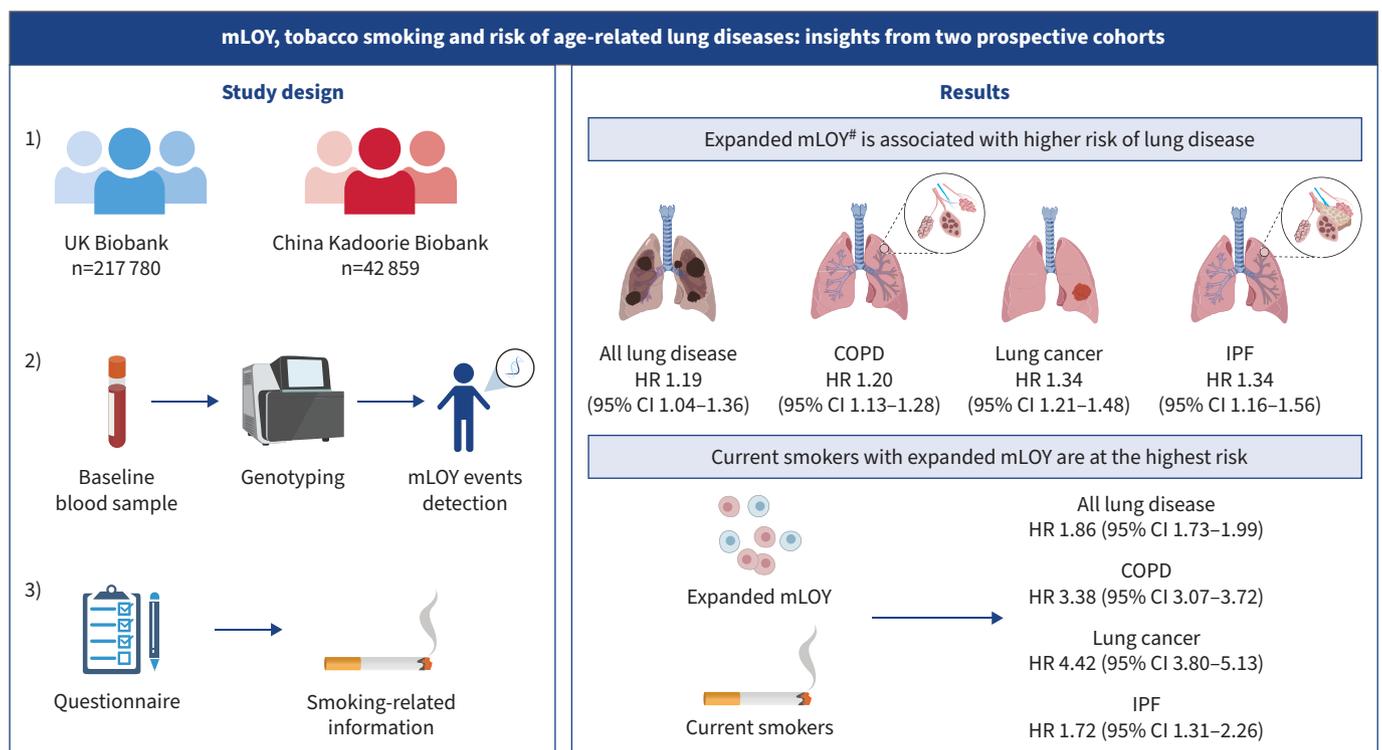




# Mosaic loss of chromosome Y, tobacco smoking and risk of age-related lung diseases: insights from two prospective cohorts

Chenghao Weng, Yuxuan Zhao, Mingyu Song, Zilun Shao, Yuanjie Pang , Canqing Yu , Pei Pei, Ling Yang, Iona Y. Millwood, Robin G. Walters, Yiping Chen, Huaidong Du, Junshi Chen, Zhengming Chen, Giulio Genovese , Chikashi Terao, Jun Lv, Liming Li and Dianjianyi Sun on behalf of the China Kadoorie Biobank Collaborative Group



**GRAPHICAL ABSTRACT** Overview of the study design and findings. mLOY: mosaic loss of chromosome Y; IPF: idiopathic pulmonary fibrosis. <sup>#</sup>: expanded mLOY refers to mLOY carriers with affected cell fraction  $\geq 10\%$ .



# Mosaic loss of chromosome Y, tobacco smoking and risk of age-related lung diseases: insights from two prospective cohorts

Chenghao Weng<sup>1,13</sup>, Yuxuan Zhao<sup>2,13</sup>, Mingyu Song<sup>2</sup>, Zilun Shao<sup>2</sup>, Yuanjie Pang<sup>2</sup>, Canqing Yu<sup>2,3,4</sup>, Pei Pei<sup>3</sup>, Ling Yang<sup>5</sup>, Iona Y. Millwood<sup>5</sup>, Robin G. Walters<sup>5</sup>, Yiping Chen<sup>5</sup>, Huaidong Du<sup>5</sup>, Junshi Chen<sup>6</sup>, Zhengming Chen<sup>5</sup>, Giulio Genovese<sup>7,8,9</sup>, Chikashi Terao<sup>10,11,12</sup>, Jun Lv<sup>2,3,4</sup>, Liming Li<sup>2,3,4</sup> and Dianjianyi Sun<sup>2,3,4</sup> on behalf of the China Kadoorie Biobank Collaborative Group

<sup>1</sup>School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China. <sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China. <sup>3</sup>Peking University Center for Public Health and Epidemic Preparedness and Response, Beijing, China. <sup>4</sup>Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China. <sup>5</sup>Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK. <sup>6</sup>China National Center for Food Safety Risk Assessment, Beijing, China. <sup>7</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>8</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>9</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA. <sup>10</sup>Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan. <sup>11</sup>Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan. <sup>12</sup>Department of Applied Genetics, School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan. <sup>13</sup>These authors contributed equally to this work.

Corresponding author: Dianjianyi Sun (dsun1@bjmu.edu.cn)



Shareable abstract (@ERSpublications)

Evidence from European and East Asian populations suggests that mosaic loss of chromosome Y events are associated with an elevated risk of age-related lung diseases, and that current smokers may be at greatest risk <https://bit.ly/4gx0cIT>

Cite this article as: Weng C, Zhao Y, Song M, *et al.* Mosaic loss of chromosome Y, tobacco smoking and risk of age-related lung diseases: insights from two prospective cohorts. *Eur Respir J* 2024; 64: 2400968 [DOI: 10.1183/13993003.00968-2024].

Copyright ©The authors 2024.

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact [permissions@ersnet.org](mailto:permissions@ersnet.org)

Received: 20 May 2024  
Accepted: 11 Sept 2024

## Abstract

**Background** Little is known about the underlying relationship between mosaic loss of chromosome Y (mLOY), the most common chromosomal alterations in older men, and the risk of age-related lung diseases.

**Methods** We included 217 780 participants from the UK Biobank (UKB) and 42 859 participants from the China Kadoorie Biobank. The mLOY events were detected using the Mosaic Chromosomal Alterations (MoChA) pipeline. Outcomes included all lung diseases, COPD, lung cancer and idiopathic pulmonary fibrosis (IPF). Cox proportional hazard models were fitted to estimate the hazard ratios and 95% confidence intervals of mLOY with lung diseases in both cohorts. The combined hazard ratios were derived from meta-analysis.

**Results** Results from the two cohorts showed that expanded mLOY was associated with increased risks of all lung diseases (HR 1.19 (95% CI 1.04–1.36)), COPD (HR 1.20 (95% CI 1.13–1.28)), lung cancer (HR 1.34 (95% CI 1.21–1.48)) and IPF (HR 1.34 (95% CI 1.16–1.56)) in the UKB). There was evidence of positive interactions between mLOY and smoking behaviour (relative excess risk due to interaction (97.5% CI) >0). Additionally, we observed that current smokers with expanded mLOY had the highest risk of incident lung diseases in both cohorts.

**Conclusions** mLOY may be a novel predictor for age-related lung diseases. For current smokers carrying mLOY, adopting quitting smoking behaviour may contribute to substantially reducing their risk of incident lung diseases.

## Introduction

Ageing leads to reduced lung function and increased vulnerability to environmental stressors [1], which increase one's susceptibility to age-related lung diseases including COPD, lung cancer and idiopathic pulmonary fibrosis (IPF) [1–3]. These diseases contribute significantly to the burden of diseases worldwide [4, 5], with COPD and lung cancer among the 10 leading causes of disability-adjusted life-years in



populations aged  $\geq 50$  years in 2019. Unfortunately, IPF patients have a median survival period of only 2–4 years post-diagnosis due to a lack of treatment options [6]. Therefore, identifying potential risk factors for age-related lung diseases and implementing effective interventions for high-risk individuals is crucial.

Mosaic chromosomal alterations (mCAs) are structural changes on chromosomes that are considered a hallmark of human ageing [7–9]. Somatic mosaic loss of chromosome Y (mLOY) is the most common type of mCA in older men, affecting  $>40\%$  of men aged  $>70$  years [10, 11]. Although once considered a normal age-related process, in recent years mLOY has been shown to be associated with reduced life expectancy and an increased risk of age-related diseases such as cardiovascular disease, solid tumours and Alzheimer's disease [10–15]. However, the underlying relationship of mLOY with the risk of lung diseases is still unclear. Previous studies examining the correlation between mLOY and lung cancer have yielded conflicting findings [16–18], while the relationship between mLOY and other lung diseases remains unexplored. In addition, it is well known that tobacco smoking increases the risks of both age-related lung diseases and mLOY [19, 20]. Thus, whether there is a synergistic interaction between mLOY and smoking behaviour with lung diseases is largely unknown.

In this study, we used genotyping data from the UK Biobank (UKB) and the China Kadoorie Biobank (CKB) to estimate the prospective associations of mLOY with the risk of lung diseases. We also investigated the joint association of mLOY and tobacco smoking with the risk of lung diseases and examined the potential modifying effect of smoking behaviour.

## Methods

### Study population

The UKB is a large prospective cohort that recruited over 500 000 participants aged 40–69 years from 22 assessment centres across the UK during 2006–2010 [21]. In the CKB study, 512 891 adults aged 30–79 years were recruited in 10 geographically defined regions (five urban and five rural) in China between 2004 and 2008 [22]. Baseline characteristics were collected through questionnaires, physical measurements and blood samples in both studies. Detailed information about the UKB and the CKB has been described previously [21, 22]. The UKB has approvals from the North West Multi-Centre Research Ethics Committee, and the CKB was approved by the Ethical Review Committee of the Chinese Center for Disease Control and Prevention (Beijing, China) and the Oxford Tropical Research Ethics Committee, University of Oxford (Oxford, UK). All participants provided informed consent at baseline. After exclusion, a total of 217 780 UKB participants (supplementary figure S1) and 42 859 CKB participants (supplementary figure S2) were included in the analysis.

### Detection of mLOY events

The mCA call sets for UKB participants were obtained from previous studies [8, 9, 11]. We followed similar procedures to identify mLOY carriers in the CKB population, as described elsewhere [11]. Briefly, we used the Mosaic Chromosomal Alterations (MoChA) pipeline (<https://github.com/freeseek/mocha>) to detect mLOY events among male participants with the log R ratio (LRR) and B-allele frequency (BAF) values, which were calculated according to the allele-specific genotyping intensities in the pseudoautosomal region 1 (PAR1) of the sex chromosomes [11]. The fraction of cells carrying mLOY was determined using the deviation of BAF on the PAR1 region. Based on this, detectable mLOY was further classified as non-expanded (cell fraction  $<10\%$ ) or expanded (cell fraction  $\geq 10\%$ ).

### Ascertainment of outcome

The incident health events of UKB participants were determined by linking to hospital admission records and the national mortality register. In the CKB study, the disease incidence information was ascertained via Chinese national health insurance databases and linkage with local disease and death registries. Incident cases were determined using International Classification of Diseases, 10th Revision codes in both studies, and the outcomes of interest included the incidence of all lung diseases (C33–C34, J09–J29, J40–J98), COPD (J41–J44), lung cancer (C33–C34) and IPF (J84.1). However, IPF was not included as an outcome in the CKB study due to the limited number of IPF cases recorded. The end of follow-up was defined as the onset of certain outcomes, loss to follow-up, death or censoring, whichever occurred first. In the UKB, the end of follow-up was 30 September 2021 for England, 31 July 2021 for Scotland and 28 February 2018 for Wales. The end of follow-up for the CKB was 31 December 2018.

### Tobacco smoking and other covariates

According to self-reported smoking status, participants were categorised into never-smokers, former smokers and current smokers. Smoking intensity was measured by the proportion of the lifespan exposed

to smoking, defined as pack-years divided by the number of years between age at recruitment and the age of 16 years.

The following covariates were included in the UKB analyses: 1) sociodemographic characteristics, including age, age squared, assessment centre (England, Scotland or Wales), ethnic background (White or non-White), socioeconomic status (measured by the Townsend deprivation index), qualification (college or other) and occupation (agricultural and related work, industrial and related work, non-agricultural or industrial work, retirement, or other); 2) lifestyle and health factors, including smoking status, smoking intensity, alcohol consumption, total Metabolic Equivalent Task (MET) minutes per week, healthy diet (defined as daily intake of fruits and vegetables), body mass index (BMI) and prevalent health conditions (diabetes and hypertension); and 3) genotyping batch and top 10 principal components. In the CKB cohort, we adjusted for the region (urban or rural areas) instead of the assessment centre, and socioeconomic status was measured by household income (CNY <2500, 2500–4999, 5000–9999, 10 000–19 999, 20 000–34 999 or  $\geq$ 35 000).

### Statistical analysis

Baseline characteristics were presented as mean with standard deviation or median (interquartile range) for continuous variables and number (percentage) for categorical variables, stratified by the status of mLOY. Both ANOVA and the Kruskal–Wallis H-test were utilised to obtain p-values for continuous variables, while the Chi-squared test was employed for categorical variables.

Multiple imputation by chained equations was employed to generate 10 complete datasets in the UKB, with a maximum of 10 iterations [23]. Cox proportional hazard models were used to estimate the hazard ratios and 95% confidence intervals of mLOY and its subtypes with the risk of lung diseases. Participants without mLOY were set as the reference group. Estimates from each dataset in the UKB were pooled according to Rubin's rules. We adjusted for potential confounders and constructed the following models: 1) Model 1 included age, age squared and assessment centre or region; 2) Model 2 further adjusted for ethnic background (in the UKB only), socioeconomic status, qualification, occupation, alcohol consumption, total MET minutes per week for all activities, healthy diet, BMI, prevalent health conditions (diabetes and hypertension), genotyping batch and top 10 principal components; and 3) Model 3 additionally adjusted for smoking status and smoking intensity. Schoenfeld residuals were used to assess the proportional hazard assumption, and no violation was detected. In addition, results from the UKB and the CKB were meta-analysed using a random effects model.

To investigate the joint association between mLOY and smoking status on the risk of lung diseases, we divided participants into six groups according to their joint categories, with non-smokers without mLOY as the reference group. We performed a similar analysis by further dividing mLOY into non-expanded and expanded mLOY, with smoking status categorised as non-current smokers and current smokers. The Chi-squared test was used to test for multiplicative interaction between mLOY and smoking status in relation to lung diseases by comparing models with and without a cross-product term in the fully adjusted model. To test for additive interaction, the relative excess risk due to interaction (RERI) and the attributable proportion due to interaction (AP) were calculated using the R package "epiR". We provided a 97.5% confidence interval for RERI and AP to avoid false positives due to multiple tests.

To test the robustness of our results, we performed several sensitivity analyses: 1) by excluding incident cases during the first 2 years of follow-up; 2) by adopting a broader definition of IPF as J84.0, J84.1, J84.8 and J84.9 in the UKB cohort [24]; 3) by further adjusting for several additional smoking-related characteristics, including time since quitting, type of tobacco currently smoked and exposure to second-hand smoke; and 4) by excluded those with missing information at baseline to ensure the imputation datasets were unbiased. As an additional analysis, we also investigated the association of mLOY events with the co-occurrence of lung disease. All analyses were performed using R version 4.1.2 (www.r-project.org). The statistical significance level was set at a two-tailed  $p < 0.05$ .

## Results

### Baseline characteristics

In total, 217 780 UKB participants (mean age 56.7 years) were included in the current study, 43 994 of whom (20.2%) were identified as mLOY carriers. Overall, mLOY carriers were more likely to be older, less deprived, White, less educated, less engaged in agriculture and industry, current smokers, exposed to higher levels of tobacco, daily drinkers, have higher levels of physical activity, healthier eating habits, lower BMI, and have a higher proportion of diabetes and hypertension (supplementary table S1).

Similarly, 2458 mLOY carriers (5.7%) were detected among 42 859 male participants (mean age 54.8 years) in the CKB cohort. Compared with non-mLOY carriers, those with mLOY were more likely to be older, living in urban areas, better educated, less engaged in agriculture and industry, current smokers, exposed to higher levels of tobacco, and daily drinkers. Additionally, individuals with mLOY had lower levels of physical activity, lower BMI, and a higher proportion of diabetes and hypertension (supplementary table S2).

### Associations of mLOY with the risk of lung diseases

Before the final analysis of each specific outcome, we excluded individuals with the corresponding disease at baseline. During a median follow-up period of 12.5 years, 27 817 incident lung disease cases, 8734 incident COPD cases, 2662 incident lung cancer cases and 1606 incident IPF cases were documented in the UKB cohort. The CKB study recorded 5848 cases of incident lung diseases, 2358 cases of incident COPD and 1037 cases of incident lung cancer during a median follow-up period of 12.2 years. The distribution of cell fraction of mLOY in both cohorts is presented in supplementary figure S3.

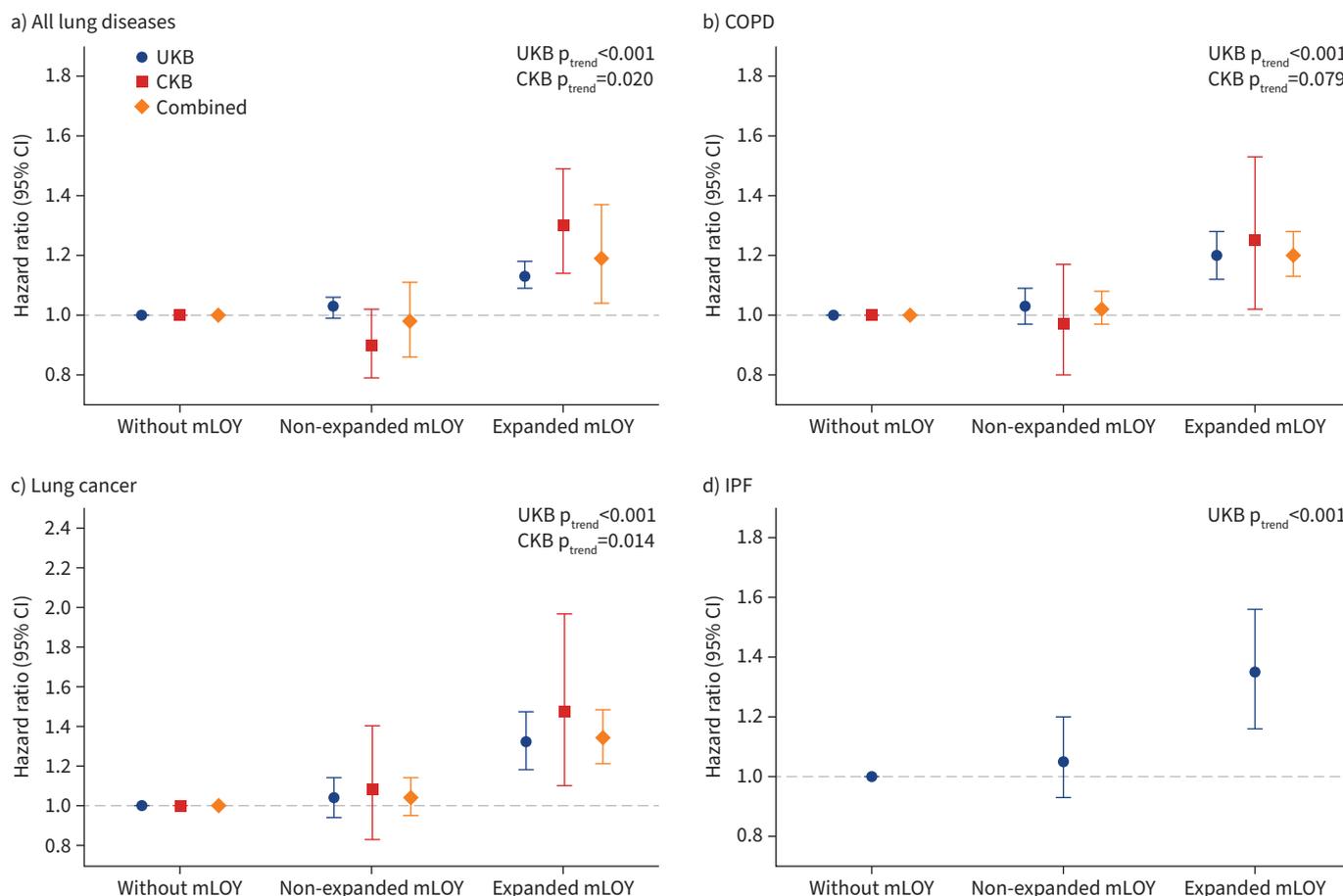
After adjusting for sociodemographic characteristics, lifestyle and health factors (except for smoking status and intensity) in the Cox regression model (Model 2), participants with mLOY had higher risks of all outcomes in the UKB cohort. However, in the CKB cohort, mLOY was only associated with a higher risk of lung cancer (table 1). The associations were greatly attenuated after adjusting for smoking status and intensity (table 1). In the fully adjusted model, mLOY carriers had a 6% higher risk of all lung diseases (HR 1.06 (95% CI 1.03–1.09)), a 9% higher risk of COPD (HR 1.09 (95% CI 1.04–1.14)), a 14% higher risk of lung cancer (HR 1.14 (95% CI 1.05–1.24)) and a 15% higher risk of IPF (HR 1.15 (95% CI 1.04–1.28)) in the UKB, and a 22% higher risk of lung cancer (HR 1.22 (95% CI 1.00–1.50) in the CKB (table 1). The random effects meta-analysis yielded similar results to those in the UKB cohort (table 1).

When classifying mLOY by cell fraction, we found that increased cell fraction was associated with an elevated risk of all outcomes in both cohorts ( $p_{\text{trend}} < 0.05$ ) (figure 1), except for COPD in the CKB study ( $p_{\text{trend}} = 0.079$ ) (figure 1). Only expanded mLOY was associated with an increased risk of all outcomes (figure 1 and supplementary table S3). The combined results showed that expanded mLOY was associated with an increased risk of all lung diseases (HR 1.19 (95% CI 1.04–1.36)), COPD (HR 1.20 (95% CI 1.13–1.28)) and lung cancer (HR 1.34 (95% CI 1.21–1.48)) (figure 1 and supplementary table S3). In the UKB cohort, expanded mLOY was also associated with a 34% increased risk of IPF (HR 1.34 (95% CI 1.16–

TABLE 1 Associations of mosaic loss of chromosome Y (mLOY) with the risk of age-related lung diseases

	mLOY carrier		Hazard ratio (95% CI)		
	Cases/mLOY carriers (n/N)	Incident cases per 1000 person-years	Model 1 <sup>#</sup>	Model 2 <sup>¶</sup>	Model 3 <sup>*</sup>
<b>All lung diseases</b>					
UKB	7938/38 079	18.7	1.14 (1.11–1.17)	1.15 (1.12–1.18)	1.06 (1.03–1.09)
CKB	479/2014	24.5	1.07 (0.97–1.18)	1.08 (0.98–1.19)	1.05 (0.95–1.15)
Combined			1.12 (1.07–1.18)	1.13 (1.08–1.19)	1.06 (1.03–1.09)
$p_{\text{heterogeneity}}$ -value			0.221	0.272	0.852
<b>COPD</b>					
UKB	2963/42 132	6.1	1.39 (1.33–1.46)	1.37 (1.31–1.44)	1.09 (1.04–1.14)
CKB	218/2037	10.6	1.07 (0.93–1.24)	1.13 (0.98–1.30)	1.08 (0.94–1.25)
Combined			1.23 (0.95–1.59)	1.26 (1.04–1.52)	1.09 (1.04–1.14)
$p_{\text{heterogeneity}}$ -value			0.001	0.011	0.904
<b>Lung cancer</b>					
UKB	1017/43 923	2.0	1.55 (1.42–1.68)	1.49 (1.37–1.62)	1.14 (1.05–1.24)
CKB	111/2442	4.5	1.33 (1.09–1.63)	1.31 (1.07–1.60)	1.22 (1.00–1.50)
Combined			1.48 (1.29–1.70)	1.44 (1.31–1.59)	1.15 (1.07–1.24)
$p_{\text{heterogeneity}}$ -value			0.169	0.310	0.497
<b>IPF</b>					
UKB	563/43 940	1.1	1.22 (1.10–1.35)	1.23 (1.11–1.37)	1.15 (1.04–1.28)

UKB: UK Biobank; CKB: China Kadoorie Biobank; IPF: idiopathic pulmonary fibrosis. <sup>#</sup>: Model 1 adjusted for age, age squared and assessment centre or region; <sup>¶</sup>: Model 2 further adjusted for ethnic background (in the UKB only), socioeconomic status, qualification, occupation, alcohol consumption, total Metabolic Equivalent Task minutes per week for all activities, healthy diet, body mass index, prevalent health conditions (diabetes and hypertension), genotyping batch and top 10 principal components; <sup>\*</sup>: Model 3 additionally adjusted for smoking status and smoking intensity.



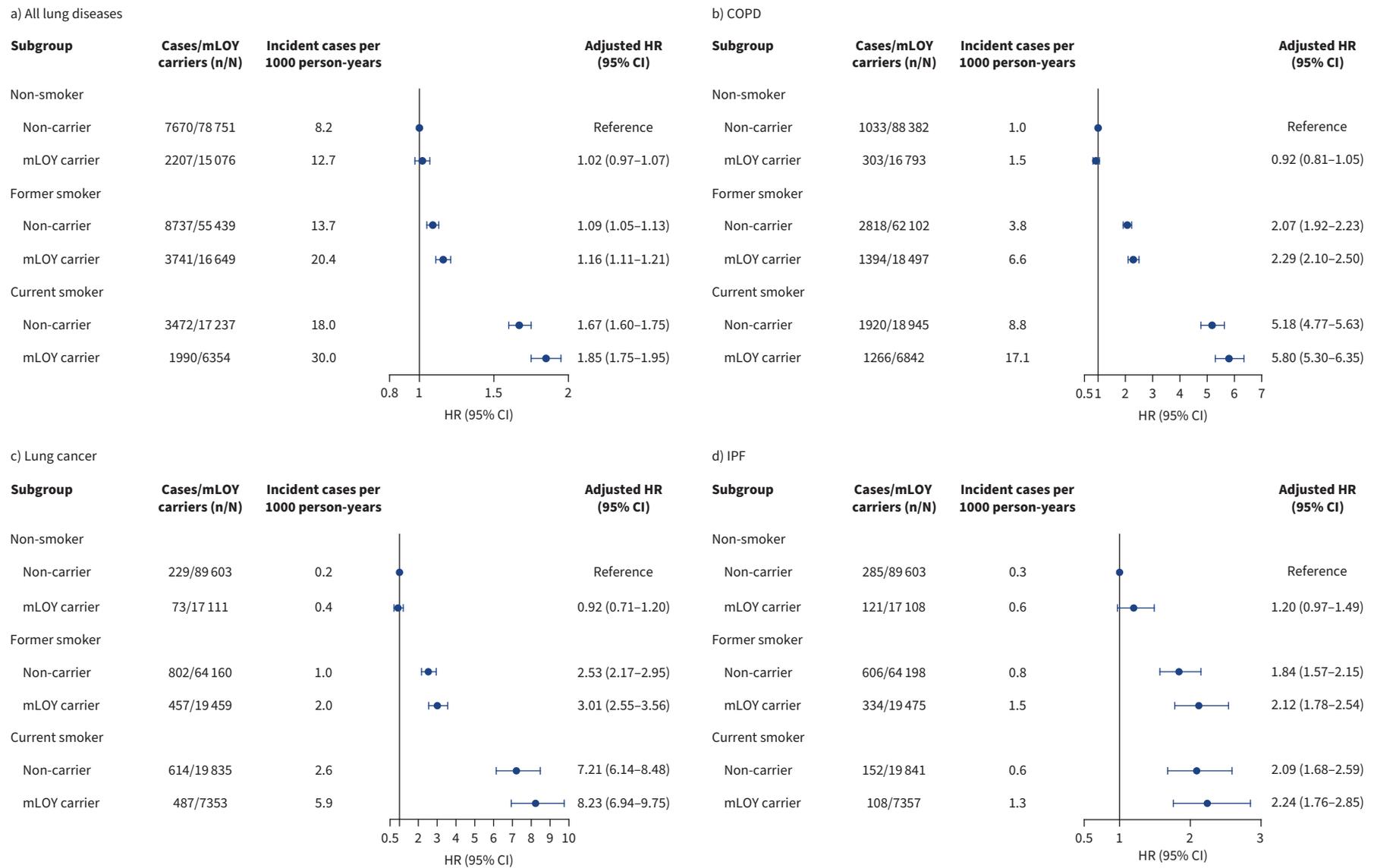
**FIGURE 1** Associations between mosaic loss of chromosome Y (mLOY) subtypes and lung diseases: a) all lung diseases; b) COPD, c) lung cancer and d) idiopathic pulmonary fibrosis (IPF). UKB: UK Biobank; CKB: China Kadoorie Biobank.

1.56)) (figure 1 and supplementary table S3). Similar conclusions were drawn from sensitivity analyses (supplementary tables S4–S10). In addition, expanded mLOY was associated with more lung conditions in both cohorts (supplementary table S11). Considering that asthma and pneumonia are also common among the elderly, we additionally analysed the association between mLOY and these conditions, and found that expanded mLOY is associated with a higher risk of both asthma (HR 1.17 (95% CI 1.04–1.31)) (supplementary table S12) and pneumonia (HR 1.14 (95% CI 1.08–1.20)) (supplementary table S13).

#### Joint association and interaction of mLOY and smoking behaviour with the risk of lung diseases

Furthermore, we investigated the joint effect of mLOY and smoking status on the risk of incident lung diseases. Compared with non-smokers without mLOY, we observed that current smokers carrying mLOY had the highest risk of all outcomes in the UKB cohort: all lung diseases (HR 1.85 (95% CI 1.75–1.95)), COPD (HR 5.80 (95% CI 5.30–6.35)), lung cancer (HR 8.23 (95% CI 6.94–9.75)) and IPF (HR 2.24 (95% CI 1.76–2.85)) (figure 2). Similar associations were observed in the CKB cohort for all lung diseases (HR 1.21 (95% CI 1.06–1.38)), COPD (HR 1.48 (95% CI 1.22–1.80)) and lung cancer (HR 2.13 (95% CI 1.59–2.85)) (supplementary figure S4). Positive additive interactions (RERI (97.5% CI) >0) were found between mLOY and smoking status on the risk of all lung diseases, COPD and lung cancer in the UKB but not in the CKB (table 2). In addition, multiplicative interactions were also found in the UKB cohort (all lung diseases  $p_{interaction}$  < 0.001 and COPD  $p_{interaction}$  = 0.013) (table 2).

We further classified mLOY according to cell fraction and repeated the analysis. Our findings showed that current smokers carrying expanded mLOY had the highest risk of all outcomes in the UKB cohort when compared to non-current smokers without mLOY: all lung diseases (HR 1.86 (95% CI 1.73–1.99)), COPD (HR 3.38 (95% CI 3.07–3.72)), lung cancer (HR 4.42 (95% CI 3.80–5.13)) and IPF (HR 1.72 (95% CI 1.31–2.26)) (figure 3). Similar associations were observed in the CKB cohort: all lung diseases (HR 1.50



**FIGURE 2** Joint association of mosaicism loss of chromosome Y (mLOY) and smoking status with the risk of lung diseases in the UK Biobank: **a)** all lung diseases; **b)** COPD, **c)** lung cancer and **d)** idiopathic pulmonary fibrosis (IPF).

**TABLE 2** Interactions between mosaic loss of chromosome Y and smoking status on the effect of lung diseases in the UK Biobank (UKB) and the China Kadoorie Biobank (CKB)

	Former smoker		Current smoker		P <sub>interaction</sub> -value
	RERI (97.5% CI)	AP (97.5% CI)	RERI (97.5% CI)	AP (97.5% CI)	
<b>All lung diseases</b>					
UKB	0.06 (−0.02–0.13)	0.05 (−0.01–0.11)	0.15 (0.03–0.28)	0.08 (0.02–0.15)	<0.001
CKB	−0.26 (−0.72–0.19)	−0.29 (−0.88–0.30)	−0.10 (−0.41–0.22)	−0.08 (−0.35–0.18)	0.444
<b>COPD</b>					
UKB	0.30 (0.10–0.51)	0.13 (0.04–0.22)	0.70 (0.22–1.18)	0.12 (0.04–0.20)	0.013
CKB	−0.42 (−1.29–0.46)	−0.40 (−1.42–0.63)	−0.33 (−0.93–0.27)	−0.22 (−0.64–0.19)	0.323
<b>Lung cancer</b>					
UKB	0.55 (0.10–1.01)	0.18 (0.04–0.33)	1.10 (−0.01–2.22)	0.13 (0.01–0.26)	0.390
CKB	−0.65 (−1.96–0.66)	−0.73 (−2.74–1.28)	0.18 (−0.79–1.20)	0.08 (−0.37–0.54)	0.400
<b>IPF</b>					
UKB	0.09 (−0.33–0.51)	0.04 (−0.15–0.23)	−0.04 (−0.73–0.64)	−0.02 (−0.32–0.29)	0.269

RERI: relative excess risk due to interaction; AP: attributable proportion due to interaction; IPF: idiopathic pulmonary fibrosis.

(95% CI 1.27–1.77)), COPD (HR 1.64 (95% CI 1.29–2.09)) and lung cancer (HR 2.38 (95% CI 1.70–3.33)) (supplementary figure S5). There was a positive additive interaction between expanded mLOY and current smoking status on the risk of all lung diseases (RERI 0.16 (97.5% CI 0.00–0.32)) (supplementary table S14) in the UKB cohort.

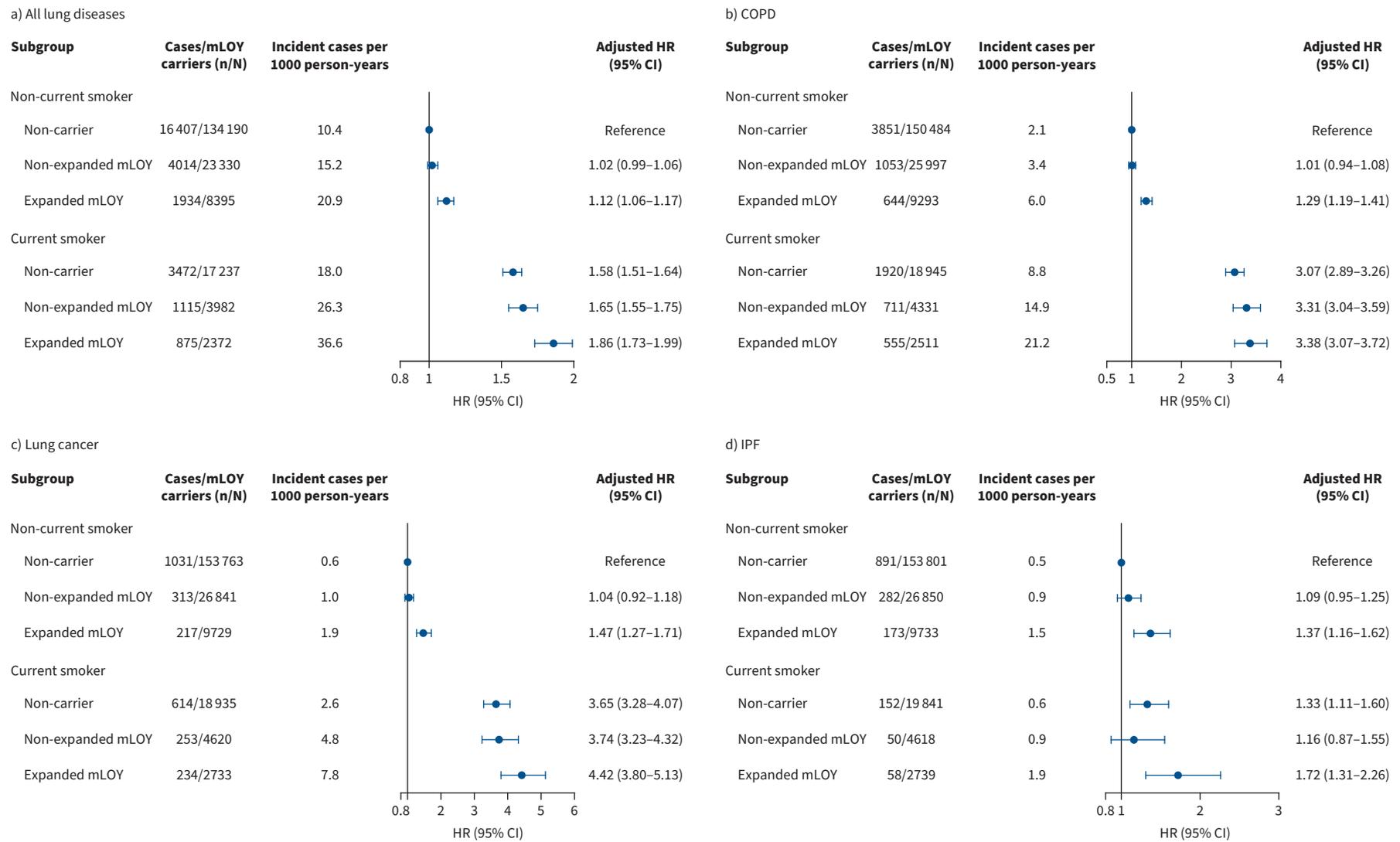
As an additional analysis, we performed a similar analysis by dividing participants into nine groups based on mLOY subtypes and smoking status (supplementary table S15). The sample size of some groups was too small (especially in the CKB) (supplementary table S16), thus the results should be interpreted with caution.

## Discussion

In these two large-scale prospective cohorts, we discovered that peripheral blood leukocyte mLOY was linked to a higher risk of incident lung diseases, including all lung diseases, COPD, lung cancer and IPF. These associations were primarily driven by expanded mLOY. In the UKB cohort, we observed evidence of interactions between mLOY and smoking status on diseases on both multiplicative and additive scales. Additionally, the cohorts showed the highest risks of lung diseases in current smokers with expanded mLOY.

Our study of the two large-scale cohorts is the first to report significant associations of mLOY with incident COPD and IPF, which has extended the disease spectrum of mLOY to a wider range of lung diseases besides lung cancer. The associations of mLOY with COPD and lung cancer have also been assessed in a phenome-wide association study conducted in the UKB population, but no significant result was generated [18]. This may be due to a stricter significance threshold ( $p < 4.39 \times 10^{-5}$ ). Conversely, the classification rules by which individuals with diseases diagnosed before enrolment were assigned into the control group could have masked the correlation to some extent since significant associations were reported between mLOY and prevalent lung diseases. Our study excluded baseline cases from the main analysis to eliminate such an effect. For lung cancer, while results from the UKB and Biobank Japan support the findings in our study [14, 16], inconsistent results from the latest study in the INTEGRAL-ILCCO cohort have also been found [25]. A Mendelian randomisation study conducted in a Chinese population even found that genetically predicted mLOY may protect against lung cancer [17]. Plausible explanations for these inconsistent results are still lacking. Therefore, more prospective research or validations in different studies or races are still needed.

To further elucidate these associations, we divided mLOY carriers into subgroups by the fraction of cells affected. We found that expanded mLOY was the main driving factor for increased disease risk, suggesting the accumulating effects of mutated cells on human lungs. Similar associations of expanded mosaic events of chromosome Y with lung cancer were reported by ZEKAVAT *et al.* [26]. In another study using median LRR (mLRR) thresholds to identify mLOY carriers, increased incident lung cancer risk was only found in carriers with lower mLRR values ( $mLRR < -0.40$ ), which also reflected a higher fraction of cells carrying mLOY mutations [14]. These results indicate that when examining the health effects of mLOY, it is important to identify mLOY mutations and infer the average levels of cells affected by mLOY.



**FIGURE 3** Joint association of mosaic loss of chromosome Y (mLOY) subtypes and smoking status with the risk of lung diseases the UK Biobank: a) all lung diseases, b) COPD, c) lung cancer and d) idiopathic pulmonary fibrosis (IPF).

The mechanisms underlying the associations between mLOY and age-related lung diseases are not fully understood. However, mLOY is thought to accelerate the ageing process and thus promote age-related diseases [7]. For instance, mLOY is linked to the dysregulation of around 500 autosomal genes, indicating the potential role of mLOY as a genome ageing inducer that ultimately results in chronic tissue decline [27]. The mLOY tumour tissue exhibited a higher level of genomic instability [28]. Furthermore, lung cancer tissues with a low Y gene expression phenotype showed significant DNA methylation loss, indicating that mLOY may contribute to epigenetic ageing in affected cells [29]. More importantly, the decline of the body's immune function is another shared pathogenesis of COPD, lung cancer and IPF [30–32], and mLOY mutations are capable of disrupting immune homeostasis and function by inducing the senescence of haematopoietic stem cells and immune cells and dysregulating the expression of immune-related genes. The effects of mLOY on tumours through the immune system have been verified in a recent experimental study by establishing Y<sup>-</sup> cancer cell models in mice, which showed a more aggressive development of Y<sup>-</sup> cancer cells and an immunosuppressive tumour micro-environment with significant T-cell exhaustion in Y<sup>-</sup> mouse models [33]. All these mechanisms may contribute to understanding the associations between mLOY and lung ageing, and the immune suppression function may account for part of the reasons why mLOY detected in blood cells could affect organs elsewhere.

Notably, the novel association of mLOY with IPF, along with previous studies focusing on cardiac fibrosis, may stimulate interest in examining the effects of mLOY on fibrotic diseases. IPF and cardiac fibrosis are characterised by the abnormal accumulation of collagen fibres. A recent study has revealed the mechanism behind accelerated cardiac fibrosis in mLOY carriers. Through the enhancement of the transforming growth factor (TGF)- $\beta$ -dependent feed-forward signalling loop, peripheral blood mLOY macrophages localised to the heart cause the activation and proliferation of fibroblasts, eventually leading to cardiac fibrosis and heart failure [12]. Meanwhile, the downregulation of genes associated with TGF- $\beta$  signalling (*TGIF1*, *TGIF2*, *SMAD7*) and the upregulation of genes related to fibrosis and TGF- $\beta$  signalling observed in circulating peripheral blood mLOY monocytes further supported such mechanism [34]. Given that fibrosis also occurred in the lung interstitium of mLOY mice and the results yielded by our study [12], we hypothesise that similar mechanisms may be the key linking mLOY with IPF. Further investigation is needed to validate this hypothesis in future biological studies, examining the dose–response effect of cell fraction.

Mechanistically, how expanded mLOY contributes to a higher risk of chronic lung disease remains elusive. A higher proportion of mCA-affected cells signifies a more severe state of the ageing genome and epigenome [7]. Meanwhile, high-frequency mCAs may also be an indicator of a reduction in the stem cell pool [35]. Disease manifests when mutated cells proliferate enough to alter physiological processes, potentially leading to serious health conditions [35]. Further research is needed to uncover the potential mechanism.

Our study has important implications for public health. Tobacco smoking is the major shared risk factor between mLOY and lung diseases [19, 20], and the associations observed were largely attenuated after adjusting for smoking status and smoking intensity. Interestingly, further investigation revealed evidence of positive interactions between mLOY and smoking behaviour with the risk of incident lung diseases, which is biologically plausible. mLOY and tobacco smoking may cause lung diseases through common mechanistic pathways such as inducing genomic instability and chronic inflammation [7, 20]. Additive interactions were observed in both former and current smokers, suggesting that ever-smokers may face greater damage from mLOY. However, former smokers suffer a relatively lower excess risk than current smokers. This implies that adopting quitting smoking behaviour may contribute to lowering the risk for mLOY carriers. However, additive interactions were not observed in the CKB population, indicating that such effects might differ in different ethnic backgrounds. In addition, our findings also support the potential clinical utility of mLOY, especially expanded mLOY, as a strong predictor of age-related lung diseases as well as a risk stratification tool for older men. Given that the presence of mLOY events is also associated with other age-related diseases, including cardiovascular disease, dementia and cancer [10–15], we believe that mLOY may serve as a powerful predictor of men's long-term health.

Our study has notable strengths. Based on genotyping and phenotyping data from two large-scale cohorts, we were able to provide the prospective association between peripheral blood leukocyte mLOY and the risk of age-related lung diseases, with comprehensive adjustment for potential confounders. Moreover, we present evidence from both British and Chinese populations, enhancing the robustness and generalisability of our findings.

Several limitations should be acknowledged. First, although we adjusted for smoking status and smoking intensity, concerns are that the residual confounding from smoking behaviour which was not fully considered

could bias our results. However, our sensitivity analysis adjusting for detailed smoking history yielded similar results, and we also found an additive interaction between mLOY and smoking, suggesting that mLOY is likely to have a similar but separate role in age-related lung diseases as tobacco smoking. Therefore, the significant associations observed may not be entirely due to residual confounding. Second, the genetic data used to detect mLOY were collected from baseline blood samples in both cohorts, which makes it impossible to assess the dynamic changes of mLOY. Given that the prevalence of mLOY increases with age, the association between mLOY and lung diseases may have been underestimated. Third, other potential risk factors, such as exposure to polycyclic aromatic hydrocarbons and glyphosate use, are difficult to define and subsequently adjust [36, 37]. However, we attempted to capture as much of these occupation-related pollutant exposure histories as possible by adjusting for occupations.

In summary, our study indicates that mLOY is associated with an increased risk of lung diseases, particularly age-related lung diseases. These associations were primarily observed in individuals with expanded mLOY. With the advancement and implementation of next-generation sequencing technology, detecting mLOY in the future will no longer be challenging. Although it is still difficult to determine whether there is a causal association between mLOY and diseases, mLOY can serve as a biomarker of great potential or even be considered a pre-disease state. Experiments aiming at identifying the mechanisms might be considered to provide more evidence and gain a deeper understanding of how mLOY contributes to disease. Future research on the development of mLOY-targeted drugs and their potential health benefits is also urgent. Additionally, this study again emphasises the importance of tobacco control, especially for mLOY carriers.

**Acknowledgements:** We thank all the participants in the UK Biobank and the China Kadoorie Biobank. The most important acknowledgement is to the participants in the China Kadoorie Biobank study and the members of the survey teams in each of the 10 regional centres, as well as to the project development and management teams based at Beijing, Oxford and the 10 regional centres. This research was conducted using the UK Biobank Resource under application number 86473.

**Data availability:** Researchers can apply for access to UK Biobank resources and related data. Details of how to access China Kadoorie Biobank data are available at <https://www.ckbiobank.org/data-access>.

**Ethics statement:** The UK Biobank has approvals from the North West Multi-Centre Research Ethics Committee, and the China Kadoorie Biobank was approved by the Ethical Review Committee of the Chinese Center for Disease Control and Prevention (Beijing, China) and the Oxford Tropical Research Ethics Committee, University of Oxford (Oxford, UK). All participants provided informed consent at baseline.

**Author contributions:** Study concept and design: D. Sun. Statistical analysis and interpretation of data: C. Weng and Y. Zhao. Data validation: Y. Zhao and M. Song. Drafting of manuscript: C. Weng and Y. Zhao. Contributed to study design, sample collection and critical revision of the manuscript for important intellectual content: C. Weng, Y. Zhao, M. Song, Z. Shao, Y. Pang, C. Yu, P. Pei, L. Yang, I.Y. Millwood, R.G. Walters, Y. Chen, H. Du, J. Chen, Z. Chen, G. Genovese, C. Terao, J. Lv, L. Li and D. Sun. Study supervision: D. Sun.

**Conflict of interest:** The authors have no potential conflicts of interest to disclose.

**Support statement:** This work was supported by National Natural Science Foundation of China (82103920, 82192901, 82192904, 82192900, 81941018, 82388102) and National Key R&D Program (2020YFC2003405, 2023YFC2509400). The China Kadoorie Biobank baseline survey and the first re-survey were supported by the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up has been supported by Wellcome grants to Oxford University (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z) and grants from the National Key R&D Program of China (2016YFC0900500), National Natural Science Foundation of China (81390540, 91846303, 81941018), and Chinese Ministry of Science and Technology (2011BAI09B01). DNA extraction and genotyping were supported by GlaxoSmithKline and the UK Medical Research Council (MC-PC-13049, MC-PC-14135). Funding information for this article has been deposited with the Crossref Funder Registry.

## References

- 1 Schneider JL, Rowe JH, Garcia-de-Alba C, *et al*. The aging lung: physiology, disease, and immunity. *Cell* 2021; 184: 1990–2019.
- 2 Cho SJ, Stout-Delgado HW. Aging and lung disease. *Annu Rev Physiol* 2020; 82: 433–459.
- 3 Arimura-Omori M, Kiyohara C, Yanagihara T, *et al*. Association between telomere-related polymorphisms and the risk of IPF and COPD as a precursor lesion of lung cancer: findings from the Fukuoka Tobacco-Related Lung Disease (FOLD) registry. *Asian Pac J Cancer Prev* 2020; 21: 667–673.

- 4 GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir Med* 2020; 8: 585–596.
- 5 GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020; 396: 1204–1222.
- 6 Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet* 2017; 389: 1941–1952.
- 7 Dai X, Guo X. Decoding and rejuvenating human ageing genomes: lessons from mosaic chromosomal alterations. *Age Res Rev* 2021; 68: 101342.
- 8 Loh PR, Genovese G, Handsaker RE, et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* 2018; 559: 350–355.
- 9 Loh PR, Genovese G, McCarroll SA. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* 2020; 584: 136–141.
- 10 Forsberg LA, Rasi C, Malmqvist N, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet* 2014; 46: 624–628.
- 11 Thompson DJ, Genovese G, Halvardson J, et al. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* 2019; 575: 652–657.
- 12 Sano S, Horitani K, Ogawa H, et al. Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality. *Science* 2022; 377: 292–297.
- 13 Garcia-Gonzalez P, de Rojas I, Moreno-Grau S, et al. Mendelian randomisation confirms the role of Y-chromosome loss in Alzheimer’s disease aetiopathogenesis in men. *Int J Mol Sci* 2023; 24: 898.
- 14 Loftfield E, Zhou W, Yeager M, et al. Mosaic Y loss is moderately associated with solid tumor risk. *Cancer Res* 2019; 79: 461–466.
- 15 Duan Q, Gao Y, Cao X, et al. Mosaic loss of chromosome Y in peripheral blood cells is associated with age-related macular degeneration in men. *Cell Biosci* 2022; 12: 73.
- 16 Terao C, Momozawa Y, Ishigaki K, et al. GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation. *Nat Commun* 2019; 10: 4719.
- 17 Qin N, Li N, Wang C, et al. Association of mosaic loss of chromosome Y with lung cancer risk and prognosis in a Chinese population. *J Thorac Oncol* 2019; 14: 37–44.
- 18 Lin SH, Brown DW, Rose B, et al. Incident disease associations with mosaic chromosomal alterations on autosomes, X and Y chromosomes: insights from a phenome-wide association study in the UK Biobank. *Cell Biosci* 2021; 11: 143.
- 19 Dumanski JP, Rasi C, Lönn M, et al. Smoking is associated with mosaic loss of chromosome Y. *Science* 2015; 347: 81–83.
- 20 Lugg ST, Scott A, Parekh D, et al. Cigarette smoke exposure and alveolar macrophages: mechanisms for lung disease. *Thorax* 2022; 77: 94–101.
- 21 Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; 12: e1001779.
- 22 Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011; 40: 1652–1666.
- 23 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011; 30: 377–399.
- 24 Duckworth A, Gibbons MA, Allen RJ, et al. Telomere length and risk of idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease: a Mendelian randomisation study. *Lancet Respir Med* 2021; 9: 285–294.
- 25 Cheng C, Hong W, Li Y, et al. Mosaic chromosomal alterations are associated with increased lung cancer risk: insight from the INTEGRAL-ILCCO cohort analysis. *J Thorac Oncol* 2023; 18: 1003–1016.
- 26 Zekavat SM, Lin SH, Bick AG, et al. Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. *Nat Med* 2021; 27: 1012–1024.
- 27 Dumanski JP, Halvardson J, Davies H, et al. Immune cells lacking Y chromosome show dysregulation of autosomal gene expression. *Cell Mol Life Sci* 2021; 78: 4019–4033.
- 28 Muller P, Velazquez Camacho O, Yazbeck AM, et al. Why loss of Y? A pan-cancer genome analysis of tumors with loss of Y chromosome. *Comput Struct Biotechnol J* 2023; 21: 1573–1583.
- 29 Willis-Owen SAG, Domingo-Sabugo C, Starren E, et al. Y disruption, autosomal hypomethylation and poor male lung cancer survival. *Sci Rep* 2021; 11: 12453.
- 30 Kheradmand F, Zhang Y, Corry DB. Contribution of adaptive immunity to human COPD and experimental models of emphysema. *Physiol Rev* 2023; 103: 1059–1093.
- 31 Lahiri A, Maji A, Potdar PD, et al. Lung cancer immunotherapy: progress, pitfalls, and promises. *Mol Cancer* 2023; 22: 40.
- 32 Shenderov K, Collins SL, Powell JD, et al. Immune dysregulation as a driver of idiopathic pulmonary fibrosis. *J Clin Invest* 2021; 131: e143226.

- 33 Abdel-Hafiz HA, Schafer JM, Chen X, *et al.* Y chromosome loss in cancer drives growth by evasion of adaptive immunity. *Nature* 2023; 619: 624–631.
- 34 Mas-Peiro S, Abplanalp WT, Rasper T, *et al.* Mosaic loss of Y chromosome in monocytes is associated with lower survival after transcatheter aortic valve replacement. *Eur Heart J* 2023; 44: 1943–1952.
- 35 Hubbard AK, Brown DW, Machiela MJ. Clonal hematopoiesis due to mosaic chromosomal alterations: impact on disease risk and mortality. *Leuk Res* 2023; 126: 107022.
- 36 Bai Y, Guan X, Wei W, *et al.* Effects of polycyclic aromatic hydrocarbons and multiple metals co-exposure on the mosaic loss of chromosome Y in peripheral blood. *J Hazard Mater* 2021; 414: 125519.
- 37 Chang VC, Zhou W, Berndt SI, *et al.* Glyphosate use and mosaic loss of chromosome Y among male farmers in the Agricultural Health Study. *Environ Health Perspect* 2023; 131: 127006.