Distinct characteristics of lymphoid and myeloid clonal hematopoiesis in Word Trade Center first responders

Myvizhi Esai Selvan[1], Pei-Fen Kuan[2], Xiaohua Yang[3], John Mascarenhas[4], Robert J. Klein[1], Benjamin J. Luft[3,5], Paolo Boffetta[5,6,7], Zeynep H. Gümüş[1,4,8]

[1] Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[2] Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA
[3] World Trade Center Health and Wellness Program, Stony Brook University, Stony Brook, NY, USA
[4] Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[5] Division of Infectious Diseases, Renaissance School of Medicine, Stony Brook University, Stony Brook, NY, USA
[6] Stony Brook Department of Family, Population and Preventive Medicine, Renaissance School of Medicine, Stony Brook, NY, USA
[7] Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy
[8] Marc and Jennifer Lipschultz Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Corresponding author. Zeynep H. Gümüş, Mailing address: One Gustave Levy Pl, Box 1498, New York, NY, USA. Email: zeynep.gumus@gmail.com

Conflict of Interest. The authors declare no potential conflicts of interest.

Abstract

BACKGROUND
Clonal hematopoiesis of indeterminate potential (CHIP) is a condition when healthy individuals harbor clonal mutations in myeloid (M-CHIP) and/or lymphoid (L-CHIP) cells at variant allele fraction (VAF) ≥0.02. While CHIP is associated with an increased risk of hematologic malignancy and cardiovascular disease, its association with airborne carcinogens is largely unknown.

OBJECTIVES
Here, we studied M/L-CHIP in responders to the 9/11 terrorist attacks on the World Trade Center (WTC), who were exposed to a complex mix of airborne carcinogens. Then we explored the association of CHIP mutations with phenotypes such as age, ancestry, exposure, HLA zygosity, and other clinical, laboratory, mental and cognitive data. Finally, we compared CHIP prevalence in WTC responders to 293 unexposed controls.

METHODS
Using banked peripheral blood and ultra-deep whole-exome sequencing at 250X, we characterized CHIP mutations and their interaction with clinical, mental and cognitive characteristics, exposure, peripheral blood counts, and HLA zygosity in 350 WTC responders. We used Fisher's exact test for categorical variables; Wilcoxon rank sum test for continuous variables; and logistic regression for multivariate analysis.

RESULTS
Among WTC participants, M-CHIP prevalence was 16.2% and L-CHIP 21.4%. M-CHIP prevalence increased with age (p=0.02), was elevated in previous-smokers (p=0.01), and associated with lower platelet counts (p=0.03). The most frequently occurring genes for M-CHIP were *DNMT3A*, *TET2*, *PPM1D* and for L-CHIP were *EEF1A1*, *DDX11* and *KMT2D*. Notably, harboring a *DDX11* mutation associated with a lower Montreal Cognitive Assessment score (p=6.57e-03). Overall, M/L-CHIP was more prevalent in WTC responders versus controls.

DISCUSSION
Study results will inform the development of personalized risk-adapted CHIP and cancer screening programs in individuals exposed to airborne carcinogens.

Introduction

In the aftermath of the 9/11 terrorist attacks on the World Trade Center (WTC), over 91,000 individuals were involved in the rescue and recovery efforts, clean-up of debris and restoration of essential services (Rescue & Recovery Workers - 9/11 Health). These included first responders such as firefighters, police officers, and paramedics, in addition to operating engineers, steel workers, railway tunnel workers, telecommunications workers, sanitation workers, medical examiners, and volunteers. Many had no prior training in civil disaster response (Herbert et al. 2006). Due to their unprecedented exposure to a complex mix of known or suspected airborne carcinogens, there were justified concerns regarding their elevated cancer risk. WTC responders were exposed to benzene, formaldehyde, asbestos, silica, cement dust, glass fibers, heavy metals, polycyclic aromatic hydrocarbons, and polychlorinated biphenyls, polychlorinated dibenzofurans, and dioxins (Lioy and Georgopoulos 2006). Exposures to these agents may increase cancer risk. However, the association between exposure to this complex mix and the development of hematologic malignancies is not well characterized.

Recently, it has been discovered that a condition called clonal hematopoiesis (CH) elevates risk for hematological neoplasm, cytopenia, cardiovascular disease (CVD), infection and all-cause mortality (Dawoud et al. 2021; Jaiswal et al. 2014, 2017; Jaiswal and Ebert 2019; Niroula et al. 2021; Sperling et al. 2017; Zekavat et al. 2021). Individuals with CH harbor clonal genomic mutations in their blood cells associated with certain hematologic malignancies, yet with no detectable hematologic disorders nor unexplained persistent cytopenia (Heuser et al. 2016). A subset of CH, termed clonal hematopoiesis of indeterminate potential (CHIP), involves having a clonal population of blood cells that carry a point mutation or short insertion/deletion with a variant allele fraction (VAF) ≥ 2% in a gene recurrently mutated in blood cancers. CHIP onset is strongly associated with age, reflecting the gradual accumulation of somatic mutations and quiescent clones which accompany cellular senescence (Genovese et al. 2014). CHIP has been shown to occur in the myeloid lineage (M-CHIP) (Abelson et al. 2018; Desai et al. 2018), and studies have mostly focused on genes known to be recurrently mutated in myeloid malignancies (Bick et al. 2020b; Jaiswal et al. 2017). However, CHIP can also occur in the lymphoid lineage (L-CHIP), and contribute to the risk of lymphoid malignancies, an area that is less documented (Agathangelidis et al. 2018; Condoluci and Rossi 2018; Niroula et al. 2021; Singh et al. 2020; von Beck et al. 2023; Weigert et al. 2012).

Here, we studied the prevalence of M- and L-CHIP (M/L-CHIP) in 350 WTC responders using ultra deep (250X) whole-exome sequencing (WES). Our study has several unique aspects. First, we studied participants from the WTC Health Program General Responders Cohort (GRC), which is a relatively diverse group of workers with respect to their occupational titles. Second, we leveraged the wealth of additional data we collected on these responders, including demographics (gender, race/ethnicity), WTC exposure, blood counts (including various immune cell subsets), smoking history, body mass index (BMI), mental, cognitive and general health characteristics, as well as HLA alleles derived from the WES datasets, to study their association with M/L-CHIP. Third, utilizing WES enabled us to study the full spectrum of both M/L-CHIP

mutations without being *a priori* restricted to a particular gene-set. This is critical, as the literature and the definition of what constitutes CHIP genes and mutations continues to evolve, our raw data files will continue providing a rich resource for future investigations. Overall, our results demonstrate that deep sequencing of M/L-CHIP mutations, integrated with other metrics, enables a more personalized risk-adapted assessment for CHIP associated etiologies in persons exposed to airborne carcinogens. We anticipate that the study findings will inform the development of CHIP-related diagnostic measures which can be incorporated into screening programs for cancer and other inflammation related conditions.

Methods

The World Trade Center Health Program (WTCHP) General Responder Cohort (GRC) includes participants in the rescue, recovery, and cleanup efforts at the WTC site after 9/11/2001 on the basis of eligibility criteria, which included type of tasks, site location, and dates and hours worked/volunteered (Dasaro et al. 2017; Solan et al. 2013). A subset of participants located in Long Island were referred to the WTC Clinical Center located at Stony Brook University (SBU, B. Luft, PI) (referred to here as the WTC-SBU). This Center was established in 2002 to enlist, monitor and treat WTC-related conditions in individuals with documented WTC-response experience. Monitoring protocol included self-administered physical and mental health questionnaires followed by a physical examination, laboratory tests, spirometry, and a chest radiograph. This population has been characterized very carefully and included in previous studies of various health conditions associated with response to the WTC disaster, including post-traumatic stress disorder (PTSD), prostate cancer, cognitive impairment, and COVID-19 (Clouston et al. 2016, 2019, 2022; Morozova et al. 2021). Routine monitoring visits were performed every 12–18 months. Compared with the whole WTCHP GRC, the WTC-SBU cohort includes relatively more law enforcement personnel and men and fewer individuals with a low level of education (i.e. without a high-school degree) (Clouston et al. 2016). The study was approved annually under IRB #604113 by the Committees on Research Involving Human Subjects at SBU. More than 95% provided written informed consent for their data to be used for research.

Sample Collection. Consented participants provided whole blood samples during their annual clinical checkup visits between 2016 and January 2019. Samples were collected into Vacutainer Plastic K2EDTA (containing ethylenediaminetetraacetic acid) Tubes (BD, 367527) and stored at -80C until analysis.

Demographics. From the WTC-SBU cohort, we randomly selected 350 samples for WES from 345 unique participants. Detailed phenotypic characteristics are in Table 1A. Over 90% were male and white; 7.5% were female, and 9% were non-white. Less than 8% of participants were ≤ 55 years old, 49.6% were 56-60 years old, 24.3% were 61-65 years old, 9.3% were 66-70 years old and 9% were older than 70. Of these, 4.3% were current smokers, 44.6% were former smokers, and 51% were never-smokers.

Clinical data include WTC exposure history, body mass index (BMI), and mental, cognitive and general health characteristics (including cholesterol and triglyceride levels) assessed at each monitoring visit. WTC exposure history was assessed during the enrollment interview at the WTC Health Program and has been described in detail (Wisnivesky et al. 2011). Briefly, an exposure variable was created using total time spent working at Ground Zero or on the debris pile (Wisnivesky et al. 2011). Post-traumatic stress disorder (PTSD) symptoms were assessed using the PTSD Checklist (PCL) a 20-item self-report measure modified to assess symptoms over the past month, which has excellent psychometric properties, convergent validity and internal consistency (Wilkins et al. 2011). Mild cognitive impairment (MCI) was measured using the Montreal Cognitive Assessment (MoCA), a widely used objective multidomain test (Nasreddine et al. 2005).

Blood and lipid markers were measured concurrently with blood collection for CHIP analysis from 294 (out of 345) participants. Blood markers include platelet count, basophil count, lymphocyte count, neutrophil count, segmented neutrophil count, eosinophil count, monocyte count, LMR (ratio lymphocyte/monocyte), white blood cell count (WBC), red blood cell count (RBC), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), MCH concentration (MCHC), and red cell distribution width (RDW). Lipid markers include total cholesterol, triglycerides, HDL cholesterol, LDL cholesterol, and very low LDL (VLDL).

Sequencing Analysis. Deep WES for 350 WTC-SBU samples was performed at Azenta Inc (Burlington, MA, USA) using the HiSeq 2500 system (Illumina, San Diego, CA, USA) using standard protocols. Sample quality control (QC), library construction, and sequencing were performed in line with industry quality of process standards. To allow for ample sensitivity and accuracy in CHIP mutation calling even at low clonal frequencies, a median 250X coverage was used. Sequencing generated high-quality 150-bp, pair-end read data in standard fastq format.

Data Pre-Processing for Variant Discovery. To pre-process and QC raw sequence reads, we used fastp (Chen et al. 2018). Specifically, we trimmed the adapters and filtered out bad reads (low quality, too short or too many unknown bases). Next, we adhered to the Genome Analysis Toolkit (GATK, https://software.broadinstitute.org/gatk) best practices for data pre-processing to generate analysis-ready bam files from the fastq files. Briefly, we aligned the sequence reads to the Genome Reference Consortium Human Build GrCh38 using BWA-MEM (Li 2013) followed by duplicate marking using Picard (http://broadinstitute.github.io/picard) and base quality score recalibration using GATK. We used these bam files for calling germline variants and somatic mutations.

Germline Variant Calling and Kinship Analysis. We called germline variants using HaplotypeCaller in GATK in GVCF mode, jointly genotyped individual gVCF files for all autosomes, filtered variants by variant quality score recalibration in GATK and finally removed sites with ≥ 20% missing data. Next, to exclude any genetic duplicates among WTC participants, we performed kinship analysis with germline variants using KING software (Manichaikul et al. 2010) and identified three duplicate pairs (kinship coefficient > 0.354). One

pair was from the same individual at different time points. From these 6 samples, we included the most recent sample collected from the participant with a duplicate pair.

Somatic Variant Calling. After kinship analysis, we performed somatic variant calling on the bam files of the remaining 345 samples using the GATK Mutect2 pipeline in tumor-only mode. To exclude likely germline calls and sequencing artifacts, we provided external reference of germline variants from gnomAD and a Panel Of Normals (PON) to Mutect2. To create the PON, we used WES data from the publicly available Genotype-Tissue Expression (GTEx) cohort (phs000424), where we filtered for 70 young individuals (aged ≤ 40 years) as they would be less likely to harbor CHIP. To identify somatic mutations with high confidence, we applied the orientation bias and then PASS filters of Mutect2.

Somatic Variant Filtering to identify M/L-CHIP carriers. We considered 76 somatic driver genes for M-CHIP (Bick et al. 2020b; Kar et al. 2022), and 235 genes for L-CHIP (Niroula et al. 2021) (Supplementary Table S1), and filtered for predefined CHIP variants at a Mutect2 VAF >2%. For L-CHIP somatic variants, we specifically filtered for *pathogenic* (variants curated from cBioPortal) or *putative* (variants that alter canonical protein sequence) (Niroula et al. 2021).

To remove likely artifacts, we applied additional QC filters. Briefly, we filtered for somatic variants at a minimum depth of 20 reads, minimum 3 reads (supporting the mutant allele) and at least one read in both forward and reverse directions (supporting the reference and mutant alleles). In addition, we excluded somatic variants observed in gnomAD with allele frequency ≥ 0.1% and with an observed frequency >1% in the cohort (unless previously reported to be involved in hematologic malignancies). Next, we annotated the identified variants for pathogenicity using Combined Annotation Dependent Depletion (CADD) score and excluded variants with a scaled CADD score < 10. Finally, we implemented additional filters for *putative* L-CHIP mutations (Supplementary Table S1), including a minimum alternate allele read of 5, maximum VAF of 0.2 and at least two reads in both forward and reverse direction supporting the alternate allele.

We included two filtering exceptions. The first one is, based on reports that M-CHIP variants in the *U2AF1* gene cannot be reliably identified in the human GrCh38 reference genome due to unintended replication of the *U2AF1* locus in chromosome 21 (Miller et al. 2022). Therefore, we realigned *U2AF1* using build GrCh37 and repeated somatic variant calling. Second, it was reported that the variant *ASXL1*- G646Wfs*12 with a VAF ≥ 10% was a true CHIP variant (Vlasschaert et al. 2023) and not a sequencing artifact. Yet, we did not identify any WTC responders with CHIP mutations in *U2AF1* gene or at *ASXL1-G646Wfs*12*. After filtering, we considered all responders who harbored at least one known M/L-CHIP somatic variant as having CHIP. We list the identified M/L-CHIP mutations in Supplementary Table S2.

CHIP Prevalence in WTC Responders and Unexposed Controls. We next compared M/L-CHIP prevalence in the WTC cohort to unexposed controls. For controls, we used M/L-CHIP calls from 293 healthy controls recruited within the Mount Sinai Crohn's and Colitis Registry (MSCCR) cohort (STUDY-11-01669) which was processed using the same analytical pipeline

as described above (Nathan et al. 2023). We provide control phenotypic characteristics in Table 1B. However, to make the median coverage of the WTC cohort comparable to that of the controls, we downsampled the WTC data using Picard's DownsampleSam tool to keep 41.5% of the total reads so that both WTC and MSCCR cohorts would be comparable as the total number of aligned bases (PF_ALIGNED_BASES).

HLA zygosity. To test whether the zygosity of the HLA alleles was associated with the prevalence of CHIP, we performed HLA-typing using HLA-HD (Kawaguchi et al. 2017), and determined HLA class I and class II alleles of each participant with precision up to 6-digits.

Statistical Analysis. To understand the association of CHIP with all available factors, we used standard statistics. Briefly, we summarized categorical variables as counts and percentages and used Fisher's exact tests for analyses. We summarized continuous variables using median and median absolute deviation, and used Wilcoxon rank sum tests for analyses. For association analyses, as the majority of the responders were White, we collapsed the race variable into two categories: White and non-White/ unknown. We collapsed WTC exposure as a variable into three categories, including very low/low, intermediate and high/very high. We excluded the missing values from the association analysis.

We considered several WTC outcome subgroups, including a) CHIP positive (at least one M or L-CHIP mutation, N=118), b) M-CHIP positive, (at least one M-CHIP mutation, N=56), c) L-CHIP positive (at least one L-CHIP mutation, N=74), d) *DNMT3A* mutation (at least one mutation on *DNMT3A* gene, N=22), e) *TET2* mutation (at least one mutation in *TET2* gene, N=15), f) *PPM1D* mutation (at least one mutation in *PPM1D* gene, N=11), g) *EEF1A1* mutation (at least one mutation in *EEF1A1* gene, N=18), and h) *DDX11* mutation (at least one mutation in *DDX11* gene, N=13). For all these cases, we defined participants who did not harbor any M/ L-CHIP mutation as CHIP-negative (N=227).

For association analysis, we first performed univariate analysis using the individual factors as covariates. We provide false discovery rate adjusted p-values in Supplementary Table S3. Then, among the characteristics that were significant from marginal associations, we fitted a multivariate logistic regression using presence of CHIP as outcome variable and the significant characteristics as covariates. We considered a p-value < 0.05 as statistically significant.

Results

To study the prevalence of CHIP mutations in WTC responders, we performed deep WES at 250X on 350 banked blood samples from 345 participants (Figure 1A). The participants were aged 48-90 years (median, 59 years), with no previous hematologic malignancy at enrollment, with demographics detailed in Table 1A. We applied rigorous QC metrics and filters (Figure 1B, Supplementary Table S1), to reveal M/L-CHIP prevalence, which we then associated with age, ancestry, exposure, HLA zygosity, and other clinical, laboratory, mental and cognitive data.

Finally, we compared CHIP prevalence in the WTC cohort to 293 unexposed controls (demographics in Table 1B) from the New York City area.

Prevalence and characteristics of L/M-CHIP mutations. After sample QC, 16.2% (56 participants) of the 345 WTC participants harbored 71 M-CHIP mutations in13 genes, while 21.4% (74 participants) harbored 85 L-CHIP mutations in 43 genes (Supplementary Table S2). Overall, 34.2% (118/345) of the WTC participants harbored at least one M- or L-CHIP mutation. We observed greater clonal complexity, identified by the presence of more than one mutation in 7.0% (24/345) of WTC participants, where 3.5% (12 participants) carried both mutation types. The majority (>80%) of the participants who harbored M/L-CHIP mutations carried only a single mutation (Figure 2-A, D). Of the M-CHIP mutations, 39% were non-synonymous, 30% were stop-gain, 23% were frameshift deletions and the rest were frameshift insertions and splicing (Figure 2B). For L-CHIP, 87% were non-synonymous, 9% were stop-gain and the rest were frameshift indels (Figure 2E). The top M-CHIP genes were *DNMT3A*, *TET2*, *PPM1D* and *ASXL1* (Figure 2C). The top L-CHIP genes were *EEF1A1*, *DDX11*, *KMT2D*, *ATM* and *FAT2* (Figure 2F). Figure 3A shows the VAF distribution of the M/L-CHIP mutations. The highest VAF was in *TET2* mutations (Figure 3C).

Additional factors on CHIP prevalence. We studied the associations between M- and/or L-CHIP mutations and i) clinical data - age, gender, race, smoking status, cardiovascular diseases (CVD), stroke, body mass index (BMI), Montreal Cognitive Assessment (MoCA), PTSD Checklist (PCL); ii) laboratory data (blood and lipid counts) and iii) HLA zygosity (class I and class II alleles) (Supplementary Table S3, Figure 4). Further, we studied the associations driven by the top mutated M-CHIP (*DNMT3A*, *TET2*, *PPM1D*) and L-CHIP (*EEF1A1*, *DDX11*) genes. Figure 4 summarizes the factors with at least one significant association in the univariate analysis. We did not observe associations in L-CHIP positive WTC cases with the factors considered in this study. In Supplementary Table S3 we list all associations with the participant characteristics we considered.

Clinical data. Several studies (Bick et al. 2020b; Kar et al. 2022) have observed that older age, smoking and other clinical characteristics are associated with CHIP. We report the distribution of M/L-CHIP in the WTC cohort with respect to age in Table 2. Consistent with literature on CHIP and aging (Bick et al. 2020b; Niroula et al. 2021), participants with CHIP mutations, specifically M-CHIP mutations and the top M-CHIP genes, *DNMT3A* and *TET2,* were associated with older age compared to CHIP negative (individuals without M/L-CHIP mutations) cases. Age remained a significant factor in the multivariate logistic regression model too. The median age of participants with CHIP, M-CHIP, *DNMT3A*, *TET2* were 61, 62, 66.5 and 64 years, respectively. Whereas the median age of CHIP negative participants was 59 years. Former smokers were 44.6% of the WTC cohort, and a smoking history was associated with M-CHIP in both uni- and multivariate regression models. M-CHIP mutations and particularly *DNMT3A* were also associated with lower BMI compared to CHIP negative participants. Finally with regards to the mental health and cognitive characteristics, participants with *DDX11* mutations (a top L-CHIP gene), were associated with higher PCL scores and lower MoCA scores. MoCA scores remained significant in the multivariate logistic regression model. The median MoCA score for

participants with *DDX11* mutation was 22, which was within the range indicative of mild cognitive impairment.

Laboratory data. Previous studies have revealed associations of M-CHIP with myeloid cell parameters (e.g. platelet, red blood cell, neutrophil and monocyte counts), and L-CHIP with elevated lymphocyte counts (Niroula et al. 2021). In the WTC cohort, M-CHIP positive cases had lower platelet counts (uni- and multivariate regression models) as compared to CHIP negative cases. Participants with the most frequently observed M-CHIP mutations, *DNMT3A* and *TET2*, were associated with lower absolute lymphocyte counts. *TET2* mutation carriers were additionally associated with lower RBC counts, and higher segmented neutrophils. At the same time, *PPM1D* mutation carriers were associated with lower platelet counts and higher mean corpuscular hemoglobin (MCH). Further, *DDX11* mutation carriers were associated with higher absolute lymphocytes, mean corpuscular volume (MCV), lymphocyte monocyte ratio (LMR), and lower segmented neutrophils.

HLA zygosity. We tested potential major histocompatibility complex (MHC) determinants that might be associated with CHIP positivity, and observed an association with a higher proportion of HLA-DMB homozygosity, which remained significant in the multivariate logistic regression model. At the same time, *PPM1D* mutation status was associated with a higher proportion of DQA1 homozygosity. Finally, participants with the top L-CHIP mutation gene, *EEF1A1*, were associated with a lower proportion of DPA1 homozygosity.

<u>M/L-CHIP prevalence was higher in WTC cohort versus unexposed controls.</u> To understand the impact of WTC debris-exposure on CHIP prevalence, we compared the WTC cohort to 293 healthy unexposed controls from New York area (Nathan et al. 2023). After downsampling, in the WTC cohort the prevalence of detectable M-CHIP mutations decreased to 7.5% (26/345) and L-CHIP to 9.9% (34/345). Yet, for unexposed controls, M-CHIP prevalence was 3.1% (9/293) and L-CHIP 2.0% (6/293). Grouping the participants into age strata, both CHIP mutation types were still generally more prevalent in the WTC cohort (Figure 5A and B, Supplementary Table S4). L-CHIP mutations were statistically significantly more prevalent in WTC participants ≤ 55 years old in comparison to unexposed controls (p=0.04).

<u>*TET2* M-CHIP and *EEF1A1* and *DDX11* L-CHIP mutations were more prevalent in WTC cohort versus unexposed controls.</u> In terms of frequency of the top mutated genes (Figure 5C and D), we observed a significant difference between the number of M-CHIP *TET2* mutations in the WTC cohort versus controls (p=0.02). Furthermore, we observed a significantly higher number of L-CHIP *EEF1A1* (p< 0.0001) and *DDX11* (p=0.04) mutations in the WTC cohort vs controls.

Discussion

CHIP is an early marker of risk for numerous maladies, including cardiovascular diseases (CVD, both ischemic heart disease and stroke), hematologic malignancies, severe COVID-19 outcomes, several solid cancers, infection, and all-cause mortality (Bhattacharya et al. 2022;

Bick et al. 2020a; Bolton et al. 2021; Jaiswal et al. 2014, 2017; Jaiswal and Ebert 2019). There is also increasing appreciation that some CHIP mutations are tied to inflammatory consequences such as anemia (Girelli and Busti 2019). While CHIP presence is not a prerequisite for progression to malignancy, it significantly increases risk. The transformation rate from M-CHIP to blood cancer is estimated around 0.5-1.0% per year, roughly 13 times greater than the incidence in the general population (Genovese et al. 2014; Heuser et al. 2016). L-CHIP was also associated with higher incidence of lymphoid malignancy (Hazard Ratio ~4) (Niroula et al. 2021). Here, we report M and/or L-CHIP mutation prevalence and patterns in a population of WTC responders. This is arguably the first study to document L-CHIP mutations within a WTC cohort and within a cohort of occupationally exposed individuals in general. In addition, we uniquely leveraged a rich dataset of clinical, laboratory, mental, cognitive and HLA zygosity information to study their associations with CHIP.

Despite the severe toxic inhalation exposures of WTC responders, to date, only one study has assessed their CHIP prevalence. Focused on FDNY first responders, this study used targeted sequencing on 237 genes known to be recurrently mutated in hematologic malignancies, and reported elevated CHIP in comparison to unexposed controls (Jasra et al. 2022). However, increased cancer mortality has also been reported for WTC responders who were not part of the FDNY (Li et al. 2023). Indeed, the combined WTC population (pooled from FDNY, WTC Health Registry and GRC) has a documented elevated risk for melanoma, as well as cancers of the prostate, thyroid and tonsil compared to the general population (Li et al. 2022), and WTC GRC cohort observed a modest elevation of hematologic malignancies (Shapiro et al. 2019). Here, our deep WEX of the WTC GRC, which includes a more diverse group of first responders than the FDNY, revealed a high prevalence of both M-CHIP (16.2%) and L-CHIP (21.4%) mutations, with an overall prevalence of 34.2%. While the elevation in M-CHIP with respect to WTC debris exposure in the WTC GRC was consistent with the previous report on targeted sequencing of samples from WTC FDNY participants, the genes, mutations and filtering criteria were different between the studies, thus preventing direct comparisons.

Currently, there are no standard definitions of CHIP, and the continual identification of specific driver mutations and their subsequent inclusion (and exclusion) in CHIP gene lists changes the definition of the condition. For this reason, the 2022 World Health Organization definition of CHIP is not gene or variant specific, only referring to somatic mutations of myeloid malignancy-associated genes detected in the blood or bone marrow at a VAF of ≥ 2% (≥4% for X-linked gene mutations in males) in individuals without a diagnosed hematologic disorder or unexplained cytopenia (Khoury et al. 2022; Steensma et al. 2015). As a result of this evolving landscape, different research studies report results on different variants and genes, without an international consensus on how CHIP should be defined (Bick et al. 2020b; Niroula et al. 2021). Furthermore, there are no established ‘gold standard’ analysis pipelines devoted to CHIP analysis. For example, while a VAF of ≥2 % is recommended, there are no guidelines on what the corresponding read depth would be, as more mutations with a true VAF of 2% will be missed at a read depth of 30X than 250X. In addition, the filters to remove technical artifacts vary between studies. Hence, both the biological and technical definitions of CHIP vary between studies, making comparisons between studies and with population level controls challenging.

Because of the differences described above, instead of directly comparing CHIP prevalence in our study with population level cohorts (ie. UK BioBank and TOPMED), we analyzed an unexposed control cohort using the same analytical pipeline (Figure 1B). Furthermore, we downsampled the WTC GRC findings to match sequencing depth of the control cohort. We observed that both M- and L-CHIP were generally more prevalent in WTC responders than unexposed controls. This suggests that WTC exposure or other characteristics of WTC responders, such as chronic stress following the WTC experience (Rogers et al. 2022) is associated with elevated M/L-CHIP, especially for those 55 and younger.

Our findings on age-associated increase in M-CHIP prevalence in the WTC cohort were consistent with its well-documented association with age in other population-level cohorts (Bick et al. 2020b; Niroula et al. 2021). These studies have documented that lifestyle factors like smoking history, BMI were also positively associated with M-CHIP prevalence (Komic et al. 2023). While we observed a similar association for smokers in the WTC cohort, a larger cohort size is needed to understand the association of BMI with M-CHIP in this population.

Genes with mutations most commonly associated with M-CHIP in the literature are involved in epigenetic regulation (*TET2, ASXL1, DNMT3A,* and *IDH2*), RNA splicing (*SF3B1, SRSF2, U2AF1*), cell signaling (*JAK2* and *NRAS*), and DNA repair and cell cycle regulation (*TP53*) (Heuser et al. 2016; Sperling et al. 2017; Steensma et al. 2015). Similar to these trends, we observed M-CHIP mutations in most of these genes (see Figure 2C). For L-CHIP mutations, prior studies have reported an even distribution across a larger number of genes (Niroula et al. 2021). While we did observe some of the previously reported L-CHIP gene mutations, 21% of L-CHIP mutations were in *EEF1A1* and 15% in *DDX11* (see Figure 2F). This signal was driven by three putative mutations, *DDX11*:P368S, *EEF1A1*:E293K and *EE1F1A*:V315L, which have known associations with hematologic malignancies (Ma et al. 2022; Papaemmanuil et al. 2013, 2016; Tyner et al. 2018). L-CHIP mutations in *EEF1A1* and *DDX11* remained significantly more prevalent in the WTC cohort in comparison to unexposed controls as well. Note that, *EEF1A1* and *DDX11* were not among the genes considered in the previous targeted sequencing WTC-FDNY study (Jasra et al. 2022).

As the WTC responder population ages, multiple studies are reporting an increased risk for neurocognitive and motor dysfunction that resembles neurodegenerative diseases, in addition to the presence of cortical atrophy and cognitive impairment at midlife within this population. These risks have been associated not only with physical exposures at the WTC site, but also with chronic post-traumatic stress disorder (PTSD) (Clouston et al. 2022). We therefore investigated the association of the available mental and cognitive data on the WTC cohort with CHIP, and observed an association for L-CHIP *DDX11* gene with higher PCL scores (indicative of PTSD severity) and lower MoCA scores (indicative of mild cognitive impairment). *DDX11* is a helicase that participates in cellular processes that alter RNA secondary structure, and helicase dysfunction has been implicated in multiple syndromes (Uchiumi et al. 2015), including neurodegeneration (Lovell et al. 2000).

As CHIP can associate with inflammatory diseases, we leveraged our WEX data to further perform Human Leukocyte Antigen (HLA) typing (both classes I and II) and test the diversity in the HLA genes in association with M/L-CHIP, in the context of the heterozygote advantage hypothesis. This hypothesis posits that individuals who harbor heterozygous genotypes at the HLA genes are able to display a greater variety of antigenic peptides than those with homozygous genotypes at the HLA genes, leading to an immune response to a broader range of antigens (Doherty and Zinkernagel 1975; Pagliuca et al. 2022; Penn et al. 2002). Testing for homozygosity in Class I and Class II HLA genes with CHIP, overall CHIP prevalence was statistically significantly associated with higher proportion of *HLA-DMB* homozygosity. We also observed a weaker signal for *PPM1D* mutations with higher proportion of *HLA-DQA1* homozygosity; and *EEF1A1* mutations with a lower proportion of *HLA-DPA1* homozygosity. These are HLA Class II genes expressed in antigen presenting cells, including lymphocytes, and future studies are needed to understand the interplay between homozygosity in these genes, immune deficiencies and cancer (Ogobuiro et al. 2023; Pagliuca et al. 2022; Planelles et al. 2006).

In exploratory analyses, we compared the association of blood counts data in WTC responders with the UK BioBank population. Specifically, we investigated M/L-CHIP with myeloid parameters (e.g. platelet, red blood cell, neutrophil and monocyte counts), and lymphocyte counts (Niroula et al. 2021). In the WTC cohort, consistent with the population-level UK Biobank cohort, *TET2* mutations were negatively associated with lymphocyte counts (Kar et al. 2022) and *PPM1D* mutations were negatively associated with platelet counts (Kamphuis et al. 2023). However, most associations we observed between M-CHIP mutations and blood count were not statistically significant in a multivariate regression analysis, when age and other factors were taken into account. We also observed opposite trends from the UK Biobank cohort in the association of M-CHIP with platelet counts, and *TET2* mutations with segmented neutrophil counts (though UK Biobank uses neutrophil counts). To understand whether this discrepancy is due to WTC-debris exposure, further investigation is warranted.

Recent studies on the UK BioBank report that in individuals without detectable CHIP at the time of lab measurements, abnormal complete blood count (CBC) labs can still be predictive of risk for future CHIP (Gu et al. 2023; Weeks et al. 2023), and that those that harbor abnormal myeloid blood cell parameters and CHIP are at the highest risk for developing myeloid malignancies. Similarly, those with elevated lymphocyte counts were reported to be at elevated risk for lymphoid malignancies, and those additionally with L-CHIP mutations at highest risk. These reports suggest that the WTC participants with elevated myeloid and/or lymphoid blood counts, together with CHIP mutations, may be at highest risk for future malignancies. Currently, there is interest in treating CHIP patients with activated inflammatory pathways with IL-1β inhibitor and *NLRP3* inhibitors in addition to *IRAK1* inhibitors (Kanagal-Shamanna et al. 2024; Wang et al. 2021). Thus, it may be appropriate to investigate such interventions in WTC responders at the highest risk to prevent or delay overt myeloid neoplasm, CVD, major bleeding, and infection.

Overall, our study focused on the impact of exposure from 9/11 attacks on WTC first responders, and yet has implications beyond this population. While the WTC exposure included a complex mix of carcinogenic compounds specific to the site, exposure itself was a main determinant of risk for CHIP. This warrants future risk studies on populations exposed to debris (both from building fires and collapses) from modern warfare in populated cities.

Study findings will inform future research for risk stratification for those at highest risk, who may need to undergo more frequent screening and monitoring to prevent or delay cancer development. Despite a higher relative risk of malignancies in CHIP positive individuals, the absolute incidence remains low in the CHIP positive (Heuser et al. 2016). Yet, to identify individuals at highest risk for cancers, genetic information on CHIP can additionally be combined with demographic, clinical and laboratory data to build risk models. Efforts towards this end are already starting to gain traction (Weeks et al. 2023). We anticipate that with the establishment of CHIP clinics at many academic centers, longitudinal monitoring of CHIP individuals will inform improved risk models. Those at highest risk can then be followed, have risk assessment and eventually participate in clinical trials in CHIP clinics.

This study should be considered in the context of its limitations. Here, we compared the prevalence of CHIP in WTC rescue and recovery workers to unexposed controls. While the control participants were from the New York Area, they were not first responders. Previous studies on CHIP have additionally considered mosaic chromosomal alterations (mCAs) in the peripheral blood of individuals. We were unable investigate the prevalence of mCAs, as WES typically does not include enough heterozygous sites to allow their detection at low cell fractions (Loh et al. 2020). The study size is another weakness; the limited number of study participants makes it harder to detect modest effects. Thus, we cannot distinguish between no evidence for association due to limited power and an actual lack of association (e.g. between age and L-CHIP). This is especially important in understanding the differences in the association trends for M-CHIP positive cases with clinical and lab data in the WTC cohort versus the UK BioBank. Finally, the multiplicity of analyses we conducted may lead to results which are statistically significant at the 5% level, yet due to chance (false positives): in this respect, our results need confirmation in independent populations.

Data availability. We will deposit the whole exome sequencing data of the WTC responders in the database of Genotypes and Phenotypes (dbGaP).

Code Availability. All analyses were performed utilizing standard publicly available software. Any specific analysis code details are available from the authors upon request.

Acknowledgements
This work was supported by grants to P.B., J.M. and Z.H.G. from the Centers for Disease Control and Prevention, National Institute of Occupational; Safety and Health (award # 1U01OH012187-01); to Z.H.G from Cancer Moonshot R33 award # CA263705-01; to B.J.L. and

X.Y. from the Centers for Disease Control and Prevention (CDC/NIOSH 75D301-22-C-15522); and in part through the computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences.

Authors' contributions

P.B and Z.H.G. conceived and designed the study. M.E.S., P.K., R.J.K., and Z.H.G. wrote the manuscript. B.J.L. and X.Y. recruited participants and handled sample and data collection. Z.H.G led sample sequencing at Azenta Inc. M.E.S. performed sequence analyses and P.K. performed statistical analyses. All authors were involved in the interpretation of the results. B.J.L., J.M. and P.B. edited the manuscript. Z.H.G. supervised the study. All authors approved the final manuscript.

References

Abelson S, Collord G, Ng SWK, Weissbrod O, Mendelson Cohen N, Niemeyer E, et al. 2018. Prediction of acute myeloid leukaemia risk in healthy individuals. Nature 559:400–404; doi:10.1038/S41586-018-0317-6.

Agathangelidis A, Ljungström V, Scarfò L, Fazi C, Gounari M, Pandzic T, et al. 2018. Highly similar genomic landscapes in monoclonal B-cell lymphocytosis and ultra-stable chronic lymphocytic leukemia with low frequency of driver mutations. Haematologica 103:865–873; doi:10.3324/HAEMATOL.2017.177212.

Bhattacharya R, Zekavat SM, Haessler J, Fornage M, Raffield L, Uddin MM, et al. 2022. Clonal Hematopoiesis Is Associated With Higher Risk of Stroke. Stroke 53:788–797; doi:10.1161/STROKEAHA.121.037388.

Bick AG, Pirruccello JP, Griffin GK, Gupta N, Gabriel S, Saleheen D, et al. 2020a. Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. Circulation 141:124–131; doi:10.1161/CIRCULATIONAHA.119.044362.

Bick AG, Weinstock JS, Nandakumar SK, Fulco CP, Bao EL, Zekavat SM, et al. 2020b. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature 2020 586:7831 586:763–768; doi:10.1038/s41586-020-2819-2.

Bolton KL, Koh Y, Foote MB, Im H, Jee J, Sun CH, et al. 2021. Clonal hematopoiesis is associated with risk of severe Covid-19. Nat Commun 12; doi:10.1038/S41467-021-26138-6.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890; doi:10.1093/BIOINFORMATICS/BTY560.

Clouston SAP, Hall CB, Kritikos M, Bennett DA, DeKosky S, Edwards J, et al. 2022. Cognitive impairment and World Trade Centre-related exposures. Nat Rev Neurol 18:103–116; doi:10.1038/S41582-021-00576-8.

Clouston SAP, Kotov R, Pietrzak RH, Luft BJ, Gonzalez A, Richards M, et al. 2016. Cognitive impairment among World Trade Center responders: Long-term implications of re-experiencing the 9/11 terrorist attacks. Alzheimers Dement (Amst) 4:67–75; doi:10.1016/J.DADM.2016.08.001.

Clouston SAP, Kuan P, Kotov R, Mukherjee S, Thompson-Carino P, Bromet EJ, et al. 2019. Risk factors for incident prostate cancer in a cohort of world trade center responders. BMC Psychiatry 19; doi:10.1186/S12888-019-2383-1.

Condoluci A, Rossi D. 2018. Age-related clonal hematopoiesis and monoclonal B-cell lymphocytosis/chronic lymphocytic leukemia: a new association? Haematologica 103:751–752; doi:10.3324/HAEMATOL.2018.191098.

Dasaro CR, Holden WL, Berman KD, Crane MA, Kaplan JR, Lucchini RG, et al. 2017. Cohort Profile: World Trade Center Health Program General Responder Cohort. Int J Epidemiol 46; doi:10.1093/IJE/DYV099.

Dawoud AAZ, Gilbert RD, Tapper WJ, Cross NCP. 2021. Clonal myelopoiesis promotes adverse outcomes in chronic kidney disease. Leukemia 2021 36:2 36:507–515; doi:10.1038/s41375-021-01382-3.

Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S, et al. 2018. Somatic mutations precede acute myeloid leukemia years before diagnosis. Nat Med 24:1015–1023; doi:10.1038/S41591-018-0081-Z.

Doherty PC, Zinkernagel RM. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256:50–52; doi:10.1038/256050A0.

Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. 2014. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 371:2477–2487; doi:10.1056/NEJMOA1409405.

Girelli D, Busti F. 2019. Anemia and adverse outcomes in the elderly: a detrimental inflammatory loop? Haematologica 104:417–419; doi:10.3324/HAEMATOL.2018.208066.
Gu M, Kovilakam SC, Dunn W, Marando L, Barcena C, Mohorianu I, et al. 2023. S121: MULTIPARAMETER PREDICTION OF MYELOID NEOPLASIA RISK. Hemasphere 7:e53186b7; doi:10.1097/01.HS9.0000967396.53186.B7.
Herbert R, Moline J, Skloot G, Metzger K, Baron S, Luft B, et al. 2006. The World Trade Center Disaster and the Health of Workers: Five-Year Assessment of a Unique Medical Screening Program. Environ Health Perspect 114:1853; doi:10.1289/EHP.9592.
Heuser M, Thol F, Ganser A. 2016. M E D I C I N E Clonal Hematopoiesis of Indeterminate Potential A Risk Factor for Hematologic Neoplasms. Dtsch Arztebl Int 113:317–22; doi:10.3238/arztebl.2016.0317.
Jaiswal S, Ebert BL. 2019. Clonal hematopoiesis in human aging and disease. Science 366; doi:10.1126/SCIENCE.AAN4673.
Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V., Mar BG, et al. 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. New England Journal of Medicine 371:2488–2498; doi:10.1056/NEJMOA1408617/SUPPL_FILE/NEJMOA1408617_DISCLOSURES.PDF.
Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. 2017. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. N Engl J Med 377:111–121; doi:10.1056/NEJMOA1701719.
Jasra S, Giricz O, Zeig-Owens R, Pradhan K, Goldfarb DG, Barreto-Galvez A, et al. 2022. High burden of clonal hematopoiesis in first responders exposed to the World Trade Center disaster. Nat Med 28:468–471; doi:10.1038/S41591-022-01708-3.
Kamphuis P, Van Bergen MGJM, Van Zeventer IA, De Graaf AO, Dinmohamed AG, Salzbrunn JB, et al. 2023. Abnormal Platelet Counts and Clonal Hematopoiesis in the General Population. Hemasphere 7:E821; doi:10.1097/HS9.0000000000000821.
Kanagal-Shamanna R, Beck DB, Calvo KR. 2024. Clonal Hematopoiesis, Inflammation, and Hematologic Malignancy. Annu Rev Pathol 19:479–506; doi:10.1146/ANNUREV-PATHMECHDIS-051222-122724.
Kar SP, Quiros PM, Gu M, Jiang T, Mitchell J, Langdon R, et al. 2022. Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. Nat Genet 54:1155–1166; doi:10.1038/S41588-022-01121-Z.
Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. 2017. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. Hum Mutat 38:788–797; doi:10.1002/HUMU.23230.
Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF, et al. 2022. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia 36:1703–1719; doi:10.1038/S41375-022-01613-1.
Komic L, Kumric M, Urlic H, Rizikalo A, Grahovac M, Kelam J, et al. 2023. Obesity and Clonal Hematopoiesis of Indeterminate Potential: Allies in Cardiovascular Diseases and Malignancies. Life (Basel) 13; doi:10.3390/LIFE13061365.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997; doi:10.48550/arxiv.1303.3997.
Li J, Hall CB, Yung J, Kehm RD, Zeig-Owens R, Singh A, et al. 2023. A 15-year follow-up study of mortality in a pooled cohort of World Trade Center rescue and recovery workers. Environ Res 219; doi:10.1016/J.ENVRES.2022.115116.
Li J, Yung J, Qiao B, Takemoto E, Goldfarb DG, Zeig-Owens R, et al. 2022. Cancer Incidence in World Trade Center Rescue and Recovery Workers: 14 Years of Follow-Up. J Natl Cancer Inst 114:210–219; doi:10.1093/JNCI/DJAB165.

Lioy PJ, Georgopoulos P. 2006. The anatomy of the exposures that occurred around the World Trade Center site: 9/11 and beyond. Ann N Y Acad Sci 1076:54–79; doi:10.1196/ANNALS.1371.002.
Loh PR, Genovese G, McCarroll SA. 2020. Monogenic and polygenic inheritance become instruments for clonal selection. Nature 584:136–141; doi:10.1038/S41586-020-2430-6.
Lovell MA, Xie C, Markesbery WR. 2000. Decreased base excision repair and increased helicase activity in Alzheimer's disease brain. Brain Res 855:116–123; doi:10.1016/S0006-8993(99)02335-5.
Ma MCJ, Tadros S, Bouska A, Heavican T, Yang H, Deng Q, et al. 2022. Subtype-specific and co-occurring genetic alterations in B-cell non-Hodgkin lymphoma. Haematologica 107:690–701; doi:10.3324/HAEMATOL.2020.274258.
Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. 2010. Robust relationship inference in genome-wide association studies. Bioinformatics 26: 2867–2873.
Miller CA, Walker JR, Jensen TL, Hooper WF, Fulton RS, Painter JS, et al. 2022. Failure to Detect Mutations in U2AF1 due to Changes in the GRCh38 Reference Sequence. J Mol Diagn 24:219–223; doi:10.1016/J.JMOLDX.2021.10.013.
Morozova O, Clouston SAP, Valentine J, Newman A, Carr M, Luft BJ. 2021. COVID-19 cumulative incidence, asymptomatic infections, and fatality in Long Island, NY, January-August 2020: A cohort of World Trade Center responders. PLoS One 16; doi:10.1371/JOURNAL.PONE.0254713.
Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 53:695–699; doi:10.1111/J.1532-5415.2005.53221.X.
Nathan D, Selvan ME, Guisado D, Collatuzzo G, Boffetta P, Cohen L, et al. 2023. MDS-472 Clonal Hematopoiesis in Inflammatory Bowel Diseases. Clin Lymphoma Myeloma Leuk 23:S368; doi:10.1016/s2152-2650(23)01192-8.
Niroula A, Sekar A, Murakami MA, Trinder M, Agrawal M, Wong WJ, et al. 2021. Distinction of lymphoid and myeloid clonal hematopoiesis. Nat Med 27:1921–1927; doi:10.1038/S41591-021-01521-4.
Ogobuiro I, Baca Y, Ribeiro JR, Walker P, Wilson GC, Gulhati P, et al. 2023. Multiomic Characterization Reveals a Distinct Molecular Landscape in Young-Onset Pancreatic Cancer. JCO Precis Oncol 7; doi:10.1200/PO.23.00152.
Pagliuca S, Gurnari C, Rubio MT, Visconte V, Lenz TL. 2022. Individual HLA heterogeneity and its implications for cellular immune evasion in cancer and beyond. Front Immunol 13; doi:10.3389/FIMMU.2022.944872.
Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. 2016. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med 374:2209–2221; doi:10.1056/NEJMOA1516192.
Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. 2013. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122:3616–3627; doi:10.1182/BLOOD-2013-08-518886.
Penn DJ, Damjanovich K, Potts WK. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc Natl Acad Sci U S A 99:11260–11264; doi:10.1073/PNAS.162006499.
Planelles D, Nagore E, Moret A, Botella-Estrada R, Vila E, Guillén C, et al. 2006. HLA class II polymorphisms in Spanish melanoma patients: homozygosity for HLA-DQA1 locus can be a potential melanoma risk factor. Br J Dermatol 154:261–266; doi:10.1111/J.1365-2133.2005.06896.X.
Rescue & Recovery Workers - 9/11 Health. Available: https://www.nyc.gov/site/911health/enrollees/rescue-recovery-workers.page.

Rogers AH, Zvolensky MJ, Vujanovic AA, Ruggero CJ, Oltmanns J, Waszczuk MA, et al. 2022. Anxiety sensitivity and Pain Experience: a prospective investigation among World Trade Center Responders. J Behav Med 45:947–953; doi:10.1007/S10865-022-00336-Z.

Shapiro MZ, Wallenstein SR, Dasaro CR, Lucchini RG, Sacks HS, Teitelbaum SL, et al. 2019. Cancer in General Responders Participating in World Trade Center Health Programs, 2003-2013. JNCI Cancer Spectr 4; doi:10.1093/JNCICS/PKZ090.

Singh M, Jackson KJL, Wang JJ, Schofield P, Field MA, Koppstein D, et al. 2020. Lymphoma Driver Mutations in the Pathogenic Evolution of an Iconic Human Autoantibody. Cell 180:878-894.e19; doi:10.1016/J.CELL.2020.01.029.

Solan S, Wallenstein S, Shapiro M, Teitelbaum SL, Stevenson L, Kochman A, et al. 2013. Cancer incidence in world trade center rescue and recovery workers, 2001-2008. Environ Health Perspect 121:699–704; doi:10.1289/EHP.1205894.

Sperling AS, Gibson CJ, Ebert BL. 2017. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. Nat Rev Cancer 17:5–19; doi:10.1038/NRC.2016.112.

Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. 2015. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood 126:9–16; doi:10.1182/BLOOD-2015-03-631747.

Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. 2018. Functional genomic landscape of acute myeloid leukaemia. Nature 562:526–531; doi:10.1038/S41586-018-0623-Z.

Uchiumi F, Seki M, Furuichi Y. 2015. Helicases and human diseases. Front Genet 6; doi:10.3389/FGENE.2015.00039.

Vlasschaert C, Mack T, Heimlich JB, Niroula A, Uddin MM, Weinstock J, et al. 2023. A practical approach to curate clonal hematopoiesis of indeterminate potential in human genetic data sets. Blood 141; doi:10.1182/BLOOD.2022018825.

von Beck K, von Beck T, Ferrell PB, Bick AG, Kishtagari A. 2023. Lymphoid clonal hematopoiesis: implications for malignancy, immunity, and treatment. Blood Cancer J 13; doi:10.1038/S41408-022-00773-8.

Wang T, Pine AR, Kotini AG, Yuan H, Zamparo L, Starczynowski DT, et al. 2021. Sequential CRISPR gene editing in human iPSCs charts the clonal evolution of myeloid leukemia and identifies early disease targets. Cell Stem Cell 28:1074-1089.e7; doi:10.1016/J.STEM.2021.01.011.

Weeks LD, Niroula A, Neuberg D, Wong W, Lindsley RC, Luskin MR, et al. 2023. Prediction of risk for myeloid malignancy in clonal hematopoiesis. NEJM evidence 2; doi:10.1056/EVIDOA2200310.

Weigert O, Kopp N, Lane AA, Yoda A, Dahlberg SE, Neuberg D, et al. 2012. Molecular ontogeny of donor-derived follicular lymphomas occurring after hematopoietic cell transplantation. Cancer Discov 2:47–55; doi:10.1158/2159-8290.CD-11-0208.

Wilkins KC, Lang AJ, Norman SB. 2011. Synthesis of the psychometric properties of the PTSD checklist (PCL) military, civilian, and specific versions. Depress Anxiety 28:596–606; doi:10.1002/DA.20837.

Wisnivesky JP, Teitelbaum SL, Todd AC, Boffetta P, Crane M, Crowley L, et al. 2011. Persistence of multiple illnesses in World Trade Center rescue and recovery workers: a cohort study. Lancet 378:888–897; doi:10.1016/S0140-6736(11)61180-X.

Zekavat SM, Lin SH, Bick AG, Liu A, Paruchuri K, Wang C, et al. 2021. Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. Nat Med 27:1012–1024; doi:10.1038/S41591-021-01371-0.

Tables.

Table 1A. Characteristics of the World Trade Center Health Program (WTCHP) General Responders Cohort (GRC) members selected for WES in this study (the WTC cohort), which, after sample QC that removed duplicates, totaled 345 individuals.

Phenotype	Category	# Samples (%)
Sex (M/F =12.3)	Male (M)	319 (92.5%)
	Female (F)	26 (7.5%)
Smoking status	Current smoker	15 (4.3%)
	Previous smoker	154 (44.6%)
	Never smoker	176 (51.0%)
Race	White/ Caucasian	314 (91.0%)
	Black/ African-American	12 (3.5%)
	Other race	13 (3.8%)
	Unknown	6 (1.7%)
Age (yrs) Median= 59 SD= 5.90	≤ 55	27 (7.8%)
	56 - 60	171 (49.6%)
	61 - 65	84 (24.3%)
	66 - 70	32 (9.3%)
	> 70	31 (9.0%)
Derived exposure level	Very high	15 (4.3%)
	High	55 (15.9%)
	Intermediate	210 (60.9%)
	Low	50 (14.5%)
	Very low	6 (1.7%)
	Missing/ No record	9 (2.6%)

Table 1B. Characteristics of the 293 unexposed controls from the MSCCR cohort.

Phenotype	Category	# Samples (%)
Sex (M/F =1.05)	Male (M)	150 (51.2%)
	Female (F)	143 (48.8%)
Smoking status	Smoker	93 (31.7%)
	Never smoker	200 (68.3%)
Race	European	151 (51.5%)
	African-American	65 (22.2%)
	AMI	68 (23.2%)
	East Asian	9 (3.1%)
Age (yrs) Median= 54 SD= 11.20	≤ 55	160 (54.6%)
	56 - 60	67 (22.9%)
	61 - 65	40 (13.7%)
	> 65	26 (8.9%)

Table 2. Prevalence of M-CHIP and L-CHIP mutations

Age (Years)	# Samples	# Samples with M-CHIP mutation	# Samples with L-CHIP mutation
≤ 55	27	2 (7.4%)	6 (22.2%)
56-60	171	17 (9.9%)	33 (19.3%)
61-65	84	16 (19.0%)	23 (27.4%)
66-70	32	6 (18.8%)	7 (21.9%)
> 70	31	15 (48.4%)	5 (16.1%)

Figures.

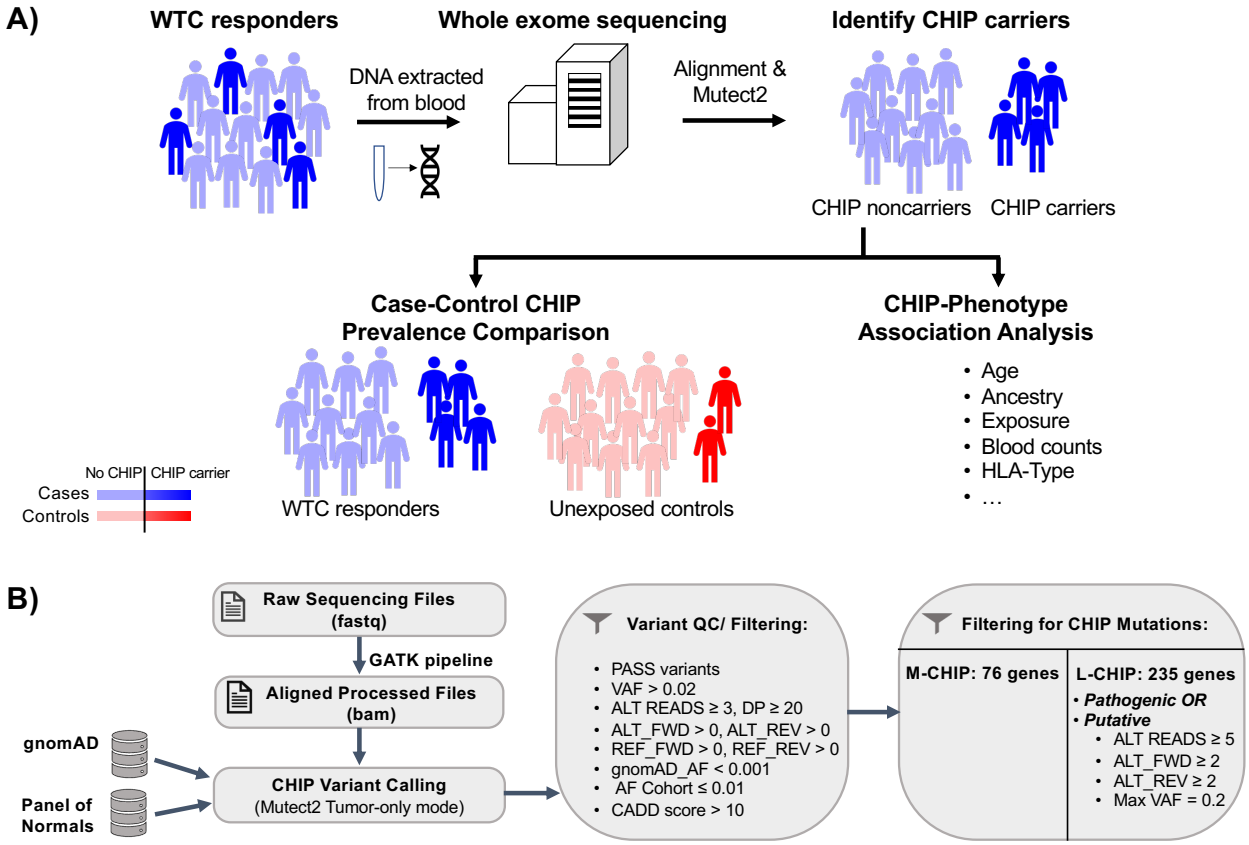

Figure 1. Study design and pipeline. A) Study outline - To understand L/M-CHIP prevalence in WTC debris exposed first responders, we collected blood from participants, extracted DNA, and performed WES on 350 samples. Next, we aligned the exomes to the human genome and called somatic variants using Mutect2 to identify L/-M-CHIP carriers, the specific CHIP mutations and VAF. Additionally, we compared the prevalence of L/-M-CHIP in WTC debris exposed first responders to healthy controls from the New York City area. Finally, we associated CHIP prevalence with respect to age, ancestry, exposure, cognitive/metal data, HLA zygosity, blood counts and other clinical variables. **B)** WES data analysis pipeline to identify M/L-CHIP mutations.

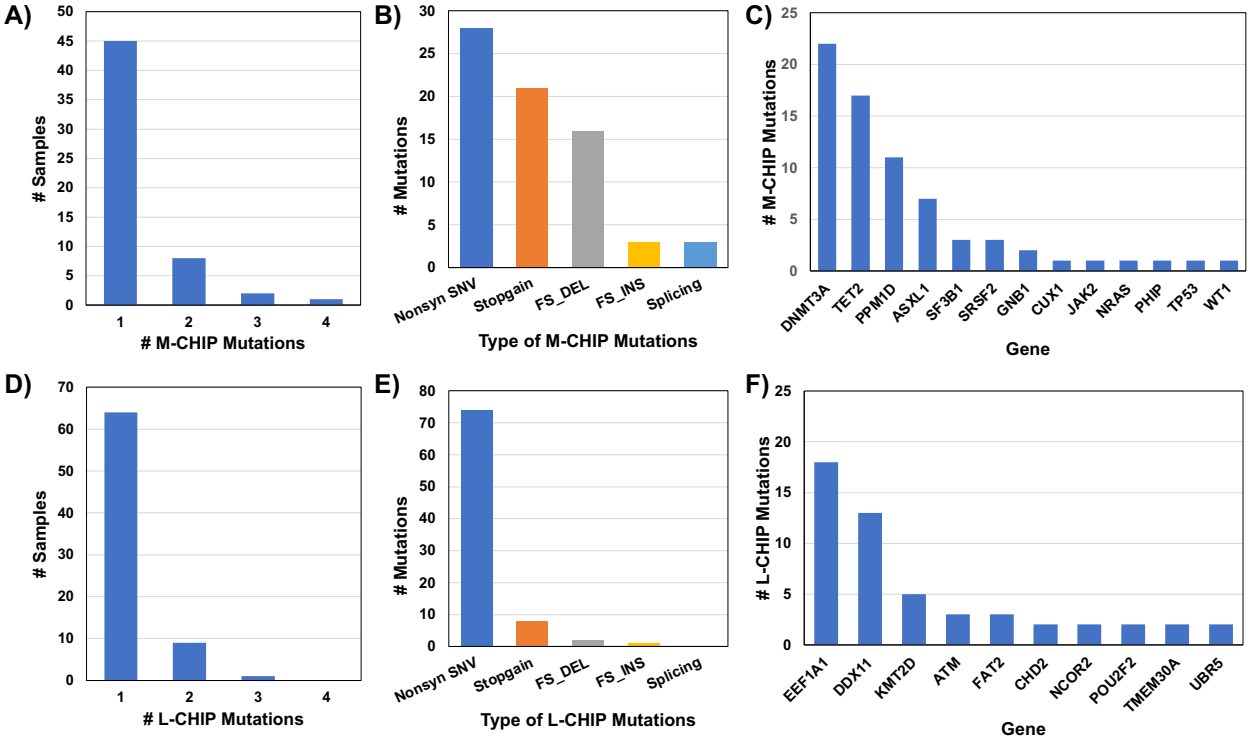

Figure 2. Characteristics of M-CHIP and L-CHIP mutations in 345 WTC-debris-exposed first-responders. A) Number of participants with 1,2, 3 and 4 M-CHIP mutations, **B)** Number of different types of M-CHIP mutations, **C)** Number of M-CHIP mutations observed in each gene, **D)** Number of participants with 1,2, 3 and 4 L-CHIP mutations, **E)** Number of different types of M-CHIP mutations, and **F)** Number of L-CHIP mutations observed in top genes (#Mutations > 2).

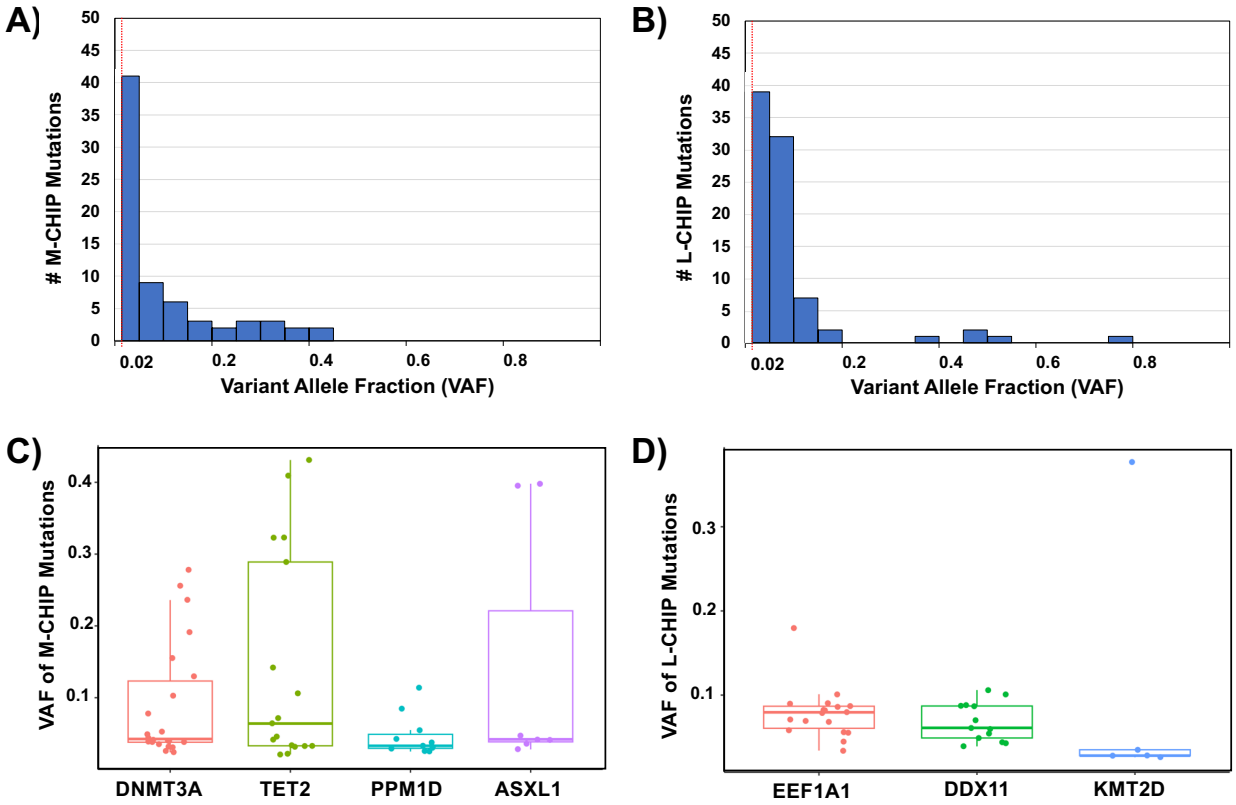

Figure 3. Variant allele fraction (VAF) of M-CHIP and L-CHIP mutations. A) Histogram of the VAF of the M-CHIP mutations, **B)** Histogram of the VAF of the M-CHIP mutations, **C)** Variant allele fraction (VAF) distribution of mutations in the top M-CHIP genes **D)** VAF distribution of mutations in the top L-CHIP genes. Based on VAF, *TET2* has the largest M-CHIP clones, while *EEF1A1* and *DDX11* have the largest L-CHIP clones.

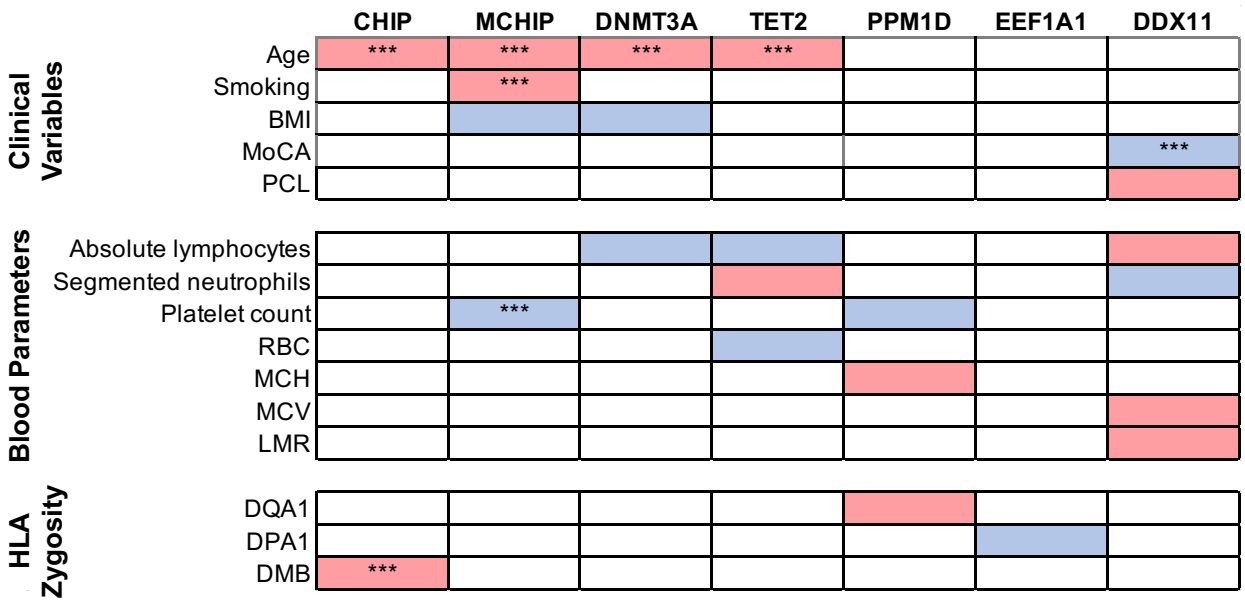

Figure 4: CHIP-Phenotype associations. Heatmaps showing significant associations between overall CHIP, M-CHIP, M-CHIP driven by *DNMT3A*, *TET2*, *PPM1D* mutations and L-CHIP driven by *EEF1A1*, *DDX11* mutations with clinical variables, blood parameters and HLA zygosity. Blue and red correspond to significant ($p \leq 0.05$) negative (OR < 1) and positive (OR > 1) correlation. White represents absence of significant correlation. Asterisk represents significant associations ($p \leq 0.05$) in the multivariate regression analysis.

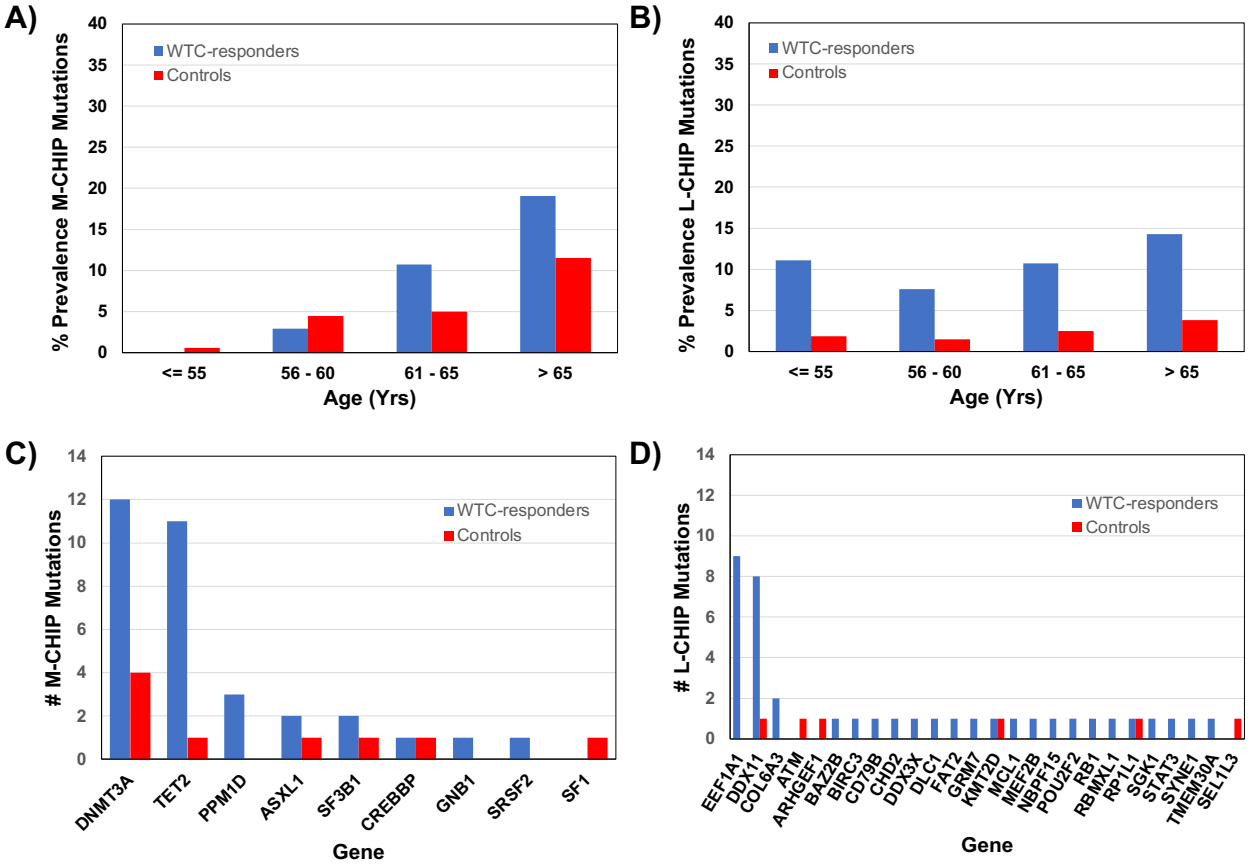

Figure 5. Prevalence and characteristics of M-CHIP and L-CHIP mutations in 345 WTC-debris-exposed individuals (blue) and 293 unexposed controls (red). A) % Prevalence of M-CHIP mutations in the cohorts as a function of age **B)** % Prevalence of L-CHIP mutations in the cohorts as a function of age **C)** Genes identified with M-CHIP mutations **D)** Genes identified with L-CHIP mutations

Supplementary Tables

Supplementary Table S1. List of myeloid associated (M-CHIP) mutations and lymphoid associated (L-CHIP) mutations considered in this study.

Supplementary Table S2. List of M/L-CHIP mutations identified among the 345 WTC samples.

Supplementary Table S3. CHIP-Phenotype associations.

Supplementary Table S4. Prevalence of M-CHIP and L-CHIP mutations in downsampled 345 WTC-debris-exposed first responders and 293 unexposed controls.

Age (Years)	# Samples		M-CHIP mutation			L-CHIP mutation		
	WTC	Control	# WTC	# Control	*p*-value	# WTC	# Control	*p*-value
≤ 55	27	160	0 (0.0%)	1 (0.6%)	1	3 (11.1%)	3 (1.9%)	0.04
56-60	171	67	5 (2.9%)	3 (4.5%)	0.69	13 (7.6%)	1 (1.5%)	0.12
61-65	84	40	9 (10.7%)	2 (5.0%)	0.50	9 (10.7%)	1 (2.5%)	0.17
> 65	63	26	12 (19.0%)	3 (11.5%)	0.54	9 (14.3%)	1 (3.8%)	0.27