# Genetic modification of inflammation and clonal hematopoiesis-associated coronary artery disease

Zhi Yu (1, 2), Trevor P. Fidler (3), Yunfeng Ruan (1), Caitlyn Vlasschaert (4), Tetsushi Nakao (1, 2, 5, 6), Md Mesbah Uddin (1, 2), Taralynn Mack (7), Abhishek Niroula (1, 5, 8), J. Brett Heimlich (7), Seyedeh M. Zekavat (1, 2, 9), Christopher J. Gibson (5), Gabriel K. Griffin (1, 6, 10), Yuxuan Wang (11), Gina M. Peloso (11), Nancy Heard-Costa (12, 13), Daniel Levy (13, 14), Ramachandran S. Vasan (12, 13, 15), François Aguet (1), Kristin Ardlie (1), Kent D. Taylor (16), Stephen S. Rich (17), Jerome I. Rotter (16), Peter Libby (18), Siddhartha Jaiswal (19), Benjamin L. Ebert (1, 5), Alexander G. Bick (1, 7), Alan R. Tall (3), Pradeep Natarajan (1, 2, 20)

**Affiliations:**

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. Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA

4. Department of Medicine, Queen's University, Kingston, ON, Canada

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

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

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

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

9. Computational Biology & Bioinformatics Program, Yale University, New Haven, CT, USA

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

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

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

13. Framingham Heart Study, Framingham, MA, USA

14. Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

15. Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

16. 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

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

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

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

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

Abstract word counts: 139

Text word counts: 5198

No. of references: 81

No. of figures: 5

No. of tables: 2

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**

Clonal hematopoiesis of indeterminate potential (CHIP) is associated with an increased risk of cardiovascular diseases (CVD), putatively via inflammasome activation. We pursued an inflammatory gene modifier scan for CHIP-associated CVD risk among 424,651 UK Biobank participants. CHIP was identified using whole exome sequencing data of blood DNA and modeled both as a composite and for common drivers (*DNMT3A*, *TET2*, *ASXL1*, and *JAK2*) separately. We developed predicted gene expression scores for 26 inflammasome-related genes and assessed how they modify CHIP-associated CVD risk. We identify *IL1RAP* as a potential key molecule for CHIP-associated CVD risk across genes and increased *AIM2* gene expression leading to heightened *JAK2*- and *ASXL1*-associated CVD risks. We show that CRISPR-induced *Asxl1* mutated murine macrophages have a particularly heightened inflammatory response to AIM2 agonism. Our study provides new evidence to support gene-specific strategies to address CHIP-associated CVD risk.

## Introduction

Clonal hematopoiesis (CH) of indeterminate potential (CHIP) is the age-related acquisition and expansion of somatic mutations of genes frequently mutated in hematologic malignancies (e.g., *DNMT3A*, *TET2*, *ASXL1*, or *JAK2*)[1] detected from sequencing blood DNA among asymptomatic individuals. CHIP is common among older adults, affecting at least 1 in 10 adults over 70 years[2-5]. CHIP is associated with an increased risk of hematologic malignancy and all-cause mortality[3,4], as well as a range of cardiovascular diseases (CVD)[6-10]. Recent evidence, primarily from murine and cell-based studies, suggests that dysregulated inflammation may be a key contributor to the augmented risk of CVD conferred by certain CHIP mutations [6,11-14].

Heightened interleukin (IL)-1β signaling, a key inflammatory pathway, , promotes the development of CHIP-associated atherosclerosis in *Tet2* CHIP as initially disclosed largely by murine studies[6,11]. Inhibition of the NOD-, LRR- and pyrin domain-containing protein 3 (Nlrp3) inflammasome abrogates accelerated atherosclerosis in atherogenic mice with hematopoietic *Tet2* deficiency versus wild-type[11,15]. In humans, CHIP is associated with increased gene expression and circulating concentrations of NLRP3 downstream products, particularly in the context of *TET2* CHIP[16-18]. Humans harboring *IL6R* p.Asp358Ala, a common variant known to disrupt *IL6R* and associate with modestly reduced CVD risk reduction in the general population, had greater reductions in CVD risk among individuals with *DNMT3A* or *TET*2 CHIP mutations versus those without[7]. However, recent murine work indicates that different CHIP genes may confer CVD risk differentially. For example, among atherogenic transgenic mice expressing *Jak2*$^{VF}$, bone marrow genetic deficiency of the absent in melanoma 2 (*Aim2*) inflammasome mitigated atherosclerotic lesion development [15]. Whether these findings extend to humans is currently not well understood. In general, the range of inflammatory cytokines

differentially influencing CVD risk by CHIP genes in humans requires further study.. Prioritization by human genetics may yield or bolster new approaches to CVD precision medicine[19].

To overcome risks of confounding from biomarker correlation analyses, we leveraged genetics to pursue a broader inflammatory gene modifier scan for CHIP -associated CVD among 424,651 UK Biobank participants with blood DNA exome sequencing for CHIP genotyping, array-derived genome-wide genotyping for transcriptomic imputation, and baseline and incident clinical outcomes. We developed predicted gene expression scores for genes related to the NLRP3 and AIM2 inflammasomes based on externally trained data and conducted independent validation. Then we assessed whether and to what extent the predicted gene expression modifies CHIP-associated CVD risk. Lastly, we validated a human genomics-based discovery in a murine model. Broadly, we demonstrate a systematic approach to prioritize potential therapeutic strategies for CHIP-associated disease.

## Results

### Baseline characteristics of the UK Biobank cohort

The schematic of this study is shown in **Figure 1**. Among the 424,651 unrelated participants enrolled in the UK Biobank study who underwent exome sequencing and were free of hematologic cancers and CVD at baseline, the mean (SD) age was 56.4 (8.1) years, and 190,019 (44.7%) were men. We identified 26,353 (6.2%) individuals with CHIP mutations, of whom 14,580 (55.3%) had mutations in *DNMT3A*, 5,230 (19.8%) in *TET2*, and 2,436 (9.2%) in *ASXL1*. Two hundred and fifty-five participants (1.0%) had *JAK2* mutations, 229 (89.8%) of whom had *JAK2* p.V617F and 248 (97.3%) were large clones defined as a variant allele fraction (VAF)

>10%. Consistent with previous reports, participants with CHIP versus those without were on average four years older, more likely to be White, have higher BMI, be ever smokers, and have a higher prevalence of cardiovascular comorbidities including hypertension, hyperlipidemia, and type 2 diabetes mellitus (**Table 1**).

**Associations between CHIP mutations and incident CVD**

During the 11.0-year median follow-up, 44962 (10.6%) incident CVD events (a composite of myocardial infarction, coronary artery disease or revascularization, stroke, or death)[7] were observed. The presence of any CHIP associated with increased CVD event risk independent of potential confounders (age, sex, white British ancestry, body mass index [BMI] at the time of enrollment, ever-smoker status, diagnoses of type 2 diabetes mellitus at the time of enrollment, and the first ten principal components of genetic ancestry) with a composite effect of hazard ratio (HR) 1.18 (95% confidence interval [CI]: 1.14-1.22, *P*: $1.5\times10^{-21}$). Among the top CHIP genes, CVD effects varied by genes with *JAK2* 2.81-fold (95% CI: 2.25-3.51, P-value: $8.5\times10^{-20}$), *ASXL1* 1.41-fold (95% CI: 1.29-1.54, *P*: $3.5\times10^{-14}$), *TET2* 1.11-fold (95% CI: 1.03-1.19, *P*: $4.5\times10^{-3}$), and *DNMT3A* 1.06-fold (95% CI: 1.01-1.11, *P*: 0.01). In addition, other CHIP genes also showed significant associations with CVD incidence with *SRSF2* 2.6-fold (95% CI: 2.18-3.09, P-value: $6.8\times10^{-27}$), *SF3B1* 1.47-fold (95% CI: 1.14-1.89, *P*: $2.9\times10^{-3}$), *TP53* 1.43-fold (95% CI: 1.18-1.72, *P*: $2.2\times10^{-4}$), and *PPM1D* 1.39-fold (95% CI: 1.18-1.64, *P*: $7.6\times10^{-5}$). Large clones generally demonstrated larger effects, with large CHIP associated with 1.29-fold (95% CI: 1.24-1.35, P: $8.6\text{x}10^{-29}$) incident CVD risk. (**Table 2**). Sensitivity analyses restricting the outcome to coronary artery disease (CAD) alone yielded attenuated results in the same directions (**Supplemental Table 1**)

**Predicted expression of inflammatory genes**

We expanded the examination for CHIP modifiers through two dimensions. (1) In addition to a composite of all CHIP mutations at any driver genes, we examined the most commonly mutated CHIP genes individually[6], such as *DNMT3A*, *TET2*, *ASXL1*, and *JAK2*. (2) In addition to *IL6R*, we generated predicted expression levels of all other inflammatory genes that are implicated in or closely related to the NLRP3 and AIM2 inflammasome pathways, including *NLRP3*, *IL1B*, *IFNG*, *IL18*, *CARD8*, *CASP1*, *CASP5*, *DHX33*, *IFNGR1*, *IFNGR2*, *IL1R1*, *IL1R2*, *IL1RAP*, *IL6*, *IL6ST*, *IL10*, *IL18BP*, *IL18R1*, *IL18RAP*, *IRF1*, *JAK1*, *JAK2*, *JAK3*, *NEK7*, *NLRC4*, *SOCS*, *STAT1*, *STAT3*, *STAT4*, *STAT5A*, *STAT6*, *TNF*, and *TYK2* (**Methods** and **Supplemental Figure 1**).

We developed predicted expression scores based on summary statistics of the whole blood or peripheral blood mononuclear cells (PBMC) cis-expression quantitative trait locus (eQTL) results for the corresponding genes from the eQTLGen Consortium[20]. For each selected gene, we used both the pruning and thresholding method (P+T)[21] and the polygenic risk score-continuous shrinkage (PRS-CS) method[22] to generate a series of candidate scores for participants with European ancestry (EA) and non-European ancestry separately; they were then tuned using non-overlapping individual-level RNA-sequencing data from the Framingham Study (FHS; whole blood) and Multi-Ethnic Study of Atherosclerosis (MESA; PBMC)[23,24]. The final predicted expression score of each gene was selected based on the proportion of the variance ($R^2$) of experimentally measured expression levels that can be explained by the candidate scores (**Methods**). For most genes, P+T method generated better performed score than PRS-CS (**Supplemental Table 2**). For this analysis, we kept genes whose selected best-performing

predicted expression scores had $R^2$ >1% among EA participants, resulting in scores for 26 (of 35 total evaluated) genes. The predicted expression scores explained a median of 3.5% [interquartile range 1.8%-6.3%] of the adjusted variance of corresponding gene expression levels among EA participants. The score for *IL18RAP* explained the largest proportion of phenotypic variance (34.7%), and that for *IL1B* explaining the least (1.05%) among analyzed genes with $R^2$>1% (**Figure 2 and Supplemental Table 2**).

**Modification of CHIP-associated CVD risk by predicted expression of inflammatory genes**

We observed significant associations between predicted expression scores of several inflammatory genes and incident CVD risk with the presence of CHIP or specific CHIP gene(s) (collectively called CHIP variables), while the corresponding associations for those without CHIP were all non-significant. For predicted expression scores that were significantly associated with incident CVD risk at $P$<0.05 level only in the presence of CHIP variable(s), we carried forward to evaluate the interactions between those scores and the corresponding CHIP variables (N=9 pairs) on the primary CVD outcome (**Figures 3 and 4**).

Regarding specific modification pairs, first, we found evidence supporting recent murine findings[15] in humans by observing that a genetic predisposition to higher *AIM2* expression was associated with amplified risk for incident CVD for those with *JAK2* CHIP. One standard deviation (SD) increase in predicted expression score for *AIM2* was associated with an almost 2-fold increased risk in CVD incidence (HR: 1.85, 95% CI: 1.12-3.07, $P$: 0.02) among participants with *JAK2* mutations. In contrast, the predicted expression score for *AIM2* was not associated with incident CVD event risk in those without *JAK2* mutations (HR, 0.99; 95% CI: 0.98–1.00; $P$: 0.16), which was significantly different for those with *JAK2* CHIP (false discovery rate [FDR]

for interaction: 0.04). A recent murine study demonstrated that mice expressing *Jak2*$^{VF}$ displayed a 2-fold increase in atherosclerotic lesion development and that increased risk was markedly reduced in the presence of *Aim2* bone marrow genetic deficiency[15].

Second, we observed modification effects of the predicted expression level of *IL1RAP* on incident CVD risk associated with any CHIP, *DNMT3A*, and *JAK2* CHIP mutations. *IL1RAP* encodes IL-1 receptor accessory protein (IL-1RAP), a coreceptor involved in several inflammatory signaling pathways and the lack of which completely abrogates cellular response to IL-1[25-28]. For one standard deviation (SD) increase in the predicted expression score for *IL1RAP*, HRs (95% CI) were 1.04 (1.01, 1.07) in the presence of any CHIP mutations, 1.06 (1.02, 1.11) in the presence of *DNMT3A* mutation, and 1.38 (1.13, 1.69) in the presence of *JAK2* mutations, in contrast with HRs (95% CI) of 1.00 (0.99, 1.01), 1.00 (0.99, 1.01), and 1.00 (0.99, 1.01) among participants without these mutations (FDR for interaction: 0.04, 0.04, and 0.04, respectively). While the relationship for *TET2* was directionally consistent, no significant association was observed. This result implicates IL1RAP as a potentially key IL-1β/IL-6 pathway-related molecule for CHIP-associated CVD risk across genes[7,11].

Third, we identified novel potential modification effects by the predicted expression of *AIM2* and *IL10* on *ASXL1*-associated CVD risk. In addition to it *AIM2*'s aforementioned interaction with *JAK2* on CVD risk, *AIM2* predicted expression showed a similar modification effect on the *ASXL1*-associated CVD disease risk (*ASXL1* mutation present: HR: 1.14, 95% CI: 1.02-1.28; *ASXL1* mutation absent: HR: 0.99, 95% CI: 0.98-1.00; FDR for interaction: 0.04). Similar effects were not observed for *DNMT3A* or *TET2*-associated CVD. *IL10* is expressed in atherosclerotic plaques, and its encoded protein, IL-10, is an anti-inflammatory cytokine that inhibits many cellular processes that advance human artherosclerosis[29-37]. The protective effect

of IL-10 is pronounced in the presence of *ASXL1* mutation, with its predicted expression score associated with a significantly decreased risk of incident CVD (HR, 0.91; 95% CI: 0.83–0.99; *P*: 0.04) in the presence of *ASXL1* mutation but a null effect (HR, 1.00; 95% CI: 0.99–1.01; *P*: 0.91) when it is absent (FDR for interaction: 0.06). Another molecule implicated was *IL18RAP* which encodes IL-18 receptor accessory protein (IL-18RAP). IL-18RAP enhances the IL-18-binding activity of the IL-18 receptor and plays a role in signaling by the inflammatory cytokine IL-18[38]. However, we observed attenuated CVD risk associated with the predicted expression score of *IL18RAP* among participants with *ASXL1* mutation (HR: 0.90, 95% CI: 0.83-0.98, *P*: 0.02) but not those without (HR: 1.00, 95% CI: 0.99-1.01, *P*: 0.41; FDR for interaction: 0.04). These results are shown in **Figure 3, Figure 4, Supplemental Table 3,** and **Supplemental Table 4**. These identified inflammatory expression scores that modify CHIP variable-associated CVD risk were not associated with the corresponding CHIP variable, with *JAK2* gene expression and *JAK2* CHIP mutation (FDR=$6.1\times10^{-6}$) as the exception.

**Aim2 inflammasome activation from *Asxl1* mutations**

Motivated by the aforementioned findings, we sought to explore the inflammatory response of *Asxl1* murine macrophages. Our results validated a recent murine study that demonstrated that the bone marrow genetic deficiency of *Aim2* mitigated atherosclerotic lesion development to a greater extent among atherogenic *Jak2*$^{VF}$ versus atherogenic mice wild-type for *Jak2*[15], lending support to our analytical approach. To determine if *Asxl1* mutations promote AIM2 inflammasome activation in macrophages, we introduced a frameshift mutation in exon 12 of *Asxl1* in murine LT-HSCs to mimic human mutations (**Figure 5A**) and compared the IL-1β production in the presence of AIM2 agonist (double-stranded DNA fragments [pdAdT]) between

*Asxl1* mutated murine bone marrow-derived macrophages (BMDMCs) and wild-type murine BMDMs. In line with our human genetics findings, *Asxl1* mutant BMDMs had potentiated AIM2 inflammasome activation with a doubling of the Il-1β concentration in response to pdAdT at both high and low concentrations. In contrast, this inflammatory amplification effect by the presence of *Asxl1* mutation was not heightened in response to the NLRP3 agonist, ATP (**Figure 5B**).

**Associations with hematopoietic traits and cardiometabolic biomarkers**

For the eight CHIP mutation-predicted expression score pairs that showed significant modification effects on CVD incidence, we examined the associations between those predicted expression scores with 31 hematopoietic traits and five common cardiometabolic biomarkers among participants with the corresponding CHIP mutations. After accounting for multiple hypothesis testing [N=248 (8*31) for hematopoietic traits and N=40 (8*5) for cardiometabolic biomarkers], we did not observe any significant associations passing the FDR=0.05 threshold. The suggestive nominal associations were observed between the predicted expression score of *IL18RAP* and reduced eosinophil count and eosinophil percentage among individuals with *ASXL1* mutations (P=0.002 and 0.003, respectively). This is in line with previous analysis of cap analysis of gene expression (CAGE) sequencing data showed that *IL18RAP* is highly expressed in eosinophils, neutrophils, and natural killer (NK) cells[39] **(Supplemental Figure 2**, **Supplemental Table 5**, and **Supplemental Table 6**).

**Discussion**

Leveraging validated human genetic instruments, we show that specific inflammatory genes may influence incident CVD risk in a manner that is specific to the presence of mutations in key CHIP genes. Our observations are consistent with the notions that reduced *AIM2* expression could specifically mitigate *JAK2* mutation-associated CVD risk and with IL1RAP as a key molecule for CHIP-associated CVD risk across multiple CHIP genes; these findings agree with prior murine studies. Furthermore, we discovered potential modification by *AIM2* expression on *ASXL1*-associated CVD risk in humans with corroboration in CRISPR-induced *Asxl1* mutated mice. Our observations provide human genetic and pre-clinical support toward new precision medicine paradigms for CVD meriting assessment in prospective studies.

Our study has three key implications. First, our findings further show that CVD prognosis and mechanism are distinct by implicated CHIP gene. Prior studies showed that NLRP3 inflammasome inhibition mitigates the heightened atherogenesis observed in *Tet2* chimeric atherogenic mice compared to atherogenic mice wild-type for *Tet2*[11]. Correspondingly, a common disruptive coding variant in *IL6R* (a downstream mediator of NLRP3) *TET2* or *DNMT3A* CHIP among humans[7,40]. A *post hoc* exploratory analysis of a completed clinical trial of a monoclonal antibody targeting IL-1B (also a downstream mediator of NLRP3) supports this finding[41]. Recently, it was observed that atherogenic mice expressing $Jak2^{VF}$ displayed a 2-fold increase in atherosclerotic lesion area with increased features of plaque instability that were reduced in the presence of hematopoietic *Aim2* deficiency. The present study used human genetics as instruments and observed similar attenuation effects by genetically-predicted lower expression levels of *AIM2* on *JAK2*-associated CVD risk. These pre-clinical data lend support for AIM2 inflammasome inhibition to address *JAK2*-associated increased CVD risk.

Furthermore, we discovered AIM2's potential modulatory role for *ASXL1*-associated CVD risk in humans and validated it by demonstrating increased AIM2 inflammasome activation in bone marrow-derived macrophages harboring CRISPR-induced *Asxl1* mutation. Our analyses indicate that the inflammation response is heightened to a greater degree than the NLRP3 inflammasome, which is implicated in *TET2*-associated CAD. Prior studies showed that *Asxl1* mutant knock-in mice had elevated reactive oxygen species and increased DNA damage that could potentially lead to AIM2 inflammasome activation[42].

Second, we observed that increased genetic predisposition to *IL1RAP* expression yields increased incident CVD risk for participants with both *DNMT3A* or *JAK2* CHIP mutations. IL-1RAP is a transmembrane protein that potentiates multiple inflammatory signaling pathways, including IL-1, IL-33, IL-36G, and stem cell factor[27,28], and it has a unique feature of being expressed at higher levels in stem and progenitor cells from myeloid leukemia patients compared to normal hematopoietic stem and progenitor cells (HSPC)[43-46]. These properties of IL-1RAP led to several studies investigating the targetability of IL-1RAP as a treatment strategy for myeloid leukemia[25,45,47,48] and may underlie its modification effects of CHIP-associated CVD and, potentially, other disease risks. These observations agree with the aforementioned human genetic observations using a common missense variant in *IL6R*[7]. Furthermore, *Dnmt3a*-inactivated lineage-negative bone marrow cells versus wild-type cells transplanted into mice had greater IL-6 concentrations[49], and humans with *DNMT3A* mutations had greater expression of NLRP3-related cytokines among peripheral blood mononuclear cells[18]. While the results above and a prior murine study supports the role of AIM2 in *JAK2* CHIP, the IL-1β inflammasome was shown to also influence indices related to plaque stability in *Jak2*$^{VF}$ transgenic mice[15]. Given the significant impact of IL-1RAP predicted expression across all CHIP associated CAD risk,

whether IL-1RAP represents a more effective therapeutic target than individual inflammasomes or their downstream effectors warrants further study.

Third, our approach of using genetically predicted expression as therapeutic instruments in humans can potentially advance precision medicine for CVD and beyond. Precision medicine aims to identify and implement therapies that are maximally efficacious based on key features[50]. We leverage prior insights showing the value of human genetics for therapeutic development prioritization[19]. Prior studies have similarly used genotype-imputed transcriptomics to nominate therapeutic targets[51-53]. Given the overall relatively low heritability of inflammatory gene expression, we use both summary- and individual-level training data to impute gene expression perturbations from human genetics. We now compare effects by strata to identify subgroups that may maximally clinically benefit from inflammation modulation. Our subsequent murine validation lends overall support to this framework.

Our study has important limitations. First, the predicted expression scores for inflammatory genes are genetic proxies for expression levels from birth, which is well before the acquisition of age-related CHIP mutations. Thus, our analyses do not capture the modification effects after CHIP is manifest, which would more closely mimic clinical trials. However, our approach was corroborated by murine modeling introducing an inflammatory stimulus after a CHIP mutation is introduced. Second, CHIP mutations remain uncommon in the unselected population and so power is limited for interaction analyses. Third, our framework is similarly dependent on suitable heritabilities of the gene expression instruments and we are thus underpowered to detect associations for instruments with low heritabilities. Since we used individual-level validation data, we were able to exclude instruments with very low heritabilities to optimize multiple hypothesis testing. Lastly, the majority of participants in our study

population, as well as the eQTLGen Consortium, which we used for generating the predicted expression score, were of European ancestry[20,54]; therefore, our findings may not be generalizable to other ancestries.

In conclusion, validating the approach used, our study replicated murine findings in humans that *JAK2* CHIP mutation enhances CVD risk, and the unexpected finding that genetically reduced *Aim2* expression specifically reduces this risk. Examinations across the modification of other CHIP mutation-associated CVD risk by the predicted expression levels of other inflammatory genes yield several novel findings, including modification by *AIM2* expression in *ASXL1*-associated CVD risk, which we corroborated in CRISPR-induced *Asxl1* mutant mice. Our results may contribute to developing CHIP-type specific CVD therapies and advance precision medicine goals.

## Methods

### Study Population

In the current analysis, we included the first 424,651 unrelated participants enrolled in the UK Biobank study who underwent exome sequencing of blood DNA and were free of hematologic cancer and CVD at baseline[55,56]. Between 2006 and 2010, approximately 500,000 residents of the UK aged 40-69 years were recruited at one of 22 assessment centers across the UK and had samples, including blood-derived DNA, collected at baseline as well as baseline clinical characteristics, biomarkers, and subsequently incident clinical events through medical history and linkage to data on hospital admissions and mortality. Details regarding this cohort have been described elsewhere in detail[54]. Relatedness was defined as one individual in each pair within the

third degree of relatedness determined based on kinship coefficients centrally calculated by UK Biobank[54]. Secondary use of data for the present analysis was approved by the Massachusetts General Hospital Institutional Review Board (protocol 2021P002228) and facilitated through UK Biobank Application 7089.

**Whole exome sequencing and CHIP detection**

Exomes of approximately 450,000 UK Biobank participants were sequenced from blood-derived DNA at the Regeneron Genetics Center, as reported previously[56]. Briefly, exomes were captured by IDT's xGen probe library and sequenced on the Illumina Novaseq platform. Sample-specific FASTQ. files were aligned to the GRCh38 reference. The resultant binary alignment file (BAM) containing the genomic information was evaluated for duplicate reads using Picard3 MarkDuplicates tool and then converted by samtools to CRAM files that, after going through quality controls, were submitted to the UK Biobank data repository for distribution. CHIP detection was conducted through using GATK Mutect2 software (https://software.broadinstitute.org/gatk) as previously performed[7,57,58]. Participants were annotated as having putative CHIP if the output contained at least one of a pre-specified list of putative CHIP variants in 74 genes anticipated to cause myeloid malignancy at a VAF>2% (**Supplemental Table 7**)[3,6,59]. Common sequencing artifacts and germline variants were excluded as described elsewhere[7].

**RNA sequencing data**

RNA sequencing (RNA-seq) data was obtained from two TransOmics in Precision Medicine (TOPMed) cohorts: Multi-Ethnic Study of Atherosclerosis (MESA) and Framingham Heart Study (FHS).

MESA is a multi-ancestry prospective cohort of 6,814 self-identified White, Black, Hispanic, or Asian men and women free of clinical cardiovascular disease at recruitment in 2000-2002[60]. Included in this study are 889 individuals who had RNA-seq data in peripheral blood mononuclear cells (PBMCs) measured at baseline. A total of 889 participants were randomly selected from the MESA cohort for RNA sequencing for PBMC following standard protocol. The technical details for the sample acquisition and RNA sequencing can be found at Liu et al[61].

FHS is a multi-generational cohort initiated in 1948[62]. The Framingham Offspring cohort (Gen 2) was recruited in 1971 (N=5,124), and Generation 3 (Gen 3) cohort was recruited in 2002-2005 (N=4,095)[63,64]. The participants were predominantly self-identified White. Included in this study are 2,622 individuals from the Offspring and Gen 3 cohorts who had their peripheral whole blood samples collected and blood RNA sequenced at exam 9 and 2, respectively. The technical details for the blood draw and RNA sequencing can be found at Liu et al[65].

**Gene selection and predicted expression score generation**

We examined pairs of common CHIP mutations that are associated with CVD risk[6], including *DNMT3A*, *TET2*, *ASXL1*, and *JAK2*, and genetically-predicted expression levels of inflammatory genes that are biologically closely related to the NLRP3 or AIM2 inflammasomes; these genes were selected based on established biological pathways[66,67] and protein-protein interactions[68].

Specifically, both regulated by IFN-γ [66,69], AIM2 and NLRP3 inflammasome activation lead to cleavage of IL-1β and IL-18 to produce their mature forms[70,71]. IL-1β and IL-18 in their active forms then exert diverse biological functions related to inflammation[72], including inducing the production of IL-6, a strong independent predictor of cardiovascular outcomes[73,74]. We, therefore, include genes encoding these key proteins, namely *IFNG*, *AIM2*, *NLRP3*, *IL1B*, *IL18*, and *IL6R*. Based on the protein-protein interaction networks provided by STRING (https://string-db.org/), we further extended our study to genes that encode proteins with the top 10 highest interaction scores with each of the key proteins (since AIM2 and NLRP3 highly interact, we only kept one of them, NLRP3, as a key protein for selecting genes in the extended list). This resulted in a total of 29 additional genes, namely *CARD8, CASP1, CASP5, DHX33, IFNGR1, IFNGR2, IL10, IL18BP, IL18R1, IL18RAP, IL1R1, IL1R2, IL1RAP, IL6, IL6ST, IRF1, JAK1, JAK2, JAK3, NEK7, NLRC4, SOCS, STAT1, STAT3, STAT4, STAT5A, STAT6, TNF,* and *TYK2*.

For all selected genes, we used genotyping array data from the UK Biobank participants to generate predicted expression scores. The details on quality control and imputation of genotypic data in UK Biobank have been described elsewhere in detail[54]. Briefly, genotypic data were obtained using either UK BiLEVE Axiom arrays (Affymetrix Research Service Laboratory) or UK Biobank Axiom and then imputed to either the Haplotype Reference Consortium (HRC) or the merged UK10K+1000 Genomes as reference panel. Principal component analysis (PCA) was performed using fastPCA[75] based on a pruned set of 147,604 single nucleotide variations (SNVs) among unrelated individuals[76].

We calculated the predicted expression score as weighted sums of expression-increase allele counts among selected single nucleotide polymorphisms (SNPs), weighted by their raw or

posterior effect sizes on the expression levels of the corresponding genes (beta coefficient)[22,77]. Raw beta coefficient estimates based on summary statistics of the whole blood (85% of the Consortium) and peripheral blood mononuclear cells (PBMC; 15% of the Consortium) cis-eQTL results from the eQTLGen Consortium (N: 31,684; https://www.eqtlgen.org/)[20], with *cis* being defined as within +/-500,000 bp around the transcriptional start site (TSS) of the encoding gene of the target protein. The majority of participants included in the eQTLGen Consortium are of European descent, which is similar to our study population[20]. We used two methods to calculate the scores among EA and non-EA participants separately. (1) The pruning + thresholding (P+T) approach, where we used the raw effect size as weights for SNPs and conducted SNPs selection based on the following:

$$\textbf{Formula}: V_i(r_c^2, w_c, p_T) = \sum_{j \in S_{clumping}(r_c^2, w_c)} I(p_j < p_r)\widehat{\beta_j} \cdot G_{ij},$$

where for individual i, $\widehat{\beta_j}$ and $p_j$ are the effect size and *P* of variant j estimated from the summary statistics, respectively, $G_{ij}$ is the genotype dosage for that individual i and variant j, the set of $S_{clumping}(r_c^2, w_c)$ means restricting to variants remained after clumping at squared correlation threshold of $r_c^2$ and clumping window size of $w_c$, and $I(p_j < p_r)$ is a binary indicator function with 1 indicating *P* of variant j less than the specific *P* cutoff $p_r$ and 0 the other way[21]. For each gene, we created 30 candidate P+T-based predicted expression scores based on three $r^2$ levels (0.1, 0.01, and 0.001), five *P*-value thresholds ($5 \times 10^{-8}$, $1 \times 10^{-5}$, 0.001, 0.01, and 0.1), and two clumping window size (within 250kb and 5mb to both ends of the index SNP). (2) The PRS-CS approach, which uses a continuous shrinkage Bayesian framework to calculate the posterior mean of effect sizes, used as weights, across all SNPs[22]. For each gene, we also created four candidate PRS-CS-based predicted expression scores using four candidate global shrinkage

parameters ($1 \times 10^{-6}$, $1 \times 10^{-4}$, 0.01, and 1). For both approaches, we used a set of unrelated individuals from phase 3 1000 Genomes Project as the LD reference panel[78]. Since eQTLGen summary statistics were from both whole bold and peripheral blood mononuclear cells (PBMC) samples, we used genotypes and transcriptome concentrations from both FHS (whole blood) and MESA (PBMC) for score tunning[61]. For each gene, we selected the optimal method and parameters for generating the score based on the largest $R^2$ of the measured transcriptome levels in either FHS or MESA, since the eQTL source data was from either whole blood or PBMC. The best predicted expression scores were all standardized to zero-mean and unit-variance and were approximately normally distributed in the population. In the current study, we kept genes whose final-selected best-performed predicted expression scores had $R^2$ >1% among EA participants, resulting in suitable scores for 26 genes (**Figure 2** and **Supplemental Table 2**).

**Study outcomes**

The primary outcome, CVD event, was a composite of composite of myocardial infarction, coronary artery revascularization, stroke, or death as before[7]. We also secondarily used CAD for sensitivity analysis, which was defined as myocardial infarction, percutaneous transluminal coronary angioplasty or coronary artery bypass grafting, chronic ischemic heart disease, and angina. Both disease outcomes were defined by a combination of inpatient hospital billing International Classification of Diseases (ICD) codes and UK death registries listed in **Supplemental Table 8**[7]. The exploratory outcomes included 31 hematopoietic cell count indices and indices and five cardiometabolic biomarkers (C-reactive protein [CRP], total cholesterol, high-density lipoprotein [HDL] cholesterol, low-density lipoprotein [LDL] cholesterol, and

triglycerides). These conventionally measured biomarkers were analyzed as quantitative traits and were log2-transformed (with adding one across all measurements to avoid zero values for CRP), standardized to zero-mean and unit-variance, and normalized in the population. Blood samples of UK Biobank participants were collected into 4-mL EDTA vacutainers by vacuum draw, stored at 4 degrees centigrade, and then transported to the UK Biocentre in temperature-controlled shipping boxes[79]. Full blood counts were measured among all participants using clinical hematology analyzers at the centralized processing laboratory (Stockport, UK). Serum CRP level was measured by immunoturbidimetric high-sensitivity analysis on a Beckman Coulter AU5800. Lipid measurements were performed on the Beckman Coulter AU5800 platform and run using an immune-turbidimetric approach.

### ***Asxl1*-chimeric mice**

Bone marrows from CD45.2$^{+}$ Cas9 transgenic mice (The Jackson Laboratory, 026179) were harvested and enriched for c-Kit$^{+}$ cells using magnetic beads (Miltenyi Biotec, Cat # 130-091-224). LT-HSCs (Lin$^{-}$c-Kit$^{+}$Sca1$^{+}$CD48$^{-}$CD150$^{+}$)[80] were then harvested by flow cytometry sorting. LT-HSCs were then spinfected with 6ug/mL Polybreen (MilliporeSigma, TR-1003-G) and Lentiviruses containing non-targeting guides (Nmt4) or guides targeted to Asxl1 in exon 12 (*Asxl1*-G623*). LT-HSCs were washed and then incubated for three days. LT-HSCs were then mixed with 1 million supporting cells from CD45.1$^{+}$ wild type mice and transplanted into irradiated *Ldlr*$^{-/-}$ recipient mice.

### ***Asxl1*-CRISPR validation**

CRISPR guides designed to exon 12 of *Asxl1* were designed by CHOPCHOP[81] and screened in skin-derived fibroblasts from Cas9 transgenic mice. Guide sequence AGTGGTAACCTCTCGCCCCTCGG was evaluated by sanger sequencing of PCR amplification of flanking regions using Forward GCAGCATAAAATGGCTCTTGAT and reverse GCTGAGTCTTCTCTTCTGGCTC primers.

**Murine inflammasome activation studies**

Five weeks after transplantation, bone marrow was harvested and cultured in L-Cell Media for 5 days to generate bone marrow-derived macrophages (BMDMs). 20,000 BMDMs/well were seeded into 96 well plates and allowed to recover overnight. BMDMs were then primed with 20ng/mL LPS (Cell Signaling, 14011) for three hours and then stimulated with the indicated concentrations of ATP (Sigma) for 1 hour. For AIM2 inflammasome activation, BMDMs were primed for 1 hour with 20ng/mL LPS (Cell Signaling, 14011) and then incubated with lipofectamine 2000 (Thermofisher, Cat# 11668019) and poly (deoxyadenylic-deoxythymidylic) acid sodium salt (pdAdT) (Invivogen, tlrl-patn) for 6 hours. Following incubations, supernatants were collected, spun down at 3000g for 10 minutes, then assessed for IL-1β protein by ELISA (R&D Systems, DY401) **(Supplemental Table 9**).

**Statistical analyses**

We evaluated the association between CHIP mutations and incident CVD, as well as the modification effects by predicted expression levels of inflammatory genes measured as predicted

expression scores. Using Cox proportional hazard models, we first estimated the hazard ratios (HRs) and associated 95% confidence intervals (CI) of (1) the presence of CHIP mutations and (2) the presence of large clones, defined as a variant allele fraction (VAF) >10%, of CHIP mutations for incident CVD events. Then we conducted stratified analyses evaluating the associations between the predicted expression scores of selected inflammatory genes on the incidence of the primary outcome (i.e., CVD) with or without the presence of CHIP variables. For predicted expression scores that were associated with incident CVD risk (defined as $P<0.05$) only in the presence of CHIP variables(s), we carried forward to evaluate the interactions between those scores and the corresponding CHIP variables on the primary outcome. We considered time at risk to start at the enrollment of the study and continue until the event of interest, death, loss to follow-up, or the end of follow-up. Models were adjusted for age at the time of enrollment, sex, self-reported white British race/ethnicity, BMI, diagnoses of type 2 diabetes mellitus at the time of enrollment, ever-smoker status, and the first ten principal components of genetic ancestry. Since only less than 2% of the study population has missingness for any of the adjusted covariates, we removed those individuals from our regression models.

For significant interactions (FDR <0.05) discovered in the above analysis, we evaluated their associations across 31 hematological and five cardiometabolic traits using the same Cox proportional hazard models with adjusting for the same sets of covariates. All hematological and lipid traits were log2-transformed, standardized to zero-mean and unit-variance, and were approximately normally distributed in the population. Analyses used R version 4.0.0 software (The R Foundation, Vienna, Austra), two-tailed *P*-values, as well a statistical significance level of 0.05 for other analyses.

## Data availability

UK Biobank individual-level data are available by request via the application at https://www.ukbiobank.ac.uk. TOPMed individual-level DNA and RNA sequencing data used in this analysis are available through restricted access via the dbGaP.


## 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 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 fill list of participating MESA investigators and institutes can be found at http://www.mesa-nhlbi.org. This study was also supported in part by the National Institutes of Health, National Heart, Lung, Long and Blood Institute (NHLBI) contract 1R01HL151855 and the National Institute of Diabetes and Digestive and Kidney Diseases contract UM1DK078616. The Framingham Heart Study (FHS) acknowledges the support of contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible. Dr. Vasan is supported in part by the Evans Medical Foundation and the Jay and Louis Coffman Endowment from the Department of Medicine, Boston University School of Medicine.

## 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.T. is supported by Leducq Foundation (TNE-18CVD04) and NIH (HL155431). A.N. is supported by funding from the Knut and Alice Wallenberg Foundation (KAW 2017.0436). 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.L. receives funding support from the National Heart, Lung, and Blood Institute (1R01HL134892, 1R01HL163099-01, and 1R01HL163099-01), the American Heart Association

(18CSA34080399), 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). T.F. is supported by NHLBI (K99HL157649). Z.Y. is supported by NHLBI (5T32HL007604-37).

**Disclosures:**

A.R.T. is a scientific advisory board member and shareholder for Staten Biotech, TenSixteen Bio, Beren Pharmaceuticals and a consultant for CSL and Eli Lilly. 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. 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, Norvo 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, Moderna, Novartis, PlaqueTec, TenSixteen Bio, Soley Thereapeutics, and XBiotech, Inc. P.L.'s laboratory has received research funding in the last 2 years from

Novartis. P.L. is on the Board of Directors of XBiotech, Inc. PL 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. and A.R.T. are advisors to TenSixteen Bio. TenSixteen Bio is a company focused on clonal hematopoiesis but had no role in the present work. 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. Natarajan, P., Jaiswal, S. & Kathiresan, S. Clonal Hematopoiesis: Somatic Mutations in Blood Cells and Atherosclerosis. *Circ Genom Precis Med* **11**, e001926 (2018).
3. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488-98 (2014).
4. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477-87 (2014).
5. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* **20**, 1472-8 (2014).
6. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* **377**, 111-121 (2017).
7. Bick, A.G. *et al.* Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation* **141**, 124-131 (2020).
8. Bhattacharya, R. *et al.* Clonal Hematopoiesis Is Associated With Higher Risk of Stroke. *Stroke*, Strokeaha121037388 (2021).
9. Zekavat, S.M. *et al.* TP53-mediated clonal hematopoiesis confers increased risk for incident peripheral artery disease. *medRxiv*, 2021.08.22.21262430 (2021).
10. Yu, B. *et al.* Supplemental Association of Clonal Hematopoiesis With Incident Heart Failure. *J Am Coll Cardiol* **78**, 42-52 (2021).
11. Fuster, J.J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842-847 (2017).

12. Meisel, M. *et al.* Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature* **557**, 580-584 (2018).

13. Cai, Z. *et al.* Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis. *Cell Stem Cell* **23**, 833-849.e5 (2018).

14. Sano, S. *et al.* Tet2-Mediated Clonal Hematopoiesis Accelerates Heart Failure Through a Mechanism Involving the IL-1β/NLRP3 Inflammasome. *J Am Coll Cardiol* **71**, 875-886 (2018).

15. Fidler, T.P. *et al.* The AIM2 inflammasome exacerbates atherosclerosis in clonal haematopoiesis. *Nature* **592**, 296-301 (2021).

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

17. 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).

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

19. Plenge, R.M., Scolnick, E.M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* **12**, 581-94 (2013).

20. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300-1310 (2021).

21. Choi, S.W., Mak, T.S. & O'Reilly, P.F. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* **15**, 2759-2772 (2020).

22. Ge, T., Chen, C.Y., Ni, Y., Feng, Y.A. & Smoller, J.W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1776 (2019).

23. Liu, Y. *et al.* Methylomics of gene expression in human monocytes. *Hum Mol Genet* **22**, 5065-74 (2013).

24. Joehanes, R. *et al.* Gene expression analysis of whole blood, peripheral blood mononuclear cells, and lymphoblastoid cell lines from the Framingham Heart Study. *Physiol Genomics* **44**, 59-75 (2012).

25. Ågerstam, H. *et al.* IL1RAP antibodies block IL-1-induced expansion of candidate CML stem cells and mediate cell killing in xenograft models. *Blood* **128**, 2683-2693 (2016).

26. Cullinan, E.B. *et al.* IL-1 receptor accessory protein is an essential component of the IL-1 receptor. *J Immunol* **161**, 5614-20 (1998).

27. Drube, S. *et al.* The receptor tyrosine kinase c-Kit controls IL-33 receptor signaling in mast cells. *Blood* **115**, 3899-906 (2010).

28. Dinarello, C.A. Immunological and inflammatory functions of the interleukin-1 family. *Annu Rev Immunol* **27**, 519-50 (2009).

29. Mallat, Z. *et al.* Protective role of interleukin-10 in atherosclerosis. *Circ Res* **85**, e17-24 (1999).

30. Wang, P., Wu, P., Siegel, M.I., Egan, R.W. & Billah, M.M. Interleukin (IL)-10 inhibits nuclear factor kappa B (NF kappa B) activation in human monocytes. IL-10 and IL-4 suppress cytokine synthesis by different mechanisms. *J Biol Chem* **270**, 9558-63 (1995).

31. de Vries, J.E. Immunosuppressive and anti-inflammatory properties of interleukin 10. *Ann Med* **27**, 537-41 (1995).

32. Lacraz, S., Nicod, L.P., Chicheportiche, R., Welgus, H.G. & Dayer, JM IL-10 inhibits metalloproteinase and stimulates TIMP-1 production in human mononuclear phagocytes. *J Clin Invest* **96**, 2304-10 (1995).

33. de Waal Malefyt, R., Abrams, J., Bennett, B., Figdor, C.G. & de Vries, J.E. Interleukin 10(IL-10) inhibits cytokine synthesis by human monocytes: an autoregulatory role of IL-10 produced by monocytes. *J Exp Med* **174**, 1209-20 (1991).

34. Jungi, T.W., Brcic, M., Eperon, S. & Albrecht, S. Transforming growth factor-beta and interleukin-10, but not interleukin-4, down-regulate procoagulant activity and tissue factor expression in human monocyte-derived macrophages. *Thromb Res* **76**, 463-74 (1994).

35. Mertz, P.M., DeWitt, D.L., Stetler-Stevenson, W.G. & Wahl, L.M. Interleukin 10 suppression of monocyte prostaglandin H synthase-2. Mechanism of inhibition of prostaglandin-dependent matrix metalloproteinase production. *J Biol Chem* **269**, 21322-9 (1994).

36. Arai, T. *et al.* Endogenous interleukin 10 prevents apoptosis in macrophages during Salmonella infection. *Biochem Biophys Res Commun* **213**, 600-7 (1995).

37. Cohen, S.B., Crawley, J.B., Kahan, M.C., Feldmann, M. & Foxwell, B.M. Interleukin-10 rescues T cells from apoptotic cell death: association with an upregulation of Bcl-2. *Immunology* **92**, 1-5 (1997).

38. GENCARDS. (2022).

39. Forrest, A.R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-70 (2014).

40. Vlasschaert, C. *et al.* A practical approach to curate clonal hematopoiesis of indeterminate potential in human genetic datasets. *medRxiv*, 2022.10.21.22281368 (2022).

41. Svensson, E.C. *et al.* TET2-Driven Clonal Hematopoiesis and Response to Canakinumab: An Exploratory Analysis of the CANTOS Randomized Clinical Trial. *JAMA Cardiol* **7**, 521-528 (2022).

42. Fujino, T. *et al.* Mutant ASXL1 induces age-related expansion of phenotypic hematopoietic stem cells through activation of Akt/mTOR pathway. *Nat Commun* **12**, 1826 (2021).

43. de Boer, B. *et al.* Prospective Isolation and Characterization of Genetically and Functionally Distinct AML Subclones. *Cancer Cell* **34**, 674-689.e8 (2018).

44. Bonardi, F. *et al.* A proteomics and transcriptomics approach to identify leukemic stem cell (LSC) markers. *Mol Cell Proteomics* **12**, 626-37 (2013).

45. Järås, M. *et al.* Isolation and killing of candidate chronic myeloid leukemia stem cells by antibody targeting of IL-1 receptor accessory protein. *Proc Natl Acad Sci U S A* **107**, 16280-5 (2010).

46. Barreyro, L. *et al*. Overexpression of IL-1 receptor accessory protein in stem and progenitor cells and outcome correlation in AML and MDS. *Blood* **120**, 1290-8 (2012).

47. Askmyr, M. *et al.* Selective killing of candidate AML stem cells by antibody targeting of IL1RAP. *Blood* **121**, 3709-13 (2013).

48. Ågerstam, H. *et al.* Antibodies targeting human IL1RAP (IL1R3) show therapeutic effects in xenograft models of acute myeloid leukemia. *Proc Natl Acad Sci U S A* **112**, 10786-91 (2015).

49. Sano, S. *et al.* CRISPR-Mediated Gene Editing to Assess the Roles of Tet2 and Dnmt3a in Clonal Hematopoiesis and Cardiovascular Disease. *Circ Res* **123**, 335-341 (2018).

50. Collins, F.S. & Varmus, H. A new initiative on precision medicine. *N Engl J Med* **372**, 793-5 (2015).

51. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).

52. Baird, D.A. *et al.* Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLoS Genet* **17**, e1009224 (2021).

53. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* **9**, 2282 (2018).

54. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

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

56. Van Hout, C.V. *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749-756 (2020).

57. Van der Auwera, GAOC, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, (O'Reilly Media, 2020).

58. Nakao, T. *et al.* Bidirectional Mendelian randomization supports bidirectional causality between telomere length and clonal hematopoiesis of intermediate potential. *medRxiv*, 2021.02.26.21252199 (2021).

59. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-9 (2013).

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

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

62. Dawber, T.R. & Kannel, W.B. The Framingham study. An epidemiological approach to coronary heart disease. *Circulation* **34**, 553-5 (1966).

63. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. & Castelli, W.P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**, 518-25 (1975).

64. Splansky, G.L. *et al.* The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* **165**, 1328-35 (2007).

65. Liu, C. *et al.* Whole Genome DNA and RNA Sequencing of Whole Blood Elucidates the Genetic Architecture of Gene Expression Underlying a Wide Range of Diseases. *medRxiv*, 2022.04.13.22273841 (2022).

66. DeYoung, K.L. *et al.* Cloning a novel member of the human interferon-inducible gene family associated with control of tumorigenicity in a model of human melanoma. *Oncogene* **15**, 453-7 (1997).

67. Abbate, A. *et al*. Interleukin-1 and the Inflammasome as Therapeutic Targets in Cardiovascular Disease. *Circ Res* **126**, 1260-1280 (2020).

68. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605-d612 (2021).

69. Mishra, B.B. *et al.* Nitric oxide controls the immunopathology of tuberculosis by inhibiting NLRP3 inflammasome-dependent processing of IL-1β. *Nat Immunol* **14**, 52-60 (2013).

70. Guo, H., Callaway, J.B. & Ting, J.P. Inflammasomes: mechanism of action, role in disease, and therapeutics. *Nat Med* **21**, 677-87 (2015).

71. Liu, Z., Han, C. & Fu, Y.X. Targeting innate sensing in the tumor microenvironment to improve immunotherapy. *Cell Mol Immunol* **17**, 13-26 (2020).

72. Dinarello, C.A. Biologic basis for interleukin-1 in disease. *Blood* **87**, 2095-147 (1996).

73. Ridker, P.M., MacFadyen, J.G., Thuren, T. & Libby, P. Residual inflammatory risk associated with interleukin-18 and interleukin-6 after successful interleukin-1β inhibition with canakinumab: further rationale for the development of targeted anti-cytokine therapies for the treatment of atherothrombosis. *Eur Heart J* **41**, 2153-2163 (2020).

74. Ridker, P.M. *et al.* Modulation of the interleukin-6 signalling pathway and incidence rates of atherosclerotic events and all-cause mortality: analyses from the Canakinumab Anti-Inflammatory Thrombosis Outcomes Study (CANTOS). *Eur Heart J* **39**, 3499-3507 (2018).

75. Sittel, F., Filk, T. & Stock, G. Principal component analysis on a torus: Theory and application to protein dynamics. *J Chem Phys* **147**, 244101 (2017).

76. Galinsky, K.J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am J Hum Genet* **98**, 456-472 (2016).

77. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466-8 (2015).

78. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

79. Elliott, P. & Peakman, T.C. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* **37**, 234-44 (2008).

80. Kiel, M.J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109-21 (2005).

81. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174 (2019).

**Table 1. Characteristics of the study population in the UK Biobank (N= 424,203).** Study population was restricted to unrelated individuals in the UK Biobank who had exome sequencing data and were free of hematological cancer and CVD at baseline with unrelatedness defined as less than 3rd-degree relatedness.

| **Metric** [a] | **No CHIP** | **CHIP** | **P-value** [b] |
|---|---|---|---|
| | **N= 39,7937** | **N= 26,266** | |
| Age (years) | 56.2 (8.1) | 59.8 (7.1) | <0.001 |
| Male | 178,064 (44.7) | 11,712 (44.6) | 0.63 |
| White British ancestry | 333,065 (83.7) | 22,227 (84.6) | <0.001 |
| Body mass index (kg/m$^2$) | 27.4 (4.8) | 27.4 (4.6) | 0.013 |
| Ever smoked | 175,335 (44.1) | 12,926 (49.2) | <0.001 |
| Hypertension [c] | 110,675 (27.8) | 8,711 (33.2) | <0.001 |
| Hypercholesterolemia [c] | 53,162 (13.4) | 4,240 (16.1) | <0.001 |
| Type 2 diabetes [c] | 8,186 (2.1) | 702 (2.7) | <0.001 |
| Among CHIP carriers: | | | |
| *DNMT3A* | NA | 14,565 (55.5) | NA |
| Large *DNMT3A* [d] | NA | 6,104 (23.2) | NA |
| *TET2* | NA | 5,217 (19.9) | NA |
| Large *TET2* | NA | 2,259 (8.6) | NA |
| *ASXL1* | NA | 2,422 (9.2) | NA |
| Large *ASXL1* | NA | 1,198 (4.6) | NA |
| *JAK2* | NA | 235 (0.9) | NA |
| Large *JAK2* | NA | 228 (0.9) | NA |

[a] Metrics are represented as mean (standard deviation) for continuous variables and % (n) for categorical variables.

[b] P values calculated with a 2-sample t-test for continuous traits or Fisher exact test for categorical traits.

[c] Clinical conditions are those occurring prior to enrollment.

[d] Large CHIP is defined as VAF >10%.

CHIP, clonal hematopoiesis of indeterminate potential; CVD: cardiovascular disease; VAF: variant allele fraction.

**Table 2. Associations between CHIP mutation and incidence of CVD event.** CVD event outcome is defined as a composite of myocardial infarction, coronary artery disease or revascularization, stroke, or death. b Models were adjusted for age at the time of enrollment, sex, white British ancestry, body mass index, diagnoses of type 2 diabetes mellitus at the time of enrollment, diagnoses of hypertension at the time of enrollment, ever-smoker status, and the first ten principal components of genetic ancestry. Participants with prevalent hematological cancers or CVD were removed from the analyses.

| | **Presence of CHIP** | | **Presence of large CHIP** [a] | |
|---|---|---|---|---|
| | **HR (95% CI)** | **P-value** | **HR (95% CI)** | **P-value** |
| **CHIP** | 1.18 (1.14, 1.22) | $1.5\times10^{-21}$ | 1.29 (1.24, 1.35) | $8.6\times10^{-29}$ |
| ***DNMT3A*** | 1.06 (1.01, 1.11) | 0.01 | 1.13 (1.06, 1.21) | $3.9\times10^{-4}$ |
| ***TET2*** | 1.11 (1.03, 1.19) | $4.5\times10^{-3}$ | 1.28 (1.16, 1.41) | $5.8\times10^{-7}$ |
| ***ASXL1*** | 1.41 (1.29, 1.54) | $3.5\times10^{-14}$ | 1.52 (1.35, 1.71) | $4.1\times10^{-12}$ |
| ***JAK2*** | 2.81 (2.25, 3.51) | $8.5\times10^{-20}$ | 2.80 (2.23, 3.50) | $3.3\times10^{-19}$ |
| ***PPM1D*** | 1.39 (1.18, 1.64) | $7.6\times10^{-5}$ | 1.46 (1.13, 1.88) | $3.9\times10^{-3}$ |
| ***TP53*** | 1.43 (1.18, 1.72) | $2.2\times10^{-4}$ | 2.04 (1.57, 2.65) | $1.0\times10^{-7}$ |
| ***SRSF2*** | 2.60 (2.18, 3.09) | $6.8\times10^{-27}$ | 3.32 (2.72, 4.04) | $1.1\times10^{-32}$ |
| ***SF3B1*** | 1.47 (1.14, 1.89) | $2.9\times10^{-3}$ | 1.55 (1.16, 2.06) | $2.9\times10^{-3}$ |

[a] Large CHIP is defined as VAF >10%.

CHIP, clonal hematopoiesis of indeterminate potential; CVD: cardiovascular disease; VAF: variant allele fraction.

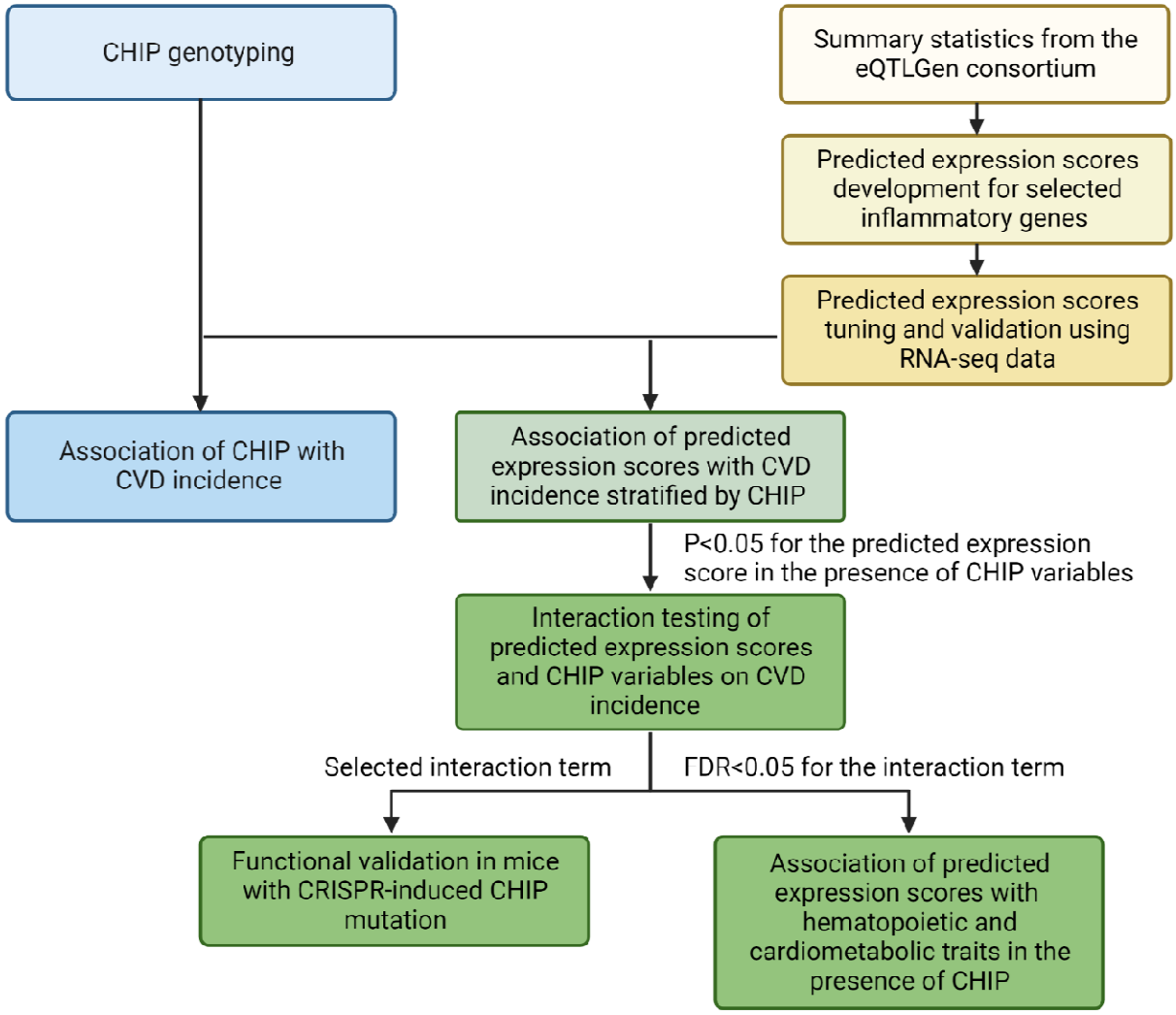


**Figure 1. Study schematics.** CHIP was identified using whole exome sequencing data of blood DNA. Predicted expression score for inflammatory genes was developed based on *cis*-expression quantitative trait locus results and validated using measured RNA-sequencing data, and then examined whether they modified CHIP-associated CVD risk. Predicted expression scores that significantly modify CHIP-associated CVD risk were further validated in mice model and evaluated for their associations with hematopoietic and cardiometabolic traits. CHIP: clonal hematopoiesis of indeterminate potential; CVD: cardiovascular disease; FDR: false discovery rate; HR: hazard ratio.

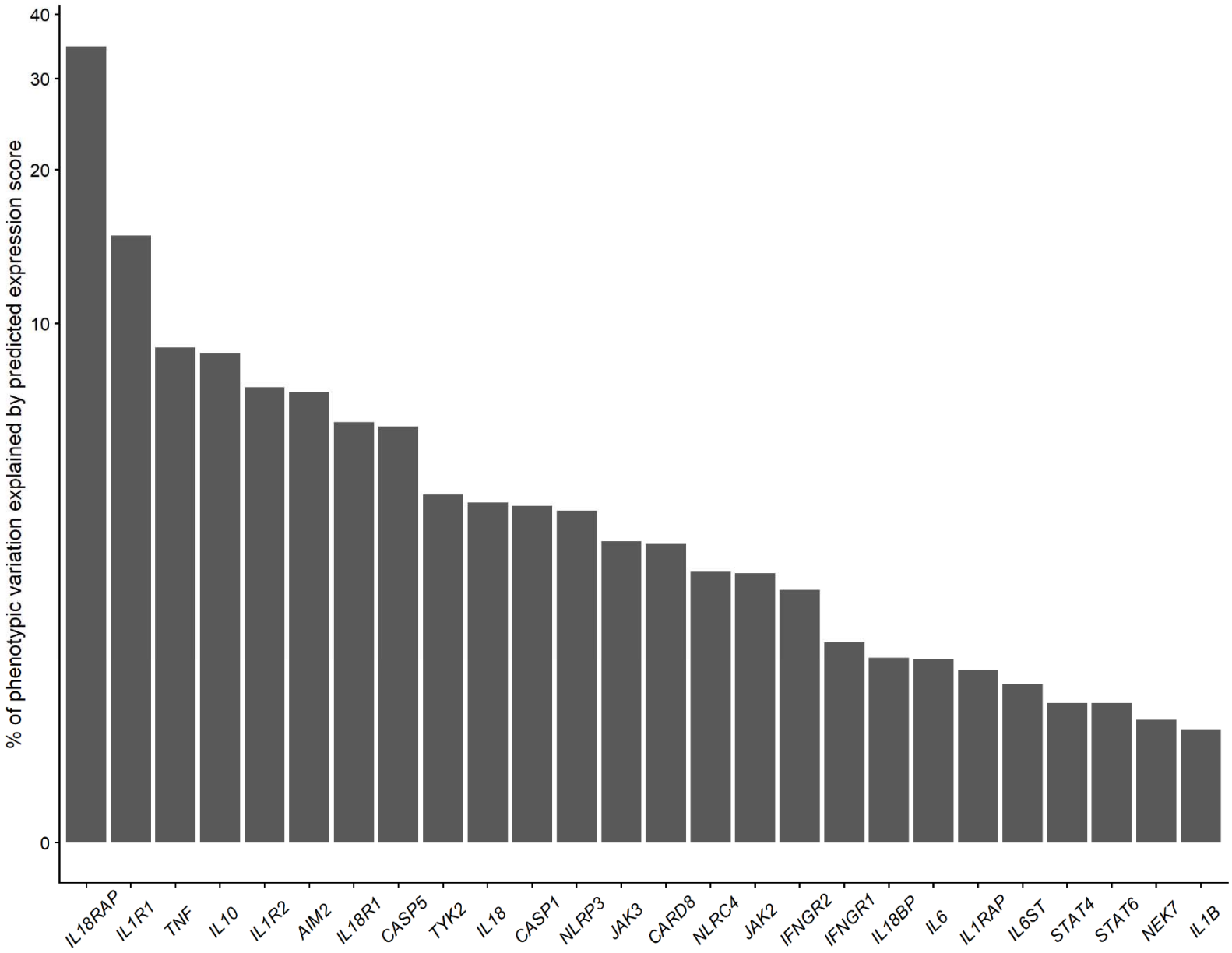


**Figure 2. Proportion of the variance of experimentally measured expression levels that can be explained by predicted expression scores for inflammatory genes among participants with European ancestry.** Inflammatory genes were identified through canonical pathways and protein-

protein interactions based on STRING (https://string-db.org/). Predicted expression scores of examined genes were calculated by applying either P+T or PRS-CS methods to the summary statistics of the eQTL for those genes from the eQTLGen consortium (https://www.eqtlgen.org/) and validated using experimental measured RNA-sequencing data in the Multi-Ethnic Study of Atherosclerosis (peripheral blood mononuclear cells [PBMC]) and Framingham Heart Study (whole blood). Since the eQTL source data was from either PBMC or whole blood, we report the largest $R^2$ of the measured transcriptome levels in either FHS or MESA. eQTL: expression quantitative trait loci; PRS-CS: polygenic risk score-continuous shrinkage; P+T: pruning and thresholding

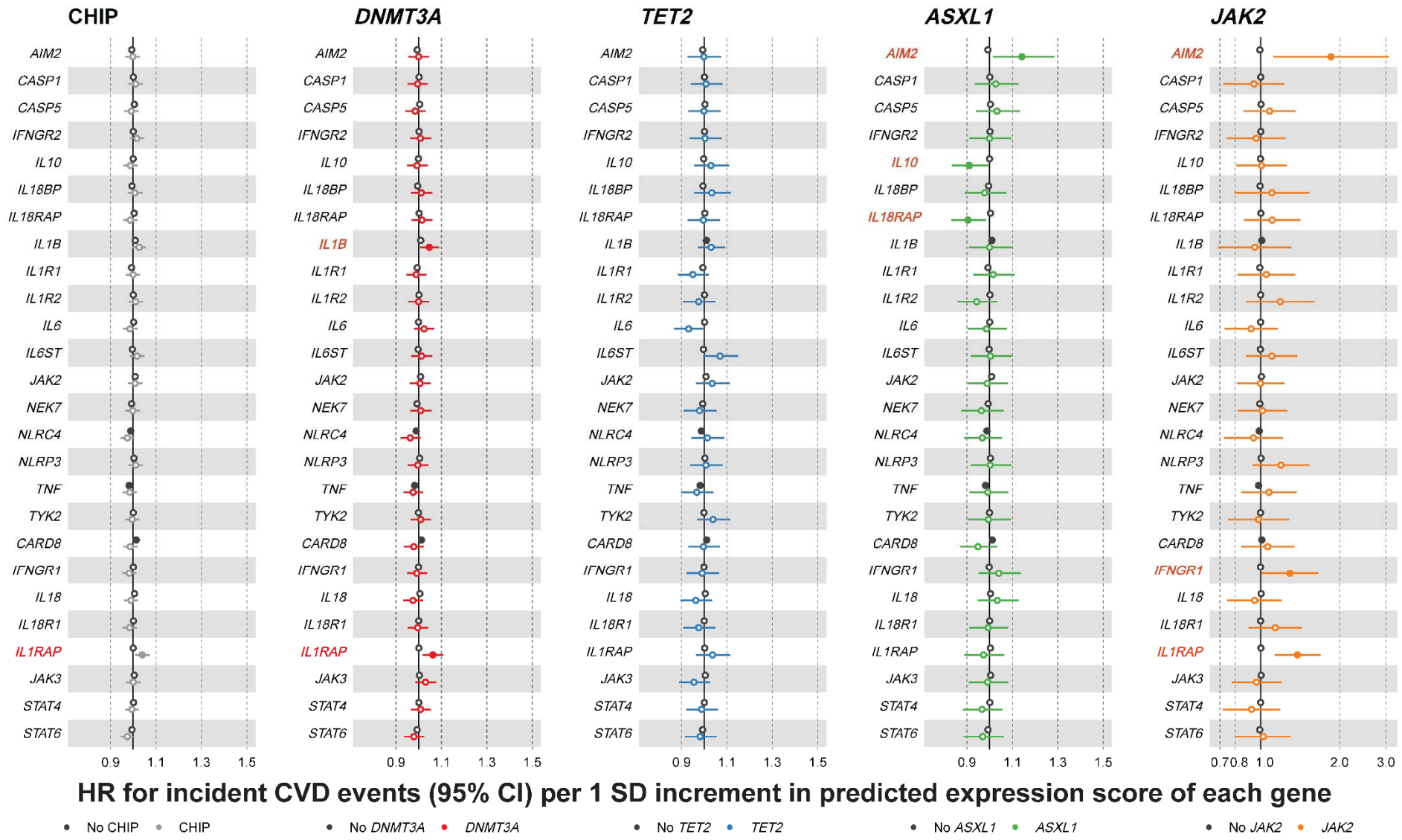


**Figure 3. HR of predicted expression scores of inflammatory genes on CVD event incidence stratified by CHIP mutation status.** Inflammatory genes were identified through canonical pathways and protein-protein interactions based on STRING (https://string-db.org/). Predicted expression scores of examined genes were calculated by applying either P+T or PRS-CS methods to the summary statistics of the eQTL for those genes from the eQTLGen consortium (https://www.eqtlgen.org/) and validated using experimental measured RNA-sequencing data in the Multi-Ethnic Study of Atherosclerosis (peripheral blood mononuclear cells) and Framingham Heart Study (whole blood)**.** CVD event outcome is defined as a composite of

myocardial infarction, coronary artery disease or revascularization, stroke, or death. Black color indicates the absence of CHIP mutations, and non-black colors indicate the presence of CHIP mutations. A solid circle indicates a significant association at $P<0.05$ level. Red colored gene name indicates significant association between corresponding expression score and CVD outcome in the presence of CHIP mutation at $P<0.05$ level. CHIP, clonal hematopoiesis of indeterminate potential; CVD, cardiovascular disease; eQTL: expression quantitative trait loci; HR: hazard ratio; PRS-CS: polygenic risk score-continuous shrinkage; P+T: pruning and thresholding

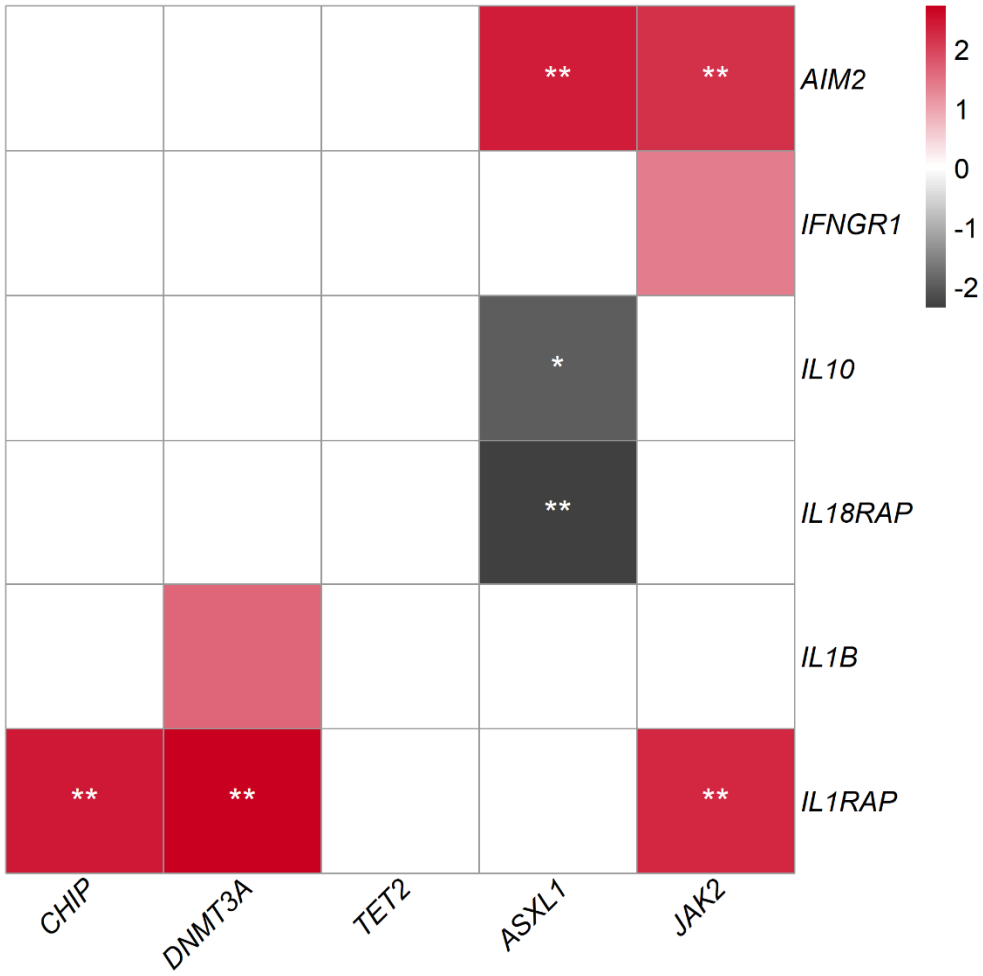


**Figure 4. Heatmap for the Z-scores of interactions between CHIP mutations and predicted expression scores of inflammatory genes on CVD event incidence.** Only predicted expression scores significantly associated with CVD event incidence among participants with CHIP mutations were examined for their interactions in this step. Inflammatory genes were identified through canonical pathways and protein-protein interactions based on STRING (https://string-db.org/). Predicted expression scores of examined genes were calculated by applying either P+T or PRS-CS methods to the summary statistics of the eQTL for those genes from the eQTLGen consortium (https://www.eqtlgen.org/)**.** CVD event outcome is defined as a composite of myocardial infarction, coronary artery disease or revascularization, stroke, or death. Black color indicates a negative Z-score, and red indicates a positive Z-score. Two white stars indicate statistical significance at an FDR=0.05 level; single white star indicate statistical significance at a FDR=0.1 level. The darker the color, the stronger the effects. CHIP, clonal hematopoiesis of indeterminate potential; CVD, cardiovascular disease; eQTL: expression quantitative trait loci**;** FDR: false discovery rate; PRS-CS: polygenic risk score-continuous shrinkage; P+T: pruning and thresholding

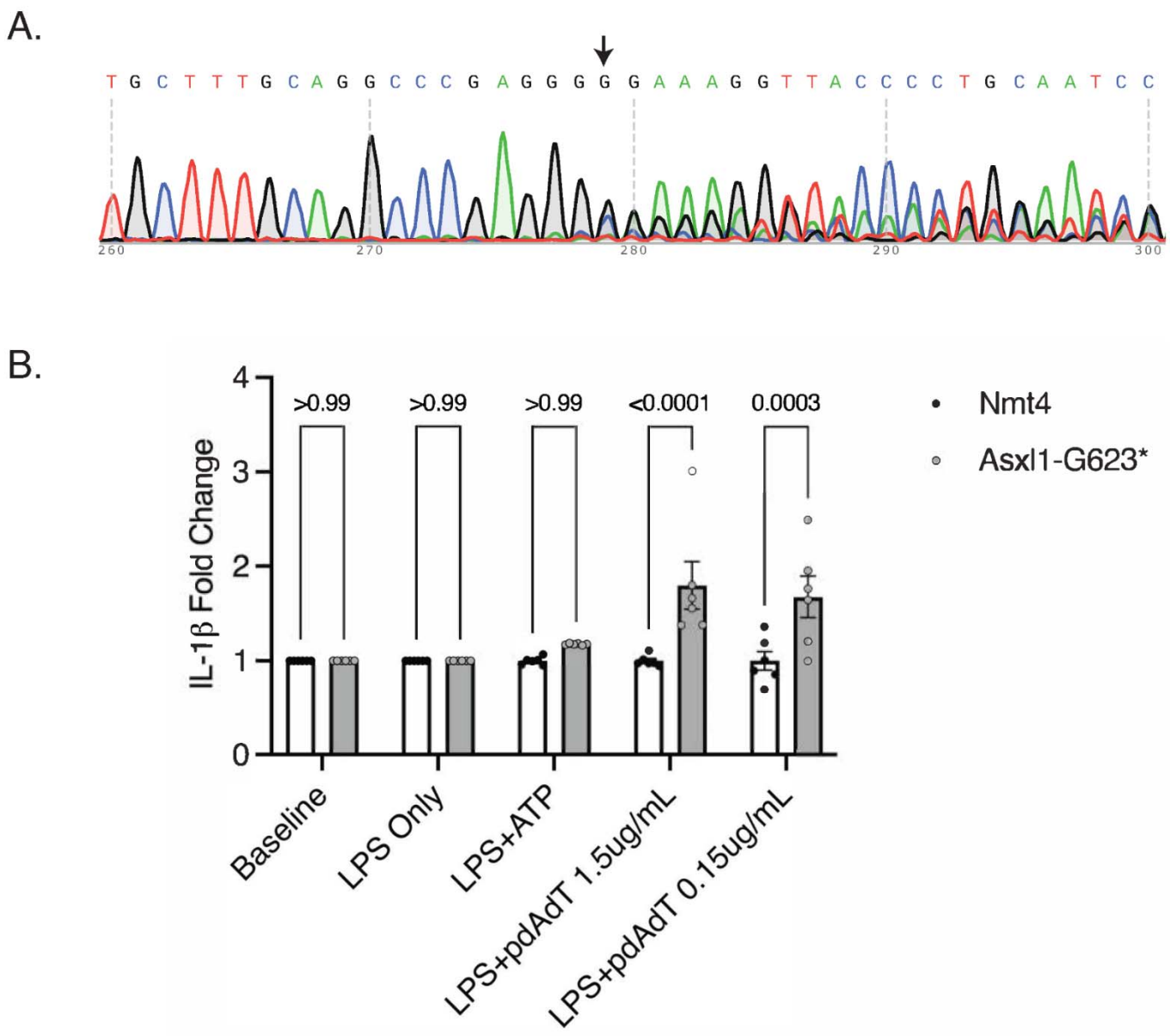


**Figure 5. Inflammasome activation in bone marrow derived macrophages harboring *Asxl1* mutations.** Bone Marrow Derived Macrophages (BMDMs) were harvested from mice harboring a mixture of wild-type control (Nmt4) or *Asxl1* mutated bone marrow (Asxl1-G623*). A. Sanger Sequencing of Cas9 transgenic murine fibroblast transfected with lentiviruses containing *Asxl1*-guides targeting exon 12, arrow indicates target site.

B. BMDM inflammasome activation marked by IL-1β in supernatant. BMDMs were primed with LPS for potentiating IL-1β. ATP was used for stimulating NLRP3 inflammasome and pdAdT was used for stimulating AIM2 inflammasome. Data are mean±SE of fold change. Two-way ANOVA followed by Bonferroni post hoc test in B.