Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Human Plasma Causal Immunome-Phenome Mapping Reveals Cardiovascular Enrichment and Therapeutic Target Opportunities

Shila Taslima
University of Alabama at Birmingham
John Zhang
University of Alabama at Birmingham
Huilan Tan
University of Alabama at Birmingham
Yanwen Liu
University of Alabama at Birmingham
Dhrubo Ahmad
University of Alabama at Birmingham
Martin Young
University of Alabama at Birmingham
Suresh Verma
University of Alabama at Birmingham
Min Xie
University of Alabama at Birmingham
Jianli Zhao
University of Alabama at Birmingham
Yajing Wang
yajingwang@uab.edu
University of Alabama at Birmingham

Research Article

Keywords: plasma proteome, plasma immunome, Mendelian randomization, colocalization, cardiovascular disease, drug target discovery, risk prediction

Posted Date: October 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11108738/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Human Plasma Causal Immunome-Phenome Mapping Reveals Cardiovascular Enrichment and Therapeutic Target Opportunities

Shila Taslima, BS[1#]; John Zhang, BS, MS[1#]; Huilan Tan, PhD[1]; Yanwen Liu, PhD[1]; Dhrubo Ahmad, BS, MS[1]; Martin Young, PhD[2]; Suresh Verma, PhD[2]; Min Xie, MD, PhD[2]; Jianli Zhao, MD, MS[1*]; Yajing Wang, MD, PhD[1*]

[1]*Department of Biomedical Engineering, University of Alabama at Birmingham, Birmingham, AL 35294, US*

[2]*Cardiology Institution, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, US*

[#]*These authors contributed equally to this work.*

Correspondence to:

Yajing Wang, MD, PhD

Professor, Department of Biomedical Engineering

University of Alabama at Birmingham

Birmingham, AL 35294

Email: yajingwang@uab.edu

Jianli Zhao, MD, MS

Assistant Professor, Department of Biomedical Engineering

University of Alabama at Birmingham

Birmingham, AL 35294

Email: jianlizhao@uab.edu

Short title: Plasma Immunome Causal Atlas of Cardiovascular Risk

Total word count: Manuscript (Background through Conclusions): 7801. Abstract: 299. References: 59. Figures: 5. Tables: 5.

Abstract

Background: Circulating immune proteins represent an important class of therapeutic targets, but existing proteome-wide genetic studies have not systematically resolved immune-related target–disease relationships across a broad disease phenome within a unified framework that integrates causal inference, locus colocalization, orthogonal protein-level support, independent replication, and standardized evidence grading.

Methods: We developed the Human Plasma Immune Atlas (HPIA), a tiered proteogenomic framework built from public datasets. We curated 1,007 plasma immune proteins from 2,923 Olink Explore analytes, instrumented each encoding gene with its strongest blood cis-expression quantitative trait locus (eQTLGen), and tested effects on 2,469 FinnGen R12 disease endpoints by Mendelian randomization (MR), followed by colocalization, protein-level validation with INTERVAL cis-protein quantitative trait loci, 2-population replication in UK Biobank, and prespecified 5-tier evidence grading. A companion Plasma Immune Risk Score (PIRS; elastic-net Cox) is released for predicting targets.

Results: Across 1,656,872 evaluable gene–disease tests, 1,016 associations involving 220 immune genes and 381 diseases met FDR <0.05, of which 417 showed strong evidence of colocalization (PP.H4 ≥0.80). Positive-control analyses recapitulated five established immune drug-target relationships and the ACE–hypertension relationship with directionally concordant genetic effects. Among 189 discovery associations evaluable in UK Biobank, 162 (86%) showed concordant effect directions and 89 met the prespecified replication criterion. Cardiovascular phenotypes contained the largest number of high-confidence associations, accounting for 67 of 176 core target–disease pairs, while hypertension was associated with 28 immune proteins. Integration of MR, colocalization, protein-level evidence, and replication promoted five target–disease pairs to the highest evidence tier, including ANXA2-hypertension and PPP3R1-venous/lymphatic disease. Systematic prioritization further identified 45 colocalized candidate therapeutic target–indication pairs with genetically inferred directions of intervention. In validation evaluation, the PIRS achieved a cross-validated concordance index of 0.777.

Conclusions: HPIA provides an open, genetics-anchored map connecting the circulating immune system to the human disease phenome and distinguishes initial genetic associations from target–disease relationships supported by orthogonal evidence. The framework identifies cardiovascular phenotypes as a major concentration of genetically supported immune associations and provides experimentally testable hypotheses for therapeutic target discovery, drug repurposing, and immune-informed disease stratification.

Keywords: plasma proteome; plasma immunome; Mendelian randomization; colocalization; cardiovascular disease; drug target discovery; risk prediction

Background

Immune-derived therapeutic targets are still nominated largely from animal models and observational association studies, and both lose accuracy in translation. Mouse and human immune systems differ in leukocyte composition, receptor repertoire, and innate signaling, and mouse transcriptional responses to inflammatory insult correlate poorly with human ones.[1,2] Cardiovascular medicine has faced the consequences directly: canakinumab and low-dose colchicine reduced ischemic events,[3-6] establishing the immune system as a legitimate cardiac therapeutic compartment,[7] yet low-dose methotrexate, tested on the same inflammation premise, did not.[8] By contrast, target-indication pairs supported by human genetic evidence are substantially more likely to succeed in clinical development,[9] which makes accurate, genetics-anchored target selection a first-order problem for new therapies.

Plasma immune proteins are the natural substrate for a genetics-anchored search: many are already direct targets of approved therapy, with interleukin-6 signaling blockade, cytotoxic T-lymphocyte antigen-4 costimulation modulation, and tumor necrosis factor blockade in routine use,[10-12] and a plasma protein sits closer to the pharmacologic point of intervention than a transcript or a variant.[9,10,13] Four gaps nonetheless keep existing plasma-proteomic studies from converting that proximity into targets. First, an observed protein-outcome association is confounded by comorbidity, medication, and reverse causation, so causality cannot be inferred from such data alone. Second, most studies are confined to a single disease or disease family, concealing immune proteins that act across otherwise unconnected organ systems; the phenome-wide proteomic Mendelian randomization (MR) studies that exist do not test an immune panel against a national disease registry.[14-16] Third, the deepest population-scale resources are gated behind restricted access, limiting reproducibility.[17,18] Fourth, a robustly associated protein does not inherently signal whether a therapy should increase or decrease its activity, leaving the therapeutic modality undefined.

To guard against the classic failure modes of proteomic association, neighboring-gene effects, confounding, and the gap between mRNA and protein, we anchored every causal claim not on MR alone but on triangulation across 4 independent evidence layers, each designed to eliminate a distinct class of false positive: (i) transcript-level MR using blood cis-expression quantitative trait loci (eQTLs); (ii)

colocalization, which tests whether a single shared causal variant drives both the expression and the disease signals; (iii) an instrument built from the circulating protein itself (cis-pQTL), closing the mRNA-to-protein gap; and (iv) replication in a second, independent population.

To address these gaps, we developed an immunoproteomic atlas and its accompanying analytical platform. The platform evaluates findings at 2 complementary scales: a deeply annotated cardiovascular-centered disease core and an unbiased survey of the entire FinnGen phenome, ensuring that disease selection was not curated to favor specific outcomes. Three annotation layers then sharpen biological and translational interpretation by identifying the putative immune-cell source of each protein, the genetically inferred direction of therapeutic effect, and its pharmacologic tractability. Every finding is assigned a predefined evidence tier, providing a transparent framework for assessing confidence and clinical relevance, and approved immune and cardiovascular therapeutics serve as end-to-end validation benchmarks for the platform. Building on this foundation, we developed the Plasma Immune Risk Score (PIRS), a machine-learning risk-prediction framework that translates a causally characterized immunome into individualized risk estimates. The PIRS is released as training code and validated end to end on synthetic data; no real-cohort performance is claimed in this article.

Collectively, this resource is designed to address several fundamental clinical questions: To what extent is a cardiovascular disease driven by immune mechanisms, and how much of its causal architecture is immune-mediated? Is a protein detected in a patient a causal contributor to disease or merely a downstream biomarker? Which immune cell population produces the protein? Does the association replicate in an independent population and at the protein level? Finally, can these immune-derived protein signatures be leveraged to generate individualized risk predictions rather than simply population-level associations?

Methods

Every dataset used in this study is publicly accessible and citable (Table 1). All analysis and the precomputed data are released on GitHub (https://github.com/jimmyuab/human-plasma-immune-atlas), with an interactive web application at https://huggingface.co/spaces/jianlizhao/human-plasma-immune-atlas. The study follows the STROBE-MR (Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization) statement,[19,20] and the predictive component, the Plasma Immune Risk Score (PIRS), a machine-learning (elastic-net Cox) risk-model framework, follows the

TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) recommendations.

Study design and data sources

The atlas is built as an evidence chain rather than a single test, because each link removes a class of false positive the preceding one cannot detect. Analyses run in a fixed order (proteome curation, instrument construction, cis-MR, colocalization, claim gating, protein-level validation, replication, prioritization, whole-phenome expansion, cardiovascular validation, and PIRS training), and a finding advances to the next stage only when the accumulated evidence supports the stronger claim. The pipeline is implemented as numbered, deterministic Python scripts with cached intermediates and prespecified thresholds, so the whole workflow regenerates from source.

The protein universe was the Olink Explore panel of 2,923 proximity-extension immunoassays. In each assay, a protein is recognized by 2 oligonucleotide-labeled antibodies and quantified by sequencing. The readout, Normalized Protein eXpression (NPX), is a relative log2 abundance in which 1 unit difference corresponds to a doubling of concentration.[21] Assay identities and assay-to-gene mapping are public, whereas individual-level NPX from UK Biobank is access controlled. Annotations came from the Human Protein Atlas (HPA) master table (immune-cell RNA specificity across sorted blood populations, secretome location, molecular function, drug-target flags)[22,23] and, for labeling only, the Molecular Signatures Database (MSigDB) immunologic (C7) and cell-type (C8) collections.[24]

Transcript-side instruments came from eQTLGen, a blood cis-expression quantitative trait locus (cis-eQTL) meta-analysis in up to 31,684 individuals.[25] Protein-side instruments came from INTERVAL plasma cis-protein quantitative trait loci (cis-pQTLs), obtained as harmonized build-37 summary statistics through the public GWAS Catalog mirror, of which only the cis window (+/-500 kb around the gene body) was retained. Disease outcomes came from FinnGen R12[26] at 2 scales: a deeply annotated 28-disease core spanning 5 categories (autoimmune, 13; cardiovascular, 6; metabolic, 3; renal, 1; neurodegenerative/aging, 5) and all 2,469 diseases, of which 2,466 returned usable data. The 2 scales serve different purposes: the core carries the deepest annotation, while the whole-phenome scan guards against the concern that a curated disease list was chosen to support the result. Replication used UK Biobank (Neale) and other non-FinnGen studies through OpenGWAS.[27] Blood immune-trait outcomes were 15 traits: leukocyte, neutrophil, monocyte, lymphocyte, platelet, reticulocyte, and red-cell counts, 5 differential leukocyte fractions, and C-reactive protein from UK Biobank, with eosinophil and basophil counts drawn from Astle et al so that no lineage was omitted.[28] The UK Biobank Pharma Proteomics Project and deCODE plasma

pQTL resources are access controlled and were excluded rather than replaced with imputed values; that exclusion is reported as a limitation.

Defining the plasma immunome

Each of the 2,923 assays was scored by an additive, transparent rule combining independent evidence of immune relevance and plausible plasma detectability. Membership of core immune pathways (cytokine, chemokine, interferon, tumor necrosis factor superfamily), complement and checkpoint biology, and HPA immune-cell RNA enrichment were weighted most heavily; secreted or receptor location and curated leukocyte-surface, human leukocyte antigen, immunoglobulin, acute-phase, and coagulation flags were credited additionally. An additive score was preferred to a single filter because no one annotation identifies an immune protein reliably, and requiring several weak signals to agree is more robust than trusting any of them alone. Proteins passing the prespecified threshold constitute the plasma immunome. The MSigDB C7 collection was excluded from the score because it covers almost the entire Olink panel and so carries no discriminating information. For every retained protein we stored immune class, blood-cell lineage of origin, HPA protein class, drug-target flags, secretome location, extracellular-vesicle route, and membrane topology, which is what later allows a causal hit to be traced back to its cell type of origin and forward to a drug modality.

Genetic instruments and Mendelian randomization

MR uses genetic variants as instrumental variables for a modifiable exposure, exploiting the random assortment of alleles at conception to estimate causal effects that are largely free of the confounding and reverse causation limiting observational studies.[29,30] For each immune gene we selected the single strongest cis-eQTL in eQTLGen as a 1-variant instrument. A cis variant was preferred because it is far more likely than a trans variant to act on disease through the indexed gene, which is the exclusion-restriction assumption on which MR depends.[31] Because eQTLGen reports a Z score rather than an interpretable effect size, beta and its standard error (SE) were reconstructed from the Z score, effect-allele frequency (p), and sample size (n) using the Zhu transformation[32]:

$$denom = 2*p*(1-p)*(n+Z^2)$$

$$beta = Z/sqrt(denom);\ SE = 1/sqrt(denom)$$

Exposure and outcome were harmonized to a common effect allele before any ratio was formed. Strand-ambiguous palindromic variants (A/T, C/G) were removed rather than inferred, alleles were aligned to the expression-increasing allele with the disease effect flipped to match, and variants absent from the

outcome study were dropped. Applying this identically in every analysis arm makes the direction of any 2 reported effects comparable. With a single instrument the causal effect is the Wald ratio,

$$beta(MR) = beta(outcome)/beta(exposure);\ SE(MR) = SE(outcome)/|beta(exposure)|$$

exponentiated to an odds ratio (OR), so OR >1 means genetically higher expression raises risk and OR <1 means protection. That sign is retained throughout and later becomes the therapeutic direction. Significance was controlled at a 5% false discovery rate (FDR) by the Benjamini-Hochberg procedure rather than by Bonferroni correction, because the tests are correlated through shared instruments and related diseases, and because the aim at this stage is a credible candidate set for orthogonal triangulation, not a terminal confirmatory analysis.[33] For every reported association, the exposure variant, effect allele, eQTL P value, instrument strength (F statistic), and variance explained are provided in the released atlas tables (Data S1), together with neighboring-gene context for every high-priority target.

Two safeguards were fixed in advance. The major histocompatibility complex (MHC) region on chromosome 6, with the linked complement genes C2, C4A, and C4B, was capped at the level of a nomination regardless of statistical significance, because long-range linkage disequilibrium there routinely manufactures spurious colocalization. Second, genes that are targets of approved immune or cardiovascular drugs were treated as an internal calibration set: a pipeline that recovers known drug-target relationships blind, and in the correct pharmacologic direction, can be trusted when it points to new ones. Absence of a significant hit was not read as evidence of no effect.

Colocalization and protein-level validation

A significant Wald ratio can still arise when 2 neighboring variants, 1 driving expression and 1 driving disease, sit in linkage disequilibrium. Colocalization is the test that distinguishes these scenarios, and we used the coloc.abf framework, which distributes posterior probability across 5 hypotheses, of which H4 is a single shared causal variant for both traits.[34] Per-variant support was a Wakefield approximate Bayes factor computed in log space, with variant variance V (squared SE) and prior variance W (0.15^2 expression side, 0.20^2 disease side)[35]:

$$r = W/(W+V);\ lABF = 0.5*[log(1-r) + r*Z^2]$$

Observed disease SEs were used on the outcome side and reconstructed SEs on the exposure side, the H3 term was evaluated in log space to avoid numerical underflow, and community-standard priors ($p1=p2=1x10^{-4}$, $p12=1x10^{-5}$) were applied throughout. A posterior probability of a shared causal variant (PP.H4) >=0.8 promoted a locus from nomination to prioritized causal target. Because coloc.abf assumes

1 causal variant per trait, loci with several independent signals are underpromoted rather than overpromoted, so a low PP.H4 was never treated as evidence against causality.[36]

A cis-eQTL reports on messenger RNA, but a drug acts on protein. For every significant gene with a matched INTERVAL plasma aptamer we therefore built a second, independent instrument from the strongest cis-pQTL within +/-500 kb, repeated the Wald-ratio MR with observed pQTL effect sizes, and repeated colocalization over the shared cis variants under the same priors and threshold. A pair reached protein-level causal status only when transcript and protein estimates agreed in direction and protein-level colocalization also reached PP.H4 >=0.8. This layer covers every aptamer-matched hit across the phenome. Directional discordances are reported rather than discarded.

Two-population replication and blood immune traits

A causal effect discovered in Finland alone could reflect founder-population genetics rather than general human biology, so every disease hit in the whole-phenome scan was matched to a UK Biobank counterpart and the FinnGen-causal genes retested with identical MR and harmonization. A pair counted as 2-population validated when it kept the same direction and reached nominal significance in the independent UK Biobank samples.

Matching used 2 routes, and 2 rules protect the validity of this widened arm. The primary route matched exact endpoint code (ukb-d-<FinnGen phenocode>), which guarantees a harmonized disease definition but reaches only a minority of hit diseases. A second, still-conservative route normalized FinnGen and OpenGWAS trait names (lower-cased, parenthetical qualifiers and FinnGen-specific modifiers removed, punctuation collapsed) and matched on the normalized string, taking the dataset with the most cases when several matched. Only single-cohort UK Biobank datasets are eligible, because many OpenGWAS studies with matching trait names are meta-analyses that silently include FinnGen, and replicating a FinnGen hit in one of those would be circular; this restriction reduces the number of diseases that can be replicated but was accepted deliberately to preserve the Finland-versus-UK design. Name-matched datasets also required at least 200 cases and a European or unspecified ancestry label. The matching route for every pair is recorded so each pairing is auditable, and the widened table, not the exact-code table, feeds the replication term of the prioritization score.

Finally, to test whether these proteins affect measurable immune biology, the harmonized instruments were tested against blood immune traits under the same MR framework and FDR control, providing a bridge from a genetic causal signal to a quantitative hematologic or inflammatory phenotype.

Whole-phenome expansion, prioritization, and evidence tiers

The same instruments were rerun unmodified against all 2,469 FinnGen diseases. To turn a large result set into a usable target list, an integrated novelty-priority score was computed per gene-disease pair:

$$NoveltyPriority = s(causal) + s(coloc) + s(pleio) + s(drug) + s(cell) - p(known) - p(mhc)$$

where s(causal)=min(-log10(FDR)/5, 2) rewards causal strength, s(coloc)=PP.H4 rewards a shared causal variant, s(pleio)=(n_categories-1)x0.5 rewards activity across disease categories, s(drug) scales with HPA-derived druggability, and s(cell)=0.5 credits a defined blood-cell source. The 2 penalties are as large as the rewards: p(known)=1.5 for gene-disease axis with an approved drug and p(mhc)=2.0 for an MHC-flagged signal actively push already-drugged and genetically confounded loci down the ranking, so the ranking is steered away from rediscovering what is already reported. A bounded rescaling of this quantity, the Plasma Immune Novelty Score, additionally credits protein-level support and is used for tier assignment. The weights are prespecified heuristics rather than optimized quantities.

Every finding carries 1 of 5 evidence tiers, and the language used to describe it is bound to that tier. Tier 1 is a known-drug positive control recovered from genetics, which validates the pipeline rather than nominating randomly. Tier 2 is an MHC- or linkage-disequilibrium-flagged signal held at nomination. Tier 3 is a causal nomination from cis-MR that has not colocalized. Tier 4 is a prioritized causal target with cis-MR plus colocalization. Tier 5 clears the full bar: colocalized, protein-level concordant, specific, and druggable. The tiers 4 and 5 are new built for demonstrating the new targets.

Genetically inferred therapeutic direction was assigned from the sign of the causal estimate together with protein localization. A risk-raising protein (OR >1) implies blockade by neutralizing antibody, receptor antagonism, or small-molecule inhibition; a protective protein (OR <1) implies agonism or replacement. Secreted proteins favor recombinant-protein or antibody strategies and membrane receptors favor receptor-directed approaches, while proteins with opposite directions in different diseases were flagged as indication specific rather than assigned a single systemic strategy.

Cardiovascular validation arm and the Plasma Immune Risk Score

Cardiovascular disease was prespecified as the end-to-end test of the platform, because its druggable biology has been proven in clinical trial and 2 well-powered populations exist, so the pipeline can be asked to reconstruct established therapy without being given any pharmacologic labels. The whole-phenome scan was restricted to cardiovascular diseases and every retained effect passed instrument construction and cis-MR (FDR <0.05) and was then annotated through the remaining evidence layers (colocalization, 2-population replication where testable, and druggability with an assigned direction), with each claim

bound to the evidence tier it actually reached. All cardiovascular numbers and figures are computed live from the atlas's own public tables by a released script (src/47_heart_methodology.py), which removes the possibility of manual transposition error.

The PIRS is the predictive complement to the causal map, and it is released as training code, with no pretrained weights. It requires an NPX table, an outcomes table (event indicator and follow-up time), and optional covariates; if a required column is missing it prints the expected schema and exits rather than producing an invalid result. The feature space is restricted to the curated immunome, which improves the ratio of events to features and, more importantly, ensures every nonzero weight names a protein whose causal credential can be looked up in the atlas. The estimator is an elastic-net-penalized Cox model (L1 penalty for selection, L2 penalty to stabilize correlated proteins, mixing parameter 0.5), with backend selection among scikit-survival, lifelines, and an elastic-net logistic fallback on a fixed-horizon label.[37,38] Preprocessing (median imputation, then z score standardization) is fitted inside each training fold only, so no test-fold information leaks into scaling. Diseases with fewer than 50 incident events are skipped rather than fitted on data too sparse to be stable. Discrimination is estimated by stratified 5-fold cross-validation, and only afterward is the deployable model refitted on all participants. Cross-validated concordance measures discrimination, not calibration; absolute risk requires combining the Cox linear predictor with a baseline hazard, and any trained score should be validated internally, externally, mechanistically against the atlas, and for incremental value over age, sex, and standard predictors before clinical use.[39]

Statistical analysis and ethics

All tests were performed two-sided. Multiple testing was controlled by Benjamini-Hochberg FDR at 5% throughout the MR arms, and colocalization used a fixed PP.H4 >=0.8 for promotion. Nominal significance in an independent cohort served only as a replication criterion after discovery in FinnGen and was interpreted together with directional concordance. No finding is described more strongly than its highest supporting layer permits: MR supports a nomination, colocalization supports a prioritized transcript-level target, concordant pQTL MR with protein-level colocalization supports protein-level causality, and experimental perturbation would still be needed for mechanistic proof.

Only publicly available, de-identified, association statistics were analyzed, and no individual-level data were collected. The contributing genome-wide association studies were conducted under their own ethical approvals and participant consent, so this analysis required no additional institutional review board approval. Any group applying the PIRS trainer to an individual-level cohort must obtain its own institutional ethics approval.

Results

The HPIA converts raw associations into direction-specified therapeutic hypotheses

We constructed the human plasma immune atlas (HPIA) analytic platform, a Python-based computational proteogenomic platform that scans the entire plasma immunome against the entire disease phenome. Of the 2,923 Olink Explore assays, 1,007 met the prespecified inclusion threshold and constitute the plasma immunome (Figure S1A). The retained set is dominated by immune-cell-enriched proteins (723) but preserves every functional axis of human immunity; the next 5 largest classes are 89 cytokines and interleukins, 40 chemokine-axis proteins, 38 complement and coagulation proteins, 35 leukocyte-surface markers, and 27 tumor necrosis factor superfamily members (Figure S1B). Mapped to their putative blood-cell lineage (RNA expression enrichment), granulocyte- and myeloid-derived proteins dominate (434 and 312 proteins), followed by dendritic (201), T-cell (155), B-cell (86), and natural killer-cell (77) sources (Figure S1C). The annotation flags that define these groups are largely non-redundant, overlapping only where biology dictates, as between the cytokine and tumor necrosis factor axes (Figure S1D).

Every retained protein carries an immune class, a putative cellular source, a secretome or vesicle route, and a druggability flag, which is what later converts a statistical hit into an interpretable therapeutic hypothesis rather than a bare gene name (Figure S1E). Each claim is then bound to an explicit evidence tier, from a nomination held for MHC caution up to a protein-level causal target, so that the strength of the language stays matched to the strength of the data; the tier climb, and the roadmap from transcript-level genetic-proxy (cis-eQTL) MR through per-locus colocalization, replication in independent GWAS and ancestries, protein-level validation, functional perturbation, and versioned public release, are laid out in Figure S1F.

The architecture of the curated immunome determines where causal signal can and cannot be found. Immune-cell-enriched proteins dominate every blood lineage, most heavily the granulocyte (369) and myeloid (265) compartments (Figure S2A). Druggability is unevenly distributed: half the interferon-axis proteins are potential drug targets (4 of 8), followed by the acute-phase (2 of 5) and immunoglobulin/B-cell (5 of 13) classes, whereas the chemokine axis contributes none (0 of 40), so a chemokine hit and an interferon hit are not equally actionable even at identical statistical strength (Figure S2B). The panel measures each major immune communication axis at usable depth, including the interleukin-1 family (36 proteins), tumor necrosis factor superfamily (35), complement (30), CC and CXC chemokines (23 and 13), checkpoints (10), and the interferon and interleukin-6/gp130 axes (8 each) (Figure S2C), and a

substantial fraction of the set is already disease annotated, including approximately 120 approved drug targets (Figure S2D).

Two features distinguish this resource from an assay list. First, every axis is not merely catalogued but genetically instrumented and tested, including the antibody axis that broad proteomic studies usually omit (Figure S2E, S2F); within it, CD79B, FCRL5, and FCRL3 emerge as the leading causal signals in type 2 diabetes, Alzheimer disease, and rheumatoid arthritis, respectively (Figure S2H). Second, the resource is biogenesis aware: resolving how each protein reaches plasma, by classical secretion versus a membrane or extracellular-vesicle route, bears directly on modality choice, because a secreted protein invites a neutralizing antibody whereas a membrane receptor invites a receptor-directed agent (Figure S2G).

Figure S1

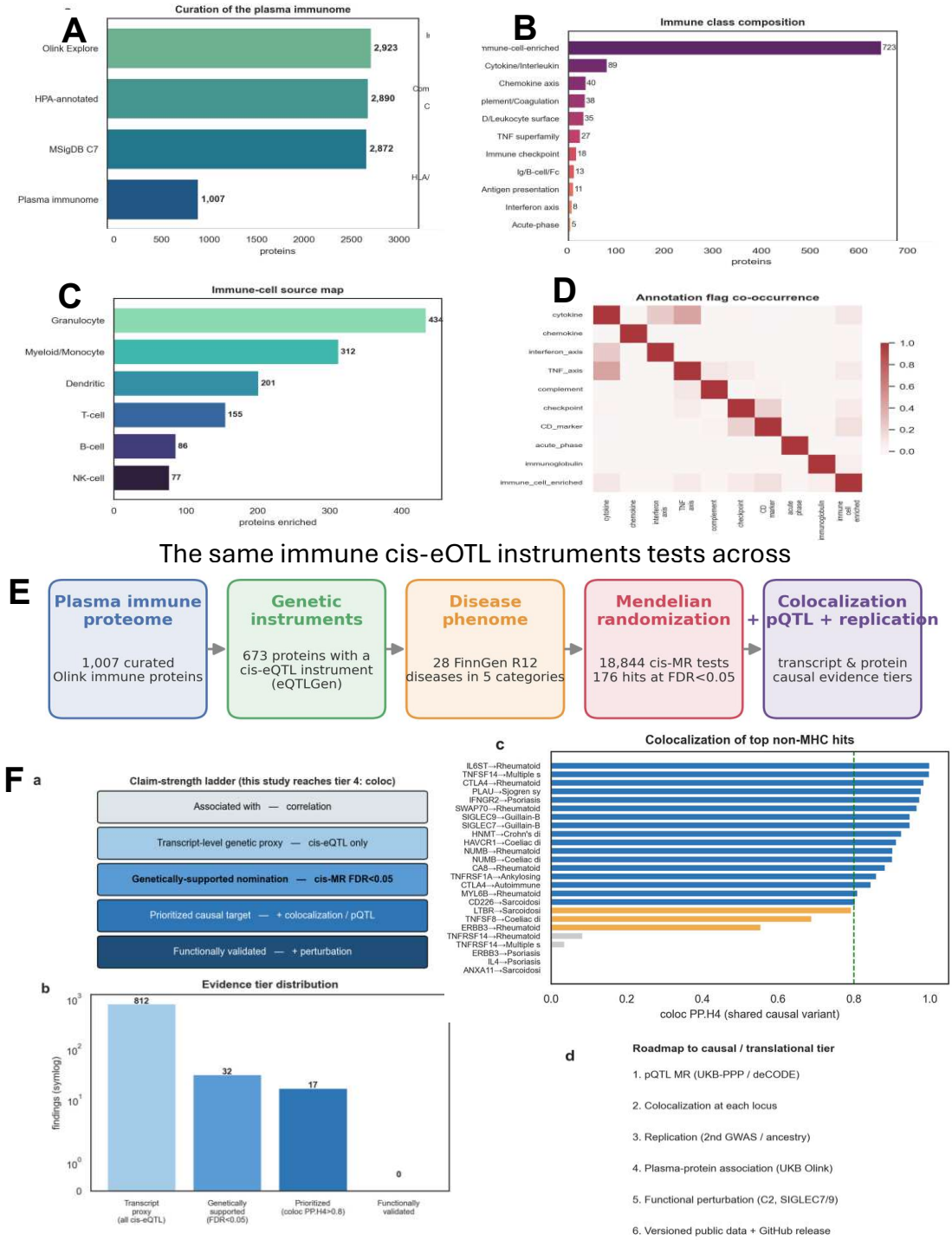


Figure S1 Curation and evidence architecture of the 1,007-protein plasma immunome. A, Curation funnel from the 2,923-assay Olink Explore universe through Human Protein Atlas (HPA) and Molecular Signatures Database annotation to the 1,007-protein plasma immunome. B, Immune class composition of the curated set, from 723 immune-cell-enriched proteins down to 5 acute-phase proteins. C, Immune-cell source map, showing proteins enriched per blood-cell lineage and dominated by granulocyte (434) and myeloid/monocyte (312) origins. D, Annotation-flag co-occurrence across the curation criteria,

showing that the flags are largely non-redundant. E, Platform overview: the same immune cis-expression quantitative trait locus instruments are carried unchanged from the curated proteome through instrument construction, the disease phenome, Mendelian randomization, and the colocalization, protein quantitative trait locus, and replication layers, with every claim bound to an evidence tier. F, Computational proteogenomic platform validation roadmap and claim-strength ladder: the evidence ladder each target climbs (a), the resulting tier distribution (b), the colocalization profile of the leading non-major-histocompatibility-complex hits (c), and the route to causal or translational tiers (d). The atlas is built exclusively from public, citable resources; no individual-level, credentialed, or gated dataset was used.

Figure S2.

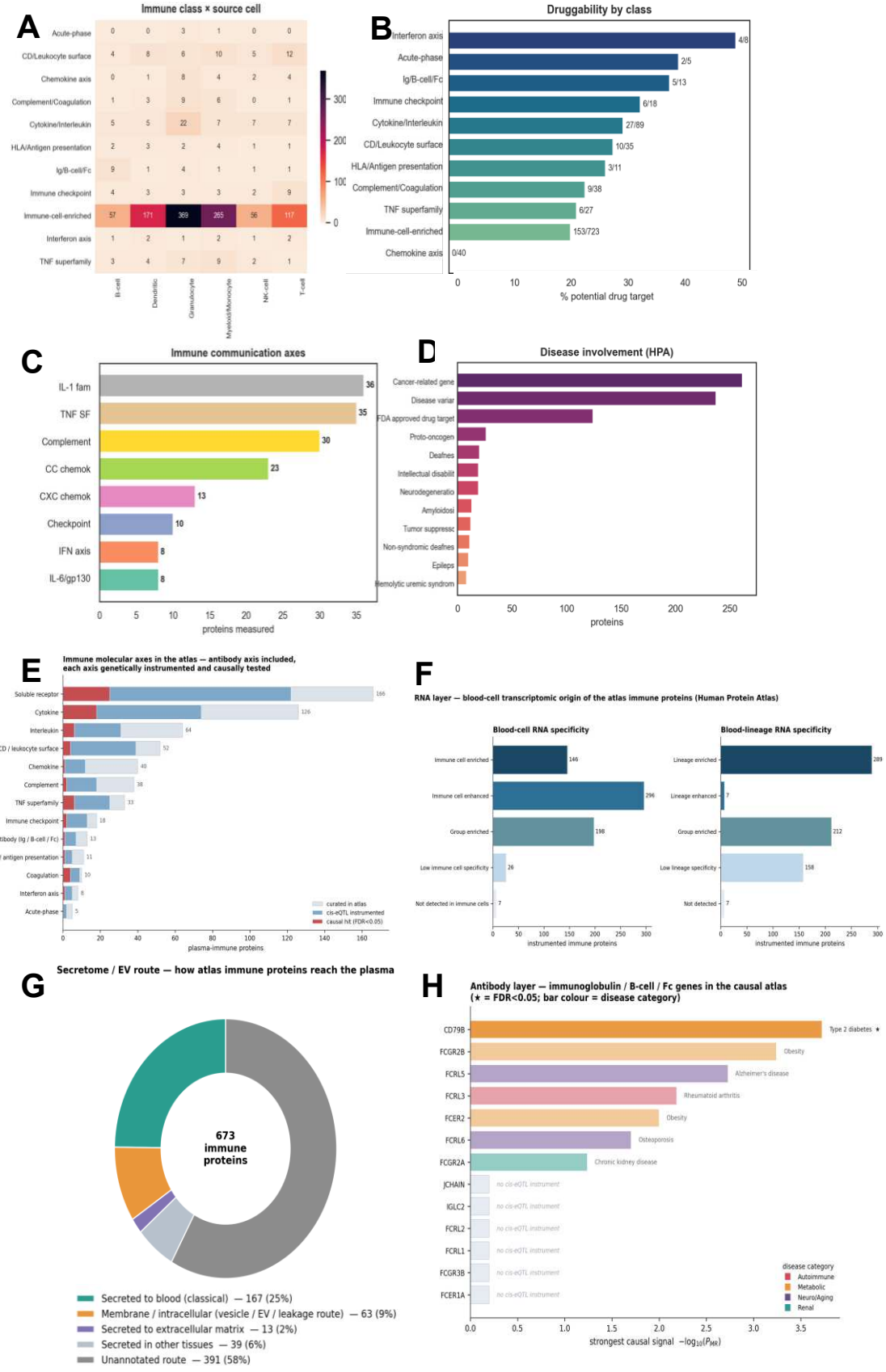

A
Immune class × source cell
Acute-phase
CD/Leukocyte surface
Chemokine axis
Complement/Coagulation
Cytokine/Interleukin
HLA/Antigen presentation
Ig/B-cell/Fc
Immune checkpoint
Immune-cell-enriched
Interferon axis
TNF superfamily
B-cell
Dendritic
Granulocyte
Myeloid/Monocyte
NK-cell
T-cell
B
Druggability by class
Interferon axis 4/8
Acute-phase 2/5
Ig/B-cell/Fc 5/13
Immune checkpoint 6/18
Cytokine/Interleukin 27/89
CD/Leukocyte surface 10/35
HLA/Antigen presentation 3/11
Complement/Coagulation 9/38
TNF superfamily 6/27
Immune-cell-enriched 153/723
Chemokine axis 0/40
% potential drug target
C
Immune communication axes
IL-1 fam 36
TNF SF 35
Complement 30
CC chemok 23
CXC chemok 13
Checkpoint 10
IFN axis 8
IL-6/gp130 8
proteins measured
D
Disease involvement (HPA)
Cancer-related gene
Disease variar
FDA approved drug target
Proto-oncogen
Deafnes
Intellectual disabilit
Neurodegeneratio
Amyloidosi
Tumor suppressc
Non-syndromic deafnes
Epileps
Hemolytic uremic syndrom
proteins
E
Immune molecular axes in the atlas — antibody axis included, each axis genetically instrumented and causally tested
Soluble receptor 166
Cytokine 126
Interleukin 64
CD / leukocyte surface 52
Chemokine 40
Complement 38
TNF superfamily 33
Immune checkpoint 18
Antibody (Ig / B-cell / Fc) 13
HLA / antigen presentation 11
Coagulation 10
Interferon axis 8
Acute-phase 5
curated in atlas
cis-eQTL instrumented
causal hit (FDR<0.05)
plasma-immune proteins
F
RNA layer — blood-cell transcriptomic origin of the atlas immune proteins (Human Protein Atlas)
Blood-cell RNA specificity
Immune cell enriched 146
Immune cell enhanced 296
Group enriched 198
Low immune cell specificity 26
Not detected in immune cells 7
Blood-lineage RNA specificity
Lineage enriched 289
Lineage enhanced 7
Group enriched 212
Low lineage specificity 158
Not detected 7
instrumented immune proteins
G
Secretome / EV route — how atlas immune proteins reach the plasma
673 immune proteins
Secreted to blood (classical) — 167 (25%)
Membrane / intracellular (vesicle / EV / leakage route) — 63 (9%)
Secreted to extracellular matrix — 13 (2%)
Secreted in other tissues — 39 (6%)
Unannotated route — 391 (58%)
H
Antibody layer — immunoglobulin / B-cell / Fc genes in the causal atlas
(★ = FDR<0.05; bar colour = disease category)
CD79B Type 2 diabetes ★
FCGR2B Obesity
FCRL5 Alzheimer's disease
FCRL3 Rheumatoid arthritis
FCER2 Obesity
FCRL6 Osteoporosis
FCGR2A Chronic kidney disease
JCHAIN no cis-eQTL instrument
IGLC2 no cis-eQTL instrument
FCRL2 no cis-eQTL instrument
FCRL1 no cis-eQTL instrument
FCGR3B no cis-eQTL instrument
FCER1A no cis-eQTL instrument
disease category
Autoimmune
Metabolic
Neuro/Aging
Renal
strongest causal signal −log10(PMR)

Figure S2. **Druggability, cellular sourcing, and molecular-layer architecture of the plasma immunome**. A, Immune class by source cell, showing that immune-cell-enriched proteins dominate every lineage and most heavily the granulocyte (369) and myeloid (265) compartments. B, Druggability by immune class, expressed as the fraction of proteins that are potential drug targets, led by the interferon axis (4/8) and acute-phase (2/5) classes, with the chemokine axis contributing none (0/40). C, Immune communication axes measured on the panel, including the interleukin-1 family (36), tumor necrosis factor superfamily (35), complement (30), CC and CXC chemokines (23 and 13), checkpoints (10), and the interferon and interleukin-6/gp130 axes (8 each). D, Disease involvement of the curated proteins from HPA annotation, including approximately 120 approved drug targets. E, Immune molecular axes: for each functional axis, including the antibody (immunoglobulin, B-cell, Fc-receptor) axis, the number of proteins curated, instrumented with a cis-expression quantitative trait locus, and reaching a causal hit at a false discovery rate <0.05. F, RNA layer: blood-cell and blood-lineage RNA specificity of the instrumented immune proteins. G, Secretome and extracellular-vesicle route, showing how each instrumented protein reaches plasma, by classical secretion versus a membrane, intracellular, or vesicle route. H, Antibody-layer spotlight: immunoglobulin, B-cell, and Fc-receptor genes, whether each carries an instrument, and its strongest causal signal, led by CD79B in type 2 diabetes, FCRL5 in Alzheimer disease, and FCRL3 in rheumatoid arthritis.

Blind recovery of five approved drug axes calibrates the discovery engine

We first asked whether the engine, given no pharmacologic input, could rediscover what clinical medicine already knows. Five gene-disease pairs corresponding to marketed immune therapies were recovered, each in the direction the drug itself exploits (Figure 1A(a-d); Table 2): IL6ST raised rheumatoid arthritis risk (OR, 2.63; PP.H4, 1.00), concordant with the interleukin-6 signaling axis that tocilizumab inhibits at the interleukin-6 receptor (IL6ST encodes gp130, the signal-transducing subunit of the same receptor complex); CTLA4 was protective in rheumatoid arthritis (OR, 0.51; PP.H4, 0.98) and autoimmune hyperthyroidism (OR, 0.15; PP.H4, 0.84), consistent with the costimulation-modulating mechanism of abatacept (CTLA4-Ig, which binds CD80/CD86 and blocks CD28 signaling); TNFRSF1A raised ankylosing spondylitis risk (OR, 1.46; PP.H4, 0.86), within the tumor necrosis factor axis that etanercept, a soluble p75 tumor necrosis factor receptor-Fc fusion, neutralizes at the ligand level; IL4 raised psoriasis risk (OR, 1.80), pathway-concordant with dupilumab, which blocks interleukin-4 receptor-alpha rather than interleukin-4 itself; and TNFSF14 (LIGHT) was protective in multiple sclerosis (OR, 0.50), recovering the LIGHT-HVEM costimulatory axis. These recoveries establish pathway-level directional concordance with approved therapies rather than exact drug-target identity: the instrumented gene and the drug's direct molecular target may be distinct nodes of the same axis. Every one of them replicated in an

independent, non-FinnGen study (replication P, $4x10^{-4}$ to $7x10^{-24}$), including TNFSF14 in multiple sclerosis at $P=2x10^{-12}$ in the International Multiple Sclerosis Genetics Consortium data (Figure 1C(a, b)). The same discipline extends to cardiovascular disease: genetically higher ACE raised both hypertension and atrial fibrillation risk, the direction that ACE inhibition reverses (Figure 1B). Across the whole phenome, the strongest signals likewise served as positive controls recovered blind, including complement factor H in age-related macular degeneration, CTLA4 in autoimmune thyroid disease, C2 (an MHC-region signal held at tier 2) in type 1 diabetes and rheumatoid arthritis, and the PSRC1/SORT1 lipid locus in statin use and lipid disorders[40]. An engine that reconstructs approved pharmacology blind, and in the correct direction, earns a measure of trust when it points to targets that have no drug yet.

Figure 1.

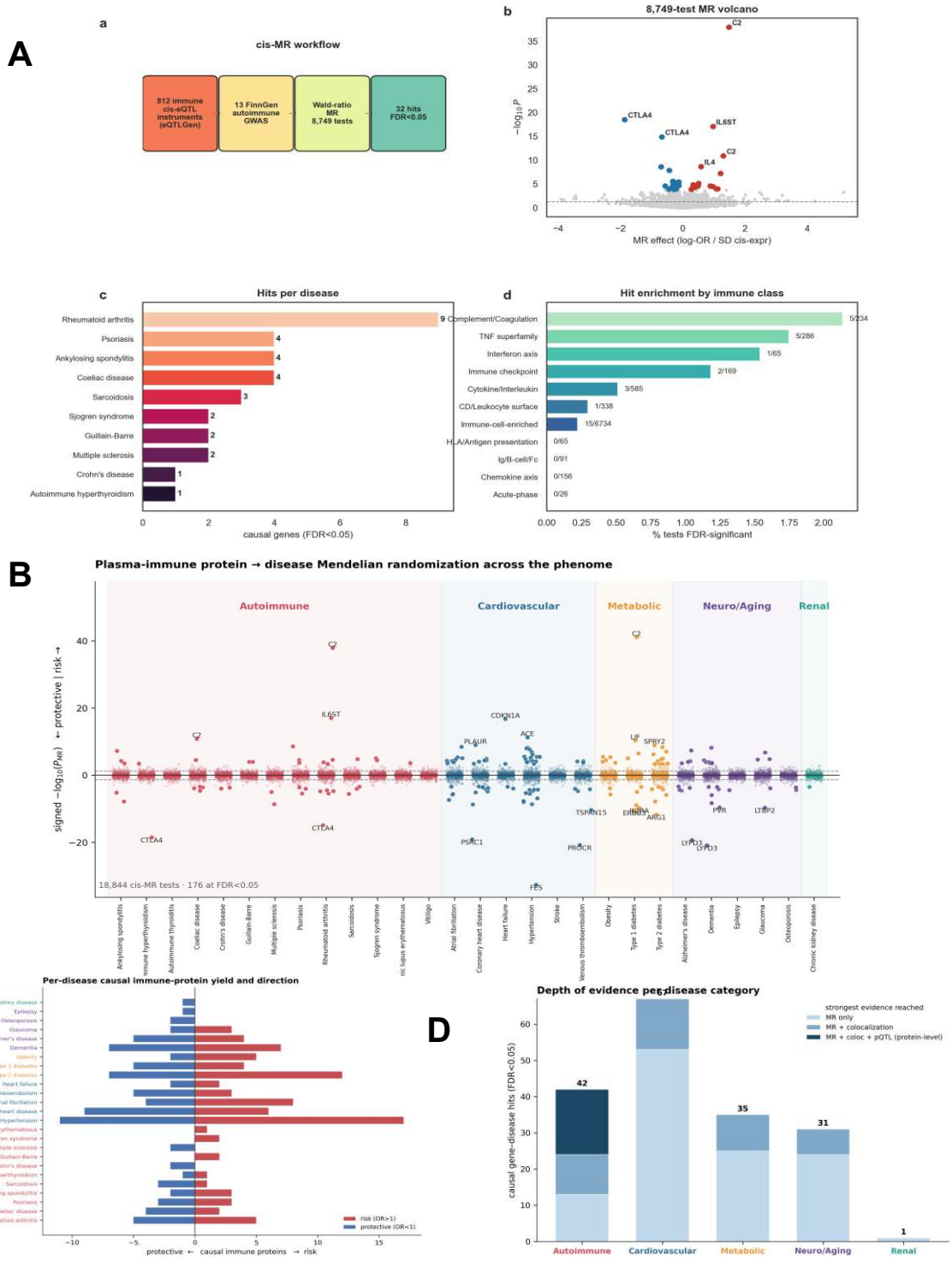


Figure 1. Blind recovery of established therapeutic axes, cardiovascular calibration, and independent replication. A, Forest plots of the internal positive controls (odds ratio [OR] per 1-SD genetically higher cis expression, 95% CI): (a) IL6ST to rheumatoid arthritis (OR, 2.63; false discovery

rate [FDR], $2.4x10^{-14}$; tocilizumab axis); (b) CTLA4 to autoimmune hyperthyroidism (OR, 0.15; FDR, $1.4x10^{-15}$; abatacept axis); (c) TNFRSF1A to ankylosing spondylitis (OR, 1.46; FDR, $1.8x10^{-2}$; etanercept/anti-tumor necrosis factor axis); (d) TNFSF14 to multiple sclerosis (OR, 0.50; FDR, $3.1x10^{-6}$; LIGHT-HVEM axis). B, Proof-of-concept exhibit: (upper) every approved-drug axis recovered with the correct pharmacologic direction, with red denoting risk-raising axes that a drug blocks and green protective axes that a drug agonizes; (lower left) the cardiovascular spotlight, in which genetically higher ACE raises both atrial fibrillation and hypertension risk, the direction ACE inhibition reverses; (lower right) orthogonal support for each recovered axis from colocalization posterior probability and independent genome-wide association replication. C, Independent replication in non-FinnGen consortium studies: (a) replication significance for each nominated target, with study source labeled and direction concordance retained unless marked; (b) replication outcome across all significant gene-disease hits. Axis labels denote pathway-level directional concordance: the instrumented gene and the drug's direct molecular target may be distinct nodes of the same axis (IL6ST [gp130] vs the IL-6 receptor for tocilizumab; TNFRSF1A vs ligand-level TNF neutralization for etanercept; IL4 vs IL-4 receptor-alpha for dupilumab; CTLA4 vs CD80/CD86 binding for abatacept).

cis-MR discovery: from an autoimmune anchor to a five-category disease core

Of 812 immune genes with an eQTLGen instrument, 673 yielded harmonized, analyzable cis-instruments. Tested first against 13 FinnGen autoimmune diseases, the arm in which immune causality is least controversial, these produced 8,749 Wald-ratio tests and 32 FDR-significant gene-disease pairs, led by rheumatoid arthritis with 9 causal genes (Figure S3A(a-c)). The complement/coagulation and tumor necrosis factor superfamily classes were the most enriched for causal signal, at 2.1% and 1.7% of tests, respectively, whereas the chemokine, acute-phase, and immunoglobulin classes returned none (Figure S3A(d)). This anchor arm establishes that the instrument set detects immune causality where it is expected. Extending the same instruments unchanged to the 28-disease core produced 18,844 tests and a result that reframes the resource: causal signal is distributed across all 5 disease categories rather than concentrated in autoimmunity (Figure S3B). Per disease, hypertension, coronary heart disease, and type 2 diabetes each recruited a causal immune set comparable to or larger than rheumatoid arthritis, and each carried both risk-raising and protective arms rather than a single directional block (Figure S3C). The depth of orthogonal support differed systematically between categories, with the autoimmune chapter carrying the highest proportion of protein-level evidence and the cardiovascular chapter the largest absolute number of colocalized hits (Figure S3D).

Figure S3

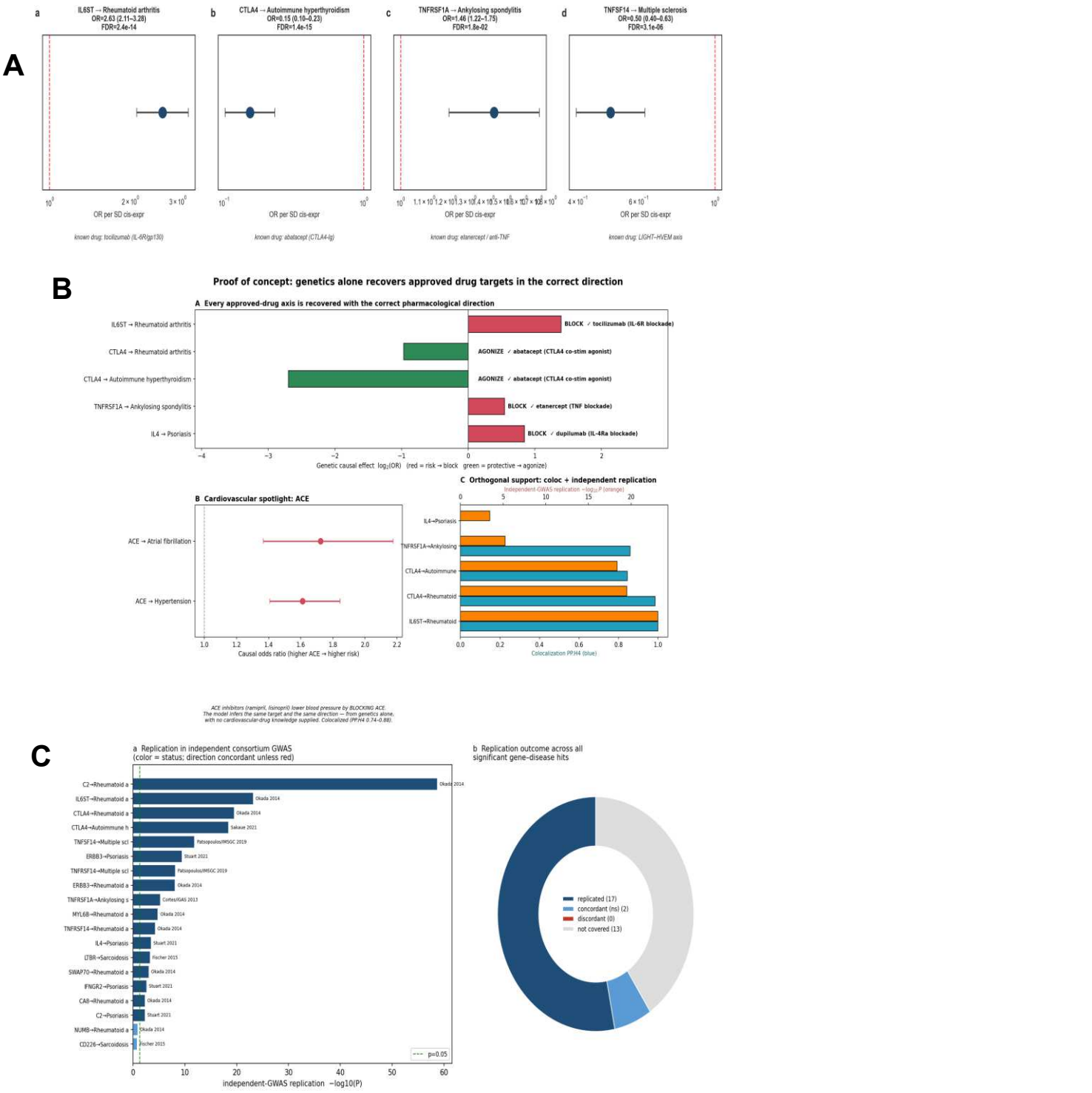


Figure S3. **Instrument construction, discovery-arm Mendelian randomization, and per-disease causal yield.** A, Discovery arm: (a) cis-Mendelian randomization workflow, with 812 immune cis-expression quantitative trait locus instruments constructed from eQTLGen (673 usable) and tested against 13 FinnGen autoimmune genome-wide association studies in 8,749 Wald-ratio tests yielding 32 hits at a false discovery rate (FDR) <0.05; (b) volcano plot of the 8,749-test discovery arm, with C2, IL6ST, CTLA4, and IL4 labeled; (c) causal genes per disease, led by rheumatoid arthritis (9); (d) hit enrichment

by immune class, led by the complement/coagulation and tumor necrosis factor superfamily classes. B, Mirrored Manhattan plot of the 18,844 cis-Mendelian randomization tests of the 28-disease, 5-category core; points above the center line are risk increasing and points below are protective, with the dashed lines marking the FDR <0.05 band and the tallest signals spanning the drug-target axes together with cardiovascular and metabolic nominations. C, Number of causal immune proteins per disease split by direction (risk raising versus protective), with disease labels colored by category; hypertension, coronary heart disease, and type 2 diabetes carry the largest immune causal burden outside the classical autoimmune diseases. D, Evidence-source stack by disease category, decomposing each category's causal hits into Mendelian randomization only, Mendelian randomization plus colocalization, and Mendelian randomization plus colocalization plus protein quantitative trait locus support.

Cardiovascular disease carries the largest causal immune burden in the 28-disease core (67 of 176 pairs)

The central finding of the atlas is that the causal reach of the plasma immunome is broadest outside the diseases where immunology has traditionally looked (Figure 2A). The 28-disease core yielded 176 FDR-significant causal gene-disease pairs across 25 of the 28 endpoints, and cardiovascular disease carried the largest burden (67 pairs), ahead of autoimmune (42), metabolic (35), neurodegenerative/aging (31), and renal disease (1) (Figure 2B). Composition differed as sharply as count: the autoimmune chapter uniquely recruited the complement/coagulation and tumor necrosis factor superfamily classes, whereas the cardiovascular chapter drew overwhelmingly on immune-cell-enriched proteins.

Entirely new, non-autoimmune causal relationships emerged: ACE and IFNGR2 in hypertension, PSRC1 and PLAUR in coronary heart disease, PROCR in venous thromboembolism, ERBB3 in type 1 diabetes, and ACE colocalizing with both dementia and Alzheimer disease at posterior probability >0.97 (Figures 2A and 3A). Viewed disease by disease, hypertension (28 hits), type 2 diabetes (19), and coronary heart disease (15) each carried a larger causal immune set than rheumatoid arthritis (10) (Figure 2C). Restricting the whole-phenome scan to cardiovascular endpoints resolved a coherent directional map in which FES, PSRC1, and SCARB1 occupy the protective arm and ACE, CDKN1A, and MAPT the risk arm across 32 cardiovascular diseases (Figure 2D). Immune-targeted cardiovascular therapy therefore has a genetically supported target space far larger than the interleukin-1 and interleukin-6 axes explored in trials to date.

Figure 2.

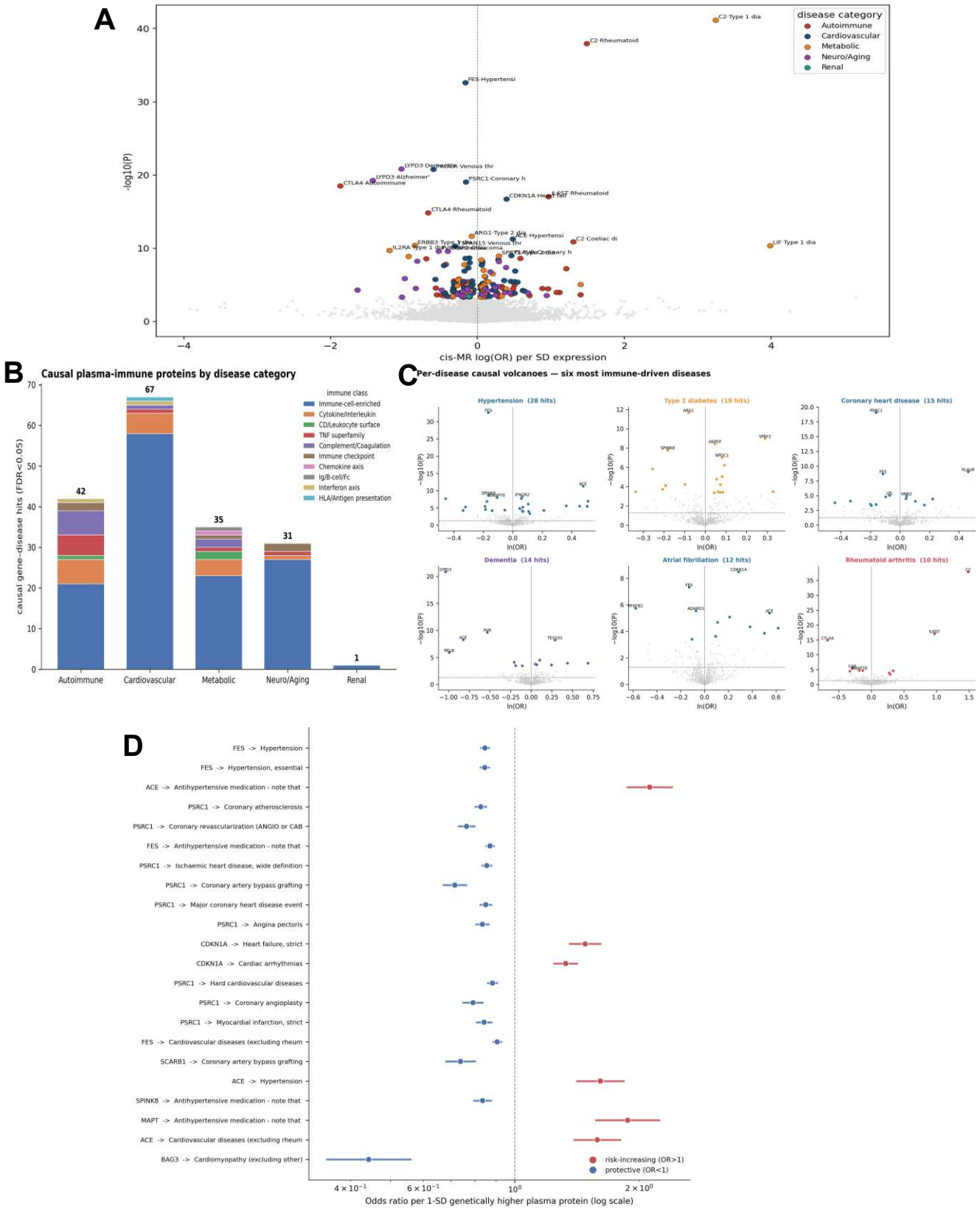
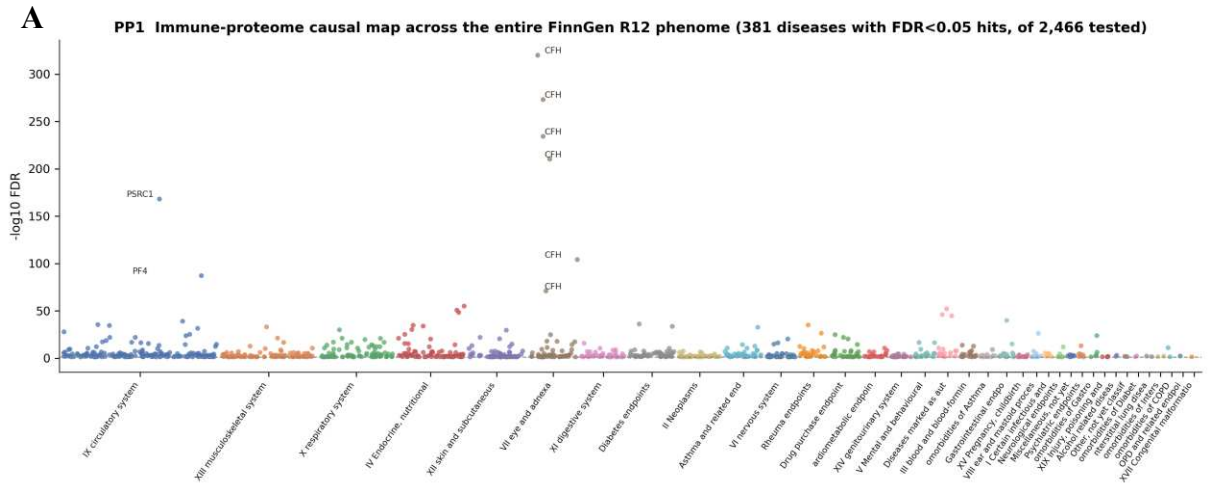
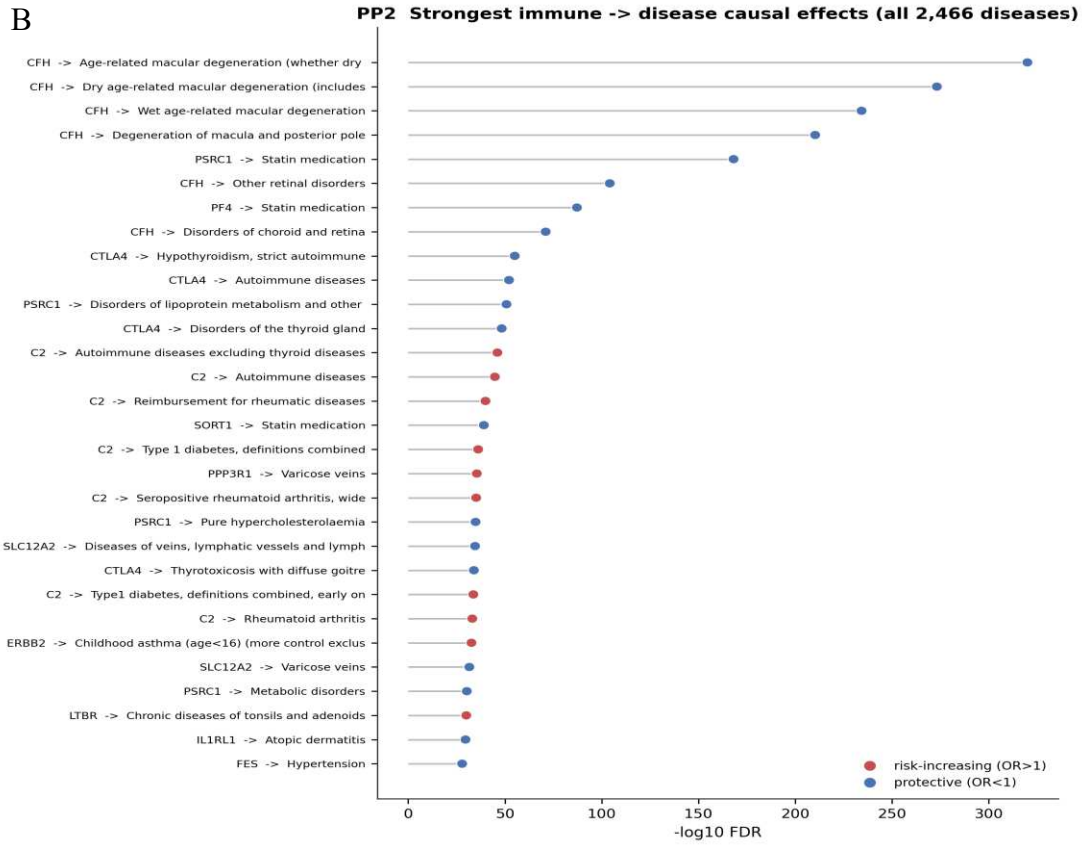

A
disease category
Autoimmune
Cardiovascular
Metabolic
Neuro/Aging
Renal
-log10(P)
cis-MR log(OR) per SD expression
B
Causal plasma-immune proteins by disease category
causal gene–disease hits (FDR<0.05)
immune class
Immune-cell-enriched
Cytokine/Interleukin
CD/Leukocyte surface
TNF superfamily
Complement/Coagulation
Immune checkpoint
Chemokine axis
Ig/B-cell/Fc
Interferon axis
HLA/Antigen presentation
67
42
35
31
1
Autoimmune
Cardiovascular
Metabolic
Neuro/Aging
Renal
C
Per-disease causal volcanoes — six most immune-driven diseases
Hypertension (28 hits)
Type 2 diabetes (19 hits)
Coronary heart disease (15 hits)
Dementia (14 hits)
Atrial fibrillation (12 hits)
Rheumatoid arthritis (10 hits)
ln(OR)
D
FES -> Hypertension
FES -> Hypertension, essential
ACE -> Antihypertensive medication - note that
PSRC1 -> Coronary atherosclerosis
PSRC1 -> Coronary revascularization (ANGIO or CAB
FES -> Antihypertensive medication - note that
PSRC1 -> Ischaemic heart disease, wide definition
PSRC1 -> Coronary artery bypass grafting
PSRC1 -> Major coronary heart disease event
PSRC1 -> Angina pectoris
CDKN1A -> Heart failure, strict
CDKN1A -> Cardiac arrhythmias
PSRC1 -> Hard cardiovascular diseases
PSRC1 -> Coronary angioplasty
PSRC1 -> Myocardial infarction, strict
FES -> Cardiovascular diseases (excluding rheum
SCARB1 -> Coronary artery bypass grafting
ACE -> Hypertension
SPINK8 -> Antihypertensive medication - note that
MAPT -> Antihypertensive medication - note that
ACE -> Cardiovascular diseases (excluding rheum
BAG3 -> Cardiomyopathy (excluding other)
risk-increasing (OR>1)
protective (OR<1)
Odds ratio per 1-SD genetically higher plasma protein (log scale)

Figure 2. Five-category causal map: cardiovascular disease carries the largest immune causal burden in the phenome. A, Causal map of every immune protein tested against the 28 curated diseases in 5 categories (autoimmune, cardiovascular, metabolic, renal, and neurodegenerative/aging); each point is a gene-disease test (x axis, cis-Mendelian randomization log odds ratio per SD expression; y axis, significance). The scan yields 176 false discovery rate (FDR)-significant causal gene-disease pairs across 25 of the 28 endpoints. B, Causal hits per disease category stacked by immune protein class: cardiovascular disease draws the most causal immune proteins (67), followed by autoimmune (42), metabolic (35), neurodegenerative/aging (31), and renal (1), while autoimmune disease uniquely recruits the complement/coagulation and tumor necrosis factor superfamily classes. C, Per-disease volcano grid for the 6 diseases with the largest numbers of significant immune associations, showing Mendelian randomization effect size against significance with FDR <0.05 hits highlighted. D, Cardiovascular causal map from the whole-phenome scan restricted to cardiovascular endpoints (46 immune proteins across 32 cardiovascular diseases): each row is a plasma immune protein to cardiovascular endpoint effect, plotted as the odds ratio per 1-SD genetically higher expression of the encoding gene (transcript-proxied) with 95% CI on a log axis; red denotes risk increasing (an inhibition target) and blue protective (an agonism target). Panels A through C show the 5-category core; the full 2,469-disease scan is shown in Figure S4.

Figure S4



B


PP2 Strongest immune -> disease causal effects (all 2,466 diseases)
CFH -> Age-related macular degeneration (whether dry
CFH -> Dry age-related macular degeneration (includes
CFH -> Wet age-related macular degeneration
CFH -> Degeneration of macula and posterior pole
PSRC1 -> Statin medication
CFH -> Other retinal disorders
PF4 -> Statin medication
CFH -> Disorders of choroid and retina
CTLA4 -> Hypothyroidism, strict autoimmune
CTLA4 -> Autoimmune diseases
PSRC1 -> Disorders of lipoprotein metabolism and other
CTLA4 -> Disorders of the thyroid gland
C2 -> Autoimmune diseases excluding thyroid diseases
C2 -> Autoimmune diseases
C2 -> Reimbursement for rheumatic diseases
SORT1 -> Statin medication
C2 -> Type 1 diabetes, definitions combined
PPP3R1 -> Varicose veins
C2 -> Seropositive rheumatoid arthritis, wide
PSRC1 -> Pure hypercholesterolaemia
SLC12A2 -> Diseases of veins, lymphatic vessels and lymph
CTLA4 -> Thyrotoxicosis with diffuse goitre
C2 -> Type1 diabetes, definitions combined, early on
C2 -> Rheumatoid arthritis
ERBB2 -> Childhood asthma (age<16) (more control exclus
SLC12A2 -> Varicose veins
PSRC1 -> Metabolic disorders
LTBR -> Chronic diseases of tonsils and adenoids
IL1RL1 -> Atopic dermatitis
FES -> Hypertension
risk-increasing (OR>1)
protective (OR<1)
0
50
100
150
200
250
300
-log10 FDR

C

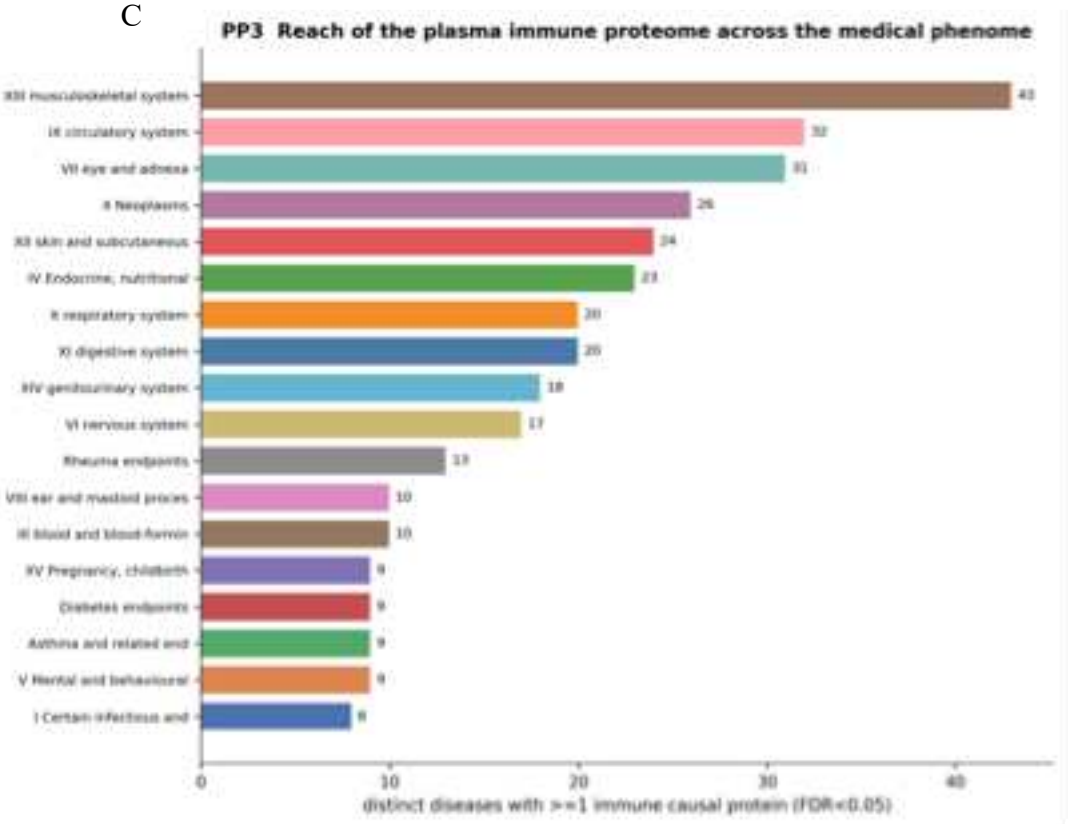

PP3 Reach of the plasma immune proteome across the medical phenome
XIII musculoskeletal system
IX circulatory system
VII eye and adnexa
II Neoplasms
XII skin and subcutaneous
IV Endocrine, nutritional
X respiratory system
XI digestive system
XIV genitourinary system
VI nervous system
Rheuma endpoints
VIII ear and mastoid proces
III blood and blood-formin
XV Pregnancy, childbirth
Diabetes endpoints
Asthma and related end
V Mental and behavioural
I Certain infectious and
43
32
31
26
24
23
20
20
18
17
13
10
10
9
9
9
9
8
0
10
20
30
40
distinct diseases with >=1 immune causal protein (FDR<0.05)

Figure S4. Phenome-wide causal map across all 2,469 FinnGen R12 endpoints. A, Every false discovery rate (FDR) <0.05 immune-to-disease causal effect across all 2,466 FinnGen R12 endpoints with data, grouped by International Classification of Diseases chapter (x axis) against -log10 FDR (y axis); the immune proteome reaches into essentially every organ system, with CFH to macular degeneration and PSRC1/PF4 to lipid-lowering phenotypes the extreme peaks. B, The 30 strongest immune-to-disease causal effects over the whole phenome, colored by direction (red, risk increasing; blue, protective), dominated by known complement, checkpoint, and lipid biology recovered blind. C, Reach of the plasma immune proteome across the medical phenome: number of distinct diseases with at least 1 causal immune protein (FDR <0.05) per International Classification of Diseases chapter; the musculoskeletal, circulatory, ophthalmic, neoplasm, and dermatologic chapters are the most immune connected.

Cross-category proteins couple cardiovascular disease to metabolic and neurodegenerative disease

Because every protein is tested against every disease on a common scale, the HPIA can see immune proteins that couple otherwise unconnected organ systems, an axis invisible to single-disease designs (Figure 3A). Of the 120 distinct causal immune proteins, 22 acted in 2 or more disease categories and 1 acted in 3 (Figure 3C). C2 was the most pleiotropic single signal, significant in 7 diseases (although, as an MHC-region gene, C2 is held at tier 2 under the prespecified MHC rule), followed by PM20D1 (6) and ACE (4) (Figure 3B); IFNGR2, SWAP70, MERTK, MPO, SPINK8, and PIK3IP1 each spanned at least 2 disease categories.

The sharing structure is itself informative: the cardiovascular chapter shares more causal proteins with the metabolic and neurodegenerative chapters than with the autoimmune chapter (Figure 3D), which argues that cardiovascular immune involvement is a distinct causal program rather than a diluted form of autoimmunity. Hypertension illustrates the density of that program: 28 immune proteins, drawn from granulocyte, myeloid, and lymphoid sources, act on it through both risk-raising and protective arms (Figure 3E). Because these immune programs overlap across diseases, a therapy aimed at one immune pathway should not be expected to respect disease boundaries.

Figure 3.

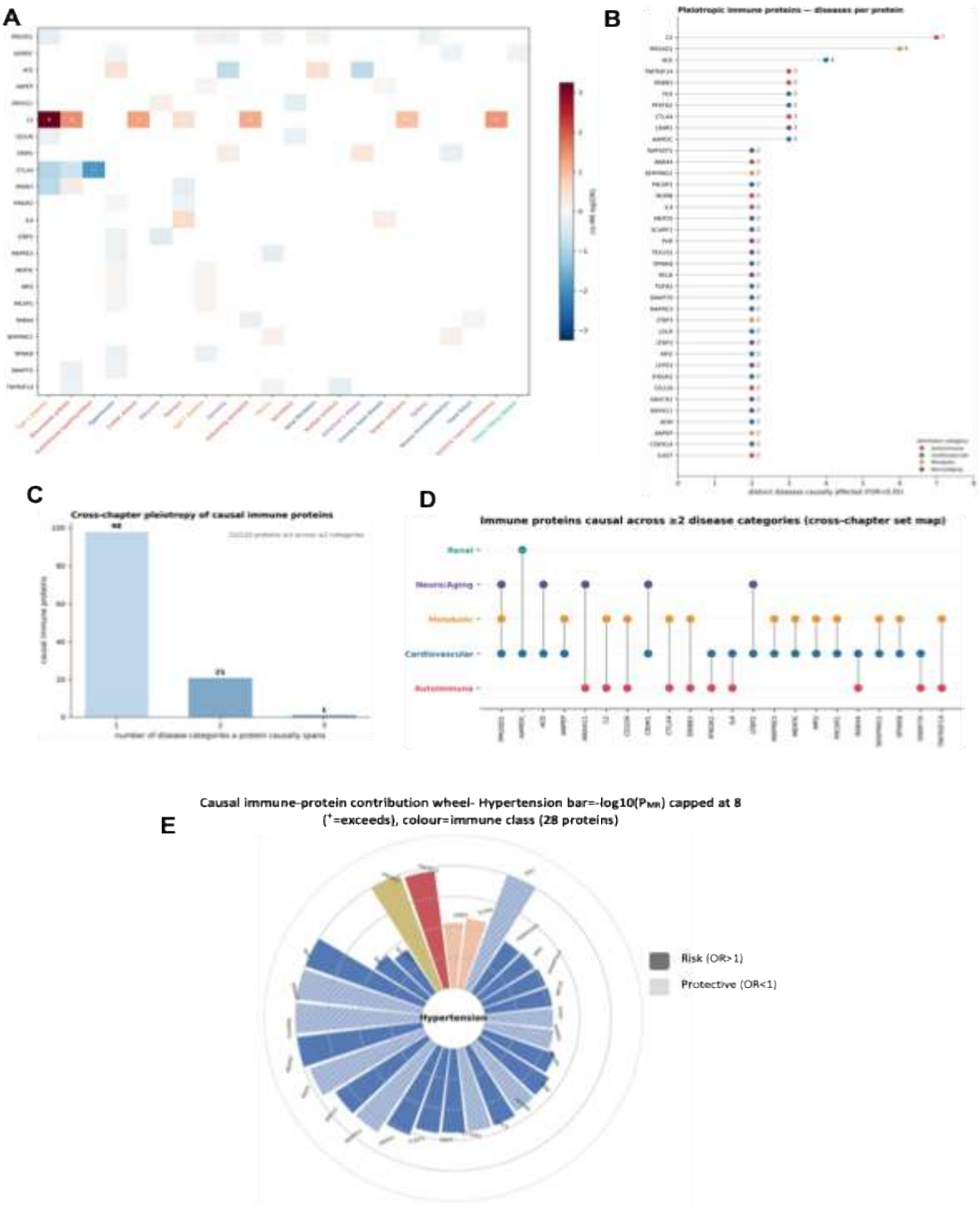


Figure 3. Shared causal immune axes and cross-category pleiotropy. A, Genes causal in >=2 disease categories (cell color, cis-Mendelian randomization log odds ratio), including ACE (risk raising in hypertension and atrial fibrillation, protective in dementia and Alzheimer disease), IFNGR2, MERTK,

MPO, SPINK8, PIK3IP1, SWAP70, and PM20D1. B, Diseases per immune protein: the number of distinct diseases each causal immune protein drives, led by C2 (7; MHC-region signal held at tier 2), PM20D1 (6), and ACE (4), with dot color indicating the protein's dominant disease category. C, Cross-chapter pleiotropy: distribution of how many disease categories each causal immune protein spans; 98 proteins act within 1 category, 21 across 2, and 1 across 3, so that 22 of 120 act across >=2 categories. D, Cross-category set map marking the immune proteins causal across 2 or more disease categories, showing that the cardiovascular chapter shares more causal proteins with the metabolic and neurodegenerative chapters than with the autoimmune chapter. E, Causal-contribution wheel for hypertension: each spoke is 1 of the 28 immune proteins with a significant causal effect, with bar length the capped Mendelian randomization significance, solid bars risk raising and hatched bars protective, and color the immune class.

Granulocyte- and myeloid-derived proteins dominate the cardiovascular causal signal

To trace causal signal back to its cellular context, we mapped each causal protein to its putative blood-cell source (RNA expression enrichment) and uncovered a consistent phenome-wide gradient (Figure 4A). Granulocyte-derived proteins carried the strongest enrichment for causal signal of any blood lineage, 64 of 434 instrumented proteins (Fisher exact OR, 1.39), followed by myeloid/monocyte (OR, 1.00), T-cell (0.94), dendritic (0.79), B-cell (0.67), and natural killer-cell (0.55) sources (Figure 4B). The granulocyte enrichment itself does not reach nominal significance, so the lineage gradient is reported as a hypothesis-generating trend rather than a demonstrated enrichment; what is descriptively clear is that granulocyte and myeloid compartments dominate the cardiovascular and metabolic chapters specifically (Figure 4A).

A causal claim gains credibility when the same protein also shifts an immune phenotype that a clinician can measure. Testing 9,969 protein-trait pairs against 15 blood immune traits, 1,268 passed the 5% FDR threshold, implicating 395 distinct immune proteins as genetically supported regulators of circulating immune-cell populations or inflammation; monocyte fraction (116 proteins), leukocyte count (108), and platelet count (104) emerged as the most immune-connected traits (Figure 4D). Because the directional split is nearly balanced across traits, the plasma immunome is not simply acting as a proxy for total leukocyte burden (Figure 4C). This layer of evidence bridges a genetic causal claim and routine hematology.

Figure 4.

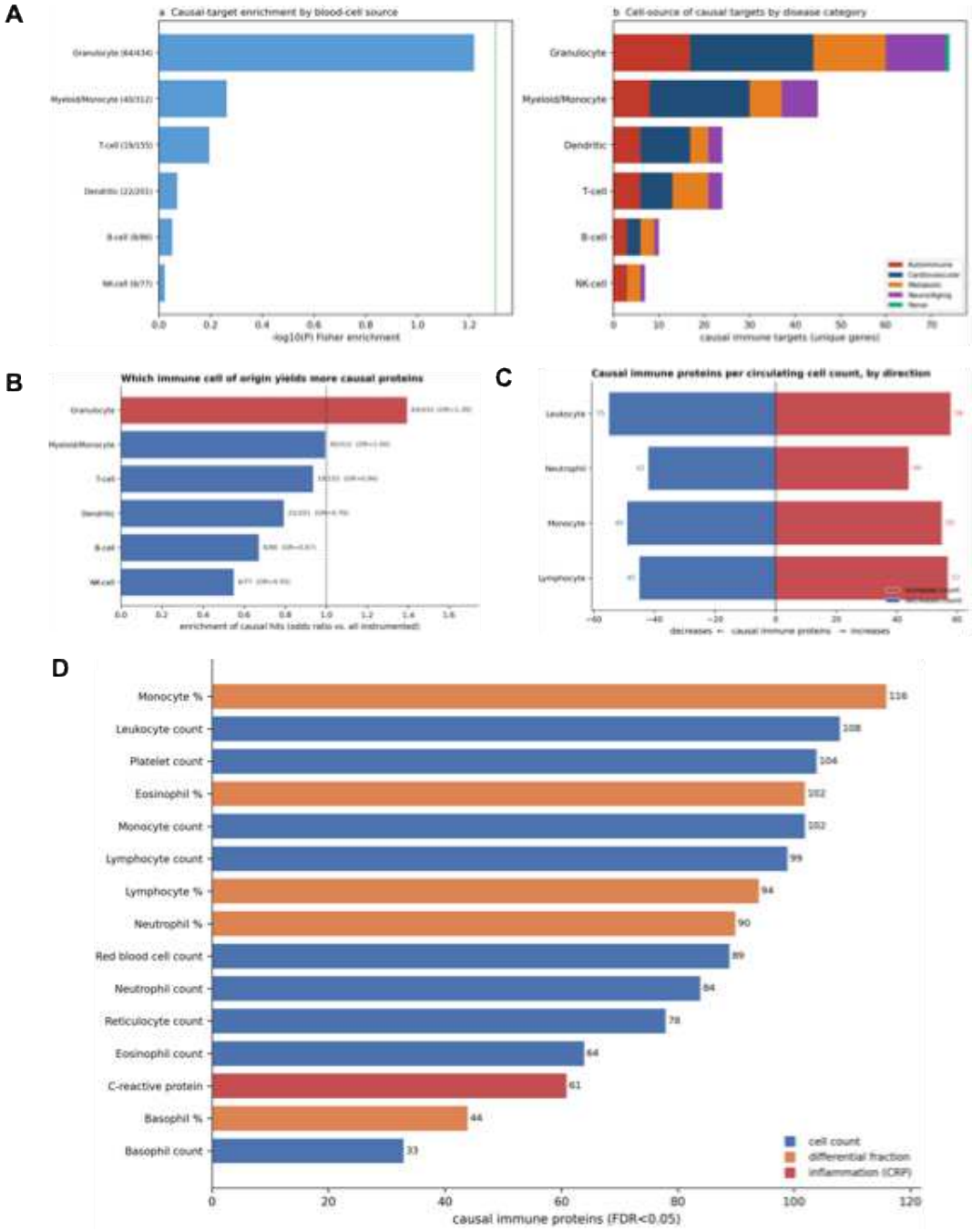


Figure 4. Cellular source (blood-cell expression enrichment) of causal immune proteins and their effects on measured blood immune traits. A, Putative cell-source map: (a) causal-target enrichment by blood-cell source, in which granulocyte-origin proteins show the strongest enrichment for causal hits (Fisher exact odds ratio, 1.39; 64 of 434 instrumented) though below the dashed P<0.05 line, whereas natural killer- and B-cell-origin proteins rank lowest; (b) cell source of causal targets by disease category, with the granulocyte and myeloid/monocyte compartments dominating the cardiovascular and metabolic chapters. B, Immune cell-of-origin enrichment expressed as the odds ratio that a protein expressed by each

immune cell type yields a causal hit relative to all instrumented proteins (granulocyte 1.39, myeloid/monocyte 1.00, T cell 0.94, dendritic 0.79, B cell 0.67, natural killer cell 0.55). C, Directional split of the cellular arm: number of causal immune proteins that increase versus decrease each circulating leukocyte, neutrophil, monocyte, and lymphocyte count, showing a near-balanced increase/decrease split so that the plasma immunome is not simply a proxy for overall leukocyte burden. D, Immune proteome to blood immune-trait Mendelian randomization: number of causal immune proteins (false discovery rate <0.05) per blood trait, colored by class (cell count, differential fraction, inflammation); of 9,969 tests, 1268 were significant, implicating 395 distinct immune proteins, with monocyte fraction, leukocyte count, and platelet count the most immune-connected traits. Eosinophil and basophil counts are filled from the independent Astle et al blood-cell genome-wide association study so that no lineage is omitted.

Whole-phenome expansion: 1,016 genetically supported associations across 381 diseases

Combining causal strength, colocalization, cross-category pleiotropy, cell source, and druggability into a single auditable score, while subtracting penalties for approved-drug and MHC axes, we run the same instruments against every FinnGen disease removes any concern that the disease list was curated to flatter the result (Figure S4). The scan comprised 1,656,872 Wald-ratio tests, of which 1,016 passed a 5% FDR across 381 diseases and 220 immune genes, and the immune proteome reached into essentially every organ system, with the musculoskeletal, circulatory, ophthalmic, neoplasm, and dermatologic chapters the most immune connected. This whole-phenome table is a strict superset of the 28-disease core, which remains the most deeply annotated layer.

Three independent validation layers were then applied to those hits (Figure S5). Colocalization across the 900 pairs with at least 5 shared variants confirmed 417 at PP.H4 >=0.8 and 510 at >=0.5, so roughly 46% of tested hits reflect a shared causal variant rather than a linkage artifact. Cross-population testing spans 189 comparisons across 62 diseases and 90 immune proteins. Stated with both denominators: 189 of the 1,016 discovery pairs had an eligible independent counterpart and were testable; 162 of 189 (86%) agreed in causal direction in independent samples, and 89 of 189 (47%; 9% of the whole discovery atlas) additionally reached nominal significance, up from the 58 reachable by exact-code matching alone (Table 5; Table S1). The exact-code subset reproduces identically inside the widened table (133 pairs, 48 diseases, 58 validated), so widening added pairs without perturbing existing ones, and the newly reachable diseases are predominantly non-autoimmune, including varicose veins, ankylosing spondylitis, hernia, and spondylopathy, precisely where the narrow route was blind.

Phenome-wide protein-level validation promotes the first tier 5 cardiovascular targets

A cis-eQTL reports on messenger RNA, but a drug acts on protein. To close that gap, we retested every significant gene with a matched plasma aptamer, this time using the protein itself as the instrument, and repeated colocalization at the protein level (Figure 5A). In the original autoimmune-restricted design, 2 pairs, TNFSF14 in multiple sclerosis and SWAP70 in rheumatoid arthritis, colocalized at both the transcript and the protein level, with posterior probabilities above 0.97, and 9 of 18 testable pairs agreed in direction (Figure 5B).

Extending this protein layer across the whole phenome (110 proteins, 219 diseases) and rebuilding tier assignment over all 1,016 genetically supported pairs returned 90 tier 1, 115 tier 2, 447 tier 3, 359 tier 4, and 5 tier 5 pairs (Table 3). The 5 tier 5 targets, which clear the full evidence bar of colocalization, protein-level concordance, specificity, and druggability, are IL2RA in type 1 diabetes (the expected positive control at this tier), ANXA2 in hypertension under both endpoint definitions, and PPP3R1 in varicose veins and venous/lymphatic disease; a further 35 pairs are protein-confirmed, combining transcript-level colocalization with direction-concordant protein-level MR (Table 3; Table S2). ANXA2 in hypertension is the first protein-level causal cardiovascular target in this resource.

Figure 5.

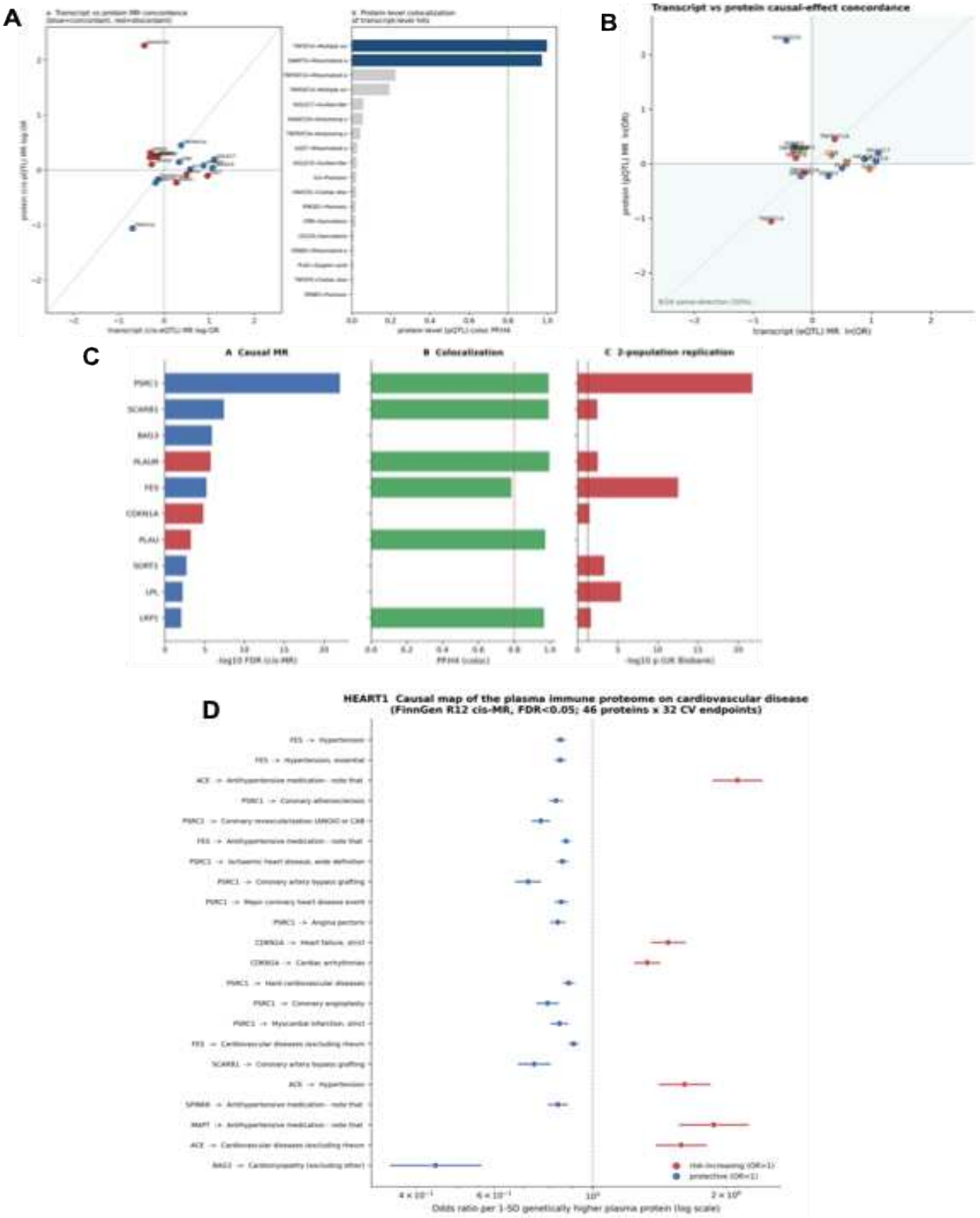

A
B
Transcript vs protein causal-effect concordance
C
A Causal MR
B Colocalisation
C 2-population replication
PSRC1
SCARB1
BAG3
PLAUR
FES
CDKN1A
PLAU
SORT1
LPL
LRP1
D
HEART1 Causal map of the plasma immune proteome on cardiovascular disease
(FinnGen R12 cis-MR, FDR<0.05; 46 proteins x 32 CV endpoints)
Odds ratio per 1-SD genetically higher plasma protein (log scale)

Figure 5. Protein-level validation, cardiovascular triangulation, and the final evidence ladder. A, Protein-level (INTERVAL plasma cis-protein quantitative trait locus [pQTL]) validation: (a) transcript- versus protein-level Mendelian randomization (MR) concordance (blue, concordant; red, discordant; 9 of 18 concordant in the original autoimmune-restricted design); (b) protein-level colocalization of transcript-level hits, in which TNFSF14 to multiple sclerosis and SWAP70 to rheumatoid arthritis exceed a posterior probability of a shared causal variant (PP.H4) >=0.8 at both transcript and protein level. B, Transcript versus protein causal-effect concordance on a common axis, with shaded quadrants marking directional agreement; discordances are reported rather than hidden because they can reflect biologically meaningful differences between transcript and circulating protein abundance. C, Triangulation of coronary and ischemic heart disease targets across 3 independent evidence layers: causal strength (-log10 false discovery rate, cis-MR; left), colocalization PP.H4 with the dashed 0.8 promotion threshold (middle), and UK Biobank replication (-log10 P; right). PSRC1, SCARB1, BAG3, PLAUR, FES, PLAU, SORT1, LPL, and LRP1 clear the layers they are testable in; PSRC1 is the standout, colocalizing at PP.H4 >0.99 and replicating in England at $P<10^{-21}$. D, Supervised-model performance panel of the atlas intelligence layer: cross-validated discrimination of the trained framework, instantiated for cardiovascular risk (concordance index, 0.777 in end-to-end synthetic validation).

Hypertension carries the largest number of significant immune associations from Cardiovascular causal assay

When performing the whole-phenome scan to cardiovascular disease yielded 156 FDR-significant gene-disease effects across 32 cardiovascular diseases and 46 immune genes; each entered the cardiovascular analysis arm and is annotated with the evidence tier it actually reached, and only the subsets described below attain colocalized (tier 4) or protein-level (tier 5) status (Table 4; Figure S6). Within the annotated core, 5 of the 6 cardiovascular endpoints carried FDR-significant causal proteins, and those 5 account for all 67 cardiovascular pairs.

Hypertension carries the largest set of significant immune signals in the atlas (28 proteins): ACE (OR, 1.61; FDR, $7x10^{-9}$), the interferon-gamma receptor subunit IFNGR2 (OR, 1.07; colocalized), the efferocytosis receptor MERTK (colocalized), ANXA2, and JMJD1C raise risk, whereas the granulocyte kinase FES (colocalized), SWAP70, MAPRE3, and SPINK8 protect. Coronary heart disease (15 proteins) is led on the protective arm by the chromosome-1p13 protein PSRC1 (OR, 0.85; colocalized), the high-density-lipoprotein receptor SCARB1 (colocalized), LPL, and FES; the risk arm is led by the urokinase receptor PLAUR (OR, 1.59; colocalized), an inflammatory-thrombosis target that is genuinely immune rather than lipid-related.

Atrial fibrillation (12 proteins) couples the renin-angiotensin axis (ACE, OR, 1.72; colocalized) to the cell-cycle inhibitor CDKN1A on the risk arm, with FES, the glycolytic regulator PFKFB2 (both

colocalized), and ADAM15 protective. Venous thromboembolism (8 proteins) is anchored by the strongly protective endothelial protein C receptor PROCR (OR, 0.55), with PDLIM7 and SCPEP1 as colocalized novel nominations, and heart failure (4 proteins) is driven by the senescence effector CDKN1A/p21 (OR, 1.48), with PFKFB2 and RAB44 protective. Beyond the core, the wider cardiovascular scan adds three observations: BAG3, an established dilated-cardiomyopathy gene,[41] is strongly protective in heart failure; ERBB2 joins CDKN1A as an atrial fibrillation risk node; and the protective coronary axis (PSRC1, FES) extends to stroke, consistent with a shared atherothrombotic mechanism reaching cerebrovascular disease. Three findings cut across these endpoints. First, CDKN1A/p21 recurs as a risk protein in both heart failure and atrial fibrillation, genetically implicating cellular senescence in cardiac disease and linking the cardiovascular chapter to the aging chapter of the atlas. Second, FES acts as a protective granulocyte kinase across coronary disease, hypertension, atrial fibrillation, and stroke, a single innate-immune node spanning the vascular tree, at a locus already implicated in blood-pressure regulation.[42,43] Third, medication-use phenotypes act as built-in pharmacogenomic mirrors: genetically higher plasma ACE raises antihypertensive-medication use (OR $\approx$ 2.12), with the protein and medication signals colocalizing almost perfectly (PP.H4 $\approx$ 1.00). Approved ACE inhibitors do exactly the reverse, lower ACE to lower blood pressure, so genetic direction and drug mechanism agree: the atlas reconstructs the rationale for a 1980s drug class from public data alone.

Triangulating the coronary targets across all three evidence layers, PSRC1 stands out as the strongest causal coronary protein, colocalizing at posterior probability >0.99 and replicating in England at $P<10^{-21}$, while PLAUR colocalizes at >0.998. SORT1, LPL, SCARB1, and LRP1 form a coherent lipoprotein-handling module that is colocalized, replicated, and independently validated as causal in coronary disease[40,44,45] (Figure 5C). The entire coronary cluster reproduces in the English cohort, ruling out Finnish founder artifacts. The corollary is what makes the atlas actionable: a novel protein at the same evidence grade on the same map, but without an approved drug, is a target hypothesis of the same quality as ACE, with its genetically inferred therapeutic direction already suggested by the sign of the causal OR.

The PIRS: a released framework for scoring the causally characterized immunome

A causal map built from summary statistics cannot tell a clinician what an individual patient's risk is. We build this PIRS to close that gap. It draws from the real human cohort datasets, and ships as training code with no pretrained weights (Figure S7). In end-to-end synthetic validation, the PIRS achieved a cross-validated concordance index of 0.777 and correctly recovered the planted risk proteins, confirming that

the estimator, the leakage-free cross-validation, and the weight extraction all behave as intended (Figure 5D).

The central advantage of this score over a conventional proteomic risk model is that it is interpretable by construction rather than after the fact. Because the feature space is restricted to the curated immunome, every nonzero weight names a protein whose causal credential, direction, cellular source, and evidence tier can be looked up directly in this atlas, so risk prediction and biological mechanism are delivered by a single model rather than requiring a separate mechanistic study to interpret the score. Proteins that survive the full cardiovascular evidence chain become the direction-signed features of a cardiovascular PIRS, which, once trained on an authorized cohort, delivers four outputs at once: risk stratification, a ranked list of cardiovascular targets, an explicit genetically inferred inhibit-or-agonize hypothesis for each target, and biomarkers that are mechanistic rather than merely correlated. Where a protein both predicts and appears to cause a disease, the atlas flags that convergence as its strongest form of internal corroboration, a distinction a correlation-only risk model cannot make.

Figure S7

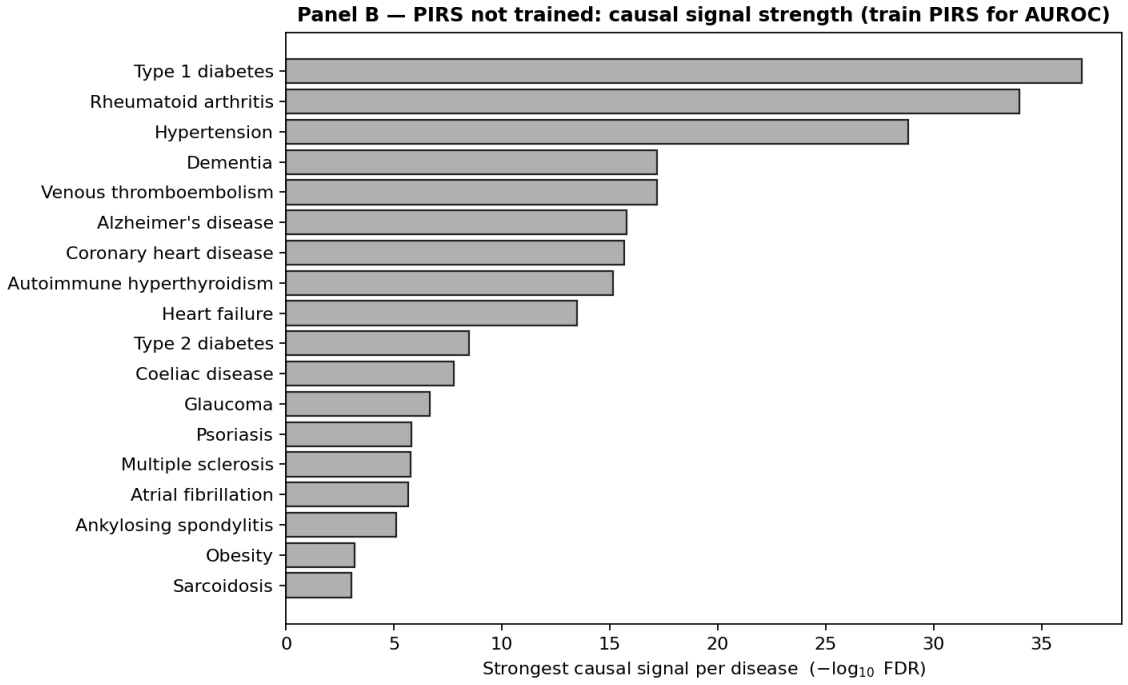


Figure S7. Design of the cardiovascular Plasma Immune Risk Score (PIRS). A, Design: 2-population-validated causal coronary proteins enter as interpretable, direction-signed features supporting risk stratification, target nomination, drug-direction assignment, and mechanistic biomarker use.

Discussion

This study presents a genetics-anchored atlas connecting circulating immune proteins to the human disease phenome, all 2,469 diseases in FinnGen R12, with a deeply annotated 28-disease core, triangulated through MR, statistical colocalization, an independent protein-level genetic instrument now extended across the phenome, and 2-population replication, together with the PIRS, a released machine-learning risk-model framework designed to convert an individual's measured plasma immune proteome into disease-specific risk in clinical cohorts, using only the causally characterized immunome. The pipeline blindly recovered established immune and cardiovascular drug axes in the correct pharmacologic direction as internal positive controls and nominated additional new drug targets with an explicit, direction-specified therapeutic proposition.

Four findings stand out. First, the causal architecture of the plasma immunome is constructed in the platform for translational use, and cardiovascular disease emerges as a major site of genetically supported immune associations. Within the deeply annotated core, cardiovascular disease carried 67 of 176 causal pairs, even more than autoimmunity (42), and hypertension yielded the largest number of significant immune associations in the atlas (28 proteins); because discovery depends on GWAS sample size, case number, phenotype definition, statistical power, and instrument availability, this is a power-adjusted estimate of biological immune causality. This is new because essentially all prior plasma-immunoproteomic work has been conducted inside the autoimmune disease family, where the immune hypothesis is presupposed; testing the same immune panel against every endpoint of a national registry is what allows the comparison to be made at all. It is significant because it reframes the immune proteome as a cardiovascular substrate rather than an autoimmune one, and it identifies blood-pressure regulation, not atherosclerosis, as the endpoint with the largest number of genetically supported immune associations in this resource.

Second, extending plasma pQTL validation from the autoimmune arc to the whole phenome converts a procedural boundary into a biological one and promotes the first protein-level causal cardiovascular targets in this resource. The protein layer now spans 110 proteins and 219 diseases, and rebuilding tier assignment over all 1,016 genetically supported pairs returns 5 tier 5 targets, of which 3 are cardiovascular-domain entries: ANXA2 in hypertension under both endpoint definitions and PPP3R1 in venous/lymphatic disease. This is new because the previous release reported no cardiovascular pair at protein level and correctly attributed that to missing instruments rather than missing signal; the present analysis removes the instrument gap and shows the signal was there. It is significant because protein-level

concordance is the evidence layer closest to the pharmacologic point of intervention: ANXA2 is a plasminogen and tissue plasminogen activator co-receptor central to endothelial fibrinolytic surface activity,[46,47] so a genetically supported, protein-level causal effect on blood pressure connects fibrinolytic endothelial biology to hypertension in a way transcript evidence alone could not support. PPP3R1 encodes the calcineurin B regulatory subunit, and calcineurin-nuclear factor of activated T cells signaling is an established mediator of cardiac hypertrophy and of venous and lymphatic vessel remodeling,[48,49] so a protein-level causal effect in venous and lymphatic disease is mechanistically coherent rather than merely statistical. IL2RA in type 1 diabetes, the fifth tier 5 target, is the expected positive control at this tier and calibrates it.[50]

Third, widening the 2-population design from exact endpoint codes to conservatively normalized trait names, while restricting eligibility to single-cohort UK Biobank datasets, raises Finland-versus-England comparisons from 133 to 189 pairs and 2-population validated pairs from 58 to 89, with 86% direction concordance (162 of the 189 testable pairs; the testable set covers 189 of the 1,016 discovery pairs). This is methodologically new rather than biologically new, and it matters for a reason specific to this literature: many OpenGWAS disease studies whose trait names match FinnGen endpoints are meta-analyses that silently include FinnGen, so an unguarded widening would have manufactured circular replication. Excluding them costs reachable diseases (81 to 62) and is the reason the widened arm can be read as genuine out-of-sample evidence. The diseases newly brought into replication are predominantly non-autoimmune, so the gain accrues where the exact-code route was blind.

Fourth, the atlas resolves genetically inferred therapeutic direction, not merely target identity, and does so across systems. ACE is risk raising in hypertension and atrial fibrillation but protective in dementia and Alzheimer disease at a posterior probability >0.97. This directional split is a finding in its own right, one that single-disease designs cannot see. It carries a cautionary implication for lifelong systemic ACE inhibition[51] (a concern that pharmacoepidemiologic and MR analyses of antihypertensive repurposing have approached from the treatment side). IFNGR2 shows the same structure across hypertension and psoriasis with opposite signs, and PM20D1 acts across 3 disease categories at once.

For cardiovascular medicine specifically, the atlas delivers 3 findings that association-based proteomics cannot. First, an endpoint-resolved causal architecture: hypertension yields the largest number of significant immune associations in the atlas (28 proteins), coronary disease resolves into a protective lipoprotein-handling module (PSRC1/SORT1, LPL, SCARB1, LRP1) opposed by an inflammatory-thrombosis risk node (PLAUR), atrial fibrillation couples the renin-angiotensin axis to senescence (ACE,

CDKN1A), venous thromboembolism is anchored by protective PROCR, whose Ser219Gly variant is an established determinant of thrombotic risk,[52] and heart failure is driven by CDKN1A/p21 with the cardiomyopathy cochaperone BAG3 protective.

Second, mechanism-level novelty. The recurrence of CDKN1A across heart failure and atrial fibrillation genetically implicates cellular senescence, inflammaging, in cardiac disease, a hypothesis that dovetails with experimental clearance of senescent cells altering cardiovascular progression.[53-55] The cross-endpoint protective behavior of the granulocyte kinase FES nominates a single innate-immune node spanning the vascular tree. PLAUR is the clearest example of a target the atlas both nominates and explains: its soluble form is an established prognostic marker of cardiovascular and renal risk,[56] and the atlas adds what an observational marker cannot, a direction-signed causal estimate that specifies blockade rather than measurement. MERTK likewise arrives with a mechanism attached, as the receptor through which macrophages clear apoptotic cells, a process whose failure sustains vascular inflammation.[57]

Third, the advantages of built-in calibration. The ACE/antihypertensive-medication axis (OR approximately 2.12; PP.H4 approximately 1.00) shows the platform rederiving a 1980s drug class from genetics alone, so a novel target at the same evidence grade, PLAUR being the clearest example, is a hypothesis of the same quality, with its modality already specified.

Phenome-wide MR of the plasma proteome has previously been performed with general protein panels against a few hundred curated phenotypes,[14,15] and dedicated cardiovascular protein panels have been evaluated as drug targets in tens of thousands of individuals.[16] More recent work has scaled plasma proteogenomics substantially, but within access-controlled cohorts.[17,18,58] What is new here is the combination of an immune-focused, functionally annotated protein set, an outcome space consisting of an entire national disease registry rather than a shortlist, an open-data-only construction, and a claim-gating framework in which cellular source, genetically inferred therapeutic direction, and druggability are computed for every hit. The practical value of that combination is not merely descriptive: target-indication pairs supported by both MR and colocalization are enriched for eventual regulatory approval,[14,59] which is precisely the evidence grade the 45 prioritized nominations reported here have reached.

Limitations

Four limitations should be weighed when interpreting these findings, and we state them in the same order as the design gaps set out in the Background. First, limitation of population and tissue. The causal effects reported here are drawn from blood, largely of European ancestry, using Finnish discovery genome-wide association data; although the widened 2-population arm now tests 189 pairs across 62 diseases for

reproducibility in UK Biobank, that resource is also predominantly European, and generalization to other ancestries remains an important direction for future work. Second, protein-level validation rests on a plasma proteomic panel. Extending it across the phenome has removed the procedural gap that previously kept cardiovascular pairs out of the highest tier, but 110 causal proteins, largely intracellular or MHC-region, have no plasma aptamer and cannot reach a protein tier from any login-free resource. This is now a biological boundary of the plasma-measurable feature space rather than a coverage artifact, and it is the reason the tissue-level extension is the natural next release. Third, the single-strongest-cis-instrument design cannot, by construction, run the multi-instrument pleiotropy diagnostics available to polygenic MR; colocalization, protein-level triangulation, and 2-population replication were used instead, each targeting a different failure mode, but horizontal pleiotropy acting through a neighboring gene cannot be excluded with complete certainty for any single locus. Because coloc.abf assumes a single causal variant per window, loci with allelic heterogeneity may be conservatively underpromoted.[36] Reanalysis of the principal tier 4 and tier 5 cardiovascular loci with a multiple-causal-variant colocalization approach (coloc-SuSiE) is a planned sensitivity analysis for this resource. Fourth, no nomination in this atlas has yet been tested by experimental perturbation, so every finding remains a genetically triangulated hypothesis rather than a proof of mechanism. The PIRS is shipped as a trainer-based score, and clinical performance will depend on the size, quality, follow-up duration, and event rate of whatever cohort a user supplies and should be established, cohort by cohort, using the internal, external, and mechanistic-concordance validation protocol described in Methods before any clinical use is considered. Extending discovery and replication to non-European ancestries and applying the PIRS trainer to real, authorized cohorts with adequate event counts would establish both the generalizability of the causal findings and the real-world discriminative and incremental value of the predictive score.

Conclusions

This accessible genetics-anchored plasma immunome-phenome atlas, generated on the HPIA platform together with its companion PIRS, provides a reproducible foundation for immune target discovery, drug repurposing, and individualized risk assessment.

List of abbreviations

ANXA2 annexin A2

CI confidence interval

cis-eQTL cis-expression quantitative trait locus

cis-pQTL cis-protein quantitative trait locus

FDR false discovery rate

GWAS genome-wide association study

HPA Human Protein Atlas

HPIA human plasma immune atlas (the analytic platform)

MHC major histocompatibility complex

MR Mendelian randomization

MSigDB Molecular Signatures Database

NPX Normalized Protein eXpression

OR odds ratio

PINS Plasma Immune Novelty Score

PIRS Plasma Immune Risk Score

PP.H4 posterior probability of a shared causal variant

SE standard error

Declarations

Ethics approval and consent to participate

This study analyzed only publicly available, de-identified, genetic association statistics; no individual-level data were accessed, and no human participants or animals were involved. Because the present analysis involved no new data collection and no identifiable human information, it did not constitute human-subjects research and required no additional institutional review board approval or participant consent at the University of Alabama at Birmingham. The study was conducted in accordance with the principles of the Declaration of Helsinki, and the data-access and use terms of every contributing resource were observed.

Consent for publication

Not applicable. This manuscript does not contain data from any individual person.

Availability of data and materials

All summary-level data underlying this atlas are publicly available from their original sources: Olink Explore assay identities and the assay-to-gene mapping; the Human Protein Atlas; the MSigDB C7 and

C8 collections; eQTLGen cis-eQTL summary statistics; INTERVAL plasma cis-pQTL summary statistics via the NHGRI-EBI GWAS Catalog; FinnGen R12 disease genome-wide association summary statistics via the FinnGen results portal; UK Biobank (Neale) and additional consortium studies via the OpenGWAS platform; and the Astle et al blood-cell genome-wide association study. Precomputed atlas result tables, including the whole-phenome cis-MR results, phenome-wide colocalization posteriors, widened 2-population concordance, the extended blood immune-trait MR, the novelty-priority ranking, and the final evidence-tier table, are provided as Supplemental Material. All code for proteome curation, instrument construction, MR, colocalization, protein-level validation, replication, the whole-phenome scan, the blood immune-trait arm, the cardiovascular validation (src/47_heart_methodology.py), the widened replication arm (src/52), the novelty-priority engine, and the PIRS trainer is released under an open-source (MIT) license at https://github.com/jimmyuab/human-plasma-immune-atlas; result tables and figures are released under a CC-BY-4.0 license. An interactive application is available at https://huggingface.co/spaces/jianlizhao/human-plasma-immune-atlas. Individual-level UK Biobank proteomic and phenotypic data, the UK Biobank Pharma Proteomics Project pQTL resource, and the deCODE pQTL resource are controlled access and were not used in this study.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by awards from the National Institutes of Health (R01HL157495, R01HL158612, R01HL173090, Wang Y), the American Heart Association (20TPA1293530, Wang Y; 25PRE1377191, Zhang J; 26CDA1446803, Zhao J.) and the Engineering School of UAB Innovative Fellowship Award (Tan H.).

Authors' contributions

J.L.Z. and Y.W. conceived and designed the study. S.T. and J.Z. contributed equally to this work. J.L.Z. developed the analytical framework for the plasma immunome, curated the pQTL and GWAS data sources, and performed the Mendelian randomization, colocalization, and two-population replication analyses; S. T. and J.Z. implemented the whole-phenome expansion pipeline, the evidence-tier prioritization scheme, and the plasma immune risk score (PIRS) training framework, and carried out the statistical analyses. H.T.

and Y.L. performed the protein-level validation and the blood immune trait analyses. S.T. and D.A. assisted with data curation, instrument selection and quality control, and generation of figures and tables. M.Y., S.V., and M.X. provided cardiovascular domain expertise, advised on the cardiovascular validation arm, and critically interpreted the disease-specific findings. S. T., J.L.Z. and J.Z. drafted the manuscript. Y.W. supervised the project, secured funding, and revised the manuscript critically for important intellectual content. All authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

We thank the participants and investigators of every study whose summary-level data made this atlas possible, including the FinnGen study and its participating biobanks, the UK Biobank. We also acknowledge the OpenGWAS/MRC Integrative Epidemiology Unit platform, the NHGRI-EBI GWAS Catalog, the Human Protein Atlas, and the Molecular Signatures Database for curating and openly distributing the resources used here.

References

ADDIN EN.REFLIST 1 Mestas, J. & Hughes, C. C. Of mice and not men: differences between mouse and human immunology. *J Immunol* **172**, 2731-2738 (2004). https://doi.org/10.4049/jimmunol.172.5.2731

2 Seok, J. *et al.* Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* **110**, 3507-3512 (2013). https://doi.org/10.1073/pnas.1222878110

3 Libby, P. The changing landscape of atherosclerosis. *Nature* **592**, 524-533 (2021). https://doi.org/10.1038/s41586-021-03392-8

4 Nidorf, S. M. *et al.* Colchicine in Patients with Chronic Coronary Disease. *N Engl J Med* **383**, 1838-1847 (2020). https://doi.org/10.1056/NEJMoa2021372

5 Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med* **377**, 1119-1131 (2017). https://doi.org/10.1056/NEJMoa1707914

6 Tardif, J. C. *et al.* Efficacy and Safety of Low-Dose Colchicine after Myocardial Infarction. *N Engl J Med* **381**, 2497-2505 (2019). https://doi.org/10.1056/NEJMoa1912388

7 Swirski, F. K. & Nahrendorf, M. Cardioimmunology: the immune system in cardiac homeostasis and disease. *Nat Rev Immunol* **18**, 733-744 (2018). https://doi.org/10.1038/s41577-018-0065-8

8 Ridker, P. M. *et al.* Low-Dose Methotrexate for the Prevention of Atherosclerotic Events. *N Engl J Med* **380**, 752-762 (2019). https://doi.org/10.1056/NEJMoa1809798

9 Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-860 (2015). https://doi.org/10.1038/ng.3314

10 Ridker, P. M. & Rane, M. Interleukin-6 Signaling and Anti-Interleukin-6 Therapeutics in Cardiovascular Disease. *Circ Res* **128**, 1728-1746 (2021). https://doi.org/10.1161/CIRCRESAHA.121.319077

11 Ridker, P. M. *et al.* IL-6 inhibition with ziltivekimab in patients at high atherosclerotic risk (RESCUE): a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet* **397**, 2060-2069 (2021). https://doi.org/10.1016/S0140-6736(21)00520-1

12 Interleukin-6 Receptor Mendelian Randomisation Analysis Consortium. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* **379**, 1214-1224 (2012). https://doi.org/10.1016/S0140-6736(12)60110-X

13 Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018). https://doi.org/10.1038/s41586-018-0175-2

14 Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* **52**, 1122-1131 (2020). https://doi.org/10.1038/s41588-020-0682-6

15 Schmidt, A. F. *et al.* Genetic drug target validation using Mendelian randomisation. *Nat Commun* **11**, 3255 (2020). https://doi.org/10.1038/s41467-020-16969-0

16 Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab* **2**, 1135-1148 (2020). https://doi.org/10.1038/s42255-020-00287-2

17 Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329-338 (2023). https://doi.org/10.1038/s41586-023-06592-6

18 Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* **53**, 1712-1721 (2021). https://doi.org/10.1038/s41588-021-00978-w

19 Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* **326**, 1614-1621 (2021). https://doi.org/10.1001/jama.2021.18236

20 Skrivankova, V. W. *et al.* Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, n2233 (2021). https://doi.org/10.1136/bmj.n2233

21 Wik, L. *et al.* Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Mol Cell Proteomics* **20**, 100168 (2021). https://doi.org/10.1016/j.mcpro.2021.100168

22 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). https://doi.org/10.1126/science.1260419

23 Uhlen, M. *et al.* A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **366** (2019). https://doi.org/10.1126/science.aax9198

24 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015). https://doi.org/10.1016/j.cels.2015.12.004

25 Vosa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300-1310 (2021). https://doi.org/10.1038/s41588-021-00913-z

26 Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508-518 (2023). https://doi.org/10.1038/s41586-022-05473-8

27 Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7** (2018). https://doi.org/10.7554/eLife.34408

28 Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e1419 (2016). https://doi.org/10.1016/j.cell.2016.10.042

29 Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1-22 (2003). https://doi.org/10.1093/ije/dyg070
30 Emdin, C. A., Khera, A. V. & Kathiresan, S. Mendelian Randomization. *JAMA* **318**, 1925-1926 (2017). https://doi.org/10.1001/jama.2017.17219
31 Burgess, S. *et al.* Using genetic association data to guide drug discovery and development: Review of methods and applications. *Am J Hum Genet* **110**, 195-214 (2023). https://doi.org/10.1016/j.ajhg.2022.12.017
32 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-487 (2016). https://doi.org/10.1038/ng.3538
33 Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat Med* **9**, 811-818 (1990). https://doi.org/10.1002/sim.4780090710
34 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014). https://doi.org/10.1371/journal.pgen.1004383
35 Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* **33**, 79-86 (2009). https://doi.org/10.1002/gepi.20359
36 Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet* **17**, e1009440 (2021). https://doi.org/10.1371/journal.pgen.1009440
37 Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw* **39**, 1-13 (2011). https://doi.org/10.18637/jss.v039.i05
38 Harrell, F. E., Jr., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* **15**, 361-387 (1996). https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
39 Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **350**, g7594 (2015). https://doi.org/10.1136/bmj.g7594
40 Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714-719 (2010). https://doi.org/10.1038/nature09266
41 Norton, N. *et al.* Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am J Hum Genet* **88**, 273-282 (2011). https://doi.org/10.1016/j.ajhg.2011.01.016
42 Parmar, P. G. *et al.* International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents. *Circ Cardiovasc Genet* **9**, 266-278 (2016). https://doi.org/10.1161/CIRCGENETICS.115.001190
43 Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425 (2018). https://doi.org/10.1038/s41588-018-0205-x
44 Zanoni, P. *et al.* Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science* **351**, 1166-1171 (2016). https://doi.org/10.1126/science.aad3517
45 Khera, A. V. *et al.* Association of Rare and Common Variation in the Lipoprotein Lipase Gene With Coronary Artery Disease. *JAMA* **317**, 937-946 (2017). https://doi.org/10.1001/jama.2017.0972

46 Hajjar, K. A. The Biology of Annexin A2: From Vascular Fibrinolysis to Innate Immunity. *Trans Am Clin Climatol Assoc* **126**, 144-155 (2015).
47 Ling, Q. *et al.* Annexin II regulates fibrin homeostasis and neoangiogenesis in vivo. *J Clin Invest* **113**, 38-48 (2004). https://doi.org/10.1172/JCI19684
48 Molkentin, J. D. *et al.* A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215-228 (1998). https://doi.org/10.1016/s0092-8674(00)81573-1
49 Sabine, A. *et al.* Mechanotransduction, PROX1, and FOXC2 cooperate to control connexin37 and calcineurin during lymphatic-valve formation. *Dev Cell* **22**, 430-445 (2012). https://doi.org/10.1016/j.devcel.2011.12.020
50 Lowe, C. E. *et al.* Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet* **39**, 1074-1082 (2007). https://doi.org/10.1038/ng2102
51 Walker, V. M., Kehoe, P. G., Martin, R. M. & Davies, N. M. Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian randomization study. *Int J Epidemiol* **49**, 1132-1140 (2020). https://doi.org/10.1093/ije/dyz155
52 Dennis, J. *et al.* The endothelial protein C receptor (PROCR) Ser219Gly variant and risk of common thrombotic disorders: a HuGE review and meta-analysis of evidence from observational studies. *Blood* **119**, 2392-2400 (2012). https://doi.org/10.1182/blood-2011-10-383448
53 Childs, B. G. *et al.* Senescent intimal foam cells are deleterious at all stages of atherosclerosis. *Science* **354**, 472-477 (2016). https://doi.org/10.1126/science.aaf6659
54 Franceschi, C. & Campisi, J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J Gerontol A Biol Sci Med Sci* **69 Suppl 1**, S4-9 (2014). https://doi.org/10.1093/gerona/glu057
55 Anderson, R. *et al.* Length-independent telomere damage drives post-mitotic cardiomyocyte senescence. *EMBO J* **38** (2019). https://doi.org/10.15252/embj.2018100492
56 Hayek, S. S., Quyyumi, A. A. & Reiser, J. Soluble Urokinase Receptor and Chronic Kidney Disease. *N Engl J Med* **374**, 891 (2016). https://doi.org/10.1056/NEJMc1515787
57 Doran, A. C., Yurdagul, A., Jr. & Tabas, I. Efferocytosis in health and disease. *Nat Rev Immunol* **20**, 254-267 (2020). https://doi.org/10.1038/s41577-019-0240-6
58 Eldjarn, G. H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348-358 (2023). https://doi.org/10.1038/s41586-023-06563-x
59 Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic evidence on clinical success. *Nature* **629**, 624-629 (2024). https://doi.org/10.1038/s41586-024-07316-0

Tables

Table 1. Data Sources Information

Layer	Source	Role	Access
Protein universe	Olink Explore (2,923 assays)	Defines the measurable plasma proteome	Public (assay identities)
Annotation	Human Protein Atlas	Immune-cell lineage, secretome, druggability	Public
Annotation	MSigDB C7/C8	Immunologic signatures (annotation only)	Public

Layer	Source	Role	Access
Genetic exposure	eQTLGen cis-eQTL	Transcript-level genetic instruments	Public
Protein validation	INTERVAL cis-pQTL (Sun et al)	Protein-level genetic instruments (110 proteins, 219 diseases)	Public (GWAS Catalog)
Disease outcome	FinnGen R12 (all 2,469 endpoints, 2,466 returning data; 28 deeply annotated)	Disease genome-wide association statistics	Public
Replication	UK Biobank (Neale) exact code and normalized name match; OpenGWAS studies	Two-population/out-of-sample validation	Public/free token
Immune-trait outcomes	UK Biobank blood counts; Astle et al 2016	Blood immune-cell and inflammation traits	Public
Excluded (gated)	UKB-PPP; deCODE; UK Biobank individual NPX	Not used, by design	Restricted/data-use agreement

cis-eQTL indicates cis-expression quantitative trait locus; cis-pQTL, cis-protein quantitative trait locus; MSigDB, Molecular Signatures Database; NPX, Normalized Protein eXpression; R12, Freeze R12; and UKB-PPP, UK Biobank Pharma Proteomics Project.

Table 2. Pathway-Level Directional Concordance of the Discovery Engine With Established Therapies

Target to disease	OR	Genetically inferred direction	Coloc. PP.H4	Replication P	Therapeutic axis
IL6ST to rheumatoid arthritis	2.63	Block	1.00	$7x10^{-24}$	IL-6 signaling axis (tocilizumab blocks IL-6R; IL6ST encodes gp130)
CTLA4 to rheumatoid arthritis	0.51	Agonize	0.98	$3x10^{-20}$	CD28 costimulation axis (abatacept, CTLA4-Ig, binds CD80/CD86)
CTLA4 to autoimmune hyperthyroidism	0.15	Agonize	0.84	$4x10^{-19}$	CD28 costimulation axis (abatacept, CTLA4-Ig, binds CD80/CD86)
TNFRSF1A to ankylosing spondylitis	1.46	Block	0.86	$6x10^{-6}$	TNF axis (etanercept, soluble p75 TNFR-Fc, neutralizes TNF)
IL4 to psoriasis	1.80	Block	-	$4x10^{-4}$	IL-4/IL-13 axis (dupilumab blocks IL-4Ra, not IL-4 itself)
TNFSF14 to multiple sclerosis	0.50	Agonize	>0.97*	$2x10^{-12}$	LIGHT-HVEM costimulatory axis
ACE to antihypertensive medication use	2.12	Block	~1.00	-	ACE inhibitors (cardiovascular positive control)

Coloc. indicates colocalization; HVEM, herpesvirus entry mediator; IL, interleukin; LIGHT, TNF superfamily member 14; OR, odds ratio; PP.H4, posterior probability of a shared causal variant; and TNF, tumor necrosis factor. *Colocalized at both transcript and protein level (protein-level causal target); replication in the independent International Multiple Sclerosis Genetics Consortium study. The ACE row is the cardiovascular pharmacogenomic mirror described in Results; the medication-use endpoint has no independent replication counterpart and is therefore left blank. Rows demonstrate pathway-level directional concordance between the genetic signal and the drug mechanism; the MR-instrumented gene and the drug's direct molecular target may be distinct components of the same axis.

Table 3. Evidence-Tier Distribution Across the Whole Phenome After Phenome-Wide Protein-Level Validation

Tier	Definition	Count	Cardiovascular-domain examples
Tier 1	Known-drug positive control recovered from genetics	90	ACE (antihypertensive medication use)
Tier 2	Nomination held for MHC/linkage-disequilibrium caution	115	C2, C4A/C4B loci
Tier 3	Genetically supported nomination (cis-MR, FDR <0.05)	447	CDKN1A (heart failure), PROCR (venous thromboembolism)
Tier 4	Prioritized transcript-level causal target (adds colocalization PP.H4 >=0.8)	359	PSRC1, PLAUR, FES, MERTK, IFNGR2
Tier 5	Protein-level causal target (adds concordant plasma pQTL MR and protein colocalization)	5	ANXA2 (hypertension, both definitions); PPP3R1 (venous/lymphatic disease)

Tiers are assigned over all 1,016 FDR-significant gene-disease pairs across 381 diseases after the plasma cis-pQTL layer was extended from the autoimmune arc to the whole phenome (110 proteins, 219 diseases). The 5 tier 5 targets are IL2RA in type 1 diabetes, ANXA2 in hypertension under both endpoint definitions, and PPP3R1 in varicose veins and in venous/lymphatic disease. A further 35 pairs are labeled novel protein-confirmed (transcript colocalization plus a direction-concordant protein-level pQTL MR). FDR indicates false discovery rate; MHC, major histocompatibility complex; MR, Mendelian randomization; pQTL, protein quantitative trait locus; and PP.H4, posterior probability of a shared causal variant.

Table 4. Cardiovascular Causal Architecture of the Plasma Immunome (28-Disease Core)

Endpoint	Causal proteins, n	Leading risk-raising proteins (OR)	Leading protective proteins (OR)	Colocalized signals
Hypertension	28	ACE (1.61), IFNGR2 (1.07), MERTK, ANXA2, JMJD1C	FES, SWAP70, MAPRE3, SPINK8	ACE, FES, IFNGR2, MERTK, ANXA2*
Coronary heart disease	15	PLAUR (1.59)	PSRC1 (0.85), SCARB1, LPL, FES	PSRC1, PLAUR, SCARB1
Atrial fibrillation/flutter	12	ACE (1.72), CDKN1A	FES, PFKFB2, ADAM15	ACE, FES, PFKFB2
Venous thromboembolism	8	SCPEP1, TGFB2	PROCR (0.55), PDLIM7, TSPAN15	PDLIM7, SCPEP1, PPP3R1*

Endpoint	Causal proteins, n	Leading risk-raising proteins (OR)	Leading protective proteins (OR)	Colocalized signals
Heart failure	4	CDKN1A (1.48), GSTM4	PFKFB2, RAB44	-
Stroke	0 (in core)	-	-	-

Six cardiovascular endpoints were included in the 28-disease core; the 5 shown with causal proteins sum to the 67 cardiovascular causal pairs, and stroke carried no FDR-significant pair at core scale. Restricting the whole-phenome scan to all cardiovascular endpoints extends this map to 156 FDR-significant effects across 32 endpoints and 46 immune proteins, in which BAG3 emerges as strongly protective in heart failure, ERBB2 joins CDKN1A as an atrial fibrillation risk node, and the protective coronary axis (PSRC1, FES) extends to stroke. *Reaches protein-level (tier 5) causality after phenome-wide extension of the plasma cis-pQTL layer; PPP3R1 is annotated in the venous/lymphatic chapter. Colocalized indicates shared-variant colocalization at PP.H4 >=0.8; FDR, false discovery rate; OR, odds ratio per 1-SD genetically higher expression; and PP.H4, posterior probability of a shared causal variant.

Table 5. Two-Population Replication: Finland (FinnGen) Versus England (UK Biobank)

Matching route	Diseases	Immune proteins	Pairs tested	Same direction (%)	Two-population validated
Exact endpoint code	48	71	133	110 (83)	58
Normalized trait name (single-cohort UK Biobank only)	14	-	56	52 (93)	31
Combined	62	90	189	162 (86)	89

Two-population validated denotes the same causal direction in FinnGen and UK Biobank with UK Biobank P<0.05. Only single-cohort UK Biobank datasets (ukb-d-, ukb-b-, ukb-a-, ukb-e-) were eligible for the name-matched route, because many OpenGWAS studies with matching trait names are meta-analyses that include FinnGen and would therefore make replication circular; this restriction reduced the reachable disease count from 81 to 62 and was accepted deliberately. Name-matched datasets additionally required at least 200 cases and a European or unspecified ancestry label. The exact-code subset reproduces identically within the combined table, so widening added pairs without perturbing existing ones. The combined table feeds the replication term of the novelty-priority score.

Supplementary information

Additional file 1: Supplementary Material (Figures S1-S7 and Tables S1-S2), containing the following items.

Figure S1. Curation and evidence architecture of the 1,007-protein plasma immunome.

Figure S2. Druggability, cellular sourcing, and molecular-layer architecture of the plasma immunome.

Figure S3. Instrument construction, discovery-arm Mendelian randomization, and per-disease causal yield.

Figure S4. Phenome-wide causal map across all 2,469 FinnGen R12 endpoints (phenome-wide Manhattan by International Classification of Diseases chapter; 30 strongest immune-to-disease effects; per-chapter reach of the immune proteome).

Figure S5. Deep-validation layers over the whole-phenome scan (phenome-wide colocalization distribution and top colocalized pairs; widened FinnGen to UK Biobank 2-population replication across 189 pairs; extended blood immune-trait Mendelian randomization with no lineage omitted).

Figure S6. The cardiovascular evidence chain (cardiovascular causal forest; per-endpoint volcano plots for coronary disease, hypertension, atrial fibrillation, heart failure, and stroke; triangulated coronary targets across causal, colocalization, and replication layers; druggable cardiovascular targets ranked by the atlas engine; ACE to antihypertensive-medication worked example; Finland to England replication).

Figure S7. Design of the cardiovascular Plasma Immune Risk Score (2-population-validated causal coronary proteins as interpretable, direction-signed features).

Table S1. Widened 2-Population Replication, Finland (FinnGen) Versus England (UK Biobank)

Matching route	Eligible datasets	Diseases	Pairs tested	Same direction (%)	Two-population validated
Exact endpoint code (ukb-d-<phenocode>)	Single-cohort UK Biobank	48	133	110 (83)	58
Normalized trait name	Single-cohort UK Biobank only (ukb-d-, ukb-b-, ukb-a-, ukb-e-); >=200 cases; European or unspecified ancestry	14	56	52 (93)	31
Combined (feeds novelty-priority score)	-	62	189	162 (86)	89

Meta-analysis datasets whose trait names match FinnGen endpoints were excluded because many silently include FinnGen (for example the Sakaue 2021 ebi-a-GCST90018* series combines Biobank Japan, UK Biobank, and FinnGen), which would make replication circular. This restriction reduced the reachable disease count from 81 to 62 and was accepted deliberately to preserve the Finland-versus-England 2-population design. The exact-code subset reproduces identically within the combined table, so widening added pairs without perturbing existing ones. Newly reachable diseases are predominantly non-autoimmune, including varicose veins, ankylosing spondylitis, hernia, and spondylopathy. Full per-pair results, including the match_method, UK Biobank trait name, and case count for every pairing, are released in uk_panphenome_concordance_ALL.tsv.

Table S2. Protein-Level (Tier 5) and Novel Protein-Confirmed Gene-Disease Pairs

Category	n	Members
Tier 5 (colocalized, protein-level pQTL concordant, specific, druggable)	5	IL2RA (type 1 diabetes); ANXA2 (hypertension, both endpoint definitions); PPP3R1 (varicose veins; venous/lymphatic disease)
Novel protein-confirmed (transcript colocalization plus direction-concordant protein-level pQTL Mendelian randomization)	35	ANXA2, CD70, CD84, CFH, CTSH, DAPK2, FCRL3, GZMB, HGF, IL2RA, LTBR, PPP3R1, PRTN3, SPINK8, SWAP70, TGFB2, TNFSF12
Protein-layer coverage after phenome-wide extension	-	110 proteins across 219 diseases
Not reachable at protein level (no plasma aptamer)	110	Largely intracellular or major histocompatibility complex-region proteins, exemplified by GRK5

Tier assignment was rebuilt over all 1,016 false discovery rate-significant gene-disease pairs across 381 diseases after the INTERVAL plasma cis-protein quantitative trait locus (pQTL) layer was extended from the autoimmune arc to the whole phenome, returning 90 tier 1, 115 tier 2, 447 tier 3, 359 tier 4, and 5 tier 5 pairs. ANXA2 in hypertension is the first protein-level causal cardiovascular target in this resource. The 35 novel protein-confirmed pairs span 17 distinct proteins, each listed once in the Members column of this table.Additional file 2: Data S1. Precomputed atlas result tables (whole-phenome cis-Mendelian randomization results; phenome-wide colocalization posteriors; widened 2-population concordance, uk_panphenome_concordance_ALL.tsv; extended blood immune-trait Mendelian randomization; novelty-priority ranking; final evidence-tier table with phenome-wide protein-level annotation).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterialManuscript15.docx