

A Transcriptional Progression Model for Head and Neck Cancer¹

Patrick K. Ha, Nicole E. Benoit, Robert Yochem, James Sciubba, Marianna Zahurak, David Sidransky, Jonathan Pevsner, William H. Westra, and Joseph Califano²

Departments of Otolaryngology-Head and Neck Surgery, Head and Neck Cancer Research Division [P. K. H., N. E. B., R. Y., J. S., D. S., and J. C.], Oncology Biostatistics [M. Z.], and Pathology [W. H. W.], Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, and Department of Neurology, Kennedy Krieger Institute, Baltimore, Maryland [J. P.]

ABSTRACT

Purpose: A genetic progression model for head and neck squamous cell carcinoma (HNSC) has been established and implies the presence of transcriptional dysregulation as a consequence of accumulation of genetic alterations. Although expression array data have been provided for HNSC, the timing of transcriptional dysregulation in the progression from normal mucosa to dysplastic epithelium to invasive HNSC has not been described. Here, we describe a transcriptional progression model of HNSC.

Experimental Design: Expression arrays representing >12,000 genes and expressed sequence tags were used to examine malignant lesions (M), premalignant lesions (PM), distant, histopathologically normal mucosa from patients with premalignant or malignant lesions (MN), and normal mucosa from the upper aerodigestive tract of patients with noncancer diagnoses (N). Significance analysis of microarrays, hierarchical clustering, and principal components analysis was used to identify genes with differential expression patterns.

Results: Using a false discovery rate of <5% for significance analysis of microarray, the M group revealed 965 up-regulated and 1106 down-regulated genes relative to the N group. The PM group demonstrated 108 up-regulated and 226 down-regulated genes relative to the N group, whereas the M group demonstrated only 5 up-regulated and 13 down-regulated genes relative to the PM group. Both hierarchical cluster analysis and principal components analysis revealed a consistent separation between the N, PM, and M

groups, with a closer association between the PM and M groups. To provide independent validation of the microarray data, quantitative reverse transcription-PCR was performed for a significantly up-regulated gene, integrin α 6, correlating well with microarray data (linear regression analysis, $P < 0.0001$).

Conclusions: Similarly to the genetic progression model of HNSC, this transcriptional model shows that the majority of alterations occurs before the development of malignancy and identifies key targets of transcriptional dysregulation during progression from a normal to a premalignant state and from a premalignant to a malignant state.

INTRODUCTION

The progression of HNSC³ has been described from a genetic perspective, with a distinct pattern and timing of genetic alterations along a continuum of malignant transformation. These data suggest that premalignant head and neck lesions possess many of the alterations found in cancer before the development of a malignant phenotype and that most genetic alterations occur before the phenotypic expression of malignancy (1). Alterations in gene expression during tumor progression are also of significant interest, but the broad picture of altered gene expression and timing of these alterations during head and neck tumorigenesis has not yet been characterized.

The advent of gene expression profiling allows for rapid, genome-wide analysis in a standardized fashion. Many methods of normalization of gene expression, analysis of differences in gene expression among diverse groups that represent different biological states, and attempts to group different phenotypic entities based on similarities have been used to interpret the large amounts of data yielded from expression array data. In this study, we have used the SAM method as one of the effective ways of identifying particular gene products involved in differentiating transcriptional profiles between two groups (2). To examine the similarities in expression profiles between groups defined by a normal, premalignant, and malignant phenotype, we have used hierarchical clustering analysis, as well as PCA, two commonly used techniques (3).

The purpose of this study is to describe a transcriptional progression model for premalignant and M of the head and neck using array-based gene expression profiling. The characterization of the timing and nature of these events in the early and late stages of HNSC development yields insights into head and neck tumorigenesis, identifies novel gene expression and regulatory

Received 11/21/02; revised 3/10/03; accepted 3/21/03.

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.

¹ Supported in part by the Traylor Research Fund and Clinical Innovator Award from the Flight Attendant Medical Research Institute (to J. C.). Dr. Califano is a Damon Runyon-Lilly Clinical Investigator supported by the Damon Runyon Cancer Research Foundation.

² To whom requests for reprints should be addressed, at Department of Otolaryngology-Head and Neck Surgery, 601 North Caroline Street, 6th Floor, Baltimore, MD 21287-0910. Phone: (410) 955-6420; Fax: (410) 614-1411; E-mail: jcalifa@jhmi.edu.

³ The abbreviations used are: HNSC, head and neck squamous cell carcinoma; M, malignant lesion; PM, premalignant lesion; SAM, significance analysis of microarray; FDR, false discovery rate; MN, premalignant or malignant lesion; N, noncancer diagnosis; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; MAP, mitogen-activated protein; VEGF, vascular endothelial growth factor; PCA, principal components analysis; RT-PCR, reverse transcription-PCR.

Table 1 Summary of the site and pathologic diagnosis of the lesions sampled for this study

ID	Site	Diagnosis
1	p.o. cavity	Normal control mucosa
2	p.o. cavity	Normal control mucosa
3	Oropharynx	Normal control mucosa
4	Hypopharynx	Normal control mucosa
5	Buccal mucosa	Normal control mucosa
6	Supraglottic larynx	Normal control mucosa
7	Buccal mucosa	Matched normal mucosa
8	Buccal mucosa	Matched normal mucosa
9	Buccal mucosa	Matched normal mucosa
10	Floor of mouth	Matched normal mucosa
11	Buccal mucosa	Matched normal mucosa
12	Buccal mucosa	Epithelial hyperplasia
13	p.o. tongue	Mild focal dysplasia
14	p.o. tongue	Moderate epithelial dysplasia
15	p.o. tongue	Moderate epithelial dysplasia
16	p.o. tongue	Moderate epithelial dysplasia
17	Pyramidal sinus	Moderate epithelial dysplasia
18	Buccal mucosa	Severe epithelial dysplasia/CIS
19	Buccal mucosa	Severe epithelial dysplasia/CIS
20	True vocal cord	Invasive carcinoma
21	p.o. tongue	Invasive carcinoma
22	p.o. tongue	Invasive carcinoma
23	Retromolar trigone	Invasive carcinoma
24	Floor of mouth	Invasive carcinoma
25	Base of tongue	Invasive carcinoma
26	Buccal mucosa	Invasive carcinoma

pathway alterations, and characterizes global transcriptional progression patterns for premalignant and malignant phenotypes. In turn, these insights may identify novel therapeutic targets that may lead to useful insights into diagnosis, surveillance, and treatment of HNSC.

MATERIALS AND METHODS

Appropriate informed consent was obtained from patients with suspected head and neck neoplasms and/or premalignant disease after approval by the Joint Committee on Clinical Investigation of the Johns Hopkins Medical Institutions. Snap-frozen mucosal tissue was obtained from 6 normal, nonsmoking, control patients without cancer diagnoses, 8 patients without a diagnosis of invasive cancer with clinically determined, leukoplakic/PM (1 squamous hyperplasia without atypia in the setting of tobacco exposure, 1 mild dysplasia, 4 moderate dysplasia, and 2 severe dysplasia/carcinoma *in situ*), and 7 patients with invasive carcinoma (Table 1). In addition, five samples were obtained from anatomically distinct, contralateral, unaffected mucosa from patients with lesions (3 with premalignant disease and 2 with invasive cancer). An additional pleomorphic adenoma of parotid gland origin was also used as a control.

The tissue was snap frozen, and a 6- μ m H&E section was taken to confirm the pathologic diagnosis and presence of >85% lesional tissue verified by a head and neck pathologist (W. H. W.). The degree of inflammation was characterized (none, moderate, and severe), and no correlation was found with results from any of the analyses (data not shown). Total RNA was extracted from tissue samples using the Invitrogen (Carlsbad, CA) TRIzol reagent. The tissue was kept frozen in liquid nitrogen, whereas TRIzol was added. The samples were care-

fully and mechanically homogenized, and the mixture was transferred into a clean 1.5-ml tube using a sterile cell scraper. Chloroform was then added, and the samples were centrifuged. The top, aqueous layer was transferred to a new tube for RNA precipitation with isopropanol. The isolated RNA was washed once with 80% ethanol and resuspended in water. The RNA samples were then further purified using an RNeasy kit (Qiagen, Valencia, CA) as per protocol.

Samples were processed as described in the Affymetrix GeneChip (Santa Clara, CA) Expression Analysis Technical Manual. Starting with 5 μ g of total RNA, cDNA was synthesized using a T7-(dT)₂₄ primer [Genset (Paris, France), high-performance liquid chromatography purified] and the Invitrogen Superscript Choice kit. The cDNA was extracted with a phenol:chloroform extraction, followed by ethanol precipitation, and then resuspended in 10–12 μ l. Using the whole cDNA sample, biotinylated cRNA was synthesized using an Enzo (New York, NY) BioArray HighYield RNA Transcript Labeling Kit. The biotinylated cRNA was extracted with the Qiagen RNeasy columns and then precipitated overnight with ethanol and ammonium acetate. The cRNA was resuspended in a small volume and quantitated. cRNA (15 μ g) was fragmented by metal-induced hydrolysis with a solution containing magnesium and potassium. An aliquot was then run on an agarose gel to ensure that the distribution of fragmented RNA was between 35 and 200 bp. The 300- μ l hybridization cocktail was made, using all of the fragmented cRNA, along with Affymetrix controls, buffer, and acetylated BSA. The hybridization cocktail (200 μ l) was loaded onto each Affymetrix Hu95A.v2 GeneChip, and the chips were incubated in a rotating oven for 16 h at 45°C.

After 16 h, the hybridization solution was removed from the chip, and the chips were washed, using an Affymetrix GeneChip Fluidics station and solutions. The biotinylated cRNA now hybridized to the chips was visualized with a GeneArray scanner, using an antibody-enhanced staining with streptavidin conjugated to phycoerythrin. The chips were initially analyzed with the Affymetrix MicroArray Suite software, version 4.0. For quality control purposes, we repeated three samples with different chips with excellent reproducibility (data not shown).

Scaled values were then calculated, and negative average difference values were adjusted to a value of 1 (the mean for all values for all analyzed arrays after this normalization step was 259.7). A logarithmic transformation was then performed using SNOMAD software (4). Briefly, SNOMAD includes two nonlinear transformations that correct for bias and variance, which are nonuniformly distributed across the range of microarray element signal intensities: (a) local mean normalization and (b) local variance correction (Z-score generation using a locally calculated SD). This process attempts to eliminate the inherent nonuniform variability that may exist across the microarray by comparing values within smaller ranges of microarray element signal intensities. This normalization will reduce the overestimation of fold changes for genes with low expression values. This normalization was performed using results from an independently performed array hybridization for a normal mucosal sample as a control.

Final values for each sample, expressed as a Z-score, were analyzed for significance using SAM as described. Briefly, this

method assigns a score to each gene on the basis of gene expression relative to the SD of repeated measurements, in this case, measurements for each of four groups (N, MN, PM, and M). For genes with scores greater than an adjustable threshold, permutations of the repeated measurements are generated to estimate the percentage of genes identified by chance, the FDR. We set a cutoff for the FDR to be <5% in an attempt to restrict the possibility of false positives to <5% of the genes described. The four groups (N, MN, PM, and M) were compared with each other, and the number of significant genes, as well as their loci, was noted. PCA using Partek Pro (Partek, St. Charles, MO) and unsupervised hierarchical cluster analysis using J Express (Molmine, Bergen, Norway) were performed using complete linkage and a Pearson correlation for a distance measure in an effort to determine global similarities in gene expression for individual samples. These analyses serve to display the similarity of aggregate global expression patterns in a spatial pattern; those samples located closer in space are considered to possess a greater degree of similarity in gene expression patterns.

To perform an independent validation of the expression array data, we selected one gene noted to be up-regulated by SAM analysis (integrin $\alpha 6$) and performed quantitative RT-PCR on the same quantity of RNA purified from each sample. The forward primer was designed to span the gap between exons 25 and 26, and the reverse primer and fluorescent probe resided within exon 26. The amplicon size was 151 bp. One round of reverse transcription was performed to create the cDNA, followed by 50 cycles of amplification (95°C for 15 s and 60°C for 1 min) using the Applied Biosystems 7900 thermal cycler and software for analysis. Primers and probes to a housekeeping gene (GAPDH) were also used to standardize the input RNA and provide a reference for our gene of interest. Primer and probe sequences are as follows:

Integrin $\alpha 6$: forward: TCTGTAATTGTGTGGATTCTT-TAAACG,

reverse: GTTCCACTTTGTGATCCACTGTTTT,

probe: Fam-TAGGTACGATGACAGTGTCCCCGAT-ACCA-Tamra

GAPDH: forward: GTCCCCACCACACTGAATC

reverse: GCCCTCCCCTCTTCAAG

probe: Fam-CCCCTCCTCACAGTTGCCATGTAG-ACC-Tamra

The cycle difference between the gene of interest and GAPDH was determined and converted to a fold difference in expression level (integrin $\alpha 6$ versus GAPDH). This ratio was also calculated using the raw data obtained from the Affymetrix Gene Chip analysis. The relationship between these two measures was quantified using bivariate Pearson correlation and simple linear regression. For the linear regression, the dependent variable was integrin $\alpha 6$ RT-PCR. Regression coefficients were estimated using least squares, and individual data points were studied for their influence on parameter estimates using standard influence diagnostics (5). The *P* reported is two sided.

RESULTS

Twenty-six samples including 6 normal control patients, 8 patients with clinically determined, leukoplakic/PM, and 7 patients with invasive carcinoma as well as 5 samples from ana-

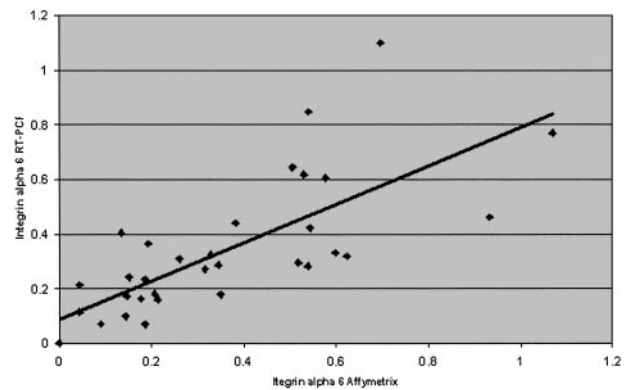


Fig. 1 A scatterplot of the ratio of fold difference between the expression levels of integrin $\alpha 6$ and GAPDH as determined by quantitative RT-PCR and the fold difference between these two genes as determined by the Affymetrix Gene Chip. The correlation between these two measures was statistically significant ($P < 0.0001$).

tomically distinct, contralateral, unaffected mucosa from patients with MN were analyzed by hybridization of cRNA to gene expression arrays containing >12,000 genes. Initial normalization using SNOMAD was performed as described above, followed by analysis for genes with significantly different expression patterns between the four groups of normal, matched normal, premalignant, and malignant samples using SAM with an FDR < 5%. In addition, hierarchical cluster and PCA were performed as described above, to delineate grouping patterns based on expression data.

To provide a means of verifying our array data and normalization, we selected a gene found to be statistically up-regulated in the progression from normal to PM, integrin $\alpha 6$, as well as a housekeeping gene, GAPDH. We independently performed quantitative RT-PCR for both genes on total RNA extracted from each sample and compared the ratio of integrin $\alpha 6$ /GAPDH as measured by quantitative RT-PCR to the ratio of the normalized values for these genes derived from array analysis. Using logistic regression, the Pearson coefficient (0.7) was found to have a statistically significant correlation ($P < 0.0001$) between these two different methods of detection, thereby suggesting that our normalization and array data were reliable (Fig. 1).

Using the SAM method, the four different groups were compared: (a) normal; (b) matched normal; (c) premalignant; and (d) invasive cancer. There were no significant differences in gene expression between normal and matched normal. When the normal and invasive cancer groups were compared with an FDR of <5%, 965 up-regulated and 1106 down-regulated genes were detected in invasive cancer when compared with normal. Premalignant groups yielded 108 up-regulated and 226 down-regulated genes in comparison with normal. Finally, only 5 up-regulated and 13 down-regulated genes were found to be significant when the invasive cancer groups were compared with premalignant (Fig. 2). These data demonstrate that a greater proportion of transcriptional alterations occur during the transition from normal to premalignant mucosa than in the transition from premalignant to malignant tumor. Specific genes

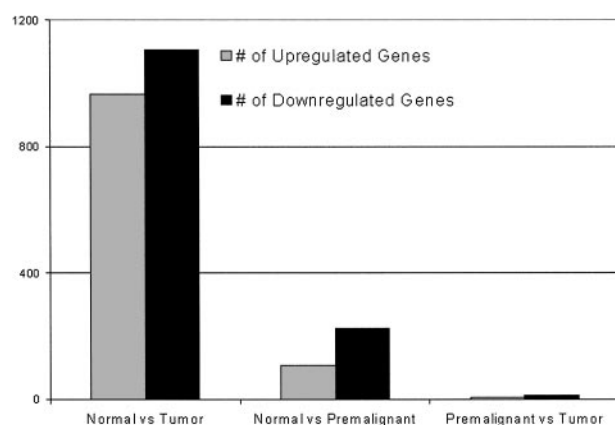


Fig. 2 Graphical representation of the number of genes and expressed sequence tags demonstrating statistically significant alterations in expression level as determined by the SAM package. It is evident that there are many more alterations between the normal to premalignant group than there are between the premalignant and malignant group.

identified and SAM output files can be found listed on the Internet.⁴ These include genes and gene pathways often implicated in carcinogenesis, including those genes involved in processes, such as cell adhesion, invasion, cell cycle, DNA repair, transcription, growth factor signaling, immune recognition, as well as other processes. A few select genes known to be altered in head and neck cancer that were up-regulated in M group when compared with N are listed in Table 2, and genes that were down-regulated in the M group when compared with N are listed in Table 3. Of note, many of these genes have been shown to have similar expression profiles in previous microarray studies of HNSC reported in the literature (6–9). In particular, similar alteration patterns are noted in cytokeratin expression, cell adhesion alterations, *wnt* and *notch* pathway alterations, angiogenic and other cytokine pathways, as well as other cellular systems like the MAP kinase pathway.

Results from the hierarchical clustering analysis are shown in Fig. 3a. The N and MN samples tended to cluster together, and PM and M samples formed related but distinct clusters. However, one premalignant sample 12 did cluster with these normal ones. This sample was a biopsy specimen from a patient with spit tobacco-related leukoplakia but with a histological appearance that lacked atypia. The premalignant samples and invasive cancer samples also tended to segregate with the histological diagnoses; however, one of the premalignant samples 19 clustered with the malignant group. The sample had the histological appearance of carcinoma *in situ* and was the most advanced of the PM analyzed in this group. Of note, when included in this analysis, a parotid gland pleomorphic adenoma sample clustered independently of all of the premalignant, malignant, and normal head and neck samples (data not shown). N and MN samples clustered together, which is of interest because the phenomenon of field cancerization often results in wide-

Table 2 Selected up-regulated gene products in malignant HNSC vs. normal mucosa^a

GenBank accession no.	Gene name	Function
AF016098	Vascular endothelial cell growth factor 165 receptor 2	Angiogenesis
AF024710	VEGF	Angiogenesis
U09278	Fibroblast activation protein	Angiogenesis
M63193	Platelet-derived endothelial cell growth factor	Angiogenesis
AF024710	VEGF	Angiogenesis
AF024710	VEGF	Angiogenesis
M63193	Platelet-derived endothelial cell growth factor	Angiogenesis
L22475	Bax γ	Apoptosis
AF016266	TRAIL receptor 2	Apoptosis
U50062	RIP protein kinase	Apoptosis
X07820	Metalloproteinase stromelysin-2	Cell adhesion
X63629	p cadherin	Cell adhesion
D21255	OB-cadherin-2	Cell adhesion
D21254	OB-cadherin-1	Cell adhesion
L23808	Human metalloproteinase	Cell adhesion
X53586	Integrin α 6	Cell adhesion
U37022	Cyclin-dependent kinase 4	Cell cycle
X51688	Cyclin A	Cell cycle
X00588	Epidermal growth factor receptor	Cytokine
X06374	Platelet-derived growth factor	Cytokine
M38449	Transforming growth factor- β	Cytokine
M34539	FK506-binding protein	Cytokine modulation
X63547	tre oncogene	Degradation
U09759	JNK2	MAP kinase pathway
U12779	MAP kinase-activated protein kinase 2	MAP kinase pathway
U25265	MEK5	MAP kinase pathway
X53416	Actin-binding protein (filamin)	Structural
M13509	Human skin collagenase	Structural
D78577	14-3-3 protein	Transcription
M29039	jun-B	Transcriptional regulation
V00568	<i>c-myc</i>	Transcriptional regulation
L37882	Frizzled	wnt pathway
AF003837	Jagged1	wnt pathway
U95299	Human Notch4	wnt pathway
U81787	Wnt10B	wnt pathway
U07563	ABL	

^a TRAIL, tumor necrosis factor-related apoptosis-inducing ligand; JNK, c-Jun NH₂-terminal kinase; MEK, MAP/extracellular signal-regulated kinase kinase; ABL, Abelson; RIP, receptor interacting protein.

spread genetic alteration in head and neck mucosa from which a malignancy has developed. The MN samples consisted of normal mucosa harvested from the contralateral side of the upper aerodigestive tract, usually the buccal mucosa, to ensure geographic separation from clonally expanded, genetically altered cell populations surrounding a primary lesion.

PCA revealed a similar finding to the cluster analysis. A graphical schematic shown in Fig. 3b demonstrates a two-dimensional representation of PCA between the four groups. The proximity of the colored boxes in this space represents the degree of relatedness between the samples. Both N and MN groups clustered together, in a similar manner noted to that found with cluster analysis, whereas PM and M groups appear

⁴ Internet address: <http://www.hopkinsmedicine.org/headneckcancer/manuscript.html>.

Table 3 Selected down-regulated genes in malignant HNSC vs. normal

GenBank accession no.	Gene name	Function
AF047826	Cadherin-7	Adhesion
D31784	Cadherin-6	Adhesion
M32304	Metalloproteinase inhibitor	Adhesion
L34059	Cadherin-4	Adhesion
AJ011497	Claudin-7	Adhesion
U95032	Growth arrest-specific protein 2	Apoptosis
X89101	Fas	Apoptosis
U29656	DR-nm23	Apoptosis
X07695	Cytokeratin 4	Cell structure
X07696	Cytokeratin 15	Cell structure
U77604	Microsomal glutathione S-transferase 2	Detoxification
J05459	Glutathione transferase M3	Detoxification
AF043105	Glutathione S-transferase mu 3	Detoxification

to be closely related in terms of expression pattern, with tight clustering within PM and M groups. Again, similar outliers were noted in PCA when compared with cluster analysis, in that hyperplasia sample 12 segregated with normal mucosal samples, and premalignant sample 19 segregated with malignant samples.

DISCUSSION

This study is the first report to broadly define the initial transcriptional events in HNSC carcinogenesis that result in development of a premalignant phenotype and compare them with those that occur later in carcinogenesis. Gene expression microarray technology has allowed for the rapid screening of several thousand gene products and the prospect of identifying those genes responsible for tumor initiation and progression. However, such a large quantity of data has proven to be difficult to manipulate and analyze. The SAM method has been shown to be an effective means of selecting genes significant in categorizing groups of tissue samples (2). We therefore selected this method to identify the number of significant genes responsible for the progression of head and neck lesions to invasive cancer and to identify early and late transcriptional alterations in the progression of HNSC.

Convincing evidence suggests that the majority of genetic alterations that accumulate during the progression of head and neck cancer occur early; so-called PM possess many of the mutations and loss of genetic material that exist in invasive cancers. It follows that the same pattern appears to be true on a transcriptional level. A group of 334 genes was significantly altered (either up- or down-regulated) when comparing the normal control tissue with the premalignant set of tissues, whereas only 23 genes were altered when comparing the premalignant with malignant group. When comparing the normal *versus* tumor group, many more genes were identified (965 up-regulated and 1106 down-regulated) in comparison with those identified in a stepwise fashion. This is not surprising, because the comparison between the groups with greatest genetic and phenotypic difference across a continuum is more likely to produce significant differences, in contrast to a comparison between groups that are more phenotypically and genetically similar. A larger number of transcriptional alterations was identified than in other studies of HNSC (10) but may

reflect a difference in sample size, cancer stage/site, analysis methods, and a number of other factors. However, remarkable consistency was observed in that similar alteration patterns are noted in pathways found to be altered in these previous studies, in particular, alterations in cytokeratin expression, cell adhesion molecules, the MAP kinase pathway, as well as cytokine alterations, including VEGF overexpression (8). This is offered as further validation of the methods used to determine transcriptional alterations in this study, in that consistent alterations are seen using diverse techniques and different modes of analysis. The actual gene product alterations include some overlap with previous published studies, such as with the down-regulation of Claudin-7 (8) and the up-regulation of genes involved in the *notch* and *wnt* pathways (9). Many other gene products were identified, but additional studies will need to be performed to validate these findings and confirm the role of these genes in head and neck cancer progression.

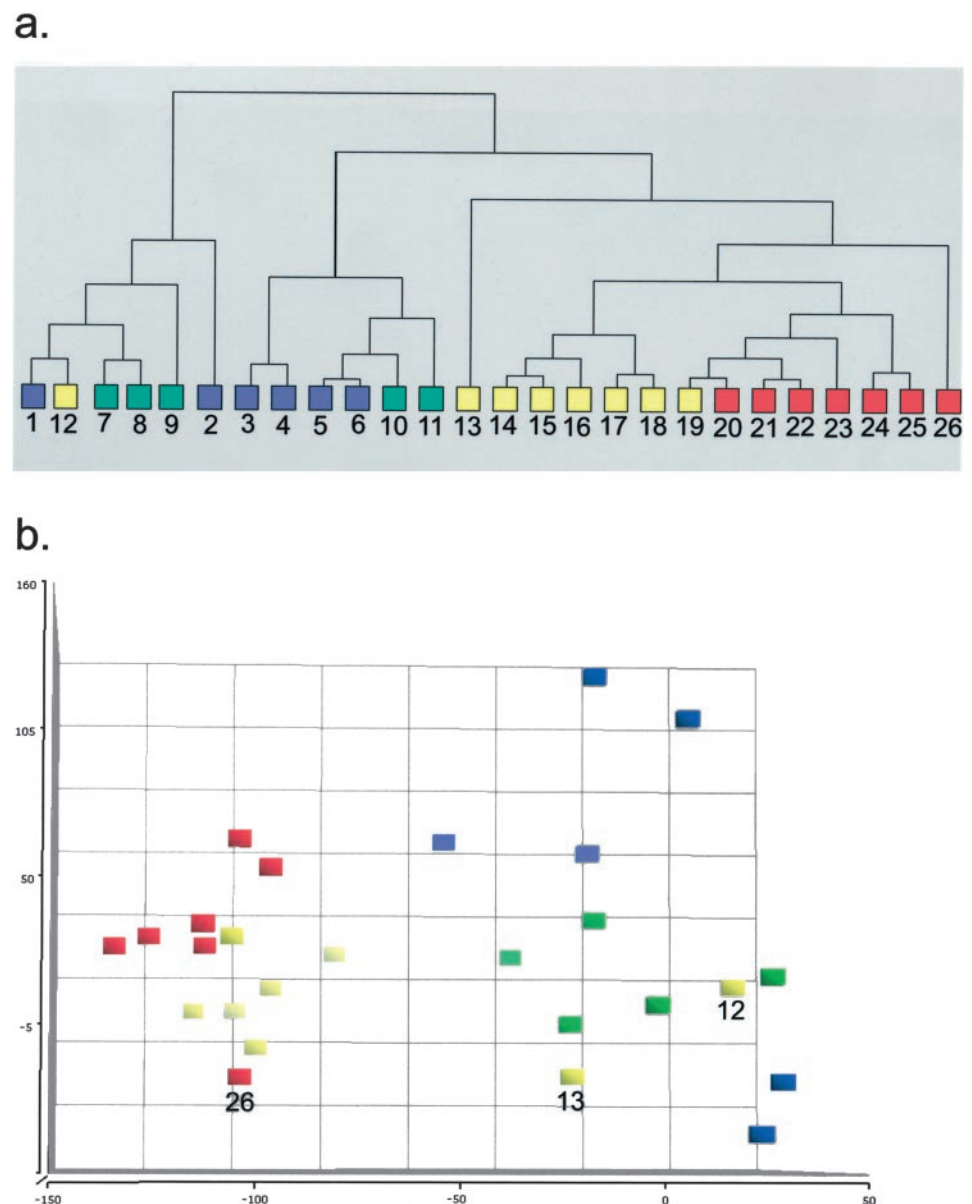
This pattern of early transcriptional alteration in HNSC carcinogenesis suggests that there is relatively little difference between premalignant and malignant state, as compared with the difference between a normal and premalignant state. This would be consistent with previous genetic progression models derived from DNA-based alterations. However, it is possible that there are a number of genes and pathways that have a significant effect on the cellular function but fall below the level of detection of our methods. Alternatively, modest increases or decreases in a series of transcripts from the same pathway could still lead to a profound effect on the translational level. Furthermore, the transformation of initial translational sequence to the ultimate protein product may not necessarily mirror the level of transcription because of post-translational modification and protein interactions within the cell.

In our study, we also wanted to examine the overall similarity between these histological classification groups. Using both PCA and cluster analysis, there was a distinct separation between the normal, premalignant, and invasive cancer groups when analyzing across all 12,500 gene products. This finding supports a consistent process of differential alteration in transcription specific to each phenotype that seems to occur during head and neck cancer progression. The ability to group such PM has the potential use for diagnosis and possibly prognosis; some preliminary studies have been already attempted to categorize subgroups within squamous cell head and neck neoplasms (6). Other studies have been performed using microarray technology to predict the radioresponsiveness of tumors with some success (7).

Our data support the segregation of lesions according to histopathologic diagnosis rather than site of origin. Although a significant proportion of the samples tested was from the oral cavity, samples from normal, premalignant, and malignant categories were obtained from other head and neck sites, with no pattern of segregation related to site of origin. The difference in genetic and expression alterations in HNSC related to site of origin has been debated; although limited in size, this study suggests that the transcriptional profile of premalignant and M is predominantly related to histopathologic grade.

Additionally of interest is the lack of significant difference detected between normal and matched normal groups.

Fig. 3 Analysis of histological groups and expression. Individual samples are labeled according to Table 1. *a*, hierarchical clustering analysis graphic demonstrating the separation of the four groups based solely on gene expression data. The segregation of the groups demonstrates that there are global expression differences that occur as head and neck lesions develop a malignant phenotype. *Blue*, normal; *green*, matched normal; *yellow*, premalignant; *red*, malignant. *b*, PCA graphic depicting a two-dimensional representation of the four histological groups. The distance in space between the colored boxes represents the degree of relatedness between the samples. Close proximity of malignant and premalignant groups is noted, with separation from normal and matched normal samples. *Blue*, normal; *green*, matched normal; *yellow*, premalignant; *red*, malignant.



We and others have found a pattern of clonally related spread both adjacent and distant to areas of premalignant mucosa and malignancy in the upper aerodigestive tract (1). The selection of contralateral, distant mucosa for the MN group was intentional, in that we attempted to obtain mucosa unaffected by this phenomenon of lateral clonal expansion. This finding supports the interpretation of the phenomenon of local field cancerization as a manifestation of lateral clonal spread, because distant, normal mucosa from patients with premalignant disease without genetic alterations does not demonstrate any significant expression differences from normal mucosa from patients without premalignant or malignant disease.

The use of gene expression microarrays to examine lesions from different stages of head and neck cancer progression provides a powerful tool for the identification of

gene products that are potentially significant in bringing about this transformation. Analyses consistently identify a larger number of transcriptional alterations in the transition from the normal to premalignant state, than in comparison from the premalignant to malignant state. This information can be used to provide further insight into HNSC tumorigenesis and a plethora of targets for diagnosis and therapy. In particular, these observations may be applied to early detection and chemopreventive strategies.

REFERENCES

1. Califano, J., van der Riet, P., Westra, W., Nawroz, H., Clayman, G., Piantadosi, S., Corio, R., Lee, D., Greenberg, B., Koch, W., and Sidransky, D. Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer Res.*, 56: 2488–2492, 1996.

2. Tusher, V. G., Tibshirani, R., and Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA*, *98*: 5116–5121, 2001.
3. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, *95*: 14863–14868, 1998.
4. Colantuoni, C., Henry, G., Zeger, S., and Pevsner, J. Local mean normalization of microarray element signal intensities across an array surface: quality control and correction of spatially systematic artifacts. *Biotechniques*, *32*: 1316–1320, 2002.
5. Draper, N., and Smith, H. (eds.). *Applied Regression Analysis*, Ed. 2, pp. 169–176. New York: Wiley, 1981.
6. Belbin, T. J., Singh, B., Barber, I., Socci, N., Wenig, B., Smith, R., Prystowsky, M. B., and Childs, G. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Res.*, *62*: 1184–1190, 2002.
7. Hanna, E., Shrieve, D. C., Ratanatharathorn, V., Xia, X., Breau, R., Suen, J., and Li, S. A novel alternative approach for prediction of radiation response of squamous cell carcinoma of head and neck. *Cancer Res.*, *61*: 2376–2380, 2001.
8. Al Moustafa, A. E., Alaoui-Jamali, M. A., Batist, G., Hernandez-Perez, M., Serruya, C., Alpert, L., Black, M. J., Sladek, R., and Foulkes, W. D. Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous carcinoma cells. *Oncogene*, *21*: 2634–2640, 2002.
9. Leethanakul, C., Patel, V., Gillespie, J., Pallente, M., Ensley, J. F., Koontongkaew, S., Liotta, L. A., Emmert-Buck, M., and Gutkind, J. S. Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. *Oncogene*, *19*: 3220–3224, 2000.
10. Alevizos, I., Mahadevappa, M., Zhang, X., Ohyama, H., Kohno, Y., Posner, M., Gallagher, G. T., Varvares, M., Cohen, D., Kim, D., Kent, R., Donoff, R. B., Todd, R., Yung, C. M., Warrington, J. A., and Wong, D. T. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene*, *20*: 6196–6204, 2001.