



Published in final edited form as:

*Clin Cancer Res.* 2015 November 15; 21(22): 5173–5180. doi:10.1158/1078-0432.CCR-15-0009.

## ***TERT* Polymorphism rs2736100-C Is Associated with EGFR Mutation-Positive Non-Small Cell Lung Cancer**

Rongrong Wei, Ph.D.<sup>#1</sup>, Lan Cao, Ph.D.<sup>#2</sup>, Hengying Pu, Ph.D.<sup>#3</sup>, Hongwei Wang, Ph.D.<sup>#4</sup>, Yonglan Zheng, Ph.D.<sup>4</sup>, Xiaomin Niu, M.D.<sup>5</sup>, Xiaoling Weng, M.D.<sup>6</sup>, Hong Zhang, M.D.<sup>6</sup>, Murray Favus, M.D.<sup>4</sup>, Lanjun Zhang, M.D.<sup>7</sup>, Weihua Jia, Ph.D.<sup>3</sup>, Yixin Zeng, Ph.D.<sup>3</sup>, Christopher I Amos, Ph.D.<sup>8</sup>, Shun Lu, M.D.<sup>5,\*</sup>, Hui-Yun Wang, Ph.D.<sup>3,\*</sup>, Yun Liu, Ph.D.<sup>6,\*</sup>, and Wanqing Liu, Ph.D.<sup>1,\*</sup>

<sup>1</sup>Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, West Lafayette, IN, USA

<sup>2</sup>Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai, PR China

<sup>3</sup>State Key Laboratory of Oncology in South China, Sun Yat-Sen University Cancer Center, Guangzhou, P. R. China

<sup>4</sup>Department of Medicine, The University of Chicago, Chicago, IL, USA

<sup>5</sup>Department of Shanghai Lung Cancer Center, Shanghai Chest Hospital, Shanghai Jiaotong University, Shanghai, P. R. China

<sup>6</sup>Institutes of Biomedical Sciences, Fudan University, Shanghai, P. R. China

<sup>7</sup>Department of Thoracic Surgery, Sun Yat-Sen University Cancer Centre, Guangzhou, P. R. China

<sup>8</sup>Center for Genomic Medicine, Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA

# These authors contributed equally to this work.

### **Abstract**

**Purpose**—Epidermal growth factor receptor (EGFR) mutation-positive (EGFRmut+) non-small cell lung cancer (NSCLC) may be a unique orphan disease. Previous studies suggested that the telomerase reverse transcriptase (*TERT*) gene polymorphism is associated with demographic and clinical features strongly associated with EGFR mutations, e.g. adenocarcinoma histology, never-smoking history and female gender. We aim to test the association between *TERT* polymorphism and EGFRmut+ NSCLC.

**Experimental Design**—We conducted a genetic association study in Chinese NSCLC patients (n=714) and healthy controls (n=2,520), between the rs2736100 polymorphism and EGFRmut+

**Correspondence to:** Wanqing Liu, PhD. Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University. RHPH224B, 575 Stadium Mall Drive, IN 47907, USA. Tel: 765-496-6389; liu781@purdue.edu..

\*Joint senior authors

**Conflict of interest:** All authors disclose no potential conflicts of interest.

NSCLC. We further tested the association between the EGFR mutation status and mean leukocyte telomere length (LTL). The potential function of rs2736100 in lung epithelial cells was also explored.

**Results**—The rs2736100-C allele was significantly associated with EGFRmut+ NSCLC (OR=1.52, 95%CI=1.28–1.80,  $p=1.6\times 10^{-6}$ ) but not EGFRmut– NSCLC (OR=1.07, 95%CI=0.92–1.24,  $p=0.4$ ). While NSCLC patients as a whole have significantly longer LTL compared to healthy controls ( $p=10^{-13}$ ), the EGFRmut+ patients have even longer LTL compared to EGFRmut–patients ( $p=0.008$ ). Meanwhile, rs2736100 was significantly associated with *TERT* mRNA expression in both normal and tumor lung tissues. All results remained significant after controlling for age, gender, smoking status and histology ( $p<0.05$  for all tests). Moreover, the rs2736100 DNA sequence has an allele-specific affinity to nuclear proteins extracted from lung epithelial cells, which led to an altered enhancer activity of the sequence in vitro.

**Conclusion**—Our study suggests that telomerase and telomere function may be essential for carcinogenesis of EGFRmut+ NSCLC. Further investigation for the underlying mechanism is warranted.

### Keywords

EGFR mutation; NSCLC; TERT; rs2736100; genetic association

## INTRODUCTION

One of the most important findings in non-small cell lung cancer (NSCLC) research in the past decade is the discovery of somatic mutations in the *epidermal growth factor receptor* (*EGFR*) gene (1, 2). These mutations are located in the exons (18 to 21) encoding the EGFR tyrosine kinase (TK) domain (3, 4). More specifically, the missense point mutation L858R in exon 21 and in-frame microdeletions in exon 19 represent approximately 90% of all mutations (3, 4). The majority of these mutations has been characterized to be gain-of-function with enhanced EGFR signaling, and have demonstrated to be driver mutations for NSCLC (5–7). Transgenic murine models have shown that ectopic expression of mutant EGFR in the lung induces lung adenocarcinoma (5, 6). More importantly, these mutations are significantly associated with outcomes of EGFR-targeting therapy (8–10). These lines of evidence strongly support the notion that *EGFR* mutations play a critical role in both the development and treatment of NSCLC, and that NSCLC can be further defined by its *EGFR* somatic mutation status into unique subtypes (11, 12).

The mechanism underlying *EGFR* mutagenesis remains largely unknown. Thus far, no significant environmental mutagenic factors have been associated with these mutations. EGFR mutations were significantly associated with a never-smoking history in NSCLC patients (3, 13), excluding the involvement of tobacco smoke carcinogens. It has also not been associated with other known air pollution factors such as radon (14). However, the incidence of *EGFR* mutations has a distinct geographic distribution in human populations. *EGFR* somatic mutations were detected in 30%–50% of East Asian NSCLC patients, but in less than 20% of patients of other ancestries (3, 4). Moreover, the prevalence of these mutations in East Asians who migrate to other countries remains high, suggesting that the

development of these mutations is related to genetic background rather than geographic or environmental factors (3, 4). These observations strongly suggest a germline susceptibility to EGFRmut+ NSCLC. Therefore, identifying risk alleles for this orphan disease will not only reveal the mechanism underlying carcinogenesis of NSCLC, but will also potentially identify the high risk population for preventive plans.

In understanding genetic susceptibility to lung cancer, a number of loci have been identified in genome-wide association studies (GWAS) to date. Among these loci, the *telomerase reverse transcriptase (TERT)* gene was consistently associated with NSCLC in multiple GWAS and replication studies (15–22). More specifically, a common polymorphism rs2736100 in intron 2 of *TERT* has been strongly linked to lung adenocarcinoma, particularly in never-smoked women (16–18, 21, 23). Interestingly, among extensive epidemiological studies, EGFR mutations in NSCLC were also strongly associated with adenocarcinoma histology, never-smoking history and female gender (3, 4, 13). The overlap in association with these histological and demographic features between *EGFR* mutations and the rs2736100 polymorphism prompted us to hypothesize that rs2736100 may be a risk factor for EGFRmut+ NSCLC. Meanwhile, rs2736100 has been also strongly associated with leukocyte telomere length (LTL) in several GWAS (24–27). Our recent GWAS analysis of the LTL also confirmed the same association in a Han Chinese population (27). How the telomere biology is involved in the development of lung cancer remains incompletely understood. In particular, no study thus far has been performed to explore the relationship between TERT function and EGFR mutagenesis. In this study, we set out to test our hypothesis by conducting a genetic association study in a Han Chinese population. We further examined the association between LTL and EGFR mutation status. The potential role of rs2736100 in regulating TERT function was also investigated.

## MATERIALS AND METHODS

### Patient samples

This study included Han Chinese patients with NSCLC who were diagnosed and treated in Shanghai Chest Hospital and Sun Yat-Sen University Cancer Centre (Guangzhou), China, between 2008 and 2013, with written informed consent from all patients obtained. Patients were diagnosed and sample histology was reviewed in each hospital according to the World Health Organization tumor classification criteria (28). Biospecimens of a total of 714 patients were collected, including peripheral blood and their matched tumor tissue (n=351) or paired fresh frozen tumor and adjacent normal tissue (n=363). Venous blood samples were anti-coagulated with EDTA and stored in –80°C, while tissue samples were flash frozen and also stored in –80°C until use. Information about patient demographic characteristics and clinical data, including age, gender, smoking status (yes or no), TNM stage, histological classification (adenocarcinoma, squamous cell carcinoma or large cell carcinoma) were also collected for all patients. Blood samples of healthy control samples (n=2,520) were also collected from the above mentioned hospitals, excluding individuals with any lung disease or cancer. Again, demographic data including age, gender, disease status, and smoking history were obtained using questionnaires. Distribution of the collected information and comparison between groups (NSCLC vs control, and EGFRmut+ vs

EGFRmut-) were demonstrated in Table 1. Collection of samples and performance of this study were approved by the Institutional Review Boards (IRBs) of Shanghai Fudan University, Shanghai Chest Hospital, Shanghai Jiaotong University, Sun Yat-Sen University and Purdue University.

### Genotyping and EGFR mutation detection

Germline DNA was extracted from either whole blood or normal lung tissue, while tumor DNA was extracted from the matched tumor tissue. Genotyping of the rs2736100 was conducted using a Taqman-based assay (Life Technologies, CA, USA) in a PRISM 7900HT real-time PCR system (Life Technologies, CA, USA) according to the manufacturer's instruction. EGFR mutations (exons 18 to 21) were detected using Sanger sequencing with a protocol previously established in the lab (29). EGFRmut+ NSCLC patients were defined as individuals with any somatic mutation detected in exons 18–21 of the tumor DNA, while EGFRmut- patients were individuals with wild-type EGFR in their tumor.

### Quantification of TERT mRNA expression in lung tissue

Total RNA was extracted from paired cancer and adjacent normal lung tissue using a TRIzol® Plus RNA Purification Kit (Life Technologies, USA). For quantitative RT-PCR, total RNA (1µg) was reversely transcribed with random primers using moloney murine leukaemia virus (MMLV) reverse transcriptase (Promega, Madison, WI, USA). The quantitative PCR reactions were carried out with Platinum SYBR Green qPCR SuperMix-UDG reagents (Life Technologies, CA, USA) in a PRISM 7900HT system (Life Technologies, CA, USA) with the β-Actin gene (*ACTB*) as the internal control. Primer sequences for *TERT* gene amplification were TERT-F: 5'-GGCGTACAGGTTTCACGCA-3' and TERT-R: 5'-CGACATCCCTGCGTTCTTG-3'. Primer sequences for *ACTB* gene was described previously (29). Relative expression of *TERT* was defined as the difference (Ct) of Ct values between *ACTB* and *TERT* ( $Ct = Ct_{ACTB} - Ct_{TERT}$ ).

### Quantification of mean leukocyte telomere length (LTL)

Blood DNA of NSCLC patients (n=351) and part of healthy controls (n=343) were used to quantify the mean LTL using our previously established real-time PCR based protocol (27). Briefly, mean LTL telomere length was quantified as the quantity of telomere repeats relative to that of the *RNase P* gene as a reference, using the primers and conditions described before (27). Reactions were performed in duplicates in 10 µl reactions in the same plate with a PRISM 7900HT real-time PCR system (Life Technologies). The mean LTL was defined as a T/S ratio between telomere repeats and the *RNase P* gene for each sample. The T/S ratios were further log transformed (+log10) for subsequent analyses.

### Electrophoretic mobility shift assay (EMSA)

EMSA assay was performed based on a protocol previously established in the lab (30). Briefly, the A549 (ATCC) and 16HBE (a gift from Professor Dieter Gruenert at the University of California at San Francisco) were cultured in standard conditions and collected for nuclear protein extraction using a NE-PER Nuclear and Cytoplasmic Extraction Kit commercial kit (Thermal Scientific, IL, USA). EMSA was performed with total nuclear

extracts and probes with or without competitors using a Light Shift Chemiluminescent EMSA Kit (Thermal Scientific). Single-strand oligonucleotides and their complementary strands spanning the rs2736100 sequence were synthesized and annealed to double-strand DNA according to our previously published protocol(30). The oligonucleotide sequences were 5'-GGGCGGGGGCAAAGCTACAGAAACTCAACACGG-3' (C allele) and 5'-GGGCGGGGGCAAAGCTAAGAAACTCAACACGG (A allele). The oligos were either end-labeled with biotin as probes or non-labeled as competitors for biotin-labeled oligonucleotides. Briefly for EMSA reactions, 20 µl of binding reaction containing 20 pmol probes and 2 µg nuclear extract from A549 or 16HBE cells were incubated with or without competitors (200×) at room temperature for 20 minutes. Complexes were then resolved on 4% acrylamide gels (29:1 acrylamide:bisacrylamide). After electrophoresis, DNA and DNA/protein complexes were electrophoretically transferred to a nylon membrane. The transferred DNA was then cross-linked to the nylon membrane, and the biotin-labeled DNA-protein complex was detected by chemiluminescence kit (Pierce, Inc., Rockford, IL). The assay was repeated multiple times and the representative result was presented.

### Luciferase assays

A 217bp DNA fragment spanning the human *TERT* rs2736100 region was generated using PCR from a heterozygous DNA sample using the following primers: 5'-CTGGGTACCCTGCTGACTTAGTCC-3' and 5'-TTTGCTAGCAATAACAAGACAGAAGA-3'. A few nucleotides (underlined) were modified to create restriction enzyme digestion sites for KpnI and NheI, respectively. PCR products were gel-purified with a Gel Extraction Kit (Qiagen, Valencia, CA). The fragment was first cloned into a pCR2.1 vector using a T-A Cloning Kit (Invitrogen, Carlsbad, CA) and sequenced thereafter to obtain the A and C allele fragment, respectively. The plasmids containing A or C alleles were then amplified, digested with KpnI and NheI (New England Biolabs, MA, USA), gel-purified and cloned into the upstream multiple cloning site of the PGL3-Promoter Luciferase reporter plasmid (Promega). Subsequently, the constructed PGL3-A-P-Luc or PGL3-C-P-LuLuciferase plasmids were respectively co-transfected with the pCMV-beta-gal vector (Clontech, CA, USA) into the 16HBE and A549 cells with Lipofectamine 2000 Reagent (Life Technologies) following the manufacturer's protocol. The empty PGL3-Promoter vector was used as controls. Twenty-four hours after transfection, the cells were lysed with Tropic Lysis Buffer (Life Technologies). The luciferase and β-Galactosidase activities were measured using a Luciferase Assay System (Promega) and a β-Galactosidase Reporter Gene Assay System (Life Technologies) according to the manufacturers' instructions. β-Galactosidase activity was used to normalize transfection efficiency. The relative luciferase activities of the PGL3-TERT-P enhancer vectors were further normalized to the empty PGL3-Promoter vector. The experiment was performed in triplicates and repeated three times.

### Data analysis and statistics

The difference in the distribution of covariates (gender, smoking history and histology) between NSCLC and controls and between EGFRmut+ and EGFRmut- groups were examined using a Chi-squared test (CST), respectively, while the difference in age between groups was tested using a t-test. The CST was also used to test the associations between

rs2736100 and each of the phenotypes (NSCLC, EGFRmut+ NSCLC and EGFRmut- NSCLC). Odds ratio, 95%CI and p values were calculated. Given the strong association between covariates and EGFR mutation status (Table 1), a logistic multivariate regression analysis was further performed to test the association between the polymorphism and each of phenotypes by controlling age (constant), gender (binary), smoking history (binary, yes or no) and histology (binary, adenocarcinoma or non-adenocarcinoma), assuming an additive effect of the rs2736100 C allele. The corrected OR, 95%CI and p values were also calculated.

Difference of mean LTL between groups (healthy control vs NSCLC, EGFRmut+ vs EGFRmut-) was tested using a t-test, followed by a multivariate linear model with age, gender, smoking and histology data controlled. Comparison of *TERT* expression between paired samples were conducted using a paired t-test, while association between rs2736100 and *TERT* mRNA expression was performed using a multivariate linear model with all covariates controlled. Luciferase activity was compared using a t-test. All data analyses were performed using the SPSS 20.0 (Chicago, USA) and plotted using the Graphpad Prism 6.0 (CA, USA).

## RESULTS

### Association between demographic and clinical features and EGFR mutation

In order to test our hypothesis, we first detected mutations in EGFR in 714 Chinese NSCLC patients. The mutation rate was 42.4% (303/714), which is concordant with other reports (3, 4, 13). We then compared the difference in the distribution of age, gender, smoking history, histological information between NSCLC patients and healthy controls (n=2,520), and between EGFRmut+ (n=303) and EGFRmut- (n=414) groups. We found that there were significant differences in age, gender and smoking status between healthy controls and NSCLC patients. Consistent with previous reports (3, 4, 13), EGFR mutations were significantly more prevalent in younger patients (p=0.01), females (p<0.0001), never smokers (p<0.0001) and adenocarcinoma tumors (p<0.0001) (Table 1).

### Association between the rs2736100-C allele and EGFRmut+ NSCLC

To test whether the rs2736100 polymorphism was associated with EGFR mutations, we performed a genetic association study between the rs2736100 polymorphism and NSCLC as a single phenotype, as well as EGFRmut+ and EGFRmut- NSCLC as separate phenotypes, respectively. For the allelic association, we found that the rs2736100-C allele was significantly associated with NSCLC overall [odds ratio (OR)=1.24, 95% confidence interval (CI)=1.10-1.39, p=4×10<sup>-4</sup>]. However, when the patients were divided into EGFRmut+ and EGFRmut-, the same allele was more significantly associated with EGFRmut+ NSCLC (OR=1.52, 95%CI=1.28-1.80, p=1.6×10<sup>-6</sup>), but not with EGFRmut- NSCLC (OR=1.07, 95%CI=0.92-1.24, p=0.4). When compared between EGFRmut+ and EGFRmut- lung cancer patients, we also found that the C allele was significantly associated with EGFRmut+ NSCLC (OR=1.42, 95%CI=1.15-1.76, p = 1.1×10<sup>-3</sup>) (Table 2). We further investigated the genotypic risk of rs2736100 for each phenotype, using the A/A genotype as a reference genotype. As a result, the C/C genotype possessed a statistically

significant association with the highest risk (OR=2.35, 95%CI=1.66–3.32,  $p=7.4\times 10^{-7}$ ) for EGFRmut+ NSCLC compared to the non-significant association of the C/C genotype among EGFRmut- NSCLC (OR=1.18, 95%CI=0.84–1.56,  $p=0.38$ ) and the relatively lower risk for the C/C genotype for overall NSCLC (OR=1.56, 95%CI=1.23–1.98,  $p=3\times 10^{-4}$ ) (Table 3). There appeared to be a trend of additive effect of the C allele in all groups, with the C/A genotype conferring a mildly increased risk among EGFR+ cancers (OR=1.47, 95%CI=1.08–1.99,  $p=0.013$ ) and no significant associations for the C/A heterozygotes for the EGFR- cancers of lung cancer overall. Given the significant associations between EGFR mutations and age, gender, smoking status and histology (Table 1), we re-examined the aforementioned associations by controlling these covariates in a logistic regression model, with an assumption of additive effect of the C allele. We found that the polymorphism was still significantly associated with EGFRmut+ NSCLC when comparing between EGFRmut+ patients and healthy controls (corrected OR=1.52, corrected  $p=3.2\times 10^{-6}$ ), between EGFRmut+ and EGFRmut- patients (corrected OR=1.30, corrected  $p=0.035$ ), and between all NSCLC patients and healthy controls (corrected OR=1.29, corrected  $p=1.09\times 10^{-4}$ ) (Table 3).

#### Association between rs2736100, mean LTL and EGFR mutation status

Given the role of TERT in maintaining the telomere length, we further investigated the inter-relationship between rs2736100, mean leukocyte telomere length (LTL) and EGFR mutations. We quantified mean LTL using real-time PCR in NSCLC samples ( $n=351$ ) and healthy controls ( $n=343$ ) with enough blood DNA available. We found that the rs2736100-C allele was significantly associated with longer mean LTL in quantified samples, after adjusting for age, gender and smoking status (corrected  $p=0.002$ , Fig 1A). We further compared the mean LTL between healthy controls and EGFRmut+ and EGFRmut- patients. It was shown that while NSCLC patients as one population have significantly longer LTL than healthy controls (t-test,  $p=10^{-13}$ ), EGFRmut+ patients have even significantly longer LTL than the EGFRmut- patients (t-test,  $p=0.008$ ) (Fig 1B). Again, these differences remained significant after adjusting for age, gender and smoking status ( $p<0.043$  for both).

#### Inter-relationship between rs2736100, TERT gene expression in normal lung and NSCLC

In order to further understand the function of TERT and rs2736100 in NSCLC, we quantified mRNA levels of the *TERT* gene in both normal and tumor lung tissue samples with high quality RNA available ( $n=62$  for normal tissues and 52 of which with matched tumor RNA). We compared the *TERT* expression between the paired samples and found that there was a significant increase in *TERT* transcription in NSCLC tumor tissues compared with their adjacent normal tissue samples (Paired t-test,  $p=0.013$ , Fig 2A). We also tested the associations between rs2736100 and *TERT* mRNA levels among normal and tumor lung tissues, respectively. The rs2736100-C allele was significantly associated with increased *TERT* mRNA levels in both normal and tumor tissue samples with an additive effect (regression coefficient  $r=0.289$ ,  $p=0.023$  in tumor and  $r=0.287$ ,  $p=0.037$  in normal tissue, Fig 2B and 2C). After controlling age, gender, smoking and histology information, these associations remained significant (corrected  $p=0.027$  and 0.047, respectively).

### Allele-specific affinity between rs2736100 and nuclear proteins of lung epithelial cells

To explore the mechanism underlying the hypothesis that rs2736100 may regulate *TERT* transcription, we carried out an electrophoresis mobility shift assay (EMSA) to test whether rs2736100 has allelic interaction with nuclear protein factors. We tested this using nuclear extracts of a normal primary lung epithelial cell (16HBE) and a lung adenocarcinoma cell line (A549). We found that rs2736100 interacted with unknown protein factors, with two specific protein-DNA complex bands formed, and more strongly with the A-allele probe as compared to the C allele. Notably, there was also a stronger affinity between the rs2736100-A allele probe and the nuclear proteins extracted from 16HBE, as compared to that in A549 (Fig3A). The experiments were repeated multiple times with similar results observed.

### Enhancer activity of the DNA sequence flanking rs2736100

Next, we asked whether the DNA sequence around the rs2736100 locus played a role as a regulatory element for gene transcription. We cloned a 217bp DNA sequence spanning the rs2736100 polymorphism and tested its activity as a potential enhancer in regulating the luciferase reporter gene. We found that both the A and C allele-containing sequences were significantly associated with increased luciferase activity in both 16HBE and A549 cells, compared to the empty vector ( $p < 0.0001$ ), indicating an enhancer activity of the DNA sequence flanking rs2736100 (Fig 3B). The C allele-containing sequence exerted a significantly higher enhancer activity compared to the A allele in 16HBE ( $p = 0.0008$ ), but not in A549 ( $p = 0.26$ ).

## DISCUSSION

Our study for the first time linked a germline allele in the *TERT* gene to EGFRmut+ NSCLC, which potentially reveals an important mechanism underlying EGFR mutagenesis and lung cancer development. Previous studies have consistently observed associations between the C allele of rs2736100 and lung adenocarcinoma, lung cancer in women, and in never-smokers; our data suggests that these previous associations might be actually attributed to the underlying association between this allele and EGFRmut+ NSCLC as a unique orphan disease. Given the limited access to samples with both tumor and germline DNA available, our study only tested the genetic association in one population. While independent confirmation of our results will be essential, our findings based on the mechanistic studies consistently supported our hypothesis outlined above.

Our findings suggest that the interaction between EGFR and TERT pathways may play a critical role in NSCLC development. As a critical component of the telomerase, TERT is essential for maintaining telomeres, which protects chromosomal ends from degradation and prevents inappropriate DNA fusion and rearrangements. Specifically in lung cancer, *TERT* gene expression, activity and gene copy number are significantly increased in lung cancer cells (31, 32). While these observations indicate a direct involvement of TERT function in the physiology of somatic cells, other studies also suggest that increased TERT function confers a germline susceptibility to lung cancer, possibly by providing a necessary genetic background for cancer development. It was shown that LTL of NSCLC patients are significantly longer than that of healthy controls (32). While the rs2736100-C allele has





cancer subtype can have its unique genetic susceptibility. Indeed, previous studies have identified germline susceptibility alleles for a few somatic mutation-defined cancers, including a *MC1R* polymorphism and BRAF-mutant melanoma (44, 45), a *JAK2* germline polymorphism and the *JAK2* V617F-mutant myelo proliferative neoplasms (46) and a *FGFR3* 5'-distal germline polymorphism and *FGFR3* somatic mutations in urinary bladder cancer (47). We found in our study that the rs2736100-C allele exerted a higher relative risk (OR=1.52) among the EGFRmut+ NSCLC compared to general NSCLC as a single group (OR=1.24). This may also possibly explain the small effect size of cancer risk alleles observed in many GWAS over the past several years. The majority of these GWAS was focused on cancers classified based on histological information as a general phenotype, which actually includes multiple diseases attributed to different genetic risk alleles.

There are several questions that remain unaddressed. First, the EGFR mutations are significantly more common in NSCLC patients of East Asian origin than Caucasians and other ethnic groups. The mechanism underlying this ethnic difference is still unclear. The rs2736100-C allele was significantly associated with lung cancer risk in Caucasian, East Asian and African-American populations (16, 17, 21, 22, 35–39). While our study suggests the association between rs2736100 and EGFR mutations among Asian population, whether this allele confers risk for EGFR mutation in other populations needs to be further validated. Meanwhile, the C allele frequency in East Asian population (~37%–41%) is actually lower than in Caucasian (53%) but similar to that in African population (38%) according the HapMap data, which is not correlated with the distribution of EGFR mutations in these populations (East Asian>Caucasian>African-American). This reflects that other alleles contributing to the risk for EGFRmut+ lung cancer may still be yet to be identified. A genome-wide association analysis is thus warranted to discover other risk factors related to EGFRmut+ NSCLC. Second, a detailed intermediate mechanism underlying the interaction between EGFR and TERT is still largely unknown. Thus far, there are only limited studies focused on the crosstalk between these two pathways. The reason that increased germline TERT activity confers risks specifically to EGFR mutations as opposed to other somatic mutations needs to be further explored. Elucidating this detailed mechanism will be of importance to the understanding of lung cancer pathogenesis and development of new drugs for lung cancer treatment.

In conclusion, our study observed a significant association between *TERT* rs2736100-C allele and EGFRmut+ NSCLC via increasing *TERT* transcription and activity, which reveals insight into the pathogenesis of an important lung cancer subtype. Our data thus sheds new light on understanding the etiology of this lung cancer subtype. Validations in different populations and mechanistic studies exploring the detailed relationship between *TERT* and *EGFR* are warranted.

## GRANT SUPPORT

This study was supported in part by American Cancer Society-IL Division (Grant#: 189273) (W.L.); start-up fund of the College of Pharmacy, Purdue University (W.L.) and Research Fund of National Laboratory of Oncology in South China (H.W.). National Natural Science Foundation of China (NSFC) (Grant#: 81302005), Project of Shanghai Municipality Science & Technology Commission (Grant#: 13ZR1438500), Youth Foundation of Shanghai Municipal Public Health Bureau (Grant#: 20124Y114), Sino-Swiss Lung Cancer Clinic Center joint translational medicine research (Grant#: 2012DFG31320).

## REFERENCES

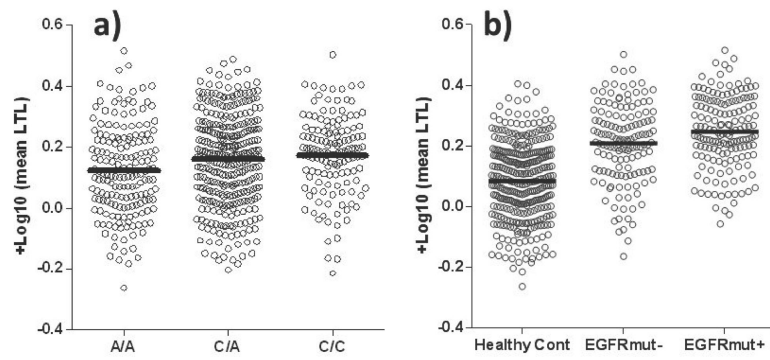
1. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*. 2004; 304:1497–500. [PubMed: 15118125]
2. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *The New England journal of medicine*. 2004; 350:2129–39. [PubMed: 15118073]
3. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *Journal of the National Cancer Institute*. 2005; 97:339–46. [PubMed: 15741570]
4. Siegelin MD, Borczuk AC. Epidermal growth factor receptor mutations in lung adenocarcinoma. *Laboratory investigation; a journal of technical methods and pathology*. 2014; 94:129–37.
5. Ji H, Li D, Chen L, Shimamura T, Kobayashi S, McNamara K, et al. The impact of human EGFR kinase domain mutations on lung tumorigenesis and in vivo sensitivity to EGFR-targeted therapies. *Cancer cell*. 2006; 9:485–95. [PubMed: 16730237]
6. Politi K, Zakowski MF, Fan PD, Schonfeld EA, Pao W, Varmus HE. Lung adenocarcinomas induced in mice by mutant EGF receptors found in human lung cancers respond to a tyrosine kinase inhibitor or to down-regulation of the receptors. *Genes & development*. 2006; 20:1496–510. [PubMed: 16705038]
7. Greulich H, Chen TH, Feng W, Janne PA, Alvarez JV, Zappaterra M, et al. Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS medicine*. 2005; 2:e313. [PubMed: 16187797]
8. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *The New England journal of medicine*. 2009; 361:947–57. [PubMed: 19692680]
9. Tsao MS, Sakurada A, Cutz JC, Zhu CQ, Kamel-Reid S, Squire J, et al. Erlotinib in lung cancer - molecular and clinical predictors of outcome. *The New England journal of medicine*. 2005; 353:133–44. [PubMed: 16014883]
10. Wu YL, Zhou C, Hu CP, Feng J, Lu S, Huang Y, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *The lancet oncology*. 2014; 15:213–22. [PubMed: 24439929]
11. Planchard D. Identification of driver mutations in lung cancer: first step in personalized cancer. *Targeted oncology*. 2013; 8:3–14. [PubMed: 23371030]
12. An SJ, Chen ZH, Su J, Zhang XC, Zhong WZ, Yang JJ, et al. Identification of enriched driver gene alterations in subgroups of non-small cell lung cancer patients based on histology and smoking status. *PloS one*. 2012; 7:e40109. [PubMed: 22768234]
13. Tokumo M, Toyooka S, Kiura K, Shigematsu H, Tomii K, Aoe M, et al. The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2005; 11:1167–73. [PubMed: 15709185]
14. Taga M, Mechanic LE, Hagiwara N, Vahakangas KH, Bennett WP, Alavanja MC, et al. EGFR somatic mutations in lung tumors: radon exposure and passive smoking in former- and never-smoking U.S. women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2012; 21:988–92.
15. Brenner DR, Brennan P, Boffetta P, Amos CI, Spitz MR, Chen C, et al. Hierarchical modeling identifies novel lung cancer susceptibility variants in inflammation pathways among 10,140 cases and 11,012 controls. *Human genetics*. 2013; 132:579–89. [PubMed: 23370545]
16. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nature genetics*. 2011; 43:792–6. [PubMed: 21725308]

17. Miki D, Kubo M, Takahashi A, Yoon KA, Kim J, Lee GK, et al. Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nature genetics*. 2010; 42:893–6. [PubMed: 20871597]
18. Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS, et al. The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS genetics*. 2010; 6
19. Truong T, Hung RJ, Amos CI, Wu X, Bickeboller H, Rosenberger A, et al. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *Journal of the National Cancer Institute*. 2010; 102:959–71. [PubMed: 20548021]
20. Wang Y, Broderick P, Matakidou A, Eisen T, Houlston RS. Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. *Carcinogenesis*. 2010; 31:234–8. [PubMed: 19955392]
21. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet*. 2009; 85:679–91. [PubMed: 19836008]
22. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, et al. Lung cancer susceptibility locus at 5p15.33. *Nature genetics*. 2008; 40:1404–6. [PubMed: 18978790]
23. Mocellin S, Verdi D, Pooley KA, Landi MT, Egan KM, Baird DM, et al. Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. *Journal of the National Cancer Institute*. 2012; 104:840–54. [PubMed: 22523397]
24. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet*. 2013; 45:422–7. [PubMed: 23535734]
25. Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Human molecular genetics*. 2013; 22:5056–64. [PubMed: 23900074]
26. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nature genetics*. 2013; 45:371–84. [PubMed: 23535731]
27. Liu Y, Cao L, Li Z, Zhou D, Liu W, Shen Q, et al. A Genome-Wide Association Study Identifies a Locus on TERT for Mean Telomere Length in Han Chinese. *PloS one*. 2014; 9:e85043. [PubMed: 24465473]
28. Teng XD. World Health Organization classification of tumours, pathology and genetics of tumours of the lung. *Zhonghua bing li xue za zhi Chinese journal of pathology*. 2005; 34:544–6. [PubMed: 16383307]
29. Liu W, Wu X, Zhang W, Montenegro RC, Fackenthal DL, Spitz JA, et al. Relationship of EGFR mutations, expression, amplification, and polymorphisms to epidermal growth factor receptor inhibitors in the NCI60 cell lines. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2007; 13:6788–95. [PubMed: 18006781]
30. Liu W, Innocenti F, Wu MH, Desai AA, Dolan ME, Cook EH Jr. et al. A functional common polymorphism in a Sp1 recognition site of the epidermal growth factor receptor gene promoter. *Cancer research*. 2005; 65:46–53. [PubMed: 15665278]
31. Kang JU, Koo SH, Kwon KC, Park JW, Kim JM. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer genetics and cytogenetics*. 2008; 182:1–11. [PubMed: 18328944]
32. Lan Q, Cawthon R, Gao Y, Hu W, Hosgood HD 3rd, Barone-Adesi F, et al. Longer telomere length in peripheral white blood cells is associated with risk of lung cancer and the rs2736100 (CLPTM1L-TERT) polymorphism in a prospective cohort study among women in China. *PloS one*. 2013; 8:e59230. [PubMed: 23555636]
33. Piao CQ, Liu L, Zhao YL, Balajee AS, Suzuki M, Hei TK. Immortalization of human small airway epithelial cells by ectopic expression of telomerase. *Carcinogenesis*. 2005; 26:725–31. [PubMed: 15677631]

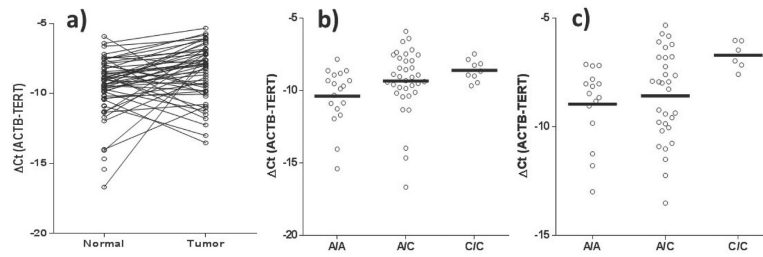
34. Hu Y, Shen Y, Ji B, Wang L, Zhang Z, Zhang Y. Combinational RNAi gene therapy of hepatocellular carcinoma by targeting human EGFR and TERT. *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences*. 2011; 42:387–91. [PubMed: 21238587]
35. Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nature genetics*. 2010; 42:604–7. [PubMed: 20543847]
36. Kinnersley B, Migliorini G, Broderick P, Whiffin N, Dobbins SE, Casey G, et al. The TERT variant rs2736100 is associated with colorectal cancer risk. *British journal of cancer*. 2012; 107:1001–8. [PubMed: 22878375]
37. Rajaraman P, Melin BS, Wang Z, McKean-Cowdin R, Michaud DS, Wang SS, et al. Genome-wide association study of glioma and meta-analysis. *Human genetics*. 2012; 131:1877–88. [PubMed: 22886559]
38. Gago-Dominguez M, Jiang X, Conti DV, Castela JE, Stern MC, Cortessis VK, et al. Genetic variations on chromosomes 5p15 and 15q25 and bladder cancer risk: findings from the Los Angeles-Shanghai bladder case-control study. *Carcinogenesis*. 2011; 32:197–202. [PubMed: 21081471]
39. Nan H, Qureshi AA, Prescott J, De Vivo I, Han J. Genetic variants in telomere-maintaining genes and skin cancer risk. *Human genetics*. 2011; 129:247–53. [PubMed: 21116649]
40. Zou P, Gu A, Ji G, Zhao L, Zhao P, Lu A. The TERT rs2736100 polymorphism and cancer risk: a meta-analysis based on 25 case-control studies. *BMC cancer*. 2012; 12:7. [PubMed: 22221621]
41. Shames DS, Wistuba II. The evolving genomic classification of lung cancer. *The Journal of pathology*. 2014; 232:121–33. [PubMed: 24114583]
42. Previati M, Manfrini M, Galasso M, Zerbinati C, Palatini J, Gasparini P, et al. Next generation analysis of breast cancer genomes for precision medicine. *Cancer letters*. 2013; 339:1–7. [PubMed: 23879964]
43. Mehnert JM, Kluger HM. Driver mutations in melanoma: lessons learned from bench-to-bedside studies. *Current oncology reports*. 2012; 14:449–57. [PubMed: 22723080]
44. Landi MT, Bauer J, Pfeiffer RM, Elder DE, Hulley B, Minghetti P, et al. MC1R germline variants confer risk for BRAF-mutant melanoma. *Science*. 2006; 313:521–2. [PubMed: 16809487]
45. Fargnoli MC, Pike K, Pfeiffer RM, Tsang S, Rozenblum E, Munroe DJ, et al. MC1R variants increase risk of melanomas harboring BRAF mutations. *The Journal of investigative dermatology*. 2008; 128:2485–90. [PubMed: 18368129]
46. Kilpivaara O, Mukherjee S, Schram AM, Wadleigh M, Mullally A, Ebert BL, et al. A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nature genetics*. 2009; 41:455–9. [PubMed: 19287384]
47. Kiemeny LA, Sulem P, Besenbacher S, Vermeulen SH, Sigurdsson A, Thorleifsson G, et al. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nature genetics*. 2010; 42:415–9. [PubMed: 20348956]

### TRANSLATIONAL RELEVANCE

The C allele of the rs2736100 polymorphism in human *TERT* gene has been associated with increased risk for lung cancer, in particular for the subtypes related to adenocarcinoma, female gender, and nonsmoking history, a group of unique demographic and histologic features. The reason for this preferable association remains unclear. Our study observed a strong inter-relationship between rs2736100-C, longer telomere and EGFR mutation-driven lung cancer subtype. This is particularly consistent with previously well-documented associations between EGFR mutations and the aforementioned demographic and histologic features, between these features and longer telomere as well as between longer telomere and rs2736100-C. Our study further suggests that rs2736100-C may at least in part lead to the altered TERT function. Our data for the first time connected these associations at the molecular level, and revealed the molecular basis underlying the lung cancer subtype driven by EGFR mutations, which provides important insights into the lung cancer etiology.



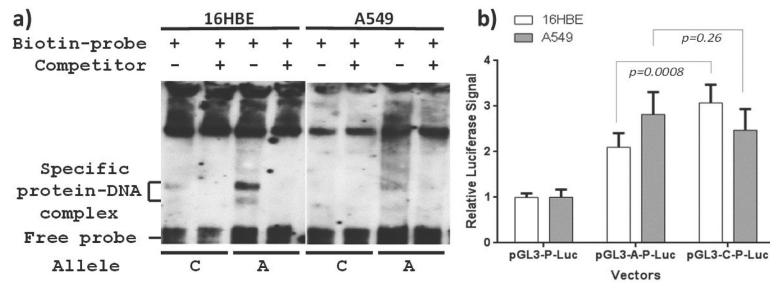
**Fig 1.** Association between rs2736100 and mean LTL among all quantified samples (n=691) (corrected  $p=0.002$ ) (A). Comparison of mean LTL between NSCLC and healthy controls (t-test corrected  $p=10^{-13}$ ), and between EGFRmut+ and EGFRmut- (t-test, corrected  $p=0.043$ ) (B). Horizontal bars refer to the mean.



**Fig 2.**

Comparison of *TERT* mRNA expression between paired tumor and adjacent normal tissue (paired t-test,  $p=0.013$ ) (A). Association between rs2736100 and *TERT* mRNA expression in normal (corrected  $p=0.047$ ) (B) and tumor tissue (corrected  $p=0.027$ ) (C). Horizontal bars refer to the mean.



**Fig 3.**

EMSA assay testing the affinity between rs2736100 allele and nuclear extracts of two lung epithelial cell lines 16HBE and A549 (**A**). Luciferase activities of rs2736100 flanking sequences contains A (PGL3-A-P-Luc) and C allele (PGL3-C-P-Luc). Luciferase activities of both alleles in the two cell lines were significantly higher than that of the empty vector ( $p < 0.0001$ ). There was also a significantly higher luciferase activity of the C allele compared to the A allele in 16HBE ( $p = 0.0008$ ), but not in A549 ( $p = 0.26$ ) (**B**).

**Table 1**

Distribution of demographic and histological information among NSCLC patients and healthy controls.

Cofactors	NSCLC			P Value <sup>a</sup>	
	EGFRmut+	EGFRmut-	Healthy Control	NSCLC vs. Control	EGFRmut+ vs. EGFRmut-
Age (Mean±SD)	56.7±10.5	58.7±10.5	60.5±10.3	<0.0001	0.012
Gender (Male)	144	315	874	<0.0001	<0.0001
Smoking (Yes)	85	273	496	<0.0001	<0.0001
Histology (ADC <sup>b</sup> )	295	288	-	-	<0.0001
Total	303	411	2520		

<sup>a</sup> Age data were compared using t-test, while gender, smoking and histological information were compared using Chi-squared test.

<sup>b</sup> ADC refers to adenocarcinoma.

**Table 2**

Allelic association between rs2736100 and NSCLC and subtypes.

<b>Genotype</b>	<b>Case (N)</b>	<b>%</b>	<b>Cont (N)</b>	<b>%</b>	<b>OR</b>	<b>95% CI</b>	<b>P value</b>
<i>All NSCLC vs Healthy Cont</i>							
C	671	47.8	2143	42.5	1.24	1.10–1.39	4×10 <sup>-4</sup>
A	733	52.2	2897	57.5			
<i>EGFRmut+ vs Healthy Cont</i>							
C	313	52.9	2143	42.5	1.52	1.28–1.80	1.6×10 <sup>-6</sup>
A	279	47.1	2897	57.5			
<i>EGFRmut- vs Healthy Cont</i>							
C	358	44.1	2143	42.5	1.07	0.92–1.24	0.4
A	454	55.9	2897	57.5			
<i>EGFRmut+ vs EGFRmut-</i>							
C	313	52.9	358	44.1	1.42	1.15–1.76	1.1×10 <sup>-3</sup>
A	279	47.1	454	55.9			

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Genotypic association between rs2736100 and NSCLC and subtypes.

Genotype	Case (N)	%	Cont (N)	%	OR	95% CI	P value	Corrected OR	Corrected P value
<i>All NSCLC vs Healthy Controls</i>									
C/C	159	22.6%	437	17.3%	1.56	1.23–1.98	3.0E–04	1.29	1.09×10 <sup>–4</sup>
C/A	353	50.3%	1269	50.4%	1.19	0.98–1.45	0.08		
A/A	190	27.1%	814	32.3%	referent				
<i>EGFRmut+ vs Healthy Controls</i>									
C/C	82	27.7%	437	17.3%	2.35	1.66–3.32	7.4E–07	1.52	3.2×10 <sup>–6</sup>
C/A	149	50.3%	1269	50.4%	1.47	1.08–1.99	0.013		
A/A	65	22.0%	814	32.3%	referent				
<i>EGFRmut– vs Healthy Controls</i>									
C/C	77	19.0%	437	17.3%	1.18	0.84–1.56	0.38	1.11	0.22
C/A	204	50.2%	1269	50.4%	1.03	0.81–1.30	0.71		
A/A	125	30.8%	814	32.3%	referent				
<i>EGFRmut+ vs EGFRmut–</i>									
C/C	82	27.7%	77	19.0%	2.05	1.33–3.16	1.1E–03	1.30	0.035
C/A	149	50.3%	204	50.2%	1.41	0.97–2.03	0.069		
A/A	65	22.0%	125	30.8%	referent				