OncoTargets and Therapy

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Exploring the Association Between PRC2 Genes Variants and Lung Cancer Risk in Chinese Han Population

Min Gao[1,2,*], Yongwen Li[3,*], Hua Huang[4,*], Yaguang Fan[3], Ruifeng Shi[4], Lianchun Su[5], Chen Chen[3], Xuanguang Li[4], Guangsheng Zhu[4], Di Wu[4], Peijun Cao[4], Hongyu Liu[3], Jun Chen[3–5], Shirong Kang[1]

[1]Department of Thoracic Surgery, the Affiliated Hospital of Inner Mongolia Medical University, Hohhot, 010010, People's Republic of China; [2]Inner Mongolia Medical University, Hohhot, 010010, People's Republic of China; [3]Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, People's Republic of China; [4]Department of Lung Cancer Surgery, Tianjin Medical University General Hospital, Tianjin, 300052, People's Republic of China; [5]Department of Thoracic Surgery, First Affiliated Hospital of Shihezi University School of Medicine, Shihezi, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jun Chen; Shirong Kang, Email huntercj2004@qq.com; bssbzzk2013@163.com

Background: Genetic susceptibilities play a large role in the pathogenesis of lung cancer (LC). The polycomb repressive complex 2 (PRC2) is a conserved chromatin-associated complex that represses gene expression and is crucial for proper organismal development and gene expression patterns. Despite PRC2 dysregulation has been observed in various human cancers, the relationship between PRC2 genes variants and lung cancer risk remains largely unexplored.

Methods: To investigate the association between single nucleotide polymorphisms (SNPs) in PRC2 genes and the risk of developing LC, we genotyped blood genomic DNA from 270 LC patients and 452 healthy individuals of Chinese Han ethnicity using the TaqMan™ genotyping technique.

Results: We found that rs17171119T>G(adjusted odds ratio (OR) = 0.662, 95% CI: 0.467–0.938, $P < 0.05$), rs10898459 T>C(adjusted OR = 0.615, 95% CI: 0.4–0.947, $P < 0.05$), and rs1136258 C>T(adjusted OR = 0.273, 95% CI: 0.186–0.401, $P < 0.001$) were significantly associated with a reduced risk of LC. Stratified analysis revealed a protective effect of rs17171119 in both male and female patients, specifically those with lung adenocarcinoma (LUAD). Additionally, rs1391221 showed a protective effect in both the LUAD and lung squamous cell carcinoma (LUSC) groups, while rs1136258 exhibited a protective effect in both females and males, as well as in both LUAD and LUSC groups. Furthermore, analysis of The Cancer Genome Atlas (TCGA) dataset revealed expression levels of EED and RBBP4 in both LUAD and LUSC.

Conclusion: This study provides evidence that allelic variants in EZH2, EED, and RBBP4 may act as protective factors against LC development and could serve as genetic markers associated with susceptibility to LC.

Keywords: the polycomb repressive complex 2, single nucleotide polymorphism, lung cancer

Introduction

Lung cancer (LC) is a highly prevalent malignancy worldwide and stands as the foremost cause of cancer-related mortality.[1] The five-year survival rate of LC remains dishearteningly low, hovering around 20%.[2] Histologically, LC is categorized into small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC accounts for approximately 10%–15% of LC cases, while NSCLC encompasses about 80%–85% of cases, including adenocarcinoma, squamous cell carcinoma, and more rarely, large cell lung cancer.[3] Certain subtypes of LC, such as squamous cell carcinoma and SCLC, are primarily associated with the most significant risk factor—cigarette smoking. Tobacco smoke consists of over 7000 chemicals, including more than 60 known or suspected carcinogens.[4,5]. However, a recent report from the United States has indicated a higher incidence of LC among young women compared to young men,[6] implying

Received: 4 May 2023
Accepted: 27 June 2023
Published: 3 July 2023

that smoking alone does not entirely account for the occurrence of LC. In addition to active smoking, other risk factors, including passive smoking, radon exposure, and dust pollution, can contribute to LC.[7,8] Moreover, genetic factors also influence the incidence of this disease.[9] In this regard, it has been shown that variations at specific gene site can elevate the risk of developing LC in carriers. In recent years, exploration of the molecular atlas of LC has enhanced our comprehension of its occurrence and progression, offering valuable guidance for targeted therapies. Targeted therapies, such as those directed against EGFR, KRAS, and BRAF mutations or amplifications, and ALK alterations, which have significantly improved the prognosis of patients with advanced LC.[7,10]

Genetic susceptibilities play a large role in the risk of developing LC. These susceptibilities influence various aspects, including smoking behavior (which affects dopamine reward mechanisms and nicotine metabolism), metabolism and detoxification of carcinogens, DNA repair, cell cycle control, and other cellular responses.[11] Genome-wide association studies (GWASs) have identified specific chromosomal regions associated with increased susceptibility to developing cancer. Among various types of genetic variants, single nucleotide polymorphisms (SNPs) are particularly prevalent and have a notable impact on cancer susceptibility and individual responses to drugs.[12] Therefore, investigating SNPs provides a robust approach to understanding the etiology, treatment, and prevention of human diseases.[13] SNPs can occur throughout the genome, encompassing coding and non-coding regions (eg, enhancers and promoters), and influence gene transcription and protein expression.[14,15] Therefore, it is crucial to identify cancer-relevant genetic variations and explore their interactions with functional genes to enhance our comprehension of cancer pathogenesis. For example, in our previous study, we revealed a significant association between the SNPs rs920778 and rs1899663 in the long noncoding RNA (lncRNA) HOX transcript intergenic antisense RNA (HOTAIR) and primary LC susceptibility.[16] The distribution frequency of the variant at the rs920778 site of HOTAIR is higher in males and smoking patients with squamous cell carcinoma. Since rs920778 is located in the enhancer region of HOTAIR and carriers of the variant exhibit increased HOTAIR expression levels, there may be an augmented susceptibility to squamous cell carcinoma.

The polycomb repressive complex 2 (PRC2) is a conserved repressive chromatin-associated complex that is essential for the maintenance of organismal development and gene expression patterns to uphold cell identity.[17,18] PRC2 comprises a trimeric core of enhancer of zeste homolog 1 or 2 (EZH1/2), trimers of suppressor of zeste 12 protein homolog (SUZ12), and embryonic ectoderm development (EED).[19] These three proteins, together with retinoblastoma-binding protein 4 or 7 (RBBP4 and RBBP7; also known as RbAp48 and RbAp46), constitute the core subunits of PRC2, which mediates monomethylation, demethylation, and trimethylation ((H3K27 me1, H3K27 me2, and H3K27 me3) on histone H3 and plays key roles in regulating gene expression by ensuring appropriate gene silencing.[20,21] These histone methyltransferase activities of PRC2 are deregulated in several human cancers and certain developmental disorders, such as Weaver syndrome.[17]

Given the importance of PRC2 in the development and progression of many tumors, genetic variants within PRC2 have been implicated in susceptibility to different cancers, including prostate[22] and triple-negative breast cancer.[23] A previous study reported that the C/C genotype of rs6950683 and the C/C genotype of rs3757441 in *EZH2* reduced susceptibility to oral squamous cell carcinoma.[24] Additionally, Hyuna Sung et al[25] identified a protective effect of the T variant at the rs10898459 site in *EED* against esophageal squamous cell carcinoma. Despite these findings, investigations on gene polymorphisms within PRC2 and their association with tumor susceptibility remain scarce. Furthermore, the relationship between genetic variations in PRC2 and susceptibility to LC has not been conclusively established to date. In order to evaluate the effects of SNPs in PRC2 genes on the genetic susceptibility to LC, we selected and investigated 10 SNPs from 4 core PRC2 genes to determine whether these specific SNPs are associated with a genetic predisposition to the pathogenesis of LC.

Materials and Methods

Study Population

Participants in this study included 270 primary LC patients and 452 healthy volunteers. The LC patients were diagnosed based on histopathology at Tianjin Medical University General Hospital from 2021 to 2022. The control group consisted

Table 1 General Information of Research Cohorts

Clinical Factors		Case(N=270)	Control (N=452)
Gender	Male	143(53.0%)	205 (45.4%)
	Female	127(47.0%)	247(54.6%)
Age (year)	≤64	154(57.0%)	384 (85.0%)
	>64	116(43.0%)	68(15.0%)
Smoking	Never	134(49.6%)	
	Ever	136(50.4%)	
Pathological type	LUAD	203(75.2%)	
	LUSC	47(17.4%)	
	SCLC	14(5.2%)	
	Others	6(2.2%)	
Clinical stage	I–II	159(58.9%)	
	III–IV	103(38.1%)	
	Missing	8(3.0%)	
Lymph node metastasis	Yes	77(28.5%)	
	No	169(62.6%)	
	Missing	24(8.9%)	
Pleural spread	Yes	129(47.8%)	
	No	110(40.7%)	
	Missing	31(11.5%)	

Abbreviations: LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; SCLC, small cell lung cancer.

of 452 healthy volunteers registered at the Physical Examination Center of the Affiliated Hospital of Inner Mongolia Medical University in 2021. The following information about the patients was also collected: age, gender, smoking history, clinical stage, lymph node status, pleural status, and pathological diagnosis. The inclusion criteria for LC patients were Chinese Han LC patients without having received any antitumor treatment, whereas the inclusion criteria for the control group were Chinese Han people without cancer or other serious diseases. All participants provided written consent and the research was approved by the Institutional Ethics Committee of the General Hospital of Tianjin Medical University and the Institutional Ethics Committee of the Affiliated Hospital of Inner Mongolia Medical University. Comprehensive dermatological and clinical characteristics of the research cohorts are presented in Table 1.

Blood samples from patients and control volunteers were collected in EDTA-coated tubes and genomic DNA was extracted from peripheral blood lymphocytes using a DNA extraction kit (Qiagen) following the instructions provided by the manufacturer.

SNP Selection and Genotyping

We selected high-quality SNP sites using the following criteria: 1) minor allele frequency of ≥5% in the Chinese Han population based on the NCBI dbSNPs (http://www.ncbi.nlm.nih.gov), of which the top 50 were selected based on their Regulome DB scores; 2) sites with HapMap Chinese Han population (HapMap-HCB) frequencies of ≥8% according to the SNPinfo web server (http://www.snpinfo.niepfunc.htm); 3) common sites that were selected as candidate SNPs based on the above two results. As a result, 10 candidate SNPs of PRC2 were identified and chosen for our analysis; ie, *EZH2* (rs1880357C>G, rs3757441T>C, and rs17171119T>G), *EED* (rs1391221G>C, rs7952481C>G, and rs10898459T>C), *SUZ12* (rs578635T>G and rs508192A>G), and *RBBP4* (rs1136258C>T and rs12407673C>T).

Gene polymorphism analysis was performed using pre-designed TaqMan™ SNP genotyping probes (Applied Biological Systems, Foster City, CA, USA), PCR primers, and master mix (Applied Biosystems). The genotyping procedure was performed using 384-well plates and the ABI7900 real-time PCR system (Applied Biological Systems). Each reaction system included genotyping master mix (2 units, 2.5 μL), genotyping assay reagent (40 units, 0.125 μL), and template DNA (2.5 μL). To ensure quality control, each test included a no-template (NTC) and a positive control.

The PCR conditions consisted of an initial denaturation at 95°C for 2min, followed by 40 cycles of denaturation at 95°C for 15s and annealing/extension at 60°C for 1min. The readings were recalculated for each plate and the results were analyzed using SDS2.4 software. During genotyping, to ensure experimental accuracy and reproducibility, approximately 10% of the samples were subjected to repeated testing, and consistency was detected.

Bioinformatics Analysis

The mRNA expression data of LUAD and LUSC were downloaded from the TCGA database (https://portal.gdc.cancer.gov/repository). LUAD consisted of a total of 54 paracancerous and 497 carcinoma samples, and LUSC included a total of 49 paracancerous and 502 carcinoma samples. Clinical data on LC in the KM-plotter database (https://kmplot.com/analysis/index.php?p=service) were examined to determine whether EZH2, EED, and RBBP4 are associated with the prognosis of patients with LC.

Statistical Analysis

All data analyses were performed using SPSS24.0 software. *T*-test was used for intergroup comparisons, and a *P* value of <0.05 indicated statistical significance. Hardy–Weinberg equilibrium (HWE) for each SNP among controls was tested using a goodness-of-ft χ2-test. Logistic regression analysis was used to calculate the risk factors and odds ratios (ORs), which were adjusted by age and gender, as well as 95% confidence intervals (95% CIs), with $P < 0.05$ indicating statistical significance. Linkage disequilibrium (LD) analysis between different genetic polymorphic sites was performed using the SHEsis online software provided at (http://analysis.bio-x.cn/myAnalysis.php). LD analysis of the SNP polymorphism sites was performed to valuate their nonrandom association in our research cohorts. The degree of LD between SNPs was further evaluated using the correlation factor D' and ≥0.5 as the degree of LD threshold.

Results

Characteristics of the Study Cohorts

A total of 722 participants were enrolled in the study, including 270 individuals diagnosed with LC and 452 healthy volunteers. Among the cases, there were 143 males and 127 females, with a median age of 64 years. Out of the LC patients, 203 were diagnosed with LUAD, 47 with LUSC, 14 with SCLC, and 6 with other pathological types of cancer. According to the eighth edition of the tumor node metastasis staging criteria of NSCLC and SCLC, the LC group included 159 cases classified as stages I and II, and 103 cases classified as stages III and IV. A total of 77 cases (28.5%)

Table 2 Single Nucleotide Polymorphisms (SNPs) and HWE *P*-values in the Control Group

Gene	Species	SNP ID	Position	Allele	HWE *P* value
EZH2	Human	rs17171119	7,148,820,364	T/G	*<0.05
EZH2	Human	rs1880357	7,148,880,518	C/G	0.139
EZH2	Human	rs3757441	7,148,827,660	T/C	0.086
EED	Human	rs7952481	1,186,243,380	G/C	*<0.05
EED	Human	rs1391221	1,186,244,237	G/C	0.143
EED	Human	rs10898459	1,186,261,897	T/C	0.348
SUZ12	Human	rs508192	1,731,988,128	A/G	0.944
SUZ12	Human	rs578635	1,731,962,878	T/G	0.775
RBBP4	Human	rs12407673	132,651,491	C/T	0.920
RBBP4	Human	rs1136258	132,677,093	C/T	0.111

Note: *$P \leq 0.05$.

had lymph node metastasis and 129 cases had pleural metastasis. Detailed information regarding the research cohorts presented in Table 1. Table 2 presents the site information of SNPs and the HWE *P* value of the control group.

Distribution Frequency of PRC2 Gene SNPs

The distribution frequencies and corresponding OR values of the SNP sites of PRC2 genes in the healthy and LC patient groups are shown in Table 3. The OR value was calibrated according to gender and age. The distribution frequency of the TG genotype (TG/GG) at the rs17171119T>G site of *EZH2* was significantly lower in the LC group than that of the control group (adjusted OR = 0.662, 95% CI: 0.467–0.938, $P < 0.05$), and we found that the genotype and allele frequencies of rs1880357C>G and rs3757441T>C of EZH2 were not significantly different between healthy controls and LC patients.

In addition, the distribution frequency of the CC genotype at the rs1391221(G > C) site of *EED* was significantly higher in the cancer group than that in the control group (adjusted OR = 3.268, 95% CI: 1.005–10.63, $P < 0.05$). The

Table 3 Logistic Regression Analysis of Single Nucleotide Polymorphisms (SNPs) and Lung Cancer Susceptibility

Genes	SNP	Genotype	Case (N=270)	Control(N=452)	OR[a](95% CI)	*P value*
EZH2	rs1880357					
		CC	188 (69.6%)	302 (66.8%)	1	–
		C/G	70 (25.9%)	129 (28.5%)	0.826 (0.56, 1.219)	0.336
		GG	12 (4.4%)	21 (4.6%)	0.643 (0.283, 1.46)	0.290
		C/G & GG	82 (30.4%)	150 (33.2%)	0.796 (0.551, 1.151)	0.225
		C	446 (82.6%)	733 (81.1%)	1	–
		G	94 (17.4%)	171 (18.9%)	0.801 (0.585, 1.095)	0.163
	rs3757441					
		TT	149(55.2%)	245(54.2%)	1	–
		T/C	103(38.1%)	166(36.7%)	0.907 (0.630, 1.306)	0.601
		CC	18(6.7%)	41(9.1%)	0.659 (0.340, 1.278)	0.217
		T/C & CC	121(44.8%)	207(45.8%)	0.858 (0.607, 1.212)	0.386
		T	401(74.3%)	657(72.7%)	1	–
		C	139(25.7%)	247(27.3%)	0.846 (0.643, 1.114)	0.234
	rs17171119					
		TT	160 (59.3%)	240 (53.1%)	1	–
		T/G	94 (34.8%)	130 (28.8%)	0.943 (0.644, 1.38)	0.762
		GG	16 (5.9%)	82 (18.1%)	0.231 (0.123, 0.435)	<0.001
		T/G &GG	110 (40.7%)	212 (46.9%)	0.662 (0.467, 0.938)	0.02
		T	414 (76.7%)	610 (67.5%)	1	–
		G	126 (23.3%)	292 (32.3%)	0.547 (0.416, 0.719)	<0.001
EED	rs1391221					
		GG	184(68.2%)	324(71.7%)	1	–
		G/C	77(28.5%)	122(27.0%)	1.270 (0.864, 1.866)	0.224
		CC	9(3.3%)	6(1.3%)	3.268(1.005, 10.63)	*0.049
		G/C &CC	86(31.9%)	128(28.3%)	1.361 (0.936, 1.979)	0.107
		G	444(82.2%)	770(85.2%)	1	–
		C	96(17.8%)	134(14.8%)	1.385 (0.999, 1.919)	*0.050
	rs10898459					
		TT	80 (29.6%)	103(22.8)	1	–
		T/C	101 (37.4%)	215 (47.6%)	0.615 (0.400, 0.947)	*0.027
		CC	89 (33.0%)	134 (29.6%)	0.829 (0.529, 1.301)	0.414
		T/C &CC	190 (70.4%)	349 (77.2%)	0.703 (0.475, 1.038)	0.076
		T	261 (48.3%)	421 (46.6%)	1	–
		C	279 (51.7%)	483 (53.4%)	0.915 (0.717, 1.167)	0.472

(*Continued*)

Table 3 (Continued).

Genes	SNP	Genotype	Case (N=270)	Control(N=452)	OR[a](95% CI)	*P value*
	rs7952481					
		GG	55 (20.4%)	80 (17.7%)	1	–
		G/C	98 (36.3%)	174 (38.5%)	0.931 (0.574, 1.509)	0.771
		CC	117 (43.3%)	198 (43.8%)	0.925 (0.578, 1.48)	0.745
		G/C &CC	215 (79.6%)	372 (82.3%)	0.928 (0.6, 1.434)	0.735
		G	208 (38.5%)	334 (36.9%)	1	–
		C	332 (61.5%)	570 (63.1%)	0.96 (0.747, 1.233)	0.749
SUZ12	rs578635					
		TT	148 (54.8%)	239 (52.9%)	1	–
		T/G	101 (37.4%)	181 (40.0%)	0.806 (0.56, 1.159)	0.244
		GG	21 (7.8%)	32 (7.1%)	0.962 (0.49, 1.89)	0.910
		T/G &GG	122 (45.2%)	213 (47.1%)	0.829 (0.586, 1.172)	0.287
		T	397 (73.5%)	659 (72.9%)	1	–
		G	143 (26.5%)	245 (27.1%)	0.896 (0.68, 1.18)	0.433
	rs508192					
		AA	268 (99.3%)	449 (99.3%)	1	–
		A/G	2 (0.7%)	3 (0.7%)	0.637 (0.092, 4.393)	0.646
		GG	0	0	–	–
		A/G &GG	2 (0.7%)	3 (0.7%)	0.637 (0.092, 4.393)	0.646
		A	538 (99.6%)	901 (99.7%)	1	–
		G	2 (0.4%)	3 (0.3%)	0.638 (0.093, 4.371)	0.647
RBBP4	rs1136258					
		CC	206 (76.3%)	226 (50.0%)	1	–
		C/T	40 (14.8%)	177 (39.2%)	0.227 (0.146, 0.354)	*<0.001
		TT	24 (8.9%)	49 (10.8%)	0.417 (0.23, 0.756)	*0.004
		C/T &TT	64 (23.7%)	226 (50.0%)	0.273 (0.186, 0.401)	*<0.001
		C	452 (83.7%)	629 (69.6%)	1	–
		T	88 (16.3%)	275 (30.4%)	0.388 (0.286, 0.526)	*<0.001
	rs12407673					
		CC	207 (67.7%)	360 (79.6%)	1	–
		C/T	57 (21.1%)	87 (19.2%)	0.964 (0.627, 1.482)	0.868
		TT	6 (2.2%)	5 (1.1%)	2.148 (0.533, 8.65)	0.281
		C/T &TT	63 (23.3%)	92 (20.4%)	1.022 (0.674, 1.549)	0.918
		C	471 (87.2%)	807 (89.3%)	1	–
		T	69 (12.8%)	97 (10.7%)	1.079 (0.74, 1.573)	0.692

Notes: Data were calculated by logistic regression with adjustment for gender and age as covariates. a, Adjusted by age and gender. *$P \leq 0.05$.
Abbreviations: OR, odds ratio; CI, confidence interval.

distribution frequency of the T/C genotype at the rs10898459(T > C) site of *EED* was significantly lower in the LC group than in the control group (adjusted OR =0.615, 95% CI: 0.4—0.947, $P < 0.05$).

Among the 10 SNP sites, rs1136258 of *RBBP4* was highly associated with LC. In the LC and healthy control groups, the distribution frequency of the C/T (C/T + TT) genotype in the LC group was lower than that in the healthy control group (adjusted OR = 0.273, 95% CI: 0.186–0.401, $P < 0.001$). The distribution frequencies of other SNP sites were not significantly different between the LC and healthy control groups.

3.3. Variants at rs17171119 of EZH2, rs1391221 of *EED*, and rs1136258 of *RBBP4* were significantly lower in LC patients compared than in healthy volunteers stratified by clinical factors.

A stratified analysis was performed for the rs17171119 variant of *EZH2* considering gender, age, and histology type. The distribution frequency of the GG genotype in the LC group was significantly lower than that in the healthy control group among males and females, both above and below the medium age group, as well as in the LUAD group. Notably,

no difference was observed in the LUSC group, suggesting that the GG phenotype was a protective factor for LUAD patients, but not for LUSC patients.

Similarly, a stratified analysis was also performed for the rs1136258 variant of *RBBP4* considering gender, age, and histology type. The distribution frequency of the C/T (C/T + TT) genotype in the LC group was lower than that in the healthy control group for both males and females, across both age groups, and in both the LUAD and LUSC groups. This indicated that the C/T (adjusted OR = 0.227, 95% CI: 0.146–0.354, $P < 0.001$) and C/T + TT (adjusted OR = 0.273, 95% CI: 0.186–0.401, $P < 0.001$) genotypes act as protective factors for LC.

Furthermore, the stratified analysis revealed that the distribution frequency of the (CC/GC) phenotype at the rs1391221(G > C) site of *EED* in the LC group was lower than that in the healthy control group for both theLUAD and LUSC groups, indicating that this phenotype serve as a protective factor for primary LC. For the rs10898459 site, the distribution frequency of the T/C genotype among females was significantly lower than that in the healthy control group, younger medium-age group, and the LUAD group. This indicated that the T/C phenotype of rs10898459 act as a protective factor for these sub-groups. Please refer to Table 4 for further details.

Table 4 Stratified Analyses of rs17171119 in *EZH2*, rs10898459 and rs1391221 in *EED*, and rs1136258 in *RBBP4* for Association with LC Risk

Gene	SNP	Stratify	Genotype	Case	Control	OR[a](95% CI)	*P* value
EZH2	rs17171119	Male	TT	85 (59.4%)	112 (54.6%)	1	–
			T/G	48 (33.6%)	55 (26.8%)	0.976 (0.568, 1.679)	0.931
			GG	10 (7.0%)	38 (18.5%)	0.369 (0.16, 0.847)	*0.018
			T/G&GG	58 (40.6%)	93 (45.4%)	0.751 (0.459, 1.229)	0.253
		Female	TT	75 (59.1%)	128 (51.8%)	1	–
			T/G	46 (36.2%)	75 (30.4%)	0.907 (0.529, 1.553)	0.72
			GG	6 (4.7%)	44 (17.8%)	0.131 (0.048, 0.359)	*<0.001
			T/G&GG	52 (40.9%)	119 (48.2%)	0.581 (0.353, 0.956)	*0.032
		Age	TT	91 (59.1%)	206 (53.6%)	1	–
		<=64	T/G	52 (33.8%)	108 (28.1%)	0.866 (0.535, 1.402)	0.557
			GG	11 (7.1%)	70 (18.2%)	0.215 (0.101, 0.456)	*<0.001
			T/G&GG	63 (40.9%)	178 (46.4%)	0.584 (0.376, 0.906)	*0.016
		Age	TT	69 (59.5%)	34 (50.0%)	1	–
		>64	T/G	42 (36.2%)	22 (32.4%)	0.909 (0.465, 1.776)	0.778
			GG	5 (4.3%)	12 (17.6%)	0.207 (0.066, 0.648)	*0.006
			T/G&GG	47 (40.5%)	34 (50.0%)	0.665 (0.36, 1.228)	0.189
		LUAD	TT	124 (61.1%)	240 (53.1%)	1	–
			T/G	67 (33.0%)	130 (28.8%)	0.876 (0.582, 1.317)	0.522
			GG	12 (5.9%)	82 (18.1%)	0.218 (0.108, 0.44)	*<0.001
			T/G&GG	79 (38.9%)	212 (46.9%)	0.616 (0.423, 0.897)	*0.011
		LUSC	TT	26 (55.3%)	240 (53.1%)	1	–
			T/G	19 (40.4%)	130 (28.8%)	1.337 (0.65, 2.75)	0.429
			GG	2 (4.3%)	82 (18.1%)	0.239 (0.052, 1.101)	0.066
			T/G&GG	21 (44.7%)	212 (46.9%)	0.924 (0.466, 1.831)	0.820
EED	rs10898459	Male	TT	36 (25.2%)	46 (22.4%)	1	–
			T/C	60 (42.0%)	102 (49.8%)	0.747 (0.402, 1.389)	0.355
			CC	47 (32.8%)	57 (27.8%)	0.964 (0.499, 1.863)	0.913
			T/C&CC	107(74.8%)	159 (77.6%)	0.832 (0.47, 1.472)	0.525
		Female	TT	44 (34.6%)	57 (23.1%)	1	–
			T/C	41 (32.3%)	113 (45.7%)	0.513 (0.28, 0.941)	*0.030
			CC	42 (33.1%)	77 (31.2%)	0.732 (0.394, 1.358)	0.320
			T/C&CC	83 (65.4%)	190 (76.9%)	0.606 (0.354, 1.037)	0.067
		Age	TT	51 (33.1%)	83 (21.6%)	1	–

(*Continued*)

Table 4 (Continued).

Gene	SNP	Stratify	Genotype	Case	Control	OR[a](95% CI)	*P* value
		≤64	T/C	52 (33.8%)	185 (48.2%)	0.453 (0.284, 0.724)	*0.001
			CC	51 (33.1%)	116 (30.2%)	0.721 (0.445, 1.168)	0.182
			T/C&CC	103(66.9%)	301 (78.4%)	0.556 (0.366, 0.843)	*0.006
		Age	TT	29 (25.0%)	20 (29.4%)	1	–
		>64	T/C	49 (42.2%)	30 (44.1%)	1.112 (0.527, 2.344)	0.780
			CC	38 (32.8%)	18 (26.5%)	1.443 (0.642, 3.243)	0.372
			T/C&CC	87 (75.0%)	48 (70.6%)	1.239 (0.625, 2.456)	0.537
		LUAD	TT	62 (30.5%)	103 (22.8%)	1	–
			T/C	72 (35.5%)	215 (47.6%)	0.592 (0.373, 0.939)	*0.026
			CC	69 (34.0%)	134 (29.6%)	0.859 (0.534, 1.379)	0.528
			T/C&CC	141 (69.5%)	349 (77.2%)	0.701 (0.464, 1.06)	0.092
		LUSC	TT	11 (23.4%)	103 (22.8%)	1	–
			T/C	21 (44.7%)	215 (47.6%)	1.041 (0.432, 2.508)	0.928
			CC	15 (31.9%)	134 (29.6%)	1.269 (0.499, 3.224)	0.616
			T/C&CC	36 (76.6%)	349 (77.2%)	1.129 (0.500, 2.548)	0.769
	rs1391221	Male	GG	104(72.7%)	156 (76.1%)	1	–
			G/C	34 (23.8%)	47 (22.9%)	1.131 (0.638, 2.003)	0.673
			CC	5 (3.5%)	2 (1.0%)	4.891(0.801,29.887)	0.085
			G/C&CC	39 (27.3%)	49 (23.9%)	1.272 (0.732, 2.209)	0.391
		Female	GG	80 (63.0%)	168 (68.0%)	1	–
			G/C	43 (33.9%)	75 (30.4%)	1.395 (0.826, 2.357)	0.211
			CC	4 (3.1%)	4 (1.6%)	2.324(0.456,11.838)	0.308
			G/C &CC	47 (37.0%)	79 (32.0%)	1.443 (0.864, 2.409)	0.160
		Age	GG	102 (66.2%)	270 (70.3%)	1	–
		≤64	G/C	45 (29.2%)	109 (28.4%)	1.135 (0.746, 1.727)	0.553
			CC	7 (4.5%)	5 (1.3%)	3.737(1.152,12.124)	*0.028
			G/C&CC	52 (33.8%)	114 (29.7%)	1.253 (0.837, 1.876)	0.273
		>64	GG	82 (70.7%)	54 (79.4%)	1	–
			G/C	32 (27.6%)	13 (19.1%)	1.637 (0.783, 3.423)	0.187
			CC	2 (1.7%)	1 (1.5%)	1.35 (0.117, 15.578)	0.809
			G/C&CC	34 (29.3%)	14 (20.6%)	1.617 (0.788, 3.317)	0.187
		LUAD	GG	135 (66.5%)	324 (71.7%)	1	–
			G/C	62 (30.5%)	122 (27.0%)	0.025 (0.01, 0.063)	*<0.001
			CC	6 (3.0%)	6 (1.3%)	0.001 (0, 0.002)	*<0.001
			G/C&CC	68 (33.5%)	128 (28.3%)	0.007 (0.003, 0.018)	*<0.001
		LUSC	GG	32 (68.1%)	324 (71.7%)	1	–
			G/C	12 (25.5%)	122 (27.0%)	0.036 (0.01, 0.128)	*<0.001
			CC	3 (6.4%)	6 (1.3%)	0.002 (0, 0.011)	*<0.001
			G/C&CC	15 (31.9%)	128 (28.3%)	0.01 (0.003, 0.032)	*<0.001
RBBP4	rs1136258	Male	CC	116(81.1%)	102 (49.8%)	1	–
			C/T	14 (9.8%)	77 (37.6%)	0.106 (0.051, 0.22)	*<0.001
			TT	13 (9.1%)	26 (12.6%)	0.333 (0.147, 0.753)	*0.008
			C/T &TT	27 (18.9%)	103 (50.2%)	0.164 (0.091, 0.295)	*<0.001
		Female	CC	90 (70.9%)	124 (50.2%)	1	–
			C/T	26 (20.5%)	100 (40.5%)	0.4 (0.226, 0.708)	*0.002
			TT	11 (8.6%)	23 (9.3%)	0.517 (0.214, 1.249)	0.141
			C/T &TT	37 (29.1%)	123 (49.8%)	0.426 (0.254, 0.714)	*0.001
		Age	CC	122(79.2%)	200 (52.1%)	1	–
		≤64	C/T	19 (12.3%)	146 (38.0%)	0.218 (0.128, 0.371)	*<0.001
			TT	13 (8.5%)	38 (9.9%)	0.552 (0.282, 1.081)	0.083
			C/T &TT	32 (20.8%)	184 (47.9%)	0.289 (0.186, 0.449)	*<0.001

(*Continued*)

Table 4 (Continued).

Gene	SNP	Stratify	Genotype	Case	Control	OR[a](95% CI)	*P* value
		Age	CC	84 (72.4%)	26 (38.2%)	1	–
		>64	C/T	21 (18.1%)	31 (45.6%)	0.21 (0.103, 0.428)	*<0.001
			TT	11 (9.5%)	11 (16.2%)	0.31 (0.12, 0.803)	*0.015
			C/T &TT	32 (27.6%)	42 (61.8%)	0.236 (0.124, 0.448)	*<0.001
		LUAD	CC	151 (74.4%)	226 (50%)	1	–
			C/T	32 (15.8%)	177 (39.2%)	0.25 (0.156, 0.4)	*<0.001
			TT	20 (9.9%)	49 (10.8%)	0.492 (0.264, 0.917)	*0.025
			C/T &TT	52 (25.6%)	226 (50.0%)	0.307 (0.204, 0.461)	*<0.001
		LUSC	CC	37 (78.7%)	226 (50.0%)	1	–
			C/T	6 (12.8%)	177 (39.2%)	0.15 (0.055, 0.408)	*<0.001
			TT	4 (8.5%)	49 (10.8%)	0.32 (0.098, 1.047)	0.059
			C/T &TT	10 (21.3%)	226 (50.0%)	0.194 (0.085, 0.442)	*<0.001

Notes: Data were calculated by logistic regression with adjustment for gender and age as covariates. a Adjusted by age and gender. *$P \leq 0.05$.
Abbreviations: OR, odds ratio; CI, confidence interval; LUAD, lung adenocarcinoma, LUSC, lung squamous cell carcinoma, SCLC, small cell lung cancer.

Haplotype Association Analysis

Since the genotypes at rs17171119T>G of *EZH2* and rs7952481C>G of *EED* not statistically confirmed to Hardy-Weinberg equilibrium ($P < 0.05$ respectively) in Table 2, we did not analyze the haplotype association further.

Gene LD and haplotype analyses were further performed to investigate the relationship between gene polymorphisms and LC risk. We performed LD and haplotype analyses on four genes of PRC2. At the rs508192 and rs578635 sites of *SUZ12*, there was a high LD (D' = 1, $r^2 = 0.009$, Figure 1a). Additionally, a certain level of LD was found between the rs12407673 and rs1136258 sites of RBBP4 (D' = 0.672, $r^2 = 0.176$, Figure 1b). However, there was no significant LD between the rs3757441 and rs1880357 sites of EZH2 (D' = 0.295, $r^2 = 0.010$, Figure 1c), as well as between the rs10898459 and rs1391221 sites of *EED* (D' = 0.050, $r^2 = 0.001$, Figure 1d).

The possible haplotypes were examined in the cancer and control groups (Table 5). We found that the haplotype of C(rs1391221) T(rs10898459) in *EED* was related to an increased LC risk (adjusted OR = 2.647, 95% CI = 1.539–4.552, $P < 0.001$). Moreover, the population carrying the haplotype C(rs1136258) T(rs12407673) exhibited a significantly decreased LC risk (adjusted OR = 0.033, 95% CI = 0.012–0.088, $P < 0.0001$). Similarly, the haplotype T(rs1136258) C(rs12407673) of RBBP4 was associated with a reduced LC risk (adjusted OR = 0.379, 95% CI = 0.23–0.624, $P < 0.001$).

Analysis of EED and RBBP4 Expression in Patients with LC Patients Using the TCGA Database

Our previous analysis showed that the T/G (TG + GG) genotype atrs17171119 of *EZH2* and the T/C (TC + CC) genotype at rs10898459 of *EED* were associated with a lower incidence of LC in the Chinese Han population. The G/C(GC + CC) genotype at rs1391221 and C/T (CT + TT) mutant genotypes at rs1136258 of *RBBP4* appear to confer protection against LC compared with the wild-type genotypes. It has been reported that some variants of SNPs are associated with decreased mRNA levels, suggesting that allelic variations at these sites may influence expression levels and functions. To further investigate this, we analyzed RNA sequencing data from the TCGA database and found that the expression levels of EZH2, EED, and RBBP4 in LUAD and LUSC tissues were significantly higher than those in tissues adjacent to cancer (Figure 2d and e). Moreover, patients with elevated EZH2, EED, and RBBP4 expression showed worse prognosis (Figure 2a-c).

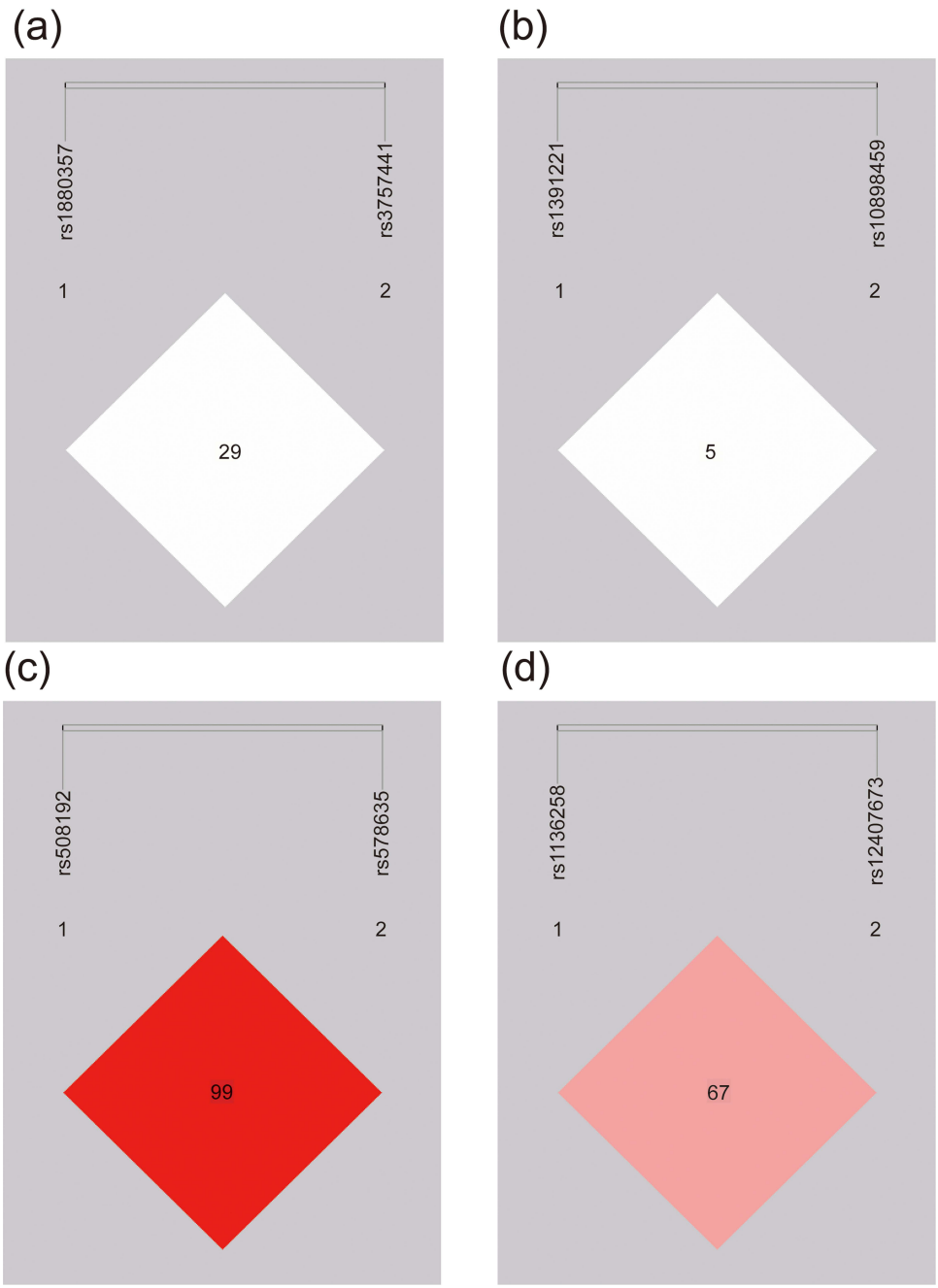

Figure 1 Linkage disequilibrium (LD) in *EZH2* (rs1880357, rs3757441), *EED* (rs1391221, rs10898459), *SUZ12* (rs578635, rs508192), and *RBBP4* (rs1136258, rs12407673). Figures show the LD plots of the SNPs in *SUZ12* (**a**), *RBBP4* (**b**), *EZH2* (**c**), and *EED* (**d**). Statistically significant SNPs are indicated by a red box.

Discussion

LC remains the leading cause of cancer-related mortality worldwide.[26] Genetic susceptibility to LC is substantial and is supported by a great amount of evidence.[9,27] Studies have demonstrated familial aggregation of LC and identified some chromosomal regions associated with LC risk.[28,29] Therefore, understanding the genetic factors associated with LC susceptibility is crucial for comprehending its pathogenesis.

PRC2 is one of the most studied complexes in cancer, and its dysregulation is frequently associated with poor prognosis.[30,31] Recent research has focused on rare genetic variants of PRC2, revealing loss-of-function and gain-of-function mutations in PRC2 member genes in various cancers.[32] Additionally, genetic variations in PRC2 have been

Table 5 Haplotype Analysis of the Control Group

Gene	SNP	SNP	Control	Case	OR[a] (95% CI)	*P* value
EZH2	rs1880357	rs3757441				
	C C G G	T C T C	337 (62.4%) 109 (20.2%) 64 (11.8%) 30 (5.6%)	547 (60.4%) 186 (20.6%) 109 (12.1%) 62 (6.9%)	1 0.951 (0.724, 1.25) 0.953 (0.68, 1.336) 0.785 (0.497, 1.24)	– 0.719 0.78 0.3
SUZ12	rs578635	rs508192				
	T G T G	A A G G	397 (73.5%) 141 (26.1%) 0 (0%) 2 (0.4%)	659 (72.9%) 242 (26.8%) 0 (0%) 3 (0.3%)	1 0.967 (0.759, 1.232) – 1.107 (0.184, 6.662)	– 0.787 – 0.912
EED	rs1391221	rs10898459				
	G G C C	T C T C	225 (41.7%) 220 (40.7%) 36 (6.7%) 59 (10.9%)	397 (43.8%) 373 (41.3%) 24 (2.7%) 110 (12.2%)	1 1.041 (0.824, 1.315) 2.647 (1.539, 4.552) 0.946 (0.663, 1.352)	– 0.738 *<0.001 0.762
RBBP4	rs1136258	rs12407673				
	C C T T	C T C T	449 (83.1%) 4 (0.8%) 21 (3.9%) 66 (12.2%)	599 (66.3%) 164 (18.1%) 74 (8.2%) 67 (7.4%)	1 0.033 (0.012, 0.088) 0.379 (0.23, 0.624) 1.314 (0.915, 1.887)	– *<0.001 *<0.001 0.138

Notes: Data were calculated by logistic regression with adjustment for gender and age as covariates. *$P \leq 0.05$.
Abbreviations: OR, odds ratio; CI, confidence interval.

implicated in individual susceptibility to cancer. In this study, we systematically examined the genetic variations of 10 SNP sites in 4 core genes of PRC2 (*EZH2, EED, SUZ12*, and *RBBP4*) among LC and healthy populations. Our aim was to identify gene effects and provide information on the relationship between these sites and genetic susceptibility to LC in a high-risk north China population.

Our findings demonstrated a significant association between PRC2 gene polymorphisms and LC in this case-controlled study. Specifically, we observed a reduced risk of LC associated with variant genotypes of rs17171119, rs10898459, and rs1136258. Stratified analysis further revealed that the protective effect of rs17171119 was observed in both males and females, but exclusively in LUAD patients. The protective effect of rs1391221 was evident in both the LUAD and LUSC groups, while the protective effect of rs1136258 was observed in both males and females, as well as in both the LUAD and LUSC groups. However, the likelihood of obtaining false positive results increases when multiple statistical tests are performed. The possible limitation of multiple testing and its impact on the interpretation of our results might not be generalisable. In the future, we hope to replicate our findings using larger sample sizes.

Over the last two decades, the International Lung Cancer Consortium (ILCCO) has collected more than 50,000 cases and controls and has identified several genes/SNPs associated with LC. For example, two tyrosine-DNA phosphodiesterase SNPs (rs942190 and rs2401863) in 890 patients from 10 ILCCO studies were analyzed, and the rs942190 GG genotype was found to be associated with a relatively low survival rate in SCLC patients with SCLC.[33] Another interdisciplinary study from the ILCCO using sequences derived from LUAD cells found an SNP (rs12614710) associated with NSCLC in EPAS1 (OR = 1.50; 95% CI: 1.31–1.72; p = 7.75×10-9).[34] Moreover, a GWAS exploring the relationship between genetic mutations and exposure to asbestos in LC revealed that individuals with heterozygous and homozygous SNPs (rs13053856, rs11090910, rs11703832, and rs12170325) for the MIRLET7B gene were

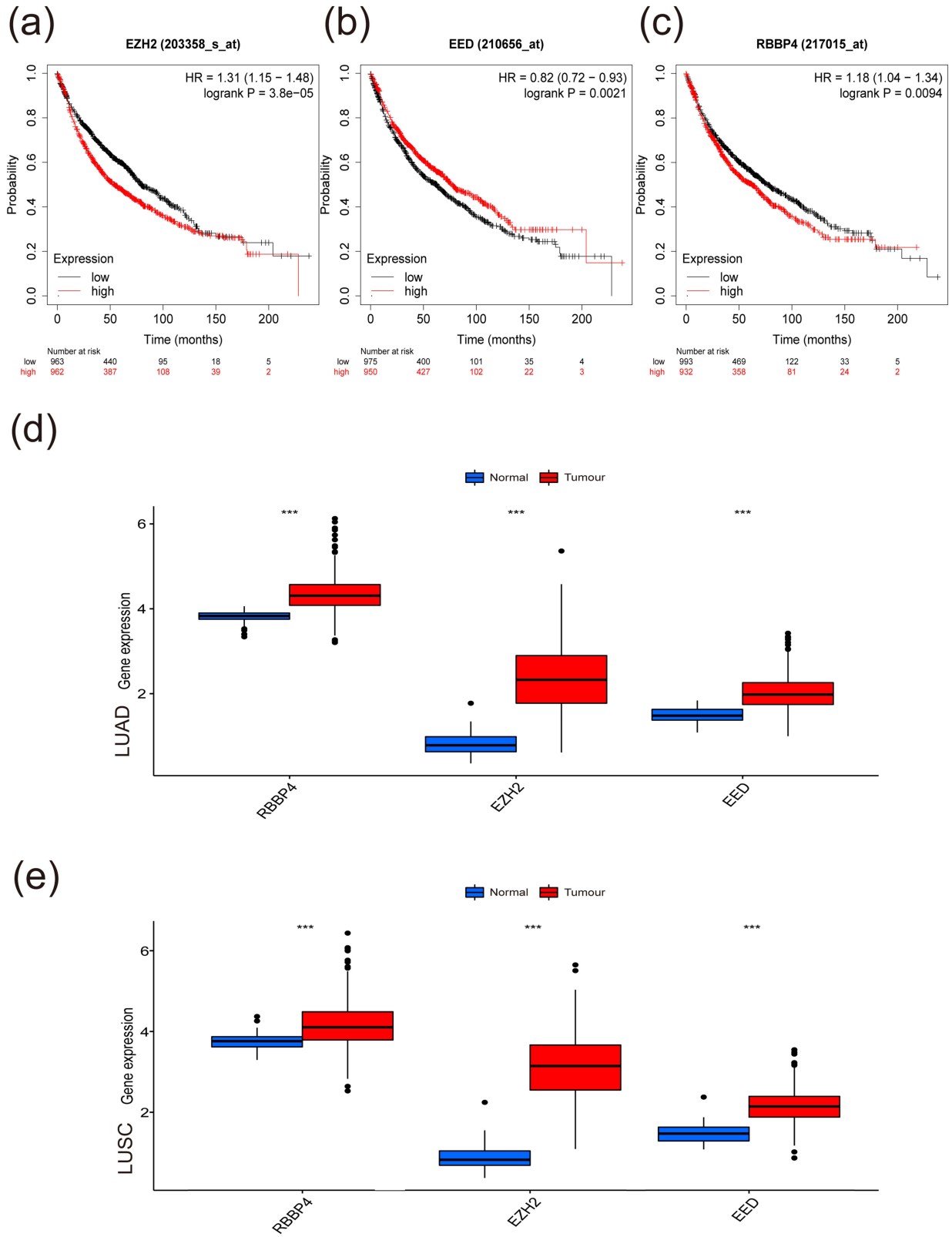

Figure 2 Analyses of PRC2 gene expression, clinical characteristics, and prognostic conditions in lung cancer patients. **(a-c)** Overall survival and differential expression levels of *EZH2* **(a)**, *EED* **(b)**, and *RBBP4* **(c)** in lung cancer cells were analyzed using KM-plotter databases. **(d)** Differential expression levels of *EZH2*, *EED*, and *RBBP4* in lung adenocarcinoma cells and adjacent tissues. **(e)** Differential expression levels of *EZH2*, *EED*, and *RBBP4* in lung squamous carcinoma cells and adjacent tissues. ***$P<0.001$.

significantly associated with an increased risk of LC (likelihood ratio test, $P < 5 \times 10^{-7}$, df = 1).[35] This study included 1539 cases and 1761 controls.

Certain SNPs have been shown to impact gene expression, subsequently influencing gene function. For example, the T variant of rs10898459 of *EED* is associated with a protective effect against esophageal squamous cell carcinoma and is correlated with reduced *EED* mRNA levels in normal esophageal tissues.[25] Promoter polymorphisms of *EED* are associated with susceptibility to ulcerative colitis due to changes in EED expression levels.[36] Consistent with these findings, our analysis of EZH2, EED, and RBBP4 expressions in LC based on the TCGA data revealed elevated expression levels in LUAD and LUSC tissues compared with normal tissues. Moreover, high expression levels of EZH2, EED, and RBBP4 were associated with a worse prognosis. Collectively, these findings are consistent with the possibility that SNPs in PRC2 genes may impact the function of genes and thus play a role in LC development. However, as a limitation of this study, gene expression levels were not evaluated in the LC group or control group. Thus, it is premature to infer that SNPs of PRC2 genes influence PRC2 gene expression and susceptibility to LC. Further investigation will be performed in the future.

To the best of our knowledge, this is the first study to elucidate the association between PRC2 polymorphisms and LC risk, but it has some limitations. First, future research should explore a wider array of potentially functional SNPs within PRC2. Second, our findings cannot be generalized due to sample size limitations and geographic bias. Further studies with larger sample sizes and more diverse human populations than those in the present study are needed to confirm the findings. Additionally, while smoking is a well-known critical factor in LC, the absence of clinical data and smoking history in the healthy control group may lead to incomplete study outcomes. Lastly, although our data may help develop risk prediction models, the precise mechanisms influencing cancer occurrence remain unclear and warrant further investigation through functional and biological experiments.

Conclusions

Our study revealed a noteworthy association between polymorphisms in the core genes of PRC2 and the risk of LC. Our findings indicate that variant alleles of EZH2, EED, and RBBP4 might act as protective factors against LC development and potentially serve as genetic markers associated with susceptibility to LC.

Data Sharing Statement

All data needed to evaluate the conclusions of this study are presented in this article.

Ethics Approval and Consent to Participate

Our study was approved by the Institutional Ethics Committee of the General Hospital of Tianjin Medical University and the Institutional Ethics Committee of the Affiliated Hospital of Inner Mongolia Medical University. This study was conducted in accordance with the Declaration of Helsinki.

Author Contributions

SRK, JC, and HYL designed the research studies. MG, YWL, HH, LCS, and PJC performed the experiments. YGF, RFS, CC, XGL, GSZ, and DW analyzed data. MG, YWL, and HYL wrote the manuscript. JC provided financial support. SRK, YWL, and HYL supervised the project. All authors have jointly participated in the drafting, revision or critical review of the article and final approval of the version to be published. All authors were informed about the purpose of the study and have agreed on the journal to which the article is to be submitted and have agreed to take responsibility for all aspects of the work.

Funding

This work was supported by the National Natural Science Foundation of China (82172569, 82072595, and 61973232), Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-061B and TJWJ2022XK005). Tianjin Health Science and Technology Project (ZC20179) and Beijing Science and Technology Innovation Medical Development Fund (KC2021-JX-0186-57).

Disclosure

The authors have declared that no competing interest exists.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249.
2. Brody H. Lung cancer. *Nature*. 2020;587(7834):S7.
3. Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. *Hum Genet*. 2011;130(1):59–78.
4. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *Tobacco smoke and involuntary smoking*. 2004;1–1438.
5. Czoli CD, Hammond D. TSNA exposure: levels of NNAL among Canadian tobacco users. *Nicotine Tob Res*. 2015;17(7):825–830.
6. Jemal A, Miller KD, Ma J, et al. Higher lung cancer incidence in young women than young men in the United States. *N Engl J Med*. 2018;378 (21):1999–2009.
7. Wheeler DA, Wang L. From human genome to cancer genome: the first decade. *Genome Res*. 2013;23(7):1054–1062.
8. Hamra GB, Guha N, Cohen A, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environ Health Perspect*. 2014;122(9):906–911.
9. Benusiglio PR, Fallet V, Sanchis-Borja M, Coulet F, Cadranel J. Lung cancer is also a hereditary disease. *Eur Respir Rev*. 2021;30(162):564.
10. Fois SS, Paliogiannis P, Zinellu A, Fois AG, Cossu A, Palmieri G. Molecular epidemiology of the main druggable genetic alterations in non-small cell lung cancer. *Int J Mol Sci*. 2021;22(2):34.
11. Shields PG. Molecular epidemiology of smoking and lung cancer. *Oncogene*. 2002;21(45):6870–6876.
12. Shastry BS. SNP alleles in human disease and evolution. *J Hum Genet*. 2002;47(11):561–566.
13. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*. 1998;280(5366):1077–1082.
14. Horn S, Figl A, Rachakonda PS, et al. TERT promoter mutations in familial and sporadic melanoma. *Science*. 2013;339(6122):959–961.
15. Sur I, Tuupanen S, Whitington T, Aaltonen LA. Taipale J: lessons from functional analysis of genome-wide association studies. *Cancer Res*. 2013;73(14):4180–4184.
16. Wang C, Li Y, Li YW, et al. HOTAIR lncRNA SNPs rs920778 and rs1899663 are associated with smoking, male gender, and squamous cell carcinoma in a Chinese lung cancer population. *Acta Pharmacol Sin*. 2018;39(11):1797–1803.
17. Conway E, Healy E, Bracken AP. PRC2 mediated H3K27 methylations in cellular identity and cancer. *Curr Opin Cell Biol*. 2015;37:42–48.
18. van Mierlo G, Veenstra GJC, Vermeulen M, Marks H. The complexity of PRC2 subcomplexes. *Trends Cell Biol*. 2019;29(8):660–671.
19. Jiao L, Liu X. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science*. 2015;350(6258):aac4383.
20. Højfeldt JW, Laugesen A, Willumsen BM, et al. Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat Struct Mol Biol*. 2018;25(3):225–232.
21. Pengelly AR, Copur Ö, Jäckle H, Herzig A, Müller J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science*. 2013;339(6120):698–699.
22. Raspin K, FitzGerald LM, Marthick JR, et al. A rare variant in EZH2 is associated with prostate cancer risk. *Int J Cancer*. 2021;149(5):1089–1099.
23. Liu LC, Chien YC, Wu GW, et al. Analysis of EZH2 genetic variants on triple-negative breast cancer susceptibility and pathology. *Int J Med Sci*. 2022;19(6):1023–1028.
24. Su KJ, Lin CW, Chen MK, Yang SF, Yu YL. Effects of EZH2 promoter polymorphisms and methylation status on oral squamous cell carcinoma susceptibility and pathology. *Am J Cancer Res*. 2015;5(11):3475–3484.
25. Sung H, Yang HH, Zhang H, et al. Common genetic variants in epigenetic machinery genes and risk of upper gastrointestinal cancers. *Int J Epidemiol*. 2015;44(4):1341–1352.
26. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7–33.
27. Byun J, Han Y, Li Y, et al. Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237 controls identifies new susceptibility loci contributing to lung cancer. *Nat Genet*. 2022;54(8):1167–1177.
28. Musolf AM, Moiz BA, Sun H, et al. Whole exome sequencing of highly aggregated lung cancer families reveals linked loci for increased cancer risk on chromosomes 12q, 7p, and 4q. *Cancer Epidemiol Biomarkers Prev*. 2020;29(2):434–442.
29. Liu P, Vikis HG, Wang D, et al. Familial aggregation of common sequence variants on 15q24-25.1 in lung cancer. *J Natl Cancer Inst*. 2008;100 (18):1326–1330.
30. Deb G, Singh AK, Gupta S. EZH2: not EZHY (easy) to deal. *Mol Cancer Res*. 2014;12(5):639–653.
31. Jiang T, Wang Y, Zhou F, Gao G, Ren S, Zhou C. Prognostic value of high EZH2 expression in patients with different types of cancer: a systematic review with meta-analysis. *Oncotarget*. 2016;7(4):4584–4597.
32. Laugesen A, Højfeldt JW, Helin K. Role of the polycomb repressive complex 2 (PRC2) in transcriptional regulation and cancer. *Cold Spring Harb Perspect Med*. 2016;6(9):86.
33. Lohavanichbutr P, Sakoda LC, Amos CI, et al. Common TDP1 polymorphisms in relation to survival among small cell lung cancer patients: a multicenter study from the international lung cancer consortium. *Clin Cancer Res*. 2017;23(24):7550–7557.
34. Wang Z, Wei Y, Zhang R, et al. Multi-omics analysis reveals a HIF network and hub gene EPAS1 associated with lung adenocarcinoma. *EBioMedicine*. 2018;32:93–101.
35. Liu CY, Stücker I, Chen C, et al. Genome-wide gene-asbestos exposure interaction association study identifies a common susceptibility variant on 22q13.31 associated with lung cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2015;24(10):1564–1573.
36. Yu JI, Kang IH, Seo GS, Choi SC, Yun KJ, Chae SC. Promoter polymorphism of the EED gene is associated with the susceptibility to ulcerative colitis. *Dig Dis Sci*. 2012;57(6):1537–1543.