

DOI : 10.3760/cma.j.cn112338-20230412-00228

收稿日期 : 2023-04-12

本文编辑 : 万玉立

大型队列研究

中国 10 个地区成年男性 Y 染色体嵌合缺失流行病学分布及其与肺癌的前瞻性关联研究

赵禹焯¹ 宋明钰¹ 吕筠^{1,2,3} 余灿清^{1,2,3} 裴培² 杜怀东^{4,5} 陈君石⁶ 陈铮鸣⁵ 李立明^{1,2,3} 孙点剑^{1,2,3} 代表中国慢性病前瞻性研究项目协作组

¹北京大学公共卫生学院流行病学与卫生统计学系, 北京 100191; ²北京大学公众健康与重大疫情防控战略研究中心, 北京 100191; ³重大疾病流行病学教育部重点实验室 (北京大学), 北京 100191; ⁴英国牛津大学医学研究委员会人口健康研究组, 牛津 OX3 7LF; ⁵英国牛津大学临床与流行病学研究中心纳菲尔德人群健康系, 牛津 OX3 7LF; ⁶国家食品安全风险评估中心, 北京 100022

通信作者 : 孙点剑一, Email : dsun1@bjmu.edu.cn

【摘要】目的 描述和分析中国 10 个地区成年男性 Y 染色体嵌合缺失 (mLOY) 携带者流行病学分布特征, 并探究 mLOY 与肺癌的前瞻性关联。**方法** 数据来源于 CKB 研究基线调查 (人口学特征、生活方式、体格指标等)、遗传分型及随访监测 (截至 2018 年 12 月 31 日的健康结局), 采用嵌合染色体变异 workflow 检测识别 mLOY 携带者, 分年龄、地区、基线生活方式及疾病特征描述其流行病学分布特征, 采用多因素 logistic 回归模型分析 mLOY 的影响因素, 采用 Cox 比例风险回归模型探究 mLOY 和肺癌的前瞻性关联, 在不同年龄段及不同吸烟状况人群中进行分层研究, 并采用乘积分布法分析 mLOY 在吸烟与肺癌关联间潜在的中介作用。**结果** 研究纳入 42 859 名成年男性, 共检出 2 458 名 mLOY 携带者 (检出率为 5.7%), 检出率随年龄增长而增加 ($P < 0.05$), 城市地区检出率 ($7.3\% \pm 0.2\%$) 高于农村地区 ($4.7\% \pm 0.1\%$), 吸烟可能是 mLOY 检出的危险因素 ($OR = 1.49$, $95\%CI : 1.36 \sim 1.64$)。经过平均 11.1 年随访, 共累计观察到 1 041 名肺癌新发病例, mLOY 携带者新发肺癌的风险增加了 24% ($HR = 1.24$, $95\%CI : 1.01 \sim 1.52$), 扩增性 mLOY 携带者 (携带 mLOY 细胞分数 $\geq 10\%$) 肺癌发生风险增加了 50% ($HR = 1.50$, $95\%CI : 1.13 \sim 2.00$), 未发现年龄或吸烟状况的修饰效应作用 (交互作用 $P > 0.05$)。mLOY 在吸烟和肺癌的关联中的中介效应显著, 中介效应的点估计值 ($95\%CI$) 为 0.09 (0.01~0.17)。**结论** 我国 10 个地区成年男性 mLOY 检出率存在明显的社会经济及生活方式分布差异, 携带 mLOY 尤其是扩增性 mLOY 会显著增加肺癌的发病风险, 且 mLOY 可能在吸烟与肺癌的关联中起中介作用。

【关键词】 Y 染色体嵌合缺失; 流行病学分布; 肺癌; 前瞻性关联

基金项目 : 国家自然科学基金 (82103920, 82192901, 82192904, 82192900, 81941018); 国家重点研发计划 (2020YFC2003405); 中国香港 Kadoorie Charitable 基金; 英国 Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z); 英国医学研究理事会基金 (MCPC_14135, MC_PC_13049)

Epidemiological distribution of mosaic loss of chromosome Y in adult men in ten study regions in China and its prospective association with lung cancer

Zhao Yuxuan¹, Song Mingyu¹, Lyu Jun^{1,2,3}, Yu Canqing^{1,2,3}, Pei Pei², Du Huaidong^{4,5}, Chen Junshi⁶, Chen Zhengming⁵, Li Liming^{1,2,3}, Sun Dianjianyi^{1,2,3}, for the China Kadoorie Biobank Collaborative Group

¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; ²Peking University Center for Public Health and Epidemic Preparedness

& Response, Beijing 100191, China;³Key Laboratory of Epidemiology of Major Diseases(Peking University), Ministry of Education,Beijing 100191, China; ⁴ Medical Research Council Population Health Research Unit at the University of Oxford, Oxford OX3 7LF, United Kingdom;⁵Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, United Kingdom; ⁶ China National Center for Food Safety Risk Assessment, Beijing 100022, China

Corresponding author: Sun Dianjianyi, Email: dsun1@bjmu.edu.cn

【Abstract】 Objective To detect the mosaic loss of chromosome Y events in adult men in ten study regions in China, describe the epidemiological distribution of mLOY carriers and assess its prospective association with lung cancer. **Methods** Based on the baseline, genetic and follow-up (as of December 31, 2018) data from China Kadoorie Biobank, we used mosaic Chromosomal Alterations pipeline to detect mLOY carriers and described the epidemiological characteristics of mLOY carriers in adult men, including age, area distribution, lifestyle and disease history, in ten regions in China. We used multivariate logistic regression model to find the potential relevant factor of mLOY. Cox proportional hazard regression model was fitted to assess the prospective association of mLOY with lung cancer. Stratification analysis were conducted to evaluate the potential modification effects of smoking and age. We also conducted mediation analysis to assess the mediating effect of mLOY in the association between smoking and lung cancer. **Results** A total of 42 859 adult men were included in our analysis, in whom 2 458 mLOY carriers were detected (5.7%). The detection rate increased with age ($P<0.05$). The detection rate was higher in urban area ($7.3\%\pm 0.2\%$) than that in rural area ($4.7\%\pm 0.1\%$) . The results of logistic regression analysis indicated that smoking might be a risk factor for the detection of mLOY ($OR=1.49$, $95\%CI: 1.36-1.64$). After an average of 11.1 years of follow-up, 1 041 lung cancer cases were observed. The prospective analysis showed that mLOY carriers had increased risk for lung cancer by 24% compared with non-mLOY carriers ($HR=1.24$, $95\%CI: 1.01-1.52$). Expanded mLOY carriers (mLOY cell fraction $\geq 10\%$) were associated with 50% higher risk for lung cancer ($HR=1.50$, $95\%CI : 1.13\sim 2.00$). Stratification analysis showed there were no modification effects of smoking and age in the association between mLOY and lung cancer (all $P > 0.05$). Mediation analysis showed that mLOY could be a mediating factor in the association between smoking and lung cancer, the estimated effect was 0.090 (0.006-0.170). **Conclusions** There were significant differences in detection rate of mLOY among people with different social-economic characteristics and lifestyles. Besides, mLOY carriers, especially expanded mLOY carriers, had increased risk for lung cancer and mLOY could be a mediating factor in the association between smoking and lung cancer.

【Key words】 Mosaic loss of chromosome Y; Epidemiological distribution; Lung cancer; Prospective study

Fund programs: National Natural Science Foundation of China (82103920, 82192901, 82192904, 82192900, 81941018), National Key Research and Development Program of China (2020YFC2003405); Kadoorie Charitable Foundation in Hong Kong of China; United Kingdom Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z); United Kingdom Medical Research Council (MCPC_14135,MC_PC_13049)

Y 染色体嵌合缺失 (mosaic loss of chromosome Y, mLOY) 是指体细胞部分细胞 Y 染色体发生缺失突变 (染色体部分或者整体) 而造成体内产生多种核型细胞系的现象^[1], 是最常见的嵌合染色体变异 (mosaic chromosome alteration, mCA) 类型^[2]。Y 染色体是性别决

定及促进男性发育的染色体，而 mLOY 在衰老、细胞癌变等生物学过程可能起到重要作用^[3-5]，可作为基因组失稳或衰老相关疾病的生物学标志^[6-7]。目前 mLOY 研究多集中于欧洲地区、美洲地区或日本等，基于中国大型人群队列的研究较少^[2]。

肺癌是我国男性发病率最高的癌症^[8]，也是非血液系统癌症中与 mLOY 关联研究结果较多的癌症之一^[7]，但现有基于人群队列的研究结果还并未得出统一的结论。数项基于英国生物银行 (UK Biobank, UKB)、日本生物银行 (Biobank Japan, BBJ) 的前瞻性研究均发现研究对象体内携带 mLOY 会显著增加肺癌的发病风险^[5, 9-10]，而 Lin 等^[11]利用 UKB 数据进行的全表型组关联研究 (phenome-wide association study, PheWAS) 却未发现 mLOY 和肺癌之间的关联，近期一项研究显示遗传预测的 mLOY 会对于肺癌起保护作用，但吸烟会使得此效应衰减^[12]。本研究利用中国慢性病前瞻性研究 (China Kadoorie Biobank, CKB) 基线遗传数据及调查数据，检测中国 10 个地区成年男性 mLOY 携带者，对其流行病学分布特征进行描述，并探究其与肺癌的前瞻性关联。

对象与方法

1. **研究人群：**以 CKB 数据为基础，CKB 于 2004-2008 年开展基线研究，其项目地涉及 5 个城市地区 (山东省青岛市、黑龙江省哈尔滨市、海南省海口市、江苏省苏州市、广西壮族自治区柳州市) 和 5 个农村地区 (四川省彭州市、甘肃省天水市、河南省辉县市、浙江省桐乡市、湖南省浏阳市)，共招募 512 724 名 30~79 岁的研究对象完成基线调查。CKB 项目获得英国牛津大学和中国 CDC 伦理委员会审查批准，所有研究对象签署知情同意书。有关 CKB 项目的详细介绍见文献^[13-15]。具有基因分型和对应基线数据的 CKB 研究对象共 100 640 名，本研究在剔除基因分型质量不佳个体 ($n=314$)、问卷记录性别和基因型计算性别不匹配 ($n=28$)、所有女性研究对象 ($n=57 438$)，以及 BMI 信息缺失者 ($n=1$) 后，共纳入 42 859 名成年男性。

2. 研究内容：

(1) 数据收集：在 CKB 基线调查中，一般人口学特征 (性别、年龄、地区、最高文化程度、职业等)、生活方式特征 (吸烟状况、饮酒状况、体力活动水平、是否接受过输血、是否有偿献血等)、疾病史 (呼吸系统疾病、癌症等)、服药史、疾病家族史、恶性肿瘤家族史等信息均由受过统一培训的调查员通过面对面访谈电子问卷调查获得。根据从事的不同体力活动水平 (工作、交通、休闲及家务) 的累积时间 (h/d) 以及对应的代谢当量 (MET) 相乘加和，得到研究对象每天的体力活动水平 (MET-h/d)，具体 MET 测量方法见文献^[16]。体格检查项目 (身高、体重、血压、心率等) 均统一由受过培训的调查员按照标准化操作手册进行。BMI (kg/m^2) 由体重 (kg) 除以身高 (m) 的平方计算得到。CKB 项目的基因分型由华大基因科技服务有限公司完成，共使用了两套针对中国汉族人定制化设计的芯片 (Affymetrix Axiom[®] CKB Array 1 & 2)。

(2) mLOY 的识别检测：研究采用 Loh 等^[17]于 2018 年建立的嵌合染色体变异 workflow (mosaic Chromosomal Alterations pipeline, MoChA) 来识别检测研究人群的 mLOY 携带者。MoChA 通过基因芯片原始的基因强度数据计算 R 比对数值 (LRR) 及 B 等位基因频率 (BAF) 两项指标来判断研究对象是否携带 mCA 及变异类别。mLOY 的检出率为检出的 mLOY 携带者人数除以总人数。另外，MoChA 还将计算不同个体携带 mLOY 的细胞分数，依据不同的细胞分数研究对象又可分为非携带者、非扩增性 mLOY 携带者 (携带 mLOY 细胞分数 < 10%) 及扩增性 mLOY 携带者 (携带 mLOY 细胞分数 $\geq 10\%$)^[9]。

(3) 结局：CKB 的长期随访从研究对象完成基线调查之日开始算起，直到出现结局事件、死亡、失访或截至 2018 年 12 月 31 日。研究对象的发病和死亡信息通过多种途径获取，包括各调查点的死亡和常规疾病监测系统、全民医疗保险数据库，以及主动的定向监测。发病或死因分类采用《国际疾病分类 (ICD) 》第十版 (ICD-10) 临床修订版本进行统一编

码。本研究的主要结局为肺癌（ICD-10：C33和C34）。

3. 统计学分析：描述和比较 mLOY 携带者和非携带者的基线特征。采用一般线性模型（连续变量）、二元或多元 logistic 回归（分组变量），在调整年龄、10 个地区后，描述 mLOY 携带者和非携带者的基线特征。采用 logistic 回归计算不同年龄组（调整 10 个地区）、不同地区（调整年龄组）人群的 mLOY 检出率，并绘制折线图和柱状图。将差异有统计学意义（ $P < 0.05$ ）的变量以及既往文献发现的 mLOY 可能影响因素纳入多因素 logistic 回归探究 mLOY 的相关因素。排除基线患有肺癌的研究对象后（ $n=18$ ），使用 Cox 比例风险回归模型，以随访时间为时间轴，以 mLOY 非携带者为参照组，探究携带 mLOY 和肺癌发生之间的前瞻性关联，计算风险比（HR）及 95%CI，对于已知或可能影响肺癌发生风险的因素进行逐步调整：模型 1：调整年龄（10 岁 1 组）、10 个地区、前 10 个遗传主成分及芯片类型；模型 2：在模型 1 的基础上调整其他人口学和生活方式变量（婚姻状况、文化程度、职业、BMI、膳食摄入情况、吸烟状况、被动吸烟状况、饮酒状况、体力活动水平）；模型 3：在模型 2 的基础上调整慢性阻塞性肺疾病（COPD）史和恶性肿瘤家族史。本研究还利用 Cox 比例风险回归模型进一步探究非扩增性 mLOY 携带者及扩增性 mLOY 携带者肺癌的发生风险。在交互分析中，通过在模型 3 中加入 mLOY 和所探究修饰因素的乘法交互项，并使用 χ^2 检验比较加入交互项后模型是否得到改善。在中介效应分析中，通过拟合 logistic 回归模型检验 mLOY 对吸烟回归的回归系数显著性，通过拟合 Cox 比例风险回归模型检验肺癌发病对 mLOY 和吸烟回归的回归系数显著性，通过 R 包“RMediation”，采用乘积分布法检验中介效应显著性^[18]。采用 R4.0.3 软件进行统计学分析，双侧检验，检验水准 $\alpha=0.05$ 。

结果

1. 研究对象：共 42 859 名男性，年龄（ 54.8 ± 11.2 ）岁，其中城市人口占 41.9%，共检出 2 458 名 mLOY（5.7%）携带者。mLOY 携带者和非携带者基线特征见表 1。与非 mLOY 携带者相比，mLOY 携带者平均年龄更大，城市人口比例更高。在调整年龄及 10 个地区后，mLOY 携带者中退休、当前吸烟、当前每日饮酒比例更高，与非携带者相比，差异有统计学意义。两组在已婚、文化程度、被动吸烟、每日摄入新鲜蔬菜/水果、体力活动水平、接受过输血、有偿献血以及疾病史等基线特征方面的差异无统计学意义。

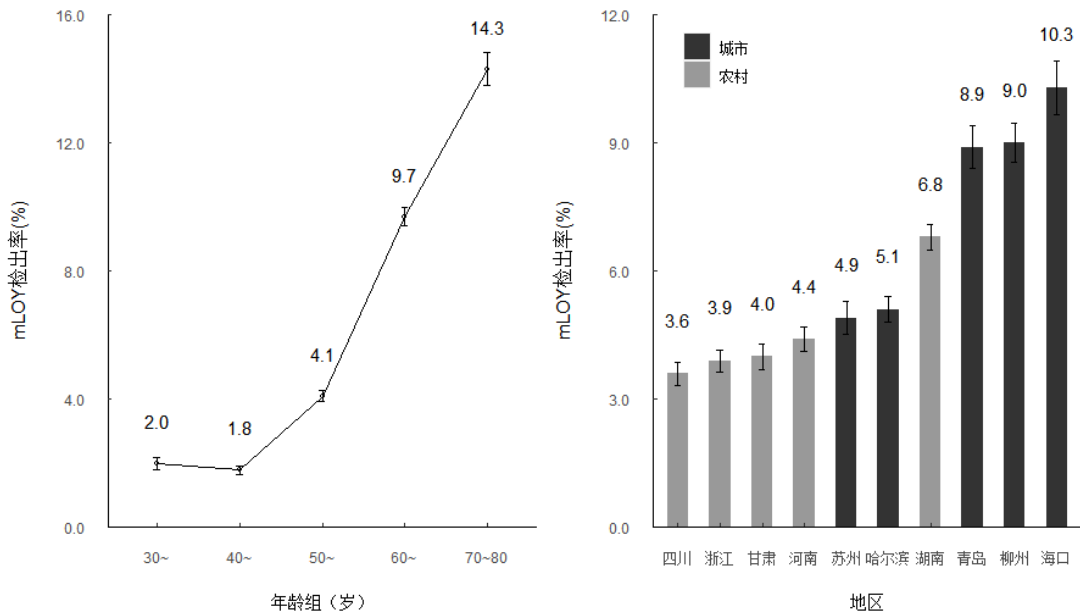
表 1 中国 10 个地区成年男性 Y 染色体嵌合缺失（mLOY）携带者和非携带者基线特征

特征	非携带者	mLOY 携带者	P 值
样本量 (%)	40 401 (94.3)	2 458 (5.7)	
基线年龄 ($\bar{x} \pm s$)	54.30 \pm 11.11	62.72 \pm 9.83	<0.001
城市地区 (%)	16 678 (41.3)	1 264 (51.4)	<0.001
已婚 (%)	37 119 (91.9)	2 185 (88.9)	0.943
职业 (%)			
农业及其相关行业	18 163 (45.0)	885 (36.0)	
工业及其相关行业	6 630 (16.4)	177 (7.2)	0.099
非工农业	5 952 (14.7)	204 (8.3)	0.866
退休	6 361 (15.7)	952 (38.7)	<0.001
其他	3 295 (8.2)	240 (9.8)	0.285
文化程度 (%)			
小学及以下	18 372 (45.5)	1 337 (54.4)	
初高中	18 890 (46.7)	910 (37.0)	0.234

大专及以上	3 139 (7.8)	211 (8.6)	0.685
当前吸烟 (%)	27 444 (67.9)	1 747 (71.1)	<0.001
被动吸烟 (%)	34 225 (84.7)	1 871 (76.1)	0.242
当前每日饮酒 (%)	12 159 (30.1)	838 (34.1)	0.003
每日摄入新鲜蔬菜/水果 (%)	38 279 (94.7)	2 342 (95.3)	0.311
BMI (%)			
<18.5	2 205 (5.5)	183 (7.4)	
18.5~	21 506 (53.2)	1 358 (55.2)	0.614
24.0~	12 983 (32.1)	729 (29.7)	0.362
≥28.0	3 707 (9.2)	188 (7.6)	0.133
体力活动水平 (MET-h/d, $\bar{x} \pm s$)	20.88±15.21	14.67±12.96	0.075
接受过输血 (%)	1 394 (3.4)	119 (4.8)	0.336
有偿献血 (%)	153 (0.4)	11 (0.4)	0.343
疾病史			
高血压 (%)	17 148 (42.4)	1 212 (49.3)	0.054
冠心病 (%)	993 (2.5)	130 (5.3)	0.078
卒中或短暂性脑缺血发作 (%)	694 (1.7)	72 (2.9)	0.502
COPD (%)	4 659 (11.5)	421 (17.1)	0.111
糖尿病 (%)	1 335 (3.3)	120 (4.9)	0.783
癌症家族史 (%)	6 761 (16.7)	350 (14.2)	0.418
芯片 2 (%)	25 904 (64.1)	1 433 (58.3)	0.613

注：MET：代谢当量；COPD：慢性阻塞性肺疾病；年龄的 *P* 值调整 10 个地区，地区分布的 *P* 值调整年龄，其余基线特征和 *P* 值均调整年龄和地区

2. mLOY 检出率及可能相关因素：男性研究人群中总体 mLOY 检出率为 5.7%，分别调整地区和年龄后的不同年龄组、不同地区人群 mLOY 检出率见图 1。调整地区分布后，mLOY 的检出率随年龄增长而增加 ($P < 0.05$)，且在 50 岁之后增长迅速，70~80 岁男性人群 mLOY 检出率为 14.3±0.5%。调整年龄分布后，10 个地区中城市地区的 mLOY 的检出率整体高于农村地区（城市：7.3%±0.2%；农村：4.7%±0.1%； $P < 0.001$ ），城市地区中检出率最高为海口项目点（10.3%±0.6%），农村地区检出率最高为湖南项目点（6.8%±0.3%），四川项目点检出率最低（3.6%±0.3%）。



注：以省份名称代替农村项目点，以城市名称代替城市项目点

图 1 调整后中国 10 个地区成年男性 mLOY 携带者的年龄、地区检出率

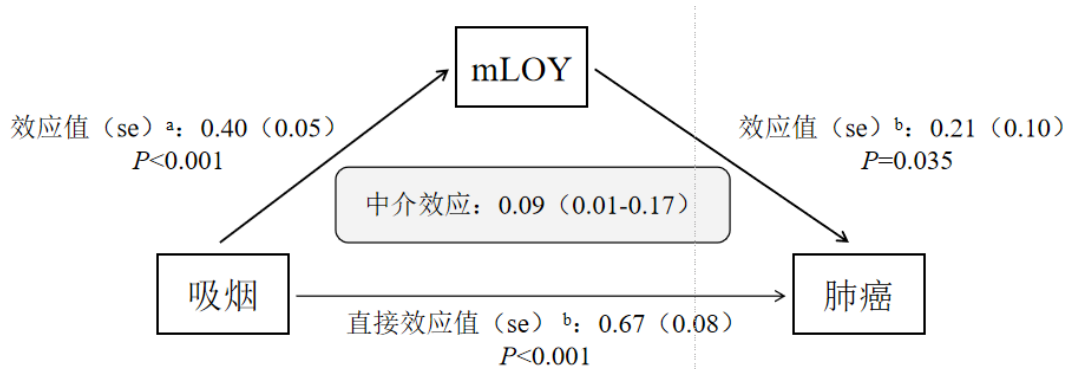
将基线年龄、10 个地区、职业、吸烟状况、被动吸烟状况、饮酒状况、疾病史（高血压、冠心病、卒中或短暂性脑缺血、糖尿病、COPD、前 10 个遗传主成分以及芯片类型纳入多因素 logistic 回归结果显示，退休者检出 mLOY 的可能性更高 ($OR=1.37$, $95\%CI: 1.18\sim1.60$)；吸烟可能是 mLOY 检出的危险因素 ($OR=1.49$, $95\%CI: 1.36\sim1.64$)；未发现当前每日饮酒、被动吸烟状况、高血压疾病史、冠心病疾病史、卒中或短暂性脑缺血发作疾病史、糖尿病病史、COPD 史及芯片类型与 mLOY 检出之间存在统计学关联。

3. mLOY 和肺癌的关联：在剔除基线患有肺癌 ($n=18$) 的研究对象后，前瞻性关联研究共纳入 42 841 名研究对象，平均随访 11.1 年，累计观察到 1 041 名肺癌新发病例，粗发病密度为 2.2/1 000 人年。mLOY 携带者新发肺癌的风险增加了 24% ($HR=1.24$, $95\%CI: 1.01\sim1.52$, 表 2)。进一步采用模型 3 探究扩增性和非扩增性 mLOY 携带者相比 mLOY 非携带者肺癌的发生风险，扩增性 mLOY 携带者肺癌发生风险增加了 50% ($HR=1.50$, $95\%CI: 1.13\sim2.00$)，非扩增性 mLOY 携带与肺癌发生风险无统计学关联 ($HR=1.08$, $95\%CI: 0.83\sim1.41$)。交互作用及分层分析结果未发现年龄或吸烟状况对 mLOY 和肺癌之间的关联起修饰效应 (交互作用 $P>0.05$)，但当前吸烟组 mLOY 和肺癌关联的 HR 值整体略高于非吸烟组。采用乘积分布检验 mLOY 的中介效应结果见图 2，mLOY 在吸烟和肺癌的关联中的中介效应显著，中介效应的点估计值 ($95\% CI$) 为 0.09 (0.01~0.17)。

表 2 中国 10 个地区成年男性 mLOY 携带者和肺癌的前瞻性关联研究

组别	样本量	发病密度 (/1 000 人年)	模型 1 ^a		模型 2 ^b		模型 3 ^c		交互 <i>P</i> 值
			<i>HR</i> 值 (95% <i>CI</i>)	<i>P</i> 值	<i>HR</i> 值 (95% <i>CI</i>)	<i>P</i> 值	<i>HR</i> 值 (95% <i>CI</i>)	<i>P</i> 值	
总体	42 841	2.2	1.34 (1.10~1.64)	0.004	1.24 (1.02~1.52)	0.035	1.24 (1.01~1.52)	0.035	0.640
年龄组 (岁)									
<60	27 971	1.2	1.58 (0.98~2.53)	0.060	1.38 (0.86~2.22)	0.186	1.38 (0.86~2.22)	0.188	
≥60	14 870	4.5	1.28 (1.02~1.60)	0.030	1.27 (1.02~1.59)	0.034	1.27 (1.01~1.59)	0.037	0.546
吸烟状况									
当前不吸	13 663	1.4	1.10 (0.66~1.82)	0.724	1.11 (0.67~1.84)	0.696	1.09 (0.66~1.82)	0.728	
当前吸	29 178	2.6	1.28 (1.03~1.60)	0.027	1.26 (1.01~1.57)	0.041	1.26 (1.01~1.57)	0.040	

注：^a模型 1：采用 Cox 风险比例回归模型，以 mLOY 非携带者为参照组，调整年龄（10 岁 1 组）、10 个地区、前 10 个遗传主成分、芯片类型，计算 *HR* 值；^b模型 2：在模型 1 的基础上调整其他人口学变量（婚姻状况、文化程度、职业、BMI、膳食状况、吸烟状况、饮酒状况、体力活动水平）；^c模型 3：在模型 2 的基础上调整慢性阻塞性肺疾病史和恶性肿瘤家族史



注：^a吸烟与 mLOY 的关联：拟合多因素 logistic 回归模型，纳入基线年龄（10 岁 1 组）、10 个地区、职业、吸烟状况、被动吸烟状况、饮酒状况、疾病史（高血压、冠心病、卒中或短暂性脑缺血发作、糖尿病、慢性阻塞性肺疾病史、前 10 个遗传主成分以及芯片类型作为协变量；^b吸烟、mLOY 与肺癌的关联：拟合 Cox 比例风险回归模型，纳入年龄（10 岁 1 组）、10 个地区、人口学变量（婚姻状况、文化程度、职业、BMI、膳食状况、吸烟状况、被动吸烟状况、饮酒状况、体力活动水平）、慢性阻塞性肺疾病史、恶性肿瘤家族史、前 10 个遗传主成分以及芯片类型作为协变量；效应值均为 β 值

图 2 mLOY 在吸烟与肺癌关联中的中介作用

敏感性分析中剔除随访时间 < 2 年的人群 ($n=1186$)，前瞻性关联分析结果显示 mLOY 与总体肺癌直接存在关联 ($HR=1.24$, $95\%CI: 1.00\sim 1.54$), $P=0.046$ 。

讨论

本研究利用 CKB 遗传数据，检出了中国成年男性中 mLOY 事件并对其流行病学分布进行描述，同时检验 mLOY 与肺癌的前瞻性关联并探寻可能的修饰因素。本研究结果显示，mLOY 的检出存在显著的人群分布差异， ≥ 50 岁男性人群 mLOY 检出率显著增加，城市男性、当前吸烟及每日饮酒者检出率更高，且携带 mLOY 尤其是扩增性 mLOY 会显著增加肺癌发病风险。

本研究结果显示，成年男性人群 mLOY 检出率随年龄变化趋势与既往人群队列研究结果均一致^[5, 7, 9, 19-22]，证明其与衰老的强关联，但相比于欧洲地区、美洲地区人群，CKB 男性人群 mLOY 检出率更低。Zekavat 等^[9]在其研究中汇总了 UKB 欧洲自然人群及 BBJ 日本病例人群 mCA 检出事件，并使用 MoChA 在麻总百翰生物银行 (MGBB, 美国病例人群)、芬兰基因研究工程 (FinnGen, 欧洲自然人群)、哥伦比亚大学欧文医学中心新型冠状病毒肺炎生物银行 (CUBCOVID-19, 美国新型冠状病毒感染人群) 中检测 mCA，UKB、MGBB、FinnGen 及 CUBCOVID-19 人群 mLOY 检出率分别为 20.1%、22.0%、18.0% 及 7.7%，在 MGBB 及 FinnGen 中， >70 岁男性 mLOY 检出率 $>40\%$ ，而 CKB ≥ 70 岁男性 mLOY 检出率为 14.3%。另有一项同样基于 UKB 的研究在 206 353 例男性自然人群中检出 19.3% 的 mLOY 携带者^[11]，其人群平均年龄与 CKB 男性人群平均年龄相仿，人群分布的结果发现相比于非携带者，mLOY 携带者中白种人比例相对于黄种人 ($OR=2.82$, $95\%CI: 2.56\sim 3.12$) 及黑种人 ($OR=4.01$, $95\%CI: 3.46\sim 4.68$) 比例更高。目前基于中国其他大型人群队列的 mLOY 报道尚缺乏，因此难以直接进行比较，但以上结果也提示 mLOY 发生频率可能在不同人种间存在差异。既往少有研究报道 mLOY 检出的城乡分布差异，而本研究结果显示我国城市地区成年男性 mLOY 检出率高于农村地区，且海口项目点 mLOY 检出率最高，可能与紫外线强^[23]、男性每日吸烟量在 10 个地区中最高有关^[24]。

吸烟可能是 mLOY 的危险因素。吸烟与 mLOY 的关联曾在多个国家多个地区人群队列

研究报道^[9, 25-26]，Wright 等^[22]还利用孟德尔随机化方法发现了吸烟与 mLOY 之间的因果关联，本研究则提供了基于中国人群的证据。吸烟是公认的致癌因素之一，烟草及烟草烟雾中包含的甲醛、苯并芘等均为致癌物质^[27]，基于细胞层面的研究也证实吸烟会增加肺部细胞的突变频率，还可能观察到支气管上皮细胞中复杂的染色体异常^[28]。Zhou 等^[26]在其研究中发现，尽管 mLOY 和当前吸烟的关联效应很强，但戒烟后这种关联会减弱，提示戒烟可在一定程度上延缓突变。

前瞻性关联研究结果显示携带 mLOY 尤其是携带超过一定比例的 mLOY 细胞会显著增加肺癌的发生风险，基于 UKB、BBJ 的前瞻性研究也报道过一致结果^[5, 9-10]，但另有一项基于 UKB 的 PheWAS 研究并未发现携带 mLOY 和肺癌之间的统计学显著关联^[11]。mLOY 作为一种染色体异常，其发生即可能导致细胞内 Y 染色体区域的基因数目或基因表达异常^[2]，分子层面的研究也发现 Y 染色体非同源重组区域存在 DDX3Y、EIF1AY 等细胞周期相关基因^[29]，mLOY 可能影响细胞周期的调控从而更易引发癌变。本研究还发现了 mLOY 在吸烟与肺癌间起到的中介效应，尽管效应值相对较弱，但也在一定程度上说明烟草当中致癌物可引起细胞复制时 DNA 链的断裂和损伤及碱基对的改变^[30]，累积形成大规模的染色体变异，从而引起基因组的衰老而提升患肺癌的风险。另有一项研究发现，遗传预测的 mLOY 对肺癌可能起保护作用^[12]，而吸烟会使这种关联减弱，此研究为肺癌和 mLOY 的关联提供新视角，也说明此领域仍需更多的研究。

本研究报道了基于中国大型人群队列的 mLOY 携带者流行病学分布特征并进行了 mLOY 和肺癌之间的前瞻性关联分析，使用的样本量大，涉及地区广泛，用于检测 mLOY 的遗传数据也是由针对中国人设计的芯片由专门公司进行统一检测，结果具有可靠性。

本研究存在局限性。首先，目前所使用的遗传数据和问卷信息均来自于 CKB 基线调查，因此本研究所描述的是基线时期 mLOY 的流行病学分布，mLOY 是一种随年龄不断发展的变异现象，本研究难以对 mLOY 的动态发展进行描述；其次，CKB 研究招募参与成员采用的是非概率抽样，对于基线 mLOY 的流行病学人群分布描述可能存在一定的人群偏倚，但在描述分布和检出率时对于年龄及地区进行了调整；再次，一些环境污染物暴露如 PM₁₀、多环芳烃等可能影响 mLOY 和肺癌的发生^[31]，但本研究对于此类环境污染的定义相对困难，统计分析时仅对于与环境暴露相关的生活地区、职业等因素进行了研究；最后，本研究未能对于肺癌不同的亚型进行研究。

综上所述，本研究描述了中国 10 个地区成年男性的 mLOY 携带者流行病学分布特征，并发现携带 mLOY 会增加肺癌的发病风险。随着 DNA 信息检测技术的更新迭代，今后的 mLOY 等 mCA 的检测识别技术可能朝着更便捷、快速的方向发展，且已有研究列举将 mLOY 作为生物标志具有男性中检出率高、多组织中均能发现识别等优势^[7]，若能探明 mLOY 和疾病之间的关联或时序关系，mLOY 可能将在今后的疾病生物标志研究中提供新思路。

致谢 感谢所有参加中国慢性病前瞻性研究项目的队列成员和各项目地区的现场调查队调查员；感谢项目管理委员会、国家项目办公室、牛津协作中心和 10 个项目地区办公室的工作人员

利益冲突 所有作者声明无利益冲突

作者贡献声明 赵禹焯：分析数据、结果解释、撰写文章；宋明钰：分析与复核数据；吕筠、余灿清、裴培：实施研究、采集数据；陈君石、陈铮鸣、李立明：项目设计、方案制定、经费支持；孙点剑一：构思研究、结果解释、论文修改、经费支持

参考文献

- [1] Biasecker LG, Spinner NB. A genomic view of mosaicism and human disease[J]. *Nat Rev Genet*, 2013, 14(5): 307-320. DOI: 10.1038/nrg3424.

- [2] Dai XQ, Guo XH. Decoding and rejuvenating human ageing genomes: Lessons from mosaic chromosomal alterations[J]. *Ageing Res Rev*, 2021, 68: 101342. DOI: 10.1016/j.arr.2021.101342.
- [3] Guo XH, Dai XQ, Zhou T, et al. Mosaic loss of human Y chromosome: what, how and why[J]. *Hum Genet*, 2020, 139(4): 421-446. DOI: 10.1007/s00439-020-02114-w.
- [4] Fukami M, Miyado M. Mosaic loss of the Y chromosome and men's health[J]. *Reprod Med Biol*, 2022, 21(1): e12445. DOI: 10.1002/rmb2.12445.
- [5] Lofffield E, Zhou WY, Yeager M, et al. Mosaic Y loss is moderately associated with solid tumor risk[J]. *Cancer Res*, 2019, 79(3): 461-466. DOI: 10.1158/0008-5472.Can-18-2566.
- [6] López-Otín C, Blasco MA, Partridge L, et al. The hallmarks of aging[J]. *Cell*, 2013, 153(6): 1194-1217. DOI: 10.1016/j.cell.2013.05.039.
- [7] Guo XH, Li JF, Xue JL, et al. Loss of Y chromosome: An emerging next-generation biomarker for disease prediction and early detection?[J]. *Mutat Res Rev Mutat Res*, 2021, 788: 108389. DOI: 10.1016/j.mrrev.2021.108389.
- [8] 邱海波, 曹素梅, 徐瑞华. 基于 2020 年全球流行病学数据分析中国癌症发病率、死亡率和负担的时间趋势及与美国和英国数据的比较[J]. *癌症*, 2022, 41(4): 165-177.
- Qiu HB, Cao SM, Xu RH. Analyze the time trend of cancer incidence rate, mortality and burden in China based on the global epidemiological data in 2020 and compare it with the data from the United States and the United Kingdom[J]. *Cancer*, 2022, 41(4): 165-177.
- [9] Zekavat SM, Lin SH, Bick AG, et al. Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection[J]. *Nat Med*, 2021, 27(6): 1012-1024. DOI: 10.1038/s41591-021-01371-0.
- [10] Terao C, Momozawa Y, Ishigaki K, et al. GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation[J]. *Nat Commun*, 2019, 10(1): 4719. DOI: 10.1038/s41467-019-12705-5.
- [11] 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[J]. *Cell Biosci*, 2021, 11(1): 143. DOI: 10.1186/s13578-021-00651-z.
- [12] 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]. *J Thorac Oncol*, 2019, 14(1): 37-44. DOI: 10.1016/j.jtho.2018.09.013.
- [13] Chen ZM, Lee L, Chen JS, et al. Cohort profile: the kadoorie study of chronic disease in China (KSCDC)[J]. *Int J Epidemiol*, 2005, 34(6): 1243-1249. DOI: 10.1093/ije/dyi174.
- [14] Chen ZM, Chen JS, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up[J]. *Int J Epidemiol*, 2011, 40(6): 1652-1666. DOI: 10.1093/ije/dyr120.
- [15] 李立明, 吕筠, 郭彧, 等. 中国慢性病前瞻性研究: 研究方法和调查对象的基线特征[J]. *中华流行病学杂志*, 2012, 33(3): 249-255. DOI: 10.3760/cma.j.issn.0254-6450.2012.03.001.
- Li LM, Lv J, Guo Y, et al. The China Kadoorie Biobank: related methodology and baseline characteristics of the participants[J]. *Chin J Epidemiol*, 2012, 33(3): 249-255. DOI: 10.3760/cma.j.issn.0254-6450.2012.03.001.
- [16] Du HD, Bennett D, Li LM, et al. Physical activity and sedentary leisure time and their

- associations with BMI, waist circumference, and percentage body fat in 0.5 million adults: the China Kadoorie Biobank study[J]. *Am J Clin Nutr*, 2013, 97(3): 487-496. DOI: 10.3945/ajcn.112.046854.
- [17] Loh PR, Genovese G, Handsaker RE, et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations[J]. *Nature*, 2018, 559(7714): 350-355. DOI: 10.1038/s41586-018-0321-x.
- [18] 方杰, 温忠麟, 张敏强. 类别变量的中介效应分析[J]. *心理科学*, 2017, 40(2): 471-477. DOI: 10.16719/j.cnki.1671-6981.20170233.
- Fang J, Wen ZL, Zhang MQ. Mediation analysis of categorical variables[J]. *J Psychol Sci*, 2017, 40(2): 471-477. DOI: 10.16719/j.cnki.1671-6981.20170233.
- [19] Forsberg LA. Loss of chromosome Y (LOY) in blood cells is associated with increased risk for disease and mortality in aging men[J]. *Hum Genet*, 2017, 136(5): 657-663. DOI: 10.1007/s00439-017-1799-2.
- [20] Thompson DJ, Genovese G, Halvardson J, et al. Genetic predisposition to mosaic Y chromosome loss in blood[J]. *Nature*, 2019, 575(7784): 652-657. DOI: 10.1038/s41586-019-1765-3.
- [21] Loftfield E, Zhou WY, Graubard BI, et al. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank[J]. *Sci Rep*, 2018, 8(1): 12316. DOI: 10.1038/s41598-018-30759-1.
- [22] Wright DJ, Day FR, Kerrison ND, et al. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility[J]. *Nat Genet*, 2017, 49(5): 674-679. DOI: 10.1038/ng.3821.
- [23] 吴胜安, 邢彩盈, 朱晶晶. 海南岛气候特征分析[J]. *热带生物学报*, 2022, 13(4): 315-323. DOI: 10.15886/j.cnki.rds wxb.2022.04.001.
- Wu SA, Xing CY, Zhu JJ. Analysis of climate characteristics in Hainan island[J]. *J Trop Biol*, 2022, 13(4): 315-323. DOI: 10.15886/j.cnki.rds wxb.2022.04.001.
- [24] 王昕, 吕筠, 郭彧, 等. 中国慢性病前瞻性研究: 10个项目地区成年人群吸烟行为特征差异分析[J]. *中华流行病学杂志*, 2015, 36(11): 1200-1204. DOI: 10.3760/cma.j.issn.0254-6450.2015.11.004.
- Wang X, Lv J, Guo Y, et al. Regional differences in adults' smoking pattern: findings from China Kadoorie Biobank study in 10 areas in China[J]. *Chin J Epidemiol*, 2015, 36(11): 1200-1204. DOI: 10.3760/cma.j.issn.0254-6450.2015.11.004.
- [25] Dumanski JP, Rasi C, Lönn M, et al. Smoking is associated with mosaic loss of chromosome Y[J]. *Science*, 2014, 347(6217): 81-83. DOI: 10.1126/science.1262092.
- [26] Zhou WY, Machiela MJ, Freedman ND, et al. Mosaic loss of chromosome Y is associated with common variation near *TCL1A*[J]. *Nat Genet*, 2016, 48(5): 563-568. DOI: 10.1038/ng.3545.
- [27] Steliga MA, Dresler CM. Epidemiology of lung cancer: smoking, secondhand smoke, and genetics[J]. *Surg Oncol Clin*, 2011, 20(4): 605-618. DOI: 10.1016/j.soc.2011.07.003.
- [28] Yoshida K, Gowers KHC, Lee-Six H, et al. Tobacco smoking and somatic mutations in human bronchial epithelium[J]. *Nature*, 2020, 578(7794): 266-272. DOI: 10.1038/s41586-020-1961-1.
- [29] Cáceres A, Jene A, Esko T, et al. Extreme downregulation of chromosome Y and cancer risk in men[J]. *J Natl Cancer Inst*, 2020, 112(9): 913-920. DOI: 10.1093/jnci/djz232.

- [30] Alexandrov LB, Ju YS, Haase K, et al. Mutational signatures associated with tobacco smoking in human cancer[J]. *Science*, 2016, 354(6312): 618-622. DOI: 10.1126/science.aag0299.
- [31] Wong JYY, Margolis HG, Machiela M, et al. Outdoor air pollution and mosaic loss of chromosome Y in older men from the Cardiovascular Health Study[J]. *Environ Int*, 2018, 116: 239-247. DOI: 10.1016/j.envint.2018.04.030.