

# The breast cancer susceptibility-related polymorphisms at the TOX3/LOC643714 locus associated with lung cancer risk in a Han Chinese population

Chaowen Jiang<sup>1,\*</sup>, Shilong Yu<sup>1,\*</sup>, Pin Qian<sup>1,\*</sup>, Ruiling Guo<sup>2</sup>, Ruijie Zhang<sup>3</sup>, Zhi Ao<sup>1</sup>, Qi Li<sup>1</sup>, Guoming Wu<sup>1</sup>, Yan Chen<sup>1</sup>, Jin Li<sup>1</sup>, Changzheng Wang<sup>1</sup>, Wei Yao<sup>1</sup>, Jiancheng Xu<sup>1</sup>, Guisheng Qian<sup>1</sup>, Fuyun Ji<sup>1</sup>

<sup>1</sup>Institute of Human Respiratory Disease, Xinqiao Hospital, The Third Military Medical University, Chongqing 400037, China

<sup>2</sup>Department of Respiratory Diseases, 324<sup>th</sup> Hospital of People's Liberation Army (No.324 Hospital of PLA), Chongqing 400020, China

<sup>3</sup>Department of Respiratory Medicine, The Third Xiangya Hospital, Central South University, Changsha, Hunan 410013, China

\*These authors have contributed equally to this work

**Correspondence to:** Fuyun Ji, **email:** Jif@email.chop.edu, jifuyun@263.net  
Guisheng Qian, **email:** qiangs1220@163.com

**Keywords:** lung cancer, risk, TOX3/LOC643714, polymorphisms, breast cancer

**Received:** February 02, 2016

**Accepted:** July 09, 2016

**Published:** July 28, 2016

## ABSTRACT

It has been well established that besides environmental factors, genetic factors are also associated with lung cancer risk. However, to date, the prior identified genetic variants and loci only explain a small fraction of the familial risk of lung cancer. Hence it is vital to investigate the remaining missing heritability to understand the development and process of lung cancer. In the study, to test our hypothesis that the previously identified breast cancer risk-associated genetic polymorphisms at the TOX3/LOC643714 locus might contribute to lung cancer risk, 16 SNPs at the TOX3/LOC643714 locus were evaluated in a Han Chinese population based on a case-control study. Pearson's chi-square test or Fisher's exact test revealed that rs9933638, rs12443621, and rs3104746 were significantly associated with lung cancer risk ( $P < 0.001$ ,  $P < 0.001$ , and  $P = 0.005$ , respectively). Logistic regression analyses displayed that lung cancer risk of individuals with rs9933638(GG+GA) were 1.89 times higher than that of rs9933638AA carriers (OR = 1.893, 95% CI = 1.308-2.741,  $P = 0.001$ ). Similar findings were manifested for rs12443621 (OR = 1.824, 95% CI = 1.272-2.616,  $P = 0.001$ , rs12443621(GG+GA) carriers vs. rs12443621AA carriers) and rs3104746 (OR = 1.665, 95% CI = 1.243-2.230,  $P = 0.001$ , rs3104746TT carriers vs. rs3104746(TA+AA) carriers). The study discovered for the first time that three SNPs (rs9933638, rs12443621, and rs3104746) at the TOX3/LOC643714 locus contributed to lung cancer risk, providing new evidences that lung cancer and breast cancer are linked at the molecular and genetic level to a certain extent.

## INTRODUCTION

Lung cancer is a major public health concern worldwide, causing as many deaths as next four most deadly cancers combined (breast, prostate, colon, and pancreas). Non-small cell lung cancer (NSCLC) is the commonest lung cancer histology, comprised primarily of adenocarcinoma and squamous cell carcinoma [1]. In China, lung cancer has replaced liver cancer to become the leading cause of cancer-related deaths and accounted

for 29% of all male cancer deaths and 23% of all female cancer deaths, totally approximately 220,5200 deaths, in 2014 (World Health Statistics, WHO, 2014). It has been established that multiple environment (mainly cigarette smoking and asbestos) and genetic factors are involved and gene-environment interactions exist in the development and process of lung cancer [2–4].

In the past few years, several genetic variants and loci have been identified to be the genetic risk factors of lung cancer [5–9]. However, to date, these known common

loci only explain a small fraction of the familial risk of lung cancer. Hence it is vital to investigate the remaining missing heritability to understand the development and process of lung cancer. The TOX3/LOC643714 locus on chromosome 16q12.1 was one of the first breast cancer regions identified through genome-wide association study (GWAS) in populations of European and East Asian origin [10]. The chromosomal region spanning the 5' end of TOX3, the intergenic region between TOX3 and LOC643714, and the entire coding part of LOC643714 are located in a 133 kb linkage disequilibrium (LD) block [11]. LOC643714 is an uncharacterized gene of unknown function [<http://www.ncbi.nlm.nih.gov/gene/643714>]. Identified first in a screen for transcripts containing trinucleotide repeat expansions, TOX3 gene, also termed trinucleotide repeat containing 9 (TNRC9), belongs to the high-mobility group (HMG) family of nonhistone chromatin proteins, indicating its potential role as a transcript factor [12] and involvement in bending and unwinding of DNA and alteration of chromatin structure [13]. TOX3 is largely expressed within the brain in the frontal and occipital lobe, the central nervous system (CNS), and the ileum in normal human tissues. Through interaction with the cAMP-response-element-binding protein (CREB), TOX3 regulates Ca<sup>2+</sup>-dependent neuronal transcription [14]. The overexpression of TOX3 induces transcription involving isolated estrogen responsive elements and estrogen-responsive promoters, and protects neuronal cells from cell death caused by endoplasmic reticulum stress or BCL2-Associated X Protein (BAX) overexpression through the induction of anti-apoptotic transcripts and repression of pro-apoptotic transcripts [15].

In the several GWAS, a number of novel genetic variants and loci at the TOX3/LOC643714 locus were identified to be independently associated with elevated risk of breast cancer and rs3803662 was highlighted for its potential biological contribution to the development of breast cancer. Rs3803662 was first reported to be related with breast cancer risk in an Icelandic population [13]. The following studies demonstrated that rs3803662 was not only associated with increased risk of breast cancer in both BRCA1 and BRCA2-mutation carriers and estrogen receptor (ER) positive patients [16], male breast cancer [17, 18], triple-negative (TN) breast cancer [19, 20], but also with overall survival (OS) of breast cancer [21, 22]. The other SNPs at the TOX3/LOC643714 locus reported to be associated with risk of breast cancer included rs3104746 [23, 24], rs8051542 [25–30], rs4784227 [25, 31–34], rs12443621 [25–27, 35–37], rs3112612 [25, 38], rs3112562 [23], rs3104793 [23, 39], and rs8046994 [23]. Among them, a slice of SNPs were shown to be related with risk of breast cancer in certain ethnic populations. For example, rs3104746, rs3112562, rs3104793, and rs8046994 were indicated to be associated with risk of breast cancer in an African-American women population [23]. Rs8051542 was found to be significantly associated

with breast cancer risk in Tunisians [27] and rs3112612 yielded moderate predictive power in Ashkenazi Jewish women with strong family histories but no identifiable BRCA1/2 mutation [37]. Moreover, rs8051542, rs3803662, rs4784227, and rs12443621 were found to be associated with breast cancer risk of Chinese population [25, 28, 37]. Besides breast cancer, rs3104788 and rs3104767 were indicated to be susceptible to periodic leg movements (PLM) [40, 41] and restless legs syndrome susceptibility [42, 43], respectively.

Though the known etiology and carcinogenesis of lung cancer are different from that of breast cancer, the patients of the two diseases could be treated by some common chemotherapeutic agents such as taxanes (paclitaxel and taxotere), vincristine (Navelbine, NVB), and platinum-containing anticancer drugs (cisplatin and carboplatin). Most importantly and interestingly, a recent meta-analysis based on four lung cancer GWAS in populations of European ancestry, the MD Anderson Cancer Center (MDACC) GWAS, the Institute of Cancer Research (ICR) GWAS, the National Cancer Institute (NCI) GWAS, and the International Agency for Research on Cancer (IARC) GWAS, identified a rare variant of *BRCA2* gene, a well-known risk factor for breast, ovarian, and aggressive prostate cancers, to an increased risk of squamous cell lung cancer among cigarette smokers, suggesting that lung cancer and breast cancer are linked at a molecular and genetic level to a certain extent [44].

Due to the findings of a molecular and genetic linkage between lung cancer and breast cancer and the potential involvement of TOX3 in bending and unwinding of DNA and alteration of chromatin structure, we deduced that the previously identified breast cancer susceptibility-associated variants and loci at the TOX3/LOC643714 locus may contribute to lung cancer risk. To test the hypothesis, 16 SNPs at the TOX3/LOC643714 locus were selected and genotyped in a Han Chinese population from Southwestern China based on a case-control study. The genotyping data demonstrated that three SNPs (rs9933638, rs12443621, and rs3104746) at the TOX3/LOC643714 locus were associated with elevated risk of lung cancer and might be potentially biologically relevant to lung carcinogenesis.

## RESULTS

### Subject characteristics

Totally, 352 unrelated patients and 407 unrelated controls were recruited from Southwestern China for the case-control study. No female cigarette smokers were gathered. The general descriptive characteristics of the study population were given in Table 1. The median number of pack-years of combined cases and controls was utilized as the cut-point to stratify the cigarette smoking subjects. As shown in Table 1, there were no significant

**Table 1: General characteristics of controls and lung cancer cases in a Han Chinese population**

| Characteristics                    | Controls (n = 407) (%) | Cases (n = 352) (%) | p-Value             |
|------------------------------------|------------------------|---------------------|---------------------|
| Gender                             |                        |                     |                     |
| Male                               | 320 (78.62)            | 272 (77.27)         | 0.654 <sup>a</sup>  |
| Female                             | 87(21.38)              | 80 (22.73)          |                     |
| Age at diagnosis (years)           |                        |                     |                     |
| <39                                | 23 (5.65)              | 18(5.11)            | 0.819 <sup>a</sup>  |
| 40-49                              | 73(17.94)              | 54 (15.34)          |                     |
| 50-59                              | 125 (30.71)            | 112 (31.82)         |                     |
| 60-69                              | 116 (28.50)            | 104 (29.55)         |                     |
| >70                                | 70 (17.20)             | 63 (17.90)          |                     |
| Mean age ± SD                      | 57.90 ±10.01           | 58.88 ±11.40        | 0.916 <sup>b</sup>  |
| Histology                          |                        |                     |                     |
| Adenocarcinoma                     | -                      | 148 (42.05)         | -                   |
| Squamous cell carcinoma            | -                      | 99 (28.13)          |                     |
| Other non-small cell carcinoma     | -                      | 57 (16.19)          |                     |
| Small cell carcinoma               | -                      | 48 (13.64)          |                     |
| Pack-years of smoking <sup>c</sup> |                        |                     |                     |
| 0-30                               | 359 (88.21)            | 228 (64.77)         | 0.0001 <sup>a</sup> |
| >30                                | 48 (11.79)             | 124 (35.23)         |                     |
| Mean pack-year ± SD                | 8.07 ± 16.65           | 22.44 ±19.53        | 0.008 <sup>b</sup>  |
| Mean pack-year ± SD <sup>d</sup>   | 12.04 ± 20.81          | 31.63 ± 21.13       | 0.001 <sup>b</sup>  |

<sup>a</sup>χ<sup>2</sup>-Test or Fisher's exact test.<sup>b</sup>t-Test.<sup>c</sup> The median number of pack years of combined cases and controls were utilized as the cut-point.<sup>d</sup> Restricted to males only.

difference in gender and age between the controls and cases. As expected, cases smoked more cigarettes ( $P < 0.001$ ). The distribution of tumour types among the patients was as follows: adenocarcinoma, 42.05%; squamous cell carcinoma, 28.13%; other non-small cell carcinoma, 16.19%; and small cell carcinoma, 13.64%.

### Association of the alleles of the 16 SNPs with lung cancer risk

The basic information regarding the 16 SNPs at the TOX3/LOC643714 locus was demonstrated in Supplementary Table S3. The 16 SNPs were genotyped in all of the lung cancer patients and healthy controls and qualified according to Hardy-Weinberg equilibrium (HWE) in the study population (Supplementary Table S3). As shown in Table 2, Pearson's chi-square test or Fisher's exact test demonstrated that of the 16 SNPs, three

SNPs (rs9933638, rs12443621, and rs3104746) were found to be significantly associated with lung cancer risk ( $P < 0.001$ ,  $P < 0.001$ , and  $P = 0.005$ , respectively) and rs3095661 displayed a marginally significance ( $P = 0.041$ ) (Table 2). Among the four SNPs, rs9933638, representing a haplotype block covering 12 SNPs including rs12443621, is located at intron 2 of TOX3. Rs3095661 and rs3104746 are located at intron 4 of TOX3 and intron 2 of LOC643714, respectively (Supplementary Table S3).

The stratification by gender demonstrated that rs9933638, rs12443621, and rs3104746 were significantly associated with lung cancer risk of both male ( $P < 0.001$ ,  $P = 0.003$ , and  $P = 0.038$ , respectively) and female subjects ( $P = 0.007$ ,  $P = 0.005$ , and  $P = 0.049$ , respectively). Additionally, the analysis of the 16 SNPs among patients with adenocarcinoma and squamous cell carcinoma, the two most common types of NSCLC, respectively, revealed that rs9933638, rs12443621, and rs3104746 were

**Table 2: Distribution of the alleles of the 16 SNPs at the TOX3/LOC643714 locus between controls and cases in the study**

| SNPs       | Allele | Control<br>(n=407)<br>(%) | Lung<br>cancer<br>(n=352)<br>(%) | P<br>Value <sup>a</sup> | Adeno-<br>carcinoma<br>(n=148)<br>(%) | P<br>Value <sup>a</sup> | Squamous<br>cell<br>carcinoma<br>(n=99)(%) | P<br>Value <sup>a</sup> | Male                      |                                  |                         | Female                   |                                 |                         |
|------------|--------|---------------------------|----------------------------------|-------------------------|---------------------------------------|-------------------------|--|-------------------------|---------------------------|----------------------------------|-------------------------|--------------------------|---------------------------------|-------------------------|
|            |        |                           |                                  |                         |                                       |                         |  |                         | Control<br>(n=320)<br>(%) | Lung<br>cancer<br>(n=272)<br>(%) | P<br>Value <sup>a</sup> | Control<br>(n=87)<br>(%) | Lung<br>cancer<br>(n=80)<br>(%) | P<br>Value <sup>a</sup> |
| rs3095661  | G      | 728<br>(89.43)            | 605<br>(85.94)                   | 0.041                   | 247<br>(83.45)                        | 0.009                   | 182<br>(92.11)                             | 0.357                   | 573<br>(89.53)            | 469<br>(86.21)                   | 0.088                   | 155<br>(89.08)           | 136<br>(85.00)                  | 0.327                   |
|            | C      | 86<br>(10.57)             | 99<br>(14.06)                    |                         | 49<br>(16.55)                         |                         | 16<br>(7.89)                               |                         | 67<br>(10.47)             | 75<br>(13.79)                    |                         | 19<br>(10.92)            | 24<br>(15.00)                   |                         |
| rs10852413 | T      | 665<br>(81.70)            | 570<br>(81.20)                   | 0.842                   | 237<br>(80.07)                        | 0.543                   | 163<br>(82.32)                             | 0.918                   | 520<br>(81.25)            | 444<br>(81.92)                   | 0.821                   | 145<br>(83.33)           | 126<br>(78.75)                  | 0.328                   |
|            | C      | 149<br>(18.30)            | 132<br>(18.80)                   |                         | 59<br>(19.92)                         |                         | 35<br>(17.68)                              |                         | 120<br>(18.75)            | 98<br>(18.08)                    |                         | 29<br>(16.67)            | 34<br>(21.25)                   |                         |
| rs16951204 | G      | 536<br>(65.85)            | 485<br>(68.89)                   | 0.228                   | 186<br>(62.84)                        | 0.356                   | 131<br>(66.16)                             | 1.000                   | 418<br>(65.31)            | 380<br>(69.85)                   | 0.106                   | 118<br>(67.82)           | 105<br>(65.63)                  | 0.728                   |
|            | C      | 278<br>(34.15)            | 219<br>(31.11)                   |                         | 110<br>(37.16)                        |                         | 67<br>(33.84)                              |                         | 222<br>(34.69)            | 164<br>(30.15)                   |                         | 56<br>(32.18)            | 55<br>(34.38)                   |                         |
| rs4784219  | A      | 595<br>(73.10)            | 523<br>(74.50)                   | 0.559                   | 229<br>(77.37)                        | 0.163                   | 149<br>(75.25)                             | 0.590                   | 469<br>(73.28)            | 402<br>(74.17)                   | 0.741                   | 126<br>(72.41)           | 121<br>(75.63)                  | 0.534                   |
|            | T      | 219<br>(26.90)            | 179<br>(25.50)                   |                         | 67<br>(22.63)                         |                         | 49<br>(24.75)                              |                         | 171<br>(26.72)            | 140<br>(25.83)                   |                         | 48<br>(27.59)            | 39<br>(24.38)                   |                         |
| rs9302555  | T      | 630<br>(77.40)            | 557<br>(79.12)                   | 0.455                   | 241<br>(81.42)                        | 0.161                   | 159<br>(80.26)                             | 0.391                   | 500<br>(78.13)            | 445<br>(81.80)                   | 0.652                   | 130<br>(74.71)           | 112<br>(70.00)                  | 0.391                   |
|            | G      | 184<br>(22.60)            | 147<br>(20.88)                   |                         | 55<br>(18.58)                         |                         | 39<br>(19.74)                              |                         | 140<br>(21.88)            | 99<br>(18.20)                    |                         | 44<br>(25.29)            | 48<br>(30.00)                   |                         |
| rs8051542  | C      | 657<br>(83.38)            | 577<br>(81.96)                   | 0.493                   | 247<br>(83.33)                        | 1.000                   | 177<br>(89.47)                             | 0.803                   | 519<br>(83.57)            | 442<br>(81.25)                   | 0.316                   | 138<br>(82.63)           | 135<br>(84.38)                  | 0.766                   |
|            | T      | 131<br>(16.62)            | 127<br>(18.04)                   |                         | 49<br>(16.67)                         |                         | 21<br>(10.53)                              |                         | 102<br>(16.43)            | 102<br>(18.75)                   |                         | 29<br>(17.37)            | 25<br>(15.62)                   |                         |
| rs9933638  | G      | 445<br>(54.80)            | 465<br>(66.05)                   | 0.000                   | 200<br>(67.57)                        | 0.0001                  | 130<br>(65.79)                             | 0.006                   | 352<br>(55.17)            | 356<br>(65.44)                   | 0.000                   | 93<br>(53.45)            | 109<br>(68.13)                  | 0.007                   |
|            | A      | 367<br>(45.20)            | 239<br>(33.95)                   |                         | 96<br>(32.43)                         |                         | 68<br>(34.21)                              |                         | 286<br>(44.83)            | 188<br>(34.563)                  |                         | 81<br>(46.55)            | 51<br>(31.88)                   |                         |
| rs12443621 | G      | 450<br>(55.28)            | 459<br>(65.20)                   | 0.000                   | 197<br>(66.67)                        | 0.0007                  | 128<br>(64.47)                             | 0.020                   | 358<br>(55.94)            | 350<br>(64.34)                   | 0.003                   | 92<br>(52.87)            | 109<br>(68.13)                  | 0.005                   |
|            | A      | 364<br>(44.72)            | 245<br>(34.80)                   |                         | 99<br>(33.33)                         |                         | 70<br>(35.53)                              |                         | 282<br>(44.06)            | 194<br>(35.66)                   |                         | 82<br>(47.13)            | 51<br>(31.88)                   |                         |
| rs3095604  | C      | 556<br>(68.30)            | 470<br>(66.76)                   | 0.545                   | 200<br>(67.57)                        | 0.827                   | 128<br>(64.42)                             | 0.352                   | 430<br>(67.19)            | 368<br>(67.65)                   | 0.901                   | 126<br>(72.41)           | 102<br>(63.75)                  | 0.100                   |
|            | G      | 258<br>(31.70)            | 234<br>(33.24)                   |                         | 96<br>(32.43)                         |                         | 70<br>(35.53)                              |                         | 210<br>(32.81)            | 176<br>(32.35)                   |                         | 48<br>(27.59)            | 58<br>(36.25)                   |                         |
| rs1362550  | G      | 652<br>(80.10)            | 576<br>(81.82)                   | 0.432                   | 244<br>(82.43)                        | 0.439                   | 151<br>(76.32)                             | 0.241                   | 506<br>(79.06)            | 440<br>(83.91)                   | 0.467                   | 146<br>(83.91)           | 136<br>(85.00)                  | 0.880                   |
|            | C      | 162<br>(19.90)            | 128<br>(18.18)                   |                         | 52(17.57)                             |                         | 47(23.68)                                  |                         | 134<br>(20.94)            | 104<br>(16.09)                   |                         | 28<br>(16.09)            | 24<br>(15.00)                   |                         |
| rs28463809 | T      | 558<br>(68.55)            | 466<br>(66.19)                   | 0.351                   | 194<br>(65.54)                        | 0.346                   | 128<br>(64.47)                             | 0.309                   | 439<br>(68.59)            | 368<br>(67.65)                   | 0.754                   | 119<br>(68.39)           | 98<br>(61.25)                   | 0.201                   |
|            | G      | 256<br>(31.45)            | 238<br>(33.81)                   |                         | 102<br>(34.46)                        |                         | 70<br>(35.53)                              |                         | 201<br>(31.41)            | 176<br>(32.35)                   |                         | 55<br>(31.61)            | 62<br>(38.75)                   |                         |
| rs4784226  | C      | 616<br>(75.68)            | 527<br>(74.86)                   | 0.721                   | 220<br>(74.32)                        | 0.638                   | 154<br>(75.63)                             | 0.578                   | 488<br>(76.25)            | 406<br>(74.63)                   | 0.542                   | 128<br>(73.56)           | 121<br>(75.63)                  | 0.707                   |

(Continued)

| SNPs      | Allele | Control<br>(n=407)<br>(%) | Lung<br>cancer<br>(n=352)<br>(%) | P<br>Value <sup>a</sup> | Adeno-<br>carcinoma<br>(n=148)<br>(%) | P<br>Value <sup>a</sup> | Squamous<br>cell<br>carcinoma<br>(n=99)(%) | P<br>Value <sup>a</sup> | Male                      |                                  |                         | Female                   |                                 |                         |
|-----------|--------|---------------------------|----------------------------------|-------------------------|---------------------------------------|-------------------------|--|-------------------------|---------------------------|----------------------------------|-------------------------|--------------------------|---------------------------------|-------------------------|
|           |        |                           |                                  |                         |                                       |                         |  |                         | Control<br>(n=320)<br>(%) | Lung<br>cancer<br>(n=272)<br>(%) | P<br>Value <sup>a</sup> | Control<br>(n=87)<br>(%) | Lung<br>cancer<br>(n=80)<br>(%) | P<br>Value <sup>a</sup> |
| rs3803662 | T      | 198<br>(24.32)            | 177<br>(25.14)                   | 0.869                   | 76<br>(25.68)                         | 0.829                   | 44<br>(22.37)                              | 0.449                   | 152<br>(23.75)            | 138<br>(25.37)                   | 0.619                   | 46<br>(26.44)            | 39<br>(24.38)                   | 0.161                   |
|           | A      | 550<br>(67.57)            | 472<br>(67.05)                   |                         | 198<br>(66.89)                        |                         | 128<br>(64.47)                             |                         | 426<br>(67.30)            | 370<br>(68.01)                   |                         | 124<br>(71.26)           | 102<br>(63.75)                  |                         |
|           | G      | 264<br>(32.43)            | 232<br>(32.95)                   |                         | 98<br>(31.11)                         |                         | 70<br>(35.53)                              |                         | 214<br>(32.70)            | 174<br>(31.99)                   |                         | 50<br>(28.74)            | 58<br>(36.25)                   |                         |
| rs4784227 | C      | 616<br>(75.68)            | 530<br>(75.28)                   | 0.905                   | 220<br>(74.44)                        | 0.638                   | 156<br>(78.95)                             | 0.402                   | 488<br>(76.25)            | 408<br>(75.00)                   | 0.635                   | 128<br>(73.56)           | 122<br>(76.25)                  | 0.615                   |
|           | T      | 198<br>(24.32)            | 174<br>(24.72)                   |                         | 76<br>(25.56)                         |                         | 42<br>(21.05)                              |                         | 152<br>(23.75)            | 136<br>(25.00)                   |                         | 46<br>(26.44)            | 38<br>(23.75)                   |                         |
| rs3104746 | T      | 665<br>(81.70)            | 613<br>(87.07)                   | 0.005                   | 257<br>(86.67)                        | 0.047                   | 177<br>(89.39)                             | 0.008                   | 525<br>(82.03)            | 471<br>(86.58)                   | 0.038                   | 140<br>(80.46)           | 142<br>(88.75)                  | 0.049                   |
|           | A      | 149<br>(18.30)            | 91<br>(12.93)                    |                         | 39<br>(13.33)                         |                         | 21<br>(10.61)                              |                         | 115<br>(17.97)            | 73<br>(13.42)                    |                         | 34<br>(19.54)            | 18<br>(11.25)                   |                         |
| rs3112562 | C      | 562<br>(69.21)            | 506<br>(71.88)                   | 0.260                   | 210<br>(71.11)                        | 0.606                   | 149<br>(75.25)                             | 0.100                   | 436<br>(68.13)            | 398<br>(73.16)                   | 0.064                   | 126<br>(73.26)           | 108<br>(67.50)                  | 0.279                   |
|           | G      | 250<br>(30.79)            | 198<br>(28.13)                   |                         | 86<br>(28.89)                         |                         | 49<br>(24.75)                              |                         | 204<br>(31.88)            | 146<br>(26.84)                   |                         | 46<br>(26.74)            | 52<br>(32.50)                   |                         |

<sup>a</sup>  $\chi^2$ -Test or Fisher's exact test.

significantly associated with risk of both adenocarcinoma ( $P < 0.001$ ,  $P < 0.001$ , and  $P = 0.047$ , respectively) and squamous cell carcinoma ( $P = 0.006$ ,  $P = 0.020$ , and  $P = 0.008$ , respectively), while rs3095661 was found to be only related with risk of adenocarcinoma ( $P = 0.009$ ) (Table 2).

#### Association of the genotypes of the four SNPs (rs9933638, rs12443621, rs3104746, and rs3095661) with lung cancer risk

As shown in Table 3, consistent with the association between the alleles of the three SNPs (rs9933638, rs12443621, and rs3104746) and lung cancer risk, there was a significantly different distribution of the genotypes of the three SNPs between lung cancer cases and controls ( $P < 0.001$ ,  $P < 0.001$ , and  $P = 0.003$ , respectively). The multivariate logistic regression analyses with adjustment of age, gender, and smoking revealed that individuals with rs9933638GG had an elevated risk of lung cancer compared with rs9933638GA and rs9933638AA carriers (codominant model, OR = 2.571, 95%CI = 1.710-3.867,  $P < 0.001$ , and OR = 1.509, 95%CI = 1.022-2.229,  $P = 0.038$ , respectively). The dichotomic analysis further demonstrated that individuals with rs9933638GG showed an increased risk of lung cancer compared with rs9933638(GA+AA) carriers (dominant model, OR = 1.877, 95%CI = 1.423-2.476,  $P < 0.001$ ). Additionally, individuals with rs9933638(GG+GA) also displayed an

increased risk of lung cancer compared with rs9933638AA carriers (recessive model, OR = 1.893, 95%CI = 1.308-2.741,  $P = 0.001$ ), suggesting that individuals with the allele G of rs9933638 were susceptible to lung cancer in a dose-dependent manner. Similar findings were discovered for rs12443621. For rs3104746, the dichotomic analysis demonstrated that compared with individuals with rs3104746(TA+AA), rs3104746TT carriers had an increased risk of lung cancer (dominant model, OR = 1.665, 95%CI = 1.243-2.230,  $P = 0.001$ ), suggesting that individuals with rs3104746TT were susceptible to lung cancer compared with rs3104746(TA+AA) carriers. Moreover, though a marginally significant difference of rs3095661 alleles was found between controls and cases ( $P = 0.041$ , Table 2), no significant difference of rs3095661 genotypes was found in the four models between controls and cases ( $P = 0.362$ ,  $P = 0.066$ ,  $P = 0.999$ ,  $P = 0.197$ , respectively, Table 3). Notably, the detection of rs3095661CC carriers only in lung cancer patients (6/352, 1.70%) but not in controls strongly suggested that rs3095661CC might be risk factor of lung cancer.

#### Distribution of rs9933638, rs12443621, and rs3104746 among lung cancer patients stratified by cigarette smoking

Because rs9933638, rs12443621, and rs3104746 were found to be associated with lung cancer risk, the



**Table 3: Association of the genotypes of the four SNPs (rs9933638, rs12443621, rs3104746, and rs3095661) at the TOX3/LOC643714 locus with lung cancer risk**

| SNPs       | Model        | Genotype | Control (n = 407) | Case (n = 352) | P-value <sup>a</sup> | Adjusted P-value <sup>b</sup> | OR (95% CI)        |
|------------|--------------|----------|-------------------|----------------|----------------------|-------------------------------|--------------------|
| rs3095661  | Codominant   | G/G      | 321(78.87)        | 259(73.58)     | 0.194                | 0.362                         | Ref.               |
|            |              | G/C      | 86(21.13)         | 87(24.72)      | 0.008                | 0.999                         | 0.000(0.000-.)     |
|            |              | C/C      | 0(0.00)           | 6(1.70)        | 0.029                | 0.999                         | 0.000(0.000-.)     |
|            | Dominant     | G/G      | 321(78.87)        | 259(73.58)     | 0.103                | 0.066                         | 0.749(0.550-1.020) |
|            |              | G/C-C/C  | 86(21.13)         | 93(26.42)      |                      |                               |                    |
|            | Recessive    | G/G-G/C  | 407(100.00)       | 346(98.30)     | 0.010                | 0.999                         | 0.000(0.000-.)     |
|            |              | C/C      | 0(0.00)           | 6(1.70)        |                      |                               |                    |
|            | Overdominant | G/G-C/C  | 321(78.87)        | 265(75.28)     | 0.260                | 0.197                         | 0.815(0.597-1.112) |
|            |              | G/C      | 86(21.13)         | 87(24.72)      |                      |                               |                    |
| rs9933638  | Codominant   | G/G      | 117(28.82)        | 152(43.18)     | 0.001                | 0.000                         | Ref.               |
|            |              | G/A      | 211(51.97)        | 161(45.74)     | 0.000                | 0.000                         | 2.571(1.710-3.867) |
|            |              | A/A      | 78(19.21)         | 39(11.08)      | 0.067                | 0.038                         | 1.509(1.022-2.229) |
|            | Dominant     | G/G      | 117(28.82)        | 152(43.18)     | 0.000                | 0.000                         | 1.877(1.423-2.476) |
|            |              | G/A-A/A  | 289(71.18)        | 200(56.82)     |                      |                               |                    |
|            | Recessive    | G/G-G/A  | 328(80.79)        | 313(88.92)     | 0.002                | 0.001                         | 1.893(1.308-2.741) |
|            |              | A/A      | 78(19.21)         | 39(11.08)      |                      |                               |                    |
|            | Overdominant | G/G-A/A  | 195(48.03)        | 191(54.26)     | 0.094                | 0.059                         | 1.287(0.991-1.673) |
|            |              | G/A      | 211(51.97)        | 161(45.74)     |                      |                               |                    |
| rs12443621 | Codominant   | G/G      | 130(31.86)        | 153(43.47)     | 0.020                | 0.000                         | Ref.               |
|            |              | G/A      | 191(46.81)        | 153(43.47)     | 0.000                | 0.000                         | 2.250(1.512-3.348) |
|            |              | A/A      | 87(21.32)         | 46(13.07)      | 0.062                | 0.029                         | 1.534(1.045-2.253) |
|            | Dominant     | G/G      | 130(31.86)        | 153(43.47)     | 0.001                | 0.000                         | 1.646(1.244-2.178) |
|            |              | G/A-A/A  | 278(68.14)        | 199(56.53)     |                      |                               |                    |
|            | Recessive    | G/G-G/A  | 321(78.68)        | 306(86.93)     | 0.003                | 0.001                         | 1.824(1.272-2.616) |
|            |              | A/A      | 87(21.32)         | 46(13.07)      |                      |                               |                    |
|            | Overdominant | G/G-A/A  | 217(53.19)        | 199(56.53)     | 0.381                | 0.338                         | 1.141(0.871-1.493) |
|            |              | G/A      | 191(46.81)        | 153(43.47)     |                      |                               |                    |
| rs3104746  | Codominant   | T/T      | 266(65.36)        | 267(75.85)     | 0.002                | 0.003                         | Ref.               |
|            |              | T/A      | 133(32.68)        | 79(22.44)      | 0.788                | 0.448                         | 1.469(0.544-3.964) |
|            |              | A/A      | 8(1.97)           | 6(1.70)        | 0.778                | 0.796                         | 0.875(0.318-2.408) |
|            | Dominant     | T/T      | 266(65.36)        | 267(75.85)     | 0.002                | 0.001                         | 1.665(1.243-2.230) |
|            |              | T/A-A/A  | 141(34.64)        | 85(24.15)      |                      |                               |                    |
|            | Recessive    | T/T-T/A  | 399(98.03)        | 346(98.30)     | 1.000                | 0.635                         | 1.271(0.473-3.417) |
|            |              | A/A      | 8(1.97)           | 6(1.70)        |                      |                               |                    |
|            | Overdominant | T/T-A/A  | 274(67.32)        | 273(77.56)     | 0.002                | 0.001                         | 1.662(1.234-2.240) |
|            |              | T/A      | 133(32.68)        | 79(22.44)      |                      |                               |                    |

<sup>a</sup>  $\chi^2$ -Test or Fisher's exact test.

<sup>b</sup> Adjusted by age, gender, and cigarette smoking.

distribution of the three SNPs was analysed among lung cancer patients stratified by median number of pack-years of cigarette smoking to investigate whether the gene-environment interaction exists. The linear-by-linear association test revealed that except rs3104746 ( $P = 0.638$ ), the risk allele G of both rs9933638 and rs12443621 showed an increasing trend from light smoking to heavy smoking groups ( $P = 0.004$  and  $P < 0.001$ , respectively,  $p$  trend  $< 0.001$ , Table 4). Multivariate logistic regression analyses adjusted for age and gender revealed that there was a significantly different distribution of genotypes of both rs9933638 and rs12443621 between light smoking patients and heavy smoking patients. Compared with patients with rs9933638GA and 9933638AA, patients with rs9933638GG were enriched and demonstrated an increased risk of lung cancer in heavy smoking subjects (codominant model, OR = 1.714, 95%CI = 1.155-2.542,  $P = 0.007$ ; OR = 2.811, 95%CI = 1.405-5.625,  $P = 0.003$ , respectively, Table 4). In the dominant model, patients with rs9933638GG carriers were enriched and showed an elevated risk of lung cancer in heavy-smoking patients compared with rs9933638(GA+AA) carriers (OR = 1.878, 95%CI = 1.290-2.736,  $P = 0.001$ ). In the recessive model, lung cancer risk of individuals with rs9933638(GG+GA) in heavy-smoking patients was 2.87 times higher than that of rs9933638AA carriers (OR = 2.868, 95%CI = 1.573-5.227,  $P = 0.001$ ). Similarly, patients with rs12443621GG were also enriched in heavy smoking patients and demonstrated an increase risk of lung cancer compared with rs12443621GA and rs12443621AA carriers (codominant model, OR = 1.821, 95%CI = 1.222-2.713,  $P = 0.003$ , and OR = 3.067, 95%CI = 1.541-6.104,  $P = 0.001$ , respectively, Table 4). In the dominant model, individuals with rs12443621GG carriers were enriched in heavy-smoking patients and showed an elevated risk of lung cancer compared with rs12443621(GA+AA) carriers (OR = 2.018, 95%CI = 1.381-2.947,  $P < 0.001$ ). In the recessive model, lung cancer risk of individuals with rs12443621(GG+GA) in heavy-smoking patients was 2.87 times higher than that of rs12443621AA carriers (OR = 2.868, 95%CI = 1.599-5.143,  $P < 0.001$ ).

## DISCUSSION

Lung cancer is one of the major causes of cancer-related death worldwide. Besides environmental factors, inherited genetic variants or polymorphisms are also involved in lung cancer risk. In the case-control study, the genotyping of the 16 SNPs at the TOX3/LOC643714 locus in a Han Chinese population revealed that rs9933638/rs12443621 and rs3104746 might contribute to risk of lung cancer. To our best knowledge, the study discovered for the first time that the previously identified breast cancer susceptibility-associated SNPs at the TOX3/LOC643714 locus were risk factors of lung cancer.

The TOX3/LOC643714 locus was one of the first breast cancer regions identified through GWAS in populations of European and East Asian origin [10]. In the recent years, quite a few of SNPs at the TOX3/LOC643714 locus including rs3803662, rs3104746, rs8051542, rs4784227, rs12443621, rs3112612, rs3112562, rs3104793, rs8046994, rs3104788, and rs3104767, were demonstrated to be independently associated with elevated risk of breast cancer and the other human diseases [16–43]. Of the SNPs mentioned above, only rs12443621 and rs3104746 were found to be significantly associated with increased risk of lung cancer in the study. Initially, Pearson's chi-square test or Fisher's exact test revealed that rs9933638 were strongly associated with lung cancer risk (Table 2), confirmed by logistic regression analyses in which lung cancer risk of individuals with rs9933638GG was shown to be 2.57 and 1.51 times higher than that of individuals with rs9933638GA and rs9933638AA, respectively. In addition, lung cancer risk of rs9933638(GG+GA) carriers was demonstrated to be 1.89 times higher than that of rs9933638AA carriers (Table 3). The findings suggested that individuals with the allele G of rs9933638 were susceptible to lung cancer in a dose-dependent manner. To validate the findings, rs12443621, the previously identified breast cancer risk factor of Chinese population [25, 28, 37] and covered by the haplotype block represented by rs9933638, was genotyped in the study. The genotyping data showed that the allele G of rs12443621 was also a risk factor for lung cancer (Table 2), verified by the stratification analysis which revealed that subjects with rs12443621GG and rs12443621GA had an increased risk of lung cancer compared with individuals with rs12443621AA (Table 3). The validation results strongly confirmed that rs9933638, representing a haplotype block covering rs12443621, was associated with risk of lung cancer. Most importantly, rs12443621, the previously determined breast cancer risk factor of Chinese population [25, 28, 37], was discovered to be a lung cancer risk factor of Chinese population in the study provided new evidences that lung cancer and breast cancer are linked at a molecular and genetic level at least in part in Chinese population, which may help to explore the novel carcinogenesis mechanisms.

Rs3803662 was reported to be associated with breast cancer risk of both male and female subjects [13, 16–20]. Consistent with rs3803662, the stratification by gender revealed that all of the three SNPs (rs9933638, rs12443621, and rs3104746) were all associated with increased lung cancer risk of both male and female individuals. Furthermore, the three SNPs were all indicated to be related with elevated lung cancer risk of patients with adenocarcinoma and squamous cell carcinoma. The findings suggested that the association of the three SNPs with lung cancer risk was gender- and histology-independent. Moreover, the analysis of the three

**Table 4: Distribution of genotypes of rs9933638, rs12443621, and rs3104746 among lung cancer patients stratified by cigarette smoking**

| SNPs       | Model        | Genotype   | light-smokings (n = 228) (%) <sup>a</sup> | Heavy-smokings (n = 124)(%) <sup>a</sup> | P-value <sup>b</sup> | Adjusted P-value <sup>c</sup> | OR (95% CI)         |
|------------|--------------|------------|---|--|----------------------|-------------------------------|---------------------|
| rs9933638  | Allele       | G          | 282(61.84)                                | 181(72.98)                               | 0.004                |                               |                     |
|            |              | A          | 174(38.16)                                | 67(27.02)                                |                      |                               |                     |
|            | Codominant   | G/G        | 86(37.72)                                 | 66(53.23)                                | 0.026                | 0.002                         | Ref.                |
|            |              | G/A        | 110(48.25)                                | 49(39.52)                                | 0.012                | 0.007                         | 1.714(1.155-2.542)  |
|            |              | A/A        | 32(14.04)                                 | 9(7.26)                                  | 0.336                | 0.003                         | 2.811(1.405-5.625)  |
|            |              | G/G        | 86(37.72)                                 | 66(53.23)                                | 0.007                | 0.001                         | 1.878(1.290-2.736)  |
|            | Recessive    | G/A-A/A    | 142(62.28)                                | 58(46.77)                                | 0.081                | 0.001                         | 2.868(1.573-5.227)  |
|            |              | G/G-G/A    | 196(85.96)                                | 115(92.74)                               |                      |                               |                     |
|            |              | A/A        | 32(14.04)                                 | 9(7.26)                                  |                      |                               |                     |
|            | Overdominant | G/G-A/A    | 118(51.75)                                | 75(60.48)                                | 0.119                | 0.212                         | 1.256(0.878-1.796)  |
| G/A        |              | 110(48.25) | 49(39.52)                                 |  |                      |                               |                     |
| rs12443621 | Allele       | G          | 282(61.30)                                | 180(73.77)                               | 0.000                |                               |                     |
|            |              | A          | 178(38.70)                                | 64(26.23)                                |                      |                               |                     |
|            | Codominant   | G/G        | 85(37.28)                                 | 66(54.55)                                | 0.013                | 0.001                         | Ref.                |
|            |              | G/A        | 109(47.81)                                | 46(38.02)                                | 0.008                | 0.003                         | 1.821(1.222-2.713)  |
|            |              | A/A        | 34(14.91)                                 | 9(7.44)                                  | 0.259                | 0.001                         | 3.067(1.541-6.104)  |
|            |              | G/G        | 85(37.28)                                 | 66(54.55)                                | 0.002                | 0.000                         | 2.018(1.381-2.947)  |
|            | Recessive    | G/A-A/A    | 143(62.72)                                | 55(45.45)                                | 0.059                | 0.000                         | 2.868(1.599-5.143)  |
|            |              | G/G-G/A    | 194(85.09)                                | 112(92.56)                               |                      |                               |                     |
|            |              | A/A        | 34(14.91)                                 | 9(7.44)                                  |                      |                               |                     |
|            | Overdominant | G/G-A/A    | 119(52.19)                                | 75(61.98)                                | 0.090                | 0.176                         | 1.281(0.895-1.832)  |
| G/A        |              | 109(47.81) | 46(38.02)                                 |  |                      |                               |                     |
| rs3104746  | Allele       | T          | 404(66.01)                                | 58(63.04)                                | 0.638                |                               |                     |
|            |              | A          | 208(33.99)                                | 34(36.96)                                |                      |                               |                     |
|            | Codominant   | T/T        | 178(78.07)                                | 87(71.90)                                | 0.111                | 0.140                         | Ref.                |
|            |              | T/A        | 45(19.74)                                 | 34(28.10)                                | 0.181                | 0.147                         | 0.645(0.418-0.995)  |
|            |              | A/A        | 5(2.19)                                   | 0(0.00)                                  | 1.000                | 0.999                         | 7.817E8(0.000-.)    |
|            |              | T/T        | 178(78.07)                                | 87(71.90)                                | 0.236                | 0.140                         | 0.726(0.474-1.111)  |
|            | Recessive    | T/A-A/A    | 50(21.93)                                 | 34(28.10)                                | 0.168                | 0.056                         | 7.802(0.953-63.898) |
|            |              | T/T-T/A    | 223(97.81)                                | 121(100.00)                              |                      |                               |                     |
|            |              | A/A        | 5(2.19)                                   | 0(0.00)                                  |                      |                               |                     |
|            | Overdominant | T/T-A/A    | 183(80.26)                                | 87(71.90)                                | 0.082                | 0.098                         | 0.697(0.455-1.069)  |
| T/A        |              | 45(19.74)  | 34(28.10)                                 |  |                      |                               |                     |

<sup>a</sup> Lung cancer patients stratified by median number of pack-years of smoking (light smoking, Pack-years of smoking were 0-30; heavy smoking groups, Pack-years of smoking were more than 30).

<sup>b</sup>  $\chi^2$ -Test or Fisher's exact test.

<sup>c</sup> Adjusted by age and gender.



SNPs among lung cancer patients stratified by cigarette smoking revealed that except rs3104746, the risk allele G of both rs9933638 and rs12443621 was enriched in lung cancer patients with heavy cigarette smoking and lung cancer risk of individuals with rs9933638/rs12443621(GG+GA) was 2.87 times higher than that of rs9933638/rs12443621AA carriers among heavy smoking patients (Table 4), demonstrating that individuals carrying risk genotypes and with heavy cigarette smoking may have a higher risk of lung cancer. The findings strongly confirmed that the gene-environment interaction exists in the development and process of lung cancer.

There are several limitations in the study. First, due to the fact that most of lung cancer patients recruited were characterized to be poorly differentiated (316/352, 89.77%) and most of patients with NSCLC were identified at advanced stages (276/304, 90.79%), we tried but failed to clarify the association of the 16 SNPs with histological grade and stage of patients with lung cancer. Second, the subjects of the study were recruited only from Southwestern China, large-scale studies are required to clarify the association of the SNPs at the TOX3/LOC643714 locus with lung cancer risk in the other Han Chinese populations.

In summary, the present study discovered for the first time that rs9933638/rs12443621 and rs3104746, the previously identified breast cancer susceptibility-related SNPs at the TOX3/LOC643714 locus, contributed to the individual's risk to lung cancer in the Southwestern Han Chinese population. The findings provided additional evidences that lung cancer and breast cancer are correlated at a molecular and genetic level at least in part in Chinese population.

## MATERIALS AND METHODS

### Study population

Patients (n = 352) with primary lung cancer diagnosed from September 2007 to December 2008 were recruited from the Institute of Human Respiratory Disease of Xinqiao Hospital, the Third Military Medical University. All patients were newly diagnosed, histologically confirmed and previously untreated. 407 age- and sex frequency-matched healthy control samples were collected from individuals at the Centre of Physical Examination of Xinqiao Hospital between November 2007 and December 2008. The exclusion criterion for the control group was any history of cancer. All of the subjects were unrelated at least within three generations. After explaining the purpose and procedures of the study, all participants signed a written informed consent form, completed a detailed questionnaire regarding their smoking habits, and donated 5 ml peripheral blood. Blood samples were drawn into Na-EDTA tubes from all subjects and stored at -70°C for genomic DNA extraction. The

study was approved by the Ethical Committee of Xinqiao Hospital, the Third Military Medical University.

### Selection of SNPs

Totally, 16 SNPs at the TOX3/LOC643714 locus were selected in the study. Of the 16 SNPs, six SNPs (rs8051542, rs12443621, rs3803662, rs3104746, rs3112562, and rs4784227) were selected based on the published references in which these SNPs were suggested to be susceptible to breast cancer or the other human diseases and the other 10 SNPs were selected from the genetic variation data for TOX3 gene obtained from the HapMap project for 45 healthy Chinese Han Beijing (CHB) adults ([www.hapmap.org](http://www.hapmap.org)). Haplotype blocks, representing regions inherited without substantial recombination in the ancestors of the current population, were constructed throughout the entire TOX3 gene using Haploview (version 4.0, Broad Institute of MIT and Harvard, Cambridge, MA) [45]. The history of recombination between a pair of SNPs can be estimated with the use of the normalized measure of allelic association  $D'$  (value of  $D$  prime between the two loci) [46, 47]. The criterion for the selected SNPs to construct a haplotype block is that all SNPs in one region must be in strong LD with  $D' > 0.98$  for the upper 95% confidence bound and  $> 0.7$  for the lower bound. A maximally informative htSNP was then selected from each block using the software Tagger program (<http://www.broad.mit.edu/mpg/haploview>). This algorithm selects a subset of variants that capture all known common genetic variations in the TOX3 gene based on a LD threshold of  $r^2 \geq 0.8$ . The inverse of  $r^2$  represents the ratio of sample size needed to detect an indirect association with an un-analyzed SNP to direct association at the same power.

### Genotyping analysis

Genomic DNA was extracted from whole blood using the QIAamp DNA Blood Mini Kit according to the manufacturer's instructions (QIAGEN, Maryland, USA). The SNP genotyping was performed using an improved multiplex ligation detection reaction (iMLDR) technique (Genesky Biotechnologies Inc., Shanghai, China). In brief, the selected SNP loci were genotyped in one ligation reaction. Two multiplex PCR reactions were designed to amplify fragments covering all SNP loci. The primer information of the two reaction mixtures is described in Supplementary Table S1 and S2, respectively. The PCR for both reactions was 95°C, 2 min; 11 cycles (94°C, 20s; 65°C -0.5°C/cycle, 40s; 72°C, 1 min 30s); 24 cycles (94°C, 20s; 59°C, 30s; 72°C, 1 min 30s); 72°C, 2 min; hold at 4°C. The ligation cycling programme was 95°C, 2 min; 38 cycles (94°C, 1 min; 56°C, 4 min); hold at 4°C. Half a microlitre of ligation product was loaded

into the ABI 3730XL and the raw data were analysed by GeneMapper 4.1.

## Data analyses

Cigarette smoking was stratified by the median number of pack-years of combined cases and controls (1 pack-year = 20 cigarettes per day for 1 year). Cases and controls were compared by Student's *t*-test for continuous variables and Pearson's chi-square test or Fisher's exact test for categorical variables. The Hardy-Weinberg equilibrium of each SNP was tested by SNPStats (<http://bioinfo.iconcologia.net/snpstats/start.htm>). Each component of the model was: codominant model (major allele homozygotes vs. heterozygotes vs. minor allele homozygotes), dominant model (major allele homozygotes vs. heterozygotes + minor allele homozygotes), recessive model (major allele homozygotes + heterozygotes vs. minor allele homozygotes), and overdominant model (major allele homozygotes + minor allele homozygotes vs. heterozygotes). To assess the independent effect of each SNP, the multivariate logistic regression analyses with adjustments for possible confounding factors (age, gender, and smoking habits) were performed to estimate the association between the SNPs and cancer risk as well as the possible gene-environment interactions. All associations were presented as odds ratios (ORs) with the corresponding 95% confidence intervals (95%CI). All statistical analyses were performed using the Statistical Package for Social Science 15 for Windows (SPSS Inc, Chicago, IL, USA). In the statistical analysis, all statistical tests were two-sided and  $P < 0.05$  was considered significant.

## ACKNOWLEDGMENTS

The authors thank the patients and healthy controls who participated in this study for their willingness to contribute to the advancement of science.

## CONFLICTS OF INTEREST

None declared.

## FUNDING

The study was supported by National Natural Science Foundation of China (Nos. 81170044 and 81372499) and Natural Science Foundation of the Third Military Medical University (No. 2010XLC29 and 2012D289) awarded to Dr. Fuyun Ji.

## Author Contributors

The author contributions were as follows: funding, FYJ; concept and design, FYJ, GSQ, RLG, and JCX;

performing the experiments, CWJ and PQ; data collection and analysis, CWJ, SLY, PQ, RLG, CZW, RJZ, GMW, WY, and FYJ; sample collection and storage, ZA, QL, JL, and YC; contributed reagents/materials/analysis tools, FYJ; writing of the paper, FYJ.

## REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin.* 2013; 63:11-30. doi:10.3322/caac.21166.
2. Schabath MB, Spitz MR, Hong WK, Delclos GL, Reynolds WF, Gunn GB, Whitehead LW, Wu X. A myeloperoxidase polymorphism associated with reduced risk of lung cancer. *Lung Cancer.* 2002; 37:35-40. doi:10.1016/S0169-5002(02)00034-X.
3. Kiyohara C, Yoshimasu K, Shirakawa T, Hopkin JM. Genetic polymorphisms and environmental risk of lung cancer: a review. *Rev Environ Health.* 2004; 19:15-38. doi:10.1515/REVEH.2004.19.1.15.
4. Cote ML, Chen W, Smith DW, Benhamou S, Bouchardy C, Butkiewicz D, Fong KM, Gené M, Hirvonen A, Kiyohara C, Larsen JE, Lin P, Raaschou-Nielsen O, Povey AC, et al. Meta- and pooled analysis of GSTP1 polymorphism and lung cancer: a HuGE-GSEC review. *Am J Epidemiol.* 2009; 169:802-14. doi:10.1093/aje/kwn417.
5. De Ruyck K, Szaumkessel M, De Rudder I, Dehoorne A, Vral A, Claes K, Velghe A, Van Meerbeeck J, Thierens H. Polymorphisms in base-excision repair and nucleotide-excision repair genes in relation to lung cancer risk. *Mutat Res.* 2007; 631: 101-10. doi:10.1016/j.mrgentox.2007.03.010.
6. Wang Y, Yang H, Li L, Wang H, Zhang C, Yin G, Zhu B. Association between CYP2E1 genetic polymorphisms and lung cancer risk: a meta-analysis. *Eur J Cancer.* 2010; 46:758-64. doi:10.1016/j.ejca.2009.12.010.
7. Chen Z, Li Z, Niu X, Ye X, Yu Y, Lu S, Chen Z. The effect of CYP1A1 polymorphisms on the risk of lung cancer: a global meta-analysis based on 71 case-control studies. *Mutagenesis.* 2011; 26:437-46. doi:10.1093/mutage/ger002.
8. Rafnar T, Sulem P, Besenbacher S, Gudbjartsson DF, Zanon C, Gudmundsson J, Stacey SN, Kostic JP, Thorgeirsson TE, Thorleifsson G, Bjarnason H, Skuladottir H, Gudbjartsson T, et al. Genome-wide significant association between a sequence variant at 15q15.2 and lung cancer risk. *Cancer Res.* 2011; 71:1356-61. doi:10.1158/0008-5472.CAN-10-2852.
9. Zheng S, Qian P, Li F, Qian G, Wang C, Wu G, Li Q, Chen Y, Li J, Li H, He B, Ji F. Association of mitochondrial DNA variations with lung cancer risk in a Han Chinese population from southwestern China. *PLoS One.* 2012; 7:e31322. doi:10.1371/journal.pone.0031322.
10. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, et al.

Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007; 447:1087-93. doi:10.1038/nature05887.

11. Udler MS, Ahmed S, Healey CS, Meyer K, Struewing J, Maranian M, Kwon EM, Zhang J, Tyrer J, Karlins E, Platte R, Kalmyrzaev B, Dicks E, et al. Fine scale mapping of the breast cancer 16q12 locus. *Hum Mol Genet*. 2010; 19:2507-15. doi:10.1093/hmg/ddq122.
12. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2007; 39: 865-69. doi:10.1038/ng2064.
13. Yuan SH, Qiu Z, Ghosh A. TOX3 regulates calcium-dependent transcription in neurons. *Proc Natl Acad Sci USA*. 2009; 106:2909-14. doi:10.1073/pnas.0805555106.
14. Dittmer S, Kovacs Z, Yuan SH, Siszler G, Kögl M, Summer H, Geerts A, Golz S, Shioda T, Methner A. TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. *J Cell Sci*. 2011; 124:252-60. doi:10.1242/jcs.068759.
15. Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, Schmutzler RK, Versmold B, Engel C, Meindl A, Arnold N, Hofmann W, Sutter C, Niederacher D, et al. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet*. 2008; 82: 937-48. doi:10.1016/j.ajhg.2008.02.008.
16. Orr N, Cooke R, Jones M, Fletcher O, Dudbridge F, Chilcott-Burns S, Tomczyk K, Broderick P, Houlston R, Ashworth A, Swerdlow A. Genetic variants at chromosomes 2q35, 5p12, 6q25.1, 10q26.13, and 16q12.1 influence the risk of breast cancer in men. *PLoS Genet*. 2011; 7:e1002290. doi:10.1371/journal.pgen.1002290.
17. Orr N, Lemnrau A, Cooke R, Fletcher O, Tomczyk K, Jones M, Johnson N, Lord CJ, Mitsopoulos C, Zvelebil M, McDade SS, Buck G, Blancher C, et al. Genome-wide association study identifies a common variant in RAD51B associated with male breast cancer risk. *Nat Genet*. 2012; 44:1182-4. doi:10.1038/ng.2417.
18. Stevens KN, Vachon CM, Lee AM, Slager S, Lesnick T, Olswold C, Fasching PA, Miron P, Eccles D, Carpenter JE, Godwin AK, Ambrosone C, Winqvist R, et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res*. 2011; 71:6240-9. doi:10.1158/0008-5472.CAN-11-1266.
19. Purrington KS, Slager S, Eccles D, Yannoukakos D, Fasching PA, Miron P, Carpenter J, Chang-Claude J, Martin NG, Montgomery GW, Kristensen V, Anton-Culver H, Goodfellow P, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis*. 2014; 35:1012-9. doi:10.1093/carcin/bgt404.
20. Gudmundsdottir ET, Barkardottir RB, Arason A, Gunnarsson H, Amundadottir LT, Agnarsson BA, Johannsson OT, Reynisdottir I. The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. *BMC Cancer*. 2012; 12:621. doi:10.1186/1471-2407-12-621.
21. Fasching PA, Pharoah PD, Cox A, Nevanlinna H, Bojesen SE, Karn T, Broeks A, van Leeuwen FE, van't Veer LJ, Udo R, Dunning AM, Greco D, Aittomäki K, et al. The role of genetic breast cancer susceptibility variants as prognostic factors. *Hum Mol Genet*. 2012; 21:3926-39. doi:10.1093/hmg/ddc159.
22. Ruiz-Narváez EA, Rosenberg L, Cozier YC, Cupples LA, Adams-Campbell LL, Palmer JR. Polymorphisms in the TOX3/LOC643714 Locus and Risk of Breast Cancer in African-American Women. *Cancer Epidemiology Biomarkers & Prevention*. 2010; 19:1320-7. doi:10.1158/1055-9965.EPI-09-1250.
23. O'Brien KM, Cole SR, Poole C, Bensen JT, Herring AH, Engel LS, Millikan RC. Replication of breast cancer susceptibility loci in whites and African Americans using a Bayesian approach. *Am J Epidemiol*. 2014; 179:382-94. doi:10.1093/aje/kwt258.
24. He X, Yao G, Li F, Li M, Yang X. Risk-association of five SNPs in TOX3/LOC643714 with breast cancer in southern China. *Int J Mol Sci*. 2014; 15:2130-41. doi:10.3390/ijms15022130.
25. Zhang L, Long X. Association of three SNPs in TOX3 and breast cancer risk: Evidence from 97275 cases and 128686 controls. *Sci Rep*. 2015; 5:12773. doi:10.1038/srep12773.
26. Shan J, Mahfoudh W, Dsouza SP, Hassen E, Bouaouina N, Abdelhak S, Benhadjayed A, Memmi H, Mathew RA, Aigha II, Gabbouj S, Remadi Y, Chouchane L. Genome-Wide Association Studies (GWAS) breast cancer susceptibility loci in Arabs: susceptibility and prognostic implications in Tunisians. *Breast Cancer Res Treat*. 2012; 135:715-24. doi:10.1007/s10549-012-2202-6.
27. Long J, Shu XO, Cai Q, Gao YT, Zheng Y, Li G, Li C, Gu K, Wen W, Xiang YB, Lu W, Zheng W. Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev*. 2010; 19:2357-65. doi:10.1158/1055-9965.
28. Zheng W, Wen W, Gao YT, Shyr Y, Zheng Y, Long J, Li G, Li C, Gu K, Cai Q, Shu XO, Lu W. Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *J Natl Cancer Inst*. 2010; 102:972-81. doi:10.1093/jnci/djq170.
29. Liang J, Chen P, Hu Z, Shen H, Wang F, Chen L, Li M, Tang J, Wang H, Shen H. Genetic variants in trinucleotide repeat-containing 9 (TNRC9) are associated with risk of estrogen



- receptor positive breast cancer in a Chinese population. *Breast Cancer Res Treat.* 2010; 124:237-41. doi:10.1007/s10549-010-0809-z.
30. Long J, Cai Q, Shu XO, Qu S, Li C, Zheng Y, Gu K, Wang W, Xiang YB, Cheng J, Chen K, Zhang L, Zheng H, et al. Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet.* 2010; 6:e1001002. doi:10.1371/journal.pgen.1001002.
  31. Peng S, Lü B, Ruan W, Zhu Y, Sheng H, Lai M. Genetic polymorphisms and breast cancer risk: evidence from meta-analyses, pooled analyses, and genome-wide association studies. *Breast Cancer Res Treat.* 2011; 127:309-24. doi:10.1007/s10549-011-1459-5.
  32. Cowper-Salari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eekhoutte J, Moore JH, Lupien M. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet.* 2012; 44:1191-8. doi:10.1038/ng.2416.
  33. O'Brien KM, Cole SR, Engel LS, Bensen JT, Herring AH, Engel LS, Engel LS, Millikan RC. Replication of Breast Cancer Susceptibility Loci in Whites and African Americans Using a Bayesian Approach. *American Journal of Epidemiology.* 2014; 179:382-94. doi:10.1158/1055-9965.
  34. Bayraktar S, Thompson PA, Yoo SY, Do KA, Sahin AA, Arun BK, Bondy ML, Brewster AM. The relationship between eight GWAS-identified single-nucleotide polymorphisms and primary breast cancer outcomes. *Oncologist.* 2013; 18:493-500. doi:10.1634/theoncologist.2012-0419.
  35. Chen MB, Wu XY, Shen W, Wei MX, Li C, Cai B, Tao GQ, Lu PH. Association between polymorphisms of trinucleotide repeat containing 9 gene and breast cancer risk: evidence from 62,005 subjects. *Breast Cancer Res Treat.* 2011; 126:177-83. doi:10.1007/s10549-010-1114-6.
  36. Rinella ES, Shao Y, Yackowski L, Pramanik S, Oratz R, Schnabel F, Guha S, LeDuc C, Campbell CL, Klugman SD, Terry MB, Senie RT, Andrulis IL, et al. Genetic variants associated with breast cancer risk for Ashkenazi Jewish women with strong family histories but no identifiable BRCA1/2 mutation. *Hum Genet.* 2013; 132:523-36. doi:10.1007/s00439-013-1269-4.
  37. Huo D, Zheng Y, Ogundiran TO, Adebamowo C, Nathanson KL, Domchek SM, Rebbeck TR, Simon MS, John EM, Hennis A, Nemesure B, Wu SY, Leske MC, et al. Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis.* 2012; 33:835-40. doi:10.1093/carcin/bgs093.
  38. Haba-Rubio J, Marti-Soler H, Marques-Vidal P, Tobback N, Andries D, Preisig M, Waeber G, Vollenweider P, Kutalik Z, Tafti M, Heinzer R. Prevalence and Determinants of Periodic Limb Movements in the General Population. *Ann Neurol.* 2016; 79:464-74. doi: 10.1002/ana.24593.
  39. Moore H 4th, Winkelmann J, Lin L, Finn L, Peppard P, Mignot E. Periodic leg movements during sleep are associated with polymorphisms in BTBD9, TOX3/BC034767, MEIS1, MAP2K5/SKOR1, and PTPRD. *Sleep.* 2014; 37:1535-42. doi:10.5665/sleep.4006.
  40. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, Zong X, Laplana M, Wei Y, Han Y, Lloyd A, Delahaye-Sourdeix M, Chubb D, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet.* 2014; 46:736-41. doi:10.1038/ng.3002.
  41. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21:263-65. doi:10.1093/bioinformatics/bth457.
  42. Gabriel SB, Schaffner SF, Nguyen H. The structure of haplotype blocks in the human genome. *Science.* 2002; 296:2225-2229. doi:10.1126/science.1069424.
  43. Daly MJ, Rioux JD, Schaffner SF. High-resolution haplotype structure in the human genome. *Nature Genetics.* 2001; 29:229-32. doi:10.1038/ng1001-229.
  44. Margolis RL, Abraham MR, Gatchell SB, Li SH, Kidwai AS, Breschel TS, Stine OC, Callahan C, McInnis MG, Ross CA. cDNAs with long CAG trinucleotide repeats from human brain. *Hum Genet.* 1997; 100:114-22. doi:10.1007/s004390050476.
  45. Lin Y, Fu F, Chen M, Huang M, Wang C. Associations of two common genetic variants with breast cancer risk in a Chinese population: a stratified interaction analysis. *PLoS One.* 2014; 9:e115707. doi:10.1371/journal.pone.0115707.
  46. Winkelmann J, Czamara D, Schormair B, Knauf F, Schulte EC, Trenkwalder C, Dauvilliers Y, Polo O, Högl B, Berger K, Fuhs A, Gross N, Stiasny-Kolster K, et al. Genome-Wide Association Study Identifies Novel Restless Legs Syndrome Susceptibility Loci on 2p14 and 16q12.1. *Plos Genetics.* 2011; 7:e1002171. doi:10.1371/journal.pgen.1002171.
  47. Lin CH, , Chen ML, Wu VC, Li WY, Sy HN, Wu SL, Chang CC, Chiu PF, Liou HH, Lin CY, Chang HW, Lin SY, Wu KD, et al. Association of candidate genetic variants with restless legs syndrome in end stage renal disease: a multicenter case-control study in Taiwan. *Eur J Neurol.* 2014; 21:492-98. doi:10.1111/ene.12337.