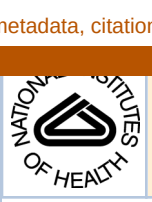



Published in final edited form as:

*Cancer*. 2011 February 1; 117(3): 606–617. doi:10.1002/cncr.25472.

# Methylation of the Candidate Biomarker *TCF21* Is Very Frequent Across A Spectrum of Early Stage Non-Small Cell Lung Cancers

**Kristy L. Richards, PhD, MD**[1,2,3], **Baili Zhang, MS**[1], **Menghong Sun, PhD**[4], **Wenli Dong, PhD**[5], **Jennifer Churchill, BS**[1], **Linda L. Bachinski, PhD**[1], **Charmaine D. Wilson, MS**[1], **Keith A. Baggerly, PhD**[6], **Guosheng Yin, PhD**[5], **D. Neil Hayes, MD, MPH**[2], **Ignacio I. Wistuba, MD**[4,5], and **Ralf Krahe, PhD**[1,7,*]

[1]Department of Genetics, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

[2]Department of Medicine, The Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA

[3]Department of Genetics, The Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA

[4]Department of Pathology, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

[5]Department of Biostatistics, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

[6]Department of Bioinformatics and Computational Biology, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

[7]Department of Thoracic/Head and Neck Medical Oncology, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

## Abstract

**Background—**The transcription factor *TCF21* is involved in mesenchymal-to-epithelial differentiation and was shown to be aberrantly hypermethylated in lung and head and neck cancers. Because of its reported high frequency of hypermethylation in lung cancer, we sought to characterize the stages and types of non-small cell lung cancer (NSCLC) that are hypermethylated and to define the frequency of hypermethylation and associated “second hits”.

**Methods—**We determined *TCF21* promoter hypermethylation in 105 NSCLC including various stages and histologies in smokers and nonsmokers. Additionally, we examined *TCF21* loss-of-heterozygosity and mutational status. We also assayed 22 cancer cell lines from varied tissue origins. We validated and expanded our NSCLC results by examining TCF21 immunohistochemical expression on a tissue microarray containing 300 NSCLC cases.

**Results—**Overall, 81% of NSCLC samples showed *TCF21* promoter hypermethylation and 84% showed decreased TCF21 protein expression. Multivariate analysis showed that TCF21 expression, although below normal in both histologies, was lower in adenocarcinoma than squamous cell carcinoma, and was not independently correlated with gender, smoking and EGFR mutation status, or clinical outcome. Cell lines from other cancer types also showed frequent *TCF21* promoter hypermethylation.

*Corresponding author: Ralf Krahe, PhD, Department of Genetics, Unit 1010, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030; USA; Phone: +1-713-834-6345; Fax: +1-713-834-6319; rkrahe@mdanderson.org.

**Conflict of Interest Disclosures**: None declared.

**Conclusions—**Hypermethylation and decreased expression of *TCF21* were tumor-specific and very frequent in all NSCLC, even early-stage disease, thus making *TCF21* a potential candidate methylation biomarker for early-stage NSCLC screening. *TCF21* hypermethylation in a variety of tumor cell lines suggests it may also be a valuable methylation biomarker in other tumor types.

### Keywords

TCF21; methylation; biomarker; lung cancer; screening

## Introduction

Lung cancer is the number one cause of cancer mortality worldwide, and kills more people than breast, colon, and prostate cancer combined.[1] Unlike these other common cancers, however, there is no effective screening strategy to detect early-stage lung cancer at a time when surgery may be curative. The need for such a strategy is obvious, and many attempts to detect lung cancer early have so far failed to show clinical benefit.[2-4] These include screening CT scans, sputum cytology, screening chest X-rays, serum markers.

Recently, promoter hypermethylation has been recognized as an important mechanism by which genes regulating cellular proliferation are silenced during cancer development[5, 6] Promoter hypermethylation involves DNA methylation of CpG islands in or near the promoter region of certain genes, rendering them transcriptionally silent. This downregulation of gene expression of important cellular growth control genes has been shown to be important for cancer progression and outcome, with poorer outcomes associated with promoter hypermethylation of such important genes as *RASSF1A*, *RARB*, and *HIF1*.[7-9]

*TCF21* is a recently recognized target of aberrant promoter hypermethylation in cancer, discovered in a genomic screen for regions of DNA that are hypermethylated in cancer.[10] It was reported to be frequently hypermethylated in head and neck and lung cancer, and restoration of *TCF21* expression inhibited tumor growth, both in a lung cancer cell line and in a mouse xenograft model. *TCF21* is widely expressed; its normal function is to promote mesenchymal transition into epithelial cells.[11] Reversal of this process, known as the epithelial-to-mesenchymal transition (EMT), has been implicated in tumor invasion and metastasis;[12, 13] Therefore, silencing of *TCF21* may be a mechanism for tumor cells to gain these aggressive characteristics during the course of tumor progression. Given that *TCF21* was reported to be frequently hypermethylated and silenced in NSCLC, as well as its plausible biologic role in tumor progression, we sought to more precisely define the frequency of *TCF21* promoter hypermethylation in NSCLC. We were especially interested in defining its frequency among different cancer stages and histologic subtypes. Here, we show that *TCF21* is very frequently hypermethylated in a variety of NSCLC, and that protein expression of TCF21 is also very frequently reduced, either of which could be used for screening and/or diagnostic purposes as a biomarker of early disease.

## Materials and Methods

### Frozen Tumor Specimens, Cell Lines, and DNA Extraction

Patient NSCLC specimens were obtained from surgical specimens at both the University of Texas M. D. Anderson Cancer Center (42 matched tumor/normal samples, 7 unpaired tumor samples) as well as from the University of North Carolina Lineberger Comprehensive Cancer Center tumor bank of surgical specimens (56 unpaired tumor samples). In both institutions, informed consent was obtained prior to surgery for the use of specimens as part of an IRB-approved protocol, in accord with the Helsinki Declaration. Tissue was snap-frozen and used for later DNA extraction. Genomic DNA was extracted from the DNA-

protein phase of TriZol-extracted tissues according to the manufacturer's suggestions (Invitrogen). DNA was extracted using the PureGene kit (Gentra) on cell pellets from four HNSCC cell lines (SCC-4, SCC-9, SCC-15 and SCC-25), five lung cancer cell lines (H1395, H520, H2170, SK-MES-1 and SW-900), one breast cancer cell line (MCF7), one cervical cancer cell line (HeLa), two brain cancer cell lines (SK-N-AS and M059K), one uterine cancer cell line (AN3CA), one sarcoma cell line (HT1080), one kidney cancer cell line (HEK293), and six colon cancer cell lines (LoVo, SW48, HCT-15, DLD-1, COLO 320DM and RKO) according to the manufacturer's suggestions. All cell lines are available from ATCC (Manassas, VA). Four normal pools, each comprised of DNA from peripheral blood mononuclear cells (PBMCs) of six individuals were generated representing different genders and ages (females ≤40 yrs of age, females age >40 yrs, males ≤40 yrs and males ≥40 yrs).

### *TCF21* Promoter Methylation

PCR and sequencing primers were designed using the PSQ Assay Design software (Qiagen). PCR was performed in a 25 μl reactions containing Qiagen HotStart Taq master mix (Qiagen) using 1 μl bisulfate-converted DNA (about 10 ng/μl). Bisulfite conversion of genomic DNA was performed as previously reported.[14] Briefly, 0.5-1.0 μg of genomic DNA was treated using the EZ-96 DNA Methylation Gold Kit (Zymo Research), including DNA sulfonation, deamination, desalting, desulfonation and recovery. Bisulfite-treated DNA was stored at −80°C until use. To reduce the cost per assay, an amplification protocol was developed using a biotinylated universal primer approach.[14] Final primer concentrations were 10 nM of the reverse primer tailed with the universal primer (5′-GACGGGACACCGCTGATCGTTTACCAAAAAAAAACCCCCTAA-3′), 100 nM of the untailed forward primer (5′-GGTAGGGTGGTTTTGAGTT-3′), and 90 nM of the universal biotinylated primer (5′-GGGACACCGCTGATCGTTTA-3′) in each reaction. The universal primer sequence is underlined. The predicted amplicon size was 153 bp. Amplification was carried out as follows: denaturation at 95°C for 5 min, followed by 50 cycles at 95°C for 30 sec, 51°C for 1 min, 72°C for 45 sec, and a final extension at 72°C for 7 min.

Following PCR amplification, Pyrosequencing was performed on a PSQ96HS system (Qiagen) according to the manufacturer's protocol including the use of single strand binding protein (PyroGold reagents). The Pyrosequencing primer was (5′-TTGAGTTTGGAGAAGG-3′). The results were analyzed using Q-CpG software (Qiagen), which calculates the methylation percentage ($^{m}C/(^{m}C+C)$) for each CpG site, allowing quantitative comparisons. The methylation index (MI) was calculated as the average value of $^{m}C/(^{m}C+C)$ for all nine of the interrogated CpG sites in the assay. Genomic DNA treated with M.*Sss*I (New England Biolabs) was used as a universally methylated positive control; the same untreated genomic DNA amplified by whole genome amplification (GenomiPhi, GE Healthcare) was used as a universally unmethylated negative control.

### Decitabine Treatment and Quantitative Real-Time RT-PCR

Three colon cancer cell lines (DLD-1, HCT-15 and RKO) with high levels (>85%) of *TCF21* promoter hypermethylation were plated at a density of 500,000 cells/T75 flask. DLD-1 and HCT-15 cells were grown in RPMI-1640 supplemented with 10% FBS and 1% penicillin/streptomycin, RKO cells in EMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Drug treatment with 1 μM decitabine (Sigma-Aldrich, St. Louis, MO) was started 3 hrs after seeding. Culture medium and drug were changed daily for treated and untreated cells. Cultures were grown for a minimum of four days until 80% confluency. Total cellular RNA was isolated using TRIZol reagent (Invitrogen). Input RNA (1 μg) was reverse-transcribed using the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA). TCF21 expression was assessed by TaqMan qRT-PCR using assays Hs00162646_m1

and Hs01546814_m1 (Applied Biosystems, Foster City, CA, USA) covering exon 1-2 and exon 2-3, respectively. qRT-PCR was carried out as follows in 20-μl final reaction volume using 55 ng of RNA-equivalents as cDNA input: initial denaturation at 95°C for 8.5 min, followed by 45 cycles at 95°C for 15 sec and 60°C for 1 min according to the manufacturer's suggestions. GUSB (Hs99999908_m1) was used as endogenous housekeeping gene control for normalization. Each assay was performed in triplicate. Relative expression levels were calculated using the $\Delta\Delta C_t$ method and scaled.

## LOH and Mutation Detection

Primers were designed for detection of four microsatellites within and flanking *TCF21*. Primer sequences are shown in Table 1. All forward primers were 5′-tailed with 5′-GACGGGACACCGCTGATCGTTTA-3′ and all reverse primers were 5′-tailed with 5′-GTTTCTT-3′. A universal primer with the sequence 5′-GGGACACCGCTGATCGTTTA-3′ end-labeled with either FAM, HEX, or NED was used in all microsatellite amplifications. PCR conditions for the three primer reactions were as described above for amplification using the universal biotinylated primer. Amplification products were pooled as appropriate and analyzed by capillary electrophoresis on an ABI 3100 Genetic Analyzer (Applied Biosystems).

The coding region of *TCF21* (exons 1 and 2) was sequenced in both directions in four fragments. In all, 45 lung cancer samples showing zero or one hit were sequenced. Samples which had already been scored as having two hits were not sequenced. Primer sequences are shown in Table 1. All forward primers were 5′-tailed with M13 forward sequence 5-TGTAAAACGACGGCCAGT-3′, and all reverse primers with M13 reverse 5′-CAGGAAACAGCTATGACC-3′. After amplification, samples were treated with Exo-SAP (Amersham), sequenced using Big Dye Terminator v3.1 (Applied Biosystems) under standard conditions and products purified by ethanol precipitation, dehydrated in a vacuum centrifuge, and resuspended in 20 μl formamide before capillary electrophoresis on an ABI 3100 Genetic Analyzer. Sequences were aligned and visualized using Sequencher software (Gene Codes). Fragment 1 contained a polymorphic $(CT)_n$ simple tandem repeat of 8 to 12 units, which, when polymorphic, was used to confirm retention-of-heterozygosity identified by the microsatellites.

## Archival NSCLC Case Selection and Tissue Microarray Construction

We obtained archival, formalin-fixed and paraffin-embedded (FFPE) material from surgically resected lung cancer specimens containing tumor and adjacent lung tissues from the Lung Cancer Specialized Program of Research Excellence (SPORE) Tissue Bank at The University of Texas M. D. Anderson Cancer Center, which was approved by the Institutional Review Board. Tumor tissue specimens from 300 NSCLCs (191 adenocarcinomas, and 109 squamous cell carcinomas) were histologically examined, classified using the 2004 World Health Organization (WHO) classification system,[15] and selected for tissue microarray (TMA) construction. After histologic examination, TMAs were constructed using triplicate 1-mm diameter cores from each tumor. Detailed clinical and pathological information, including demographic data, smoking history (never- and ever-smokers) and status (never, former, and current smokers), pathologic TNM staging,[16] overall survival, and time of recurrence, was available in most cases (Table 2). Patients who had smoked at least 100 cigarettes in their lifetime were defined as smokers, and smokers who quit smoking at least 12 months before lung cancer diagnosis were defined as former smokers.

### Immunohistochemical Staining and Evaluation

An anti-human TCF21 antibody was used for immunostaining (ab32981, Abcam). FFPE tissue histology sections (5-μm thick) were deparaffinized, hydrated, heated in a steamer for 10 min with 10 mM sodium citrate (pH 6.0) for antigen retrieval. Peroxide blocking was performed with 3% $H_2O_2$ in methanol at room temperature for 15 min, followed by 10% bovine serum albumin in TBS-t for 30 min. Slides were incubated with primary antibody at 1:200 dilution for 65 minutes at room temperature. After washing with TBS-t, incubation with biotin-labelled secondary antibody for 30 min followed. Finally, samples were incubated with a 1:40 solution of streptavidin-peroxidase for 30 min. The staining was then developed with 0.05% 3′,3-diaminobenzidine tetrahydrochloride prepared in 0.05 mol/l Tris buffer at pH 7.6 containing 0.024% $H_2O_2$ and counterstained with hematoxylin. FFPE lung tissues having normal bronchial epithelia were used as positive control. For a negative control, we used the same specimens used for the positive controls, replacing the primary antibody with PBS.

TCF21 immunostaining was detected in the cytoplasm of epithelial and tumor cells. Immunohistochemical expression was quantified by microscope observation by two pathologists (M.S. and I.W.) using a four-value intensity score (0, 1+, 2+ and 3+) and the percentage of the reactivity extent. A final score was obtained by multiplying both intensity and extension values (range 0-300), and four levels of expression were arbitrarily calculated based on that score: (a) negative (score 0-9); (b) low (score 10-100); (c) intermediate (score 100 to 199); and (d) and high (score 200-300). Levels and scores were used for analysis.

### *EGFR* Mutation Analysis

Exons 18 through 21 of *EGFR* were PCR amplified using intron-based primers as previously described.[17, 18] From microdissected FFPE cells, ~200 cells were used for each PCR amplification. All PCR products were directly sequenced using the PRISM dye-terminator cycle sequencing method (Applied Biosystems). All sequence variants were confirmed by independent PCR amplifications from at least two independent microdissections and DNA extraction, and sequenced in both directions, as previously reported.

### Statistical Analysis

The clinical and pathological data were summarized using descriptive statistics and frequency tabulations. Wilcoxon rank-sum and Kruskal-Wallis tests were used to compare biomarker expression among different prognostic factor levels. The generalized linear model was used to assess the effect of prognostic factors on TCF21 expression in the multivariable setting. Fisher's exact test was used to compare the association between categorical variables. We examined the association between overall survival (OS) and recurrence-free survival (RFS) rates and TCF21 expression in NSCLC patients with stage I or II disease, who had not undergone adjuvant chemotherapy. OS was defined as the time from surgery to death or the end of the study; RFS was defined as the time from surgery to recurrence or the end of the study. Univariate and multivariate Cox proportional hazards models were used to assess the effects of TCF21 protein expression on survivals. Two-sided $p$-values$<0.05$ were considered statistically significant. All analyses were conducted using SAS (v 9.1, Cary, NC) and S-plus (v 8.0, Seattle, WA) software.

## Results

### *TCF21* is Highly Methylated in Nearly All Cancer Cell Lines

To characterize *TCF21* methylation levels in normal and malignant states, we examined various cancer cell lines from a spectrum of tissue types (brain, breast, cervix, colon, connective tissue, head and neck, kidney, lung, and uterus). We also assayed *TCF21*

methylation in normal PBMCs from younger and older individuals of both sexes, since methylation levels can be influenced by age and/or sex. Universally methylated control DNA and genetically matched unmethylated control DNA defined the boundaries of detection of our assay (3-93% methylation). Using Pyrosequencing-based Methylation Analysis (PMA) we analyzed *TCF21* methylation by averaging methylation levels of nine promoter CpG sites. All but one cell line (SK-N-AS, a neuroblastoma cell line, 38%) was highly methylated, with levels at or approaching the upper limit of detection (Fig. 1). Normal PBMCs were essentially identical regardless of age or gender, and demonstrated moderate levels of baseline methylation at ~20%.

### *TCF21* is Hypermethylated in >80% of NSCLC

To define the threshold for hypermethylation positivity, we began our analysis using genetically matched NSCLC and adjacent normal tissue pairs from the same patient. To assess the baseline levels of *TCF21* methylation in lung tissue, we examined both normal adjacent tissue (NAT) from the tumor/normal (T/N) pairs (n=42) comparing them to PBMC. Average methylation levels in NAT were 21.5% (SD=4.6; n=42), and in the normal PBMC 20.1%. Average *TCF21* methylation levels in T samples were 41.3% (SD=11.6; n=42) (Fig. 2A). The difference between the average methylation levels in N and T tissues was highly significant ($p$-value$<1\times10^{-13}$).

Using a threshold of 30% methylation, we found that 37 of 42 tumors (88%) were hypermethylated, while 41 of 42 matched normal samples (98%) were not. Using this cutoff to define hypermethylation, we then assayed second set of 63 unpaired NSCLC samples. This second set of tumors contained a small number of large cell histologic subtypes, and some mixed histologic types (mostly adeno-squamous). We found that 48 (76%) of them were hypermethylated (Fig. 2B). Overall, the average methylation levels of all the tumor samples combined was 39.2% (SD=11.7; n=105). Using the threshold of 30% methylation, the overall frequency of hypermethylation in NSCLC was 81% (85/105).

### Reactivation of *TCF21* Expression by Demethylating Agent

To show that *TCF21* promoter hypermethylation correlates with transcriptional silencing of the gene, we treated three colorectal cancer cell lines with high methylation levels (>85%) with the demethylating drug decitabine. We performed quantitative TaqMan mRNA real-time PCR to determine relative *TCF21* expression levels with and without treatment. For all three cell lines culturing in the presence of the demethylating agent led to reactivation of *TCF21* expression at the mRNA level as assayed by two distinct quantitative real-time PCR assays (Fig. 3).

### Reduced Expression of TCF21 Protein in NSCLC

To determine whether *TCF21* promoter hypermethylation also resulted in decreased TCF21 protein expression, we used a NSCLC TMA containing tumor samples from 300 patients. The microarray was stained with a TCF21 antibody, and protein levels were scored as none, low, intermediate, or high (Fig. 3A). While normal adjacent lung tissue stained strongly for TCF21, 253 of 300 (84%) NSCLC samples showed reduced (either low or none) staining (Fig. 3B).

Similar frequencies of *TCF21* hypermethylation and decreased protein expression suggested that hypermethylation leads to reduced protein levels, which would be consistent with previously reported decreased mRNA levels resulting from *TCF21* promoter hypermethylation.[10] Because our TMA included only 9 overlapping samples between the TMA and *TCF21* methylation sets, we assembled a smaller TMA with 31 samples overlapping (Table 3). Interestingly, *TCF21* hypermethylation and reduced TCF21 protein

expression were sometimes discordant (Table 3), suggesting that mechanisms other than hypermethylation could result in decreased protein expression.

### *TCF21* Loss-of-Heterozygosity and Sequence Analysis

Because some NSCLC samples showed loss of TCF21 protein expression without hypermethylation and the average levels of *TCF21* hypermethylation were ~40%, which might not be expected to completely abolish protein expression, we examined potential "second hits" at the *TCF21* locus (Table 3). First, we examined loss-of-heterozygosity (LOH) in 33 of the paired samples, using four microsatellite markers spanning the *TCF21* locus and closely flanking region. LOH was seen in 14 (42%) of these samples, with no significant differences comparing samples with and without hypermethylation (*p*-value=0.172). In addition to LOH, we sequenced the *TCF21* coding region in 45 lung cancer samples that showed either zero or one hit by methylation or LOH analysis. Samples with both hypermethylation and LOH were not sequenced. No *TCF21* coding mutations were found.

### Reduced TCF21 Protein Expression is Widespread and Independent of Stage and Other Clinical Features, but Correlated with Histology

To determine whether TCF21 expression was correlated with clinical features such as gender, race, stage, smoking status, histology, or prognosis, we performed univariate analysis (Table 2). Histology and TCF21 expression showed significant correlation (*p*-value=0.003), as did smoking status (*p*-value=0.048) and gender (*p*-value=0.021). In a multivariate analysis with histology, gender, and smoking status, only histology was statistically significantly (*p*-value = 0.007) associated with TCF21 levels, while smoking history and gender were not independently associated. Cox proportional hazards analysis was performed to assess association between TCF21 and overall survival and recurrence-free survival, but neither association was significant, in either a multivariate model or a univariate model (data not shown).

Given previously reported associations between smoking, gender, and histology with *EGFR* status,[19] we then analyzed the 202 patient subset for which *EGFR* status was known, for associations with TCF21 expression. When only adenocarcinomas were considered (n = 172), *EGFR* status was not associated with TCF21 expression level (*p*-value = 0.138), nor was *EGFR* status associated with TCF21 expression level in a univariate analysis with all 202 patients (*p*-value = 0.241). Therefore, the only significant correlation (*p*-value = 0.007) is that adenocarcinomas have lower levels of TCF21 expression than SCCs, although all histologies have significantly lower TCF21 levels than those in normal tissue.

## Discussion

### *TCF21* Has the Highest Frequency of Promoter Hypermethylation in NSCLC of Any Gene Known to Date

Many genes have been reported to be hypermethylated in NSCLC.[20-24] However, the frequency of these events has not been high enough in all NSCLC subtypes for utilization as a screening tool, requiring combinations of genes to approach a sensitivity high enough for a screening test. Despite numerous reports of hypermethylated genes in NSCLC, identified by a variety of approaches, none has a reported frequency of hypermethylation as high as *TCF21*, except one that also examined *TCF21* itself, and a recent publication limited to only the SCC subtype of NSCLC.[20-23] This study was specifically focused on *TCF21* in NSCLC and the susceptibility locus at 6q23-q25. Among 43 genes selected in the region, *TCF21* had the highest rates of cancer-specific hypermethylation (81%),[23] exactly matching our rates of *TCF21* hypermethylation.

The high rates (80-85%) of *TCF21* promoter hypermethylation and decreased protein expression are high enough for *TCF21* to be considered for development as a screening biomarker, either by increased methylation or decreased protein levels. The sensitivity of *TCF21* hypermethylation/decreased TCF21 protein expression compares favorably with that of prostate-specific antigen (PSA), the current screening biomarker for prostate cancer, which has been shown to be <4 (i.e., in the normal range) in 15% of men with prostate cancer, a sensitivity of 85%.[25] Of course, one of the main difficulties in lung cancer screening remains in the acquisition of relevant tissue (in this case early lung tumors), but detection of *TCF21* hypermethylation has been reported in biopsies and sputum samples, which is promising.[26] If its sensitivity in sputum/bronchial brushings were not high enough to be used alone, *TCF21* could be used as part of a panel of screening biomarkers.

### Detection of *TCF21* Hypermethylation by Highly Quantitative Method

One significant advantage of methylation detection by Pyrosequencing-based Methylation Analysis (PMA) following bisulfite conversion is that quantitative levels can be measured across multiple sites, rather than the more qualitative output obtained with methylation-specific PCR (MS-PCR) or other qualitative or semiquantitative methods (e.g., COBRA). PMA enabled us to reliably detect a difference between the 20% average methylation in N tissue, and 40% average methylation in T tissue. This difference would likely not have been detected with less quantitative methylation detection strategies. It is possible that other genes known to be hypermethylated in NSCLC may prove to be more sensitive and/or specific, if more quantitative methods such as Pyrosequencing are routinely applied. The 40% methylation levels in NSCLC tissue raises the question of whether only one of the two *TCF21* alleles is silenced by hypermethylation, or whether 40% of cells have both alleles silenced, either of which could produce the observed result. It is interesting that hypermethylation of 40% of alleles is frequently associated with completely absent TCF21 protein expression, suggesting either that the second allele is silenced by a different mechanism than hypermethylation, or that there is a threshold level of gene expression necessary to produce detectable TCF21 protein levels.

### Reduction of TCF21 Protein Levels Similar to *TCF21* Hypermethylation Rates

In addition to *TCF21* hypermethylation, we also examined the downstream effect of this hypermethylation by examining protein expression directly. In both cases we found *TCF21* hypermethylation/decreased TCF21 protein levels at similar rates--81% and 84%, respectively. Given that decreased mRNA expression of *TCF21* has been shown to result from promoter hypermethylation,[10] the similar rates of hypermethylation and decreased protein expression are consistent with the notion that decreased mRNA expression results in decreased protein expression. However, since there were cases with low/absent protein expression despite normal *TCF21* methylation levels, other regulatory mechanisms likely are in effect. LOH occurs at a rate of 42%. Since LOH occurs in at least a few cases without *TCF21* hypermethylation, this implies inactivation of *TCF21* in other ways. Since we did not detect any coding mutations, these could be promoter or other regulatory region DNA mutations. Alternatively, dysregulation by micro-RNAs could be a factor. Interestingly, the sole predicted regulator of *TCF21* is miR-92a,[27] which is overexpressed in a variety of cancers.[28, 29]

### TCF21 is an Excellent Candidate Biomarker for Early Lung Cancer Detection

Several characteristics of *TCF21* make it an attractive target for screening efforts in NSCLC. First, it is hypermethylated at similar frequencies in all histologic subtypes of NSCLC examined, including early- and late-stage cancers. Second, it has a higher frequency of hypermethylation than any gene published to date in NSCLC, without subdivision by histologic subtype.[10, 20-23] This high sensitivity is combined with a high specificity as well.

We detected a false-positive rate of only 1 in 42 samples with NAT, for a specificity of 98%. In other reported control tissues, such as PBMCs and human bronchial epithelial cells (HBECs) from smokers, there were no false-positives (n=20 in each case).[23] The high specificity in normal adjacent tissue is especially noteworthy in that there appears to be no evidence for a "field-effect", which can complicate screening in smokers who often have cancers arising in a field of premalignant lesions, leading to false-positive screening results. Instead, the very low prevalence of *TCF21* hypermethylation in NAT that we report suggests that *TCF21* hypermethylation is restricted to cancerous tissue only.

In summary, we have established that *TCF21* hypermethylation and reduced TCF21 protein are ubiquitous in NSCLC, occurring in 80-85% of tumors across a wide variety of stages, histologies, and other clinical characteristics. Given the high rate of increased methylation and decreased protein expression, combined with their lack in normal adjacent tissue, we propose that *TCF21* is an excellent candidate biomarker for further development as a lung cancer screening tool.

## Acknowledgments

We would like to thank Tamer Ahmed for technical assistance and Mario Sirito for helpful discussions.

**Sources of Support:** This work was supported in part by the Kleberg Foundation, DoD W81XWH-05-2-0027, and NIH-NCI P01 CA34936 to RK. These agencies had no involvement in the study design, in the collection, analysis and interpretation of data, in writing of the manuscript and the decision to submit the manuscript for publication.

## References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin. 2005; 55:74–108. [PubMed: 15761078]
2. Humphrey LL, Teutsch S, Johnson M. Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: an update for the U.S. Preventive Services Task Force. Ann Intern Med. 2004; 140:740–753. [PubMed: 15126259]
3. Rossi A, Maione P, Colantuoni G, et al. Screening for lung cancer: New horizons? Crit Rev Oncol Hematol. 2005; 56:311–320. [PubMed: 15978829]
4. Chorostowska-Wynimko J, Szpechcinski A. The impact of genetic markers on the diagnosis of lung cancer: a current perspective. J Thorac Oncol. 2007; 2:1044–1051. [PubMed: 17975498]
5. Vucic EA, Brown CJ, Lam WL. Epigenetics of cancer progression. Pharmacogenomics. 2008; 9:215–234. [PubMed: 18370850]
6. Duffy MJ, Napieralski R, Martens JW, et al. Methylated genes as new cancer biomarkers. Eur J Cancer. 2009; 45:335–346. [PubMed: 19138839]
7. Kim YT, Park SJ, Lee SH, et al. Prognostic implication of aberrant promoter hypermethylation of CpG islands in adenocarcinoma of the lung. J Thorac Cardiovasc Surg. 2005; 130:1378. [PubMed: 16256792]
8. Maruyama R, Sugio K, Yoshino I, Maehara Y, Gazdar AF. Hypermethylation of FHIT as a prognostic marker in nonsmall cell lung carcinoma. Cancer. 2004; 100:1472–1477. [PubMed: 15042681]
9. Safar AM, Spencer H 3rd, Su X, et al. Methylation profiling of archived non-small cell lung cancer: a promising prognostic system. Clin Cancer Res. 2005; 11:4400–4405. [PubMed: 15958624]
10. Smith LT, Lin M, Brena RM, et al. Epigenetic regulation of the tumor suppressor gene TCF21 on 6q23-q24 in lung and head and neck cancer. Proc Natl Acad Sci USA. 2006; 103:982–987. [PubMed: 16415157]
11. Quaggin SE, Schwartz L, Cui S, et al. The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. Development. 1999; 126:5771–5783. [PubMed: 10572052]

12. Baum B, Settleman J, Quinlan MP. Transitions between epithelial and mesenchymal states in development and disease. Semin Cell Dev Biol. 2008; 19:294–308. [PubMed: 18343170]
13. Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. J Cell Biol. 2006; 172:973–981. [PubMed: 16567498]
14. Colella S, Shen L, Baggerly KA, Issa JP, Krahe R. Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. Biotechniques. 2003; 35:146–150. [PubMed: 12866414]
15. Travis, WD.; Brambilla, E.; Muller-Hermelink, HK.; Harris, CC. Tumours of the lung. In: Travis, WD.; Brambilla, E.; Muller-Hermelink, HK.; Harris, CC., editors. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart, World Health Orgnaziation Classification of Tumours. Lyon: International Agency for Research on Cancer (IARC); 2004. p. 9-124.
16. Mountain CF. Revisions in the International System for Staging Lung Cancer. Chest. 1997; 111:1710–1717. [PubMed: 9187198]
17. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst. 2005; 97:339–346. [PubMed: 15741570]
18. Tang X, Shigematsu H, Bekele BN, et al. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. Cancer Res. 2005; 65:7568–7572. [PubMed: 16140919]
19. Kris MG, Natale RB, Herbst RS, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. JAMA. 2003; 290:2149–2158. [PubMed: 14570950]
20. Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. Cancer Res. 2001; 61:3225–3229. [PubMed: 11309270]
21. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer. 2004; 4:707–717. [PubMed: 15343277]
22. Hoque MO, Kim MS, Ostrow KL, et al. Genome-wide promoter analysis uncovers portions of the cancer methylome. Cancer Res. 2008; 68:2661–2670. [PubMed: 18413733]
23. Tessema M, Willink R, Do K, et al. Promoter methylation of genes in and around the candidate lung cancer susceptibility locus 6q23-25. Cancer Res. 2008; 68:1707–1714. [PubMed: 18339850]
24. Anglim PP, Galler JS, Koss MN, et al. Identification of a panel of sensitive and specific DNA methylation markers for squamous cell lung cancer. Mol Cancer. 2008; 7:62. [PubMed: 18616821]
25. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med. 2004; 350:2239–2246. [PubMed: 15163773]
26. Shivapurkar N, Stastny V, Xie Y, et al. Differential methylation of a short CpG-rich sequence within exon 1 of TCF21 gene: a promising cancer biomarker assay. Cancer Epidemiol Biomarkers Prev. 2008; 17:995–1000. [PubMed: 18398044]
27. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006; 34:D140–144. [PubMed: 16381832]
28. Matsubara H, Takeuchi T, Nishikawa E, et al. Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92. Oncogene. 2007; 26:6099–6105. [PubMed: 17384677]
29. Volinia S, Calin GA, Liu CG, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci USA. 2006; 103:2257–2261. [PubMed: 16461460]

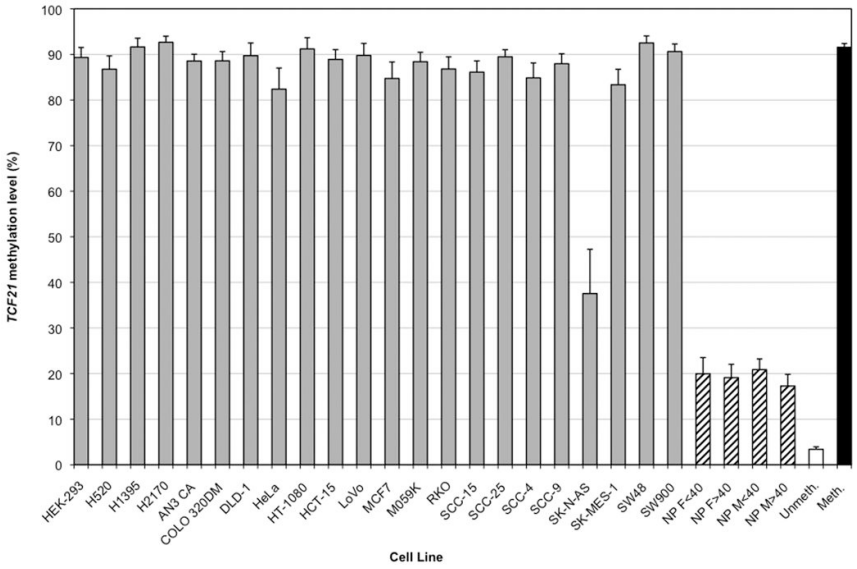


**Figure 1.**

Methylation levels of individual cancer cell lines, normal PBMCs, and positive and negative methylation controls. *TCF21* promoter methylation levels are shown for 22 cancer cell lines, four pools of PBMCs of different sex (male, female) and age (≤40, >40 yrs). Control samples, fully methylated by treatment with *Sss*I methylase or fully unmethylated by whole genome amplification, are also shown.

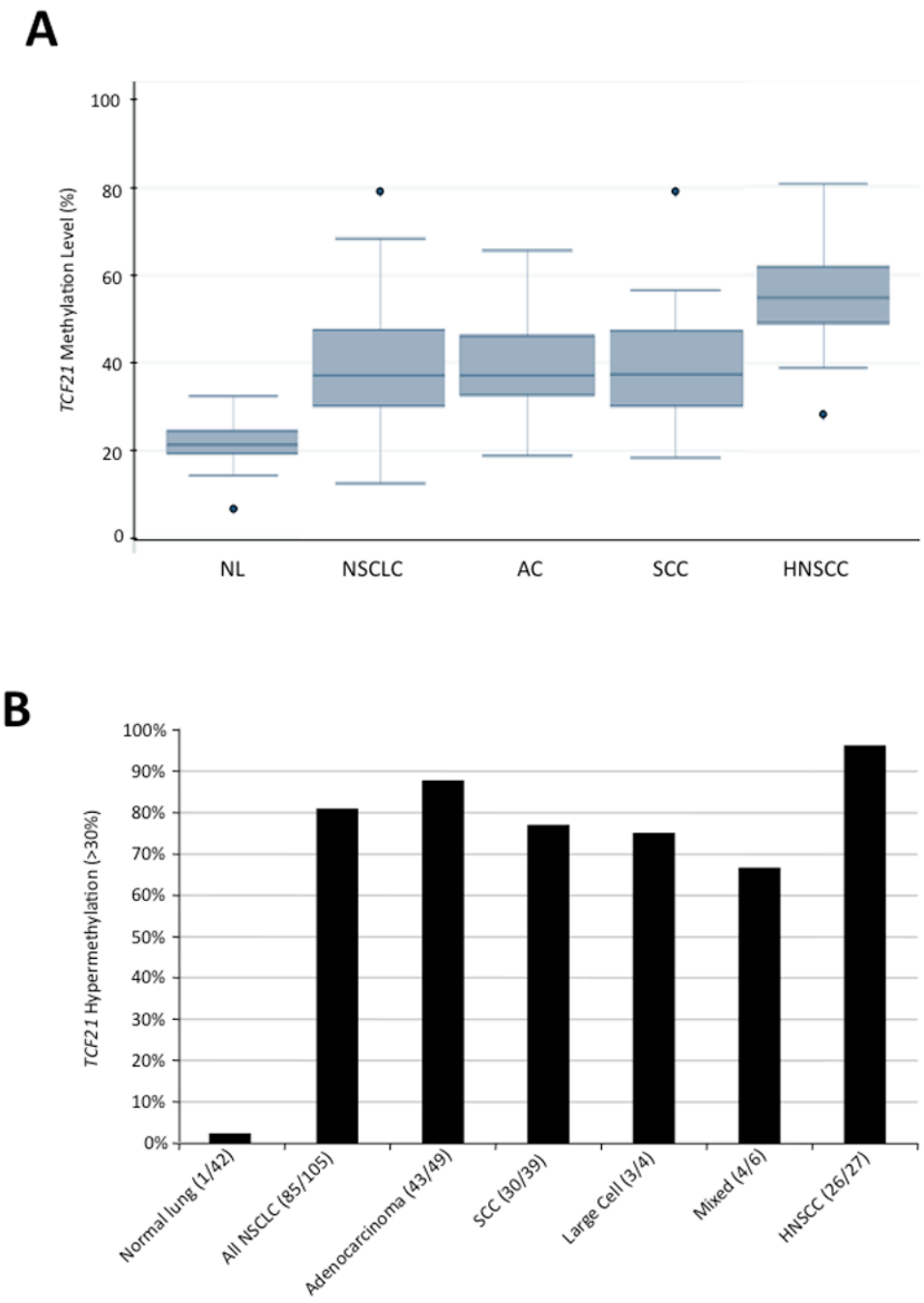


**Figure 2.**

Methylation levels and percent of tumors with methylation levels >30% threshold. **(A)** *TCF21* promoter methylation levels are shown in a box and whisker plot for 42 normal adjacent lung tissues, 42 NSCLC tumors, 63 additional NSCLC tumors, all 105 NSCLC tumors combined, and 24 HNSCC tumors. **(B)** Bar graph representing the number of NSCLC and HNSCC tumors exceeding the 30% threshold for hypermethylation.

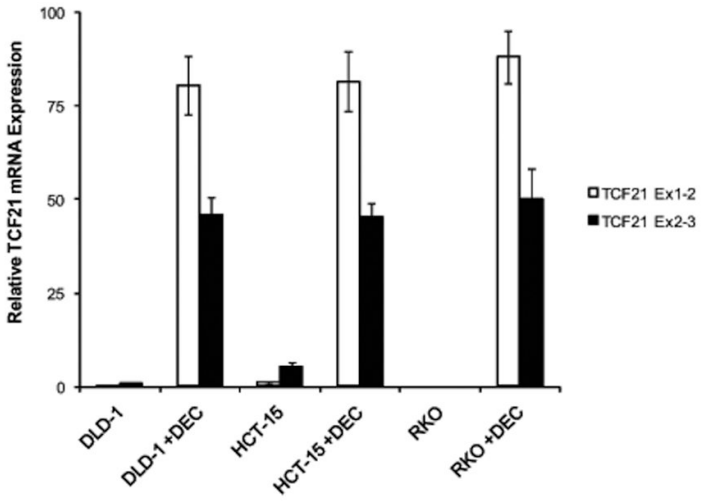


**Figure 3.**

Reactivation of *TCF21* expression by demethylating agent decitabine. Quantitative TaqMan mRNA real-time PCR results interrogating exon junctions of exons 1/2 and exons 2/3 in three colorectal cancer cell lines with high methylation levels (>85%) with and without decitabine treatment are shown.

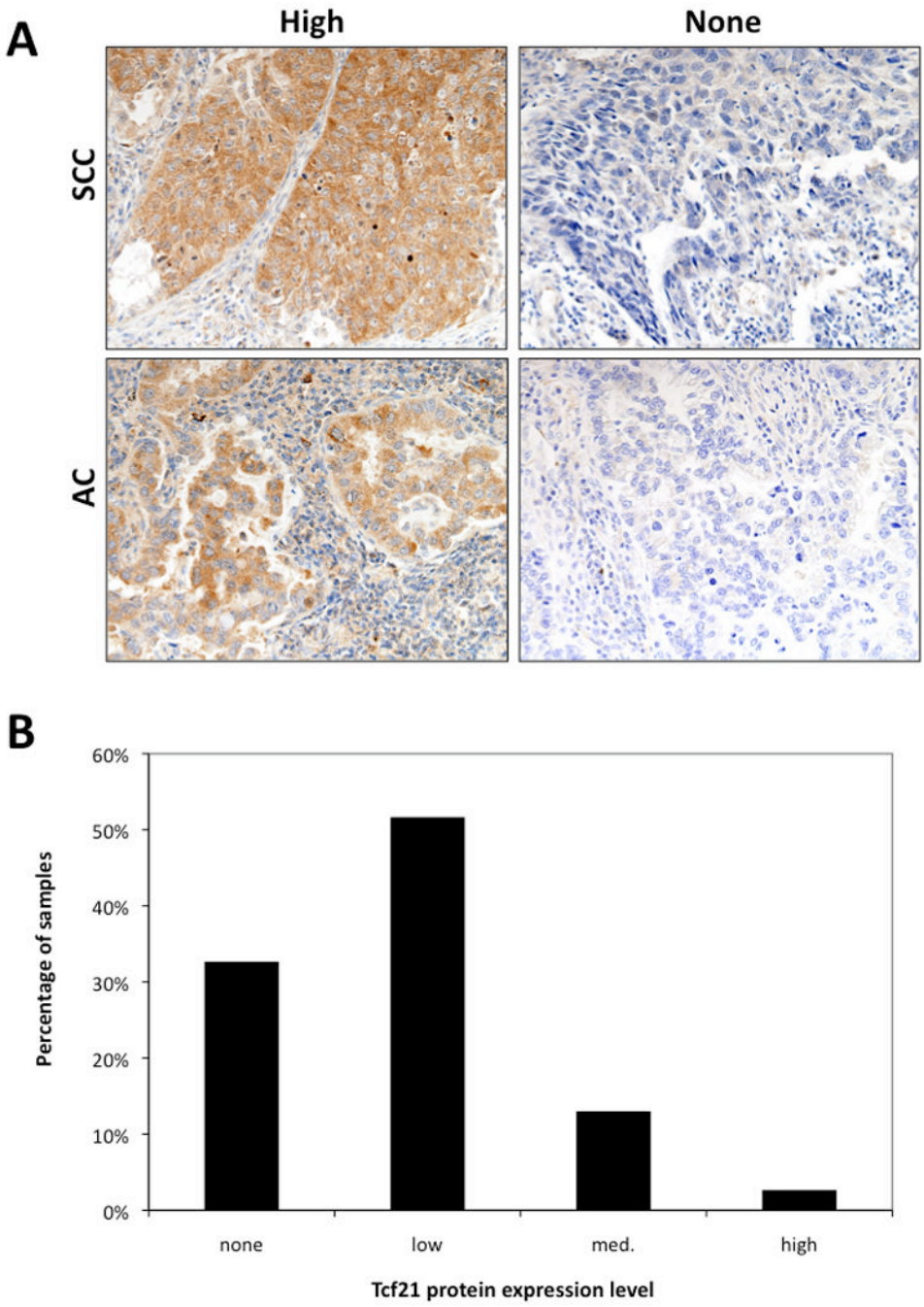


**Figure 4.**

TCF21 protein expression. **(A)** TCF21 immunohistochemical expression in lung cancer. NSCLC samples were stained with an anti-TCF21 antibody and scored as none, low, medium, or high. Representative examples of SCC and adenocarcinoma samples with high TCF21 expression (left) and no TCF21 staining (right) are shown. **(B)** Frequency of TCF21 expression in NSCLC on the TMA (300 patients). The percentage of samples in each expression category is shown. Reduced expression was defined as “none” or “low”.

Table 1

**Primers for LOH and Mutation Detection in *TCF21***

| Analysis | Forward Primer (5′—3′) | Reverse Primer (5′—3′) | Amplicon Size (bp) |
|---|---|---|---|
| **LOH** | | | |
| TCF21_5′_GT19 | CATGCCTGGGCGACAG | GGCCCTTAAGCTGACAACTT | 131 |
| TCF21_CA14 | ATGCAGTAAGGCCATAGTTTG | CAGCAGTGAGAGAACACCC | 294 |
| TCF21_E3_CT13 | GTGTTTCAAGTAAGCGAGTCT | GGGAAAGCCTGCTAGAAT | 163 |
| TCF21_3′_GT12-AT8 | AAGCTTAAGACTTTGGCTAA | CTAGTCTGGTTCCCTAGAGTT | 208 |
| **Sequence** | | | |
| TCF21_E1-1 | ATTGAGTTTCCCTCCGGTTGTGAA | CTCGCAGTTGGAGCTCTCCTCGGTG | 547 |
| TCF21_E1-2 | TCAGCGATGTGGAGGACCTTCAAGA | GCGGTGGTCGAGATGTGTAAGTCA | 563 |
| TCF21_E2-1 | CCCCCTTCCTTTCATCTCAG | AATTACATATTGCACTTGGACCAGC | 449 |
| TCF21_E2-2 | AAGATTCCCATCTATTTAACTTTA | ATCTGCATCTTCATTATGAAACTCA | 419 |

**Table 2**

**Clinical Characteristics of Patient Samples on the NSCLC Tissue Microarray and Correlation with TCF21 Expression**

| Variable | Categories | N | % | Mean TCF21 Expression | *p*-value |
|---|---|---|---|---|---|
| Histology | Adenocarcinoma | 191 | 63.7 | 34.8 | |
| | Squamous Cell Carcinoma | 109 | 36.3 | 54.9 | **0.003** |
| Gender | Female | 158 | 52.7 | 36.2 | |
| | Male | 142 | 47.3 | 48.7 | **0.021** |
| Race[a] | Caucasian, non-hispanic | 269 | 89.7 | 41.6 | |
| | Other | 31 | 10.3 | 46.1 | 0.209 |
| Tobacco History | Yes | 247 | 82.3 | 43.5 | |
| | No | 53 | 17.7 | 35.8 | 0.078 |
| Smoker | Current | 98 | 32.7 | 50.1 | |
| | Former | 149 | 49.7 | 39.1 | |
| | Never | 53 | 17.7 | 35.8 | **0.048** |
| Pathological T Stage | T1 | 111 | 37.0 | 42.1 | |
| | T2 | 159 | 53.0 | 41.5 | |
| | T3 | 16 | 5.3 | 68.0 | |
| | T4 | 14 | 4.7 | 19.3 | 0.086 |
| Pathological N Stage | N0 | 211 | 70.3 | 42.2 | |
| | N1 | 53 | 17.7 | 44.3 | |
| | N2 | 36 | 12.0 | 38.5 | 0.969 |
| Pathological M Stage | M0 | 291 | 97.0 | 42.6 | |
| | M1 | 9 | 3.0 | 27.3 | 0.290 |
| Pathological Stage | I | 189 | 63.0 | 41.9 | |
| | II | 58 | 19.3 | 48.4 | |
| | III | 44 | 14.7 | 37.6 | |
| | IV | 9 | 3.0 | 27.3 | 0.722 |
| Vital Status | Alive | 155 | 51.7 | 42.8 | |
| | Dead | 145 | 48.3 | 41.4 | 0.481 |
| Recurrence | Yes | 107 | 37.2 | 44.6 | |
| | No | 181 | 62.8 | 40.2 | 0.230 |

| Variable | Categories | N | % | Mean TCF21 Expression | *p*-value |
|---|---|---|---|---|---|
| | Missing | 12 | | | |
| Adjuvant Therapy | Yes | 7 | 2.8 | 23.3 | |
| | No | 240 | 97.2 | 44.0 | 0.314 |
| | Missing | 53 | | | |

[a] Race: "Other" includes 13 African-Americans, 8 Asians, 9 Hispanics, and 1 Native American.

**Table 3**

**Methylation, LOH, and Protein Expression Data for TCF21 in Lung Cancer Samples**

| Sample ID | Histology | Methylation (%) | Hypermethylated (yes/no) | LOH | No. of Hits | TCF21 protein expression on TMA |
|---|---|---|---|---|---|---|
| 462 | Adeno | 37.21% | yes | nd | | negative |
| 612 | Adeno | 32.74% | yes | nd | | high |
| 645 | Adeno | 31.39% | yes | nd | | negative |
| 649 | Adeno | 39.24% | yes | ROH | 1 | low |
| 759 | Adeno | 35.89% | yes | nd | | low |
| 782 | Adeno | 30.23% | yes | ROH | 1 | intermediate |
| 798 | Adeno | 33.43% | yes | LOH | 2 | low |
| 801 | Adeno | 51.64% | yes | nd | | low |
| 840 | Adeno | 35.06% | yes | LOH | 2 | intermediate |
| 842 | Adeno | 28.46% | no | nd | | low |
| 846 | Adeno | 22.59% | no | LOH | 1 | negative |
| 848 | Adeno | 44.62% | yes | LOH | 2 | intermediate |
| 1114 | Adeno | 62.16% | yes | ROH | 1 | nd |
| 1147 | Adeno | 56.92% | yes | LOH | 2 | low |
| 1153 | Adeno | 38.89% | yes | ROH | 1 | low |
| 1274 | Adeno | 44.23% | yes | ROH | 1 | nd |
| 1289 | Adeno | 39.44% | yes | ROH | 1 | low |
| 1323 | Adeno | 33.34% | yes | ROH | 1 | nd |
| 1352 | Adeno | 65.74% | yes | LOH | 2 | nd |
| 1355 | Adeno | 43.28% | yes | ROH | 1 | nd |
| 1406 | Adeno | 50.94% | yes | ROH | 1 | nd |
| 1416 | Adeno | 46.27% | yes | LOH | 2 | nd |
| 1436 | Adeno | 30.04% | yes | ROH | 1 | nd |
| 756 | SCC | 39.34% | yes | ROH | 1 | negative |
| 758 | SCC | 43.11% | yes | nd | | intermediate |
| 794 | SCC | 51.55% | yes | LOH | 2 | negative |
| 799 | SCC | 54.54% | yes | LOH | 2 | intermediate |
| 870 | SCC | 20.62% | no | nd | | high |
| 879 | SCC | 79.12% | yes | ROH | 1 | intermediate |

| Sample ID | Histology | Methylation (%) | Hypermethylated (yes/no) | LOH | No. of Hits | TCF21 protein expression on TMA |
|---|---|---|---|---|---|---|
| 913 | SCC | 59.04% | yes | ROH | 1 | intermediate |
| 931 | SCC | 48.34% | yes | LOH | 2 | low |
| 1140 | SCC | 40.10% | yes | ROH | 1 | intermediate |
| 1200 | SCC | 47.69% | yes | ROH | 1 | negative |
| 1234 | SCC | 40.23% | yes | ROH | 1 | low |
| 1237 | SCC | 39.78% | yes | LOH | 2 | intermediate |
| 1244 | SCC | 42.65% | yes | LOH | 2 | low |
| 1453 | SCC | 45.70% | yes | ROH | 1 | nd |
| 771_3 | SCC | 18.20% | no | LOH | 1 | intermediate |
| 786_2 | SCC | 29.73% | yes | LOH | 2 | low |
| 793_2 | SCC | 22.67% | no | nd | | intermediate |
| 878_2 | SCC | 33.74% | yes | ROH | 1 | nd |
| 905_1 | SCC | 31.90% | yes | ROH | 1 | nd |

Adeno, adenocarcinoma; SCC, squamous cell carcinoma; nd, not done; ROH, retention-of-heterozygosity; LOH, loss-of-heterozygosity; No. of Hits, total of hypermethylation and LOH events (if both were assayed): 0 is neither, 1 is either hypermethylation or LOH, and 2 is both; TMA, tissue microarray.

*Cancer*. Author manuscript; available in PMC 2012 February 1.