Identification of two novel susceptibility variants on chromosome 5p15.33 for non-small cell lung cancer in Chinese population

Zhi Ji

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Ying Li

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Cheng Xiang

Xiaogan Centreal Hospital

Xia Luo

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Peihong Yuan

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Jieyi Long

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Na Shen

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Ying Shen

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Lingyan Deng

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Jiaoyuan Li

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Liming Cheng (✉ chengliming2015@163.com)

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

<https://orcid.org/0000-0001-6444-5090>

Research article

Keywords: non-small cell lung cancer, 5p15.33, low frequency, genetic variant, susceptibility

Posted Date: July 19th, 2019

DOI: <https://doi.org/10.21203/rs.2.11710/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Abstract Background Genome-wide association studies (GWAS) have identified 5p15.33 as a susceptibility locus for lung cancer. However, for non-small cell lung cancer (NSCLC), low-frequency risk variants in this region have not been systematically identified and studied. We have screened low-frequency variants on chromosome 5p15.33 in Chinese NSCLC patients by targeted sequencing before. **Methods** In this study, we selected candidate variants from low-frequency variants by comparing the genotype frequency of 400 NSCLC patients with 1,008 East Asians from The genome Aggregation Database (gnomAD). The associations between candidate variants and NSCLC were then discovered and replicated in two case-control sets: stage I with 960 cases and 916 controls, and stage II with 1,596 cases and 1,614 controls in total. **Results** Among the 5 candidates, 2 polymorphisms were significantly associated with risk of NSCLC, including rs33963617 (OR = 0.63, 95% CI: 0.53-0.76, $P = 3.80 \times 10^{-7}$) in *TERT* and rs77518573 (OR = 0.73, 95% CI: 0.63-0.84, $P = 2.00 \times 10^{-5}$) in *CLPTM1L*. When stratified by histologic subtype, a significant association was only investigated in adenocarcinoma for rs77518573. We also observed an obvious cumulative effect of the two significant variants. **Conclusions** We newly identified two NSCLC related variants on chromosome 5p15.33. Both *TERT*-rs33963617 and *CLPTM1L*-rs77518573 conferred reduced risk for NSCLC in Chinese population. Our study gives new insight into the power of low-frequency variants in explaining cancer heritability and expands our understanding about the essential role of *TERT*-*CLPTM1L* region in lung cancer development. **Keywords**: non-small cell lung cancer, 5p15.33, low frequency, genetic variant, susceptibility

Background

Lung cancer is the most common diagnosed malignancy around the world.[1] In China, the morbidity and mortality of lung cancer ranks the first among various malignancies, with an estimated 733,300 new cases and 610,200 deaths in 2015.[2] Environmental factors such as unhealthy lifestyle, fine-particulate air pollution and occupational exposure play the primary role in causing sporadic lung cancer.[3-5] Epidemiological investigations have showed that about one fifth lung cancers were attributable to cigarette smoking.[6] However, only a small proportion of the smokers would be suffered from lung cancer during whole lifetime, suggesting that genetic background might be a noticeable reason influencing the individual susceptibility to this disease.[7]

In recent years, genome-wide association studies (GWAS) have revealed amounts of gene loci predisposing for multiple malignancies in different populations. Chromosome 5p15.33 has been reported as a pleiotropic locus which associated with a variety of tumors such as lung cancer, glioma, breast cancer, ovarian cancer, and pancreatic cancer.[8-11] The discovery of the lung cancer susceptibility variants such as rs2736100 and rs401681 indicated this region might play an important role in lung cancer development.[12, 13] There are two plausible genes in chromosome 5p15.33, namely, the telomerase reverse transcriptase (*TERT*) gene and the cleft and palate transmembrane1 like (*CLPTM1L*) gene. The gene *TERT* encodes a catalytic subunit of telomerase reverse transcriptase which participated in maintaining the stability of chromosomes and ensuring the integrity of genetic information.[14] Telomerase was usually suppressed and inactive in most of the matured somatic cells, while about 85% of tumor cells could detect telomerase activity in varying degrees, suggesting a crucial role of *TERT* in emergence of tumors.[15, 16] *CLPTM1L* encodes a highly conserved transmembrane protein which was implicated in cell apoptosis and over-expressed in lung cancer tissues.[17, 18] Although systematic fine-mapping has been conducted to prioritize and characterize the causal variants in this region, the underling carcinogenic mechanisms have not been fully understood until now.[19, 20]

In our previous study, we performed a targeted sequencing based approach on *TERT-CLPTM1L* region to identify NSCLC associated common (minor allele frequency (MAF) $\geq 5.0\%$) variants and found that single nucleotide polymorphism (SNP) rs7726159 conferred lung cancer risk might by affecting MYC binding and inducing telomere lengthening.[21] However, evidences suggested that in addition to common variants, polymorphisms with low-frequency can also contribute to cancer susceptibility.[22, 23] To uncover the NSCLC related low frequency variants, we systematically assessed the association of low-frequency genetic variants in *TERT-CLPTM1L* region with the risk of NSCLC in a Chinese population in this study. Candidate variants were selected by comparing the frequency of low-frequency variants in 400 NSCLC patients scanned by previous targeted sequencing with that of East Asians (EAS) from The genome Aggregation Database (gnomAD). The relationships between candidate variants and NSCLC were further estimated by conducting a two-stage case-control study with a total of 2,556 NSCLC patients and 2,530 healthy controls.

Methods

2.1. Study subjects

As mentioned before, we enrolled 400 newly diagnosed NSCLC patients for targeted sequencing and acquired the spectrum of variation on chromosome 5p15.33.[21] In this study, we recruited a total of 2,556 patients with NSCLC and 2,530 healthy controls, partly contained in our previous study, to discover and validate the population associations. Patients were enrolled continuously from the Tongji Hospital of Huazhong University of Science and Technology (HUST), Wuhan, China during Jan 2014 to Jan 2018. Primary NSCLC cases with histological or cytological confirmation were eligible, while patients with previous cancer history, metastatic NSCLC from other organs, and previous treatment with radiotherapy or chemotherapy were excluded. Healthy controls with no history of malignant tumors and major diseases were randomly selected from subjects who went to the check-up center at the same hospital during the same time as the cases recruited. All the controls were frequency matched to the cases by age and gender. Three milliliter peripheral venous blood of all participants were collected in tubes with EDTA anticoagulant and stored freezingly at -80°C . All the participants received written informed consent and interviewed face-to-face on a standard questionnaire about basic demographic characteristics. This study was approved by the Institutional Review Committee of the Tongji Hospital, Tongji Medical College, HUST.

1.2. Selection of the candidate SNPs

Based on the results of targeted sequencing in our previous research,[21] we captured a total of 1,478 variants on chromosome 5p15.33 in 400 NSCLC patients. One thousand two hundred and fifteen variants were removed according to their rare frequency (MAF $\leq 1.0\%$). Among the remaining SNPs, 237 variants got through the quality control and 31 of them were defined as low-frequency variants with MAFs range from 1% to 5%. We thus downloaded the genotype frequency data of these 31 variants in 1,008 East Asians from gnomAD and compared each variant with frequency of our NSCLC patients. Significant variants in comparisons were selected as candidates.

1.3. DNA extraction and genotyping

The RelaxGene Blood DNA system DP349-02 (TIANGEN, Beijing, China) was used to extract genomic DNA from peripheral blood. The candidate SNPs were genotyped using the TaqMan SNP genotyping analysis on the ABI ViiATM7 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). All assays were performed on 384-well

plates with mixed cases and controls. Positive and negative samples were used for quality control on each plate. Ten percent samples were randomly selected for repetition, and the coincidence rate reached 98%.

1.4. Statistical analyses

Unpaired student's t-test was utilized to compare the distribution of continuous variable (age) between cases and controls. And Pearson's χ -square test was used to evaluate the difference of categorical variables about demographic characteristics (gender and smoking status) as well as genotypes in two groups. Hardy-Weinberg equilibrium (HWE) was calculated for each variant by using goodness-of-fit χ -square test among controls. Unconditional multivariate logistic regression was used to evaluate the effect of candidate variants on NSCLC susceptibility in total and subgroups analysis. All statistical analyses were performed in SPSS 21.0, and two-sided *P* value less than 0.05 was considered statistically significant.

Results

3.1. Characteristics of the study subjects

The demographic characteristics of all subjects and the basic clinical features about the NSCLC patients are shown in Table 1. A total of 2,556 NSCLC cases and 2,530 healthy controls were included in this study. In discovery stage with 960 cases and 916 controls, 69.4% of the patients diagnosed with NSCLC were males, and in healthy controls, males occupied a percentage of 68.1%. The mean (SD) age of the case and control groups were 58.0 (9.8) years and 58.6 (9.5) years, respectively. In the replication stage, we enrolled 1,596 NSCLC cases and 1,614 healthy controls. The mean (SD) age was 54.5 (7.4) in cases and 54.8 (7.6) in controls. Males accounted for 64.7% among cases compared with 65.8% among controls. When combining the two stages, there were no significant differences in distribution of gender (66.5% males in cases, 66.6% males in controls, *P* = 0.898) and age (55.8 ± 8.5 years for cases, 56.2 ± 8.6 years for controls, *P* = 0.087) between two groups. It's worth noting that more smokers were observed in NSCLC patients than healthy controls in both the two stages (*P* values being 0.006 for discovery stage and 3.07×10^{-4} for replication stage) as well as the combined samples (*P* = 0.002). Besides, more than 60% of the tumors were confirmed as adenocarcinomas and about 30% were squamous cell carcinomas (SCC), while the remaining cancers were other non-small cell histological types. Based on the Lung Cancer TNM classification criteria (IASLC 2009), 35.5% of the total NSCLC patients were diagnosed in an early stage (0, I or II stages) and 58.9% cases were in late stage (III and IV stages).

1.5. Screening of the candidate variants

Among the 31 low-frequency variants identified by sequencing in our NSCLC patients, seven variants were statistically significant when comparing their frequency with that of the East Asians from gnomAD (Supplementary Table 1). We therefore selected these seven SNPs as our candidate variants. However, two of the candidates which located in segmentally duplicated regions were failed in prob design, thus the remaining 5 polymorphisms were successfully genotyped and further analyzed.

1.6. Association between candidate SNPs and NSCLC susceptibility

In the discovery stage, we assessed the associations between 5 candidate variants and NSCLC susceptibility in 960 cases and 916 controls. As shown in Table 2, two polymorphisms were significantly associated with NSCLC risk, including rs33963617 (OR = 0.71, 95%CI: 0.53-0.94, *P* = 0.018 in dominant model) and rs77518573 (OR =

0.68, 95%CI: 0.53-0.88, $P = 0.003$ in dominant model). No significant association was observed between the other three variants and risk of NSCLC in this population. Based on these findings, we further validated the two promising associations in a replicated case-control study with a total of 1,596 cases and 1,614 controls. As a result, the two SNPs exhibited consistent significant relationships with NSCLC, with the estimated ORs being 0.59 (95%CI: 0.47-0.74, $P = 4.00 \times 10^{-6}$) for rs33963617 and 0.75 (95%CI: 0.63-0.90, $P = 0.002$) for rs77518573 under a dominant model. The associations retained significant when combining the samples of the two stages together (Table 3). People carrying the rs33963617 AG or AA genotypes had a reduced risk for NSCLC comparing with subjects harboring the GG homozygotes (OR = 0.63, 95% CI: 0.53-0.76, $P = 3.80 \times 10^{-7}$). As for rs77518573, a protective effect towards NSCLC was also observed in subjects carrying the minor allele: comparing with the wild CC genotype, the variant genotypes (AC or AA) conferred a significant decline in NSCLC risk (OR = 0.73, 95%CI: 0.630-0.84, $P = 2.00 \times 10^{-5}$). However, the impact of the mutant homozygotes seemed non-significant for both two variants owing to their scarcity in population.

Since the two significant variants, rs33963617 and rs77518573, were adjacent with a physical distance of about 67 kb on chromosome location, we next examined whether these variants have cumulative association by counting the number of associated alleles in each subject. Individuals carrying more than one minor allele conferred significant reduced risk than those carrying no minor allele at the two variants, and there was a gradual decrease in OR with a greater number of minor alleles (Figure 1). Specifically, compared with individuals carrying no minor allele, individuals carrying one, two or more minor alleles had an adjusted OR of 0.75 (95% CI: 0.66-0.85, $P = 1.10 \times 10^{-5}$) and 0.34 (95% CI: 0.23-0.49, $P = 1.06 \times 10^{-8}$), respectively.

Stratification was primarily implemented by smoking status according to the deleterious influence of smoking on lung cancer (Table 4). Among non-smokers, both the variants showed protective effects against NSCLC, with the adjusted ORs being 0.65 (95%CI: 0.51-0.83, $P = 0.001$) for rs33963617 and 0.68 (95%CI: 0.56-0.83, $P = 1.82 \times 10^{-4}$) for rs77518573. Smokers with the variant genotypes of rs33963617 exhibited significant reduced risk of NSCLC compared with smokers harboring the wild-type genotype (OR = 0.69, 95% CI: 0.54-0.89, $P = 0.005$). While for rs77518573, we observed a marginal effect in smokers, with OR of 0.82 (95%CI: 0.67-1.02, $P = 0.075$). However, no prominent interaction was found between smoking and rs77518573 ($P = 0.404$) or rs33963617 ($P = 0.853$). When stratified the NSCLC patients by pathological type, we found remarkable and consistent associations for rs33963617 with both adenocarcinoma (OR = 0.66, 95%CI: 0.56-0.79, $P = 2.00 \times 10^{-6}$) and SCC (OR = 0.68, 95%CI: 0.52-0.90, $P = 0.007$) (Table 4). The SNP rs77518573 was confirmed significantly associated with only adenocarcinoma (OR = 0.60, 95%CI: 0.49-0.74, $P = 1.00 \times 10^{-6}$) but not SCC (OR = 0.85, 95%CI: 0.67-1.06, $P = 0.148$).

Discussion

Growing evidence showed that common variants in the *TERT-CLPTM1L* region account for the bulk of malignancies including NSCLC, while comprehensive investigation is limited about disease-related low-frequency variants in this region. In current study, a targeted-sequencing based approach was performed to identify low frequency variants contributing to NSCLC. Genotyping in two stage case-control studies revealed that two genetic variants, rs33963617 and rs77518573 were significantly associated with NSCLC in Chinese population. Moreover, a significant cumulative effect with the risk of NSCLC reducing was illustrated as the number of variant genotypes increased.

During the past decade, GWASs have demonstrated dozens of genetic loci which predisposing to lung cancer. However, the known common variants identified by GWASs are collectively responsible for less than 30% of lung cancer heritability, indicating much of the genetic architecture to lung cancer remains unexplained.[24] Except for common variants, low frequency variants have been proposed as a portion source of heritability and therefore may be associated with the risk of complex diseases.[25, 26] Exome-wide scan and sequencing based approaches are commonly used strategies to assess the relationship of low-frequency and rare variants with cancer or other complex traits.[27, 28] To date, several low-frequency variants have been uncovered to be associated with lung cancer susceptibility and even prognosis.[29, 30] Fine-mapping based on deep sequencing of the chromosome 5p15.33 has been conducted to define the causal lung cancer susceptibility loci.[19, 20, 31] However, little attention has been drawn to the characterization of the disease associated low-frequency variants in this region. Our study provides compelling evidence that low-frequency variants within *TERT-CLPTM1L*, a well known GWAS locus on chromosome 5p15.33, can modulate risk of NSCLC in a Chinese population. These findings support an important role of low-frequency variants in lung cancer development and shed light on their capacity in explicating the missing heritability of complex traits.

The beneficial effects of rs33963617 and rs77518573 against lung cancer were confirmed for the first time in this study. SNP rs33963617 (c.2097G>A, p.Ala699=) is a synonymous variant located in the 5th exon of *TERT* gene. In particular, it is a low-frequency (MAF: 4.9%) variant in our NSCLC patients while commonly (MAF: 7.4%) arised in healthy controls. According to data from dbSNP, the frequency of this variant is much higher in East Asians (8.0%) than the other ethnic populations such as Europeans (0.7%), Africans (0.2%) and Americans (0.3%).[32] Therefore, the prominent association between rs33963617 and NSCLC risk may be specific in Chinese population considering the allelic discrepancy among different populations. Although no significant association with cancer has been reported previously, we found that rs33963617 was recorded in ClinVar database and annotated in relation to an inheritable disease called dyskeratosis congenital.[33] Intriguingly, dyskeratosis congenital is referred to as a telomere biology disorder accompanied with aberrant telomerase activity and telomeric imbalance which were generally driven by mutations linked to telomere maintenance.[34, 35] Telomere dysfunction can trigger cellular senescence and constitute a driving force for tumorigenesis by inducing limitless replicative potential and causing extensive genome instability, both of which were the hallmarks of cancer.[36, 37] Without changing the amino acid sequence, synonymous variant can also mediate gene regulation via transcriptional or post-transcriptional manner.[38-40] Hence, we speculate that rs33963617 may play a part in telomere biology and thus contribute to lung cancer susceptibility. However, further *in vivo* and *in vitro* experiments are required to arbitrate the concrete function of this variant and elucidate the potential mechanisms in tumorigenesis.

CLPTM1L is the other candidate gene at chromosome 5p15.33, albeit not as famous as *TERT*. The expression of this gene has been found increased in lung adenocarcinomas compared to matched normal lung tissues.[20] Meanwhile, over-expression of *CLPTM1L* predicted a poor survival of lung cancer patients.[41] Functional characterization established a protumorigenic role of *CLPTM1L* for Ras-driven lung cancer due to its regulation on cell survival signaling.[42] Several SNPs in the *CLPTM1L* gene have been reported to be associated with decreased risk of lung cancer.[43] And the risk genotype at this locus rendered a cis-regulation on *CLPTM1L* expression, suggesting a conceivable causative role of *CLPTM1L* in cancer development.[42] In our study, we newly identified a protective effect of non-coding variant rs77518573 against NSCLC. Notably, the frequency of rs77518573 was also disparate among different populations (9.4% in East Asians, less than 1% in Africans and Europeans and absent in Americans), insinuating a particular relationship with NSCLC in Chinese population. Since situates merely 1.6 kb upstream the transcription start site (TSS) of the *CLPTM1L* gene, the SNP [rs77518573](#) is supposed

to be a potential regulatory variant residing in a traditional promoter element.[44] Promoter is an essential component for transcription initiation and the assembly of RNA polymerase and associated regulators. Variants in this element thus may modulate the expression of its target gene by affecting the binding of transcriptional activators or repressors. Nevertheless, whether the SNP rs77518573 mediates a regulation to *CLPTM1L* and the underlie mechanisms need further elaboration in future.

We detected an obvious cumulative effect between the significant variants *TERT*-rs33963617 and *CLPTM1L*-rs77518573. This finding highlights the importance of both the *TERT* and *CLPTM1L* genes in lung cancer occurrence and prompts an additive or synergetic effect for non-linkaged variants in the same locus. In subgroup analysis, we observed consistent conspicuous associations with adenocarcinoma and SCC for *TERT*-rs33963617, while only significant relation to adenocarcinoma for *CLPTM1L*-rs77518573. Coincidentally, a previous study investigated similar results which showing concurrent associations with different histologic types of NSCLC for a common variant in *TERT*, whereas just significant correlation with SCC for another variant in *CLPTM1L*. [45] More than corroborate the pathological heterogeneity of lung cancers, the disparity may also remind the distinct carcinogenic mechanisms of the two genes.

Our study comprehensively investigated the relationship between low-frequency variants on chr5p15.33 and NSCLC susceptibility in Chinese population. Based on two stage case-control study representing a total of more than 5,000 subjects, we identified two novel NSCLC risk associated variants. Owing to the relative high variation frequency in East Asians and divergent genetic background among different populations, the significant associations with NSCLC may be specific in Chinese population. While considering the lack of the two protective alleles in races beyond yellow, people in west countries may be confronted with an elevated hazard for NSCLC by nature comparing with the people in East Asian. However, limitations should be addressed. Firstly, we scanned potential low-frequency variants in only 400 NSCLC patients which may lead to frequency departure. In fact, the significant variants were proved to be common in healthy controls though their frequency were much lower, even rare or absent, in other ethnicities. Replication of the associations may be thus required in other ethnic populations. Second, despite the prominent cumulative effect of the two significant variants, we didn't discuss the interaction between them due to the relative low frequency and the limited sample size. Our findings should be validated in large population associations. Finally, *in vivo* and *in vitro* experiments are warranted to explore the potential function of the significant variants and figure out their role in lung cancer tumorigenesis.

Conclusions

In summary, we newly identified two NSCLC related variants on chromosome 5p15.33. Both *TERT*-rs33963617 and *CLPTM1L*-rs77518573 conferred reduced risk for NSCLC in Chinese population. Our study gives new insight into the power of low-frequency variants in explaining cancer heritability and expands our understanding about the essential role of *TERT-CLPTM1L* region in lung cancer development.

Abbreviations

GWAS, genome wide association study; NSCLC, non-small cell lung cancer; TERT, telomerase reverse transcriptase; CLPTM1L, cleft-lip and palate transmembrane protein 1-like; LD, linkage disequilibrium; HUST, Huazhong University of Science and Technology; Q-PCR, quantitative polymerase chain reaction; SNP, single nucleotide polymorphism; MAF, minor allele frequency; HWE, Hardy-Weinberg equilibrium; OR, oddsratio; CI, confidence intervals; SCC, squamous cell carcinomas.

Declarations

Conflicts of interest

The authors declare that they have no potential conflicts of interest.

Authors' Contributions

Conception and design: Z. Ji, J. Li, L. Cheng

Development of methodology: Z. Ji, C. Xiang, L. Cheng

Acquisition of data (provided samples, acquired and managed patients, provided facilities, etc.): Z. Ji, Y. Li, X. Luo, P. Yuan, J. Long, N. Shen, Y. Shen, L. Deng, J. Li, L. Cheng

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Z. Ji, Y. Li, J. Li, L. Cheng

Writing, review, and/or revision of the manuscript: Z. Ji, J. Li, L. Cheng

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y. Li, C. Xiang, J. Li, L. Cheng

Study supervision: J. Li, L. Cheng

Funding information

This work was supported by the National Natural Science Foundation of China (Grant numbers: 81572071).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81572071).

We thank all the staffs of laboratory medicine whose contributions made this work possible.

Data availability statement

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

References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2019**. *CA Cancer J Clin* 2019, **69**(1):7-34.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J: **Cancer statistics in China, 2015**. *CA Cancer J Clin* 2016, **66**(2):115-132.
3. Biesalski HK, Bueno DMB, Chesson A, Chytil F, Grimble R, Hermus RJ, Kohrle J, Lotan R, Norpoth K, Pastorino U *et al*: **European Consensus Statement on Lung Cancer: risk factors and prevention**. *Lung Cancer Panel. CA Cancer J Clin* 1998, **48**(3):167-176, 164-166.
4. Pope CR, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD: **Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution**. *JAMA* 2002, **287**(9):1132-1141.

5. Pelucchi C, Pira E, Piolatto G, Coggiola M, Carta P, La Vecchia C: **Occupational silica exposure and lung cancer risk: a review of epidemiological studies 1996-2005.** *ANN ONCOL* 2006, **17**(7):1039-1050.
6. Islami F, Goding SA, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, McCullough ML, Patel AV, Ma J, Soerjomataram I *et al.* **Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States.** *CA Cancer J Clin* 2018, **68**(1):31-54.
7. Brennan P, Hainaut P, Boffetta P: **Genetics of lung-cancer susceptibility.** *LANCET ONCOL* 2011, **12**(4):399-408.
8. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV *et al.* **Common 5p15.33 and 6p21.33 variants influence lung cancer risk.** *NAT GENET* 2008, **40**(12):1407-1409.
9. Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, Simon M, Marie Y, Boisselier B, Delattre JY *et al.* **Genome-wide association study identifies five susceptibility loci for glioma.** *NAT GENET* 2009, **41**(8):899-904.
10. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, Edwards SL, Pickett HA, Shen HC, Smart CE *et al.* **Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer.** *NAT GENET* 2013, **45**(4):371-384, 381e-384e.
11. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, Arslan AA, Bueno-de-Mesquita HB, Gallinger S, Gross M *et al.* **A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33.** *NAT GENET* 2010, **42**(3):224-228.
12. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, Yang L, Dai J, Hu L, Tan W *et al.* **A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese.** *NAT GENET* 2011, **43**(8):792-796.
13. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J *et al.* **Lung cancer susceptibility locus at 5p15.33.** *NAT GENET* 2008, **40**(12):1404-1406.
14. Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR: **Telomerase catalytic subunit homologs from fission yeast and human.** *SCIENCE* 1997, **277**(5328):955-959.
15. Akincilar SC, Unal B, Tergaonkar V: **Reactivation of telomerase in cancer.** *CELL MOL LIFE SCI* 2016, **73**(8):1659-1670.
16. Shay JW, Bacchetti S: **A survey of telomerase activity in human cancer.** *EUR J CANCER* 1997, **33**(5):787-791.
17. Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y: **A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis.** *Biochem Biophys Res Commun* 2001, **280**(4):1148-1154.
18. Ni Z, Tao K, Chen G, Chen Q, Tang J, Luo X, Yin P, Tang J, Wang X: **CLPTM1L is overexpressed in lung cancer and associated with apoptosis.** *PLOS ONE* 2012, **7**(12):e52598.
19. Dong J, Cheng Y, Zhu M, Wen Y, Wang C, Wang Y, Geng L, Shen W, Liu J, Li Z *et al.* **Fine mapping of chromosome 5p15.33 identifies novel lung cancer susceptibility loci in Han Chinese.** *INT J CANCER* 2017, **141**(3):447-456.
20. Kachuri L, Amos CI, McKay JD, Johansson M, Vineis P, Bueno-de-Mesquita HB, Boutron-Ruault MC, Johansson M, Quiros JR, Sieri S *et al.* **Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci.** *CARCINOGENESIS* 2016, **37**(1):96-105.
21. Li Y, Xiang C, Shen N, Deng L, Luo X, Yuan P, Ji Z, Li J, Cheng L: **Functional polymorphisms on chromosome 5p15.33 disturb telomere biology and confer the risk of non-small cell lung cancer in Chinese population.** *Mol Carcinog* 2019, **58**(6):913-921.

22. Chang J, Zhong R, Tian J, Li J, Zhai K, Ke J, Lou J, Chen W, Zhu B, Shen N *et al*: **Exome-wide analyses identify low-frequency variant in CYP26B1 and additional coding variants associated with esophageal squamous cell carcinoma.** *NAT GENET* 2018, **50**(3):338-343.
23. Li J, Zou L, Zhou Y, Li L, Zhu Y, Yang Y, Gong Y, Lou J, Ke J, Zhang Y *et al*: **A low-frequency variant in SMAD7 modulates TGF-beta signaling and confers risk for colorectal cancer in Chinese population.** *Mol Carcinog* 2017, **56**(7):1798-1807.
24. Dai J, Shen W, Wen W, Chang J, Wang T, Chen H, Jin G, Ma H, Wu C, Li L *et al*: **Estimation of heritability for nine common cancers using data from genome-wide association studies in Chinese population.** *INT J CANCER* 2017, **140**(2):329-336.
25. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al*: **Finding the missing heritability of complex diseases.** *NATURE* 2009, **461**(7265):747-753.
26. Gazal S, Loh PR, Finucane HK, Ganna A, Schoech A, Sunyaev S, Price AL: **Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations.** *NAT GENET* 2018, **50**(11):1600-1607.
27. Gudmundsson J, Sulem P, Gudbjartsson DF, Masson G, Agnarsson BA, Benediktsdottir KR, Sigurdsson A, Magnusson OT, Gudjonsson SA, Magnusdottir DN *et al*: **A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer.** *NAT GENET* 2012, **44**(12):1326-1329.
28. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, Fine RS, Lu Y, Schurmann C, Highland HM *et al*: **Rare and low-frequency coding variants alter human adult height.** *NATURE* 2017, **542**(7640):186-190.
29. Jin G, Zhu M, Yin R, Shen W, Liu J, Sun J, Wang C, Dai J, Ma H, Wu C *et al*: **Low-frequency coding variants at 6p21.33 and 20q11.21 are associated with lung cancer risk in Chinese populations.** *AM J HUM GENET* 2015, **96**(5):832-840.
30. Zhu M, Geng L, Shen W, Wang Y, Liu J, Cheng Y, Wang C, Dai J, Jin G, Hu Z *et al*: **Exome-Wide Association Study Identifies Low-Frequency Coding Variants in 2p23.2 and 7p11.2 Associated with Survival of Non-Small Cell Lung Cancer Patients.** *J THORAC ONCOL* 2017, **12**(4):644-656.
31. Zhao Y, Chen G, Zhao Y, Song X, Chen H, Mao Y, Lu D: **Fine-mapping of a region of chromosome 5p15.33 (TERT-CLPTM1L) suggests a novel locus in TERT and a CLPTM1L haplotype are associated with glioma susceptibility in a Chinese population.** *INT J CANCER* 2012, **131**(7):1569-1576.
32. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K: **dbSNP: the NCBI database of genetic variation.** *NUCLEIC ACIDS RES* 2001, **29**(1):308-311.
33. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J *et al*: **ClinVar: public archive of interpretations of clinically relevant variants.** *NUCLEIC ACIDS RES* 2016, **44**(D1):D862-D868.
34. Martinez P, Blasco MA: **Telomere-driven diseases and telomere-targeting therapies.** *J CELL BIOL* 2017, **216**(4):875-887.
35. Tummala H, Walne A, Collopy L, Cardoso S, de la Fuente J, Lawson S, Powell J, Cooper N, Foster A, Mohammed S *et al*: **Poly(A)-specific ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita.** *J CLIN INVEST* 2015, **125**(5):2151-2160.
36. Shay JW: **Role of Telomeres and Telomerase in Aging and Cancer.** *CANCER DISCOV* 2016, **6**(6):584-593.
37. Maciejowski J, de Lange T: **Telomeres in cancer: tumour suppression and genome instability.** *Nat Rev Mol Cell Biol* 2017, **18**(3):175-186.

38. Sauna ZE, Kimchi-Sarfaty C: **Understanding the contribution of synonymous mutations to human disease.** *NAT REV GENET* 2011, **12**(10):683-691.
39. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, Mari B, Barbry P, Mosnier JF, Hebuterne X *et al*: **A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease.** *NAT GENET* 2011, **43**(3):242-245.
40. Biasoli D, Compston-Garnett L, Ricketts SL, Birand Z, Courtay-Cahen C, Fineberg E, Arendt M, Boerkamp K, Melin M, Koltookian M *et al*: **A synonymous germline variant in a gene encoding a cell adhesion molecule is associated with cutaneous mast cell tumour development in Labrador and Golden Retrievers.** *PLOS GENET* 2019, **15**(3):e1007967.
41. Ni Z, Chen Q, Lai Y, Wang Z, Sun L, Luo X, Wang X: **Prognostic significance of CLPTM1L expression and its effects on migration and invasion of human lung cancer cells.** *CANCER BIOMARK* 2016, **16**(3):445-452.
42. James MA, Vikis HG, Tate E, Rymaszewski AL, You M: **CRR9/CLPTM1L regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis.** *CANCER RES* 2014, **74**(4):1116-1127.
43. Jin T, Li B, He N, Zhang Y, Xia R, Kang L, Ding Y, Yuan D: **CLPTM1L polymorphism as a protective factor for lung cancer: a case-control study in southern Chinese population.** *Tumour Biol* 2016, **37**(8):10533-10538.
44. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG *et al*: **Ensembl 2018.** *NUCLEIC ACIDS RES* 2018, **46**(D1):D754-D761.
45. Myneni AA, Chang SC, Niu R, Liu L, Ochs-Balcom HM, Li Y, Zhang C, Zhao B, Shi J, Han X *et al*: **Genetic polymorphisms of TERT and CLPTM1L and risk of lung cancer—a case-control study in a Chinese population.** *LUNG CANCER* 2013, **80**(2):131-137.

Tables

Table 1. Selected characteristics of NSCLC cases and controls in this study.

	Stage I			Stage II			Combined Samples		
	Cases	Controls	<i>P</i>	Cases	Controls	<i>P</i>	Cases	Controls	<i>P</i>
	(n=960)	(n=916)		(n=1596)	(n=1614)		(n=2556)	(n=2530)	
Age, Mean (SD)	58.0 (9.8)	58.6 (9.5)	0.165	54.5 (7.4)	54.8 (7.6)	0.428	55.8 (8.5)	56.2 (8.6)	0.087
Gender, N (%)			0.558			0.522			0.898
Male	666 (69.4)	624 (68.1)		1033 (64.7)	1062 (65.8)		1699 (66.5)	1686 (66.6)	
Female	294 (30.6)	292 (31.9)		563 (35.3)	552 (34.2)		857 (33.5)	844 (33.4)	
Smoking status, N (%)			0.006			3.07×10 ⁻⁴			0.002
Smoking	474 (49.4)	407 (44.4)		803 (50.3)	706 (43.7)		1277 (50.0)	1113 (44.0)	
Non-smoking	455 (47.4)	505 (55.1)		776 (48.6)	866 (53.7)		1231 (48.1)	1371 (54.2)	
Unknown	31 (3.2)	4 (0.4)		17 (1.1)	42 (2.6)		48 (1.9)	46 (1.8)	
Pathological type, N (%)									
AC	583 (60.7)			1016 (63.7)			1599 (62.6)		
SCC	309 (32.2)			415 (26.0)			724 (28.3)		
Others	53 (5.5)			138 (8.6)			191 (7.5)		
Unknown	15 (1.6)			27 (1.7)			42 (1.6)		
Clinical stage, N (%)									
Early stage*	332 (34.6)			576 (36.1)			908 (35.5)		
Advanced stage [†]	593 (61.8)			912 (57.1)			1505 (58.9)		
Unknown	35 (3.6)			108 (6.8)			143 (5.6)		

AC, adenocarcinoma; SCC, squamous cell carcinomas

*: Early stage included stage 0, I and II.

†: Advanced stage included stage III and stage IV.

Table 2. Association between candidate variants and lung cancer risk.

Position	SNP	Gene	Location	Ref/Alt	MAF		OR (95% CI) [*]	<i>P</i> [†]
					Cases	Controls		
chr5:1268624	rs180675821	TERT	intronic	G/A	0.016	0.011	1.46 (0.82-2.61)	0.201
chr5:1279439	rs33963617	TERT	synonymous	G/A	0.050	0.072	0.71 (0.53-0.94)	0.018
chr5:1320880	rs74368690	CLPTM1L	intronic	G/A	0.011	0.015	0.77 (0.43-1.38)	0.376
chr5:1343620	rs552616298	CLPTM1L	intronic	A/T	0.014	0.019	0.80 (0.48-1.36)	0.417
chr5:1346773	rs77518573	CLPTM1L	Upstream	C/A	0.069	0.097	0.68 (0.53-0.88)	0.003

*: Adjusted for age, gender and smoking status.

†: *P* values alculated by a dominant model in logistic regression adjusted by sex, age and smoking status.

Table 3. The associations between rs33963617 and rs77518573 with NSCLC risk in Stage I, Stage II and combined samples.

rs33963617						rs77518573				
	Genotype	Cases	Controls	OR (95% CI)*	<i>P</i>	Genotype	Cases	Controls	OR (95% CI)*	<i>P</i>
Stage I	GG	861	785	1		CC	832	735	1	
	AG	91	117	0.72 (0.54- 0.97)	0.030	AC	121	157	0.69 (0.53- 0.89)	0.040
	AA	3	7	0.44 (0.11- 1.72)	0.239	AA	6	9	0.62 (0.22- 1.75)	0.363
	Dominant model			0.71 (0.53- 0.94)	0.018	Dominant model			0.68 (0.53- 0.88)	0.003
Stage II	GG	1426	1311	1		CC	1293	1245	1	
	AG	148	223	0.59 (0.48- 0.75)	6.00×10^{-6}	AC	285	354	0.76 (0.64- 0.91)	0.003
	AA	2	4	0.45 (0.08- 2.45)	0.354	AA	1	5	0.18 (0.02- 1.58)	0.122
	Dominant model			0.59 (0.47- 0.74)	4.00×10^{-6}	Dominant model			0.75 (0.63- 0.90)	0.002
Stage I +II	GG	2287	2096	1		CC	2125	1980	1	
	AG	239	340	0.64 (0.54- 0.77)	9.12×10^{-7}	AC	406	511	0.74 (0.64- 0.85)	4.20×10^{-5}
	AA	5	11	0.43 (0.15-	0.122	AA	7	14	0.47 (0.19-	0.100

		1.25)			1.16)
Dominant	0.63	3.80×10 ⁻⁷	Dominant	0.73	2.00×10 ⁻⁵
model	(0.53-		model	(0.63-	
	0.76)			0.84)	

*: Adjusted for age, gender and smoking status.

Table 4. Subgroup analysis about the association between the identified variants and lung cancer.

Subgroups	rs33963617				rs77518573			
	Cases [†]	Controls [†]	OR (95% CI) [‡]	<i>P</i>	Cases [†]	Controls [†]	OR (95% CI) [‡]	<i>P</i>
Smoking status								
NO	1131/118	1164/192	0.65 (0.51- 0.83)	0.001	1055/195	1099/295	0.68 (0.56- 0.83)	1.82×10 ⁻⁴
YES	1113/124	959/159	0.69 (0.54- 0.89)	0.005	1030/212	900/237	0.82 (0.67- 1.02)	0.075
Pathological type								
AC	1442/144	2096/351	0.60 (0.49- 0.74)	1.00×10 ⁻⁶	1349/239	1980/525	0.66 (0.56- 0.79)	2.00×10 ⁻⁶
SCC	642/74	2096/351	0.68 (0.52- 0.90)	0.007	592/127	1980/525	0.85 (0.67- 1.06)	0.148
Others	168/19	2096/351	0.67 (0.41- 1.09)	0.104	151/39	1980/525	0.90 (0.62- 1.30)	0.568
Clinical stage [*]								
Early stage	813/84	2096/351	0.62 (0.48- 0.80)	2.27×10 ⁻⁴	747/157	1980/525	0.81 (0.67- 0.99)	0.044
Advanced stage	1354/140	2096/351	0.62 (0.51- 0.77)	8.00×10 ⁻⁶	1260/233	1980/525	0.69 (0.58- 0.81)	1.60×10 ⁻⁵

AC, adenocarcinoma; SCC, squamous cell carcinomas

*: Early stage included stage 0, I and II, and advanced stage included stage III and IV.

†: Wild-type homozygote/Heterozygote and variant homozygote.

‡: adjusted for age, gender and smoking status under dominant model.

Figures

Figure 1

The cumulative effect of the two NSCLC associated polymorphisms rs33963617 and rs77518573.

Supplementary Files

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

- [SupplementTable1.docx](#)