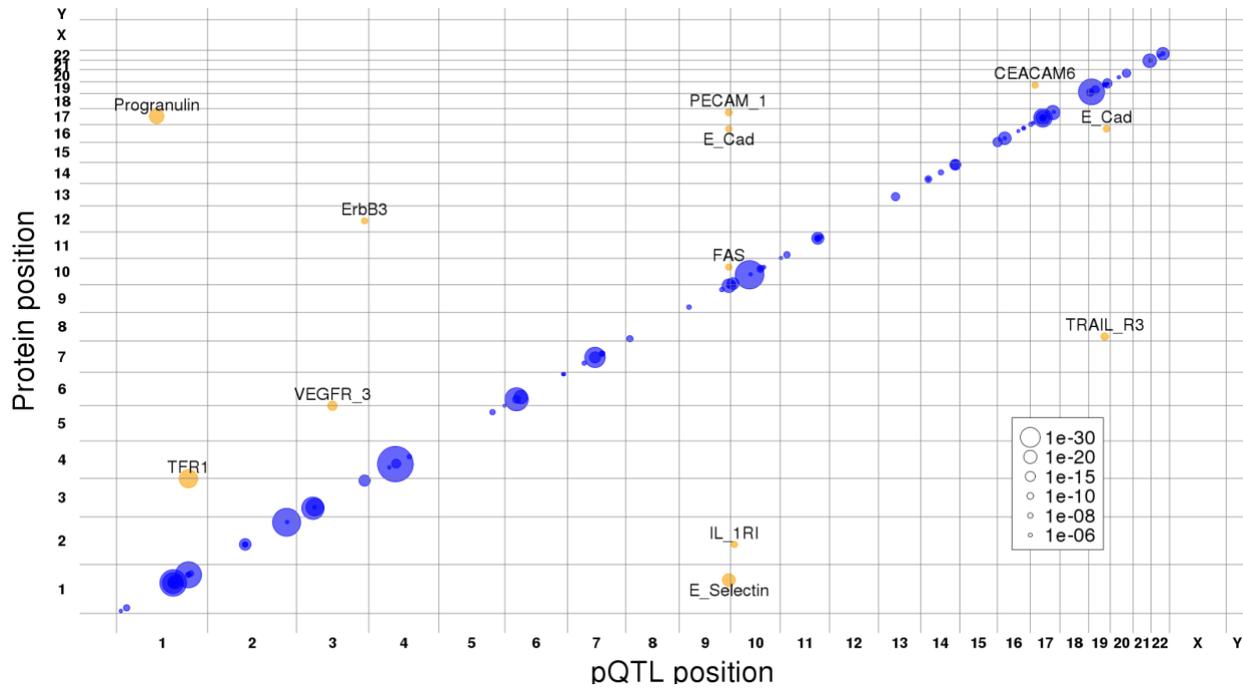


Figure 4: Replication of *cis*-pQTLs

The percentage of replication of previously reported *cis*-regulated proteins between our analysis and three previous studies: Sun et al., 2018²¹; Suhre et al., 2017²² and Deming et al., 2016²³. For each dataset, the percentage of replication (y-axis) as a function of the significance threshold (x-axis) was computed as the number of *cis*-regulated proteins reported in this work that were also reported in the corresponding dataset as *cis*-regulated (the “replicated” proteins) divided by the total number of proteins reported as *cis*-regulated in a previous study that were analyzed in our work (the “replicable” proteins). The dashed vertical lines represent the p-value significance threshold corresponding to the FDR of *cis*-pQTLs (red) and to the nominal replication threshold (orange).

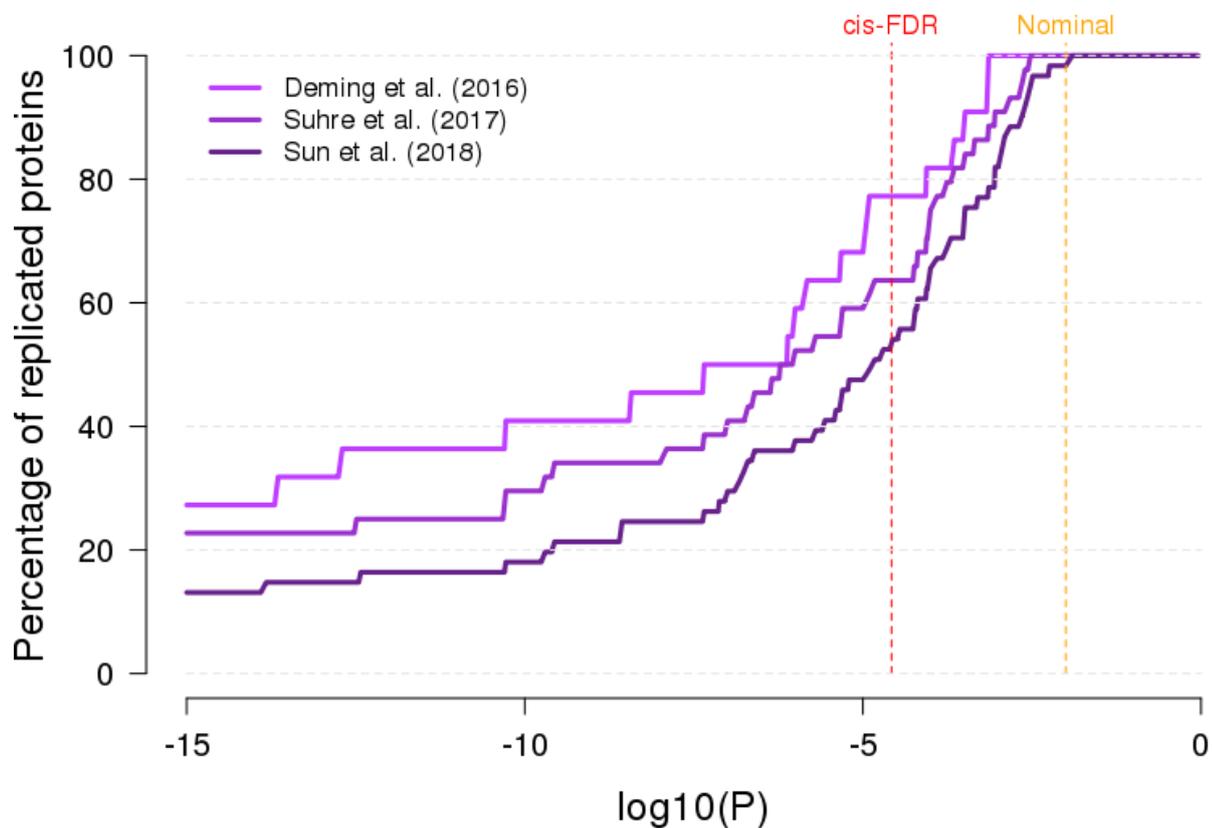


Figure 5: Clinical relevance of pQTLs and associated genes

Enrichment in GWAS hits (orange) of *cis*-pQTLs (**A** - left) and *trans*-pQTLs (**B** - left) in comparison with, respectively, 10,000 randomly sampled set of *cis*-SNPs matched by MAF (bins of 5%) (**A** - right) and 10,000 randomly sampled set of *trans*-SNPs matched by MAF (bins of 5%) (**B** - right). Empirical resampling p-values are shown.

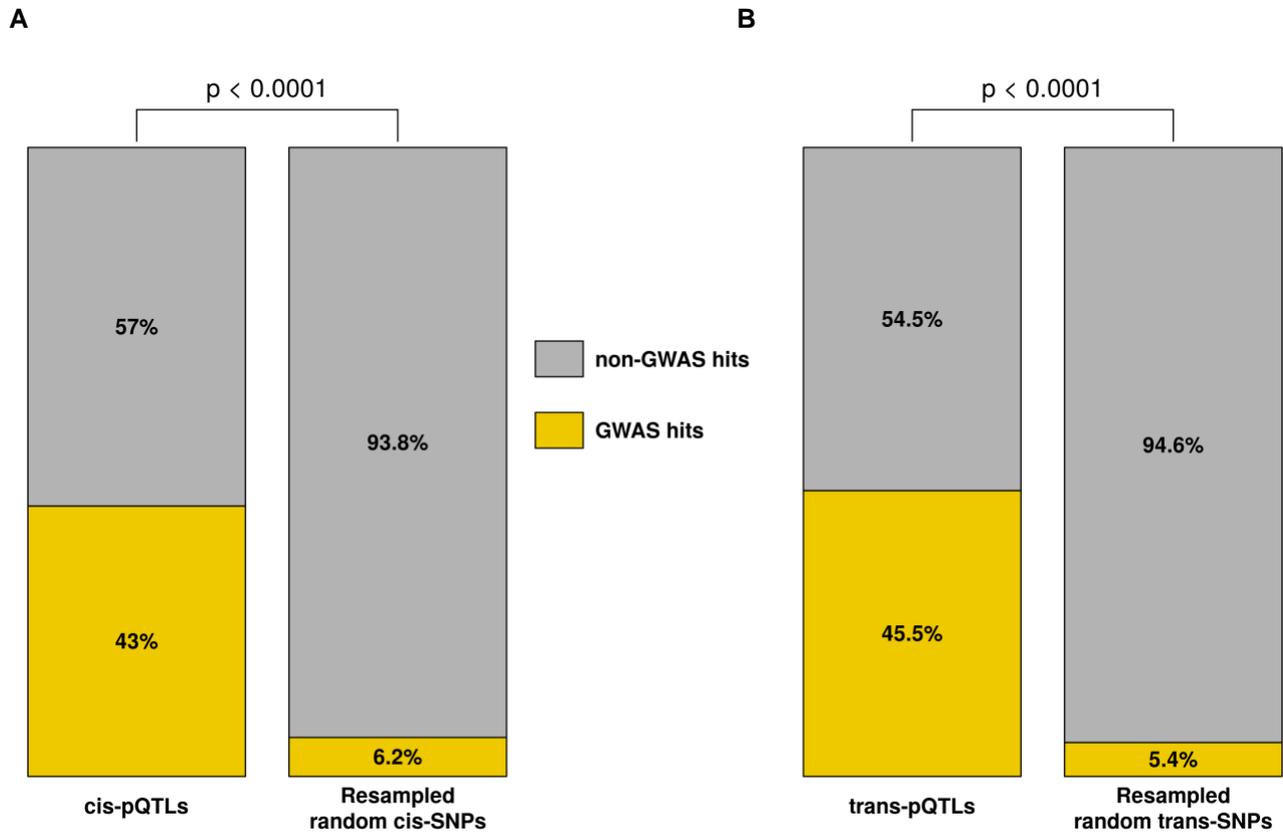
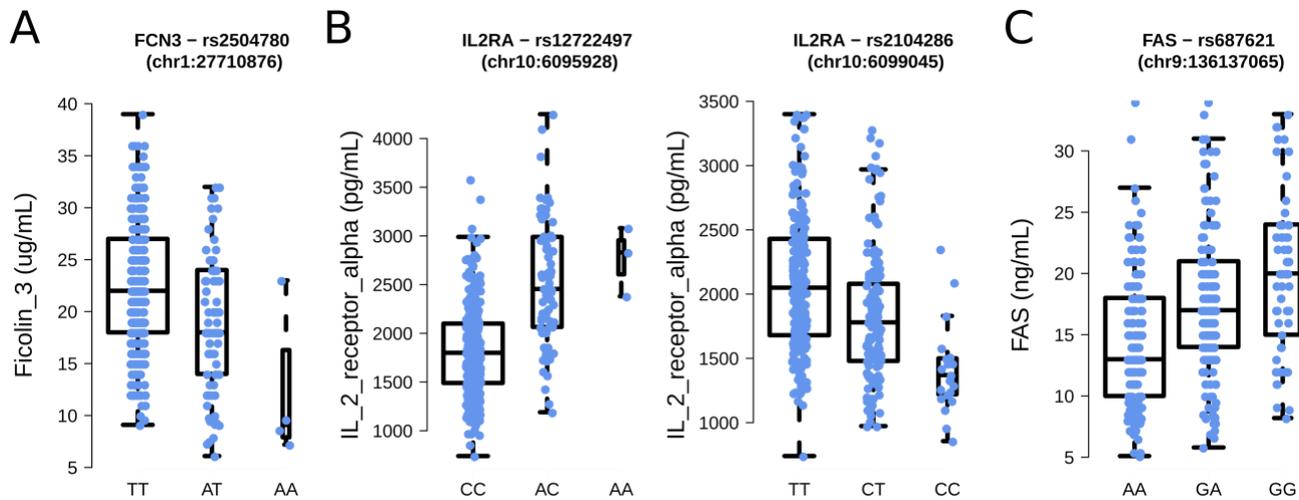


Figure 6: Impact of the novel pQTLs identified on the expression of the 3 target primary immunodeficiency genes

Expression levels of the two homozygous states and the heterozygous state of (A) Ficolin 3 x rs2504780, (B) Interleukine 2 receptor alpha x rs12722497 (**left**) and rs2104286 (**right**) and (C) Ficolin 3. Each dot corresponding to the non-transformed plasma levels of an individual.



Supplementary notes

Supplementary note 1. Genome-wide analysis of plasma proteins excluding blood-cell fractions

For the purpose of evaluating the contribution of blood-cell fractions in the pQTL assessment, the genome-wide association analysis was performed as previously detailed, while removing the 7 blood-cell fractions from the model (Methods). This approach identified 115 protein quantitative trait loci (pQTLs) collectively involving 94 proteins and 113 SNPs ($FDR \leq 0.05$). Among them, 103 were defined as *cis*-pQTLs. In addition, 12 pQTLs were identified in *trans*, *i.e.* located further than 1MB far from the gene boundaries, or located on another chromosome. 73 proteins were associated with only one SNP, while 21 were associated with two independent SNPs. Among the 94 proteins with significant pQTLs, 83 proteins were associated exclusively with *cis*-pQTLs, 9 exclusively with *trans*-pQTLs and 2 with both *cis* and *trans* pQTLs. In comparison with the first analysis, 92 pQTLs were reproduced, 81 in *cis* and 11 in *trans*. An association was considered as reproduced when the SNP, or a SNP in linkage disequilibrium with $R^2 \geq 0.8$, was significantly associated with the same protein at $FDR \leq 0.05$. 12 *cis*-pQTLs were no longer associating with the same SNPs or to SNPs in high LD ($R^2 \geq 0.8$) with it, but with other *cis* SNPs; and 8 pQTLs (7 *cis*, 1 *trans*) were not reproduced. On the opposite, 10 additional *cis*- and 1 *trans*- pQTLs were obtained.

Supplementary Note 2. Primary immune deficiencies caused by genes for which a novel pQTL association was identified in our study: potential role of pQTL as modifier variants.

Ficolin-3 autosomal recessive primary immuno deficiency (OMIM 613860)

Ficolin 3 is secreted from by the liver and the lungs⁸³, and, among other protein from the ficolin family, is responsible for the activation of the lectin pathway of the complement system⁹¹. A heterozygous frameshift mutation of FCN3, the gene coding for Ficolin-3, is found at a frequency of 1% in the healthy population, and leads to a 50% decrease in Ficolin-3 plasma levels, while homozygous state leads to complete Ficolin-3 deficiency. Ficolin-3 deficiency is an autosomal recessive primary immunodeficiency (OMIM 613860) associated with bacterial infections^{83,85,92,93}, and, as of today, only reported in males. It was shown that the activation of the Lectin pathway by Ficolin-3 was dose-dependent^{70,94}, and that lower activation of the Ficolin-3 lectin pathway was associated with post-infection mortality⁹⁵. The common variant (AF = 10.7%) identified in this work, rs2504780 (1:27710876, T>A), is located 9.5kb upstream of FCN3, and associated with a diminution of Ficolin-3 levels (size effect = -3.79 $\mu\text{g/mL}$ per alternative allele, **Figure 6A**) comparable with heterozygous FCN3 loss-of-function⁷⁰. Interestingly, this effect was substantially higher in male than in

females (size effect = -4.42 and -3.05 $\mu\text{g/ml}$ per alternative allele, respectively). This variant could be a risk factor for Ficolin-3 deficiency and might play a role in the observed etiology of both complete Ficolin-3 deficiency or Ficolin-3 lectin pathway reduced activation in the response to infection, and also in the development of autoimmune diseases caused by a reduced clearance of apoptotic cells. Future analysis of auto-antibodies in our cohort may allow us to directly test this hypothesis.

IL2RA deficiency and immunodeficiency 41 with lymphoproliferation and autoimmunity (OMIM 606367)

IL2RA is expressed at the surface of regulatory T cells (Tregs) and is a part of the Interleukin 2 receptor (CD25), which induce the growth and proliferation of Tregs. IL2RA expression follows a co-dominant pattern, where heterozygous missense or nonsense mutations found in healthy individuals lead to the expression of only half of IL2RA at the membrane of Tregs⁷². However, cases of homozygous or compound heterozygous missense and nonsense IL2RA variants were reported in immunodeficient patients, also affected by lymphoproliferation and autoimmunity (OMIM 606367)^{19,72-74}. Carriers of such variants did not express CD25 at the plasma membrane, which resulted in the failure of the patients' Tregs to respond to IL2 stimulation. Family members with heterozygous mutations were healthy, although their Tregs only presented half the amount of cell surface CD25, as compared with homozygous wild-type individuals, and showed a decreased ability to control the proliferation of responder T cells⁷¹. We identified two independent *cis*-pQTLs associated to interleukin 2 receptor alpha (IL2RA) levels, *i.e.* rs12722497 (6:6095928; C>A, AF = 10%) and rs2104286 (6:6099045; T>C, AF = 24%), both located in the first intron of IL2RA. rs12722497 associated with an increased expression of IL2RA (size effect = +699 pg/mL per alternative allele; **Figure 6B left**), contributing to 17.8% of the total observed variability of IL2RA plasma levels, and rs2104286 associated with a decreased expression of IL2RA (size effect = -200 pg/mL per alternative allele; **Figure 6B right**), further contributing to 7% of IL2RA variability. These variants might potentially alter the normal ability of regulatory T cells to control the proliferation of responder T cells, in a way similar to individuals carrying heterozygous *IL2RA* loss-of-function variants⁷¹. In addition, rs2104286 is in partial linkage disequilibrium ($R^2 = 0.55$) with rs12722489, a SNP previously reported as an eQTL and a pQTL for IL2RA^{21,96} and also as a susceptibility factor to multiple sclerosis^{97,98} and Crohn's disease⁹⁹. Interestingly, in our cohort, we previously reported rs12722497, but not rs2104286, as an eQTL after whole blood stimulation with whole microbes²⁶. Whether healthy microbiome species act on the same genetic variant to impact plasma levels of the IL2RA protein will be an interesting area for future study.

FAS deficiency and Autoimmune Lymphoproliferative Syndrome (ALPS, OMIM 601859)

FAS, a member of the tumor necrosis factor receptor superfamily, is expressed at the membrane of T cells and other immune cells. In normal conditions, the binding of FAS ligand (FASL) to FAS triggers apoptosis through the caspase cascade in a dose-dependent manner, preventing the accumulation of peripheral T cells in organs¹⁰⁰. Heterozygous missense or nonsense variants in FAS extracellular domain result in the expression of only half of FAS receptors at the cellular membrane, reducing the sensitivity of T cells to FASL binding and allowing the proliferation and the accumulation of auto-reactive CD4⁺- CD8⁻ T cells in organs. Several patients affected by the autoimmune lymphoproliferative syndrome (ALPS, OMIM 601859), characterized by autoimmunity, multi-lineage cytopenia and lymphadenopathy, have been shown to carry such heterozygous FAS mutations, causing the observed symptoms. However, these haploinsufficient mutations are not fully penetrant, as healthy relatives of the patients were identified^{101–105}, suggesting that epistatic factors could either prevent or promote the development of ALPS in heterozygous carriers. We identified a common variant, rs687621 (9:136137065, A>G, AF = 36%), contributing to 8.4% of FAS plasma protein level variability and associating with an increased expression of FAS (size effect = +2.79 ng/mL per alternative allele; **Figure 6C**). Our results raise the question of whether rs687621 could contribute to a protective role against ALPS by increasing the expression of the wild-type allele in heterozygous carriers and consequently increasing the sensitivity of peripheral T cells to FASL. However, the rs687621 polymorphism is located at the ABO locus, which is known to associate with the expression of many plasma proteins^{21,24,41,42}. Such an association hotspot could be explained by the glycosyltransferase activity of ABO proteins⁷⁵, which by transferring glycosyl residuals on target proteins may potentially alter its binding affinity of the associated antibody in immunoassays, thus constituting a technical artifact. In light of these potential caveats the biological relevance of the FAS-associated *trans*-pQTL identified should be taken with caution, prior to replication.