DOI：10.12287/j.issn.2096-8965.20250311

环境因素与肺癌关联的暴露组范围孟德尔随机化研究

李迎晨[1#]，那方谏[2#]，赵明芳[1*]

（1.中国医科大学附属第一医院肿瘤内科，辽宁 沈阳 110001；
2.中国医科大学网络信息中心，辽宁 沈阳 110021）

【摘要】目的 通过孟德尔随机化（MR）分析，系统探讨暴露组范围环境因素与肺癌及其亚型的因果关联。**方法** 基于大规模全基因组关联研究（GWAS）汇总数据，通过MR框架整合逆方差加权、MR-Egger等方法，分析62种环境因素与肺癌及其亚型的因果关系，并进行敏感性分析。**结果** 发现35组环境因素存在显著因果关联，其中14组为先前未被充分认识的关联，包括皮肤日光晒黑倾向与肺鳞癌及整体肺癌风险呈正相关关系；职业性重体力劳动显著增加肺鳞癌、小细胞肺癌及整体肺癌风险等。这些因素在肺癌不同亚型中呈现共享或亚型特异性关联。此外，验证分析进一步证实了11组稳健关联，分层分析揭示了吸烟状态对部分因果效应的调节作用，最后，多变量MR分析进一步明确了每日吸烟量、BMI和教育水平等因素的独立因果效应。**结论** 识别了特定环境因素与肺癌之间的因果关联，为精准预防提供遗传流行病学依据。

【关键词】肺癌；环境因素；孟德尔随机化；工具变量

中图分类号：R734.2　　**文献标识码：**A　　**文章编号：**2096-8965（2025）03-0088-13

Exposome-wide Mendelian randomization study on the association between environmental factors and lung cancer

Li Yingchen[1#], Na Fangjian[2#], Zhao Mingfang[1*]

(1. Department of Medical Oncology, The First Hospital of China Medical University, Shenyang 110001, Liaoning, China;
2. Network Center, China Medical University, Shenyang 110021, Liaoning, China)

【Abstract】 Objective Systematically investigate the causal relationships between exposome-wide environmental factors and lung cancer and its subtypes through Mendelian randomization (MR) analysis. **Methods** Leveraging large-scale genome-wide association study (GWAS) summary data, inverse-variance weighted, MR-Egger, and other MR methods were integrated within the MR framework to analyze causal associations between 62 environmental factors and lung cancer and its subtypes, complemented by sensitivity analyses. **Results** A total of 35 environmental factors were identified with significant causal associations with lung cancer and its subtypes, including 14 previously underrecognized associations. For example, susceptibility to tan was positively associated with squamous cell carcinoma and overall lung cancer

收稿日期：2025-03-20；**修回日期：**2025-05-30

基金项目：国家重点研发计划项目（2023YFC2508602）；国家自然科学基金项目（82373413）；社会科学规划基金项目（L21BGL059）；辽宁省中央引导地方科技发展资金计划项目（2022010127-JH6/1001）

通信作者：赵明芳（1971-），女，辽宁沈阳人，教授，主任医师，博士研究生导师，主要从事肺癌精准诊疗与分子转化研究，
E-mail：zhaomf618@126.com
#为共同第一作者

risk, while occupational heavy physical labor significantly increased the risk of squamous cell carcinoma, small cell lung cancer, and overall lung cancer. These factors showed both shared and subtype-specific effects across different lung cancer subtypes. Additionally, validation analyses confirmed 11 robust associations. Stratified analyses revealed that smoking status modified the causal effects of certain exposures. Finally, multivariable MR further clarified the independent causal effects of daily cigarette consumption, body mass index, and education level. **Conclusions** This study identifies specific environmental factors causally associated with lung cancer, providing genetic epidemiological evidence to support precision prevention strategies.

【**Keywords**】Lung cancer; Environmental factors; Mendelian randomization; Instrumental variables

0 引言

肺癌是全球癌症相关死亡的主要原因之一，其疾病负担持续加重。世界卫生组织统计数据显示，2022年肺癌以近250万例的新发病例成为全球最常见的癌症[1]，防控形势严峻。尽管吸烟仍是肺癌的主要危险因素[2-3]，但仍有不少病例发生于非吸烟人群，且不同病理亚型（如肺腺癌、肺鳞癌和小细胞肺癌）的病因异质性尚未完全阐明。近年来，环境因素在肺癌发生中的作用日益受到关注，流行病学研究陆续报道了各种环境因素与肺癌风险的潜在关联[4-6]。然而，当前研究仍存在诸多争议：观察性研究易受残余混杂、反向因果关系及测量偏倚的干扰，导致不同队列间结果存在显著异质性，难以区分真实的因果效应与虚假关联；多种暴露因素与不同肺癌亚型之间的因果关系尚不明确，这也严重制约了基于分子分型的精准预防策略制定。

孟德尔随机化（Mendelian Randomization，MR）作为因果推断的方法学革新，通过利用遗传变异作为工具变量（Instrumental Variable，IV），为破解这一研究困境提供了方法论突破[7]。该方法基于减数分裂过程中遗传变异的随机分配特性[8]，模拟自然随机对照试验，可有效规避传统流行病学研究中的混杂偏倚和反向因果关系，从而为解析环境暴露与肺癌的因果关联提供高阶证据。然而，既往MR研究多聚焦于单一或几个暴露因素的分析模式[9-11]，缺乏对多维度环境暴露的系统评估。

因此，在这项更新的MR研究中，我们采用了环境因素框架，涵盖社会经济因素、行为因素、代谢因素、身体测量及生活和工作环境等暴露因素[12]，并将肺癌及其不同亚型作为结局，旨在系统地调查与不同肺癌亚型具有因果关联的环境因素；同时，综合识别相同环境因素对不同肺癌亚型的因果效应差异，以期为肺癌的风险评估提供新的见解。

1 方法

1.1 研究设计

如图1所示，本研究基于MR分析的3个核心假

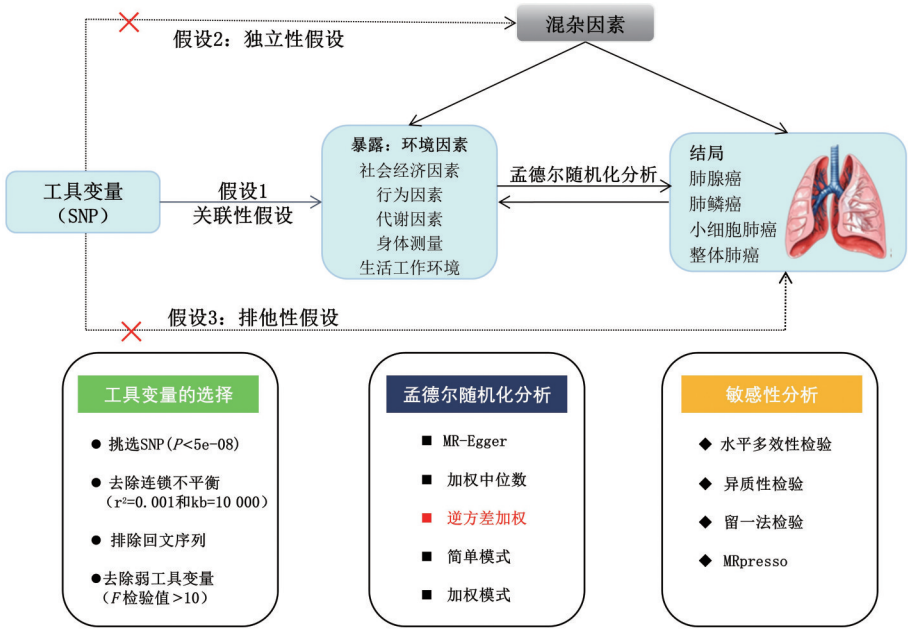


图1 MR的研究设计

Figure 1 Study design of MR analysis

设：（1）遗传IV与常见环境因素有强烈关联；（2）遗传变异与潜在混杂因素独立；（3）遗传变异仅通过常见环境因素影响肺癌的发生，而不是通过其他途径[13]。通过对全基因组关联分析（Genome-Wide Association Study，GWAS）的汇总统计数据进行分析，从中提取了与目标暴露相关的遗传变异，并根据严格标准筛选合适的IV。采用5种不同的MR方法开展双向双样本的MR分析。同时进行一系列敏感性分析，包括异质性检验、多效性评估和留一法检验，以确保研究结果的稳健性。

1.2 数据来源

暴露组范围的环境因素被系统划分为5大类别：社会经济因素、行为因素、代谢特征、身体测量指标及生活工作环境。本研究筛选出样本量充足且更新时效性较高的GWAS数据作为环境暴露源。为提升统计效力并挖掘潜在新关联，我们整合了肺癌跨学科研究的肺癌数据集作为结局，涵盖肺腺癌、肺鳞癌、小细胞肺癌和整体肺癌表型。随后利用国际肺癌联盟和芬兰数据对发现的因果关联进行了验证。通过表1系统整理了各个GWAS数据的信息，包括分类归属、样本规模及GWAS标识码。从上述GWAS的汇总统计数据中筛选出单核苷酸多态性（Single Nucleotide Polymorphism，SNP）作为IV，研究人群限定为欧洲血统群体。由于所有分

表1 MR研究的数据源

Table 1 Data sources for MR analysis

	GWAS性状	GWAS_ID	样本量（例）
社会经济因素	受教育程度	ebi-a-GCST90029013	1 131 438
	家庭收入	ukb-b-7408	397 751
	孤独感	ukb-b-8476	455 364
行为因素	吸烟的起始年龄	ieu-b-4877	607 291
	每日吸烟支数	ieu-b-25	337 334
	过去吸烟史	ukb-b-2134	424 960
	每周酒精饮料摄入量（杯/周）	ieu-b-73	335 394
	酒精摄入频率（次/周）	ukb-b-5779	462 346
	每周烈酒摄入量（杯/周）	ukb-b-1707	326 565
	每周红酒摄入量（杯/周）	ukb-b-5239	327 026
	每日咖啡摄入量（杯/日）	ebi-a-GCST008028	15 837
	咖啡摄入	ukb-b-5237	428 860
	茶饮摄入	ebi-a-GCST90096926	434 171
	睡眠时长	ukb-b-4424	460 099
	早起习惯	ebi-a-GCST90029029	452 896
	白天小睡	ukb-b-5776	452 633
	打鼾	ukb-b-17400	430 438
	中等到剧烈身体活动	ebi-a-GCST006097	337 234
	剧烈身体活动	ebi-a-GCST006098	261 055
	剧烈运动或其他锻炼	ebi-a-GCST006100	350 492
	计算机使用时间	ukb-b-4522	360 895
	手机使用时长	ukb-b-4094	456 972
代谢因素	低密度脂蛋白胆固醇	ieu-b-110	440 546
	高密度脂蛋白胆固醇	ieu-b-109	403 943
	甘油三酯	ieu-b-111	441 016
	载脂蛋白B	ieu-b-108	439 214
	餐后2 h血糖	ebi-a-GCST90002227	63 396
	空腹血糖	ebi-a-GCST90002232	200 622
	空腹胰岛素	ebi-a-GCST90002238	151 013

续表

	GWAS性状	GWAS_ID	样本量（例）
	糖化血红蛋白	ebi-a-GCST90014006	389 889
	血清 25OHD	ebi-a-GCST90000618	496 946
	尿酸水平	ebi-a-GCST90025965	437 354
	C 反应蛋白	prot-c-4337_49_2	501 428
身体测量	收缩压	ebi-a-GCST90000062	810 865
	舒张压	ebi-a-GCST90000059	810 865
	腰围	ieu-a-61	232 101
	臀围	ieu-a-49	213 038
	腰臀比	ieu-a-79	210 082
	BMI	ieu-b-40	681 275
	肺活量	ebi-a-GCST90029027	422 876
	第一秒用力呼气量	ukb-b-19657	421 986
	FEV1/FVC	ebi-a-GCST90029026	446 811
	呼气峰流速	ukb-b-12019	421 986
生活工作环境	$PM_{2.5}$	ukb-b-11312	423 796
	$PM_{2.5\sim10}$	ukb-b-12963	423 796
	PM_{10}	ukb-b-589	455 314
	氮氧化物	ukb-b-12417	456 380
	24 h平均噪声水平	ukb-b-19490	456 380
	16 h平均噪声水平	ukb-b-11532	456 380
	需要体力劳动的工作	ukb-b-2002	263 615
	皮肤晒黑的容易程度	ukb-b-533	453 065
	嘈杂的工作场所	ukb-b-2091	151 624
	背景噪声问题	ukb-a-261	330 759
	工作场所非常热：很少/从不	ukb-d-22608_0	90 644
	工作场所非常热：有时	ukb-d-22608_1	90 544
	工作场所非常冷：有时	ukb-d-22607_1	90 500
	工作场所非常冷：经常	ukb-d-22607_2	90 188
	工作时很少/从不接触油漆、稀释剂或胶水	ukb-d-22613_0	90 389
	工作场所充满化学或其他烟雾：很少/从不	ukb-d-22610_0	90 344
	工作场所充满化学或其他烟雾：有时	ukb-d-22610_1	89 060
	工作场所充满化学或其他烟雾：经常	ukb-d-22610_2	88 735
	辐射暴露	finn-b-L12_RADIATIONRELATEDSKIN	218 792
结局	肺腺癌	ieu-a-984	65 864
	肺鳞癌	ieu-a-989	62 467
	小细胞肺癌	ieu-a-988	23 371
	肺癌	ieu-a-987	85 449
	肺腺癌	ieu-a-965	18 336
	肺鳞癌	ieu-a-967	18 313
	小细胞肺癌	finn-b-C3_SCLC	218 792
	肺癌	ieu-a-966	27 209
	从未吸烟者肺癌	ebi-a-GCST004747	9 859
	曾吸烟者肺癌	ebi-a-GCST004749	40 187

析数据均源于已发表研究和公开数据库的GWAS汇总数据，因此无需另行申请伦理审批或受试者知情同意。

1.3 IV的选择

首先基于全基因组显著性水平（P<5e-08）和次等位基因频率（MAF>0.01）筛选显著SNP；随后，基于1 000基因组的欧洲人群参考面板，以连锁不平衡参数（r^2=0.001，kb=10 000）进行聚类分析，保留各独立位点内最显著SNP；进一步通过等位基因频率比对排除回文结构SNP，以消除方向性歧义；此外，还计算了各SNP的F统计量以评估其效能[14]，F统计量估算了基因变异与暴露之间的关联强度，提示IV的有效性。具体而言，F统计量越大，工具越强，统计功效越高；而F统计量小于10则表明工具效能较低，我们严格剔除F值小于10的弱IV。最后，协调了每组暴露因素和结局的GWAS数据集，确保两者的效应大小与相同的效应等位基因一致。总体而言，该筛选流程通过分阶段质控确保IV满足核心假设，既保证遗传变异与暴露的强相关性，也通过连锁不平衡聚类和回文SNP剔除实现IV的独立性及方向一致性。

1.4 统计分析

为了确定暴露组范围的环境因素与各种类型肺癌之间的关系，本研究使用了5种分析方法：逆方差加权、MR-Egger、加权中位数、简单模式和加权模式。其中逆方差加权法[15]是主要分析方法，它是基于Wald比率计算单个遗传IV的效应值，然后整合多IV效应量获得主要的因果估计；为验证结果的稳健性，进一步采用加权中位数法和MR-Egger回归[16]进行敏感性分析，前者在半数以上有效IV时保持估计可靠性，后者通过截距项检测定向多效性偏倚。仅当逆方差加权法的P值具有统计学显著性，且所有方法所得的效应方向一致时，才将该结果视为具有潜在的因果关联。对于敏感性分析，使用MR-PRESSO评估水平多效性，而MR-Egger避免将回归线强制通过零解释定向遗传多效性的存在。在回归截距非零且P<0.05的情况下，将它们视为遗传多效性的统计学显著指标。此外，通过Cochran Q的异质性检验，认为P<0.05存在异质性[17]。最后，进一步施行留一法敏感性检验，以确定单个IV是否对因果效应有显著影响。

为评估吸烟这一重要混杂因素对环境与行为暴露因素与肺癌因果关系的影响，在主分析基础上，我们进一步开展了基于吸烟状态的分层MR分析。具体地，分别采用吸烟者与非吸烟者的肺癌GWAS数据作为结局变量，评估候选暴露因素在不同吸烟状态人群中的因果效应。考虑到吸烟频率、每日吸烟量等变量与吸烟状态高度相关，可能引入混杂或交互作用，在分层分析中剔除了此类与吸烟高度重合的暴露因素，仅保留在主分析中与肺癌显著相关且不与吸烟直接重叠的变量。

为了评估多种暴露因素在不同肺癌亚型中的独立因果效应，我们分别在肺腺癌、肺鳞癌及整体肺癌中开展了多变量孟德尔随机化（Multivariable Mendelian Randomization，MVMR）分析。MVMR模型中纳入了在发现阶段和验证阶段均显著的暴露因素，并优先选择具有生物学合理性和较强遗传IV的因素。

所有分析在R4.3.1平台完成，依托“TwoSampleMR”和“MRPRESSO”包执行MR核心算法，并且采用Bonferroni校正方法进行矫正，使用“forestploter”包和“pheatmap”包进行可视化。

2 结果

在这项研究中，我们全面探讨了在环境因素框架下，62种暴露因素与4类肺癌结局之间的因果关联，包括肺腺癌、肺鳞癌、小细胞肺癌和整体肺癌。总体而言，我们的研究通过分析和错误发现率（False Discovery Rate，FDR）矫正，共识别出35组暴露—结果间的统计学显著关联（见图2），其中包括23组未经多重校正但显著的提示性关联，以及12组在Bonferroni校正后仍保持显著性的强证据关联，提示这些结果在多重检验控制下仍具稳健性。值得一提的是，有14组先前被低估和新发现的关联：皮肤日光晒黑倾向与肺鳞癌及整体肺癌风险；职业性重体力劳动显著与肺鳞癌、小细胞肺癌及整体肺癌；血清尿酸盐水平与肺腺癌及整体肺癌；睡眠时长延长与肺鳞癌及整体肺癌；噪声暴露与肺鳞癌和肺腺癌；收缩压水平与肺鳞癌；使用电脑的时间与小细胞肺癌和整体肺癌，图3显示了每个SNP与暴露及肺癌之间的关联效应。值得注意的是，由于工作场所噪声与肺腺癌只有1个SNP，所以不展示其散点图。基于更新的大规模GWAS数据，我们的MR研究也增强了一些先前已识别的因果关联的稳健性。涉及环境因素的SNP数量从1个到456个不等，每个保留SNP的F统计量均高于经验阈值10，表明所有参与此次MR研究的SNP均具有足够的有效性。

2.1 暴露组范围的环境因素与肺腺癌

首先，我们聚焦于肺腺癌，这是最常见的肺癌

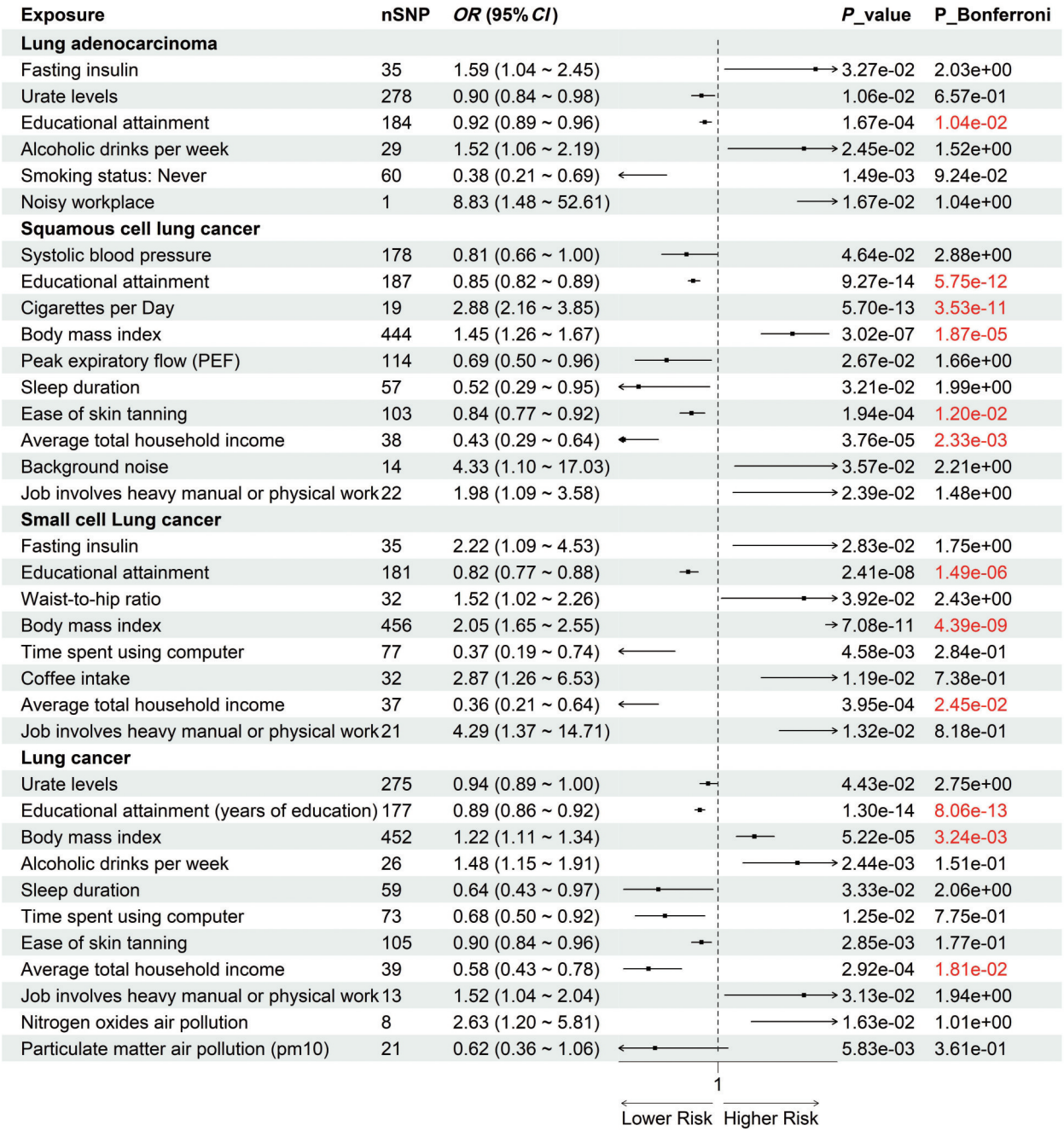


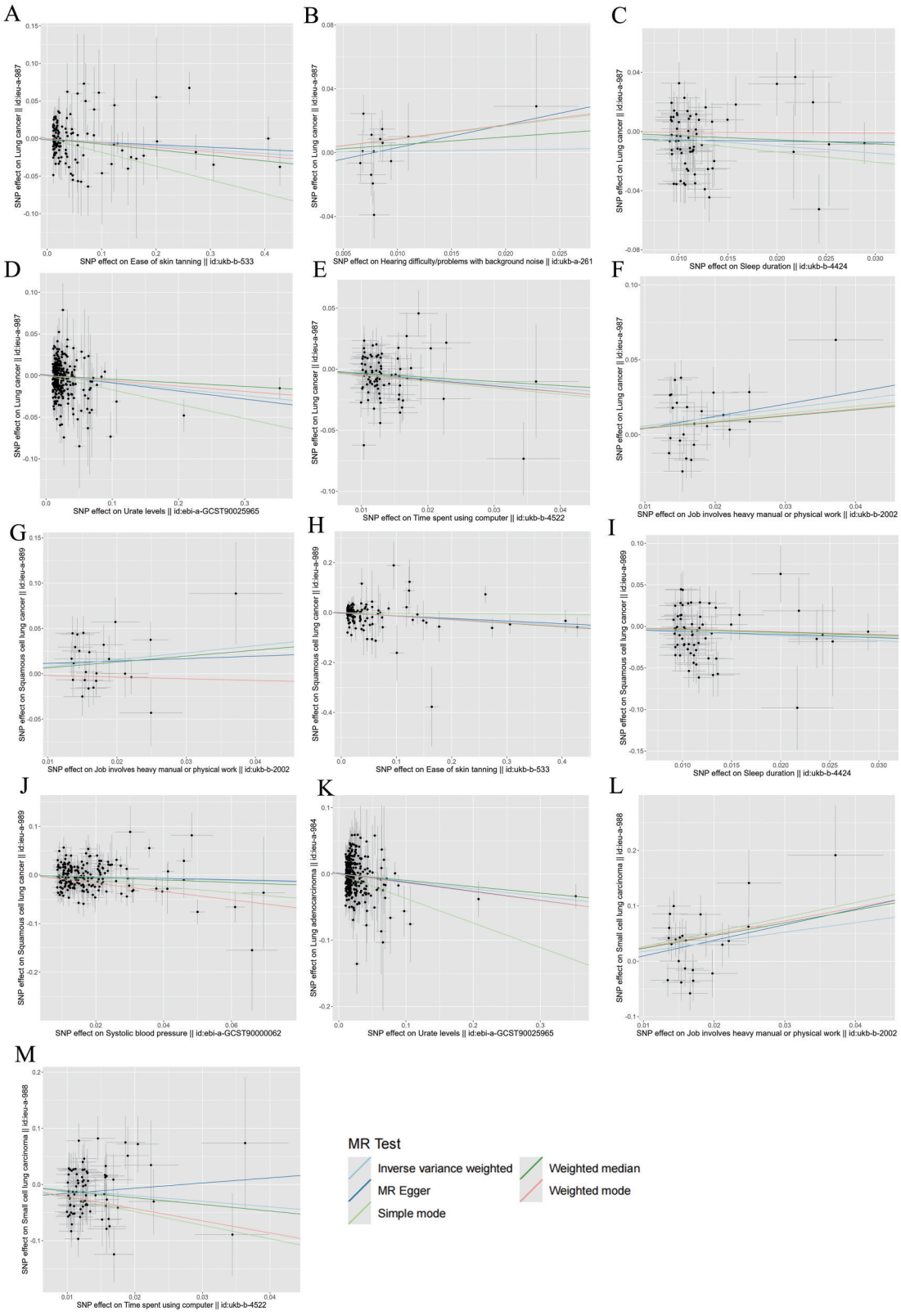
图2　基于TRICL数据集环境因素与各种肺癌亚型的MR分析结果森林图

Figure 2　Forest plot of MR analysis results showing associations between environmental factors and different lung cancer subtypes in TRICL dataset

亚型之一。通过逆方差加权的方法，并结合Bonferroni多重检验校正，发现教育水平提升在校正后仍显著与肺腺癌风险降低相关（*OR*=0.92，95%*CI*: 0.89~0.96，$P=1.67\times10^{-4}$），提示其为一个稳健且具有广泛保护效应的社会经济因素。尽管未通过多重校正，空腹胰岛素水平升高、酒精摄入增加以及嘈杂工作环境暴露仍呈现出明显的风险倾向；相反，较高的尿酸水平和非吸烟状态则表现出潜在的保护作用。总体来看，肺腺癌的风险谱受多种因素共同影响，包括教育程度、代谢状态、生活方式和职业环境。敏感性分析进一步验证了结果的稳健性，未观察到显著的异质性或水平多效性干扰。

2.2　暴露组范围的环境因素与肺鳞癌

继肺腺癌之后，我们进一步探讨了肺鳞癌的环境风险因素，结果显示，其风险模式具有明显的特异性。经过Bonferroni多重检验校正后，多个因素与肺鳞癌风险的关联仍具有统计学显著性。具体而言，教育水平提升（*OR*=0.85，95%*CI*：0.82~0.89，



A：皮肤日光晒黑倾向相关的SNP与肺癌之间的散点图；B：背景噪声下听力问题相关的SNP与肺癌之间的散点图；C：睡眠时长相关的SNP与肺癌之间的散点图；D：血尿酸盐水平相关的SNP与肺癌之间的散点图；E：使用电脑时长相关的SNP与肺癌之间的散点图；F：重体力或体力工作相关的SNP与肺癌之间的散点图；G：重体力或体力工作相关的SNP与肺鳞癌之间的散点图；H：皮肤日光晒黑倾向相关的SNP与肺鳞癌之间的散点图；I：与睡眠时长相关的SNP与肺鳞癌之间的散点图；J：与收缩压水平相关的SNP与肺鳞癌之间的散点图；K：血尿酸盐水平相关的SNP与肺腺癌之间的散点图；L：重体力或体力工作相关的SNP与小细胞肺癌之间的散点图；M：使用电脑时长相关的SNP与肺癌之间的散点图。

图3　环境暴露与肺癌MR分析的散点图

Figure 3　Scatter plot of Mendelian randomization analysis showing associations between environmental exposures and lung cancer

P=9.27×10^{-14}）、家庭收入增加（OR=0.43，95%CI：0.29~0.64，P=3.76×10^{-5}）以及日光易晒黑倾向（OR=0.84，95%CI：0.77~0.92，P=1.94×10^{-4}）等社会经济和生物性因素显著降低肺鳞癌发病风险；相反，吸烟量（每日香烟数）（OR=2.88，95%CI：2.16~3.85，P=5.70×10^{-13}）、BMI升高（OR=1.45，95%CI：1.26~1.67，P=3.02×10^{-7}）和家庭背景噪声暴露（OR=4.33，95%CI：1.10~17.03，P=3.57×10^{-2}）则显著增加肺鳞癌风险。此外，尽管收缩压升高和体力劳动职业暴露亦表现出一定的趋势性关联，但在Bonferroni校正后差异无统计学意义。综上，肺鳞癌风险受到吸烟强度、肥胖、社会经济地位下降以及噪声暴露等多重因素的综合影响，而日光反应特征可能为一种潜在的保护性生物标志物。敏感性分析未发现显著的异质性或水平多效性，提示上述关联具有较强的稳健性。

2.3 暴露组范围的环境因素与小细胞肺癌

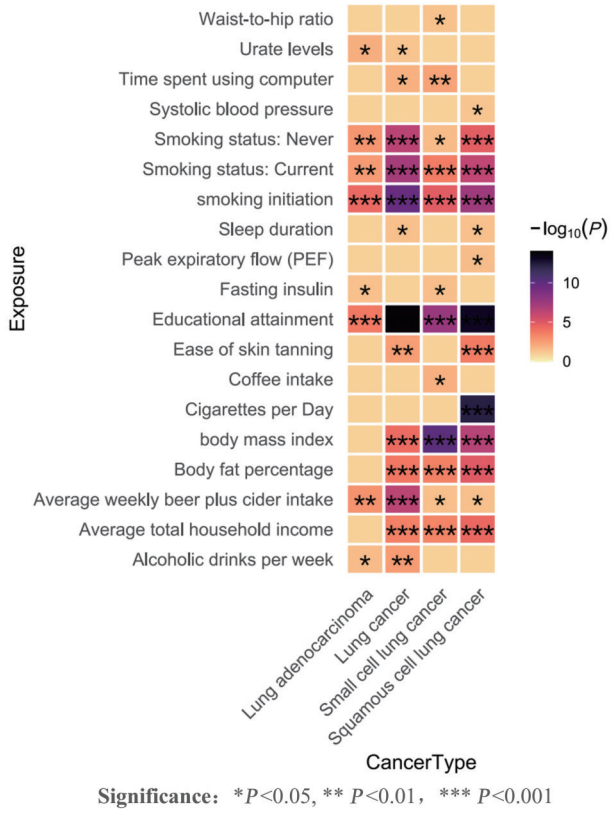
接下来，我们转向小细胞肺癌，这种高度侵袭性的亚型展现出一组具有特异性的环境风险因素。经过Bonferroni多重检验校正后，受教育水平提升（OR=0.82，95%CI：0.77~0.88，P=2.41×10^{-8}）、BMI升高（OR=2.05，95%CI：1.65~2.55，P=7.08×10^{-11}）以及家庭收入增加（OR=0.36，95%CI：0.21~0.64，P=3.95×10^{-4}）仍与小细胞肺癌风险显著相关，提示社会经济因素和肥胖在该亚型的发生中具有关键作用。其余因素如空腹胰岛素水平、腰臀比、计算机使用时间、体力劳动职业和咖啡摄入等也表现出一定的关联趋势，但在多重校正后差异无统计学意义。总体来看，小细胞肺癌的风险模式在一定程度上延续了其他亚型的共性因素。敏感性分析结果支持上述关联的稳健性，未观察到明显的异质性或水平多效性。

2.4 暴露组范围的环境因素与整体肺癌

最后，为了识别跨越不同组织学亚型的共性环境风险因素，我们对整体肺癌进行了分析。经过Bonferroni多重检验校正后，教育水平提升（OR=0.89，95%CI：0.86~0.92，P=1.30×10^{-14}）、BMI升高（OR=1.22，95%CI：1.11~1.34，P=5.22×10^{-5}）和家庭收入增加（OR=0.58，95%CI：0.43~0.78，P=2.92×10^{-4}）仍与肺癌风险显著相关，分别表现出稳健的保护作用或风险增强效应。此外，尿酸水平、计算机使用时间、酒精摄入、氮氧化物暴露和日晒易晒黑等因素也显示出一定关联趋势，但在多重校正后差异无统计学意义。总体而言，较高的社会经济地位为肺癌的广泛保护因素，而肥胖则是贯穿多个亚型的重要风险因素。敏感性分析支持这些结果的稳健性，未发现明显的异质性或水平多效性。

2.5 环境因素与肺癌及其不同亚型的共享且相互独立的因果关联

总体而言，肺癌及其亚型的因果关联呈现共享性与特异性并存的特征（见图4）。具体来说，受教育程度、吸烟量和BMI是肺腺癌、肺鳞癌、小细胞肺癌及整体肺癌共同关联的共享因素。此外，家庭平均收入与肺鳞癌、小细胞肺癌和整体肺癌存在因果关联；酒精摄入、空腹胰岛素及尿酸水平在肺腺癌与总体肺癌中共享关联；睡眠时长、皮肤易晒黑程度分别在肺鳞癌与总体肺癌中重叠；计算机使用时间和体力工作暴露则与小细胞肺癌及整体肺癌相关。部分环境因素表现出亚型特异性独立关联，例如咖啡摄入仅与小细胞肺癌、噪声暴露与肺腺癌、空气污染物与总体肺癌等具有亚型特异性独立关联，提示肺癌不同亚型既存在共同风险驱动因素，也受到特定环境暴露因素的差异化影响。



Significance：*P<0.05, ** P<0.01，*** P<0.001

颜色梯度表示经负对数转换后的P值（$-\log_{10}P$），色度越深表明统计学显著性越强，即关联关系越显著。星号标记表示不同的显著性等级。

图4 暴露组范围的环境因素与各种肺癌亚型之间的因果效应

Figure 4 The causal effects of exposome-wide environmental factors on different lung cancer subtypes

2.6 跨阶段验证揭示关键环境因素与肺癌亚型的稳定因果关系

在验证阶段的分析中，我们总体上证实了发现阶段所识别的主要环境暴露因素与肺癌及其亚型之间的因果关联（见图5）。综合验证结果，共识别出11对具有稳健因果关系的暴露–结局对，涉及9种环境暴露因素。具体而言，受教育水平在肺腺癌、肺鳞癌以及整体肺癌中均表现出显著的保护效应；BMI在肺鳞癌与整体肺癌中均显示出正向关联，进一步支持其作为肺癌危险因素的作用；每日吸烟量与日晒易黑倾向在肺鳞癌中得到验证；此外，每周酒精摄入量、睡眠时长、家庭总收入以及计算机使用时间与整体肺癌的因果关联亦在验证阶段得到复制。值得注意的是，验证阶段还识别出若干潜在的新风险因素，如运动强度与吸烟起始年龄等，提示这些变量可能在肺癌发生发展中发挥作用。总体而言，跨阶段验证显著增强了发现阶段结论的可靠性，提升了因果推断的稳健性，并为肺癌防控提供了更为坚实的科学依据。

2.7 环境因素与吸烟状态分层分析与MVMR分析结果

在整合发现阶段与验证组的单变量MR结果后，共识别出9个与肺癌显著相关的环境/暴露因素，其中包括2个与吸烟行为密切相关的变量。为减少吸烟带来的混杂影响，在分层分析中剔除了高度与吸烟行为重合的变量，并采用吸烟者与不吸烟

Exposure	nSNP	*OR* (95% *CI*)	*P*_value	*P*_Bonferroni
Lung adenocarcinoma				
Educational attainment	202	0.93 (0.88 ~ 0.99)	1.70e-02	1.05e+00
smoking initiation	83	1.48 (1.12 ~ 1.96)	6.00e-03	3.72e-01
Alcoholic drinks per week	31	1.94 (1.24 ~ 3.03)	4.00e-03	2.48e-01
Smoking status: Never	70	0.22 (0.07 ~ 0.72)	1.20e-02	7.44e-01
Cigarettes per Day	21	2.49 (1.82 ~ 3.41)	1.25e-08	7.75e-07
Squamous cell lung cancer				
Strenuous sports or other exercises	12	0.05 (0.00 ~ 0.69)	2.60e-02	1.61e+00
Educational attainment	202	0.86 (0.81 ~ 0.91)	1.08e-07	6.70e-06
Cigarettes per Day	21	3.10 (2.08 ~ 4.63)	3.26e-08	2.02e-06
Body mass index	488	1.53 (1.29 ~ 1.81)	1.06e-06	6.57e-05
smoking initiation	83	1.76 (1.32 ~ 2.35)	1.35e-04	8.37e-03
Ease of skin tanning	110	0.85 (0.74 ~ 0.98)	2.17e-02	1.35e+00
Average total household income	43	0.41 (0.25 ~ 0.69)	1.00e-03	6.20e-02
Body fat percentage	368	1.33 (1.02 ~ 1.73)	3.20e-02	1.98e+00
Smoking status: Never	70	0.15 (0.05 ~ 0.42)	3.57e-04	2.21e-02
Small cell Lung cancer				
Serum 25-Hydroxyvitamin D levels	113	0.35 (0.15 ~ 0.83)	1.80e-02	1.12e+00
Urate levels	297	0.63 (0.41 ~ 0.98)	4.00e-02	2.48e+00
Cigarettes per Day	21	4.09 (1.26 ~ 13.29)	1.90e-02	1.18e+00
Average weekly beer plus cider intake	18	414.18 (3.66 ~ 46815.22)	1.25e-02	7.75e-01
Lung cancer				
Strenuous sports or other exercises	12	0.13 (0.02 ~ 0.79)	2.69e-02	1.67e+00
Educational attainment	202	0.88 (0.85 ~ 0.92)	1.38e-10	8.56e-09
body mass index	488	1.23 (1.09 ~ 1.39)	1.00e-03	6.20e-02
smoking initiation	83	1.61 (1.34 ~ 1.94)	3.15e-07	1.95e-05
Alcoholic drinks per week	31	1.52 (1.11 ~ 2.08)	9.00e-03	5.58e-01
Smoking status: Current	14	54.10 (6.02 ~ 485.92)	3.66e-04	2.27e-02
Sleep duration	64	0.45 (0.26 ~ 0.77)	3.00e-03	1.86e-01
Time spent using computer	81	0.64 (0.43 ~ 0.96)	3.18e-02	1.97e+00
Average weekly beer plus cider intake	19	3.11 (1.57 ~ 6.15)	1.15e-03	7.13e-02
Average total household income	43	0.44 (0.30 ~ 0.66)	6.18e-05	3.83e-03
Body fat percentage	367	1.26 (1.06 ~ 1.50)	1.00e-02	6.20e-01
Smoking status: Never	70	0.18 (0.09 ~ 0.36)	1.12e-06	6.94e-05
Cigarettes per Day	21	2.63 (1.88 ~ 3.68)	1.41e-08	8.74e-07

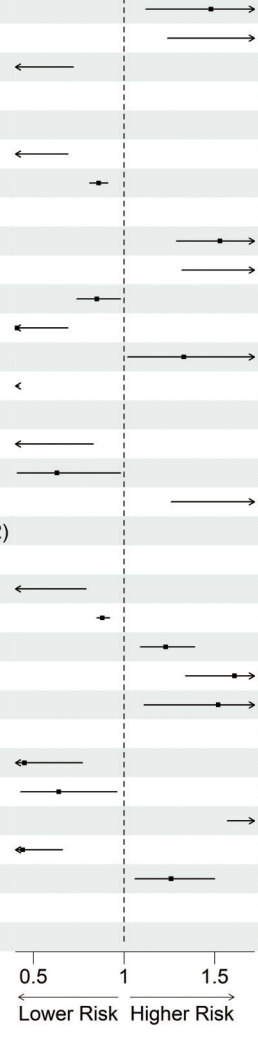
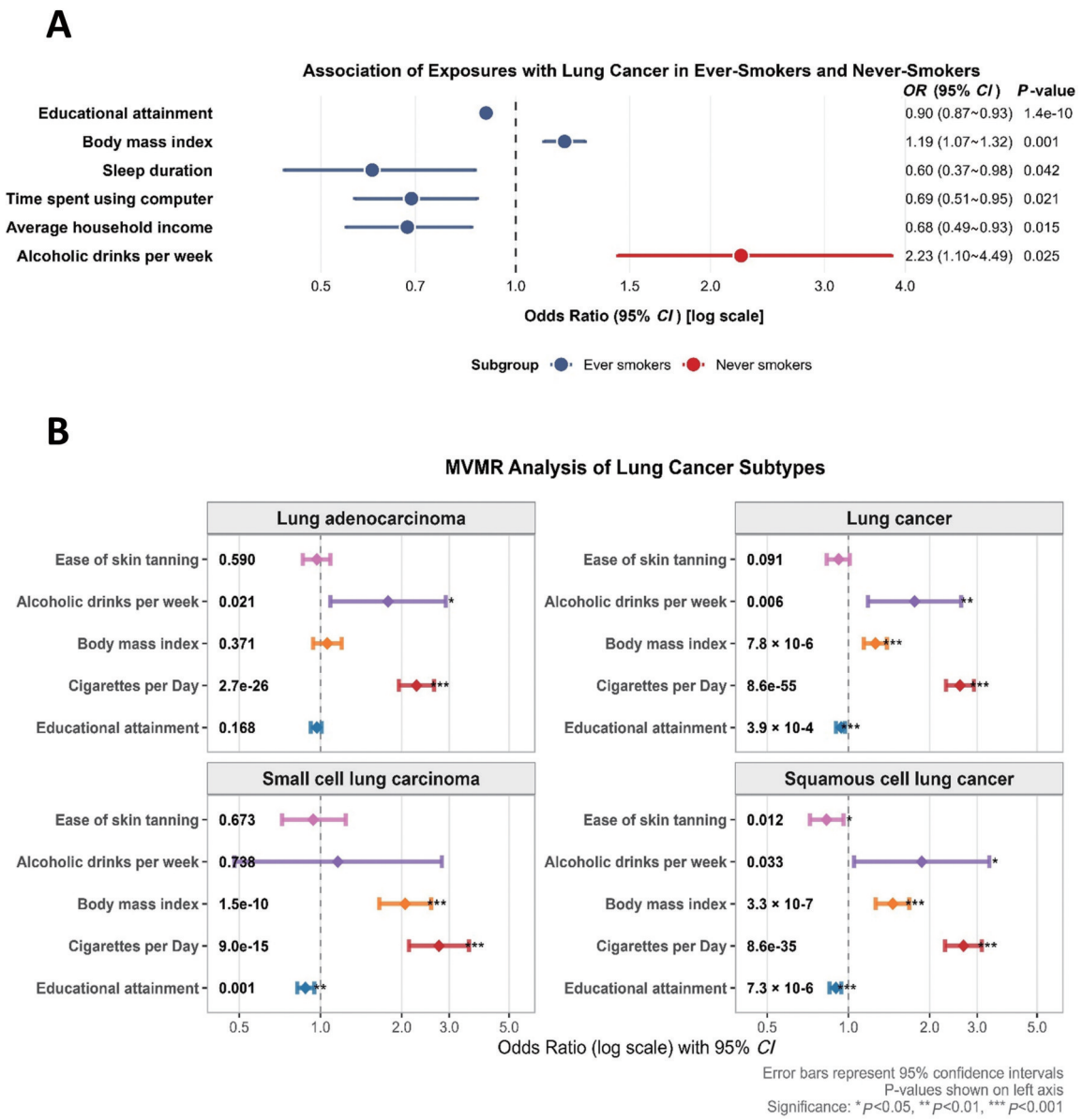


图5 环境因素与各种肺癌亚型的MR分析结果森林图（验证阶段）

Figure 5 Forest plot of MR analysis results showing associations between environmental factors and different lung cancer subtypes (validation stage)

者的肺癌GWAS数据分别进行了亚组MR分析（见图6A）。分析结果显示，部分暴露因素在不同吸烟状态下的因果效应存在异质性。在吸烟者中，教育程度、睡眠时长、计算机使用时间、家庭收入等因素与肺癌风险呈显著负相关关系，而BMI升高则与肺癌风险增加相关。相对地，在非吸烟者中，仅每周饮酒量与肺癌风险升高显著相关。

多变量MR在控制多种暴露因素间潜在混杂的基础上，评估其对肺癌及不同亚型的独立因果效应。结果显示（见图6B），每日吸烟量在所有肺癌类型中均具有较强的正向因果关联，提示其为最主要的独立风险因素。相比之下，教育程度与小细胞肺癌、鳞状细胞肺癌及整体肺癌风险呈显著负相关关系，表明其可能具有独立的保护作用。BMI升高则与非腺癌型肺癌风险增加相关。值得注意的是，尽管吸烟和BMI为主要风险因素，每周饮酒量与总肺癌风险呈负相关关系，日光易晒黑倾向也仅与鳞状细胞肺癌风险呈显著降低相关。以上结果表明，在控制混杂因素后，不同暴露因素对肺癌各亚型存在差异化的独立因果效应。



A：曾经吸烟者与从不吸烟者中遗传预测暴露因素与肺癌风险的关联；B：多变量孟德尔随机化分析揭示各因素在肺腺癌、鳞状细胞癌、小细胞癌和整体肺癌中的特异性效应。

图6　环境暴露因素与肺癌风险的亚组和多变量MR分析结果

Figure 6　Subgroup and multivariable Mendelian randomization analyses showing association between environmental exposures and lung cancer risk

3 讨论

本研究基于大规模的暴露组范围MR分析，系统评估了62种环境因素与肺癌及其主要亚型（包括肺腺癌、肺鳞癌、小细胞肺癌）之间的因果关系。在发现阶段，共识别出35组显著的暴露—结局关联，验证了吸烟、BMI、受教育水平等已知风险因素的广泛影响。随后结合验证阶段结果确认了其中11组稳健的因果关联，涵盖9种环境因素，具体为受教育水平在肺腺癌、肺鳞癌及整体肺癌中均表现出显著的保护性作用；BMI在肺鳞癌及整体肺癌中呈正向关联，进一步巩固了其作为肺癌危险因素的证据；每日吸烟量显著增加肺鳞癌风险，而日光易晒黑倾向则显示出对肺鳞癌的潜在保护作用。此外，每周酒精摄入量与整体肺癌风险呈正相关关系，属于危险因素；相反，睡眠时长、家庭总收入及计算机使用时间均表现出保护性作用，且与整体肺癌具有显著因果关联。进一步的吸烟状态分层分析表明，部分暴露因素的因果效应在吸烟者与非吸烟者中存在异质性；多变量MR分析则明确了每日吸烟量、BMI等因素对肺癌及其不同亚型的独立因果作用。本研究不仅揭示了肺癌亚型间环境因素的共享机制，还识别出若干亚型特异的风险谱系，为肺癌的病因学理解及精准防控提供了新的科学依据。

值得注意的是，我们在环境暴露因素分析中发现了具有显著亚型特异性关联的新证据，这些发现可能为肺癌异质性机制研究提供新方向。其中，皮肤日光易晒黑倾向与肺鳞癌的因果关联在本研究中首次被揭示，前瞻性研究表明适度紫外线暴露是肺鳞癌的保护因素[18]，且其作用具有一定的滞后效应[19]。这一关联可能通过多种生物学机制共同驱动。首先，阳光暴露可促进皮肤维生素D合成，其活性形式1,25-二羟维生素D_3具有抗增殖、促分化和免疫调节作用，有助于抑制肺癌的发生与发展[20]。其次，紫外线辐射还可能通过影响皮肤免疫细胞（如朗格汉斯细胞）功能和炎症因子的释放，间接调节全身免疫状态，增强机体对肿瘤的免疫监视能力[21]。第三，适度的紫外线暴露可能诱导抗氧化应激反应，从而维持氧化应激与抗氧化防御之间的动态平衡；当暴露水平适中时，有助于抵御DNA损伤和降低肿瘤发生的风险[22]。我们推测，“日光易晒黑倾向”可能代表个体对紫外线刺激的一种适度适应，或是反映了更规律、温和的光照习惯，而非剧烈晒伤，从而呈现出一定的保护效应。此外，本研究还发现了噪声与肺癌的潜在关联，尽管既往研究已证实其与心血管疾病的正相关关系[23-24]，且噪声通过激活下丘脑–垂体–肾上腺轴（HPA轴）[25]，促进皮质醇释放及慢性炎症[26]，影响免疫微环境而促进癌症的发生。睡眠时长延长对肺癌的保护作用可能与褪黑素抑制NF-κB通路相关[27-28]，提示调整作息是潜在的预防策略。此外，还有与通常认知不同的结果，收缩压升高本身可能并非保护性因素，而高血压患者的长期用药（如ACE抑制剂/ARB类药物）可能通过抑制血管紧张素Ⅱ介导的炎症反应和血管生成[29]，间接降低肺癌风险。

本研究还证实了受教育水平[30-31]、BMI升高[32-34]及吸烟[35-36]等共性风险因素在肺癌各亚型中的跨亚型显著性关联，与既往流行病学研究结论高度一致。进一步对这3种跨亚型共性风险因素的因果效应强度进行了比较，结果显示，吸烟对各亚型均具有较强的正向因果效应，其次为BMI，教育水平的影响相对较弱。上述排序提示，在公共卫生资源分配中，应优先聚焦控烟干预，其次强化体重管理与教育公平，以最大程度降低整体肺癌负担。此外，环境暴露的剂量–反应关系解析对于揭示肺癌风险动态变化及优化预防策略至关重要。BMI与肺癌风险呈线性正相关关系，确立BMI≥25 kg/m²作为干预临界值，为健康管理和公共卫生政策提供依据。紫外线暴露则表现出“J型”剂量–反应曲线，适度暴露通过维生素D合成和免疫调节发挥保护作用，高剂量或不足暴露均可能增加风险[22]。未来研究应结合动态监测、生物标志物及统计模型，精准量化暴露水平并解析其非线性关系，同时通过机制研究明确关键阈值，推动肺癌预防向精准化、个体化方向发展。

然而，本研究的MR分析在提供有价值因果推论的同时，也存在一些局限性。首先，目前可用的肺癌GWAS数据主要采用二分类设计，限制了对性别、年龄及肺癌不同发展阶段的精细分层分析，例如，由于当前缺乏公开可用的性别分层肺癌GWAS数据，暂无法开展性别特异性的MR分析，未来研究应基于更细致的分层人群数据，系统探讨性别、年龄及肺癌不同发展阶段等变量对环境暴露与肺癌风险关系的异质性影响。此外，由于GWAS数据的可用性限制，本研究无法评估部分在观察性研究中与肺癌相关的环境暴露因素，因缺乏合适的IV，无法全面检测潜在关联，同时也不能排除因样本量不足而未能发现真实因果关系的可能性。尽管存在上述局限性，本研究仍具有显著优

势。首先，MR方法通过减少混杂因素和反向因果关系的影响，提高了暴露因素与肺癌间因果推断的可靠性。其次，本研究并未局限于单一类别的环境因素，而是涵盖多个亚组分析，提供了更全面的因果推断视角。第三，基于最新的GWAS数据和严格且稳健的分析方法，本研究在识别和排除环境暴露对肺癌的潜在因果效应方面取得了重要进展。与以往MR研究可能存在的高估或低估情况，本研究的结果更加严谨可靠，为深入探索环境因素对肺癌及其不同亚型的影响提供了坚实的数据支持，有助于未来制定更精准的肺癌预防策略。

综上所述，本研究系统揭示了暴露组范围内的环境因素与不同肺癌亚型之间的共享且相互独立的因果关联，为肺癌的精准预防提供了重要的理论依据。

参考文献

[1] BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2024, 74(3): 229-263.

[2] ELISIA I, LAM V, CHO B, et al. The effect of smoking on chronic inflammation, immune function and blood cell composition[J]. Scientific Reports, 2020, 10(1): 19480.

[3] SIEGEL R L, KRATZER T B, GIAQUINTO A N, et al. Cancer statistics[J]. CA: A Cancer Journal for Clinicians, 2025, 75(1): 10-45.

[4] KUANG Z Y, WANG J X, LIU K X, et al. Global, regional, and national burden of tracheal, bronchus, and lung cancer and its risk factors from 1990 to 2021: Findings from the global burden of disease study 2021[J]. eClinicalMedicine, 2024, 75: 102804.

[5] MOMEN N C, BAKER M G, DRISCOLL T, et al. The effect of occupational exposure to welding fumes on trachea, bronchus, and lung cancer: A supplementary analysis of regular occupational exposure and of occasional occupational exposure based on the systematic review and meta-analysis from the WHO / ILO Joint estimates of the work-related burden of disease and injury[J]. Environment International, 2025, 196: 109216.

[6] ZHU M, HAN Y T, MOU Y L, et al. Effect of long-term fine particulate matter exposure on lung cancer incidence and mortality in Chinese nonsmokers[J]. American Journal of Respiratory and Critical Care Medicine, 2025, 211(4): 600-609.

[7] BURGESS S, DAVEY SMITH G, DAVIES N M, et al. Guidelines for performing mendelian randomization investigations: Update for summer[J]. Wellcome Open Research, 2023, 4:186.

[8] EMDIN C A, KHERA A V, KATHIRESAN S. Mendelian randomization[J]. The Journal of the American Medical Association, 2017, 318(19): 1925-1926.

[9] LAN G H, XIE M Y, LAN J L, et al. Association and mediation between educational attainment and respiratory diseases: A mendelian randomization study[J]. Respiratory Research, 2024, 25(1):115.

[10] LU J M, LIN Y, JIANG J F, et al. Investigating the potential causal association between consumption of green tea and risk of lung cancer: A study utilizing Mendelian randomization[J]. Frontiers in Nutrition, 2024, 11: 1265878.

[11] WU H, YANG J, WANG H, et al. Mendelian randomization to explore the direct or mediating associations between socioeconomic status and lung cancer[J]. Frontiers in Oncology, 2023, 13: 1143059.

[12] LI D, ZHOU L H, CAO Z, et al. Associations of environmental factors with neurodegeneration: An exposome-wide mendelian randomization investigation[J]. Ageing Research Reviews, 2024, 95: 102254.

[13] CARTER A R, ANDERSON E L. Correct illustration of assumptions in mendelian randomization[J]. International Journal of Epidemiology, 2024, 53(2): dyae050.

[14] PALMER T M, LAWLOR D A, HARBORD R M, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors[J]. Statistical Methods in Medical Research, 2012, 21(3): 223-242.

[15] BURGESS S, BUTTERWORTH A, THOMPSON S G. Mendelian randomization analysis with multiple genetic variants using summarized data[J]. Genetic Epidemiology, 2013, 37(7): 658-665.

[16] BOWDEN J. Misconceptions on the use of MR-Egger regression and the evaluation of the InSIDE assumption [J]. International Journal of Epidemiology, 2017, 46(6): 2097-2099.

[17] COHEN J F, CHALUMEAU M, COHEN R, et al. Cochran's Q test was useful to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy[J]. Journal of Clinical Epidemiology, 2015, 68(3): 299-306.

[18] LIN S W, WHEELER D C, PARK Y, et al. Prospective study of ultraviolet radiation exposure and risk of cancer in the United States[J]. International Journal of Cancer, 2012, 131(6): E1015-E1023.

[19] LU X Z, WANG R C, LI J F, et al. Exposure-lag response of surface net solar radiation on lung cancer incidence: A global time-series analysis[J]. Translational Lung Cancer Research, 2024, 13(10): 2524-2537.

[20] JEON S M, SHIN E A. Exploring vitamin D metabolism and function in cancer[J]. Experimental and Molecular Medicine, 2018, 50(4): 1-14.

[21] MODI B G, NEUSTADTER J, BINDA E, et al. Langerhans cells facilitate epithelial DNA damage and squamous cell carcinoma[J]. Science, 2012, 335(6064): 104-108.

[22] YU Z W, ZHENG M, FAN H Y, et al. Ultraviolet (UV) radiation: A double-edged sword in cancer development and therapy[J]. Molecular Biomedicine, 2024, 5(1): 49.

[23] HANSELL A L, BLANGIARDO M, FORTUNATO L, et al. Aircraft noise and cardiovascular disease near heathrow airport in London: Small area study[J]. British Medical Journal, 2013, 347: f5432.

[24] HUSS A, SPOERRI A, EGGER M, et al. Aircraft noise, air pollution, and mortality from myocardial infarction[J]. Epidemiology (Cambridge, Mass.), 2010, 21(6): 829-836.

[25] LEE S, LEE Y, LEE W, et al. Environmental noise exposure and autoimmune diseases: Evidence from a national cohort study[J]. Ecotoxicology and Environmental Safety, 2025, 289: 117674.

[26] JIANG Q H, ZHANG Y, YE T, et al. Metabonomics and transcriptomics analyses reveal the underlying HPA-axis-related mechanisms of lethality in larimichthys polyactis exposed to underwater noise pollution[J]. International Journal of Molecular Sciences, 2024, 25(23): 12610.

[27] LIU S, YANG T N, WANG Y X, et al. Parkin-TLR4-NLRP3 axis directs melatonin to alleviate atrazine-induced immune impairment in splenic macrophages[J]. Journal of Pineal Research, 2024, 76(8): e70014.

[28] KIM R, KIM M, JEONG S, et al. Melatonin alleviates myocardial dysfunction through inhibition of endothelial-to-mesenchymal transition via the NF - κB pathway[J]. Journal of Pineal Research, 2024, 76(4): e12958.

[29] YARMOLINSKY J, DÍEZ-OBRERO V, RICHARDSON T G, et al. Genetically proxied therapeutic inhibition of antihypertensive drug targets and risk of common cancers: A mendelian randomization analysis[J]. PLoS Medicine, 2022, 19(2): e1003897.

[30] MUGUTSO E, BESSON C, HAMED R, et al. Educational attainment and loss to follow-up in a quebec lung cancer screening pilot program[J]. Annals of the American Thoracic Society, 2023, 20(12): 1819-1822.

[31] SCHWARTZ R M, BEVILACQUA K G, ALPERT N, et al. Educational attainment and quality of life among older adults before a lung cancer diagnosis[J]. Journal of Palliative Medicine, 2020, 23(4): 498-505.

[32] FANG Z, SONG M Y, LEE D H, et al. The role of mendelian randomization studies in deciphering the effect of obesity on cancer[J]. Journal of the National Cancer Institute, 2022, 114(3): 361-371.

[33] VITHAYATHIL M, CARTER P, KAR S, et al. Body size and composition and risk of site-specific cancers in the UK Biobank and large international consortia: a mendelian randomisation study[J]. PLoS Medicine, 2021, 18(7): e1003706.

[34] WEISS J, BERNATZ S, JOHNSON J, et al. Opportunistic assessment of steatotic liver disease in lung cancer screening eligible individuals[J]. Journal of Internal Medicine, 2025, 297(3): 276-288.

[35] DU M L, XIN J Y, ZHENG R, et al. CYP2A6 activity and cigarette consumption interact in smoking-related lung cancer susceptibility[J]. Cancer Research, 2024, 84 (4): 616-625.

[36] LARSSON S C, BURGESS S. Appraising the causal role of smoking in multiple diseases: A systematic review and meta-analysis of Mendelian randomization studies[J]. EBioMedicine, 2022, 82(8): 104154.

（上接第87页）

[11] ELDRIDGE R C, WENTZENSEN N, PFEIFFER R M, et al. Endogenous estradiol and inflammation biomarkers: potential interacting mechanisms of obesity-related disease [J]. Cancer Causes and Control, 2020, 31(4): 309-320.

[12] CAMBIER S, GOUWY M, PROOST P. The chemokines CXCL8 and CXCL12: molecular and functional properties, role in disease and efforts towards pharmacological intervention[J]. Cellular and Molecular Immunology, 2023, 20(3): 217-251.

[13] PIOLI P A, JENSEN A L, WEAVER L K, et al. Estradiol attenuates lipopolysaccharide-induced CXC chemokine ligand 8 production by human peripheral blood monocytes [J]. The Journal of Immunology, 2007, 179(9): 6284-6290.

[14] HAIM K, WEITZENFELD P, MESHEL T, et al. Epidermal growth factor and estrogen act by independent pathways to additively promote the release of the angiogenic chemokine CXCL8 by breast tumor cells[J]. Neoplasia, 2011, 13(3): 230-243.

[15] KOZIEŁ M J, PIASTOWSKA-CIESIELSKA A W. Estrogens, estrogen receptors and tumor microenvironment in ovarian cancer[J]. International Journal of Molecular Sciences, 2023, 24(19): 14673.