Molecular Profiling of Premalignant Lesions in Lung Squamous Cell Carcinomas Identifies Mechanisms Involved in Stepwise Carcinogenesis

Aik T. Ooi1, Adam C. Gower3, Kelvin X. Zhang4, Jessica L. Vick5, Longsheng Hong3, Brian Nagao3, W. Dean Wallace3, David A. Elashoff6,7, Tonya C. Walser2,7, Steven M. Dubinett2,3,7, Matteo Pellegrini6,8, Marc E. Lenburg9, Avrum Spira9, and Brigitte N. Gomperts1,2,7,8

Abstract

Lung squamous cell carcinoma (SCC) is thought to arise from premalignant lesions in the airway epithelium; therefore, studying these lesions is critical for understanding lung carcinogenesis. Previous microarray and sequencing studies designed to discover early biomarkers and therapeutic targets for lung SCC had limited success identifying key driver events in lung carcinogenesis, mostly due to the cellular heterogeneity of patient samples examined and the interindividual variability associated with difficult to obtain airway premalignant lesions and appropriate normal control samples within the same patient. We performed RNA sequencing on laser-microdissected representative cell populations along the SCC pathologic continuum of patient-matched normal basal cells, premalignant lesions, and tumor cells. We discovered transcriptomic changes and identified genomic pathways altered with initiation and progression of SCC within individual patients. We used immunofluorescent staining to confirm gene expression changes in premalignant lesions and tumor cells, including increased expression of SLC2A1, CEACAM5, and PTBP3 at the protein level and increased activation of MYC via nuclear translocation. Cytoband enrichment analysis revealed coordinated loss and gain of expression in chromosome 3p and 3q regions, respectively, during carcinogenesis. This is the first gene expression profiling study of airway premalignant lesions with patient-matched SCC tumor samples. Our results provide much needed information about the biology of premalignant lesions and the molecular changes that occur during stepwise carcinogenesis of SCC, and it highlights a novel approach for identifying some of the earliest molecular changes associated with initiation and progression of lung carcinogenesis within individual patients. Cancer Prev Res; 7(5); 487–95. ©2014 AACR.

Introduction

Lung cancer is the most deadly cancer worldwide, accounting for more deaths than prostate cancer, breast cancer, pancreatic cancer, and colon cancer combined (1). Squamous cell carcinoma (SCC) is a common type of non–small cell lung cancer that accounts for 30% of all lung cancers and is frequently associated with smoking (2). In general, despite current therapeutic strategies of chemotherapy, radiotherapy, and trials with targeted therapies, the overall survival of patients with lung cancer, including SCC, is still very poor with a 5-year survival rate of 15.9% (3).

SCC often arises centrally from a large airway, usually a bronchus. Ongoing injury of airway epithelia leads to repair and regeneration that can give rise to a phenotype of squamous metaplasia and subsequently to dysplasia, both of which are histologic features seen in the airways of smokers (4, 5). It is believed that SCC develops through a series of genetic and epigenetic changes that alter the epithelium from squamous metaplasia, then to dysplasia, carcinoma in situ and finally to invasive carcinoma (6).

Although there have been studies devoted to discovering the genetic and molecular changes observed in lung cancer, few studies have directly investigated changes associated with squamous metaplasia or dysplasia (7–9). In fact, it is not known with certainty whether premalignant lesions of the airway are the direct progenitors of invasive SCC. This is mainly due to the challenge inherent in following airway...
premalignant lesions serially over time in the large airways to determine whether a particular lesion is destined to develop into SCC. To better understand the process of carcinogenesis leading to SCC, especially those steps involved in the early and precancerous stages, a comprehensive study of the molecular alterations that characterize premalignant lesions is needed along with a direct comparison with the molecular changes found in SCC from that same individual.

Basal cells of the airway are known to be stem/progenitor cells required for airway epithelial repair (10), and we hypothesize that premalignant lesions arise from aberrant repair in these cells (11). Therefore, we profiled the transcriptome of airway basal cell, premalignant lesions, and tumors from the same patients to improve our understanding of the stepwise carcinogenesis in SCC and to aid in the identification of new diagnostic and therapeutic approaches for SCC and novel chemopreventive strategies.

Materials and Methods

Case selection and histology review

Resected tissue blocks from SCC cases were reviewed with two pathologists to identify regions of normal airway epithelium, squamous metaplasia or dysplasia, or carcinomas. Patients with fresh-frozen or formalin-fixed paraffin-embedded (FFPE) tissue blocks containing all three regions were selected for the study. Immunofluorescent staining of KRT5 was performed to validate the identification of selected lesions. Fresh-frozen tissues were used for RNA sequencing, whereas FFPE tissues were used for validation of independent cases with quantitative real-time PCR (qRT-PCR) and immunofluorescent staining.

Laser capture microdissection

Tissues were sectioned at a 7-μm thickness and mounted on regular uncharged glass slides for patients 1, 2, and 3, and on polyethylene naphthalate membrane slides (Leica) for patient 4, followed by hematoxylin and eosin staining. Laser capture microdissection was performed using the Arcturus eLiX for patient 1 and 2, Zeiss PALM for patient 3, and Leica LMD7000 for patient 4. A tissue area of 800,000 to 1,200,000 μm² was dissected and collected from each lesion.

RNA extraction and sequencing library preparation

RNA was extracted from laser-microdissected cells using the RNeasy Micro Kit (Qiagen). The cDNA was generated using the Ovation RNA-Seq System (NuGEN) for patients 1 and 2 and the Ovation RNA-Seq V2 System (NuGEN) for patients 3 and 4. For patients 1 and 2, cDNA of approximately 200 bp was selected by gel purification. For patients 3 and 4, the cDNA was sheared to 140 to 180 bp using the Covaris focused-ultrasonicator with the following settings: duty cycle 10%; intensity 5; cycles per burst 200; total time 6 minutes. The size range of the sheared cDNA was confirmed by Bioanalyzer analysis before library construction using the Encore Library System (NuGEN). The average size of each library was estimated by Bioanalyzer analysis, and the concentration of each was measured on the Qubit fluorometer (Invitrogen).

Sequence analysis

Sequencing libraries from patients 1 and 2 were each sequenced on a single flow cell lane of an Illumina Genome Analyzer IIx, generating 36-base single-end reads, and libraries from patients 3 and 4 were each sequenced on a single flow cell lane of an Illumina HiSeq 2000, generating 50-base single-end reads. All reads were trimmed to 35 bases before alignment. In the case of patient 1, the first base of each read was also trimmed-off due to a problem with the first sequencing cycle. Reads that failed the chastity filter of Illumina [brightest intensity/(brightest intensity + second brightest intensity) < 0.6 for at least two of the first 25 cycles] were automatically removed during preprocessing. The remaining reads were aligned to the human genome (build hg19) using Bowtie v0.12.7 (12), allowing only unique alignments and up to two mismatches per read. Reads aligning to the mitochondrial genome were removed from further analysis. Gene expression estimates were then computed by measuring the coverage of each of 55,841 Ensembl Gene loci (Ensembl build 69) using the BEDTools software suite (13). The coverage for each Ensembl Gene locus in each sample was then normalized to the size of the locus and the total number of reads mapping uniquely to the nuclear genome to obtain an RPKM (14) value for each gene in each sample. RPKM values were seventh root-transformed before analysis to produce an approximately normal distribution of (nonzero) gene expression values.

Statistical analysis

All models were created using the R environment for statistical computing (version 2.12.0). Linear mixed-effects models were created using the nlme R package (version 3.1-97) and negative binomial models were created using the MASS R package (version 7.3-7). The Student two-sample t test with equal variance (or, in the case of GDS1312, Student paired t test) was used to assess the significance of differential expression of candidate genes in Gene Expression Omnibus (GEO) datasets. Analysis of GEO datasets was performed using the preprocessed expression levels generated using default Affymetrix probe sets (averaging across multiple probesets to obtain a single expression value for each gene).

Gene set enrichment analysis

Positionally defined (cytoband) Ensembl Gene sets were created using the biomaRt R package to extract chromosomal band annotation for Ensembl Gene identifiers using Ensembl version 69. These gene sets were then used to perform preranked GSEA (gene set enrichment analysis; ref. 15) using lists of all Ensembl Genes ranked by the t statistics from the linear mixed-effects models, to identify cytobands that were overrepresented among genes coordinately up- or downregulated in premalignant or tumor cells compared with normal basal cell. Analysis was performed using GSEA v2.0.8 (build 14) with 1,000 permutations, removal of gene sets with >500 genes, and a random seed of 1,234.
Quantitative real-time PCR

Amplified cDNA generated during the library preparation for patients 3 and 4 was used for qRT-PCR analysis. In addition, normal basal cell and premalignant lesions from 4 independent patients were laser microdissected from FFPE tissues, and cDNA was generated using the Ovation RNASeq FFPE System. TaqMan Gene Expression Assays (Life Technologies) were used to examine the expression levels of selected candidate genes (CEACAM5, SLC2A1) in normal basal cell and premalignant lesions. β2-microglobulin (B2M) was used as an endogenous control. Statistical analysis was performed using the sign test.

Immunofluorescent staining

FFPE tissues were sectioned at a 5-μm thickness and stained as previously described (11). Antibodies used are: rabbit anti-KRT5, mouse anti-KRT5, mouse anti-PTBP3, and rabbit anti-c-myc (Abcam); mouse anti-CEACAM5 (Pro-Mab Biotechnologies Inc.); rabbit anti-SLC2A1 (Alpha Diagnostic International); anti-rabbit Cy3, anti-mouse Alexa Fluor 647, anti-mouse Cy3, and anti-rabbit Alexa Fluor 647 (Jackson ImmunoResearch). Immunostained tissues were visualized on an Axiocam system (Zeiss), and images were taken using the Axiovision software.

Results

Study population, sample acquisition, and sequence alignment

Fresh-frozen tissue blocks were obtained from 4 individuals with lung SCC (patients 1–4) at the time of tumor resection, and regions of normal basal cell, premalignant (squamous metaplastic and dysplastic) cells, and tumor cells were successfully captured from sectioned tissues by laser microdissection (Supplementary Fig. S1). The demographic information and clinical characteristics of these patients, as well as a description of the histology of each microdissected premalignant region, are presented in Supplementary Table S1. Sequencing libraries of the expected concentration and cDNA size ranges were generated from RNA isolated from the microdissected cells. All sequenced samples produced reads with mean Phred quality scores above 25, indicating that it was possible to generate sequencing libraries of good quality with our method of isolating RNA from laser-microdissected materials.

A table of the number of reads that aligned uniquely within each sample is shown in Supplementary Table S2. The fraction of reads aligning to the mitochondrial genome varied considerably among samples. In patients 1 and 2, this fraction varied from 7% to 28%, but in patients 3 and 4, mitochondrial reads comprised from 22% to 65% of uniquely aligned reads, with the highest fraction found in the tumor samples from patient 4 (57%–65%). Because of this large amount of variability, reads aligning to the mitochondrial genome were discarded from analysis after alignment, and RPKM (reads/kb/millions of reads) values were computed relative to the total number of reads aligning uniquely to the nuclear genome.

Identification of genes associated with carcinogenesis

To identify SCC-associated genes whose expression is also associated with progression from normal airway basal cell to premalignant (metaplastic or dysplastic) lesions, a multistep procedure was used as outlined in Fig. 1A. First, Ensembl Genes with zero aligned reads in all samples from at least 1 patient were removed from analysis (to ensure that all patients contributed evidence to each result), leaving 20,817 genes for analysis. This list was then filtered to consider only those genes with substantial evidence of expression (median of greater than 50 uniquely aligned reads across all samples), leaving 7,025 genes for analysis. Using linear mixed-effects models and negative binomial generalized linear models (see Supplementary Methods for details), we then identified 626 early-stage genes (significantly differentially expressed in a similar manner in both premalignant lesions and tumor compared with normal basal cell), 730 late-stage genes (significantly differentially expressed in a similar manner in tumor compared with both premalignant lesions and normal basal cell), and 68 “stepwise” genes (significantly differentially expressed in both of the described stages of carcinogenesis; Fig. 1B; Supplementary Table S3).

Experimental and computational validation of candidate genes

Three genes were selected for further validation: CEACAM5, SLC2A1, and PTBP3. These genes, whose expression was upregulated in premalignant lesions and tumor cells compared with normal basal cell, were chosen because of their potential roles in the biology of lung carcinogenesis. The expression of CEACAM5 and SLC2A1 was measured by performing qRT-PCR on remaining material from the sequencing libraries of patients 3 and 4, as well as on laser-microdissected RNA from four additional independent cases (patients 5–8). In each case, the mRNA level of each gene was significantly higher (sign test $P < 0.05$) in the premalignant lesion than in normal basal cell (Fig. 2A).

Because mRNA and protein levels may not always be well correlated (16–18), immunofluorescent staining was performed in sections of normal epithelium, premalignant lesion, and carcinoma from 2 independent cases (patients 9 and 10). CEACAM5 and SLC2A1 were not detectable in the normal epithelia, but they were highly expressed in cells within both metaplastic lesions and the SCC tumors (Fig. 2B and C). SLC2A1 was expressed throughout the KRT5+ component of the tumor, whereas CEACAM5 was expressed in some, but not all, KRT5+ tumor cells. PTBP3 was strongly expressed in premalignant lesions and tumor cells, and although it was strongly expressed in columnar KRT5+ cells of normal airway epithelium, its expression was undetectable in normal basal cell (Supplementary Fig. S2).

To better understand the biologic role that these genes may play in the development of lung SCC, the significance of the differential expression of each gene was assessed in several GEO datasets with respect to experimental parameters relevant to lung SCC carcinogenesis. First, SLC2A1 and PTBP3 were confirmed to be significantly upregulated (SLC2A1,
P = 0.004; PTBP3, P = 0.017) in an independent set of SCC tumors (n = 5) with respect to paired samples of adjacent normal tissue (GEO dataset GDS1312; ref. 19); however, the expression of CEACAM5 was unchanged (P = 0.64). Next, a collection of SCC (n = 18) and adenocarcinoma (ADC; n = 40) lung tumors (GDS3627; refs. 20, 21) was interrogated to determine the specificity of the expression of these genes with respect to the SCC tumor type. The expression of SLC2A1 and PTBP3 was again strongly increased in SCC tumors compared with ADC tumors (SLC2A1, \( P = 1.1 \times 10^{-7} \); PTBP3, \( P = 0.0004 \)); however, CEACAM5 was moderately downregulated in SCC relative to ADC (P = 0.08). Finally, because premalignant lesions in large central airways are believed to arise from injury caused by cigarette smoking, the expression levels of these genes were examined in a study of bronchoscopic brushings of healthy current (n = 34), former (n = 18), and never (n = 23) smokers (GDS534; ref. 21). In this study, CEACAM5 and SLC2A1 were significantly upregulated in brushings from current smokers compared with those from never smokers (CEACAM5, \( P = 0.0001 \); SLC2A1, \( P = 0.016 \)), although PTBP3 was not (P = 0.66).

Prediction of chromosomal gains and losses during carcinogenesis

GSEA performed using positionally defined gene sets (cytobands) revealed that late-stage (but not early-stage) carcinogenesis is associated with a coordinate loss of expression in the p arm of chromosome 3 and an attendant gain of expression in 3q26.33-3q29 (Fig. 3A), which corresponds to previously reported observations of frequent 3p deletion and 3q amplification in squamous tumors (22, 23). In particular, chromosomal band 3q26.33 has been reported to be consistently amplified in lung SCC (24).

Identification of biologic changes in early- and late-stage carcinogenesis

Ingenuity pathway analysis (IPA; Ingenuity Systems) was used to further characterize the changes in biologic functions resulting from the differential expression of genes associated with early-stage events, which contribute to the initiation and formation of premalignant lesions, or with late-stage events, which are involved in the progression...
from premalignant lesions to tumor. This analysis revealed that the early-stage carcinogenesis was characterized uniquely by increased protein ubiquitination and cell-cycle progression, whereas the late-stage events were marked primarily by increased transcriptional and translational activity and cellular migration and transformation (Fig. 3B; Supplementary Table S4). In addition, an increase in cell survival and proliferation and a corresponding down-regulation of cell death mechanisms was observed throughout both stages of carcinogenesis. IPA was also used to determine whether the genes identified to be differentially expressed either early or late in carcinogenesis are enriched in known targets of various transcription factors. This approach revealed that the set of genes that is differentially expressed early in carcinogenesis and remains dysregulated in tumor cells is enriched in previously reported targets of MYC and TP53 (Fig. 4A; Supplementary Tables S5 and S6). As MYC and TP53 are predicted to activate or repress the expression of these targets, respectively, this suggests that MYC activity is significantly induced ($P = 2.41 \times 10^{-8}$, $Z$-score $= 3.789$) and TP53 activity is potentially repressed ($P = 9.30 \times 10^{-8}$, $Z$-score $= -1.034$) during early carcinogenesis, and that their activity remains altered throughout tumorigenesis. Importantly, the gene expression levels of TP53 and MYC did not change significantly with respect to the pathologic continuum from normal to tumor, suggesting that the predicted changes in their activity are due to posttranscriptional regulation.

**Figure 2.** Experimental validation of CEACAM5 and SLC2A1 expression. A, qRT-PCR. Box plots represent RNA levels of CEACAM5 and SLC2A1 in normal basal cell and premalignant lesions from 6 patients, showing increased expression of both genes in premalignant lesions compared with normal basal cell. $B2M$ was used as the endogenous control. B and C, Immunofluorescent staining of CEACAM5 and SLC2A1. Protein staining shows increased expression of CEACAM5 and SLC2A1 in premalignant lesions and SCC compared with basal cell in the normal epithelium. Top rows, normal airway epithelium; middle rows, premalignant lesions; bottom rows, SCC. Left columns, KRT5, stained in red as marker for basal cell, premalignant lesions, and tumor cells; middle columns, SLC2A1 or CEACAM5, stained in green; right columns, merged images of left and middle columns. DAPI (4',6-diamidino-2-phenylindole), stained in blue, as nuclear marker. White scale bars, 50 μm. Insets show close-up views of the boxed regions.
To test the hypothesis that MYC activity is induced during early SCC carcinogenesis, immunofluorescent staining of MYC was performed to examine its nuclear and cytoplasmic localization in normal basal cell, premalignant lesions, and tumor cells (Fig. 4B). MYC staining was exclusive to the nuclei of premalignant lesions and tumor cells. In the histologically normal basal cell of the airways from patients with lung cancer, however, MYC was localized predominantly in the cytoplasm, although some areas of nuclear staining were also seen. The increased expression of MYC targets in the premalignant lesions and tumor cells, together with a concomitant increase in the nuclear localization of MYC, is strong evidence for a carcinogenesis-associated increase in MYC activity without a significant increase in gene expression.

Discussion

Little is known about the development of premalignant lesions and their progression to SCC because of a lack of appropriate in vitro and in vivo stepwise models of SCC tumorigenesis. The current practice of profiling whole-tissue biopsies has inherent limitations in the study of airway premalignancy, as such biopsy samples are highly heterogeneous (7–9) and are, therefore, potentially subject to confounding cell-type-specific effects. The approach described here allows the examination of specific cell populations along the continuum of lung carcinogenesis and the study of relationships between each of these populations. Furthermore, as the premalignant lesions are in close proximity to SCC within the same patients, it is reasonable to
expect that alterations in gene expression shared between premalignant and tumor cells reflect molecular changes that occur during carcinogenesis.

We focused specifically on the expression patterns of three genes, CEACAM5, SLC2A1, and PTBP3, that are upregulated in the premalignant lesions (and, in the case of SLC2A1, further upregulated in tumor cells). CEACAM5, a cell surface glycoprotein that plays a role in cell adhesion and intracellular signaling, has been shown to be important in other epithelial cell cancers, such as colon cancer (25). SLC2A1 (also known as glucose transporter 1, or GLUT1) is a facilitative glucose transporter associated with hepatocellular cancer and head and neck SCC (26, 27). PTBP3 (also known as regulator of differentiation 1, or ROD1) is an RNA-binding protein that regulates pre-mRNA alternative splicing and plays a role in the regulation of cell proliferation and differentiation (28, 29). The protein level expression of each gene was substantially increased in premalignant lesions and tumor cells, although the expression of CEACAM5 within the tumor cells was more heterogeneous than that of the other genes. In addition, although PTBP3 was strongly expressed in normal airway epithelium, its expression was restricted to columnar KRT5<sup>+</sup> cells.

We also examined the expression of these genes in publicly available microarray datasets related to SCC carcinogenesis. In one such experiment, the genes SLC2A1 and PTBP3 were significantly upregulated in lung SCC tumors relative to matched adjacent normal tissue, but unexpectedly, CEACAM5 was not. However, that study profiled tumor biopsies, which often contain significant stromal contamination; moreover, we observed substantial heterogeneity of CEACAM5 immunostaining in SCC tumor tissue in this study. The identification of CEACAM5 as an early-stage marker of squamous lung carcinogenesis in this study may, therefore, be attributable to the careful laser microdissection of SCC tumor cells from the surrounding stroma.

Because lung SCC is strongly associated with a history of tobacco smoking, we examined the relationship between smoking history and the expression of these genes in a previous study of bronchoscopic brushings. In that study, CEACAM5 and SLC2A1 were significantly upregulated in brushings from current smokers compared with those from never smokers. In a subsequent study from the same authors (30), the expression of CEACAM5 was reported to be irreversibly altered in former smokers for up to several decades after smoking cessation, suggesting that a smoking-associated increase in CEACAM5 expression in histologically normal airway epithelium may be an early event associated with carcinogenesis in these individuals.

We used GSEA to identify chromosomal regions that were enriched in differentially expressed genes, which suggested

![Figure 4. Identification of MYC as a dysregulated transcription factor in early carcinogenesis. A, identification of dysregulated transcription factors by IPA. Expression patterns of known downstream targets of TP53 and MYC suggest the inhibition of TP53 and the activation of MYC during early carcinogenesis. B, experimental validation of MYC activation in premalignant lesions and SCC. Immunofluorescent staining shows detection of MYC expression in the nuclei of premalignant lesions and SCC compared with cytoplasmic localization in basal cell. Top rows, normal airway epithelium; middle rows, premalignant lesion; bottom rows, SCC. Left columns, KRT5, stained in green as marker for basal cell, premalignant lesion, and tumor cells; middle columns, MYC, stained in red; right columns, merged images of left and middle columns. DAPI, stained in blue, as nuclear marker in all images. White scale bars, 50 μm. Insets show close-up views of the boxed regions.](image)
that the frequent 3p loss and 3q amplification that are characteristic of SCC (and rare in ADC; refs. 24, 31, 32) are late-stage events in SCC carcinogenesis. A relevant work by van Boerdonk and colleagues presented a longitudinal study of 6 patients with squamous metaplastic lesions that showed carcinoma in situ or carcinoma at follow-up bronchoscopy (33). These lesions showed 3p loss and 3q gain when compared with 23 lesions from subjects with no sign of cancer in their follow-up bronchoscopy. The potential discordant result about the timing of this genomic amplification of chromosome 3 (early event vs. late event in lung carcinogenesis) may be a result of the different experimental designs of both studies. The van Boerdonk study found chromosomal changes in squamous metaplastic lesions of subjects that had follow-up carcinomas when compared with lesions from subjects with no follow-up carcinomas, but they did not compare the premalignant lesions to normal airway epithelium or SCC from these same individuals to establish the potential stepwise chronology of this molecular event. In our study, we compared premalignant lesions with matched normal basal cells (early-stage) and matched carcinomas (late-stage), all within individual patients with SCC. Our data show chromosome 3 abnormalities during late-stage carcinogenesis, suggesting the possibility that the observed 3p loss and 3q gain could have happened at any time point during the progression from squamous metaplasia to carcinomas. We also used IPA to identify biologic functions and regulators that were overrepresented among the genes associated with early- or late-stage carcinogenesis. This analysis revealed that early-stage carcinogenesis is marked primarily by increased flux through the cell cycle, but that cellular proliferation continues throughout late-stage carcinogenesis.

Finally, we used IPA to make predictions about the upstream regulators that might be responsible for these changes, and identified the transcription factors TP53 and MYC as likely candidates based on the coordinate differential expression of their target genes. Nuclear expression of MYC in premalignant lesions and tumor cells, suggesting that activation of MYC by nuclear translocation could be an important event contributing to dysregulated cell-cycle progression during SCC carcinogenesis. Previous reports have also identified the potential importance of MYC in premalignant lesions of lung carcinoma (31) and breast cancer (34). Further analysis of other datasets is needed to validate the results.

Although this study represents a novel approach for identifying driver molecular events associated with squamous cell lung carcinogenesis, there are a number of important limitations to the work. Our model assumes that there is a molecular relationship between the premalignant and tumor cells found within the airway of the same patient, although the lesions may develop from disparate clonal populations, thereby limiting the interpretation of those changes as reflecting a stepwise change between lesions. Longitudinal studies of premalignant lesions resampled over time are needed to identify molecular alterations associated with progression or regression within a clonal population of cells. Furthermore, our group and others have previously reported molecular alterations throughout the histologically normal airway of smokers with lung cancer (35). Those molecular events in the histologically normal “field of injury” may reflect some of the earliest events in carcinogenesis and will not be captured directly by our approach. Finally, our study did not evaluate the potential role of stromal cells in initiation and progression of SCC.

In summary, we present a novel approach to identify the molecular alterations that characterize premalignant lesions and carcinogenesis in lung SCC. By isolating and transcriptome profiling a progenitor cell population within normal airway epithelium, premalignant lesions and SCC from the same individual, we were able to provide unique insight into stepwise molecular alterations that occur during lung carcinogenesis. Our analysis identified coordinate changes in the activity of upstream regulators and the expression of downstream genes within the same patient during early- and late-stage carcinogenesis. Further studies that profile molecular alterations within an individual premalignant lesion followed serially over time (as it progresses or regresses) will provide further resolution to the molecular events associated with lung carcinogenesis. Additional work will also be necessary to determine whether any of the genes identified in our study can be used to distinguish premalignant lesions that will progress to cancer from those that will regress. Genes identified and validated in this manner might serve as early biomarkers for SCC detection and targets for SCC chemoprevention.

Disclosure of Potential Conflicts of Interest
M.E. Lenburg has received a commercial research grant from Johnson and Johnson, has ownership interest (including patents) and is a consultant/advisory board member of Allegro Diagnostics. A.E. Spira has ownership interest (including patents) and is a consultant/advisory board member of Allegro Diagnostics. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions
Conception and design: A.T. Ooi, D.A. Elashoff, T.C. Walser, S.M. Dubinett, A. Spira, B.N. Gomperts
Development of methodology: A.T. Ooi, A.C. Gower, B.N. Gomperts
Acquisition of data (acquired and managed patients, provided facilities, etc.): A.T. Ooi, J.L. Vick, B. Nagao, T.C. Walser, A. Spira
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.T. Ooi, A.C. Gower, K.X. Zhang, W.D. Wallace, D.A. Elashoff, M. Pellegrini, M.E. Lenburg, B.N. Gomperts
Writing, review, and/or revision of the manuscript: A.T. Ooi, A.C. Gower, W.D. Wallace, D.A. Elashoff, M. Pellegrini, M.E. Lenburg, B.N. Gomperts
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.T. Ooi, A.C. Gower, I. Hong
Study supervision: A.T. Ooi, M.E. Lenburg, A. Spira, B.N. Gomperts

Acknowledgments
The authors thank Gang Liu and Linggi Luo for performing the RNA-seq of patients 1, 2 and 4. UCLA Broad Stem Cell Research Center High-Throughput Sequencing Core Resource for performing the RNA-seq for patient 3, as well as Frank Schembri for technical assistance in optimizing RNA extraction and Josh Campbell and Nacho Caballero for assistance with negative binomial analysis. Histologic services were performed by the UCLA Translational Pathology Core Laboratory (TPCL) and Laser microdissection was performed through the TPCL and the Advanced Light Microscopy Core. The statistical analysis for qRT-PCR was performed by Tristan Grogan and supported by the NIH/National Center for Advancing Translational Science (NCATS) UCLA CTSI grant number UL1TR000124.
Grant Support
This work was supported by the Department of Defense (DOD) CTRA LC090615 (to S.M. Dubinett, A. Spira, B.N. Gomperts, and T.C. Walser), Tobacco-Related Disease Research Program (TRDRP, 19FT-0046; to A.T. Ooi), National Heart, Lung, and Blood Institute (NHLBI) R01HL094561-01, CIRM RN2-009-04 (to B.N. Gomperts), the National Cancer Institute (NCI; #101-CA152751), to S.M. Dubinett, A. Spira, B.N. Gomperts, and T.C. Walser), the Department of Defense (DOD; #W81XWH-10-1-0106; to S.M. Dubinett, A. Spira, and T.C. Walser), Department of Veteran Affairs (VA; #501BX000359; to S.M. Dubinett and T.C. Walser), and TRDRP (#20KT-0055; to T.C. Walser).

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 October 23, 2013; revised February 24, 2014; accepted February 25, 2014; published OnlineFirst March 11, 2014.

References
Molecular Profiling of Premalignant Lesions in Lung Squamous Cell Carcinomas Identifies Mechanisms Involved in Stepwise Carcinogenesis

Aik T. Ooi, Adam C. Gower, Kelvin X. Zhang, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-13-0372

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2014/03/12/1940-6207.CAPR-13-0372.DC1

Cited articles
This article cites 35 articles, 16 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/7/5/487.full.html#ref-list-1

Citing articles
This article has been cited by 9 HighWire-hosted articles. Access the articles at:
/content/7/5/487.full.html#related-urls

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.