Eleven genomic loci affect plasma levels of chronic inflammation marker soluble urokinase-type plasminogen activator receptor

Soluble urokinase-type plasminogen activator receptor (suPAR) is a chronic inflammation marker associated with the development of a range of diseases, including cancer and cardiovascular disease. The genetics of suPAR remain unexplored but may shed light on the biology of the marker and its connection to outcomes. We report a heritability estimate of 60% for the variation in suPAR and performed a genome-wide association meta-analysis on suPAR levels measured in Iceland (N = 35,559) and in Denmark (N = 12,177). We identified 13 independently genome-wide significant sequence variants associated with suPAR across 11 distinct loci. Associated variants were found in and around genes encoding uPAR (PLAUR), its ligand uPA (PLAU), the kidney-disease-associated gene PLA2R1 as well as genes with relations to glycosylation, glycoprotein biosynthesis, and the immune response. These findings provide new insight into the causes of variation in suPAR plasma levels, which may clarify suPAR’s potential role in associated diseases, as well as the underlying mechanisms that give suPAR its prognostic value as a unique marker of chronic inflammation.
The plasma protein soluble urokinase-type plasminogen activator receptor (suPAR) is a non-specific biomarker for chronic inflammation (also termed low-grade inflammation) and was recently identified as a key molecule of senescent cells. It structurally consists of three domains (D_{1–D_{III}}) and is the soluble form of the membrane-bound receptor uPAR, which is bound to a variety of immune cells, smooth muscle cells, and podocytes by a glycosyl-phosphatidylinositol (GPI) anchor. suPAR is a receptor for urokinase-type plasminogen activator (uPA), an enzyme known for activating plasminogen into plasmin. Plasmin (uPA), an enzyme known for activating plasminogen into plasmin, is bound to a variety of immune cells, smooth muscle cells, and the soluble form of the membrane-bound receptor uPAR, which plays a role in various cellular processes including cell adhesion, migration, proliferation, angiogenesis, and chemotaxis. uPAR can be cleaved into its soluble form, suPAR, by several proteases, including uPA, GPI-specific phospholipase D, matrix metalloproteinases (MMPs), cathepsin G, neutrophil elastase, and plasmin.

In the general population, an elevated plasma suPAR level has been found to predict various health conditions, including incident cancer, cardiovascular disease, diabetes, depression, as well as early mortality. Elevated suPAR levels are also associated with pulmonary diseases including asthma and chronic obstructive pulmonary disease (COPD). It is also known that increased suPAR levels have strong associations with chronic kidney disease (CKD) across populations, and have been able to independently predict declining eGFR (estimated glomerular filtration rate) and incident CKD. Mouse models have indicated that suPAR may not only be associated with acute kidney injury, but may be causative in the development of this. In acute medically ill patients, increased suPAR is associated with readmissions and mortality, independent of clinical presentation, and suPAR is used in clinical routine in some European emergency departments for patient risk assessment.

Lifestyle factors associated with suPAR have been extensively studied. In particular, smoking has been found to be strongly associated with higher suPAR levels, and suPAR can be lowered by smoking cessation. Unhealthy diet, inactive lifestyle, and obesity have substantial impacts on suPAR levels in the general population. In addition, longitudinal research shows that multiple childhood risk factors (including exposure to adverse experiences, low IQ, and poor self-control) are associated with elevated suPAR in adulthood.

The genetics of suPAR remain unexplored but may shed light on the biology of the marker and its connection to outcomes. A recent genome-wide association analysis (GWAS) meta-analysis of the well-known inflammatory marker C-reactive protein (CRP) identified 58 associated genetic loci and consequently provided new insight into the genetic etiology of chronic inflammation. However, CRP and suPAR reflect different aspects of chronic inflammation despite both being used as inflammatory biomarkers. In addition, unlike CRP, suPAR is a stable biomarker as circadian changes in plasma suPAR are minimal, and suPAR measurements in individuals have been shown to be correlated across five and seven years. We aimed to investigate whether suPAR plasma levels are under the genetic influence and if so, identify associated genetic variants that may facilitate our understanding of suPAR’s biology and its links to associated diseases. Moreover, gaining new insight into suPAR through genetics may potentially improve the marker’s current prognostic capabilities.

Therefore, we performed a heritability analysis in a sample of British twins to estimate the genetic contribution to suPAR levels for the first time. We then performed a GWAS on suPAR levels in a general Icelandic population cohort and in a population of healthy Danish blood donors and combined these in a meta-analysis to identify genetic variants that affect this chronic inflammation marker’s plasma levels. Significant findings were followed-up in two independent cohorts, a sample from Great Britain and another from New Zealand. Furthermore, a pathway-based analysis as well as phenome-wide association studies (pheWASs) were performed to examine the suPAR-associated variants and their predicted genes further.

**Results**

**Heritability of suPAR.** We tested if variation in suPAR levels at age 18 years was genetically influenced in the Environmental Risk (E-Risk) Longitudinal Twin Study sample (Great Britain). Within-pair correlations for suPAR levels were $r = 0.69$ (95% CI: 0.65–0.73) for MZ twin pairs, and $r = 0.39$ (95% CI: 0.32–0.46) for DZ twin pairs. Using a univariate twin model, we found that additive genetic effects accounted for 60% (95% CI: 38–82%) of the variation in suPAR levels, while shared environmental influences accounted for 10% (95% CI: 0–31%) of the variance and nonshared environmental influences accounted for 30% (95% CI: 26–35%) of the variance in suPAR levels. We additionally calculated the SNP-based heritability based on the general Icelandic population cohort. The SNP heritability estimate was calculated to be 12.5% (SD: 4.8%).

**GWAS meta-analysis.** We performed GWASs on plasma suPAR levels in the general Icelandic population cohort ($N = 35,559$) as well as in the Danish Blood Donor Study (DBDS) ($N = 12,177$). We performed a meta-analysis of the two GWASs ($N = 47,736$) and employed a weighted Bonferroni adjustment to determine statistical significance as previously described. The $P$-value significance thresholds were $2.0 \times 10^{-8}$ for high-impact variants (including stop-gained, frameshift, splice-acceptor, or splice-donor variants, $N = 11,723$), $4.0 \times 10^{-8}$ for ‘moderate-impact’ variants (including missense, splice-region variants, and in-frame indels, $N = 202,336$), $3.7 \times 10^{-9}$ for ‘low-impact’ variants (including upstream and downstream variants, $N = 2,896,354$), and $6.1 \times 10^{-10}$ for the ‘lowest-impact’ variants (including intron and intergenic variants, $N = 37,239,641$). Our GWAS meta-analysis identified 13 independent genome-wide significant genetic variants associated with suPAR across 11 distinct loci in the genome (Fig. 1, Table 1, Supplementary Data 1). The variants were tested for heterogeneity between the two cohorts. Only one variant, the rs71311394 intron variant in $ST3GAL6$, shows evidence of heterogeneity at $P < 0.05$. However, the direction of effects for rs71311394 are consistent between the two cohorts and the association with suPAR levels is significant in each cohort ($\text{Effect}_{\text{M}} = 0.11$, $P_{\text{M}} = 5.3 \times 10^{-12}$; $\text{Effect}_{\text{D}} = 0.21$, $P_{\text{D}} = 2.1 \times 10^{-18}$). Two of the 13 genetic variants (rs114821641 and rs755902185 located in the $PLA2R1$/$LY75$ locus) were identified via conditional analysis using the Icelandic data exclusively, where linkage disequilibrium (LD) data is available for the same population.

**Comparison of genetic variants’ effects for suPAR unadjusted vs adjusted for smoking.** As suPAR levels have strong associations with smoking, we investigated whether smoking status would affect the outcome of the suPAR GWAS results. Using the Icelandic cohort, we performed two GWASs; the 30,469 individuals with available information on smoking status, unadjusted for smoking; and the same 30,469 individuals, adjusted for...
smoking. Smokers had higher suPAR levels than non-smokers (Effect = 0.12 SD, \( P = 2.2 \times 10^{-22} \) from \( t \)-test). A test of difference in the GWAS results between the two above-mentioned Icelandic GWASs revealed no difference when adjusting for smoking, with heterogeneity \( p \)-values ranging from 0.94 to 1.00 (Supplementary Data 2).

**Validation of variants from independent cohorts.** The lead genetic variants for each suPAR-associated locus were examined in two independent validation cohorts (Table 2 and Supplementary Data 3). We used a sample consisting of 837 individuals of white European-descent non-Maori descent from The Dunedin Multidisciplinary Health and Development Study (Dunedin) cohort, of which eight of the 13 variants were available for replication. A sample of 1444 E-Risk members of white European-descent was also used as a validation cohort, of which eight of the 13 variants were available. \( P \)-values < 0.05 were considered statistically significant in the validation phase. In the Dunedin cohort, five out of the eight available variants were confirmed and all eight variants had effect estimates in the same direction. In the E-Risk cohort, three out of the six available variants were confirmed and five variants had effect estimates in the same direction.

**Summary of suPAR-associated loci.** The following section lists the eleven loci in further detail, describing the variants’ predicted genes based on their position within a given gene or the closest gene, associations from previous studies, and selected annotated gene ontology (GO) biological processes for each gene.

**Chr1.q31.3.** The missense variant rs1061170-C on chromosome 1 in the gene CFH is associated with an increase in suPAR (effect estimate = 0.048 units of a standard deviation per copy increment in the effect allele; \( P = 2.79 \times 10^{-11} \), effect allele frequency (EAF) = 0.39). The variant causes a missense mutation (His > Tyr) in the gene CFH, which encodes the glycoprotein Complement Factor H; a protein that regulates complement activation in an immune response. The variant is known to be associated with age-related macular degeneration\(^{32,33} \). GO terms biological processes associated with the gene include complement activation, regulation of complement-dependent cytotoxicity, and viral process.

**Chr2.q24.2.** Three independently significant variants at a locus on chromosome 2 were found, resulting in two candidate genes in this locus. The missense variant rs3828323-C in the gene PLA2R1 (phospholipase A2 receptor 1) is associated with an increase in suPAR (effect = 0.118; \( P = 7.5 \times 10^{-65} \), EAF = 0.48). Several variants in PLA2R1 have previously been associated to membranous nephropathy, and serum anti-PLA2R1 antibody associates with loss of kidney function\(^{34,35} \). GO terms biological processes associated with the gene include cytokine production, negative regulation of phospholipase A2 activity, and receptor-mediated endocytosis among others.

The remaining two significant variants in this locus are less common and are located in the nearby gene LY75 encoding the protein lymphocyte antigen 75. rs114821641-T causes a stop-gain mutation and an increase in suPAR (effect = 0.400; \( P = 1.08 \times 10^{-13} \), EAF = 0.003), and rs755902185 is a deletion which causes a frameshift mutation in LY75 and an increase in suPAR (effect = 0.445; \( P = 1.05 \times 10^{-9} \), EAF = 0.0004). Variants in LY75 have previously been associated with Inflammatory Bowel Syndrome and Crohn’s disease\(^{36,37} \). GO terms biological processes associated with the gene include endocytosis, immune response, and inflammatory response.

**Chr3.q12.1.** The 3 prime untranslated region (UTR) variant rs71311394-G on chromosome 3 in the gene ST3GAL6 (ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 6) is associated with an increase in suPAR (effect = 0.137; \( P = 5.31 \times 10^{-26} \)). GO terms biological processes associated with the gene include glycolipid biosynthetic processes, cellular response to interleukin-6, and glycosylation among others\(^{38} \).

**Chr6.p21.32.** The deletion variant rs200185927 on chromosome 6 downstream from HLA-DRA is associated with an increase in...
Table 1 Summary statistics for the 13 independently genome-wide significant variants from the meta-analysis (N = 47,736).

| Variant ID | Chr     | Position (Hg38) | Effect allele | Other allele | Effect-allele frequency | Gene     | Variant type | Combined effect | Combined P-value | 95% CI          | Comb. P-value |
|------------|---------|-----------------|---------------|--------------|-------------------------|----------|--------------|-----------------|-----------------|----------------|---------------|
| rs1061770  | 1       | 1,196,690,107   | C             | T            | 0.39                    | CFH      | Missense    | 0.048           | 0.034, 0.062    | 2.79 × 10^{-11} | 0.034, 0.062 |
| rs1047132 | 1       | 1,149,821,641   | C             | T            | 0.017                   | PLAU     | Stop gained | 0.011           | 0.003, 0.019    | 3.73 × 10^{-12} | 0.003, 0.019 |
| rs3162317 | 2       | 1,599,955,647   | C             | T            | 0.01                     | LYS5     | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs1140912 | 2       | 1,598,648,976   | C             | T            | 0.012                   | LYS5     | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs7559021 | 2       | 3,701,082,959   | C             | T            | 0.012                   | ST3GAL6  | Missense    | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs7131139 | 3       | 9,879,376       | A             | G            | 0.013                   | ABO      | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs2001859 | 6       | 3,701,082,959   | C             | T            | 0.012                   | ST3GAL6  | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs7517984 | 9       | 13,325,759      | C             | T            | 0.013                   | ABO      | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs7559021 | 9       | 13,325,759      | C             | T            | 0.013                   | ABO      | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs2227566 | 10      | 7,391,359       | C             | T            | 0.012                   | ST3GAL4  | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs7952602 | 11      | 12,636,974      | C             | T            | 0.013                   | ST3GAL4  | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs3439293 | 18      | 27,137,930      | C             | T            | 0.013                   | ST3GAL4  | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |
| rs601338  | 19      | 48,703,417      | C             | T            | 0.013                   | ST3GAL4  | Frameshift  | 0.043           | 0.029, 0.064    | 5.38 × 10^{-6}  | 0.029, 0.064 |

Allele frequencies reported are based on European populations (Supplementary Data 1 reports the cohorts-specific effect allele frequencies). Table is ordered by genomic locus. Full table available in Supplementary Data 1.

In addition, we identified PLAU missense variant rs2302524-C to be independently significant in both the Danish and Icelandic GWASs, but with opposite effect directions (Effect_{ICE} = -0.24; P_{ICE} = 3.29 × 10^{-11}, EAF_{ICE} = 0.18; Effect_{DK} = 0.09, P_{DK} = 4.68 × 10^{-8}, EAF_{DK} = 0.17). Due to the inconsistent direction, it has not been included as a valid 14th suPAR-associated signal and hence is not shown in Table 1. Validation cohorts from New Zealand and Great Britain indicate that rs2302524 is significantly associated with increased suPAR levels as found in the Danish cohort (Supplementary Data 4). The

Chr9.q34.2. Intron variant rs75179845-C on chromosome 9 in the gene ABO is associated with an increase in suPAR (effect = 0.148; P = 8.37 × 10^{-27}, EAF = 0.10). The gene determines the blood group of the individual, and GO terms biological processes associated with the gene include lipid and protein glycosylation.

Chr10.q22.2. Splice-site variant rs2227566-C on chromosome 10 in the gene PLA2R1 is associated with a decrease in suPAR (effect = -0.124; P = 1.36 × 10^{-70}, EAF = 0.46). The variant is located in the splice region of PLA2R1, the gene encoding the protein uPA, which can bind to uPAR and consequently release the receptor into the bloodstream as suPAR. The protein uPA cleaves plasminogen to form the active form of plasmin and GO terms biological processes associated with this gene include blood coagulation, chemotaxis, fibrinolysis, regulation of smooth muscle cell-matrix adhesion, and migration among others. The lead variant rs2227566-C has been previously associated with asthma and airway hyperresponsiveness.

Chr11.q24.2. Intron variant rs7952602-C on chromosome 11 in the gene ST3GAL4 (ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 4) is associated with a decrease in suPAR (effect = -0.131; P = 2.33 × 10^{-34}, EAF = 0.14). GO terms biological processes of this gene include glycolipid biosynthetic processes and glycosylation among others.

Chr17.p13.1. Deletion variant rs758641530 on chromosome 17 in the gene ASGR1 (asialoglycoprotein receptor) is associated with an increase in suPAR (effect = 1.089; P = 1.66 × 10^{-132}, EAF = 0.004). ASGR1 is a lectin that mediates the endocytosis of plasma glycoproteins, and a previous study found that a deletion in this gene is associated with reduced levels of non-HDL cholesterol and a reduced risk of coronary artery disease.

Chr18.q11.2. Intron variant rs34392939 on chromosome 18 in the gene CHST9 (Carbohydrate Sulfotransferase 9) is associated with a decrease in suPAR (effect = -0.065; P = 1.58 × 10^{-18}, EAF = 0.30). A GO term biological process associated with this gene is proteoglycan biosynthetic process, and copy number variations of CHST9 associate with hematologic malignancies.
### Table 2 Results from the validation cohorts of the suPAR-associated variants.

| Variant ID | Chr | Position (Hg38) | EA | EAF | Gene | Comb. effect | Comb. P value | N | EA | EAF | N | EA | EAF | N | EA | EAF | N | EA | EAF |
|------------|-----|----------------|----|-----|------|-------------|--------------|---|----|-----|---|----|-----|---|----|-----|---|----|-----|
| rs12031707 | 1   | 1869403097     | C  | 0.38 | PLAU | 0.048 | 0.092 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs3828323  | 2   | 159951564      | C  | 0.092 | CFH  | 2.79E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs114821641| 2   | 159858447      | T  | 0.48  | PLA2R1| 7.50E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs755902185| 2   | 159864896      | A  | 0.0004| PLA2R1| 1.53E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs71311394 | 3   | 98793766       | G  | 0.46  | ST3GAL6| 1.05E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs717988452| 6   | 324494596      | A  | 0.06  | FUT2  | 1.03E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs573641344| 7   | 126387774      | C  | 0.14  | ASGR1 | 1.02E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs36229266  | 9   | 355362924      | T  | 0.0003| PLA2R1| 1.53E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |
| rs36229204  | 19  | 48703477       | G  | 0.05  | FUT2  | 1.03E−07 | 0.050 | 11  | 829 | C   | 0.37  | 0.07  | 0.092 | 0.050 | 1395 | C | 0.48 |

suPAR polygenic risk scores (PRs) and pheWASs. To investigate whether the combined effect of suPAR-associated genetic variations were associated with specific phenotypes, we performed pheWASs using suPAR PRSs as the exposure. We calculated PRSs for the individuals in the Icelandic population cohort based on the summary statistics from the Danish cohort’s suPAR GWAS. The PRs explained 0.94% of the suPAR variance in Icelandic individuals.

A total of 14,493 case/control phenotypes and 28,389 quantitative phenotypes in the Icelandic population cohort were tested. After Bonferroni multiple testing correction ($P < 7.86 \times 10^{-7}$), in the case-control pheWAS, we found that suPAR PRSs were associated with type 1 diabetes (effect = $2.212; P = 5.11 \times 10^{-18}$), autoimmune diseases (effect = $0.354; P = 4.30 \times 10^{-10}$) and obesity (effect = $0.634; P = 1.04 \times 10^{-7}$). In the quantitative phenotype pheWAS, we found that suPAR PRSs were associated with increased levels of plasma PLAU (effect = $0.649; P = 1.99 \times 10^{-10}$), increased levels of B12 (effect = $0.125; P = 7.30 \times 10^{-13}$), decreased high-density lipoprotein cholesterol (effect = $-0.107; P = 3.96 \times 10^{-10}$), increased fasting plasma glucose (effect = $0.089; P = 6.04 \times 10^{-9}$), increased alkaline phosphatase (effect = $0.100; P = 1.82 \times 10^{-8}$), increased potassium (effect = $0.085; P = 2.61 \times 10^{-9}$), increased BMI (effect = $0.109; P = 1.03 \times 10^{-7}$) and increased eosinophils (effect = $0.056; P = 7.01 \times 10^{-8}$) (Supplementary Data 6-7).
We additionally calculated PRSs for the individuals in the UK Biobank based on the summary statistics from the suPAR GWAS meta-analysis (Icelandic cohort + Danish cohort). 15,120 case/control phenotypes and 5,609 quantitative phenotypes were available in the UK Biobank pheWAS. No significant case/control phenotypes associated with suPAR after Bonferroni multiple test correction, but for the quantitative phenotypes, increased mean corpuscular hemoglobin (effect = 0.006; \( P = 2.40 \times 10^{-11} \)) and increased mean corpuscular volume (effect = 0.005; \( P = 3.80 \times 10^{-9} \)) were significant (Supplementary Data 8). Single-variant pheWASs were also performed for each of the 13 independently significant variants separately in both the Icelandic and UK Biobank dataset using the same methods and the significant results are available in Supplementary Data 9–12.

Mendelian randomization analyses for suPAR vs. the identified pheWAS findings. We performed Mendelian randomization analyses for suPAR and the identified pheWAS findings using phenotypes available in the Icelandic population cohort. No significant results were found and the removal of outliers detected using MR-PRESSO did not improve the estimates significantly (Supplementary Data 13).

Cis-expression quantitative trait loci (eQTL) analysis of the suPAR-associated variants. We tested if our variants were in high LD \( (r^2 > 0.8) \) with one more top cis-eQTL based on various tissues and 17 sources including GTEx and Icelandic data. Results are reported in Supplementary Data 14 and sources listed in Supplementary Data 15.

Genetic Correlation between suPAR and CRP. We performed a genetic correlation analysis between suPAR and CRP. The genetic correlation between suPAR and CRP was calculated to be 0.2351 (SD = 0.0533, \( P = 1.03 \times 10^{-5} \)), using suPAR measurements from the Icelandic/Danish meta-analysis and CRP measurements from UK Biobank. We have additionally checked for CRP variants from the GWAS-catalog for our suPAR-associated variants and their LD-classes (all variants with \( r^2 > 0.80 \)). Only the FUT2 variant (rs601338) was found to be an overlapping variant, where the same variant was reported in Han et al.\(^43\). We have additionally searched all the genes that we report to see if they have been reported with CRP, of which the only overlap we find in addition to FUT2 is ABO, where there is a variant at chr9:133266942 (rs643434) reported to associate with CRP in the CRP GWAS meta-analysis by Ligthart et al.\(^25\). This variant has \( r^2 = 0.17 \) with the ABO variant found in our study.

Discussion

Eleven genome-wide significant loci driven by 13 variants were associated with suPAR in our GWAS meta-analysis based on 47,736 individuals. These findings, including our heritability estimate of 60%, provide an updated explanation for the inter-individual variation in suPAR plasma levels.

We identified suPAR-associated variants localized in and near the genes encoding uPA \((PLAUR)\) and the ligand uPA \((PLAU)\), although how they affect suPAR plasma levels is uncertain. The lead variant rs36229204 at the PLAUR locus is in complete LD with rs4251805 \((r^2 = 1; D^2 = 1)\) which is located in PLAUR’s 5’ UTR, a region known for its importance in the regulation of a gene’s translation. Similarly, the lead variant rs2227566 at the PLAU locus is located in the splice region of the gene PLAU, but it is also in LD \( (r^2 = 0.44; D^2 = 1) \) with missense variant rs2227564-T (meta-analysis effect = -0.131; \( P = 1.57 \times 10^{-62} \), EAF = 0.26) which is the lead variant in the Danish cohort’s suPAR GWAS. This PLAU variant has previously been associated with an increased risk of colorectal cancer, asthma, oral tongue squamous cell carcinoma, and poor coronary collateral circulation in coronary artery disease patients, as well as Alzheimer’s Disease\(^39,46–49\). The variant causes a missense mutation in the kringle domain of uPA—the important domain necessary for protein–protein interactions including integrins\(^50\). The binding of uPA and suPAR is stabilized by the kringle domain\(^51,52\), suggesting that this missense mutation may produce a conformational change that reduces uPA’s ability to bind and/or to cleave uPAR, thereby reducing suPAR levels as seen in the results from this study.
Our results may also provide some insight into suPAR’s possible role in chronic and acute kidney disease. Significant variants were found in and near PLA2R1, where the top variant rs3828323 increases suPAR levels. As a missense variant, the resulting change from non-polar to polar residue may alter the binding of PLA2R1’s ligand PL2A (secretory phospholipase A2). The residue change is located in the extracellular part of the receptor between two C-type lectin domains, of which one is part of the receptor-binding region for its ligand PL2A. Interestingly, several variants in PLA2R1 have previously been associated to membranous nephropathy, and serum anti-PLA2R1 antibody associates with loss of kidney function. Two studies on membranous nephropathy have identified associated genetic variants, of which one found an association with two variants: rs17831251 as the lead associated variant. The two variants that are not reported in our study, i.e., rs35771982 and rs17831251 associate with suPAR in our meta-analysis but these associations do not hold up after adjusting for our lead variant (rs71311394). This would indicate that the signal observed in the previous membranous nephropathy genetic studies at the PL2A locus is the same as for suPAR in our study. It is well-reported that increased suPAR has strong associations with CKD, though it is unknown how suPAR and PLA2R1 are related to each other in respect to kidney disease. It is possible that when anti-PLA2R1 attaches to podocyte-bound PLA2R1 due to the altered structure of the receptor caused by the genetic variant, it forms immune complexes which consequently activates the immune system. As a result, the inflammatory response would increase suPAR levels and allow suPAR to activate beta3 integrins on the podocytes as shown in the previous studies, and this activation would produce the podocye conformational change that causes membranous nephropathy. It may therefore be hypothesized that it is a variant in PL2A2R1 that gives suPAR the appearance of a causal role in the development of membranous nephropathy and other kidney diseases previously shown to have strong associations with suPAR levels. Though rs3828323 in PL2A2R1 was the lead variant at this locus in our meta-analysis, it should be noted that two additional independent but rare variants were identified in this locus, located in the gene LYT75. The PL2A2R1 variant rs3828323 exists in an LD-class which overlaps into both genes. LYT75 and PL2A2R1 may therefore both be considered candidate genes in this locus. However, from the PRS-based quantitative phENAS, we found that a higher suPAR PRS was strongly associated with increased PL2A2R1 plasma levels (effect = 0.649; \( P = 1.99 \times 10^{-105} \)), further supporting PL2A2R1 as a gene of interest for future studies.

Another noteworthy observation from our study, supported by the results from the BiNGO pathway-based analysis, is that six of the suPAR-associated loci have significant variants in genes encoding proteins that are in some form involved in processes related to glycoprotein biosynthesis or protein glycosylation. These genes include ST3GAL4, ST3GAL6, ABO, CHST9, ASGR1, and FUT2. From the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, it is also evident that the proteins ST3GAL4, ST3GAL6, ABO and FUT2 participate in the glycosphingolipid biosynthesis (lacto and neolacto series) pathway. Glycosphingolipids are often localized in glycosphingolipid-enriched microdomains called lipid rafts where they have a role in mediating cell–cell interactions and regulating proteins in the same plasma membrane. Glycosphingolipids may therefore have a regulatory function on suPAR as it has previously been found that uPA-induced uPAR cleavage is strongly accelerated when uPAR is localized in lipid rafts. ASGR1 is a lectin that mediates the endocytosis of plasma glycoproteins which may also impact uPAR’s role on the plasma membrane. It is well known that uPAR is highly glycosylated but the function of the glycosylation is not completely understood. However, glycosylation has been found to increase uPAR’s affinity to uPA and enhance suPAR’s solubility. Though the results from our study suggest that uPAR glycosylation may have a molecular function that impacts suPAR plasma levels, the effects of glycosylation and deglycosylation on suPAR detection for the two methods used in this study are not known, and this may affect the results. Nevertheless, the results support previous studies that glycosylation increases uPAR-uPA affinity and suPAR solubility and therefore the genes identified in our study may be considered in future candidate gene studies to investigate their role in affecting the amount of suPAR present in plasma.

Two different suPAR detection methods were used in our study; a proteomics-based assay in the Icelandic population and ELISA in the Danish population. Although a direct comparison between the two different assays using the same samples was not possible, a comparison of the results of the GWAS findings between the Danish and Icelandic population cohorts was performed, of which the results showed high concordance. Only one variant, the rs71311394 variant in ST3GAL6, shows evidence of heterogeneity (\( P = 0.003 \)). However, the direction of effects for rs71311394 are consistent between the two populations (EffectIce = 0.11 vs. EffectDK = 0.21) and the association with suPAR levels is genome-wide significant in both populations (\( P_{\text{Ice}} = 5.3 \times 10^{-12} \) vs. \( P_{\text{DK}} = 2.1 \times 10^{-18} \)). Given this high degree of similarity in the GWAS findings we believe that a direct comparison between the assays would not add any further genetic insights into our findings.

As our findings indicate that individuals can be genetically predisposed to higher or lower suPAR levels, it may have implications for future precision or personalized medicine practices by potentially improving suPAR’s current prognostic capabilities. It is known that increased suPAR is associated with readmissions and mortality in acute medically ill patients, and suPAR is used in patient risk assessments in some Danish hospitals. Genetic profiling of patients may contribute to optimized patient treatment by identifying patients genetically predisposed to higher or lower levels of chronic inflammation, thereby enabling risk assessment of a patient at an earlier stage before they reach an acute medically ill stage of their disease progress. However, it must be noted that the SNP-based heritability based on the Icelandic cohort was calculated to be 12.5% (SD: 4.9%). As our univariate twin model estimated that additive genetic effects account for 60% of the variation in suPAR levels, we are unable to account for much of the heritability of suPAR, a situation frequently reported in GWASs of other phenotypes. Larger studies will likely enable the discovery of more variants that explain some of this missing heritability, as well as future studies focusing on other forms of genetic variation such as copy number variants.

In conclusion, we provide evidence that suPAR plasma levels are under the substantial genetic influence and that 13 independently significant genetic variants at 11 distinct loci influence suPAR levels in Icelandic and Danish individuals. Our data further support genetic links between suPAR-measured chronic inflammation and phenotypes such as diabetes and obesity. Our results indicate that suPAR’s strong associations with chronic kidney disease may be related to a suPAR-associated missense variant in the gene PL2A2R1, and that variants in many genes related to glycosylation and glycoprotein biosynthesis pathways affect suPAR levels. Genes identified in this study may be examined as candidate genes in future functional studies to help clarify suPAR’s potential role in the causation of associated diseases, as well as the underlying mechanisms that give suPAR its prognostic value as a unique marker of chronic inflammation.
Methods

Participants. The meta-analysis is based on data from two Northern European population cohorts: a Danish cohort consisting of healthy blood donors, and a general Icelandic population cohort.

Danish cohort. The Danish cohort is based on participants originating from the DBDS, a nationwide research platform utilizing the existing infrastructure in the Danish blood banks56. Participants must be generally healthy and not on medication or be eligible to donate. Upon enrollment, participants gave informed consent, whole blood, plasma, and answered a comprehensive questionnaire. So far, ~110,000 adult DBDS participants have been enrolled with informed consent, whole blood, plasma samples, questionnaire data, and genome-wide genotype data gathered from each. suPAR was measured in 14,367 consecutive DBDS participants from 1 March 2010 until 10 December 2010, of which 12,177 (84.8%) participated in the GWAS after fulfilling quality control requirements. The project is approved by the Research Ethics Committees by the following three protocols: The DBDS (M-20090257), Genetics of healthy ageing (CVK-1700407), Family study on the genetics of healthy ageing (NVK-1803847). The project is approved by the Danish Data Protection Agency under the combined approval for health care research at The Capital Region of Denmark (P-2019-99).

Icelandic cohort. Plasma samples from 40,014 Icelanders were collected during 2000–2019. Fifty-two percent of the samples were collected as part of the Icelandic Cancer Project (ICP), while the remaining samples (48%) were collected as part of various other studies; decode genetics, Reykjavik, Iceland. In the ICP, all prevalent and newly diagnosed Icelandic cancer cases and their relatives were invited to participate in a comprehensive study of cancer, along with a control population, randomly selected from the National Registry. The median collection date for each sample was 1 July 2002, whereas the median collection date for other samples was 15 May 2015. All samples were measured using the SOMAselect platform (SomaLogic), containing 5284 aptamers providing a measurement of relative binding of the plasma sample to each of the aptamers in relative fluorescence units (RFU), corresponding to 4792 proteins, of which suPAR is included. After quality control, unique measurements for N = 35,559 individuals (88.9%) were used for GWAS. All participants who donated samples gave informed consent and the National Bioethics Committee of Iceland approved the study (VSN-15-198) which was conducted in agreement with conditions issued by the Data Protection Authority of Iceland. Personal identities of the participants’ data and biological samples were encrypted by a third-party system (Identity Protection System), approved, and monitored by the Data Protection Authority.

Two independent cohorts agreed to validate the findings from this study: the Environmental Risk Longitudinal (E-Risk) Twin Study from Great Britain, as well as The Dunedin Longitudinal Study from New Zealand, of which the former was additionally used for the twin/heritability analysis.

Environmental risk (E-Risk) longitudinal twin study. Participants were members of the E-Risk longitudinal twin study, which tracks the development of a 1994–95 birth cohort of 2,232 British children60. Briefly, the E-Risk sample was constructed in 1999–2000, when 11,166 families (93% of those eligible) with same-sex 5-year-old twins participated in the first wave of assessments. The sample represents socioeconomic conditions in Great Britain, as reflected in the families’ distribution on a neighborhood-level socioeconomic index (nLACOR Index of Residential Neighborhoods, developed by CACI Inc. for commercial use); 25.6% of E-Risk families live in “wealthier achiever” neighborhoods compared to 25.3% nationwide; 5.3% vs. 11.6% live in “urban prosperity” neighborhoods because such households are often childless). Home visits were conducted when participants were aged 5, 7, 10, 12, and most recently, 18 years (93% participation). At age 18, each twin was interviewed by a different interviewer. Whole blood was collected from 82% (n = 1700) of the participants. Plasma was available for 1448 participants. The Joint South London and Maudsley, and the Institute of Psychiatry Research Ethics Committee approved each phase of the study. Parents gave informed consent and twins gave assent between 5 and 12 years old.

The Dunedin multidisciplinary health and development study. Participants were members of a study, a longitudinal investigation of health and behavior in a representative birth cohort. Participants (n = 1037; 91% of eligible birth; 52% male) were all individuals born between April 1972 and March 1973 in Dunedin, New Zealand (NZ), who were eligible based on residence in the province and who participated in the first assessment at age 3 years62. The cohort represented the full range of socioeconomic status (SES) in the general population of NZ’s South Island and as such enriched the NZ National Health and Nutrition Survey key adult health indicators (e.g., body mass index, smoking, GP visits) and the NZ Census of citizens of the same age on educational attainment. The cohort is primarily white (93%), matching South Island demographics63. Assumptions were carried out at birth and ages 3, 5, 7, 9, 11, 13, 15, 18, 21, 26, 32, and 38 years. At age 38 years, 95% (n = 961) of the 1007 participants still alive took part. At each assessment, each participant was brought to a research unit for testing. Blood from participants of Maori ancestry was not transported to Duke University for cultural reasons, and plasma samples were not available for participants who did not provide blood or due to phlebotomy or defrost cycle problems. The relevant ethics committees approved each phase of the study and written informed consent was obtained from all participants.

suPAR assessment. Plasma suPAR levels were measured in the DBDS cohort and two validation cohorts (Dunedin and E-Risk) using the CE/IVD-approved suPARnostic AUTO Flex ELISA (ViroGates A/S, Birkerød, Denmark) following the manufacturer’s instructions. The suPARnostic assay utilizes two monoclonal antibodies: a capture antibody directed towards the DIII subunit and a detection antibody against the DIII subunit. Full-length suPAR (DIDIII)DIDIII can be cleaved into DIII and DIDIII, and the assay can capture free full-length suPAR as well as the suPAR fragment (DIDIII)DIDIII but not the DIII fragment. The DIDIII full-length suPAR molecule can bind urokinase plasminogen activator (uPA) and DIII-DIDIII/uPA complexes will not be detected in the suPARnostic assay64. suPAR levels were measured in 14,367 participants in DBDS, 1444 in E-Risk, and 837 in the Dunedin cohort. suPAR levels have been described in previous publications64 whereas suPAR levels were measured at age 38 in the Dunedin Study, as previously described64. For the Icelandic cohort, suPAR is one of the plasma proteins measured using the SOMAselect platform as described above.

Genotyping and imputation. Genotyping and imputation of the 110,000 DBDS Genetic Cohort is described in Hansen et al65. Briefly, DNA purification is performed on the whole blood sample immediately stored in −80 °C. The samples were genotyped using the Global Screening Array by Illumina, which includes >650,000 variants with custom chip content optimized for comparison with the Illumina Omni Express chip. All genotype data are processed simultaneously for genotyping calling, quality control, and imputation. Quality control was conducted in both populations, including using a minimum allele count of 5, the exclusion of individuals or variants with more than 10% missingness, and individuals deviating more than three standard deviations (SDs) from the population heterozygosity (correcting for individuals carrying large copy number variations, >100 Kbp). We performed imputation using a reference panel backbone consisting of DIII and DIDIII reference sequences and a reference panel comprised of a subset of Illumina 1 Mbp probes with minor allele frequency (MAF) > 1%, and (2) an in-house dataset consisting of >6000 Danish whole-genome sequences to improve the prediction of variations with a MAF down to around 0.01%.

The process used to whole-genome sequence the 49,708 Icelanders, as well as the subsequent imputation, has been described in previous publications76. Briefly, we sequenced the whole genomes of 49,708 Icelanders using Illumina technology to a mean depth of at least 10x (median 32x). SNPs and indels were identified and their genotypes called using joint calling with Graphyper2. In total, 166,281 Icelanders were genotyped using Illumina SNP chips and their genotypes were phased using long-range phasing. All sequenced individuals were also chip- and long-range phased, providing information about haplotype sharing that was subsequently used to improve genotype calls. Genotypes of the 32 million high-quality sequence variants were imputed into all chip-typed Icelanders. Using genealogical information, the sequence variants were also imputed into relatives of the chip-typed further increasing the sample size for association analysis and the power to detect associations. All the variants tested had imputation information over 0.8.

For the two validation cohorts, we used Illumina HumanOmniExpress BeadChip arrays (Illumina CA, USA) to assay common single nucleotide polymorphism (SNP) variation in the genomes of participants of the E-Risk and Dunedin studies, as previously described74. The resulting dataset was restricted to SNPs called successfully in >98% of each cohort and in Hardy-Weinberg equilibrium (p > 0.001). Additional SNPs were imputed using the IMPUTE2 software (version 2.3.1, https://mathgen.stats.ox.ac.uk/impute/IMPUTE2.html)75 and 1000 Genomes Project (https://www.internationalgenome.org/). The resulting genotype dataset was included in the analysis conducted on autosomal SNPs appearing in dbSNP (version 140; http://www.ncbi.nlm.nih.gov/SNP/)77 that were called in >98% of each sample. Invariant SNPs were excluded. Prephasing and imputation were conducted using a 50 M base pair sliding window. The resulting genotype database included genotyped SNPs and SNPs imputed with a 90% probability of a specific genotype among European-descent E-Risk members (n = 1998 children in 1011 families) and among the non-Maori members of the Dunedin cohort (n = 918).

Twin/heritability analysis. To test the genetic contribution to suPAR levels at age 18 in the E-Risk study, we used a univariate twin model comparing correlations between MZ and DZ twins to decompose the phenotypic variation in plasma levels of suPAR into additive genetic, shared environmental, and unique environmental components. We used Mplus version 7.4 (Muthén and Muthén, Los Angeles, CA, USA) for the analysis. We additionally calculated SNP heritability based on the Icelandic cohort. We estimated the narrow-sense heritability of suPAR with LD score
regression\textsuperscript{39}, using an LD score map calculated with high-quality markers from the Icelandic population.

**GWAS statistical analysis.** The suPAR measurements were each rank-based inverse normal transformed to a standard normal distribution (separately for each sex) and adjusted for age using a generalized additive model. A linear mixed model implemented by BOLT-LMM\textsuperscript{39}, was used to test for association between a variant and suPAR levels, assuming an additive genetic model. Thirty-five million variants are tested in Iceland, while 26 million variants are tested in Denmark. BOLT-LMM accounts for cryptic relatedness and population stratification\textsuperscript{39}, and we additionally used LD score regression to account for distribution inflation in the dataset due to cryptic relatedness and population stratification\textsuperscript{39}. The Danish and Icelandic datasets were combined using a fixed-effect inverse-variance weighted meta-analysis, allowing the populations to have different frequencies for alleles and genotypes but assuming them to have a common effect. Heterogeneity in effect estimates was assessed using a likelihood-ratio test. Effects are given in units of SDs. In total 40 million variants are tested either in Iceland or Denmark, of which 21 million variants are tested in both datasets. Rare variants may therefore be present only in one discovery study.

We accounted for multiple testing by means of a weighted Bonferroni correction, taking into account the higher prior probability of association of certain variant annotations while controlling the family-wise error rate (FWER) at 0.05\textsuperscript{31}. The method has been described previously\textsuperscript{31} and results in stricter multiple testing correction than the commonly used threshold of $5 \times 10^{-8}$ (which would not control FWER at 0.05 given that 40 million markers were tested) while being more powerful than simply correcting for 40 million tests using a fixed threshold of 0.05/40,000,000 $= 1.25 \times 10^{-7}$. The resulting significance thresholds were 2.0 $\times 10^{-7}$ for high-confidence (HC), 0.5 $\times 10^{-7}$ for high-confidence + stringent, 1.0 $\times 10^{-7}$ for high-confidence + stringent + step-down, 1.0 $\times 10^{-10}$ for moderate-impact variants (including missense, splice-region variants, and in-frame indels, N = 11,723), 4.0 $\times 10^{-9}$ for 'low-impact' variants (including upstream and downstream variants, N = 2,896,354), and 6.1 $\times 10^{-11}$ for the 'lowest-impact' variants (including intron and intergenic variants, N = 37,239,641). To identify whether several variants in a single locus are independently associated with suPAR, we performed conditional analysis using Icelandic individual-level data, where LD data are available from the same population. This is in contrast with methods such as GCTA which use summary-level data only. The variants' predicted genes are based on their actual position within a given gene or the closest gene (from Ensembl variant effect annotations) and calculates the biological processes that are statistically overrepresented in our set of 12 suPAR-associated variants. To quantitatively investigate biological processes associated with each gene, we used the Biological Network Gene Ontology Analysis (BiNGO) bioinformatics tool\textsuperscript{44} to assess whether there are GO terms biological processes that are statistically overrepresented in our set of 12 suPAR-associated genes. The BiNGO tool uses GO terms from the Gene Ontology database (www.geneontology.org) and calculates the P values for overrepresentation of biological processes in our set of genes using the hypergeometric test. This takes into account both the total number of genes from the input dataset and the total number of genes for the specific GO term biological process. A total of 14,306 variants were tested in each of the five million UK Biobank subjects based on the meta-analysis summary statistics (Denmark + Iceland). Briefly, to generate the suPAR-PRS for the UK Biobank sample we used 63,000 informative variants across the genome and constructed locus allele-specific weightings by applying LDpred to the summary data from the subset meta-analysis GWAS\textsuperscript{39}. Constructing individual weightings, we were able to calculate an aggregated score of suPAR in all included individuals. Subsequently, we assessed the impact of suPAR-PRS on 63,609 traits (binary and quantitative) using a Bonferroni significance threshold of $P < 7.86 \times 10^{-7}$. More specifically, a pheWAS was performed in a comprehensive genotype dataset within the Icelandic population consisting of 14,493 case/control phenotypes and 28,389 quantitative phenotypes, as well as pheWAS in UK Biobank with 15,120 case/control phenotypes and 5609 quantitative phenotypes\textsuperscript{85}. In addition, single-variant pheWASs were performed in the same datasets using the same Bonferroni significance threshold.

**Mendelian randomization analyses for suPAR vs. the identified pheWAS findings.** We performed Mendelian randomization analyses for suPAR and the identified pheWAS findings using ukb-variants available in the Icelandic population cohort. The analyses were performed using the R Package "MendelianRandomization"\textsuperscript{84}, using the inverse variance weighted (IVW) and MR-Egger methods. The MR-PRESSO global test\textsuperscript{85} was additionally used to detect dependencies between instruments and remove them. In the instances where outliers were found, the outliers were removed and the IV method was reutilized. The instrumental variables (IVs) used are the variants we report significant for suPAR, i.e., the lead variants. Mendelian randomization for suPAR and 10 phenotypes were performed, including type 1 diabetes, autoimmune diseases as a general category, obesity, rheumatoid arthritis, B12, fasting glucose, alkaline phosphatase, and potassium.

**cis-eQTL analysis of the suPAR-associated variants.** We tested if our variants were in high LD ($r^2 > 0.8$) with one or more top cis-eQTL based on various tissues and 17 sources including GTEx and Icelandic data (See Supplementary Data 15 for list of sources). For the Icelandic data, RNA sequencing and estimation of the association between sequence variants and gene expression have been described in a recent publication\textsuperscript{86}.

**Genetic correlation between suPAR and CRP.** The genetic correlation between suPAR and CRP was calculated using suPAR from our Iceland/Denmark meta-analysis and CRP from UK biobank. We additionally investigated the overlap of findings between the suPAR-associated variants identified in our study and CRP-associated variants from the GWAS-catalog (https://www.ebi.ac.uk/gwas/). CRP-associated variants reported in the GWAS-catalog were checked for overlap with our suPAR-associated variants + LD-classes (all variants with $r^2 > 0.80$).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** The suPAR meta-analysis summary statistics will be made available at https://www.decode.com/summary-data/

Sequence variants passing GATK filters have been deposited in the European Variation Archive, accession number PRJEB15197.

For information on further access to data included in the meta-analysis, please contact the following authors of the respective cohorts: Hreinn Stefansson for data from the Icelandic cohort (hreinn.stefansson@decode.is) and Sisse Rye Ostrowski for data from the Danish Blood Donor Study (Sisse.Rye.Ostrowski@region.dk). The Dunedin Study data and E Risk study data are not publicly available but are available on request by qualified scientists. Requests require a concept paper describing the purpose of data access, ethical approval at the applicant’s institution, and provision for secure data access. Secure access is possible on the Duke University, Otago University, and King’s College London campuses. For UK Biobank please register on https://bbams.ndph.ox.ac.uk/sams/ and apply for the data through there.

**Code availability.** Variants in the Icelandic and Danish cohorts were imputed using software developed at Decode genetics based on the IMPUTE HMM model\textsuperscript{87} as previously described\textsuperscript{38}. A linear mixed model implemented by BOLT-LMM\textsuperscript{39} was used to test for association between sequence variants and suPAR levels. We used publicly available software (URLs listed below) in conjunction with the above described algorithms in the sequencing processing pipeline (Whole-genome sequencing, Association testing, RNA-seq mapping, and analysis):
ARTICLE

References

1. Wagner, V. G. & Gil, J. T cells engineered to target senescence. Nature 583, 37–38 (2020).
2. Thuno, M., Macho, B. & Eugen-Olsen, J. suPAR: the molecular crystal ball. Dis. Markers 27, 157–172 (2009).
3. Desmedt, S., Desmedt, V., Delanghe, J. R., Speeckaert, R. & Speeckaert, M. M. The intriguing role of soluble urokinase receptor in inflammatory diseases. Crit. Rev. Clin. Lab. Sci. 54, 117–133 (2017).
4. We used R extensively to analyze data and create plots.

Received: 7 August 2020; Accepted: 23 April 2021; Published online: 02 June 2021

BWA 0.7.10 mem, https://github.com/lh3/bwa
GenomeAnalysisTKLite 2.3.9, https://github.com/broadgsa/gatk/
Picard tools 1.117, https://broadinstitute.github.io/picard/
SAMtools 1.3, http://samtools.github.io/
Bedtools v2.25.0-g5675696, https://github.com/arq5x/bedtools2/
Variant Effect Predictor https://github.com/Ensembl/ensembl-vep
BOLT-LMM https://data.broadinstitute.org/alkesgroup/BOLT-LMM/downloads/
IMPUTE2 v2.31.1 http://www.sph.umich.edu/csg/abecasis/impute/impute_v2.html
dbsNP v140; https://www.ncbi.nlm.nih.gov/SNP/
LD Score Regression software; https://github.com/bulik/bdsc
BiNGO v3.0.3 https://www-ps.ugent.be/hd/papers/BiNGO/Download.html
Cytoscape v3.7.1 https://cytoscape.org/download.html
49. Wu, W. L., Jiang, H., Wang, M. Y. & Zhang, D. F. Meta-analysis of the association between urokinase-plasminogen activator gene rs2227564 polymorphism and Alzheimer’s disease. *Ann. J. Alzheimers Dis. Other Dement.* **28**, 517–523 (2013).

50. Tarui, T. et al. Direct interaction of the kringle domain of urokinase-type plasminogen activator. *Thromb. Haemost.* **95**, 522–534 (2006).

51. Appella, E. & Blasi, F. The growth-factor module of Urokinase is the binding sequence for its receptor. *Ann. N. Y. Acad. Sci.* **511**, 192–195 (1987).

52. Beder, K. et al. The kringle stabilizes urokinase binding to the urokinase receptor. *Blood* **102**, 3600–3608 (2003).

53. Xie, J. et al. The genetic architecture of membranous nephropathy and its potential to improve non-invasive diagnosis. *Nat. Commun.* **11**, 1690 (2020).

54. Maile, L. A. et al. Blocking ligand occupancy of the alphaVbeta3 integrin inhibits the development of nephropathy in diabetic pigs. *Endocrinology* **155**, 4665–4675 (2014).

55. Hayek, S. S. et al. A tripartite complex of suPAR, APOL1 risk variants and alphavbeta3 integrin on podocytes mediates chronic kidney disease. *Nat. Med.* **25**, 545–553 (2019).

56. Kanekisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

57. Schnaar, R. L. & Kinoshita, T. Glycosphingolipids. in *Niemann-Pick Disease, Type B* (eds. F. Chinnici et al.) 125-135 (Cold Spring Harbor Laboratory Press Copyright (2011)).

58. Cunningham, O. et al. Dimerization controls the lipid raft partitioning of uPAR/CD87 and regulates its biological functions. *J. Biol. Chem.* **285**, 29453–29460 (2010).

59. Behrendt, N. et al. The human receptor for urokinase plasminogen activator. NH2-terminal amino acid sequence and glycosylation variants. *J. Biol. Chem.* **265**, 6453–6460 (1990).

60. Ploug, M., Rahbek-Nielsen, H., Nielsen, P. F., Roepstorff, P. & Dano, K. Glycosylation profile of a recombinant urokinase-type plasminogen activator receptor expressed in Chinese hamster ovary cells. *J. Biol. Chem.* **273**, 13923–13931 (1998).

61. Møller, L. B., Pollanen, J., Ronne, E., Pedersen, N. & Blasi, F. N-linked glycosylation of the ligand-binding domain of the human urokinase receptor contributes to the affinity for its ligand. *J. Biol. Chem.* **268**, 11152–11159 (1993).

62. Gardsvoll, H., Werner, F., Sondergaard, L., Dano, K. & Ploug, M. Characterization of low-glycosylated forms of soluble human urokinase receptor expressed in Drosophila Schneider 2 cells after deletion of glycosylation-sites. *Protein Expr. Purif.* **54**, 284–295 (2004).

63. Blanco-Gómez, A. et al. Missing heritability of complex diseases: enlightenment by genetic variants from intermediate phenotypes. *BioEssays: N. Rev. Mol., Cell. Dev. Biol.* **38**, 664–673 (2016).

64. Bargarth, K. et al. The hospital questionnaire platform in the Danish Blood Donor Study. *Comput. Methods Programs Biochem.* **135**, 101–104 (2016).

65. Hansen, T. F. et al. DBDS Genomic Cohort, a prospective and comprehensive resource for integrative and temporal analysis of genetic, environmental and lifestyle factors affecting health of blood donors. *Bmj Open* **9** (2019).

66. Moffitt, T. E. Teen-aged mothers in contemporary Britain. *J. Child Psychol. Psychiatry* **43**, 727–742 (2002).

67. Poulton, R., Moffitt, T. E. & Silva, P. A. A. The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. *Soc Psychiatry Psychiatr. Epidemiol.* **50**, 679–693 (2015).

68. Chew-Harris, J., Appleby, S., Richards, A. M., Troughton, R. W. & Pemberton, C. J. Analytical, biochemical and clearance considerations of soluble urokinase plasminogen activator receptor (suPAR) in healthy individuals. *Clin. Biochem.* **69**, 36–44 (2019).

69. Rasmussen, I. et al. Cumulative childhood risk is associated with a new polymorphism and Alzheimer’s disease. *Nat. Genet.* **50**, 693–698 (2018).

70. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Nordisk Foundation (NNF1700027594). S.B. reports grants from Innovation Fund Denmark, grants from Novo Nordisk Foundation during the conduct of the study; and personal fees from Intomics A/S and Proscion A/S, outside the submitted work. S.B. and K. Stefánsson, and H.U. to perform Mendelian randomization analyses using summarized data. *Int J. Epidemiol.* **46**, 1734–1739 (2017).

71. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomics data. *Nature* **562**, 203–209 (2018).

72. Vayorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J. Epidemiol.* **49**, 122 (2020).

73. Venkataraman, K., Kao, C.-Y., Je, L. & Neale, B. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).

74. Sævarsdottir, S. et al. FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune thyroid disease. *Nature* **584**, 619–623 (2020).

75. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).

76. Gudbjartsson, D. F. et al. Sequence variants from whole genome sequencing a large group of Icelanders. *Sci. Data* **2**, 150011 (2015).

Acknowledgements

We thank participants in all the included cohorts. Similarly, we thank personnel employed in all blood banks across Denmark for making DBDS inclusion a part of their work routine. This study was supported by the Danish Council for Independent Research (09-096412 and 0602-026348), Aase og Eirik Danielsens Fond, AP Møller Fonden, and the Danish Regions (02/2611). The DBDS genetic infrastructure was supported by the Novo Nordisk Foundation (NNF1700027594). S.B. reports grants from Innovation Fund Denmark, grants from Novo Nordisk Foundation during the conduct of the study; and personal fees from Intomics A/S and Proscion A/S, outside the submitted work. S.B. and K. Stefánsson, and H.U. to perform Mendelian randomization analyses using summarized data. *Int J. Epidemiol.* **46**, 1734–1739 (2017).

Verbanck, M., Chen, C.-Y., Neale, B. & Do, D. O. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).

Sævarsdottir, S. et al. FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune thyroid disease. *Nature* **584**, 619–623 (2020).

Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).

Gudbjartsson, D. F. et al. Sequence variants from whole genome sequencing a large group of Icelanders. *Sci. Data* **2**, 150011 (2015).

Author contributions

J.D. contributed to the conception, design, analysis, interpretation of data and drafted the work. E.F., M.K.M., L.S., and B.G. contributed to the manuscript. All authors approved the submitted version of the manuscript.

Competing interests

The authors E.F., G.T., M.F., M.K.M., V.T., S.H.L., L.S., B.G., D.G., H.S., and K. Stefánsson are affiliated with deCODE genetics/AMGEN, declare competing interests as employees. J.E.O. is named inventor on patents on suPAR owned by Copenhagen University Hospital Hvidovre, Denmark. J.E.O. is co-founder, shareholder, and board member.
member of ViroGates A/S, Denmark, the company that developed the suPARnostic ELISA. ViroGates A/S had no role in the study design, data collection, analysis, or interpretation of study findings, and/or in the decision to submit the manuscript for publication. The remaining authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02144-8.

Correspondence and requests for materials should be addressed to J.D.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

DBDS Genomic Consortium

**Denmark** Steffen Andersen, Karina Banasik, Søren Brunak, Kristoffer Sølvsten Burgdorf, Christian Erikstrup, Thomas Folkmann Hansen, Gregor Jemec, Poul Jennum, Rene Kasper Nielsen, Mette Nyegaard, Helene Martina Paarup, Ole Birger Pedersen, Mikkel Petersen, Erik Sørensen, Henrik Ullum & Thomas Werge

**Iceland** Daniel Gudbjartsson, Kari Stefansson, Hreinn Stefánsson & Unnur Þorsteinsdóttir

16Department of Finance, Copenhagen Business School, Copenhagen, Denmark. 17Department of Clinical Medicine, Zealand University Hospital, Roskilde, Denmark. 18Department of Clinical Neurophysiology, University of Copenhagen, Copenhagen, Denmark. 19Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark. 20Department of Biomedicine, Aarhus University, Aarhus, Denmark. 21Department of Clinical Immunology, Odense University Hospital, Odense, Denmark. 22Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark. 23Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Copenhagen University Hospital, Roskilde, Denmark.