IMPACT OF DNA REPAIR, FOLATE AND GLUTATHIONE GENE POLYMORPHISMS ON RISK OF NON-SMALL CELL LUNG CANCER

Title: IMPACT OF DNA REPAIR, FOLATE AND GLUTATHIONE GENE POLYMORPHISMS ON RISK OF NON-SMALL CELL LUNG CANCER.

Authors: Cristina Pérez-Ramírez, Marisa Cañadas-Garre, Ahmed Alnatsha, Eduardo Villar, Juan Ramón Delgado, Miguel Ángel Calleja-Hernández, María José Faus-Dáder

PII: S0344-0338(17)30895-6
DOI: https://doi.org/10.1016/j.prp.2017.11.015
Reference: PRP 51949

To appear in:

Received date: 5-9-2017
Revised date: 8-11-2017
Accepted date: 16-11-2017

1 TITLE

IMPACT OF DNA REPAIR, FOLATE AND GLUTATHIONE GENE POLYMORPHISMS ON RISK OF NON-SMALL CELL LUNG CANCER.

AUTHORS:

Cristina Pérez-Ramírez[1,2], Marisa Cañadas-Garre[1,3], Ahmed Alnatsha[1,3], Eduardo Villar[4], Juan Ramón Delgado[5], Miguel Ángel Calleja-Hernández[1,6], María José Faus-Dáder[2].

Author for correspondence

Marisa Cañadas-Garre
Nephrology Research Group
Centre for Public Health
Queen's University of Belfast
c/o Regional Genetics Centre, Level A, Tower Block
Belfast City Hospital, Lisburn Road, Belfast, BT9 7AB
Telephone: 02890 638460
Fax: 02890 235900
http://go.qub.ac.uk/Marisa
E-mail: marisacgarre@gmail.com

Affiliations

1 **Pharmacogenetics Unit**
UGC Provincial de Farmacia de Granada
Instituto de Investigación Biosanitaria de Granada
Complejo Hospitalario Universitario de Granada
Avda. Fuerzas Armadas, 2
Telephone: +34958020108
Fax: +34901021804

2 **Department of Biochemistry**
Faculty of Pharmacy
University of Granada
Campus Universitario de Cartuja, s/n
18071 Granada, Spain
Telephone: +34958243838

3 **Nephrology Research Group**
Centre for Public Health
Queen's University of Belfast
c/o Regional Genetics Centre, Level A, Tower Block
Belfast City Hospital, Lisburn Road, Belfast, BT9 7AB
Telephone: 02890 638460
Fax: 02890 235900

3 **Department of Molecular Medicine**
Faculty of Medicine
University of Tübingen
Geissweg 5
72076 Tübingen
Telephone: +49 07071 29-73682

4 Pathology Service
UGC Anatomía Patológica
Instituto de Investigación Biosanitaria de Granada
Complejo Hospitalario Universitario de Granada
Avda. Fuerzas Armadas, 2
18014 Granada, Spain
Telephone: +34958020374

5 Medical Oncology Service
UGC Oncología Médica
Instituto de Investigación Biosanitaria de Granada
Complejo Hospitalario Universitario de Granada
Avda. Fuerzas Armadas, 2
18014 Granada, Spain
Telephone: +34958020346

6 Department of Pharmacology
Faculty of Pharmacy
University of Granada
Campus Universitario de Cartuja, s/n
18071 Granada, Spain
Telephone: +34958243538

Mails

Cristina Pérez-Ramírez	cperezramirez87@gmail.com
Ahmed Alnatsha	r.j.b-man@hotmail.com
Marisa Cañadas-Garre	marisacgarre@gmail.com
Eduardo Villar	eduardovillar6@gmail.com
Juan Ramón Delgado	juanramondelgado@gmail.com
María José Faus-Dáder	mfaus@ugr.es
Miguel Ángel Calleja Hernández	mangel.calleja.sspa@juntadeandalucia.es

Keywords:

NON-SMALL CELL LUNG CANCER
RISK
POLYMORPHISMS
DNA REPAIR PATHWAY
FOLATE METABOLISM PATHWAY
GLUTATHIONE METABOLISM PATHWAY

WORD COUNT (excluding references and figure legends):

Abstract: 98 words
Main Text: 2801 words

FUNDING

This work was partly supported by a contract for Marisa Cañadas-Garre (Técnicos de Apoyo Subprogram. CA12/00097) from Instituto de Salud Carlos III, Ministerio de Economía y Competitividad and a research grant for Cristina Pérez-Ramírez (FPU12/04722), from Ministerio de Educación, Cultura y Deporte. The Hospital Universitario Virgen de las Nieves Biobank was supported by grants cofinanced by FEDER funds (EU) from the Instituto de Salud Carlos III (PT13/0010/0039).

CONFLICT OF INTEREST

The authors declare that there is not conflict of interest that could be perceived as prejudicing the impartiality of the research reported and there is not any competing financial interest in relation to the work described in this article.

2 ABSTRACT

Lung cancer, particularly non-small cell lung cancer (NSCLC) subtype, is the leading cause of cancer-related death related worldwide. Numerous gene polymorphisms in DNA repair, folate and glutathione pathways have been associated with susceptibility of NSCLC. We conducted this study to evaluate the effects of ERCC1, ERCC2, ERCC5, XRCC1, XRCC3, MTHFR, MTR, MTHFD1, SLC19A1 and GSTP1 gene polymorphisms on risk of NSCLC.

No association between these gene polymorphisms and susceptibility of NSCLC were found in our patients, suggesting that genetic variations in genes involved in DNA repair, folate and glutathione metabolism pathways may not influence the risk of NSCLC.

3 INTRODUCTION

Lung cancer is one of the most common and serious types of cancer in the world, being the main cause of cancer mortality worldwide [1]. The incidence of lung cancer is approximately 14% in both genders (second after prostate cancer in men and breast cancer in women)[1]. The latest statistics of cancer in the United States have estimated 224,300 new cases and 158,000 deaths in 2016 [1].

Lung cancer is classified into two major types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), which is divided into three types: squamous cell carcinoma, adenocarcinoma and large cell carcinoma [2, 3]. NSCLC cases account for approximately 80% of all lung cancers [4-6]. Most patients are diagnosed in older age (≅65 years) and in late-stage (IIIB-IV), according to the guidelines of the American Joint Committee on Cancer (AJCC) showing low overall survival rates at 5 years (5% for IIIB and 1% for IV stages) [4, 6, 7].

Smoking is by far the biggest cause of lung cancer (It causes more than 80% of the cases), although different genetic factors may have an increased risk of lung cancer. Recently, genetic factors, such as single nucleotide polymorphisms (SNPs) have showed a strong association with risk of lung cancer [2, 8]. One of the most studied pathway is repaired DNA, which plays an essential role on carcinogenesis [9]. The DNA may be damaged through many factors such as tobacco smoke, dietary factors, infectious agents and radiation [10]. These modifications may introduce errors during the replication and the transcription, that induce dysfunction of repair DNA machinery [10]. The leading systems involved in repairing DNA damage are nucleotide-excision repair (NER), base excision repair (BER) and double-strand break repair (DSB) [11, 12]. NER pathway is involved in repairing helix-distorting DNA lesions through ERCC1 (excision repair cross-complementing group 1) and ERCC2 (excision repair cross-complementation group 2), ERCC5 (excision repair cross-complementation group 5) genes [11]. Defective repair in Chinese hamster cells 1 (XRCC1) is the key gene in BER pathway, that removes damaged DNA by small, non-helix-distorting lesions [13]. Most serious modifications, such as deletions and chromosomal aberrations, are repaired by DBS pathway, through X-ray repair complementing defective repair in Chinese hamster cells 3 (XRCC3) gene [14]. Therefore, polymorphisms in these genes may play a crucial role on development of lung cancer [15-23].

Other pathways involved in risk of lung cancer are folate and glutathione metabolism. Folate metabolism is involved in DNA synthesis and DNA methylation [24]. The MTHFR (methylenetetrahydrofolate reductase), MTHFD1 (methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1) and MTR (5-methyltetrahydrofolate-homocysteine methyltransferase) enzymes are involved in methylation of DNA. Genetic alterations in these genes are associated with aberrant methylation of tumor-suppressor genes and hypomethylation of proto-oncogenes [25]. SLC19A1 (reduced folate carrier 1) is also one of the key genes in folate pathway, because of its function as folate transporter [26]. In fact, polymorphisms in this gene are related with DNA synthesis disorders that lead to genomic instability [27]. Therefore, genetic variations in these genes may lead to alteration in the DNA methylation and synthesis, inducing tumorigenesis [28-32]. The damage caused by cytotoxic and carcinogenic agents is catalysed by glutathione transferases enzymes, such as GSTP1 (glutathione S-transferase P1) [33]. This gene, which is located in chromosome 11q13, is mainly expressed in lung tissues and exerts an important function on inhaled carcinogens, such as tobacco carcinogens [34-37]. Thus, polymorphisms in GSTP1 are believed to be associated with risk of lung cancer [37, 38].

Based on above, we conducted this study to evaluate the effects of ERCC1, ERCC2, ERCC5, XRCC1, XRCC3, MTHFR, MTR, MTHFD1, SLC19A1 and GSTP1 gene polymorphisms on risk of NSCLC. To prepare this paper a thorough search was made. These SNPs were selected based on their effects on gene or protein function/expression, etc confirmed by data from biochemical/function studies

4 MATERIAL AND METHODS

A retrospective case-control study was performed.

1.1 Study Subjects

This study involved 174 NSCLC cases and 298 controls of South Spanish Caucasian origin. The eligible cases included patients histologically or cytologically diagnosed as NSCLC between 2003-2015 at the University Hospital Virgen de las Nieves, Granada, Spain. The participants of control group were Caucasian individuals older than 18 years from the same geographic area (South Spain), with available follow-up records at the same hospital. Individuals with prior cancer history were excluded from this study. The subjects enrolled in this case-control study provided their written informed consent form for blood/saliva sample collection and donation to the Biobank. The identification of samples was based on non-patient identifiers and they were treated confidentially.

This case-control study was conducted in accordance with the declaration of Helsinki under the approval of the Ethics and Research Committee of the Sistema Sanitario Público de Andalucía Biobank.

1.2 Sociodemographic and clinical variables

Sociodemographic (gender, smoking status and age at diagnosis) and clinical variables (previous lung disease) were collected from clinical records. Smoking status was classified as current-smokers, former-smokers and never-smokers. Individuals were categorized as never smokers if they smoked less than 100 cigarettes; former smokers had quit at least 5 years prior to the recruitment and current smokers continued to smoke or quit less than 5 years prior to the recruitment. Tumor histology and stage were also collected for the case group. Tumor staging was categorized according to the guidelines of the American Joint Committee on Cancer (AJCC) [39].

1.3 Genetic variables

1.3.1 DNA isolation

DNA samples isolated from blood or saliva were obtained from the Hospital Universitario Virgen de las Nieves Biobank, a part of the Sistema Sanitario Público de Andalucía Biobank.

1.3.2 Detection of gene polymorphisms

ERCC1 (rs11615), ERCC1 (rs3212986), ERCC2 (rs13181), ERCC2 (rs1799793), ERCC2 (rs50872), ERCC2 (rs238405), ERCC2 (rs238416), ERCC5 (rs104778), ERCC5 (rs17655), ERCC5 (rs2296147), ERCC5 (rs2094258), ERCC5 (rs873601), XRCC1 (rs1799782), XRCC1 (rs25487), XRCC3 (rs861539), MTHFR (rs1801131), MTHFR (rs1801133), MTHFD1 (rs2236225), MTR (rs1805087), SLC19A1 (rs1051266) and GSTP1 (rs1695) gene polymorphisms were analyzed by Real-Time PCR using TaqMan® probes. Genotyping methodology was previously described [40]. Briefly, the presence of wild-type and variant alleles was defined by comparing the relative end-point fluorescence created by the degradation of each fluorescently labelled TaqMan® probe (FAM/VIC). Allelic calls were determined using StepOne™ Software v2.3. To evaluate internal consistency, 86% of samples were duplicated, and 15% of the results were confirmed by sequencing analysis.

1.4 Statistical Analysis

Descriptive analysis was performed using R 3.0.1 [49]. Quantitative data were expressed as the mean (± standard deviation) for normally-distributed variables or medians and percentiles (25 and 75) for non-normal distributed variables. The Shapiro-Wilks test was performed to assess normality.

Hardy-Weinberg equilibrium and pairwise haplotype frequencies were estimated, and Lewontin's D prime (D') and the linkage disequilibrium coefficient (r2) were calculated. The bivariate association between NSCLC risk and polymorphisms was assessed for multiple models (genotypic, additive, allelic, dominant and recessive), using the Pearson's chi-square and Fisher's exact test, and evaluated by odds ratio (OR) and their corresponding 95% confidence intervals (CI). The models were defined as follows: allelic (D vs d), dominant ((DD, Dd) vs dd), recessive (DD vs (Dd, dd)) and genotypic (DD vs Dd vs dd) and additive, being D the minor allele and d the major allele. Permutation analysis (EMP1) was used for multiple comparisons, that compares the observed statistic with 100,000 statistics obtained in permutations in the correspondent SNP. To control the family-wise error rate when testing multiple SNPs, the observed statistic was compared with the maximum of permuted statistics over all SNPs (EMP2). Unconditional multiple logistic regression models (genotypic, dominant and recessive) were considered to determine the influence of potential confounding variables on the risk of lung cancer. All tests were two-sided with a significant level of $p<0.05$, and were performed using the free, open-source whole genome association analysis toolset R 3.2.2 or PLINK [41, 42].

2 RESULTS

2.1 Patients characteristics

Clinical, sociodemographic and pathologic characteristics of the 174 NSCLC patients and 298 controls are described in Table 1. The patients had an average age of 60.67 ± 10.58 and consisted of 109 adenocarcinomas (109/174; 62.64%) and 61 squamous cell carcinoma (61/174; 35.06%), most of them with advanced stage (119/174; 68.39%). The median age for the controls was 72.90 [67,80], and the proportion of current smokers was lower in this group ($p<2.2\cdot10^{-16}$; Table 1). Distribution of previous lung disease was similar in both groups ($p=0.618$; Table 1).

2.2 Genotype distribution

Genotype frequencies were in agreement with the values expected under the Hardy-Weinberg equilibrium model. Linkage disequilibrium values D' and r^2 are shown in Table S1. No linkage disequilibrium was showed in any case. All SNPs presented minor allele frequencies over 1%, therefore none of them were excluded from the analysis (Table S2).

2.3 Influence of gene polymorphisms on risk of NSCLC

The bivariate analysis was performed in multiple models: genotypic, additive, allelic, dominant and recessive (Table 2). ERCC5 rs104778 was the only gene polymorphism associated with risk of NSCLC in the recessive model. Carriers of the CC genotype were in higher risk of NSCLC compared with those with T allele ($p=0.04507$; OR=1.624; $CI_{95\%}=1.005, 2.617$). This significant association was confirmed after adjusting by permutation analysis (Table 3). However, no significant association was found in the logistic regression analysis (Table 4). Similarly, no significant differences for other gene polymorphisms (rs11615, rs3212986, rs13181, rs1799793, rs50872, rs238405, rs238416, rs17655, rs2296147, rs2094258, rs873601, rs1799782, rs25487, rs861539, rs1695, rs1801131, rs1801133, rs2236225, rs1805087 and rs1051266) in allele or genotype frequencies were shown between NSCLC patients and controls (Table 2).

5 DISCUSSION

Smoking is the main risk factor for lung cancer development. However, genetic alterations in genes involved in DNA repair pathway, folate and glutathione metabolisms have been reported to play a crucial role in risk of NSCLC. We conducted this study to evaluate the effect of variants in genes involved in DNA repair, folate and glutathione metabolism pathways on NSCLC risk.

DNA repair pathway plays a crucial function in protecting against gene mutation caused by carcinogenesis. ERCC1 is the key enzyme in the NER pathway [43]. ERCC1 rs11615 and rs3212986 gene polymorphisms are the most studied [20]. A meta-analysis, which evaluated 11 studies from Caucasian and Asian population (5009 cases/5542 controls) reported that in the overall population individuals carrying at least one C allele of ERCC1 rs11615 gene polymorphism were in lower risk of NSCLC (OR=0.90; $CI_{95\%}$=0.81, 0.99; I^2=56.3%; $P_{heterogeneity}$=0.011; C vs T) [20]. However, stratified analysis by ethnicity showed no significant association between rs11615 and lung cancer risk in either Asian or Caucasian subgroups [20]. Similarly, a meta-analysis that included 3 studies with 3060 cases and 2729 controls from Caucasian and Asian population failed to find an association between ERCC1 rs3212986 and NSCLC incidence (OR=1.03; $CI_{95\%}$=0.95, 1.11; I^2=0%; $P_{heterogeneity}$=0.746; A vs C) [20]. In our study, we did not find a significant association between neither of those gene polymorphisms and risk of NSCLC, which was consistent with the results of both meta-analyses in Caucasian population. ERCC2 also plays an essential function on DNA repair pathway [44]. Numerous SNPs in ERCC2 gene have been described, being ERCC2 rs13181 and rs1799793 the most investigated [16, 21]. A meta-analysis comprising 28 studies, with 23370 subjects (10242 cases/13128 controls) from Caucasian and Asian origin has recently evaluated the role of ERCC2 Lys751Gln polymorphism in risk of lung cancer [21]. This meta-analysis presented the C allele of ERCC2 rs13181 as an increased risk factor for developing lung cancer, both in overall population (OR=1.186; $CI_{95\%}$=1.089, 1.292; $P_{heterogeneity}$=0.000; A vs GG) and subgroup analysis (Caucasian: OR=1.113; $CI_{95\%}$=1.033, 1.199; $P_{heterogeneity}$=0.005; C vs AA and Asian: OR=1.252; $CI_{95\%}$=1.015, 1.544; $P_{heterogeneity}$=0.238; C vs AA) [21]. For ERCC2 rs1799793, a meta-analysis analyzed a total of 17 studies involving 5820 lung cancer cases and 7378 controls showed that Asian and smoking carriers of ERCC2 rs1799793-A allele were in higher risk for developing NSCLC (OR=1.84; $CI_{95\%}$=1.31, 2.27; $P_{heterogeneity}$=0.238; A vs GG and OR=1.56; $CI_{95\%}$=1.31, 1.87; $P_{heterogeneity}$=0.528; A vs GG, respectively) [16]. However, in our study we did not find a significant association between ERCC2 rs13181 and rs1799793 gene polymorphisms and risk of NSCLC. We also evaluated the role of other ERCC2 gene polymorphisms (rs50872, rs238405 and rs238416) on risk of NSCLC, but no significant association was detected either. To date, no other studies have explored the association between these polymorphisms and NSCLC incidence. Other gene with an important role on DNA repair pathway is ERCC5 [45]. ERCC5 rs17655 and rs104778 gene polymorphisms are the most studied [17]. Recently, a meta-analysis has evaluated the effect of ERCC5 rs17655 on risk of lung cancer [17, 18, 22]. This meta-analysis gathered 6 studies from Caucasian and Asian population including a total of 2293 patients and 2586 controls, and reported higher risk of NSCLC for CC genotype (OR=1.24; $CI_{95\%}$=1.04, 1.48; I^2=40.8%; $P_{heterogeneity}$=0.13; CC vs GG) [17]. No significant association was observed in our patients between ERCC5 rs17655 and risk of NSCLC. The ERCC5 rs104778-CC genotype has been associated with a significantly increased risk of lung cancer (OR= 1.79, $CI_{95\%}$= 1.19, 2.63; CC vs TT) in Caucasian population but not in Asian (OR= 1.23, $CI_{95\%}$= 0.53, 2.85; CC vs TT) [18, 22]. In our study, the recessive model was in consonance with these results. However, the logistic regression analysis did not show significant association. We also analyzed ERCC5 rs2296147, rs2094258 and rs873601 gene polymorphisms, but they did not show any effect on NSCLC risk. To date, no other studies have evaluated the role of these gene polymorphisms on risk of NSCLC. XRCC1 is the main gene in DSB pathway [13]. XRCC1 rs1799782 and rs25487 gene polymorphisms are the most studied [15, 23]. For XRCC1 rs1799782 polymorphism, a meta-analysis, which comprised a total of 16 studies with 4848 cases and 6592 controls from Caucasian and Asian origin, has reported that individuals carrying the AG genotype had lower risk of lung cancer (OR=0.88; $CI_{95\%}$=0.79-0.97; $P_{heterogeneity}$=0.37; AG vs AA) [15]. No other associations were significant (OR=1.07; $CI_{95\%}$=0.85-1.33; $P_{heterogeneity}$=0.21; GG vs AA and OR=0.91; $CI_{95\%}$=0.83-1.00; $P_{heterogeneity}$=0.21; G vs AA) [15]. Recently, a study that recruited 222 healthy controls and 102 patients with lung cancer from Caucasian origin has described increased risk for lung cancer for the GG genotype (OR=10.667; $CI_{95\%}$=1.309-86.933; p=0.007; GG vs A)[23]. The XRCC1 rs25487-A allele was proposed as protective for lung cancer in a meta-analysis that included a total of 30 studies with 10214 cases and 12599 controls (OR=0.86; $CI_{95\%}$=0.77-0.97; $P_{heterogeneity}$=0.08; G vs AA) [15]. However, in our study, no significant association for none of these gene polymorphisms was found. XRCC3 is the most important gene in DBS pathway [14, 46]. XRCC3 rs861539 gene polymorphism is the most relevant [19]. A recent meta-analysis has been performed to evaluate its role on risk of lung cancer [19]. A total of 21 studies, including 6880 lung cancer cases and 8329 controls from Caucasian and Asian population, were included, but no significant association was found (OR=1.02; $CI_{95\%}$=0.91-1.13; I^2=59%; $P_{heterogeneity}$=0.0006; T vs C) [19]. Our study is in consonance with this result, since we did not find significant association between XRCC3 rs861539 and risk of NSCLC.

Other pathways involved in risk of NSCLC are folate and glutathione metabolism. Several proteins are crucial in folate metabolisms such as MTHFR, MTHFD1, MTR and SLC19A1 and have been found to be associated with risk of lung cancer [28-32]. Two polymorphisms in MTHFR, rs1801131 and rs1801133 are the most studied [29, 30]. For MTHFR rs1801131, a meta-analysis that included a total of 7 studies with 5087 cases and 6232 controls from Asian origin, reported no significant association with risk of lung cancer (OR=0.99; $CI_{95\%}$=0.93-1.05; I^2=24.4%; $P_{heterogeneity}$=0.24; C vs A) [29]. Similarly, we did not found a significant association between MTHFR rs1801131 and risk of lung cancer. The MTHFR rs1801133-TT genotype has been associated with higher risk of lung cancer in a meta-analysis that comprised a total of 19 studies with 11644 cases and 12024 controls from Caucasian and Asian population (OR=1.19; $CI_{95\%}$=1.03-1.37; I^2=61%; $P_{heterogeneity}$=0.0003; TT vs C) [30]. However, we did not find a significant association between MTHFR rs1801133 and risk of lung cancer. One of the most investigated polymorphisms in MTHFD1 gene is rs2236225. The importance of MTHFD1 rs2236225 has been evaluated in a previous study composed by 500 lung cancer patients and 517 controls from Asian origin [31]. Although the T allele of rs2236225 polymorphism in MTHFD1 was associated with lower risk of lung cancer [31], this effect was not confirmed in our study (OR=0.76; $CI_{95\%}$=0.59-0.99; T vs CC). The role of SLC19A1 and MTR gene polymorphisms on risk of lung cancer has also been previously studied. In particular, SLC19A1 rs1051266 and MTR rs1805087 have been investigated, but no significant results have been reported [28, 31, 32]. Thus, our results were in agreement with these previous studies. GSTP1 is the key enzyme involved in glutathione metabolism [47]. The most studied polymorphism in GSTP1 gene is rs1695 [37, 38]. A meta-analysis that enrolled 13 studies with 2026 cases and 2451 controls from Asian origin has reported higher risk of lung cancer for GSTP1 rs1695-GG genotype (OR=1.36; $CI_{95\%}$=1.01-1.84; I^2=31.7%; $P_{heterogeneity}$=0.137; GG vs AA)[38]. However, this effect has not been confirmed in Caucasian population. A meta-analysis, which included a total of 14 studies with 2054 case series and 5525 controls from Caucasian origin has reported no association between GSTP1 rs1695-G allele/GG genotype and susceptibility of squamous cell carcinoma, a subtype of NSCLC (OR=1.06;

$CI_{95\%}$=0.93-1.21; I^2=53%; $P_{heterogeneity}$=0.01; G vs A and OR=1.20; $CI_{95\%}$=0.94-1.53; I^2=37%; $P_{heterogeneity}$=0.09) [37]. These results in Caucasian population are in accordance with our study [37].

The main limitation of this study was the limited size of the sample compared to other studies, particularly in the cases group. This may have led to a lack of power to detect associations in some polymorphisms. The strengths of our study include a very homogeneous cohort of cases, only composed by NSCLC patients diagnosed by the same team of pathologists, recruited from the same geographic area, which increases their uniformity.

In summary, we investigated the effect of ERCC1 (rs11615), ERCC1 (rs3212986), ERCC2 (rs13181), ERCC2 (rs1799793), ERCC2 (rs50872), ERCC2 (rs238405), ERCC2 (rs238416), ERCC5 (rs17655), ERCC5 (rs104778), ERCC5 (rs2296147), ERCC5 (rs2094258), ERCC5 (rs873601), XRCC1 (rs1799782), XRCC1 (rs25487), XRCC3 (rs861539), MTHFR (rs1801131), MTHFR (rs1801133), MTHFD1 (rs2236225), MTR (rs1805087), SLC19A1 (rs1051266) and GSTP1 (rs1695) gene polymorphisms on NSCLC risk. The results showed no association between these gene polymorphisms and susceptibility of NSCLC, suggesting that genetic variations in genes involved in DNA repair, folate and glutathione metabolism pathways may not influence the risk of NSCLC.

6 CONCLUSIONS & FUTURE PERSPECTIVE

This study failed to identify an association between ERCC1, ERCC2, ERCC5, XRCC1, XRCC3, MTHFR, MTR, MTHFD1, SLC19A1 and GSTP1 gene polymorphisms and susceptibility to NSCLC. Further studies including a wider spectrum of gene polymorphisms and larger samples will be required to definitely rule out the influence of these genes on risk of NSCLC.

7 EXECUTIVE SUMMARY

Relationship between gene polymorphisms in DNA repair, folate and glutathione pathways and NSCLC risk.

- DNA repair genes such as *ERCC1, ERCC2* and *XRCC1* play a crucial role on development of lung cancer.
- Polymorphisms in folate metabolism are related with DNA synthesis disorders that lead to genomic instability and subsequently cancer susceptibility.
- *GSTP1* is expressed mainly in lung tissues and exerts an important function on inhaled carcinogens, such as tobacco carcinogens. Thus, polymorphisms in this gene are believed to be associated with risk of lung cancer.

Conclusions

Genetic variations in genes involved in DNA repair, folate and glutathione metabolism pathways may not influence the risk of NSCLC.

8 ACKNOWLEDGMENTS

The results of this investigation are part of the doctoral thesis presented by Cristina Pérez-Ramírez at the University of Granada, developed within the doctoral program "Medicina Clínica y Salud Pública".

9 AUTHOR CONTRIBUTIONS

Cristina Pérez-Ramírez revised the bibliography and wrote the protocol for this study and acted as the primary lead in the conception, design and implementation of the project, data collection, molecular analysis, statistical analysis and interpretation, as well as all aspects of the development and writing of the article and responses to internal and external reviewers.

Marisa Cañadas-Garre, PhD, supervised the planning of the study protocol and contributed to the conception, design and implementation of the project, statistical analysis and interpretation as well as all aspects of the development and writing of the article and responses to internal and external reviewers.

Ahmed Alnatsha performed literature review, molecular analysis and interpretation, clinical data collection and interpretation and collaborated in drafting the manuscript.

Eduardo Villar, MD, PhD, performed molecular and pathological analysis and interpretation, clinical data collection and interpretation and collaborated in drafting the manuscript.

Juan Ramón Delgado, MD participated in the clinical data collection and interpretation, and provided clinical guidance and feedback for this study and critical review of the manuscript.

María José Faus-Dáder, PhD and **Miguel Ángel Calleja-Hernández**, PhD, participated in article selection, study quality assessment, critical review of the manuscript and contributed to revisions and responses to internal and external reviewers.

10 REFERENCES

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2016. *CA: a cancer journal for clinicians* 66(1), 7-30 (2016).
2. Marshall AL, Christiani DC: Genetic susceptibility to lung cancer—light at the end of the tunnel? *Carcinogenesis* 34(3), 487-502 (2013).
3. Ansari J, Shackelford RE, El-Osta H: Epigenetics in non-small cell lung cancer: from basics to therapeutics*. Translational lung cancer research* 5(2), 155-171 (2016).
4. Edge S, Byrd, D.R., Compton, C.C., Fritz, A.G., Greene, F.L., Trotti, A: AJCC Cancer Staging Manual. (7th). 649 (2010).
5. Herbst Rs HJ, Lippman Sm.: Lung cancer. *N Engl J Med.* 359(13), 1367-1380 (2008).
6. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA: Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship*. Mayo Clinic proceedings* 83(5), 584-594 (2008).
7. Herbst RS, Heymach JV, Lippman SM: Lung cancer*. The New England journal of medicine* 359(13), 1367-1380 (2008).
8. Risch A, Plass C: Lung cancer epigenetics and genetics*. International Journal of Cancer* 123(1), 1-7 (2008).

9. Dixon K, Kopras E: Genetic alterations and DNA repair in human carcinogenesis. *Semin Cancer Biol* 14(6), 441-448 (2004).

10. Kiyohara C, Yoshimasu K: Genetic polymorphisms in the nucleotide excision repair pathway and lung cancer risk: a meta-analysis. *International journal of medical sciences* 4(2), 59-71 (2007).

11. Sancar A: DNA repair in humans. *Annu Rev Genet* 29, 69-105 (1995).

12. Garcia-Campelo R, Alonso-Curbera G, Anton Aparicio LM, Rosell R: Pharmacogenomics in lung cancer: an analysis of DNA repair gene expression in patients treated with platinum-based chemotherapy. *Expert Opin Pharmacother* 6(12), 2015-2026 (2005).

13. Whitehouse CJ, Taylor RM, Thistlethwaite A *et al.*: XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* 104(1), 107-117 (2001).

14. Brenneman MA, Wagener BM, Miller CA, Allen C, Nickoloff JA: XRCC3 controls the fidelity of homologous recombination: roles for XRCC3 in late stages of recombination. *Molecular cell* 10(2), 387-395 (2002).

15. Wang Y, Yang H, Li H *et al.*: Association between X-ray repair cross complementing group 1 codon 399 and 194 polymorphisms and lung cancer risk: a meta-analysis. *Cancer letters* 285(2), 134-140 (2009).

16. Zhan P, Wang Q, Wei SZ *et al.*: ERCC2/XPD Lys751Gln and Asp312Asn gene polymorphism and lung cancer risk: a meta-analysis involving 22 case-control studies. *J Thorac Oncol* 5(9), 1337-1345 (2010).

17. Zhou B, Hu XM, Wu GY: Association between the XPG gene Asp1104His polymorphism and lung cancer risk. *Genetics and molecular research : GMR* 15(2), (2016).

18. Zienolddiny S, Campa D, Lind H *et al.*: Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis* 27(3), 560-567 (2006).

19. Bei L, Xiao-Dong T, Yu-Fang G, Jian-Ping S, Zhao-Yu Y: DNA repair gene XRCC3 Thr241Met polymorphisms and lung cancer risk: a meta-analysis. *Bulletin du cancer* 102(4), 332-339 (2015).

20. Xu TP, Shen H, Liu LX, Shu YQ: Association of ERCC1-C118T and -C8092A polymorphisms with lung cancer risk and survival of advanced-stage non-small cell lung cancer patients receiving platinum-based chemotherapy: a pooled analysis based on 39 reports. *Gene* 526(2), 265-274 (2013).

21. Tan X, Xian L, Chen X *et al.*: Association between ERCC2 Lys751Gln polymorphism and lung cancer risk: a meta-analysis involving 23,370 subjects. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 17(2), 99-107 (2014).

22. Shen M, Berndt SI, Rothman N *et al.*: Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. *International journal of cancer. Journal international du cancer* 116(5), 768-773 (2005).

23. Catana A, Pop M, Hincu BD *et al.*: The XRCC1 Arg194Trp polymorphism is significantly associated with lung adenocarcinoma: a case-control study in an Eastern European Caucasian group. *OncoTargets and therapy* 8, 3533-3538 (2015).

24. Wei Q, Shen H, Wang LE *et al.*: Association between low dietary folate intake and suboptimal cellular DNA repair capacity. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 12(10), 963-969 (2003).

25. Jacques PF, Bostom AG, Williams RR *et al.*: Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations. *Circulation* 93(1), 7-9 (1996).

26. Hou Z, Matherly LH: Biology of the major facilitative folate transporters SLC19A1 and SLC46A1. *Current topics in membranes* 73, 175-204 (2014).

27. Huang X, Gao Y, He J *et al.*: The association between RFC1 G80A polymorphism and cancer susceptibility: Evidence from 33 studies. *Journal of Cancer* 7(2), 144-152 (2016).

28. Shen M, Rothman N, Berndt SI *et al.*: Polymorphisms in folate metabolic genes and lung cancer risk in Xuan Wei, China. *Lung cancer* 49(3), 299-309 (2005).

29. Mao R, Fan Y, Jin Y, Bai J, Fu S: Methylenetetrahydrofolate reductase gene polymorphisms and lung cancer: a meta-analysis. *Journal of human genetics* 53(4), 340-348 (2008).

30. Zhu N, Gong Y, He J, Xia J, Chen X: Influence of methylenetetrahydrofolate reductase C677T polymorphism on the risk of lung cancer and the clinical response to platinum-based chemotherapy for advanced non-small cell lung cancer: an updated meta-analysis. *Yonsei medical journal* 54(6), 1384-1393 (2013).

31. Liu H, Jin G, Wang H *et al.*: Association of polymorphisms in one-carbon metabolizing genes and lung cancer risk: a case-control study in Chinese population. *Lung cancer* 61(1), 21-29 (2008).

32. Cui Lh YZ, Zhang Tt, Shin Mh, Kim Hn, Choi Js.: Influence of polymorphisms in MTHFR 677 C→T, TYMS 3R→2R and MTR 2756 A→G on NSCLC risk and response to platinum-based chemotherapy in advanced NSCLC. *Pharmacogenomics.* 12(6), 797-808 (2011).

33. Wang Y, Spitz MR, Schabath MB, Ali-Osman F, Mata H, Wu X: Association between glutathione S-transferase p1 polymorphisms and lung cancer risk in Caucasians: a case-control study. *Lung Cancer* 40(1), 25-32 (2003).

34. Saarikoski ST, Voho A, Reinikainen M *et al.*: Combined effect of polymorphic GST genes on individual susceptibility to lung cancer. *International Journal of Cancer* 77(4), 516-521 (1998).

35. Ryberg D, Skaug V, Hewer A *et al.*: Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. *Carcinogenesis* 18(7), 1285-1289 (1997).

36. Anttila S, Hirvonen A, Vainio H, Husgafvel-Pursiainen K, Hayes JD, Ketterer B: Immunohistochemical localization of glutathione S-transferases in human lung. *Cancer research* 53(23), 5643-5648 (1993).

37. Feng X, Zhou HF, Zheng BS, Shi JJ, Luo C, Qin JJ: Association of glutathione S-transferase P1 gene polymorphism with the histological types of lung cancer: a meta-analysis. *Molecular biology reports* 40(3), 2439-2447 (2013).

38. Li XM, Yu XW, Yuan Y *et al.*: Glutathione S-transferase P1, gene-gene interaction, and lung cancer susceptibility in the Chinese population: An updated meta-analysis and review. *Journal of cancer research and therapeutics* 11(3), 565-570 (2015).

39. Edge Sb BR, Compton Cc, Fritz Ag, Green Fl, Trotti a, Editors: AJCC Cancer Staging Manual. 7 ed. *Springer-Verlag New York*, pp. XV–648 (2010).

40. Jimenez-Varo E, Canadas-Garre M, Henriques CI, Pinheiro AM, Gutierrez-Pimentel MJ, Calleja-Hernandez MA: Pharmacogenetics role in the safety of acenocoumarol therapy. *Thrombosis and haemostasis* 112(3), 522-536 (2014).

41. Team RC. R: A language and environment for statistical computing. [Internet]. 2013 Oct. Available from: http://www.r-project.org/.

42. Purcell S, Neale B, Todd-Brown K *et al.*: PLINK: a toolset for whole-genome association and population-based linkage analysis. *American journal of human genetics* 81, (2007).

43. De Laat WL, Appeldoorn E, Jaspers NG, Hoeijmakers JH: DNA structural elements required for ERCC1-XPF endonuclease activity. *The Journal of biological chemistry* 273(14), 7835-7842 (1998).

44. Coin F, Marinoni JC, Rodolfo C, Fribourg S, Pedrini AM, Egly JM: Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nature genetics* 20(2), 184-188 (1998).

45. O'donovan A, Davies AA, Moggs JG, West SC, Wood RD: XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* 371(6496), 432-435 (1994).

46. Shrivastav M, De Haro LP, Nickoloff JA: Regulation of DNA double-strand break repair pathway choice. *Cell research* 18(1), 134-147 (2008).

47. Singh S: Cytoprotective and regulatory functions of glutathione S-transferases in cancer cell proliferation and cell death. *Cancer chemotherapy and pharmacology* 75(1), 1-15 (2015).

Table 1. Clinico-pathologic characteristics of NSCLC cases and controls cases.

Chr	SNP	Minor Allele	Mayor Allele	Model	Cases	Controls	χ²	p-value χ²	p-value Fisher

	Cases	Controls	χ²	p-value	Reference	OR	IC95%
	n (%)	n (%)					
Gender							
Female	46 (26.44)	135 (45.3)	15.7488	$7.233 \cdot 10^{-05}$	Female	2.3	1.53-3.46
Male	128 (73.56)	163 (54.7)					
Previous lung disease							
YES	44 (25.29)	83 (27.85)	0.2486	0.618			
NO	130 (74.71)	215 (72.15)					
Smoking status							
Current-Smokers	85 (48.85)	17 (6.32)	123.5797	$< 2.2 \cdot 10^{-16}$			
Former-smokers	62 (35.63)	107 (39.78)					
Non-smokers	27 (15.52)	145 (53.90)					
Age at NSCLC diagnosis	60.67±10.58	72.90 [67,80]	73.4962	$< 2.2 \cdot 10^{-16}$			
Histology							
Adenocarcinoma	109 (62.64)	-	-	-	-	-	-
Squamous cell carcinoma	61 (35.06)	-					
Unknown	4 (2.30)	-					
Tumor stage							
I, II or IIIA	55 (31.61)	-	-	-	-	-	-
IIIB or IV	119 (68.39)	-					

Qualitative variables: number (percentage)
Quantitative variables:
Normal distribution: mean ± standard deviation.
Non-normal distribution: P_{50} [P_{25}, P_{75}].

*Test Wilcoxon

Table 2. Polymorphisms and association with risk of NSCLC

1	rs1801131	C	A	Genotypic	13/67/94	20/121/157	0.2492	0.8829	0.8783
				Additive	93/255	161/435	0.009469	0.9225	0.9225
				Allelic	93/255	161/435	0.009349	0.923	0.9395
				Dominant	80/94	141/157	0.07904	0.7786	0.8484
				Recessive	13/161	20/278	0.09754	0.7548	0.8519
1	rs1801133	T	C	Genotypic	30/97/47	55/138/105	4.363	0.1129	0.1142
				Additive	157/191	248/348	1.12	0.29	0.29
				Allelic	157/191	248/348	1.101	0.294	0.307
				Dominant	127/47	193/105	3.403	0.06509	0.06708
				Recessive	30/144	55/243	0.1098	0.7403	0.8044
1	rs1805087	G	A	Genotypic	6/48/120	7/76/215	0.8201	0.6636	0.6592
				Additive	60/288	90/506	0.7406	0.3895	0.3895
				Allelic	60/288	90/506	0.7534	0.3854	0.4067
				Dominant	54/120	83/215	0.54	0.4624	0.464
				Recessive	6/168	7/291	0.4956	0.4814	0.5631
11	rs1695	G	A	Genotypic	13/81/80	34/122/142	2.581	0.2751	0.2718
				Additive	107/241	190/406	0.1302	0.7182	0.7182
				Allelic	107/241	190/406	0.1306	0.7178	0.7714
				Dominant	94/80	156/142	0.1236	0.7252	0.7745
				Recessive	13/161	34/264	1.9	0.1681	0.203
13	rs2094258	A	G	Genotypic	8/59/107	11/97/190	0.375	0.829	0.8033
				Additive	75/273	119/477	0.3423	0.5585	0.5585
				Allelic	75/273	119/477	0.3382	0.5609	0.5601
				Dominant	67/107	108/190	0.2414	0.6232	0.6231
				Recessive	8/166	11/287	0.2336	0.6288	0.6341
13	rs2296147	C	T	Genotypic	38/84/52	64/165/69	2.996	0.2236	0.2242
				Additive	160/188	293/303	0.9462	0.3307	0.3307
				Allelic	160/188	293/303	0.8924	0.3448	0.3801

				Dominant	122/52	229/69	2.611	0.1062	0.1259
				Recessive	38/136	64/234	0.008525	0.9264	1
13	rs1047768	T	C	Genotypic	39/85/50	45/154/99	4.175	0.124	0.13
				Additive	163/185	244/352	3.222	0.07266	0.07266
				Allelic	163/185	244/352	3.118	0.07744	0.08853
				Dominant	124/50	199/99	1.023	0.3117	0.3557
				Recessive	39/135	45/253	4.016	0.04507	0.04711
13	rs17655	G	C	Genotypic	14/66/94	17/120/161	1.069	0.5859	0.572
				Additive	94/254	154/442	0.1587	0.6904	0.6904
				Allelic	94/254	154/442	0.156	0.6929	0.702
				Dominant	80/94	137/161	6.58×10^{-07}	0.9994	1
				Recessive	14/160	17/281	0.9813	0.3219	0.3397
13	rs873601	G	A	Genotypic	15/59/100	23/118/157	1.522	0.4673	0.4549
				Additive	89/259	164/432	0.4045	0.5248	0.5248
				Allelic	89/259	164/432	0.4224	0.5157	0.5428
				Dominant	74/100	141/157	1.015	0.3137	0.3387
				Recessive	15/159	23/275	0.1209	0.7281	0.7287
14	rs2236225	T	C	Genotypic	32/81/61	64/137/97	0.7285	0.6947	0.7107
				Additive	145/203	265/331	0.6597	0.4167	0.4167
				Allelic	145/203	265/331	0.6993	0.403	0.4147
				Dominant	113/61	201/97	0.3101	0.5776	0.6136
				Recessive	32/142	64/234	0.6456	0.4217	0.4775
14	rs861539	T	C	Genotypic	16/86/72	43/148/107	3.277	0.1943	0.1924
				Additive	118/230	234/362	2.865	0.09052	0.09052
				Allelic	118/230	234/362	2.693	0.1008	0.1087
				Dominant	102/72	191/107	1.398	0.2371	0.24
				Recessive	16/158	43/255	2.752	0.09715	0.1129
19	rs25487	A	G	Genotypic	23/84/67	42/131/125	0.8301	0.6603	0.6605

				Additive	130/218	215/381	0.1531	0.6956	0.6956
				Allelic	130/218	215/381	0.1558	0.693	0.7262
				Dominant	107/67	173/125	0.5389	0.4629	0.4972
				Recessive	23/151	42/256	0.07092	0.79	0.8901
19	rs1799782	T	C	Genotypic	1/25/148	3/46/249	NA	NA	0.9221
				Additive	27/321	52/544	0.2624	0.6084	0.6084
				Allelic	27/321	52/544	0.2675	0.605	0.6286
				Dominant	26/148	49/249	NA	NA	0.6975
				Recessive	1/173	3/295	NA	NA	1
19	rs13181	G	T	Genotypic	30/68/76	33/134/131	4.022	0.1339	0.1389
				Additive	128/220	200/396	0.954	0.3287	0.3287
				Allelic	128/220	200/396	1.008	0.3155	0.322
				Dominant	98/76	167/131	0.003537	0.9526	1
				Recessive	30/144	33/265	3.613	0.05733	0.06804
19	rs238416	A	G	Genotypic	25/85/64	55/141/102	1.341	0.5114	0.5273
				Additive	135/213	251/345	0.9931	0.319	0.319
				Allelic	135/213	251/345	1.003	0.3167	0.337
				Dominant	110/64	196/102	0.3141	0.5752	0.6177
				Recessive	25/149	55/243	1.305	0.2534	0.309
19	rs50872	T	C	Genotypic	6/63/105	9/92/197	1.585	0.4526	0.462
				Additive	75/273	110/486	1.395	0.2376	0.2376
				Allelic	75/273	110/486	1.336	0.2477	0.2694
				Dominant	69/105	101/197	1.583	0.2083	0.2332
				Recessive	6/168	9/289	0.06544	0.7981	0.7913
19	rs1799793	A	G	Genotypic	27/69/78	32/128/138	2.346	0.3094	0.3141
				Additive	123/225	192/404	0.9121	0.3395	0.3395
				Allelic	123/225	192/404	0.9681	0.3251	0.3524
				Dominant	96/78	160/138	0.0971	0.7553	0.7746

				Recessive	27/147	32/266	2.294	0.1299	0.1493
19	rs238405	A	T	Genotypic	38/84/52	79/137/82	1.308	0.5199	0.5172
				Additive	160/188	295/301	1.026	0.3111	0.3111
				Allelic	160/188	295/301	1.09	0.2964	0.3116
				Dominant	122/52	216/82	0.3031	0.582	0.5979
				Recessive	38/136	79/219	1.286	0.2569	0.2711
19	rs3212986	T	G	Genotypic	14/68/92	21/117/160	0.1625	0.9219	0.9121
				Additive	96/252	159/437	0.0914	0.7624	0.7624
				Allelic	96/252	159/437	0.09195	0.7617	0.7618
				Dominant	82/92	138/160	0.02952	0.8636	0.9238
				Recessive	14/160	21/277	0.1597	0.6894	0.7177
19	rs11615	C	T	Genotypic	30/85/59	47/144/107	0.2764	0.8709	0.8709
				Additive	145/203	238/358	0.2756	0.5996	0.5996
				Allelic	145/203	238/358	0.2739	0.6007	0.6307
				Dominant	115/59	191/107	0.1923	0.661	0.6902
				Recessive	30/144	47/251	0.1738	0.6768	0.6995
21	rs1051266	A	G	Genotypic	43/80/51	71/162/65	4.056	0.1316	0.1358
				Additive	166/182	304/292	0.9854	0.3209	0.3209
				Allelic	166/182	304/292	0.9603	0.3271	0.3452
				Dominant	123/51	233/65	3.332	0.06794	0.07644
				Recessive	43/131	71/227	0.0472	0.828	0.8246
1	rs1801131	C	A	Genotypic	13/67/94	20/121/157	0.2492	0.8829	0.8783
				Additive	93/255	161/435	0.009469	0.9225	0.9225
				Allelic	93/255	161/435	0.009349	0.923	0.9395
				Dominant	80/94	141/157	0.07904	0.7786	0.8484
				Recessive	13/161	20/278	0.09754	0.7548	0.8519
1	rs1801133	T	C	Genotypic	30/97/47	55/138/105	4.363	0.1129	0.1142
				Additive	157/191	248/348	1.12	0.29	0.29

				Allelic	157/191	248/348	1.101	0.294	0.307
				Dominant	127/47	193/105	3.403	0.06509	0.06708
				Recessive	30/144	55/243	0.1098	0.7403	0.8044
1	rs1805087	G	A	Genotypic	6/48/120	7/76/215	0.8201	0.6636	0.6592
				Additive	60/288	90/506	0.7406	0.3895	0.3895
				Allelic	60/288	90/506	0.7534	0.3854	0.4067
				Dominant	54/120	83/215	0.54	0.4624	0.464
				Recessive	6/168	7/291	0.4956	0.4814	0.5631
11	rs1695	G	A	Genotypic	13/81/80	34/122/142	2.581	0.2751	0.2718
				Additive	107/241	190/406	0.1302	0.7182	0.7182
				Allelic	107/241	190/406	0.1306	0.7178	0.7714
				Dominant	94/80	156/142	0.1236	0.7252	0.7745
				Recessive	13/161	34/264	1.9	0.1681	0.203
13	rs2094258	A	G	Genotypic	8/59/107	11/97/190	0.375	0.829	0.8033
				Additive	75/273	119/477	0.3423	0.5585	0.5585
				Allelic	75/273	119/477	0.3382	0.5609	0.5601
				Dominant	67/107	108/190	0.2414	0.6232	0.6231
				Recessive	8/166	11/287	0.2336	0.6288	0.6341
13	rs2296147	C	T	Genotypic	38/84/52	64/165/69	2.996	0.2236	0.2242
				Additive	160/188	293/303	0.9462	0.3307	0.3307
				Allelic	160/188	293/303	0.8924	0.3448	0.3801
				Dominant	122/52	229/69	2.611	0.1062	0.1259
				Recessive	38/136	64/234	0.008525	0.9264	1
13	rs1047768	T	C	Genotypic	39/85/50	45/154/99	4.175	0.124	0.13
				Additive	163/185	244/352	3.222	0.07266	0.07266
				Allelic	163/185	244/352	3.118	0.07744	0.08853
				Dominant	124/50	199/99	1.023	0.3117	0.3557
				Recessive	39/135	45/253	4.016	0.04507	0.04711

13	rs17655	G	C	Genotypic	14/66/94	17/120/161	1.069	0.5859	0.572
				Additive	94/254	154/442	0.1587	0.6904	0.6904
				Allelic	94/254	154/442	0.156	0.6929	0.702
13	rs17655	G	C	Dominant	80/94	137/161	6.58e-07	0.9994	1
				Recessive	14/160	17/281	0.9813	0.3219	0.3397
13	rs873601	G	A	Genotypic	15/59/100	23/118/157	1.522	0.4673	0.4549
				Additive	89/259	164/432	0.4045	0.5248	0.5248
				Allelic	89/259	164/432	0.4224	0.5157	0.5428
				Dominant	74/100	141/157	1.015	0.3137	0.3387
				Recessive	15/159	23/275	0.1209	0.7281	0.7287
14	rs2236225	T	C	Genotypic	32/81/61	64/137/97	0.7285	0.6947	0.7107
				Additive	145/203	265/331	0.6597	0.4167	0.4167
				Allelic	145/203	265/331	0.6993	0.403	0.4147
				Dominant	113/61	201/97	0.3101	0.5776	0.6136
				Recessive	32/142	64/234	0.6456	0.4217	0.4775
14	rs861539	T	C	Genotypic	16/86/72	43/148/107	3.277	0.1943	0.1924
				Additive	118/230	234/362	2.865	0.09052	0.09052
				Allelic	118/230	234/362	2.693	0.1008	0.1087
				Dominant	102/72	191/107	1.398	0.2371	0.24
				Recessive	16/158	43/255	2.752	0.09715	0.1129
19	rs25487	A	G	Genotypic	23/84/67	42/131/125	0.8301	0.6603	0.6605
				Additive	130/218	215/381	0.1531	0.6956	0.6956
				Allelic	130/218	215/381	0.1558	0.693	0.7262
				Dominant	107/67	173/125	0.5389	0.4629	0.4972
				Recessive	23/151	42/256	0.07092	0.79	0.8901
19	rs1799782	T	C	Genotypic	1/25/148	3/46/249	NA	NA	0.9221
				Additive	27/321	52/544	0.2624	0.6084	0.6084
				Allelic	27/321	52/544	0.2675	0.605	0.6286

				Dominant	26/148	49/249	NA	NA	0.6975
				Recessive	1/173	3/295	NA	NA	1
19	rs13181	G	T	Genotypic	30/68/76	33/134/131	4.022	0.1339	0.1389
				Additive	128/220	200/396	0.954	0.3287	0.3287
				Allelic	128/220	200/396	1.008	0.3155	0.322
				Dominant	98/76	167/131	0.003537	0.9526	1
				Recessive	30/144	33/265	3.613	0.05733	0.06804
19	rs238416	A	G	Genotypic	25/85/64	55/141/102	1.341	0.5114	0.5273
				Additive	135/213	251/345	0.9931	0.319	0.319
				Allelic	135/213	251/345	1.003	0.3167	0.337
				Dominant	110/64	196/102	0.3141	0.5752	0.6177
				Recessive	25/149	55/243	1.305	0.2534	0.309
19	rs50872	T	C	Genotypic	6/63/105	9/92/197	1.585	0.4526	0.462
				Additive	75/273	110/486	1.395	0.2376	0.2376
				Allelic	75/273	110/486	1.336	0.2477	0.2694
				Dominant	69/105	101/197	1.583	0.2083	0.2332
				Recessive	6/168	9/289	0.06544	0.7981	0.7913
19	rs1799793	A	G	Genotypic	27/69/78	32/128/138	2.346	0.3094	0.3141
				Additive	123/225	192/404	0.9121	0.3395	0.3395
				Allelic	123/225	192/404	0.9681	0.3251	0.3524
				Dominant	96/78	160/138	0.0971	0.7553	0.7746
				Recessive	27/147	32/266	2.294	0.1299	0.1493
19	rs238405	A	T	Genotypic	38/84/52	79/137/82	1.308	0.5199	0.5172
				Additive	160/188	295/301	1.026	0.3111	0.3111
				Allelic	160/188	295/301	1.09	0.2964	0.3116
				Dominant	122/52	216/82	0.3031	0.582	0.5979
				Recessive	38/136	79/219	1.286	0.2569	0.2711
19	rs3212986	T	G	Genotypic	14/68/92	21/117/160	0.1625	0.9219	0.9121

				Additive	96/252	159/437	0.0914	0.7624	0.7624
				Allelic	96/252	159/437	0.09195	0.7617	0.7618
				Dominant	82/92	138/160	0.02952	0.8636	0.9238
				Recessive	14/160	21/277	0.1597	0.6894	0.7177
19	rs11615	C	T	Genotypic	30/85/59	47/144/107	0.2764	0.8709	0.8709
				Additive	145/203	238/358	0.2756	0.5996	0.5996
				Allelic	145/203	238/358	0.2739	0.6007	0.6307
				Dominant	115/59	191/107	0.1923	0.661	0.6902
				Recessive	30/144	47/251	0.1738	0.6768	0.6995
21	rs1051266	A	G	Genotypic	43/80/51	71/162/65	4.056	0.1316	0.1358
				Additive	166/182	304/292	0.9854	0.3209	0.3209
				Allelic	166/182	304/292	0.9603	0.3271	0.3452
				Dominant	123/51	233/65	3.332	0.06794	0.07644
				Recessive	43/131	71/227	0.0472	0.828	0.8246

Chr: chromosome; NA: not applicable

Table 3. Permutation analysis

SNP	Genotypic		Additive		Recessive		Dominant	
	EMP1	EMP2	EMP1	EMP2	EMP1	EMP2	EMP1	EMP2
rs1801131	0.8774	1	0.9324	1	0.8502	1	0.8478	1
rs1801133	0.1133	0.8579	0.2846	0.9973	0.8046	1	0.06727	0.6692
rs1805087	0.3312	1	0.4109	0.9998	0.3171	1	0.2832	1
rs1695	0.2744	0.9942	0.7636	1	0.2045	0.9548	0.7743	1
rs2094258	0.7184	1	0.5978	1	0.7013	1	0.6119	1
rs2296147	0.2274	0.9833	0.364	0.9991	1	1	0.1275	0.8479

rs1047768	0.1271	0.8826	0.08115	0.7392	0.04669	0.517	0.3549	0.9979

	Genotypic	Dominant	Recessive

rs17655	0.5838	1	0.6915	1	0.3355	0.9984	0.9978	1
rs873601	0.4638	1	0.5174	1	0.8598	1	0.3385	0.998
rs2236225	0.7126	1	0.4227	0.9999	0.4763	0.9999	0.6155	1
rs861539	0.1923	0.9715	0.08657	0.8038	0.1113	0.8246	0.2397	0.9878
rs25487	0.655	1	0.7299	1	0.8903	1	0.4972	1
rs1799782	1	1	0.6322	1	1	1	1	1
rs13181	0.1369	0.904	0.3173	0.9989	0.06758	0.6392	1	1
rs238416	0.5175	1	0.3405	0.9985	0.3089	0.9908	0.6167	1
rs50872	0.2754	0.9999	0.2372	0.9912	0.7015	1	0.1625	0.9766
rs1799793	0.3097	0.9974	0.3668	0.9992	0.1493	0.8963	0.773	1
rs238405	0.5193	1	0.3182	0.9981	0.2705	0.9913	0.598	1
rs3212986	0.9327	1	0.7831	1	0.7188	1	0.9228	1
rs11615	0.8629	1	0.6151	1	0.7008	1	0.6922	1
rs1051266	0.1349	0.8991	0.3411	0.9988	0.9117	1	0.07669	0.7009

EMP1: point-wise empirical p-value obtained by permutation; EMP2: empirical p-value that controls for multiple SNPs

	TT vs CC			CT vs CC			T- vs CC			TT vs C-			
	p-value	OR	CI95%	p-value	OR	IC95%	p-value	OR	IC95%	p-value	OR	IC95%	p-
Age													
≤60	<0.001	4.85	2.80-8.54	<0.001	4.85	2.80-8.54	<0.001	4.88	2.83-8.59	<0.001	4.84	2.80-8.53	<0
Smoking status													
Current-Smokers	<0.001	6.52	3.52-12.59	<0.001	6.52	3.52-12.59	<0.001	6.54	3.53-12.60	<0.001	6.49	3.50-12.52	<0
Non-Smokers	<0.001	0.31	0.18-0.53	<0.001	0.31	0.18- 0.53	<0.001	0.31	0.18-0.53	<0.001	0.31	0.18-0.53	<0
ERCC5 rs1047768	0.549	1.23	0.62-2.46	0.616	0.87	0.51- 1.49	0.859	0.95	0.58-1.59	0.053	1.34	0.73-2.46	0.6

Table 4. Influence of clinical characteristic and ERCC5 rs1047768 gene polymorphisms on risk of NSCLC