The rs28757157 and rs59429575 polymorphisms in CYP19A1 are associated with lung cancer in Chinese Han population

Chan Zhang
The First People's Hospital of Yunnan Province
Yujing Cheng
The First People's Hospital of Yunnan Province
Wanlu Chen
The Frist People's Hospital of Yunnan Province
Qi Li
The First People's Hospital of Yunnan Province
Run Dai
The Frist People's Hospital of Yunnan Province
Yajie Wang
The Frist People's Hospital od Yunnan Province
Tonghua Yang (tonghua__yang@163.com)
the Frist People's Hospital of Yunnan Province

Research article

Keywords: CYP19A1, Polymorphisms, Lung cancer, Chinese Han population

DOI: https://doi.org/10.21203/rs.3.rs-25248/v3

Abstract

Background: Lung cancer is the leading cause of cancer death globally. Recent studies have revealed that the *CYP19A1* gene played a crucial role in cancer initiation and development. The aim of this study was to assess the association of *CYP19A1* genetic polymorphisms with the risk of lung cancer in the Chinese Han population.

Method: This study randomly recruited 507 lung cancer patients and 505 healthy controls. The genotypes of four SNPs of *CYP19A1* gene were identified by Agena MassARRY technique. Genetic model analysis was used to assess the association between genetic variation and lung cancer risk. Odds ratio (OR) values and 95% confidence intervals (CIs) were provided for the evaluation of the lung cancer risk effect.

Results: Rs28757157 and rs59429575 polymorphisms of *CYP19A1* were significantly correlated with the risk of lung cancer. In stratified analysis, rs28757157 was associated with increased cancer risk in smokers and individuals aged ≤ 60 years. Meanwhile, rs59429575 was identified as a risk biomarker in females and lung adenocarcinoma patients ($p < 0.05$). While rs28757157 exerted a protective role among people with a BMI greater than 24 ($p = 0.033$).

Conclusions: This study identified two new SNPs (rs28757157 and rs59429575) of *CYP19A1* associated with lung cancer in Chinese Han population. These findings provide data support for further functional studies of CYP19A1 in lung cancer.

Background

Lung cancer is a kind of malignant tumor with high morbidity and mortality [1]. In China, this malignant tumor has the highest mortality rate, accounting for about 25% of cancer-related deaths [2]. At present, there are many risk factors were found to increase the risk of lung cancer. Among them, smoking seems to be most strongly associated with lung cancer risk [3]. However, new research shows that about 10% to 15% of people diagnosed with lung cancer are non-smokers, indicating the importance of other risk factors such as exogenous air pollution, environmental factors and genetic factors. According to the latest statistics, about 8% of lung cancer cases are estimated to be caused by genetics alone [4, 5]. The fact that the risk of lung cancer in the immediate family of a patient with lung cancer increased by 2.4 times further confirms the decisive role of genetic factors in disease risk [6].

In addition, among non-smoking related lung cancer patients, women have a higher risk of lung cancer than men, especially adenocarcinoma. The incidence of lung cancer in non-smokers is 15% in women and only 10% in men. These findings have drawn attention to the effects of the estrogen on lung cancer risk [7]. It has been reported that both estrogen receptor and aromatase are present in human lung tumors [8-10]. These results suggest that estrogen may play a role in the biological behavior of human lung cancer.

Cytochrome p450 (CYP450) enzymes are pivotal for biological homeostasis. CYP450 enzymes also play a key role in the metabolism of many endogenous substrates and exogenous carcinogens as well as aromatic and heterocyclic amines. They then covalently combine with DNA to form DNA adducts, which in turn cause cancer [11, 12]. The CYP450 family 19, subfamily A, polypeptide 1 (*CYP19A1*) gene encodes aromatase, which is a member of the CYP450 superfamily of enzymes and a key enzyme in oestradiol biosynthesis. Mutations in the *CYP19A1* gene can result in either increased or decreased aromatase activity, and aromatase plays an important role in lung cancer [13, 14]. This suggests that *CYP19A1* genetic variation may indirectly affect the occurrence of lung cancer, but the exact mechanism is unclear. At the same time, many literatures have reported the inseparable relationship between *CYP19A1* gene and lung-related diseases, including lung cancer [15-17].

Based on previous research, the present study explored the relationship between four single nucleotide polymorphisms (SNPs) of *CYP19A1* gene and lung cancer susceptibility through case-control study, with particular attention to sex and smoking status of the participants.

Methods

Participants

From May 2015 to February 2018, we recruited 507 pathologically confirmed lung cancer patients from the Shaanxi Provincial Cancer Hospital. All cases were diagnosed as adenocarcinoma, squamous cell carcinoma or small-cell lung cancer by histological examination , according to the World Health Organization tumor classification system; diagnoses were confirmed by two independent patholohists. The exclusion criteria of patients were as follows: 1) history of other tumors; 2) family history of lung cancer; 3) chemotherapy or radiotherapy treatment; 4) hypertension, diabetes mellitus, or any endocrine metabolic diseases; 5) other lung diseases. The control group was composed of 505 healthy subjects who were volunteer blood donors from the physical examination center at the same hospital as the cases. Controls with a history of any cancers, other endocrine metabolic diseases, and other lung diseases should be excluded. Eligible study participants were screened by completing a specialized questionnaire, including demographic characteristics, disease history, lung status, and family history of other types of tumors. All participants were Chinese Han ancestry from northwest China.

SNP selection

Based on 1000 genome database (http://www.internationalgenome.org/), the dbSNP database (http://www.bioinfo.org.cn) and the Han Chinese in Beijing (CHB) population data, we screened four SNPs (rs28757157, rs3751592, rs3751591, and rs59429575) in *CYP19A1* of which the minor allele frequencies more than 5% to ensure the valid statistical analysis.

SNP genotyping

Genomic DNA was extracted from collected peripheral blood samples using a DNA purification extraction kit (GoldMag Xi'an, China). The concentration and purity of DNA were determined quantitatively by ultraviolet spectrophotometer (Nanodrop 2000, Thermo, USA). Multiplexed SNP MassEXTEND assay was designed with the Agena Bioscience Assay Design Suite software, version 3.0 (Agena Bioscience, USA). SNP genotyping was conducted utilizing the MassARRAY platform (Agena Bioscience, USA). Data processing carried out with Agena Bioscience TYPER software, version 4.0 (Agena Bioscience, San Diego, CA, USA) [18].

Statistical analysis and Bioinformatics analysis

SPSS software (SPSS 22.0, USA) and Microsoft Excel were used for statistical analysis. The differences in gender and age distribution between the case group and the control group were determined by the χ^2 test and independent sample *t*-test, respectively. χ^2 test was used to determine whether individual polymorphisms were in Hardy-Weinberg equilibrium (HWE). In addition, χ2 test was used to detect the difference of allele and genotype frequencies between the cases and controls. The PLINK software (*http://www.cog-genomics.org/plink2/*) was adopted to define the relationship between the polymorphisms and risk of lung cancer among Chinese Han nationality population in different genetic model analysis (genotype model, dominant model, recessive models and additive model). Logistic regression analysis was used to calculate the odds ratio (OR) and 95% confidence interval (CIs) to evaluate the lung cancer risk effect [19-21]. The $p < 0.05$ was considered statistically significant in all tests. The functionality of candidate SNPs were annotated using the HaploReg v4.1 database. (*https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php*).

Results

Study population

In this study, a random sample of 507 lung cancer patients (353 males and 154 females) was involved as well as 505 healthy control people (354 males and 151 females), with an average age of 60.75 ± 9.98 and 60.40 ± 7.39 (Table 1). There was no significant difference in gender and age distribution between the case group and the control group ($p > 0.05$). In addition, the characteristics of the study population were collected for subsequent studies, including body mass index (BMI), smoking and drinking history, pathological type, pathological stage, and lymph node metastasis (LNM).

Table 1 Characteristics of the study population

	Cases	Controls	*p*
Total	507	505	
Age			0.524†
Mean ± SD	60.75 ± 9.98	60.40 ± 7.39	
≤ 60	236(47%)	235(47%)	
> 60	271(53%)	270(53%)	
Gender			0.891‡
Male	353(70%)	354(70%)	
Female	154(30%)	151(30%)	
BMI			
≤ 24	316(62%)	146(29%)	
> 24	177(35%)	161(32%)	
Absence	74(3%)	198(39%)	
Smoking			
Yes	251(50%)	136(27%)	
No	250(49%)	140(28%)	
Absence	6(1%)	229(45%)	
Drinking			
Yes	114(22%)	109(22%)	
No	356(70%)	135(27%)	
Absence	37(8%)	261(51%)	
Pathological type			
Squamous cell carcinoma.	119(23%)		
Adenocarcinoma	187(37%)		
Absence	201(40%)		
Stage			
I-II	83(16%)		
III-IV	260(51%)		
Absence	146(33%)		
LNM			
Negative	84(17%)		
Positive	214(42%)		
Absence	209(41%)		

SD: standard deviation; BMI: body mass index; LNM: lymph node metastasis.

†*p* values were calculated from independent sample *t*-test.

‡*p* values were calculated from two-sided χ^2 test.

Basic information of the selected SNPs

Four SNPs in *CYP19A1* were genotyped among the subjects. The basic information of all candidate SNPs is listed in Table 2. All SNPs are located on chromosome 15 and in the different position of *CYP19A1* gene. The deviation of Hardy-Weinberg equilibrium in the control group was evaluated, and the results showed that the candidate SNPs all met the expected *p* value ($p > 0.05$), and all SNPs satisfied further study. In addition, under the allele risk model, there was no significant difference in the distribution of alleles of each SNP between the lung cancer cases and the control group. ($p > 0.05$). Functional prediction of SNPs was conducted in HaploReg v4.1 database to explore their regulatory effect. The results showed that the four SNPs exhibited potential biological functions in gene regulation.

Table 2 Basic information of candidate SNPs in the study

SNP ID	Genes	Chr.	Position	Alleles A/B	MAF		p-HWE†	p‡	OR (95%CI)	HaploReg
					Case	Control				
rs28757157	CYP19A1	Chr15	51253204	T/C	0.225	0.208	0.501	0.359	1.11(0.11-0.89)	Enhancer histone marks, Motifs changed, Selected eQTL hits
rs3751592	CYP19A1	Chr15	51314381	C/T	0.135	0.126	1.000	0.532	1.09(0.13-0.84)	Promoter histone marks, Enhancer histone marks, DNAse, Motifs changed, Selected eQTL hits
rs3751591	CYP19A1	Chr15	51314513	G/A	0.143	0.131	0.695	0.458	1.10(0.13-0.85)	Promoter histone marks, Enhancer histone marks, DNAse, Motifs changed
rs59429575	CYP19A1	Chr15	51314880	T/C	0.172	0.161	0.869	0.511	1.08(0.12-0.86)	Promoter histone marks, Enhancer histone marks, DNAse, Proteins bound, Motifs changed

Genetic model analyses of the selected SNPs

Four genetic models analysis for the relationship between *CYP19A*1 polymorphism and risk of lung cancer are listed in Table 3. Our results revealed an association between rs28757157 and increased risk of lung cancer in the genotype model (OR = 1.33, 95% CI: 1.03-1.73, p = 0.032). Rs59429575 was also associated with an increased risk of lung cancer in recessive model (OR = 2.03, 95% CI: 1.00-4.11, p = 0.049).

Table 3 Analysis of association between CYP19A1 polymorphism and risk of lung cancer

SNP ID	Model	Genotype	Case(*N*)	Control(*N*)	Crude		Adjusted†	
					OR(95%CI)	*p*‡	OR(95%CI)	*p*‡
rs28757157	Genotype	CC	276	313	1			
		TC	201	172	1.33(1.02-1.72)	**0.034***	1.33(1.03-1.73)	**0.032***
		TT	9	19	0.54(0.24-1.21)	0.132	0.53(0.24-1.20)	0.130
	Dominant	CC	276	313	1		1	
		TT-TC	210	191	1.25(0.97-1.61)	0.089	1.25(0.97-1.61)	0.085
	Recessive	TC-CC	477	485	1		1	
		TT	9	19	0.48(0.22-1.08)	0.075	0.48(0.21-1.07)	0.072
	Additive	—	—	—	1.12(0.89-1.40)	0.332	1.12(0.89-1.41)	0.324
rs59429575	Genotype	CC	357	354	1			
		TC	126	138	0.91(0.68-1.20)	0.491	0.91(0.68-1.20)	0.496
		TT	24	12	1.98(0.98-4.03)	0.058	1.98(0.97-4.01)	0.060
	Dominant	CC	357	354	1		1	
		TT-TC	150	150	0.99(0.76-1.3)	0.951	0.99(0.76-1.30)	0.955
	Recessive	TC-CC	483	492	1		1	
		TT	24	12	2.04(1.01-4.12)	**0.048***	2.03(1.00-4.11)	**0.049***
	Additive	—	—	—	1.08(0.86-1.35)	0.523	1.08(0.86-1.35)	0.523
rs3751592	Genotype	TT	385	386	1		1	
		CT	107	111	0.97(0.72-1.31)	0.824	0.96(0.71-1.30)	0.815
		CC	15	8	1.88(0.79-4.49)	0.155	1.87(0.78-4.47)	0.158
	Dominant	TT	385	119	1		1	
		CC-CT	122	386	1.03(0.77-1.37)	0.852	1.03(0.77-1.37)	0.864
	Recessive	CT-TT	492	497	1		1	
		CC	15	8	1.89(0.80-4.51)	0.149	1.89(0.79-4.49)	0.152
	Additive	—	—	—	1.08(0.84-1.39)	0.542	1.08(0.84-1.39)	0.553
rs3751591	Genotype	AA	372	377	1		1	
		GA	120	118	1.03(0.77-1.38)	0.839	1.04(0.77-1.39)	0.817
		GG	12	7	1.74(0.68-4.46)	0.251	1.75(0.68-4.49)	0.246
	Dominant	AA	372	125	1		1	
		GG-GA	132	377	1.07(0.81-1.42)	0.639	1.08(0.81-1.43)	0.617
	Recessive	GA-AA	492	495	1		1	
		GG	12	7	1.73(0.67-4.42)	0.256	1.73(0.68-4.44)	0.252
	Additive	—	—	—	1.10(0.85-1.42)	0.458	1.11(0.86-1.43)	0.440

†Adjusted for age and sex in a logistic regression model.

‡*p* values were calculated from logistic regression.

*Bold values indicate statistical significance ($p < 0.05$).

In addition, we conducted a stratified analysis to explore the risk effects of these SNPS in specific groups of people. Stratified analysis of the clinical characteristics of rs28757157 polymorphism are shown in Table 4. The results indicated that rs28757157 heterozygote genotype (TC) was associated with increased susceptibility to lung cancer in people under 60 years of age (OR = 1.6, 95% CI: 1.09-2.35, $p = 0.016$). While the TT genotype of rs28757157 exerted a protective role in the development of lung cancer among people under 60 years and people with a BMI greater than 24.(age ≤ 60: OR = 0.11, 95% CI: 0.01-0.87, $p = 0.036$; BMI > 24: OR = 0.19, 95% CI: 0.04-0.87, $p = 0.033$). In addition, rs28757157 was identified as a genetic risk factor for lung cancer susceptibility in the allele, dominant, and additive models in the smoking population (allele model: OR =

1.48, 95% CI: 1.00-2.18, p = 0.048; dominant model: OR = 1.58, 95% CI: 1.01-2.47, p = 0.045; additive model: OR = 1.57, 95% CI: 1.03-2.4, p = 0.036).

Table 4 Stratification analyses by clinical characteristics about the rs28757157 polymorphism

Variables		p†, OR (95% CI)					
		Allele	Homozygote	Heterozygote	Dominant	Recessive	Additive
Age	≤ 60	0.521, 1.11(0.81-1.50)	0.059, 1.14(0.92-2.08)	**0.016, 1.60(1.09-2.35)**	0.051, 1.45(0.99-2.12)	**0.036, 0.11(0.01-0.87)**	0.283, 1.21(0.86-1.70)
	> 60	0.542, 1.10(0.82-1.49)	0.959, 0.97(0.36-2.64)	0.367, 1.19(0.82-1.72)	0.837, 1.17(0.81-1.67)	0.863, 0.92(0.34-2.46)	0.498, 1.12(0.81-1.53)
BMI	≤ 24	0.221, 1.26(0.89-1.80)	0.516, 1.70(0.34-8.46)	0.215, 1.30(0.86-1.98)	0.188, 1.32(0.87-1.99)	0.596, 1.54(0.31-7.60)	0.177, 1.30(0.89-1.91)
	> 24	0.980, 1.01(0.70-1.44)	0.052, 0.21(0.04-1.02)	0.133, 1.42(0.90-2.25)	0.351, 1.24(0.79-1.93)	**0.033, 0.19(0.04-0.87)**	0.973, 1.01(0.68-1.48)
Smoking	Yes	**0.048, 1.48(1.00-2.18)**	0.283, 3.27(0.38-28.49)	0.062, 1.54(0.98-2.41)	**0.045, 1.58(1.01-2.47)**	0.348, 2.81(0.32-24.30)	**0.036, 1.57(1.03-2.40)**
	No	0.736, 0.39(1.17-0.82)	0.677, 0.74(0.17-3.11)	0.163, 1.37(0.88-2.12)	0.204, 1.32(0.86-2.04)	0.557, 0.65(0.16-2.72)	0.315, 1.23(0.82-1.82)
Drinking	Yes	0.293, 1.30(0.80-2.11)	0.585, 0.51(0.04-5.83)	0.142, 1.54(0.87-2.73)	0.180, 1.47(0.84-2.60)	0.504, 0.44(0.04-4.98)	0.272, 1.35(0.79-2.30)
	No	0.320, 1.19(0.84-1.69)	0.863, 0.90(0.26-3.08)	0.169, 1.34(0.88-2.04)	0.201, 1.31(0.87-1.97)	0.725, 0.80(0.24-2.73)	0.292, 1.22(0.84-1.77)

BMI: body mass index.
Allele: T > C; Homozygote: TT; Heterozygote: TC; Dominant: CC vs TT-TC; Recessive: TC-CC vs TT.
†p values were calculated from logistic regression.
Bold values indicate statistical significance ($p < 0.05$).

Stratification analyses by clinical characteristics of the rs59429575 polymorphism are summarized in Table 5. Stratified analysis by gender demonstrated a remarkable relationship between enhanced lung cancer risk and the TT genotype of rs59429575 (OR = 5.18, 95% CI: 1.12-24.05, p = 0.036). In addition, in the patients with lung adenocarcinoma, rs59429575 was identified as a risk factor for lung cancer development (Homozygote: OR = 2.37, 95% CI: 1.01-5.56, p = 0.047; recessive model: OR = 2.55, 95% CI: 1.09-5.95, p = 0.03).

Table 5 Stratification analyses by clinical characteristics about the rs59429575 polymorphism
LUAD: Lung Adenocarcinoma; LUSC: Lung Squamous cell carcinoma; LNM: lymph node metastasis.
Allele: T > C; Homozygote: TT; Heterozygote: TC; Dominant: CC vs TC-TT; Recessive: TC-CC vs TT.
†p values were calculated from logistic regression.
Bold values indicate statistical significance ($p < 0.05$).

Discussion

In this study, we examined four SNPs of *CYP19A1* gene in correlation to the susceptibility of lung cancer in a Chinese Han cohort. Statistical and bioinformatics results revealed the important roles of rs28757157 and rs59429575 in the outset of lung cancer in total or stratified population, which helped improve our understanding of *CYP19A1* in this disease.

ıbles		$p^{\dagger}$, OR (95% CI)					
		Allele	Homozygote	Heterozygote	Dominant	Recessive	Additive
er	Male	0.514, 1.10(0.83-1.46)	0.417, 1.41(0.61-3.24)	0.834, 1.04(0.74-1.46)	0.663, 1.08(0.78-1.49)	0.427, 1.40(0.61-3.19)	0.523, 1.09(0.83-1.44)
	Female	0.839, 1.04(0.69-1.58)	0.051, 4.64(0.99-21.68)	0.738, 0.68(0.41-1.13)	0.464, 0.84(0.52-1.35)	**0.036, 5.18(1.12-24.05)**	0.843, 1.04(0.70-1.56)
ɔr	LUAD	0.729, 1.06(0.77-1.46)	**0.047, 2.37(1.01-5.56)**	0.160, 0.75(0.50-1.12)	0.488, 0.88(0.6-1.28)	**0.030, 2.55(1.09-5.95)**	0.873, 1.03(0.75-1.41)
	LUSC	0.367, 1.18(0.82-1.71)	0.212, 1.96(0.68-5.65)	0.701, 1.10(0.69-1.75)	0.476, 1.18(0.75-1.83)	0.226, 1.91(0.67-5.47)	0.311, 1.21(0.83-1.77)
	(–)	1.00	1.00	1.00	1.00	1.00	1.00
	(+)	0.256, 1.34(0.81-2.22)	0.365, 2.04(0.44-9.60)	0.621, 1.17(0.64-2.13)	0.450, 1.25(0.70-2.23)	0.389, 1.97(0.42-9.18)	0.349, 1.26(0.78-2.06)
ɛ	I~II	1.00	1.00	1.00	1.00	1.00	1.00
	III~IV	0.122, 1.48(0.90-2.44)	0.077, 6.31(0.82-48.41)	0.957, 1.02(0.57-1.83)	0.410, 1.27(0.72-2.23)	0.077, 6.28(0.82-48.00)	0.153, 1.41(0.88-2.24)

CYP19A1 gene encoding aromatase responsible for the final step in the biosynthesis of estrogens, estradiol (E2) and estrone (E1), have been intensively investigated [22, 23]. SNPs in the intron region of *CYP19A1* have been identified to play an important role in the transcriptional regulation and splicing of *CYP19A1*, producing some different enzymes with diverse enzyme activity compared with normal gene products. Variations in the allele frequency of several *CYP19A1* SNPs have been documented in different populations and ethnic groups around the world. SNPs in *CYP19A1* were found to be associated with cancer risk in European Americans, North Indians, and Chinese [24, 25]. In particular, *CYP19A1* SNPs have been shown to be significantly associated with lung-related diseases[16,17].

A previous study has shown that SNP rs3764221 was significantly correlated with *CYP19A1* expression in non-cancerous lung tissues, and affects the susceptibility of lung adenocarcinoma. The authors suggested that *CYP19A1* polymorphisms may lead to elevated levels of local estrogen in the surrounding lung, and that this local excess estrogen production may be one of the factors associated with the polycentric development of adenocarcinoma [15]. Recent result has suggested that *CYP19A1* polymorphism is involved in lung bronchioloalveolar carcinoma and atypical adenomatous hyperplasia risk by causing differences in estrogen levels [16]. It is clear that *CYP19A1* polymorphism may cause changes in estrogen levels around the lungs, which in turn can affect the susceptibility of lung cancer.

Zhang et al. have shown that *CYP19A1*-rs727479 was significantly associated with the incidence of lung cancer. In particular, haplotype rs727479|rs730154|rs10046-ACA was only significantly associated with nonsmokers and female nonsmokers [17]. These results indicated that *CYP19A1* polymorphism is significantly associated with the susceptibility to lung cancer, and is gender-specific. Our research showed that *CYP19A1*-rs28757157 was associated with increased cancer risk in smokers and population under 60 years, and *CYP19A1*-rs59429575 was identified as risk biomarker in females and lung adenocarcinoma patients. These two SNPs are located in the intron region of the *CYP19A1* gene. Combined with previous studies and database predictions, we speculated that *CYP19A1* intron single nucleotide polymorphisms may alter mRNA splicing, leading to changes in the activity of *CYP19A1* and related estrogens, and may affect disease susceptibility. Since the statistical significance of the correlation between *CYP19A1* gene polymorphism and the risk of lung cancer is slightly weak, further experimental studies are needed to verify the results of this study.

Our study has several limitations. All subjects were enrolled from the same hospital and the limitations of sample selection may affect the accuracy of this experiment. Additional studies that encompass more geographical regions, additional ethnic groups, and larger sample size should be performed. In order to verify the results of this study, it is necessary to clarify the relationship between *CYP19A1* gene and lung cancer through subsequent functional studies.

Conclusions

In summary, our study defined two new SNPs of *CYP19A1* (rs28757157 and rs59429575) that were significantly associated with lung cancer susceptibility. These variants may be considered as markers in lung cancer risk assessment for Chinese Han population.

Abbreviations

Cytochrome p450, CYP450; The CYP450 family 19, subfamily A, polypeptide 1, *CYP19A1*; single nucleotide polymorphisms, SNPs; body mass index, BMI; lymph node metastasis, LNM.

Declarations

Ethical approval and consent to participate

Our study was conducted with appropriate approval of ethnics committee from the Shaanxi Provincial Cancer Hospital. All procedures performed in this study were in accordance with the ethical standards of the ethics committee from the Shaanxi Provincial Cancer Hospital and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from each participant at recruitment after fully describing our research to them.

Consent for publication

Not applicable

Availability of data and materials

The datasets generated during the current study are available.

Competing interests

The authors declare that they have no competing interests.

Funding

NA.

Authors' Contributions

TY designed the experiment, CZ, YC, and WC performed the experiment, QL and RD processed the data, CZ wrote the manuscript, YW revised the manuscript. All authors have read and approved the manuscript.

Acknowledgments

We thank all auta

Ethical approval and consent to participate

Our study was conducted with appropriate approval of ethnics committee from the Shaanxi Provincial Cancer Hospital. All procedures performed in this study were in accordance with the ethical standards of the ethics committee from the Shaanxi Provincial Cancer Hospital and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from each participant at recruitment after fully describing our research to them.

Consent for publication

Not applicable

Availability of data and materials

The datasets generated during the current study are available.

Competing interests

The authors declare that they have no competing interests.

Funding

NA.

Authors' Contributions

TY designed the experiment, CZ, YC, and WC performed the experiment, QL and RD processed the data, CZ wrote the manuscript, YW revised the manuscript. All authors have read and approved the manuscript.

Acknowledgments

We thank all authors for their contributions and supports. We are also grateful to all participants for providing blood samples.

hors for their contributions and supports. We are also grateful to all participants for providing blood samples.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.
2. Chen W, Zheng R, Zeng H, Zhang S, He J. Annual report on status of cancer in China, 2011. Chinese journal of cancer research = Chung-kuo yen cheng yen chiu. 2015;27(1):2-12.
3. Polanski J, Jankowska-Polanska B, Rosinczuk J, Chabowski M, Szymanska-Chabowska A. Quality of life of patients with lung cancer. OncoTargets and therapy. 2016;9:1023-8.
4. Rahal Z, El Nemr S, Sinjab A, Chami H, Tfayli A, Kadara H. Smoking and Lung Cancer: A Geo-Regional Perspective. Frontiers in oncology. 2017;7:194.
5. El-Telbany A, Ma PC. Cancer genes in lung cancer: racial disparities: are there any? Genes & cancer. 2012;3(7-8):467-80.
6. Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e1S-e29S.
7. Subramanian J, Govindan R. Lung cancer in never smokers: a review. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25(5):561-70.
8. Raso MG, Behrens C, Herynk MH, Liu S, Prudkin L, Ozburn NC, et al. Immunohistochemical expression of estrogen and progesterone receptors identifies a subset of NSCLCs and correlates with EGFR mutation. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(17):5359-68.
9. Abe K, Miki Y, Ono K, Mori M, Kakinuma H, Kou Y, et al. Highly concordant coexpression of aromatase and estrogen receptor beta in non-small cell lung cancer. Human pathology. 2010;41(2):190-8.
10. Mah V, Seligson DB, Li A, Marquez DC, Wistuba, II, Elshimali Y, et al. Aromatase expression predicts survival in women with early-stage non small cell lung cancer. Cancer research. 2007;67(21):10484-90.
11. Gonzalez FJ, Gelboin HV. Role of human cytochromes P450 in the metabolic activation of chemical carcinogens and toxins. Drug metabolism reviews. 1994;26(1-2):165-83.
12. Bartsch H, Nair U, Risch A, Rojas M, Wikman H, Alexandrov K. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. Cancer epidemiology, biomarkers & prevention : a publication of

the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2000;9(1):3-28.

13. Timofeeva MN, Kropp S, Sauter W, Beckmann L, Rosenberger A, Illig T, et al. CYP450 polymorphisms as risk factors for early-onset lung cancer: gender-specific differences. Carcinogenesis. 2009;30(7):1161-9.
14. Verma MK, Miki Y, Sasano H. Aromatase in human lung carcinoma. Steroids. 2011;76(8):759-64.
15. Ikeda K, Shiraishi K, Eguchi A, Osumi H, Matsuishi K, Matsubara E, et al. Association of a genetic variant of CYP19A1 with multicentric development of lung adenocarcinomas. Annals of surgical oncology. 2014;21(3):939-45.
16. Kohno T, Kakinuma R, Iwasaki M, Yamaji T, Kunitoh H, Suzuki K, et al. Association of CYP19A1 polymorphisms with risks for atypical adenomatous hyperplasia and bronchioloalveolar carcinoma in the lungs. Carcinogenesis. 2010;31(10):1794-9.
17. Zhang J, Yin Y, Niu XM, Liu Y, Garfield D, Chen SF, et al. CYP19A1 gene polymorphisms and risk of lung cancer. The Journal of international medical research. 2013;41(3):735-42.
18. Liang Y, Thakur A, Gao L, Wang T, Zhang S, Ren H, et al. Correlation of CLPTM1L polymorphisms with lung cancer susceptibility and response to cisplatin-based chemotherapy in a Chinese Han population. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2014;35(12):12075-82.
19. Gao L, Thakur A, Liang Y, Zhang S, Wang T, Chen T, et al. Polymorphisms in the TERT gene are associated with lung cancer risk in the Chinese Han population. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2014;23(6):497-501.
20. Hu QY, Jin TB, Wang L, Zhang L, Geng T, Liang G, et al. Genetic variation in the TP63 gene is associated with lung cancer risk in the Han population. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2014;35(3):1863-6.
21. Xun X, Wang H, Yang H, Wang H, Wang B, Kang L, et al. CLPTM1L genetic polymorphisms and interaction with smoking and alcohol drinking in lung cancer risk: a case-control study in the Han population from northwest China. Medicine. 2014;93(28):e289.
22. Simpson ER, Clyne C, Rubin G, Boon WC, Robertson K, Britt K, et al. Aromatase--a brief overview. Annual review of physiology. 2002;64:93-127.
23. Olson SH, Bandera EV, Orlow I. Variants in estrogen biosynthesis genes, sex steroid hormone levels, and endometrial cancer: a HuGE review. American journal of epidemiology. 2007;165(3):235-45.
24. Chattopadhyay S, Siddiqui S, Akhtar MS, Najm MZ, Deo SV, Shukla NK, et al. Genetic polymorphisms of ESR1, ESR2, CYP17A1, and CYP19A1 and the risk of breast cancer: a case control study from North India. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2014;35(5):4517-27.
25. Cai Q, Kataoka N, Li C, Wen W, Smith JR, Gao YT, et al. Haplotype analyses of CYP19A1 gene variants and breast cancer risk: results from the Shanghai Breast Cancer Study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17(1):27-32.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.pdf