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

Rongrong Wei1, Lan Cao2, Hengying Pu3, Hongwei Wang4, Yonglan Zheng4, Xiaomin Niu5, Xiaoling Weng6, Hong Zhang6, Murray J. Favus3, Lanjun Zhang7, Weihua Jia3, Yixin Zeng3, Christopher I. Amos8, Shun Lu5, Hui-Yun Wang3, Yun Liu6, and Wanqing Liu1

**Abstract**

**Purpose:** EGF 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, for example, 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 patients with NSCLC (n = 714) and healthy controls (n = 2,520), between the rs2736100 polymorphism and EGFRmut+ 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% confidence interval [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 patients with NSCLC as a whole have significantly longer LTL than healthy controls (\( P \leq 10^{-14} \)), the EGFRmut+ patients have even longer LTL than 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.

**Conclusions:** Our study suggests that telomerase and telomere function may be essential for carcinogenesis of EGFRmut+ NSCLC. Further investigation for the underlying mechanism is warranted. *Clin Cancer Res; 21(22); 5173–80. ©2015 AACR.*

**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 *EGF receptor (EGFR)* gene (1, 2). These mutations are located in the exons (18–21) 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 have 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 patients with NSCLC (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% to 50% of East Asian patients with NSCLC but in less than 20% of patients of other ancestries (3, 4). Moreover, the
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 interrelationship between rs2736100-C, longer telomere, and EGFR receptor (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.

Genotyping and EGFR mutation detection

Germline DNA was extracted from either whole blood or normal lung tissue, whereas tumor DNA was extracted from the matched tumor tissue. Genotyping of the rs2736100 was conducted using a TaqMan-based assay (Life Technologies) in a PRISM 7900HT real-time PCR system (Life Technologies) according to the manufacturer’s instructions. EGFR mutations (exons 18–21) were detected using Sanger sequencing with a protocol

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 interrelationship between rs2736100-C, longer telomere, and EGFR receptor (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.

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 WHO 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 anticoagulated with EDTA and stored in −80°C, whereas 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), tumor-node-metastasis (TNM) stage, and histologic 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−) are demonstrated in Table 1. Collection of samples and performance of this study were approved by the Institutional Review Boards (IRB) of Shanghai Fudan University, Shanghai Chest Hospital, Shanghai Jiaotong University (Shanghai, PR China), Sun Yat-Sen University (Guangzhou, PR China), and Purdue University (West Lafayette, IN).

Table 1. Distribution of demographic and histologic information among NSCLC patients and healthy controls

<table>
<thead>
<tr>
<th>Cofactors</th>
<th>NSCLC</th>
<th>Healthy control</th>
<th>NSCLC vs. control</th>
<th>EGFRmut+ vs. EGFRmut−</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD)</td>
<td>56.7 ± 10.5</td>
<td>58.7 ± 10.5</td>
<td>60.5 ± 10.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gender (Male)</td>
<td>144</td>
<td>315</td>
<td>874</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Smoking (Yes)</td>
<td>85</td>
<td>273</td>
<td>496</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Histology (ADC)</td>
<td>295</td>
<td>288</td>
<td>—</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total</td>
<td>503</td>
<td>411</td>
<td>2,520</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: ADC, adenocarcinoma.

*Age data were compared using a t test, whereas gender, smoking, and histologic information were compared using CST.
previously established in the laboratory (29). EGFRmut− NSCLC patients were defined as individuals with any somatic mutation detected in exons 18 to 21 of the tumor DNA, whereas 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). For quantitative RT-PCR, total RNA (1 μg) was reversely transcribed with random primers using moloney murine leukemia virus (MMLV) reverse transcriptase (Promega). The reversely transcribed RNA was amplified with TaqMan Universal PCR Master Mix (Life Technologies). For quantitative RT-PCR, total RNA (1 μg) was reversely transcribed with random primers using moloney murine leukemia virus (MMLV) reverse transcriptase (Promega). The reversely transcribed RNA was amplified with TaqMan Universal PCR Master Mix (Life Technologies). Primer sets for TERT were as follows: TERT-F: 5′-CTCGGTACAGGTTTCACGCA-3′ and TERT-R: 5′-CGACATCCCGCTGTCTTGTG-3′. Primer sequences for ACTB gene were described previously (29). The mean LTL was defined as the difference (∆Ct) of Ct values between ACTB and TERT (∆Ct = Ct_ACTB − Ct_TERT).

Quantification of mean LTL

Blood DNA of NSCLC patients (n = 351) and part of healthy controls (n = 343) was 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 luciferase activity 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. The 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 3 times.

Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) was performed on the basis of a protocol previously established in the laboratory (30). Briefly, A549 (ATCC) and 16HBE (a gift from Professor Dieter Gruenert at the University of California at San Francisco, San Francisco, CA) were cultured in standard conditions and collected for nuclear protein extraction using a NE-PER Nuclear and Cytoplasmic Extraction Kit commercial kit (Thermo Scientific). 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′-GGGCGGCGGCGCCACGACTCA-GAACACTCAACACGG-3′ (C allele) and 5′-GGGCGGCGGCGCCACGACTCAACACGG-3′ (A allele). The oligos were either end-labeled with biotin as probes or nonlabeled 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 with or without competitors (200×) at room temperature for 20 minutes. Complexes were then resolved on 4% acrylamide gels. After electrophoresis, DNA and RNA−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.). The assay was repeated multiple times and the representative result was presented.

Luciferase assays

A 217-bp DNA fragment spanning the human TERT rs2736100 region was generated using PCR from a heterozygous DNA sample using the following primers: 5′-CTCGGTACAGGTTTCACGCA-3′ and 5′-TTTGGCTACGAGTAGCTAGTCGCC-3′. A few nucleotides (underlined) were modified to create restriction enzyme digestion sites for KpnI and Nhel, respectively. PCR products were gel-purified with a Gel Extraction Kit (Qiagen). The fragment was first cloned into a pCR2.1 vector using a T-A Cloning Kit (Invitrogen) and sequenced thereafter to obtain the A and C allele fragments, respectively. The plasmids containing A or C alleles were then amplified, digested with KpnI and Nhel (New England Biolabs), 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-Luc luciferase plasmids were respectively cotransfected with the pCMV-beta-gal vector (Clontech) into the 16HBE and A549 cells with Lipofectamine 2000 Reagent (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. The β-galactosidase activity was used to normalize transfection efficiency. The 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 3 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 was examined using a χ2 test (CST), respectively, whereas the differences 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). ORs, 95% confidence intervals (CI) and P values were calculated. 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 the phenotypes by controlling age (constant), gender (binary), smoking history (binary, yes or no), and histology (binary, adenocarcinoma, or nonadenocarcinoma), 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 was conducted using a paired t test, whereas 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 and plotted using the GraphPad Prism 6.0.
Genotypic association between rs2736100 and NSCLC and subtypes

**Table 2.** Allelic association between rs2736100 and NSCLC and subtypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>All NSCLC vs. healthy controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>671 (47.8)</td>
<td>2,143 (42.5)</td>
<td>1.24 (1.10–1.39)</td>
<td>4 × 10⁻⁴</td>
</tr>
<tr>
<td>A</td>
<td>733 (52.2)</td>
<td>2,897 (57.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFRmut⁺ vs. healthy controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>315 (52.9)</td>
<td>2,143 (42.5)</td>
<td>1.52 (1.28–1.80)</td>
<td>1.6 × 10⁻⁶</td>
</tr>
<tr>
<td>A</td>
<td>279 (47.1)</td>
<td>2,897 (57.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFRmut⁺ vs. healthy controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>358 (44.1)</td>
<td>2,143 (42.5)</td>
<td>1.07 (0.92–1.24)</td>
<td>0.4</td>
</tr>
<tr>
<td>A</td>
<td>454 (55.9)</td>
<td>2,897 (57.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFRmut⁺ vs. EGFRmut⁻</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>315 (52.9)</td>
<td>358 (44.1)</td>
<td>1.42 (1.15–1.76)</td>
<td>1.1 × 10⁻³</td>
</tr>
<tr>
<td>A</td>
<td>279 (47.1)</td>
<td>454 (55.9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results

Association between demographic and clinical features and EGFR mutation

To test our hypothesis, we first detected mutations in EGFR in 714 Chinese patients with NSCLC. The mutation rate was 42.4% (303 of 714), which is concordant with other reports (3, 4, 13). We then compared the difference in the distribution of age, gender, smoking history, histologic 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 (OR, 1.52; 95% CI, 1.28–1.80; P = 1.6 × 10⁻⁴) 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 × 10⁻³; 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, 1.47–3.32; P = 7.4 × 10⁻²) for EGFRmut⁺ NSCLC compared with the nonsignificant 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 × 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 EGFRmut⁺ cancers (OR, 1.47; 95% CI, 1.08–1.99; P = 0.013) and no significant associations for the C/A heterozygotes for the EGFRmut⁻ 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.

![Table 3. Genotypic association between rs2736100 and NSCLC and subtypes](image-url)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>Corrected OR</th>
<th>Corrected P</th>
</tr>
</thead>
<tbody>
<tr>
<td>All NSCLC vs. healthy controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>555 (50.3%)</td>
<td>1,269 (50.4%)</td>
<td>1.19 (0.98–1.45)</td>
<td>0.08</td>
<td>0.68</td>
<td>0.45</td>
</tr>
<tr>
<td>C/A</td>
<td>353 (22.6%)</td>
<td>437 (17.3%)</td>
<td>1.56 (1.23–1.98)</td>
<td>3.0 × 10⁻⁴</td>
<td>1.29</td>
<td>1.09 × 10⁻⁴</td>
</tr>
<tr>
<td>A/A</td>
<td>190 (27.7%)</td>
<td>814 (32.3%)</td>
<td>Referent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFRmut⁺ vs. healthy controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>82 (27.7%)</td>
<td>437 (17.3%)</td>
<td>2.35 (1.66–3.32)</td>
<td>7.4 × 10⁻⁷</td>
<td>1.52</td>
<td>3.2 × 10⁻⁶</td>
</tr>
<tr>
<td>C/A</td>
<td>149 (50.3%)</td>
<td>1,269 (50.4%)</td>
<td>1.47 (1.08–1.99)</td>
<td>0.013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>65 (22.0%)</td>
<td>814 (32.3%)</td>
<td>Referent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFRmut⁺ vs. healthy controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>77 (19.0%)</td>
<td>437 (17.3%)</td>
<td>1.18 (0.84–1.56)</td>
<td>0.38</td>
<td>0.71</td>
<td>0.22</td>
</tr>
<tr>
<td>C/A</td>
<td>204 (50.2%)</td>
<td>1,269 (50.4%)</td>
<td>1.03 (0.81–1.30)</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>125 (30.8%)</td>
<td>814 (32.3%)</td>
<td>Referent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFRmut⁺ vs. EGFRmut⁻</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>82 (27.7%)</td>
<td>77 (19.0%)</td>
<td>2.05 (1.33–3.36)</td>
<td>1.1 × 10⁻³</td>
<td>1.30</td>
<td>0.035</td>
</tr>
<tr>
<td>C/A</td>
<td>149 (50.3%)</td>
<td>204 (50.2%)</td>
<td>1.41 (0.97–2.03)</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>65 (22.0%)</td>
<td>125 (30.8%)</td>
<td>Referent</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Association between rs2736100, mean LTL, and EGFR mutation status

Given the role of TERT in maintaining the telomere length, we further investigated the interrelationship between rs2736100, mean 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 patients with NSCLC as one population have significantly longer LTL than healthy controls (t test, P < 10−10), 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).

Interrelationship between rs2736100, TERT gene expression in normal lung and NSCLC

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 C). 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 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 2 specific protein–DNA complex bands formed, and more strongly with the A-allele probe as compared with the C allele. Notably, there was also a stronger

![Figure 1](https://www.aacrjournals.org/clin-cancer-research/article-pdf/21/22/5177/5188411/5177.pdf)

**Figure 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 = 10−15) and between EGFRmut+ and EGFRmut− (t test, corrected P = 0.043; B). Horizontal bars refer to the mean.

![Figure 2](https://www.aacrjournals.org/clin-cancer-research/article-pdf/21/22/5177/5188411/5177.pdf)

**Figure 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.
affinity between the rs2736100-A allele probe and the nuclear proteins extracted from 16HBE, as compared with that in A549 (Fig. 3A). 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 217-bp 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 with 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 with 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 suggest 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 patients with NSCLC is significantly longer than that of healthy controls (32). While the rs2736100-C allele has been associated with lung cancer risk, it was also consistently associated with longer LTL among general populations in multiple GWAS (24–27), including our recent GWAS in a large Han Chinese population (27). While our current study systematically confirmed these observations, we further observed that EGFRmut+ patients have even longer LTL than the EGFRmut− patients, suggesting that increased TERT activity is essential for EGFRmut+ NSCLC. Furthermore, previous studies demonstrated that ectopic expression of the TERT gene immortalized primary lung epithelial cells (33), whereas expression of EGFR mutants in TERT-immortalized lung epithelial cells led to cell transformation (7). These lines of evidence together suggest that an increased TERT activity attributed to genetic variation may lead to an elevated ability of cell proliferation and immortalization, which further provides a prerequisite condition for EGFR mutant–induced tumorigenesis. This potential dependence of the development of EGFRmut+ NSCLC on increased TERT function may have important clinical implications. For example, combination therapies of inhibiting both EGFR and TERT activity may be synergistic to NSCLC treatment. Indeed, additive effect has been previously observed in combinational RNAi treatment of these 2 genes in hepatocellular carcinoma cells (34).

Our data also suggest that rs2736100 may, at least in part, play a causal role in conferring lung cancer risk. The C allele of rs2736100 has been associated with increased risk for multiple cancers in several GWAS and follow-up meta-analyses (16, 17, 21, 22, 35–39). This allele, as opposed to other alleles in linkage disequilibrium, with its functions and increased cancer risk, remains incompletely understood. A previous bioinformatics analysis suggested that rs2736100 is located in a regulatory region of the human TERT gene (40). Our analyses indicated that while the DNA sequence flanking rs2736100 may play a role in enhancing TERT transcription, the C allele sequence has a significantly higher capacity for this regulation compared with the A allele. Interestingly, the A allele sequence demonstrated a specific affinity to nuclear proteins extracted from primary lung epithelia cells. Meanwhile, both the allelic affinity to nuclear proteins and regulation for transcription tended to be stronger in the primary lung epithelial cell (16HBE) than in a lung adenocarcinoma-derived cell line (A549). This suggests that the unknown nuclear protein(s) may act as a suppressor(s) in regulating the TERT transcription. Indeed, the C allele of rs2736100 was significantly associated with increased TERT mRNA expression in both normal and cancer lung cells. Of course, despite these significant observations, other polymorphisms within the region that are in linkage
discerning with rs2736100 may also play regulatory roles. Fully elucidating the molecular mechanism underlying the association between rs2736100 and various phenotypes requires a detailed screening and characterization for other common and rare variants across the entire region.

Our study further highlighted the importance of driver mutations in the classification of cancer subtypes. Cancer arises from accumulation of somatic mutations. Genome-wide sequencing of cancer genomes thus far has revealed a detailed landscape of various driver mutations during cancer development. Meanwhile, research in these driver mutations in both basic and clinical settings have suggested that cancer cells carrying the same driver mutations should be classified into the same subtype, as they share the same molecular cause and respond similarly to targeted therapy (41–43). Therefore, such a molecular-classified cancer subtype can have its unique genetic susceptibility. Indeed, previous studies have identified germline susceptibility alleles for a few somatic mutations—defined cancers, including a MCI R polymorphism and BRAF-mutant melanoma (44, 45), a JAK2 germline polymorphism, and the JAK2 V617F–mutant myeloproliferative neoplasms (46) and an 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 with 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 on the basis of histologic 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 patients with NSCLC 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 2 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 shed 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.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: R. Wei, H. Pu, X. Niu, Y. Zeng, C.I. Amos, S. Lu, H.-Y. Wang, Y. Liu, W. Liu

Development of methodology: R. Wei, L. Cao, H. Pu, H. Wang, X. Niu, S. Lu, Y. Liu

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Cao, H. Pu, H. Wang, X. Niu, M.J. Favus, I. Zhang, W. Jia, S. Lu, H.-Y. Wang, Y. Liu, W. Liu

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Wei, L. Cao, H. Pu, Y. Zheng, X. Niu, X. Weng, C.I. Amos, S. Lu, W. Liu

Writing, review, and/or revision of the manuscript: R. Wei, L. Cao, H. Pu, Y. Zheng, X. Niu, M.J. Favus, C.I. Amos, S. Lu, H.-Y. Wang, Y. Liu, W. Liu

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R. Wei, L. Cao, H. Pu, X. Niu, S. Lu, W. Liu

Study supervision: H. Pu, X. Niu, S. Lu, Y. Liu, W. Liu

Grant Support

This study was supported in part by the American Cancer Society-II Division (grant 189273; to W. Liu), a start-up fund of the College of Pharmacy, Purdue University (W. Liu); and the Research Fund of National Laboratory of Oncology in South China (H. Wang), the National Natural Science Foundation of China (NSFC; grant 61302005), the Project of Shanghai Municipality Science & Technology Commission (grant 15ZR1438500), Youth Foundation of Shanghai Municipal Public Health Bureau (grant 20124Y114), and the Sino-Swiss Lung Cancer Clinic Center joint translational medicine research (grant 2012DFC31320).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 6, 2015; revised May 18, 2015; accepted June 11, 2015; published OnlineFirst July 6, 2015.

References


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

Rongrong Wei, Lan Cao, Hengying Pu, et al.


Updated version
Access the most recent version of this article at:

Cited articles
This article cites 47 articles, 14 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/21/22/5173.full.html#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.