Human Plasma Proteomic Profile of Clonal Hematopoiesis

Zhi Yu[1,2], Amélie Vromman[3], Ngoc Quynh H. Nguyen[4,5], Shamsudheen K. Vellarikkal[3], Md Mesbah Uddin[1,2], Abhishek Niroula[1,6,7], Gabriel Griffin[1,8,9], Michael C. Honigberg[1,2,10], Amy E. Lin[1,3,6], Christopher J. Gibson[1,6], Daniel H. Katz[11,12], Usman Tahir[10,11], Sara Haidermota[1,2], Shriienidhie Ganesh[1,2], Tajmara Antoine[1,2], Joshua Weinstock[13,14], Thomas R. Austin[15,16], Vasan S. Ramachandran[17,18], Gina M. Peloso[19], Whitney Hornsby[1,2], Peter Ganz[20], JoAnn E. Manson[21,22], Bernhard Haring[23,24], Charles L. Kooperberg[25], Alexander P. Reiner[25,26], Joshua C. Bis[14,27], Bruce M. Psaty[15,16,26,27], Yuan-I Min[28], Adolfo Correa[28], Leslie A. Lange[29], Wendy S. Post[30], Jerome I. Rotter[31], Stephen S. Rich[32], James G. Wilson[11], Benjamin L. Ebert[1,6,10], Bing Yu[4,5], Christie M. Ballantyne[33], Josef Coresh[34], Vijay G Sankaran[1,35,36,37], Alexander G. Bick[38], Siddhartha Jaiswal[39], Peter Libby[3], Robert E. Gerszten[10,11], NHLBI Trans-Omics for Precision Medicine, Rajat M Gupta*[3], Pradeep Natarajan*[1,2,10]

*Indicate co-senior authorship

1. Broad Institute of MIT and Harvard, Cambridge, MA, USA

2. Cardiovascular Research Center and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

3. Cardiovascular Division, Brigham and Women's Hospital Heart & Vascular Center, Boston, MA, USA

4. Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA

5. Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

6. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

7. Department of Laboratory Medicine, Lund University, Lund, Sweden

8. Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA

9. Department of Pathology, Dana-Farber Cancer Institute, Boston, MA, USA

10. Department of Medicine, Harvard Medical School, Boston, MA, USA

11. Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

12. Department of Cardiovascular Medicine, Stanford University, Stanford, CA, USA

13. Department of Genetics, Stanford University, Stanford, CA, USA

14. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

15. Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA.

16. Department of Epidemiology, University of Washington, Seattle, WA, USA.

17. Department of Medicine, School of Medicine, Boston University, Boston, MA, USA

18. Framingham Heart Study, Framingham, MA, USA

19. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

20. Division of Cardiology, Zuckerberg San Francisco General Hospital and Department of Medicine, University of California, San Francisco, CA, USA

21. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

22. Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

23. Department of Medicine III, Saarland University Hospital, Homburg, Germany

24. Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA

25. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

26. Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA

27. Department of Medicine, University of Washington, Seattle, WA, USA

28. Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA

29. Department of Medicine, Division of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus, Denver, CO, USA

30. Division of Cardiology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

31. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA

32. Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

33. Department of Medicine, Baylor College of Medicine, Houston, TX, USA

34. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

35. Division of Hematology and Oncology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.

36. Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA.

37. Harvard Stem Cell Institute, Cambridge, MA, USA.

38. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

39. Department of Pathology and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

Please address correspondence to:

Pradeep Natarajan, MD MMSc

185 Cambridge Street, CPZN 3.184, Boston, MA 02114

617-726-1843

pnatarajan@mgh.harvard.edu

Abstract

Plasma proteomic profiles associated with subclinical somatic mutations in blood cells may offer novel insights in downstream clinical consequences. Here, we explore such patterns in clonal hematopoiesis of indeterminate potential (CHIP), which links to several cancer and non-cancer outcomes, including coronary artery disease. Among 12,911 ancestrally diverse participants (682 with CHIP) from NHLBI TOPMed with blood-based DNA sequencing and 1,148 common proteins measured by SomaScan, we identified 32 unique proteins associated with the most prevalent driver genes (*DNMT3A*, *TET2*, and *ASXL1*) after multiple testing corrections. These associations showed substantial heterogeneity by driver genes, sex, and race, were enriched for immune response and inflammation pathways, and were moderately replicated in UK Biobank (N=48,922) that used Olink for proteomics measurement. Murine single-cell RNA-sequencing data from aortic arch cells, inclusive of resident hematologic cells, in mice with *Tet2*$^{-/-}$ bone marrow and wild-type mice revealed corroborating differential expression of *TET2*-associated protein-encoding genes. Lastly, we apply these observations to identify 68 plasma proteins shared between CHIP and coronary artery disease.

Introduction

Clonal hematopoiesis of indeterminate potential (CHIP) is a common age-related phenomenon defined as the presence of expanded hematopoietic stem cell (HSC) clones caused by acquired leukemogenic mutations (e.g., *DNMT3A*, *TET2*, *ASXL1*, and *JAK2*) in persons without clinical hematologic abnormalities[1,2]. CHIP is a pre-cancerous lesion strongly predictive of hematologic malignancy[3,4]. In addition, CHIP predisposes an individual to other age-related human diseases, chiefly cardiovascular diseases, in both human genetic and murine experimental studies[4-12].

Characterizing the consequences of CHIP mutations on the plasma proteome may facilitate an improved understanding of how CHIP influences clinical outcomes. Recent studies have associated CHIP with germline DNA variation[13-15], bulk RNA transcript concentrations[16,17], and epigenomic profiles[18,19] for such insights. While proteins represent key downstream effector gene products, their associations with CHIP remain largely unknown. The circulating proteins are involved in numerous biological processes; surveying the proteome might offer new insights into CHIP and its mechanistic link to disease phenotypes[20].

Leveraging paired DNA sequencing and proteomic profiling from multi-ancestry participants of four Trans-Omics for Precision Medicine (TOPMed) cohorts (N=12,911) and UK Biobank (UKB; N= 48,922), we explored the proteomic signatures of CHIP and its most common or most disease-promoting driver genes (*DNMT3A*, *TET2*, *ASXL1*, and *JAK2*). We validated our population-based observations with a murine model and orthogonal profiling. We generated single-cell RNA-sequencing (scRNA-seq) data from mice with *Tet2*-deficient or wild-type myeloid cells, and the objective was to determine whether there was alignment between differentially expressed genes in mice and the genes encoding CHIP-associated proteins in

humans. Lastly, we explored the functional implications of CHIP-associated proteins through pathway analyses and by examining the shared and non-shared pathways between CHIP and CAD (**Figure 1**).

Results

CHIP and Proteomics Characterization in Participants Across Multiple Cohorts

Our discovery study population comprised 12,911 participants with CHIP genotyping from deep-coverage whole genome or exome sequencing of blood DNA and concurrent plasma proteomics data from four TOPMed cohorts, utilizing SomaScan assay for proteomics measurements[13,21,22]. The four cohorts are the Jackson Heart Study (JHS; N=2,058)[23], Multi-Ethnic Study of Atherosclerosis (MESA; N=976)[24], Cardiovascular Health Study (CHS; N=1,689)[25,26], and Atherosclerosis Risk in Communities (ARIC; N=8,188) Study[27]. We also used data from 48,922 participants in the UKB with proteomics measured through Olink for replication[28].

Overviews of the study cohorts are described in detail in Materials and Methods. In the samples obtained from the five cohorts, 3,881 (6.0%) individuals were identified as having CHIP. Consistent with previous reports[13,29], CHIP was robustly associated with age. Across all cohorts, approximately 90% of individuals with CHIP driver mutations had only one identified mutation. The most commonly mutated driver genes, *DNMT3A*, *TET2*, and *ASXL1*, accounted for >75% of individuals with CHIP mutations. The variant allele fraction (VAF) distributions of mutations in each driver mutation were relatively consistent across participating cohorts. Plasma proteomics was measured using SomaScan[21] in the discovery cohorts (i.e., ARIC, CHS, JHS, and MESA) and was measured using Olink in UKB[30], which was used for cross-platform replication in our study (**Figure 2**, **Supplemental Tables 1-2)**.

Diverse Proteomic Associations Across CHIP Driver Genes

CHIP was modeled both as a composite and separately for the most common or pathogenic drivers (*DNMT3A*, *TET2*, *ASXL1*, and *JAK2*), and defined both using the conventional thresholds for all mutations (VAF ≥2%) and for the expanded (large) clones (VAF ≥10%)[1], resulting in ten CHIP exposure variables. The cross-sectional associations between CHIP mutations and 1,148 plasma proteins present in all discovery cohorts were estimated within each cohort and then meta-analyzed. Consistent with prior modeling of proteomic analyses of ARIC[31], we separated ARIC into two subpopulations: ARIC European Ancestry (EA) and ARIC African Ancestry (AA). *JAK2* analyses were only conducted in cohorts with greater than 5 participants with *JAK2* mutations which only retained CHS and ARIC EA for these analyses; thus, *JAK2* analyses were considered secondary.

Since the associations between proteins and all CHIP mutations (i.e., VAF ≥2%) are highly correlated to those with their corresponding expanded mutations (i.e., VAF ≥10%) (**Supplemental Table 3**), we retained the one with the stronger association (i.e., larger absolute Z score) to maximize power. This led to the identification of 35 significant CHIP variable-protein pairs (false discovery rate [FDR]<0.05, 4,592 testings), representing 32 unique proteins, independent of potential confounders described in detail in Materials and Methods (**Figure 3** and **Supplemental Table 4**). Adding *JAK2* increased the number of significant pairs to 107 (**Supplemental Figure 1** and **Supplementary Table 4**).

Consistent with prior work implicating heightened interleukin (IL)-1β, NOD-, LRR- and pyrin domain-containing protein 3 (NLRP3), IL-6R pathways in CHIP biology[9,13,16,17,32], the 107 proteins associated with examined CHIP mutations (including *JAK2*) at FDR=0.05 level were

similarly enriched in these pathways. For example, several interleukin receptors were implicated, with CHIP mutations positively associated with the IL-23 receptor, IL-1 receptor type 2 (IL-1R2), and cytokine receptor common subunit gamma, and negatively associated with IL-22 receptor subunit alpha-2, an IL-22 antagonist that inhibits inflammation. We also observed positive associations with tumor necrosis factor superfamily (TNFSF) members, such as TNFRSF10D and EDAR. In addition, a number of C-X-C motif chemokines were all positively associated with CHIP mutations, such as C-X-C motif chemokine ligand (CXCL) 5, CXCL11, and CXCL13, as well as C-C motif chemokines, including C-C motif chemokine ligand (CCL) 5, CCL17, CCL22, and CCL28. These chemokines play a role in immune cell recruitment and activation during inflammation and also contribute to the production and regulation of IL-1β and IL-6.

Mutations in individual CHIP genes exhibited distinct proteomic associations. Among the primary CHIP variables, *TET2* demonstrated the largest number of associations with plasma proteins, associating with 16 proteins at the FDR<0.05 threshold and 26 at the FDR<0.1 threshold. *ASXL1* was associated with 11 proteins at the FDR<0.05 threshold. In contrast, despite being the most prevalent, *DNMT3A* was significantly associated with only four proteins. Furthermore, proteins associated with composite CHIP were typically driven by individual mutant genes (**Figure 3** and **Supplementary Table 4**). Despite its infrequency and restricted sample size for secondary analysis, *JAK2* was associated with more proteins than all primary CHIP mutations examined, with 54 proteins associated. Proteins associated with *JAK2* also generally differed from other examined driver genes (**Supplemental Figure 5** and **Supplementary Table 4**).

TET2 was associated with proteins involved in extracellular matrix (ECM) remodeling and immune regulation. For example, the top two associated proteins, pappalysin-1 (PAPPA) and secreted protein, acidic and rich in cysteine (SPARC), both participate in ECM remodeling [33-39]. Detailed protein functions are discussed in the **Supplemental Text**. For participants with *TET2* mutations, the plasma PAPPA and SPARC levels were 22% and 8.2% lower, respectively, than those without *TET2* mutations (FDR = 4.6×10^{-8} and 3.8×10^{-4}, respectively). We observed significant associations between *TET2* CHIP with several proteins related to immune regulation. For example, the third and fifth strongest associated proteins CXCL13 and CCL5 (both positive associations; FDR = 3.8×10^{-4} and 6.7×10^{-3}, respectively) are implicated in regulating IL-1β and IL-6 levels as mentioned above, and a few other proteins are involved in innate immunity, such as neutrophil gelatinase-associated lipocalin (NGAL; negative association; FDR = 1.1×10^{-3}) and myeloperoxidase (positive association; FDR = 0.02) (**Figure 3** and **Supplementary Table 4**).

In addition to immune regulation, *ASXL1*-associated proteins were enriched in metabolic regulation, cell signaling, and cellular function. For example, the top protein associated with *ASXL1* was carbonic anhydrase 1 (CA1)[40]. CA1 is not only the strongest associated protein for *ASXL1*, but it is also significantly associated with *TET2* and overall CHIP, with its levels being 15.9% (FDR = 2.4×10^{-7}), 8.2% (FDR = 0.01), and 4.3% (FDR = 6.7×10^{-3}) lower among participants with *ASXL1*, *TET2*, and composite CHIP, respectively, than those without CHIP. *ASXL1* was also associated with resistin, which is primarily expressed in myeloid cells and has been previously linked to diabetes (negative association; FDR = 6.7×10^{-3})[41]. Sphingosine kinase 1 (SPHK1; negative association; FDR = 2.0×10^{-6}), which catalyzes the phosphorylation of sphingosine-1-phosphate, a key regulator of cellular proliferation and inflammation, was also associated. Additionally, lymphatic vessel endothelial hyaluronic acid receptor 1 (LYVE1;

positive association; FDR = 2.8×10^{-6}), which plays a key role in the trafficking of immunity-related cells during inflammation, wound healing, and neoplasia, was also associated (**Figure 3** and **Supplementary Table 4**)[42,43].

In the secondary analysis, *JAK2* associated with 54 unique proteins exhibiting diverse functions. Top associated proteins, including P-selectin, platelet glycoprotein 1b alpha chain, and thrombospondin-1, play crucial roles in cell adhesion and platelet function had greater concentrations among those with *JAK2*[44,45] similar to knock-in mice with inducible *JAK2*^{V617F}[13,46]. Additionally, individuals with *JAK2* CHIP had reduced erythropoietin concentrations, which is also observed among individuals with *JAK2* myeloproliferative neoplasms (**Supplemental Figure 1** and **Supplemental Table 4**)[47]. Given that many *JAK2*-associated proteins were related to blood cells, we conducted sensitivity analyses in ARIC, additionally adjusting for platelet and white blood cell (WBC) count. The results were largely robust. The associations between *JAK2* and a few proteins directly related to platelet, such as platelet glycoprotein 1b alpha chain, were dampened but remained firmly statistically significant (**Supplemental Table 5** and **Supplemental Figure 6**).

For the abovementioned analysis, not adjusting for probabilistic estimation of expression residuals (PEER) factors for accounting for latent factors in proteomics measurements or additionally adjusting for estimated glomerular filtration rates (eGFR) yielded consistent results (**Supplemental Table 6**, **Supplemental Table 7,** and **Supplemental Figure 7-8**)

Sex-specific and Race-specific Differences in CHIP Variable-Protein Associations

We conducted stratified analyses by sex and race, revealing distinct CHIP variable-protein associations. While there was no difference in the prevalence of composite CHIP and each

examined driver gene by sex, more proteins were associated with CHIP mutations in males than in females despite similar counts, with minimal overlap between significantly associated proteins across sex groups. The strong association of *ASXL1* with CA1 was only present in male-only analysis, with an FDR of 8.1×10^{-7}, in contrast to an FDR of 0.23 in females. Among women, more significant associations were from *TET2*, with the top *TET2*-associated protein in combined analyses, PAPPA, SPARC, CXCL13, and NGAL, remaining significant in the female-only analysis. In contrast, the association with PAPPA was much dampened, and the associations with SPARC, CXCL13, and NGAL were not significant in males (**Supplemental Table 8-9** and **Figure 3B-3C**). For the 40 CHIP variable-protein pairs that are only significant in male or female stratified analysis, we introduced and tested for interaction terms between the corresponding CHIP variables and sex in the combined analysis across all discovery cohorts. Of these, 15 pairs displayed statistically significant interactions at an FDR = 0.05 (**Supplemental Table 10**).

Proteins associated with CHIP mutations also differ by self-reported race. Among individual driver genes, only *DNMT3A* demonstrated significant proteomic associations in the Black-only analysis, with two out of three associated proteins, namely sialic acid-binding Ig-like lectin 6 (Siglec-6; positive association; FDR = 0.01) and mitogen-activated protein kinase 1 (MAK1; positive association; FDR = 0.03), not observed in combined analyses. In contrast, significant associations in the White-only analyses were primarily driven by *TET2* and *ASXL1* and largely reflected findings in combined analyses. And these significant associations were not present in Black-only analyses (**Supplemental Tables 11-12** and **Figures 3D-3E**). We tested for interaction terms between CHIP variables and race in combined analysis in ARIC for the 18

proteins that are only significant in Black or White stratified analysis. Three pairs displayed statistically significant interactions at an FDR = 0.05 (**Supplemental Table 13**). *JAK2* analyses yielded similar sex and race-specific patterns (**Supplemental Tables 8**, **9, 11, and 12** and **Supplemental Figures 2-5**).

Enriched Biological Pathways and Protein Networks

We next performed pathway analyses to investigate the biological processes and regulatory mechanisms linked to the collective function of proteins associated with each CHIP driver gene. The examined CHIP driver genes were broadly enriched in immune response and inflammation-related pathways and disease processes (**Supplemental Figures 9-12**). However, the significantly modulated pathways of different driver genes were involved in different immune activities and exhibited divergent effects. In addition to activating the cardiac hypertrophy signaling pathway that aligns with the observations of *DNMT3A*-mediated CH in heart failure[17], *DNMT3A*-associated proteins activated pathways involved in acute responses to wound healing signaling and pathogen-induced cytokine storm signaling pathways[48]. In contrast, *TET2*-associated proteins modulated pathways implicated in autoimmunity and promoted chronic inflammation, activating the IL-17, STAT3, and IL-22 signaling pathways and inhibiting LXR/RXR activation. While *DNMT3A* and *TET2* appeared pro-inflammatory, *ASXL1* was linked to a number of reduced pro-inflammatory pathways, such as the STAT3 pathway that was predicted to be activated in the *TET2* pathway analysis and established IL-6 signaling pathway (**Figure 4**). In secondary analyses, *JAK2*-associated proteins modulated tissue remodeling pathways, such as cardiac hypertrophy and pulmonary fibrosis idiopathic signaling pathways (**Supplemental Figures 13**).

Mouse Single-cell Transcriptomic Validations of Human Proteomic Findings

To validate the mechanistic links between the human *TET2*-proteomic associations in a controlled experimental setting, we performed single cell RNA-sequencing (scRNA-seq) of 160,495 cells from 20 aortic arches of 8 atherosclerosis-prone mice, four of which were transplanted with *Tet2*$^{-/-}$ bone marrow and the other four with *Tet2*$^{+/+}$ bone marrow[32]. Using unsupervised clustering, we partitioned the cells into groups and visualized them via Uniform Manifold Approximation and Projection (UMAP). Individual clusters were labeled for cell type using known marker genes (**Supplemental Table 14**)[49]. We identified all major vascular cell subpopulations, including vascular smooth muscle cells (VSMCs; N_{cell}=96,857), fibroblasts (N_{cell}=14,596), myofibroblasts (N_{cell}=7,956), macrophages (N_{cell}=15,265), endothelial cells (ECs; N_{cell}=7,316), B cells (N_{cell}=3,513), T cells (N_{cell}=4,810), erythrocytes (N_{cell}=6,436), neuron cells (N_{cell}=3,057), and monocytes (N_{cell}=689) (**Figure 5**). Furthermore, we identified cellular subpopulations within aortic cell types, where VSMCs, fibroblasts, myofibroblasts, macrophages, ECs, and neuron cells all formed multiple subpopulations that exhibited greater similarity to each other than other aortic cell types (**Supplemental Figure 14**). We noted an elevated atherosclerotic plaque burden and increased cell numbers for most cell types in mice with *Tet2*$^{-/-}$ bone marrow compared to wild-type mice (**Supplemental Figure 15**). Among all cell types, B cells did not increase and T cells exhibited the smallest percentage increase in cell numbers (**Supplemental Table 15**).

We pursued two approaches to validate our human proteomic findings using murine scRNA-seq data: (1) Assess differential expression of mouse genes encoding proteins significantly associated with *TET2* in humans, and (2) examine potential enrichment of the most

differentially expressed genes between *Tet2*$^{-/-}$ and control mice in *TET2*-associated human proteins. For the first approach, we extracted genes encoding *TET2*-associated proteins at FDR = 0.10, matched them to their corresponding murine cognate gene orthologue (**Supplemental Table 11**), and examined their expression levels across the ten cell types by *Tet2* genotype. We examined well-expressed genes (defined as evidence of expression in greater than 30% of cells using a read count metric for expression; **Supplemental Table 12**) in the corresponding cell types. We generally observed concordant effect directions as with human proteomic associations (**Figure 6**). scRNA-seq allowed analysis of cell-type specific effects, which was important to identify the association for three genes that had expression in only one or two cell types. We observed a 9.8-fold and 11.1-fold increase in *Cxcl13* and *Cfp* expression in monocytes and a 0.46-fold decrease in *Fcer2a* expression in B cells in *Tet2*$^{-/-}$ mice compared to control mice, consistent with human proteomic associations. For genes broadly expressed across all cell types, including *Sparc*, *Lyz2*, *Eef1b2*, *Timp3*, and *Rps27a*, the large majority of cell types (particularly blood cells) exhibited significant differential expressions directionally aligned with human proteomic associations (**Figure 6** and **Supplemental Table 14**).

In reverse fashion, we compiled the top 25 differentially expressed genes in Tet2-/- versus wild-type mice of each cell type (119 unique genes), and investigated whether their corresponding proteins were associated with *TET2* CHIP in humans (**Supplemental Table 14**). Among the 16 genes with corresponding proteins in SomaScan, only SPARC was significantly associated with *TET2* at the FDR=0.05 threshold (**Supplemental Table 15**).

Comparative Analysis of Proteomic Associations Across Platforms

We compared CHIP variable-protein associations between results from the main SomaScan-based analysis (N=12,911) and results from UKB (N=48,922), which utilized the Olink platform for proteomics measurements[21,30]. There were 471 unique proteins shared between the two platforms. Among the 1,884 CHIP variable-protein pairs being compared (471 proteins * 4 CHIP variables), 35% was nominally significant in at least one of the SomaScan-based and Olink-based results. Among them, 97 was nominally significant in both sets of results, with 12 of them being significant after correcting for multiple-testing (FDR=0.05, 1,884 testings) in both SomaScan-based and Olink-based results. Those pairs include the strong signals of *ASXL1*, composite CHIP, and *TET2* with CA1 (SomaScan results: FDR = 9.8×10^{-8}, 3.7×10^{-3}, and 5.8×10^{-3}, respectively; Olink results: FDR = 3.7×10^{-5}, 1.9×10^{-4}, and 1.9×10^{-4}, respectively), as well as the *TET2* and low affinity immunoglobulin epsilon Fc receptor (FCER2; SomaScan results: FDR = 3.7×10^{-3}; Olink results: FDR = 5.4×10^{-4}) association that was validated in mice with *Tet2*$^{-/-}$ bone marrow, and others (**Supplemental Figure 16**). Since only 19% of overlapping proteins are highly correlated between the two platforms in prior work[50], we can not rule out the possibility of false positives and false negatives of this cross-platform replication (**Supplemental Table 16**).

Shared Proteomic Associations in CHIP and CAD

We next investigated the shared proteomic associations between CHIP mutations and CAD. We analyzed the cross-sectional associations between prevalent CAD, which were assessed at the visits of blood draws to maintain temporal consistency with CHIP measurement, and proteomics. Top CAD-associated proteins include known CAD biomarkers, such as N-terminal pro-BNP (NT pro-BNP; FDR = 5.7×10^{-13}), C-reactive protein (CRP; FDR = 2.7×10^{-4}), and troponin I and T

(FDR = 5.2×10^{-3} and 7.4×10^{-3}, respectively), consistent with prior studies (**Supplemental Table 17**)[51-54]. A total of 68 proteins were also associated with composite CHIP, *DNMT3A*, *ASXL1*, or *TET2* and CAD at the nominal significance threshold. These shared proteins have diverse functions, primarily enriched in inflammation and immune response pathways. Specifically, a number of the shared proteins were implicated in the regulation of IL-1β/NLRP3/IL-6 pathways, including proteins belonging to TNFRSF (such as TNFRSF1B and TNFRSF10D), members or receptors of the IL-1 cytokine family (such as IL-36α, IL-1Ra, IL-1RL1, and IL-1R2), and other chemokines (such as CXCL13 and CXCL10). Notably, *Cxcl13* and several genes encoding IL-1-related proteins have been shown to have increased expressed in *Tet2*$^{-/-}$ peritoneal macrophages exposed to lipopolysaccharide and interferon-γ in vitro[32]. Other important inflammatory proteins are implicated, such as protein S100-A9, myeloperoxidase[55,56], as well as proteins validated in *Tet2*$^{-/-}$ Ldlr$^{-/-}$ mice that consumed an atherogenic diet, such as CXCL13 and lysozyme C. Consistent with the functions of proteins associated with CHIP, proteins related to both *TET2* and CAD are involved in the ECM, cell adhesion, and signaling, such as metalloproteinase inhibitor 3 (TIMP-3)[57]. Those associated with both *ASXL1* and CAD participate in enzyme and metabolism processes, such as proprotein convertase subtilisin/kexin type 9 (PCSK9). Proteins associated with both *DNMT3A* and CAD are involved with diverse functions, with several proteins involved in the signaling and adhesion of neural cells, such as contactin-1[58-60]. Furthermore, 24 proteins were found to be associated with *JAK2*, including key proteins involved in hematopoietic traits, such as erythropoietin and ferritin (**Figure 7** and **Supplemental Table 18**).

Discussion

The various common mutations that drive CHIP yield differential associations with clinical outcomes[3,10,13], and intermediate molecular phenotypes may facilitate a better understanding of these distinctions. Leveraging large-scale proteomics data from multiple diverse cohorts across two major platforms and murine scRNA-seq data, our study identified and validated distinct proteomic signatures for common CHIP driver genes. Furthermore, our findings demonstrated the potential implications of these data to prioritize mechanisms in CHIP-associated coronary artery disease. More broadly, our study provides mechanistic insights offering novel CHIP driver gene-specific therapeutic strategies toward precision prevention of CHIP-associated clinical outcomes.

Despite convergence on clonal hematopoiesis[2-4], mutations by CHIP driver genes had varied proteomic associations. Our study demonstrated that the three common driver genes linked to unique proteins with diverse functions, which subsequently aggregated into distinct pathways, suggesting these genes may associate with diseases through separate cellular pathways. For instance, although all examined driver mutations related to some proteins involved in immune responses, with particular enrichment in regulating IL-1β/NLRP3/IL-6 pathways, in line with existing knowledge that CHIP arising from several mutations induces an altered inflammatory state, different genes act differently. *DNMT3A*, the most prevalent CHIP driver gene, exhibited a relatively indolent profile with few and weak associations, primarily implicating first-line defense pathways. In contrast, *TET2*, associated with proteins predominantly involved in innate immunity, inflammation, and extracellular matrix remodeling, and *ASXL1*, enriched in proteins related to cell signaling and function and metabolic regulation, appeared to modulate chronic inflammation pathways, such as STAT3 and IL-6 signaling pathways. *JAK2*, the least prevalent CHIP mutated gene examined, was associated with the

largest number of proteins exhibiting the strongest associations, enriched in hematopoietic traits. In particular, our study contributed evidence in humans of the anti-inflammatory role of *ASXL1* CHIP mutations, with pathway analysis showing predicted reduction across a large number of pro-inflammatory pathways, in direct contrast with other examined CHIP driver genes. This finding corresponds with the observed coexistence of pro- and anti-inflammatory characteristics in both zebrafish and murine macrophages with *Asxl1* mutations, and similarly in humans with *ASXL1* CHIP mutations[61,62]. Moreover, leveraging multi-ancestry data from TOPMed, our study highlighted sex and race differences in CHIP-proteomic associations. Several of these findings agree closely with existing observations at the phenotypic level in humans or mice, substantiating the validity of our results and offering potential molecular mechanisms to explain these differential phenotypic associations[13,63,64].

By combining paired DNA and protein data from large human populations with scRNA-seq data from atherogenic mice transplanted with *Tet2*$^{-/-}$ or with *Tet2*$^{+/+}$ bone marrow, our study examined whether results from the two species align. We demonstrated significant differential expression in mice across all well-expressed (percent expressed>30%) genes that encode *TET2*-associated proteins in humans, with the direction of effects also being consistent between the two species. For example, CXCL13 orchestrates B-cell trafficking to the site of inflammation, and its increased levels may participate in the development of several immune-related disorders[65-67]. Properdin is involved in innate immunity and is the only known positive regulator of the alternative pathway of the complement system[68]. Both proteins were positively associated with *TET2 i*n humans, and the expressions of genes encoding these proteins were markedly increased in the monocytes of *Tet2*$^{-/-}$ mice. Similarly, FCER2 is another immune-related protein that plays several critical roles in IgE-mediated immunity, including negatively regulating IgE and IgG

antibody responses in B cells[69,70], etc. It was negatively associated with *TET2* in humans, which was concordant with a significantly decreased expression of *Fcer2a* in B cells of *Tet2*$^{-/-}$ mice. In contrast to that strong alignment, only a few proteins that are encoded by the strongest differentially expressed genes in mice were significantly associated with *TET2* in humans. Inherent differences in animal experiments with humans, including rapid clonal expansion not typically observed in humans[71], may contribute to incomplete overlap.

By examining shared proteomic associations between CHIP and CAD, in addition to proteins implicated in regulating the established CHIP-related IL-1β/NLRP3/IL-6 pathways, we implicate new potential associations. For instance, both *TET2* and CAD were linked to ECM-related proteins, such as TIMP3, an ECM-bound protein inhibiting a broad range of substrates, including matrix metalloproteinase[72,73]. It is a potent angiogenesis inhibitor[74] and has been shown to play crucial roles in cardiac remodeling and cardiomyopathy[75]. Recent studies also suggest its therapeutic potential for heart failure[76]. Similarly, *ASXL1* and CAD were associated with metabolism-related proteins like PCSK9, a key regulator of lipid metabolism. Inhibitors targeting PCSK9 lower LDL levels and reduce the risk of cardiovascular diseases[77]. Furthermore, *DNMT3A* and CAD shared associations with signaling and adhesion proteins in neural cells, such as contactin-1. This cell adhesion molecule plays a critical role in various aspects of neural development and function[58-60], and has also been identified as a cardiac biomarker by several studies[78,79].

We used cross-platform validation to address the previously reported considerable variation in the quantitative agreement between the aptamer-based (SomaScan) and proximity-extension immunoassay (Olink) platforms[50,80]. We used SomaScan-measured proteins for discovery in our study because SomaScan demonstrated greater measurement precision with

small coefficients of variation (CV)) for enabling the integration of results from multiple cohorts[80], greater analytic breadth across the proteome for facilitating novel discoveries, and availability in multi-ancestry cohorts for identifying any potential group-specific effects. For example, Walker et al. identified significant associational and causal links between Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1 (SVEP1) and incident dementia and demonstrated further that its plasma protein levels relate strongly to neuroimaging-defined dementia endophenotypes[81]. This novel discovery was feasible by using SomaScan-measured proteins, as Olink uses targeted panels, each of which consists of proteins that are known to be associated with the preselected disease of that panel. However, Olink-measured proteins are implicated in more phenotypic associations[80], suggesting Olink-measured proteins may yield more revealing associations with known proteins.

Our study has limitations. Firstly, both molecular and environmental confounders might affect the associations between CHIP and the plasma proteome. While we adjusted for recognized confounders, applied PEER factors to mitigate technical noise, and performed sensitivity analyses accounting for eGFR and blood cells, the potential for residual and unmeasured confounding remains. Secondly, our study primarily utilized linear models to explore associations between CHIP driver genes and proteins. However, there may be interactions and non-linear effects at play. To address this, we assessed interactions between CHIP variables and both sex and race for proteins demonstrating differential associations in our stratified analysis. Thirdly, the cross-sectional design of our study constrains our ability to establish temporality and causality. Nevertheless, our experimental murine data provides some ability to infer associations resulting from *TET2* mutations. Lastly, our mice with *Tet2*$^{-/-}$ bone marrow exhibited increased atherosclerotic burden compared to wild-type mice in our

experiments, meaning we cannot exclude the possibility that differential expression is due to atherosclerosis rather than solely to *Tet2*$^{-/-}$ mutation. However, we observed decrease expression in mice with *Tet2*$^{-/-}$ bone marrow for genes whose encoded proteins are negatively associated with *TET2* in human. Also, the number of B cells and T cells, the cell types where we observed many differential expression that were consistent with human findings, did not differ much between the two mice groups.

These results, taken togther, provide new mechanistic insights into how different CHIP mutations may link to cardiovascular outcomes and, in particular, to the increase in atherothrombotic events associated with most CHIP mutations. These novel findings may inform the development and testing of interventions to mitigate the excess cardiovascular risk observed in carriers of various CHIP mutations.

Materials and Methods

Human study participants

TOPMed is a research program generating genomic data from DNA isolated from blood and other omics data for more than 80 NHLBI-funded research studies with extensive phenotype data[22]. Our current study includes five community-based cohorts: JHS, MESA, CHS, FHS, ARIC, and UKB. The secondary use of data for this analysis was approved by the Massachusetts General Hospital Institutional Review Board (protocol 2016P001308 and protocol 2021P002228) and, for the UKB data, facilitated through UKB Application 7089.

JHS is a longitudinal cohort study of 5,306 self-identified Black men and women recruited in 2000-04 from Jackson, Mississippi[23]. Our study included 2,058 individuals with whole-genome sequencing (WGS) and plasma proteomic profiling data[22]. MESA is a multi-

ancestry prospective cohort of 6,814 self-identified White, Black, Hispanic, or Asian (predominately of Chinese descent) men and women recruited in 2000-02[24]. We included 976 participants randomly selected for WGS and plasma proteomic profiling analysis[82]. CHS is a bi-ancestry (White and Black) longitudinal study of 5,888 men and women 65 years or older at recruitment (1989-90 or 1992-93)[25]. We analyzed data from 1,689 participants with WGS and plasma proteomics measurements who consented to genetics study[26]. ARIC is an ongoing longitudinal cohort of 15,792 middle-aged, mostly black and white participants recruited in 1987-89[83]. We included 8,188 participants with valid whole exome sequencing (WES) data and plasma proteomics measurements at Visit 2 or 3 in our study. These four studies served as discovery cohorts, with data from a total of 12,911 participants being used for the current analysis. UKB is a comprehensive, prospective cohort study consisting of approximately 500,000 men and women, predominantly of EA, who were aged between 40 and 69 years at the time of recruitment from 2006 to 2010 across the UK. Our analysis included a subset of 48,922 participants for whom both WES and plasma proteomic profiling data were available[28,84].

The mean age of the participants at the time of DNA sample collection was 57.4 years (55.8 years for JHS, 60.2 years for MESA, 73.3 years for CHS, 57.8 years for ARIC, and 56.8 years for UKB). Except for JHS, which includes only Black participants, all cohorts are bi-ancestry or multi-ancestry. All cohorts included participants of both sexes, with the percentage of men ranging from 39% to 47%.

CHIP calling

WGS and CHIP calling in JHS, MESA, and CHS were previously performed and have been described in detail elsewhere[13]. The same procedure was applied for WES data in ARIC and UK

Biobank[13,19]. In brief, whole blood-derived DNA was sequenced at an average depth of 38× using Illumina HiSeq X Ten instruments. All sequences in CRAM files were remapped to the hs38DH 1000 Genomes build 38 human genome references, following the protocol published previously[85]. Single nucleotide polymorphisms (SNPs) and short indels were jointly discovered and genotyped across the TOPMed samples using the GotCloud pipeline [86]. CHIP mutations were identified using Mutect2 software if one or more of a prespecified list of pathogenic somatic variants in 74 genes that drive clonal hematopoiesis and myeloid malignancies were present[13,87]. A Panel of Normals (PON) minimized sequencing artifacts and Genome Aggregation Database (gnomAD) filtered likely germline variants from the putative somatic mutations call set[88]. Each Variant Call Format (VCF) file was annotated using ANNOVAR software[89]. Variants were retained for further curation if they met the following criteria: total depth of coverage ≥ 10, number of reads supporting the alternate allele ≥ 3, ≥ 1 read in both forward and reverse direction supporting the alternate allele, and variant allele fraction (VAF) ≥ 0.02. In particular, variants with VAF ≥ 0.1 were defined as expanded CHIP clones.

Human proteomic measurements

The relative concentrations of plasma proteins or protein complexes from the blood samples of JHS, MESA, CHS, and ARIC were measured by the SomaScan (SomaLogic; Boulder, CO) using an aptamer (SOMAmer)-based approach, while proteomics of WHI was measured by Olink (Olink Proteomics; Uppsala, Sweden) using a proximity-extension immunoassay-based method. Detailed information on these technologies can be found in the corresponding manufacturer's protocols[30,90]. JHS and MESA utilized a 1.3K human proteins platform, while CHS and ARIC used a 5K human protein platform. We focused on 1,148 proteins shared by both

SomaScan platforms for discovery. Protein measurements were reported as relative fluorescence units (RFUs)[21]. There were no missing values in the proteomic data, and details of the quality control of the proteins were described elsewhere[91]. Proteomics measured by Olink in WHI included 552 proteins from cardiovascular, inflammation, cardiometabolic, neurology, and oncology panels. Proteins with ≥ 10% missingness were excluded. In this study, all proteins underwent log2 transformation.

Human coronary artery disease ascertainment

The five TOPMed cohorts have conducted active surveillance for coronary artery disease (CAD) events through annual follow-up by phone calls, surveys, and/or interviews, and abstracting medical records, hospitalization records, and death certificates[92-96]. In JHS, CAD was defined as myocardial infarction (MI), death due to CAD, or cardiac procedures, including percutaneous transluminal coronary angioplasty, stent placement, coronary artery bypass grafting, or other coronary revascularization[92]. In MESA, CAD included MI, death due to CAD, resuscitated cardiac arrest, and revascularization[93]. In CHS, CAD included MI, death due to CAD, angina pectoris, and cardiac procedures, including angioplasty and coronary artery bypass graft[97]. In FHS, CAD included MI, death due to CAD, angina pectoris, and coronary insufficiency[98]. In ARIC, CAD included MI observed on ECG, self-reported doctor-diagnosed heart attack, or self-reported cardiovascular surgery or coronary angioplasty, as well as study-adjudicated CAD cases between visit 1 and visit 2 or 3[99]. In our study, we defined prevalent CAD cases as those occurring before the blood sample collection visit, where CHIP measurements were also taken. By aligning the time points for prevalent CAD and CHIP assessments, we ensured a fair comparison between them in later analyses.

Regression model and meta-analysis for human analysis

We examined the cross-sectional associations between CHIP mutations and 1,148 plasma proteins within four discovery cohorts, which were then meta-analyzed. CHIP was modeled both as a composite and individually for the most common or pathogenic drivers (*DNMT3A*, *TET2*, *ASXL1*, and *JAK2*), using conventional thresholds for all mutations (VAF ≥2%) and expanded forms (VAF ≥10%), resulting in a total of ten CHIP mutatioins. Consistent with a recent analysis[31], ARIC was separated into two subcohorts: ARIC EA and ARIC AA. To reduce redundant information and multiple testing burden, we collapsed the composite CHIP and each driver gene with their corresponding expanded forms, retaining the one with stronger associations (larger absolute Z score). Within each cohort or subcohort (ARIC), linear regression models were fitted with CHIP mutations as exposures, log-transformed proteins as outcomes, and various covariates, including age at sequencing, sex, race, batch, center, diagnoses of type 2 diabetes mellitus at the time of enrollment, ever-smoker status, first ten PCs of genetic ancestry, and PEER factors as covariates. PEER factors were adjusted to account for hidden confounders, such as batches, that may influence clusters of proteins[100]. The number of PEER factors varied by cohorts based on study population size: 50 for JHS, MESA, and CHS; 70 for ARIC AA; 120 for ARIC EA[31]. *JAK2* analyses were only conducted in cohorts with more than 5 participants with JAK2 mutations (i.e., CHS and ARIC EA only) to optimize power and generate more reliable effect estimates; given the required subsetting, *JAK2* analyses were considered secondary analyses. Next, linear regression results from each discovery cohort were meta-analyzed using inverse-variance weighted fixed-effect meta-analysis. We conducted stratified analysis by sex (female and male) and by race (Black and White). For CHIP variable-protein

pairs that are only significant in either male or female stratified analysis, we introduced and tested interaction terms between the corresponding CHIP variables and sex across all discovery cohorts. Likewise, for pairs that are only significant in Black or White stratified analysis, we tested for interaction terms between the CHIP variables and race in the combined analysis of ARIC, as ARIC has good number of and relatively balanced White and Black participants. We also conducted secondary analyses with removing PEER factors from covariates adjustment as previous reports indicated its relevance to the power of discovery [31,101], as well as secondary analyses additionally adjusting for eGFR or blood cells (platelet and WBC) in ARIC[102]. Linear regression models were performed using R function 'glm' while fixed-effects meta-analysis and tests of heterogeneity were conducted using the R package 'meta'. We controlled the FDR using the Benjamini-Hochberg procedure and set an FDR threshold of 0.05 for significance.

Pathway analysis for human proteomic association results

We conducted pathway analyses of proteins associated with each CHIP driver gene at a P=0.05 threshold. We applied a nominal threshold (P=0.05) for selecting proteomic associations for pathway analysis, ensuring a comprehensive view of pathway enrichment by including a wide range of associated proteins. We input the sets of Z-scores of 102 *DNMT3A*-associated proteins, 123 *TET2*-associated proteins, and 140 *ASXL1*-associated proteins, which were organized into canonical pathways by the Ingenuity Pathway Analysis (IPA) tool. IPA pathways were constructed within the Ingenuity Knowledge Base, a large structured collection of findings containing nearly 5 million entries manually curated from the biomedical literature or integrated from third-party databases[103]. The network comprises ~40 000 nodes connected by ~1,480,000 edges representing experimentally observed cause-effect relationships related to expression,

transcription, activation, molecular modification, transportation, and binding events. IPA utilizes a right-tailed Fisher's exact test to evaluate the enrichment of CHIP driver genes-associated proteins in each pathway, as well as to infer their potential cause-effect relationships.

Experimental mice and induction of atherosclerosis

Our experiments utilized eight female and male C57Bl/6J *Ldlr*$^{-/-}$ and C57Bl/6J *Tet2*$^{-/-}$ mice strains. Mice were assigned randomly to experimental groups, and results were analyzed in a blinded fashion. To generate animals with *Tet2-/-* hematopoietic cells, we crossed mice with constitutive expression of Cre recombinase under control of Vav1 promoter (B6.Cg-Tg(Vav1-cre)A2Kio/J [Jax Cat. No. 008610]) with the *Tet2*-floxed line (B6;129S-Tet2tm1.1Iaai/J [Jax Cat. No. 017573]). Wild-type Vav1-cre were used as control. Bone marrow cells were isolated from femurs and tibias of *Tet2*$^{+/+}$ or *Tet2*$^{-/-}$ CD45.2 mice after euthanasia. Eight weeks old recipients *Ldlr*$^{-/-}$ CD45.1 mice were lethally irradiated with two doses of 460cGy gamma irradiation, separated by 4 hours. Two million bone marrow cells were transplanted into recipient mice, four with *Tet2*$^{+/+}$ and four with *Tet2*$^{-/-}$ cells, through retro-orbital injection. The mice were then housed in sterilized cages and provided with antibiotic-containing food and water (sulfatrim). Four weeks post-bone marrow transplantation, blood was collected via retro-orbital bleeding to confirm chimerism. Chimeric *Ldlr*$^{-/-}$ mice were fed a Western diet containing 0.2% cholesterol (Harlan Teklad, TD.88137) for nine weeks. Following CO_2 euthanasia, mice were perfused through left ventricular cannulation with heparinized PBS before tissue harvesting. Genotyping was performed by Transnetyx (Cordova, TN, USA). All experiments were approved by the Institutional Animal Care and Use Committees and were conducted in accordance with the guidelines of the American Association for Accreditation of Laboratory Animal Care and the National Institutes of Health.

Mouse aortic cell dissociation and single-cell RNA sequencing

Cells from 20 aortic arches were dissociated using an enzyme cocktail containing 0.5mg/ml elastase, 450U/ml collagenase I, 125U/ml collagenase XI, 60U/ml DNase I, and 20µM HEPES for 45 minutes at 37°C and 750rpm as previously described[49]. Two to three arches were pooled per sample, and cell suspensions were passed through a 40µm cell strainer. After red blood cell lysis (ACK buffer), cell suspensions were washed twice with PBS. Dissociated cells were resuspended in PBS containing 0.4% BSA. Approximately ten thousand cells per sample were partitioned into gel beads in emulsion using the 10X Chromium controller, and subsequent libraries were generated using the 3' gene expression v3 kit according to the manufacturer's instructions (10X Genomics Inc.). Libraries were sequenced by Novogene using an Illumina NovaSeq S4 flow cell, following the manufacturer's protocol (Illumina; San Diego, CA).

Mouse single-cell RNA data analysis

The Cell Ranger v3.1 count module (10X Genomics; Pleasanton, CA) was used to generate a raw gene expression count matrix from Fastq files. To eliminate ambient RNA, Cellbender v0.2.0 was applied with default parameters to the raw, unfiltered counts. The resulting filtered counts were then analyzed using Seurat v4.0.1. Cells with less than 300 UMIs and more than 25% mitochondrial RNA were excluded, and potential doublet cells were predicted and removed using DoubletFinder v2.0 before sample merging. Each sample was normalized with SCTransform, and shared features across samples were identified with the SelectIntegrationFeatures module to use as variable features for calculating PCA in the merged

Seurat object. Batch correction for the PCA embedding was performed using Harmony v0.1 with the sample as a key. The resulting batch-corrected object was used for downstream analysis, including cluster annotation with a previously described method[49]. Differential expression between *Tet2*$^{-/-}$ and wild-type control mice was performed using edgeR. Most analyses and figure generation in this step were conducted using the 'Seurat' package unless otherwise described[104].

Cross-platform replication for human proteomic associations

We assessed the compatibility of associations between CHIP mutations and proteomic data measured using two highly multiplexed technologies for large-scale proteomics measurements: aptamer-based (SomaScan 1.3K) and proximity-extension immunoassay (Olink 1.5K) platforms[21,30]. We examined the associations between CHIP mutations (composite CHIP, *DNMT3A*, *TET2*, and *ASXL1*) and proteomics measured using Olink among 48,922 participants in the UKB, employing the same model and covariates as that used in the main analysis associating CHIP mutations with SomaScan-measured proteomics. A total of 471 were overlapped between the two platforms as matched by UniProt IDs. We compared the findings of these proteins between UKB and meta-analysis results from the main discovery analyses based on SomaScan-measured proteins, with assessing shared CHIP variable-proteomic pairs between both groups.

Shared proteomic associations between CHIP and CAD in human

We investigated the shared proteomic associations between CHIP mutations and CAD using four discovery TOPMed cohorts (N=12,911), JHS, MESA, CHS, and ARIC, same study population for examining the associations between CHIP mutations and proteomics in the main analysis. We

first examined the cross-sectional associations between prevalent CAD, assessed at the same visits as blood draws to maintain temporal consistency with CHIP measurements, and proteomics. For the associations between CAD and proteomics, we employed the same linear models used for studying the associations between CHIP mutations and proteomics. Specifically, we again used processed proteomics as the outcome, replaced CHIP with CAD as the exposure, and adjusted for the same set of covariates excluding PEER factors (age at sequencing, sex, race, batch, center, type 2 diabetes mellitus diagnoses at enrollment, ever-smoker status, and the first ten PCs of genetic ancestry). Our decision not to adjust for PEER factors in the CAD and proteomics association analysis was based on a previous report indicating that PEER factors can capture proteomic variation related to disease mechanisms[81]. Additionally, empirical evidence from our own analysis showed that several PEER factors were associated with CAD, which could remove relevant signals. However, PEER factors were generally not associated with CHIP mutations, and with or without adjusting for PEER factors yielded consistent associations between CHIP mutations and proteomics. Subsequently, we identified the intersection of proteins associated with both CHIP mutations and CAD ascertained at the same visits at a P=0.05 level. We then categorized these proteins according to the type of CHIP mutations and investigated any differential enrichment by distinct types of CHIP mutations.

Use of artificial intelligence tools

We used Perplexity AI (https://www.perplexity.ai/) for searching protein functions, ChatGPT4 (https://chat.openai.com/) for refining code through the debugging process, facilitating protein function categorization, and enhancing the linguistic clarity of complex sentences, and Grammarly plugin for grammar error correction.

Data availability

TOPMed individual-level DNA and proteomics data used in this analysis are available through restricted access via the dbGaP. UK Biobank individual-level data are available for request by application (https://www.ukbiobank.ac.uk). All code used for the described analysis will be uploaded to GitHub once the manuscript is accepted for publication.

Acknowledgments

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for "NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)" (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1), and TOPMed MESA Multi-Omics (HHSN2682015000031/HSN26800004). The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, 75N92022D00005). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for "Building on GWAS for NHLBI-diseases: the U.S.

CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Metabolomics measurements were sponsored by the National Human Genome Research Institute (3U01HG004402-02S1). SomaLogic Inc. conducted the SomaScan assays in exchange for use of ARIC data. This work was supported in part by NIH/NHLBI grant R01 HL134320. Cardiovascular Health Study: This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, 75N92021D00006, and grants U01HL080295, HL105756, and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS. The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for the Multi-Ethnic Study of Atherosclerosis (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute

(NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutes can be found at http://www.mesa-nhlbi.org. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. We also thank Mrs. Leslie Gaffney from the Broad Research Communication Lab for her valuable assistance in improving the display items.

Funding

A.G.B. is supported by a Burroughs Wellcome Foundation Career Award for Medical Scientists and the National Institute of Health (NIH) Director's Early Independence Award (DP5-OD029586). A.R. is supported by NIH R01 HL148565. B.L.E. is supported by Leducq Foundation. G.G. is supported by NIH grants R01 MH104964 and R01 MH123451, and Stanley Center for Psychiatric Research. P.G. is supported by NIH grants 1R01 HL159081, R01 HL153499, and 5U01DK108809. P.L. receives funding support from the National Heart, Lung, and Blood Institute (1R01HL134892 and 1R01HL163099-01), the RRM Charitable Fund and the Simard Fund. P.N. is supported by grants from the NHLBI (R01HL142711, R01HL127564,

R01HL148050, R01HL151283, R01HL148565, R01HL135242, and R01HL151152), National Institute of Diabetes and Digestive and Kidney Diseases (R01DK125782), Fondation Leducq (TNE-18CVD04), and Massachusetts General Hospital (Paul and Phyllis Fireman Endowed Chair in Vascular Medicine). S.J. is supported by the Burroughs Wellcome Fund Career Award for Medical Scientists, Fondation Leducq (TNE-18CVD04), the Ludwig Center for Cancer Stem Cell Research at Stanford University, and the National Institutes of Health (DP2-HL157540). Z.Y. is supported by NHLBI (5T32HL007604-37).

Disclosures:

A.E.L. is currently a member of TenSixteen Bio, outside of the submitted work. B.L.E. has received research funding from Celgene, Deerfield, Novartis, and Calico and consulting fees from GRAIL. He is a member of the scientific advisory board and shareholder for Neomorph Inc., TenSixteen Bio, Skyhawk Therapeutics, and Exo Therapeutics. B.M.P serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. J.C. is a scientific advisor to SomaLogic. P.L. is an unpaid consultant to, or involved in clinical trials for Amgen, AstraZeneca, Baim Institute, Beren Therapeutics, Esperion Therapeutics, Genentech, Kancera, Kowa Pharmaceuticals, Medimmune, Merck, Moderna, Novo Nordisk, Novartis, Pfizer, and Sanofi-Regeneron. P.L. is a member of the scientific advisory board for Amgen, Caristo Diagnostics, Cartesian Therapeutics, CSL Behring, DalCor Pharmaceuticals, Dewpoint Therapeutics, Eulicid Bioimaging, Kancera, Kowa Pharmaceuticals, Olatec Therapeutics, Medimmune, Novartis, PlaqueTec, TenSixteen Bio, Soley Thereapeutics, and XBiotech, Inc. P.L.'s laboratory has received research funding in the last 2 years from Novartis, Novo Nordisk and Genentech. P.L. is on the Board of Directors of XBiotech, Inc. P.L. has a financial interest in

Xbiotech, a company developing therapeutic human antibodies, in TenSixteen Bio, a company targeting somatic mosaicism and clonal hematopoiesis of indeterminate potential (CHIP) to discover and develop novel therapeutics to treat age-related diseases, and in Soley Therapeutics, a biotechnology company that is combining artificial intelligence with molecular and cellular response detection for discovering and developing new drugs, currently focusing on cancer therapeutics. P.L.'s interests were reviewed and are managed by Brigham and Women's Hospital and Mass General Brigham in accordance with their conflict-of-interest policies. P.N. reports investigator-initiated grants from Amgen, Apple, Boston Scientific, Novartis, and AstraZeneca, personal fees from Allelica, Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs, Genentech, and Novartis, scientific board membership for Esperion Therapeutics, geneXwell, and TenSixteen Bio, and spousal employment at Vertex, all unrelated to the present work. P.N., A.G.B., S.J., and B.L.E. are scientific co-founders of TenSixteen Bio, and P.L. is an advisor to TenSixteen Bio. TenSixteen Bio is a company focused on clonal hematopoiesis but had no role in the present work. S.J. is on advisory boards for Novartis, AVRO Bio, and Roche Genentech, reports speaking fees and a honorarium from GSK, and is on the scientific advisory board of Bitterroot Bio. The other authors report no conflicts.

References:

1. Steensma, D.P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9-16 (2015).
2. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* **20**, 1472-8 (2014).
3. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477-87 (2014).
4. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488-98 (2014).
5. Kim, P.G. *et al.* Dnmt3a-mutated clonal hematopoiesis promotes osteoporosis. *J Exp Med* **218**(2021).
6. Yu, B. *et al.* Supplemental Association of Clonal Hematopoiesis With Incident Heart Failure. *J Am Coll Cardiol* **78**, 42-52 (2021).
7. Miller, P.G. *et al.* Association of clonal hematopoiesis with chronic obstructive pulmonary disease. *Blood* **139**, 357-368 (2022).
8. Bhattacharya, R. *et al.* Clonal Hematopoiesis Is Associated With Higher Risk of Stroke. *Stroke* **53**, 788-797 (2022).
9. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* **377**, 111-121 (2017).
10. Jaiswal, S. Clonal hematopoiesis and nonhematologic disorders. *Blood* **136**, 1606-1614 (2020).
11. Wong, W.J. *et al.* Clonal haematopoiesis and risk of chronic liver disease. *Nature* (2023).

12. Natarajan, P., Jaiswal, S. & Kathiresan, S. Clonal Hematopoiesis: Somatic Mutations in Blood Cells and Atherosclerosis. *Circ Genom Precis Med* **11**, e001926 (2018).

13. Bick, A.G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763-768 (2020).

14. Kar, S.P. *et al.* Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. *Nat Genet* **54**, 1155-1166 (2022).

15. Uddin, M.M. *et al.* Germline genomic and phenomic landscape of clonal hematopoiesis in 323,112 individuals. *medRxiv*, 2022.07.29.22278015 (2022).

16. Abplanalp, W.T. *et al.* Association of Clonal Hematopoiesis of Indeterminate Potential With Inflammatory Gene Expression in Patients With Severe Degenerative Aortic Valve Stenosis or Chronic Postischemic Heart Failure. *JAMA Cardiol* **5**, 1170-1175 (2020).

17. Abplanalp, W.T. *et al.* Clonal Hematopoiesis-Driver DNMT3A Mutations Alter Immune Cells in Heart Failure. *Circ Res* **128**, 216-228 (2021).

18. Nachun, D. *et al.* Clonal hematopoiesis associated with epigenetic aging and clinical outcomes. *Aging Cell* **20**, e13366 (2021).

19. Uddin, M.D.M. *et al.* Clonal hematopoiesis of indeterminate potential, DNA methylation, and risk for coronary artery disease. *Nat Commun* **13**, 5350 (2022).

20. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).

21. Rohloff, J.C. *et al.* Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol Ther Nucleic Acids* **3**, e201 (2014).

22. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).

23. Taylor, H.A., Jr. *et al.* Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis* **15**, S6-4-17 (2005).

24. Bild, D.E. *et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol* **156**, 871-81 (2002).

25. Fried, L.P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).

26. Austin, T.R. *et al.* Proteomics and Population Biology in the Cardiovascular Health Study (CHS): design of a study with mentored access and active data sharing. *Eur J Epidemiol* **37**, 755-765 (2022).

27. Wright, J.D. *et al.* The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol* **77**, 2939-2959 (2021).

28. Sun, B.B. *et al.* Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. *bioRxiv*, 2022.06.17.496443 (2022).

29. Kessler, M.D. *et al.* Common and rare variant associations with clonal haematopoiesis phenotypes. *Nature* **612**, 301-309 (2022).

30. Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* **9**, e95192 (2014).

31. Zhang, J. *et al.* Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet* **54**, 593-602 (2022).

32. Fuster, J.J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842-847 (2017).

33. Lin, T.M., Galbert, S.P., Kiefer, D., Spellacy, W.N. & Gall, S. Characterization of four human pregnancy-associated plasma proteins. *Am J Obstet Gynecol* **118**, 223-36 (1974).

34. Wald, N. *et al.* First trimester concentrations of pregnancy associated plasma protein A and placental protein 14 in Down's syndrome. *Bmj* **305**, 28 (1992).

35. Kirkegaard, I., Uldbjerg, N. & Oxvig, C. Biology of pregnancy-associated plasma protein-A in relation to prenatal diagnostics: an overview. *Acta Obstet Gynecol Scand* **89**, 1118-25 (2010).

36. Smith, G.C. *et al.* Early-pregnancy origins of low birth weight. *Nature* **417**, 916 (2002).

37. Brekken, R.A. & Sage, E.H. SPARC, a matricellular protein: at the crossroads of cell-matrix communication. *Matrix Biol* **19**, 816-27 (2001).

38. Bradshaw, A.D. & Sage, E.H. SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury. *J Clin Invest* **107**, 1049-54 (2001).

39. Tai, I.T. & Tang, M.J. SPARC in cancer biology: its role in cancer progression and potential for therapy. *Drug Resist Updat* **11**, 231-46 (2008).

40. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol Ther* **74**, 1-20 (1997).

41. Steppan, C.M. *et al.* The hormone resistin links obesity to diabetes. *Nature* **409**, 307-12 (2001).

42. Jackson, D.G. Hyaluronan in the lymphatics: The key role of the hyaluronan receptor LYVE-1 in leucocyte trafficking. *Matrix Biol* **78-79**, 219-235 (2019).

43. Lee, J.Y. & Spicer, A.P. Hyaluronan: a multifunctional, megaDalton, stealth molecule. *Curr Opin Cell Biol* **12**, 581-6 (2000).

44. Quach, M.E. & Li, R. Structure-function of platelet glycoprotein Ib-IX. *J Thromb Haemost* **18**, 3131-3141 (2020).

45. Furie, B. & Furie, B.C. Role of platelet P-selectin and microparticle PSGL-1 in thrombus formation. *Trends Mol Med* **10**, 171-8 (2004).

46. Hobbs, C.M. *et al.* JAK2V617F leads to intrinsic changes in platelet formation and reactivity in a knock-in mouse model of essential thrombocythemia. *Blood* **122**, 3787-97 (2013).

47. Chin-Yee, B. *et al.* Serum erythropoietin levels in 696 patients investigated for erythrocytosis with JAK2 mutation analysis. *Am J Hematol* **97**, E150-e153 (2022).

48. Fajgenbaum, D.C. & June, C.H. Cytokine Storm. *N Engl J Med* **383**, 2255-2273 (2020).

49. Kalluri, A.S. *et al.* Single-Cell Analysis of the Normal Mouse Aorta Reveals Functionally Distinct Endothelial Cell Populations. *Circulation* **140**, 147-163 (2019).

50. Rooney, M.R. *et al.* Comparison of Proteomic Measurements Across Platforms in the Atherosclerosis Risk in Communities (ARIC) Study. *Clin Chem* **69**, 68-79 (2023).

51. Thygesen, K. *et al.* Fourth Universal Definition of Myocardial Infarction (2018). *Circulation* **138**, e618-e651 (2018).

52. Roberts, W.L. CDC/AHA Workshop on Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: laboratory tests available to assess inflammation--performance and standardization: a background paper. *Circulation* **110**, e572-6 (2004).

53. Pearson, T.A. *et al.* Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* **107**, 499-511 (2003).

54. Atherton, J.J. *et al.* National Heart Foundation of Australia and Cardiac Society of Australia and New Zealand: Guidelines for the Prevention, Detection, and Management of Heart Failure in Australia 2018. *Heart Lung Circ* **27**, 1123-1208 (2018).

55. Arnhold, J. The Dual Role of Myeloperoxidase in Immune Response. *Int J Mol Sci* **21**(2020).

56. Averill, M.M., Kerkhoff, C. & Bornfeldt, K.E. S100A8 and S100A9 in cardiovascular biology and disease. *Arterioscler Thromb Vasc Biol* **32**, 223-9 (2012).

57. Cabral-Pacheco, G.A. *et al.* The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases. *Int J Mol Sci* **21**(2020).

58. Bizzoca, A. *et al.* Transgenic mice expressing F3/contactin from the TAG-1 promoter exhibit developmentally regulated changes in the differentiation of cerebellar neurons. *Development* **130**, 29-43 (2003).

59. Bizzoca, A. *et al.* F3/Contactin acts as a modulator of neurogenesis during cerebral cortex development. *Dev Biol* **365**, 133-51 (2012).

60. Hu, Q.D. *et al.* F3/contactin acts as a functional ligand for Notch during oligodendrocyte maturation. *Cell* **115**, 163-75 (2003).

61. Avagyan, S. *et al.* Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. *Science* **374**, 768-772 (2021).

62. Yu, Z. *et al.* Genetic modification of inflammation and clonal hematopoiesis-associated cardiovascular risk. *medRxiv*, 2022.12.08.22283237 (2023).

63. Fujino, T. & Kitamura, T. ASXL1 mutation in clonal hematopoiesis. *Exp Hematol* **83**, 74-84 (2020).

64. Montagner, S. *et al.* TET2 Regulates Mast Cell Differentiation and Proliferation through Catalytic and Non-catalytic Activities. *Cell Rep* **15**, 1566-1579 (2016).

65. Müller, G., Höpken, U.E. & Lipp, M. The impact of CCR7 and CXCR5 on lymphoid organ development and systemic immunity. *Immunol Rev* **195**, 117-35 (2003).

66. Meraouna, A. *et al.* The chemokine CXCL13 is a key molecule in autoimmune myasthenia gravis. *Blood* **108**, 432-40 (2006).

67. Lee, H.T. *et al.* Serum BLC/CXCL13 concentrations and renal expression of CXCL13/CXCR5 in patients with systemic lupus erythematosus and lupus nephritis. *J Rheumatol* **37**, 45-52 (2010).

68. Fearon, D.T. & Austen, K.F. Properdin: binding to C3b and stabilization of the C3b-dependent C3 convertase. *J Exp Med* **142**, 856-63 (1975).

69. Campbell, K.A. *et al.* Induction of B cell apoptosis by co-cross-linking CD23 and sIg involves aberrant regulation of c-myc and is inhibited by bcl-2. *Int Immunol* **9**, 1131-40 (1997).

70. Payet, M.E., Woodward, E.C. & Conrad, D.H. Humoral response suppression observed with CD23 transgenics. *J Immunol* **163**, 217-23 (1999).

71. Uddin, M.M. *et al.* Longitudinal profiling of clonal hematopoiesis provides insight into clonal dynamics. *Immun Ageing* **19**, 23 (2022).

72. Galis, Z.S., Sukhova, G.K., Lark, M.W. & Libby, P. Increased expression of matrix metalloproteinases and matrix degrading activity in vulnerable regions of human atherosclerotic plaques. *J Clin Invest* **94**, 2493-503 (1994).

73. Fan, D. & Kassiri, Z. Biology of Tissue Inhibitor of Metalloproteinase 3 (TIMP3), and Its Therapeutic Implications in Cardiovascular Pathology. *Front Physiol* **11**, 661 (2020).

74. Qi, J.H. *et al.* A novel function for tissue inhibitor of metalloproteinases-3 (TIMP3): inhibition of angiogenesis by blockage of VEGF binding to VEGF receptor-2. *Nat Med* **9**, 407-15 (2003).

75. Moore, L., Fan, D., Basu, R., Kandalam, V. & Kassiri, Z. Tissue inhibitor of metalloproteinases (TIMPs) in heart failure. *Heart Fail Rev* **17**, 693-706 (2012).

76. Takawale, A. *et al.* Myocardial overexpression of TIMP3 after myocardial infarction exerts beneficial effects by promoting angiogenesis and suppressing early proteolysis. *Am J Physiol Heart Circ Physiol* **313**, H224-h236 (2017).

77. Sabatine, M.S. *et al.* Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med* **376**, 1713-1722 (2017).

78. Yin, X. *et al.* Protein biomarkers of new-onset cardiovascular disease: prospective study from the systems approach to biomarker research in cardiovascular disease initiative. *Arterioscler Thromb Vasc Biol* **34**, 939-45 (2014).

79. Lau, E.S. *et al.* Cardiovascular Biomarkers of Obesity and Overlap With Cardiometabolic Dysfunction. *J Am Heart Assoc* **10**, e020215 (2021).

80. Katz, D.H. *et al.* Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci Adv* **8**, eabm5164 (2022).

81. Walker, K.A. *et al.* Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. *Nature Aging* **1**, 473-489 (2021).

82. Liu, Y. *et al.* Blood monocyte transcriptome and epigenome analyses reveal loci associated with human atherosclerosis. *Nat Commun* **8**, 393 (2017).

83. ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol* **129**, 687-702 (1989).

84. Backman, J.D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628-634 (2021).

85. Regier, A.A. *et al.* Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun* **9**, 4038 (2018).

86. Jun, G., Wing, M.K., Abecasis, G.R. & Kang, H.M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res* **25**, 918-25 (2015).

87. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv*, 861054 (2019).

88. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).

89. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

90. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).

91. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).

92. Keku, E. *et al.* Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. *Ethn Dis* **15**, S6-62-70 (2005).

93. McClelland, R.L. *et al.* 10-Year Coronary Heart Disease Risk Prediction Using Coronary Artery Calcium and Traditional Risk Factors: Derivation in the MESA (Multi-Ethnic Study of Atherosclerosis) With Validation in the HNR (Heinz Nixdorf Recall) Study and the DHS (Dallas Heart Study). *J Am Coll Cardiol* **66**, 1643-53 (2015).

94. Newman, A.B. *et al.* Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort. *J Am Geriatr Soc* **53**, 1101-7 (2005).

95. Dawber, T.R., Meadors, G.F. & Moore, F.E., Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* **41**, 279-81 (1951).

96. Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J. & Castelli, W.P. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* **110**, 281-90 (1979).

97. Psaty, B.M. *et al.* Study of Cardiovascular Health Outcomes in the Era of Claims Data: The Cardiovascular Health Study. *Circulation* **133**, 156-64 (2016).

98. Wulsin, L.R. *et al.* Depressive symptoms, coronary heart disease, and overall mortality in the Framingham Heart Study. *Psychosom Med* **67**, 697-702 (2005).

99. White, A.D. *et al.* Community surveillance of coronary heart disease in the Atherosclerosis Risk in Communities (ARIC) Study: methods and initial two years' experience. *J Clin Epidemiol* **49**, 223-33 (1996).

100. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).

101. Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).

102. Inker, L.A. *et al.* New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med* **385**, 1737-1749 (2021).

103. Krämer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-30 (2014).

104. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

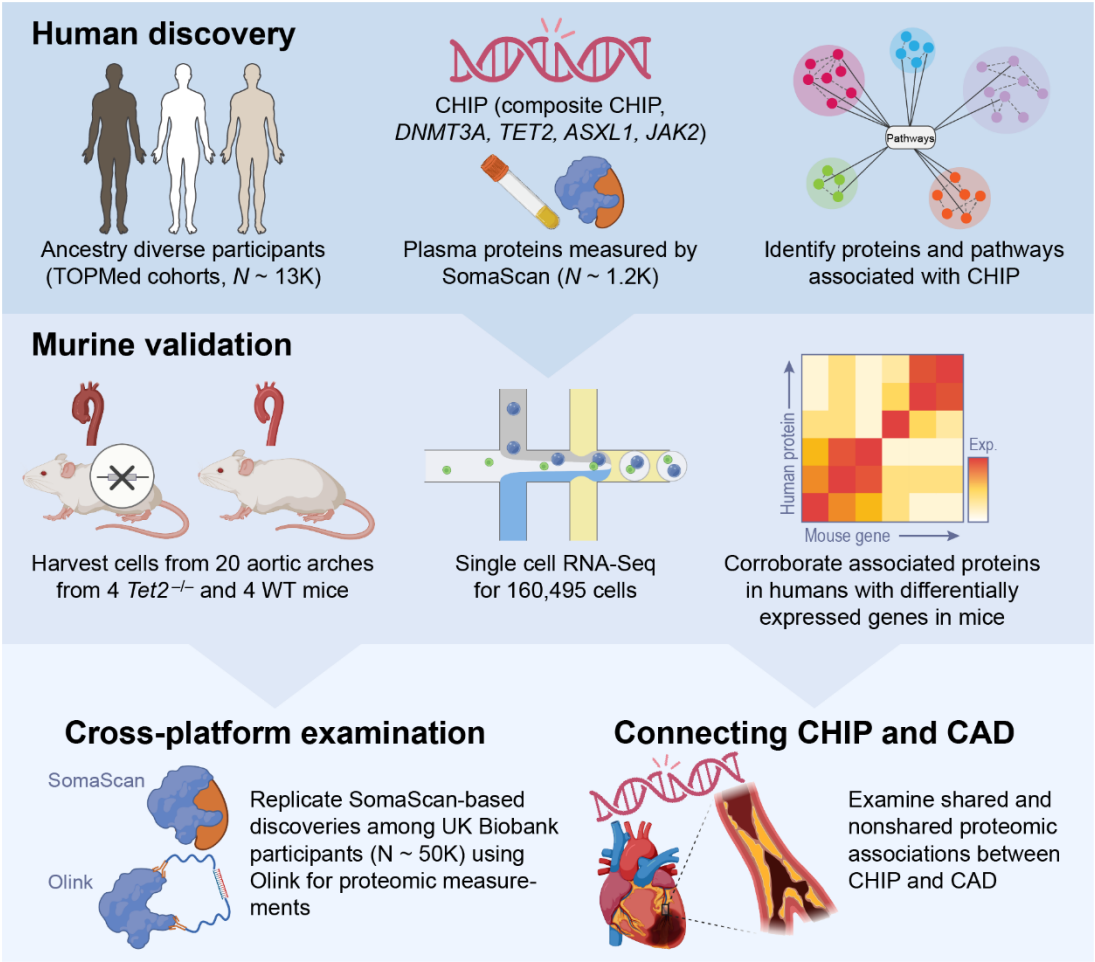
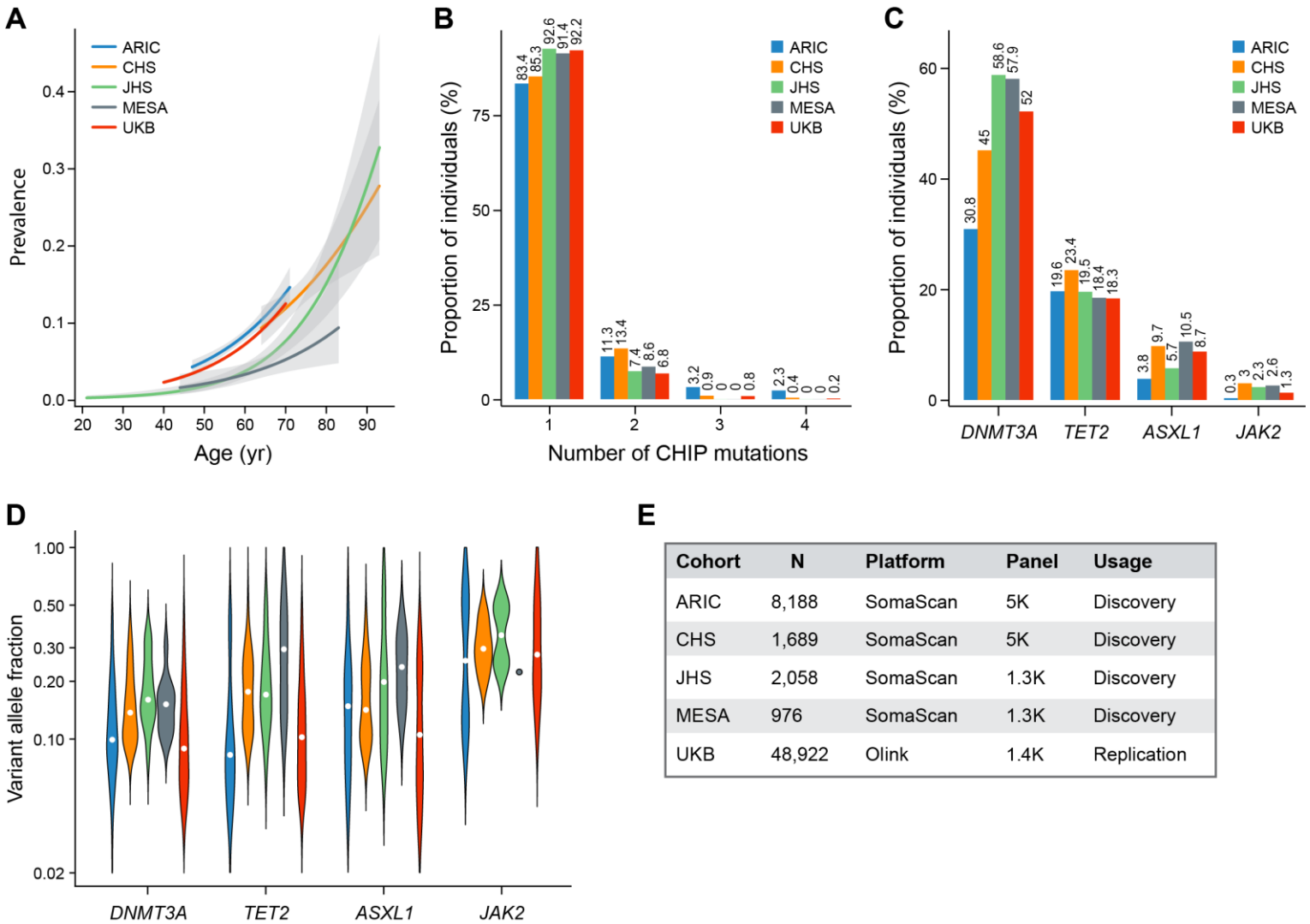


Figure 1: Scheme of the study design. We assessed the associations of CHIP and driver gene-specific CHIP subtypes (*DNMT3A, TET2, ASXL1*, and *JAK2*) with 1,148 circulating proteins measred by the SomaScan platform in 12,911 participants from TOPMed cohorts. Pathway analyses were conducted using IPA tools. We performed bone marrow transplantation of *Tet2*$^{-/-}$ and *Tet2*$^{+/+}$ myeloid cells in atherogenic mice, followed by scRNA-seq of aortic cells and data analysis using Seurat and other tools. CHIP-proteomic associations were examined in UK Biobank using Olink for protein measurements and compared with SomaScan-based results. Finally, we investigated the associations between prevalent CAD and proteomics, identifying shared proteins associated with both CAD and any examined CHIP variable. CAD: Coronary aterty disease. CHIP: Clonal hematopoiesis of indeterminate potential. TOPMed: Trans-Omics for Precision Medicine. Parts of this figure have been created with BioRender.com.



E

Cohort	N	Platform	Panel	Usage
ARIC	8,188	SomaScan	5K	Discovery
CHS	1,689	SomaScan	5K	Discovery
JHS	2,058	SomaScan	1.3K	Discovery
MESA	976	SomaScan	1.3K	Discovery
UKB	48,922	Olink	1.4K	Replication

Figure 2. CHIP and proteomics in TOPMed cohorts and UK Biobank. A, CHIP prevalence increased with donor age at the time of blood sampling. The center line represents the general additive model spline, and the shaded region is the 95% confidence interval (N_{ARIC}=8,188; N_{CHS}=1,689; N_{JHS}=2,058; N_{MESA}=976; N_{UKB}=48,922). B. More than 90% of individuals with CHIP had only one somatic CHIP driver mutation variant identified. C. Counts for four driver genes, *DNMT3A, TET2, ASXL1*, and *JAK2*, of CHIP mutations. D. CHIP clone size heterogeneity as measured by variant allele fraction by CHIP driver gene. Violin plot spanning minimum and maximum values. E. Platform and panel used for proteomics measurement by each cohort. CHIP: Clonal hematopoiesis of indeterminate potential.

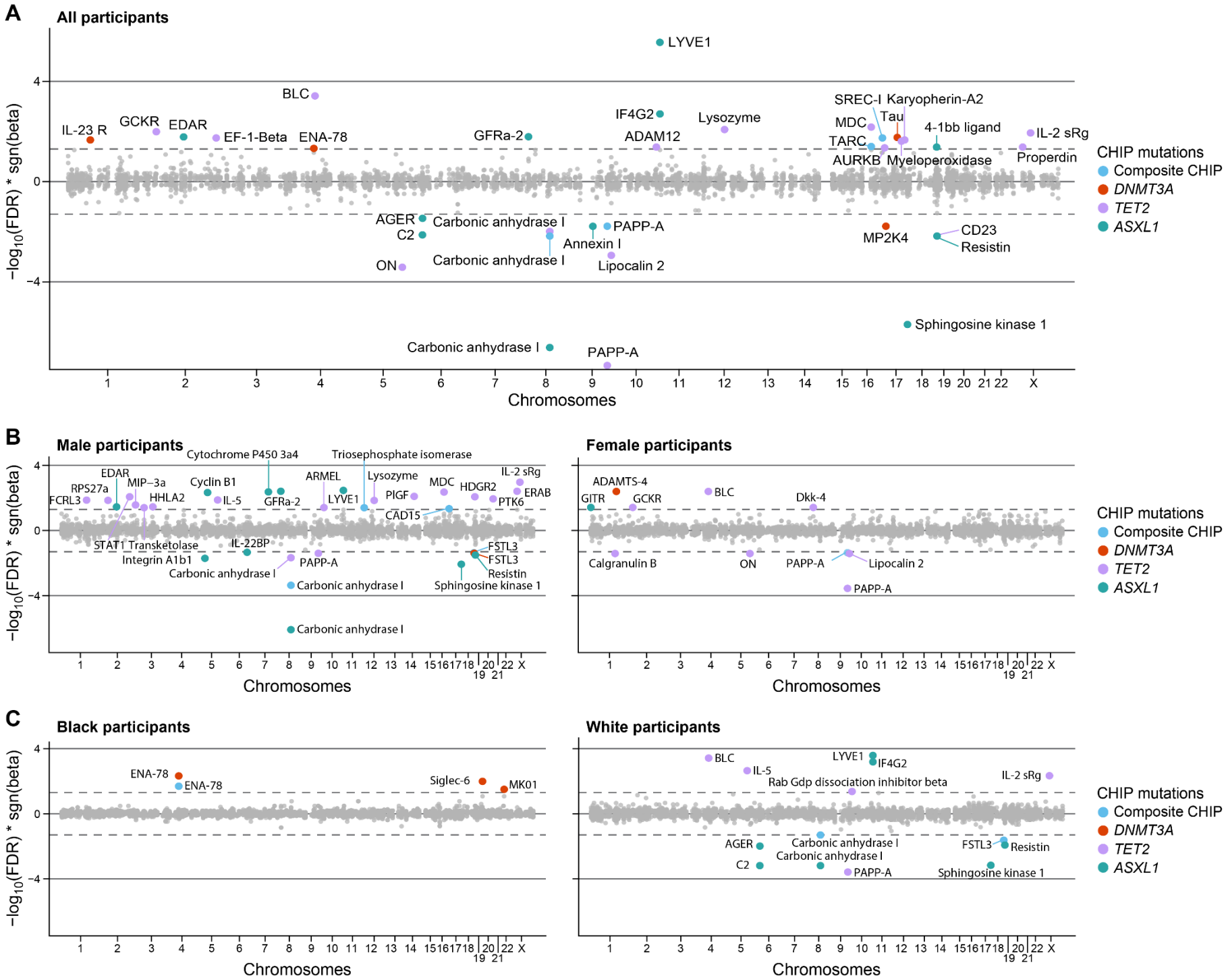

A
All participants
LYVE1
BLC
IL-23 R
GCKR
EDAR
EF-1-Beta
ENA-78
GFRa-2
IF4G2
ADAM12
Lysozyme
SREC-I
Karyopherin-A2
Tau
MDC
TARC
AURKB
Myeloperoxidase
4-1bb ligand
IL-2 sRg
Properdin
AGER
C2
Carbonic anhydrase I
PAPP-A
Annexin I
Lipocalin 2
ON
MP2K4
CD23
Resistin
Sphingosine kinase 1
Carbonic anhydrase I
PAPP-A
−log10(FDR) * sgn(beta)
Chromosomes
CHIP mutations
Composite CHIP
DNMT3A
TET2
ASXL1
B
Male participants
Cytochrome P450 3a4
Triosephosphate isomerase
EDAR
RPS27a
MIP−3a
Cyclin B1
ARMEL
Lysozyme
MDC
HDGR2
IL-2 sRg
ERAB
FCRL3
HHLA2
IL-5
GFRa-2
LYVE1
PIGF
PTK6
CAD15
STAT1
Transketolase
IL-22BP
FSTL3
Integrin A1b1
PAPP-A
Carbonic anhydrase I
Resistin
Carbonic anhydrase I
Sphingosine kinase 1
Carbonic anhydrase I
Female participants
ADAMTS-4
GITR
GCKR
BLC
Dkk-4
Calgranulin B
ON
PAPP-A
Lipocalin 2
PAPP-A
C
Black participants
ENA-78
ENA-78
Siglec-6
MK01
White participants
BLC
LYVE1
IF4G2
IL-5
Rab Gdp dissociation inhibitor beta
IL-2 sRg
AGER
C2
Carbonic anhydrase I
Carbonic anhydrase I
PAPP-A
FSTL3
Resistin
Sphingosine kinase 1

Figure 3. Meta-analyzed associations between CHIP mutations and circulating proteome in TOPMed cohorts. A. All participants (N=12,911). B. Male participants only (N=5,616) vs. Female participants only (N=7,295). C. Black participants only (N=4,452) vs. White participants only (N=8,076). Proteins that are associated at FDR=0.05 level (for 4,560 testings) are labeled with the corresponding SomaScan targets and colored in blue, red, green, and orange, indicating significant associations with composite CHIP, *DNMT3A*, *TET2*, and *ASXL1*, respectively. Associations were assessed through linear regression models adjusting for age at sequencing, sex (if applicable), race (if applicable), batch (if applicable), type 2 diabetes status, smoker status, first ten principal components of genetic ancestry, and PEER factors (the number of PEER factors varies by cohorts based on the sizes of study populations: 50 for JHS, MESA, and CHS; 70 for ARIC AA; 120 for ARIC EA). AA: African Ancestry; ARIC: Atherosclerosis Risk in Communities; CHIP: Clonal hematopoiesis of indeterminate potential; CHS: Cardiovascular Heart Study; EA: European Ancestry; FDR: False discovery rate; JHS: Jackson Heart Study; MESA: Multi-Ethnic Study of Atherosclerosis; PEER: Probabilistic estimation of expression residuals.

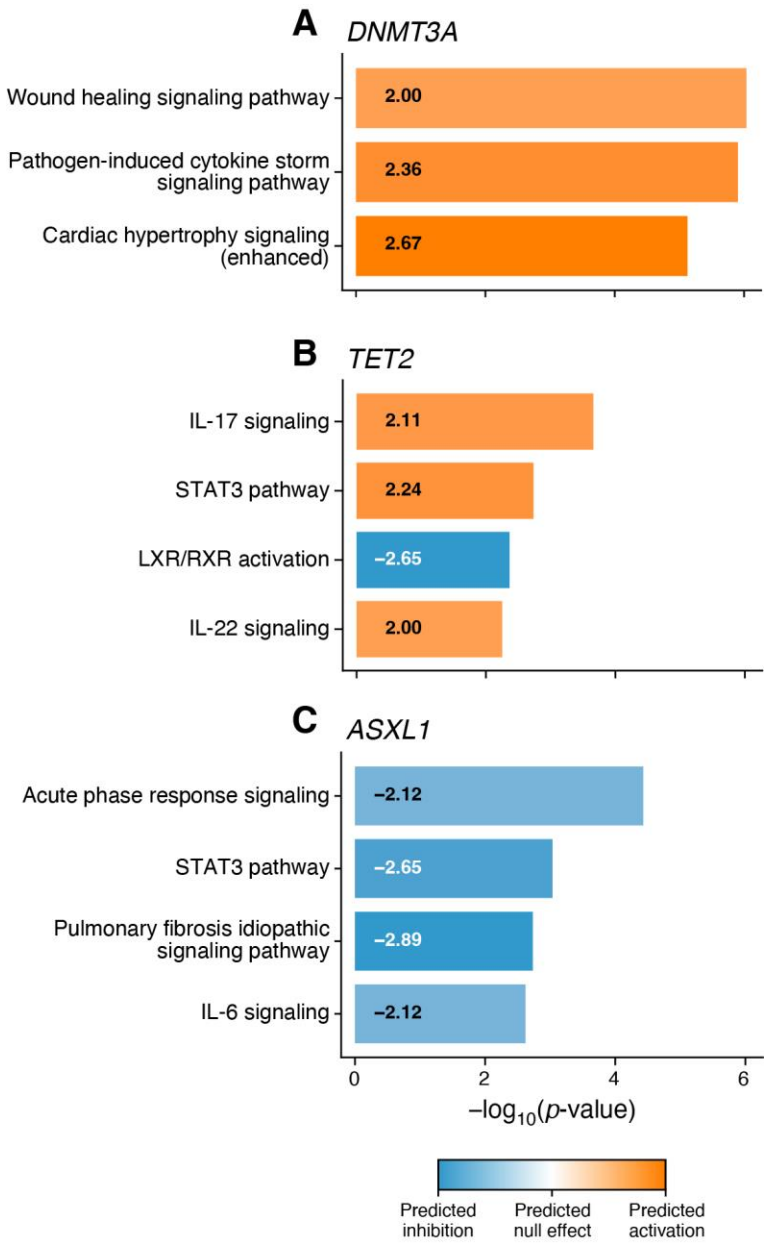


Figure 4. Significantly enriched and modulated pathways identified among proteins associated with CHIP driver genes. Significantly enriched and modulated pathways corresponding to CHIP-associated proteins were derived based on known genetic and molecular relationships

using IPA. The input was the Z-scores of the associations between major CHIP driver genes, i.e., *DNMT3A*, *TET2*, and *ASXL1*, and proteins that were significant at the P=0.05 level. The listed pathways fulfill two criteria: (1) within the top 30 most significantly enriched pathways by input proteins based on IPA analysis (P<0.05) and (2) being significantly modulated, either inhibited or activated, based on IPA analysis (Z>1.96). The orange indicates predicted activation, and the blue indicates predicted inhibition. The darker the color, the stronger the modulation effect. A. Significantly modulated canonical pathways implicated among proteins associated with *DNMT3A*. B. Significantly modulated canonical pathways implicated among proteins associated with *TET2*. C. Significantly modulated canonical pathways implicated among proteins associated with *ASXL1*. CHIP: Clonal hematopoiesis of indeterminate potential; IPA: Ingenuity Pathway Analysis

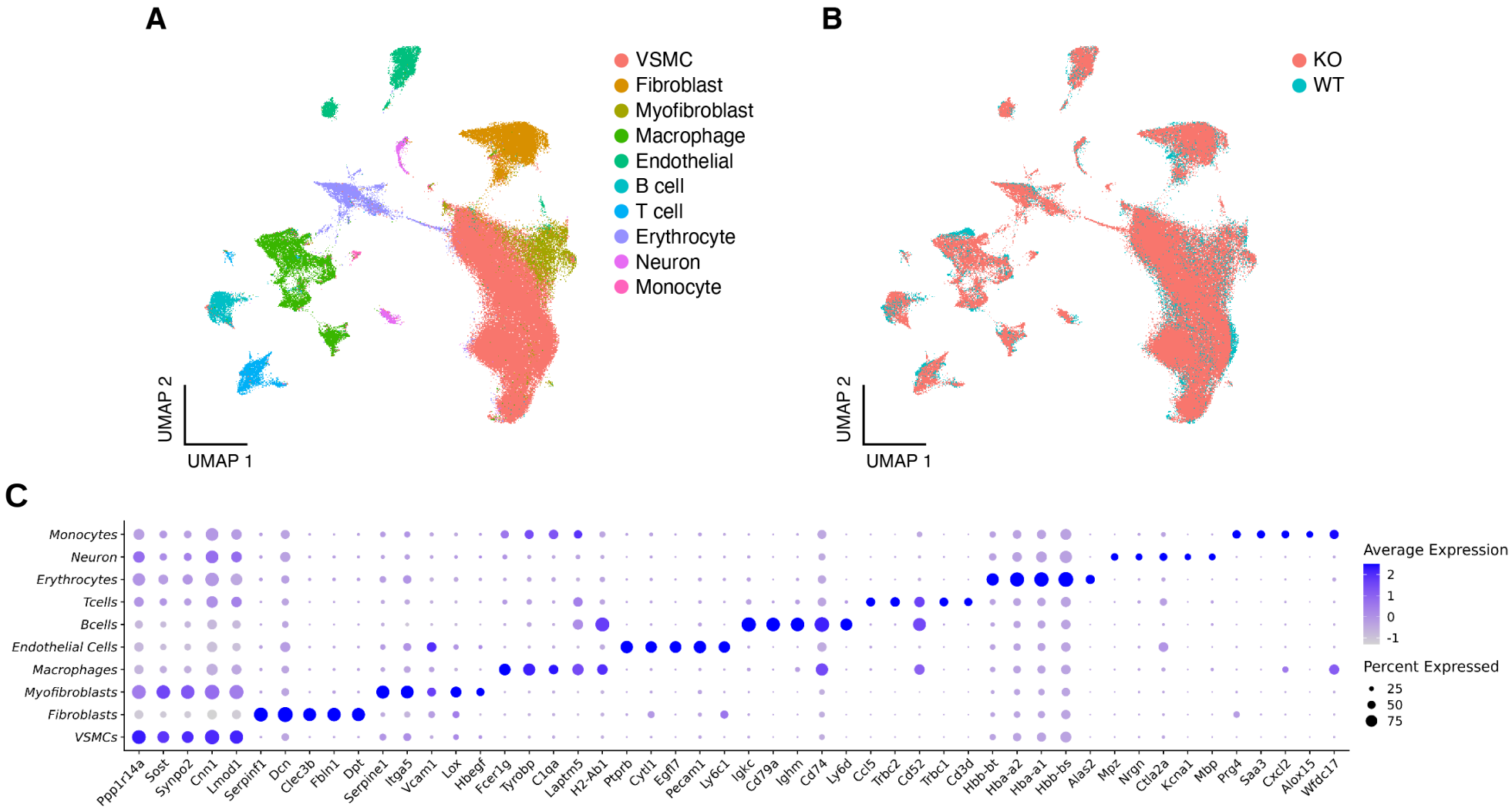


Figure 5. UMAP embeds differential gene expression. A. By cell types. B. By *Tet2*$^{-/-}$ and WT mice status. *Tet2*$^{-/-}$ mice were transplanted with suspensions of bone marrow cells with 100% *Tet2*$^{-/-}$ cells. C. Markers genes for each cell type. KO: Knock-out; VSMC: Vascular smooth muscle cells; WT: Wild type.

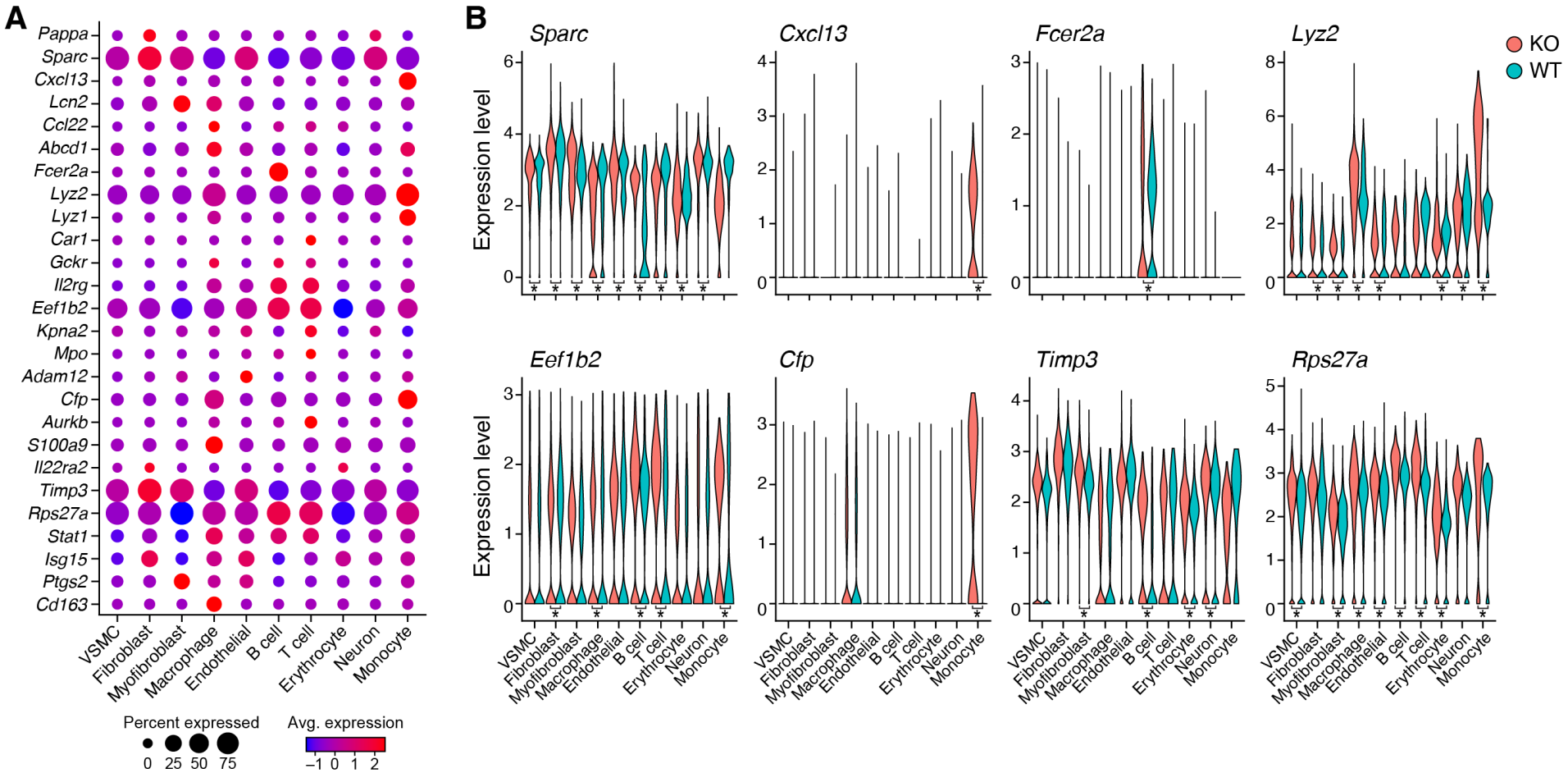


Figure 6: Single-cell RNA expression in *Tet2*$^{-/-}$ and WT mice for genes encoding plasma proteins associated with *TET2*. *Tet2*$^{-/-}$ mice were transplanted with suspensions of bone marrow cells with 100% *Tet2*$^{-/-}$ cells. For genes encoding proteins associated with *TET2* at FDR<0.1 level (for 4,560 testings), single-cell RNA expression levels in CHIP-related cell types were examined and compared between *Tet2*$^{-/-}$ mice and WT mice. CHIP-related cell types include VSMCs, fibroblasts, myofibroblasts, macrophages, endothelial cells, B cells, T cells, erythrocytes, neurons, and monocytes. The expression level was log2 transformed after adding a pseudo-count of 1 to serve as the background. The mapping of human proteins to mice genes is based on information from Uniprot and listed in Supplemental Table X. CHIP: Clonal hematopoiesis of indeterminate potential; FDR: False discovery rate; KO: Knock-out; VSMC: Vascular smooth muscle cells; WT: Wild type

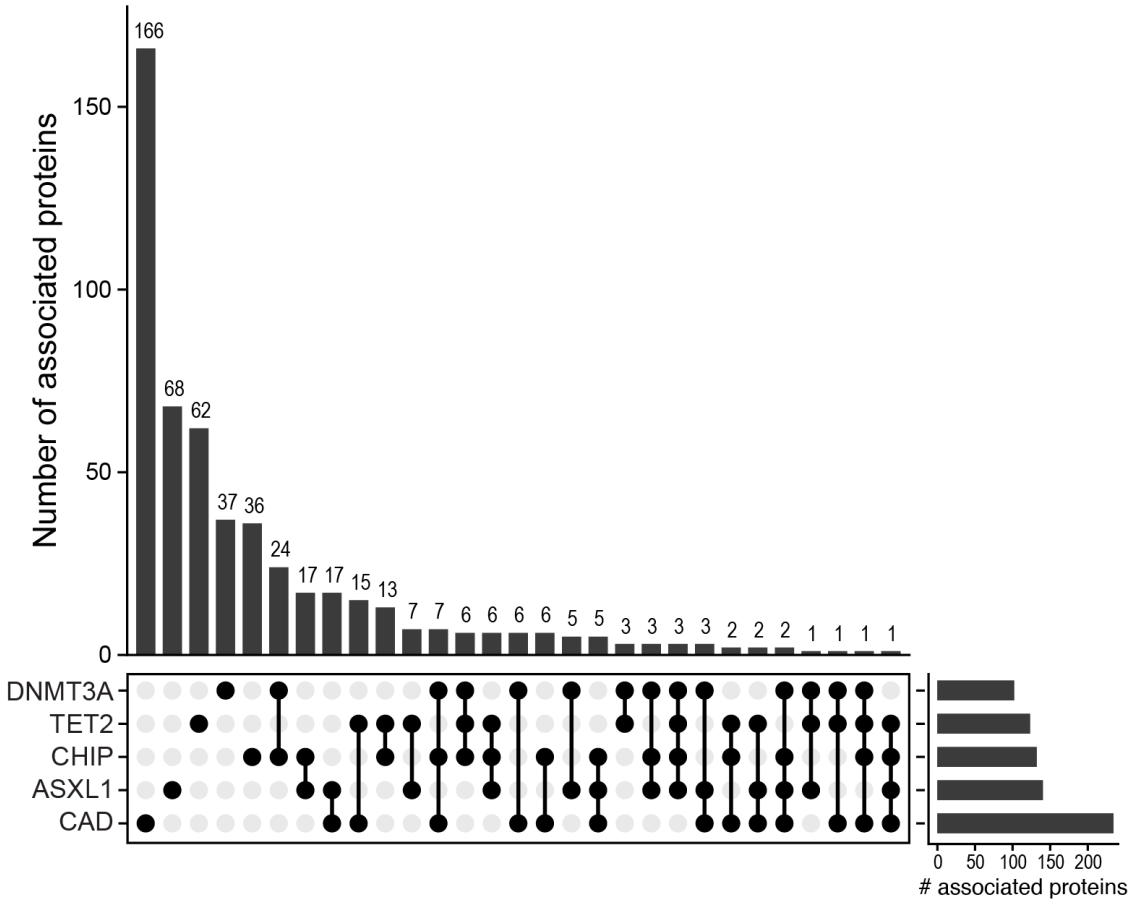


Figure 7. Upset plot showing overlapped and non-overlapped associated proteins between CHIP variables and CAD. A total of 68 proteins were associated with both prevalent CAD and any of the CHIP variables (composite CHIP, *DNMT3A, TET2*, and *ASXL1*) at P=0.05 level. For both CHIP variables and CAD, associations were assessed through linear regression models adjusting for age at sequencing, sex, race, batch (if applicable), type 2 diabetes status, smoker status, and the first ten principal components of genetic ancestry. PEER factors (the number of PEER factors varies by cohorts based on the sizes of study populations: 50 for JHS, MESA, and CHS; 70 for ARIC AA; 120 for ARIC EA) were adjusted in CHIP analysis only but not CAD analysis; this is because around 1/3 of them were associated with CAD, but in general not with CHIP. AA: African ancestry; ARIC: Atherosclerosis Risk in Communities; CAD: Coronary artery disease; CHIP, clonal hematopoiesis of indeterminate potential; CHS: Cardiovascular Heart Study; EA: European ancestry; FDR: False discovery rate; JHS: Jackson Heart Study; MESA: Multi-Ethnic Study of Atherosclerosis; PEER: Probabilistic estimation of expression residuals.