Lung cancer is the leading cause of cancer-related death worldwide. Cigarette smoking causes lung cancer in the majority of patients. However, the disease also arises in patients who are lifelong never smokers, i.e., individuals who smoked less than 100 cigarettes in their lifetime. In the United States, an estimated 10% of lung cancers occur in never smokers (1). Higher percentages of never smokers, up to 30%, have been observed in East Asian countries (2). Compared with other patients with lung cancer, “never smokers” with lung cancer have a unique clinical course (3, 4): they have a better prognosis (4), their tumors are more likely to harbor somatic mutations in the gene encoding the epidermal growth factor receptor (EGFR), and their tumors are more likely to respond to EGFR tyrosine kinase inhibitors (5, 6).

The etiology of lung cancer in never smokers is poorly understood. Second-hand smoke and radon may account for as many as 50% of cases (7, 8). Outdoor air pollution, cooking oil fumes, coal fumes, and asbestos may also contribute as risk factors (9). Given the lack of direct exposure to known carcinogens, never smokers with lung cancer may represent a subgroup for which predisposing genetic factors might be prominent and distinct from those of smokers.

Several reports, including recent genome-wide association studies, have identified germ line genetic variants associated with an increased risk of lung cancer (10–15). Most variants occur in genes that encode key proteins involved in the metabolism of tobacco-related products, including cytochrome P4501A1 (10), glutathione-S-transferase (11), myeloperoxidase (12), or nicotinic acetylcholine receptor (nAchR) subunits (13–15). The influence of cigarette
Materials and Methods

Patients. Patient samples from cases were obtained on IRB-approved protocols from three different sources: (a) an Internet-based protocol that collected samples from patients throughout the United States, (b) the Memorial Sloan-Kettering Cancer Center (MSKCC) in New York, NY, and (c) the Aichi Cancer Center in Nagoya, Japan. DNA specimens from age-matched never-smoker controls without lung cancer (controls) were collected under various IRB-approved protocols at the MSKCC (120 cases and 95 controls) or at the Aichi Cancer Center (125 cases and 247 controls; Table 1).

Internet-based enrollment at MSKCC. A prospective study to collect clinical data and biological specimens from never-smokers with lung cancer was approved by the MSKCC IRB and recorded in the ClinicalTrials.gov database (NCT00745160). The study opened in September 2008. Eligible patients were (a) >18 y old, (b) had histologically and/or cytologically proven NSCLC (adenocarcinoma, squamous cell carcinoma, and large cell carcinoma; ref. 18), and (c) were never smokers. Patients had to complete a screening questionnaire and give written informed consent for participation. Patients were excluded if they had any previous history of invasive malignancy (other than lung cancer), lived outside of the United States, or were not English speaking (as protocol documents and consent forms are only written in English).

Patients at MSKCC were allowed to enroll on the same protocol directly in the clinic. A retrospective review of the clinical and pathologic characteristics of these patients, including results of routine genotyping for the most common types of EGFR activating mutations, was done with IRB approval.

Internet-based collection of clinical data and blood. Patients read about the study via an IRB-approved web site (Fig. 1) and then provided contact information either through the web site or by sending an e-mail requesting information. A screening kit was directly sent to patients by e-mail or by regular mail. This included a detailed information packet, a screening questionnaire, consent forms, and a letter for release of pathology records. The questionnaire aimed at determining study eligibility and consisted of a detailed smoking survey, based on the Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance System Survey Questionnaire, along with additional questions regarding diagnosis, sex, age, ethnicity, mutational status of the tumor, treatment history, environmental exposure, and history of previous malignancies. A toll-free number was available for questions about the study. Documents were shipped back to study investigators using a prepaid shipping label. Eligible patients were shipped materials for blood collection, including a letter to a local laboratory testing provider, two sodium heparinized 10-mL tubes, biohazard bags, and a prepaid overnight delivery bubble envelope. For patients who did not have regularly scheduled blood tests,
saliva self-collection kits (Oragene OG-250, DNA Genotek) were sent. To ensure patient confidentiality, prior to sending kits to patients, collection materials were labeled with unique identifiers. Files linking patients’ names and protocol ID codes were accessible to only one person, who was not directly involved in the processing of the samples or subsequently in the data analysis. Bloods were drawn as part of a routine venipuncture (step 5), returned to study investigators, and stored at –80°C (step 6).

DNA extraction and genotyping of genetic variants. DNA was extracted from blood and saliva specimens using the Puregene DNA blood kit (Qiagen) or standard methods. EGFR intron 1 and exon 20 were analyzed by direct sequencing, using the following primers: ex1-F-5′-AGGGCTGAGAGAGAGAC-3′ and ex1-R-5′-GGGGAGAAATCTAGCTA-3′, and ex20-F-5′-CTCCA-CAGCCCCAGTGTC-3′ and ex20-R-5′-GCCGACGTGGTGTCAAGG-3′, respectively.

SNPs in the nAchR were genotyped as part of a pilot genome-wide association study. We genotyped 672 DNA samples using the Illumina 610-Quadv1.0 genotyping system. We first removed random samples used to assess the quality of DNA derived from our protocol (see below). Following evaluation of call rate and presence of monomorphs, we removed samples with a <98% call rate per individual, 50% identity-by-descent to another sample in the study, and/or with a gender discordance (X homozygosity). SNPs with >5% missing genotypes and monomorphs were also removed. This left us with 624 individuals genotyped at 582,871 SNPs each. We used principal components analysis, as implemented in the EIGENSOFT software package (20), separately on samples of Asian and European ancestry to remove population outliers and determine significant principal components to correct for residual population stratification. Among the Asian ancestry samples, all nonoutlier samples came from Japan. Three significant principal components were found in the Asian ancestry samples, whereas five significant principal components were found in the European ancestry samples.

Statistical analyses. For EGFR exon 20, categorical variables were compared using Fisher’s exact test. Results were considered significant at the 0.05 level. Statistical analyses were done using SPSS, version 17.0.

For the other SNPs, logistic regression was used to assess allele frequency differences between cases and controls as implemented in the software package PLINK (21). For the nAchR SNPs, significant principal components were included as covariates to account for population substructure.

Results

Determination of DNA quality from blood specimens. To determine the feasibility of extracting high-quality DNA from blood stored at room temperature for varying periods of times, we tested blood samples from 10 volunteers processed 24, 48, or 72 hours after venipuncture. In all cases, processing of extracted DNA on Human 610-Quadv1.0 arrays (Illumina) resulted in call rates higher than 99.99%.

We allowed blood collection from patients at any time during their treatment course. To ensure that this approach was valid, we showed that SNP discordance was lower than 10−5 pre- and post-chemotherapy in four patients.

Clinical data and blood sample collection. The protocol web site went online on September 15, 2008. As of May 15, 2009, 122 patients expressed interest in the study via the Internet (n = 119) or the phone (n = 3). Patients learned about the program through Internet searches (n = 49), blogs (n = 48), personal doctors (n = 15), or nonprofit organizations’ newsletters (n = 10). A significant number of patients contacted us after the web site was highlighted on cancer-related web sites; subsequently, accrual was quite steady. Five patients resided outside the United States and were excluded. Five individuals sent e-mail messages

<table>
<thead>
<tr>
<th>Table 1. Population characteristics</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>369 (100%)</td>
<td>342 (100%)</td>
</tr>
<tr>
<td>Source</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Internet protocol</td>
<td>124 (34%)</td>
<td>NA</td>
</tr>
<tr>
<td>Internet</td>
<td>45 (12%)</td>
<td>NA</td>
</tr>
<tr>
<td>MSKCC</td>
<td>79 (21%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>MSKCC blood bank</td>
<td>120 (33%)</td>
<td>95 (17%)</td>
</tr>
<tr>
<td>Aichi Cancer Center</td>
<td>125 (34%)</td>
<td>247 (83%)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>201 (54%)</td>
<td>95 (17%)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>197 (53%)</td>
<td>95 (17%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4 (1%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Asian</td>
<td>152 (41%)</td>
<td>247 (83%)</td>
</tr>
<tr>
<td>Black</td>
<td>16 (4%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>324 (88%)</td>
<td>342 (100%)</td>
</tr>
<tr>
<td>Men</td>
<td>45 (12%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>367 (99%)</td>
<td>NA</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>2 (1%)</td>
<td>NA</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>0 (0%)</td>
<td>NA</td>
</tr>
<tr>
<td>Tumor EGFR status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutant</td>
<td>131 (36%)</td>
<td>NA</td>
</tr>
<tr>
<td>Exon 19 deletion</td>
<td>75 (20%)</td>
<td>NA</td>
</tr>
<tr>
<td>L858R mutation</td>
<td>51 (14%)</td>
<td>NA</td>
</tr>
<tr>
<td>Exon 19 insertion</td>
<td>2 (0.5%)</td>
<td>NA</td>
</tr>
<tr>
<td>L861Q</td>
<td>1 (0.2%)</td>
<td>NA</td>
</tr>
<tr>
<td>T790M mutation</td>
<td>2 (0.5%)</td>
<td>NA</td>
</tr>
<tr>
<td>Wild-type</td>
<td>95 (26%)</td>
<td>NA</td>
</tr>
<tr>
<td>Unknown</td>
<td>143 (39%)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable

www.aacjrournals.org
only to share personal experiences. Overall, screening documents were sent to 111 patients.

Preferred initial contact with study investigators was e-mail (n = 89), regular mail (n = 11), and phone call (n = 11). Patients were from 28 different states, with the highest representations from Florida (n = 19) and California (n = 11; Fig. 2).

Sixty-three (57%) patients returned the screening documents. Eligibility was confirmed for 55 (87%) patients. Main reasons for exclusion were (a) a previous history of invasive cancer (n = 6), and (b) a histologic type other than NSCLC (carcinoid tumor in one case and small cell lung cancer in another case). The majority of patients (n = 52) had a regular blood test, and saliva collection kits were sent to only three patients. As of May 15, 2009, a total of 45 patients returned a blood/saliva specimen. Median time between first e-mail contact and receipt of blood was 33 days (range, 9-200 days).

During the same period, an additional 79 patients were recruited directly in MSKCC clinics. The clinical characteristics of all 124 patients whose biospecimens were collected on this protocol are shown in Table 1. There were 100 women and 24 men. Most were white and had tumors with adenocarcinoma histology (98% of cases). Tumor EGFR mutational status was known for 67 patients and consisted of exon 19 deletions in 41 (61%) cases.

**Association analyses.** Genotyping for the germ line EGFR T790M variant (C → T at nucleotide 2369 in exon 20) was done on DNA from 661 never smokers, including 369 patients with NSCLC, mostly of the adenocarcinoma subtype (99% of cases), and 292 controls, mostly of Asian ethnicity (83% of controls). Fifty controls from our cohort (45 white and 5 Asian) were not included in these analyses due to insufficient quantity of DNA after processing on SNP arrays. Median age was not significantly different in the two groups (61 and 62 years old; P = 0.822, Fisher’s exact test). Results from routine tumor genotyping for EGFR somatic mutations were available for 226 cases, 131 (58%) of whom had an activating EGFR mutation. The T790M variant was identified in 2 (0.54%; 95% confidence interval, 0.21-1.29%) of the 369 patients with NSCLC, and in none of the 292 controls (P = 0.21 at Fisher’s exact test).

SNP rs763317 in intron 1 was investigated for association with lung cancer, adjusting for ethnicity. Due to the limited number of individuals of black or Hispanic ancestry, these individuals were excluded from the analysis. No difference was observed in allele frequency between cases and controls (odds ratio, 0.93; 95% confidence interval, 0.70-1.2; P = 0.58). Similar results were found after the removal of
40 males from the data set (odds ratio, 0.94; 95% confidence interval, 0.71-1.2; \(P = 0.67\)). After restricting the cases to individuals known to have an EGFR mutant tumor, still no association was observed (odds ratio, 0.94; 95% confidence interval, 0.65-1.4; \(P = 0.73\)). Similar negative results were found with dominant and recessive models.

Data for two SNPs in the nAchR subunit gene \(\text{CHRNA3}\) on chromosome 15q25 (rs8034191 and rs1051730) were extracted from an ongoing pilot genome-wide association study, including 217 of the cases and 342 of the controls of the cohort. This allowed us to consider population substructure within those two ethnic groups, using other data from SNP arrays. The nAchR SNPs were tested for association with lung cancer in the white and Japanese cohorts separately, adjusting for significant principal components in each cohort. Neither rs8034191 nor rs1051730 was associated with lung cancer risk (Table 2).

**Pedigrees of two patients with germ line EGFR T790M mutation.** The first patient with a germ line EGFR T790M mutation was a 66-year-old woman of Asian descent (from India) who presented with a 2 cm ground-glass nodule in the left lower lobe, associated with multiple bilateral subcentimeter ground-glass opacities. The patient’s mother had a history of NSCLC (Fig. 3A). Left lower lobectomy revealed mixed adenocarcinoma with acinar and bronchioloalveolar features, harboring a somatic \(\text{EGFR L858R}\) mutation (Supplementary Fig. S1).

The second patient was a 58-year-old man of Eastern European ancestry who was diagnosed with NSCLC with bone and liver metastases. Family history was assessed by an expert genetic counselor (M. Robson). The patient had a significant family history for NSCLC; most family members were smokers (Fig. 3B). Core needle biopsy of a liver...
lesion revealed poorly differentiated acinar and solid ade-
nocarcinoma, which was found to contain the EGFR
L858R mutation at routine genotyping (Supplementary
Fig. S1). Unfortunately, tumor histology and mutational
status could not be determined for any of the family
members with NSCLC.

Discussion

The identification of genetic risk variants that may pre-
dispose certain patients to disease requires the analysis of
germ line DNA from affected cases. In this study, we show
that the Internet can provide a secure, confidential, and
convenient way to bolster the accrual of target patients
from across the country.

Our protocol had a similar design to the Harvard
Myeloproliferative Disorders Study. That protocol used
the Internet to collect blood from 345 participants over
a 1-year period (i.e., 0.41%/mo of the 7,000 newly
diagnosed cases of myeloproliferative disorders in
the United States; ref. 22). We had hoped to accrue
patients at a similar rate. However, our recruitment
was significantly lower, as we only recruited 45 patients

Fig. 3. Pedigrees of patients found to have a germ line EGFR T790M mutation. Numbers, ages of family members; arrows, probands; Lung, patients with lung cancer (followed by age at onset); Leuk, leukemia; CLL, chronic lymphoid leukemia; Gastric, gastric cancer.
over an 8-month period (i.e., 0.11%/mo of the 14,000 estimated newly diagnosed cases over the 8-month period). Limiting factors for enrollment may include an older age at onset of lung cancer and differences in socioeconomic characteristics compared with those who develop myeloproliferative diseases (23); these factors ultimately may reduce these patients’ familiarity/access to the Internet. Socioeconomic factors have previously been reported to significantly influence the completion of an Internet-based behavioral study in smokers (24) and may also exist in never smokers (25). Second, disease-free survival in NSCLC is far shorter than in myeloproliferative disorders, limiting the extra time and effort patients are willing to allot to research. This reason was given by several patients who received screening documents but did not return them. Interestingly, the majority of patients who signed the consent form ultimately completed the questionnaire, had their blood drawn, and sent the blood specimens. This success rate suggests that the study participation was not overly complicated and that a more significant limitation of the study was awareness. Thus, in the future, we will expand efforts to connect with individuals appropriate to this study, through patients’ groups, development of collaboration with other institutions, and presentation of the study at oncology meetings.

Although somatic EGFR T790M mutations are common in patients with acquired resistance to EGFR inhibitors (16), germ line EGFR T790M mutations are rare. This variant was initially reported in 2005 in a family of European descent, in which six family members in three generations had lung cancer (26). We found the germ line EGFR T790M variant in 2 of 369 cases (0.54%) of never smokers with NSCLC; both patients had family histories significant for lung cancer. In a separate study, no germ line mutation was identified in a series of 237 individuals with three or more first-degree relatives with lung cancer and 32 patients with bronchioloalveolar carcinoma (27). One germ line EGFR T790M mutation was identified in a cohort of 240 patients with previously untreated lung adenocarcinoma (28). Including our two patients, a total of five patients have now been reported to have the germ line EGFR T790M variant (Table 3). In these patients, lung cancer was diagnosed from ages 50 to 72 years, and was metastatic to the lung or other locations at the time of diagnosis. At least three patients were never smokers. Histology was adenocarcinoma and/or bronchioloalveolar carcinoma in all cases. EGFR was genotyped in all five germ line EGFR T790M cases. Somatic EGFR-activating mutations, which are known to be more frequent in never smokers with lung adenocarcinoma (6), were identified in four of the five patients. Collectively, these data indicate that a germ line T790M alteration is a genetic risk variant for lung cancer in never smokers. However, tumors develop only after a relatively long latency, suggesting that other genetic alterations, such as additional EGFR mutations, may collaborate with T790M (29).

We were unable to confirm a reported association between rs763317 in the EGFR gene and lung cancer (17). In the previous report, female never smokers with lung adenocarcinoma heterozygous at the locus had a 1.2-fold increased risk of lung cancer relative to individuals homozygous for the G allele. Individuals homozygous for the A allele were 3.6 times more likely to develop lung adenocarcinoma. Based on this effect size and the frequency of this SNP in our control population, we estimated 82% power to detect this association in our Asian cohort, and 96% power to detect the association in our white cohort. Therefore, our failure to replicate this association was not due to low power. One possible explanation is that this association was a false positive. Alternatively, the initial report could have overestimated the effect size at this locus, due to the “winner’s curse”, and therefore, our power to replicate the association was lower than we calculated.

We were also unable to replicate the previously reported association between SNPs rs8034191 and rs1051730 in CHRNA3 and lung cancer risk (15–15). Unlike EGFR rs763317, these associations are much more modest. Moreover, the minor allele frequency in Asian populations at these SNPs is much lower than in the European populations in which the association was first discovered. Our power to detect an association in our cohort, therefore, was limited. However, as it has been

<table>
<thead>
<tr>
<th>Table 3. Lung cancer patients with germ line T790M mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnicity</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Bell et al. (26)</td>
</tr>
<tr>
<td>Proband</td>
</tr>
<tr>
<td>Brother</td>
</tr>
<tr>
<td>Prudkin et al. (28)</td>
</tr>
<tr>
<td>Proband</td>
</tr>
<tr>
<td>Present study</td>
</tr>
<tr>
<td>Patient 1</td>
</tr>
<tr>
<td>Patient 2</td>
</tr>
</tbody>
</table>

Abbreviations: M, male; F, female.

*Two primary tumors.
suggested that these SNPs affect smoking behavior as well as lung cancer risk, and as we focused on never smokers with lung cancers, we cannot discount the possibility that the SNP only has an effect in smokers (or influences smoking behavior) and is truly not associated with lung cancer in never smokers.

To conclude, this study shows the feasibility of bolstering accrual for germ line predisposition studies in never smokers with lung cancer by collecting clinical information and blood sample specimens from appropriate patients throughout the nation using an Internet-based protocol. We plan in future studies to conduct genome-wide association studies with more cases and controls to identify genetic risk variants in this disease population.

Disclosures of Potential Conflicts of Interest

W. Pao: consultant/advisory board, MolecularMD. W. Pao and V. Miller: EGFR T790M patent, MolecularMD. The other authors disclosed no potential conflicts of interest.

Acknowledgments

We thank R. Levine for protocol assistance/reviewing the manuscript; K. Liao for processing specimens on the Illumina platform; T. Aliff for referring patients; H. West for discussing the study on his lung cancer blog; and S. Mantel for discussing this study in the monthly newsletter of Joan’s Legacy: Uniting Against Lung Cancer. We are grateful to all the patients who participated.

Grant Support

HOPP Lung Cancer Research Fund (W. Pao), the Rosalind Warren Memorial Fund (W. Pao), the MSKCC Geoffrey Beene Cancer Research Center (W. Pao), the NIH/National Cancer Institute (R01-CA121210; W. Pao), the Society of MSKCC (I. Orloff), Steps for Breath (I. Orloff), and the Labrecque Foundation (I. Orloff). Services provided by the Genomics Core Facility were partially supported by a National Cancer Institute CCSG award (P30-CA008748). N. Girard is a recipient of travel grants from the Collège des Professeurs de Pneumologie/AstraZeneca, the National Federation of French Comprehensive Cancer Centers/Fondation de France, and the Philippe Foundation.

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 9/6/09; revised 10/25/09; accepted 11/8/09; published OnlineFirst 1/12/10.

References


Clinical Cancer Research

Analysis of Genetic Variants in Never-Smokers with Lung Cancer Facilitated by an Internet-Based Blood Collection Protocol: A Preliminary Report


Clin Cancer Res  Published OnlineFirst January 12, 2010.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-09-2437

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2010/01/18/1078-0432.CCR-09-2437.DC1

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.