Identification of a Gene Expression Signature Associated with Recurrent Disease in Squamous Cell Carcinoma of the Head and Neck

Matthew A. Ginos,1 Grier P. Page,2 Bryan S. Michalowicz,3 Ketan J. Patel,1 Sonja E. Volker,1 Stefan E. Pambuccian,4 Frank G. Ondrey,4 George L. Adams,5 and Patrick M. Gaffney1

1Division of Hematology, Oncology and Transplantation, Department of Medicine, University of Minnesota School of Medicine, Minneapolis, Minnesota; 2Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama; 3Department of Preventive Sciences, University of Minnesota School of Dentistry, Minneapolis, Minnesota; 4Department of Laboratory Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, Minnesota; and 5Department of Otolaryngology and Head and Neck Surgery, University of Minnesota School of Medicine, Minneapolis, Minnesota

ABSTRACT

Molecular studies of squamous cell carcinoma of the head and neck (HNSCC) have demonstrated multiple genetic abnormalities such as activation of various oncogenes (Ras, Myc, epidermal growth factor receptor, and cyclin D1), tumor suppressor gene inactivation (TP53 and p16), and loss of heterozygosity at numerous chromosomal locations. Despite these observations, accurate and reliable biomarkers that predict patients at highest risk for local recurrence have yet to be defined. To identify gene expression signatures that may serve as biomarkers, we studied 41 squamous cell carcinoma tumors (25 primary and 16 locally recurrent) from various anatomical sites and 13 normal oral mucosal biopsy samples from healthy volunteers with microarray analysis using Affymetrix U133A GeneChip arrays. Differentially expressed genes were identified by calculating generalized t tests (P < 0.001) and applying a series of filtering criteria to yield a highly discriminant list of 2890 genes. Hierarchical clustering and image generation using standard software were used to visualize gene expression signatures. Several gene expression signatures were readily identifiable in the HNSCC tumors, including signatures associated with proliferation, extracellular matrix production, cytokine/chemokine expression, and immune response. Of particular interest was the association of a gene expression signature enriched for genes involved in tumor invasion and metastasis with patients experiencing locally recurrent disease. Notably, these tumors also demonstrated a marked absence of an immune response signature suggesting that modulation of tumor-specific immune responses may play a role in local treatment failure. These data provide evidence for a new gene expression-based biomarker of local treatment failure in HNSCC.

INTRODUCTION

Head and neck malignancies account for 6% of all cancers diagnosed in the United States and result in an estimated 14,000 deaths annually (1). Although improvements in local control and survival have been achieved with the use of combined modality therapies, 5-year survival rates for head and neck cancers have not improved significantly over the past 20 years (2, 3). Local-regional relapse after definitive therapy is a major cause of morbidity and mortality in patients with head and neck squamous cell carcinoma (HNSCC) and has prompted substantial efforts in identifying molecular biomarkers that accurately predict patients at risk for disease recurrence (4). Identification of biomarkers that signal increased risk of treatment failure in HNSCC would have a major impact on treatment planning decisions. For instance, biomarkers could be used to identify surgically treated patients at highest risk for local regional recurrence, thus allowing improved accuracy in selecting patients most likely to benefit from postoperative radiation therapy. In addition, biomarkers that predict local recurrence may be useful in identifying patients who may be spared postradiation neck dissection or those likely to benefit from concomitant chemoradiation approaches, altered fractionation radiation schedules, or new molecular-targeted therapies.

Many clinical and pathological prognostic factors have been described in HNSCC (tumor stage, lymph node involvement, postsurgical margin, and histological grade). However, as prognostic biomarkers, clinical or pathological factors lack sensitivity and accuracy in the clinical setting and, with the exception of disease stage, are infrequently used to guide treatment decisions. More recent work has focused on defining tumor-specific molecular abnormalities that predict patient outcome (4, 5). These studies have revealed numerous molecular abnormalities in HNSCC, including activation of various oncogenes (Ras, Myc, epidermal growth factor receptor, and cyclin D1;Refs. 6–14), tumor suppressor gene inactivation (TP53 and p16; Refs. 15–19), expression of angiogenic factors (20–22), DNA ploidy (23, 24), and loss of heterozygosity at numerous chromosomal locations (25–28). Despite these molecular-based observations, biomarkers that precisely identify patients at highest risk for local recurrence have yet to be defined. This may be due, in part, to the molecular heterogeneity of HNSCC and the limited capacity of any single molecular biomarker to accurately and reliably predict outcome in individual patients. A more comprehensive screen of the molecular defects in HNSCC such as that obtained through microarray analysis may reveal biomarkers with enhanced sensitivity and specificity in the clinical setting.

High-throughput gene expression profiling techniques offer a unique mechanism for interrogating transcriptome-wide levels of gene expression and have proven value in defining gene expression signatures for clinically important subsets of patients who would otherwise be undetected by conventional prognostic strategies (29, 30). In addition, microarray approaches have been very successful in identifying subsets of tumors that correlate with clinical parameters such as survival, histological grade, and response (31–35). In this study, we applied gene expression microarray technology to a group of 41 HNSCC tumors resected at surgery and compared the gene expression to 13 normal oral mucosal biopsies from healthy volunteers. Our aims were to obtain a comprehensive view of gene expression differences in HNSCC and normal oral mucosa (NOM) and to identify gene expression signatures that correlate with defined clinical parameters with the hope of identifying new prognostic gene expression-based biomarkers in HNSCC.

MATERIALS AND METHODS

Patient Characteristics and Biopsy Samples

Consent from patients and volunteer control subjects was obtained in accord with guidelines set forth by the Institutional Review Board of the Human Subjects Protection Committee at the University of Minnesota. The University of Minnesota Cancer Center Tissue Procurement Facility obtained tumor samples from surgical resection specimens from patients undergoing surgery for HNSCC using standardized procedures. All samples were immediately
GENE EXPRESSION PROFILING IN HEAD AND NECK CANCER

placed on ice and within 30 min of devascularization frozen in liquid nitrogen after removal of portions needed for pathological diagnosis. Histological analyses were performed to ensure that each specimen contained >50% tumor tissue and <10% necrotic debris, and those samples not meeting these criteria were rejected. Healthy control subjects without a history of oral cancer, premalignant lesions, or periodontal disease were recruited through the University of Minnesota School of Dentistry. After administration of a local anesthetic, a 6-mm punch biopsy of tissue was obtained from the buccal mucosa in the region adjacent to the third molar. Tissue specimens were flash frozen in liquid nitrogen and stored until extraction of mRNA. On average these tissues contained 40% epithelial mucosa and 60% submucosa tissue.

Extraction of Total RNA and Probe Preparation

Fifty to 100 mg of tissue were submerged in 1 ml of Trizol reagent (Life Technologies, Inc., Gaithersburg, MD) and immediately homogenized using a rotor-stator homogenizer (Powergen 700; Fischer Scientific) under RNAse-free conditions. Total RNA was extracted from the samples using the Trizol extraction protocol after a 1-min spin at 12,000 × g to pellet particulate matter. Total RNA was precipitated by incubating with 0.5 ml of isopropl alcohol for 10 min followed by centrifugation at 12,000 × g for 10 min at 4°C. The pellet was washed twice with 75% ethanol, dissolved in RNAse-free water, and stored at −80°C until further use. Agarose gel electrophoresis was performed on each sample before additional analysis to confirm the presence of nondegraded RNA.

Five to 10 μg of total RNA were used to prepare biotinylated cRNAs for hybridization using the standard Affymetrix protocol (Affymetrix, Santa Clara, CA). Briefly, RNA was converted to first strand cdNA using a T7-linked oligodeoxythymidylic acid primer (Genset, La Jolla, CA), followed by second strand synthesis (Invitrogen Corporation, Carlsbad, CA). The double-stranded cdNA was then used as template for labeled in vitro transcription reactions using biotinylated ribonucleotides (Enzo, Farmingdale, NY). Fifteen μg of each labeled cRNA were hybridized to Affymetrix U133A GeneChips (Affymetrix, Santa Clara, CA) using standard conditions in an Affymetrix fluids station.

Analysis of Microarray Data

Preprocessing of Microarray Data. Scanned Affymetrix array data were uploaded into the GeneData Cobi 4.0 database maintained by the Supercom- puting Institute’s Computational Genomics Laboratory at the University of Minnesota. Preprocessing of the Affymetrix arrays was carried out using GeneData Refiner 3.0 software to correct for variations in hybridization intensity because of gradient effects, dust specks, or scratches. Gene expression intensity for each array was scaled to an arbitrary value of 1500 intensity units because of gradient effects, dust specks, or scratches. Gene expression intensity for each array was scaled to an arbitrary value of 1500 intensity units to allow comparisons across all arrays. Expression intensity values for each gene were derived using Refiner by applying the Microarray Suite 5.0 algorithm.

Statistical Analysis. Genes differentially expressed between the 41 HNSCC and 13 normal oral mucosal biopsy specimens were identified using a Satter-thwaite test (36) to robustly estimate significance despite unequal variance among groups. Genes that met additional investigation met the following criteria: P < 0.001, absolute value of the difference in mean expression between the two groups of samples (Δ) > 100 intensity units, and fold difference in mean expression ≥ 2.0 or ≤ 0.5. Unsupervised hierarchical clustering was performed with Cluster (37) using Pearson’s correlation distance metric and average linkage followed by visualization in Treeview (37). Because of the large number of statistical tests performed we estimated the Bayesian posterior true positive, posterior true negative, and posterior false negative rates for each of the genes using the mix-o-matic method of Allison et al. (38) and to illustrate that more genes were observed to be differentially expressed than expected (supplemental data Figure 1; http://www.gaffney.umn.edu/suppl).

Immunohistochemistry

Immunohistochemical analysis was performed to validate the differential expression of selected genes in tissue sections and to localize the expression of the genes. On the basis of their gene expression profiles, 4 of the HNSCC (2 samples with high expression and 2 with low expression of inflammation and immune response-related genes) were selected. Sections of HNSCC were formalin-fixed and paraffin-embedded, sectioned at 4-μm and immunostained according to standard protocols. Immunostaining with antibodies against CD1 (PSL, prediluted; Ventana Medical Systems, Inc., Oro Valley, AZ) and cyto- keratin (AE1 AE3, prediluted; Dako Corporation, Carpinteria, CA) was performed on the automated immunostainer NexES IHC (Ventana Medical Systems, Inc.).

Real-Time PCR Analysis

Changes in mRNA levels were compared by quantitative real-time PCR analysis, using the Light Cycler (Roche Diagnostics Corp). One μg of total RNA was converted to cDNA using Superscript II (Invitrogen Corporation) according to manufacturers specifications. PCR reaction mixtures consisted of 2 μl of Faststart DNA Master SYBR Green I mixture [containing TaqDNA polymerase, reaction buffer, deoxynucleotide triphosphate mix (with dUTP instead of dTTP), SYBR Green I dye, and 10 mM MgCl2], 0.5 μM of each target primer stock, 2 or 4 mM MgCl2 [glyceraldehyde-3-phosphate dehydro- genase (GAPDH), PTEN, PTGTA, 4 μM SNAI2, and 2 mM MET] in a final reaction volume of 20 μl. Cumulative fluorescence was measured at the end of the extension phase of each cycle. Product-specific amplification was confirmed by a melting curve analysis and agarose gel electrophoresis analysis. Quantifi- cation was performed at the log-linear phase of the reaction and cycle numbers obtained at this point were plotted against a standard curve prepared with serially diluted samples. Results were normalized to GAPDH. Primer sequences and reaction parameters are as follows: GAPDH, forward, 5’- ACCACAGTCCTCAGATCCATC-3’ and reverse, 5’-TCCACCACCTGTTCGTGTGA-3’; MET, forward, 5’-GAAGGAGGCAAGCTGTCAC-3’ and reverse, 5’-ATGGCCAAGACCAAGCTGTC-3’; PTEN, forward, 5’-GCTCAGGACCCCTGGGAAAG-3’ and reverse, 5’-AGCGCCTTACGCGCAAGACCAAGCTGTC-3’; and SNAI2, forward, 5’-CATCTTTGGGCGGAGTGTCC-3’ and reverse, 5’-GGCCAGCCACAGAAGTTGAAT-3’.

RESULTS

To identify differentially expressed genes present in HNSCC and NOM, we obtained specimens from 41 patients undergoing surgical resection for HNSCC. Clinical characteristics of the 41 patients are shown in Table 1 and supplemental data Table 1. This study population closely parallels the general population of patients with HNSCC having a median age of 64 years and greater percentage of males with disease. The majority of cancers were located in the oral cavity (44%)
and larynx (37%) and over two-thirds of the cancers were advanced stage (stage III/IV) consistent with referral patterns to the University of Minnesota. Ninety percent of patients smoked tobacco and one-third drank alcohol in excess of 4 drinks/day. We chose to compare gene expression in HNSCC to that of NOM from healthy volunteers because these tissues would likely provide the widest spectrum of gene expression differences and would not be subject to the field cancerization effect observed in this malignancy when using normal mucosa adjacent to resected tumor. No attempt was made to control for tobacco or alcohol consumption in the control group assuming that the potential influence of these factors on gene expression would likely be overshadowed by the differences between normal and neoplastic tissue. Of the 13 normal subjects, 1 was currently a smoker, 5 had smoked at some time in the past, whereas the remaining 7 had never smoked. None of the 13 normal subjects reported excessive alcohol use defined as ≥4 drinks/day.

To visualize the gene expression data, hierarchical clustering was performed using genes that satisfied stringent filtering criteria

---

Fig. 1. Gene expression profiles of 41 HNSCC and 13 normal oral mucosal biopsy samples. Differentially expressed genes were identified by parametric T tests with correction for unequal sample variance. A total of 2890 genes were identified using the following filter criteria: t test \( P < 0.001 \), difference in mean intensity ≥ 100, fold change in mean intensity ≥ 2.0 or ≤ 0.5. To visualize the data, genes and samples were grouped using Cluster and graphically represented in Treeview. Prominent gene expression signatures are highlighted with colored bars to the right of the figure. B, immunohistochemistry staining of two representative tumors with and without the immune response signature using antibodies against cytokeratin and CD3, markers for squamous carcinoma and T lymphocytes, respectively.

---

7 Internet address: http://rana.lbl.gov.
Table 2 Functional gene classes demonstrating significant differential expression between HNSCC and NOM

<table>
<thead>
<tr>
<th>Function Name</th>
<th>Total*</th>
<th>$P^b$</th>
<th>FDR$^c$</th>
<th>Bonferroni$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflammatory response</td>
<td>71</td>
<td>1.11E-16</td>
<td>4.72E-14</td>
<td>4.72E-14</td>
</tr>
<tr>
<td>Immune response</td>
<td>95</td>
<td>8.44E-15</td>
<td>1.79E-12</td>
<td>3.59E-12</td>
</tr>
<tr>
<td>Epidermal differentiation</td>
<td>38</td>
<td>1.65E-11</td>
<td>2.34E-09</td>
<td>7.02E-09</td>
</tr>
<tr>
<td>Cell-cell signaling</td>
<td>100</td>
<td>3.14E-10</td>
<td>3.34E-08</td>
<td>1.34E-07</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td>77</td>
<td>5.72E-09</td>
<td>4.86E-07</td>
<td>2.43E-06</td>
</tr>
<tr>
<td>Chemokine</td>
<td>43</td>
<td>8.73E-07</td>
<td>6.18E-06</td>
<td>3.71E-06</td>
</tr>
<tr>
<td>Cellular defense response</td>
<td>40</td>
<td>1.74E-08</td>
<td>1.06E-06</td>
<td>7.39E-06</td>
</tr>
<tr>
<td>Development</td>
<td>80</td>
<td>3.44E-08</td>
<td>1.83E-06</td>
<td>1.46E-05</td>
</tr>
<tr>
<td>Antimicrobial humoral response</td>
<td>45</td>
<td>9.90E-08</td>
<td>4.68E-06</td>
<td>4.21E-05</td>
</tr>
<tr>
<td>Response to viruses</td>
<td>18</td>
<td>7.16E-07</td>
<td>3.04E-05</td>
<td>3.04E-04</td>
</tr>
<tr>
<td>Cell surface receptor linked signal transduction</td>
<td>54</td>
<td>3.29E-06</td>
<td>1.27E-04</td>
<td>1.40E-03</td>
</tr>
<tr>
<td>Cell motility</td>
<td>47</td>
<td>3.55E-06</td>
<td>1.26E-04</td>
<td>1.51E-03</td>
</tr>
<tr>
<td>Cell proliferation</td>
<td>79</td>
<td>1.81E-05</td>
<td>5.90E-04</td>
<td>7.67E-03</td>
</tr>
<tr>
<td>Protein biosynthesis</td>
<td>6</td>
<td>1.81E-05</td>
<td>5.49E-04</td>
<td>7.69E-03</td>
</tr>
<tr>
<td>Skeletal development</td>
<td>36</td>
<td>2.59E-05</td>
<td>7.34E-04</td>
<td>1.10E-02</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>221</td>
<td>6.40E-05</td>
<td>1.70E-03</td>
<td>2.72E-02</td>
</tr>
<tr>
<td><strong>Cellular component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracellular matrix</td>
<td>42</td>
<td>1.21E-10</td>
<td>1.59E-08</td>
<td>1.59E-08</td>
</tr>
<tr>
<td>Extracellular space</td>
<td>105</td>
<td>1.70E-09</td>
<td>1.12E-07</td>
<td>2.23E-07</td>
</tr>
<tr>
<td>Extracellular</td>
<td>40</td>
<td>5.69E-08</td>
<td>2.49E-06</td>
<td>7.46E-06</td>
</tr>
<tr>
<td>Integrin complex</td>
<td>13</td>
<td>3.22E-05</td>
<td>1.06E-03</td>
<td>4.22E-03</td>
</tr>
<tr>
<td>Intermediate filament</td>
<td>18</td>
<td>4.22E-05</td>
<td>1.11E-03</td>
<td>5.53E-03</td>
</tr>
<tr>
<td>Plasma membrane</td>
<td>135</td>
<td>8.73E-05</td>
<td>1.91E-03</td>
<td>1.14E-02</td>
</tr>
<tr>
<td>Collagen type IV</td>
<td>8</td>
<td>9.89E-05</td>
<td>1.85E-03</td>
<td>1.30E-02</td>
</tr>
<tr>
<td>Cytosol</td>
<td>30</td>
<td>1.96E-04</td>
<td>3.21E-03</td>
<td>2.57E-02</td>
</tr>
<tr>
<td><strong>Molecular function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA binding</td>
<td>9</td>
<td>6.38E-12</td>
<td>3.33E-09</td>
<td>3.33E-09</td>
</tr>
<tr>
<td>Chemokine</td>
<td>29</td>
<td>3.23E-11</td>
<td>8.43E-09</td>
<td>1.69E-08</td>
</tr>
<tr>
<td>Collagen</td>
<td>29</td>
<td>1.30E-10</td>
<td>2.26E-08</td>
<td>6.97E-08</td>
</tr>
<tr>
<td>Cell adhesion molecule</td>
<td>59</td>
<td>1.43E-09</td>
<td>1.86E-07</td>
<td>7.44E-07</td>
</tr>
<tr>
<td>Structural constituent of muscle</td>
<td>24</td>
<td>4.48E-08</td>
<td>4.68E-06</td>
<td>2.34E-05</td>
</tr>
<tr>
<td>Cytokine</td>
<td>18</td>
<td>9.40E-06</td>
<td>7.01E-04</td>
<td>4.91E-03</td>
</tr>
<tr>
<td>Hyaluronic acid binding</td>
<td>8</td>
<td>1.34E-05</td>
<td>8.74E-04</td>
<td>6.99E-03</td>
</tr>
<tr>
<td>Cell adhesion receptor</td>
<td>21</td>
<td>1.69E-05</td>
<td>9.79E-04</td>
<td>8.81E-03</td>
</tr>
<tr>
<td>Proteoglycan</td>
<td>9</td>
<td>1.72E-05</td>
<td>8.94E-04</td>
<td>8.98E-03</td>
</tr>
<tr>
<td>Structural constituent of cytoskeleton</td>
<td>37</td>
<td>1.90E-05</td>
<td>9.00E-04</td>
<td>9.89E-03</td>
</tr>
<tr>
<td>Endopeptidase inhibitor</td>
<td>17</td>
<td>2.34E-05</td>
<td>1.02E-03</td>
<td>1.22E-02</td>
</tr>
<tr>
<td>Tropomyosin binding</td>
<td>8</td>
<td>2.80E-05</td>
<td>1.12E-03</td>
<td>1.46E-02</td>
</tr>
<tr>
<td>Extracellular matrix glycoprotein</td>
<td>12</td>
<td>5.91E-05</td>
<td>2.20E-03</td>
<td>3.08E-02</td>
</tr>
<tr>
<td>Collagenase</td>
<td>5</td>
<td>7.87E-05</td>
<td>2.74E-03</td>
<td>4.11E-02</td>
</tr>
</tbody>
</table>

* Total number of genes identified in list of 2890 differentially expressed genes corresponding to specific Gene Ontology classification.

Binomial distribution $P$-value.

Adjusted $P$-value using False Discovery Rate (FDR) correction (Ref. 47).

Adjusted $P$-value using Bonferroni correction ($P$-value × total genes for that function).

($P < 0.001$, $\Delta > 100$, fold difference $\geq 2.0$ or $\leq 0.5$) yielding a highly discriminant set of 2890 genes (Fig. 1A and supplemental data Table 2). A number of gene expression signatures are apparent by visual inspection mostly associated with genes over expressed in HNSCC. The expression pattern of genes down-regulated in HNSCC compared with NOM is relatively uniform, although a few clusters of functionally relevant genes can be defined. A cluster of genes mapping to a region of chromosome 1q21 (involutin, SPRR 1A, 1B, 2B, 3, NICE-1, S100 A1, 6, 8, 9, 14), termed the epidermal differentiation complex, is down-regulated in the majority of HNSCC samples. Genes located in this region are highly expressed in terminally differentiated squamous epithelium and have been shown to play an important role in maintenance of the epithelial phenotype (39). Down-regulation of these genes in HNSCC is consistent with dedifferentiation of squamous epithelial cells associated with malignant transformation. Another cluster of genes down-regulated in HNSCC includes several extracellular matrix proteins regulating epithelial adhesion and collagen fibril formation [tenasin X, dermatopontin, and collagen 14 type, α1 (undulin)]. No specific role in cancer progression has been ascribed to dermatopontin or collagen 14. However, tenasin X knockout mice demonstrate increased susceptibility to tumor invasion and metastasis after inoculation of footpads with B16-BL6 melanoma cells (40).

Genes up-regulated in the HNSCC samples formed several signatures enriched for genes associated with an array of biological processes, including extracellular matrix functions, interferon response, proliferation, and immune response. The proliferation signature included many genes involved in cell cycle control and chromosome maintenance and has been identified in transcriptional profiling studies of other cancers and likely represents a common cancer-specific gene expression program (41). A prominent signature associated with genes expressed in inflammation and immune response was present in over half of the squamous carcinomas and likely represents a robust tumor immune response. To investigate whether this signature was in part because of tumor infiltrating immune cells and not because of aberrant expression of these genes by HNSCC, we selected two tumors expressing and not expressing this signature and performed immunohistochemistry staining using antibodies that identify squamous carcinoma (cytokeratin) and T lymphocytes (CD3). Although the percentage of squamous carcinoma in each specimen is roughly equivalent, tumor-infiltrating T lymphocytes are clearly more abundant in tumors expressing the immune response signature, thus verifying the immune response signature is due, in part, to tumor-infiltrating T lymphocytes (Fig. 1B).

Another prominent gene expression signature highly expressed in a subset of HNSCC includes genes implicated in invasion (snail homologue 2, met proto-oncogene; Refs. 42, 43), RAS signaling, (r-ras homologue 2, cell surface/extracellular matrix interactions (collagen type IV α5 and α6, laminin γ2, α5, β1; integrin α5, α6, β4, β5), and angiogenesis (vascular endothelial growth factor C, placental growth factor). The high expression of these genes in this subset of tumors...
suggests a potentially aggressive phenotype prone to invasion and metastasis.

To more thoroughly characterize sets of functionally related genes differentially expressed between HNSCC and NOM, we used Onto-Express (44, 45) to classify genes according to the following Gene Ontology (GO) (46) categories: biological process; cellular role; and molecular function. The numbers of genes corresponding to each GO category among the 2890 differentially expressed genes was tallied and compared with the number of genes expected for each GO category based on their representation on the Affymetrix U133A array. Significant differences from the expected were calculated with a two-sided binomial distribution. False discovery rates (47) and Bonferroni adjustments were also calculated based upon the number of GO categories having at least 1 gene in the list of 2890 differentially expressed genes. Table 2 shows all GO functional classes with a Bonferroni-corrected significance of $P < 0.05$, the significance of each class, the false discovery rate for that class, as well as the number of genes corresponding to each GO functional class identified in our differentially expressed gene list. The functional gene groups demonstrating the most significant representation in our set of differentially expressed genes appear under the biological process ontology and map to the inflammatory response, immune response, and epidermal differentiation categories. Genes involved in inflammatory and immune response are highly expressed in subsets of HNSCC and correlate with the presence of tumor infiltrating immune cells (Fig. 1, A and B). In contrast, genes corresponding to the epidermal differentiation category are highly expressed in NOM compared with HNSCC, reflecting the loss of normal epithelial architecture associated with malignant transformation (Fig. 1A). Functional categories significantly represented under the cellular component and molecular function ontologies include genes involved in extracellular matrix functions, integrin complexes, RNA binding, chemokine signaling, and cell adhesion.

A major aim of this study was to identify correlations between gene expression signatures and clinical parameters with the goal of identifying new gene expression-based biomarkers of HNSCC. To identify associations between genes and clinical parameters, we calculated correlation coefficients for each gene across all 41 HNSCC tumor samples with stage of disease, tumor grade (well, moderate, or poorly differentiated), primary tumor site, and presence of locally recurrent disease. The significance of the correlation was determined by permuting the class labels (5000 permutations) pertaining to each clinical parameter and calculating a level at which the correlation would possess a type 1 error of $1\%$. The resulting correlation curves are plotted alongside the corresponding gene as a moving average (window size $n = 30$ genes; Fig. 2A). Interestingly, no correlations exceeding the $1\%$ type one error threshold were seen between gene expression signatures and
tumor stage, tumor grade, or anatomical location of the primary tumor. However, a correlation (correlation coefficient = 0.42, \( P < 0.01 \)) exceeding this threshold was detected between patients who experienced recurrent disease and the signature previously identified as enriched for genes associated with tumor invasion and metastasis (Fig. 1A). Importantly, the signature associated with recurrent disease was present in 7 primary tumors procured before administration of definitive therapy (Fig. 2B). Therefore, it is quite likely that the recurrent disease signature is an intrinsic feature of a subset of HNSCC tumors and not simply a biased result secondary to analyzing a collection of recurrent tumors. Furthermore, all 7 patients whose tumors expressed the recurrent disease signature recurred within 12 months (median time to recurrence = 8 months) of completing therapy, indicating that this signature may function as a predictor of early local treatment failure in HNSCC.

To confirm the findings of the microarray analysis we performed real-time PCR (quantitative-PCR) analysis using primers specific for MET, SNAI2, and PTHLH and compared the fold differences in expression predicted by microarray to that obtained with quantitative-PCR. Twelve tumor samples demonstrating a wide range of expression values for MET, SNAI2, and PTHLH by microarray were selected for analysis. These genes were selected for confirmation because of their prominent overexpression in the recurrent disease signature. For expression values from microarray, the mean intensity for each probe set [two for MET (Affymetrix probe IDs: 203510_at, 213807_x_at), one for SNAI2 (213139_at), and three for PTHLH (206300_s_at, 210355_at, 211756_at)] was normalized to the mean GAPDH intensity [three probe sets (M33197_5_at, M33197_M_at, and M33197_3_at)] for each of the 12 tumor samples. We selected GAPDH as the normalization standard because it was uniformly expressed across all tumor samples by microarray. Normalized expression values from microarrays were then ranked ordered from highest to lowest for each of the transcripts and fold changes calculated by dividing the sample with the highest expression by the sample with the lowest expression. This was repeated sequentially using the next highest and lowest values until all fold change calculations were made. The procedure was then repeated using the corresponding normalized expression levels for each transcript for all 12 samples derived by quantitative-PCR. Fold change concordance for each transcript is shown in Table 3. Quantitative-PCR confirmed the direction of fold change for nearly all of the samples where the microarray fold change was >2.0 with one exception. Overall, these results confirm our findings of differential gene expression by microarray analysis.

**DISCUSSION**

The objectives of this study were to investigate the range of gene expression differences between HNSCC and NOM to identify gene expression programs disregulated in HNSCC and to correlate gene expression signatures with clinical parameters to identify potentially new biomarkers of HNSCC. Using statistical and data filtering criteria, we identified 2890 genes differentially expressed between HNSCC and NOM. Functional gene classes (defined by GO annotation) highly represented in HNSCC include those involved in inflammatory response, immune response, epidermal differentiation, cell adhesion, and extracellular matrix functions.

The presence of a robust immune response in a subset of HNSCC tumors is associated with infiltrating T lymphocytes within the tumor (Fig. 1B), a phenomenon well documented in the literature for HNSCC as well as other cancers (48). Earlier studies on HNSCC reported that a peritumoral lymphocyte infiltrate was associated with a better prognosis when compared with tumors not demonstrating immune infiltrate (49–53). In addition, greater peritumoral lymphocyte infiltrate demonstrates a much stronger correlation with E-cadherin transcriptional repression than Snail (64).

Table 3 Concordance in fold change direction between microarray and real-time PCR (QPCR) expression levels

<table>
<thead>
<tr>
<th>MET</th>
<th>PTHLH</th>
<th>SNAI2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microarray</td>
<td>QPCR</td>
<td>Microarray</td>
</tr>
<tr>
<td>37.7</td>
<td>8.2</td>
<td>172.2</td>
</tr>
<tr>
<td>8.7</td>
<td>14.9</td>
<td>115.3</td>
</tr>
<tr>
<td>1.9</td>
<td>0.3</td>
<td>36.3</td>
</tr>
<tr>
<td>1.5</td>
<td>2.2</td>
<td>10.6</td>
</tr>
<tr>
<td>1.4</td>
<td>1.3</td>
<td>7.3</td>
</tr>
<tr>
<td>1.1</td>
<td>0.7</td>
<td>1.7</td>
</tr>
</tbody>
</table>

- **MET** and **PTHLH** have been shown to be expressed in HNSCC and NOM. **SNAI2** is a known transcriptional repressor of E-cadherin gene expression and triggers a complete epithelial/mesenchymal transition in MDCK cell lines (42). In breast cancer cell lines, both Snail and SNAI2 repress E-cadherin gene expression, however, **SNAI2** demonstrates a much stronger correlation with E-cadherin transcriptional repression than Snail (64). **SNAI2** also ap-
pears to modulate sensitivity to apoptotic stimuli because it is able to protect hematopoietic progenitor cells from radiation-induced apoptosis in vivo (65). To our knowledge, this is the first evidence demonstrating overexpression of SNAI2 in HNSCC.

The recurrent disease signature also contains several genes important in tumor cell/extracellular matrix interactions. For instance, laminin α3, β3, and γ2 and integrin α6 and β4 are all present in this signature. The three laminin subunits together comprise laminin 5, which serves as the ligand for the α6β4 integrin. Laminin 5 and integrin α6β4 are up-regulated in squamous carcinoma of the skin, colon, esophagus, and larynx, and higher expression is correlated with increased invasiveness (66). The importance of laminin 5–α6β4 integrin interactions in promoting invasive squamous carcinoma was recently demonstrated using a model in which human epithelial keratinocytes stably transplanted with oncogenic RAS and Kras were transplanted into scid/scid mice (67). The identification of laminin 5 subunits α6β4 expressed in the recurrent disease signature from our study provides additional evidence that this signature is associated with a highly invasive subset of HNSCC tumors.

Several gene expression profiling studies in head and neck cancer have been performed recently (68–79). These studies have used a variety of array platforms, tissue types, and experimental designs, making direct comparisons with our data difficult. However, the results of two studies, performed on primary tumor tissues with reasonably large sample sizes, can be compared with our study. The study by Belbin et al. (68) evaluated gene expression signatures in 17 patients with HNSCC. Their study was a two-color cDNA array analysis using mRNA from normal human epithelial keratinocytes as a control. Two groups of squamous cell carcinomas were identified based on supervised clustering with confirmation of the clustering results performed using a bootstrap resampling method. Modest association was seen with poorer cause-specific survival at 2 years in tumors expressing a group 1 signature.

In other work, Mendez et al. (69) evaluated 26 oral squamous carcinomas, 2 premalignant lesions, and 18 normal mucosal samples, 9 of which were adjacent to oral cancers. They used the Affymetrix HuGene FL platform for their analyses and a series of regression modeling techniques and unsupervised clustering methodologies to determine differentially expressed genes. The aim of this study was to identify gene expression signatures that differentiate early-stage from late-stage disease or metastatic from nonmetastatic cancers. No significant differences in gene expression were seen that defined early-stage versus late-stage disease or metastatic from nonmetastatic cancers. Our data concurs with their finding that gene expression signatures do not provide substantial amounts of information related not only to the tumorigenic process and may serve as an important biomarker of local treatment failure in HNSCC. Future studies will focus on confirming these findings in an independent, prospectively collected cohort of HNSCC patients with the goal of implementing gene expression-based biomarkers into the clinical management of patients with HNSCC.

ACKNOWLEDGMENTS

We thank Drs. Timothy Behrens, Kathy Moser, Erik Peterson, Brian Van Ness, and Jennifer Hall for critical review and helpful discussions during the preparation of this article. We also thank Sarah Bowell and Diane Rauch of the University of Minnesota Cancer Center Tissue Procurement Facility, Aaron Becker and Jill Plumb-Smith of the University of Minnesota Affymetrix Core Facility, and SeonaCherie Hanson of the Department of Laboratory Medicine and Pathology Immunohistochemistry Laboratory for their expert technical assistance.

REFERENCES

GENE EXPRESSION PROFILING IN HEAD AND NECK CANCER


Downloaded from cancers.cancer.org journals on April 15, 2017. © 2004 American Association for Cancer Research.


Identification of a Gene Expression Signature Associated with Recurrent Disease in Squamous Cell Carcinoma of the Head and Neck
