A Preliminary Transcriptome Map of Non-Small Cell Lung Cancer

Takeshi Fujii, Tatiana Dracheva, Audrey Player, Susan Chacko, Robert Clifford, Robert L. Strausberg, Kenneth Buetow, Norio Azumi, William D. Travis, and Jin Jen

Laboratory of Population Genetics, Center for Cancer Research [T. F., T. D., A. P., R. C., K. B., J. J.], Cancer Genomics Office [R. L. S.], National Cancer Institute, and Center for Information Technology [S. C.], NIH, Bethesda, Maryland 20892; Department of Pathology, Georgetown University Medical Center, Washington, D.C. 20007 [N. A.]; and Department of Pulmonary and Mediastinal Pathology, Armed Forces Institute of Pathology, Washington, D.C. 20306-6000 [W. D. T.]

Abstract

We constructed a genome-wide transcriptome map of non-small cell lung carcinomas based on gene-expression profiles generated by serial analysis of gene expression (SAGE) using primary tumors and bronchial epithelial cells of the lung. Using the human genome working draft and the public databases, 25,135 nonredundant UniGene clusters were mapped onto unambiguous chromosomal positions. Of the 23,056 SAGE tags that appeared more than once among the nine SAGE libraries, 11,156 tags representing 7,897 UniGene clusters were positioned onto chromosomes. A total of 43 and 55 clusters of differentially expressed genes were observed in squamous cell carcinoma and adenocarcinoma, respectively. The number of genes in each cluster ranged from 18 to 78 in squamous cell carcinomas and from 20 to 165 in adenocarcinomas. The size of these clusters varied from 1.8 Mb to 65.5 Mb in squamous cell carcinomas and from 1.6 Mb to 98.1 Mb in adenocarcinomas. Overall, the clusters with genes over-represented in tumors had an average of 3–4-fold increase in gene expression compared with the normal control. In contrast, clusters of genes with reduced expression had about 50–65% of the gene expression level compared with the normal. Examination of clusters identified in squamous cell lung cancer suggested that 9 of 15 clusters with over-expressed genes and 13 of 28 clusters with underexpressed genes were concordant with previously reported cytogenetic, comparative genomic hybridization or loss of heterozygosity studies. Therefore, at least a portion of the gene clusters identified via the transcriptome map most likely represented the transcriptional or genetic alterations occurred in the tumors. Integrating chromosomal mapping information with gene expression profiles may help reveal novel molecular changes associated with human lung cancer.

Introduction

Lung cancer accounts for >150,000 cancer deaths each year in the United States alone. NSCLC is the predominant form of lung cancer and consists of two major histological subtypes: squamous cell carcinoma and adenocarcinoma. Cytogenetic studies have shown that most NSCLCs displayed complex karyotypes with multiple numerical and structural alterations (1, 2). LOH (3–5) and CGH studies (6–11) showed most NSCLCs displayed complex karyotypes with multiple numerical and structural alterations (1, 2). LOH (3–5) and CGH studies (6–11) showed that gene alterations in cancer, Pollack et al. (14) developed a cDNA microarray-based CGH method that used 3360 cDNAs representing 3195 RH-mapped genes as probes in an effort to determine DNA copy-number variation in cancer. Caron et al. (15) constructed a genome-wide mRNA expression profile (transcriptome map) of several tumors using the publicly available SAGE tag-mapping database. By assigning genes to the RH-based chromosomal position, clusterings of highly expressed genes to specific chromosomal regions were observed. Completion of the initial human genome working draft (16, 17) has now made it possible to directly assign the genome-wide high-throughput gene expression profiles to the human genome working draft sequence. Using nine different SAGE libraries generated from primary lung tumors and normal respiratory epithelial cells, we have integrated gene expression profiles with the human genome working draft sequence. We observed clusters of genes that were differentially expressed between the tumors and the corresponding normal epithelial cells on the resulting transcriptome map. Many of these clusters corresponded to regions of gene amplification or deletion demonstrated previously by CGH and/or LOH analyses, indicating the clustering of gene expression changes in the specific chromosome regions may represent the underlying genetic and transcriptional alterations in lung cancer.

Materials and Methods

Tissue and SAGE Libraries. Primary lung tumor tissues used for SAGE were obtained after surgical resection at Johns Hopkins Hospital, Baltimore, MD, as described previously (12). SAEC and NHBE cells were purchased from Clonetics/BioWhittaker (Walkersville, MD) and propagated following the manufacturer’s instruction. SAGE libraries were generated using a total of nine different samples including two of each from primary cultures of normal small or large airway epithelial cells (SAEC and NHBE, respectively), two of each from primary squamous cell carcinomas and adenocarcinomas of the lung, and a ninth library derived from the A549 lung adenocarcinoma cell line. A total of 374,643 tags were sequenced. Approximately 55,000 tags were analyzed for squamous cell carcinomas, NHBE cells, and A549 cell libraries, whereas about 22,000 tags were generated from each adenocarcinoma and SAEC libraries (12, 13).

Public Databases and Software. Gene identity and UniGene Cluster assignment of each SAGE tag were obtained using the SAGEmap reliable tag-to-gene mapping table (dated June 19, 2001). The expression level of each transcript was normalized in each experiment to represent the occurrences of each transcript per 100,000 transcripts in the library. The public databases used in this study were the April 1, 2001 freeze of the UCSC human genome working draft sequence and its annotation database (dated July 16, 2001).4 UniGene Clusters (build #138, dated July 29, 2001),5 and RefSeq (dated June 15, 2001).6 Sequence-matching searches over the human genome working draft sequence were performed using a BLAT program (courtesy of W. James Kent, UCSC, Santa Cruz, CA) that was implemented onto the NIH Biowulf/ Lobos3 Beowulf superclusters. Construction of the transcriptome-mapping

Received 1/3/02; accepted 4/25/02.

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 1754 solely to indicate this fact.

1 To whom requests for reprints should be addressed, at Laboratory of Population Genetics, Center for Cancer Research, National Cancer Institute, NIH, 41 Library Drive, Building 41/Room D702, Bethesda, MD 20892-5060. Phone: (301) 435-8958; Fax: (301) 435-8963; E-mail: jenj@mail.nih.gov.

2 The abbreviations used are: NSCLC, non-small cell lung cancer; SAGE, serial analysis of gene expression; LOH, loss of heterozygosity; CGH, comparative genomic hybridization; EST, expressed sequence tag; RH, radiation-hybrid; NHBE, normal human bronchial epithelial; SAEC, small airway epithelial cell; UCSC, University of California Santa Cruz; RefSeq, Reference Sequences.


4 http://genome.ucsc.edu/.


Downloaded from cancerres.aacrjournals.org on July 13, 2017. © 2002 American Association for Cancer Research.
Results and Discussion

Assignment of UniGene Clusters on the Human Genome Working Draft. An outline for the construction of the transcriptome map is shown in Fig. 1. We assigned UniGene clusters onto chromosome positions, as chromosomal coordinates, based on the UCSC human genome working draft sequence. A total of 1,271,925 accessions consist of Known Genes, mRNA, or spliced EST entities in the annotation of the working draft sequence were used. Additionally, 371,123 UniGene and RefSeq sequences not included in the UCSC annotations at the time of this study were sequence-matched over the human genome working draft sequence using the BLAT program that was implemented on the NIH Biowulf/LOBOS3 Beowulf supercluster. Only the accession entry with the highest matching score was chosen when BLAT outputs were redundant. To reduce redundancy, accessions that had a shared UniGene cluster identity and a positional overlap were joined as a single positional cluster. Positional clusters that shared the same strand orientation and at least one common exon were also joined. Those that were represented by only one EST accession were removed to reduce the possibility of contaminant genomic DNA segment. When UniGene clusters matched to multiple positions, the positional clusters derived from the most reliable category (i.e., RefSeq > mRNA > spliced EST) were used. Clusters having the same UniGene cluster identity and located within a 1-Mb distance were joined together. In doing so, 28,223 unique chromosomal positions were assigned to represent 27,542 UniGene clusters, and 26,992 of these clusters had a unique correspondence to chromosomal positions. These figures were considered as reasonable given the current estimate of 25,000–40,000 genes predicted in the human genome (16, 17).

Generation of the Transcriptome Map for NSCLC on the Basis of SAGE. As summarized in Fig. 1, a total of 374,643 tags from nine lung SAGE libraries were included for the construction of NSCLC transcriptome map. SAGE analysis of five cancer tissues and four normal respiratory epithelial cells identified 66,501 distinct tags. After removing tags that matched to mitochondrial DNA or ribosomal RNA sequences, 23,056 of the unique tags appearing more than once among all nine of the libraries were subjected to UniGene cluster assignment using the SAGEmap reliable tag-to-gene mapping table. A total of 18,595 tags were assigned to one or more UniGene clusters, and 11,156 of these tags, representing 7,097 UniGene clusters, were assigned to unique chromosomal positions. The remaining 7,439 tags were excluded because of their multiple assignments to the genome. Of these clusters, 6,501 UniGene clusters were expressed in either squamous cell carcinoma and/or SAEC cells. The discrepancy among the numbers of distinct SAGE tags, UniGene clusters, and chromosomal positions could be accounted for in part by alternatively spliced transcripts and the presence of multiple polyadenylation sites, which can result in multiple SAGE tags for the same gene (18). The resulting transcriptome map from SAGE in NSCLC is shown in Fig. 2.

Clustering of Differentially Expressed Genes. Similar to previous reports (12, 19), a majority of the genes were expressed at compatible levels between the tumors and the corresponding normal cells based on SAGE tag counts. However, clustering of genes with similar expression pattern was present along segments of many chromosomes. To identify chromosomal clustering of highly differentially expressed genes, we used an unbiased approach based on a moving-median of the gene expression levels and then tested the window sizes and folds of differential expression based on existing cytogenetic, CGH, and LOH data. Not surprisingly, most chromosomes showed clustering of genes differentially expressed between normal and tumor tissues. For squamous cell carcinoma, 43 clusters were observed among the 6,501 UniGene clusters analyzed, whereas 55 clusters of genes were observed among the 5,512 UniGene clusters used in adenocarcinoma. The statistical significance of the resulted gene clustering was evaluated using the Monte-Carlo simulation to determine whether the observed clustering of differentially expressed genes was a random event (Table 1). In this evaluation, we regarded all of the chromosomes as a continuum of ordered genes and permuted the genomic order of all of the genes for squamous cell carcinoma or adenocarcinoma. A total of 10,000 simulations were performed to determine the incidence of clustering for each cancer type. For squamous cell carcinoma, we identified 16 clusters with increased gene expression and 30 clusters with reduced gene expression in the tumors. In contrast, Monte-Carlo simulation expected an average of 6.5 and 15.0 clusters, respectively, by random permutation. Therefore, it was highly unlikely that the clusters identified via the transcriptome map resulted from random variation in the gene clustering over the genome ($P = 0.000064$ and $P = 0.000006$ for over- and underexpressed clusters, respectively).

Overall, the genes increased in tumor had an average of 3–4-fold increase in expression, whereas clusters with decreased expression contained genes with an average of 50–65% the normal level of expression. For adenocarcinoma, 41 and 23 clusters were observed for over- and underexpressed gene clusters, respectively, compared with 38.7 and 11.5 clusters by Monte-Carlo simulation. This result suggested that the observed clustering could be observed by chance alone for genes overexpressed in adenocarcinomas ($P = 0.5652$).
Several possible explanations for this observation are: (a) adenocarcinomas are intrinsically more similar to normal lung parenchyma because of their peripheral location; (b) the number of SAGE tags available in this tumor type may be insufficient for accurate, reliable analyses of differential expression along each chromosome; and (c) the variable histological subtyping of lung adenocarcinoma. Although the clusters with decreased gene expression in adenocarcinoma were reliable ($P = 0.00004$), we focused our analysis on squamous tumors. When each chromosome was considered separately, there were 15 clusters of genes with increased expression levels and 28 clusters of genes with decreased expression levels. No clustering was observed on chromosomes 6, 15, 18, 21, 22, and X.

**Clustering of Genes with Increased Expression in Squamous Cell Carcinoma.** Gene amplification is generally associated with tumor progression, occurring often in a subset of late-stage cancers, and has prognostic significance (20). Some of the amplicons are known to harbor oncogenes or growth-regulatory genes. Previous CGH analyses have shown that chromosomal arms 3q, 5p, 7p, and 8q were the most frequently over-represented in NSCLC, and gene amplification was often present in chromosomal regions 3q13, 3q26, 3q28-qter, 7q11.2, 8p11-12, 8q24, 12p12, and 19q13.1-13.2 (6). Of the 15 clusters containing genes over-represented in squamous cell carcinoma (Table 2), 9 were located at 1q, 3q, 5p, 7p, and 8q, and were consistent with previous reports (6, 7, 9–11).

Chromosome region 3q24-qter is one of the most frequent targets.
for amplification in squamous cell carcinomas of the lung, head and neck, and the cervix (21–24). On the basis of the literature, amplification occurs at several locations on 3q and is especially frequent at 3q26.2-q26.3. Several candidate targets for 3q26 amplicon in squamous cell carcinomas of the head and neck, and esophagus include PIK3CA (25), SKIL (SNO; Ref. 26), and the RNA component of the telomerase (27). Our transcriptome map revealed two distinct regions, 3q25-q26.3 and 3q27-q29 (clusters #13–14 and #15). On the basis of our SAGE data, the PIK3CA gene was not expressed in either normal or tumor tissues, whereas the SKIL gene was minimally expressed only in squamous cell carcinoma but not in NHBE cells. Therefore, other target gene(s) may remain to be discovered for 3q25-q26 amplicon. In contrast, the p53 homologue p63 gene, mapped at 3q28, is one of the likely candidates responsible for our observed amplification of 3q27-q29 (cluster #15) as it was 33-fold higher in the tumor by SAGE analysis (13) and is consistent with the published reports (28).

**Clustering of Genes with Decreased Expression in Squamous Cell Carcinoma.** To identify potential tumor suppressor loci involved in the pathogenesis of lung cancer, high-resolution (10 cM), genome-wide LOH analyses has been reported (5). Among 36 NSCLC cell lines, 45 minimal areas of allelic loss were identified including nine hotspots where >60% of the cell lines had LOH at one or more of these locations. These areas included 1p36.12-qter, 8p21.3-q23.1, 9p21.3-p22, 13q12-q14, 17p12-p13, 19p13.3, Xp-q21, and Xq22.1. Previous CGH analyses in NSCLC (7) and squamous cell carcinoma of the lung (9) showed frequent DNA copy-number decreases on chromosomes 1p21-p31, 2q34-q36, 3p, 4p, 4q, 5q, 6q14-q24, 8p, 9p, 10q, 13q13-qter, 18q12-qter, and 21q21. These regions...
Table 1 Validation of differently expressed gene clusters by Monte-Carlo simulation

<table>
<thead>
<tr>
<th>Cluster no.</th>
<th>Cyogenetic band</th>
<th>Chromosomal position</th>
<th>Status</th>
<th>Number of genes</th>
<th>Median T/N ratio</th>
<th>Notable genes/tumor suppressor/oncogenes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1p36.2–p36.3</td>
<td>Chr1:2205808–7866192</td>
<td>–</td>
<td>20</td>
<td>0.43</td>
<td>[TP73]</td>
<td>(5, 7, 9, 42)</td>
</tr>
<tr>
<td>2</td>
<td>1p36.1</td>
<td>Chr1:17749951–23736394</td>
<td>–</td>
<td>20</td>
<td>0.59</td>
<td>(5)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2p22–p23</td>
<td>Chr2:27865382–37322776</td>
<td>–</td>
<td>23</td>
<td>0.73</td>
<td>(7, 9)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2p11.2–q11.1</td>
<td>Chr2:80547384–96027377</td>
<td>–</td>
<td>25</td>
<td>0.52</td>
<td>(7, 9, 43)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2q14.2–q15.2</td>
<td>Chr2:121891208–1368535134</td>
<td>–</td>
<td>22</td>
<td>0.75</td>
<td>(7, 9)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2q35–q36</td>
<td>Chr2:223459509–231867469</td>
<td>–</td>
<td>20</td>
<td>0.70</td>
<td>(7, 9)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2q37</td>
<td>Chr2:238670940–247114424</td>
<td>–</td>
<td>22</td>
<td>0.60</td>
<td>(7, 9, 43)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3p22–p23</td>
<td>Chr3:29156374–34720761</td>
<td>–</td>
<td>22</td>
<td>0.76</td>
<td>MLHI</td>
<td>(5, 7, 9, 32–36, 44)</td>
</tr>
<tr>
<td>11</td>
<td>3p21</td>
<td>Chr3:50161501–53381366</td>
<td>–</td>
<td>21</td>
<td>0.74</td>
<td>SEMA3B, FUSI, RASSF1</td>
<td>(5, 7, 9, 32–36, 44)</td>
</tr>
<tr>
<td>12</td>
<td>3p21</td>
<td>Chr3:55712479–58799957</td>
<td>–</td>
<td>28</td>
<td>0.78</td>
<td>[0.9 Mb centromeric to APOC1]</td>
<td>(5, 7, 11)</td>
</tr>
<tr>
<td>16</td>
<td>4q15.3–p16</td>
<td>Chr4:6772157–85065969</td>
<td>–</td>
<td>16</td>
<td>0.58</td>
<td>(7, 9, 43)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5q35</td>
<td>Chr5:185901250–191841574</td>
<td>–</td>
<td>26</td>
<td>0.64</td>
<td>(5–7, 9, 11, 42)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>9p13–p21</td>
<td>Chr9:35863848–380391975</td>
<td>–</td>
<td>20</td>
<td>0.62</td>
<td>[11.5 Mb telomeric to CDKN2A]</td>
<td>(5, 9, 38, 39)</td>
</tr>
<tr>
<td>27</td>
<td>10q24</td>
<td>Chr10:104155005–111136770</td>
<td>–</td>
<td>39</td>
<td>0.68</td>
<td>[2.8 Mb telomeric to PTEN]</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>11q13</td>
<td>Chr11:79065935–84562231</td>
<td>–</td>
<td>23</td>
<td>0.70</td>
<td>KRT7, KRT5, KRT6A, KRT1, KRT4, KRT8, HOXC, PPN15, MTH1, MMP15, [2.6 Mb centromeric to CDH11]</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>12q13</td>
<td>Chr12:55578725–59469204</td>
<td>–</td>
<td>29</td>
<td>0.70</td>
<td>KRT7, KRT5, KRT14, KRT16, [BRC1], WNT3, HOBX7, [TOC], SEC14L1, TIMP2</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>16q12.2–q21</td>
<td>Chr16:64645711–68405819</td>
<td>–</td>
<td>25</td>
<td>0.66</td>
<td>(5–7, 9, 11, 42, 44)</td>
<td></td>
</tr>
</tbody>
</table>

33–35d 17q21 Chr17:42070147–54387502 – 78 7 41 30 0.72 [0.9 Mb centromeric to APOC1] (5, 7, 11)

36 17q25 Chr17:81703388–85614916 – 18 4 12 2 0.57 [PKI1, SKI, SKIL, TP63] (6, 7, 9–11, 44)

37 19p13.3 Chr19:3149667–5728422 – 25 4 14 7 0.74 [6, 7, 9–11, 44]

38 19p13.1 Chr19:17718229–21340745 – 20 2 12 6 0.76 [6, 7, 9–11, 44]

39 19q13.1 Chr19:36880234–46441554 – 40 7 25 8 0.60 [6, 7, 9–11, 44]

40 19q13.1–q13.2 Chr19:51112826–54555604 – 22 5 12 5 0.59 [6, 7, 9–11, 44]

41 19q13.3 Chr19:59444162–63343735 – 41 0 26 15 0.60 [6, 7, 9–11, 44]

42 20q12–q13.1 Chr20:41509642–49168634 – 38 10 22 6 0.61 [6, 7, 9–11, 44]

43 20q13 Chr20:62621763–64468901 – 17 5 10 2 0.58 [6, 7, 9–11, 44]

44 3q21–q43 Chr3:259270782–274578830 + 23 12 12 9 3.20 [6, 7, 9–11, 44]

45 7q21–q23 Chr7:12570929–194025818 + 29 15 11 13 2.76 [6, 7, 9–11, 44]

46 13–14 Chr13:62192762–103593987 + 37 20 3 14 3.52 [6, 7, 9–11, 44]

47 3q27–q29 Chr3:197917861–215639073 + 53 31 4 18 3.52 [6, 7, 9–11, 44]

48 4q21–q24 Chr4:80267829–105875989 + 30 16 7 7 3.05 [6, 7, 9–11, 44]

49 18–19 Chr18:663141–72945158 + 64 39 7 18 3.41 [6, 7, 9–11, 44]

50 21 Chr21:4316341–30793763 + 39 23 4 12 3.29 [6, 7, 9–11, 44]

22–23 8q21.1–q24.1 Chr8:77518537–128248010 + 56 30 4 22 2.96 [6, 7, 9–11, 44]

51 9p12–q22 Chr9:40952402–91352574 + 30 16 2 12 3.16 [6, 7, 9–11, 44]

52 9q31–q32 Chr9:107340066–115631474 + 21 13 3 5 3.17 [6, 7, 9–11, 44]

53 11q21–q14 Chr11:23162216–239892274 + 27 14 6 7 2.81 [6, 7, 9–11, 44]

54 14q23–q24 Chr14:5510283–63535623 + 25 13 2 10 2.97 [6, 7, 9–11, 44]

**Notes:**
- a: Chromosomal positions are indicated as nucleotide coordinates based on the UCSC Human Genome Working Draft (April 1, 2001 freeze).
- b: Genes in brackets are within the cluster but not observed in the lung SAGE libraries.
- c: Publications with results consistent with those observed in this study.
- d: Genes between parentheses indicate clusters within the cluster but not observed in the lung SAGE libraries.
- e: Gaps up to 1.3 Mb exist among the clusters.
- f: The values shown are higher than the actual number of gene clusters because all of the chromosomes were regarded as a continuum and a few clusters had boundaries between chromosomes.
were common regions of allelic losses believed to harbor either known or unidentified tumor suppressor genes. Our results showed 28 clusters of underrepresented genes throughout the genome in squamous cell carcinoma. Thirteen of these 28 clusters were concordant with previous reports by either LOH and/or CGH analyses.

Chromosome arm 1p is one of the most frequently deleted chromosomal regions in various neoplasms including neuroblastoma, breast, and lung cancers. Although TP73 is located at 1p36, no inactivation of TP73 by either somatic mutation or DNA methylation has been observed (29). In squamous cell carcinoma, we observed cluster #1 that corresponded to the shortest region of deletion located between D15S507 and TP73 at 1p36.2 (29). This cluster included TP73 but was located outside of the PRDM2 (RIZ1) gene inactivated by promoter methylation in liver and breast cancers (30).

Our SAGE-derived transcriptome map displayed three clusters of decreased gene expression on chromosome arm 3p (clusters #10–12). Cluster #10 at 3p24-p22 contained the MLH1 gene, which has been reported as deleted in association with the presence of p53 mutations and tobacco exposure (31). Although there was no difference in the expression levels of the MLH1 gene between normal and tumor tissues, genes surrounding MLH1 were underexpressed in our SAGE data. Cluster #12 corresponded to 3p21, a region most commonly deleted in lung cancer (32). On the basis of our SAGE data, the MST1R and RASSF1A genes at 3p21 were overexpressed in squamous cell carcinoma, whereas the SEMA3F, SEMA3B, IFRD2, FUS1, PL6, and MAPKAPK3 genes were underrepresented (32–36). Consistent with this observation, both SEMA3B and FUS1 genes have been shown to be epigenetically silenced by promoter methylation in lung cancers (32, 35, 36).

Allelic loss on chromosome arm 10q, especially at 10q21-qter, has been observed by LOH and CGH studies in various malignancies including renal cell carcinoma, bladder, endometrial as well as prostatic cancers, glioma, malignant melanoma, and lymphoma, as well as squamous cell carcinomas of the head and neck and NSCLCs (5, 9, 37, 38). LOH at 10q has been associated with a metastatic phenotype and poor prognosis in tumors of the upper respiratory tract (38, 39). The PTEN/MMAC1 gene is a candidate tumor suppressor responsible for the 10q23.3 deletion because it is mutated in multiple advanced cancers including renal cell carcinoma, prostatic cancer, breast cancer, and glioma. The PTEN gene was associated with the tumorigenesis of squamous cell carcinoma in the lung, and head and neck (37). However, other studies (39, 40) indicated that the PTEN gene was normally transcribed and expressed despite the presence of LOH close to the locus (40). In addition, genetic alterations were reported to be infrequent in squamous cell carcinomas of the head and neck (38), and lung (39). These results suggested that tumor-suppressor gene(s) other than PTEN, involved in the malignant progression of tumors of the upper respiratory tract, might exist in 10q. Our transcriptome map showed that the PTEN gene was located 8.2 Mb telomeric to the cluster at 10q24 (cluster #27), and PTEN gene expression level was 2-fold higher in squamous tumor than in normal cells. Therefore, other genes more proximal to PTEN may be targeted for loss in NSCLC.

**Implications.** We have constructed a transcriptome map based on gene expression profiles generated by SAGE in squamous cell carcinomas, adenocarcinomas, and normal lung epithelial cells. To validate the lung transcriptome map that was generated using the human genome working draft sequence, we compared it to the previous tag assignments of the Cancer Genome Anatomy Project, which was developed based on the Genebridge4 RH-map (42). Physical assignment of the genes to different chromosomes was observed in 1.9% of the UniGene clusters present on the RH-based Cancer Genome Anatomy Project map indicating the chromosome assignment of SAGE tags was at least 98% accurate. Furthermore, similar to the RH-based transcriptome map (15), we observed an uneven distribution of differentially expressed genes throughout the genome as represented by clusterings of highly expressed genes to specific chromosomal regions. This observation is consistent with the fact that genetic alterations often affect a set of genes closely positioned in the genome. The knowledge of such clustering along the chromosomes may provide alternative markers for cancer detection and prognosis. However, the identification of a particular gene cluster can only be suggestive of a potential chromosomal region with relatively consistent gene expression patterns because every cluster contained at least some genes with expression patterns different or unchanged from others in the same cluster.

In squamous tumors, about half of the clusters identified were consistent with the previous reports by CGH and/or LOH analyses. It is highly likely that some of the remaining clusters may represent novel gene expression changes at chromosome regions not implicated previously in NSCLCs. It is also possible that some clusters may have been identified by chance alone. Another possible caveat is the fact that not all highly expressed genes are amplified, nor are all amplified genes highly expressed (14). Nonetheless, our analyses suggest that a substantial number of gene clusters were supported by other studies, and some of them are likely involved in tumorigenesis and progression of lung cancers.

It is worth noting that the transcriptome map we have generated appeared to be more sensitive in detecting genes with increased expression levels, because no clustering was seen in several regions that were known to be deleted in squamous cell lung cancers. This lower sensitivity in detecting deleted regions on the transcriptome map could result if the normal expression levels of the genes were already low or close to the background. This fact is also shown in Table 1 where the ratio of gene expression was 3–4-fold for overexpressed gene clusters but only <50% for those with reduced expression. Therefore, it appears that overexpressed genes were more likely identified, whereas those having reduced expression could have been under-represented on the transcriptome map. Additional studies using LOH, CGH, or fluorescence in situ hybridization will be needed to examine the candidate cluster regions to better understand the positional clustering of gene expression changes with the chromosomal alterations at the corresponding region.

Finally, the accuracy of the present transcriptome map is entirely dependent on and subject to the completeness of the human genome working draft and its annotations. It is limited by the methods of choice for tag to UniGene assignment and criteria for redundancy reduction of the UniGene clusters. The detection of gene clusters with the most differential gene expression is also subject to the method of analysis, the size and reliabilities of the SAGE libraries, and the number of genes, as well as the genomic length of the clusters. Although the window size and the cluster identification criteria could affect the resulting clusters, altering the parameters almost always resulted in statistically significant clustering of genes for squamous cell carcinomas. In contrast, clusters of genes overexpressed in adenocarcinomas appeared to be much less reliable. Identification of gene clusters for this tumor type may rely on the identification of additional SAGE tags from the respected libraries or the inclusion of additional samples. Nevertheless, the results presented here suggest that the ability to map gene expression profiles onto specific chromosome locations will likely facilitate the identification of novel genetic changes that underlie lung tumorigenesis and the use of this knowledge to guide clinical management of the cancer patients.
References


A Preliminary Transcriptome Map of Non-Small Cell Lung Cancer

Takeshi Fujii, Tatiana Dracheva, Audrey Player, et al.


Updated version  Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/62/12/3340

Cited articles  This article cites 45 articles, 22 of which you can access for free at: http://cancerres.aacrjournals.org/content/62/12/3340.full#ref-list-1

Citing articles  This article has been cited by 8 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/62/12/3340.full#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.