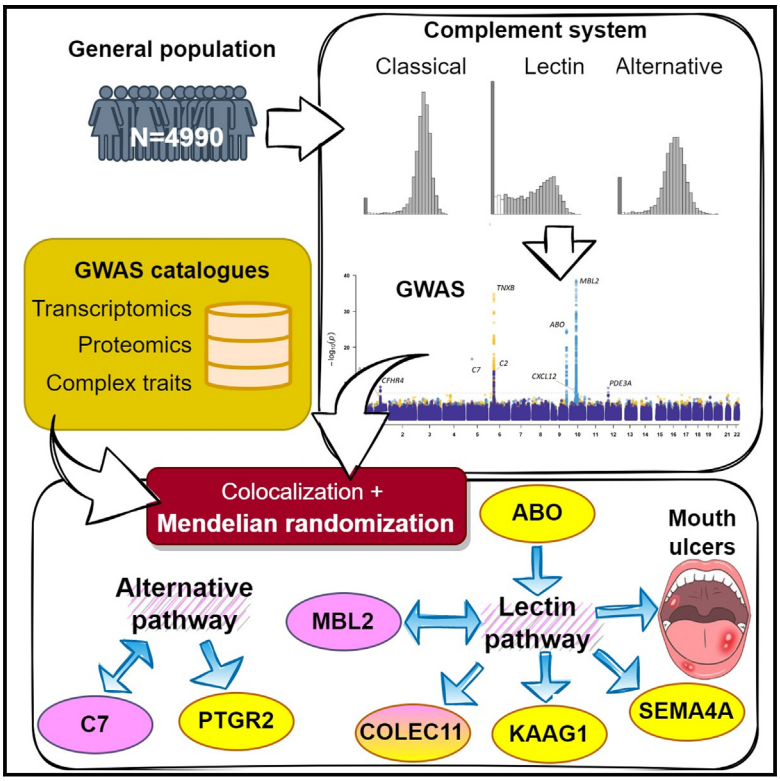
Cell Reports

Article

Genetic determinants of complement activation in the general population

Graphical abstract



Highlights

- Genome-wide association study of the functional activity of the three complement pathways
- Identified genetic loci are pathway specific, not shared between pathways
- ABO, KAAG1, SEMA4A, and PTGR2 proteins are causally involved with complement pathways
- The lectin pathway is causally associated with mouth ulcer risk

Authors

Damia Noce, Luisa Foco, Dorothea Orth-Höller, ..., Florian Kronenberg, Reinhard Würzner, Cristian Pattaro

Correspondence

florian.kronenberg@i-med.ac.at (F.K.), reinhard.wuerzner@i-med.ac.at (R.W.), cristian.pattaro@eurac.edu (C.P.)

In brief

Noce et al. conduct a general-population genome-wide association study of the functional activity of complement pathways. Identified loci explain a large genetic heritability fraction, enabling detection of causal relations with proteins and establishing a causal role of the lectin pathway on mouth ulcer risk.

Noce et al., 2024, Cell Reports *43*, 113611
January 23, 2024
https://doi.org/10.1016/j.celrep.2023.113611

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Genetic determinants of complement activation in the general population

Damia Noce,[1,2,9] Luisa Foco,[1,9] Dorothea Orth-Höller,[2,3,9] Eva König,[1] Giulia Barbieri,[1,4] Maik Pietzner,[5,6] Dariush Ghasemi-Semeskandeh,[1,8] Stefan Coassin,[7] Christian Fuchsberger,[1] Martin Gögele,[1] Fabiola Del Greco M.,[1] Alessandro De Grandi,[1] Monika Summerer,[7] Eleanor Wheeler,[6] Claudia Langenberg,[5] Cornelia Lass-Flörl,[2] Peter Paul Pramstaller,[1] Florian Kronenberg,[7,10,*] Reinhard Würzner,[2,10,*] and Cristian Pattaro[1,10,11,*]

[1]Institute for Biomedicine (affiliated to the University of Lübeck), Eurac Research, Via Volta 21, 39100 Bolzano, Italy
[2]Institute of Hygiene & Medical Microbiology, Department of Hygiene, Microbiology and Public Health, Medical University of Innsbruck, Schöpfstr. 41, 6020 Innsbruck, Austria
[3]MB-LAB – Clinical Microbiology Laboratory, Franz-Fischer-Str. 7b, 6020 Innsbruck, Austria
[4]Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy
[5]Computational Medicine, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Berlin, Germany
[6]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
[7]Institute of Genetic Epidemiology, Medical University of Innsbruck, Schöpfstr. 41, 6020 Innsbruck, Austria
[8]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands
[9]These authors contributed equally
[10]Senior author
[11]Lead contact
*Correspondence: florian.kronenberg@i-med.ac.at (F.K.), reinhard.wuerzner@i-med.ac.at (R.W.), cristian.pattaro@eurac.edu (C.P.)
https://doi.org/10.1016/j.celrep.2023.113611

SUMMARY

Complement is a fundamental innate immune response component. Its alterations are associated with severe systemic diseases. To illuminate the complement's genetic underpinnings, we conduct genome-wide association studies of the functional activity of the classical (CP), lectin (LP), and alternative (AP) complement pathways in the Cooperative Health Research in South Tyrol study (n = 4,990). We identify seven loci, encompassing 13 independent, pathway-specific variants located in or near complement genes (*CFHR4*, *C7*, *C2*, *MBL2*) and non-complement genes (*PDE3A*, *TNXB*, *ABO*), explaining up to 74% of complement pathways' genetic heritability and implicating long-range haplotypes associated with LP at *MBL2*. Two-sample Mendelian randomization analyses, supported by transcriptome- and proteome-wide colocalization, confirm known causal pathways, establish within-complement feedback loops, and implicate causality of ABO on LP and of CFHR2 and C7 on AP. LP causally influences collectin-11 and KAAG1 levels and the risk of mouth ulcers. These results build a comprehensive resource to investigate the role of complement in human health.

INTRODUCTION

Complement is an essential, evolutionarily ancient[1] part of the innate immune system, with transversal roles across all human tissues.[2] Altered complement activation is associated with a broad spectrum of human diseases of both syndromic and complex nature, including autoimmune, neurodegenerative, metabolic, cardiovascular, and infectious diseases such as malaria[2] and COVID-19.[3]

The main canonical functions of complement are immune complex clearance, chemotaxis for recruitment of inflammatory cells, opsonization and phagocytosis of foreign particles, and cell lysis.[4] Complement activation occurs via three distinctly operating pathways: the classical pathway (CP), the mannose-binding lectin pathway (LP), and the alternative pathway (AP). The three pathways converge, generating the complement component C3 convertases, which lead to the formation of the terminal complement complex (TCC) or the membrane-attack complex (MAC) when formed on a cell surface.[4] Complement proteins operate in plasma, body tissues, on the cell surface, and even within the cell.[5] To maintain host defense while avoiding host damage, complement activation must be tightly controlled by several soluble and membrane-bound regulators, such as factor H, CD46, CD55, and CD59.[4]

The biological mechanisms underlying complement activation have been studied for decades, but knowledge of the degree of complement activation in the general population is limited.[6] Genome-wide association studies (GWASs) have revolutionized genetic and biological knowledge in many areas of human health.[7] Some GWASs have been conducted on individual proteins involved in complement activation[8–10] or their ratio,[11] also exploiting multiplex proteomic platforms.[12,13] However, no hypothesis-free investigation has been conducted on the three individual complement activation pathways to date. One reason for this is that to date, no good functional readouts for the three pathways have been available on a large scale.

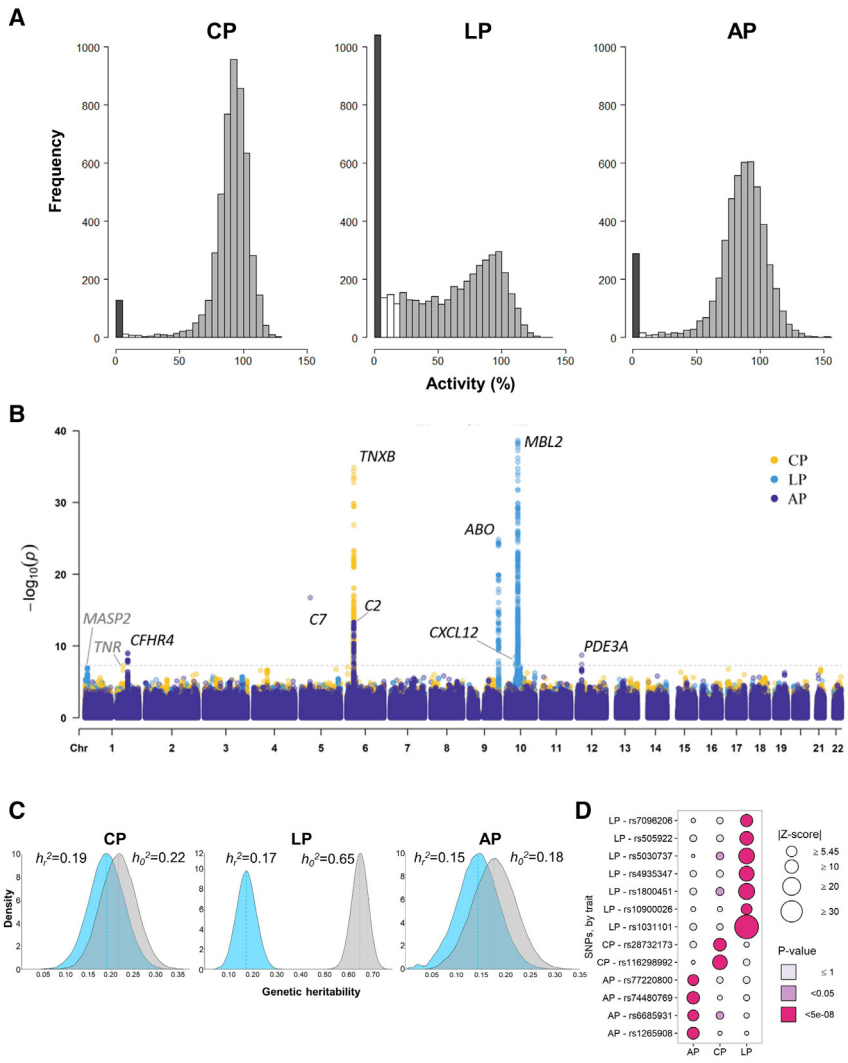


Figure 1. GWAS of the classical, lectin, and alternative pathways

(A) Distribution of classical pathway (CP), lectin pathway (LP), and alternative pathway (AP) in the study sample (Q1, 1st quartile; Q3, 3rd quartile). For CP, we observed 127 individuals with <5% activation (dark gray), 27 with activation between 5% and 20% (white), and 4,836 with >20% activation (gray). For LP, the subdivision was 1,158, 402, and 3,430, respectively. For AP, the subdivision was 289, 34, and 4,667, respectively.

(B) Manhattan plot of the $-\log_{10}$(p values) (y axis, truncated at 40) by chromosomal location (x axis) of each SNP for the three complement pathways: CP (yellow), LP (light blue), and AP (purple). The dashed line indicates the genome-wide significance threshold. Significant loci are labeled in black, along with the name of the closest gene. Borderline significant loci that are relevant to complement are labeled in gray.

(C) Genetic heritability analysis. Distributions of the pedigree-based age- and sex-adjusted genetic heritability of each trait (h_0^2, gray) and residual genetic heritability estimate after additional adjustment for the significant lead SNPs (h_r^2, light blue). The relative difference between h_0^2 and h_r^2 quantifies the proportion of additive genetic heritability explained by the identified variants.

(D) Association between lead SNPs and complement pathways are pathway specific: associations are significant only with the pathway for which the lead SNP was uncovered.

Here, we have exploited an implementation of qualitative assays, designed to detect complement deficiency and attribute the defect to a specific pathway, on a large population sample. We measured the functional activatability of CP, LP, and AP and conducted quantitative trait analyses to characterize their genetic underpinnings. Functional annotation, epidemiological data integration, and the exploitation of large transcriptomic and proteomic databases to enhance biological interpretation were instrumental in comprehensively characterizing the involvement of complement in human health.

RESULTS

GWAS of the complement activation pathways

We quantified complement activation in 4,990 general-population adults from the Cooperative Health Research in South Tyrol (CHRIS) study.[14] Pathway distributions were bimodal, reflecting previous observations (Figure 1A).[6] CP and AP were strongly correlated (Pearson's $r_{CP,AP}$ = 0.68); less correlation was observed with LP ($r_{CP,LP}$ = 0.30 and $r_{AP,LP}$ = 0.28). CP and AP, winsorized at 5%, and the inverse normally transformed LP underwent GWAS on 7,721,937 single-nucleotide polymorphisms (SNPs), imputed on the Haplotype Reference Consortium dataset (imputation quality score, Rsq $\geq$ 0.3; minor allele frequency [MAF] > 1%). Absence of genomic inflation was supportive of appropriate population structure modeling (Figure S1). We identified seven pathway-specific genome-wide significant loci ($p < 5 \times 10^{-8}$; Figure 1B and Table 1), functionally annotating all variants in strong linkage disequilibrium (LD $r^2 \geq 0.8$) within ±250 kb around each lead SNP (Figure S2).

CP was associated with two independent SNPs in the major histocompatibility complex (MHC): rs116298992 inside *TNXB* and rs28732173 next to complement genes *C4B*, *C4A*, *CFB*, and *C2* (Table 1 and Figure S3). rs116298992 is an expression quantitative trait locus (QTL) for several genes including *C4A* and is in strong LD with a likely deleterious missense variant for *DXO* (Figure S2A).

LP was associated with three loci (Table 1): one centered on the ABO blood group variant rs505922 (Figures S2C and S4A); one centered on rs10900026 in *CXCL12* (Figures S2D and S4B); and one at rs1031101 upstream of *MBL2*, which encodes the soluble mannose-binding lectin protein and is responsible for LP activation (Figure S5). rs1031101 is in strong LD with the known deleterious variant rs1800450 (Figure S2I). Iterative conditional analyses

Table 1. Loci associated with the complement activation pathways: Unconditional and conditional results and functional annotation

						Unconditional results		Conditional results[b]			
Pathway	RSID	Chr:Pos (build 37)	Nearest gene	Range of consequences to transcripts	EA, EAF	b (SE)[a]	p value	b (SE)[a]	p value	$OR_{risk\ of\ <5\%\ activity}$ (95% CI)	Other missense variants (*target gene*) in strong LD with CADD PHRED score ≥20 within ±250 kb[c]
CP	rs116298992[d]	6:32,068,495	*TNXB*	intronic ↔ upstream gene	T, 0.01	−13.5 (1.08)	1.4×10^{-35}	–	–	1.00 (0.93, 1.07)	rs45531831 (*DXO*)
	rs28732173[d]	6:32,071,010	*TNXB*	intronic	A, 0.04	−5.6 (0.65)	1.1×10^{-17}	−5.7 (0.64)	7.8×10^{-19}	1.01 (0.96, 1.05)	–
LP	rs505922	9:136,149,229	*ABO*	intronic ↔ regulatory region	C, 0.38	0.2 (0.02)	1.5×10^{-25}	–	–	0.98 (0.94, 1.03)	–
	rs10900026	10:44,854,402	*CXCL12*	intronic	A, 0.57	0.14 (0.02)	3.0×10^{-9}	0.10 (0.02)[e]	7.0×10^{-7}	0.99 (0.97, 1.02)	–
	rs35805975	10:54,528,236	*MBL2*	synonymous 136 N>N ↔ TF binding site	A, 0.03	0.3 (0.07)	1.0×10^{-4}	−0.3 (0.05)	1.9×10^{-8}	0.83 (0.72, 0.96)	–
	rs1800451	10:54,531,226	*MBL2*	missense 57 G>E	T, 0.03	−1.0 (0.06)	8.4×10^{-54}	−0.8 (0.05)	1.1×10^{-66}	2.39 (2.10, 2.72)	–
	rs5030737	10:54,531,242	*MBL2*	missense 52 R>C	A, 0.08	−0.5 (0.04)	2.8×10^{-49}	−0.9 (0.03)	1.0×10^{-193}	1.81 (1.68, 1.96)	–
	rs7096206	10:54,531,685	*MBL2*	upstream gene ↔ TF binding site	C, 0.78	0.2 (0.03)	1.7×10^{-15}	0.6 (0.04)	5.8×10^{-51}	0.80 (0.75, 0.84)	–
	rs1031101	10:54,533,360	*MBL2*	upstream gene ↔ regulatory region	G, 0.14	−1.0 (0.03)	7.5×10^{-287}	–	–	2.41 (2.27, 2.55)	rs1800450 (*MBL2*)
	rs4935347	10:54,540,783	*MBL2*	intergenic ↔ regulatory region	C, 0.78	0.3 (0.03)	2.3×10^{-39}	0.6 (0.02)	1.4×10^{-139}	0.72 (0.68, 0.76)	–
AP	rs6685931	1:196,867,233	*CFHR4*	intronic	C, 0.39	−2.2 (0.36)	8.8×10^{-10}	–	–	1.02 (1.00, 1.05)	–
	rs74480769	5:40,972,211	*C7*	intronic	G, 0.04	−8.6 (1.01)	1.8×10^{-17}	–	–	1.03 (0.97, 1.11)	–
	rs1265908	6:31,891,662	*C2*	intronic ↔ NMD transcript	C, 0.07	−5.1 (0.67)	4.7×10^{-14}	–	–	0.98 (0.94, 1.03)	rs106287 (*SKIV2L*), rs440160 (*TNXB*)
	rs77220800	12:20,712,401	*PDE3A*	intronic	A, 0.04	−5.9 (0.99)	1.8×10^{-9}	–	–	0.98 (0.92, 1.05)	–

CP, classical pathway; LP, lectin pathway; AP, alternative pathway; Chr:Pos, chromosome, position; EA, effect allele; EAF, effect allele frequency; b, coefficient of association; SE, standard error of b; OR, odds ratio; 95% CI, 95% confidence interval; LD, linkage disequilibrium; CADD, combined annotation-dependent depletion; TF, transcription factor; NMD, nonsense-mediated mRNA decay.

[a]The coefficient of association reflects the change in the phenotype per copy of the effect allele: for CP and AP, effect is expressed in terms of degree of activation (%); for LP, effect is expressed as standard deviations of a normal distribution (STAR Methods).

[b]Reported for the secondary SNPs in each locus.

[c]Extensive annotation reported in Figure S2.

[d]rs116298992 and rs28732173 are independent variants: LD $r^2 = 0.002$.

[e]Conditioned on rs1031101.

Table 2. Analysis of haplotypes associated with LP at *CXCL12* and *MBL2*

Haplotype index number	rs10900026 (*CXCL12*)	rs35805975 (*MBL2*)	rs1800451 (*MBL2*)	rs5030737 (*MBL2*)	rs7096206 (*MBL2*)	rs1031101 (*MBL2*)	rs4935347 (*MBL2*)	Frequency	b	SE	p value
1 (reference)	A	G	C	G	C	A	C	0.310	–	–	–
2	C	.	.	.	.	.	.	0.191	−0.07	0.02	0.0037
3	.	.	.	.	G	.	T	0.095	−0.65	0.03	$<1.0 \times 10^{-16}$
4	C	.	.	.	G	.	T	0.085	−0.65	0.03	$<1.0 \times 10^{-16}$
5	.	.	.	.	.	G	.	0.075	−1.36	0.03	$<1.0 \times 10^{-16}$
6	C	.	.	.	.	G	.	0.070	−1.34	0.03	$<1.0 \times 10^{-16}$
7	.	.	.	A	.	.	.	0.045	−0.97	0.04	$<1.0 \times 10^{-16}$
8	C	.	.	A	.	.	.	0.038	−0.93	0.04	$<1.0 \times 10^{-16}$
9	.	.	.	.	G	.	.	0.019	−0.68	0.07	$<1.0 \times 10^{-16}$
10	.	A	.	.	.	.	.	0.018	−0.28	0.06	1.8×10^{-6}
11	C	.	T	.	.	.	T	0.016	−1.44	0.06	$<1.0 \times 10^{-16}$
12	C	.	.	.	G	.	.	0.015	−0.54	0.07	3.8×10^{-13}
13	.	.	T	.	.	.	T	0.012	−1.35	0.06	$<1.0 \times 10^{-16}$
14	C	A	.	.	.	.	.	0.008	−0.26	0.10	0.0118
15	C	.	T	.	.	.	.	0.001	−1.36	0.20	2.3×10^{-11}

Haplotypes are sorted by decreasing sample frequency. b denotes coefficient of association, expressed in standard deviations of LP, which was transformed using the inverse normal transformation (STAR Methods). SE, standard error of b.

identified four additional independently associated SNPs at *MBL2*, including deleterious missense variants rs5030737 and rs1800451, and an additional conditionally genome-wide significant variant (rs35805975; Table 1; Figures S2E–G, and S5).

AP was associated with loci centered on four lead SNPs (Table 1 and Figure S6): rs6685931 in *CFHR4* and in LD with variants spanning the *CFH-CFHR3-CFHR1-CFHR4-CFHR2-CFHR5* gene cluster; rs1265908 in *C2*, next to *CFB*, *C4B*, and *C4A*, and in strong LD with highly deleterious missense variants targeting *SKIV2L* and *TNXB* (Figure S2M); rs77220800 at *PDE3A*; and rs74480769 in *C7* and near *C6*. Given the rs74480769 variant is isolated with no LD support, we repeated the analysis using only directly genotyped sites to exclude genotype imputation artifacts, confirming the results (Table S1A and Figure S7). All unconditional and conditional associations with $p < 5 \times 10^{-6}$ are listed in Tables S1B and S1C, respectively. The uncovered genome-wide significant variants explained 14%, 74%, and 17% of CP, LP, and AP genetic heritability, respectively (Figure 1C).

CP and AP lead SNPs were not associated with CP and AP deficiency, which we defined as levels of <5% (versus ≥20%), on the basis of previous observations[6] (STAR Methods and Table 1). In contrast, the lead SNPs for LP showed large odds ratios for LP deficiency, ranging from 0.72 to 2.39 per effect allele copy (Table 1). Alternative trait transformations did not provide additional insights (Figures S8–S10; Tables S1D and S1E).

All identified SNPs were pathway specific, that is, associated only with the pathway for which they have been uncovered and not shared across pathways (Figure 1D and Table S1F). As age and sex may affect innate immune response,[15,16] we fitted models including main as well as SNP-by-age and SNP-by-sex interaction effects (Table S1G). Females had higher CP and lower AP levels than males. LP was not associated with sex. Older age was associated with higher levels of CP and AP and lower levels of LP. After accounting for multiple testing, we did not observe SNP-by-age or SNP-by-sex interaction evidence.

Haplotype analysis of the *MBL2* locus

The conditional analyses showed that the association of LP with rs10900026 at *CXCL12* was not fully independent of rs1031101 at *MBL2* (Table 1). In addition, despite most associations being statistically independent, we observed substantial LD between all SNPs identified at *MBL2* and *CXCL12* (Figure S11). This evidence suggested the presence of long-range haplotypes. Haplotype analyses, conducted to understand the variants' combined effects, identified 15 haplotypes with a relative frequency between 0.1% and 31% (Table 2). Compared to the most common haplotype, all haplotypes were associated with lower LP levels. Haplotypes including variants disrupting the collagen-like domain (rs1800451 and rs5030737) were those with the largest effects. A large effect was also observed for haplotypes involving variations at rs1031101, whose G allele was in complete LD with the T allele of missense variant rs1800450, causing lower serum MBL. Even if not exerting a large effect, we still observed an effect of the C allele of rs10900026 at *CXCL12* when all other SNPs were unchanged.

Given variants located in the promoter (rs7095891 "P/Q," rs7096206 "X/Y," rs11003125 "L/H") and in the first exon (rs1800451 "A/C," rs1800450 "A/B," rs5030737 "A/D") of *MBL2* are known to affect serum MBL2 levels and protein function, we contextualized our analysis with respect to previously studied (legacy) haplotypes (Tables S1H and S1I).[17] The deeper granularity observed by integrating new and legacy SNPs into combined haplotypes showed that some legacy haplotypes may lose their effect on LP depending on the allelic combinations at the newly identified variants (Table S1J): this supports

CellPress
OPEN ACCESS

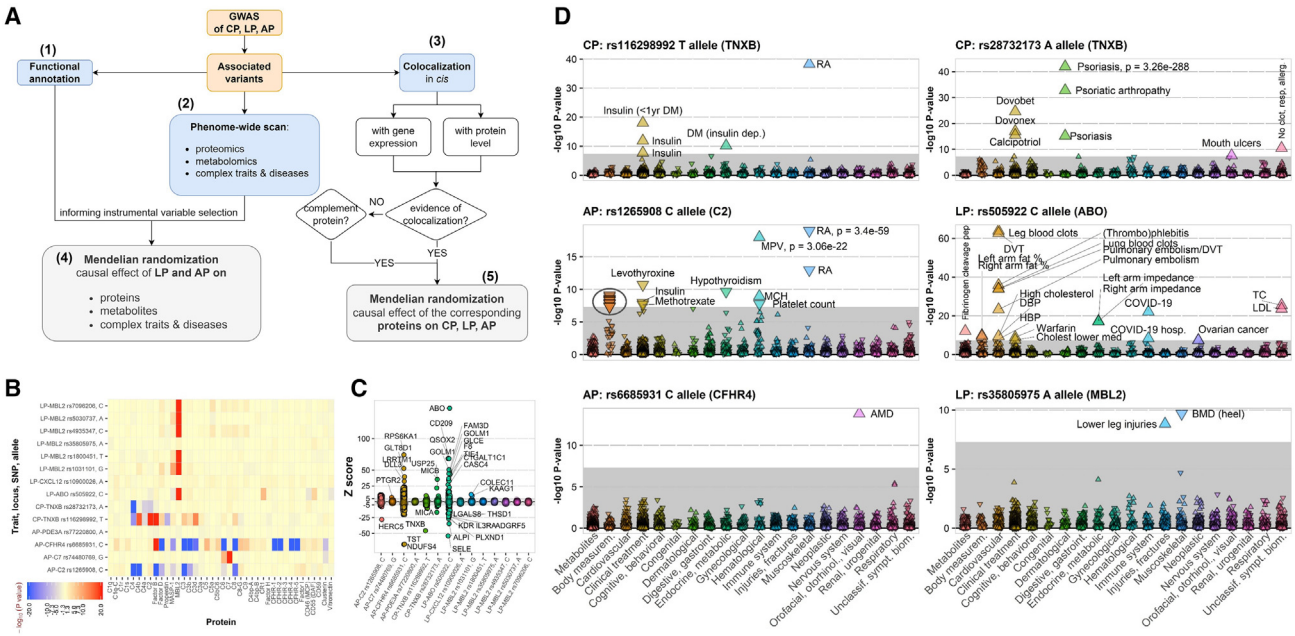


Figure 2. Flowchart of downstream analyses and phenome-wide association analyses

(A) Downstream analyses were conducted to elucidate (1) the functional role of identified variants, (2) their associations with proteins, metabolites, and complex traits and diseases, and (3) the colocalization of identified loci with gene expression and protein levels. Results informed Mendelian randomization studies to identify (4) causal effects of LP and AP on proteins, metabolites, and complex traits and diseases and (5) causal effects of complement and non-complement proteins on CP, LP, and AP.

(B) Association of identified variants with 40 complement proteins; $-\log_{10}$(p values) are indicated in red on the positive scale to reflect concordant effect of the effect allele on the trait and the protein, and in blue on the negative scale to reflect discordant effects of the effect allele on the trait and the protein (see full results in Table S1K).

(C) Association of identified variants with 4,937 other proteins. For each variant, the *Z*-scores of the variant-protein association are given with respect to the same reference allele reported in Table 1. Magnified dots outside the gray area indicate genome-wide significance ($p < 5 \times 10^{-8}$) results. Due to the elevated number of significant findings, only isolated dots are annotated with the name of the protein-encoding gene (see Table S1L for full results).

(D) PheWAS across 671 metabolites and 2,783 diseases and traits. Results are displayed by locus and trait in the form of $-\log_{10}$(p values) of association, indicating whether the effect alleles were positively (upward triangles) or negatively (downward triangles) associated with the trait. The gray area indicates the canonical significance level of 5×10^{-8}. In the rs1265908 panel, the circle encloses the following body measurements: right leg, left leg, whole body, trunk, and right arm fat mass; body, trunk, left leg, right leg, left arm, and right arm fat percentage; BMI; weight; and hip circumference. RA, rheumatoid arthritis; DM, diabetes mellitus; MPV, mean platelet volume; MCH, mean corpuscular hemoglobin; DVT, deep vein thrombosis; DBP, diastolic blood pressure; HBP, high blood pressure; TC, total cholesterol; LDL, LDL cholesterol; AMD, age-related macular degeneration; BMD, bone mineral density. Detailed results in Tables S1M (metabolites) and S1N (complex traits and diseases).

the added value of using a GWAS to identify variants associated with LP also at a very well-known and studied locus such as the *MBL2*.

Complement-associated variants are shared across proteins and complex traits

To characterize the role of the identified variants, identify the implicated genes, and elucidate biological pathways involving complement, we implemented and followed a rigid combination of downstream analyses (Figure 2A), as described in the following.

We started by conducting phenome-wide scans of molecular and disease-related traits. First, we tested association with 40 complement proteins from a recent large-scale proteomic GWAS[13] (Figure 2B and Table S1K): the observed associations were either very specific, i.e., involving only one or few proteins, or very extensive, i.e., involving multiple proteins. Specific associations were observed for all SNPs associated with LP at *MBL2*, which were associated almost exclusively with MBL2 levels, and for the AP-associated SNP rs74480769 at *C7*, which was only associated with C7. Likely reflecting pleiotropy of *ABO* and the *CFH* gene cluster, we observed multiple associations for the LP-associated SNP rs505922 at *ABO* and the AP-associated SNP rs6685931 at *CFHR4*. Finally, all MHC SNPs (rs116298992 and rs28732173 associated with CP and rs1265908 associated with AP) showed association with several complement proteins, not all relevant to the specific pathways, reflecting the extensive LD spanning this region.

A similar breakdown of the results into specific, pleiotropic, and LD-affected association patterns was observed when expanding testing to all 4,937 non-complement proteins from Pietzner et al.[13] (Figure 2C and Table S1L). The LP-associated variants at *MBL2* were not associated with any protein, except for association of rs1031101 with Collectin-11 (COLEC11) and the kidney-associated DCDC2 antisense RNA 1 (KAAG1). These findings support the LP specificity of the *MBL2* LP-associated variants.

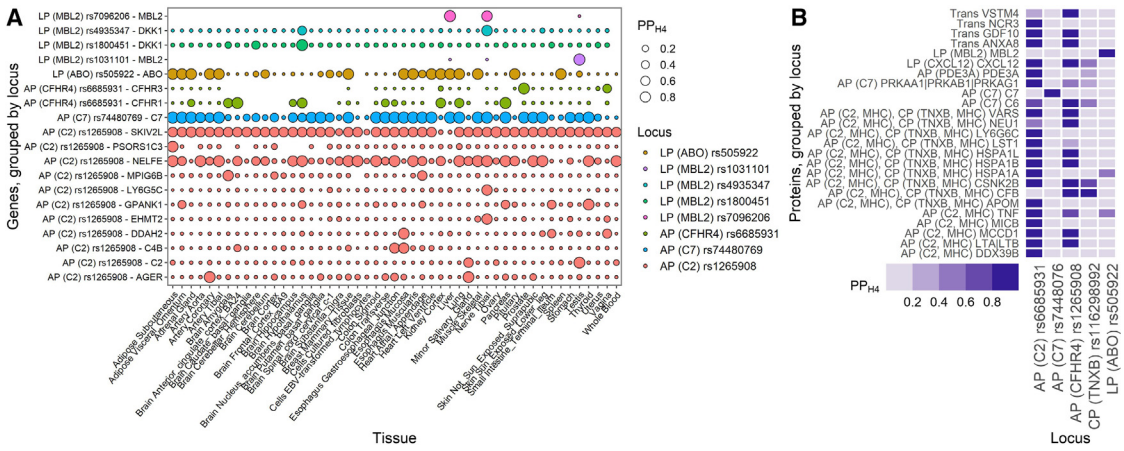


Figure 3. Colocalization with transcriptomic and proteomic levels

(A) Colocalization with gene expression from the GTEx v8 database. Displayed are genes with colocalization probability $PP_{H4} > 0.8$ in at least one tissue.

(B) Colocalization with protein levels. Displayed are proteins and loci showing at least one colocalization at $PP_{H4} > 0.80$ (full results in Table S1O).

The AP-associated variant rs77220800 at *PDE3A* was not associated with any protein, and the rs74480769 at *C7* was only associated with prostaglandin reductase 2 (PTGR2), supporting AP specificity of these loci. In contrast, the LP-associated variant rs505922 at *ABO* and the AP-associated variant rs6685931 at *CFHR4* were associated with multiple proteins, reflecting pleiotropy. Multiple associations due to the extensive LD were observed for all MHC variants associated with CP or AP.

We finally screened 671 metabolites and 2,783 complex traits and diseases (STAR Methods). Eight variants were not associated with any trait. Six were associated with at least one trait at $p < 5 \times 10^{-8}$ (Figure 2D; Tables S1M and S1N). The *MBL2* rs35805975 LP-increasing allele was associated with lower heel bone mineral density and risk of unspecified lower leg injuries. The *ABO* rs505922 LP-increasing allele was associated with a multiplicity of functions and traits. The *CFHR4* rs6685931 AP-lowering allele was associated with age-related macular degeneration (AMD) risk. The CP-lowering allele at rs116298992 (*TNXB*) was associated with risk of rheumatoid arthritis and insulin-dependent diabetes: this may reflect LD with HLA haplotypes[18] or involvement of *TNXB* on CP-related disorders.[19,20] Similar caution is needed to interpret MHC associations of the CP-associated SNP rs28732173 with autoimmune diseases and the AP-associated SNP at *C2* with multiple hematological, endocrine, metabolic, and immunological traits.

While these results put the identified loci into a general context, they do not allow conclusions on causal mechanisms. In the following, we assessed causality using Mendelian randomization (MR) after addressing the pleiotropy issue through instrumental variable (IV) selection and preventing LD confounding through colocalization analysis.

Colocalization with transcriptomic and proteomic levels

To identify implicated genes at each locus, we estimated the probability of colocalization with the expression of genes located within ±100 kb of the lead SNP associated with each specific complement pathway across 49 tissues from the Genotype-Tissue Expression (GTEx) Project,[21] claiming colocalization when $PP_{H4} > 0.80$ (Figure 3A; see Data S1, File 1 for more extensive results on each locus). Because complement-related genes can be expressed across multiple tissues, we chose to explore all available tissues. Results were independent on the window size, except for the two CP loci in the MHC, which were excluded from further analyses because the extensive LD did not guarantee the results' reliability (Figure S12). At *ABO*, we observed colocalization of LP with *ABO* expression in 15 tissues (Figure 3A). At *MBL2*, despite *MBL2* expression data being available for just three tissues, we observed colocalization of LP with *MBL2* at rs1031101 in testis ($PP_{H4} = 0.99$) and rs7096206 in liver ($PP_{H4} = 0.99$). We also observed colocalization of LP with *DKK1* in brain hypothalamus at rs1800451 and tibial nerve at rs4935347. We observed strong colocalization of *C7* expression with AP across most tissues: $PP_{H4} > 0.80$ in 24 tissues and $PP_{H4} > 0.99$ in 16 tissues (Figure 3A). At *CFHR4* (rs6685931), we observed colocalization of AP with *CFHR1* in the brain anterior cingulate cortex and pancreas, with consistent evidence in lung, kidney cortex, and other brain tissues, as well as with *CFHR3* in prostate. At the MHC rs1265908 locus, AP colocalized with *NELFE* ($PP_{H4} > 0.8$ in 18 tissues); scattered colocalizations across different tissues were also observed for *C2*, *SKIV2L*, and *AGER*.

We extended colocalization analyses to 62 proteins from Pietzner et al.[13] whose encoding genes were located within the identified loci or had shown evidence of colocalization with the transcriptomic level (Figure 3B and Table S1O). At *ABO*, LP colocalized with MBL2 almost perfectly ($PP_{H4} = 0.9988$). At *C7*, AP did colocalize perfectly with C7 ($PP_{H4} = 1$). Interestingly, at the *CFH* gene cluster, AP colocalized with multiple proteins whose encoding genes were located in other loci associated with AP (PDE3A, $PP_{H4} = 0.93$), AP and CP (several MHC proteins, including CFB), and LP (CXCL12, $PP_{H4} = 0.98$). As expected, we observed multiple colocalizations for AP at the MHC.

In summary, we identified 17 genes colocalizing with either LP or AP (Figure 3A): six were complement genes (*MBL2*, *CFHR1*, *CFHR3*, *C7*, *C4B*, and *C2*) and 11 non-complement genes. We also identified four complement proteins (MBL2, C7, C6, and

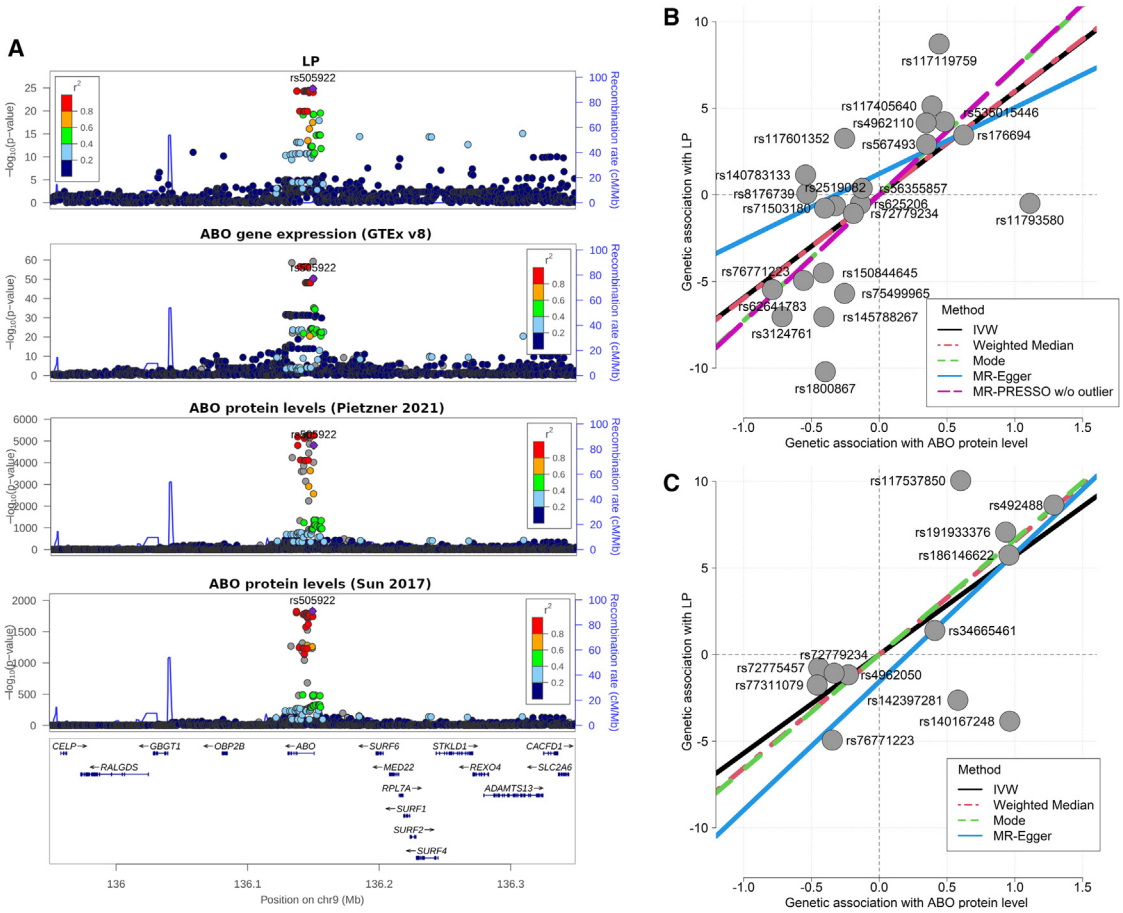


Figure 4. Colocalization and causal effect of ABO on LP

(A) Regional association plots for LP *ABO* gene expression in the visceral adipose omentum (PP_{H4} = 0.89), ABO protein level in Pietzner et al.[13] (PP_{H4} = 0.03), and ABO protein levels in Sun et al.[12] (PP_{H4} = 0.9988).

(B) Results of the *cis* MR analysis of ABO protein levels from Pietzner et al.[13] on LP.

(C) Results of the *cis* MR analysis of ABO protein levels from Sun et al.[12] on LP.

CFB) and 21 non-complement proteins colocalizing with CP, LP, or AP (Figure 3B). The 11 non-complement genes and 21 non-complement proteins corresponded to 26 unique non-complement proteins.

Causal effects of proteins on complement pathways

We assessed the causal effect of these 26 non-complement proteins and all 40 complement proteins on the three pathways using two-sample MR analysis (Figure 2A). Given that mammalian phenotypes can be driven by both *cis* and *trans* QTLs,[22] SNPs for use as IVs were selected both in *trans* and in *cis* (±500 kb) of the protein-encoding gene boundaries.[23] In addition, given many associations felt within expected biologically plausible genes and involved functional variants within the gene, e.g., *MBL2*, we also selected IVs just inside the genes. This exploratory approach should be critically evaluated in the overall context. However, it is expected that protein QTLs are less likely to influence targets through pleiotropic pathways as the distance to the functional variants decreases.[23] All IVs for all proteins are listed in Table S1P.

Among the non-complement proteins, higher ABO levels were causally associated with higher LP levels. Despite lack of colocalization between LP and ABO protein levels (Figure 4A), the *cis* MR analysis showed a strong causal effect of ABO on LP (p_{cis} = 5.8×10^{-7}; Figure 4B and Table S1Q; Data S1, File 2), with some heterogeneity not due to horizontal pleiotropy (non-significant MR-Egger intercept p value). However, given that the lack of colocalization could make this evidence weaker, we repeated all analyses using an independent proteomics dataset[12]: here, we observed clear colocalization between the ABO protein product levels and LP (PP_{H4} = 0.9988; Figure 4A), confirming the causal effect of ABO on LP (p_{cis} = 1.1×10^{-10}, with non-significant MR-Egger intercept and non-significant MR-PRESSO global test; $p_{ingene} = 4.5 \times 10^{-23}$; Figure 4C and Data S1, File 2). Despite colocalization with AP or LP, for all other non-complement proteins, MR assumptions could not be verified, preventing causal claims (Table S1Q and Data S1, File 2).

MR analyses of complement proteins on complement pathways highlighted several causal effects supported by known biology (Figure 5A; sensitivity analyses in Data S1, File 2).

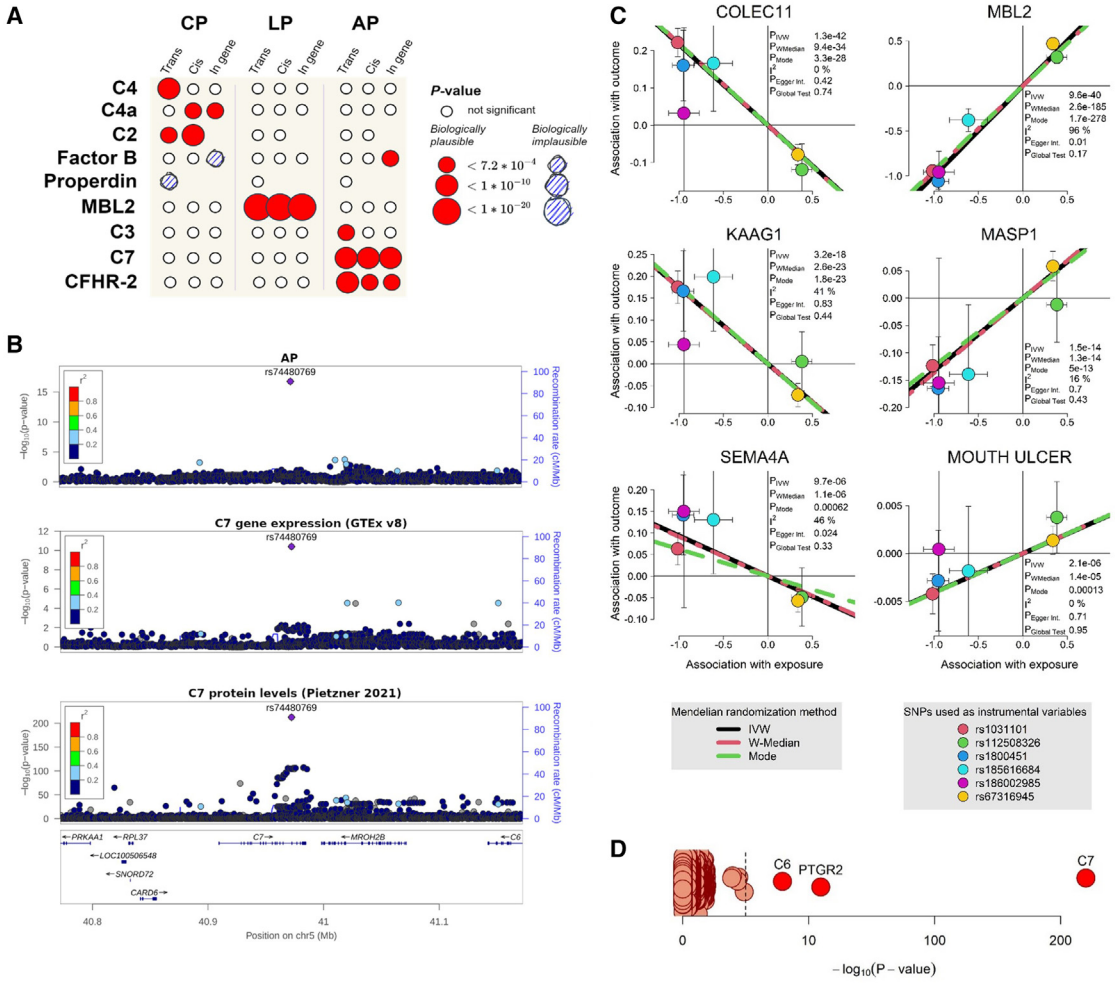


Figure 5. Causal analysis of the complement pathways

(A) Causal effects of complement proteins on complement pathways. Circle sizes are proportional to the MR analysis p values (inverse-variance-weighted [IVW] method), with IVs located in *trans*, *cis*, and within the protein-encoding gene. Reported are proteins with at least one significant association (full results in Table S1R; sensitivity analysis results in Data S1, File 2).

(B) Colocalization of genetic associations with AP levels, *C7* gene expression, and C7 protein levels at the C7 locus.

(C) Causal effects of LP on protein levels and complex diseases. Shown here are the significant findings resulting from MR analyses of LP on 4,979 proteins and 2,783 complex traits and diseases. LP showed causal effects on collectin-11 (COLEC11), mannose-binding lectin protein (MBL2), kidney-associated antigen 1 (KAAG1), mannan-binding lectin serine protease 1 (MASP1), Semaphorin-4A (SEMA4A), and mouth ulcer risk. Genetic associations with exposure and outcome are shown for each of the six SNPs used as IV. Reported are: p value (P) for the IVW method, weighted median (W-Median), and mode (Mode) estimation methods; the heterogeneity index I^2; the MR-Egger intercept p value; and the MR-PRESSO global test p value. IVW results are in Tables S1U (proteins) and S1W (complex traits); sensitivity analyses are in Data S1, File 2.

(D) AP was causally associated with C7, PTGR2, and C6 protein levels. Reported are the $-\log_{10}$(p values) from the Wald ratio method (full results in Table S1X).

MBL2 was causally associated with LP consistently across *trans*, *cis*, and within-gene analyses ($p_{trans} = 1.1 \times 10^{-35}$; $p_{cis} = 1.1 \times 10^{-44}$; $p_{ingene} = 1.1 \times 10^{-77}$; Table S1R). CFHR2 was causally related to AP consistently across all analyses ($p_{trans} = 1.4 \times 10^{-11}$; $p_{cis} = 4.4 \times 10^{-9}$; $p_{ingene} = 5.0 \times 10^{-4}$), across different MR methods, and with no evidence of horizontal pleiotropy (Table S1R). Supported by strong evidence of colocalization (Figure 5B), C7 showed a causal effect on AP across all analyses ($p_{ingene} = 4.4 \times 10^{-18}$; Table S1R). Additional causal associations supported by known biology were observed for C3 and CFB on AP and for C4, C4a, and C2 on CP. These findings were supported by consistent results obtained with robust methods (weighted median, mode estimation, MR-PRESSO). However, in some cases, horizontal pleiotropy could not be fully excluded. Contradictory findings were identified for proteins with encoding genes located in the MHC, supporting the difficulty to statistically dissect biological signals when multiple genes involving different pathways are clumped together by strong LD (Table S1R). The C3d-to-C3 ratio[11] did not show causal effects on complement pathways (Table S1S).

Causal effects of the lectin and alternative pathways on protein levels and complex traits

Supported by biological plausibility[24] and by evidence that LP-associated variants at *MBL2* were LP specific (Figures 2B and 2C; Data S1, File 1), we selected six variants at *MBL2* as IVs for MR analyses of LP on several outcomes (Table S1T). Screening for causal effects of LP on 4,979 proteins[13] showed causal effects on COLEC11 ($p = 6.3 \times 10^{-44}$, $p_{het} = 0.4639$), KAAG1 ($p = 9.0 \times 10^{-25}$, $p_{het} = 0.2695$), semaphorin-4A (SEMA4A; $p = 9.7 \times 10^{-6}$; $p_{het} = 0.0974$), MBL2 ($p = 4.1 \times 10^{-41}$; $p_{het} = 2.4 \times 10^{-41}$; MR-Egger intercept $p = 0.99$), and MASP-1 ($p = 1.0 \times 10^{-16}$; $p_{het} = 0.3995$; Figure 5C and Table S1U; Data S1, File 2). The additional screening of 671 metabolites revealed a causal effect on X-12407, an untargeted metabolite of unknown function already associated with hepatic steatosis[25] (Table S1V). Finally, the systematic MR analysis across 2,783 diseases and traits highlighted, very robustly, that higher LP levels were causally associated with the risk of developing mouth ulcers ($p = 9.5 \times 10^{-6}$, with no heterogeneity; Figure 5C and Table S1W; Data S1, File 2).

For AP, SNP rs74480769 at *C7* was a biologically plausible, valid, and strong instrument (Table S1T), specific to C7 (Figures 2B and 2C; Data S1, File 1), minimizing the risk of pleiotropy. MR analyses of 4,979 proteins suggested causal effects of AP on C7 ($p = 9.0 \times 10^{-213}$), C6 ($p = 1.2 \times 10^{-8}$), and PTGR2 ($p = 1.1 \times 10^{-11}$; Figure 5D and Table S1X). Data did not support causal effects of AP on human metabolites (Table S1Y) or complex traits or diseases (Table S1Z).

DISCUSSION

We identified 13 independent variants over seven distinct loci associated with the three complement activation pathways, explaining a large portion of complement activation genetic heritability. The tagged loci carry known complement-related genes, with *ABO* and *PDE3A* being notable exceptions. Identified variants were pathway specific, not shared across pathways. Integration of transcriptomics and proteomics enabled causal analyses that expanded our knowledge of the involvement of complement in human health. In particular, results confirmed biological knowledge and analytical evidence[26] about the causal effects of C4, C4a, and C2 on CP, of MBL2 on LP, and of C3, CFB, and CFHR2 on AP. In addition, they highlight potential feedback loops between MBL2 and LP and between C7 and AP. Finally, they implicate causal roles for LP and AP on non-complement proteins and for LP on mouth ulcers.

The identified variants explained a large genetic heritability portion of each pathway. This was particularly marked for LP because here we identified variants on *MBL2*, which encodes the main LP triggering protein. LP's heritability aligns to previous studies on MBL2 reporting 77% in adult twins,[27] but lower than the 96% observed in healthy pediatric twins[28]: the age-related decline might reflect exogenous factors coming into play with aging. For LP there is thus limited room for uncovering additional loci: one of them, *MASP2*, which is crucial for LP activation, was just below the significance threshold (Figure 1B). In contrast, CP and AP await further studies to uncover their largely missing heritability.

The main innovation and strength of our study was the simultaneous quantification of all three complement activation pathways in a large general-population sample. Until now, complement pathway assessment was limited to clinical settings or small samples, primarily focused on determining activation insufficiency rather than assessing the variability of the degree of activation among unselected individuals. Others have used GWASs together with MR to assess the involvement of most of the complement proteins in health and disease.[12,13,29] However, no similar study of the activation of the three pathways has been conducted directly to date.

The large-scale assessment allowed us to conduct a genomic screening over a densely imputed genotype panel that confirmed the association of LP activation with MBL2 known functional variants in the collagen-like domain (rs1800451 "A/C," rs1800450 "A/B," rs5030737 "A/D") and in the promoter and 5′ UTR (rs7096206 X/Y). Additional known variants, rs11003125 (L/H) and rs7095891 (P/Q), were in strong LD with the reported variants, with similar degree of association with LP. Associations of AP at *CFHR4* and *C7* and of CP at MHC near known CP-related genes such as *C4*, *C4a*, and *C2* further support the validity of the analyses.

Haplotype analysis at *MBL2* and *CXCL12* brought several pieces of evidence: first, the existence of extensive haplotypes spanning up to 10 Mb involved in LP regulation; second, the newly identified *MBL2* variants add information in terms of haplotype diversity over the legacy haplotypes studied over a long time[17]; and third, *CXCL12* might have a role in LP regulation. Despite the observed significant association between LP and rs10900026 at *CXCL12* being reduced after conditioning on rs1031101 at *MBL2*, haplotypes with variations only at *CXCL12* were still associated with LP. rs10900026 is located close to distal enhancer-like elements and falls within transcription factor binding sites, potentially affecting the binding of several transcription factors (*DMRTA1*, *DMRTA2*, *DMRTC2*). Remarkably, MBL knockout mice have defective hematopoietic reconstitution and reduced expression of *CXCL12*, also called stromal cell-derived factor 1 (*SDF1*), which plays an essential role in homing stem cells together with its receptor *CXCR4*.[30]

MR analyses identified a causal effect of MBL2 on LP, reflecting known physiology,[31] as well as causal effects of LP on MBL2 and MASP1, which suggest the existence of regulatory feedback loops in LP. Feedback loops were already described in the coagulation system, a conceptually similar cascade.[32] LP also showed causal effects on COLEC11, KAAG1, and SEMA4A.

COLEC11 is a known activator of LP upon binding with MASP1.[33] Local production of collectin-11 has been consistently involved in immunoglobulin A nephropathy, where LP activation seems to play a crucial role in the disease pathogenesis, severity, and fibrosis.[34] Related to the latter, COLEC11 is a marker of hepatic stellate cells and appears to be a key driver of liver fibrosis.[35] In hypoxia and hypothermia mouse models, collectin-11 was found to be locally overexpressed in renal proximal tubule cells, where it recognized an aberrant fucosylated antigen present on stressed cells, promoting LP activation and organ injury: this mechanism can partially explain ischemia-reperfusion injury in renal transplantation, which can severely impair graft survival.[36]

KAAG1 was described as a self-antigen recognized by cytolytic T lymphocytes in renal carcinoma cell lines, but the protein was also recognized in other normal tissues.[37] Genome browsers annotate *KAAG1* as an antisense transcript (long non-coding RNA) with very low expression levels in multiple tissues. Its biological role and the factors regulating its translation are unknown. The observed causal effect of LP on KAAG1 is information that may help elucidate *KAAG1* function.

SEMA4A drives a T helper 2 response upon binding to the immunoglobulin-like transcript 4 (ILT-4) protein. ILT-4 is also a scavenger receptor of the soluble complement fragment C4d, an end product of CP and LP and a recognized marker of organ rejection.[38,39] These mechanisms could belong to the complex modulation of T cell response operated by the complement system, which bridges innate and adaptive immunity.[40] All these relations warrant functional validation.

Our results strongly implicated a causal effect of ABO on LP. The histo-blood group ABO encodes a glycosyltransferase catalyzing the transfer on glycoproteins of *N*-acetylgalactosamine (A antigen) or galactose (B antigen), or none of them (O blood group). Given rs505922 is in complete LD with rs8176719, which determines the O blood group, our findings indicate higher LP activation in non-O blood groups. At rs505922 we further observed colocalization of ABO protein product with both the MBL2 protein levels and LP activation. The observed causal relation might mirror the involvement of ABO in the glycosylation of LP proteins. Some authors have provided immunohistochemical evidence of an increased deposition of C4d, which is produced by both the CP and the LP, in ABO-incompatible kidney allograft specimen.[41] Extensive search did not provide other apparent evidence. Investigations that could clarify whether the differential susceptibility to diseases and infections traditionally observed in different blood groups[42] may be mediated by LP activation, involving autoimmunity and senescent cell removal functions, are warranted.[43]

Particularly interesting are the findings for AP, for which we identified four loci: three including known complement genes and one at *PDE3A*. The cyclic nucleotide phosphodiesterase encoded by *PDE3A* is the most abundant isoform in platelets, modulating platelet activation by decreasing cyclic AMP.[44] Activated platelets can in turn activate the complement system.[45] PDE3A is a target for cilostazol, a vasodilator that prevents platelet aggregation and exerts an anti-inflammatory action whose mechanisms are not yet fully understood. AP association at rs1265908 on *C2* also implicated *NELFE* and *SKIV2L*. *SKIV2L* is a helicase belonging to the SKI exosome complex, which is essential for B cell development[46] and modulates the activation of RIG-I-like receptors in response to exogenous RNAs.[47]

Our results further implicated a shared genetic variant driving *C7* gene expression, C7 protein levels, and AP levels, with a causal effect of AP on C7 and a feedback effect of C7 on AP. Most complement components are synthesized in the liver and, to a lesser extent, in extrahepatic sites such as mononuclear phagocytes. A major exception is C7, which is among the very few that are not primarily synthesized by the liver,[48] conveying a regulatory function to C7: when C7 is synthesized locally, e.g., by activated granulocytes and monocytes, complement attack is markedly potentiated.[49] In light of this, further investigations of the C7 influence on AP are warranted. AP showed causal effects also on C6, which is also part of the TCC/MAC, and PTGR2, which might be related to complement through the NADPH pathway, as it catalyzes the NADPH-dependent conversion of 15-keto-prostaglandin E2 to 15-keto-13,14-dihydro-prostaglandin E2.

Globally, the results also highlight the central role of complement at both the deep biological level and the disease level. An example of how these results add evidence to known loci is the AP-associated rs6685931 variant at *CFHR4*, which was previously associated with the C3d-to-C3 ratio in patients with AMD.[11] Haplotypes spanning the *CFH-CFHR1*-to-*CFHR5* gene cluster determine factor H levels,[9,10] which are under strong genetic control[50] and are associated with C3 and C4 levels.[8] In addition to colocalization of AP with *CFHR1* and *CFHR3* expression, we observed colocalization of AP with several proteins whose encoding genes are located in loci associated with each complement pathway: PDE3A, C2, and C7 with AP; CXCL12 with LP; and TNXB with CP. Those proteins were not causally related to AP: more likely the observed colocalizations were simply reflective of the effect of the *CFH* gene cluster on multiple systems, including proteins and AP, as confirmed by the clear causal effect of CFHR2 on AP. Also, rs1265908 at *C2* was associated with proteins belonging to heterogeneous classes, but in this case the result is likely reflective of widespread LD at the MHC. In addition to proteins, identified variants exhibited association with several complex traits but almost no association with metabolite levels, consistent with the role of complement proteins that do not generate small metabolites.

MR analyses depicted a causal role of LP on mouth ulcers. Both salivary agglutinin and salivary scavenger and agglutinin (SALSA) proteins can activate LP when they are absorbed on surfaces and inhibit LP activation when they are in fluid phase, hence exerting a dual role in LP modulation.[51,52] Dysregulation of this mechanism may cause mouth ulcers, which are often observed in allergic conditions and inflammatory bowel diseases, but this hypothesis must be proved. Remarkably, *KAAG1*, whose encoded protein levels were causally affected by LP, is expressed in oral mucosa and just a few other tissues.

We did not observe causal effects of AP on any disease. One reason for this could be our removal, by design, of individuals with AP deficiency, limiting the pertinence of observed associations to the range of normal AP variations. Having only the *C7* SNP as a valid MR instrument might have further limited the statistical power. For CP, causal analyses were impeded by the only SNPs being located in the MHC. Finally, phenotypes directly related to complement dysregulation include the susceptibility to peculiar infections or uncommon diseases for which large GWASs are absent. In all cases, complement GWASs should achieve larger sample sizes to enable the effective investigation of causal pathways on the disease domain.

Our bidirectional MR analysis of the 40 complement proteins and the three complement pathways served two purposes. On the one hand, this was a positive control. On the other hand, we wanted to approach the problem with a hypothesis-free mindset, allowing the possibility that causality was not necessarily monodirectional: while complement proteins are known activators of the complement pathways, we allowed the

possibility of observing the reverse phenomenon, namely that the proteins are themselves altered as a consequence of complement activation. In fact, we identified previously undescribed feedback loops, suggesting the existence of yet unexplored mechanisms. From a more technical point of view, this screening also gave the opportunity to assess somehow the reliability of the MR approach in the presence of strong prior biological knowledge. We observed cases such as the effect of prostaglandin and CFB on CP where, in the absence of known biology confuting the results, it would have been difficult to assess the results' validity. This places a cautionary note in the interpretation of large-scale screens that use MR automatically as a discovery tool.

In conclusion, this study expands biological knowledge of the complement biological mechanisms and implications. Our results constitute a valuable resource to deepen knowledge of the causal interplay between immune system and complex traits and diseases.

Limitations of the study

Our study had several limitations. First, given the innovative approach, there was no available replication study. Nevertheless, we conducted numerous downstream bioinformatic analyses that identified known associations and mechanisms, serving as reassuring positive controls. Second, the sample size was not large enough either to investigate rare genetic variants or to conduct aggregated data analyses that would have amplified the results' impact, such as genetic correlation analysis[53] or integrative bioinformatic assessment of pathways, tissues, and cell-type enrichment.[54] Finally, some identified variants fell in the MHC, whose extensive LD prevented precise identification of causative genes and variants: recent improvements in the genetic characterization of this locus[55] will enable greater precision in future studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Study design
- METHOD DETAILS
  - Assessment of functional complement activity
  - Genotyping and genetic imputation
  - Validation of the C7 SNP rs74480769
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Genome-wide association studies
  - Association of *C7* SNP rs74480769 with AP
  - Haplotype analysis
  - Conditional analyses and interaction with sex and age
  - Genetic heritability explained by the identified variants
  - Variant annotation
  - Shared genetic background with human traits and diseases
  - Colocalization analyses
  - Mendelian randomization (MR) analyses
  - Software

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113611.

ACKNOWLEDGMENTS

We thank all participants in the CHRIS study, the general practitioners of the Vinschgau/Val Venosta district, the personnel of the hospital of Schlanders/Silandro, the field study team, and the personnel of the CHRIS Biobank (BRIF code BRIF6107) for their support and collaboration. Extensive acknowledgment is reported at https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-015-0704-9. We thank the INTERVAL study for making the summary statistics from their protein GWAS freely available. At Eurac Research, we thank Chiara Volani (technical assistance); Ilaria Bozzolan and Giulia Caprioli (biobank operations); Yuri D'Elia, Daniele Di Domizio, and Clemens Egger (IT support); Laura Eccel (graphical support); and Jonathan Mitchell (bioinformatic support). We further thank Gertraud Streiter, Maria Bertsch, and Viktoria Staudinger (Medical University of Innsbruck) for technical assistance; Zoltán Kutalik and Eleonora Porcu (University of Lausanne) for methodological discussion; and Alexander Teumer (Medical University of Greifswald) for sharing the colocalization analysis code. This study was supported by the Austrian Science Fund FWF funded doctoral program HOROS (FWF, Vienna, W-1253, to D.N., R.W., and F.K.), which also funded the open access publishing costs, and the transnational doctoral program BI-DOC between the Innsbruck Medical University, Austria, the Land of Tyrol, and the Eurac Research Institute for Biomedicine, Bolzano, Italy. The present research was partly conducted within the "PACE: Partnership to Accelerate Covid-19 rEsearch in South Tyrol" project, funded by the Department of Innovation, Research and University of the Autonomous Province of Bolzano within the 2019–2021 Research Program (unique project code: D52F20000770003, to C.P. and P.P.P.).

AUTHOR CONTRIBUTIONS

Conceptualization, C.P., D.O.-H., R.W., F.K., and D.N.; methodology, D.N., F.D.G.M., and C.P.; formal analysis, D.N., L.F., E.K., G.B., D.G.-S., and C.P.; investigation, D.N., D.O.-H., L.F., F.K., R.W., and C.P.; resources, D.O.-H., F.K., M.S., S.C., R.W., C.F., A.D.G., M.P., C.L., M.G., and P.P.P.; data curation, D.N., L.F., G.B., E.K., M.G., and C.P.; writing – original draft, D.N., L.F., D.O.-H., E.K., F.K., R.W., and C.P.; writing – review & editing, D.N., L.F., D.O.-H., E.K., G.B., M.P., D.G.-S., S.C., C.F., M.G., F.D.G.M., A.D.G., M.S., E.W., C.L., C.L.-F., P.P.P., F.K., R.W., and C.P.; visualization, D.N., L.F., E.K., and C.P.; supervision, D.O.-H., R.W., F.K., and C.P.; funding acquisition, R.W., F.K., D.O.-H., C.L.-F., and P.P.P.

DECLARATION OF INTERESTS

R.W. has no financial holdings and does not have professional affiliations or holds advisory positions or board memberships, but receives contributions from SVAR for a scientific project.

Received: September 30, 2022
Revised: September 8, 2023
Accepted: December 7, 2023

REFERENCES

1. Nonaka, M. (2014). Evolution of the Complement System. In MACPF/CDC Proteins - Agents of Defence, Attack and Invasion, G. Anderluh and R. Gilbert, eds. (Springer Netherlands), pp. 31–43.
2. McGeer, P.L., Lee, M., and McGeer, E.G. (2017). A review of human diseases caused or exacerbated by aberrant complement activation. Neurobiol. Aging *52*, 12–22.
3. Risitano, A.M., Mastellos, D.C., Huber-Lang, M., Yancopoulou, D., Garlanda, C., Ciceri, F., and Lambris, J.D. (2020). Complement as a target in COVID-19? Nat. Rev. Immunol. *20*, 343–344.
4. Pouw, R.B., and Ricklin, D. (2021). Tipping the balance: intricate roles of the complement system in disease and therapy. Semin. Immunopathol. *43*, 757–771.
5. West, E.E., Kunz, N., and Kemper, C. (2020). Complement and human T cell metabolism: Location, location, location. Immunol. Rev. *295*, 68–81.
6. Seelen, M.A., Roos, A., Wieslander, J., Mollnes, T.E., Sjöholm, A.G., Wurzner, R., Loos, M., Tedesco, F., Sim, R.B., Garred, P., et al. (2005). Functional analysis of the classical, alternative, and MBL pathways of the complement system: standardization and validation of a simple ELISA. J. Immunol. Methods *296*, 187–198.
7. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. *101*, 5–22.
8. Yang, X., Sun, J., Gao, Y., Tan, A., Zhang, H., Hu, Y., Feng, J., Qin, X., Tao, S., Chen, Z., et al. (2012). Genome-Wide Association Study for Serum Complement C3 and C4 Levels in Healthy Chinese Subjects. PLoS Genet. *8*, e1002916.
9. Zhu, L., Zhai, Y.L., Wang, F.M., Hou, P., Lv, J.C., Xu, D.M., Shi, S.F., Liu, L.J., Yu, F., Zhao, M.H., et al. (2015). Variants in Complement Factor H and Complement Factor H-Related Protein Genes, CFHR3 and CFHR1, Affect Complement Activation in IgA Nephropathy. J. Am. Soc. Nephrol. *26*, 1195–1204.
10. Alic, L., Papac-Milicevic, N., Czamara, D., Rudnick, R.B., Ozsvar-Kozma, M., Hartmann, A., Gurbisz, M., Hoermann, G., Haslinger-Hutter, S., Zipfel, P.F., et al. (2020). A genome-wide association study identifies key modulators of complement factor H binding to malondialdehyde-epitopes. Proc. Natl. Acad. Sci. USA *117*, 9942–9951.
11. Lorés-Motta, L., Paun, C.C., Corominas, J., Pauper, M., Geerlings, M.J., Altay, L., Schick, T., Daha, M.R., Fauser, S., Hoyng, C.B., et al. (2018). Genome-Wide Association Study Reveals Variants in CFH and CFHR4 Associated with Systemic Complement Activation: Implications in Age-Related Macular Degeneration. Ophthalmology *125*, 1064–1074.
12. Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., et al. (2018). Genomic atlas of the human plasma proteome. Nature *558*, 73–79.
13. Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Cortes, A., Koprulu, M., Wörheide, M.A., Oerton, E., Cook, J., Stewart, I.D., Kerrison, N.D., et al. (2021). Mapping the proteo-genomic convergence of human diseases. Science *374*, eabj1541.
14. Pattaro, C., Gögele, M., Mascalzoni, D., Melotti, R., Schwienbacher, C., De Grandi, A., Foco, L., D'Elia, Y., Linder, B., Fuchsberger, C., et al. (2015). The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J. Transl. Med. *13*, 348.
15. Shaw, A.C., Goldstein, D.R., and Montgomery, R.R. (2013). Age-dependent dysregulation of innate immunity. Nat. Rev. Immunol. *13*, 875–887.
16. Roved, J., Westerdahl, H., and Hasselquist, D. (2017). Sex differences in immune responses: Hormonal effects, antagonistic selection, and evolutionary consequences. Horm. Behav. *88*, 95–105.
17. Garred, P., Larsen, F., Seyfarth, J., Fujita, R., and Madsen, H.O. (2006). Mannose-binding lectin and its genetic variants. Gene Immun. *7*, 85–94.
18. Padyukov, L. (2022). Genetics of rheumatoid arthritis. Semin. Immunopathol. *44*, 47–62.
19. Kapferer-Seebacher, I., Pepin, M., Werner, R., Aitman, T.J., Nordgren, A., Stoiber, H., Thielens, N., Gaboriaud, C., Amberger, A., Schossig, A., et al. (2016). Periodontal Ehlers-Danlos Syndrome Is Caused by Mutations in C1R and C1S, which Encode Subcomponents C1r and C1s of Complement. Am. J. Hum. Genet. *99*, 1005–1014.
20. van Dijk, F.S., Ghali, N., Demirdas, S., and Baker, D. (1993). TNXB-Related Classical-Like Ehlers-Danlos Syndrome. In GeneReviews, M.P. Adam, et al., eds. (University of Washington, Seattle).
21. The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. https://www.science.org/doi/10.1126/science.aaz1776.
22. Xiang, R., Fang, L., Liu, S., Macleod, I.M., Liu, Z., Breen, E.J., Gao, Y., Liu, G.E., Tenesa, A., et al.; CattleGTEx Consortium (2023). Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle. Cell Genom. *3*, 100385.
23. Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P.C., Hurle, M.R., Gutteridge, A., Erola, P., Liu, Y., Luo, S., et al. (2020). Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat. Genet. *52*, 1122–1131.
24. Burgess, S., Davey Smith, G., Davies, N.M., Dudbridge, F., Gill, D., Glymour, M.M., Hartwig, F.P., Kutalik, Z., Holmes, M.V., Minelli, C., et al. (2019). Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. *4*, 186.
25. Pietzner, M., Budde, K., Homuth, G., Kastenmüller, G., Henning, A.K., Artati, A., Krumsiek, J., Völzke, H., Adamski, J., Lerch, M.M., et al. (2018). Hepatic Steatosis Is Associated With Adverse Molecular Signatures in Subjects Without Diabetes. J. Clin. Endocrinol. Metab. *103*, 3856–3868.
26. Gaya da Costa, M., Poppelaars, F., van Kooten, C., Mollnes, T.E., Tedesco, F., Würzner, R., Trouw, L.A., Truedsson, L., Daha, M.R., Roos, A., and Seelen, M.A. (2018). Age and Sex-Associated Changes of Complement Activity and Complement Levels in a Healthy Caucasian Population. Front. Immunol. *9*, 2664.
27. Sorensen, G.L., Petersen, I., Thiel, S., Fenger, M., Christensen, K., Kyvik, K.O., Sørensen, T.I.A., Holmskov, U., and Jensenius, J.C. (2007). Genetic influences on mannan-binding lectin (MBL) and mannan-binding lectin associated serine protease-2 (MASP-2) activity. Genet. Epidemiol. *31*, 31–41.
28. Husby, S., Herskind, A.M., Jensenius, J.C., and Holmskov, U. (2002). Heritability estimates for the constitutional levels of the collectins mannan-binding lectin and lung surfactant protein D. A study of unselected like-sexed mono- and dizygotic twins at the age of 6-9 years. Immunology *106*, 389–394.
29. Sun, B.B., Chiou, J., Traylor, M., Benner, C., Hsu, Y.-H., Richardson, T.G., Surendran, P., Mahajan, A., Robins, C., Vasquez-Grinnell, S.G., et al. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. Nature *622*, 329–338. https://doi.org/10.1038/s41586-023-06592-6.
30. Adamiak, M., Cymer, M., Anusz, K., Tracz, M., and Ratajczak, M.Z. (2020). A Novel Evidence That Mannan Binding Lectin (MBL) Pathway of Complement Cascade Activation is Involved in Homing and Engraftment of Hematopoietic Stem Progenitor Cells (HSPCs). Stem Cell Rev. Rep. *16*, 693–701.
31. Garred, P., Genster, N., Pilely, K., Bayarri-Olmos, R., Rosbjerg, A., Ma, Y.J., and Skjoedt, M.O. (2016). A journey through the lectin pathway of complement-MBL and beyond. Immunol. Rev. *274*, 74–97.
32. Jesty, J., and Beltrami, E. (2005). Positive feedbacks of coagulation: their role in threshold regulation. Arterioscler. Thromb. Vasc. Biol. *25*, 2463–2469.
33. Hansen, S., Selman, L., Palaniyar, N., Ziegler, K., Brandt, J., Kliem, A., Jonasson, M., Skjoedt, M.O., Nielsen, O., Hartshorn, K., et al. (2010). Collectin 11 (CL-11, CL-K1) is a MASP-1/3-associated plasma collectin with microbial-binding activity. J. Immunol. Baltim. Md *185*, 6096–6104.

34. Barratt, J., Lafayette, R.A., Zhang, H., Tesar, V., Rovin, B.H., Tumlin, J.A., Reich, H.N., and Floege, J. (2023). IgA nephropathy: the lectin pathway and implications for targeted therapy. Kidney Int. *104*, 254–264.

35. Dobie, R., Wilson-Kanamori, J.R., Henderson, B.E.P., Smith, J.R., Matchett, K.P., Portman, J.R., Wallenborg, K., Picelli, S., Zagorska, A., Pendem, S.V., et al. (2019). Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme during Liver Fibrosis. Cell Rep. *29*, 1832–1847.e8.

36. Farrar, C.A., Tran, D., Li, K., Wu, W., Peng, Q., Schwaeble, W., Zhou, W., and Sacks, S.H. (2016). Collectin-11 detects stress-induced L-fucose pattern to trigger renal epithelial injury. J. Clin. Invest. *126*, 1911–1925.

37. Van Den Eynde, B.J., Gaugler, B., Probst-Kepper, M., Michaux, L., Devuyst, O., Lorge, F., Weynants, P., and Boon, T. (1999). A new antigen recognized by cytolytic T lymphocytes on a human kidney tumor results from reverse strand transcription. J. Exp. Med. *190*, 1793–1800.

38. Lu, N., Li, Y., Zhang, Z., Xing, J., Sun, Y., Yao, S., and Chen, L. (2018). Human Semaphorin-4A drives Th2 responses by binding to receptor ILT-4. Nat. Commun. *9*, 742.

39. Murata, K., and Baldwin, W.M. (2009). Mechanisms of complement activation, C4d deposition, and their contribution to the pathogenesis of antibody-mediated rejection. Transplant. Rev. *23*, 139–150.

40. Reis, E.S., Mastellos, D.C., Hajishengallis, G., and Lambris, J.D. (2019). New insights into the immune functions of complement. Nat. Rev. Immunol. *19*, 503–516.

41. Imai, N., Nishi, S., Alchi, B., Ueno, M., Fukase, S., Arakawa, M., Saito, K., Takahashi, K., and Gejyo, F. (2006). Immunohistochemical evidence of activated lectin pathway in kidney allografts with peritubular capillary C4d deposition. Nephrol. Dial. Transplant. *21*, 2589–2595.

42. Dotz, V., and Wuhrer, M. (2016). Histo-blood group glycans in the context of personalized medicine. Biochim. Biophys. Acta *1860*, 1596–1607.

43. Nauta, A.J., Raaschou-Jensen, N., Roos, A., Daha, M.R., Madsen, H.O., Borrias-Essers, M.C., Ryder, L.P., Koch, C., and Garred, P. (2003). Mannose-binding lectin engagement with late apoptotic and necrotic cells. Eur. J. Immunol. *33*, 2853–2863.

44. Feijge, M.A.H., Ansink, K., Vanschoonbeek, K., and Heemskerk, J.W.M. (2004). Control of platelet activation by cyclic AMP turnover and cyclic nucleotide phosphodiesterase type-3. Biochem. Pharmacol. *67*, 1559–1567.

45. Del Conde, I., Crúz, M.A., Zhang, H., López, J.A., and Afshar-Kharghan, V. (2005). Platelet activation leads to activation and propagation of the complement system. J. Exp. Med. *201*, 871–879.

46. Yang, K., Han, J., Gill, J.G., Park, J.Y., Sathe, M.N., Gattineni, J., Wright, T., Wysocki, C., de la Morena, M.T., and Yan, N. (2022). The mammalian SKIV2L RNA exosome is essential for early B cell development. Sci. Immunol. *7*, eabn2888.

47. Eckard, S.C., Rice, G.I., Fabre, A., Badens, C., Gray, E.E., Hartley, J.L., Crow, Y.J., and Stetson, D.B. (2014). The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. Nat. Immunol. *15*, 839–845.

48. Würzner, R., Joysey, V.C., and Lachmann, P.J. (1994). Complement component C7. Assessment of in vivo synthesis after liver transplantation reveals that hepatocytes do not synthesize the majority of human C7. J. Immunol. Baltim. Md *152*, 4624–4629.

49. Würzner, R. (2000). Modulation of complement membrane attack by local C7 synthesis. Clin. Exp. Immunol. *121*, 8–10.

50. Esparza-Gordillo, J., Soria, J.M., Buil, A., Almasy, L., Blangero, J., Fontcuberta, J., and Rodríguez de Córdoba, S. (2004). Genetic and environmental factors influencing the human factor H plasma levels. Immunogenetics *56*, 77–82.

51. Reichhardt, M.P., Holmskov, U., and Meri, S. (2017). SALSA—A dance on a slippery floor with changing partners. Mol. Immunol. *89*, 100–110.

52. Gunput, S.T., Wouters, D., Nazmi, K., Cukkemane, N., Brouwer, M., Veerman, E.C., and Ligtenberg, A.J. (2016). Salivary agglutinin is the major component in human saliva that modulates the lectin pathway of the complement system. Innate Immun. *22*, 257–265.

53. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3; and Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

54. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. *6*, 5890.

55. Luo, Y., Kanai, M., Choi, W., Li, X., Sakaue, S., Yamamoto, K., Ogawa, K., Gutierrez-Arcelus, M., Gregersen, P.K., Stuart, P.E., et al. (2021). A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. Nat. Genet. *53*, 1504–1516.

56. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.

57. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

58. Kamat, M.A., Blackshaw, J.A., Young, R., Surendran, P., Burgess, S., Danesh, J., Butterworth, A.S., and Staley, J.R. (2019). PhenoScanner V2: an expanded tool for searching human genotype–phenotype associations. Bioinformatics *35*, 4851–4853.

59. Niemi, M.E.K., Karjalainen, J., Liao, R.G., Neale, B.M., Daly, M., Ganna, A., Pathak, G.A., Andrews, S.J., Kanai, M., Veerapen, K., et al. (2021). Mapping the human genetic architecture of COVID-19. Nature *600*, 472–477.

60. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012.

61. Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., et al. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat. Genet. *53*, 420–425.

62. Jupp, S., Burdett, T., Leroy, C., and Parkinson, H.E. (2015). A new Ontology Lookup Service at EMBL-EBI. SWAT4LS *2*, 118–119.

63. Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.

64. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.

65. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Software *67*, 1–48.

66. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354.

67. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinforma. Oxf. Engl. *26*, 2336–2337.

68. Hadfield, J.D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. J. Stat. Software *33*, 1–22.

69. Machiela, M.J., and Chanock, S.J. (2015). a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinforma. Oxf. Engl. *31*, 3555–3557.

70. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K., and Kastenmüller, G. (2015). SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics *31*, 1334–1336.

71. Broadbent, J.R., Foley, C.N., Grant, A.J., Mason, A.M., Staley, J.R., and Burgess, S. (2020). MendelianRandomization v0.5.0: updates to an R package for performing Mendelian randomization analyses using summarized data. Wellcome Open Res. *5*, 252.

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

73. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife *7*, e34408.

74. Noce, D., Gögele, M., Schwienbacher, C., Caprioli, G., De Grandi, A., Foco, L., Platzgummer, S., Pramstaller, P.P., and Pattaro, C. (2017). Sequential recruitment of study participants may inflate genetic heritability estimates. Hum. Genet. *136*, 743–757.

75. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics *55*, 997–1004.

76. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M., and Poland, G.A. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am. J. Hum. Genet. *70*, 425–434.

77. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

78. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.

79. Li, B., and Martin, E.B. (2002). An approximation to the F distribution using the chi-square distribution. Comput. Stat. Data Anal. *40*, 21–26.

80. Thomas, D.C., Lawlor, D.A., and Thompson, J.R. (2007). Estimation of Bias in Nongenetic Observational Studies Using "Mendelian Triangulation" by Bautista et al. Ann. Epidemiol. *17*, 511–513.

81. Greco M, F.D., Minelli, C., Sheehan, N.A., and Thompson, J.R. (2015). Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat. Med. *34*, 2926–2940.

82. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet. Epidemiol. *40*, 304–314.

83. Hartwig, F.P., Davey Smith, G., and Bowden, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int. J. Epidemiol. *46*, 1985–1998.

84. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int. J. Epidemiol. *44*, 512–525.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Genomic DNA stored at −20°C	Human	CHRIS study
Serum stored at −80°C	Human	CHRIS study
Critical commercial assays		
WIESLAB® complement system screen kit	SVAR, Malmö https://www.svarlifescience.com/products/compl300ruo	COMPL300RUO
TaqMan predesigned assay for rs74480769 genotyping	ThermoFisher	Assay ID: C_100391655_10 Catalog # 4351379
HumanOmniExpressExome Bead array v1.2	Illumina	WG-351-2304
rowheadDeposited data		
Complement activation measurements	This study	The CHRIS Access Committee at access.request.biomedicine@eurac.edu
GWAS summary statistics	This study	The CHRIS Access Committee at access.request.biomedicine@eurac.edu
Associations of identified variants with complement proteins	This study	Table S1K
Associations of identified variants with 4937 proteins	This study	Table S1L
Associations of identified variants with metabolite levels	This study	Table S1M
Results of phenome-wide scan of 2783 diseases and traits	This study	Table S1N
Results of colocalization analysis with protein levels	This study	Table S1O
Selected instrumental variables for proteins used as exposures in Mendelian randomization analyses of complement pathways.	This study	Table S1P
Causal effects of non-complement proteins on complement traits	This study	Table S1Q
Causal effects of complement proteins on complement traits	This study	Table S1R
Mendelian randomization results of C3d/C3 on CP, LP, and AP.	This study	Table S1S
Selection of instrumental variables for CP, LP, and AP used as exposures in Mendelian randomization analyses	This study	Table S1T
Causal effects of LP on protein levels	This study	Table S1U
Causal effects of LP on human metabolites	This study	Table S1V
Causal effects of LP on complex traits and diseases	This study	Table S1W
Causal effects of AP on protein levels	This study	Table S1X
Causal effects of AP on human metabolites	This study	Table S1Y
Causal effects of AP on complex traits and diseases	This study	Table S1Z
Haplotype Reference Consortium dataset v1.1	McCarthy et al.[56]	https://www.sanger.ac.uk/collaboration/haplotype-reference-consortium/
Ensembl VEP in GRCh37 release 104 (May 2021)	McLaren et al.[57]	https://grch37.ensembl.org/info/docs/tools/vep/index.html
PhenoScanner v2 (accessed 5 May 2021)	Kamat et al.[58]	http://www.phenoscanner.medschl.cam.ac.uk/
COVID19-hg GWAS meta-analyses round 5	Niemi et al.[59]	https://www.covid19hg.org/results/r5/
C3d-to-C3 ratio association results	Lorés-Motta et al.[11]	https://www.aaojournal.org/article/S0161-6420(17)32520-4/fulltext
Proteome GWAS summary statistics (1)	Pietzner et al.[13]	https://omicscience.org/apps/pgwas/
GWAS Catalog	Buniello et al.[60]	https://www.ebi.ac.uk/gwas/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Polygenic Score Catalog	Lambert et al.[61]	https://www.pgscatalog.org/
Ontology Lookup Service	Jupp et al.[62]	https://www.ebi.ac.uk/ols/index
ICD10 classification	Service web page	https://www.cdc.gov/nchs/icd/icd10.htm
MeSH classification	Service web page	https://meshb.nlm.nih.gov/
Genotype-Tissue Expression (GTEx) Project database v8	GTEx consortium[21]	https://www.gtexportal.org/home/
Proteome GWAS summary statistics (2)	Sun et al.[12]	https://www.nature.com/articles/s41586-018-0175-2
NCBI Gene	NCBI web page	https://www.ncbi.nlm.nih.gov/gene/
Software and algorithms		
Data analysis scripts	This study	Data S1
SHAPEIT2	Delaneau et al.[63]	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html#home
minimac3	Das et al.[64]	https://github.com/Santy-8128/Minimac3
R package 'lme4' v3.4.1	Bates et al.[65]	https://cran.r-project.org/web/packages/lme4
EMMAX	Kang et al.[66]	https://genome.sph.umich.edu/wiki/EMMAX
EPACTS v3.3.0	Software web page	https://github.com/statgen/EPACTS
R package 'CMplot' v3.6.2	Software web page	https://cran.r-project.org/web/packages/CMplot
LocusZoom v0.4.8	Pruim et al.[67]	http://locuszoom.org
Haplostats v1.8.9	Software web page	https://cran.r-project.org/web/packages/haplo.stats/index.html
R package 'coxme' v2.2.16	Software web page	https://cran.r-project.org/web/packages/coxme/index.html
R package 'MCMCglmm' v3.5.1	Hadfield et al.[68]	https://cran.r-project.org/web/packages/MCMCglmm/index.html
LDlink 5.1 release 5/25/2021	Machiela et al.[69]	https://ldlink.nih.gov/?tab=home
SNiPA v3.2	Arnold et al.[70]	https://snipa.org/snipa3/
Bioconductor package 'ontoProc' v1.18.0	Software web page	https://www.bioconductor.org/packages/release/bioc/html/ontoProc.html
ZOOMA	Software web page	https://www.ebi.ac.uk/spot/zooma/
R package 'gtx' v2.1.6	Software web page	https://github.com/tobyjohnson/gtx
Swiss v1.1.0	Software web page	https://github.com/statgen/swiss
R package 'MendelianRandomization' v0.5.1	Broadbent et al.[71]	https://CRAN.R-project.org/package=MendelianRandomization
R package 'MR-PRESSO' v1.0	Verbanck et al.[72]	https://github.com/rondolab/MR-PRESSO
R package 'TwoSampleMR' v0.5.6	Hemani et al.[73]	https://mrcieu.github.io/TwoSampleMR
R software package v4.1.0	Software web page	www.r-project.org
RStudio v1.4.1106© (2009–2021 RStudio, PBC)	Software web page	https://posit.co/download/rstudio-desktop/
draw.io v19.0.3	Software web page	https://app.diagrams.net/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Cristian Pattaro (Cristian.Pattaro@eurac.edu).

Materials availability

This study did not generate any biochemical material.

Data and code availability

- Publicly available data used in this work are detailed in the key resources table. CHRIS study data used in this work can be requested with an application to the CHRIS Access Committee at access.request.biomedicine@eurac.edu.

- The original code used to analyze the data has been included in the Data S1 file.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study design

The CHRIS study is a population-based study, whose baseline assessment was carried out in South Tyrol (Italy) between 2011 and 2018.[14,74] All participants underwent blood drawing, urine collection, anthropometric analysis, and clinical assessments in the early morning following overnight fasting. Medical history was reconstructed via interviewer- and self-administered questionnaires. The study received ethical approval by the Ethical Committee of the Healthcare System of the Autonomous Province of Bolzano. All participants gave written informed consent.

The present analysis involved the 4990 participants included in the first release of the CHRIS study (age 46 ± 16 years; 56% females).

METHOD DETAILS

Assessment of functional complement activity

We assessed activity of CP, LP, and AP on serum samples. Samples, stored at −80°C since the day of collection, were analyzed with WIESLAB complement system screen kit (SVAR, Malmö, formerly Wieslab AB, Lund, Sweden) according to manufacturers' instructions. This enzyme immunoassay measures the functional activity of CP, LP and AP using deposition of the terminal C5-9 complex in microtiter wells as a readout. The wells of the microtiter plate strips are coated with specific activators of the three pathways. Sera were thawed, vortexed and centrifuged, then diluted with a diluent containing specific blockers to ensure that only the respective pathway was activated. The positive (a pool of 5 sera from healthy individuals) and negative (sera heat-inactivated at 56°C for 20 min) controls were provided by the manufacturer. Immediately following sample thawing, complement activation was measured as absorbance (optical density, OD) at 405 nm (nm). The negative and positive controls were used in a semiquantitative way to calculate complement activity as (OD – negative control value)/(positive control value – negative control value)×100.

Genotyping and genetic imputation

Study participants were genotyped using the Illumina HumanOmniExpressExome Bead array. Genotype quality control and filtering, based on call rate ≥99%, sex mismatch check, duplicate removal, outlier detection via principal component analysis, Hardy Weinberg Equilibrium p value$\geq 1 \times 10^{-6}$, and MAF≥0.01, left 610,883 SNPs available for imputation. Haplotypes were phased with SHAPEIT2.[63] Genotype imputation was performed based on the Haplotype Reference Consortium[56] dataset release 1.1 using minimac3.[64] This resulted into 7,721,937 SNPs, expressed as allelic dosage levels, with imputation quality score Rsq≥0.3 and MAF≥0.01.

Validation of the C7 SNP rs74480769

The lead SNP rs74480769 associated with AP in the *C7* locus was is low LD with nearby variants. To exclude that the observed association was due to genotype imputation artifacts, we thus genotyped this SNP on 4605 participants with available AP measurements using a commercial ThermoFisher TaqMan assay (20 ng DNA in 3 μL reaction volume using 384-well qPCR plates). All pipetting was performed using liquid handling robotics (Tecan AG, Männedorf, CH), using a ThermoFisher QuantStudio 6 qPCR system (ThermoFisher, Waltham, MA, USA) for fluorescence reading. As a quality control assessment, we typed 250 samples (4.8%) twice, observing no between-measurement discordance. A call rate of 98.2% and no deviation from Hardy-Weinberg equilibrium (p-value = 0.283) were observed.

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome-wide association studies

The main analysis was conducted on untransformed CP and AP activation levels after removing values < 5%, to guarantee gaussianity and avoid bimodality (Figure 1A). LP was normalized using inverse-normal transformation. Each complement pathway (trait) was then regressed on sex, age, plate, and day of experiment using the following linear mixed effect model (LMM): $trait_{ijk} = \beta_0 + \beta_1 age + \beta_2 sex + \varepsilon_{ijk}$, where: i, j, and k, indicate the individual, the day, and the plate, respectively, and $\varepsilon_{ijk} = day_j + plate_{jk} + u_{ijk}$, is a normally distributed error term expressed as the sum of day effect, plate effect, and a random error u, that is, $day_j \sim N(0, \sigma_1^2)$, $plate_{jk} \sim N(0, \sigma_2^2)$, and $u_{ijk} \sim N(0, \sigma^2)$. Models were fitted using the R package 'lme4' v3.4.1 (https://cran.r-project.org/web/packages/lme4).[65]

GWAS of the fixed effect residuals were conducted on the dosage levels, assuming additive genetic effects, using EMMAX[66] as implemented in the EPACTS analysis package v3.3.0 (https://genome.sph.umich.edu/wiki/EPACTS). Genomic inflation was quantified by the genomic control (GC) factor λ.[75] Genome-wide significance was set at 5×10^{-8}. Manhattan plots were generated with the R package 'CMplot' v3.6.2 (https://cran.r-project.org/web/packages/CMplot), and regional association plots with LocusZoom

v0.4.8 (http://locuszoom.org).[67] A genome-wide significant locus was defined as a 1 Mb segment centered on the SNP with the lowest p-value of association with the trait.

Alternative modeling of the pathways' distributions were considered: (i) quantitative levels of untransformed CP and AP over their full range (bimodal distributions), using the same LMM approach described for the main analyses; and (ii) dichotomized CP, LP, and AP levels (values < 5% against those ≥20%), with GWAS conducted based on logistic regression using the same software and including the first 10 genetic principal components to account for relatedness. The 5% cutoff was defined to represent deficiency, according to Seelen et al.[6] who observed that all deficient sera had values < 5%. Results of these alternative models and comparisons with main analysis are reported in Figures S8, S9, S10, Tables S1D, and S1E.

Association of *C7* SNP rs74480769 with AP

Associations were tested by fitting linear mixed association models, including a genetic relatedness matrix, using the *lmekin* function in the R package 'coxme' v2.2.16 (https://cran.r-project.org/web/packages/coxme/index.html).

Haplotype analysis

Using the best-guessed imputed genotypes, haplotype regression analysis of LP was conducted using the 'haplo.glm' function of the R package 'haplo.stats' v1.8.9, which exploits an expectation maximization (EM) algorithm for haplotype inference.[76] Alleles were aligned to have the major allele as the reference. Haplotypes with <0.001 frequency were collapsed into a rare haplotype category. The EM control parameters were set as follows: n.try = 2 (number of times to try to maximize the log likelihood); insert.batch.size = 2 (number of loci to be inserted in a single batch); max.haps.limit = 4e6 (maximum number of haplotypes for the input genotypes); min.posterior = 1e-5 (minimum posterior probability for a haplotype pair, given the input genotypes). After the analysis, for illustrative purposes, alleles were converted to the actual reference or alternate alleles based on their non-reference allele frequency.

Conditional analyses and interaction with sex and age

Conditional analyses, to verify independence of genome-wide significant SNPs and to identify additional, conditionally significant SNPs, were performed on the individual-level data using the same LMM defined above and including significant SNPs in turn. For the 13 independent, genome-wide significant lead SNPs, SNP-by-age and SNP-by-sex interaction was assessed using the same LMM approach, including the main effects of both terms and their interaction, setting statistical significance at $\alpha = 0.05/13 = 3.8 \times 10^{-3}$.

Genetic heritability explained by the identified variants

We estimated the age- and sex-adjusted genetic heritability (h_0^2) of CP, LP, and AP, exploiting the pedigree structure of the CHRIS study, using the Bayesian analysis package 'MCMCglmm' v3.5.1 in R,[68] running 1,000,000 MCMC iterations (burn-in = 500,000 iterations; thinning lag = 1) as described previously.[74] We repeated the analysis including all independently and conditionally significant SNPs in models, to estimate the residual genetic heritability (h_r^2). The relative difference $(h_0^2 - h_r^2)/h_0^2$ estimates the proportion of genetic heritability explained by the identified variants.

Variant annotation

Prediction of the primary effect of each identified variant was performed with the Ensembl VEP tool included in the Ensembl GRCh37 release 104 (May 2021).[57] VEP was used also to retrieve the Combined Annotation Dependent Depletion (CADD) PHRED-like score[77] to assess variants' deleteriousness. Within each locus, the same annotation was extended to variants in strong LD ($r^2 \geq 0.8$) with the lead SNP and within ±250Kb around it. LD analysis was performed with LDlink 5.1 release (5/25/2021)[69] based on the 1000 Genomes phase 3 version 5 (1000Gv3p5) CEU datasets. We further used SNiPA v3.2 to integrate additional, available information.[70]

Shared genetic background with human traits and diseases

We tested whether genetic variants associated with CP, LP, and AP were also associated with other traits and diseases, through a phenome-wide scan conducted at different levels. Metabolites and diseases and traits were interrogated through PhenoScanner v2,[70] accessed on 5 May 2021. Phenotype pruning involved: removal of non-European ancestry studies; selection of the most recent and largest study among multiple studies on the same phenotype; selection of the study with the largest number of cases, when binary traits were concerned; removal of records with missing association effect, SE or p-value. GWAS on COVID-19-related phenotypes were retrieved from the COVID19-hg GWAS meta-analyses round 5 (https://www.covid19hg.org/results/r5/).[59] Genetic associations with the C3d-to-C3 ratio were retrieved from a recent work.[11] Protein-SNP associations were derived from the Pietzner et al. proteomics GWAS.[13] These interrogations resulted into 4979 proteins, including 40 complement-related proteins, 671 metabolites, and 2783 diseases and traits. Given the strong *a priori* belief and the confirmatory nature of the analysis, associations with complement-related proteins were evaluated at $\alpha = 0.05$. All other associations were tested at the genome-wide significance level $\alpha = 5 \times 10^{-8}$. Diseases and traits were mapped to the Experimental Factor Ontology (EFO) using a combination of different sources: PhenoScanner; GWAS Catalog[60] (https://www.ebi.ac.uk/gwas/) when EFO codes were not available on PhenoScanner; the 'ontoProc' Bioconductor package v1.18.0 (https://www.bioconductor.org/packages/release/bioc/html/ontoProc.html); and ZOOMA (https://www.ebi.ac.uk/spot/zooma/). The coded traits were then categorized into parent terms according to the Polygenic Score

Catalog,[61] the Ontology Lookup Service[62] (https://www.ebi.ac.uk/ols/index), and the ICD10 (https://www.cdc.gov/nchs/icd/icd10.htm) and MeSH (https://meshb.nlm.nih.gov/) classifications.

Colocalization analyses

We conducted colocalization analysis of CP, LP, and AP genetic associations with the expression levels of protein-coding genes located in *cis* of the lead SNP and the corresponding protein levels. We used the "White" samples in the Genotype-Tissue Expression (GTEx) Project database v8 as reference for genetic associations with gene expression.[21] The Pietzner et al. dataset[13] was used as reference for genetic associations with protein levels. Results from a former proteomic GWAS[12] were used for sensitivity analysis in specific cases. The effect alleles were harmonized across transcriptomic, proteomic, and GWAS analyses. For each locus, we identified tissue-gene pairs with reported gene expression data within ±100 kb of each GWAS index SNP. Sensitivity analyses were conducted at window sizes of ±500 kb. We used the R package 'gtx' (https://github.com/tobyjohnson/gtx; function: *coloc.fast*), which implements and adapts the Giambartolomei's method,[78] with default prior setting.

Mendelian randomization (MR) analyses

We conducted multiple two-sample MR analyses to fulfill different purposes. First we tested causal effects of proteins on complement pathways, exploiting GWAS results from Pietzner et al.[13] We restricted the analysis to 26 proteins that either showed colocalization ($PP_{H4} > 0.8$) with at least one complement pathway (21 instances) or for which the expression of the encoding gene showed colocalization with a complement pathway (11 instances). In addition, to corroborate known biology, we selected all 40 complement proteins available. Overall, we tested the causal effect of 66 proteins on the three pathways, setting $\alpha = 0.05/66 = 7.6 \times 10^{-4}$.

Second, we tested whether complement activation had any causal effect on any of 4979 available proteins ($\alpha = 0.05/4979 = 1.0 \times 10^{-5}$), 671 metabolites ($\alpha = 0.05/671 = 7.5 \times 10^{-5}$) or 2783 diseases and traits ($\alpha = 0.05/2783 = 1.8 \times 10^{-5}$).

Genetic association results for all involved proteins, metabolites, diseases and traits were obtained as outlined in the "*Shared genetic background with human traits and diseases*" section.

In all cases, SNPs were selected for use as IVs if their association with the exposure was genome-wide significant and had an F statistic of >10 (estimated as the square of the effect estimate over its SE[79]), provided that the SNPs were also present in the GWAS of the outcome. LD pruning was performed with Swiss v1.1.0 (https://github.com/statgen/swiss), starting by retaining the SNP with the smallest p-value and removing all non-independent SNPs (LD $r^2 < 0.01$), iteratively until saturation of the SNP set.

When proteins were chosen as the exposures, SNPs for use as IVs were chosen in *trans* (i.e.,: regardless of the location in the human genome), in *cis* (i.e.,: within ±500 kb from the encoding genes boundaries derived from the NCBI gene catalog, https://www.ncbi.nlm.nih.gov/gene/), and inside the encoding genes (Table S1P).

When using the complement pathways as exposures, IVs were selected from the unconditional analyses (Table S1T), based on literature review and specificity of the association with the relevant gene expression or protein level as observed in our experimental analyses (phenome-wide scans; colocalization analysis). SNPs showing potential horizontal pleiotropic effects were excluded. We didn't conduct MR analyses using CP as exposure because all candidate IVs were located in the MHC, making it difficult to exclude horizontal pleiotropy.

For each IV, the MR effect and its SE were estimated with the Wald estimator with simple weights.[80] When multiple IVs were available, results were pooled via fixed-effect inverse-variance-weighted (IVW) meta-analysis, assessing the between-IV homogeneity with the Cochran's Q test.[81] Sensitivity analyses that challenged the results' robustness and tested for horizontal pleiotropy were conducted using the weighted median,[82] mode-based estimation,[83] MR-Egger,[84] and MR-PRESSO[72] methods.

All MR analyses were conducted with the R packages 'MendelianRandomization' v0.5.1 (https://CRAN.R-project.org/package=MendelianRandomization)[71] and 'MR-PRESSO' v1.0 (https://github.com/rondolab/MR-PRESSO),[72] and 'TwoSampleMR' v0.5.6 (https://mrcieu.github.io/TwoSampleMR) for allele harmonization.[73]

Software

When not otherwise specified, statistical analyses were conducted with the R software package v4.1.0 (www.r-project.org) also using RStudio v1.4.1106© (2009–2021 RStudio, PBC). Figures 2A and 5A were generated with draw.io v19.0.3 (diagrams.net).