

Novel Integrative Methods for Gene Discovery Associated With Head and Neck Squamous Cell Carcinoma Development

Ian M. Smith, MD; Suhail K. Mithani, MD; Chunyan Liu, MD; Steven S. Chang, MD; Shahnaz Begum, MD; Mousiumi Dhara, BS; William Westra, MD; David Sidransky, MD; Joseph A. Califano, MD

Objective: To find head and neck squamous cell carcinoma (HNSCC) specific genetic changes using integrative genetics.

Design: Genetic analysis.

Patients: Three separate cohorts of patients with primary HNSCC were evaluated for expression-microarray of 33 000 genes (8 patients), quantitative real-time polymerase chain reaction (qRT-PCR) (36 patients), and quantitative DNA/qRT-PCR (12 patients). Controls with normal upper-aerodigestive mucosa were evaluated for expression microarray (6 patients) and qRT-PCR (7 patients).

Interventions: We utilized (1) prior reports of DNA loss and gain HNSCC accompanied by comparative genomic hybridization high-definition array data of the entire human genome, (2) a genome-wide survey of cancer-specific DNA mutations from the consensus cancer coding sequence (13 023 genes), and (3) our RNA expression microarray data of 33 000 genes to define candidate oncogenes activated by amplification or candidate tumor suppressor genes inactivated by deletion.

Main Outcome Measures: Gene expression in tissue measured by quantitative reverse transcriptase PCR. Gene copy number was measured by quantitative PCR.

Results: We found 20 genes that were in areas of demonstrated amplification or deletion overlapping with the somatic mutants from genome-wide screening of the consensus DNA cancer coding sequence reported by Sjöblom et al. Three were chosen for further study based on expression differences and proof of cancer causation from in silico study: *RUNX1T1*, *RFC4*, and *DLEC1*. From 12 patients with HNSCC, matched tumor DNA/RNA and leukocyte-derived DNA were studied. Six of 12 (50%) of the tumors demonstrated amplification of the *RUNX1T1* locus ($P = .01$), and 4 of those 6 (67%) demonstrated upregulated transcription of this gene ($P = .02$). Five of 12 (42%) of the tumors demonstrated amplification of the *RFC4* locus ($P = .03$), and 1 of those 5 (20%) demonstrated upregulated messenger RNA (mRNA) transcription of the gene ($P = .60$). Four of 12 of the tumors (33%) ($P = .05$) demonstrated deletion in the *DLEC1* locus (consistent with previously published 3p22 loss of 40%), and 3 of those 4 (75%) demonstrated reductions in mRNA expression ($P = .06$).

Conclusion: With the advent of high-throughput techniques to study cancer genetics, novel comparisons of large data sets using integrative methods may elucidate genetic alterations in HNSCC cancer.

Arch Otolaryngol Head Neck Surg. 2009;135(5):487-495

Author Affiliations:

Departments of Otolaryngology–Head and Neck Surgery (Drs Smith, Liu, Chang, Sidransky, and Califano and Ms Dhara), Surgery, Division of Plastic and Reconstructive Surgery (Dr Mithani), and Pathology (Drs Begum and Westra), Johns Hopkins Medical Institutions, Baltimore, Maryland.

HEAD AND NECK SQUAMOUS cell carcinoma (HNSCC) is the fifth most common cancer, and more than 500 000 cases are diagnosed annually worldwide. In the United States, it accounts for 3% of all cancers and 40 000 new cases each year.¹ Although considerable progress has been made in the areas of early detection, diagnosis, and treatment, the 5-year survival rate for patients has shown only modest improvement in the past 40 years.² Comprehensive analysis of clinical and treatment factors has shown tumor site-specific improvements in 5-year survival for cancers of the nasopharynx, oropharynx, and hypopharynx and late-stage laryngeal cancer.³

Exposure to tobacco and alcohol are the clinical factors most strongly associated with the development of tumors. In addition, many of the cancer genetic and epigenetic changes associated with sporadic HNSCC have been described, including *p53* (GenBank 7157), *VEGF* (GenBank 7422), and *cyclin D1* (GenBank 595).⁴ Yet, the molecular mechanisms of HNSCC carcinogenesis are still undergoing intensive investigation. Many heritable genetic alterations have been associated with cancer formation. Neoplastic cells undergo a variety of genetic alterations, from point mutations to chromosomal aberrations, that affect the function or expression of both oncogenes and tumor-suppressor genes. Chromosomal loss and

gain resulting in gene loss of heterozygosity and gene amplification, respectively, have been shown for years to result in gene silencing (Knudson's⁵ hypothesis) or gene overexpression.⁶ Amplification often results in increased messenger RNA (mRNA) transcription owing to an overabundance of promoter and template.⁷ Deletion can result in silencing via uniallelic or biallelic loss. Advances in techniques over the past few years, including array-based comparative genomic hybridization (aCGH) and array-based single-nucleotide polymorphism (aSNP), have allowed high-throughput, highly detailed studies of chromosomal loss or gain. For this study, we considered previously published areas of amplification (3q, 5p, 8q, 9q, and 20q)⁸⁻¹² and deletion (3p, 8p, 13q, and 18q)^{13,14} and performed a comprehensive search for the best source of high-definition chromosomal loss and gains in HNSCC (aCGH/aSNP).

For the past few years, efforts have been made that begin to define the consensus cancer coding sequence.¹⁵ The first effort in publishing of the human consensus cancer coding sequence showed somatic mutations in many known cancer-causing genes (eg, *p53* and *APC*), as well as discovery of somatic mutations in many new targets.¹⁶ This study group determined the sequence of well-annotated human protein-coding genes in 2 common tumor types and conducted an analysis of 13 023 genes in 11 breast and 11 colorectal cancers. A total of 189 genes (mean, 11 per tumor) were mutated at a notable frequency. Most of these genes were not previously known to be genetically altered in tumors. A wide range of cellular functions were implicated, including transcription, adhesion, and invasion.¹⁷ These types of studies find somatic DNA sequence mutations in cancers that provide a novel way to suggest function based on pressures of clonal selection in cancer. Similar efforts are under way by other organizations. New high-throughput screens of cancer genes are being developed at a rapid pace, creating the need for efficient approaches for integration of large data sets that use diverse technologies to describe genetic alterations in human cancers.¹⁶

Revelations of the human genome project and other recent advancements in technology have ushered in a new era of research. Genome-scale data sets of various forms are readily available to cancer researchers. Examples of such genome-scale data sets include aCGH (DNA loss or gain), RNA expression microarray (tissues or cell lines), small molecule cell line screens, and various proteomic approaches. Prior experience suggests that tumors may be susceptible to targeted therapies once their essential molecular alterations have been found. Integrative approaches to these genome-scale data sets allow multiple pieces of salient information to be combined in a manner that may yield novel and powerful new insights into biologic mechanisms of cancer.¹⁸ One prominent example of the usefulness of these integrative genetic approaches led to finding the oncogene *MITF* in melanoma. Researchers applied 2 genome-scale data sets (aSNP and expression microarray) to discover the oncogene *MITF* in melanoma. This gene may represent a new class of "lineage addiction oncogenes"—a fundamental tumor survival mechanism with important therapeutic implications.¹⁹ Other examples of integrative genomic ap-

proaches have also improved the genetic understanding of other cancers.²⁰⁻²³

In this study, we used an integrative genomic method to consider genes that are altered in HNSCC. We used primary data from 3 general sources: (1) genes mutated in the consensus cancer coding sequence,¹⁷ (2) DNA/chromosomal deletion or amplification in primary HNSCC, and (3) RNA gene expression differences in HNSCC and normal tissues. These were then integrated in an attempt to define novel genetic alterations (gene deletion or amplification) in HNSCC that result in important alterations in associated gene expression. Three genes were studied in detail: *RUNX1T1* (GenBank 862), a zinc finger transcription factor protein and translocation-activated oncogene in acute myeloid leukemia²⁴; *RFC4* (GenBank 5984), a gene involved in DNA replication, repair, modification, and chromatin modeling²⁵ noted to be overexpressed in other cancers,²⁶ particularly human papillomavirus-associated cancers²⁷; and *DLEC1* (GenBank 9940), a putative tumor suppressor involved in carcinogenesis of the lung, esophagus, kidney, and nasopharyngeal cancers.^{28,29}

METHODS

HISTOPATHOLOGIC FINDINGS

All samples were analyzed by the pathology department at Johns Hopkins Hospital, Baltimore, Maryland. Tissue samples were obtained via Johns Hopkins institutional review board-approved protocols under protocol No. 92-07-21-01. Normal samples were microdissected and DNA prepared from upper aerodigestive mucosa from healthy patients primarily undergoing uvulopalatopharyngoplasty procedures. Tumor samples were confirmed to be HNSCC and subsequently microdissected to separate tumor from stromal elements to yield at least 80% tumor cells. Tissue DNA was extracted as described in the subsection titled "DNA Extraction" in this section.

OLIGONUCLEOTIDE MICROARRAY ANALYSIS

Total cellular RNA was isolated using the RNeasy kit (Qiagen, Valencia, California) according to the manufacturer's instruction. We performed oligonucleotide microarray analysis using the GeneChip U133A and U133B plates on the Affymetrix (Santa Clara, California) expression microarray, which assays 33 000 genes. Samples were converted to labeled, fragmented, complementary RNA (cRNA) per the Affymetrix protocol for use on the expression microarray. Signal intensity and statistical significance were established for each transcript using dChip, version 2006 (free, downloadable software available at <http://biosun1.harvard.edu/complab/dchip>). Default settings for dChip were used, including the perfect match/mismatch difference model, invariant set normalization, and check single/probe/array outlier algorithm.

PUBLIC DATA SETS AND INTEGRATIVE GENETICS

The public databases used in this study were the University of California, Santa Cruz (UCSC), Human Genome reference sequence and the annotation database from the May 2004 freeze (hg17). For target discovery we used 3 data sets. First, we used all 1149 genes in the consensus cancer coding sequence from supplemental tables provided by Sjöblom et al¹⁷ from a survey

of 13 023 genes with DNA sequencing. High-definition detailed locations of DNA and chromosomal loss and gain were considered from the published areas in the report by Sparano et al,³⁰ which coincided with analysis from the literature of reliable areas of chromosomal loss and gain in HNSCC. This group used aCGH to develop a genome-wide molecular profile of oral squamous cell carcinoma from 21 prospectively collected fresh-frozen specimens, at a 0.9-Mb resolution to identify distinct regions of genomic alteration and their associated genes.³⁰ The consensus cancer coding sequence¹⁷ genes were located with data from UCSC genome browser and joined to areas of loss or gain based on chromosome and base pair location coinciding with areas of loss or gain. Any coding sequence overlap with the amplification or deletion was considered important. Gene RNA expression microarray analysis of 6 primary normal samples and 8 primary tumor specimens on the Affymetrix U133A and U133B platform (33 000 genes) was conducted. Microarrays were studied with dChip and invariant-set normalized. Median tumor expression, T_{exp} , and median normal expression, N_{exp} , were calculated. *P* values were determined using Stata statistical software (version 9.0; StataCorp LP, College Station, Texas). Validation of targets was initially based on putative cancer causation from a literature search and subsequently quantitative real-time polymerase chain reaction (qRT-PCR) for expression on primary tumor and normal samples, followed by quantitative polymerase chain reaction (qPCR) for precise validation of gene amplification or deletion, and/or sequencing of primary tumors (see the following 2 subsections).

DNA EXTRACTION

Samples were centrifuged and digested in a solution of detergent (sodium dodecylsulfate) and proteinase K for removal of proteins bound to the DNA. Samples were first purified and desalted with phenol/chloroform extraction. The digested sample was subjected twice to ethanol precipitation and subsequently resuspended in 500 μ L of low-salt Tris hydrochloride–ethylenediamine tetraacetic acid buffer (ethylenediamine tetraacetic acid, 2.5 mM/L, and Tris, 10 mM/L) and stored at -80°C .

qPCR AND qRT-PCR

The total DNA was measured and adjusted to the same amount for each tissue sample. The DNA was used as the templates for qRT-PCR with primers designed to specifically measure the DNA copy number of each candidate gene. β -Actin was examined to ensure accurate relative quantitation of copy number in qPCR. (Detailed PCR conditions and primer sequences are available from the authors on request.)

The total RNA was measured and adjusted to the same amount for each tissue sample, and then cDNA synthesis was performed by priming with oligo(dT) and the SuperScript First-Strand Synthesis kit (Invitrogen). The final cDNA products were used as the templates for subsequent PCR with primers designed specifically for each candidate gene. Glycerinaldehyde-3-phosphate dehydrogenase or β -actin was studied to ensure accurate relative quantitation in qRT-PCR. (Detailed PCR conditions and primer sequences are available from the authors on request.) The lesions of the 36 patients evaluated for qRT-PCR were 100% primary cancers; the patients were without prior treatment, 80.5% were male, and their median age was 58 years. There were 14 current smokers, 14 previous smokers, and 8 nonsmokers. All statistical computations were made with Stata SE software (version 9.0; StataCorp LP).

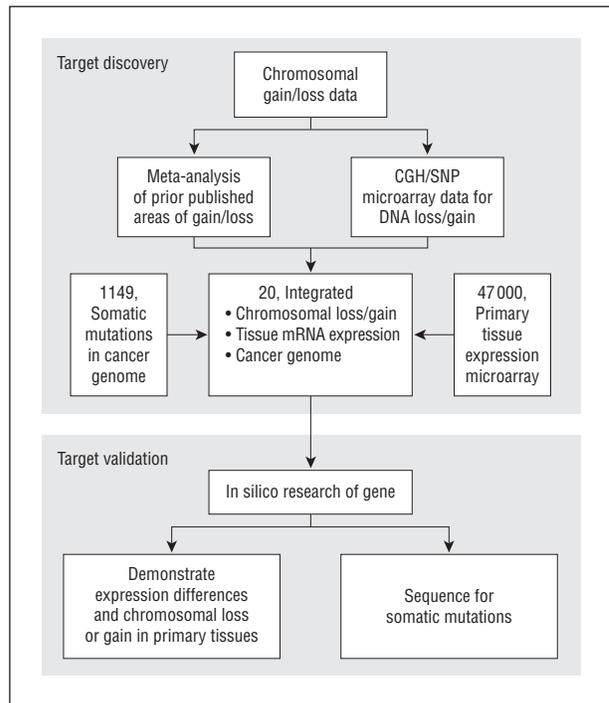


Figure 1. Integrative genetic approach. The process had 2 parts: target discovery and target validation. We used an integrative approach to identify genes in the consensus cancer coding sequence,¹⁷ located in areas of chromosomal loss or gain in head and neck squamous cell carcinoma.³⁰ We then performed gene expression microarray analysis of 6 primary normal samples and 8 primary tumor primary specimens on the U133A and U133B platform (Affymetrix, Santa Clara, California). Twenty targets were identified (see the Table). Validation of targets was initially based on putative cancer causation from a literature search, and subsequently quantitative real-time polymerase chain reaction (PCR) for expression on primary tumor and normal tissue, followed by quantitative PCR for precise validation of gene amplification or deletion, and/or sequencing of primary tumors. CGH/SNP indicates comparative genomic hybridization/single-nucleotide polymorphism; mRNA, messenger RNA.

RESULTS

We used an integrative bioinformatic approach to identify genes in the consensus cancer coding sequence,¹⁷ located in areas of chromosomal loss or gain, which had differential expression in HNSCC (**Figure 1**). The process had 2 parts: target discovery and target validation. Three data sets were incorporated into target discovery: (1) genes found in the consensus cancer coding sequence survey of 13 023 possible cancer-causing genes and DNA mutations,¹⁷ (2) regional DNA loss or gain of the whole cancer genome (using a literature search of known HNSCC amplicons and deletions with aCGH data from Sparano et al³⁰), and (3) RNA expression differences of 33 000 genes measured from expression microarray (Affymetrix, U133A and U133B plates) of 6 primary normal samples and 8 primary tumor specimens. Validation of targets was initially based on literature search, qRT-PCR for expression on primary tumor and normal samples, qPCR for validation of exact gene amplification or deletion, and/or primary tumor sequencing. The **Table** shows the 20 genes from the consensus cancer coding sequence,¹⁷ which are in areas of HNSCC chromosomal loss or gain. Of these 20 genes, 8 were part of areas of chromosomal deletion, and 12 were part of chromosomal amplification. Although all genes were validated so-

Table. Integrative Genetic Targets

| Gene Information | | | Sequencing in the Consensus Cancer Coding Sequence ^a | | | HNSCC Loss or Gain, Deleted/Amplified | HNSCC Expression | |
|------------------|---|--------------------|---|-------------------------|------------|---------------------------------------|-----------------------------|----------------------|
| Gene | Gene Name | Chromosome | Cell Type | Tumors With Mutation, % | CaMP Score | | Ratio Median T/N Expression | P Value ^b |
| <i>BRCA2</i> | Breast cancer 2, early onset | 13q13.3 – 13q14.11 | B | | | – | 0.42 | .003 |
| <i>CHL1</i> | Cell adhesion molecule with homology to L1CAM | 3p26.3 | C | 5.78 | 1.261 | – | 0.60 | .05 |
| <i>CLSTN2</i> | Calsyntenin 2 | 3q23 | C | | | + | | |
| <i>CMYA1</i> | Cardiomyopathy associated 1 | 3p22.1 – 3p22.3 | B | 7.93 | 1.360 | – | 0.92 | .44 |
| <i>DGKG</i> | Diacylglycerol kinase, gamma 90kD | 3q27.1 – 3q27.3 | B | | | + | 0.87 | .44 |
| <i>DLEC1</i> | Deleted in lung and esophageal cancer 1 isoform | 3p22.1 – 3p22.3 | B | | | – | 0.52 | .002 |
| <i>EIF4A2</i> | Eukaryotic translation initiation factor 4A | 3q27.1 – 3q27.3 | B | | | + | 0.82 | .02 |
| <i>EIF4G1</i> | Eukaryotic translation initiation factor 4 gamma, 1 | 3q27.1 – 3q27.3 | C | | | + | 1.05 | .43 |
| <i>FLJ10560</i> | Hypothetical protein LOC55171 | 3q27.1 – 3q27.3 | C | | | + | 1.31 | .52 |
| <i>ITGA9</i> | Integrin, alpha 9 precursor | 3p22.1 – 3p22.3 | B | 6.54 | 1.052 | – | 0.80 | .16 |
| <i>KCNB2</i> | Potassium voltage-gated channel, Shab-related | 8q13.3 – 8q21.11 | C | | | + | 1.23 | .20 |
| <i>KL</i> | Klotho isoform | 13q13.3 – 13q14.11 | C | | | – | | |
| <i>LIFR</i> | Leukemia inhibitory factor receptor | 5p13.1 – 5p13.2 | C | | | + | 0.84 | .61 |
| <i>MCF2L2</i> | Rho family guanine-nucleotide exchange factor | 3q26.33 | B/C | | | + | 1.00 | .90 |
| <i>MGC21688</i> | Hypothetical protein LOC131408 | 3q27.1 – 3q27.3 | B | | | + | | |
| <i>PIK3R4</i> | Phosphoinositide-3-kinase, regulatory subunit 4 | 3q22.1 – 3q22.2 | B | | | – | | |
| <i>RFC4</i> | Replication factor C 4 | 3q27.1 – 3q27.3 | B | | | + | 1.54 | .20 |
| <i>RTP1</i> | Receptor transporting protein 1 | 3q27.1 – 3q27.3 | B | | | + | 1.19 | .05 |
| <i>RUNX1T1</i> | Acute myelogenous leukemia 1 translocation 1 | 8q21.3 – 8q22.1 | C | 8.57 | 2.418 | + | | |
| <i>TGFBR2</i> | Transforming growth factor, beta receptor II | 3p23 – 3p24.1 | C | 8.72 | 2.845 | – | 0.65 | .03 |

Abbreviations: CaMP, cancer mutation prevalence; HNSCC, head and neck squamous cell carcinoma; L1CAM, neuronal cell adhesion molecule L1; N, normal; T, tumor; –, an area of chromosomal deletion; +, chromosomal amplification.

^aFrom supplemental tables provided by Sjöblom et al.¹⁷

^bP value calculated by Mann-Whitney U test.

matic mutations found in breast or colon cancer, only 5 genes were CAN genes found in the final validation set by Sjöblom et al.¹⁷ These genes and their cancer mutation prevalence score, which reflects the probability that the number of mutations observed in a gene reflects a mutation frequency that is higher than that expected to be observed by chance (frequency >1.0 was considered significant), were *CHL1* (1.26) (GenBank 10752), *CMYA1* (1.36), *ITGA9* (1.055), *RUNX1T1* (2.42), and *TGFBR2* (GenBank 7048) (2.85). These genes can be theoretically altered by mutation (missense, nonsense, etc), gene amplification or deletion, or via epigenetic activation or inactivation. All 20 genes may have functionally altering mutations in sporadic tumors, but gene amplification and deletion events tend to occur at a higher frequency in non-familial cancers, resulting in gene activation or inactivation. Five of these genes—*BRCA2* ($P=.003$), *CHL1* ($P=.05$), *DLEC1* ($P=.002$), *RTP1* (GenBank 132112) ($P=.05$), and *TGFBR2* ($P=.03$)—showed statistically significant differences in T_{exp} , compared with N_{exp} , that correlated with amplification (upregulated) or deletion (downregulated) status, based on Mann-Whitney U test. These genes are shown in **Figure 2**. One additional gene from the Table, *EIF4A2*

(GenBank 1974), is contained within a well-described 3q amplicon in HNSCC but showed statistically significant downregulation (median $T_{exp}/N_{exp}=0.82$; $P=.02$), which may reflect an area of chromosomal disruption. Also shown in Figure 2 are other genes with trends toward differential expression but which were not statistically significant: *LIFR* (median $T_{exp}/N_{exp}=0.84$; $P=.61$), *KCNB2* (median $T_{exp}/N_{exp}=1.23$; $P=.19$), *ITGA9* (median $T_{exp}/N_{exp}=0.80$; $P=.15$), and *RFC4* (median $T_{exp}/N_{exp}=1.54$; $P=.20$).

We chose a subset of targets to validate expression differences with qRT-PCR. Targets *DLEC1*, *RFC4*, and *RUNX1T1* were chosen because of expression differences and noted involvement in the pathogenesis of other cancer types. **Figure 3** shows qRT-PCR expression of *DLEC1*, *RFC4*, and *RUNX1T1*, all normalized by *GAPDH* expression in 36 primary tumor tissues, and 7 normal upper aerodigestive mucosal tissues. *DLEC1* is part of the 3p22 deletion found in HNSCC, and our hypothesis was that a subset of HNSCC primary tumors would show reduced expression. Thirteen of 36 tumors (36%) showed expression levels below all normal specimens. *RFC4* is found on an amplification of 3q27 in HNSCC, and we found that 9 of 36 tumors

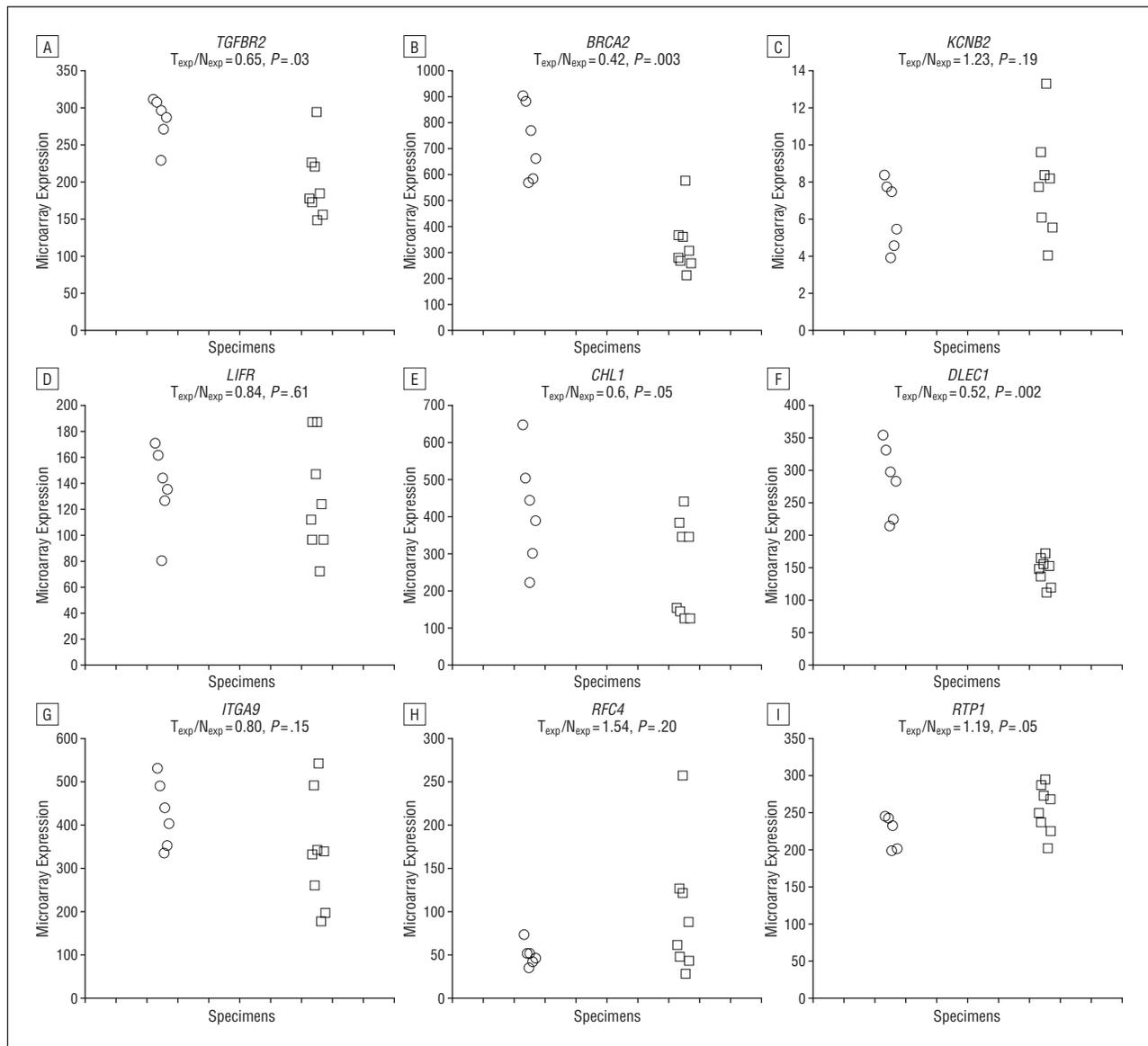


Figure 2. Expression analysis of target genes. Five of these genes, (A) *TGFBR2* ($P=.03$), (B) *BRCA2* ($P=.003$), (E) *CHL1* ($P=.05$), (F) *DLEC1* ($P=.002$), and (I) *RTP1* ($P=.05$), showed statistically significant differences in expression comparing tumor and normal expression (T_{exp} and N_{exp} , respectively) based on Mann-Whitney U test. This expression correlated with amplification (upregulation) or deletion (downregulation). Also shown are other genes with trends toward differential expression, but which were not statistically significant: (C) *KCNB2* ($P=.19$), (D) *LIFR* ($P=.61$), (G) *ITGA9* ($P=.15$), and (H) *RFC4* (median $T_{exp}/N_{exp}=1.54$; $P=.20$). Circles indicate normal specimens; squares, tumor specimens.

(25%) showed increased expression compared with the highest expression level found in the normal tissues. *RUNX1T1* is also found in an amplification, located on chromosome 8q. Eleven of the 36 tumor specimens (31%) showed expression levels higher than any normal samples.

We conducted sequencing of the *TGFBR2*. Of note, missense mutations are responsible for Lynch syndrome/HNSCC type 6, which confers an increased risk of multiple malignant lesions, primarily colon cancer, but also larynx cancer.³¹⁻³³ We were unable to find any mutations in the coding sequence of this gene in 12 primary HNSCCs. Significant expression differences were noted (see Figure 2 for P values).

Last, we tried to assess the relationship between DNA coding amplification or deletion and associated mRNA ex-

pression (**Figure 4**). Three genes were chosen for further study based on in silico genetic profiling: *RUNX1T1* (acute myelogenous leukemia translocation 1), *RFC4* (replication factor C4), and *DLEC1* (deleted in esophageal cancer isoform 1). Twelve primary HNSCC tissues had tumor DNA and RNA, matched leukocyte DNA extracted to study gene amplification and deletion, and resultant mRNA expression for each gene of interest. Six of the 12 tumors (50%) demonstrated amplification of the *RUNX1T1* locus of greater than 3.5 copies, which was not seen in the leukocyte samples ($P=.01$). Four of those 6 with amplification (67%) demonstrated upregulated transcription of this gene. Among the tissues with gene amplification, there was statistically significant overexpression of this gene ($P=.02$) by Mann-Whitney U test. Amplification of the *RFC4* locus of more than 6 copies was detected in 5 of 12 of the tumors (42%)

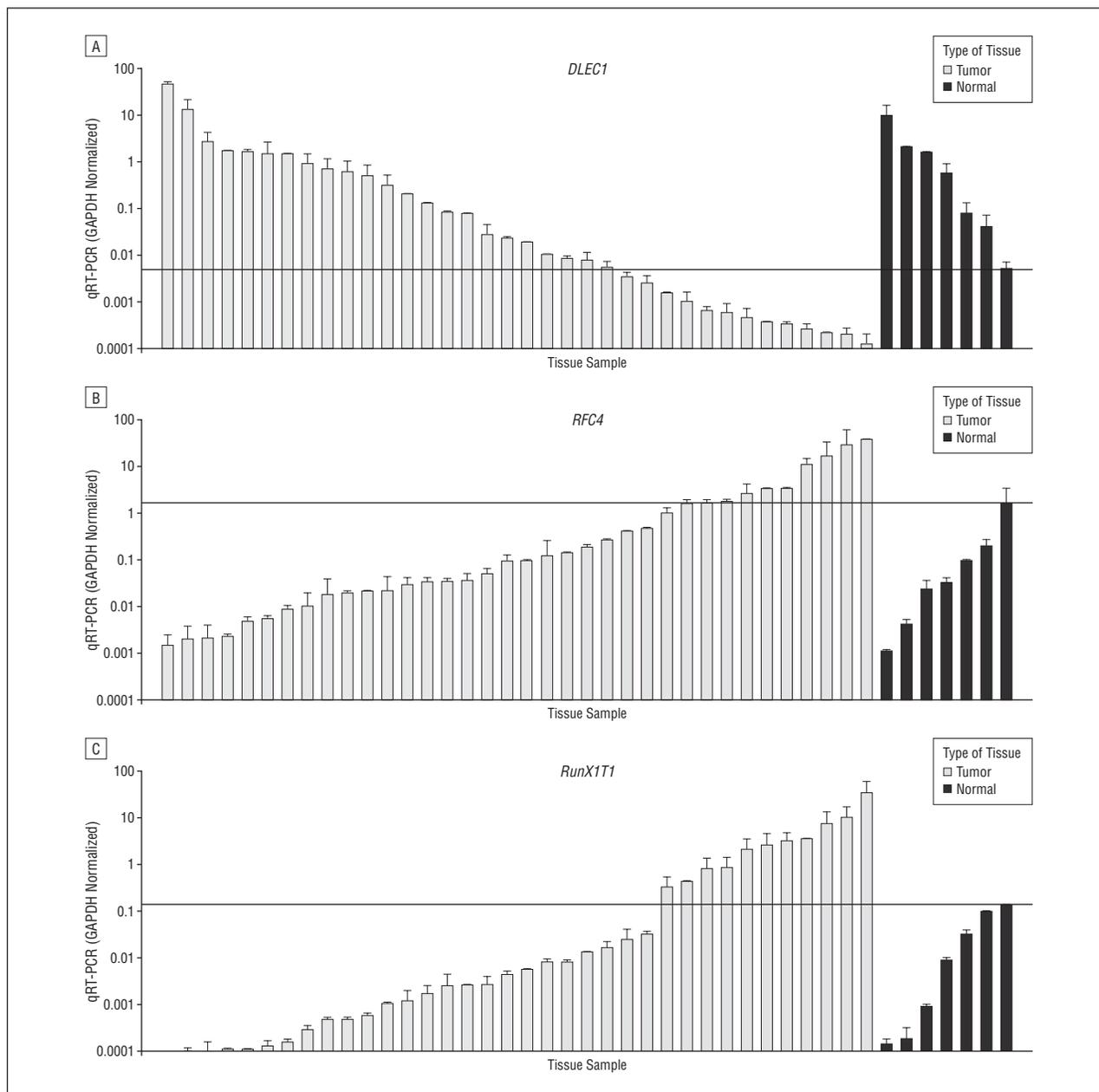


Figure 3. The quantitative real-time polymerase chain reaction (qRT-PCR) of selected targets. For initial validation of expression differences, we performed qRT-PCR on 36 primary head and neck squamous cell carcinoma (HNSCC) (tumor tissues) and 7 normal upper aerodigestive tissues (each sample is represented by a bar). Targets (A) *DLEC1*, (B) *RFC4*, and (C) *RUNX1T1* were chosen because of their novelty in the arena of HNSCC, involvement in the pathogenesis of other cancers, and expression differences noted on expression array, and the qRT-PCR expression, normalized by glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression are shown for each gene. Thirteen of 36 tumors (36%) showed expression levels below all normal specimens. *RFC4* shows increased expression compared with the highest expression level found in the normal tissues in 9 of 36 (25%). *RUNX1T1* showed overexpression in 11 of 36 of the tumor specimens (31%). The error bars indicate the standard error.

($P = .03$ by χ^2 test). One of those 5 (20%) demonstrated associated upregulated mRNA transcription. This was not statistically significant ($P = .60$). For the *DLEC1* locus, 4 of the 12 tumors (33%) vs 0 of 12 matched leukocyte DNA samples demonstrated deletion with a copy number of less than 1.5 ($P = .05$ by χ^2 test). This finding is consistent with previously published³⁴ 3p22 loss rates of 40%. Three of those 4 (75%) with copy-number reduction demonstrated mRNA expression below median levels for that gene. Samples with deletion of *DLEC1* had reduced expression of the gene that was not statistically significant ($P = .06$).

COMMENT

With the advent of novel high-throughput techniques to assay changes in tumor genomics, new and exciting methods of gene discovery in cancer are being developed. In 2005, Garraway et al¹⁹ performed an integrative comparison of expression microarray and aSNP in melanoma cell lines, which resulted in the discovery of the *MITF* oncogene that works in conjunction with the canonical *BRAF* (V600E) mutation to transform primary

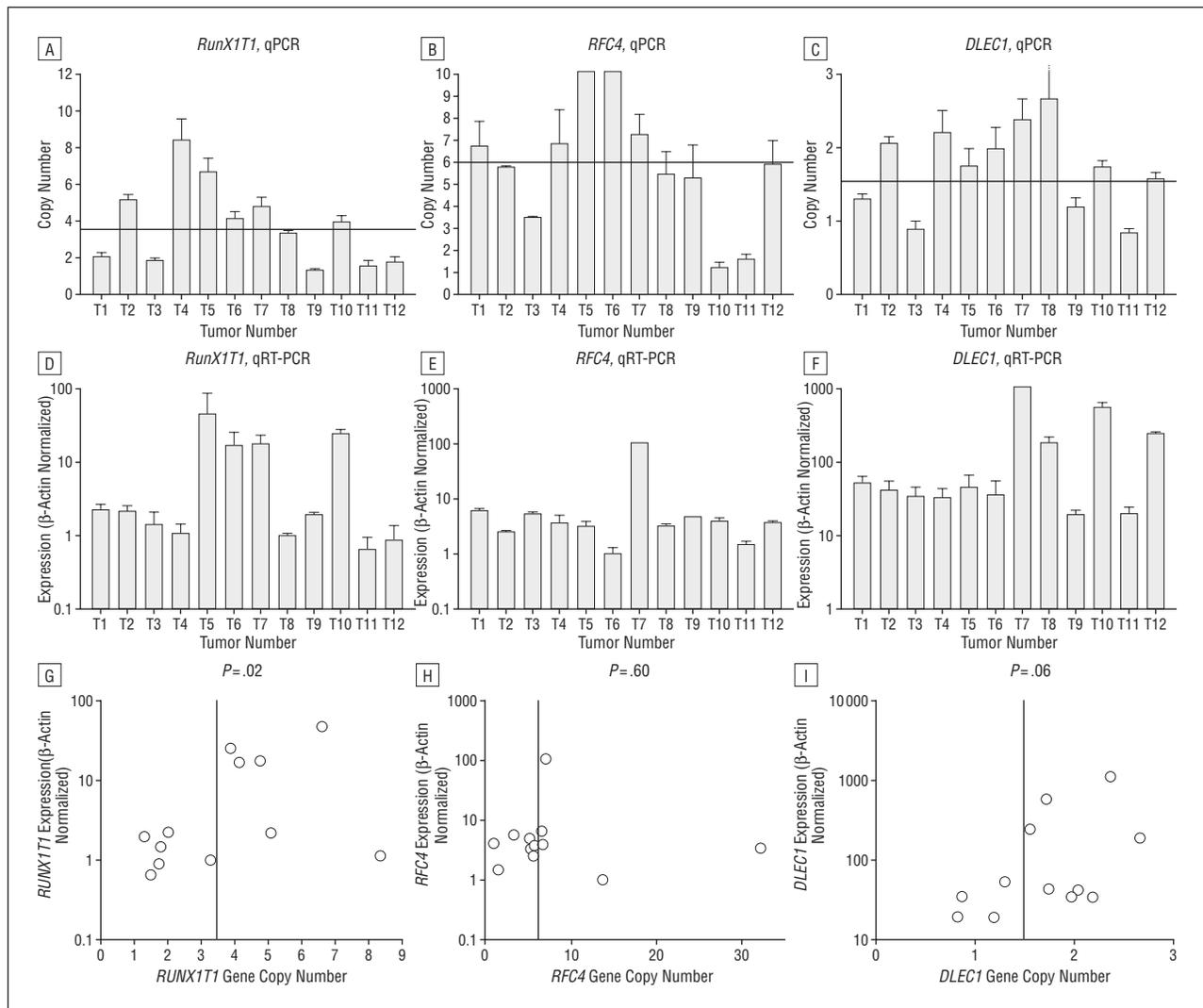


Figure 4. Relationship between amplification/deletion and messenger RNA (mRNA) expression. Twelve primary head and neck squamous cell carcinoma tumor tissues had DNA and total messenger RNA (mRNA) extracted for analysis. A-C, *RUNX1T1*, in which 6 of 12 tumors (50%) demonstrated amplification of the *RUNX1T1* locus of more than 3.5 copies, compared with 0 of 12 matched leukocyte DNA samples ($P=.01$, by χ^2 test). D-F, Four of 6 (67%) demonstrated upregulated transcription of this gene. G-I, In tissues with associated gene amplification, there was a statistically significant overexpression of the gene ($P=.02$) by Mann-Whitney U test. B, Amplification of the *RFC4* locus totaling more than 6 copies in 5 of 12 tumors (42%) compared with 0 of 12 matched leukocyte DNA samples ($P=.03$, by χ^2 test); E, 1 of 5 (20%) demonstrating upregulated mRNA transcription. H, Not a statistically significant finding ($P=.60$). C, For the *DLEC1* locus, 4 of 12 of the tumors (33%) had deletion (copy numbers <1.5), compared with 0 of 12 matched leukocyte DNA samples ($P=.05$, by χ^2 test). F, Three of 4 of the deleted samples (75%) demonstrated below-median mRNA expression. I, Samples with *DLEC1* deletions had reduced mRNA expression that was not statistically significant ($P=.06$). qPCR indicates quantitative polymerase chain reaction; qRT-PCR, quantitative real-time polymerase chain reaction.

human melanocytes. In our approach, we considered the integration of the consensus cancer coding sequence¹⁷ and gene deletion and amplification in conjunction with expression array differences. We were able to elucidate 20 genes from the putative consensus cancer coding sequence¹⁷ that were found in areas of chromosomal loss or gain. We chose to study 4 genes in detail: *TGFBR2*, *RUNX1T1*, *RFC4*, and *DLEC1*. *BRCA2* has been consistently found deleted in approximately 30% of HNSCCs, but without missense mutations.^{34,35} Three genes, *CHL1*, *RTP1*, and *EIF4A2*, showed changes in expression, but little is known regarding their function. Other genes that were not associated with expression differences and had no known previously reported oncogenic function were *CLSTN2*, *CMYA1*, *KL*, *MGC21688*, *PIK3R4*, *MCF2L2*, *FLJ10560*, *KCNB2*, and *EIF4G1*. These genes all have so-

matic mutations in human neoplasms, so they could be involved in the carcinogenesis of HNSCC in a manner that does not result in expression differences. More interestingly, several other genes with possible oncogenic function in other cancer systems but no expression differences in HNSCC were found: *DGKG*, *LIFR*, and *ITGA9*.

Of the genes we studied in detail, *RUNX1T1* is a zinc finger transcription factor protein and a demonstrated oncoprotein in acute myeloid leukemia (AML). It is the functional half of the *AML1-ETO* (*RUNX1T1*) fusion gene, under the control of the *AML1* promoter.²⁴ *RUNX1T1* is a putative zinc finger transcription factor and oncoprotein. It resides on an amplification in 8q21.3-8q22.2. *RUNX1T1* is one of the most common genetic abnormalities in AML, identified in 15% of all cases. Leukemias are often much less genetically diverse than solid

tumors and at times require only single hits for malignant cellular transformation. In vivo studies on variations of *RUNX1T1* showed rapid development of leukemia in a normal mouse subjected to retrovirus with this construct.³⁶ Initially, we showed overexpression (above that of normal controls) in 11 of 36 tumor tissues (31%). This was associated with a 50% incidence of amplification (copy number >3.5) and statistically significant overexpression in these samples ($P = .02$).

RFC4 forms a complex with *PCNA* and is involved in DNA replication, DNA repair, DNA modification, and chromatin modeling.²⁵ Recent evidence in cervical cancer demonstrates it is overexpressed.²⁶ Interestingly, *RFC4* overexpression has been found to be associated with human papillomavirus and HNSCC tumors.²⁷ We found *RFC4* upregulated in 9 of 36 tumors (25%), but only 1 of 5 tumors with significant gene amplification had marked overexpression (20%). This may reflect many confounding features in gene regulation including, but not limited to, interaction with transcription factors, promoter repression, epigenetic alterations, or transcript degradation.

DLEC1 is a putative tumor suppressor involved in carcinogenesis of the lung, esophagus, kidney, and nasopharyngeal cancers.^{28,29} Recently it was found to be the central target controlled by regional epigenetic regulation by an approach that searched for copy-number-independent gene dysregulation.³⁷ This finding was confirmed in ovarian cancer.³⁸ It is a 1755-amino acid polypeptide at 3p21.3. Functionally, transfection of this gene suppressed growth in 4 cancer cell lines by colony focus assay.³⁹ We found deletion in 4 of 12 primary HNSCC (33%), and these had concomitant reduction in expression that was not statistically significant ($P = .06$).

Although much work remains to elucidate possible contributors to that malignant phenotype of HNSCC, we did find 20 genes in the consensus cancer coding sequence¹⁷ that were associated with areas of chromosomal loss or gain, of which 5 had statistically significant differences in expression (see Table for P values). One pitfall of whole-genomic approaches is the problem of identifying the true signal from the substantial baseline noise generated by such large amounts of data. We used separate discovery and validation cohorts for expression microarray and qRT-PCR to bolster the importance of our findings. This approach represents the attempted discovery and validation of genes in HNSCC that are part of the consensus cancer coding sequence. Further work in the areas of functional experiments remains. Future work includes functional studies of *TGFBR2*, *RUNX1T1*, *RFC4*, and *DLEC1* in HNSCC systems. Small molecule inhibitors of these targets may prove to be useful for initial HNSCC tumor studies.

Submitted for Publication: May 1, 2007; final revision received March 3, 2008; accepted April 7, 2008.

Correspondence: Joseph A. Califano, MD, Department of Otolaryngology–Head and Neck Surgery, Johns Hopkins Medical Institutions, 601 N Caroline St, Sixth Floor, Baltimore, MD 21287-0910 (jcalifa@jhmi.edu).

Author Contributions: Drs Smith, Mithani, Liu, Chang, Begum, Westra, and Califano and Ms Dhara had full ac-

cess to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Smith, Sidransky, and Califano. *Acquisition of data:* Smith, Liu, Chang, Dhara, Westra, and Califano. *Analysis and interpretation of data:* Smith, Mithani, Chang, Begum, Westra, Sidransky, and Califano. *Drafting of the manuscript:* Smith, Chang, Dhara, and Califano. *Critical revision of the manuscript for important intellectual content:* Smith, Mithani, Liu, Begum, Westra, Sidransky, and Califano. *Statistical analysis:* Sidransky. *Obtained funding:* Westra and Califano. *Administrative, technical, and material support:* Smith, Liu, Begum, Dhara, Westra, and Califano. *Study supervision:* Westra and Califano.

Financial Disclosure: None reported.

Funding/Support: Dr Califano is supported by a Clinical Innovator Award from the Flight Attendant Medical Research Institute, the National Institute of Dental and Craniofacial Research (1R01DE015939-01), and the National Cancer Institute SPORE (5P50CA096784-05). Dr Smith is funded by an institutional NIH T32 grant.

Previous Presentation: This study was presented at the 2007 American Head and Neck Society Meeting at the Combined Otolaryngology Spring Meetings; April 27, 2007; San Diego, California.

REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. *CA Cancer J Clin*. 2006; 56(2):106-130.
2. Forastiere AA, Goepfert H, Maor M, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med*. 2003; 349(22):2091-2098.
3. Carvalho AL, Nishimoto IN, Califano JA, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. *Int J Cancer*. 2005;114(5):806-816.
4. Kim MM, Califano JA. Molecular pathology of head-and-neck cancer. *Int J Cancer*. 2004;112(4):545-553.
5. Knudson AG. Antioncogenes and human cancer. *Proc Natl Acad Sci U S A*. 1993; 90(23):10914-10921.
6. Stark GR, Wahl GM. Gene amplification. *Annu Rev Biochem*. 1984;53:447-491.
7. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. *Nat Genet*. 2003;34(4):369-376.
8. Nogueira CP, Dolan RW, Gooley J, et al. Inactivation of p53 and amplification of cyclin D1 correlate with clinical outcome in head and neck cancer. *Laryngoscope*. 1998;108(3):345-350.
9. Singh B, Stoffel A, Gogineni S, et al. Amplification of the 3q26.3 locus is associated with progression to invasive cancer and is a negative prognostic factor in head and neck squamous cell carcinomas. *Am J Pathol*. 2002;161(2):365-371.
10. Sorrells DL, Ghali GE, Meschonat C, et al. Competitive PCR to detect eIF4E gene amplification in head and neck cancer. *Head Neck*. 1999;21(1):60-65.
11. Wang MB, Alavi S, Engstrom M, et al. Detection of chromosome 11q13 amplification in head and neck cancer using fluorescence in situ hybridization. *Anti-cancer Res*. 1999;19(2A):925-931.
12. Weichselbaum RR, Dunphy EJ, Beckett MA, et al. Epidermal growth factor receptor gene amplification and expression in head and neck cancer cell lines. *Head Neck*. 1989;11(5):437-442.
13. Sabbir MG, Roy A, Mandal S, Dam A, Roychoudhury S, Panda CK. Deletion mapping of chromosome 13q in head and neck squamous cell carcinoma in Indian patients: correlation with prognosis of the tumour. *Int J Exp Pathol*. 2006;87(2):151-161.
14. Maestro R, Dolcetti R. Oncogenes and tumor suppressor genes in head and neck squamous cell carcinomas [in Italian]. *Acta Otorhinolaryngol Ital*. 1996;16(3) (suppl 54):21-26.
15. Collins FS, Barker AD. Mapping the cancer genome: pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am*. 2007;296(3):50-57.
16. Strauss BS. Limits to the Human Cancer Genome Project? *Science*. 2007;315(5813): 762-764.

17. Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006;314(5797):268-274.
18. Garraway LA, Weir BA, Zhao X, et al. "Lineage addiction" in human cancer: lessons from integrated genomics. *Cold Spring Harb Symp Quant Biol*. 2005;70:25-34.
19. Garraway LA, Widlund HR, Rubin MA, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature*. 2005;436(7047):117-122.
20. Chen W, Salto-Tellez M, Palanisamy N, et al. Targets of genome copy number reduction in primary breast cancers identified by integrative genomics. *Genes Chromosomes Cancer*. 2007;46(3):288-301.
21. Pujana MA, Ruiz A, Badenas C, et al. Molecular characterization of a t(9;12)(p21;q13) balanced chromosome translocation in combination with integrative genomics analysis identifies C9orf14 as a candidate tumor-suppressor. *Genes Chromosomes Cancer*. 2007;46(2):155-162.
22. Wang Q, Diskin S, Rappaport E, et al. Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. *Cancer Res*. 2006;66(12):6050-6062.
23. Dahia PL, Hao K, Rogus J, et al; Familial Pheochromocytoma Consortium. Novel pheochromocytoma susceptibility loci identified by integrative genomics. *Cancer Res*. 2005;65(21):9651-9658.
24. Peterson LF, Boyapati A, Ahn EY, et al. Acute myeloid leukemia with the 8q22;21q22 translocation: secondary mutational events and alternative t(8;21) transcripts. *Blood*. 2007;110(3):799-805.
25. Johnson A, Yao NY, Bowman GD, Kuriyan J, O'Donnell M. The replication factor C clamp loader requires arginine finger sensors to drive DNA binding and proliferating cell nuclear antigen loading. *J Biol Chem*. 2006;281(46):35531-35543.
26. Narayan G, Bourdon V, Chaganti S, et al. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. *Genes Chromosomes Cancer*. 2007;46(4):373-384.
27. Slebos RJ, Yi Y, Ely K, et al. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clin Cancer Res*. 2006;12(3, pt 1):701-709.
28. Nishiwaki T, Daigo Y, Kawasoe T, Nakamura Y. Isolation and mutational analysis of a novel human cDNA, DEC1 (Deleted in Esophageal Cancer 1), derived from the tumor suppressor locus in 9q32. *Genes Chromosomes Cancer*. 2000;27(2):169-176.
29. Kwong J, Chow LS, Wong AY, et al. Epigenetic inactivation of the Deleted in Lung and Esophageal Cancer 1 gene in nasopharyngeal carcinoma. *Genes Chromosomes Cancer*. 2007;46(2):171-180.
30. Sparano A, Quesnelle KM, Kumar MS, et al. Genome-wide profiling of oral squamous cell carcinoma by array-based comparative genomic hybridization. *Laryngoscope*. 2006;116(5):735-741.
31. Ogino S, Kawasaki T, Ogawa A, Kirkner GJ, Loda M, Fuchs CS. TGFBR2 mutation is correlated with CpG island methylator phenotype in microsatellite instability-high colorectal cancer. *Hum Pathol*. 2007;38(4):614-620.
32. Muñoz NM, Upton M, Rojas A, et al. Transforming growth factor beta receptor type II inactivation induces the malignant transformation of intestinal neoplasms initiated by Apc mutation. *Cancer Res*. 2006;66(20):9837-9844.
33. Biswas S, Chytil A, Washington K, et al. Transforming growth factor beta receptor type II inactivation promotes the establishment and progression of colon cancer. *Cancer Res*. 2004;64(14):4687-4692.
34. Kirkpatrick H, Waber P, Hoa-Thai T, et al. Infrequency of BRCA2 alterations in head and neck squamous cell carcinoma. *Oncogene*. 1997;14(18):2189-2193.
35. Nawroz-Danish HM, Koch WM, Westra WH, Yoo G, Sidransky D. Lack of BRCA2 alterations in primary head and neck squamous cell carcinoma. *Otolaryngol Head Neck Surg*. 1998;119(1):21-25.
36. Yan M, Kanbe E, Peterson LF, et al. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med*. 2006;12(8):945-949.
37. Stransky N, Vallot C, Reyat F, et al. Regional copy number-independent deregulation of transcription in cancer. *Nat Genet*. 2006;38(12):1386-1396.
38. Kwong J, Lee JY, Wong KK, et al. Candidate tumor-suppressor gene DLEC1 is frequently downregulated by promoter hypermethylation and histone hypoacetylation in human epithelial ovarian cancer. *Neoplasia*. 2006;8(4):268-278.
39. Daigo Y, Nishiwaki T, Kawasoe T, Tamari M, Tsuchiya E, Nakamura Y. Molecular cloning of a candidate tumor suppressor gene, DLC1, from chromosome 3p21.3. *Cancer Res*. 1999;59(8):1966-1972.