Germline genomic and phenomic landscape of clonal hematopoiesis in 323,112 individuals

Md Mesbah Uddin[*1,2], Zhi Yu[*1,2], Joshua S. Weinstock[3], Tetsushi Nakao[1,2,4,5], Abhishek Niroula[4,6,7], Sarah M. Urbut[1,8], Satoshi Koyama[1,2], Seyedeh M. Zekavat[1,2,9], Kaavya Paruchuri[1,2,10], Alexander J. Silver[11], Taralynn M. Mack[12], Megan Y. Wong[1,2], Sara M. Haidermota[1,2], Romit Bhattacharya[1,2,8], Saman Doroodgar Jorshery[13-15], Michael A. Raddatz[11], Michael C. Honigberg[1,2,8], Whitney E. Hornsby[16], Martin Jinye Zhang[13,17], Vijay G. Sankaran[6,18-20], Gabriel K. Griffin[21-23], Christopher J. Gibson[4], Hailey A. Kresge[11], Patrick T. Ellinor[1,2], Veterans Affairs' Million Veteran Program[24], Kelly Cho[25,26], Yan V. Sun[27,28], Peter W.F. Wilson[27,29], Saiju Pyarajan[25,26], Giulio Genovese[13,30,31], Yaomin Xu[32,33], Michael R. Savona[11], Alexander P. Reiner[34,35], Siddhartha Jaiswal[36,37], Benjamin L. Ebert[4,6,38], Alexander G. Bick[#12], and Pradeep Natarajan[#1,2,10]

[*]Contributed equally to this work
[#]Jointly supervised this work
Correspondence: alexander.bick@vumc.org (A.G.B.) and pnatarajan@mgh.harvard.edu (P.N.)

1. Program in Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
2. Cardiovascular Research Center and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
3. Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA
4. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
5. Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
6. Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
7. Department of Laboratory Medicine, Lund University, Lund, Sweden
8. Cardiology Division, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
9. Computational Biology & Bioinformatics Program, Yale University, New Haven, CT, 06510, USA
10. Department of Medicine, Harvard Medical School, Boston, MA, USA
11. Vanderbilt University School of Medicine, Nashville, TN, USA
12. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
13. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

14. Department of Computer Science, University of Toronto, Toronto, Canada
15. Department of Computer Science and Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
16. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
17. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
18. Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, USA
19. Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA
20. Harvard Stem Cell Institute, Cambridge, USA
21. Department of Pathology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
22. Division of Hematopathology, Department of Pathology, Brigham and Women's Hospital, Boston MA 02115, USA
23. Epigenomics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA
24. A full list of authors appears in the Supplementary Note
25. Veterans Affairs Boston Healthcare System, Boston, MA, USA
26. Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
27. Veterans Affairs Atlanta Healthcare Systems, Decatur, Georgia, USA
28. Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, Georgia, USA
29. Department of Medicine, Emory University School of Medicine, Atlanta, Georgia, USA
30. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
31. Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
32. Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
33. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
34. Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle WA 98109, USA
35. Department of Epidemiology, University of Washington, Seattle WA 98195, USA
36. Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA
37. Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA
38. Howard Hughes Medical Institute, Boston, MA, USA

Abstract

With age, acquired mutations can cause clonal expansion of hematopoietic stem cells (HSC). This clonal hematopoiesis of indeterminate potential (CHIP) leads to an increased predisposition to numerous diseases including blood cancer and cardiovascular disease. Here, we report multi-ancestry genome-wide association meta-analyses of CHIP among 323,112 individuals (19.5% non-European; 5.3% have CHIP). We identify 15 genome-wide significant regions and nominate additional loci through multi-trait analyses, and highlight variants in genes involved in self-renewal and proliferation of HSC, telomere maintenance, and DNA damage response pathways. We then use Mendelian randomization to establish a causal relationship between CHIP and coronary artery disease. Next, we systematically profile consequences of CHIP across the phenome, which revealed strong associations with hematopoietic, neoplastic, and circulatory conditions corroborated by polygenic enrichment of CHIP loci in immune cells and cardiomyocytes. These findings expand the genomic and phenomic landscape of CHIP.

Introduction

Self-renewing cell populations accumulate somatic mutations with aging, although most of these mutations are without functional consequence. In rare cases, these somatic mutations confer a selective advantage leading to clonal expansion[1-3]. In the case of hematopoietic stem cells, driver mutations in genes with diverse functions, including DNA methylation (*DNMT3A*, *TET2*)[4,5], RNA splicing (*SF3B1*, *U2AF1*)[6], chromatin remodeling (*ASXL1*)[7], and DNA damage response (*TP53*, *PPM1D*)[8-10] can lead to a clonal expansion of hematopoietic stem cells termed clonal hematopoiesis of indeterminate potential (CHIP) when such mutations make up >4% of peripheral blood cells (variant allele fraction, VAF≥2%). CHIP is the pre-cancerous precursor lesion for myeloid hematologic malignancy[11-13], but numerous studies in human and model systems have linked CHIP to diverse diseases of aging, including coronary artery disease[14,15], stroke[16], heart failure[17], chronic obstructive pulmonary disease[18], osteoporosis[19], and chronic liver disease[20]. However, CHIP is less commonly observed among patients with Alzheimer's disease[21]. Characterizing the germline genetic determinants of CHIP offers the opportunity to prioritize unifying features of multiple diseases of aging.

We previously performed a genome-wide association study (GWAS) of CHIP and identified three genomic regions associated with CHIP risk in a study of 3,831 CHIP cases and 61,574 controls from whole-genome sequencing of blood DNA utilizing the National Heart, Lung and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) study[22]. This effort identified three germline genetic variants, including one in the *TERT* promoter, the intronic region of *TRIM59*, and a distal enhancer of *TET2* specific to individuals of African ancestry. These findings enabled further work demonstrating that the low-frequency germline variant at *TET2* leads to locally altered methylation and decreased germline *TET2* expression, subsequently promoting the self-renewal and proliferation of hematopoietic stem cells[22]. Furthermore, bidirectional Mendelian randomization analyses of leukocyte telomere length (LTL), *TERT* locus variants, and CHIP highlight the dynamic nature of genomic models across the lifespan[23].

Here, we analyze 17,044 CHIP cases and 306,068 controls across four cohorts, including 63,442 (19.5%) individuals of non-European ancestry from whole-genome or whole-exome sequencing of blood DNA. Using this expanded dataset, we perform genome-wide discovery analyses and fine-mapping at both variant[24] and gene levels[25]. We identify multiple components of the DNA damage response pathways[26,27] that lead to CHIP, unifying epidemiological observations regarding enhanced CHIP risk in populations with specific environmental or cytotoxic exposures[8,28,29]. Multivariate Bayesian analyses with myeloproliferative neoplasm (MPN)-associated alleles[30] further prioritize several additional loci. Finally, we leverage this expanded dataset to systematically profile the disease risk of CHIP compared to another form of clonal hematopoiesis—mosaic chromosomal alterations (mCAs), and further investigate the disease associations by examining cell types[31] enriched by CHIP (**Extended Data Fig. 1**).

Results

Baseline characteristics

We analyzed 355,183 individuals across four cohorts, UK Biobank (UKB; N=200,128), TOPMed (N=87,116), Vanderbilt BioVU (N=54,583), and Mass General Brigham Biobank (MGBB; N=13,356) for CHIP mutations using previously described methods (see **Methods**). These individuals are of European (N=287,991), African (N=31,900), Asian (N=14,073), Hispanic (N=13,939), and other or unknown (N=7,280) ancestry (**Supplementary Table 1**). We identified 20,302 CHIP mutations in 18,499 individuals (5.2%). Consistent with previous reports[22], CHIP was robustly associated with age (**Fig. 1a**) and had a similar distribution of genes identified in prior studies. Across all cohorts, 90.8% of individuals with CHIP driver mutations had only one identified mutation (**Fig. 1b**). *DNMT3A*, *TET2*, and *ASXL1* are the most frequent mutated genes, accounting for more than 75% of CHIP mutations (**Fig. 1c**). Compared with other cohorts, MGBB has a relatively lower proportion of individuals with *DNMT3A* mutations but a relatively larger proportion with *TET2* mutations. The next seven most frequent genes are *PPM1D*, *TP53*, *JAK2*, *SF3B1*, *SRSF2*, *GNB1*, and *CBL*. For most of these top mutated genes, the corresponding VAF for participants in TOPMed are larger than that seen in other cohorts, which is likely explained by technical sensitivities between whole-genome (TOPMed) and whole-exome (UKB and MGBB) sequencing-based CHIP detection. The median VAF of *DNMT3A* was 0.08 in UKB, 0.13 in TOPMed, and 0.09 in MGBB; the median VAF of *TET2* was 0.11 in UKB, 0.16 in TOPMed, and 0.09 in MGBB. *JAK2* has the largest clonal fraction among the top ten genes, with a median VAF of 0.30 in UKB, 0.21 in TOPMed, and 0.32 in MGBB (**Fig. 1d**). In BioVU, CHIP was identified in four driver genes, *DNMT3A*, *JAK2*, *TET2*, and *ASXL1*, using a customized genotyping array (see Methods). The most common mutations in BioVU samples were *JAK2* V617F and *DNMT3A* R882C/H.

Fifteen genome-wide significant loci associated with CHIP categories

We performed multi-ancestry meta-analyses of overall CHIP (pooled sample size, 17,044 cases, and 306,068 controls), *DNMT3A* (8,949 cases and 307,971 controls), and *TET2* (2,851 cases and 307,527 controls) CHIP GWAS from TOPMed, UKB, MGBB, and BioVU. Here, we performed three GWAS for each of the four participating cohorts, followed by multi-ancestry meta-analyses using the fixed-effects inverse-variance-weighted approach (see **Methods**). A total of 323,112 participants and 21,455,227 variants (minor allele frequency (MAF)>=0.1%; variants present in >=two studies) remained after variant filtering and quality-control procedures. There was no evidence of artificial inflation of association statistics due to the population structure (genomic control factor $\lambda_{GC} = 1.05$ for CHIP and *DNMT3A*, 1.06 for *TET2;* **Fig. 2a-c**). The SNP heritability (h^2_{SNP}) of overall CHIP among the entire study population was estimated at 3.5% (SD=0.002) on the observed scale and 15.8% (SD=0.011) on the liability scale using the BLD-LDAK model (see **Methods** and **Supplementary Table 2**), representing a four-fold increase from the previous estimate[22].

We discovered ten loci associated with overall CHIP at genome-wide significance after correcting for multiple testing ($P < 1.67 \times 10^{-8}$, i.e., $5.0 \times 10^{-8}/3$). Seven of the loci are new and mapped to the following genes based on proximity of the lead variants: *PARP1*, *CD164/SMPD2*, *ATM*, *ITPR2*, *MSI2*, *SETBP1,* and *CHEK2* (**Fig. 2a** and **Supplementary Tables 3,4**). Consistent with previous report[22], the strongest signal remained at the *TERT* locus, with rs7705526-A as the lead variant (odds ratio (OR) (95% confidence interval (CI)) = 1.26 (1.23-1.29); $P=2.6 \times 10^{-81}$). Regional association plots are shown in **Extended Data Fig. 2**. The effects of associated SNPs were largely homogeneous across studies (i.e., heterogeneity $P>0.05$, **Supplementary Tables 3,4**). Summary-statistics-based conditional analysis yielded 12 independent SNPs that reached genome-wide significance at $P < 1.67 \times 10^{-8}$ (**Supplementary Table 5**).

CHIP driver gene-specific association analyses identified ten loci associated with *DNMT3A* and six loci associated with *TET2* (**Fig. 2bc, Extended Data Fig. 3,4, Supplementary Tables 4, 6-9**). *TCL1A* is a newly discovered locus that significantly increased the risk for *DNMT3A* CHIP (rs2887399-T; OR (95% CI) = 1.19 (1.15-1.23); $P=1.26 \times 10^{-22}$) and significantly reduced the risk for *TET2* CHIP (rs4900291-T; OR (95% CI) = 0.78 (0.73-0.84); $P=2.92 \times 10^{-11}$).

Interestingly, there was a larger overlap of loci associated with overall CHIP and *DNMT3A* CHIP: nine of the ten loci were common between the two CHIP categories (**Fig. 2d**). CHIP-associated loci ($P < 1.67 \times 10^{-8}$) also overlapped with LTL[32], expanded mCAs[33], and MPN[30] associated genome-wide significant loci ($P < 5 \times 10^{-8}$). Here, seven, five, and three of the CHIP-associated loci overlapped with LTL, expanded mCAs, and MPN-associated loci, respectively (**Fig. 2d**).

We also conducted a joint multivariate analysis across MPN[30], overall CHIP, *DNMT3A*, and *TET2* using empirical Bayes hierarchical modeling[34] (see **Methods**). This approach yielded a 3-fold boost in power for discovery. Compared with conventional GWAS, we identified 39, 39, and 46 more genetic regions for overall CHIP, *DNMT3A*, and *TET2* CHIP, respectively, with evidence for the association through this method (**Extended Data Fig. 5**).

Statistical fine-mapping identified causal variants at *ATM* and *PARP1*

Statistical fine-mapping was conducted using the results of the European meta-analysis and LD matrix derived from the UKB European population to identify the causal variants in the associated loci (see **Methods**). We identified 27 variants with posterior inclusion probability (PIP)> 10% (**Supplementary Table 10**), among which a missense variant, rs1800057-G, in the newly discovered *ATM* gene (ENST00000675843.1:c.3161C>G:p.Pro1054Arg; OR (95% CI) =1.28 (1.18-1.37); $P=1.3 \times 10^{-10}$) showed higher posterior probability of inclusion (PIP 22%) in the locus (**Fig. 3a**). The risk allele for this coding variant was predicted to be deleterious by various functional annotation tools (**Fig. 3b**). Decreased function of the *ATM* gene and increased risk for CHIP aligns with the prior observation that the loss of function and low expression of *ATM* were associated with tumorigenesis and worse prognosis in various cancers, including MPN[35,36]. In another newly discovered locus at *PARP1*, we identified a non-coding variant, rs1527365-T, with

high posterior probability (PIP=33%; OR (95% CI) = 0.88 (0.85-0.91); P=7.0 × 10^{-13}), which is the lead variant in European meta-analysis of CHIP GWAS (**Fig. 3c**). The T allele at rs1527365 is protective for overall CHIP and *DNMT3A* CHIP (**Supplementary Tables 3,6**) and associated with decreased expression of *PARP1* in whole blood (**Fig. 3d**; normalized effect size=-0.067, P=1.7 × 10^{-4}, GTEx v8). This is consistent with the observations that a higher expression of *PARP1* is a marker for worse clinical outcomes in various tumors, including myeloid leukemia, and that inhibition of *PARP1* has anti-tumor effects[37]. rs1527365 is also a cis-methylation QTL where the T allele is associated with increased DNA methylation at CpGs around *PARP1* (data from GoDMC[38]:http://mqtldb.godmc.org.uk/), further supporting the negative regulation of *PARP1* by this allele.

Similarity-based gene prioritization identified several putative causal genes

Next, to prioritize causal genes from the GWAS data, we incorporated a novel gene prioritization method, Polygenic Priority Score (PoPS)[25]. PoPS combines GWAS data with other omics data (e.g., gene expression, biological pathway, and predicted protein-protein interaction) to prioritize casual genes based on functional similarity. We applied PoPS to summary-level data from the UKB GWAS results of overall CHIP, *DNMT3A*, and *TET2* using European ancestry individuals from the UKB study population as a reference panel. Features from gene expression data, protein-protein interaction networks, and pathway membership that passed a marginal feature selection step were included in the final predictive model (see **Methods**). We identified 26 prioritized gene-trait pairs. Top putative causal genes prioritized by PoPS (PoPS score>0.3) included *PARP1*, *ITF80*, *TET2*, *TERT*, *SMPD2, ATM*, *ITPR2*, *MSI2,* and *CHEK2* for overall CHIP (**Supplementary Table 11**), *PARP1*, *ITF80*, *TERT*, *SMPD2, ATM*, *ITPR2*, *TCL1A, SETBP1,* and *CHEK2* for *DNMT3A* CHIP (**Supplementary Table 12**), and *AGTRAP, BDKRB2,* and *ZFPM1* for *TET2* CHIP (**Supplementary Table 13**). This prioritization corroborates with the fine-mapped causal variants at *PARP1*, *TERT*, and *ATM* as well as previously validated causal variants at *TET2*[22] *and TCL1A*[39]. Further, the consistency between prioritized genes by PoPS (similarity-based approach) and distance (non-similarity-based approach) suggest high confidence[25].

Mendelian randomization refined causality between CHIP and coronary artery disease

Leveraging the large CHIP meta-analysis, we used Mendelian randomization (MR) to understand the causality of CHIP and disease outcomes. Since CHIP has been associated with coronary artery disease (CAD) incidence in humans and atherosclerosis in a *Tet2* CHIP mouse model[14,40,41], we performed MR analyses of overall CHIP, *DNMT3A*, and *TET2* with CAD. While genetic predispositions for CHIP overlapped with those for LTL, previous MR studies support inverse causality for LTL on CAD, implying complex causal relationships between CHIP-LTL, and LTL-CAD[23]. Thus, we carefully selected instrumental variables (IVs) to avoid clear violations of MR assumptions by excluding IVs 1) associated with known confounders between CHIP and CAD, including hypercholesterolemia, hypertension, type 2 diabetes, body mass index, smoking

status, or 2) having more robust associations with LTL than with CAD, and/or 3) having inverse effect sizes for LTL, CHIP, and CAD, accounting for the inverse causal association between LTL and CAD. Mendelian randomization using the robust adjusted profile score (MR-RAPS) was used given the limited power for CHIP analyses due to low heritability, especially the gene-specific analyses. MR-RAPS is a novel method that accommodates weak instruments in the MR framework. The approach was used as our primary method, and a liberal p-value threshold was applied for IV selections (see **Methods**)[42]. Using summary statistics of multi-ancestry meta-analysis for CHIP, we observed significant positive causal effects of genetically-determined *TET2* CHIP mutation risk on CAD risk (P=3.2 × 10^{-4}). In contrast, the overall CHIP and *DNMT3A* CHIP mutation results were null, consistent with differential causal effects on CAD risk by CHIP mutated gene types (**Fig. 4**). However, sensitivity analyses using other MR approaches that do not allow for weak instruments yielded null results across overall CHIP, *DNMT3A*, and *TET2* CHIP mutations (**Extended Data Fig. 6 and Supplemental Tables 14-16)**.

Phenome-wide scans yielded different patterns between CHIP and mCAs

The dataset assembled now permits a comprehensive and well-powered phenome-wide association study (PheWAS) of CHIP. At the same time, we contrast CHIP with a different age-associated clonal hematopoietic phenomenon, mosaic chromosomal alterations (mCAs) (see **Methods**). We not only compared the CHIP and mCAs in the same study population where information on both is available, but we also examined the mCAs in an extended dataset (N= 1,116,579) as a secondary analysis to maximize the discovery power.

Among incident hematopoietic and neoplastic conditions (**Fig. 5a and Supplementary Table 17**), significant associations were observed with leukemias [CHIP: hazard ratio (HR) =1.78, 95% CI 1.42 - 2.24, FDR=1.76× 10^{-4}; autosomal mCAs: HR=6.87, FDR=5.81 × 10^{-89}], with the largest effect for expanded CHIP (i.e., VAF>10%) and myeloid leukemias (HR 11.56, 95% CI 6.67 - 20.06, FDR=3.67× 10^{-15}) and for expanded autosomal mCAs and lymphoid leukemias (HR 78.36, 95% CI 57.55 - 106.69, FDR=5.91 × 10^{-166}), as shown previously[2,43]. Besides hematopoietic malignancies, neoplastic associations were also detected with incident respiratory cancer (Expanded CHIP: HR 1.52, 95% CI 1.27 - 1.81, FDR=4.05× 10^{-4}), malignancies of the brain for CHIP (HR=1.84, 95% CI 1.31 - 2.58, FDR=0.025), and skin cancer for mCAs (Expanded autosomal mCAs: HR 1.45, 95% CI 1.15 - 1.82, FDR=0.024).

For CHIP, across other non-hematopoietic/neoplastic conditions, significant associations were observed with circulatory, gastrointestinal, genitourinary, renal, and infectious conditions, some of which showed consistent associations with mCAs (**Fig. 5b and Supplementary Table 17**). In particular, among circulatory conditions, association with incident arterial embolism and thrombosis was detected for expanded *TET2* (HR=2.97, 95% CI 1.58 - 5.59, FDR=0.03). While no significant association for mCAs was detected among the same sample set analyzed for CHIP, a significant association with incident arterial embolism and thrombosis was detected for expanded mCAs (HR=1.17, 95% CI 1.08 - 1.27, FDR=0.0017) (**Supplementary Table 18 and Extended**

Data Fig. 7), particularly for expanded ChrX mCAs in females (HR=3.92, 95% CI 1.92 - 8.02, FDR=0.029). Additionally, significant associations with cardiomyopathy for CHIP (HR 1.64, 95% CI 1.30 - 2.07, FDR=0.0026) but not for mCAs were observed (**Supplementary Table 18 and Extended Data Fig. 7**). The following were other notable CHIP-specific cardiovascular associations: hypertension, pulmonary heart disease, and lower extremity varicose veins. While CHIP was not strongly associated with CAD in the PheWAS analysis, using solely combined ICD-9 and ICD-10 codes per the 'phecode' algorithm, further analyses using a more comprehensive CAD phenotype (composite of myocardial infarction, coronary revascularization, stroke, and death) yielded significant associations between CHIP and incident CAD (**Supplementary Table 19**).

Other than circulatory conditions (**Fig. 5b and Supplementary Table 18**), several additional incident phenotypic associations were observed. CHIP-specific associations included peptic ulcer disease and diverticulosis. mCAs and CHIP were similarly associated with acute renal failure and pneumonia, and mCAs had a larger association with splenomegaly. Additionally, mCAs were more strongly associated with sepsis and viral infections, while CHIP was more strongly associated with bacterial infections.

Polygenic associations of CHIP at a single-cell level corroborated PheWAS findings

We used scDRS[31] to assess the polygenic enrichment of the overall CHIP, *DNMT3A*, and *TET2* GWASs in different cell populations in 2 single-cell RNA sequencing (scRNA-seq) data sets, namely the Tabula Muris Senis (TMS) mouse cell atlas[44] and Tabula Sapiens (TS) human cell atlas[45]. scDRS assesses excess expression of GWAS-associated genes in single-cell data; we considered FDR<0.3 for significant association and P<0.01 for suggestive association (see **Methods**). Results are reported in **Figure 6**. We determined the polygenic signal in the CHIP GWAS is significantly more highly expressed in ventricular myocytes and atrial myocytes from the heart (P<0.001, and P=0.004, respectively) and neuronal stem cells and oligodendrocytes from the brain non-myeloid tissue (P=0.004, and 0.01 respectively). *DNMT3A* is suggestively more highly expressed in ventricular myocytes from the heart (P=0.003). *TET2* is significantly more highly expressed in DN4 thymocytes from the thymus (P=0.002) and suggestively in T cells (P=0.007; combined from trachea, heart, brown adipose tissue, and limb muscle). Notably, the polygenic enrichment findings corroborate our PheWAS results and the existing literature[14,21]. The results are also consistent between TMS and TS (**Extended Data Fig. 8a**) and between cell type-level and tissue-level associations (**Extended Data Fig. 8b,c**). In addition, we assessed the robustness of the results by subsampling the TMS data and subsampling the putative trait gene sets. We determined that the results are consistent across the subsampling experiments (**Extended Data Figure 9**).

Discussion

In this large and diverse genetic study of CHIP, we identified ten genome-wide significant regions, including seven that are new and one that is driver gene-specific. We highlight variants in *ATM* and *PARP1* as causal CHIP variants leading to decreased DNA damage repair and a number of putative causal genes. We systematically profiled CHIP consequences phenome-wide and highlighted strong associations with hematopoietic, neoplastic, and circulatory conditions corroborated by the polygenic enrichments of CHIP in immune cells and myocytes. These findings provide an expanded genomic and phenomic landscape for clonal hematopoiesis.

Our study has several important findings. First, the relatively large sample size enabled better-powered delineations of potential mechanisms that contribute to CHIP and specific driver genes, including failure to repair DNA damage and differential clonal competitive advantage. Among the newly discovered CHIP-associated genes, *ATM* has been significantly associated with the *JAK2 V617F* clonal hematopoiesis and MPNs[46], and it is a core component of the DNA repair system, and the fine-mapped missense variant on this gene is predicted to decrease its function[47]. Also, high *PARP1* expression has been shown to exacerbate DNA damage[48], and the fine-mapped regulatory variant, which reduces CHIP risk, decreases *PARP1* expression in the human blood cells. Driver gene-specific GWAS showed differential association patterns for *DNMT3A* and *TET2* CHIP mutations. In particular, we observed significant signals at the *TCL1A* locus for both *DNMT3A* and *TET2*, with the direction of effects being opposite. *TCL1A* has been observed to implicate in B-cell and T-cell and malignancies[49,50], and its encoded protein is thought to promote Akt activity, which is central to many signaling pathways, such as cellular proliferation, growth, and survival[51,52]. Recent evidence using gene expression data suggested activation of *TCL1A* as an event driving clonal expansion for *TET2*, but not for *DNMT3A*[39] and our GWAS findings provide further evidence to support that finding. These findings indicate that while driver mutations ultimately yield CHIP, their respective heritable bases have shared and distinct heritable factors.

Second, we found that the specific somatic mutation in the hematopoietic stem cell clone was indicative of specific disease consequences. We conducted a phenome-wide scan for overall CHIP and individual driver mutations and contrasted CHIP with a distinct clonal hematopoietic phenomenon, lymphoid-biased mCAs. We observed associations of CHIP and mCAs with clonal hematopoiesis, hematologic malignancy, and non-malignant diseases linked to aging, with leukemia showing the strongest association followed by circulatory, gastrointestinal, genitourinary, renal, and infectious conditions, most of which showed consistent associations between CHIP and mCAs. However, marked heterogeneity was also observed between CHIP and mCAs. For example, unique protective associations for ChrY mCAs and cardiomyopathy in males were observed, while opposing effects between CHIP and mCAs on essential hypertension, pulmonary heart disease, and lower extremity varicose veins were observed.

Third, the application of novel statistical methods yielded new insights into CHIP biology. In alignment with the differential pheWAS associations between *DNMT3A* and *TET2* CHIP mutations, our MR analyses demonstrated their differential causal effects on CAD. We observed

a significant causal effect of genetically determined increased risk for acquiring *TET2* CHIP mutation on an increased risk for CAD, whereas the causal relationship was null for *DNMT3A*. Furthermore, our study leveraged a novel statistical approach to access the polygenic enrichment of CHIP, *DNMT3A*, and *TET2* in scRNA-seq data[31]. Corroborating the significant cardiomyopathy association in pheWAS results, CHIP had the strongest involvement with ventricular myocytes and atrial myocytes, while *DNMT3A* and *TET2* demonstrated different suggestive association patterns at the cell-type level. Lastly, using an empirical Bayes hierarchical approach for jointly modeling multiple traits[34], we leveraged the high correlation between MPN and CHIP variables and nominated three-fold additional genetic regions that contain strong signals for CHIP, *DNMT3A*, and *TET2*.

Important limitations of our study include that many of the newly analyzed samples are of European ancestry, limiting our ability to benefit from the full diversity of genetic variation present in diverse ancestries. Second, the cross-sectional nature of our CHIP analysis limits conclusions regarding the temporal evolution between CHIP and these diseases. Third, the technical sensitivity of whole genome and whole exome sequencing precludes the ability to evaluate the clinical significance of CHIP mutations at low VAF.

Overall, our analyses of the inherited basis of CHIP identified new mechanisms affecting somatic mutation acquisition and clonal fitness with an expanded appreciation for how these somatic mutations shape disease phenotypes.

Methods

Study population

UK Biobank (UKB) is a prospective cohort study of ~500,000 participants (age range 40-69 at enrollment) from across the United Kingdom with deep genetic (DNA isolated from blood) and phenotypic data[53]. The present study was conducted under UKB application ID 7089 and 50834. Secondary use of these data was approved by the Mass General Brigham Institutional Review Board (protocol 2021P002228).

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[22,54]. A total of 51 studies with diverse reported ethnicity (40% European, 32% African, 16% Hispanic/Latino, 10% Asian) were included in the Freeze 6 (https://topmed.nhlbi.nih.gov/topmed-whole-genome-sequencing-methods-freeze-6). Each of the included studies provided informed consent. Secondary analysis of the TOPMed data was approved by the Mass General Brigham Institutional Review Board (protocol 2016P001308). All relevant ethics committees approved this study, and this work is compliant with all applicable ethical regulations.

Mass General Brigham Biobank (MGBB)[55] is a volunteer biobank of patients receiving care at Mass General Brigham with electronic health records and genetic and phenotypic data on ~50,000 participants (https://biobank.massgeneralbrigham.org/). The present secondary analyses were approved by the Massachusetts General Hospital Institutional Review Board (protocol 2018P001236).

BioVU, the Vanderbilt DNA Databank, is a de-identified biorepository that includes DNA from the peripheral blood remaining from routine clinical testing of approximately 250,000 patients[56,57]. The present secondary analyses were approved by the Vanderbilt University Medical Center Institutional Review Board (IRB #201783).

CHIP detection

UKB

Whole exome sequencing of whole blood DNA of 200,628 participants was used to identify somatic mutations. We selected 500 random youngest samples for panel-of-normal (PON) and called somatic mutations on the remaining 200,128 samples using Mutect2 software[58] in the Terra platform (https://portal.firecloud.org/?return=terra#methods/gatk/mutect2-gatk4/20). PON was used to minimize sequencing artifacts, and Genome Aggregation Database (gnomAD)[59] was used to filter likely germline variants from the putative somatic mutations call set. Each Variant Call Format (VCF) file was annotated using ANNOVAR software[60], and putative CHIP mutations were identified using the pipeline described in Bick et al.[22] (https://app.terra.bio/#workspaces/terra-outreach/CHIP-Detection-Mutect2; last accessed Feb 7, 2022). For identifying CHIP, pathogenic variants were queried in 74 genes known to drive clonal hematopoiesis and myeloid malignancies (list of variants queried is presented in **Supplementary Table 20)**[22]**.** We kept variants for further curation if (i) total depth of coverage ≥10, (ii) number of reads supporting the alternate allele ≥3, (iii) ≥1 read in both forward and reverse direction supporting the alternate allele, and (iv) VAF≥0.02. Finally, CHIP mutations that passed sequence-based filtering were manually curated by a team of hematopathologists. The median depth of coverage was 77 (mean=80; SD=31.2; range 9-305), and the median number of supporting reads were 7 (mean=10; SD=8.6; range 3-101) in the final CHIP call set.

NHLBI TOPMed

Whole-genome sequencing of blood DNA was performed on 97,691 samples from Freeze 8 NHLBI TOPMed data with a mean depth of at least 30× using Illumina HiSeq X Ten instruments. All sequences in CRAM files were remapped to the hs38DH 1000 Genomes build 38 human genome reference, following the protocol published previously[61]. Single nucleotide

polymorphisms (SNP) and short indels were jointly discovered and genotyped across the TOPMed samples using the GotCloud pipeline[62]. The procedure used for CHIP and germline variants calling has been previously described[22,63].

MGBB

Whole exome sequencing with an average coverage of 55× from whole blood DNA samples was performed for 13,356 MGBB participants. The procedure used for UKB CHIP calling (described above) was also applied for whole exome sequencing data in MGBB. Here, PON was created from 100 random whole exome sequencing samples from the youngest participants (age at enrolment ≤21y). Putative CHIP mutations that passed the minimum depth of coverage (≥20) and supporting reads threshold were manually curated to arrive at the final CHIP call set.

Vanderbilt BioVU

Among 54,583 participants of the Vanderbilt BioVU, CHIP was identified in DNA genotyped on the Illumina Multi-Ethnic Genotyping Array-Expanded ($MEGA^{EX}$) for somatic mutations in four known CHIP driver genes: *DNMT3A*, *JAK2*, *TET2*, and *ASXL1*. We examined nonsense, splice site, and previously reported missense variants that are genotyped on the $MEGA^{EX}$ array, totaling 29 CHIP mutations. B-allele fractions (BAF) were calculated from the intensity of alternate allele/ (alternate intensity + wild-type intensity), following the method previously developed and validated by Hinds, et al. [46]. For validation of this method of detecting somatic mutations, we evaluated 149 $MEGA^{EX}$-genotyped patients with a putative *JAK2* mutation and performed NGS analysis on the same DNA sample. The two genotyping methods demonstrated high concordance down to the NGS limit of detection of 5% VAF (R^2=0.9931). Because the $MEGA^{EX}$-sequenced DNA samples were not available for *DNMT3A*, *TET2*, or *ASXL1* at the time of the present study, an alternative method was employed for these genes. The mean BAF and its standard deviation were calculated for the population under 40 years of age. Detectable CHIP is extremely rare before this age, allowing for the determination of the noise in BAF measurements at baseline. The BAFs of the over-40 population were normalized to the under-40 mean and standard deviation. Individuals with a normalized BAF greater than or equal to 6 standard deviations above the mean were considered to have a somatic CHIP mutation.

mCAs detection

UKB

The detection of mCAs from genome-wide array genotyping of blood DNA in the UKB has been described in detail previously[64,65]. Briefly, among 447,828 genotyped individuals in the UKB who passed sample quality control criteria, intensities from the genotyping arrays were used to yield log_2R ratio and BAF for each SNP, and the Eagle2 software[66] was used to phase SNPs. mCAs calling was performed by leveraging long-range phase information to search for allelic imbalances

between maternal and paternal allelic fractions across contiguous genomic segments. Constitutional duplications and low-quality calls were filtered out, and cell fraction was estimated as previously described[65].

MGBB

The detection of mCAs in the MGBB was previously described[33]. Briefly, among 22,143 participants who had available probe raw intensity data (IDAT) files for mCAs calling, genotype clustering was performed using the Illumina GenCall algorithm, and the resulting GTC genotype files were converted to VCF files using the bcftools gtc2vcf plugin (https://github.com/freeseek/gtc2vcf). Genotype phasing was performed using SHAPEIT4[67], and the phased genotypes were ligated across overlapping windows using bcftools concat (https://github.com/samtools/bcftools). mCAs detection in the MGBB was performed using MoChA (https://github.com/freeseek/mocha), a workflow to process raw genotypes and probe intensities to the final mCAs callset. Poor quality sample, defined as having a call rate below 0.97, BAF auto-correlation across heterozygous sites greater than 0.03, likely germline calls, or runs of homozygosity identified according to standard filtering practices described in the MoChA online documentation, were excluded from further analysis.

Vanderbilt BioVU

A total of 51,028 participants in the Vanderbilt BioVU had their blood DNA genotyped on the Illumina Multi-Ethnic Genotyping Array-Expanded (MEGA^{EX}). mCAs detection in BioVU was then performed using MoChA (https://github.com/freeseek/mocha). Samples that did not pass quality control criteria were excluded from the analysis.

Millions Veteran Program

DNA extracted from the whole blood of 613,329 participants in the Millions Veteran Program (MVP) was genotyped using a customized Affymetrix Axiom Biobank array, the MVP 1.0 Genotyping Array, followed by standard QC and imputation as described elsewhere[68]. mCAs were detected using the MoChA pipeline (https://github.com/freeseek/mocha). Samples that did not pass quality control criteria were excluded.

GWAS and Meta-analysis

GWAS for overall CHIP, *DNMT3A*, and *TET2* CHIP was performed using REGENIE[69] (in UKB, MGB, and BioVU cohort) or SAIGE[70] (in TOPMed cohort, previously reported in Bick et al.[22]). We performed GWAS in UKB, MGBB, and BioVU cohorts separately with CHIP status (VAF≥2%

as "1", "0" otherwise) as the outcome, fitting a logistic mixed model, adjusting for age at enrolment, age^2, sex, first ten principal components, self-reported ethnicity, and genotyping batch (where appropriate). Here, only unrelated samples (one sample from each pair of 1st or 2nd-degree related samples, or ≥3rd-degree relatedness) with genotype missingness<10% and non-missing outcome/covariates were included in the analysis. GWAS was performed in two steps: 1) prepare the null model using high-quality SNP in a leave-one-out cross-validation approach (REGENIE "--loocv" flag), and 2) perform the single variant association. In Step 1, we used ~500k directly genotyped SNP after excluding variants with minor allele frequency (MAF) below 5%, genotype missingness above 10%, and Hardy-Weinberg equilibrium $P>1 \times 10^{-15}$. In Step 2, we used imputed GWAS variants with a minor allele count ≥20 for the association test. To account for the case-control imbalance, we used the REGENIE "--firth" flag at GWAS P<0.01 to implement the Firth likelihood test to control Type 1 errors.

Next, we meta-analyzed GWAS results from the four cohorts using inverse-variance weighted fixed-effect meta-analysis and Cochran's *Q*-test for heterogeneity in GWAMA software[71]. Here, variants at MAF≥0.1% and present in ≥2 studies were included in the meta-analysis. Finally, variants were considered genome-wide significant at $P \leq 1.67 \times 10^{-8}$, considering 1-million independent SNP and three outcomes tested at a 5% significance level (i.e., $5 \times 10^{-8}/3$).

Conditional and Joint Analysis (COJO)

Conditional and joint (COJO)[72,73] analyses of summary statistics from multi-ancestry meta-analyses of CHIP, *DNMT3A*, and *TET2* CHIP GWAS were performed using a multi-ancestry LD reference prepared from the UKB imputed GWAS dataset that included all the 191k unrelated samples used in our UKB CHIP GWAS. The LD panel included variants with MAF≥0.1% and INFO≥0.3. In the COJO analyses, we conditioned the lead variant (false discovery rate (FDR) <5%) from each chromosome, and independent variants (FDR <5% and LD with the conditioned variant<0.8) were included iteratively. Finally, all variants were fitted simultaneously in the joint analyses when multiple independent variants were detected in the same chromosome. Independent variants at $P \leq 1.67 \times 10^{-8}$ were considered genome-wide significant.

SNP-based heritability (h^2_{SNP})

We estimated h^2_{SNP} using the summary statistics from the multi-ancestry meta-analysis of CHIP GWAS. We used the SumHer function of LDAK software[74,75] with precomputed LD tagging (BLD-LDAK model: https://genetics.ghpc.au.dk/doug/bld.ldak.genotyped.gbr.tagging.gz) prepared from directly genotyped UKB EUR sample. The BLD-LDAK EUR LD tagging was prepared using 577,457 non-ambiguous directly genotyped SNP from 2,000 white British individuals. Using 573,701 overlapping SNP from the summary statistics, h^2_{SNP} was estimated on a liability scale assuming a sample and population CHIP prevalence of 5%.

MashR

To nominate potentially associated variants, we leveraged the observation that myeloproliferative neoplasm (MPN) shared strong germline genetic susceptibility with CHIP[30]. We used MPN GWAS summary statistics from Bao, et al. [30], and applied mashR (Multivariate Adaptive Shrinkage in R)[34] to 6,749,139 variants across MPN, CHIP, *DNMT3A,* and *TET2* using the exchangeable *Z* statistic model. Briefly, we fit the empirical Bayes prior by using the maximum SNP across each of the 1,703 linkage disequilibrium (LD) blocks specified in Pickrell, et al. [76] and used a random sampling of 40,000 SNPs to estimate the relative abundance of each pattern in the overall data set. Armed with this prior information, we then estimated the likelihood and computed posteriors on all variants.

Statistical fine-mapping

To infer putative causal variants in the associated loci, we conducted LD-informed statistical fine-mapping. We derived the LD-matrix of tested variants in each locus using UK Biobank imputed dosage of European ancestries by LDstore2 software (version 2.0)[77]. Using derived LD-matrices, we applied statistical fine-mapping using the summary statistics obtained from the European meta-analysis (only included GWAS of EUR samples from UKB, TOPMed, MGBB, BioVU) by FINEMAP software (version 1.4)[24]. To obtain consistent statistical power for each variant, we only kept variants tested in at least 200,000 individuals.

Similarity-based gene polygenic prioritization of causal genes

We implemented PoPS to leverage the full genome-wide signal for nominating causal genes. Details of these methods have been described elsewhere[25]. Briefly, PoPS is a novel similarity-based gene prioritization approach that assesses the polygenic enrichments of gene features, including cell-type-specific gene expression, protein-protein interaction networks, and biological pathways, through training linear models to predict gene-level association scores from those features and converting the gene p-values from the linear models to *Z*-scores that reflect the confidence on its causal role to the given locus. We used the summary statistics of UKB CHIP, *DNMT3A*, and *TET2* GWAS and the same LD reference panel of 191k UKB samples used in COJO analysis. In total, 57,543 features were considered for analysis, and those who passed marginal feature selection were carried forward to the linear models for computation. We computed a PoPS score for all protein-coding genes within a defined 500kb window around each of the significant genomic regions and prioritized the gene with the highest PoPS score in each locus.

Mendelian randomization

All procedures for the MR analysis between CHIP and CAD were consistent with current recommendations for MR studies[78]. IVs detected in European-only CHIP GWAS were used for this causal analysis. While genetic predispositions for CHIP overlap with those for LTL, previous MR studies support inverse causality for LTL on CAD[79], implying complex causal relationships between CHIP, LTL, and CAD[23]. Thus, we carefully selected instrumental variables (IVs) to avoid clear violations of MR assumptions by excluding IVs 1) associated with known confounders between CHIP and CAD, including hypercholesterolemia, hypertension, type 2 diabetes, body mass index, smoking status, or 2) having more robust associations with LTL than with CAD, and 3) having inverse effect sizes for LTL and CAD, accounting for the inverse causal association between LTL and CAD. We used CAD GWAS summary statistics from CARDIOGRAMplusC4D[80] for the outcome (last accessed March 2022; downloaded from: http://www.cardiogramplusc4d.org/). We downloaded summary statistics from OpenGWAS project[81,82] for excluding IVs which strongly ($P<5\times10^{-8}$) associated with known confounders between CHIP and CAD, including hypercholesterolemia (ukb-b-12651), hypertension (ukb-d-I9_HYPTENS), type 2 diabetes (ebi-a-GCST007518), body mass index (ukb-b-2303), smoking status (ukb-b-20261). We used the full summary statistics for LTL GWAS from the previous study in UKB to inspect the association with LTL[79]. Since we do not have many SNPs that pass through the genome-wide significance, particularly for *TET2*, we used the multi-ancestry meta-analysis of *DNMT3A* and *TET2* CHIP summary statistics and MR using the robust adjusted profile score (MR-RAPS), which allows for the inclusion of weak IVs[42], as the primary method. We performed sensitivity analyses using inverse-variance weighted (IVW)[83], weighted median[83], weighted mode[84], MR-Egger[85], and MR-PRESSO[86] methods available in the TwoSampleMR[82] package in R. For all MR methods except for MR-RAPS, genome-wide significance ($P<5\times10^{-8}$) was considered as the criteria for the IV assumption of robust relevance (the first assumption of MR). We relaxed this assumption for MR-RAPs to $P<1\times10^{-3}$, 1×10^{-4}, and 1×10^{-5}. IVs were clumped into independent loci<10 Mb apart and in the linkage equilibrium ($R^2>0.001$ calculated in European ancestry in UKB) using PLINK (v1.90b6.24)[87]. Due to the limited availability of CHIP GWAS to date, we could not use an independent cohort for IV discovery for MR-RAPS and Steiger filtering to avoid the potential "winner's curse".

PheWAS analysis

The PheWAS of CHIP and mCAs with incident phenotypes across all disease organ system categories were performed using Cox proportional hazards models, adjusting for age, age^2, sex (not used for ChrY and ChrX mCAs analysis), smoking status (using a 25-factor smoking status adjustment in the UKB and current/prior/never smoker status in other cohorts), tobacco use disorder, and principal components 1-10 of genetic ancestry. The primary comparative analysis was conducted among individuals from UKB who had both CHIP and mCAs calls available.

Additionally, in the secondary analysis, mCAs associations were meta-analyzed across the multi-ancestry study population of UKB, MGBB, BioVU, and MVP. Time since DNA collection was used as the underlying timescale. The proportional hazards assumption was assessed using Schoenfeld residuals and was not rejected. Individuals with a history of hematological cancer before DNA collection were excluded. To address multiple testing, an association between CHIP or mCAs and incident health outcomes with FDR<0.05 was considered significant. Analyses of incident events were performed separately in each biobank using the survival package in R (version 3.5, R Foundation). Meta-analyses of the mCAs results were performed using a fixed-effects model from the "meta" R package.

CHIP associations at the cell-type level

scDRS group analysis was used to assess the polygenic associations of CHIP, *DNMT3A*, and *TET2* at cell type and tissue level. Details of the methods have been described elsewhere[31]. Briefly, scDRS is a novel statistical method that associates individual cells in a scRNA-seq data to a trait GWAS based on the aggregate expression across a set of putative GWAS trait genes, assessing statistical significance using appropriately matched control genes; furthermore, the scDRS group analysis computes a p-value for a cell group (e.g., cell type or tissue) based on the association of cells within the given cell group. Following the scDRS guideline, the putative trait gene sets were constructed as the top 1,000 MAGMA genes, where MAGMA[88], an existing gene-scoring method, was applied to GWASs from UKB participants of European ancestry using a set of 489 unrelated individuals of European ancestry from phase 3 1000 Genomes Project as an LD reference panel[89] . Two scRNA-seq data sets were used for this analysis: TMS mouse cell atlas[44] (FACS data with 110,824 cells, 23 tissues, and 103 cell types with more than 50 cells) and TS human cell atlas[45] (FACS data with 26,813 cells, 24 tissues, and 68 cell types with more than 50 cells). Multiple testing correlation (FDR) was applied to each trait separately across all cell types (103 for TMS and 68 for TS) or across all tissues (23 for TMS and 24 for TS). To increase power, associations were considered significant at FDR<0.3, and all suggestive associations with P<0.01 were reported. Sensitivity analyses were conducted by repeating the analysis using downsampled scRNA-seq data (to 50K cells) and downsampled putative trait gene sets (to 500 genes) separately, each with 20 repetitions.

Data availability

All univariable summary statistics for genotype association with CHIP will be available at the time of publication. Epigenome annotation tracks: We obtained chromHMM100 tracks from diverse primary human cells from the NIH Epigenome Roadmap (http://dcc.blueprint-epigenome.eu/#/md/secondary_analysis/Segmentation_of_ChIP-Seq_data_20140811), and additional immune-specific human primary and cell lines from the Blueprint consortium, https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final. Whole-blood eQTL summary statistics: Summary statistics from whole-blood cis-eQTL analysis from 31,683 individuals were downloaded from https://eqtlgen.org. Code used to generate all data in main and supplemental tables and figures will be provided in a publicly accessible archive at the time of publication.

Acknowledgments

We thank the investigators and their studies for contributing samples and/or data to the current work and the participants in those studies who made this research possible. This research has been conducted using the UK Biobank Resource under Application Number 7089 and 50834. This research is based on data from the Vanderbilt BioVU, which is supported by NIH Grant UL1 RR024975-01and UL1 TR000445-06. This research is also based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by Million Veteran Program-MVP000. This publication does not represent the views of the Department of Veteran Affairs or the United States Government.

Fundings

Z.Y. is supported by a grant from NHLBI (5T32HL007604-37). A.G.B. is supported by a Burroughs Wellcome Foundation Career Award for Medical Scientists and the NIH Director's Early Independence Award (DP5-OD029586). A.J.S. is supported by NIH grant F30-DK127699. A.J.S., M.A.R., and H.A.K. are supported by NIH grant T32-GM007347. A.N. is supported by Knut and Alice Wallenberg Foundation (KAW 2017.0436). A.P.R. is supported by NIH grants R01 HL148565 and R01 HL146500. 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. K.C., Y.V.S., and P.W.F.W are supported by Million Veteran Program (MVP) grant numbers I01-BX003340 and I01-BX004821. K.P. is supported by grants from the NHLBI (5-T32HL007208-43). M.C.H. is supported by the American Heart Association (940166 and 979465). P.N. is supported by grants from the NHLBI (R01HL142711, R01HL127564, R01HL148050, R01HL151283, R01HL148565, R01HL135242, 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). P.T.E. is supported by the National Institutes of Health (1RO1HL092577), the American Heart Association Strategically Focused Research Networks (18SFRN34110082), and the European Union (MAESTRIA 965286). 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.M.M. is supported by NIH grant T32-HL144446. V.G.S. is supported by NIH grants R01 DK103794 and R01 HL146500.

Author Contributions

A.G.B. and P.N. designed and conceived the study. A.G.B. and P.N. supervised the study. M.M.U., Z.Y., J.S.W., T.N., A.N., S.M.U., S.K., S.M.Z., K.P., A.J.S., T.M.M., M.Y.W., S.M.H., R.B., S.D.J., M.R., M.C.H., W.E.H., M.J.Z., V.G.S., G.K.G., C.J.G., H.A.K., P.T.E., K.C., Y.S., P.W., S.P., G.G., Y.X., M.R.S., A.P.R, S.J., B.L.E., A.G.B., and P.N. acquired, analyzed or interpreted the data. M.M.U., Z.Y., T.N., S.M.U., S.K., S.M.Z, A.B., and P.N. drafted the manuscript. All authors critically reviewed the manuscript.

Competing Interests

All unrelated to the present work: A.G.B. is a scientific co-founder and has equity in TenSixteen Bio. B.L.E. has received research funding from Celgene, Deerfield, Novartis, and Calico and consulting fees from GRAIL. B.L.E is a member of the scientific advisory board and shareholder for Neomorph Therapeutics, TenSixteen Bio, Skyhawk Therapeutics, and Exo Therapeutics. M.C.H. has received consulting fees from CRISPR Therapeutics and is on the medical advisory board of Miga Health. P.N. reports grant support from Amgen, Apple, AstraZeneca, Novartis, and Boston Scientific, consulting income from Apple, AstraZeneca, Blackstone Life Sciences, Genentech, and Novartis, and spousal employment at Vertex, all unrelated to the present work. P.N., A.G.B, S.J., and B.E. are scientific advisors with equity in TenSixteen Bio, which had no role in the present work. P.T.E. receives sponsored research support from Bayer AG and IBM Research; he has also served on advisory boards or consulted for Bayer AG, MyoKardia, and Novartis. S.J. is a consultant to Novartis, Roche Genentech, AVRO Bio, and Foresite Labs, and on the scientific advisory board for Bitterroot Bio, founder and equity holder of TenSixteen Bio. V.G.S. serves as an advisor to and/or has equity in Branch Biosciences, Ensoma, Novartis, Forma, Sana Biotechnology, and Cellarity. All other authors declare that they have no competing interests.

References

1. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* **20**, 1472-8 (2014).
2. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477-87 (2014).
3. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488-98 (2014).
4. Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-301 (2009).
5. Ley, T.J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-33 (2010).
6. Je, E.M., Yoo, N.J., Kim, Y.J., Kim, M.S. & Lee, S.H. Mutational analysis of splicing machinery genes SF3B1, U2AF1 and SRSF2 in myelodysplasia and other common tumors. *Int J Cancer* **133**, 260-5 (2013).
7. Abdel-Wahab, O. *et al.* ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* **22**, 180-93 (2012).
8. Coombs, C.C. *et al.* Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* **21**, 374-382 e4 (2017).
9. Xia, J. *et al.* Somatic mutations and clonal hematopoiesis in congenital neutropenia. *Blood* **131**, 408-416 (2018).
10. Hsu, J.I. *et al.* PPM1D Mutations Drive Clonal Hematopoiesis in Response to Cytotoxic Chemotherapy. *Cell Stem Cell* **23**, 700-713 e6 (2018).
11. Steensma, D.P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9-16 (2015).
12. Abelson, S. *et al.* Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* **559**, 400-404 (2018).
13. Desai, P. *et al.* Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat Med* **24**, 1015-1023 (2018).
14. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* **377**, 111-121 (2017).
15. Bick, A.G. *et al.* Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation* **141**, 124-131 (2020).
16. Bhattacharya, R. *et al.* Clonal Hematopoiesis Is Associated With Higher Risk of Stroke. *Stroke* **53**, 788-797 (2022).
17. Yu, B. *et al.* Association of Clonal Hematopoiesis With Incident Heart Failure. *J Am Coll Cardiol* **78**, 42-52 (2021).
18. Miller, P.G. *et al.* Association of clonal hematopoiesis with chronic obstructive pulmonary disease. *Blood* **139**, 357-368 (2022).
19. Kim, P.G. *et al.* Dnmt3a-mutated clonal hematopoiesis promotes osteoporosis. *J Exp Med* **218**(2021).

20. Wong, W.J. *et al.* Clonal hematopoiesis and risk of chronic liver disease. *medRxiv*, 2022.01.17.22269409 (2022).
21. Bouzid, H. *et al.* Clonal hematopoiesis is associated with protection from Alzheimer's disease. *medRxiv*, 2021.12.10.21267552 (2021).
22. Bick, A.G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763-768 (2020).
23. Nakao, T. *et al.* Mendelian randomization supports bidirectional causality between telomere length and clonal hematopoiesis of indeterminate potential. *Sci Adv* **8**, eabl6579 (2022).
24. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-501 (2016).
25. Weeks, E.M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *medRxiv*, 2020.09.08.20190561 (2020).
26. Ciccia, A. & Elledge, S.J. The DNA damage response: making it safe to play with knives. *Mol Cell* **40**, 179-204 (2010).
27. Huang, R.X. & Zhou, P.K. DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Signal Transduct Target Ther* **5**, 60 (2020).
28. Mencia-Trinchant, N. *et al.* Clonal Hematopoiesis Before, During, and After Human Spaceflight. *Cell Rep* **33**, 108458 (2020).
29. Jasra, S. *et al.* High burden of clonal hematopoiesis in first responders exposed to the World Trade Center disaster. *Nat Med* **28**, 468-471 (2022).
30. Bao, E.L. *et al.* Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* **586**, 769-775 (2020).
31. Zhang, M.J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *bioRxiv*, 2021.09.24.461597 (2021).
32. Codd, V. *et al.* Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet* **53**, 1425-1433 (2021).
33. Zekavat, S.M. *et al.* Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. *Nat Med* **27**, 1012-1024 (2021).
34. Urbut, S.M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet* **51**, 187-195 (2019).
35. Helgason, H. *et al.* Loss-of-function variants in ATM confer risk of gastric cancer. *Nat Genet* **47**, 906-10 (2015).
36. Karantanos, T. *et al.* ATM Germline Variant Increases the Risk of MPN Progression. *Blood* **134**, 835-835 (2019).
37. Li, X. *et al.* High PARP-1 expression predicts poor survival in acute myeloid leukemia and PARP-1 inhibitor and SAHA-bendamustine hybrid inhibitor combination treatment synergistically enhances anti-tumor effects. *EBioMedicine* **38**, 47-56 (2018).
38. Min, J.L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* **53**, 1311-1321 (2021).
39. Weinstock, J.S. *et al.* Clonal hematopoiesis is driven by aberrant activation of TCL1A. *bioRxiv*, 2021.12.10.471810 (2021).
40. Fuster, J.J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842-847 (2017).

41. Wang, Y. *et al.* Tet2-mediated clonal hematopoiesis in nonconditioned mice accelerates age-associated cardiac dysfunction. *JCI Insight* **5**(2020).
42. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D.S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *The Annals of Statistics* **48**, 1742-1769, 28 (2020).
43. Niroula, A. *et al.* Distinction of lymphoid and myeloid clonal hematopoiesis. *Nat Med* **27**, 1921-1927 (2021).
44. Tabula Muris, C. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590-595 (2020).
45. Consortium, T.T.S. & Quake, S.R. The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans. *bioRxiv*, 2021.07.19.452956 (2021).
46. Hinds, D.A. *et al.* Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood* **128**, 1121-8 (2016).
47. Shiloh, Y. & Ziv, Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol* **14**, 197-210 (2013).
48. Krishnakumar, R. & Kraus, W.L. The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol Cell* **39**, 8-24 (2010).
49. Virgilio, L. *et al.* Identification of the TCL1 gene involved in T-cell malignancies. *Proc Natl Acad Sci U S A* **91**, 12530-4 (1994).
50. Said, J.W. *et al.* TCL1 oncogene expression in B cell subsets from lymphoid hyperplasia and distinct classes of B cell lymphoma. *Lab Invest* **81**, 555-64 (2001).
51. Pekarsky, Y. *et al.* Tcl1 enhances Akt kinase activity and mediates its nuclear translocation. *Proc Natl Acad Sci U S A* **97**, 3028-33 (2000).
52. Laine, J., Kunstle, G., Obata, T., Sha, M. & Noguchi, M. The protooncogene TCL1 is an Akt kinase coactivator. *Mol Cell* **6**, 395-407 (2000).
53. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
54. Nachun, D. *et al.* Clonal hematopoiesis associated with epigenetic aging and clinical outcomes. *Aging Cell* **20**, e13366 (2021).
55. Karlson, E.W., Boutin, N.T., Hoffnagle, A.G. & Allen, N.L. Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. *J Pers Med* **6**(2016).
56. Danciu, I. *et al.* Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform* **52**, 28-35 (2014).
57. Roden, D.M. *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* **84**, 362-9 (2008).
58. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv*, 861054 (2019).
59. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
60. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
61. 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).

62. 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).
63. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
64. Loh, P.R., Genovese, G. & McCarroll, S.A. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* (2020).
65. Loh, P.R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* (2018).
66. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
67. Delaneau, O., Zagury, J.F., Robinson, M.R., Marchini, J.L. & Dermitzakis, E.T. Accurate, scalable and integrative haplotype estimation. *Nat Commun* **10**, 5436 (2019).
68. Klarin, D. *et al.* Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nat Med* **25**, 1274-1279 (2019).
69. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021).
70. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341 (2018).
71. Magi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
72. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
73. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
74. Speed, D. & Balding, D.J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet* **51**, 277-284 (2019).
75. Speed, D., Holmes, J. & Balding, D.J. Evaluating and improving heritability models using summary statistics. *Nat Genet* **52**, 458-462 (2020).
76. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).
77. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551 (2017).
78. Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* **4**, 186 (2019).
79. Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* **45**, 422-7, 427e1-2 (2013).
80. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130 (2015).
81. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv*, 2020.08.10.244293 (2020).
82. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**(2018).

83. Bowden, J., Davey Smith, G., Haycock, P.C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* **40**, 304-14 (2016).
84. Burgess, S., Foley, C.N., Allara, E., Staley, J.R. & Howson, J.M.M. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat Commun* **11**, 376 (2020).
85. Burgess, S. & Thompson, S.G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* **32**, 377-389 (2017).
86. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**, 693-698 (2018).
87. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
88. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
89. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

Figures

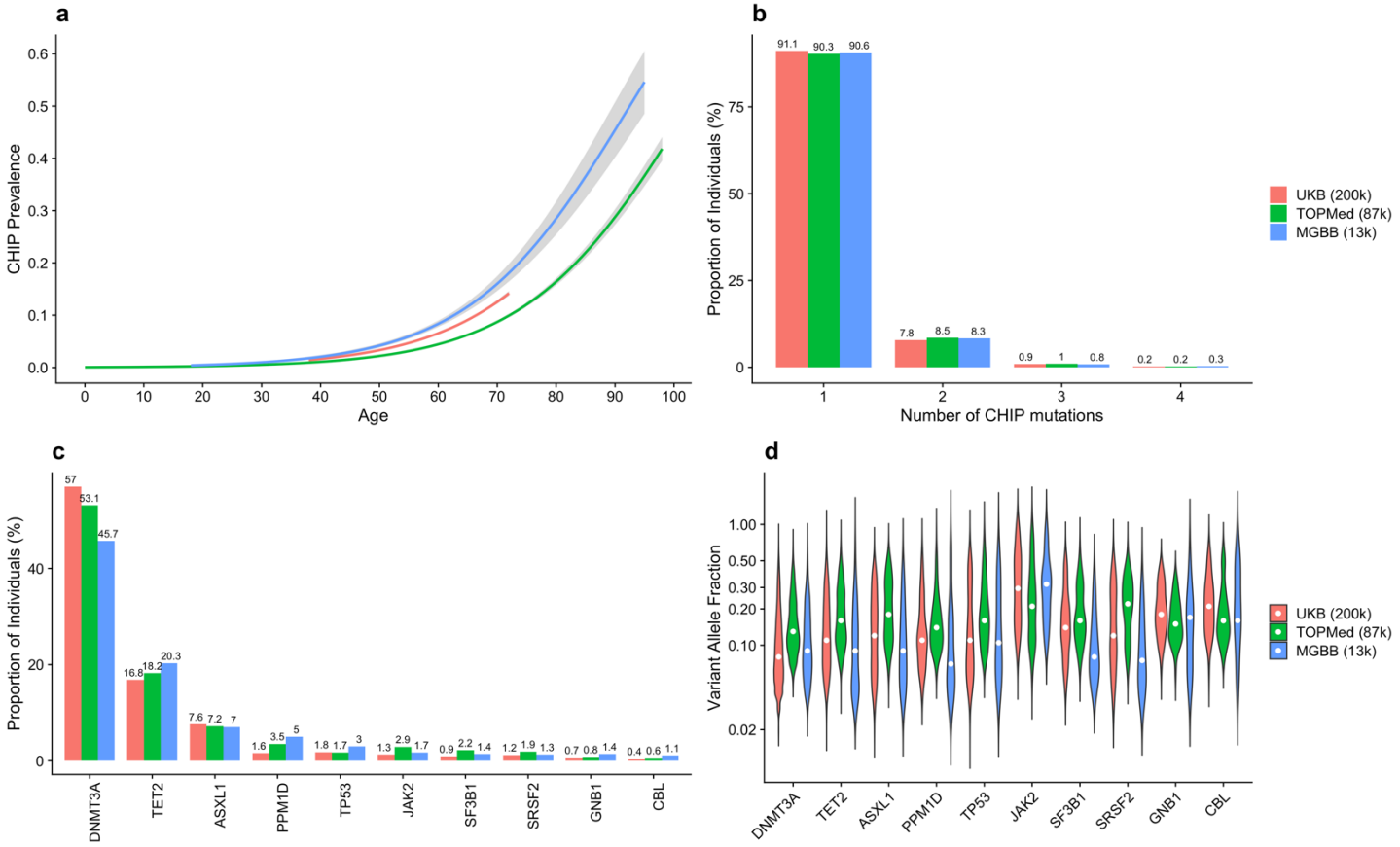

Fig. 1 | Identifying CHIP in 300,600 whole-genome or whole-exome samples. a, CHIP prevalence increased with the age of the donor at the time of blood sampling. The centre line represents the general additive model spline, and the shaded region is the 95% confidence interval (N_{UKB}=200,128 WES; N_{TOPMed}=87,116 WGS; N_{MGBB}=13,356 WES). **b**, More than 90% of individuals with CHIP had only one somatic CHIP driver mutation variant identified. **c**, Proportion of individuals carrying a driver mutation in the ten most frequently mutated genes in CHIP. **d**, There was heterogeneity in CHIP clone size as measured by variant allele fraction by CHIP driver gene. Violin plot spanning minimum and maximum values with median variant allele fraction highlighted by white circles. Information on BioVU was not included as CHIP call is array-based. CHIP: clonal hematopoiesis of indeterminate potential; WES: whole-exome samples; WGS: whole-genome samples.

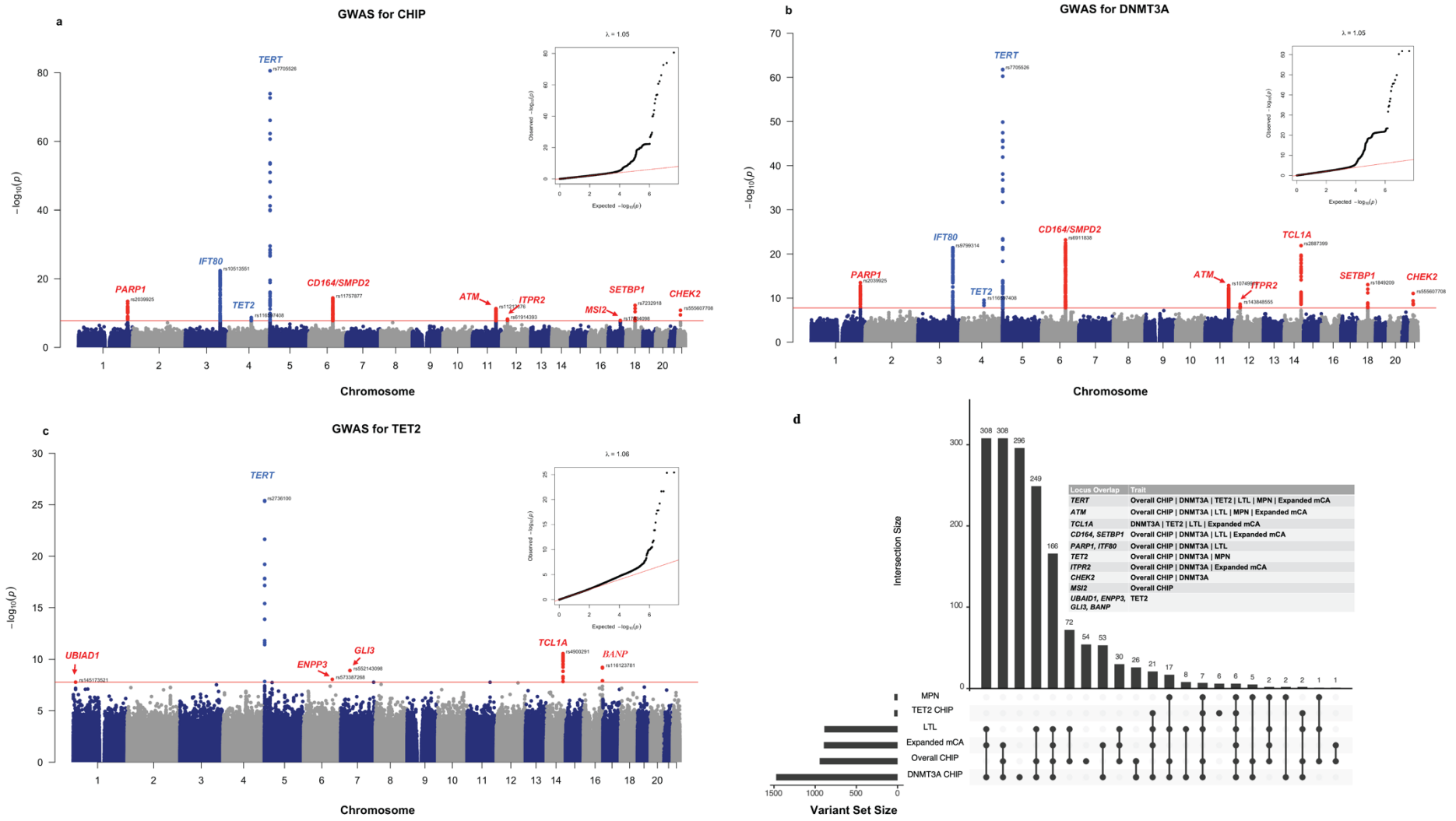

Fig. 2 | Genetic determinants of *CHIP*, *DNMT3A*, and *TET2*. Gnome-wide meta-analyses of CHIP identified (**a**) ten genome-wide-significant (P<1.67e-8) regions for overall CHIP (n=17,044 cases and n=306,068 controls), (b) ten for *DNMT3A* CHIP (n=8,949 cases and n=307,971 controls), and (**c**) six for *TET2* CHIP (n=2,851 cases and n=307,527 controls). Previously known loci in blue, and new loci in red. **d**, Overlap of significant variants from (**a-c**) with GWAS of myeloproliferative neoplasm (MPN)[30], leukocyte telomere length (LTL)[32], and expanded mosaic chromosomal alterations (mCAs)[33]. Only genome-wide significant (P<5e-8) variants from MPN, LTL, and expanded mCAs GWAS were considered to compare the overlap with variants from overall CHIP, *DNMT3A*, and *TET2* CHIP GWAS (P<1.67e-8). CHIP: clonal hematopoiesis of indeterminate potential.

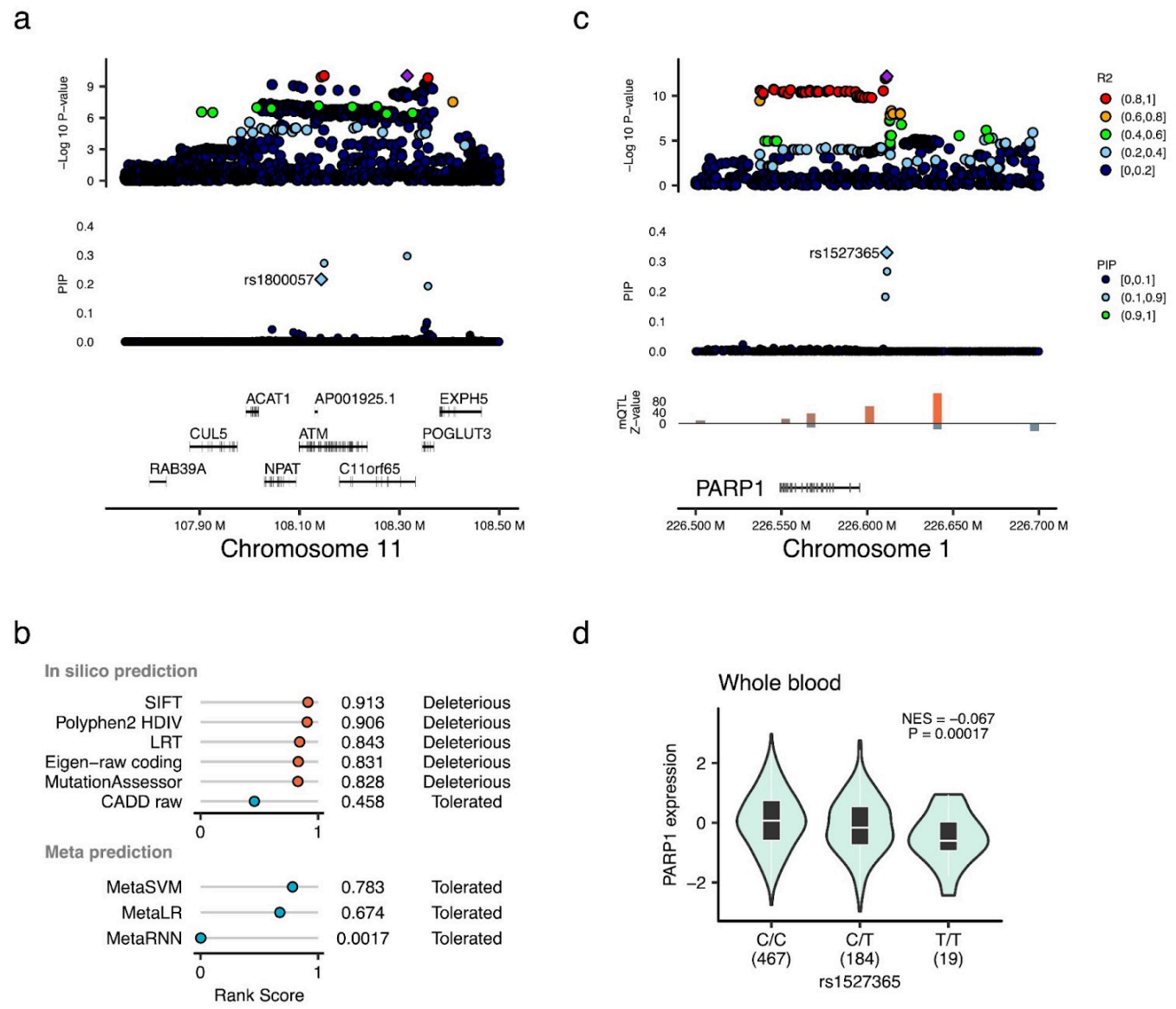

Fig. 3 | Statistical fine mapping in the genome-wide significant loci. a, Regional association plot for *ATM* locus. X-axis shows genetic coordinate. Y-axes show -log10 P-value and PPI. **b,** In-silico functional prediction for putative causal missense variant rs1800057 (ENST00000675843.1:c.3161C>G:p.Pro1054Arg) in *ATM* gene. **c**, Regional association plot for *PARP1* locus. **d**, *PARP1* expression in human whole blood by genotypes at rs1527365 (https://gtexportal.org/). The effect allele rs1527365-T is associated with lower expression of *PARP1*. NES: normalized effect size.

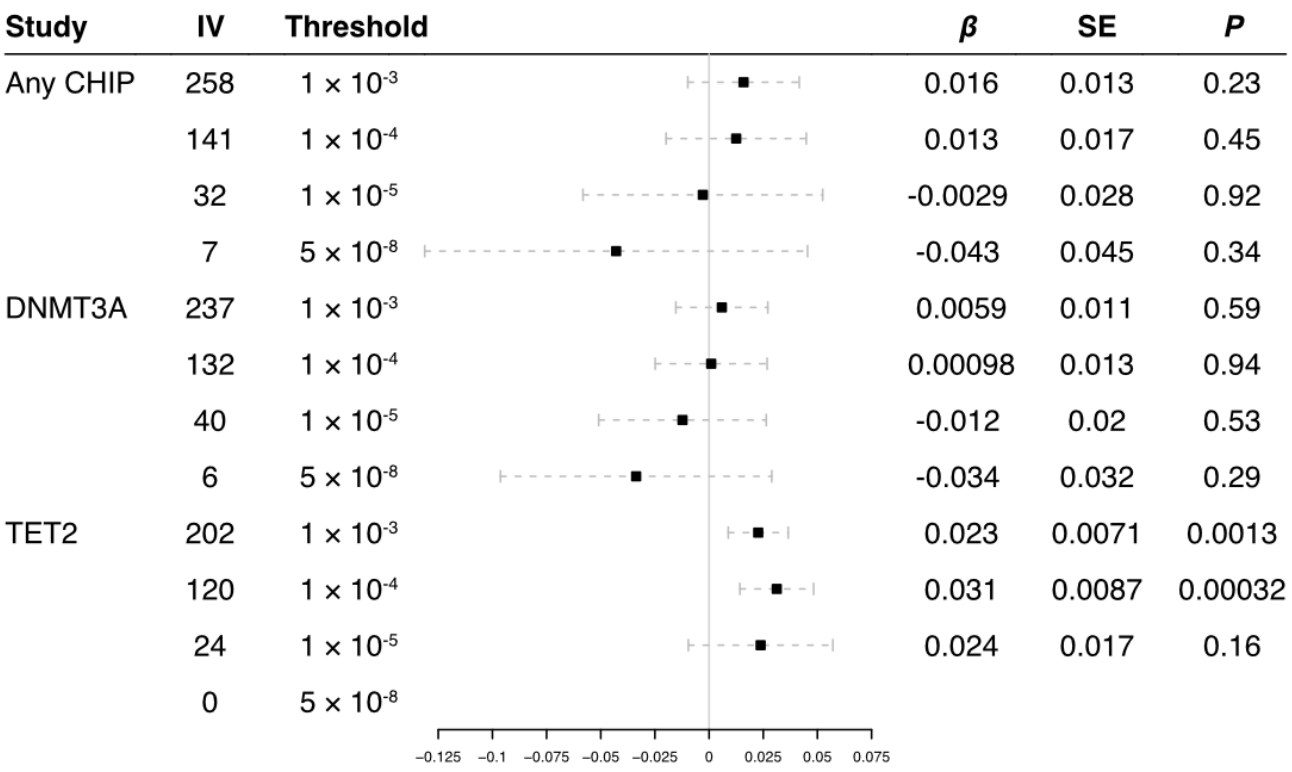

Figure 4. Mendelian randomization for CHIP and CAD using MR-RAPS. MR-RAPS for CHIP and CAD including weak IVs. We selected IVs using GWAS summary statistics for any CHIP, *DNMT3A*, and *TET2*, and filtered out the IVs with 1) associated with known confounders between CHIP and CAD, including hypercholesterolemia, hypertension, type 2 diabetes, body mass index, smoking status, or 2) having more robust associations with LTL than with CAD, and 3) having inverse effect sizes for LTL and CAD, accounting for the inverse causal association between LTL and CAD.1) more robust association with LTL than CAD, and 2) the opposite directionality of the effect for LTL and CAD, to avoid potential confounding by LTL. We used CAD GWAS summary statistics from CARDIOGRAMplusC4D for outcome. CAD: coronary artery disease, CHIP: clonal hematopoiesis of indeterminate potential, IV: instrumental variable, LTL: leukocyte telomere length, SE: standard error.

Figure 5: Heat map of significant associations between CHIP or mCAs and incident conditions from phenome-wide association analysis: (a) Hematopoietic conditions and neoplasms and (b) Non-hematopoietic/neoplastic conditions. PheCode phenotypes were filtered to those with CHIP or mCA FDR<0.05 among incident phenotypes with at least 5 incident cases. The CHIP and mCA analyses were conducted among the same subset of individuals in the UK Biobank with both CHIP and mCA calls available. All analyses were adjusted for age, age^2, sex (not used for ChrY and Chr X mCA analysis), smoking status (using a 25-factor smoking status adjustment in the UK Biobank and current/prior/never smoker status in other cohorts), tobacco use disorder, and principal components 1-10 of genetic ancestry. Colors in heat map reflect z-score (beta/se) of associations. * denotes $0.01 \leq FDR \leq 0.05$, ** denotes $0.0001 < FDR \leq 0.01$, *** denotes FDR<0.0001. CHIP: clonal hematopoiesis of indeterminate potential, mCAs: mosaic chromosomal alterations.

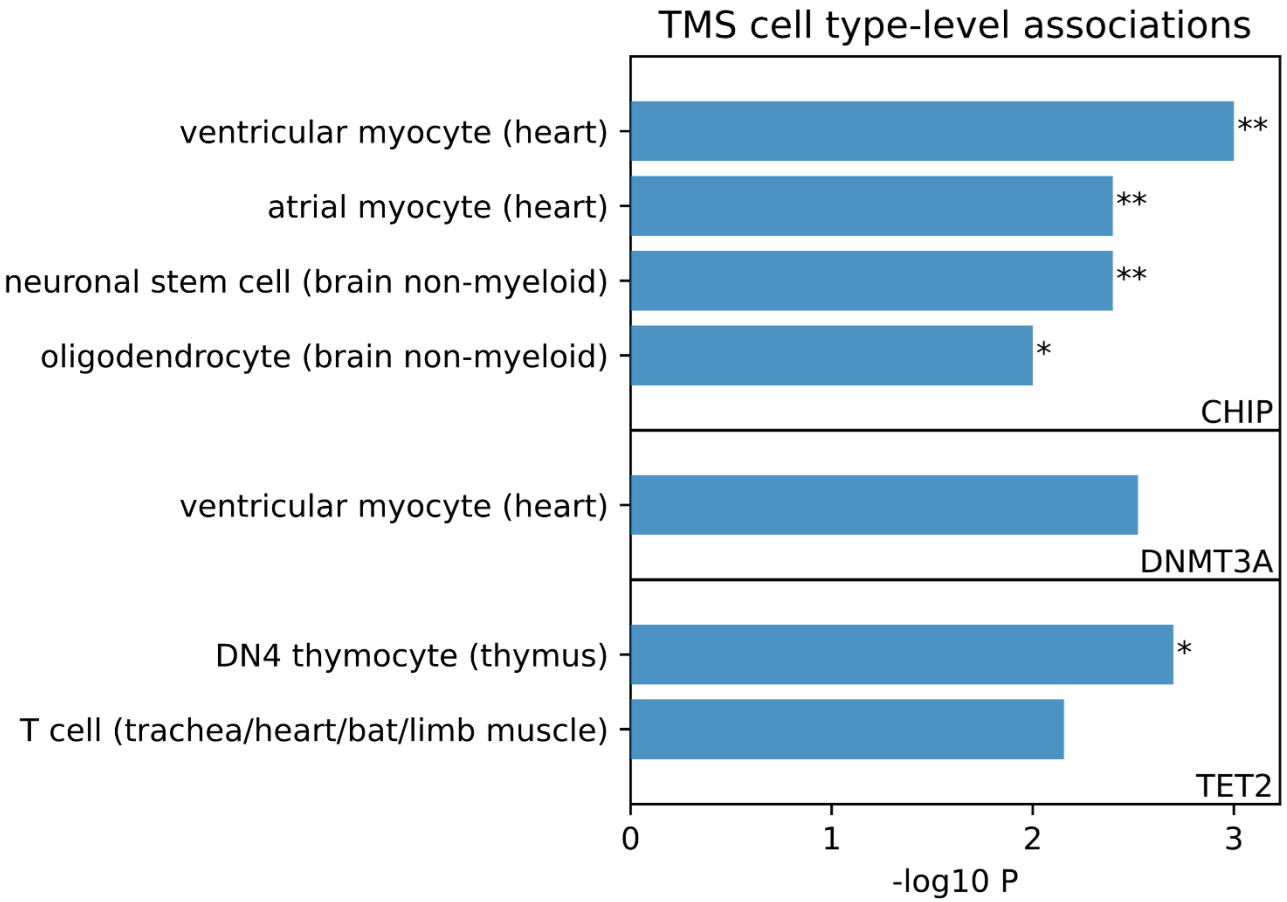

Figure 6: Cell type-level associations in TMS. Cell type-level associations in TMS. The x-axis represents scDRS –log10 cell type-level association p-values while the y-axis represents potentially associated cell types (P<0.01) for the 3 traits, ordered by significance. * denotes FDR<0.3, ** denotes FDR<0.2, and *** denotes FDR<0.1 across all cell types for a given trait. CHIP: clonal hematopoiesis of indeterminate potential. TMS: Tabula Muris Senis.