Edinburgh Research Explorer

Exploring the cross-cancer effect of smoking and its fingerprints in blood DNA methylation on multiple cancers: a mendelian randomization study

CANCER GENETICS AND EPIGENETICS

IJC INTERNATIONAL JOURNAL of CANCER

Exploring the cross-cancer effect of smoking and its fingerprints in blood DNA methylation on multiple cancers: A Mendelian randomization study

Yajing Zhou[1,2] | Xuan Zhou[2,3] | Jing Sun[2] | Lijuan Wang[2,4] | Jianhui Zhao[2] | Jie Chen[2] | Shuai Yuan[5] | Yazhou He[6] | Maria Timofeeva[7] | Athina Spiliopoulou[3] | Ines Mesa-Eguiagaray[4,8] | Susan M. Farrington[9] | Kefeng Ding[1] | Malcolm G Dunlop[8,9] | Xiao Qian[1] | Evropi Theodoratou[4,8] | Xue Li[2]

[1]Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2]Department of Big Data in Health Science, School of Public Health and The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[3]Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, UK

[4]Centre for Global Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, UK

[5]Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[6]Department of Oncology, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China

[7]Danish Institute for Advanced Study (DIAS), Epidemiology, Biostatistics and Biodemography Research Unit, Institute of Public Health, University of Southern Denmark, Odense, Denmark

[8]Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[9]Colon Cancer Genetics Group, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

Correspondence
Xiao Qian, Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China.
Email: qxiao3@zju.edu.cn

Evropi Theodoratou, Centre for Global Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, UK.
Email: e.theodoratou@ed.ac.uk

Xue Li, Department of Big Data in Health Science, School of Public Health and The Second Affiliated Hospital, Zhejiang University

Abstract

Aberrant smoking-related DNA methylation has been widely investigated as a carcinogenesis mechanism, but whether the cross-cancer epigenetic pathways exist remains unclear. We conducted two-sample Mendelian randomization (MR) analyses respectively on smoking behaviors (age of smoking initiation, smoking initiation, smoking cessation, and lifetime smoking index [LSI]) and smoking-related DNA methylation to investigate their effect on 15 site-specific cancers, based on a genome-wide association study (GWAS) of 1.2 million European individuals and an epigenome-WAS (EWAS) of 5907 blood samples of Europeans for smoking and 15 GWASs of European ancestry for multiple site-specific cancers. Significantly identified CpG sites were further used

School of Medicine, Hangzhou, China.
Email: xueli157@zju.edu.cn

for colocalization analysis, and those with cross-cancer effect were validated by overlapping with tissue-specific eQTLs. In the genomic MR, smoking measurements of smoking initiation, smoking cessation and LSI were suggested to be casually associated with risk of seven types of site-specific cancers, among which cancers at lung, cervix and colorectum were provided with strong evidence. In the epigenetic MR, methylation at 75 CpG sites were reported to be significantly associated with increased risks of multiple cancers. Eight out of 75 CpG sites were observed with cross-cancer effect, among which cg06639488 (EFNA1), cg12101586 (CYP1A1) and cg14142171 (HLA-L) were validated by eQTLs at specific cancer sites, and cg07932199 (ATXN2) had strong evidence to be associated with cancers of lung (coefficient, 0.65, 95% confidence interval [CI], 0.31-1.00), colorectum (0.90 [0.61, 1.18]), breast (0.31 [0.20, 0.43]) and endometrium (0.98 [0.68, 1.27]). These findings highlight the potential practices targeting DNA methylation-involved cross-cancer pathways.

KEYWORDS
cross-cancer effect, DNA methylation, Mendelian randomization, methylation quantitative trait loci, smoking

What's new?
Despite that smoking has been widely considered as a risk factor for multiple cancers, the smoking-related cross-cancer epigenetic pathway remains underexplored. In our study, the investigation of the genetic and epigenetic association between smoking and 15 site-specific cancers revealed the mediating effect of smoking-related aberrant methylation of CpG sites of cg06639488 (EFNA1), cg12101586 (CYP1A1), cg14142171 (HLA-L) and cg07932199 (ATXN2) on increased risks of lung, colorectal, breast, cervix cancers, highlighting underlying epigenetic pathways of carcinogenesis.

1 | INTRODUCTION

Smoking has been widely recognized as a risk factor for numerous diseases including cancer. Observational And experimental studies have confirmed the causal relationship between smoking and the risk of lung cancer,[1,2] and other common cancers like breast cancer,[3] prostate cancer,[4] ovarian cancer and cervix cancer have also been deemed as potentially consequential events associated with smoking.[5]

Studies have found aberrant DNA methylation can be induced by smoking behaviors,[6] which is considered as a potential mechanism to trigger carcinogenesis. Also, hypotheses of DNA methylation-related epigenetic modification have long been investigated and widely elaborated as one of the mechanisms of carcinogenesis.[7] DNA methylation is classically characterized by the process of forming the 5-methylcytosine in the C5 position of cytosine-phosphate-guanine (CpG) dinucleotides. This is prone to obstacle the combination of transcription complex and DNA and cause nonprogrammed alteration in downstream gene expression,[8] such as hypomethylation in activating proto-oncogenes and hypermethylation in the silencing of tumor-suppressor genes in the promoter region[9] in carcinoma tissues. A number of studies have investigated the association of smoking and single site-specific cancer via differentiated methylation level[10-12] suggesting the key role of methylation in the process of carcinogenesis. However, whether the epigenetic effect is exerted universally across multiple cancers remains unknown, and whether the mechanisms of methylation are shared through some specific CpG sites or common pathways is worth exploring.

Mendelian randomization analysis is a method that uses genetic variants, for example, single nucleotide polymorphisms (SNPs) or quantitative trait loci (QTLs) as proxies for risk factors of interest to explore the causality between exposure and disease.[13] This minimizes the unmeasured confounding effects and diminishes reverse causality. In our study, we sequentially conducted two two-sample MR analyses using instrumental variants (IVs) of SNPs as genetic proxies and methylation QTLs (mQTLs) as epigenetic proxies of methylation at CpG sites to explore the causal effect of smoking on the risk of multiple site-specific cancers. We further assessed the cross-cancer effects of smoking-related blood methylation and validated the tissue-specificity with expression-QTLs (eQTLs).

2 | MATERIALS AND METHODS

2.1 | Study design

In our study, we firstly conducted a two-sample Mendelian randomization (MR) analysis to investigate the causal effect of smoking on
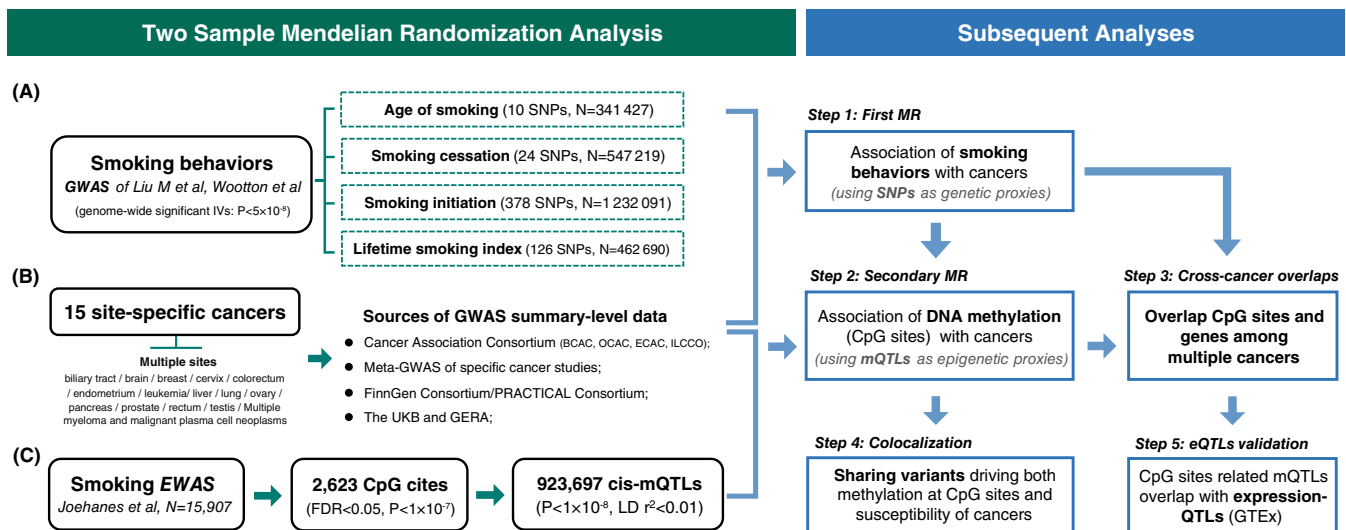
FIGURE 1 Schematic diagram of the study design. EWAS, epigenome-wide association study; GWAS, genome-wide association study; mQTLs, methylation quantitative trait loci; MR, Mendelian randomization; SNP, single nucleotide polymorphism. [Color figure can be viewed at wileyonlinelibrary.com]

genetic predisposition to 15 site-specific cancers, in which we chose three phenotypes of smoking behaviors (age of initiation, smoking initiation and smoking cessation) and an aggregative lifetime smoking index (LSI) as specific measurements of smoking. Then we performed a second two-sample MR analysis to reveal the aforementioned causality on the epigenome-wide level, using mQTLs as IVs for smoking-related blood DNA methylation at CpG sites, and further focused on those that have a cross-cancer effect and are validated with tissue-specific eQTLs. For the significant CpG cites, we also conducted a colocalization analysis to investigate the effect of sharing variants both on DNA methylation and susceptibility of cancers (Figure 1).

2.2 | Data sources

2.2.1 | Genome-wide association study summary-level data of smoking behaviors

The IVs for four smoking behaviors were extracted from two genome-wide association studies (GWASs) separately. First, SNPs associated with age of smoking initiation (10 associated variants, $N = 341\,427$), smoking initiation (378 variants, $N = 1\,232\,091$), smoking cessation (24 variants, $N = 547\,219$) at the significant threshold ($P < 5 \times 10^{-8}$) as genetic instruments were obtained from a published GWAS that identified variants associated with different aspects of smoking (initiation, cessation and heaviness) from a total of 1 232 091 individuals of European ancestry.[14] In our study, smoking initiation and cessation are both binary phenotypes comparing individuals' smoking status, with current or previous smokers coded as "2" and never smokers coded as "1" for smoking initiation, and current smokers coded as "2" and previous smokers coded as "1" for smoking cessation. Then we derived genetic variants for LSI, an aggregative indicator of smoking, from a GWAS involving 462 690 individuals of European ancestry (126 independent, genome-wide significant SNPs).[15] Linkage

disequilibrium (LD) was calculated based on 1000 Genomes European reference panel, and only genetic variants without LD ($r^2 \le 0.01$ and clump window >10 000 kb) were selected (Tables S1 and S2).

2.2.2 | Epigenome-wide data of smoking-related DNA methylation

The information of the association between smoking and DNA methylation (ie, smoking-related CpG sites) was derived from a genome-wide meta-analysis measuring 5907 blood-derived DNA samples from participants in 16 cohorts of the Cohorts for Heart and Aging Research in the Genetic Epidemiology Consortium.[6] A total of 2623 CpG sites with differentiated methylation was identified between current smokers and never smokers (false discovery rate [FDR] < 0.05, $P < 1 \times 10^{-7}$). We then obtained CpG-associated mQTLs as genetic proxies from Genetics of DNA Methylation Consortium (GoDMC) (http://mqtldb.godmc.org.uk/), a mQTL database containing genetic and methylation data from over 30 000 participants.[16] DNA methylation data were quantified from bisulfite-converted genomic blood DNA using Illumina Infinium HumanMethylation450 (HM450) BeadChip.

2.2.3 | GWAS summary-level data of 15 site-specific cancers

We obtained summary-level data of 15 site-specific cancers from publicly available datasets (Table S1): (1) endometrial cancer GWAS data were acquired from a meta-GWAS of 17 cohort studies (via the Endometrial Cancer Association Consortium [ECAC], the Epidemiology of Endometrial Cancer Consortium [E2C2] and the UK Biobank) with 12 906 cancer cases and 108 979 country-matched controls[17]; (2) colorectal cancer (CRC) GWAS data were acquired from a meta-

GWAS of 11 cohort studies of colorectal cancer with 16 871 cases and 26 328 controls[18]; (3) prostate cancer GWAS data were obtained from a meta-analysis of 8 GWAS from PRACTICAL Consortium with 79 148 prostate cancer cases and 61 112 controls[19]; (4) GWAS of breast cancer (122 977 cases vs 105 974 controls) were obtained from the Breast Cancer Association Consortium (BCAC)[20]; (5) ovarian cancer (25 509 cases vs 40 941 controls) were derived from the Ovarian Cancer Association Consortium (OCAC)[21]; (6) lung cancer (11 348 cases vs 15 861 controls) were acquired from the International Lung Cancer Consortium (ILCCO)[22]; (7) GWAS data of liver cancer (304 cases vs 218 488 controls), brain cancer (464 cases vs 218 328 controls), biliary cancer (109 cases vs 174 006 controls), leukemia (663 cases vs 218 129 controls), testis cancer (199 cases vs 95 014 controls) and multiple myeloma and malignant plasma cell neoplasms (598 cases vs 218 194 controls) were acquired from the Finngen Consortium[23]; (8) GWAS of cervix cancer (6563 cases vs 410 350 controls) and pancreatic cancer (663 cases vs 410 350 controls) were obtained from The UK Biobank[24] and GERA data.[25] All acquired data were constricted to European ancestry.

2.3 | Two-sample MR

We successively performed two two-sample MR analyses on the levels of genome and epigenome. In our first MR analysis, genetic variants (SNPs) identified for four measurements of smoking were employed as exposure to investigate its causal effect on the risk of multiple cancers. In the second phase, we identified mQTLs as epigenetic proxies for methylation at CpG cites associated with smoking. The effect allele of each mQTL was unified to be in the same direction with the effect of smoking on DNA methylation. We used Wald ratio to estimate the association when exposure had only one SNP for proxy, and inverse-variance weighted (IVW) method with random-effects to measure the combined effect for each exposure as the main method. Sensitivity analyses were additionally applied to improve the robustness of the results. The MR Egger regression and the intercept test were utilized to detect and correct for horizontal pleiotropy.[26] The Weighted Median method was used to provide consistent estimates when valid IVs weighed more than 50%.[27] The MR-PRESSO method was also employed to detect horizontal pleiotropy (global test), correct outliers by removing them (outlier test) and assess its distortion significance (distortion test).[28] The Cochrane's Q value was used to evaluate the heterogeneity of genetic variants (Q < 0.05). F-statistics were calculated to measure the strength of instruments (F < 10 was considered to be a weak instrument).[29] The beta coefficient was calculated per SD for each genetic instrument, and odds ratios (ORs) with 95% confidence interval (CI) were scaled to per one SD increase in genetically predicted smoking and one-unit increase in the log OR of liability to multiple cancers. False discovery rate was computed for multiple-testing (FDR < 0.05). MR analyses were performed by using the "TwoSampleMR" R package.

For the significant CpG sites among multiple cancers in the second MR analysis, we further identified those that had cross-cancer associations. The mQTLs of CpG sites with cross-cancer effect were then obtained and searched in the Genotype-Tissue Expression (GTEx) resource[30] to further investigate their expression effect as eQTLs in the cancer-associated tissues. The significance threshold of expression evidence was set by both P-values (eQTL effect size) and m-values (existence of eQTL-effect in the specific tissue in the cross-tissue meta-analysis),[31,32] and mQTLs that met P-value <.05/(the number of SNP-gene pairs) after Bonferroni correction and m-value >0.9 were indicated to have a statistically significant eQTL effect.

2.4 | Colocalization analysis

Among the CpG sites that were significantly associated with risk of multiple cancers, we additionally performed a colocalization analysis to investigate whether the susceptibility to site-specific cancers was driven by the same variants influencing methylation at the CpG sites. Observation of 75% or higher posterior probability of association (PPA) for both the summary effect of the CpG site and the single effect of a mQTL were deemed as evidence of colocalization. GWAS data for cancer and EWAS data for smoking-methylation (with mQTLs as proxies) were the same as those used in the MR analysis. The colocalization analysis was performed by the "coloc" R package.[33] All analyses were undertaken with R Software 4.2.1.

3 | RESULTS

3.1 | MR analysis of smoking behaviors and multiple cancers

Genetic variants for three smoking behaviors (age of smoking, smoking initiation and smoking cessation) and LSI are shown in Table S2, and the F-statistic for each IV was above 10, suggesting there was no substantial weak instrument bias.

Seven types of site-specific cancers were found significantly associated with three out of four measurements of smoking behaviors (except for age of smoking) utilizing the IVW method (Figure 2). The risk of lung cancer was indicated to be strongly affected by all three measurements of smoking: smoking initiation (OR, 1.88, 95%CI, 1.64-2.16, FDR = 3.85×10^{-18}), smoking cessation (OR, 5.86, 95% CI, 1.86-14.45, FDR = 0.014) and LSI (OR, 4.21, 95%CI, 2.91-6.09, FDR = 7.56×10^{-13}). Increased risk of CRC was also reported by smoking initiation (OR, 1.25, 95%CI, 1.12-1.40, FDR = 8.48×10^{-4}), LSI (OR, 1.33, 95%CI, 1.07-1.66, P = .011) and smoking cessation (OR, 0.66, 95%CI, 0.51-0.86, FDR = 0.015). Cervix cancer was also suggested to be causally affected by smoking initiation (OR, 1.47, 95%CI, 1.29-1.68, FDR = 1.57×10^{-7}) and LSI (OR, 1.72, 95%CI, 1.30-2.28, FDR = 0.002). Other cancers at specific sites of liver, endometrium, pancreas and ovary were also indicated to be consequently associated with smoking initiation or LSI. MR-Egger and the intercept test found horizontal pleiotropy in the analyses of some smoking behaviors with cancers (eg, smoking initiation with CRC and

Cancers	Smoking Behaviors		OR (95% CI)	P value
Endometrial Cancer				
	LSI		1.31 (1.06, 1.62)	0.013
Cervix Cancer				
	SmkInit		1.47 (1.29, 1.68)	7.83E−09
	LSI		1.72 (1.30, 2.28)	1.71E−04
Colorectal Cancer				
	SmkInit		1.25 (1.12, 1.40)	5.65E−05
	SmkCes		0.66 (0.51, 0.86)	0.002
	LSI		1.33 (1.07, 1.66)	0.011
Liver Cancer				
	SmkInit		2.66 (1.55, 4.58)	4.10E−04
Lung Cancer				
	SmkInit		1.88 (1.64, 2.16)	6.42E−20
	SmkCes		5.18 (1.86, 14.45)	0.002
	LSI		4.21 (2.91, 6.09)	2.52E−14
Ovarian Cancer				
	LSI		1.24 (1.03, 1.50)	0.024
Pancreatic Cancer				
	LSI		3.19 (1.43, 7.11)	0.004

OR (95%CI): 0.50 1.0 3.0 5.0 7.0 15.0

FIGURE 2   Forest plot of genetically predicted significant associations observed with smoking behaviors and multiple cancers. CI, confidence interval; LSI, lifetime smoking index; OR, odds ratio; SmkCes, smoking cessation; SmkInit, smoking initiation. [Color figure can be viewed at wileyonlinelibrary.com]

lung cancer), and the MR-PRESSO method indicated some outliers in the analyses (Table S3).

3.2 | MR analysis of smoking-related DNA methylation and multiple cancers

After FDR correction, 12 out of 15 specific cancers were observed to be associated with smoking-related DNA methylation with more than one mQTLs for proxies (Table 1), among which breast cancer (33 CpG sites) and prostate cancer (27 CpG sites) ranked the highest in terms of the number of CpG sites (Table S4).

A total of 75 CpG sites was found to have significant effect on the risk of 10 site-specific cancers, among which 8 CpG sites were observed to have cross-cancer effect: cg02405476 (*UBE2C*), cg06639488 (*EFNA1*), cg07932199 (*ATXN2*), cg11152384 (*RP11-554A11.8*), cg12101586 (*CYP1A1*), cg14142171 (*HLA-L*), cg22533573 (*WT1*) and cg25727671 (*HOXA7*; Table 2 and Figure S1). Most of the CpG sites have a constricted influence on the onset risk with only two types of cancer, whereas cg07932199 was associated with four cancers: breast cancer (MR estimate, 0.31, 95%CI, 0.20-0.43), colorectal cancer (MR estimate, 0.90, 95%CI, 0.61-1.18), endometrial cancer (MR estimate, 0.98, 95%CI, 0.68-1.27) and lung cancer (MR estimate, 0.65, 95%CI, 0.31-1.00). After searching the GTEx project, 4 mQTLs of 3 CpG sites with cross-cancer effect were provided with evidence of functioning as eQTLs (*P* < .05) of the same specific cancer tissues as reported in the second MR analysis.

TABLE 1   Overview of CpG sites found in the second Mendelian randomization of smoking-related DNA methylation and cancers.

	Number of CpG sites		
Category of cancer	With available mQTLs	*P* < .05	FDR < 0.05
Biliary cancer	1956	73	2
Brain cancer	1956	109	1
Breast cancer	1950	223	57
Cervix cancer	1954	110	6
Colorectal cancer	1956	157	6
Endometrial cancer	1954	140	9
Leukemia	1956	96	0
Liver cancer	1956	90	2
Lung cancer	1938	130	10
Multiple myeloma and malignant plasma cell neoplasms	1956	101	0
Ovarian cancer	1872	148	7
Pancreatic cancer	1954	78	0
Prostate cancer	1415	199	52
Rectum cancer	1955	118	1
Testis cancer	1956	109	2

Abbreviations: FDR, false discovery rate; mQTLs, methylation quantitative trait loci.

TABLE 2 Information of overlapped CpG cites among cancers.

				Adjusted associations between CpG sites (mQTLs as proxies) with cancers				
CpG cites	Chr	Position	Gene-symbol	Cancer	SNP	Beta (95%CI) (adjusted)	P-value	FDR
cg02405476	20	44 441 818	UBE2C	Breast cancer	2	0.05 (0.03, 0.08)	2.41E−04	0.018
cg02405476	20	44 441 818	UBE2C	Prostate cancer	2	0.07 (0.03, 0.10)	4.35E−04	0.014
cg06639488	1	155 103 222	EFNA1	Breast cancer	1	0.20 (0.12, 0.29)	2.48E−06	0.001
cg06639488	1	155 103 222	EFNA1	Lung cancer	1	0.52 (0.24, 0.79)	2.23E−04	0.018
cg07932199	12	112 008 034	ATXN2	Breast cancer	1	0.31 (0.20, 0.43)	1.52E−07	6.38E−05
cg07932199	12	112 008 034	ATXN2	Colorectal cancer	1	0.90 (0.61, 1.18)	7.01E−10	1.37E−06
cg07932199	12	112 008 034	ATXN2	Endometrial cancer	1	0.98 (0.68, 1.27)	1.27E−10	5.39E−08
cg07932199	12	112 008 034	ATXN2	Lung cancer	1	0.65 (0.31, 1.00)	1.95E−04	0.030
cg11152384	11	68 934 300	RP11-554A11.8	Breast cancer	1	0.24 (0.12, 0.35)	5.66E−05	0.013
cg11152384	11	68 934 300	RP11-554A11.8	Prostate cancer	1	0.95 (0.76, 1.14)	3.86E−22	9.15E−20
cg12101586	15	75 019 203	CYP1A1	Breast cancer	1	0.15 (0.09, 0.22)	6.66E−06	0.001
cg12101586	15	75 019 203	CYP1A1	Prostate cancer	1	0.14 (0.05, 0.23)	2.00E−03	0.044
cg14142171	6	30 228 101	HLA-L	Breast cancer	5	0.06 (0.03, 0.09)	1.67E−04	0.014
cg14142171	6	30 228 101	HLA-L	Cervix cancer	5	0.12 (0.05, 0.18)	5.63E−04	0.030
cg22533573	11	32 452 771	WT1	Ovarian cancer	2	0.39 (0.19, 0.59)	1.45E−04	0.023
cg22533573	11	32 452 771	WT1	Prostate cancer	2	0.23 (0.11, 0.35)	2.01E−04	0.013
cg25727671	7	27 193 351	HOXA7	Prostate cancer	1	0.23 (0.10, 0.37)	5.34E−04	0.024
cg25727671	7	27 193 351	HOXA7	Breast cancer	2	0.13 (0.06, 0.20)	3.35E−04	0.036

Abbreviations: Chr, chromosome; CI, confidence interval; FDR, false discovery rate; SNP, single nucleotide polymorphism.

rs9330263 (cg06639488) was found acting as an eQTL of lung tissue ($P = 2.10 \times 10^{-5}$), and rs11264328 (cg06639488; $P = 6.00 \times 10^{-3}$), rs2472299 (cg12101586; $P = 4.60 \times 10^{-5}$) and rs1611463 (cg14142171; $P = 5.00 \times 10^6$) were all indicated as eQTLs of breast tissue (Table S5).

3.3 | Colocalization analysis

Among the 75 CpG sites, 15 were observed with evidence of colocalization within 6 site-specific cancers, that is, breast cancer, colorectal cancer, endometrial cancer, liver cancer, lung cancer and prostate cancer. Especially, the methylation at CpG sites cg07932199 (ATXN2) was suggested with nearly 100% posterior probability of sharing causal variants (rs7310615 and rs3184504) with the susceptibility to four cancers: breast cancer, CRC, endometrial cancer and lung cancer. Six CpG sites (cg07932199, cg01765406, cg22561727, cg00867472, cg12593793 and cg26146569) only had less than 30 mQTLs available so they were excluded out of the colocalization analysis (Table S6 and Figures S2–S9).

4 | DISCUSSION

In our study, we sequentially performed two two-sample MR analyses investigating the association between smoking and multiple cancers on the genome- and epigenome-wide level, and further validated the results by tissue-specific eQTLs and colocalization analysis. The first MR suggested that smoking measurements of smoking initiation, smoking cessation and LSI were causally associated with the risk of seven site-specific cancers, in which lung, colorectal and cervix cancer were strongly indicated. The second MR revealed the effect of blood DNA methylation at CpG sites on 12 cancers, in which 8 CpG sites were observed to have cross-cancer effects. mQTLs of CpG sites cg06639488 (EFNA1), cg12101586 (CYP1A1) and cg14142171 (HLA-L) was validated by eQTLs at specific cancer tissues, and cg07932199 (ATXN2) provided colocalization evidence of both methylation and susceptibility to multiple cancers.

Our first MR analysis implied that smoking behaviors were causally associated with cancers at seven specific sites, including cervix, colorectum, endometrium, liver, lung, pancreas and ovary, among which lung cancer, cervix cancer and CRC were strongly indicated. The association of smoking and cancers have been investigated widely in a vast number of observational and experimental studies,[5,34] revealing underlying mechanisms of carcinogens interacting with body vital components, bioactive substances and genetic environment.[35,36] For instance, studies investigating causal effects of several smoking behaviors observationally or experimentally suggested the same findings as ours on lung cancer,[1,2] cervix cancer[34] and CRC.[37] Notably, our results showed that smoking cessation was significantly associated with an increased risk of CRC (OR < 1 indicating risk effect). Given that our results of smoking initiation (comparing current/former smokers with nonsmokers) indicated the overall association of smoking with increased risk of CRC, which was also supported by

subsequent epigenetic MR analysis, this finding did not affect our overall conclusions. Nonetheless, some previous studies have also implied similar clues for the contradictory findings. The genetic correlation analysis in the GWAS where we obtained three smoking measurements suggested that smoking cessation was negatively (with current smokers coded as "2" and previous smokers coded as "1") associated with inflammatory bowel disease (especially ulcerative colitis), potentially indicating an irregular association pattern between smoking cessation and intestinal diseases.[14] Nevertheless, since the relationships between smoking behaviors and cancers were complicated and the effect of smoking cessation was affected by other factors including smoking duration, smoking intensity, the age of quitting smoking and so forth,[38] further evidence is needed.

A vast number of studies have revealed the effect of smoking exerted on DNA methylation across the whole epigenome,[6] which is also responsible for increasing risk of multiple cancers. Our study found that the CpG site cg06639488 (*EFNA1*) to have cross-cancer effect on breast and lung cancer, consistent with previous findings. EFNA1 belongs to the subfamily of ephrins acting as the ligands for Eph receptors, and the interaction of EFNA1 with its most common receptor EphA2 is deemed crucial to the onset of malignant tumors, possibly via regulation of cell cytoskeleton and cell adhesion.[39,40] The upregulation of EFNA1 has already been reported in a broad variety of cancers, for instance, a study has reported a higher transcription and expression of EFNA1 in breast cancer tissues than para-cancerous tissues using the UALCAN database, elucidating the potential values of EPHA/EFNA family-related pathways in predicting breast cancer.[41] In addition, EphA2 was found overexpressed in diverse cancers, among which lung cancer, also reported in our findings, was provided with a pointed strategy targeting EPHA2 blockade.[42] Given that EFNA1-related pathways were widely observed in the pathogenesis of multiple cancers, and with a novel epigenetic perspective provided by our study that EFNA1 could possibly increase the risk of cancers through DNA methylation modification, the further mechanisms of the interaction between EFNA1 and cancers need to be investigated, and therapies targeting EFNA1-related pathways are well worth developing.

Furthermore, our studies identified cg12101586 (*CYP1A1*) to be significantly associated with increased risk of prostate cancer and breast cancer, suggesting the smoking-related methylation at cg12101586 potentially affected the expression of *CYP1A1*, thus consequently increased the risk of cancers at both sites of prostate and breast. The CYP1A1 protein is one of the members of cytochrome P450 subfamily A and participates widely in the metabolic activation of carcinogens as a catalyst, which also situates it under the searchlight of potential carcinogenic effect.[43] A cross-sectional study including 542 healthy women form TwinsUK cohort reported a higher level of expression as well as hypomethylation of CYP1A1 in the current smokers than nonsmokers, supportively showing that the expression of CYP1A1 might be associated with smoking via methylation modification.[44] Moreover, another valuable finding was provided by the same study that high methylation level of CYP1A1 could revert back after quitting smoking (especially 1 year after cessation with >50%

reversal rate), suggesting that smoking cessation could be an effective strategy to alleviate the progression of smoking-induced methylation at CYP1A1. The effect of CYP1A1 on the same cancers has been previously observed[45] reporting that CYP1A1-related pathway could exert the effect of driving cancer pathogenesis, progression and metastasis under the methylation related to smoking behaviors. Therefore, advocacy of smoking cessation needs to be addressed as an essential part of public health strategies combined with clinical treatment to provide early interventions.

The second MR analysis and overlaps of tissue-specific eQTLs additionally indicated cg14142171 (*HLA-L*) to have causal effect on cancers at breast and cervix through methylation modification. HLA is a highly polymorphic supergene which encodes the major histocompatibility complex (MHC) proteins in humans which was divided into three subregions: the HLA class I, II and III regions, in which HLA-L belongs to class I as a pseudogene. HLA is suggested to be an underlying tumor suppressor, and was found with recurring mutations among varieties of malignancies.[46] Studies have already dived into investigating the role of HLA gene variants in the pattern of cancers, suggesting the cooperative effect of multiple HLA regions of class I and II on the onset of cancers with infectious etiology or hematopoietic origin.[47] Given that limited evidence was observed with HLA-L and cancers, practices based on genetics or population yielding novel perspectives of association between HLA-L gene and cancers need further documenting.

Moreover, cg07932199 (*ATXN2*) was remarkably indicated to have potentially same driving variants as those contributing to susceptibility to cancers at breast, colorectum, endometrium and lung, in agreement with previous findings. A study using sequencing analysis of exome and mRNA-seq observed recurrent mutations of *ATXN2* in nonsmokers patients with lung cancer, revealing the nonnegligible role of *ATXN2* in the progress and prognosis of cancer patients.[48] Similarly, a study analyzing the role of m^{7}G-lncRNAs using TCGA identified *ATXN2* as a key target regulated by m7G-lncRNAs, with a higher expression in CRC.[49] Since evidence on the association of *ATXN2* and multiple cancers is relatively sparse, our study provides supportive evidence from an epigenetic perspective of methylation, which requires deeper research in the future targeting related pathways.

Notably, a previous study investigating the causal effect of mQTLs at lung cancer-related CpG sites with lung cancer suggested no confounding effects of smoking behaviors in the associations,[50] which potentially indicated that there was possibly little overlap between smoking-associated and lung cancer-associated methylation pathways, and therefore might not cause confounding to each other. This finding also highlighted the importance of future investigations on the interaction and overlap of methylation among different trait is needed.

Our study has several strengths. First, we explore the association between smoking and 15 site-specific cancers to provide a comprehensive perspective of the varying risks among different cancers responding to smoking, and to further investigate the cross-cancer effect of smoking. Also, genetic instruments for smoking behaviors and multiple cancers were derived from the newest and largest

GWASs to ensure accuracy and reliability. MR analyses were applied in our study to avoid reverse causality and to reduce the interference of confounding factors, and evidence was provided from both genetic and epigenetic perspectives by respectively utilizing SNPs and mQTLs as IVs, further highlighting the underlying role of methylation modification in carcinogenesis. Our study also has some limitations. We obtained methylation data from peripheral blood samples which could show a different methylation pattern from specific tissues,[51] and our methylation data did not include information on time-varying methylation changes. Nonetheless, we validated our main findings (eg, EFNA1, CYP1A1, HLA-L, etc) further with colocalization and tissue-specific expression evidence to enhance the reliability of the results as causal effects. The power of the analyses with some cancers (eg, biliary cancer, testis cancer, etc) might be attenuated due to their small case numbers. Horizontal pleiotropy is an inevitable problem when utilizing genetic variables, especially for phenotypes predicted by a few SNPs. However, we conducted sensitivity analyses, for example, MR Egger and MR-PRESSO which could correct and remove outliers to ensure the robustness. Also, all GWASs were derived from cohort or studies of European ancestry, which impose restrictions on the extrapolation of our conclusion.

5 | CONCLUSIONS

Our study found smoking behaviors to be genetically associated with multiple cancers, and provided further epigenetic perspective that DNA methylation at CpG sites could potentially act as a crucial part of carcinogenesis. Aberrant DNA methylation at several CpG sites related to smoking, including cg06639488 (*EFNA1*), cg12101586 (*CYP1A1*), cg14142171 (*HLA-L*) and cg07932199 (*ATXN2*), were indicated with cross-cancer carcinogenic effects.

AUTHOR CONTRIBUTIONS

The work reported in the paper has been performed by the authors, unless clearly specified in the text. *Study conceptualization and design*: Xue Li, Evropi Theodoratou and Xiao Qian; *Data collection and curation*: Yajing Zhou, Xuan Zhou, Maria Timofeeva, Athina Spiliopoulou, Evropi Theodoratou and Xue Li; *Methodology establishment*: Xuan Zhou, Yajing Zhou and Xue Li; *Formal analysis and software support*: Yajing Zhou, Xuan Zhou and Xue Li. The original draft was written by Yajing Zhou; and the manuscript was revised and edited by Xuan Zhou, Xiao Qian, Xue Li and Evropi Theodoratou. All co-authors had the opportunity to comment on the analysis and interpretation of the findings and approved the final version for publication.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The analysis results of our study are included in this published article and its supplementary information files. The UK Biobank and FinnGen dataset can be accessed through their access application process. Web resources: Genetics of DNA Methylation Consortium (GoDMC), http://mqtldb.godmc.org.uk; The Breast Cancer Association Consortium (BCAC), http://bcac.ccge.medschl.cam.ac.uk/; The Ovarian Cancer Association Consortium (OCAC), https://ocac.ccge.medschl.cam.ac.uk/; The International Lung Cancer Consortium (ILCCO), https://ilcco.iarc.fr/; Finngen Consortium, https://www.finngen.fi/en; UK Biobank, https://www.ukbiobank.ac.uk/. The data from the Genotype-Tissue Expression (GTEx) Project 353 Portal were obtained on 08/18/2022. Further details and other data that support the findings of our study are available from the corresponding authors upon request.

ETHICS STATEMENT

All studies were approved by the irrespective institutional review boards and conducted with appropriate ethical criteria in each country and in accordance with the Declaration of Helsinki.

ORCID

Yazhou He https://orcid.org/0000-0003-2358-0143
Maria Timofeeva https://orcid.org/0000-0002-2503-4253
Kefeng Ding https://orcid.org/0000-0002-2380-3717
Xue Li https://orcid.org/0000-0001-6880-2577

REFERENCES

1. Gandini S, Botteri E, Iodice S, et al. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer*. 2008;122:155-164.
2. Pirie K, Peto R, Reeves GK, Green J, Beral V. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. *Lancet*. 2013;381:133-141.
3. Duan W, Li S, Meng X, Sun Y, Jia C. Smoking and survival of breast cancer patients: a meta-analysis of cohort studies. *Breast*. 2017;33:117-124.
4. Darcey E, Boyle T. Tobacco smoking and survival after a prostate cancer diagnosis: a systematic review and meta-analysis. *Cancer Treat Rev*. 2018;70:30-40.
5. Larsson SC, Carter P, Kar S, et al. Smoking, alcohol consumption, and cancer: a Mendelian randomisation study in UK Biobank and international genetic consortia participants. *PLoS Med*. 2020;17:e1003178.

6. Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of ciga-rette smoking. *Circ Cardiovasc Genet*. 2016;9:436-447.

7. Counts JL, Goodman JI. Alterations in DNA methylation may play a variety of roles in carcinogenesis. *Cell*. 1995;83:13-15.

8. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing poten-tial and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 2007;39:457-466.

9. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer*. 2006;6:107-116.

10. Shui IM, Wong CJ, Zhao S, et al. Prostate tumor DNA methylation is associated with cigarette smoking and adverse prostate cancer out-comes. *Cancer*. 2016;122:2168-2177.

11. Teschendorff AE, Yang Z, Wong A, et al. Correlation of smoking-associated DNA methylation changes in buccal cells with DNA meth-ylation changes in epithelial cancer. *JAMA Oncol*. 2015;1:476-485.

12. Jordahl KM, Phipps AI, Randolph TW, et al. Differential DNA methyl-ation in blood as a mediator of the association between cigarette smoking and bladder cancer risk among postmenopausal women. *Epi-genetics*. 2019;14:1065-1073.

13. Burgess S, Thompson SG. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimationed*. Boca Raton: CRC Press; 2015.

14. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 mil-lion individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019;51:237-244.

15. Wootton RE, Richmond RC, Stuijfzand BG, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med*. 2020;50:2435-2443.

16. Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet*. 2021;53:1311-1321.

17. O'Mara TA, Glubb DM, Amant F, et al. Identification of nine new sus-ceptibility loci for endometrial cancer. *Nat Commun*. 2018;9:3166.

18. Li X, Timofeeva M, Spiliopoulou A, et al. Prediction of colorectal can-cer risk based on profiling with common genetic variants. *Int J Cancer*. 2020;147:3431-3437.

19. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibil-ity loci. *Nat Genet*. 2018;50:928-936.

20. Michailidou K, Lindström S, Dennis J, et al. Association analysis iden-tifies 65 new breast cancer risk loci. *Nature*. 2017;551:92-94.

21. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet*. 2017;49:680-691.

22. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet*. 2014;46:736-741.

23. *FinnGen-Tutkimushanke Vie Suomalaiset Löytöretkelle Genomitietoon*. Helsinki: FinnGen; 2017.

24. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex dis-eases of middle and old age. *PLoS Med*. 2015;12:e1001779.

25. Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk fac-tors and common diseases inferred from GWAS summary data. *Nat Commun*. 2018;9:224.

26. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44:512-525.

27. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estima-tion in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40:304-314.

28. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread hori-zontal pleiotropy in causal relationships inferred from Mendelian ran-domization between complex traits and diseases. *Nat Genet*. 2018;50:693-698.

29. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med*. 2011;30:1312-1323.

30. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45:580-585.

31. Han B, Eskin E. Random-effects model aimed at discovering associa-tions in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88:586-598.

32. Han B, Eskin E. Interpreting meta-analyses of genome-wide associa-tion studies. *PLoS Genet*. 2012;8:e1002555.

33. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using sum-mary statistics. *PLoS Genet*. 2014;10:e1004383.

34. Larsson SC, Burgess S. Appraising the causal role of smoking in multi-ple diseases: a systematic review and meta-analysis of Mendelian ran-domization studies. *EBioMedicine*. 2022;82:104154.

35. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat Rev Cancer*. 2003;3:733-744.

36. Caliri AW, Tommasi S, Besaratinia A. Relationships among smoking, oxidative stress, inflammation, macromolecular damage, and cancer. *Mutat Res Rev Mutat Res*. 2021;787:108365.

37. Dimou N, Yarmolinsky J, Bouras E, et al. Causal effects of lifetime smoking on breast and colorectal cancer risk: Mendelian randomiza-tion study. *Cancer Epidemiol Biomarkers Prev*. 2021;30:953-964.

38. Walter V, Jansen L, Hoffmeister M, Brenner H. Smoking and survival of colorectal cancer patients: systematic review and meta-analysis. *Ann Oncol*. 2014;25:1517-1525.

39. Hao Y, Li G. Role of EFNA1 in tumorigenesis and prospects for cancer therapy. *Biomed Pharmacother*. 2020;130:110567.

40. Coulthard MG, Morgan M, Woodruff TM, et al. Eph/Ephrin signaling in injury and inflammation. *Am J Pathol*. 2012;181:1493-1503.

41. Liang Z, Wang X, Dong K, Li X, Qin C, Zhou H. Expression pattern and prognostic value of EPHA/EFNA in breast cancer by bioinformat-ics analysis: revealing its importance in chemotherapy. *Biomed Res Int*. 2021;2021:5575704.

42. Amato KR, Wang S, Tan L, et al. EPHA2 blockade overcomes acquired resistance to EGFR kinase inhibitors in lung cancer. *Cancer Res*. 2016;76:305-318.

43. Androutsopoulos VP, Tsatsakis AM, Spandidos DA. Cytochrome P450 CYP1A1: wider roles in cancer progression and prevention. *BMC Cancer*. 2009;9:187.

44. Tsai PC, Glastonbury CA, Eliot MN, et al. Smoking induces coordi-nated DNA methylation and gene expression changes in adipose tis-sue with consequences for metabolic health. *Clin Epigenetics*. 2018;10:126.

45. Al-Dhfyan A, Alhoshani A, Korashy HM. Aryl hydrocarbon receptor/cytochrome P450 1A1 pathway mediates breast cancer stem cells expansion through PTEN inhibition and β-catenin and Akt activation. *Mol Cancer*. 2017;16:14.

46. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505:495-501.

47. Liu Z, Derkach A, Yu KJ, et al. Patterns of human leukocyte antigen class I and class II associations and cancer. *Cancer Res*. 2021;81:1148-1152.

48. Sun Z, Wang L, Eckloff BW, et al. Conserved recurrent gene muta-tions correlate with pathway deregulation and clinical outcomes of lung adenocarcinoma in never-smokers. *BMC Med Genomics*. 2014;7:32.

49. Chen S, Dong R, Li Y, Wu H, Liu M. m(7)G-lncRNAs are potential bio-markers for prognosis and tumor microenvironment in patients with colon cancer. *Nan Fang Yi Ke Da Xue Xue Bao*. 2022;42:681-689.

50. Battram T, Richmond RC, Baglietto L, et al. Appraising the causal rele-vance of DNA methylation for risk of lung cancer. *Int J Epidemiol*. 2019;48:1493-1504.

51. Xu Z, Sandler DP, Taylor JA. Blood DNA methylation and breast cancer: a prospective case-cohort analysis in the sister study. *J Natl Cancer Inst*. 2020;112:87-94.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

B-cell malignancies -
A new knowledge hub on the latest research in therapeutic advances

EDUCATIONAL CONTENT AVAILABLE ON THE HUB:

- On-demand Webinars - earn CME credit
- Infographics
- Patient Case Studies
- Currated Research Articles
  ...and much more

VISIT KNOWLEDGE HUB TODAY

WILEY