Gene polymorphism of cytochrome P450 significantly affects lung cancer susceptibility

Meng Li¹² | Anqi Li¹² | Ruiqing He¹² | Wenhui Dang¹² | Xinyu Liu¹² | Tian Yang¹² | Puyu Shi¹² | Xiang Bu¹² | Dan Gao¹² | Ning Zhang³ | Shuli Du⁴ | Tianbo Jin⁴ | Mingwei Chen¹²

¹The Department of Respiratory and Critical Care Medicine, First Affiliated Hospital of Xi’an Jiaotong University, Xi’an, Shaanxi, China
²Shaanxi Provincial Research Center for the Project of Prevention and Treatment of Respiratory Diseases, Xi’an, Shaanxi, China
³The Department of Clinical Laboratory, First Affiliated Hospital of Xi’an Jiaotong University, Xi’an, Shaanxi, China
⁴Ministry of Education Key Laboratory of Resource Biology and Biotechnology in Western China, Northwest University, Xi’an, Shaanxi, China

Correspondence
Mingwei Chen, PhD., The Department of Respiratory and Critical Care Medicine, the First Affiliated Hospital, Xi’an Jiaotong University, 277 West Yanta Road, Xi’an, Shaanxi 710061, People’s Republic of China.
Email: chenmw36@163.com

Funding information
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract
Background: Cytochrome P450 (CYPs) are heme proteins involved in the metabolism of a variety of endogenous and exogenous substances and play an important role in the carcinogenesis mechanisms of environmental and hereditary factors. The objective of this study was to investigate how polymorphisms of CYPs correlate with lung cancer (LC) susceptibility.

Methods: Six single nucleotide polymorphisms (SNPs) were genotyped in this study. The chi-square test and unconditional logistic regression model were used to evaluate the correlation between SNPs and LC susceptibility. The expressions and survival data of genes in patients with LC were mined using Oncomine and Kaplan-Meier Plotter database.

Results: Four SNPs were found to be significantly associated with the risk of LC development (P < 0.05). The most significant correlation was that the A allele and AA genotype of CYP2D6 rs1065852 were associated with increased risk of LC development (adjusted odds ratio [OR] = 1.35, 95% confidence interval [95%CI] = 1.13–1.60, P = 9.04e–4; OR = 1.83, 95%CI = 1.29–2.59, P = 0.001 respectively). Similar association of this variant was also found in the subgroups of male patients, cases in III-IV stages, positive lymph node, squamous cell carcinomas and adenocarcinomas. Whereas rs1065852 was considered as protective factor in females (adjusted OR = 0.33, 95% CI = 0.16–0.70, P = 0.004). In stratified analyses, the association of CYP24A1 rs2762934, CYP24A1 rs6068816, CYP20A1 rs2043449 polymorphism with LC risk appeared stronger in some subgroups. CYP2D6, CYP24A1 and CYP20A1 are overexpressed in some pathological types of LC (P < 0.05), and high levels of CYP2D6 and CYP20A1 indicate poor and good prognosis of LC, respectively.

Conclusion: This study revealed that rs1065852, rs2043449, rs2762s934, and rs6068816 of CYPs were associated with LC susceptibility in the Northwestern
1 | INTRODUCTION

Cancer constitutes a burden all over the world.\(^1\) It is estimated that nearly half of the new cases and more than half of the cancer deaths in the world will occur in Asia in 2018.\(^2\) Lung cancer (LC) is the most common cancer which accounts for 11.6% of cancer cases, and is the leading cause of male patient deaths which account for 18.4% of cancer death, especially in East Asia and Polynesia.\(^3,4\) Cigarette use remains the primary causal agent of LC,\(^4\) however, other susceptibility factors such as ionizing radiation, air pollution, and exposure to occupational and environmental carcinogens, such as radon and formaldehyde could also increase the incidence of LC.

Nowadays, an increasing number of studies show a strong link between genetic factors and carcinogenesis.\(^5,7\) Genome-wide association studies (GWAS) have been identified several susceptibility gene locus of cancer in European people, including CHRNA3/5, CHRNB4, BRCA2, CHEK2, TERT, but only a small part of LC genetic capacity can be proved by these gene loci, and most have not been systematically verified in Asian populations.\(^8-10\) Since it is the ethnic group with the largest population in East Asia, it is crucial to explore the relationship between genetic polymorphisms and susceptibility of LC in the Chinese Han population.

The cytochrome P450 superfamily (CYPs), located primarily in the liver, small intestine and kidney,\(^11\) is a large superfamily of integral membrane conserved proteins present in animals, plants, and microorganisms,\(^12\) which play a crucial role in the metabolism and activation of carcinogens.\(^13\) All these active carcinogens can combine with DNA and form DNA adducts which are capable of inducing mutations and initiating tumorigenesis. Genetic polymorphisms of CYPs have been reported to be associated with various diseases and adverse drug reactions among different populations by affecting the enzyme catalytic activity.\(^14,15\) Kiyohara C has found that the CYPs genetic polymorphism is related to the susceptibility of colorectal cancer.\(^16\) Maurya et al reported that polymorphisms of drug metabolizing CYPs showed modest associations with head and neck squamous cell carcinoma risk.\(^17\) Genetic polymorphisms have been reported for CYPs involved in the metabolic activation of polycyclic aromatic hydrocarbons (PAHs) and tobacco-specific nitrosamines,\(^18,19\) both of which are wide spreading environmental procarcinogens that induce LC and skin carcinoma.\(^20,22\) However, Kiyohara C et al have found no significant association between the genetic polymorphism of enzymes involved in xenobiotic metabolism and the risk of LC.\(^23\)

To sum up, the correlation of CYPs polymorphisms and LC risk is contradictory and inconclusive due to the diversity of ethnicity and sample size in study groups. In order to validate the association between genetic polymorphisms of CYPs and susceptibility to LC in Northwest Chinese Han population, we adopted a case-control study and selected six SNPs associated with cancer from the target enzyme system to genotype and evaluate the impact of CYPs genetic polymorphisms on the risk of LC development in general and different subgroups concerning gender, tumor stages, lymph node status, and pathologies. The gene expression and relationship between the expression level and prognosis of LC were further analyzed using Oncomine and Kaplan-Meier plotter database.

2 | MATERIALS AND METHODS

2.1 | Subject and ethics statement

Five hundred and ten Pathologically confirmed LC patients hospitalized in the First Affiliated Hospital of Xi’an Jiaotong University, Shaanxi, China, were included in this study (both SCLC and NSCLC were included). Tumor stages and pathological classifications were based on the 8th edition of TNM staging system published by the Union for International Cancer Control and pathological results respectively.\(^24\) Relevant information was extracted from medical files. Patients with other tumors and communication problems were excluded. Five hundred and four healthy subjects were recruited into the cancer-free control group in the same hospital at the same time, none of them had any history of cancers, severe endocrine and autoimmune diseases. It was made sure that there was no genetic relationship between the cases and the control subjects, the purpose of which was to minimize the environmental, hereditary and therapeutic factors affecting genetic susceptibility to LC. This study strictly complies with the Helsinki declaration of the World Medical Association. The cases and the subjects of the control group provided consent and the research was approved by the Ethics Committee of The First Affiliated Hospital of Xi’an Jiaotong University.

2.2 | SNPs selection and primer design

Six SNPs from three genes of CYPs associated with LC were selected for analysis in this study based on 1000 genome...
projects. Each of them met the criteria of the minimum allele frequency (MAF), more than 5%, in the HapMap of the Chinese Han Beijing population. All primers were designed using ASSAY DESIGN SUITE V2.0. (http://agenacx.com/online-tools, Table 1).

2.3 | SNPs genotyping and haplotype analysis

Genomic DNA was extracted from peripheral blood using GoldMag-Mini Whole Blood Genomic DNA Purification Kits (GoldMag Co. Ltd., Xi’an City, China), and quantified with a spectrophotometer (NanoDrop 2000; Thermo Fisher Scientific, Waltham, MA, United States). To have sufficient DNA for further reactions, polymerase chain reaction (PCR) was applied to each sample. Then SAP purification was performed to remove the remaining dNTP and amplified primers in PCR products. Using a MassARRAY Nanodispenser (Agena Bioscience, San Diego, CA), standardized genotyping reactions were dispensed onto a 384-well spectroCHIP. The repeated control samples were set in every genotyping plate and the concordance was more than 99%. The genotyping of these SNPs was carried out on the MassARRAY iPLEX (Agena Bioscience, San Diego, CA) platform using the allele-specific matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF-MS). Genotyping results were output by Agena Bioscience TYPER software, version 4.0. Haplovie software package (version 4.2) was used to analyze the linkage disequilibrium (LD), haplotype construction, genetic association at polymorphism loci and haplotype blocks were defined according to the criteria laid out by Gabriel and others.25

2.4 | Statistical analysis

Data were analyzed by using SPSS 18.0 statistical software (SPSS Inc, Chicago, IL) and Microsoft Excel (Microsoft Corp., Redmond, WA). All continuous data are expressed as means ± standard deviations (SDs). Pearson’s χ² test and t test were used to compare the distribution of categorical variables and continuous variables between the cases and controls respectively.

Supposing the lower frequency alleles were coded as the minor allele. Frequencies of all SNPs in both case and control groups were evaluated for the test of Hardy–Weinberg Equilibrium (HWE). The three genetic models (dominant, recessive and additive) were applied using PLINK software (http://www.cog-genomics.org/plink2/) to assess the association of single SNPs with the risk of LC development. The odds ratios (ORs) and 95% confidence intervals (95% CIs) were calculated by using logistic regression analysis and were adjusted for age and gender. To explore the possibility that the effect of a genetic polymorphism in candidate genes may
be biologically active only in some specific subgroups, we conducted stratified analyses investigating the effect of genotype within the gender, lymph node status, tumor stage and histological subtypes based on medical reports. Exploratory analyses examining the effect of genetic polymorphisms within the histological subtypes based on pathology reports were also conducted. Statistical significance was identified at \( P \leq 0.05 \) (two-side). Power and Sample Size (PS) Calculation software (http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize) was used to calculate the power of the significant difference.\(^{26}\) We estimated the association of haplotype with the susceptibility to LC using PLINK software (http://www.cog-genomics.org/plink2/). The ORs and 95% CI were also calculated using unconditional logistic regression analyses adjusted for age and sex.

### 2.5 Gene expression and survival analysis

Expression and survival data of candidate genes in LC patients were mined using ONCOMINE (https://www.oncomine.org/resource/login.html) and Kaplan-Meier Plotter (http://kmplot.com/analysis/) database. The Kaplan-Meier method and Cox regression were performed to construct survival curves and estimate hazard ratios (HRs) to assess the relationship between risk genes expression and prognosis of LC.

### 3 RESULTS

#### 3.1 Baseline characteristics

A total of 1014 participants were included in the study, 510 patients with LC in the case group (384 males and 126 females; average age: 58.08 ± 10.55 years) and 504 healthy subjects in the control group (381 males and 123 females; average age: 57.27 ± 10.85 years). Characteristics of patients in the case group and the subjects in the control group included in this study are listed in Table 2. There was no significant difference in distribution of gender and age between the two groups \((P = 0.911; 0.227 \text{ respectively})\).

#### 3.2 Linkage between candidate gene polymorphisms and LC

Six SNPs of CYPs were identified. The success ratio was >99.40% for all SNPs. Primary information of the candidate SNPs is shown in Table 3. No significant deviation of genotype frequencies for CYPs from the HWE was found in both groups (Table 3). The A allele of rs1065852 in CYP2D6 was associated with a 0.35 times increased risk of LC development in the allelic model analysis with power values of 0.937 (adjusted OR = 1.35, 95% CI = 1.13-1.60, \( P = 9.04e-4 \)).

| Table 2 Characteristics of cases and controls in the study |
|----------------------------------------------------------|
| **Case (N = 510)** | **Control (N = 504)** | **P-value** |
| **Sex** | | |
| Male | 384 | 75.3 | 381 | 75.6 | 0.911\(^a\) |
| Female | 126 | 24.7 | 123 | 24.4 | |
| **Age (y)** | | |
| Mean ± SD | 58.08 ± 10.55 | 57.27 ± 10.85 | 0.227\(^b\) |
| **TNM stage** | | |
| I-II | 129 | 25.3 | |
| III-IV | 248 | 48.6 | |
| Miss | 133 | 26.1 | |
| **Lymph node status** | | |
| Positive | 193 | 37.8 | |
| Negative | 120 | 23.5 | |
| Unknown | 197 | 38.6 | |
| **Type of lung cancer** | | |
| SCC | 169 | 33.1 | |
| AC | 161 | 31.6 | |
| SCLC | 97 | 19.0 | |
| Others | 22 | 4.3 | |
| Miss | 61 | 12.0 | |

**Notes:** \( P \leq 0.05 \) indicates statistical significance; \( P^a \) values were calculated from two-side chi-square test; \( P^b \) values were calculated by \( t \) tests; Miss indicates data loss.

**Abbreviations:** AC, adenocarcinoma; SCC, Squamous cell carcinoma; SCLC, small cell lung cancer; TNM, tumor-lymph node-metastasis.

#### 3.3 Linkage between candidate SNPs and LC development in genetic models and Haplotype analysis

We further conducted logistic regression analysis tests to analyze model associations. For SNP rs1065852 in CYP2D6, the genotype frequency distributions were different between the case group and control group \((P = 0.003, \text{ Table 4})\). We ascertained that the AA genotype of rs1065852 in CYP2D6 has a risk effect of promoting LC development compared with other genotypes. Homozygous mutations (AA) of rs1065852 in CYP2D6 increased the LC susceptibility by 1.83 times with power values of 0.998 (adjusted OR = 1.83, 95% CI = 1.29-2.59, \( P = 0.001 \)) compared with individuals carrying the wild-type (GG). The rs1065852 polymorphism significantly increased the risk of LC development in all three genetic models (Dominant, AA/AG vs GG, adjusted OR = 1.36, 95% CI = 1.03-1.80, \( P = 0.030 \); Recessive, AA vs GG/AG, adjusted OR = 1.64, 95% CI = 1.22-2.20, \( P = 0.001 \); Addictive, GG vs AA, adjusted OR = 1.34, 95% CI = 1.01-1.79, \( P = 0.040 \)).
CI = 1.13-1.60, P = 9.04e-4. No statistically significant difference in the haplotype distributions between the case group and control group was observed for CYP24A1 (P > 0.05, Figure S1, Table S1).

### Table 3 Candidate SNPs tested in cancer-related genes

| SNP ID     | Chr | Gene (S) | Position       | Allele A/B | MAF Case Control | P-HWE | Call rate (%) | OR (95%CI) | P-adj① |
|------------|-----|----------|----------------|------------|------------------|-------|---------------|------------|--------|
| rs2043449  | 2   | CYP20A1  | 203,251,967    | C/T        | 0.08             | 0.07  | 0.721         | 1.25 (0.90-1.74) | 0.179   |
| rs2762934  | 20  | CYP24A1  | 54,154,722     | A/G        | 0.09             | 0.11  | 0.818         | 0.82 (0.61-1.09) | 0.175   |
| rs1570669  | 20  | CYP24A1  | 54,157,888     | A/G        | 0.39             | 0.39  | 0.453         | 1.01 (0.85-1.21) | 0.879   |
| rs6068816  | 20  | CYP24A1  | 54,164,552     | T/C        | 0.35             | 0.33  | 0.686         | 1.11 (0.92-1.33) | 0.274   |
| rs296241   | 20  | CYP24A1  | 54,169,680     | A/G        | 0.44             | 0.44  | 0.786         | 0.98 (0.83-1.17) | 0.852   |
| rs1065852  | 22  | CYP2D6   | 42,130,692     | A/G        | 0.52             | 0.44  | 0.928         | 1.35 (1.13-1.60) | 0.94-④ | 0.937   |

Notes: P-adj were adjusted by age and sex; *P ≤ 0.05 value indicates statistical significance.

Abbreviation: CI, confidence interval; HWE, Hardy-Weinberg Equilibrium; MAF, minor allele frequency; ORs, odds ratios; SNP, single nucleotide polymorphism.

### 3.4 Linkage between candidate gene polymorphisms and LC in stratification analysis

As shown in Table 6, the AG genotype of CYP24A1 rs2762934 was associated with decreased risk of LC development (adjusted OR = 0.71, 95% CI = 0.51-1.00, P = 0.048). No statistically significant difference in the haplotype distributions between the case group and control group was observed for CYP24A1 (P > 0.05, Figure S1, Table S1).

As shown in Table 7, the AG genotype of CYP24A1 rs2762934 was associated with decreased LC risk in males (adjusted OR = 0.68; 95%CI = 0.46-0.99, P = 0.046). A similar result was obtained in the recessive model of males (adjusted OR = 0.58, 95%CI = 0.46-0.99, P = 0.046). We also identified that the additive model of Female was associated with decreased LC risk in females (adjusted OR = 0.33, 95%CI = 0.16-0.70, P = 0.004), with all power values more than 0.8 except in the recessive model of Female. In addition, similar results were also found in some genetic models of part of subgroups (Table 7).

Power calculations confirm that the sample size was large enough to discover the differences among cases and controls in candidate SNPs because the power values were more than 0.8 except in some genetic models in stratified analysis (Tables 3-8). However, no significant association was observed between other SNPs and LC in stratification analysis (Tables S2 and S3).
### Table 4

| Genotype     | Case % | Control % | P     | Crude OR (95%CI) | Adjusted OR (95%CI) | Stud power |
|--------------|--------|-----------|-------|------------------|---------------------|------------|
| rs2043449    |        |           |       |                  |                     |            |
| C/C          | 6      | 1.18      | 0.378 | 0.037            | 0.204               | 2.04       |
| C/T          | 74     | 15.5      | 0.6   | 0.21             | 1.20                | 1.19       |
| T/T          | 430    | 84.3      | 0.869 | 1.00             | 1.00                | 1.00       |
| rs2762934    |        |           |       |                  |                     |            |
| A/A          | 9      | 1.8       | 0.105 | 0.147            | 0.71                | 0.71       |
| A/G          | 73     | 14.3      | 0.06  | 0.71             | 0.71                | 0.71       |
| G/G          | 428    | 83.9      | 0.798 | 1.00             | 1.00                | 1.00       |
| rs1570669    |        |           |       |                  |                     |            |
| A/A          | 83     | 16.1      | 0.428 | 0.428            | 0.428               | 0.428      |
| A/G          | 232    | 45.6      | 0.492 | 0.492            | 0.492               | 0.492      |
| G/G          | 195    | 38.3      | 0.367 | 1.00             | 1.00                | 1.00       |
| rs6068816    |        |           |       |                  |                     |            |
| T/T          | 56     | 11.1      | 0.455 | 0.455            | 0.455               | 0.455      |
| T/C          | 245    | 48.4      | 0.452 | 0.452            | 0.452               | 0.452      |
| C/C          | 205    | 40.5      | 0.444 | 1.00             | 1.00                | 1.00       |
| rs2296241    |        |           |       |                  |                     |            |
| A/A          | 96     | 18.9      | 0.981 | 0.981            | 0.981               | 0.981      |
| A/G          | 253    | 49.7      | 0.50  | 0.50             | 0.50                | 0.50       |
| G/G          | 160    | 31.4      | 0.39  | 1.00             | 1.00                | 1.00       |
| rs1065852    |        |           |       |                  |                     |            |
| A/A          | 144    | 28.2      | 0.15  | 0.013            | 1.80                | 1.80       |
| A/G          | 240    | 47.1      | 0.498 | 0.001            | 1.83                | 1.83       |
| G/G          | 126    | 24.7      | 0.30  | 1.00             | 1.00                | 1.00       |

**Notes:**
- *P* ≤ 0.05 value indicates statistical significance; *P* adjusted for age and sex.
- Abbreviations: CI, confidence interval; ORs, odds ratios; Ref, reference category; SNP, single nucleotide polymorphism.

### 3.5 The expression and prognostic value of candidate genes in LC patients

As shown in Figure 1, we found the expressions of *CYP2D6* were significantly up-regulated in large cell lung carcinoma, AC and SCC patients compared with the normal samples (*P* < 0.05), and *CYP24A1* and *CYP20A1* were found over-expressed in the AC (*P* < 0.05). Kaplan-Meier curve and log-rank test analyses revealed that the increased *CYP2D6* level and decreased *CYP20A1* level were significantly associated with poor overall survival (OS) in all LC patients (HR = 1.42, 95% CI = 1.25-1.62, *P* = 1.1e-07; HR = 0.72, 95% CI = 0.63-0.82, *P* = 4.2e-7 respectively, Figure 2). There was no significant association between expression of *CYP24A1* and OS of LC (*P* = 0.098).

### 4 Discussion

Based on the metabolic characteristics of CYPs, we hypothesized that their polymorphisms were related to the risk of LC development. This study validated the potential relationship between four SNPs in three CYPs and risk of LC development. We found that rs2762934 and rs6068816 in *CYP24A1* decreased the risk of LC development in males and SCLC respectively, and *CYP20A1* rs2043449 was identified as a risk factor of LC development in males, III-IV stage, and SCLC subgroups. The most significant discovery is that the “A” allele and “AA” genotype of *CYP2D6* rs1065852 confer risk to LC, especially in the cases of III-IV stage AC, SCC, lymph node positive and males. These results made us assume that the susceptibility to LC may in part be defined by the individual’s genetic background of CYPs.

*CYP24A1* encodes 24-hydroxylase, the rate-limiting enzyme that catalyzes the inactivation of 1,25(OH)2D3 (1,25-D3), which is considered as a proto-oncogene. High 1,25-D3 levels have antidifferentiation and antiproliferation activities in human LC cell lines. Earlier researches reported that the gene copy number of *CYP24A1* is aberrantly amplified in several cancers and spontaneous upregulation of *CYP24A1* is a negative prognosticator of survival in lung, breast, ovarian and colon cancer.
might promote the progression of colon cancer. Wu et al found that mutated homozygous CYP24A1 rs6068816 was significantly related to the decrease of the risk of non-small cell lung cancer (NSCLC) development among Chinese people. Liu et al found CYP24A1 rs2762934 contributed to the risk of food hypersensitivity and breast cancer. In the present study, we found that rs2762934 and rs6068816 in CYP24A1 are protective factors to LC for males and in SCLC respectively. Furthermore, CYP24A1 was significantly upregulated in LC. Nithya Ramnath et al revealed that promoter DNA hypermethylation of CYP24A1 is a key mechanism regulating CYP24A1 expression in LC. CYP24A1 has a promoter region that is rich in CpG islands, and transcriptional silencing of the CYP24A1 gene is caused by promoter hypermethylation that would be conductive to 1,25-D3 anti-proliferative effects in LC. Because the amino acid sequence of CYP24A1 is not affected by rs6068816 due to synonymous polymorphisms, the SNPs, located in silencers or enhancers of splicing regions, have an effect on the phenotype of biologic activities by influencing the efficiency of mRNA splicing. The rs2762934 plays a crucial role in intron variant and 3'UTR variant. RNA-binding proteins combined with cis-acting elements in the 3'UTR region to regulate protein synthesis by influencing mRNA abundance. Both the variation of 3'UTR sequence and abnormal expression of trans-acting factors can significantly influence the transcription and expression of target genes. A possible reason for the association of rs6068816 and rs2762934 in CYP24A1 with decreased risk of SCLC and LC development in males is the alternation of posttranscription process and dysfunction of the proteins. For all we know, this has been the first clinic study to estimate the relationship between rs2762934 in CYP24A1 and LC susceptibility. CYP2D6 is a member of the CYP450 superfamily of enzymes involved in the metabolism of therapeutic drugs and is a potential susceptibility factor for certain environmental agent-induced diseases. It plays an important role particularly in the metabolism of PAH, nicotine and other carcinogens related to LC. To date, there have been studies that have shown that genetic polymorphisms of CYP2D6 increase the susceptibility to numerous cancers. Studies have indicated that polymorphisms of CYP2D6 imposed an increased risk of breast cancer and esophageal squamous cell carcinoma in those people with a family history of cancers. In China, the proportion of smoking and tobacco-attributed mortality is much higher in males than in females. SCC is one of the most common pathological type of smoking-related LC. Therefore, it could be assumed that the significant increased risk of LC in males and SCC patients by rs1065852 may be caused by the accumulation of smoking-related genetic damage. Meanwhile, high-level CYP2D6 was found in SCC and AC, and survival analysis also confirmed the poor prognosis of LC caused by CYP2D6. CYP2D6 rs1065852 is located in the intron region of CYP2D6 gene and involved in intron mutation. Intron is important for functions in RNA stability, regulation of gene expression and alternative splicing. Misregulation of alternative splicing is known contribute to tumorigenesis, and the missense variant of base near the splice site could lead to protein and amino acid change due to aberrant splicing. It might be an assumption that the polymorphism of rs1065852 may be involved in the development of LC by influencing the biological function of gene products and mRNA splicing. As we know, no study has validated the association between CYP2D6 rs1065852 and LC susceptibility, and the present study is the first of its kind to verify the correlation between CYP2D6 rs1065852 and the increased of LC in Asians. Stratified analysis also revealed significant associations between CYP20A1 rs2043449 and increased risk of LC in

| Variable     | Dominant model | Recessive model | Additive model |
|--------------|----------------|-----------------|----------------|
| rs2043449    | 1.23 (0.86-1.75) | 2.03 (0.50-8.19) | 1.24 (0.90-1.71) |
| rs2762934    | 0.76 (0.55-1.04) | 1.50 (0.53-4.26) | 0.83 (0.62-1.10) |
| rs1570669    | 0.93 (0.72-1.20) | 1.18 (0.83-1.66) | 1.01 (0.84-1.20) |
| rs6068816    | 1.17 (0.91-1.51) | 1.09 (0.73-1.62) | 1.12 (0.92-1.35) |
| rs2296241    | 0.98 (0.75-1.27) | 0.99 (0.72-1.35) | 0.98 (0.82-1.18) |
| rs1065852    | 1.36 (1.03-1.80) | 1.64 (1.22-2.20) | 1.34 (1.13-1.60) |

Notes: *P* ≤ 0.05 value indicates statistical significance; *P* adjusted for age and sex

Abbreviations: CI, confidence interval; ORs, odds ratios; SNP, single nucleotide polymorphism.
| Variable                  | CYP2D6 rs1065852 | Adjusted OR (95%CI) | $P^a$ | Study Power |
|--------------------------|------------------|---------------------|------|-------------|
|                          | AA               | AG                  | GG   | (AA/AG) vs GG | AA vs (AG/GG) | AA vs GG |
| Sex                      |                  |                     |      |              |              |          |
| Male                     | 1.60 (1.08-2.39) | 0.96                | 0.85 (0.62-1.18) | 1.20 (0.88-1.63) | 1.57 (1.11-2.21) | 0.993 | 1.25 (1.03-1.53) | 0.724 |
| Female                   | 0.33 (0.16-0.70) | 0.99                | 0.65 (0.35-1.20) | 0.339 | 0.263       | 0.993 | 0.58 (0.40-0.84) | 0.999 |
| TNM Stage                |                  |                     |      |              |              |          |
| I-II                     | 1.47 (0.87-2.47) | 0.84                | 0.84 (0.53-1.32) | 1.01 (0.66-1.54) | 1.63 (1.04-2.55) | 0.998 | 1.19 (0.91-1.56) | 0.206 |
| III-IV                   | 1.96 (1.27-3.03) | 0.99                | 1.34 (0.92-1.95) | 0.168 | 0.986       | 0.999 | 0.44 (0.24-0.83) | 0.999 |
| Lymph node status        |                  |                     |      |              |              |          |
| Positive                 | 2.08 (1.29-3.33) | 0.99                | 1.38 (0.91-2.08) | 0.131 | 0.022       | 0.003 | 1.44 (1.14-1.83) | 0.988 |
| Negative                 | 1.12 (0.65-1.94) | 0.99                | 0.81 (0.51-1.29) | 0.378 | 0.627       | 0.008 | 1.03 (0.78-1.37) | 0.818 |
| Type of cancer           |                  |                     |      |              |              |          |
| SCC                      | 1.80 (1.10-2.95) | 0.977               | 1.18 (0.77-1.81) | 0.441 | 0.139       | 0.023 | 1.34 (1.04-1.72) | 0.921 |
| AC                       | 2.18 (1.34-3.55) | 0.999               | 1.02 (0.65-1.59) | 0.942 | 0.163       | 0.023 | 1.19 (1.16-1.93) | 0.997 |
| SCLC                     | 1.28 (0.69-2.38) | 0.999               | 1.09 (0.65-1.82) | 0.744 | 0.588       | 2.15  | 1.13 (0.83-1.54) | 0.444 |

Notes: $^a$ $P \leq 0.05$ value indicates statistical significance; $P^a$ adjusted for age and sex.

Abbreviations: AC, adenocarcinoma; CI, confidence interval; ORs, odds ratios; SCC, Squamous cell carcinoma; SCLC, small cell lung cancer; TNM, tumor-lymph node-metastasis.
# Table 7
Stratified analyses between CYP24A1 rs2762934 polymorphism and lung cancer susceptibility

| Variable          | AA                                | AG                          | GG (AA/AG) vs GG | AA vs (AG/GG) | AA vs GG |
|-------------------|-----------------------------------|-----------------------------|-----------------|---------------|---------|
| **Sex**           |                                   |                             |                 |               |         |
| Male              | 0.77 (0.20-2.89) 0.693            | **0.68 (0.46-0.99) 0.046**  | 0.919           | **0.68 (0.47-0.99) 0.044**  | **0.932** |
| Female            | 4.73 (0.54-41.3) 0.160            | 0.84 (0.42-1.68) 0.630      | 1.00            | 1.02 (0.54-1.96) 0.942      | 4.86 (0.56-42.33) 0.153 | **1.18 (0.68-2.03) 0.560** |
| **TNM Stage**     |                                   |                             |                 |               |         |
| I-II              | 0.58 (0.07-4.90) 0.615            | 0.72 (0.42-1.23) 0.228      | 0.71 (0.42-1.20) 0.202 | 0.61 (0.07-5.18) 0.651      | 0.73 (0.44-1.19) 0.200 |
| III-IV            | 1.29 (0.36-4.62) 0.698            | 0.68 (0.44-1.04) 0.074      | 0.71 (0.47-1.07) 0.106 | 1.37 (0.38-4.92) 0.626      | 0.78 (0.54-1.13) 0.184 |
| **Lymph node status** |                               |                             |                 |               |         |
| Positive          | 1.23 (0.30-4.98) 0.775            | 0.69 (0.43-1.10) 0.117      | 0.72 (0.46-1.13) 0.151 | 1.31 (0.32-5.29) 0.709      | 0.78 (0.52-1.17) 0.229 |
| Negative          | 0.59 (0.07-5.01) 0.626            | 0.70 (0.40-1.22) 0.208      | 0.69 (0.40-1.19) 0.185 | 0.62 (0.07-5.31) 0.665      | 0.71 (0.43-1.18) 0.187 |
| **Type of cancer**|                                   |                             |                 |               |         |
| SCC               | 0.95 (0.18-4.88) 0.947            | 0.73 (0.45-1.18) 0.200      | 0.74 (0.46-1.19) 0.211 | 1.00 (0.19-5.14) 0.999      | 0.78 (0.50-1.20) 0.253 |
| AC                | 1.88 (0.51-6.86) 0.342            | 0.74 (0.45-1.21) 0.230      | 0.81 (0.50-1.29) 0.375 | 1.97 (0.54-7.21) 0.304      | 0.90 (0.59-1.36) 0.619 |
| SCLC              | 0.83 (0.10-7.02) 0.863            | 0.54 (0.28-1.06) 0.073      | 0.56 (0.29-1.06) 0.077 | 0.91 (0.11-7.70) 0.930      | 0.61 (0.33-1.10) 0.100 |

*Notes:* *P* ≤ 0.05 value indicates statistical significance; P* adjusted for age and sex.

Abbreviations: AC, adenocarcinoma; CI, confidence interval; ORs, odds ratios; SCC, Squamous cell carcinoma; SCLC, small cell lung cancer; TNM, tumor-lymph node-metastasis.
| Variable | CYP24A1 rs6068816 | Adjusted OR (95%CI) | P<sup>a</sup> Study Power |
|----------|-----------------|------------------|--------------------------|
| **Sex**  |                 |                  |                          |
| Male     | 1.32 (0.80-2.17) 0.284 | 1.12 (0.83-1.51) 0.465 | 1.00 1.15 (0.86-1.53) 0.340 | 1.24 (0.77-2.00) 0.376 | 1.14 (0.91-1.42) 0.255 |
| Female   | 0.91 (0.40-2.05) 0.811 | 1.36 (0.80-2.33) 0.259 | 1.00 1.25 (0.75-2.08) 0.391 | 0.77 (0.36-1.66) 0.506 | 1.06 (0.73-1.54) 0.762 |
| **TNM Stage** |                 |                  |                          |
| I-II     | 0.98 (0.48-2.02) 0.960 | 1.28 (0.85-1.94) 0.235 | 1.00 1.23 (0.83-1.83) 0.309 | 0.86 (0.43-1.71) 0.665 | 1.09 (0.81-1.48) 0.560 |
| III-IV   | 1.21 (0.72-2.02) 0.481 | 1.17 (0.84-1.62) 0.347 | 1.00 1.18 (0.86-1.60) 0.308 | 1.11 (0.68-1.81) 0.676 | 1.12 (0.89-1.41) 0.335 |
| **Lymph node status** |                 |                  |                          |
| Positive | 0.82 (0.44-1.52) 0.526 | 1.06 (0.75-1.50) 0.740 | 1.00 1.02 (0.73-1.42) 0.924 | 0.79 (0.44-1.43) 0.445 | 0.96 (0.75-1.25) 0.783 |
| Negative | 1.49 (0.77-2.90) 0.237 | 1.29 (0.84-1.99) 0.249 | 1.00 1.33 (0.88-2.01) 0.180 | 1.30 (0.70-2.42) 0.402 | 1.24 (0.92-1.68) 0.161 |
| **Type of cancer** |                 |                  |                          |
| SCC      | 1.29 (0.69-2.42) 0.420 | 1.43 (0.98-2.10) 0.063 | 1.00 1.41 (0.98-2.03) 0.066 | 1.06 (0.59-1.91) 0.845 | 1.23 (0.94-1.62) 0.135 |
| AC       | 1.19 (0.64-2.23) 0.586 | 1.45 (0.98-2.13) 0.061 | 1.00 1.40 (0.96-2.02) 0.078 | 0.97 (0.54-1.75) 0.929 | 1.20 (0.91-1.57) 0.198 |
| SCLC     | 0.94 (0.47-1.90) 0.869 | 0.58 (0.36-0.94) 0.026* 0.999 | 1.00 0.65 (0.42-1.01) 0.053 | 1.20 (0.61-2.35) 0.596 | 0.81 (0.58-1.14) 0.229 |

Notes: *P ≤ 0.05 value indicates statistical significance; P<sup>a</sup> adjusted for age and sex.
Abbreviations: AC, adenocarcinoma; CI, confidence interval; ORs, odds ratios; SCC, Squamous cell carcinoma; SCLC, Small cell lung cancer; TNM, tumor-lymph node-metastasis.
| Variable                        | CC        | CT        | TT        | CC/CT vs TT | CC vs CT/TT | CC vs TT |
|--------------------------------|-----------|-----------|-----------|-------------|-------------|----------|
| **Sex**                        |           |           |           |             |             |          |
| Male                           | 3.19 (0.64-15.94) | 1.42 (0.94-2.16) | 1.00      | **1.50 (1.00-2.24)** | **0.916**  | **1.50 (1.03-2.16)** | 0.033*   |
| Female                         | 6.33e-10 (0.00-inf) | 0.65 (0.30-1.40) | 1.00      | 0.62 (0.29-1.32) | 0.00 (0.00-inf) | 0.999    | 0.60 (0.29-1.25) | 0.174 |
| **TNM Stage**                  |           |           |           |             |             |          |
| I-II                           | 2.09e-9 (0.00-inf) | 0.95 (0.53-1.72) | 1.00      | 0.91 (0.50-1.63) | 0.00 (0.00-inf) | 0.999    | 0.87 (0.49-1.52) | 0.620 |
| III-IV                         | 2.93 (0.65-13.25) | 1.41 (0.92-2.15) | 1.00      | 1.47 (0.97-2.23) | 0.067       | 2.79 (0.62-12.60) | **1.47 (1.01-2.15)** | 0.044* |
| **Lymph node status**          |           |           |           |             |             |          |
| Positive                       | 1.85 (0.31-11.2) | 1.46 (0.92-2.32) | 1.00      | 1.50 (0.94-2.32) | 1.75 (0.29-10.57) | 1.44 (0.95-2.18) | 0.086 |
| Negative                       | 2.91 (0.46-18.22) | 0.75 (0.39-1.45) | 1.00      | 0.84 (0.45-1.57) | 3.00 (0.48-18.76) | 0.95 (0.54-1.65) | 0.841 |
| **Type of cancer**             |           |           |           |             |             |          |
| SCC                            | 0.81 (0.08-8.40) | 1.34 (0.82-2.21) | 1.00      | 1.32 (0.81-2.15) | 0.78 (0.08-8.04) | 1.26 (0.80-1.99) | 0.326 |
| AC                             | 1.07 (0.11-10.51) | 1.00 (0.58-1.72) | 1.00      | 1.00 (0.59-1.71) | 1.07 (0.11-10.50) | 1.01 (0.61-1.65) | 0.983 |
| SCLC                           | **5.36 (1.06-27.21)** | **1.14 (0.60-2.18)** | 1.00      | **1.34 (0.74-2.44)** | **5.27 (1.04-26.66)** | **0.991** | **1.46 (0.87-2.45)** | 0.148 |

Notes: *P ≤ 0.05 value indicates statistical significance; P* adjusted for age and sex.
Abbreviations: AC, adenocarcinoma; CI, confidence interval; ORs, odds ratios; SCC, Squamous cell carcinoma; SCLC, small cell lung cancer; TNM, tumor-lymph node-metastasis.
males, III-IV stage, and SCLC subgroups. Although high level of CYP20A1 predicts a better prognosis in survival analysis, the conflict between these two outcomes might be due to the sample size, territory and racial differences. Previous studies showed that CYP20A1 is expressed in the human hippocampus and substantia nigra, suggesting its involvement in brain and early development. As far as we know, CYP20A1 was considered as “orphan” CYP with no functional information. Therefore, the mechanism of rs2043449 affecting tumor susceptibility in these subgroups remains unclear; further functional analysis of CYP20A1 in these subgroups may help to clarify the relevant genetic effects of LC pathogenesis.

In this study, we identified four novel loci in three genes that show a significant linkage with LC development, and observed the expression of candidate genes in LC and the relationship between poor prognosis of LC and two genes. Although the results showed strong statistical significance, there are still several potential limitations in the present research. First, LC is a very heterogeneous disease with many other risk factors, and more genes need to be included in follow-up studies. Second, the study is conducted among only in the Chinese Han people in Northwest China, for which further investigations are needed to confirm these associations in other populations. Third, the sample size was not large enough to support some genetic models in stratified analyses. Finally, the smoking data of the samples were not collected, and further study is needed to improve the deficiencies of this research.

5 | CONCLUSION

In this study, we systematically evaluated the association of candidates genes and LC risk in a case-control study including 510 cases and 504 healthy controls. And finally we found the significant relationship of CYP2D6 rs1065852, CYP20A1 rs2043449, CYP24A1 rs2762934, and CYP24A1 rs6068816 with susceptibility to LC. In addition, we explored the overexpression of candidate genes in LC and estimated the relationship between LC prognosis and genes expression level in survival analysis using Oncomine and Kaplan-Meier Plotter database, which could potentially contribute to elucidate the etiology of LC and be used as diagnostic and prognostic molecular markers for LC in Northwest Chinese Han population.

ACKNOWLEDGMENTS

The authors thank all the researchers and patients who participated in this study.
CONFLICT OF INTEREST

The authors made no disclosures.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Mingwei Chen https://orcid.org/0000-0002-0418-1510

REFERENCES

1. Global Burden of Disease Cancer C, Fitzmaurice C, Akinyemiju TF, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016. JAMA Oncology. 2018;4(11):1553-1568.

2. 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 Cancer J Clin. 2018;68(6):394-424.

3. Wong M, Lao XQ, Ho K-F, Goggins WB, Tse S. Incidence and trends for colorectal cancer in Asia. Jpn J Clin Oncol. 2012;42:1110-1118.

4. Gapstur SM, Drope JM, Jacobs EJ, et al. A blueprint for the prevention of lung cancer. Prev Med. 2012;55:134-144.

5. Brennan P, Hainaut P, Boffetta P. Genetics of lung cancer. Lung Cancer. 2004;45:133-139.

6. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial risk and heritability of lung cancer among twins in Nordic countries. JAMA. 2016;315(15):1631-1639.

7. Carr SR, Akerley W, Cannon‐Albright LA. Genetic contribution to gender‐divergent expression of 78 cytochrome P450 mRNAs in cancer patients. Carcinogenesis. 2011;32:56-66.

8. Wong M, Lao XQ, Ho K-F, Goggins WB, Tse S. Incidence and trends for colorectal cancer in Asia. Jpn J Clin Oncol. 2012;42:1110-1118.

9. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effector genes in cancer risk. Cancer Res. 2015;75:101-109.

10. Brennan P, Hainaut P, Boffetta P. Genetics of lung cancer. Future Oncol. 2011;7:1355-1366.

11. Breneman MD, Lee DJ, Han J, et al. Familial risk and heritability of lung cancer among twins in Nordic countries. JAMA Oncology. 2016;4(1):24-30.

12. Breneman MD, Lee DJ, Han J, et al. Familial risk and heritability of lung cancer among twins in Nordic countries. JAMA Oncology. 2016;4(1):24-30.

13. Breneman MD, Lee DJ, Han J, et al. Familial risk and heritability of lung cancer among twins in Nordic countries. JAMA Oncology. 2016;4(1):24-30.

14. Breneman MD, Lee DJ, Han J, et al. Familial risk and heritability of lung cancer among twins in Nordic countries. JAMA Oncology. 2016;4(1):24-30.

15. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther. 2013;138:103-141.

16. Kiyohara C. Genetic polymorphism of enzymes involved in xenobiotic metabolism and the risk of colorectal cancer. J Epidemiol. 2000;10:349-360.

17. Maurya SS, Anand G, Dhawan A, et al. Polymorphisms in drug-metabolizing enzymes and risk to head and neck cancer: evidence for gene-gene and gene-environment interaction. Environ Mol Mutagen. 2014;55:134-144.

18. Su T, Bao Z, Zhang QY, et al. Human cytochrome P450 CYP2A13: predominant expression in the respiratory tract and its high efficiency in the metabolism of a tobacco-specific carcinogen, 4-(methyl-nitrosamino)-1-(3-pyridyl)-1-butane. Cancer Res. 2000;60:5074-5079.

19. Shimada T, Murayama N, Tanaka K, et al. Interaction of polycyclic aromatic hydrocarbons with human cytochrome P450 1B1 in inhibiting catalytic activity. Chem Res Toxicol. 2008;21:2313-2323.

20. Moorhdy B, Chu C, Carlin DJ. Polycyclic aromatic hydrocarbons: from metabolism to lung cancer. Toxicol Sci. 2015;145:5-15.

21. Li L, Megaraj V, Wei Y, Ding X. Identification of cytochrome P450 enzymes critical for lung tumorigenesis by the tobacco-specific carcinogen 4-(methyl-nitrosamino)-1-(3-pyridyl)-1-butane (NNK): insights from a novel Cyp2abfgs-null mouse. Carcinogenesis. 2014;35:2584-2591.

22. Siddens LK, Bunde KL, Harper TA, et al. Cytochrome P450 1B1 in polycyclic aromatic hydrocarbon (PAH)-induced skin carcinogenesis: tumorigenicity of individual PAHs and coal-tar extract, DNA adduction and expression of select genes in the Cyp1b1 knockout mouse. Toxicol Appl Pharmacol. 2015;287:149-160.

23. Kiyohara C, Shirakawa T, Hopkin MJ. Genetic polymorphism of enzymes involved in xenobiotic metabolism and the risk of lung cancer. Environ Health Prev Med. 2002;7:47-59.

24. Goldstraw P, Chansky K, Crowley J, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. J Thorac Oncol. 2016;11:39-51.

25. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science. 2002;296:2225-2229.

26. Dupont WD, Plummer WD, Jr. Power and sample size calculations for studies involving linear regression. Control Clin Trials. 1998;19:589-601.

27. Albertson DG, Ylstra B, Segraves R, et al. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat Genet. 2000;25:144-146.

28. Mawer EB, Hayes ME, Heys SE, et al. Constitutive synthesis of 1,25-dihydroxyvitamin D3 by a human small cell lung cancer cell line. J Clin Endocrinol Metab. 1994;79:554-560.

29. Mimori K, Tanaka Y, Yoshinaga K, et al. Clinical significance of CYP24A1 in the pathogenesis of lung cancer. Mol Med. 2014;20:246-254.

30. Parise RA, Egorin MJ, Kanterewicz B, et al. CYP24, the enzyme that catabolizes the antiproliferative agent vitamin D, is increased in non-small cell lung cancer. J Thorac Oncol. 2008;3:1041-1046.

31. Friedrich M, Rafi L, Mitschele T, et al. Analysis of the vitamin D metabolizing enzyme CYP24A1 in human lung adenocarcinoma. Cancer Res. 2015;75:1316-1323.

32. Chen G, Kim SH, King AN, et al. CYP24A1 is an independent prognostic marker of survival in patients with lung adenocarcinoma. Clin Cancer Res. 2011;17:817-826.
33. Sun H, Wang C, Hao M, et al. CYP24A1 is a potential biomarker for the progression and prognosis of human colorectal cancer. *Hum Pathol.* 2016;50:101-108.

34. Chen XQ, Mao JY, Li WB, et al. Association between CYP24A1 polymorphisms and the risk of colonic polyps and colon cancer in a Chinese population. *World J Gastroenterol.* 2017;23:5179-5186.

35. Wu X, Cheng J, Yang K. Vitamin D-related gene polymorphisms, plasma 25-hydroxy-vitamin D, cigarette smoke and non-small cell lung cancer (NSCLC) risk. *Int J Mol Sci.* 2016;17(10):1597.

36. Liu X, Wang G, Hong X, et al. Gene-vitamin D interactions on food sensitization: a prospective birth cohort study. *Allergy.* 2011;66:1442-1448.

37. Fuhrman BJ, Freedman DM, Bhatti P, et al. Sunlight, polymorphisms of vitamin D-related genes and risk of breast cancer. *Anticancer Res.* 2013;33:543-551.

38. Ramnath N, Nadal E, Jeon CK, et al. Epigenetic regulation of vitamin D metabolism in human lung adenocarcinoma. *J Thorac Oncol.* 2014;9:473-482.

39. Dever TE. Gene-specific regulation by general translation factors. *Cell.* 2002;108:545-556.

40. Wang JF, Zhang CC, Chou KC, et al. Structure of cytochrome p450s and personalized drug. *Curr Med Chem.* 2009;16:232-244.

41. Li J, Zhao M, He P, Hidalgo M, Baker SD. Differential metabolism of gefitinib and erlotinib by human cytochrome P450 enzymes. *Clin Cancer Res.* 2007;13:3731-3737.

42. Suzumura T, Kimura T, Kudoh S, et al. Reduced CYP2D6 function is associated with gefitinib-induced rash in patients with non-small cell lung cancer. *BMC Cancer.* 2012;12:568.

43. Abraham JE, Maranian MJ, Driver KE, et al. CYP2D6 gene variants and their association with breast cancer susceptibility. *Cancer Epidemiol Biomarkers Prev.* 2011;20:1255-1258.

44. Bhat GA, Shah IA, Rafiq R, et al. Family history of cancer and the risk of squamous cell carcinoma of oesophagus: a case-control study in Kashmir. *India. Br J Cancer.* 2015;113:524-532.

45. Zienolddiny S, Campa D, Lind H, et al. A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of non-small cell lung cancer in smokers. *Carcinogenesis.* 2008;29:1164-1169.

46. Lee JY, Vinayagamoorthy N, Han K, et al. Association of polymorphisms of cytochrome P450 2D6 with blood hydroxychloroquine levels in patients with systemic lupus erythematosus. *Arthritis Rheumatol.* 2016;68:184-190.

47. Bouchardy C, Benhamou S, Dayer P. The effect of tobacco on lung cancer risk depends on CYP2D6 activity. *Cancer Res.* 1996;56:251-253.

48. Crespi CL, Pennman BW, Gelboin HV, et al. A tobacco smoke-derived nitrosamine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, is activated by multiple human cytochrome P450s including the polymorphic human cytochrome P4502D6. *Carcinogenesis.* 1991;12:1197-1201.

49. Chen Z, Peto R, Zhou M, et al. Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies. *Lancet.* 2015;386:1447-1456.

50. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer.* 2001;31:139-148.

51. Anczukow O, Krainer AR. Splicing-factor alterations in cancers. *RNA.* 2016;22:1285-1301.

52. Stark K, Wu Z-L, Bartleson CJ, Guengerich FP. mRNA distribution and heterologous expression of orphan cytochrome P450 20A1. *Drug Metab Dispos.* 2008;36:1930-1937.

53. Stark K, Guengerich FP. Characterization of orphan human cytochromes P450. *Drug Metab Rev.* 2007;39:627-637.

**SUPPORTING INFORMATION**

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

*How to cite this article:* Li M, Li A, He R, et al. Gene polymorphism of cytochrome P450 significantly affects lung cancer susceptibility. *Cancer Med.* 2019;8:4892–4905. [https://doi.org/10.1002/cam4.2367](https://doi.org/10.1002/cam4.2367)