Mutated cells mediate distinct inflammatory responses in clonal hematopoiesis

Authors:

J. Brett Heimlich*[1], Pawan Bhat*[2], Alyssa C. Parker[2], Matthew T. Jenkins[2], Caitlyn Vlasschaert[3], Jessica Ulloa[4], Chad R. Potts[5], Sydney Olson[4], Alexander J. Silver[2], Ayesha Ahmad[4], Brian Sharber[4], Donovan Brown[5], Ningning Hu[4], Peter van Galen[6,7], Michael R. Savona[5,8], Alexander G. Bick[+4], P. Brent Ferrell[+5]

Affiliations:

[1]Division of Cardiovascular Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232

[2]Vanderbilt University School of Medicine, Nashville, TN 37232

[3]Department of Medicine, Queen's University, Kingston Ontario, Canada

[4]Division of Genomic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232

[5]Division of Hematology and Oncology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232

[6]Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115 USA

[7]Ludwig Center at Harvard, Harvard Medical School, Boston, MA 02115, USA

[8]Vanderbilt-Ingram Cancer Center, Program in Cancer Biology, and Center for Immunobiology Nashville, TN 37232

[+]Correspondence to Dr. Bick (alexander.bick@vumc.org) & Dr. Ferrell (brent.ferrell@vumc.org)

*Denotes equal contribution

Abstract:

Clonal hematopoiesis (CH) is an age-associated phenomenon that increases risk for hematologic malignancy and cardiovascular disease. CH is thought to increase cardiovascular disease risk through inflammation in the peripheral blood[1]. Here, we profile peripheral blood gene expression in 104,566 single cells from a cohort of 17 CH patients and 6 controls. We discovered that patients harboring *DNMT3A* and *TET2* CH mutations at baseline and in response to IL-6 stimulation confer a pro-inflammatory profile to CD14+ monocytes through previously unrecognized pathways including Macrophage Inhibitory Factor (MIF). A germline genetic variant in MIF modifies *TET2* CH cardiovascular disease risk. Using mitochondrial lineage tracing, we used a novel method to compare gene expression between mutated and non-mutated cells within individual CH patients. We found that the mutated CH monocytes, but not non-mutated monocytes are pro-inflammatory, explaining why patients with larger CH clones have increased cardiovascular disease risk.

Main Text:

With age, hematopoietic stem cells acquire mutations in leukemogenic driver genes such as DNA Methyl Transferase 3A (*DNMT3A)* and Tet Methylcytosine Dioxygenase 2 (*TET2)* resulting in a proliferative advantage and clonal hematopoiesis. Over the past decade, CH has been identified as a notable risk factor not only for hematologic malignancy but surprisingly also for multiple diseases of aging including cardiovascular disease, pulmonary disease, kidney disease, and osteoporosis[1–5]. While many of these epidemiological analyses consider CH as a single entity, a more nuanced appraisal of the literature reveals that significant associations are gene specific. For example, *TET2* is more strongly associated with a pro-inflammatory disease mechanism across multiple forms of cardiovascular disease[1,6] while *DNMT3A* CH is associated with heart failure[7,8] and osteoporosis[5].

Much attention has focused on how CH mutations lead to skewed hematopoiesis in hematopoietic stem and progenitor cells (HSPC)[9,10]. However, little attention has focused on the peripheral compartment that is the putative mediator of CH pathogenicity in non-oncologic disease. Circulating immune cells with CH mutations are morphologically and immunophenotypically similar to their non-mutated counterparts, making direct comparisons difficult in primary human tissues. Whether primary cell intrinsic transcriptional changes or secondary microenvironment effects, or both, drive pathological phenotypes is presently unknown. Further, whether the response to immune activation differs between mutated and non-mutated cells has not been clarified.

Though transcriptional profiling of single cells has become routine, it has been much more difficult to extract both genotype and transcriptional data out of the same cell simultaneously. Since mutant *DNMT3A* and *TET2* CH blood samples are a mixture of mutated and non-mutated cells, both genotyping and transcriptomic sequencing modalities are necessary to distinguish between the cell-intrinsic and cell-extrinsic consequences. Cell-intrinsic consequences may arise directly from the somatic mutation while extrinsic, indirect consequences may arise from altered cell-cell interactions or secreted immune effectors. Several technologies have sought to close this gap by selectively amplifying the mRNA transcriptome and using this to genotype cells[11–14]. This approach is effective in HSPCs that express *DNMT3A* and *TET2*, with a genotyping efficiency rate of ~1-9%; however, genotyping is less efficient in cells that do not express these genes, such as fully differentiated cells in the peripheral blood[12,13,15]. Another method, TARGET-seq, has sought to amplify specific, previously identified variants in mutated cells with high-efficiency genotyping along with gene expression, albeit lower throughput[16,17].

To overcome this, we combined single-cell RNA-sequencing (scRNA-seq) with cell-specific mitochondrial DNA barcoding to simultaneously resolve single-cell DNA mutation status for

100% of cells[15]. We focused on CH samples with either mutated *DNMT3A* or *TET2*, which represent a majority of clonal hematopoiesis of indeterminate potential cases and have divergent phenotypic consequences in CH-related diseases. Our analysis of 104,566 single cells from 17 individuals with *TET2* CH or *DNMT3A* CH and 6 age-matched controls identifies novel mechanisms of CH-driven inflammation and enables direct comparison between peripheral CH mutated and wildtype cells within the same individual.

Results:

We combined single-cell RNA-sequencing (scRNA-seq) with cell-specific mitochondrial DNA barcoding to simultaneously resolve single-cell DNA mutation status and investigate pathological mechanisms of CH (**Fig. 1A, methods**). Peripheral blood mononuclear cells (PBMCs) from 8 *TET2*, 9 *DNMT3A* and 6 age-matched controls (ages 47-89) were selected from a prospective CH observational study (**Fig. 1B, Supplementary Data Table S1**).

To trace the effects of CH mutations on peripheral blood cell type proportions, we integrated data across samples (**Fig. 1C, Extended Data Figs. 1 and 2, Supplementary Data Table S2**). Consistent with prior reports, a monocytic skew was identified among CD14+ monocytes in both *DNMT3A* and *TET2* CH samples[9,12] (**Extended Data Fig. 3**). By contrast, there were no significant differences between CH and control monocyte counts on standard clinical whole blood laboratories, indicating that removal of granulocytes during PBMC isolation increased the relative proportion of monocytes. Four patients had multiple CH mutations and concomitant cytopenias without bone marrow dysplasia, meeting diagnostic criteria for clonal cytopenias of undetermined significance (CCUS, **Extended Data Fig. 1A)**[18].

Across *TET2* and *DNMT3A* monocytes we found MHC Class II HLA genes were upregulated in both *TET2* and *DNMT3A* CH patients, suggestive of enhanced antigen presentation to CD4+ T

cells (**Fig. 1D-E and Extended Data Fig. 4A, Supplementary Data Table S3 and S4)**. We explored differential gene expression across T cell subtypes. In CD8+ T cells, both *TET2* and *DNMT3A* CH samples exhibited upregulation of genes involved in T cell activation, intracellular signaling, and immunomodulation (**Extended Data Figs. 4B-C, 5A-B, Supplementary Data Table S5 and S6)**. CD4+ T cells from both *DNMT3A* and *TET2* CH patients significantly increased levels of *CXCL3*, *CCL3*, *CD69*, and decreased expression of *IL7R*, genes associated with T cell activation (**Extended Data Figs. 5C-D, Supplementary Data Table S7 and S8)**. However, we discovered that several inflammatory ligands, including *IL1B* and *CCL7*, *CXCL1*, and *CXCL3,* were increased in only *TET2* CD14+ monocytes[19,20]. Gene Ontology pathway analysis further reinforced the differing physiological effects of *TET2* and *DNMT3A* (**Extended Data Fig. 4**).

Given the presence of augmented cell activation and antigen presentation, we hypothesized that intercellular communication might be altered in CH. Using CellChat[21] to infer cellular interactions in our scRNA-seq data, we found CH CD14+ monocytes exhibited enhanced signaling across a variety of inflammatory pathways (**Extended Data Fig. 6**). Some were present in *DNMT3A* and *TET2* CH, such as BAFF and Galectin, while others were genotype specific. For example, *DNMT3A* had heightened outgoing MHC-I and MHC-II signaling while *TET2* displayed increased communication through macrophage migration inhibitory factor (*MIF*).

Peripheral blood in CH represents a mixture of mutated and non-mutated cells. Since CH genes are not readily expressed in terminally differentiated cells in the periphery, we applied two orthogonal approaches to identify genomic mutational status in scRNAseq data using mitochondrial lineage tracing in one TET2 and one DNMT3A patient (**Fig. 2B and Extended Data Fig. 7, Supplementary Data Table S9 and S10**). In the first approach, we clonally expanded and then sequenced individual colonies from PBMCs to obtain both nuclear and MT

DNA transcripts. This provided the critical link between CH mutation status and co-occurring mitochondrial DNA mutations, enabling the identification of clonal populations in our single-cell data (**Fig. 2A, methods**). After colony sequencing in the *DNMT3A* patient (VAF 0.24), one mitochondrial mutation, 747A>G was found in 8 of the 12 colonies harboring genomic CH mutations and none of the colonies without genomic CH mutations (**Extended Data Figs. 8A, C and 9**). This suggested that the mitochondrial mutation and CH mutation were of shared ancestry. Since only 8 of the 12 colonies have this clone, we suspect this mutation arose as a subclone within the context of the *DNMT3A* mutation and therefore all cells with the MT 747A>G mutation harbor the *DNMT3A* CH mutation. As further confirmation, we performed single-cell DNA sequencing (scDNAseq) using MissionBio's Tapestri platform to identify mitochondrial mutations that align with the known genomic mutation of interest in 1,399 individual cells **(Extended Data Fig. 8B, D).** The scDNAseq revealed a similar pattern for the *TET2* patient where a single mitochondrial mutation was associated with cells that harbored *TET2* mutations. Of the 1,399 cells sequenced, 781 carried both the *TET2* mutation and MT 7754G>C mutation and 565 carried neither. Thirty carried the MT 7754G>C mutation but not the *TET2* mutation, and 23 had the *TET2* mutation but not the MT 7754G>C mutation, which we attribute to technical limitations of the assay. Using oligonucleotide conjugated antibodies along with scDNAseq, we were able to identify that the monocytes/myeloid cells were strikingly enriched for both the TET2 mutation and the mitochondrial variant **(Extended Data Fig. 9D)**.

To determine whether the alterations in inflammatory signaling occur as a cell-specific phenomenon or because of an altered microenvironment, we performed differential gene expression analysis on CD14+ monocytes comparing CH mutant cells to non-mutated cells from the same patient (see **methods**). For *TET2*, 112 genes were differentially expressed in mutant CD14+ monocytes when compared to non-mutated CD14+ monocytes (**Fig. 2C, Supplementary Data Table S11**. Notable among the upregulated genes in the *TET2* mutant

CD14+ monocytes were *IL1B*, *LGALS3*, *SOD2*, and *VIM*, all known inflammatory mediators. In contrast, *DNMT3A* had only four differentially expressed genes between the mutant and non-mutated CD14+ monocytes, none involving inflammation (**Supplementary Data Table S12**). Extending this approach to intercellular interactions, we found that *MIF* and *MPZ* signaling were significantly elevated for *TET2* mutated CD14+ monocytes compared to non-mutated CD14+ monocytes; whereas CCL, Galectin, and MHC-II were all significantly elevated in outgoing signaling patterns from *DNMT3A* mutated CD14+ monocytes, compared to non-mutated CD14+ monocytes (**Fig. 3A**).

To evaluate whether the CH gene specific inflammatory signaling pathways might have prognostic or therapeutic relevance, we performed a human genetic association study of *MIF* signaling in the context of CH. We identified *DNMT3A* and *TET2* CH mutations in 340,766 individuals from the UK Biobank with available blood whole exome sequencing and who were free of cancer at the time of sequencing blood draw. We investigated whether *MIF* modifies cardiovascular risk in CH. We incorporated individual genotypes for a common SNP in the *MIF* promoter (rs755622-G>C) that is associated with higher *MIF* expression in GTEx[22–24] (**Fig. 3B**). *MIF* rs755622 had a significant interaction with *TET2* CH (HR 1.42, 95% CI 1.06 – 1.90) but not with *DNMT3A* CH, in fully adjusted models (**Fig. 3C**). While the distribution of rs755622 alleles did not differ by CH status, 42% of all incident CAD events in *TET2* CH individuals occurred in those with at least one high risk C allele compared to 31% in those with *DNMT3A* CH and 32% in the those without any CH.

Recent evidence suggests CH increases the risk of atherosclerosis through the IL-6/IL-1B signaling pathway and that therapeutic inhibition of IL-1B decreases cardiovascular risk in TET2 CH patients[1,25]. We hypothesized that IL-6 stimulation would expose discrepant gene expression patterns between *TET2* and *DNMT3A*[26] (**Fig. 4A, methods**). After *in vitro* IL-6

stimulation, gene expression changes were consistent among canonical IL-6 pathway genes, confirming effective stimulation across individual samples (**Extended Data Fig. 10**). As a result of IL-6 exposure, mutant monocytes had remarkably divergent results based on genotype. Whereas *TET2* mutants had many upregulated genes in the inflammation, antigen presentation, and monocyte activation categories, there were no differences between *DNMT3A* mutants and their non-mutant counterparts. Further, *IL1B*, a gene known to be active downstream of IL-6 stimulation, was significantly elevated in *TET2* mutants (average log2 fold change = 0.5, adjusted p-value = 5.99e-9, **Fig. 4B)**. Although IL-6 signaling was high in both DNMT3A and TET2 samples, IL-6 provoked elevated *MIF* signaling in *TET2* samples, while TNF signaling increased in *DNMT3A* CH (**Fig. 4C**).

The heightened IL-6/IL-1B signaling response in *TET2* monocytes led us to hypothesize that the IL-6/STAT3 signaling cascade was particularly active within these cells. To test this hypothesis, we employed phosphofluorescent flow cytometry to measure response to IL-6 in high VAF *TET2* mutant, low VAF *TET2* mutant, and controls (**Fig. 4D, methods**). As expected, the basal pSTAT3+ monocyte percentage was significantly higher in *TET2*hi monocytes compared to controls. All samples showed some response to IL-6, while *TET2*hi monocytes had the highest proportion of pSTAT3+ cells, significantly higher than both control and TETlo samples (**Extended Data Fig. 11**).

Discussion:

Here we present transcriptional profiling and characterization of *DNMT3A* and *TET2* CH from human peripheral blood. By using a new approach that integrates multimodal single-cell RNA sequencing with scDNA sequencing to link mitochondrial mutations to somatic nuclear mutations, we simultaneously resolved DNA mutational status and cell state. Our study revealed CH mutation specific aberrations in cellular state permitting several conclusions.

First, we identify *TET2* mutated CD14+ monocytes as the driver of *TET2* CH-associated inflammation in the peripheral blood. Comparing mutant CD14+ monocytes directly to same-sample non-mutated cells revealed *TET2* mutated cells to be enriched for inflammatory genes and chemokines. Computationally inferred outgoing signaling among the *TET2* mutated CD14+ monocytes indicated a notable increase in *MIF* signaling. The marked differences in genetic programming between the mutated and wildtype CD14+ monocytes underscore the cell-intrinsic nature of *TET2* mutations, a finding only possible with the ability to identify mutant and wildtype cells from the same sample.

Second, our data highlights *MIF* as a mediator of *TET2* CH pathology. *MIF* resides as a pre-formed peptide in a variety of cell types and binds with its receptors CXCR2 as well as CXCR4 to promote the recruitment of monocytes and T cells to sites of tissue injury[27]. Recruitment of hyperinflammatory monocytes has been identified as the initiating event in the development of atherosclerotic plaques[28]. Moreover, a G/C single-nucleotide polymorphism (rs755622) is associated with enhanced *MIF* production in humans[29]. The SNP that increased MIF expression was significantly associated with the development of CAD in *TET2* but not *DNMT3A* CH among ~341,000 people in the UKB. Together, these data support a pathogenic role for *MIF* signaling in the setting of *TET2* CH and nominates MIF inhibitors as therapeutic candidates for TET2 CH.

Third, our observation that cell-intrinsic inflammatory responses are driven predominantly by *TET2* mutations and not *DNMT3A* mutations identifies a precision medicine therapeutic for CH. In particular, a recent analysis of the Canakinumab Anti-inflammatory Thrombosis Outcomes Trial (CANTOS) found that IL-1B antagonist, Canakinumab[30], reduced cardiovascular risk in *TET2* but not *DNMT3A* CH patients [2,31]. Our data provides a mechanistic rationale for a genotype specific approach to treat in CH.

Fourth, our study identifies cell-extrinsic effects of CH-mutations in peripheral blood. In our work, CD4+ T cells had increased expression of genes relating to TCR activation in both *TET2* and *DNMT3A* CH, however only a small minority of T-cells carry CH mutations due to the lineage bias conferred by both *DNMT3A* and *TET2*. Thus, CH-intrinsic signaling and transcriptional changes are likely intermingled with CH-extrinsic effects on non-mutated cell populations. Future investigation of T cell function and specificity could help decipher the role of adaptive immune processes in CH and define microenvironmental shifts in response to these alterations[32].

Our study has several limitations. First, while *TET2* and *DNMT3A* make up approximately 2/3 of all CH mutations, CH represents a diverse set of mutations in >70 genes. These CH mutations are likely to have divergent effects from those we describe here. Second, we cannot exclude that CH with small clones below our limit of detection are present in our control samples. However, we would expect minimal pathological effect given the very small size of these clones. Third, a shortcoming of our work is the absence of neutrophils in PBMC samples. Neutrophils have been implicated in pathologic roles in ischemia re-profusion injury and, more recently, in atherosclerotic plaque erosion and rupture[33]. Given the myeloid bias of CH mutations, it is likely that these cells also harbor mutations within the periphery.

Overall, our study provides mechanistic support for a genotype specific precision medicine approach for future CH therapeutics.

Methods:

Primary patient samples

All patients in this study consented to all study procedures under VUMC institutional review board approved research protocols (IRBs #210022, #201583) in accordance with the Treaty of Helsinki. Patients were recruited via referrals from VUMC clinics based on concern for CH. All patients were confirmed to be without active hematological malignancy at the time of enrollment and sampling. See **Extended Data Table 1** for clinical information. Fresh PBMCs were isolated using Ficoll separation. Following low-speed centrifugation, pelleted cells were resuspended in freezing media (88% FBS and 12% dimethyl sulfoxide (DMSO)) and aliquoted into cryotubes. Samples were initially stored at -80 °C for 24-48 hours and then placed in liquid nitrogen.

DNA extraction and CH variant calling

All enrolled patients underwent targeted sequencing to evaluate for the presence of CH mutations. DNA was extracted from whole blood using Qiagen Mini kits Cat #27104 according to manufacturer's recommendations. We batch sequenced samples using a custom capture panel designed to tile known CH genes on an Illumina Novaseq 6000. Somatic mutations were called using publicly available methods in workflow description language in Mutect2 on the Broad Institute's Terra Platform (https://terra.bio/). A putative variant list was formulated and then cross referenced with a list of known CH driver mutations. Variants were then filtered for read quality including sequencing depth and minimum alternate allele read depth.

IL-6 stimulation and scRNAseq library preparation

Cryopreserved PBMCs from CH patients and donors were thawed at 37 °C and washed with complete RPMI (RPMI + 10%FBS + 1%PS, cRPMI) to remove freezing media. Samples were rested in RPMI for one hour at 37 °C and 5% CO_2 prior to drug treatment/stimulation. PBMCs from each sample (500,000 cells) were plated and stimulated with either IL-6 (PeproTech #200-06) at 20 ng/mL concentration or phosphate-buffered saline (PBS) control. All cells were treated with Golgi inhibitor Brefeldin A prior to stimulation. Following stimulation, the cells were

incubated for 4.5 hours at 37 °C and 5% CO_2. Cells were pooled following staining with unique hashtag antibody oligonucleotide-conjugates (HTO) (30 minutes) staining per manufacture recommendations (Biolegend, TotalSeq-B). Pooled samples were diluted to manufacturers recommended concentration immediately run on a 10x Chromium Controller after preparation with a 10X Chromium 3' library preparation kit (10X Genomics) to create scRNA-seq libraries.

10X single-cell sequencing and data preparation

Next-generation 150-nt paired-end sequencing was conducted on an Illumina Novaseq6000. (**Supplementary Data Table S13**). The template switch oligo sequence was trimmed from the 5' end and the poly-A tail was trimmed from the 3' end. CellRanger Count (10X Genomics) was used to filter low quality reads and align to the GRCh38 reference genome using STAR as described elsewhere[34]. Resulting files from the CellRanger pipeline were then converted to Seurat[35,36] objects. Demultiplexing was performed using the HTODemux function in the R package Seurat, applying a positive quantile value of 0.99. Cells containing 15% or more reads mapping to the mitochondrial genome were filtered. Similarly, we filtered cells with less than 250 genes and 500 UMIs respectively. Any remaining doublets not removed as a part of the HTO demultiplexing were removed with the R package DoubletFinder[37], using the first 10 principal components. The doublet formation rate was set to 5%, based on standardized doublet rate expectations published by 10X Genomics[38]. Individual datasets were integrated with batch correction using a weighted nearest neighbor approach. Supervised cell annotation was performed using the R package Azimuth using standard parameters and reference UMAP (https://zenodo.org/record/4546839#.Ytlg9uzMLS4) yielding expected cell types across all samples (**Extended Data Fig. 1B**)[39]**.** Cell annotation yielded 30 clusters that were then collapsed into 22 main subsets based on expression levels of canonical markers (**Extended Data Fig. 2**).

Mitochondrial enrichment

Mitochondrial enrichment was performed using primers sequences as described in Miller et al. [15]. Briefly, the 10x Genomics 3' cDNA was amplified with mitochondrial specific primers as well as Illumina-specific adapters (i5 and i7 index barcodes, **Supplementary Data Table S14-15)**. The resulting cDNA libraries contain reads enriched for MT-DNA.

Mitochondrial DNA read processing and variant calling

MT-DNA processing and variant calling were carried out as previously published[15]. Briefly, fastq files from MAESTER enrichment were filtered for reads associated with low-frequency cell barcodes (CB) and then trimmed to remove the UMI and CB. Reads were aligned using STAR to the same genomic reference as used in the scRNAseq experiments (hg38). Next, we used a python package, maegatk, which was developed specifically to address the technical challenges in mtDNA variant calling[15,40]. Maegatk selects reads aligned with mtDNA and using combined cell barcodes from both scRNAseq and MAESTER enrichment, calls mtDNA variants with at least 3 supporting reads.

Identification and selection of informative mtDNA variants

An allele frequency matrix was constructed using all possible variants in the mitochondrial genome. Next, a table with informative variants and features for a given sample was constructed. VCFs from this process were manually reviewed and annotated for the presence or absence of the known CH mutation. VCFs were then concatenated and mtDNA variants within each colony were identified and recorded in a dataframe. A variant heatmap was derived from this data from which colony phylogenies could be formulated. mtDNA variants that were exclusively associated with known CH mutated colonies were subsequently identified. Confirmed variants were then evaluated in the single-cell data to assess relative quality and depth of reads, VAF among collected cells, and lineage bias to ensure congruent results with

initial genotype results. Once confirmed, individual cells containing more than ten reads of the mitochondrial variant were assigned mutant status.

Colony-based single-cell DNA sequencing

One million cells were suspended in 150ul of PBS (Corning cat#21-040-CV). Each sample was added to 1.5 ml of Methocult Optimum (StemCell cat#4034) and mixed by vortexing. A 1.1ml aliquot of cell mixture was plated onto individual wells of a 6-well SmartDish (Stemcell cat#27370) and were incubated for 14 days. On day 14, the plate was photographed, and colony morphology was identified using standardized STEMvision Automated Colony Counting (**Extended Data Fig. 9**). After counting, 25 colonies from each well were plucked using a pipette, washed with PBS (Corning cat#21-040-CV), and frozen at -20°C. The DNA from each individual colony was then extracted using the Qiagen QIAamp micro-DNA extraction kit (Cat# 56304). After DNA extraction, samples were batch sequenced on an Illumina Novaseq 6000 using the same custom capture panel as above with a 1:1 spike-in of a fixed content mitochondrial panel developed by Twist Biosciences cat#102040. This panel was designed to cover all 16,659 base pairs and 37 genes of the mitochondrial genome. Somatic CHIP and mitochondrial DNA mutations were called using publicly available methods as described earlier.

Single-cell DNA sequencing

Single-cell DNA sequencing was performed with MissionBio's Tapestri platform, according to manufacturer instructions and similar to previously reported[34]. Briefly, cryopreserved cells were thawed, washed with PBS, and manually counted using a hemocytometer. Cells were normalized to 10,000 cells/uL in 100 uL and incubated with 10 mg/mL dextran sulfate (Research Products International), Human TruStain FcX (BioLegend), and 1x staining buffer (MissionBio) for 3 min at ambient temperature. Next, cells were stained with the BioLegend Total-Seq D Heme Oncology Panel (Biolegend, Inc.). Following staining, cells were washed three times in

Dulbecco's PBS (Gibco), recounted, and processed as above with the addition of adding 2 µM antibody tag forward primer (Tapestri) prior to the barcoding step. Conjugated antibody oligonucleotide PCR products from supernatant after Ampure XP bead purification step were isolated with a 5 min incubation with 2 µL 5' Biotin Oligo (IDT) for 5 min at 96°C, followed by a 5 min incubation on ice. Isolated samples were washed using 2x binding and washing buffer (10 mM Tris-HCL, 1mM EDTA, 2 M NaCl) and streptavidin beads (ThermoFisher). Protein libraries were generated using the washed proteins, library template (MissionBio), and i5/i7 indices (IDT) via PCR. The antibody oligonucleotide library PCR product was cleaned again using Ampure XP beads. Encapsulated single cell genomic and mitochondrial DNA was amplified via multiplex PCR using a combination of the 20-gene, 127-amplicon catalog AML panel and a custom mitochondrial panel covering all 15 mitochondrial genes (MIssionBio, Inc.). Amplified cellular DNA was then released from individual oil droplets and purified using Ampure XP beads. The final product was quantified via Qubit fluorometer from ThermoFisher and assessed for quality on an Agilent Bioanalyzer. Samples were pooled prior to sequencing with a 25% spike-in of PhiX and run on a NovaSeq 6000 S4 flow cell from Illumina to generate 150 bp paired-end reads. Sequencing was performed at the Vanderbilt Technologies for Advanced Genomics (VANTAGE) sequencing core.

Pipeline processing and variant filtering for single-cell DNA sequencing

FASTQ files from single-cell DNA samples were processed via the Tapestri Pipeline v1.8.4. Adapters were trimmed using Cutadapt[41] and reads longer than 30 nucleotides were aligned to the hg19 reference genome with BWA-MEM[42]. For mitochondrial genotyping, hg19 locations were lifted over to hg38 locations for harmonizing data between scRNAseq (hg38) and scDNAseq. In the results we report hg38 variant locations. Cells were then called based on amplicon-read completeness in each barcode. Variants were called using GATK 3.7 with a joint-calling approach that follows GATK best practices[43]. Then, the variant lists were decomposed,

filtered, and the genotype/cell matrix loaded into the Tapestri Insights software package (v.2.2) where low-quality cells and variants were removed based on genotype quality score <20, variants genotyped in <30% of cells, read depth <10 reads, cells with <30% of genotypes present, cells with genotypes of <20% alternate allele frequencies, and variants mutated in <0.5% of cells.

Differential expression analysis, GSEA, and data visualization

To determine cell state differences in aggregated CH patients and controls, differential expression was calculated using the FindMarkers function from the R package Seurat (v4.1.1). Differences between cases and controls were quantified using the Wilcoxon rank-sum test. Genes with absolute log2 fold change greater than 0.25 and adjusted p-value less than 0.05 were considered significant. Ribosomal and mitochondrial genes were excluded. Similar analyses were performed for intra-patient comparisons between mutant and wildtype cells. Cell-cell interactions were predicted with the R package CellChat (v1.1.3), which uses a comprehensive signaling database and scRNA-seq-based differential expression testing to predict the strength of a signal between known ligand-receptor pairs. Default settings were used.

Gene set enrichment analysis was performed with the R package Clusterprofiler[44]. Evaluations were performed based on average log2 fold change for all genes with available data and all available ontologies. The minimum group size was set to 2 and the maximum group size was set to 800.

Data analysis steps were completed using R (v4.1.2) on the Terra cloud platform (https://app.terra.bio). Data visualization was performed using several R packages, including Seurat (v4.1.1)[35,36], ggpointdensity (v0.1.0)[45], ggrepel (v0.9.1.9999)[46], ggradar (v0.2)[47], UpSetR

(v1.4.0)[48], ggplot2 (v3.3.6)[49] and the Royal1 color scheme from the wesanderson package (v0.3.6)[50]. Additional packages used for file manipulation and data processing include tidyverse (v1.3.1), xlsx (v0.6.5), future (v1.26.1), and openxlsx (v4.2.5). Schematics were created with BioRender.com.

Epidemiological analyses of CHIP and cardiovascular risk

We identified *TET2* and *DNMT3A* mutations in 340,766 individuals in the UK Biobank who had no known malignancies or coronary artery disease at enrollment. Briefly, this was achieved using methodology similar to previously reported including putative somatic variant identification with Mutect2, followed by variant annotation and filtering using pre-specified criteria[1]. Variants meeting these criteria and with an allelic fraction ≥ 2% supported by ≥ 5 reads and with total site read depth ≥20 were classified as CHIP. *TET2* variants within the catalytic regions that failed a binomial test were excluded due to germline suspicion.

The primary outcome of interest was incident coronary artery disease, defined as incident myocardial infarction, coronary artery revascularization (percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG)), or death from coronary artery disease occurring in an individual with no recorded history of these events prior to enrollment. Cox proportional hazards regressions were performed to assess incident risk, and the models were adjusted for age, age^2, sex, the 10 first principal components of genetic ancestry, and where indicated, the following cardiovascular risk factors: systolic blood pressure, diabetes, smoking status, and median-imputed LDL-C levels. Individual rs75622 genotype – a proxy of *MIF* expression – was added as an interaction term in the model where indicated.

Phospho-specific flow cytometry

Cryovials of cryopreserved cells from healthy donors and CH patients were thawed in a 37°C water bath and washed with 10 mL of cRPMI. Cells were rested in cRPMI for 1 hour at 37°C/5% CO_2 prior to IL-6 stimulation. The cells were then spun down at 300xg for 5 minutes and stained for viability with AlexaFlour700 (Invitrogen, cat#P10163). Cells were counted following an additional wash with cRPMI and centrifugation at 300xg/5 min. An aliquot of 500,000 cells were plated in 200 µL of media in a 96-well plate and stimulated with 20 ng/mL of IL-6 (Peprotech) for 15 minutes. Cells were then fixed with 1.6% PFA at room temperature and spun down at 800xg for 5 minutes and permeabilized with 150 µL of methanol at -80C for at least 30 minutes. After permeabilization, cells were washed twice with 150 mL of PBS followed by a spin at 800xg for 5 minutes. Cells were resuspended in 180 µL of PBS and fluorescence cell barcoding performed as previously described[51] with serial dilutions of Pacific Blue (LifeTechnologies, cat#P10163, PB) and Pacific Orange (Invitrogen, cat#P30253, PO) dyes for 30 minutes in the dark at RT. Two concentrations of PB were prepared (20 and 4 µg/mL), while six concentrations of PO were prepared (7.00, 2.99, 1.27, 0.54, 0.23, 0.10 µg/mL). Barcoding was quenched with 80 µL of cell staining media and spun at 800xg for 5 minutes. The barcoded cells were then collected into a single tube and stained with a cocktail of antibodies: CD33 PECy7 (5 µL per 100 µL stain, Biolegend, cat#303434, clone WM53) and pSTAT3 AlexFlour488 (2.5 µL per 100 µL stain, Biolegend, cat#651006, clone 13A3-1) for 30 minutes prior to acquisition on a BD 5-laser Fortessa flow cytometer.

Data and materials availability: All filtered count matrices are available on Open Science Framework (osf.io) at https://osf.io/rac5w/ and will be made public upon publication. Raw FASTQ files and human sequence data are in the process of being submitted to dbgap. All data analysis was completed using R (v4.1.2) on the Terra.bio cloud platform. All R files used to generate the figures and tables will be made publicly available on GitHub upon publication and can be made available for reviewers upon request.

Acknowledgements

We thank Angela Jones for her assistance with sequencing efforts associated with this work and Joseph Van Amberg for his technical support with DNA extractions. 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). P.B.F is supported by a NIH K23HL138291 and a Mark Foundation Endeavor Award. P.v.G. is supported by the Ludwig Center at Harvard, the NIH (R00CA218832), Gilead Sciences, the Bertarelli Rare Cancers Fund, the Starr Cancer Consortium, the William Guy Forbeck Research Foundation, and is an awardee of the Glenn Foundation for Medical Research and American Federation for Aging Research (AFAR) Grant for Junior Faculty. M.R.S. is supported by NIH 1R01CA262287 and 1U01OH012271, an LLS Clinical Scholar Award, the Biff Ruttenburg Foundation, the Adventure Alle Fund, the Beverly and George Rawlings Research Directorship, the E.P. Evans MDS Foundation.

Author contributions

J.B.H. and P.B. designed the study, facilitated data collection, conducted formal analysis and interpretation of results, generated figures, prepared the original draft and edited the manuscript.

A.C.P., C.V., and M.T.J. collected data, conducted formal analysis and interpretation of results, generated figures, prepared the original draft and edited the manuscript

J.U., C.R.P., S.O., and N.N.H. facilitated sample curation and data collection

B.S. and A.A. provided analysis software

P.V.G. provided resources, analysis software, and edited manuscript

A.J.S. and M.R.S. facilitated sample curation, provided resources and project administration, and edited the manuscript

A.G.B. and P.B.F. conceived and supervised the study, provided funding for the study, provided resources and project administration, conducted formal analysis and interpretation of results, generated figures, prepared the original draft and edited the manuscript

Competing Interests Statement

All unrelated to the present work: M.R.S. reports personal fees from AbbVie, BMS, CTI, Sierra Oncology, Novartis, grants from Astex and Incyte, personal fees and other support from Karyopharm, Ryvu, personal fees from Sierra Oncology, grants and personal fees from Takeda, and TG Therapeutics outside the submitted work. P.B.F. reports grants from Incyte. A.G.B. is a scientific co-founder and has equity in TenSixteen Bio. All other authors declare that they have no competing interests.

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to:

Alexander G. Bick

P. Brent Ferrell

References

1. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* **377**, 111–121 (2017).
2. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N Engl J Med* **371**, 2488–2498 (2014).
3. Bick, A. G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).
4. Zink, F. *et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**, 742–752 (2017).
5. Kim, P. G. *et al. Dnmt3a* -mutated clonal hematopoiesis promotes osteoporosis. *Journal of Experimental Medicine* **218**, e20211872 (2021).
6. Bhattacharya, R. *et al.* Clonal Hematopoiesis Is Associated With Higher Risk of Stroke. *Stroke* **53**, 788–797 (2022).
7. Abplanalp, W. T. *et al.* Clonal Hematopoiesis–Driver DNMT3A Mutations Alter Immune Cells in Heart Failure. *Circulation Research* **128**, 216–228 (2021).
8. Assmus, B. *et al.* Clonal haematopoiesis in chronic ischaemic heart failure: prognostic role of clone size for DNMT3A- and TET2-driver gene mutations. *European Heart Journal* **42**, 257–265 (2021).
9. Arends, C. M. *et al.* Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. *Leukemia* **32**, 1908–1919 (2018).
10. Izzo, F. *et al.* DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat Genet* **52**, 378–387 (2020).
11. Nam, A. S. *et al.* Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **571**, 355–360 (2019).

12. Nam, A. S. *et al.* Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. 2022.01.14.476225 Preprint at https://doi.org/10.1101/2022.01.14.476225 (2022).
13. Myers, R. M. *et al.* Integrated Single-Cell Genotyping and Chromatin Accessibility Charts JAK2V617F Human Hematopoietic Differentiation. 2022.05.11.491515 Preprint at https://doi.org/10.1101/2022.05.11.491515 (2022).
14. van Galen, P. *et al.* Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. *Cell* **176**, 1265-1281.e24 (2019).
15. Miller, T. E. *et al.* Mitochondrial variant enrichment from high-throughput single-cell RNA sequencing resolves clonal populations. *Nat Biotechnol* 1–5 (2022) doi:10.1038/s41587-022-01210-8.
16. Rodriguez-Meira, A. *et al.* Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Molecular Cell* **73**, 1292-1305.e8 (2019).
17. Rodriguez-Meira, A., O'Sullivan, J., Rahman, H. & Mead, A. J. TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *STAR Protocols* **1**, 100125 (2020).
18. DeZern, A. E., Malcovati, L. & Ebert, B. L. CHIP, CCUS, and Other Acronyms: Definition, Implications, and Impact on Practice. *Am Soc Clin Oncol Educ Book* **39**, 400–410 (2019).
19. Ridker, P. M. From C-Reactive Protein to Interleukin-6 to Interleukin-1: Moving Upstream To Identify Novel Targets for Atheroprotection. *Circ Res* **118**, 145–156 (2016).
20. Veillard, N. R. *et al.* Differential influence of chemokine receptors CCR2 and CXCR3 in development of atherosclerosis in vivo. *Circulation* **112**, 870–878 (2005).
21. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).

22. Li, D.-Y. *et al.* MIF -173G/C (rs755622) polymorphism modulates coronary artery disease risk: evidence from a systematic meta-analysis. *BMC Cardiovascular Disorders* **20**, 300 (2020).
23. Luo, J.-Y. *et al.* Association between MIF gene promoter rs755622 and susceptibility to coronary artery disease and inflammatory cytokines in the Chinese Han population. *Sci Rep* **11**, 8050 (2021).
24. Du, G.-L. *et al.* MIF gene rs755622 polymorphism positively associated with acute coronary syndrome in Chinese Han population: case–control study. *Sci Rep* **10**, 140 (2020).
25. Bick, A. G. *et al.* Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation* **141**, 124–131 (2020).
26. Kartha, V. K. *et al.* Functional inference of gene regulation using single-cell multi-omics. *Cell Genomics* 100166 (2022) doi:10.1016/j.xgen.2022.100166.
27. Bernhagen, J. *et al.* MIF is a noncognate ligand of CXC chemokine receptors in inflammatory and atherogenic cell recruitment. *Nat Med* **13**, 587–596 (2007).
28. Pan, J.-H. *et al.* Macrophage Migration Inhibitory Factor Deficiency Impairs Atherosclerosis in Low-Density Lipoprotein Receptor-Deficient Mice. *Circulation* **109**, 3149–3153 (2004).
29. Illescas, O., Gomez-Verjan, J. C., García-Velázquez, L., Govezensky, T. & Rodriguez-Sosa, M. Macrophage Migration Inhibitory Factor -173 G/C Polymorphism: A Global Meta-Analysis across the Disease Spectrum. *Frontiers in Genetics* **9**, (2018).
30. Svensson, E. C. *et al.* TET2-Driven Clonal Hematopoiesis and Response to Canakinumab: An Exploratory Analysis of the CANTOS Randomized Clinical Trial. *JAMA Cardiology* **7**, 521–528 (2022).
31. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine* **371**, 2477–2487 (2014).

32. DePasquale, E. A. K. *et al.* Single-Cell Multiomics Reveals Clonal T-Cell Expansions and Exhaustion in Blastic Plasmacytoid Dendritic Cell Neoplasm. *Front Immunol* **13**, 809414 (2022).
33. Libby, P. Interleukin-1 Beta as a Target for Atherosclerosis Therapy. *Journal of the American College of Cardiology* **70**, 2278–2289 (2017).
34. Guess, T. *et al.* Distinct Patterns of Clonal Evolution Drive Myelodysplastic Syndrome Progression to Secondary Acute Myeloid Leukemia. *Blood Cancer Discovery* **3**, 316–329 (2022).
35. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
36. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
37. *DoubletFinder: DoubletFinder is a suite of tools for identifying doublets in single-cell RNA sequencing data.* (2022).
38. 10x Genomics. *Technical Note - Resolving Cell Types as a Function of Read Depth and Cell Number*. https://assets.ctfassets.net/an68im79xiti/6gDArDPBTOg4IIkYEO2Sis/803be2286bba5ca67f353e6baf68d276/CG000148_10x_Technical_Note_Resolving_Cell_Types_as_Function_of_Read_Depth_Cell_Number_RevA.pdf (2019).
39. Butler, A., Darby, C., Hao, Y., Hoffman, P. & Satija, R. *Azimuth: A shiny app demonstrating a query-reference mapping algorithm for single-cell data*. https://github.com/satijalab/azimuth (2021).
40. Lareau, C. A. *et al.* Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat Biotechnol* **39**, 451–461 (2021).
41. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).

42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
43. Auwera, G. van der & O'Connor, B. D. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. (O'Reilly Media, 2020).
44. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
45. Kremer, L. P. M. *ggpointdensity: A cross between a 2D density plot and a scatter plot*. https://github.com/LKremer/ggpointdensity (2022).
46. Slowikowski, K. *ggrepel: Automatically position non-overlapping text labels with 'ggplot2'*. https://github.com/slowkow/ggrepel (2021).
47. Bion, R. *ggradar: Create radar charts using ggplot2*. (2022).
48. Gehlenborg, N. *UpSetR: A more scalable alternative to venn and euler diagrams for visualizing intersecting sets*. http://github.com/hms-dbmi/UpSetR (2019).
49. Wickham, H. *ggplot2: Elegant graphics for data analysis*. (Springer-Verlag New York, 2016).
50. Ram, K. & Wickham, H. *wesanderson: A wes anderson palette generator*. https://github.com/karthik/wesanderson (2018).
51. Earl, D. C. *et al.* Discovery of human cell selective effector molecules using single cell multiplexed activity metabolomics. *Nat Commun* **9**, 39 (2018).

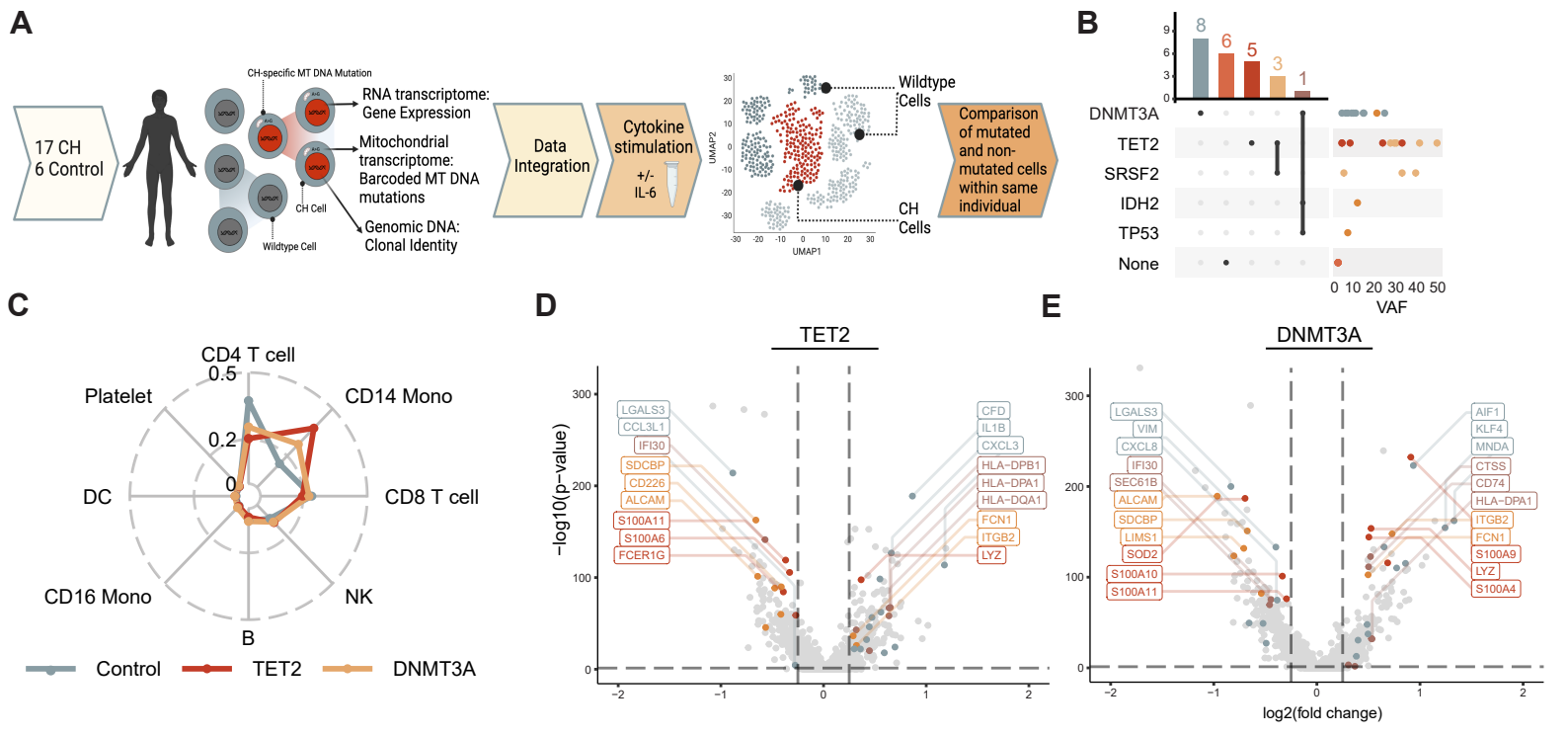

Fig. 1: Differential expression analysis reveals differing physiological effects for TET2 and DNMT3A mutant CH samples. A) Cells from 17 CH patients and 6 controls underwent scRNA-seq and a subset were stimulated with IL-6. Mutational lineage tracing was performed to link genomicvariants to scRNA-seq results. B) UpSet plot displaying CH mutations for each patient with accompanying dot plot showing variant allele frequency for each mutation. Color corresponds to mutational group. C) CD14+ monocyte proportion increased for patients with TET2 mutations and patients with DNMT3A mutations. Radar plot showing cell type proportions for controls, patients with TET2 mutations, and patients with DNMT3A mutations. D-E) Basal gene expression analysis within the CD14+ monocyte cluster identified numerous differentially expressed gene across inflammation, antigen presentation, cell adhesion, and monocyte activation pathways in both (D) TET2 (n = 8) and (E) DNMT3A (n = 9) vs control samples (n=6). Significance of differentially expressed genes tested via non-parametric Wilcoxon rank sum test (Seurat “FindMarkers” function, adjusted $p < 0.05$, abs(log2FC) > 0.25).

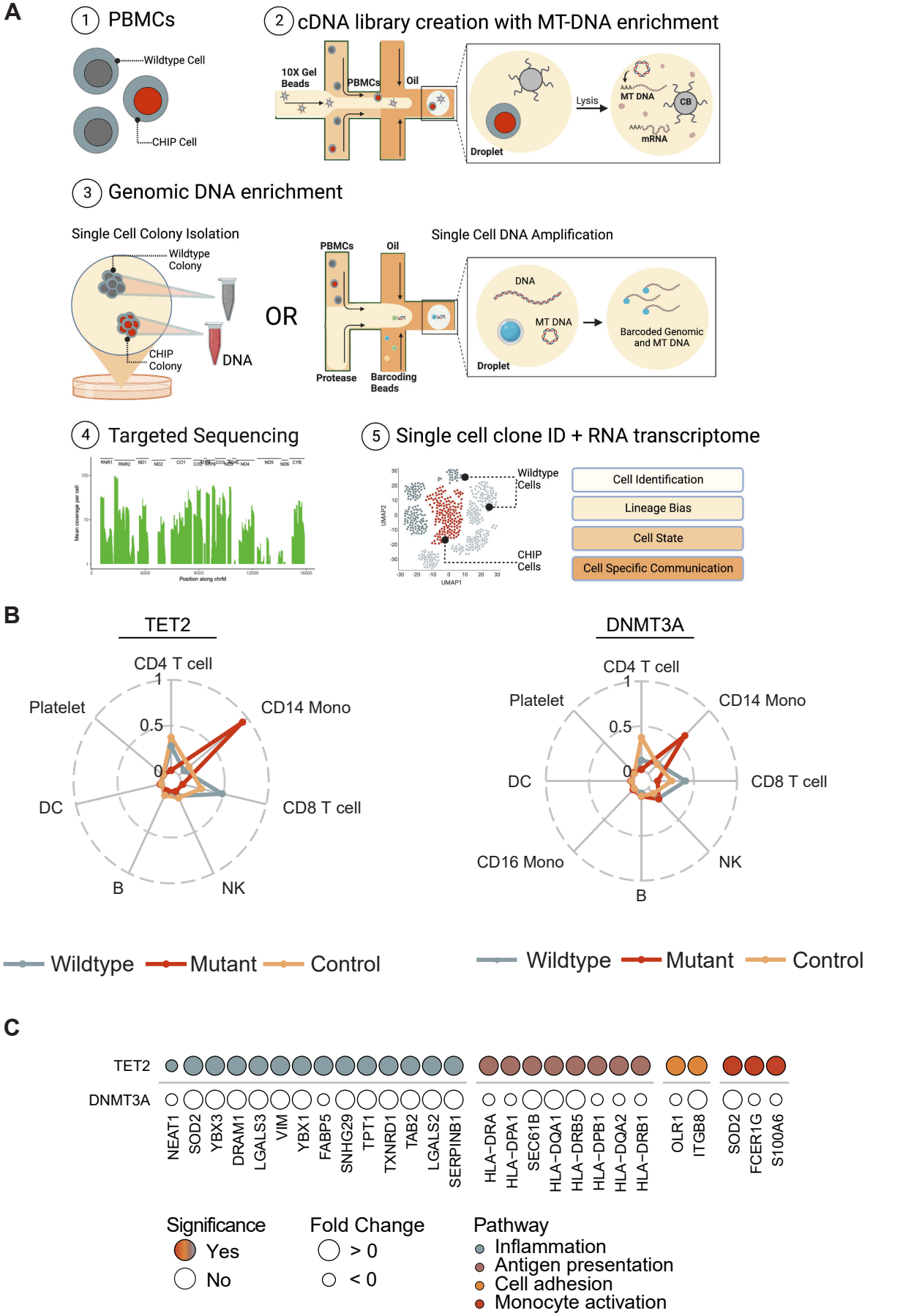

Fig. 2: Simultaneous single-cell resolution of DNA mutational status and RNA transcriptomes identifies aberrant CH gene expression programs. (A) Workflow to resolve single-cell genotype and RNA transcriptome. PBMC – peripheral blood mononuclear cells, CB – Cell Barcode. B) The proportion of both mutated TET2 and DNMT3A mutant cells was notably different from non-mutated cells ($p < 2e-16$ for both). Cell type proportions quantify monocytic skew in radar plots ($n = 4,425$ and 872, in TET2 and DNMT3A respectively) plotted against non-mutated cells from within the same sample ($n = 4,160$ and 3,585). Aggregated controls ($n = 36,373$) are plotted for reference C) Differentially expressed genes between mutated and wildtype CD14+ monocytes (Seurat "FindMarkers" function, adjusted $p < 0.05$, abs(log2FC) > 0.25). Selected top significant genes labeled and color coded by pathway. Rows represent comparisons between mutated and wildtype CD14+ monocytes from the same individual – top row is TET2, bottom is DNMT3A. Size is reflective of the adjusted p-value and darker color gradient indicates a positive fold change.

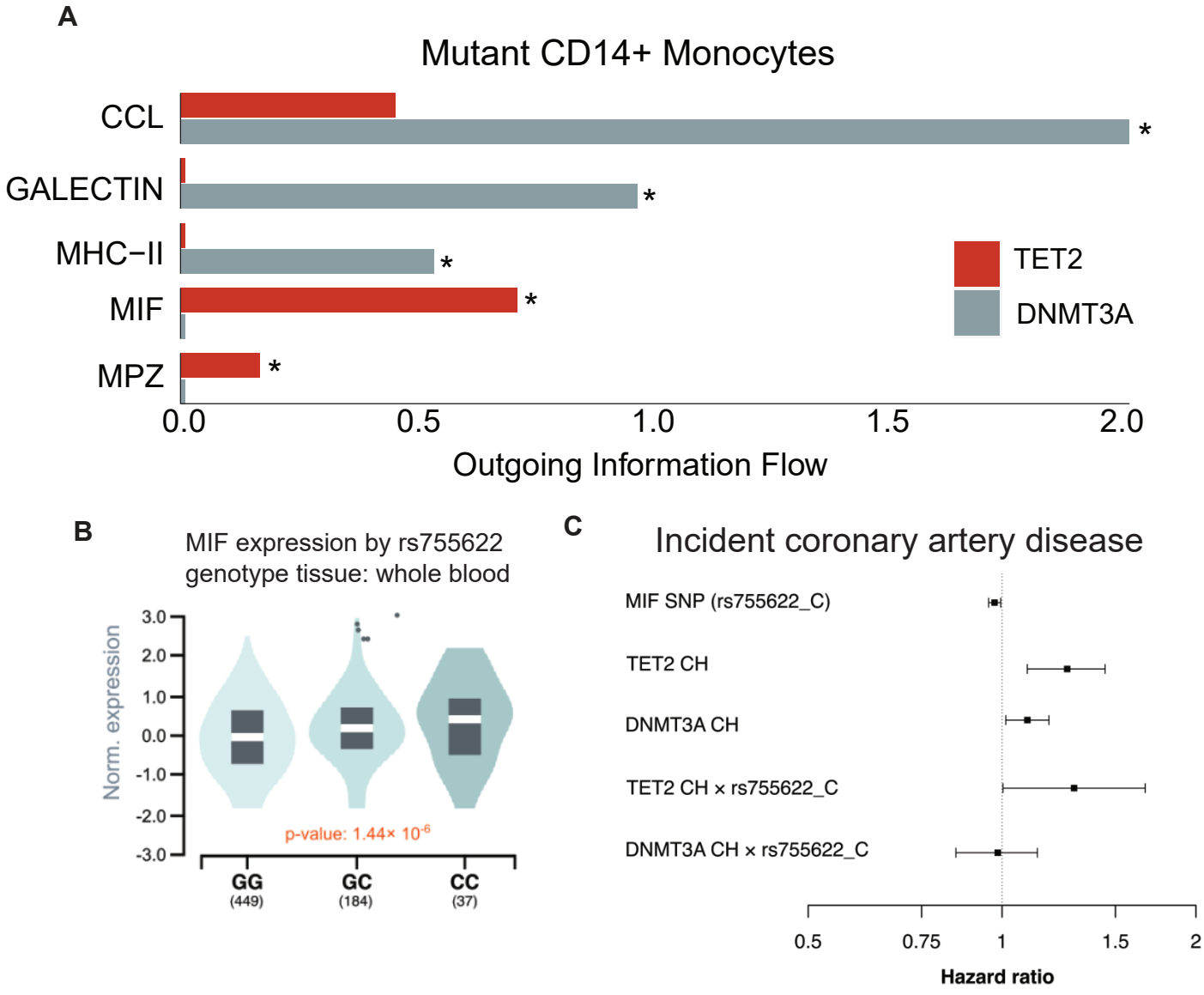

Fig. 3: MIF SNP rs755622 associates with cardiovascular disease in TET2 but not DNMT3A CH in the UK Biobank A) Quantitative outgoing signaling in mutated TET2 and DNMT3A CD14+ monocytes compared to wildtype exposes divergent utilization of signaling pathways at baseline (* = $p < 0.05$ by permutation test). B) Imputed MIF expression is enhanced with the presence of rs755622 SNP.
C) TET2 CH with the MIF risk allele (rs755622) has a significantly elevated risk of incident CAD among 340,766 individuals in the UK Biobank. The risk of incident CAD was adjusted for age, age2, sex, hypertension, diabetes, smoking history, imputed LDL levels, and 10PC of ancestry. Error bars represent 95% confidence intervals (CI).

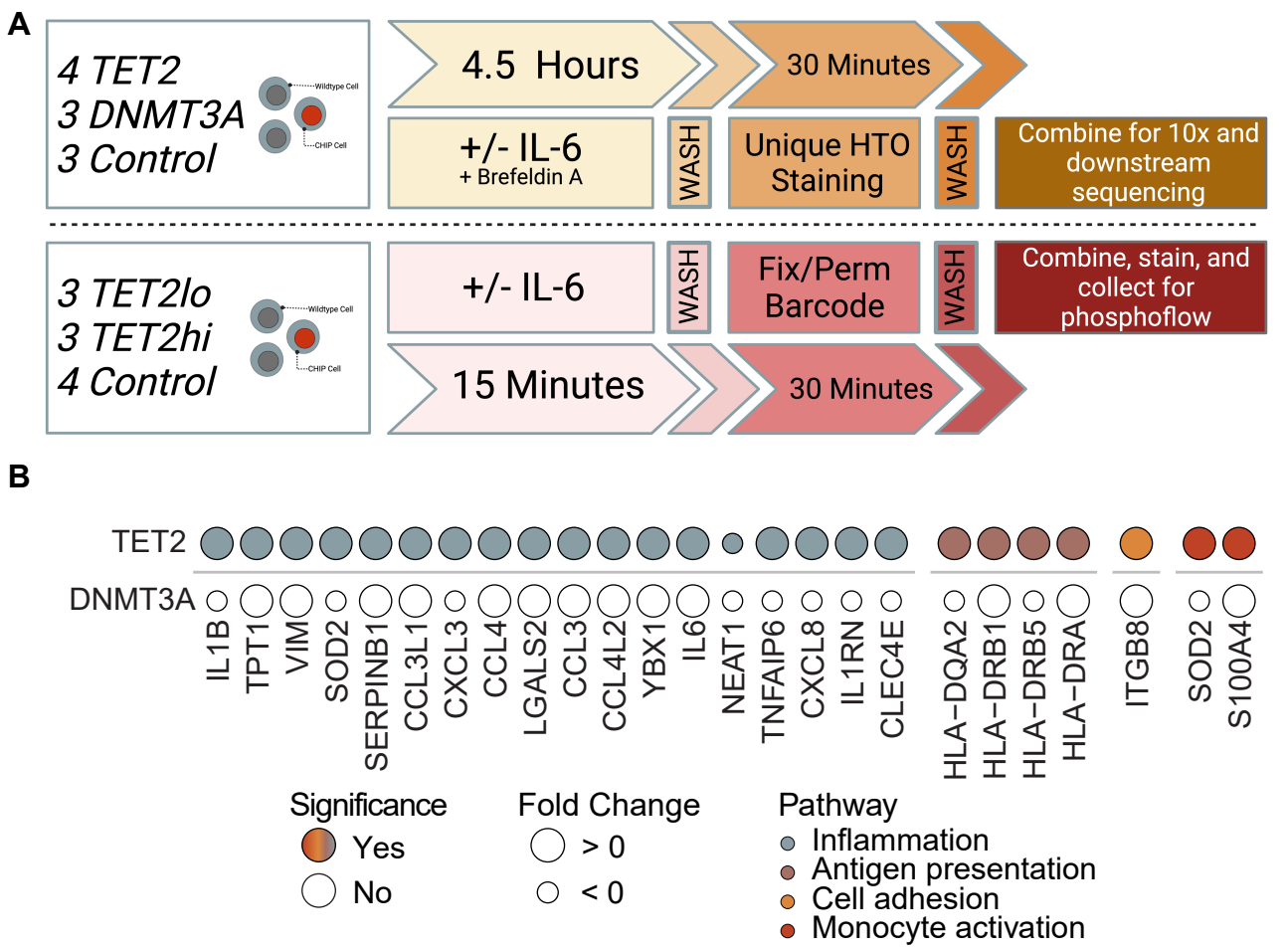

Fig. 4: IL-6 stimulation reveals cell-intrinsic differences between TET2 mutated cells and controls. A) Experimental scheme for measurement of phosphorylated STAT3 (pSTAT3) after IL-6 stimulation. VEH – vehicle, PBS – phosphate buffered saline, HTO – hashtag oligo. B) Mutant TET2 CD14+ monocytes have aberrant gene expression profiles in response to IL-6 stimulation while DNMT3A mutants exhibit minimal differences under stimulation. Differentially expressed genes via non-parametric Wilcoxon rank sum test (Seurat "FindMarkers" function, adjusted $p < 0.05$, abs(log2FC) > 0.25) in CH-mutated CD14+ monocytes compared to intra-sample wildtype controls. Rows represent comparisons between mutated and wildtype CD14+ monocytes from the same individual. The top row is TET2 (n = 4,109 mutant, n = 328 wildtype) while the bottom is DNMT3A (n = 525 mutant, n = 566 wildtype). Size reflects positive fold change and color indicates adjusted p-value. C) TET2 and DNMT3A mutants utilize different signaling pathways in response to IL-6 (* = $p < 0.05$ by permutation test). Quantitative outgoing signaling in mutated CD14+ monocytes across select pathways after IL-6 stimulation within mutated cells from single patients. D) Quantification of phospho-flow shows significantly increased pSTAT3 signaling in large VAF TET2 CH compared to both low VAF and control samples (*** = $p < 0.01$ by ANOVA with Tukey's HSD). TET2 samples were stratified by VAF. Small TET2 clones were defined as VAF < 25% (n = 3) and large as VAF > 25% (n = 3). These were compared to controls (n = 4). pSTAT3 was measured by phospho-flow cytometry after IL-6 stimulation.