Genetic Variants Modifying the Risk of Lung Cancer and Its Subtypes: A Comprehensive Meta-analysis and a Case-control Study

Debmalya Sengupta
University of Calcutta

Souradeep Banerjee
University of Calcutta

Ritabrata Mitra
Institute of Post Graduate Medical Education And Research

Tamohan Chaudhuri
Saroj Gupta Cancer Centre & Research Institute

Abhijit Sarkar
Saroj Gupta Cancer Centre & Research Institute

Gautam Battacharjee
Saroj Gupta Cancer Centre & Research Institute

Somsubhra Nath
Saroj Gupta Cancer Centre & Research Institute

Susanta Roychoudhury
Saroj Gupta Cancer Centre & Research Institute

Samshiddhi Battacharjee
National Institute of Biomedical Genomics

Mainak Sengupta ( sengupta.mainak@gmail.com )
Department of Genetics, University of Calcutta, 35, Ballygunge Circular Road, Kolkata 700019, India.

Research Article

Keywords: Risk Variants, Tobacco smoking, Association analysis, Statistical analysis, Histological subtypes, Population stratification

Posted Date: December 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-127817/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on July 16th, 2021. See the published version at https://doi.org/10.1038/s41598-021-92275-z.
Abstract

Association studies on lung cancer have often yielded conflicting and inconclusive results. We performed a comprehensive meta-analysis to dissect the precise effects of the candidate variants. We searched for association studies on lung cancer from the Indian subcontinent. Cochrane's Q-test assessed heterogeneity. Both overall and histotype-stratified meta-analysis was done using fixed-effect and random-effects models. Smoking status stratified subgroup analysis and effect modification tests were done. An associated variant with significant heterogeneity was genotyped in an eastern Indian population to investigate the contribution of potential confounders followed by a comprehensive meta-analysis across world populations. Significant heterogeneity was observed for the 8 variants. Both fixed-effect and random-effects meta-analysis of 24 variants showed FDR-corrected associations of rs3547/XRCC1 and rs1048943/CYP1A1 with lung cancer along with 5 nominal associations. del1/GSTT1, rs4646903/CYP1A1, and rs10488943/CYP1A1 were associated with adenocarcinoma, squamous cell carcinoma, and both, respectively. rs4646903/CYP1A1 was associated with lung cancer among smokers with significant effect modification by smoking. rs10488943/CYP1A1 was associated with lung adenocarcinoma in the East Indian case-control study. rs1048943/CYP1A1 was associated with lung cancer across world populations. Our work confirms the risk loci for lung cancer and its subtypes in the context of smoking and other aetiological factors, which could aid in personalised treatment.

Introduction

Despite several measures taken against tobacco smoking and consumption, lung cancer remains one of the leading causes of cancer-related mortalities worldwide with a low 5-year survival rate \(^1\). Epidemiological data suggest that the global lung cancer burden has risen to 2.1 million new cases of all cancer cases and 1.8 million deaths, which is close to 1 in 5 of all cancer deaths \(^2\). The increasing rate of incidence of lung cancer in the Indian subcontinent and East Asia is evident in a recent epidemiological report \(^3\). Lung cancer incidences show a great deal of variation across different geographical regions consisting of an admixture of diverse populations \(^4\). In India, lung cancer constitutes 5.9% of all new cancer cases and 8.1% of all cancer-related mortalities in both sexes \(^2\). The northeastern state of Mizoram accounts for the highest reported cases of lung cancer in both sexes \(^4\). Earlier reports stated that approximately one million of the total five million lung cancer deaths worldwide are contributed by India \(^5\), and the death toll is projected to rise to 1.5 million by 2020 \(^6\,\,7\). Smoking tobacco has been considered a significant factor in lung carcinogenesis \(^8\). Apart from tobacco smoking, betel quid chewing \(^9\,\,10\,\,11\) and 10–15, biofuel exposure \(^10\,\,11\,\,16\), asbestos exposure \(^10\,\,11\,\,17\,\,18\) also contribute to lung carcinogenesis. Earlier studies have revealed a rise in the incidence of lung cancer among never smokers \(^19\), particularly in women of East Asian origin \(^20\,\,21\).

Genome-wide association studies (GWAS) in the Chinese population have identified 16 susceptibility loci \((p ≤ 5.00 \times 10^{-8})\) to be associated with lung cancer risk \(^22\,\,23\) and 4 loci out of them showed evidence of association with lung cancer risk in smokers \(^22\). Similarly, another GWAS on subjects of European ancestry with 29266 lung cancer patients and 56450 controls identified 18 susceptibility loci \((p ≤ 5.00 \times 10^{-8})\), including 10 novel loci.

Interestingly, the association of the 10 novel loci varied across different histological subtypes. Out of the 10 loci, four were associated with overall lung cancer risk, while the remaining 6 loci were found to be associated with lung adenocarcinoma \(^21\). Most of the GWAS was done on subjects of European or Chinese descent, and the majority of the identified risk alleles have not been evaluated in the population of the Indian subcontinent despite several candidate gene association studies \(^24\,\,25\,\,26\). Contradictory outcomes of case-control association studies of the same polymorphism by different authors failed to identify the overall effect of the genes and the genetic variations on lung cancer susceptibility in the region. For example, one study showed no association between the GSTT1 deletion polymorphism and lung cancer in 100 cases and 76 controls \(^34\), but another study with 146 cases and 146 controls showed a significant association of GSTT1 deletion polymorphism with lung cancer \(^35\). In another study conducted, rs1048943 of CYP1A1 showed a significant association with lung cancer \(^31\), while the same variant showed no significant association with lung cancer in a different study from the Indian subcontinent \(^36\). The differences in genetic association across the geographical regions of the Indian subcontinent, comprised of distinct population groups, might be attributed to gene-gene and gene-environment interactions, which could act as potential modulators of lung cancer risk \(^37\). The contradictory outcomes of the variant between different studies could be due to small sample sizes, clinical heterogeneity between the populations, and racial/ethnic differences \(^38\) within the Indian subcontinent.

Further, the differences in socio-economic and cultural practices in different parts of the Indian subcontinent might contribute to diverse lifestyle habits like smoking, chewing of tobacco and betel quids, alcohol consumption, and exposure to air pollutants; exposure to asbestos and other occupational hazards that in turn could modify the risk of the disease. This brings forth the importance of meta-analysis, which is a powerful statistical method \(^39\) to assess the pooled effect of the variant(s) on lung cancer susceptibility in the concerned population by pooling the individual study data. The current study aimed to estimate the magnitude of the effect of the reported candidate associations on lung cancer susceptibility in the Indian subcontinent through a meta-analysis pipeline (Supplementary Information, Fig. S1) as followed in our previous pilot study \(^40\). The meta-analysis was followed by a case-control study, conducted on the lung cancer-associated polymorphic variant after multiple testing adjustments with significant heterogeneity (Supplementary Information, Fig. S2). Furthermore, the significant variants of the above analysis were meta-analysed across the world population to contrast the associations across populations of varying ethnicities.

Methods

The scheme of analysis followed in this study is explained and summarised in (Supplementary Information, Fig. S2).

Identification and Eligibility of studies
The current study followed the PRISMA guidelines. Systematic mining of the databases, such as PubMed, Scopus, and Web of Science, was done to select appropriate studies using the following keywords: (SNP/SNPs/polymorphisms/single nucleotide polymorphisms/SNVs/SNV/Mutation/Variants/Genotypes/Alleles); (Lung cancer/Lung Carcinoma/Lung malignancy/Lung neoplasm); (India/Pakistan/Nepal/Bangladesh/Bhutan/Sri-Lanka/Maldives/Afghanistan). All case-control genetic association studies on lung cancer were curated and selected manually by two authors and rechecked by the other authors.

**Inclusion and Exclusion criteria**

The selection of the studies for meta-analysis was made following the specific inclusion criteria: (a) samples should be from populations belonging to the countries of the Indian subcontinent; (b) case-control genotype data should be reported for each polymorphic variant; (c) the data source should be a full research article; (d) all association studies published till 31st December 2019 were considered; (e) each polymorphism reported in at least three independent studies on the different sample populations were considered; (f) studies should be published in English. The exclusion criteria were as follows: (a') incomplete data from review articles, letters, editorials, comments and case reports; (b') duplicated studies using the same population; (c') studies with no genotype count among cases and controls; (d') studies in languages other than English; (e') unpublished data; and (f') systematic reviews and meta-analysis were not considered.

**Data extraction**

Data extraction from the literature was done following specific inclusion and exclusion criteria. The data collected from the selected studies are (1) first author surname, (2) year of the publication, (3) mean age with standard deviation, (4) sex, (5) smoking status, (6) histological types, (7) genetic polymorphisms and (8) genotype-specific case-control data (9) geographical region of sampling done in the selected studies.

**Selection of a genetic model for meta-analysis**

Apart from del1/GSTT1, del2/GSTM1, rs3025039/VEGFA, rs1048943/CYP1A1 and rs4646903/ CYP1A1, the remaining 20 polymorphic variants were analysed in 3 different genetic models: i.e. additive, dominant and recessive models. The variants del1/GSTT1 and del2/GSTM1 were only analysed in the recessive model, while rs1048943/CYP1A1, rs4646903/CYP1A1, and rs3025039/VEGFA were analysed in a dominant model. We restricted our analysis to a particular genetic model for the mentioned variants because of the nature of the data reported in the studies selected for the analysis.

**Meta-analysis**

Inter-study heterogeneity was evaluated using Cochran's Q test ($p < 0.1$). In the presence of significant heterogeneity, the heterogeneity index ($I^2$) quantifies the degree of inconsistency across studies using the formula: $I^2 = (Q - (n - 1)n) / Q \times 100 \%$, where 'n' is the number of studies selected. The $I^2$ value is expressed as a percentage with cut-offs of 25%, 50% or 75% conventionally signifying the presence of low-, mid- or high-grade heterogeneity, respectively. Meta-analysis was conducted in R, 3.4.2 package 'metafor' on lung cancer genetic association reports from the Indian subcontinent. Logistic regression of lung cancer status on variant genotype was done using additive, recessive and dominant effect models (using R function 'glm') to obtain the study-level effect sizes (log odds ratio), standard errors and 95% confidence intervals. Both fixed-effect meta-analysis (inverse-variance weighting) and random-effects meta-analysis (DerSimonian Laird method) were used to combine the study-level estimates (using the ‘rma.uni’ function in R package ‘metafor’) to estimate cumulative odds ratios and 95% confidence intervals (95% CI) to determine the overall evidence of statistical association ($p < 0.05$) of the reported variants with lung cancer risk. Benjamini-Hochberg method was used to correct the false discovery rate (FDR) at a 5% level ($p_{FDR} < 0.05$).

**Effect on histological subtypes**

The genotype counts of the cases were stratified within the histological subtypes of lung cancer for 5 variants only. The remaining variants lacked the histological subtype-stratified genotype counts for the cases and were not included in the analysis. Similarly, a logistic regression of the status of histological subtypes of lung cancer on variant genotype was done in three genetic models (using R function 'glm'), as mentioned above, to obtain the study-level effect sizes (log odds-ratios), standard errors and 95% confidence intervals. Following this, both fixed-effect and random-effects meta-analyses were used to combine the study-level estimates (using the ‘rma.uni’ function in R package ‘metafor’) to assess the effect of variants on each lung cancer subtype.

**Effect Modification by Smoking**

For the variants nominally associated with lung cancer, we looked for smoking status-stratified summary data (genotype counts) as described earlier. The stratified study-level $\beta$-coefficient (log OR) and standard errors (SE) were calculated using logistic regression from the stratum-specific genotype counts reported in the respective studies. A fixed-effect meta-analysis obtained overall effect estimates, confidence intervals and p-values within each subgroup (i.e., for 'smoker' and 'non-smoker' groups). Finally, a meta-regression in the fixed-effect model was done with stratum (i.e., smoker/non-smoker group) as a moderator variable (using the 'rma.uni’ function in R package ‘metafor’).

**Effect of Geographical Region**

The selected studies for this meta-analysis did not report the race/ethnicity of the patients and controls. Thus, we grouped the studies into geographic regions (North, South, East and West) and conducted a meta-analysis within each subgroup. This analysis was done to assess the region-specific association of the polymorphic variants with lung cancer risk if any.

**Publication Bias**
A visual inspection of funnel plots along with Egger's regression test was done to evaluate the asymmetry (p<0.05) of the funnel plots for the estimation of publication bias if any, among the selected studies. Egger's regression test uses a weighted regression model with multiplicative dispersion for only those variants that are reported in 10 or more studies.

**Case-Control analysis in the East Indian population**

All the histological subtypes of confirmed lung cancer patients were recruited from the Saroj Gupta Cancer Centre and Research Institute, and the Department of CHEST, IPGME&R in Kolkata were considered for patient selection and recruitment for sample collection. Both males and females were considered without any specific sex bias. Patients with a recent history of tobacco smoking were selected. All individuals who had quit smoking ≥ 15 years from the date of sample collection were not considered. Similarly, controls were recruited from the Department of CHEST, IPGME&R in Kolkata, belonging to the same geographical region as the patients with confirmed smoking history. The controls were clinically-radiologically confirmed healthy smokers aged ≥ 55 years and without any history of cancer. All participants of the study were asked to fill up a detailed questionnaire, and their informed consent for voluntary participation was obtained before sample collection following the ethical guidelines of the concerned institutes and the Declaration of Helsinki, 1964. Case-Control association of rs1048943/CYP1A1 with lung cancer in the East Indian population, including 101 cases and 413 controls, was conducted. The polymorphism was selected from the current meta-analysis on lung cancer as the significant polymorphic variant after FDR correction with significant heterogeneity between the studies. The primer sequences used for the PCR of the fragment of 204 base pairs harbouring the polymorphism rs1048943 of CYP1A1 were as follows: CYP1A1-F: 5’-CTGTCCTCCTCTGTTACAGGAAGC-3’, and CYP1A1-R: 5’-TCTCACCTGTTGCAGCAGATAGC-3’. The PCR conditions followed for adequate amplification were as follows: 94°C/5 min – (94°C/40 s – 61°C/40 s – 72°C/40 s) · 30 cycles – 72°C/7 min – 4°C hold. Following PCR, a quality check of the amplicons was done in 6% PAGE. The BsrDI restriction enzyme digested the PCR amplicons at 65°C for 2 h. Covariate-adjusted (adjusted for age, sex, ethnicity, smoking intensity in pack-years, alcohol consumption, tobacco and betel quid chewing, and asbestos exposure) unconditional logistic regression of the genotype count on lung cancer status was performed using R 3.4.2 software.

**Meta-analysis of significant variants in the global population**

The variants associated with lung cancer (after FDR correction) in the sample population of the Indian subcontinent were further meta-analysed in the global population following a similar protocol and PRISMA guidelines, as mentioned above. PubMed, Scopus, and Web of Science repositories were searched to select appropriate studies using the following keywords: (CYP1A1/XRCC1); (CYP1A1*2C/rs1048943/rs3547/single nucleotide polymorphisms/SNV/Mutation/Variants/Alleles /Genotypes). Our case-control data of the case-control analysis of the East Indian population were included in this meta-analysis. Similar, inclusion and exclusion criteria to the above meta-analysis were followed, except that the selection of the studies was not confined to any particular geographical region or ethnicity. The variant, rs1048943 (CYP1A1), was analysed in the dominant model only, while rs3547 (XRCC1) was analysed in additive, dominant and recessive models. Similar categories of data were extracted from the selected studies as in the above meta-analysis. Heterogeneity between studies was assessed using Cochran’s Q test (p < 0.1). A meta-analysis following a similar protocol, as mentioned above, was conducted. The genotypes were stratified by histological subtypes and geographical regions reported in the selected studies. This was followed by subgroup analyses and an effect modification test using a fixed-effect meta-regression model for smoking status.

**Results**

**Study characteristics**

Systematic mining of the databases with the search strings as mentioned above revealed 1060 hits, which were screened down to 50 studies following the specific inclusion/exclusion criteria set for the proposed study (Figure 1). These 50 studies included 24 polymorphisms from 12 genes with 9,487 cases and 10,455 controls (Table 1). Covariate specific case-control data, particularly tobacco smoking, mean age, histological status, and geographical region of the subjects were recorded from all 50 studies selected for meta-analysis (Supplementary Information, Table S1).

**Meta-analysis for the overall association of polymorphic variants with lung cancer**

Analysis of heterogeneity between studies using I^2 metric revealed 5 variants, such as GSTP1/rs1695, rs10425222/TP53, rs9282861/SULT1A1, rs25487/XRCC1 and rs3025039/VEGFA, to show mid-grade heterogeneity in a dominant model. Similarly, rs1048943/CYP1A1 showed low grade (25% <= I^2 <= 50%) heterogeneity and rs2010963/VEGFA showed high grade (I^2 > 75%) heterogeneity in the dominant model. In the additive model, rs9282861/SULT1A1, rs25487/XRCC1, rs3547/XRCC1, rs2010963/VEGFA, and rs419558/DKK2 showed low grade heterogeneity while rs10425222/TP53 showed mid-grade (50% <= I^2 <= 75%) heterogeneity. In the recessive model, del1/GSTT1, del2/GSTM1, rs9282861/SULT1A1, rs2010963/VEGFA, and rs447722/DKK2 showed low grade heterogeneity, and rs3547/XRCC1 showed high-grade heterogeneity. Cochran’s Q test also revealed the presence of significant heterogeneity between studies (P<0.1) for 8 variants, including rs1695/GSTP1, rs10425222/TP53, rs9282861/SULT1A1, rs25487/XRCC1, rs1048943/CYP1A1, rs3025039/VEGFA, VEGFA/rs2010963, and rs25487/XRCC1 in the dominant model only. The results of the fixed-effect and random-effects meta-analysis are shown in Table 2. The fixed-effect (FE) meta-analysis gives an average effect under the assumption that all effect estimates are assessing a common underlying effect. The random-effects (RE) meta-analysis is a variation of the FE method, which incorporates the assumption that different studies estimate different yet related effects. While the FE model is more powerful in the absence of heterogeneity, the RE model gives more robust estimates of the overall effect, where large studies are unlikely to dominate, particularly in the presence of heterogeneity. Therefore, the results for both FE and RE models were discussed in this meta-analysis.

Fixed-effect meta-analysis revealed a nominal association of del1/GSTT1 (null vs. present: OR=1.36; 95% CI=1.03-1.79; p = 0.028) and del2/GSTM1 (null vs. present: OR=1.38; 95% CI=1.09-1.75; p = 0.008) with overall lung cancer risk in the in the recessive model (Table 2). The association of del1/GSTT1 and
del2/GSTM1 with overall lung cancer risk in a recessive model is depicted in forest plots (Supplementary Information, Fig. S3). Similarly, the fixed-effect meta-analysis revealed that the variant rs17037102/ DKK2 was nominally associated with the overall lung cancer risk in both additive (AA vs. AG vs. GG: OR = 1.81; 95% CI = 1.03-3.15; p = 0.038) and dominant (AA+AG vs. GG: OR = 1.82; 95% CI = 1.03-3.22; p = 0.04) models (Table 2). The association of rs17037102/ DKK2 with overall lung cancer risk is depicted in forest plots for both additive and dominant models (Supplementary Information, Fig. S4). Moreover, fixed-effect meta-analysis of rs3547/XRCC1 showed an association with the overall lung cancer risk in the sample population of the study in additive (AA vs. AG vs. GG: OR = 1.83; 95% CI = 1.23-2.90; p = 0.004) dominant (AA+AG vs. GG: OR = 2.72; 95% CI = 1.56-4.73; p = 0.0004), and recessive (AA vs. AG+GG: OR = 2.04; 95% CI = 1.19-3.51; p = 0.01) models (Table 2). The association of rs3547/XRCC1 with overall lung cancer risk is depicted in forest plots for additive and recessive models (Supplementary Information, Fig. S5). The variant rs4646903/CYP1A1 also showed a nominal association with overall lung cancer risk in the population in a dominant model (CC+CT vs. TT: OR = 1.48, 95% CI = 1.93-1.95, p = 0.005). The association is depicted in forest plots in (Supplementary Information, Fig. S6).

Interestingly for the variant rs1048943/CYP1A1, most of the signal was driven by a single study [31], which could be the reason behind the observed heterogeneity between studies. A random-effects meta-analysis revealed that rs1048943/CYP1A1 was significantly associated with the overall lung cancer risk in the Indian subcontinent in a dominant (AG+GG vs. AA: OR = 2.07; 95% CI = 1.32-3.27, p = 0.002) model (Table 2). However, adjustment for multiple testing by Benjamini-Hochberg false discovery rate (FDR) correction (P_{FDR}=0.05) revealed a significant association of rs1048943/CYP1A1 (P_{FDR} = 0.027) and rs3547/XRCC1 (P_{FDR} = 0.027) in a dominant model (Table 2). The associations of rs1048943/CYP1A1 and rs3547/XRCC1 with overall lung cancer risk in a dominant model are depicted in forest plots (Figure 2).

### Evaluation of Publication Bias

To check for potential publication bias, we looked at funnel plots constructed of odds ratios and standard errors of the analysis. Visual inspection of the funnel plots for del1/GSTT1 and del2/GSTM1 (Supplementary Information, Fig. S3). For rs4646903/CYP1A1, the funnel plots are summarised in the following Fig. (Supplementary Information, Fig. S6), and rs1048943/CYP1A1 (Figure 2 A) did not show any apparent asymmetry. Furthermore, Egger's regression test did not find significant evidence of asymmetry in the funnel plots for rs4646903/CYP1A1 (p = 0.93), and del2/GSTM1 (p = 0.07). Thus, we did not find evidence for significant publication bias for the variants studied here.

### Association based on the Geographical Region

Only 6 out of 24 variants were stratified according to the geographical zones of the Indian subcontinent. The remaining variants were reported from the same zones of the subcontinent, as mentioned above. Hence, they were not considered in this analysis. The analysis revealed del1/GSTT1, and rs1048943/CYP1A1 to be associated with lung cancer risk in the population of the "North" and "South" zone of the subcontinent in a recessive and dominant model, respectively (Supplementary Information, Table S2). The variant del2/GSTM1 was found to be associated with lung cancer risk in the population of the "North" zone of the subcontinent only in a recessive model, while the variant rs4646903/CYP1A1 has been found to show association with the population of the "East" zone only in a dominant model (Supplementary Information, Table S2). The variant rs25487/XRCC1 was found to be associated with lung cancer in the population of the "South" zone of the subcontinent in additive and dominant models, but also held association with lung cancer in the population of the "East" zone in a recessive model of analysis (Supplementary Information, Table S2). The variant rs1042522/TP53 holds an association with lung cancer in the "East" zone of the subcontinent (Supplementary Information, Table S2).

### Association of variants stratified by Histological Subtypes

The studies selected for this meta-analysis broadly reported three histological subtypes of lung cancer, such as squamous cell carcinoma, adenocarcinoma, and small cell lung carcinoma with case-control genotype count. Out of 24, only 5 variants like del1/GSTT1, del2/GSTM2, rs4646903/ CYP1A1, rs1048943/CYP1A1, and rs1042522/TP53 could be stratified based on the three histological subtypes. The remaining variants lacked relevant case-control genotype data stratified by histology. The histological subtype-stratified analysis revealed del1/GSTT1 to be associated with Adenocarcinoma (null vs. present: OR = 1.82, 95% CI = 1.06-3.11; p = 0.03) in a recessive model with no heterogeneity between studies. Similarly, a significant association of rs4646903/CYP1A1 with Squamous cell carcinoma (TC+CC vs. TT: OR = 2.14, 95% CI = 1.04-4.41; p = 0.04) and rs1048943/CYP1A1 with lung Adenocarcinoma (AG+GG vs. AA: OR = 3.38, 95% CI = 1.83-6.25; p = 0.0001) was found in a dominant model with no heterogeneity.

Similarly, rs1048943/CYP1A1 was associated with Squamous cell carcinoma (AG+GG vs. AA: OR = 3.53; 95% CI = 2.05-6.08; p = 0.00005) with no heterogeneity. However, the variant rs1042522/TP53 was not associated with any of the histological subtypes. The association of these variants with different histological subtypes in different models of analysis are depicted in forest plots (Figure 3). The detailed results of the histology-stratified meta-analysis are summarised in (Table 3). The association of variants with specific histological subtypes justifies the genetic heterogeneity of lung cancer and differences in the pathogenesis of the subtypes, which could be useful in better prognosis and therapy.

### Subgroup Analysis- Association of variants based on Tobacco Smoking status

Subgroup analysis for del1/GSTT1, del2/GSTM1, rs4646903/CYP1A1, rs1048943/CYP1A1, rs25487/XRCC1, and rs9282861/SULT1A1 was done using the smoking status-stratified summary data separately within the "Smoker" and "Non-Smoker" sub-groups. Data for the remaining variants were not available in the selected studies for this meta-analysis. The analysis revealed rs4646903/CYP1A1 and rs25487/XRCC1 to be associated with lung cancer in smokers in a dominant (TC+CC vs. TT: OR = 2.26, 95% CI = 1.44-3.53; p = 0.0004) and recessive (AA vs. AG+GG: OR = 0.48, 95% CI = 0.27-0.86, p = 0.01) models, respectively (Table 4 A). No evidence of significant heterogeneity was found. Interestingly, the variant genotype 'AA' of rs25487/XRCC1 has been found to confer protection from lung cancer in the smoker population of the Indian subcontinent. However, the variant rs1048943/CYP1A1 was found to be
associated with lung cancer in both smokers (AA vs. AG+GG: OR = 2.23, 95% CI = 1.47-3.37; \( p = 0.0002 \)) and non-smokers (AA vs. AG+GG: OR = 1.74, 95% CI = 1.10-2.73; \( p = 0.02 \)) in a dominant model. Interestingly, the magnitude of association with smokers is higher than non-smokers. No significant heterogeneity was observed for the variant in both smokers and non-smokers. The association of these variants with smoking status in different models of analysis are depicted in forest plots (Figure 4). Furthermore, we tested for effect modification of the 6 polymorphic variants by smoking status using a fixed-effect meta-regression model with smoking status as a moderator variable. The difference in effect size (log-OR) between smokers and non-smokers is denoted by the coefficient of the moderator that revealed 2 polymorphic variants, viz. del1/\textit{GSTT1} (\( p = 0.015 \)) and rs4646903/\textit{CYP1A1} (\( p = 0.01 \)), has a considerable effect on lung cancer risk modified by smoking (Table 4 B).

**Association of rs1048943/\textit{CYP1A1} with lung cancer in a case-control dataset**

The detailed demographic and clinical attributes of the study sample from East India is summarised in (Supplementary Information, Table S3). Out of the 2 variants confirmed to be associated by the meta-analysis after FDR correction, namely rs3547/\textit{XRCC1} and rs1048943/\textit{CYP1A1}, the latter showed evidence of significant heterogeneity (\( Q = 1.93, I^2 = 48.32, p = 0.092 \)). For the variant rs1048943/\textit{CYP1A1}, a consistent pattern of association with lung cancer was observed from the Northern region of the subcontinent, but lack significant association with lung cancer in the Eastern region. We hypothesised that this heterogeneity might be explained by looking at covariate-specific and subgroup-stratified analysis. One of the major reasons for heterogeneity in the crude analysis is the uneven distribution of confounder/subgroups across studies. Hence, to understand the source of heterogeneity, we genotyped this polymorphic variant in a case-control lung cancer dataset comprising 101 cases and 413 controls from a representative East Indian population, where several covariates had been measured.

The age and pack-years of smoking were converted to categorical variables based on their respective mean values. The case-control analysis revealed no significant association of rs1048943 of \textit{CYP1A1} in additive effect model (GG vs GA vs. AA: OR = 1.24, 95% CI = 0.79-1.93; \( \beta_{ADJ} = 0.35 \)), dominant effect model (GG+GA vs. AA: OR = 1.33, 95% CI = 0.825-2.16; \( \beta_{ADJ} = 0.24 \)), and recessive effect model (GG vs AG+AA: OR = 0.49, 95% CI = 0.1-1.45; \( \beta_{ADJ} = 0.53 \)) with lung cancer adjusted for pack-years of smoking, age, alcohol consumption, tobacco and betel quid chewing, and asbestos exposure (Supplementary Information, Figure S4). The representative image of the RFLP analysis is depicted in (Supplementary Information, Figure S7). No significant association of rs1048943 was found in any of the covariate subgroups, such as pack-years of smoking, age, alcohol, tobacco and betel quid chewing, and asbestos exposure (Supplementary Information, Table S5). However, stratification of genotype count based on the histological subtypes revealed rs1048943 of \textit{CYP1A1} to be nominally associated with lung adenocarcinoma (ADC) in additive (GG vs GA vs. AA: OR = 1.75, 95% CI = 1.01-3.05; \( \beta_{ADJ} = 0.047 \)) and dominant (GG+GA vs. AA: OR = 1.99, 95% CI = 1.10-3.63; \( \beta_{DOMINANT} = 0.24 \)) effect models adjusted with pack-years, alcohol consumption, tobacco and betel quid chewing, and asbestos exposure (Supplementary Information, Table S6). The results indicate the probable implication of the variant in specific histopathology of lung cancer, supporting the notion of genetic heterogeneity and variability of lung tumour subtypes, both in origin and pathogenesis.

**Meta-analysis of rs1048943 (\textit{CYP1A1}) and rs3547 (\textit{XRCC1}) in world population**

The significant finding of the above study was replicated in a worldwide sample population. Literature search with the specific keywords revealed a total of 2617 hits for the variant rs1048943 (\textit{CYP1A1}), and 2224 hits for rs3547 (\textit{XRCC1}) published till 31st December 2019, worldwide. Our case-control association was included in the pool of hits, which increased the total number of hits for rs1048943 (\textit{CYP1A1}) to 2618. Following the specific inclusion/exclusion criteria, 49 studies (including 1 study reporting both the variants) were selected for the meta-analysis (Supplementary Information, Figure S8). These 48 studies included 2 variants from 2 genes with 12393 cases and 13841 controls. All the covariate and demographic summary data for the two variants are listed in a tabular form (Supplementary Information, Table S7).

Cochran's Q test showed evidence of significant heterogeneity between studies (\( p = 0.086 \)) for rs1048943/\textit{CYP1A1} only. Fixed-effect meta-analysis revealed a significant association of rs1048943/\textit{CYP1A1} (AG+GG vs. AA: OR = 1.22; 95% CI = 1.06-1.41; \( p = 0.07 \)) with lung cancer across the world population. The random-effects meta-analysis also revealed significant association of rs1048943/\textit{CYP1A1} (AG+GG vs. AA: OR = 1.22; 95% CI = 1.001-1.483; \( p = 0.048 \)) with lung cancer across the world population (Supplementary Information, Table S8). The analysis was performed only in the dominant model due to the data presented in the selected studies. The study-level and combined association of rs1048943/\textit{CYP1A1} with lung cancer have been depicted in forest plots (Supplementary Information, Figure S9). A fixed-effect meta-analysis revealed a significant association of rs3547/\textit{XRCC1} with lung cancer across the world population in all the three genetic models (Supplementary Information, Table S8). However, a random-effects meta-analysis showed a significant association of rs3547/\textit{XRCC1} with the disease in only additive and dominant models (Supplementary Information, Table S8). The study-level and combined association of rs3547/\textit{XRCC1} with lung cancer have been depicted in forest plots (Supplementary Information, Figure S10).

Qualitative assessment of the funnel plots revealed the lack of significant publication bias for rs1048943/\textit{CYP1A1} (Supplementary Information, Figure S9). Further, Egger's regression test estimated lack of significant asymmetry of the funnel plot for rs1048943/\textit{CYP1A1} (\( p = 0.96 \)).

Stratification of the case-control genotype data by histological subtypes of lung cancer revealed a significant association of rs1048943/\textit{CYP1A1} with Squamous cell carcinoma (AG+GG vs. AA: OR = 1.39, 95% CI = 1.07-1.82; \( p = 0.015 \)) in the fixed-effect model only (Supplementary Information, Table S9 and Figure S11). The variant, rs3547/\textit{XRCC1} lacked relevant case-control genotype data stratified by histology.

Subgroup analysis revealed significant association of rs1048943/\textit{CYP1A1} with lung cancer in smokers in a dominant model (AG+GG vs. AA: OR = 1.54, 95% CI = 1.17-2.02; \( p = 0.002 \)) (Supplementary Information, Table S10 and Figure S12). Furthermore, we tested for effect modification of the variant by smoking status using a fixed-effect meta-regression model with smoking status as a moderator variable. The difference in effect size (log-OR) between
Further, rs1048943/\textit{CYP1A1} genotype was stratified by the geographical region of the selected studies that revealed significant association (\(p < 0.05\)) of the variant with lung cancer in Indian and Australian population in both fixed-effect and random-effects models (Supplementary Information, Table S12 and Fig. S13).

**Discussion**

Our study presents the first comprehensive meta-analysis of 24 variants of 12 genes across 51 studies from the Indian subcontinent that provides an insight into the combined effect of each variant on overall and covariate-stratified lung cancer risk in the region. The lack of significant publication bias confirms that the results were not overestimated under the influence of any bias in the published articles. Although GWAS data mining revealed no significant association of rs1048943/\textit{CYP1A1} with lung cancer, it showed a significant association of the \textit{CYP1A1} gene with hypertension and habitual coffee consumption. Therefore, the association of the variant with lung cancer could be modified by coffee consumption or smoking tobacco. The variant rs1048943/\textit{CYP1A1} was found to be associated with lung cancer risk in East Asians \(^{47}\), which shows the colinearity of the findings of this study to the present study as discussed here. The \textit{CYP1A1} (Cytochrome P450, family 1, member A1; 15q22-24) gene encodes a bulky phase I endoplasmic xenobiotic metabolism enzyme, present in lung tissue. The enzyme catalyses the activation of reactive electrophilic compounds, including benzo[α]pyrenes and PAHs present in tobacco smoke \(^{48}\). It promotes the formation of DNA adducts, which imparts a genotoxic effect that could lead to DNA lesions and cause lung cancer. The variant rs1048943A > G of \textit{CYP1A1} locus causes a single amino acid substitution (Ile > Val) in the heme-binding region, which increases enzyme activity, enhancing the activation of procarcinogens in tobacco smoke. It influences the metabolism of environmental carcinogens, such as tobacco smoke that modifies lung cancer susceptibility \(^{48}\).

The superfamily of glutathione-S-transferases (GSTs) comprises of multifunctional enzymes that catalyse the conjugation of reduced tripeptide glutathione to various electrophilic and hydrophobic substrates resulting in their detoxification and effective elimination from the cell. Thus, they help to reduce the carcinogenic load accumulated due to smoking from the cells. The null genotype of the deletion polymorphisms of glutathione-S-transferase theta 1 (\textit{GSTT1}) and glutathione-S-transferase mu 1 (\textit{GSTM1}) are frequently found to be associated with lung cancer with evidence of effect modification by tobacco smoking. The null genotype is responsible for the lack of the enzyme within the cell conferring higher risk to lung cancer. Inconsistent reports on the association of the null genotypes of \textit{GSTT1} (del1) and \textit{GSTM1} (del2) led to confusion regarding their correct effect on the disease pathogenesis \(^{34,35}\).

Ethnicity/racial differences in the association of \textit{GSTT1} null genotype with lung cancer has been reported where the frequency of the null genotype was found to be significantly higher in Asians than in Caucasians \(^{49}\).

The Dickkopf-related protein 2 (\textit{DKK2}) gene encodes a secretory protein belonging to the Dickkopf family. The protein, DKK2 bears two cysteine-rich regions and is involved in embryonic development through the Wnt/β-catenin signalling pathway. DKK2 exhibits a bimodal function as an agonist or antagonist of the Wnt/β-catenin signalling pathway \(^{50}\) depending on the cellular context and the presence of the co-factor kremen2. The activity of DKK2 is modulated the Wnt co-receptor, LDL-receptor related protein 5 (LRPS) and –6 (LRP6) \(^{51}\). Aberrant expression of DKK2 has been observed in many tumours, including epigenetic silencing of the expression of DKK2 in ovarian carcinoma \(^{52}\), hepatocellular carcinoma \(^{53}\), and renal carcinoma \(^{54}\). RNAi-mediated silencing of \textit{DKK2} is frequently observed in tongue squamous cell carcinoma \(^{55}\), and oesophageal adenocarcinoma \(^{56}\). These reports are suggestive of the anti-tumour effect of DKK2. However, the upregulation of DKK2 promotes cell proliferation and invasion through the Wnt signalling pathway in prostate cancer \(^{57}\), Ewing sarcoma \(^{58}\), and colorectal cancer \(^{59}\). Thus, the cellular context-dependent function of DKK2 is very complex, which is evident from the above examples. DKK2 has been found to promote angiogenesis, which is distinct from VEGF-dependent angiogenesis \(^{60}\) in the formation of closer interconnections of the vessels.

Interestingly, Dkk2-induced blood vessels consistently show higher coverage of endothelial cells (ECs) by pericytes and smooth muscle cells (SMCs), which are involved in vessel maturity and stability. Dkk2-mediated angiogenesis consists of a signalling cascade induced through LRP6-mediated APC/Asef2/Cdc42 activation. DKK2 promotes tumour progression by suppressing cytotoxic immune cell activation in colorectal carcinoma \(^{61}\) and NSCLC \(^{62}\) with \textit{APC} mutations. In a recent study \(^{26}\), the heterozygous genotype of rs17037102/\textit{DKK2} and rs419558/\textit{DKK2} confer an increased risk of lung cancer. A combination of all the 3 genotypic variants of \textit{DKK2} confers a four-fold increase to lung cancer risk.

The protein encoded by \textit{XRCC1} (X-ray repair cross-complementing 1; 19q13.31) performs an efficient repair of single-strand DNA breaks formed by the exposure to ionising radiation and alkylating agents. \textit{XRCC1} interacts with DNA ligase III, polymerase beta and poly (ADP-ribose) polymerase to participate in the base excision repair. The protein plays a role in DNA processing during meiosis and DNA recombination in germinal epithelial cells. Moreover, \textit{XRCC1} harbours a rare microsatellite polymorphism, which is associated with varying radiosensitivity in cancer \(^{63}\). Polymorphisms of \textit{XRCC1}, like Arg194Trp (exon 7), Arg280His (exon 10) and Arg399Gln (exon 11) were reported to confer increased risk to lung cancer \(^{64-66}\) with inconsistencies across different populations \(^{67-71}\).

The association of variants with different histological subtypes of lung cancer revealed del1/\textit{GSTT1} to be associated with lung adenocarcinoma, rs4646903/\textit{CYP1A1} with lung squamous cell carcinoma while rs1048943/\textit{CYP1A1} with both lung adenocarcinoma and lung squamous cell carcinoma. Thus, stratification of the genotypes based on the histological subtypes of lung cancer improved risk assessment potential. Identification of subtypes specific genetic risk markers helps to design targeted early detection and prevention strategies. Moreover, the identification of histotype associated genetic
markers may define the mechanism underlying the currently unknown origins of morphological variations that could contribute to the development of personalised treatment modalities for subtype-specific lung cancer cases.

Furthermore, subgroup analysis of 6 variants stratified by smoking status revealed rs1048943 of the CYP1A1 gene to be significantly associated with lung cancer in both smokers ($\rho = 0.0002$) and non-smokers ($\rho = 0.02$). However, the meta-regression analysis revealed the absence of any effect modification of rs1048943 on lung cancer by smoking ($\rho = 0.43$), which implies that the polymorphism has no modifier effect on lung cancer by smoking. The variant rs4646903 of the CYP1A1 gene ($\rho = 0.0004$) show an association with lung cancer in smokers only. Interestingly, significant effect modification of del1 of the GSTT1 gene ($\rho = 0.015$) and rs4646903 of the CYP1A1 gene ($\rho = 0.01$) on lung cancer by smoking was observed by meta-regression analysis, which suggested the importance of the variants in modifying the risk of lung cancer by smoking status.

Based on the meta-regression analysis, there is no significant effect modification for the remaining variants, although it can be surmised that there may be interaction in the biological mechanisms leading to lung cancer. Subgroup analysis based on covariates, such as age, sex, ethnicity, exposure types and dose was not done due to lack of sufficient reports on the population of the Indian subcontinent. The Indian subcontinent consists of a highly heterogeneous population with considerable admixture among different ethnicities, which could modify the linkage disequilibrium structure of the population. This could contribute to significant heterogeneity between the studies. Therefore, the geographical region based association of the variants was performed to assess the effect estimates of the variants in different geospatial clusters of the Indian subcontinent.

The variant rs1048943A > G of CYP1A1 locus is a non-synonymous polymorphic variant, which imparts an individual effect on lung cancer risk in various populations. On the other hand, rs3547/XRCC1 is a synonymous polymorphic variant that has shown no association with lung cancer in the Korean Chinese, Latino, and African-American population. In the Korean population, the variant shows association with lung cancer in a haplotype with other XRCC1 variants like rs3547-rs25487-rs25486-rs2293036-rs1799778-rs1001581-rs12611088-rs3213282. The meta-analysis was followed by a case-control analysis in the East Indian population, which revealed no association of rs1048943A > G of CYP1A1 with overall lung cancer risk but with lung adenocarcinoma adjusted with covariate status. Thus, our case-control study reveals rs1048943/CYP1A1 to be a histological subtype-specific variant for lung cancer in the East Indian population, which could be a potential target for personalised therapy and histology specific drug designing for lung cancer patients. The finding shows colinearity with the outcomes of the current meta-analysis. The studies, included in this meta-analysis, reported from the Eastern region of the subcontinent also shows lack of association of the risk genotype (GG) of the polymorphic variant rs1048943/CYP1A1 as summarised in (Table 1 and Supplementary Information, Table S1).

Our replication meta-analysis across the world population justifies the role of the variants rs1048943 (CYP1A1) and rs3547 (XRCC1) in conferring overall lung cancer risk, particularly in smokers, with a higher power. Interestingly, rs1048943 (CYP1A1) shows no effect modification by smoking status on lung cancer risk that is indicative of the association in smokers as a random occurrence by chance.

In the larger sample size, the variant rs1048943 (CYP1A1) shows an association with squamous cell carcinoma, which is indicative of a population-specific effect of the variant on different histological subtypes of lung cancer. The association of rs1048943 (CYP1A1) across various populations identifies the relevance of the variant in lung cancer risk in a population-specific manner, which could be critical in designing personalised treatment and precision medicine for patients of diverse populations.

Interestingly, out of 12 selected genes for meta-analysis 5 genes belong to the xenobiotic metabolism pathway, 3 belong to the DNA repair pathway, 3 belong to the Wnt/β-catenin pathway regulating various physiological aspects of lung cancer, and 1 belong to the angiogenic pathway. The xenobiotic metabolism and DNA repair pathways could be the significant ‘modifier’ and ‘driver’ pathways leading to altered gene-environment interaction and development of lung cancer. Genes of xenobiotic metabolism pathways are involved in the metabolism and detoxification of tobacco smoke components to reduce intracellular carcinogenic load. Some genes of the xenobiotic metabolism pathway also induce bio-activation of procarcinogens into potent carcinogens that can quickly form DNA-adducts and subsequent mutagenesis. Further, genes belonging to the DNA repair pathway functions to repair DNA damage induced by tobacco smoking and radiation. Detailed text mining of the available reports following the inclusion criteria, revealed the association of xenobiotic metabolism genes (XMG) and DNA repair genes (DRG) to the risk of lung cancer development.

In the current study, we were unable to perform subgroup and meta-regression analysis for other covariate risk factors for all the variants due to unavailability of sufficient data in the selected studies.

The subtype-specific polymorphic variant identification as obtained in the current meta-analysis would suffice personalised therapy and development of precision medicine. The identification of genetic variants for which there is evidence of influence on lung cancer risk through meta-analysis may provide new insights into the fundamental biological pathways involved in the development of lung cancer that could help in future research. Further, identification of lung cancer risk variants may also be beneficial in the assessment of risk scores for accurate population risk stratification and decision making, which could be of potential value in targeting primary prevention and lung cancer screening modalities in a population-specific manner.

**Declarations**

**Acknowledgement**

The authors acknowledge the support provided to Department of Genetics, University of Calcutta by The Department of Science and Technology (DST), Government of India - Promotion of University Research and Scientific Excellence (DST-PURSE). We also thank Dr Arnab Gupta, Director of Saroj Gupta Cancer Centre and Research Institute, Thakurpukur, India for collaborating and granting permission for human sample collection. The authors thank all the
patients who participated in this study and acknowledge the efforts and contribution of the doctors, nurses, and hospital administration staff of Saroj Gupta Cancer Centre and Research Institute, Thakurpukur, India, Department of CHEST, IPGME&R, Kolkata, India.

Authors’ Contributions

DS, SB\(^1\) and MS: Conceptualization. SB\(^1\), DS: Data Curation. SB\(^2\), DS, and SB\(^1\): Methodology, Software, and Investigation. Methodology and Project Administration. RM, AS, TC, GB, SN, and SR: Resources. M. Sengupta: Supervision. SB\(^2\), MS, and DS: Formal analysis and Validation. DS: Writing-Original draft. DS, SB\(^2\) and MS: Writing-Reviewing and editing. SB\(^2\), MS: Supervision. SB\(^1\): 2\(^{nd}\) Author and SB\(^2\): 9\(^{th}\) Author. All the authors have reviewed the manuscript and have agreed for publication.

Additional Information

Conflict of Interest/Competing Interests

All the authors declare that they have no direct or indirect conflict of interest and competing interests for this article.

Ethics approval

The Ethics Committee of Saroj Gupta Cancer Centre and Research Institute (IEC SGCCRI Ref No- 2017/MS/1; dated: 11.10.2017), IPGME&R (Memo No. Inst/IEC/2015/545; dated: 10.12.2015), Kolkata and the University of Calcutta (Ref No: 0024/16-117/1434; dated: 24.10.2016), Kolkata, India; approved the study with human subjects as per the regulation of the Indian Council of Medical Research (ICMR) following the Declaration of Helsinki, 1964.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

The human participants have consented to the submission of the case report to the journal.

Availability of data and materials

All data generated or analysed during this study are included in this article and its supplementary information files. The additional raw input data files will be available from the corresponding authors on request.

Funding

This research did not receive any financial assistance from funding agencies in the public, commercial, or not-for-profit sectors.

Code Availability

The R scripts/codes used for the analysis will be available from the corresponding author on request.

References

[1] Doll, R., Peto, R., Boreham, J. & Sutherland, I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. Br. J. Cancer\(^{92}\), 426-429, doi:10.1038/sj.bjc.6602359 (2005).

[2] Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J. Clin.\(^{68}\), 394-424, doi:10.3322/caac.21492 (2018).

[3] Ferlay, J. et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer\(^{127}\), 2893-2917, doi:10.1002/ijc.25516 (2010).

[4] Malik, P. S. & Raina, V. Lung cancer; prevalent trends & emerging concepts. Indian J. Med. Res.\(^{141}\), 5-7, doi:10.4103/0971-5916.154479 (2015).

[5] Thankappan, K. R. & Thresia, C. U. Tobacco use & social status in Kerala. Indian J. Med. Res.\(^{126}\), 300-308 (2007).

[6] Murray, C. J. & Lopez, A. D. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet\(^{349}\), 1498-1504, doi:10.1016/S0140-6736(96)07492-2 (1997).

[7] Doll, R. & Hill, A. B. Smoking and carcinoma of the lung; preliminary report. Br. Med. J.\(^{2}\), 739-748, doi:10.1136/bmj.2.4682.739 (1950).

[8] Ihsan, R. et al. Copy number polymorphism of glutathione-S-transferase genes (GSTM1 & GSTT1) in susceptibility to lung cancer in a high-risk population from north-east India. Indian J. Med. Res.\(^{139}\), 720-729 (2014).

[9] Ihsan, R. et al. Investigation on the role of p53 codon 72 polymorphism and interactions with tobacco, betel quid, and alcohol in susceptibility to cancers in a high-risk population from North East India. DNA Cell Biol.\(^{30}\), 163-171, doi:10.1089/dna.2010.1119 (2011).
[10] Behera, D. Epidemiology of lung cancer-Global and Indian perspective. *J. Indian Acad. Clin. Med.* **13**, 131-137 (2012).

[11] Behera, D. & Balamugesh, T. Lung cancer in India. *Indian J. Chest Dis. Allied Sci.* **46**, 269-281 (2004).

[12] Phukan, R. K. *et al.* Role of household exposure, dietary habits and glutathione S-Transferases M1, T1 polymorphisms in susceptibility to lung cancer among women in Mizoram India. *Asian Pac. J. Cancer Prev.* **15**, 3253-3260, doi:10.7314/apjcp.2014.15.7.3253 (2014).

[13] Kurmi, O. P., Arya, P. H., Lam, K. B., Sorahan, T. & Ayres, J. G. Lung cancer risk and solid fuel smoke exposure: a systematic review and meta-analysis. *Eur. Respir. J.* **40**, 1228-1237, doi:10.1183/09031936.00099511 (2012).

[14] Shen, H. *et al.* Global lung cancer risk from PAH exposure highly depends on emission sources and individual susceptibility. *Sci. Rep.* **4**, 6561, doi:10.1038/srep06561 (2014).

[15] Travis, W. D. Update on small cell carcinoma and its differentiation from squamous cell carcinoma and other non-small cell carcinomas. *Mod. Pathol.* **25** Suppl 1, S18-30, doi:10.1038/modpathol.2011.150 (2012).

[16] Doll, R. Mortality from lung cancer in asbestos workers. *Br. J. Ind. Med.* **12**, 81-86, doi:10.1136/oem.12.2.81 (1955).

[17] Cohen, A. J. Outdoor air pollution and lung cancer. *Environ. Health Perspect.* **108** Suppl 4, 743-750, doi:10.1289/ehp.00108s4743 (2000).

[18] Loomis, D., Huang, W. & Chen, G. The International Agency for Research on Cancer (IARC) evaluation of the carcinogenicity of outdoor air pollution: focus on China. *Chin. J. Cancer* **33**, 189-196, doi:10.5732/cjc.014.10028 (2014).

[19] Smolle, E. & Pichler, M. Non-Smoking-Associated Lung Cancer: A distinct Entity in Terms of Tumor Biology, Patient Characteristics and Impact of Hereditary Cancer Predisposition. *Cancers (Basel)* **11**, doi:10.3390/cancers11020204 (2019).

[20] Lan, Q. *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat. Genet.* **44**, 1330-1335, doi:10.1038/ng.2456 (2012).

[21] McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **49**, 1126-1132, doi:10.1038/ng.3892 (2017).

[22] Dong, J. *et al.* Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat. Genet.* **44**, 895-899, doi:10.1038/ng.2351 (2012).

[23] Wang, Y. *et al.* Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.* **40**, 1407-1409, doi:10.1038/ng.273 (2008).

[24] Bag, A., Bag, N., Jeena, L. M. & Jyala, N. S. Glutathione S-transferase T1 and myeloperoxidase -463 G>A genotypes in lung cancer patients of Kumaun region. *J. Nat. Sci. Biol. Med.* **5**, 293-296, doi:10.4103/0976-9668.136169 (2014).

[25] Bahl, C., Sharma, S., Singh, N. & Behera, D. Association study between genetic variations in Axin2 gene and lung cancer risk in North Indian population: A multiple interaction analysis. *Tumour Biol.* **39**, 1010428317695533, doi:10.1177/1010428317695533 (2017).

[26] Bahl, C., Singh, N., Behera, D. & Sharma, S. Association of polymorphisms in Dickopff (DKK) gene towards modulating risk for lung cancer in north Indians. *Future Oncol.* **13**, 213-232, doi:10.2217/fon-2016-0117 (2017).

[27] Bahl, C., Singh, N., Behera, D. & Sharma, S. High-order gene interactions between the genetic polymorphisms in Wnt and AhR pathway in modulating lung cancer susceptibility. *Per. Med.* **14**, 487-502, doi:10.2217/pme-2017-0018 (2017).

[28] Bahl, C., Singh, N., Behera, D. & Sharma, S. Genetic Variants in the Wingless Antagonist Genes (sFRP, DKK, and Axin2) Predict the Overall Survival and Prognosis of North Indian Lung Cancer Patients Treated with Platinum-Based Doublet Chemotherapy. *Cancer Biother. Radiopharm.*, doi:10.1089/cbr.2018.2491 (2018).

[29] Lawania, S., Singh, N., Behera, D. & Sharma, S. Xeroderma pigmentosum complementation group D polymorphism toward lung cancer susceptibility and response in patients treated with platinum chemotherapy. *Future Oncol.* **13**, 2645-2665, doi:10.2217/fon-2017-0211 (2017).

[30] Natukula, K., Jamil, K., Pingali, U. R., Attili, V. S. & Madireddy, U. R. The codon 399 Arg/Gln XRCC1 polymorphism is associated with lung cancer in Indians. *Asian Pac. J. Cancer Prev.* **14**, 5275-5279, doi:10.7314/apjcp.2013.14.9.5275 (2013).

[31] Peddireddy, V. *et al.* Association of CYP1A1, GSTM1 and GSTT1 gene polymorphisms with risk of non-small cell lung cancer in Andhra Pradesh region of South India. *Eur. J. Med. Res.* **21**, 17, doi:10.1186/s40001-016-0209-x (2016).

[32] Peddireddy, V., Badabagni, S. P., Gundimeda, S. D. & Mundiraju, H. P. Association of eNOS and ACE gene polymorphisms and plasma nitric oxide with risk of non-small cell lung cancer in South India. *Clin. Respir. J.* **12**, 207-217, doi:10.1111/crj.12517 (2018).

[33] Tilak, A. R., Kumar, S., Pant, M. C., Mathur, N. & Kumar, A. Polymorphism Arg72Pro of p53 confers susceptibility to squamous cell carcinoma of lungs in a North Indian population. *DNA Cell Biol.* **32**, 66-72, doi:10.1089/dna.2012.1792 (2013).
[34] Sobti, R. C., Sharma, S., Joshi, A., Jindal, S. K. & Janmeja, A. Genetic polymorphism of the CYP1A1, CYP2E1, GSTM1 and GSTT1 genes and lung cancer susceptibility in a north Indian population. *Mol. Cell. Biochem.* **266**, 1-9, doi:10.1023/b:mcbi.0000049127.33458.87 (2004).

[35] Sreeja, L. *et al.* Possible risk modification by CYP1A1, GSTM1 and GSTT1 gene polymorphisms in lung cancer susceptibility in a South Indian population. *J. Hum. Genet.* **50**, 618-627, doi:10.1007/s10038-005-0303-3 (2005).

[36] Girdhar, Y., Singh, N., Behera, D. & Sharma, S. Synergistic association of CYP1A1 polymorphisms with increased susceptibility to squamous cell lung cancer in north Indian smokers. *Int. J. Biol. Markers* **31**, e402-e412, doi:10.5301/jbm.5000221 (2016).

[37] Naushad, S. *M. et al.* Population-level diversity in the association of genetic polymorphisms of one-carbon metabolism with breast cancer risk. *J. Community Genet.* **7**, 279-290, doi:10.1007/s12687-016-0277-1 (2016).

[38] Pradhan, S. *et al.* Indian genetic disease database. *Nucleic Acids Res.* **39**, D933-938, doi:10.1093/nar/gkq1025 (2011).

[39] Lee, Y. H. Meta-analysis of genetic association studies. *Ann. Lab. Med.* **35**, 283-287, doi:10.3343/alm.2015.35.3.283 (2015).

[40] Sengupta, D., Guha, U., Bhattacharjee, S. & Sengupta, M. Association of 12 polymorphic variants conferring genetic risk to lung cancer in Indian population: An extensive meta-analysis. *Environ. Mol. Mutagen.* **58**, 688-700, doi:10.1002/em.22149 (2017).

[41] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & Group, P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* **339**, b2535, doi:10.1136/bmj.b2535 (2009).

[42] Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F. & Botella, J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol. Methods* **11**, 193-206, doi:10.1037/1082-989X.11.2.193 (2006).

[43] W. V. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **3**, 1-48 (2010).

[44] Sengupta, D. *et al.* Meta-Analysis of Polymorphic Variants Conferring Genetic Risk to Cervical Cancer in Indian Women Supports CYP1A1 as an Important Associated Locus. *Asian Pac. J. Cancer Prev.* **19**, 2071-2081, doi:10.22034/APJCP.2018.19.8.2071 (2018).

[45] Martorell-Marugan, J., Toro-Dominguez, D., Alarcon-Riquelme, M. E. & Carmona-Saez, P. MetaGenyo: a web tool for meta-analysis of genetic association studies. *BMC Bioinformatics* **18**, 563, doi:10.1186/s12859-017-1990-4 (2017).

[46] Sterne, J. A. & Egger, M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J. Clin. Epidemiol.* **54**, 1046-1055, doi:10.1016/s0895-4356(01)00377-8 (2001).

[47] Liu, H. X., Li, J. & Ye, B. G. Correlation between gene polymorphisms of CYP1A1, GSTP1, ERCC2, XRCC1, and XRCC3 and susceptibility to lung cancer. *Genet. Mol. Res.* **15**, doi:10.4238/gmr15048813 (2016).

[48] Guengerich, F. P. & Shimada, T. Activation of procarcinogens by human cytochrome P450 enzymes. *Mutat. Res.* **400**, 201-213, doi:10.1016/s0027-5107(98)00037-2 (1998).

[49] Parl, F. F. Glutathione S-transferase genotypes and cancer risk. *Cancer Lett.* **221**, 123-129, doi:10.1016/j.canlet.2004.06.016 (2005).

[50] Mao, B. & Niehrs, C. Kremen2 modulates Dickkopf2 activity during Wnt/LRP6 signaling. *Gene* **302**, 179-183, doi:10.1016/j.smgene.2012.01.010 (2013).

[51] Bao, J., Zheng, J. J. & Wu, D. The structural basis of DKK-mediated inhibition of Wnt/LRP signaling. *Sci. Signal.* **5**, pe22, doi:10.1126/scisignal.2003028 (2012).

[52] Zhu, J., Zhang, S., Gu, L. & Di, W. Epigenetic silencing of DKK2 and Wnt signal pathway components in human ovarian carcinoma. *Carcinogenesis* **33**, 2334-2343, doi:10.1093/carcin/bgs278 (2012).

[53] Fatima, S., Luk, J. M., Poon, R. T. & Lee, N. P. Dysregulated expression of dickkopfs for potential detection of hepatocellular carcinoma. *Expert. Rev. Mol. Diagn.* **14**, 535-548, doi:10.1586/14737159.2014.915747 (2014).

[54] Hirata, H. *et al.* Wnt antagonist gene DKK2 is epigenetically silenced and inhibits renal cancer progression through apoptotic and cell cycle pathways. *Clin. Cancer Res.* **15**, 5678-5687, doi:10.1158/1078-0432.CCR-09-0558 (2009).

[55] Kawakita, A. *et al.* MicroRNA-21 promotes oral cancer invasion via the Wnt/beta-catenin pathway by targeting DKK2. *Pathol. Oncol. Res.* **20**, 253-261, doi:10.1007/s12253-013-9689-y (2014).

[56] Wang, Y. *et al.* miR-221 Mediates Chemoresistance of Esophageal Adenocarcinoma by Direct Targeting of DKK2 Expression. *Ann. Surg.* **264**, 804-814, doi:10.1097/SLA.0000000000001928 (2016).

[57] Xu, W. *et al.* Dickkopf 2 promotes proliferation and invasion via Wnt signaling in prostate cancer. *Mol. Med. Rep.* **14**, 2283-2288, doi:10.3892/mmr.2016.5502 (2016).
[58] Hauer, K. et al. DKK2 mediates osteolysis, invasiveness, and metastatic spread in Ewing sarcoma. *Cancer Res.* **73**, 967-977, doi:10.1158/0008-5472.CAN-12-1492 (2013).

[59] Matsui, A. et al. DICKKOPF-4 and -2 genes are upregulated in human colorectal cancer. *Cancer Sci.* **100**, 1923-1930, doi:10.1111/j.1349-7006.2009.01272.x (2009).

[60] Min, J. K. et al. The WNT antagonist Dickkopf2 promotes angiogenesis in rodent and human endothelial cells. *J. Clin. Invest.* **121**, 1882-1893, doi:10.1172/JCI42556 (2011).

[61] Xiao, Q. et al. DKK2 imparts tumor immunity evasion through beta-catenin-independent suppression of cytotoxic immune-cell activation. *Nat. Med.* **24**, 262-270, doi:10.1038/nm.4496 (2018).

[62] Shen, T., Chen, Z., Qiao, J., Sun, X. & Xiao, Q. Neutralizing monoclonal antibody against Dickkopf2 impairs lung cancer progression via activating NK cells. *Cell Death Discov.* **5**, 123, doi:10.1038/s41420-019-0204-4 (2019).

[63] Chen, L., Zhuo, D., Chen, J. & Yuan, H. XRCC1 polymorphisms and lung cancer risk in Caucasian populations: a meta-analysis. *Int. J. Clin. Exp. Med.* **8**, 14969-14976 (2015).

[64] Guo, S. et al. The relationship between XRCC1 and XRCC3 gene polymorphisms and lung cancer risk in northeastern Chinese. *PLoS One* **8**, e56213, doi:10.1371/journal.pone.0056213 (2013).

[65] Kiyohara, C., Takayama, K. & Nakanishi, Y. Lung cancer risk and genetic polymorphisms in DNA repair pathways: a meta-analysis. *J. Nucleic Acids* **2010**, 701760, doi:10.4061/2010/701760 (2010).

[66] Qian, B. et al. Association of genetic polymorphisms in DNA repair pathway genes with non-small cell lung cancer risk. *Lung Cancer* **73**, 138-146, doi:10.1016/j.lungcan.2010.11.018 (2011).

[67] Gao, W. M. et al. Association of the DNA repair gene XRCC1 and XRCC3 gene polymorphisms with lung cancer risk in northeastern Chinese. *Carcinogenesis* **24**, 1671-1676, doi:10.1093/carcin/bgg115 (2003).

[68] Hou, S. M. et al. Influence of common XPD and XRCC1 variant alleles on p53 mutations in lung tumors. *Environ. Mol. Mutagen.* **41**, 37-42, doi:10.1002/em.10128 (2003).

[69] Kiyohara, C., Takayama, K. & Nakanishi, Y. Association of genetic polymorphisms in the base excision repair pathway with lung cancer risk: a meta-analysis. *Lung Cancer* **54**, 267-283, doi:10.1016/j.lungcan.2006.08.009 (2006).

[70] Li, Y. et al. Assessment of the association between XRCC1 Arg399Gln polymorphism and lung cancer in Chinese. *Tumour Biol.* **34**, 3681-3685, doi:10.1007/s13277-013-0950-5 (2013).

[71] Zheng, H., Wang, Z., Shi, X. & Wang, Z. XRCC1 polymorphisms and lung cancer risk in Chinese populations: a meta-analysis. *Lung Cancer* **65**, 268-273, doi:10.1016/j.lungcan.2009.02.002 (2009).

[72] Moorjani, P. et al. Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422-438, doi:10.1016/j.ajhg.2013.07.006 (2013).

[73] Ji, Y. N., Wang, Q. & Suo, L. J. CYP1A1 Ile462Val polymorphism contributes to lung cancer susceptibility among lung squamous carcinoma and smokers: a meta-analysis. *PLoS One* **7**, e43397, doi:10.1371/journal.pone.0043397 (2012).

[74] Ihsan, R. et al. Multiple analytical approaches reveal distinct gene-environment interactions in smokers and non smokers in lung cancer. *PLoS One* **6**, e29431, doi:10.1371/journal.pone.0029431 (2011).

[75] Islam, M. S. et al. Lung cancer risk in relation to nicotinic acetylcholine receptor, CYP2A6 and CYP1A1 genotypes in the Bangladeshi population. *Clin. Chim. Acta* **416**, 11-19, doi:10.1016/j.cca.2012.11.011 (2013).
| Selected Studies   | PMID          | Number of Cases/Controls | Mean Age±SD Cases | Mean Age±SD Controls | Sex of Cases | Sex of Controls | Number of Smokers in Cases/Controls | Zone of the Indian subcontinent |
|-------------------|---------------|--------------------------|-------------------|----------------------|--------------|----------------|-------------------------------------|----------------------------------|
| Bag et al.2014    | 25097401      | 26/33                    | 56.11±9.22        | 28.68±6.77           | 24M, 2F     | 33M            | (23/08)                             | North                            |
| Shukla et al.2013 | 23803127      | 218/238                  | 56.14±11.91       | 56.15±7.84           | 189M,29F    | 191M,47F       | 128/34                              | North                            |
| Sobti et al.2004  | 15646021      | 100/76                   | 55.5 ± 11.3       | 50.9 ± 8.1           | 95M,5F      | 73M,3F         | 86/59                               | North                            |
| Sharma et al.2015 | 26529288      | 270/270                  | 57.39 ± 10.60     | 53.23 ± 10.49        | 235M,35F    | 233M,37F       | 212/193                             | North                            |
| Kumar et al.2009  | 19009239      | 93/253                   | 42.6 ± 6.3        | 39.8 ± 5.4           | 81M,12F     | 203M,50F       | 81/101                              | North                            |
| Ihsan et al.2014  | 25027082      | 154/154                  | 59.16 ± 9.95      | 60.39 ± 10.43        | 38 M,116F   | 38 M,116F      | 105/71                              | East                             |
| Phukan et al.2014 | 24815479      | 230/460                  | 59.01±13.02       | 58.60±13.99          | 230F        | 460F           | 126/153                             | East                             |
| Ihsan et al.2011  | 22206016      | 188/209                  | 60.41±10.58       | 57.19±10.75          | 145M,43F    | 159M,50F       | 132/139                             | East                             |
| Sreeja et al.2008 | 18472644      | 111/111                  | 57.82 ± 11.74     | 56.21 ± 10.36        | 82M,29F     | 84M,27F        | 44/19                               | South                            |
| Peddireddy et al.2016 | 27090234   | 246/250                  | 57.57 ± 10.19     | 58.06 ± 9.56         | 177M,69F    | 180M,70F       | 106/63                              | South                            |
| Sreeja et al.2005 | 16228113      | 146/146                  | 58.17±10.95       | 56.06±10.67          | 133M,13F    | 128M,18F       | 102/62                              | South                            |
| Sobti et al.2008  | 18415801      | 151/151                  | 56.9 ± 10.4       | 56.4 ± 11.1          | 130M,21F    | 123M,28F       | 123/112                             | North                            |
| Shah et al.2008(a) | 18573508     | 200/200                  | 56±9               | 43±12                | 200M        | 200M           | 120/62                              | North                            |
| Singh et al.2010  | 22072123      | 200/200                  | 53 ± 10           | 51± 09               | 200M        | 200M           | 117/64                              | North                            |
| Saikia et al.2014 | 24716924      | 272/544                  | 61.96±11.91       | 61.79±12.21          | 130M,142F   | 260M,284F      | 197/341                             | East                             |
| Natukula et al.2013 | 24175813    | 100/101                  | 56.14              | 48.07                | 73M,27F     | 86M,15F        | 51/36                               | South                            |
| Uppal et al.2014  | 25584213      | 100/100                  | 63.70 ± 3.91      | 62.08 ± 2.88         | 64 M,36 F   | 76 M, 24F      | 76/56                               | North                            |
| Sreeja et al.2008(a) | 17952468    | 211/211                  | 57.82 ± 11.74     | 56.21±10.36          | 182M,29F    | 184M,27F       | 143/119                             | South                            |
| Pachouri et al.2007 | 17417947    | 103/122                  | Not Reported       | Not Reported         | 90M,13F     | 104M,18F       | 80/28                               | North                            |
| Sreeja et al.2008(b) | 18181044    | 211/211                  | 57.82 ± 1.74      | 56.21 ± 10.36        | 184M,27F    | 182M,29F       | 0/0                                 | South                            |
| Sobti et al.2009  | 19558213      | 151/151                  | 56.9±0.4          | 56.4±1.1             | 130M, 21F   | 123M,28F       | 123/112                             | North                            |
| Sobti et al.2003  | 14602525      | 100/76                   | 55.59±11.3        | 50.99±8.1            | 95M,5F      | 73M,3F         | 86/59                               | North                            |

Contd.
| Selected Studies                  | PMID            | Number of Cases/Controls | Mean Ages ± SD Cases | Mean Ages ± SD Controls | Sex of Cases | Sex of Controls | Number of Smokers in Cases/Controls | Zone of the Indian subcontinent |
|----------------------------------|-----------------|--------------------------|----------------------|-------------------------|--------------|----------------|-------------------------------|-------------------------------|
| Kotnis et al. 2008              | 18854828        | 113/198                  | 50                   | 46                      | 74M,39F      | 129M,69F       | 0/0                           | West                          |
| Pachouri et al. 2006            | 17074589        | 120/122                  | Not Reported         | Not Reported            | 90M,30F      | 104M,18F       | 80/28                         | North                         |
| Ihsan et al. 2011(a)            | 21043833        | 161/274                  | 60.24 ± 10.77        | 53.21 ± 13.37           | 120M, 41F    | 202M, 72F      | 110/139                       | East                          |
| Jain et al. 2005                | 16354872        | 40/40                    | Not Reported         | Not Reported            | 35M, 5F      | 35M, 5F       | 32/32                         | North                         |
| Kumari et al. 2016              | 27614750        | 420/420                  | Not Reported         | Not Reported            | 90M, 30F     | 104M, 18F      | 80/28                         | North                         |
| Singh et al. 2016               | 27707541        | 330/325                  | 57.9 ± 10.58         | 52.27 ± 10.84           | 285M, 45F    | 265M, 60F      | 268/223                       | North                         |
| Tiwalk et al. 2013              | 23317414        | 175/202                  | 56.5 ± 10.3          | 54.99 ± 8.1             | 169M, 06F    | 191M, 11F      | 146/159                       | North                         |
| Girdhar et al. 2016             | 27396354        | 353/351                  | 57.55 ± 10.69        | 52.84 ± 10.80           | 305M, 48F    | 300M, 51F      | 278/250                       | North                         |
| Shah et al. 2008(b)             | 18082227        | 294/263                  | 56±9                 | 43±12                   | 200M, 0F     | 200M, 0F       | 120/62                        | North                         |
| Shaffi et al. 2009              | 19827888        | 190/248                  | 52.7                 | 53.2                    | 87M, 22F     | 95M, 68F       | 84/98                         | North                         |
| Lawania et al. 2017             | 29035087        | 370/370                  | 58.11 ± 10.44        | 53.83 ± 10.18           | 319M, 51F    | 319M, 51F      | 303/271                       | North                         |
| Naik et al. 2012                | 22231433        | 100/150                  | Not Reported         | Not Reported            | Not Reported | Not Reported    | Not Reported                  | North                         |
| Naikoo et al. 2017              | 28749108        | 112/116                  | 58±11.56             | 55±12.48                | 84M, 16F     | 98M, 30F       | 79/89                         | North                         |
| Naykoo et al. 2017              | 28122267        | 199/401                  | 57.97 ± 17.04        | 56.57 ± 18.57           | 154M, 45F    | 280M, 121F     | 139/251                       | North                         |
| Phukan et al. 2014              | 25605160        | 151/151                  | 59.01 ± 10.35        | 58.49 ± 10.22           | 82M, 69F     | 82M, 69F       | 106/86                        | East                          |
| Peddireddy et al. 2018          | 27328622        | 246/250                  | 57.57 ± 10.19        | 58.06 ± 9.56            | 177M, 69F    | 180M, 70F      | 148/75                        | South                         |
| Bahl et al. 2017(a)             | 27640551        | 300/300                  | 57.38 ± 10.74        | 53.23 ± 10.44           | 258M, 42F    | 253M, 47F      | 235/208                       | North                         |
| Bahl et al. 2017(b)             | 28378643        | 303/305                  | 57.38 ± 10.74        | 53.23 ± 10.44           | 262M, 41F    | 269M, 36F      | 239/223                       | North                         |
| Bahl et al. 2018                | 30346805        | 181/31                   | 57.87 ± 10.81        | Not Reported            | 184M, 28F    | Not Reported   | 171/0                         | North                         |
| Singh et al. 2017               | 28332164        | 193/48                   | 58.71 ± 10.38        | 0                       | 215M, 26F    | 0              | 200/0                         | North                         |

Contd.
| Selected Studies       | PMID        | Number of Cases/Controls | Mean Age±SD Cases | Mean Age±SD Controls | Sex of Cases | Sex of Controls | Number of Smokers in Cases/Controls | Zone of the Indian subcontinent |
|-----------------------|-------------|--------------------------|-------------------|----------------------|--------------|----------------|------------------------------------|---------------------------------|
| Bhardwaj et al. 2017  | 29412865    | 250/237                  | 57.38 ± 10.74     | 53.23 ± 10.44        | 252M,40F    | 241M,22F        | 229/197                            | North                           |
| Budhwar et al. 2018   | 29755293    | 297/329                  | 57.87 ± 10.81     | 52.14 ± 10.42        | 254M,43F    | 265M,55F        | 233/221                            | North                           |
| Bahl et al. 2017(c)   | 29749862    | 292/263                  | 57.38 ± 10.74     | 53.23 ± 10.44        | 252M,40F    | 241M,22F        | 229/197                            | North                           |
| Tasnim et al. 2017    | 29110586    | 202/242                  | 54.87             | 43.15                | 178M,24F    | 173M,69F        | 163/144                            | East                            |
| Chowdhury et al. 2015 | 25921167    | 50/50                    | 55±10             | 46±11                | 40M,10F     | 39M,11F         | 44/Not Reported                    | East                            |
| Islam et al. 2013     | 23178447    | 106/116                  | 57.87±10.12       | 58.14±9.77           | 93M,13F     | 105M,11F        | 97/104                             | East                            |
| Mostaid et al. 2014   | 25034526    | 106/116                  | 57.87±10.12       | 58.14±9.77           | 93M,13F     | 105M,11F        | 97/104                             | East                            |
| Masood et al. 2016    | 27461642    | 252/270                  | 54±11.5           | 53±12.6              | 184M,68F    | 197M,73F        | 131/78                             | North                           |

**Table 2** A comprehensive list of meta-analysis results showing the overall association of the variants with lung cancer, with crude odds ratio (OR), 95% Confidence Interval (CI), \( P_{\text{FDR}} \), Benjamini-Hochberg False Discovery Rate (FDR) corrected \( P \)-value, Heterogeneity indices \( H^2, I^2 \), Publication Bias by Egger's Test. Both the Genetic model and model used for meta-analysis are also mentioned.
| Gene Symbol | Variant ID | Number | Genetic Model | Comparison | Model | OR (95% CI); P-value<sup>†</sup> | H² | P<sub>Het</sub><sup>‡</sup> | I² | t-stat | df | P-value<sup>§</sup> |
|-------------|------------|--------|---------------|------------|-------|---------------------------------|----|-----------------|---|------|----|----------------|
| **GSTT1**   | del 1      | 14     | Recessive     | Null (<-<-) vs Present [(+/-) + (+/+)] | Fixed | 1.36 (1.03-1.79); 0.028         | 0.5 | 1.5             | 31.1 | 0.095 | 12 | 0.9259       |
|             |            | 14     | Recessive     | Null (<-<-) vs Present [(+/-) + (+/+)] | Random| 1.36 (0.97-1.89); 0.07          |     |                 |     |       |    |               |
| **GSTM1**   | del 2      | 15     | Recessive     | Null (<-<-) vs Present [(+/-) + (+/+)] | Fixed | 1.38 (1.09-1.75); 0.008         | 0.5 | 1.9             | 24.69| 2.013 | 13 | 0.0653       |
|             |            | 15     | Recessive     | Null (<-<-) vs Present [(+/-) + (+/+)] | Random| 1.4 (1.01-1.95); 0.04           |     |                 |     |       |    |               |
| **GSTP1**   | rs1695     | 5      | Additive      | GG vs AG vs AA | Fixed | 1.05 (0.72-1.52); 0.814         | 0.91 | 1.2             | 71.3 | N/A   | N/A | N/A           |
|             |            | 5      | Additive      | GG vs AG vs AA | Random| 1.04 (0.69-1.56); 0.84          |     |                 |     |       |    |               |
|             |            | 5      | Dominant      | (GG+AG) vs AA  | Fixed | 1.39 (0.91-2.12); 0.13          | 0.47 | 3.6             | 0.001| 72    | N/A | N/A           |
|             |            | 5      | Dominant      | (GG+AG) vs AA  | Random| 1.4 (0.63-3.13); 0.406         |     |                 |     |       |    |               |
|             |            | 5      | Recessive     | GG vs (AG+AA) | Fixed | 1.06 (0.56-1.99); 0.86          | 0.94 | 1.2             | 13.7 | N/A   | N/A | N/A           |
|             |            | 5      | Recessive     | GG vs (AG+AA) | Random| 1.05 (0.53-2.07); 0.89         |     |                 |     |       |    |               |
| **TP53**    | rs1042522  | 8      | Additive      | CC vs CG vs GG | Fixed | 1.24 (0.94-1.64); 0.127         | 0.76 | 2.1             | 0.261| 52    | N/A | N/A           |
|             |            | 8      | Additive      | CC vs CG vs GG | Random| 1.23 (0.81-1.85); 0.32         |     |                 |     |       |    |               |
|             |            | 8      | Dominant      | (CC+CG) vs GG  | Fixed | 1.35 (0.96-1.92); 0.09          | 0.64 | 2.7             | 0.098| 62.8  | N/A | N/A           |
|             |            | 8      | Dominant      | (CC+CG) vs GG  | Random| 1.32 (0.74-2.37); 0.351        |     |                 |     |       |    |               |
|             |            | 8      | Recessive     | CC vs (CG+GG)  | Fixed | 1.36 (0.94-1.96); 0.103         | 0.5  | 1.8             | 0.32 | 45.2  | N/A | N/A           |
|             |            | 8      | Recessive     | CC vs (CG+GG)  | Random| 1.36 (0.82-2.26); 0.23         |     |                 |     |       |    |               |
| Gene   | SNP       | rsID       | Sample Size | Minor Allele | Genotype Model | OR (95% CI) | p-value | OR (95% CI) | p-value |
|--------|-----------|------------|-------------|--------------|----------------|-------------|---------|-------------|---------|
| CYP1A1 | rs4646903 | 11         | Dominant (TC+CC) vs TT | Fixed | 1.48 (1.93-1.95); 0.005 | 0.07 | 1.2 0.944 16 0.985 9 0.3502 |
|        |           | 11         | Dominant (TC+CC) vs TT | Random | 1.49 (1.10-2.01); 0.01 | 0.07 |        |
|        | rs1048943 | 9          | Dominant (AG+GG) vs AA | Fixed | 1.98 (1.46-2.68); 0.00001 | 0.09 | 1.9 0.092 48.3 N/A N/A N/A |
|        |           | 9          | Dominant (AG+GG) vs AA | Random | 2.07 (1.32-3.27); 0.002 | 0.03 |        |
| ERCC2  | rs13181   | 3          | Additive CC vs CA vs AA | Fixed | 1.28 (0.86-1.92); 0.225 | 0.76 | 1 0.856 1.6 N/A N/A N/A |
|        |           | 3          | Additive CC vs CA vs AA | Random | 1.28 (0.86-1.91); 0.23 | 0.76 |        |
|        |           | 3          | Dominant (CC+CA) vs AA | Fixed | 1.32 (0.84-2.08); 0.233 | 0.54 | 1.1 0.712 6.52 N/A N/A N/A |
|        |           | 3          | Dominant (CC+CA) vs AA | Random | 1.32 (0.82-2.12); 0.25 | 0.54 |        |
|        |           | 3          | Recessive CC vs (CA+AA) | Fixed | 1.49 (0.82-2.71); 0.187 | 0.5 | 1.1 0.73 5.06 N/A N/A N/A |
|        |           | 3          | Recessive CC vs (CA+AA) | Random | 1.49 (0.81-2.76); 0.2 | 0.49 |        |
| Gene Symbol | Variant ID | Number | Genetic Model | Comparison | Model | OR (95% CI); P-value | $P_{FDR}$ | $I^2_\text{Het}$ | $P_{Het}$ | t-stat | df | P-value |
|-------------|------------|--------|---------------|------------|-------|----------------------|---------|-------------|----------|--------|----|---------|
| SULT1A1     | rs9282861  | 4      | Additive      | AA vs AG vs GG | Fixed  | 1.2 (0.8-1.79); 0.38 | 0.88    | 1.5         | 0.362   | 32.3   | N/A | N/A     |
|             |            | 4      | Additive      | AA vs AG vs GG | Random | 1.19 (0.73-1.96); 0.47 | 0.88    |
|             |            | 4      | Dominant      | (AA+AG) vs GG | Fixed  | 1.23 (0.77-1.96); 0.38 | 0.73    | 2.3         | 0.085   | 56.1   | N/A | N/A     |
|             |            | 4      | Dominant      | (AA+AG) vs GG | Random | 1.23 (0.61-2.48); 0.56 | 0.73    |
|             |            | 4      | Recessive     | AA vs (AG+GG) | Fixed  | 1.53 (0.84-2.77); 0.17 | 0.5     | 1.5         | 0.49    | 33.3   | N/A | N/A     |
|             |            | 4      | Recessive     | AA vs (AG+GG) | Random | 1.61 (0.76-3.44); 0.22 | 0.49    |
| VEGFA       | rs2010963  | 3      | Additive      | GG vs GC vs CC | Fixed  | 1.27 (0.78-2.07); 0.34 | 0.88    | 1.6         | 0.235   | 37.9   | N/A | N/A     |
|             |            | 3      | Additive      | GG vs GC vs CC | Random | 1.2 (0.64-2.28); 0.57 | 0.88    |
|             |            | 3      | Dominant      | (GG+GC) vs CC | Fixed  | 2.12 (1.15-3.90); 0.02 | 0.09    | 6.2         | 0.001   | 83.8   | N/A | N/A     |
|             |            | 3      | Dominant      | (GG+GC) vs CC | Random | 1.96 (0.43-8.99); 0.39 | 0.12    |
|             |            | 3      | Recessive     | GG vs (GC+CC) | Fixed  | 0.58 (0.28-1.18); 0.13 | 0.5     | 1.4         | 0.37    | 29.8   | N/A | N/A     |
|             |            | 3      | Recessive     | GG vs (GC+CC) | Random | 0.58 (0.24-1.40); 0.23 | 0.49    |
|             | rs3025039  | 3      | Dominant      | (TT+TC) vs CC | Fixed  | 1.37 (0.79-2.36); 0.26 | 0.73    | 2.4         | 0.084   | 57.7   | N/A | N/A     |
|             |            | 3      | Dominant      | (TT+TC) vs CC | Random | 1.27 (0.54-2.96); 0.58 | 0.73    |
| DKK3        | rs3206824  | 3      | Additive      | AA vs AG vs GG | Fixed  | 1.13 (0.68-1.87); 0.64 | 0.88    | 1           | 0.821   | 2.96   | N/A | N/A     |
|             |            | 3      | Additive      | AA vs AG vs GG | Random | 1.13 (0.67-1.88); 0.65 | 0.87    |
|             |            | 3      | Dominant      | (AA+AG) vs GG | Fixed  | 1.13 (0.66-1.93); 0.65 | 0.73    | 1           | 0.803   | 3.68   | N/A | N/A     |
|             |            | 3      | Dominant      | (AA+AG) vs GG | Random | 1.13 (0.65-1.95); 0.66 | 0.73    |
|             |            | 3      | Recessive     | AA vs (AG+GG) | Fixed  | 1.27 (0.49-3.27); 0.617 | 0.83    | 1           | 0.88    | 1.18   | N/A | N/A     |
|             |            | 3      | Recessive     | AA vs (AG+GG) | Random | 1.27 (0.49-3.29); 0.62 | 0.9     | Contd.
| Gene Symbol | Variant ID | Number | Genetic Model | Comparison | Model | OR (95% CI); P-value | $P_{FDR}$ | I² | $P_{	ext{Het}}$ | $t$-stat | df | $P$-value |
|-------------|------------|--------|---------------|------------|-------|----------------------|---------|----|-------------|---------|----|-----------|
| DKK2        | rs447372   | 3      | Additive      | GG vs AG   | Fixed | 0.79 (0.50-1.24); 0.29 | 0.76    | 1.1 | 0.633       | 8.12    | N/A | N/A       |
|             |            |        | Additive      | GG vs AG   | Random| 0.79 (0.50-1.26); 0.32 | 0.76    |     |             |         |     |           |
|             |            |        | Dominant      | (GG+AG) vs AA | Fixed | 0.73 (0.44-1.22); 0.229 | 0.54    | 1   | 0.747       | 3.52    | N/A | N/A       |
|             |            |        | Dominant      | (GG+AG) vs AA | Random| 0.73 (0.43-1.23); 0.24 | 0.54    |     |             |         |     |           |
|             |            |        | Recessive     | GG vs (AG+AA) | Fixed | 0.77 (0.4-1.51); 0.453 | 0.83    | 1.6 | 0.25        | 38.9    | N/A | N/A       |
|             |            |        | Recessive     | GG vs (AG+AA) | Random| 0.77 (0.32-1.83); 0.55 | 0.88    |     |             |         |     |           |
|             | rs419558   | 3      | Additive      | TT vs TC   | Fixed | 1.45 (0.85-2.49); 0.17 | 0.76    | 1.6 | 0.352       | 37      | N/A | N/A       |
|             |            |        | Additive      | TT vs TC   | Random| 1.39 (0.69-2.79); 0.35 | 0.76    |     |             |         |     |           |
|             |            |        | Dominant      | (TT+TC) vs CC | Fixed | 1.47 (0.85-2.55); 0.17 | 0.64    | 1.6 | 0.339       | 37.8    | N/A | N/A       |
|             |            |        | Dominant      | (TT+TC) vs CC | Random| 1.41 (0.69-2.87); 0.34 | 0.64    |     |             |         |     |           |
|             | rs17037102 | 3      | Additive      | AA vs AG   | Fixed | 1.81 (1.03-3.15); 0.04 | 0.36    | 1   | 0.999       | 0       | N/A | N/A       |
|             |            |        | Additive      | AA vs AG   | Random| 1.81 (1.03-3.15); 0.04 | 0.36    |     |             |         |     |           |
|             |            |        | Dominant      | (AA+AG) vs GG | Fixed | 1.82 (1.03-3.22); 0.04 | 0.22    | 1   | 0.992       | 0       | N/A | N/A       |
|             |            |        | Dominant      | (AA+AG) vs GG | Random| 1.82 (1.03-3.22); 0.04 | 0.22    |     |             |         |     |           |
| AXIN2       | rs2240308  | 3      | Additive      | TT vs TC   | Fixed | 0.82 (0.53-1.25); 0.35 | 0.76    | 1   | 0.846       | 2.28    | N/A | N/A       |
|             |            |        | Additive      | TT vs TC   | Random| 0.82 (0.53-1.26); 0.36 | 0.76    |     |             |         |     |           |
| 3 | Dominant (TT+TC) vs CC | Fixed | 0.8 (0.47-1.36); 0.41 | 0.76 | 1 | 0.828 | 2.89 | N/A | N/A | N/A |
|---|------------------------|-------|-----------------------|------|---|-------|------|-----|-----|-----|
| 3 | Dominant (TT+TC) vs CC | Random | 0.80 (0.46-1.38); 0.42 | 0.72 | | |
| 3 | Recessive TT vs (TC+CC) | Fixed | 0.72 (0.41-1.26); 0.25 | 0.5 | 1 | 0.81 | 3.09 | N/A | N/A | N/A |
| 3 | Recessive TT vs (TC+CC) | Random | 0.72 (0.40-1.28); 0.26 | 0.49 | | |
| | rs2240307 | Additive CC vs CT vs TT | Fixed | 0.66 (0.33-1.33); 0.25 | 0.757 | 1 | 0.924 | 0.56 | N/A | N/A | N/A |
| 3 | Additive CC vs CT vs TT | Random | 0.66 (0.33-1.34); 0.25 | 0.75 | | |
| 3 | Dominant (CC+CT) vs TT | Fixed | 0.6 (0.29-1.24); 0.17 | 0.543 | 1 | 0.931 | 0.56 | N/A | N/A | N/A |
| 3 | Dominant (CC+CT) vs TT | Random | 0.60 (0.28-1.25); 0.17 | 0.53 | | |

Contd.

---
| Gene Symbol | Variant ID  | Number | Genetic Model | Comparison     | Model | OR (95% CI); P-value | $P_{FDR}$ | $H^2$ | $P_{Het}$ | $I^2$ | t-stat | df  | $P$-value |
|-------------|-------------|--------|---------------|----------------|-------|----------------------|----------|-------|-----------|-------|--------|-----|----------|
| AXIN 2      | rs35285779  | 3      | Additive      | GG vs AG vs AA | Fixed | 0.89 (0.50-1.58); 0.69 | 0.88     | 1     | 0.97      | 0.08  | N/A    | N/A | N/A      |
|             |             | 3      | Additive      | GG vs AG vs AA | Random| 0.89 (0.50-1.58); 0.69 | 0.87     |       |           |       |        |      |          |
|             |             | 3      | Dominant      | (GG+AG) vs AA  | Fixed | 0.87 (0.48-1.57); 0.64 | 0.73     | 1     | 0.984     | 0.02  | N/A    | N/A | N/A      |
|             |             | 3      | Dominant      | (GG+AG) vs AA  | Random| 0.87 (0.48-1.57); 0.64 | 0.73     |       |           |       |        |      |          |
|             |             | 2      | Recessive     | GG vs (AG+AA) | Fixed | 0.95 (0.24-3.82); 0.94 | 0.94     | 1     | 0.94      | 0     | N/A    | N/A | N/A      |
|             |             | 2      | Recessive     | GG vs (AG+AA) | Random| 0.95 (0.24-3.82); 0.94 | 0.99     |       |           |       |        |      |          |
|             | rs9915936   | 3      | Additive      | AA vs AG vs GG | Fixed | 0.66 (0.33-1.29); 0.22 | 0.76     | 1.2   | 0.545     | 18.4  | N/A    | N/A | N/A      |
|             |             | 3      | Additive      | AA vs AG vs GG | Random| 0.67 (0.31-1.43); 0.3  | 0.76     |       |           |       |        |      |          |
|             |             | 3      | Dominant      | (AA+AG) vs GG  | Fixed | 0.63 (0.31-1.26); 0.189 | 0.54     | 1.1   | 0.644     | 11.8  | N/A    | N/A | N/A      |
|             |             | 3      | Dominant      | (AA+AG) vs GG  | Random| 0.63 (0.30-1.34); 0.23 | 0.54     |       |           |       |        |      |          |
|             | rs7219582   | 3      | Additive      | TT vs TG vs GG | Fixed | 0.98 (0.52-1.85; 0.946) | 0.95     | 1     | 0.973     | 0.25  | N/A    | N/A | N/A      |
|             |             | 3      | Additive      | TT vs TG vs GG | Random| 0.98 (0.56-1.86; 0.95) | 0.95     |       |           |       |        |      |          |
|             |             | 2      | Dominant      | (TT+TG) vs GG  | Fixed | 0.99 (0.51-1.91; 0.964) | 0.96     | 1     | 0.825     | 0.12  | N/A    | N/A | N/A      |
|             |             | 2      | Dominant      | (TT+TG) vs GG  | Random| 0.99 (0.51-1.91; 0.96) | 0.96     |       |           |       |        |      |          |
|             | rs35415678  | 3      | Additive      | TT vs TC vs CC | Fixed | 1.09 (0.64-1.85; 0.75) | 0.89     | 1     | 0.935     | 0.43  | N/A    | N/A | N/A      |
|             |             | 3      | Additive      | TT vs TC vs CC | Random| 1.09 (0.64-1.86; 0.75) | 0.89     |       |           |       |        |      |          |
|             |             | 3      | Dominant      | (TT+TC) vs CC  | Fixed | 1.1 (0.62-1.96; 0.741) | 0.78     | 1     | 0.935     | 0.39  | N/A    | N/A | N/A      |
|             |             | 3      | Dominant      | (TT+TC) vs CC  | Random| 1.1 (0.62-1.97; 0.74) | 0.78     |       |           |       |        |      |          |
|             |             | 2      | Recessive     | TT vs (TC+CC) | Fixed | 1.11 (0.44-2.75; 0.83) | 0.92     | 1     | 0.85      | 0.06  | N/A    | N/A | N/A      |
|             |             | 2      | Recessive     | TT vs (TC+CC) | Random| 1.11 (0.44-2.75; 0.83) | 0.99     |       |           |       |        |      |          |
| Gene Symbol | Variant ID | Number | Genetic Model | Comparison | Model   | OR (95% CI); P-value \(\hat{a}\) | \(P_{FDR}\) \(\hat{b}\) | \(I^2\) | \(P_{Het}\) \(\hat{c}\) | \(t\)-stat | \(df\) | \(P\)-value |
|-------------|------------|--------|---------------|------------|---------|-------------------------------|-----------------|-------|-----------------|---------|--------|-------------|
| **XRCC1**   | rs25487    | 9      | Additive      | AA vs AG vs GG | Fixed   | 1.09 (0.85-1.41); 0.5       | 0.88            | 1.6   | 0.362           | 36.8    | N/A    | N/A         |
|             |            | 9      | Additive      | AA vs AG vs GG | Random  | 1.07 (0.78-1.48); 0.68      |                 |       |                 |         |         | 0.88        |
|             |            | 9      | Dominant      | (AA+AG) vs GG | Fixed   | 1.10 (0.81-1.49); 0.56      | 0.73            | 2.8   | 0.017           | 64      | N/A    | N/A         |
|             |            | 9      | Dominant      | (AA+AG) vs GG | Random  | 1.03 (0.61-1.74); 0.9       |                 |       |                 |         |         | 0.73        |
|             |            | 9      | Recessive     | AA vs (AG+GG) | Fixed   | 1.24 (0.87-1.76); 0.23      | 0.5             | 1.2   | 0.79            | 15.5    | N/A    | N/A         |
|             |            | 9      | Recessive     | AA vs (AG+GG) | Random  | 1.24 (0.84-1.82); 0.27      |                 |       |                 |         |         | 0.49        |
| rs1799782   | 4          | 4      | Additive      | TT vs CT vs CC | Fixed   | 1.16 (0.74-1.84); 0.52      | 0.88            | 1.1   | 0.8             | 8.59    | N/A    | N/A         |
|             | 4          | 4      | Additive      | TT vs CT vs CC | Random  | 1.16 (0.72-1.88); 0.54      |                 |       |                 |         |         | 0.88        |
|             | 4          | 4      | Dominant      | (TT+CT) vs CC | Fixed   | 1.2 (0.73-1.97); 0.47       | 0.73            | 1.1   | 0.858           | 4.83    | N/A    | N/A         |
|             | 4          | 4      | Dominant      | (TT+CT) vs CC | Random  | 1.20 (0.72-2.00); 0.48      |                 |       |                 |         |         | 0.73        |
|             | 3          | 3      | Recessive     | TT vs (CT+CC) | Fixed   | 1.31 (0.55-3.15); 0.54      | 0.83            | 1.3   | 0.55            | 20      | N/A    | N/A         |
|             | 3          | 3      | Recessive     | TT vs (CT+CC) | Random  | 0.99 (0.32-3.07); 0.99      |                 |       |                 |         |         | 0.99        |
| rs915927    | 3          | 3      | Additive      | GG vs AG vs AA | Fixed   | 1.04 (0.67-1.62); 0.86      | 0.91            | 1     | 0.99            | 0.01    | N/A    | N/A         |
|             | 3          | 3      | Additive      | GG vs AG vs AA | Random  | 1.04 (0.67-1.62); 0.86      |                 |       |                 |         |         | 0.91        |
|             | 3          | 3      | Dominant      | (GG+AG) vs AA | Fixed   | 1.12 (0.68-1.87); 0.65      | 0.73            | 1     | 0.939           | 0.22    | N/A    | N/A         |
|             | 3          | 3      | Dominant      | (GG+AG) vs AA | Random  | 1.12 (0.68-1.87); 0.65      |                 |       |                 |         |         | 0.73        |
|             | 3          | 3      | Recessive     | GG vs (AG+AA) | Fixed   | 0.88 (0.46-1.69); 0.706     | 0.88            | 1     | 0.95            | 0.17    | N/A    | N/A         |
|             | 3          | 3      | Recessive     | GG vs (AG+AA) | Random  | 0.88 (0.46-1.69); 0.71      |                 |       |                 |         |         | 0.94        |
| rs3547      | 3          | 3      | Additive      | AA vs AG vs GG | Fixed   | 1.89 (1.23-2.90); 0.004     | 0.36            | 1.7   | 0.298           | 39.7    | N/A    | N/A         |
|             | 3          | 3      | Additive      | AA vs AG vs GG | Random  | 1.85 (1.06-3.25); 0.03      |                 |       |                 |         |         | 0.36        |
|             | 3          | 3      | Dominant      | (AA+AG) vs GG | Fixed   | 2.72 (1.56-4.73); 0.0004    | 0.03            | 1.3   | 0.485           | 22.7    | N/A    | N/A         |
|             | 3          | 3      | Dominant      | (AA+AG) vs GG | Random  | 2.68 (1.41-5.07); 0.002     |                 |       |                 |         |         | 0.03        |
|             | 3          | 3      | Recessive     | AA vs (AG+GG) | Fixed   | 2.04 (1.19-3.51); 0.01      | 0.5             | 2     | 0.2             | 50.9    | N/A    | N/A         |
|             | 3          | 3      | Recessive     | AA vs (AG+GG) | Random  | 1.96 (0.89-4.29); 0.09      |                 |       |                 |         |         | 0.49        |

\(\hat{a}\) P-value<0.05, OR, Crude Odds Ratio, 95\% CI, 95\% Confidence Interval, \(I^2\) and \(H^2\) are measures of heterogeneity, \(\hat{b}\) \(P_{FDR}<0.05\)

- \(P_{Het}<0.1\) (Cochran's Q test), Significant associations are depicted in bold. N/A stands for Not Applicable.
Table 3 Results of the histological subtype-stratified meta-analysis of reported lung cancer variants in the Indian subcontinent.

| Variant ID/ Gene   | Histological subtypes | Number | Genetic Model | Comparison | Model | OR (95% CI); P-value | \( I^2 \) | \( H^2 \) | \( P_{het} \) |
|-------------------|-----------------------|--------|---------------|------------|-------|-----------------------|------|------|-------|
| del1/GSTT1        | Squamous cell carcinoma | 3      | Recessive     | Null (-/-) vs Present (+/-, +/+ | Fixed  | 1.38 (0.7-2.74); 0.35 | 17.69 | 1.21 | 0.49  |
|                   | Squamous cell carcinoma | 3      | Recessive     | Null (-/-) vs Present (+/-, +/+ | Random | 1.38 (0.65-2.94); 0.40 |      |      |       |
|                   | Adenocarcinoma         | 3      |               |            | Fixed  | 2.14 (1.04-4.41); 0.04 | 0.79  | 1.01 | 0.91  |
|                   | Adenocarcinoma         | 3      |               |            | Random | 2.14 (1.03-4.42); 0.04 |      |      |       |
|                   | Small Cell Lung Carcinoma | 2    |               |            | Fixed  | 1.15 (0.44-3.03); 0.77 | 26.08 | 1.36 | 0.31  |
|                   | Small Cell Lung Carcinoma | 2    |               |            | Random | 1.09 (0.34-3.51); 0.88 |      |      |       |
| del2/GSTM1        | Squamous cell carcinoma | 3      | Recessive     | Null (-/-) vs Present (+/-, +/+ | Fixed  | 1.33 (0.74-2.4); 0.34 | 0.11  | 1    | 0.96  |
|                   | Squamous cell carcinoma | 3      | Recessive     | Null (-/-) vs Present (+/-, +/+ | Random | 1.33 (0.74-2.39); 0.34 |      |      |       |
|                   | Adenocarcinoma         | 3      |               |            | Fixed  | 1.32 (0.68-2.57); 0.41 | 24.23 | 1.32 | 0.48  |
|                   | Adenocarcinoma         | 3      |               |            | Random | 1.28 (0.58-2.84); 0.54 |      |      |       |
|                   | Small Cell Lung Carcinoma | 2    |               |            | Fixed  | 1.53 (0.67-3.46); 0.31 | 0.32  | 1    | 0.77  |
|                   | Small Cell Lung Carcinoma | 2    |               |            | Random | 1.53 (0.67-3.47); 0.31 |      |      |       |
| rs4646903/CYP1A1  | Squamous cell carcinoma | 4      | Dominant      | (TC+CC) vs TT | Fixed  | 1.82 (1.06-3.11); 0.03 | 11.38 | 1.13 | 0.71  |
|                   | Squamous cell carcinoma | 4      | Dominant      | (TC+CC) vs TT | Random | 1.82 (1.03-3.23); 0.04 |      |      |       |
|                   | Adenocarcinoma         | 4      |               |            | Fixed  | 1.36 (0.68-2.7); 0.38  | 20.93 | 1.26 | 0.6   |
|                   | Adenocarcinoma         | 4      |               |            | Random | 1.28 (0.56-2.91); 0.55 |      |      |       |
|                   | Small Cell Lung Carcinoma | 2    |               |            | Fixed  | 0.88 (0.34-2.29); 0.8  | 0.001 | 1.02 | 0.98  |
|                   | Small Cell Lung Carcinoma | 2    |               |            | Random | 0.88 (0.34-2.29); 0.79 |      |      |       |
| rs1048943/CYP1A1  | Squamous cell carcinoma | 5      | Dominant      | (AG+GG) vs AA | Fixed  | 3.53 (2.05-6.08); 0.000005 | 52.24 | 2.09 | 0.11  |
|                   | Squamous cell carcinoma | 5      | Dominant      | (AG+GG) vs AA | Random | 2.67 (1.19-6.01); 0.02 |      |      |       |
|                   | Adenocarcinoma         | 4      |               |            | Fixed  | 3.38 (1.83-6.25); 0.0001 | 49.65 | 1.99 | 0.16  |
|                   | Adenocarcinoma         | 4      |               |            | Random | 2.76 (1.33-5.75); 0.007 |      |      |       |
|                   | Small Cell Lung Carcinoma | 3    |               |            | Fixed  | 1.63 (0.64-4.14); 0.31 | 5.04  | 1.05 | 0.55  |
|                   | Small Cell Lung Carcinoma | 3    |               |            | Random | 1.3 (0.51-3.33); 0.58 |      |      |       |

Contd.
| Variant ID/ Gene       | Histological subtypes          | Number | Genetic Model | Comparison      | Model    | OR (95% CI); P-value | $I^2$ | $H^2$ | $P_{het}$ |
|-----------------------|--------------------------------|--------|---------------|-----------------|----------|----------------------|------|-------|----------|
| rs1042522/TP53        | Squamous cell carcinoma        | 5      | Additive      | AA vs AG vs GG  | Fixed    | 1.12 (0.74-1.70); 0.6 | 72.7 | 3.67  | 0.03     |
|                       | Squamous cell carcinoma        | 5      |               |                 | Random   | 0.95 (0.41-2.18); 0.9 |      |       |          |
|                       | Adenocarcinoma                 | 5      |               |                 | Fixed    | 1.04 (0.67-1.64); 0.85 | 35.38| 1.55  | 0.38     |
|                       | Adenocarcinoma                 | 5      |               |                 | Random   | 1.02 (0.57-1.83); 0.94 |     |       |          |
|                       | Small Cell Lung Carcinoma      | 4      |               |                 | Fixed    | 1.15 (0.69-1.92); 0.6 | 60.09| 2.51  | 0.13     |
|                       | Small Cell Lung Carcinoma      | 4      |               |                 | Random   | 1.09 (0.45-2.63); 0.85 |     |       |          |
|                       | Squamous cell carcinoma        | 5      | Dominant      | (AA+AG) vs GG   | Fixed    | 1.27 (0.76-2.13); 0.37 | 71.25| 3.48  | 0.03     |
|                       | Squamous cell carcinoma        | 5      |               |                 | Random   | 1.12 (0.41-3.03); 0.82 |     |       |          |
|                       | Adenocarcinoma                 | 5      |               |                 | Fixed    | 1.18 (0.67-2.06); 0.56 | 39.73| 1.66  | 0.34     |
|                       | Adenocarcinoma                 | 5      |               |                 | Random   | 1.18 (0.56-2.50); 0.66 |     |       |          |
|                       | Small Cell Lung Carcinoma      | 4      |               |                 | Fixed    | 1.38 (0.72-2.64); 0.33 | 66.89| 3.02  | 0.08     |
|                       | Small Cell Lung Carcinoma      | 4      |               |                 | Random   | 1.31 (0.39-4.44); 0.66 |     |       |          |
|                       | Squamous cell carcinoma        | 5      | Recessive     | AA vs (AG+GG)  | Fixed    | 1.36 (0.78-2.37); 0.28 | 63.92| 2.77  | 0.12     |
|                       | Squamous cell carcinoma        | 5      |               |                 | Random   | 1.15 (0.43-3.04); 0.78 |     |       |          |
|                       | Adenocarcinoma                 | 5      |               |                 | Fixed    | 1.24 (0.67-2.26); 0.49 | 22.18| 1.28  | 0.55     |
|                       | Adenocarcinoma                 | 5      |               |                 | Random   | 1.23 (0.61-2.49); 0.57 |     |       |          |
|                       | Small Cell Lung Carcinoma      | 4      |               |                 | Fixed    | 1.54 (0.78-3.04); 0.22 | 63.96| 2.77  | 0.06     |
|                       | Small Cell Lung Carcinoma      | 4      |               |                 | Random   | 1.61 (0.48-5.41); 0.44 |     |       |          |

$P_{value}<0.05$*, $0.01**$, $0.001***$, OR, Crude Odds Ratio, 95% CI, 95% Confidence Interval, $I^2$ and $H^2$ are measures of heterogeneity

- $P_{het}<0.1$ (Cochran's Q test).

Significant associations are depicted in bold.
Table 4
[A] The results of the subgroup analysis stratified by smoking status. Significant associations are depicted in bold.

| Variant ID/ Gene Symbol | Sub-groups | Number | Genetic Model | OR (95% CI); $P$-value† | $I^2$ | $H^2$ | $P_{het}$§ |
|-------------------------|------------|--------|---------------|--------------------------|-------|-------|------------|
| del1/ GSTT1             | Smoker     | 9      | Recessive     | 1.43 (0.93-2.22); 0.11   | 31.58 | 1.46  | 0.38       |
|                         | Non-Smoker | 9      |               | 0.63 (0.39-1.04); 0.07   | 43.89 | 1.78  | 0.34       |
| del2/ GSTM1             | Smoker     | 9      | Recessive     | 0.97 (0.65-1.44); 0.88   | 4.49  | 1.05  | 0.98       |
|                         | Non-Smoker | 9      |               | 1.19 (0.77-1.82); 0.43   | 3.48  | 1.04  | 0.98       |
| rs4646903/CYP1A1        | Smoker     | 7      | Dominant      | 2.26 (1.44-3.53); 0.0004 | 42.63 | 1.47  | 0.54       |
|                         | Non-Smoker | 7      |               | 0.99 (0.64-1.53); 0.95   | 50.39 | 1.88  | 0.34       |
| rs1048943/CYP1A1        | Smoker     | 7      | Dominant      | 2.23 (1.47-3.37); 0.0002 | 28.78 | 1.4   | 0.41       |
|                         | Non-Smoker | 7      |               | 1.74 (1.10-2.73); 0.02   | 23.67 | 1.31  | 0.60       |
| rs25487/XRCC1           | Smoker     | 4      | Additive      | 0.67 (0.43-1.04); 0.08   | 40.52 | 1.68  | 0.19       |
|                         | Non-Smoker | 4      |               | 0.66 (0.39-1.10); 0.11   | 7.16  | 1.08  | 0.76       |
|                         | Smoker     | 4      | Dominant      | 0.76 (0.44-1.31); 0.32   | 32.07 | 1.47  | 0.31       |
|                         | Non-Smoker | 4      |               | 0.56 (0.30-1.06); 0.07   | 18.85 | 1.23  | 0.54       |
|                         | Smoker     | 4      | Recessive     | 0.48 (0.27-0.86); 0.01   | 60.85 | 2.55  | 0.1        |
|                         | Non-Smoker | 4      |               | 0.53 (0.26-1.07); 0.08   | 16.1  | 1.19  | 0.62       |
| rs9282861/SULT1A1       | Smoker     | 2      | Additive      | 0.66 (0.37-1.16); 0.15   | 0.92  | 1.01  | 0.71       |
|                         | Non-Smoker | 2      |               | 0.68 (0.34-1.36); 0.28   | 0.32  | 1     | 0.78       |
|                         | Smoker     | 2      | Dominant      | 0.66 (0.33-1.34); 0.25   | 2.16  | 1.02  | 0.64       |
|                         | Non-Smoker | 2      |               | 0.56 (0.24-1.3); 0.18    | 0.001 | 1     | 0.95       |
|                         | Smoker     | 2      | Recessive     | 0.51 (0.24-1.11); 0.09   | 57.56 | 2.36  | 0.12       |
|                         | Non-Smoker | 2      |               | 0.65 (0.24-1.76); 0.39   | 0.02  | 1     | 0.89       |
The results of the effect modification of variants on lung cancer by smoking status. *Significant associations are depicted in bold.*

| Moderator variable   | Genetic Model | Number | Moderator Effect Size (θ) | θ low | θ high | P-value† |
|----------------------|---------------|--------|---------------------------|-------|--------|----------|
| del1/ GSTT1          | Recessive     | 18     | 0.82                      | 0.16  | 1.48   | **0.015**|
| del2/ GSTM1          | Recessive     | 18     | -0.21                     | -0.79 | 0.38   | 0.49     |
| rs4646903/ CYP1A1    | Dominant      | 14     | 0.83                      | 0.2   | 1.45   | **0.01** |
| rs1048943/ CYP1A1    | Dominant      | 14     | 0.25                      | -0.37 | 0.86   | 0.43     |
| rs25487/ XRCC1       | Additive      | 8      | 0.02                      | -0.66 | 0.7    | 0.96     |
|                      | Dominant      | 8      | 0.3                       | -0.54 | 1.14   | **0.48** |
|                      | Recessive     | 8      | -0.09                     | -1.01 | 0.82   | **0.84** |
| rs9282861/ SULT1A1   | Additive      | 4      | -0.03                     | -0.93 | 0.87   | **0.94** |
|                      | Dominant      | 4      | 0.17                      | -0.93 | 1.26   | **0.77** |
|                      | Recessive     | 4      | -0.23                     | -1.49 | 1.03   | **0.72** |

†P-value<0.05*, 0.01**, 0.001***, OR, Crude Odds Ratio, 95% CI, 95% Confidence Interval, $I^2$ and $H^2$ are measures of heterogeneity. §$P_{Het}<0.1$ (Cochran’s Q test). Significant associations are depicted in bold.

**Figures**
Figure 1

The pathway of the selection of studies for meta-analysis on candidate association studies reported from the Indian subcontinent. The flow diagram is constructed as per the PRISMA guidelines (2009).
Figure 2

Forest plot depicting the odds ratios (ORs), and 95% CI of the polymorphism, rs1048943/CYP1A1 for its association with overall lung cancer risk in the Indian subcontinent in both [A] fixed-effect and [A'] random-effects model, [A'] Funnel plot that shows no evidence of publication bias between the studies reporting the polymorphism, rs1048943/CYP1A1 Forest plot depicting the odds ratios (ORs), and 95% CI of the polymorphism, rs3547/XRCC1 for its association with overall lung cancer risk in the Indian subcontinent in a [B] fixed-effect model and [B'] random-effects model. The results are obtained in a dominant model of analysis. The forest plots of the significant associations were given (pFDR < 0.05*)
Figure 3

[A] Forest plot depicting the odds ratios (ORs), and 95% CI of the polymorphism, del1/GSTT1 for its association with lung adenocarcinoma in a recessive model in the Indian subcontinent for both fixed-effect and random-effects model. [B] Forest plot depicting the odds ratios (ORs), and 95% CI of the polymorphism, rs4646903/CYP1A1 for its association with squamous cell carcinoma in a dominant model in the Indian subcontinent for both fixed-effect and random-effects model. [C] Forest plot depicting the odds ratios (ORs), and 95% CI of the polymorphism, rs1048943/CYP1A1 for its association with adenocarcinoma in a dominant model in the Indian subcontinent for both fixed-effect and random-effects model, and [D] Forest plot depicting the odds ratios (ORs), and 95% CI of the polymorphism, rs1048943/CYP1A1 for its association with squamous cell carcinoma in a dominant model in the Indian subcontinent for both fixed-effect and the random-effects model. The forest plots of the significant associations were given (p < 0.05*)
Figure 4

[A] Forest plot depicting the odds ratios (ORs), and 95% CI of, rs4646903/CYP1A1 for its association with lung cancer in smokers in a dominant model in the Indian subcontinent, [A'] Forest plot depicting the odds ratios (ORs), and 95% CI of, rs4646903/CYP1A1 for its association with lung cancer in non-smokers in a dominant model in the Indian subcontinent, [A''] Combine forest plot depicting the odds ratios (ORs), and 95% CI of, rs4646903/CYP1A1 for its association with lung cancer stratified by smoking status in a dominant model in the Indian subcontinent, [B] Forest plot depicting the odds ratios (ORs), and 95% CI of, rs25487/XRCC1 for its association with lung cancer in smokers in a recessive model in the Indian subcontinent, [B'] Forest plot depicting the odds ratios (ORs), and 95% CI of, rs25487/XRCC1 for its association with lung cancer in non-smokers in a recessive model in the Indian subcontinent, [B''] Combine forest plot depicting the odds ratios (ORs), and 95% CI of, rs25487/XRCC1 for its association with lung cancer stratified by smoking status in a recessive model in the Indian subcontinent, [C] Forest plot depicting the odds ratios (ORs), and 95% CI of, rs1048943/CYP1A1 for its association with lung cancer in smokers in a dominant model in the Indian subcontinent, [C'] Forest plot depicting the odds ratios (ORs), and 95% CI of, rs1048943/CYP1A1 for its association with lung cancer stratified by smoking status in a dominant model in the Indian subcontinent, [C''] Combine forest plot depicting the odds ratios (ORs), and 95% CI of, rs1048943/CYP1A1 for its association with lung cancer stratified by smoking status in a dominant model in the Indian subcontinent. All the forest plots were generated in a fixed-effect meta-regression model. The forest plots of the significant associations were given (p < 0.05*)

Supplementary Files

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

- SupplementaryInformation.pdf