

Clonal hematopoiesis of indeterminate potential and the risk of pulmonary embolism: an observational study



Qianwei Liu,^{a,b,*} Karin Ekström Smedby,^c Huiwen Xue,^{a,b} Tove Wästerlid,^c Jiong Li,^{d,e} Fang Fang,^f and Xinyuan Liu^{g,**}

^aDepartment of Hematology, Nanfang Hospital, Southern Medical University, Guangzhou, China

^bClinical Medical Research Center of Hematology Diseases of Guangdong Province, Guangzhou, China

^cDivision of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, and Department of Hematology, Karolinska University Hospital, Stockholm, Sweden

^dDepartment of Clinical Medicine-Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark

^eDepartment of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China

^fInstitute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

^gCenter for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, School of Life Sciences, Fudan University, Guangzhou, China



Summary

Background Pulmonary embolism causes a substantial burden of morbidity and mortality. Although there are several well-established risk factors for pulmonary embolism, a substantial proportion of cases cannot be attributed to provoked or known risk factors. Accumulating evidence has suggested an association of clonal hematopoiesis of indeterminate potential (CHIP) with the risk of arterial thromboembolism. However, the association between CHIP and the risk of pulmonary embolism remains unknown.

Methods We performed a community-based cohort study (between 2006 and 2022) including 464,417 individuals with available whole exome sequencing (WES) data in the UK biobank (UKB) to examine the association between CHIP and pulmonary embolism. CHIP was ascertained by analyzing WES data. We used Cox regression models to calculate hazard ratios (HRs) with 95% confidence intervals (CIs) to assess the association between CHIP and pulmonary embolism. In addition, we performed analyses for several types of CHIP mutations, including *DNMT3A*, *TET2*, *ASXL1*, *PPM1D*, *SRSF2*, and *JAK2*.

Findings The study included 14,418 individuals with CHIP and 449,999 individuals without CHIP. The median age at cohort entry was 58 and 63 years among individuals without and with CHIP, respectively. We observed an increased risk of pulmonary embolism (HR 1.17, 95% CI, 1.05–1.31) among individuals with CHIP. The increased risk was mainly noted for CHIP with *TET2* (HR 1.42, 95% CI 1.16–1.74) or *JAK2* (HR 4.17, 95% CI 2.09–8.35) mutation, but not for *DNMT3A* mutation (HR 1.01, 95% CI 0.86–1.19), *ASXL1* mutation (HR 1.15, 95% CI 0.83–1.60), *PPM1D* mutation (HR 1.22, 95% CI 0.66–2.27), or *SRSF2* mutation (HR 0.62, 95% CI 0.20–1.93).

Interpretation Our results highlight the association of pulmonary embolism in individuals with CHIP, especially the *TET2*-mutant or *JAK2*-mutant CHIP. If further studies will identify a causal relationship between clonal hematopoiesis and pulmonary embolism, prioritizing early screening for pulmonary embolism in individuals with CHIP could be significantly beneficial.

Funding Initial Founding for High Level Talented Scholars in Nanfang Hospital, Southern Medical University (No. 2023G001) and the Outstanding Youths Development Scheme of Nanfang Hospital, Southern Medical University (Grant No. 2023J009).

Copyright © 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Clonal hematopoiesis; Pulmonary embolism; Epidemiology; Cohort study

eClinicalMedicine
2024;74: 102753

Published Online 26 July
2024

<https://doi.org/10.1016/j.eclinm.2024.102753>

*Corresponding author. Department of Hematology, Nanfang Hospital, Southern Medical University, Guangzhou, China.

**Corresponding author. Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China.

E-mail addresses: qianweiliu@smu.edu.cn (Q. Liu), liuxinyuancsu@outlook.com (X. Liu).

Research in context

Evidence before this study

We searched PubMed for relevant research up to Jan 10, 2024 using the search terms “clonal hematopoiesis of indeterminate potential” and “clonal hematopoiesis” in combination with “pulmonary embolism”, “embolism”, “thrombosis”, “thromboembolism”, and “thrombotic”. We found two original studies on a potential link between CHIP and pulmonary embolism. One study suggested an association of CHIP with risk of pulmonary embolism based on a small sample and retrospective design, while another study reported no association between CHIP and general venous thromboembolism among patients with cancer.

Added value of this study

We performed a community-based cohort study with a large sample size, providing novel evidence of the association between CHIP and pulmonary embolism. We also revealed that an association between CHIP and pulmonary embolism was predominantly noted for *TET2*-mutant CHIP and *JAK2*-mutant CHIP.

Implications of all the available evidence

Further studies are necessary to confirm a causal relationship between clonal hematopoiesis and pulmonary embolism and suggest early screening for pulmonary embolism among individuals with, especially *TET2*-mutant CHIP and *JAK2*-mutant CHIP.

Introduction

Pulmonary embolism is the most severe manifestation of venous thromboembolism and ranks among the leading causes of morbidity and mortality worldwide.^{1–3} The occurrence of pulmonary embolism involves embolus from venous thrombus traveling to and obstructing the arterial tree of the lungs, constituting a lethal risk particularly when not recognized.⁴ Although several risk factors (e.g., cancer, major trauma and immobility) have been well-documented for pulmonary embolism, a substantial proportion of cases cannot be attributed to provoked or known risk factors,³ limiting therefore the effectiveness of early prevention and screening in vulnerable populations.

Clonal hematopoiesis of indeterminate potential (CHIP) is an aging-related condition characterized by expansion of hematopoietic cell clones that possess specific somatic mutations, and is recognized as a precursor state of hematological malignancy.⁵ CHIP has been suggested to be associated with an increased risk of diverse somatic disorders, including cardiovascular disease, stroke, osteoporosis, gout, chronic obstructive pulmonary, chronic liver disease and solid tumors.^{6–12} In contrast to the accumulating evidence of a link between CHIP and thromboembolism in artery,⁶ research on pulmonary embolism, a type of venous thromboembolism, in relation to CHIP is limited. One previous study investigated the association between CHIP and general venous thromboembolism (including pulmonary embolism and deep vein thrombosis) among cancer patients, leaving the general population understudied.¹³ Another study demonstrated a plausible link between CHIP and pulmonary embolism, but the obvious methodological weaknesses of the study (i.e., retrospective design, very small sample size, and lack adjustment for important confounders) preclude robust conclusions.¹⁴ Therefore, we aimed to determine the relationship between CHIP and risk of pulmonary embolism in a cohort study of the UK Biobank (UKB).

Methods

Study cohort

This study leveraged data from the UKB, a community-based cohort, including approximately 500,000 UK residents aged 37–73 years between 2006 and 2010.¹⁵ Recruitment of the UKB occurred at 22 centers distributed across the United Kingdom, gathering a comprehensive dataset encompassing a range of variables, including socioeconomic status, lifestyle factors, clinical diagnoses, and genetic sequencing.¹⁵ Initially, we included 469,522 participants with existing data on whole exome sequencing (WES) and time of blood sampling. We first excluded participants with pre-existing hematological malignancy ($N = 3296$) (the tenth revision of the International Classification of Diseases (ICD-10) codes for hematological malignancy were provided in [Supplementary eTable S1](#)) before the date of blood collection for WES. Then, we additionally excluded individuals with pulmonary embolism ($N = 1809$) before the date of blood sampling, resulting in a total of 464,417 individuals for the ultimate analysis ([Fig. 1](#)). Participants were followed from blood sampling until diagnosis of pulmonary embolism, death, or December 31, 2022, whichever occurred first.

Ascertainment of CHIP

The original quality functionally equivalent (OQFE) protocol was used to generate CRAM files and the GRCh38 reference genome were used as the basis for germline variant calling.¹⁶ The identification of CHIP was based on 74 driver genes known to be associated with myeloid malignancies.¹⁷ Mutect2 was deployed in ‘tumor-only’ mode to detect potential CHIP somatic variants for the exonic regions of 73 out of the 74 genes, with the exception of *U2AF1*.^{16,17} For *U2AF1*, we used Rust-HTSLIB binary (available at https://github.com/weinstockj/pileup_region) to specifically identify variants, which could mitigate potential biases in the detection of genetic variants arising from segmental

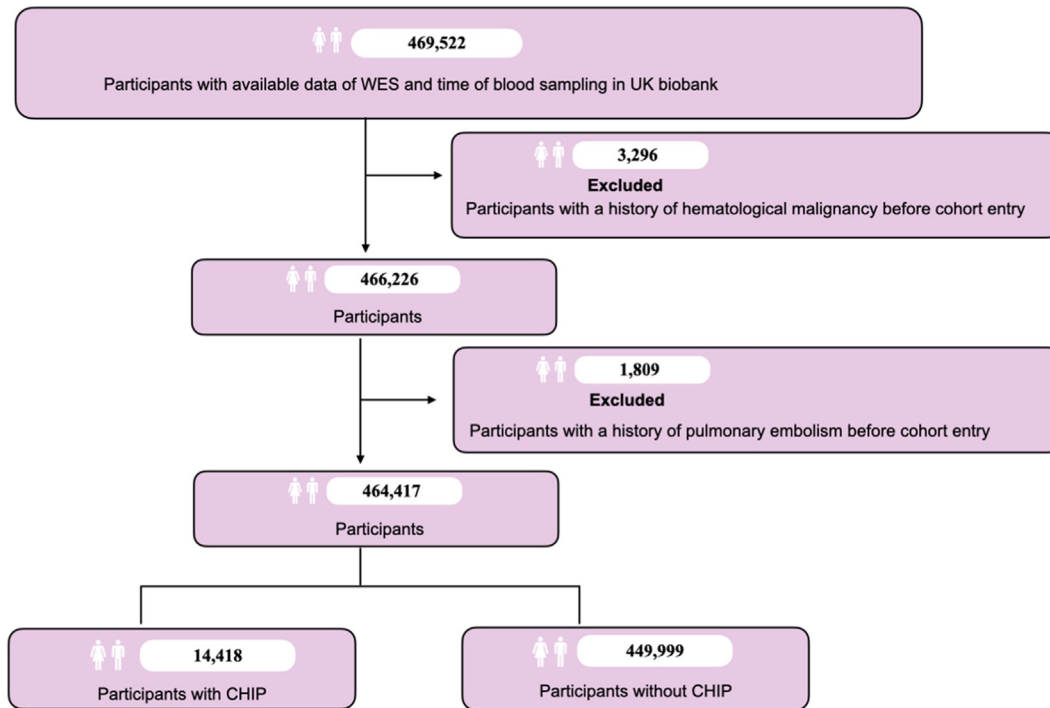


Fig. 1: Flowchart of the cohort and design in the study. WES, whole exome sequencing; CHIP, clonal hematopoiesis of indeterminate potential.

duplication errors in the GRCh38 reference genome.¹⁸ Gene annotation was conducted using ANNOVAR.¹⁹ The Genome Aggregation Database (gnomAD)²⁰ was used as a reference to preliminarily distinguish potential germline variants from putative somatic mutations, under the threshold of a minor allele frequency (MAF) < 0.01.²¹ To filter out recurrent artifacts, we established a panel of normals (PON) through randomly selecting 200 of the youngest individuals (i.e., at the age of 40 or below) in the UKB who were free of hematological malignancy at blood sampling for WES. We filtered variants identified by Mutect2 by FilterMutectCalls for the probabilities generated by LearnReadOrientationModel.^{16,22} Only putative variants marked as 'PASS' by FilterMutectCalls were considered. Furthermore, variants labeled as 'germline' or 'weak_evidence' were eligible for inclusion under these conditions: they appeared more than five times with a 'PASS' flag, and they passed a one-sided exact binomial test with a variant allele frequency (VAF) of 0.5 ($P < 0.001$). The exclusion criteria were: (1) variants with a total read depth of less than 20; (2) a minimum read depth for the alternate allele less than 5; (3) a VAF below 2%; (4) a lack of support of both forward and reverse sequencing reads. Besides, variants did not associated with age or TERT promoter (rs7705526) were also excluded.¹⁷

Ascertainment of pulmonary embolism

Pulmonary embolism was identified through inpatient hospital records, using the ICD-10 codes (Supplementary eTable S1).²³ The inpatient hospital records data incorporate admission and diagnosis information obtained from the Hospital Episode Statistics for England, Scottish Morbidity Record data for Scotland, and the Patient Episode Database for Wales. In addition, we included supplementary cases using death register data provided by the National Health Service Digital for England and Wales and the Information Services Division for Scotland, as well as the read coded from primary care data.¹⁵

Covariates

Information on sex, birth date, ethnicity, education level, Townsend deprivation index, and body mass index (BMI) determined at the time of recruitment to the UKB were included as covariates. Age was calculated as date of blood sampling minus birth date. Townsend deprivation index incorporates information on employment, car availability, and housing (none and overcrowding) to reflect socioeconomic status. BMI, calculated as weight in kilograms divided by the square of height in meters (kg/m^2), was derived from measurements obtained from the baseline assessment at the assessment centers. Pre-existing cardiovascular disease (CVD) was defined

as any diagnosis of CVD before cohort entry, identified through inpatient hospital records as well as primary healthcare data using ICD-10 codes (Supplementary eTable S1). Cancer was identified by ICD-10 codes (Supplementary eTable S1) from the Cancer Register, as a time-varying covariate during follow-up. Information of white blood cell levels, hemoglobin concentration, and platelet counts in serum were acquired through biochemical measures in UKB.

Statistics

First, we profiled the demographic and characteristics of individuals with and without CHIP, including sex, age, ethnicity, educational level, Townsend deprivation index, BMI (<18.5 , $18.5 \leq$ and <25 , $25 \leq$ and <30 , or ≥ 30), smoking and alcohol drinking. Chi-square test was used to test difference in the demographics and characteristics of individuals with or without CHIP. Second, we computed unadjusted incidence rates by dividing the number of pulmonary embolism cases by the accumulated person-years at risk among individuals with and without CHIP. Third, we used Cox regression model to calculate the hazard ratios (HRs) with 95% confidence intervals (CIs) for risk of pulmonary embolism in relation to CHIP. All analyses were conducted using time since follow-up initiation as the underlying time scale. In Model 1, we adjusted for sex, age at blood sampling (as natural cubic spline), ethnicity (white, others or unknown), educational level (college/university degree, other degree, or unknown), and the Townsend deprivation index (Q1, Q2, Q3, or Q4). In Model 2, we further adjusted for lifestyle factors, including BMI (as natural cubic spline), smoking (never, previous, current, or unknown) and alcohol drinking (never, previous, current, or unknown). Fourth, we explored potential effect modifiers of the investigated associations through subgroup analyses by sex, age at blood sampling (<60 or ≥ 60), ethnicity, Townsend deprivation index, educational level, BMI (<18.5 , $18.5 \leq$ and <25 , $25 \leq$ and <30 , or ≥ 30), smoking, and alcohol drinking. Fifth, we performed additional analyses by type of CHIP mutations, including the most common CHIP mutations as *DNMT3A*, *TET2*, *ASXL1*, *PPM1D*, and *SRSF2*.

Supplementary and sensitivity analyses

(1) To reduce potential concerns regarding reverse causality, we performed a sensitivity analysis using different lag time (1-year, 3-year, and 5-year) before initiating follow-up after blood collection. (2) As previous studies have suggested a role of *JAK2* in the pathogenesis of embolism and thrombosis,^{24,25} we further conducted stratified analysis by *JAK2* mutation. (3) We also performed an additional analysis by the VAF of CHIP mutation ($VAF < 0.1$ or $VAF \geq 0.1$) to determine whether the size of CHIP clone could influence the investigated association. (4) As some types of CVD have been suggested

to be linked with both CHIP and pulmonary embolism,^{6,26} we conducted a sensitivity analysis stratified by pre-existing CVD. (5) As cancer status was associated with both CHIP and pulmonary embolism,^{3,11} to explore the effect of cancer on the studied association, we performed an additional analysis stratified by cancer (as time-varying variable). (6) To explore the impact of counts of blood cells on the observed association, we performed additional analyses by further adjusting for blood counts, including count of white blood cell, concentration of hemoglobin, or count of platelet, as natural cubic spline. (7) To investigate the robustness of time since cohort entry as the analytical time scale, we conducted a sensitivity analysis using age as the underlying time scale. (8) To account for the competing risk of death, we performed a sensitivity analysis using Fine-Gray model to calculate the sub-distribution hazard ratio (sHR).²⁷

The analyses were executed using SAS software (Version 9.4, SAS institute Inc). Wald test was used to test differences between the HRs of different groups. The significance level was determined using two-sided tests, with results considered statistically significant at a threshold of $P < 0.05$.

Ethics

All participants in the UKB signed informed consent before information collection and ethical approval has been obtained from the NHS National Research Ethics Service (REC reference: 21/NW/0157). This research has been approved by the Ethical Review Board in Nanfang Hospital, Southern Medical University in China (reference number: NFEC-2023-559).

Role of funding source

This study was partially supported by the Initial Founding for High Level Talented Scholars in Nanfang Hospital, Southern Medical University (No. 2023G001) and the Outstanding Youths Development Scheme of Nanfang Hospital, Southern Medical University (Grant No. 2023J009). The funders for the present study had no role in study design, data collection, data analyses, interpretation, or writing of report.

Results

Primary analyses

The study included 464,417 individuals, among whom 14,418 (3.1%) had CHIP. The median age at cohort entry was 58 and 63 years among individuals without and with CHIP, respectively (Table 1). Compared to individuals without CHIP, individuals with CHIP were more likely to be male, of White ethnicity, and overweight, and have a lower educational level but a higher prevalence of previous or current smoking.

Individuals with CHIP were followed for a median duration of 13.6 years, whereas individuals without CHIP had a median follow-up of 13.8 years. Among

individuals with CHIP, there were 339 incident cases of pulmonary embolism (incidence rate [IR], 180.0 per 100,000 person-years), while 7544 cases were identified among individuals without CHIP (IR, 124.6 per 100,000 person-years) (Table 2). Compared with individuals without CHIP, individuals with CHIP had an increased risk of pulmonary embolism in both model 1 (HR 1.18, 95% CI, 1.06–1.32) and model 2 (HR 1.17, 95% CI, 1.05–1.31) (Table 2). Given the similar estimates between model 1 and model 2, all results presented below are from model 2. The overall association did not demonstrate substantial variation across demographic, socioeconomic, or lifestyle factors (Table 2).

The analyses stratified by mutation type of CHIP showed an increased risk of pulmonary embolism in relation to *TET2* mutation (HR 1.42, 95% CI 1.16–1.74), but not to *DNMT3A* mutation (HR 1.01, 95% CI 0.86–1.19), *ASXL1* mutation (HR 1.15, 95% CI 0.83–1.60), *PPM1D* mutation (HR 1.22, 95% CI 0.66–2.27), or *SRSF2* mutation (HR 0.62, 95% CI 0.20–1.93) (Fig. 2).

Supplementary and sensitivity analysis

(1) Largely comparable results were found when using a 1-year (HR 1.17, 95% CI 1.05–1.31), 3-year (HR 1.19, 95% CI 1.06–1.34), or 5-year (HR 1.16, 95% CI 1.03–1.32) lag time in the analyses. (2) We observed increased risks among individuals with (HR 4.17, 95% CI 2.09–8.35) or without *JAK2* mutation (HR 1.15, 95% CI 1.03–1.29), although the risk increase was much stronger for the former. (3) The observed association was more noticeable for individuals with a high level of VAF (≥ 0.1) of any CHIP mutation (HR 1.24, 95% CI 1.09–1.40), compared with those with VAF < 0.1 (HR 1.03, 95% CI 0.83–1.26), though the difference of HRs between the two groups was not statistically significant (P for difference = 0.14) (Supplementary eTable S2). (4) The association did not vary greatly by the history of pre-existing CVD (HR 1.13, 95% CI 0.99–1.30 for presence of pre-existing CVD vs. HR 1.24, 95% CI 1.04–1.48 for absence of pre-existing CVD, P for difference = 0.41). (5) We observed comparable results among individuals with (HR 1.17, 95% CI 1.01–1.35) or without (HR 1.09, 95% CI 0.93–1.28) cancer (P for difference = 0.52) during follow-up. (6) The overall association did not vary after adjusting for count of white blood cell, concentration of hemoglobin, or count of platelet (Supplementary eTable S3). (7) Using age as underlying time-scale yielded very similar results (HR 1.16, 95% CI, 1.04–1.29). (8) Largely similar risk estimates (sHR 1.15, 95% CI, 1.04–1.29) were observed when accounting for the competing risk of death.

Discussion

Leveraging a large community-based cohort study from UKB, we found that individuals with CHIP, especially individuals with *TET2* or *JAK2* mutation, had an

Characteristics	Individuals without CHIP; No. (%)	Individuals with CHIP; No. (%)	P value
Number of individuals	449,999	14,418	
Sex			<0.001
Female	244,645 (54.4%)	7538 (52.3%)	
Male	205,354 (45.6%)	6880 (47.7%)	
Median age at cohort entry (years)	58	63	
Age at cohort entry (years)			<0.001
<60	249,710 (55.5%)	4609 (32.0%)	
≥ 60	200,289 (44.5%)	9809 (68.0%)	
Ethnicity			<0.001
White	423,538 (94.1%)	13,690 (95.0%)	
Other	24,841 (5.5%)	675 (4.7%)	
Unknown	1620 (0.4%)	53 (0.4%)	
Education level			<0.001
College/University degree	146,205 (32.5%)	4177 (29.0%)	
Other degree	222,718 (49.5%)	6755 (46.9%)	
Unknown	81,076 (18.0%)	3486 (24.2%)	
Townsend deprivation index at recruitment			0.59
Q1	113,766 (25.3%)	3601 (25.0%)	
Q2	112,654 (25.0%)	3631 (25.2%)	
Q3	112,487 (25.0%)	3568 (24.7%)	
Q4	111,092 (24.7%)	3618 (25.1%)	
BMI			<0.001
BMI < 18.5	3994 (0.9%)	143 (1.0%)	
18.5 \leq BMI < 25	146,456 (32.5%)	4398 (30.5%)	
25 \leq BMI < 30	190,652 (42.4%)	6303 (43.7%)	
BMI ≥ 30	108,897 (24.2%)	3574 (24.8%)	
Smoking			<0.001
Never	246,244 (54.7%)	6907 (47.9%)	
Previous	154,425 (34.3%)	5672 (39.3%)	
Current	47,076 (10.5%)	1733 (12.0%)	
Unknown	2254 (0.5%)	106 (0.7%)	
Alcohol drinking			0.17
Never	19,732 (4.4%)	634 (4.4%)	
Previous	15,929 (3.5%)	561 (3.9%)	
Current	413,208 (91.8%)	13,188 (91.5%)	
Unknown	1130 (0.3%)	35 (0.2%)	

CHIP, clonal hematopoiesis of indeterminate potential; BMI, body mass index.

Table 1: Characteristics of the cohort participants.

increased risk of pulmonary embolism. The association did not vary greatly across demographic, socioeconomic, or lifestyle factors, pre-existing CVD, or cancer, and was independent of other biomarkers such as count of blood cell, concentration of hemoglobin, or count of platelet.

Consistent with our research, a prior study also observed a positive association between CHIP and pulmonary embolism.¹⁴ However, this study was limited by a retrospective design, small sample size (N = 61), and little control for confounding factors.¹⁴ In the present study, we applied a prospective design with a large sample size and comprehensive adjustment for potential confounders, therefore providing robust

Characteristics	Individuals without CHIP		Individuals with CHIP		Hazard ratio (95% confidence interval)	
	No. of cases	IR	No. of cases	IR	Model 1 ^a	Model 2 ^b
Overall	7544	124.6	339	180.0	1.18 (1.06-1.32)	1.17 (1.05-1.31)
Sex						
Female	3630	109.3	171	169.9	1.25 (1.07-1.46)	1.26 (1.08-1.47)
Male	3914	143.0	168	191.5	1.12 (0.96-1.30)	1.10 (0.94-1.29)
Age at cohort entry (years)						
<60	2761	80.8	62	99.5	1.10 (0.86-1.42)	1.12 (0.87-1.43)
≥60	4783	181.3	277	219.7	1.20 (1.06-1.35)	1.19 (1.05-1.34)
Ethnicity						
White	7218	126.5	326	182.3	1.18 (1.06-1.32)	1.17 (1.05-1.31)
Other	301	91.4	12	134.4	1.17 (0.66-2.09)	1.10 (0.60-2.01)
Unknown	25	117.7	<5	153.2	1.19 (0.16-8.87)	1.63 (0.21-12.76)
Education level						
College/University degree	1706	110.3	67	140.0	1.05 (0.82-1.34)	1.04 (0.82-1.33)
Other degree	1837	120.6	83	173.4	1.17 (0.93-1.45)	1.17 (0.94-1.45)
Unknown	1829	120.9	84	181.4	1.23 (0.99-1.54)	1.21 (0.97-1.51)
Townsend deprivation index at recruitment						
Q1	2172	147.3	105	226.5	1.25 (1.03-1.52)	1.25 (1.02-1.52)
Q2	1870	94.3	73	131.8	1.11 (0.88-1.41)	1.10 (0.87-1.39)
Q3	3631	120.9	147	165.8	1.13 (0.96-1.33)	1.11 (0.94-1.32)
Q4	2043	190.9	119	268.2	1.30 (1.08-1.56)	1.30 (1.08-1.57)
BMI						
BMI < 18.5	74	145.6	<5	115.2	0.68 (0.17-2.77)	-
18.5 ≤ BMI < 25	1548	78.1	69	119.0	1.19 (0.93-1.52)	1.16 (0.91-1.48)
25 ≤ BMI < 30	3157	122.8	145	175.4	1.18 (1.00-1.39)	1.16 (0.98-1.37)
BMI ≥ 30	2765	190.2	123	267.4	1.20 (1.00-1.43)	1.20 (1.00-1.44)
Smoking						
Never	3605	107.7	139	150.7	1.14 (0.96-1.35)	1.13 (0.95-1.34)
Previous	2939	142.4	146	199.6	1.17 (0.99-1.38)	1.18 (1.00-1.39)
Current	946	153.2	52	239.8	1.27 (0.96-1.68)	1.29 (0.98-1.71)
Unknown	54	184.1	<5	152.6	0.73 (0.18-3.04)	0.74 (0.18-3.06)
Alcohol drinking						
Never	379	144.0	12	144.1	0.86 (0.49-1.53)	0.86 (0.49-1.54)
Previous	377	181.6	20	285.2	1.36 (0.87-2.14)	1.39 (0.89-2.19)
Current	6759	121.3	306	177.3	1.19 (1.06-1.33)	1.18 (1.05-1.32)
Unknown	29	197.6	<5	229.3	1.02 (0.13-7.69)	1.13 (0.15-8.75)

CHIP, clonal hematopoiesis of indeterminate potential; BMI, body mass index; IR, incidence rate. ^aAdjusted for sex (female or male), age at cohort entry (as natural cubic spline), ethnicity (white, other, or unknown), educational level (college or university degree, other degree, or unknown), Townsend deprivation index (Q1, Q2, Q3, or Q4). ^bAdjusted for sex (female or male), age at cohort entry (as natural cubic spline), ethnicity (white, other, or unknown), educational level (college or university degree, other degree, or unknown), Townsend deprivation index (Q1, Q2, Q3, or Q4), BMI (as natural cubic spline), smoking (never, previous, current, or unknown) and alcohol drinking (never, previous, current, or unknown).

Table 2: Risk of pulmonary embolism among individuals with CHIP by demographic factors, socioeconomic status and lifestyle factors, compared to the reference group.

evidence for the studied association. Additionally, we observed an association between CHIP and pulmonary embolism among individuals with cancer, in contrast with a previous study that found no association between CHIP and general venous thromboembolism (including pulmonary embolism and deep vein thrombosis) among cancer patients.¹³ This discrepancy might be due to different study design and analytical approaches. For instance, we excluded the first year of follow-up in the

present study to mitigate concern of reverse causality, whereas the previous study focused solely on outcome events within the first year after accession.

The mechanisms underlying the observed association might be multifactorial. CHIP has been suggested to be associated with risk of cancer, which might further cause pulmonary embolism. However, our findings of comparable results among individuals with or without cancer suggested altered functions of somatic CHIP

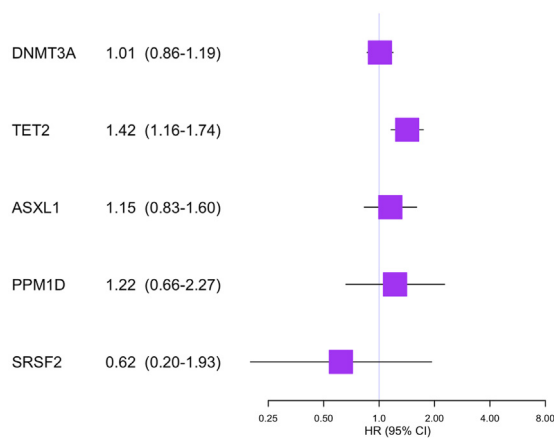


Fig. 2: Hazard ratios of pulmonary embolism among individuals of CHIP by different gene mutations. Adjusted for sex (female or male), age at cohort entry (as natural cubic spline), ethnicity (white, other, or unknown), educational level (college or university degree, other degree, or unknown), Townsend deprivation index (Q1, Q2, Q3, or Q4), BMI (as natural cubic spline), smoking (never, previous, current, or unknown) and alcohol drinking (never, previous, current, or unknown). CHIP, clonal hematopoiesis of indeterminate potential; BMI, body mass index; HR, hazard ratio; CI, confidence interval.

mutations, beyond oncogenesis, might also contributed. Corroborating with the well-documented mechanisms between *JAK2* mutation and thrombosis,^{24,25} we observed a strong association between *JAK2*-CHIP and pulmonary embolism. Potential mechanisms for such a link might include biological changes of platelet caused by *JAK2* mutation, such as hypersensitivity of thrombopoietin (TPO)/MPL (TPO receptor) pathway, increased megakaryocyte activity, and enhanced platelet functions.²⁵ An alternative explanation for the observed association between *JAK2*-CHIP and increased risk of pulmonary embolism might be the alterations of endothelial cells.²⁸ In current study, we also noted an increased risk of pulmonary embolism among individuals without *JAK2* mutation, suggesting that other mutations beyond *JAK2* might also play a role in the development of pulmonary embolism. We observed a pronounced association of *TET2* with risk of pulmonary embolism, which might be mechanistically attributed to the upregulation of proinflammatory cytokines, such as interleukin-8 and interleukin-1 β ,^{6,29} subsequently contributing to the pathogenesis of pulmonary embolism.^{30,31}

Our study exhibits a number of strengths. One major advantage is the application of prospective design with a large sample size. In addition, independent collection of WES data and diagnosis of pulmonary embolism reduce potential information bias. Moreover, the availability of detailed information on demographic, socioeconomic, and lifestyle factors provided us with an

ability to meticulously control for potential confounders. Nevertheless, this study also has some limitations. First, our study lack extensive repeated WES data to monitor time-varying exposures. Consequently, individuals who were initially without CHIP at cohort recruitment but acquired CHIP during the follow-up period remained unidentified, which might have led to a lower estimation of the association we noted. Second, we cannot fully exclude reverse causality, although largely similar results were observed in the analyses with a 1-, 3-, or 5-year lag time. Third, the sample size on different CHIP mutations is limited, leading to inadequate statistical power in this analysis. Future studies with larger sample sizes are needed to validate these findings. Fourth, this study is limited by using a single cohort without validation from external cohort. Particularly, participants of the present study were predominantly white British and were not representative of the whole UK population.³² Future studies, ideally incorporating multiple cohorts from diverse backgrounds, are essential to corroborate our initial findings and examine the generalizability of these findings in broader contexts. Fifth, given that individuals with CHIP might have a higher mortality than individuals without CHIP, the observed association might have been affected by the competing risk of death from other causes. Sixth, the present study may be subject to residual confounding. Although we made efforts to control for the effect of some potential confounders, through multivariable adjustment or stratified analysis, the observational nature of this study could not exclude residual confounding due to unmeasured or unknown confounders. Seventh, despite the considerable sample size of the present study, the statistical power remained insufficient to detect modest risk differences or to perform analyses in small subgroups (e.g., certain CHIP mutations with a limited number of cases).

In conclusion, we observed a higher risk of pulmonary embolism among individuals with CHIP, especially those with *TET2* or *JAK2* mutations. These findings suggest potential shared mechanisms between CHIP and pulmonary embolism, and the value of early screening for pulmonary embolism among individuals with CHIP should be evaluated. Future studies are warranted to verify these findings.

Contributors

QL and XL conceived the idea for the study. QL, KS, FF, and XL designed the study. QL and HX performed the analysis. XL prepared the manuscript draft. QL and HX have verified the underlying data. QL, KS, TW, HX, JL, FF, and XL interpreted data, and revised the manuscript for submission. All authors read and approved the final version of the manuscript.

Data sharing statement

This study utilized the UK Biobank Resource, under Application Number 106912. Researchers can access data from the UK Biobank by submitting an application at www.ukbiobank.ac.uk.

Declaration of interests

The authors declare no conflict of interest.

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number [106912]. This study was partially supported by the Initial Founding for High Level Talented Scholars in Nanfang Hospital, Southern Medical University (No. 2023G001) and the Outstanding Youths Development Scheme of Nanfang Hospital, Southern Medical University (Grant No. 2023J009).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2024.102753>.

References

- Goldhaber SZ, Bounameaux H. Pulmonary embolism and deep vein thrombosis. *Lancet*. 2012;379(9828):1835–1846.
- Di Nisio M, van Es N, Büller HR. Deep vein thrombosis and pulmonary embolism. *Lancet*. 2016;388(10063):3060–3073.
- Khan F, Tritschler T, Kahn SR, Rodger MA. Venous thromboembolism. *Lancet*. 2021;398(10294):64–77.
- Pulmonary embolism. *Nat Rev Dis Prim*. 2018;4:18031.
- Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9–16.
- Jaiswal S, Natarajan P, Silver AJ, et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N Engl J Med*. 2017;377(2):111–121.
- Kim PG, Niroula A, Shkolnik V, et al. Dnm3a-mutated clonal hematopoiesis promotes osteoporosis. *J Exp Med*. 2021;218(12):e20211872.
- Agrawal M, Niroula A, Cunin P, et al. TET2-mutant clonal hematopoiesis and risk of gout. *Blood*. 2022;140(10):1094–1103.
- Bhattacharya R, Zekavat SM, Haessler J, et al. Clonal hematopoiesis is associated with higher risk of stroke. *Stroke*. 2022;53(3):788–797.
- Miller PG, Qiao D, Rojas-Quintero J, et al. Association of clonal hematopoiesis with chronic obstructive pulmonary disease. *Blood*. 2022;139(3):357–368.
- Reed SC, Croessmann S, Park BH. CHIP happens: clonal hematopoiesis of indeterminate potential and its relationship to solid tumors. *Clin Cancer Res*. 2023;29(8):1403–1411.
- Wong WJ, Ermdin C, Bick AG, et al. Clonal haematopoiesis and risk of chronic liver disease. *Nature*. 2023;616(7958):747–754.
- Dunbar A, Bolton KL, Devlin SM, et al. Genomic profiling identifies somatic mutations predicting thromboembolic risk in patients with solid tumors. *Blood*. 2021;137(15):2103–2113.
- Soudet S, Jedraszak G, Evrard O, Marolleau JP, Garçon L, Pietri MAS. Is hematopoietic clonality of indetermined potential a risk factor for pulmonary embolism? *TH Open*. 2021;5(3):e338–e342.
- Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–209.
- David B, Takuto S, Kristian C, Gad G, Chip S, Lee L. Calling somatic SNVs and indels with Mutect2. *bioRxiv*. 2019. <https://doi.org/10.1101/861054>.
- Vlasschaert C, Mack T, Heimlich JB, et al. A practical approach to curate clonal hematopoiesis of indeterminate potential in human genetic data sets. *Blood*. 2023;141(18):2214–2223.
- Miller CA, Walker JR, Jensen TL, et al. Failure to detect mutations in U2AF1 due to changes in the GRCh38 reference sequence. *J Mol Diagn*. 2022;24(3):219–223.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
- Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
- Dawoud AAZ, Tapper WJ, Cross NCP. Age-related loss of chromosome Y is associated with levels of sex hormone binding globulin and clonal hematopoiesis defined by TET2, TP53, and CBL mutations. *Sci Adv*. 2023;9(16):eade9746.
- Gu M, Kovilakam SC, Dunn WG, et al. Multiparameter prediction of myeloid neoplasia risk. *Nat Genet*. 2023;55(9):1523–1530.
- Casez P, Labarère J, Sevestre MA, et al. ICD-10 hospital discharge diagnosis codes were sensitive for identifying pulmonary embolism but not deep vein thrombosis. *J Clin Epidemiol*. 2010;63(7):790–797.
- Arellano-Rodrigo E, Alvarez-Larrán A, Reverter JC, Villamor N, Colomer D, Cervantes F. Increased platelet and leukocyte activation as contributing mechanisms for thrombosis in essential thrombocythemia and correlation with the JAK2 mutational status. *Hematologica*. 2006;91(2):169–175.
- Hobbs CM, Manning H, Bennett C, et al. JAK2V617F leads to intrinsic changes in platelet formation and reactivity in a knock-in mouse model of essential thrombocythemia. *Blood*. 2013;122(23):3787–3797.
- Becattini C, Agnelli G, Prandoni P, et al. A prospective study on cardiovascular events after acute pulmonary embolism. *Eur Heart J*. 2005;26(1):77–83.
- Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391–4400.
- Sozer S, Fiel MI, Schiano T, Xu M, Mascarenhas J, Hoffman R. The presence of JAK2V617F mutation in the liver endothelial cells of patients with Budd-Chiari syndrome. *Blood*. 2009;113(21):5246–5249.
- Fuster JJ, MacLauchlan S, Zuriaga MA, et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science*. 2017;355(6327):842–847.
- Christiansen SC, Naess IA, Cannegieter SC, Hammerstrøm J, Rosendaal FR, Reitsma PH. Inflammatory cytokines as risk factors for a first venous thrombosis: a prospective population-based study. *PLoS Med*. 2006;3(8):e334.
- Bontekoe E, Brailovsky Y, Hoppensteadt D, et al. Upregulation of inflammatory cytokines in pulmonary embolism using biochip-array profiling. *Clin Appl Thromb Hemost*. 2021;27:10760296211013107.
- Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026–1034.