Neoplasia • Vol. 8, No. 11, November 2006, pp. 925–932 925

www.neoplasia.com

Global Expression-Based Classification of Lymph Node Metastasis and Extracapsular Spread of Oral Tongue Squamous Cell Carcinoma[1]

Xiaofeng Zhou[*,†,‡], *Stephane Temam*[§,¶], *Myungshin Oh*[*,#], *Nisa Pungpravat*[*], *Bau-Lin Huang*[*], *Li Mao*[§] *and David T. Wong*[*,†,**]

[*]Dental Research Institute, School of Dentistry, University of California at Los Angeles, Los Angeles, CA, USA; [†]Jonsson Comprehensive Cancer Center, University of California at Los Angeles, Los Angeles, CA, USA; [‡]Center for Molecular Biology of Oral Diseases, College of Dentistry, University of Illinois at Chicago, Chicago, IL, USA; [§]Department of Thoracic/Head and Neck Medical Oncology, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA; [¶]Department of Head and Neck Surgery, Institut Gustave-Roussy, Villejuif, France; [#]Department of Biostatistics, School of Public Health, University of California at Los Angeles, Los Angeles, CA, USA; **Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA, USA

Abstract

Regional lymph node metastasis is a critical event in oral tongue squamous cell carcinoma (OTSCC) progression. The identification of biomarkers associated with the metastatic process would provide critical prognostic information to facilitate clinical decision making for improved management of OTSCC patients. Global expressional profiles were obtained for 25 primary OTSCCs, where 11 cases showed lymph node metastasis (pN^+) histologically and 14 cases were non-metastatic (pN^-). Seven of pN^+ cases also exhibited extracapsular spread (ECS) of metastatic nodes. Multiple expression indices were used to generate signature gene sets for $pN^{+/-}$ and $ECS^{+/-}$ cases. Selected genes from signature gene sets were validated using quantitative reverse transcription–polymerase chain reaction (qRT-PCR). The classification powers of these genes were then evaluated using a logistic model, receiver operating characteristic curve analysis, and leave-one-out cross-validation. qRT-PCR validation data showed that differences at RNA levels are either statistically significant ($P < .05$) or suggestive ($P < .1$) for six of eight genes tested (*BMP2*, *CTTN*, *EEF1A1*, *GTSE1*, *MMP9*, and *EGFR*) for $pN^{+/-}$ cases, and for five of eight genes tested (*BMP2*, *CTTN*, *EEF1A1*, *MMP9*, and *EGFR*) for $ECS^{+/-}$ cases. Logistic models with specific combinations of genes (*CTTN+MMP9+EGFR* for pN and *CTTN+EEF1A1+MMP9* for ECS) achieved perfect specificity and sensitivity. Leave-one-out cross-validation showed overall accuracy rates of 85% for both pN and ECS prediction models. Our results demonstrated that the pN and the ECS of OTSCCs can be predicted by gene expression analyses of primary tumors.

Neoplasia (2006) **8,** 925–932

Keywords: Squamous cell carcinoma, lymph node metastasis, extracapsular spread, microarray, leave-one-out cross-validation.

Introduction

Head and neck squamous cell carcinomas (HNSCCs) are a heterogeneous group of tumors that arise from the epithelium of the upper aerodigestive tract. HNSCC is the sixth most common malignancy in humans and is associated with high alcohol and tobacco use. Despite tremendous improvements in surgery, radiotherapy, and chemotherapy over the last decade, the prognosis for patients with HNSCC has been more or less unchanged for the past three decades. This is because patients continue to die from metastatic diseases at regional and distant sites. Improvement in patient survival requires an increased understanding of tumor metastasis so that aggressive tumors can be detected early in the disease process and targeted therapeutic interventions can be developed. Detection of local lymph node metastasis is pivotal for choosing appropriate treatment, especially for individuals diagnosed with HNSCC in the oral cavity or oropharynx [1]. Most of these individuals have their primary tumor removed. Treatment of individuals clinically diagnosed with lymph node metastasis (pN^+ status) involves additional surgical removal of a substantial portion of the neck, including all five local lymph node levels [called radical neck dissection (RND)]. On histologic examination of removed tissues, 10% to 20% of clinically diagnosed pN^+ individuals turn out to be metastasis-free (pN^-) [2]. Clinical diagnosis of pN^- status is even less accurate. Postoperative histologic examination shows that approximately one third of clinically diagnosed pN^- individuals have metastasis-positive lymph nodes in the neck [3]. Currently, there are several different strategies for

Address all correspondence to: David T. Wong, DMD, DMSc, UCLA School of Dentistry, PO Box 951668, Los Angeles, CA 90095-1668. E-mail: dtww@ucla.edu

[1]This work was supported, in part, by National Institutes of Health grants R01 DE015970 (to D. Wong), P01 CA106451 Project 3 (to L. Mao), K22 DE014847, RO3 DE016569, and RO3 CA114688; Tobacco-Related Disease Research Program (TRDRP) grant 13KT-0028 (to X. Zhou); and a Foundation de France grant (to S. Temam).

Received 1 June 2006; Revised 21 August 2006; Accepted 23 August 2006.

DOI 10.1593/neo.06430

treating diagnosed pN^- individuals [4]. In the "watch-and-wait" strategy, diagnosed pN^- individuals do not undergo any neck dissection; this risks fatality by allowing overlooked metastases to spread further. Because false-negative rate is very high, most clinics carry out neck surgery for all diagnosed pN^- individuals. In these cases, supraomohyoid neck dissection (SOHND) is performed, which removes the three upper lymph node levels [5]. SOHND is less appropriate than RND for pN^+ individuals falsely diagnosed as pN^- and, moreover, is completely unnecessary for individuals correctly diagnosed as pN^-. Although SOHND is less rigorous than RND, the treatment causes disfigurement, long-term discomfort, and pain, and can lead to additional complications such as shoulder and neck disability [6–9]. Both strategies result in inappropriate treatment because of limitations in reliably detecting lymph node metastasis. In addition to nodal metastasis, many studies have also suggested that extracapsular spread (ECS) of lymph node metastasis is one of the most important negative prognostic factors for several different cancer types [10–13], including head and neck cancers [14,15]. The detection of ECS is currently performed by histologic examination of dissected lymph nodes, which is also prone to the same limitations as for current methods of nodal metastasis diagnosis. This points to an immediate need for new diagnostic strategies.

Currently, no molecular biomarkers have been included in clinical work-up strategies for the detection of nodal metastasis and ECS. Because several genes have been reported in retrospective trials to yield prognostic information independently of Tumor–Node–Metastasis (TNM) classification, it is reasonable to hypothesize that molecular "fingerprints" that might define subgroups of patients with significantly more aggressive disease could exist. Tumor cells may progress via the bloodstream or the lymphatic system to colonize new areas of the body. Gene expression signatures of primary tumors have been identified in several tumor types for an increased risk of metastasis [16,17]. The metastasis of HNSCCs is unique in that they metastasize mainly to regional lymph nodes through the draining lymphatics, where metastasis to distant sites is relatively uncommon. Several recent gene expression studies have suggested the existence of such fingerprints in primary tumors for the metastasis of HNSCC [18–20]. In this study, we carry out a genomewide expression analysis of oral tongue squamous cell carcinoma (OTSCC) to identify these fingerprints for nodal metastasis as well as for ECS, and to further validate them using quantitative reverse transcription–polymerase chain reaction (qRT-PCR). Transcriptional profiling capable of predicting ECS phenotypes has not been reported previously; thus, our study should have high substantive significance.

Materials and Methods

Tumor Procurement and RNA Extraction

Discarded surgically resected tissues from T4 OTSCC patients were obtained for this study. This study has been approved by the Institutional Review Boards at the University of California at Los Angeles and the Institut Gustave-Roussy. These tissues were snap-frozen. Clinical characterizations of these patients are outlined in Table 1. There are no significant statistical differences in age and gender among all patient groups ($P > .1$). Tumor stage was determined according to the designated classification of the American Joint Committee on Cancer.

Cancer tissues containing > 80% tumor cells, based on hematoxylin–eosin staining and pathological examination, were identified and selectively microdissected by a trained pathologist. Total RNA was isolated using RNeasy Mini kit (Qiagen, Valencia, CA) and was quantified by the RiboGreen RNA Quantitation Reagent (Molecular Probes, Carlsbad, CA).

Array Hybridization and Data Analysis

Fifty to 200 ng of purified total RNA was amplified by a modified T7 RNA amplification protocol, as described previously [21]. The Enzo BioArray High Yield RNA Transcript Labeling System (Enzo, Farmingdale, NY) was used for labeling the sample before hybridization. Biotinylated cRNA (IVT product) was purified using the RNeasy kit (Qiagen). The quantity and the purity of biotinylated cRNA were determined by spectrophotometry, and an aliquot of the sample was checked by gel electrophoresis. The sample was hybridized to the Affymetrix Human Genome U133 Plus 2.0 GeneChip arrays (Affymetrix, Santa Clara, CA) according to Affymetrix protocols. The arrays were scanned with a GeneChip Scanner 3000. The scanned array images were processed with GeneChip Operating software (GCOS). Microarray data were analyzed using three common microarray analysis methods, including the Affymetrix Microarray Analysis Suite version 5.0 (MAS 5.0) (which is now implemented in GCOS) [22], the Model Based Expression Index (MBEI) from Li and Wong [23] (implemented in dChip), and Robust Multiarray Analysis (RMA) from Irizarry et al. [24]. Three separate candidate gene lists of similar sizes ($\sim$100 top genes) were generated by stringent statistical criteria of *t*-test statistics and by fold change for both pN and ECS groupings (99 and 103 transcripts were selected using dChip with $P < .0033$ and

Table 1. Clinical Characterization of OTSCC Patients.

	pN^- (**n** = 14)	pN^+, ALL (**n** = 11)	pN^+, ECS^+ (**n** = 7)
Age (years)			
Median	54	65	66
Range	41–67	37–82	37–78
Mean	55	61	63
Gender (%)			
Male	86	64	43
Female	14	36	57
Tumor site (%)			
Tongue	100	100	100
Pathological T stage (%)			
Stage 4	100	100	100
Pathological N stage (%)			
Stage 0	100	0	0
Stage 1	0	9	0
Stage 2	0	91	100
Pathological M stage (%)			
Stage 0	93	100	100
Stage 1	7	0	0

$P < .00095$ and > 1.9-fold and 2.02-fold change in intensity for pN and ECS, respectively; 99 and 98 transcripts were selected using MAS 5.0 with $P < .018$ and $P < .015$ and > 2.42-fold and 2.45-fold change in intensity for pN and ECS, respectively; 102 and 99 transcripts were selected using RMA with $P < .002$ and $P < .002$ and > 1.17-fold and 1.16-fold change in log intensity for pN and ECS, respectively). To narrow down candidate genes, a signature gene list was compiled by selecting genes present in at least two of three initial candidate gene lists. The rationale is that using a combination of several methods increases statistical power to detect a true biomarker, reduces chances of false-positives, and improves reproducibility in future validation experiments.

To display the consistency of the expression pattern of signature gene sets, we performed average linkage hierarchical clustering, an unsupervised method, and multidimensional scaling (MDS). The goal of hierarchical clustering is to graphically display the similarity between genes and samples. MDS is a visualization method used to project differences in gene expression patterns among samples into a multidimensional space and to reveal underlying structures that explain observed similarities among samples.

qRT-PCR

Based on their statistical significance and biologic relevance, selected candidate genes from the signature gene list were chosen for further validation using qRT-PCR. qRT-PCR was performed on all 25 cases of OTSCC, as described previously [25]. RNA was converted to first-strand cDNA using MuLV reverse transcriptase (Applied Biosystems, Foster City, CA), and qPCR was performed using iQ SYBR Green Supermix (Bio-Rad, Hercules, CA) in a Bio-Rad iCycler iQ real-time PCR detection system. All reactions were performed in triplicate. Melting curve analyses were performed to ensure the specificity of qRT-PCR. Primer sets used to test the expression of these selected candidate genes are listed in Table 2. Data analysis was performed using the $2^{-\Delta\Delta C_t}$ method described previously [26], where β-*actin* was used as reference gene. qRT-PCR–based gene expression values between two groups were compared by nonparametric Wilcoxon test.

Prediction Models

To evaluate the classification power for each gene, receiver operating characteristic (ROC) curve analysis is performed based on a logistic model with a binary outcome of $pN^{+/-}$ (or $ECS^{+/-}$) as a dependent variable and with the qRT-PCR value of a candidate gene as an independent variable. The ROC curve shows sensitivity on the *y* axis and (1 − specificity) on the *x* axis for each possible cut point of fitted probabilities from the model. The area under the curve (AUC) was computed through numerical integration of the ROC curve, which measures the overall diagnostic/classification power. In addition, the set of sensitivity and specificity is shown using the best cut point value suggested, which minimizes the absolute value of sensitivity minus specificity. Multivariate classification models were also constructed to determine the best combination of selected candidate genes for $pN^{+/-}$ (or $ECS^{+/-}$). Using the binary outcomes of $pN^{+/-}$ (or $ECS^{+/-}$) as dependent variables, the best-fit logistic model was constructed by stepwise model selection method for $pN^{+/-}$ (or $ECS^{+/-}$) groupings [27]. ROC analysis was performed on this model. We also used leave-one-out cross-validation to evaluate the logistic regression model, as described previously [28]. In brief, this validation procedure removes one observation and finds the best logistic model using stepwise selection with the remaining cases. Then, we predict the class of the left-out case in the best logistic model. This procedure is repeated for each of the observations. The cross-validated accuracy rate is the percentage of all models that correctly predict the left-out sample.

Results

Global gene expression profiling was performed using Affymetrix GeneChip U133+2.0 array on 11 lymph node metastatic (pN^+) and 14 nonmetastatic negative (pN^-) OTSCC samples. A signature gene set for pN^+ and pN^- cases (33 genes; Table 3) was generated by integrating three microarray analyses (MAS 5.0, dChip, and RMA), as described in the Materials and Methods section. This signature gene set provides the classification value for the pN^+ and pN^- groups (Figure 1). The hierarchical clustering misclassified one for each of two groups (Figure 1*A*). Apparent separation between pN^+ and pN^- groups was observed on MDS (Figure 1*B*).

In our pN^+ patient cohort, seven had been clinically diagnosed as ECS^+. A second signature gene list for $ECS^{+/-}$ (22 genes; Table 3) was generated by integrating three microarray analysis methods (MAS 5.0, dChip, and RMA), as described in the Materials and Methods section. This signature gene set can provide superb classification power for

Table 2. Primer Used for Real-Time qRT-PCR.

Gene Symbol	Gene Name	Affymetrix Probe ID	Forward Primer	Reverse Primer
BMP2	Bone morphogenetic protein 2	205289_AT	aactctcctcctgcccctta	tgcaggttcatcgttttcct
EEF1A1	Eukaryotic translation elongation factor 1 α1	204892_X_AT	agtctggtgatgctgccatt	gcgacccaaaggtggatagt
CTTN	Cortactin	214782_AT	ctgagttcttccttccccact	taaatgtgcaggccaaacaag
GTSE1	G-2 and S-phase expressed 1	204318_S_AT	gttctaagccgaaccaaatcc	acctcagcctcccaagttcta
ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	213902_AT	ttgcctcttcgtgaactttg	accacccaaataccctgttg
MMP9	Matrix metalloproteinase 9	203936_S_AT	gcacgacgtcttccagtacc	tcaactcactccgggaactc
EGFR	Epidermal growth factor receptor	201983_S_AT	gcatctttaagggctccaca	actatcctccgtggtcatgc
MTUS1	Mitochondrial tumor suppressor 1	212096_S_AT	tatctctgctcacgcttcca	cagcagggaacaacacaaga

Table 3. Signature Gene Sets for pN Metastasis and ECS Based on the Intersection of Three Different Expression Indices.

Signature Gene Set for $pN^{+/-}$				Signature Gene Set for $ECS^{+/-}$			
Probe ID	Gene Title	Gene Symbol	Intersect*	Probe ID	Gene Title	Gene Symbol	Intersect*
Genes upregulated in pN^+ cases				**Genes upregulated in ECS^+ cases**			
226368_at	Carbohydrate sulfotransferase 11	*CHST11*	2	205100_at	Glutamine-fructose-6-phosphate transaminase 2	*GFPT2*	2
214782_at	Cortactin	*CTTN*	3	204318_s_at	G-2 and S-phase expressed 1	*GTSE1*	3
204318_s_at	G-2 and S-phase expressed 1	*GTSE1*	2	236524_at	β-Amyloid binding protein precursor	*BBP*	2
214375_at	PTPRF-interacting protein, binding protein 1	*PPFIBP1*	2	205290_s_at	Bone morphogenetic protein 2	*BMP2*	2
218414_s_at	Nude, *A. nidulans*, homolog of, 1	*NDE1*	2	226368_at	Carbohydrate sulfotransferase 11	*CHST11*	2
219557_s_at	Nuclear receptor interacting protein 3	*NRIP3*	2	214782_at	Cortactin	*CTTN*	2
223940_x_at	Metastasis-associated lung adenocarcinoma transcript 1	*MALAT1*	2	229579_s_at	Dispatched homolog 2	*DISP2*	3
210235_s_at	Protein tyrosine phosphatase, receptor type, of polypeptide, interacting protein, α1	*PPFIA1*	3	232120_at	Epidermal growth factor receptor	*EGFR*	2
242509_at	MGC17624 protein	*MGC17624*	2	203936_s_at	Matrix metalloproteinase 9	*MMP9*	2
1553185_at	RAS and EF hand domain containing	*RASEF*	3	219557_s_at	Nuclear receptor interacting protein 3	*NRIP3*	2
236947_at	Semaphorin 3C	*SEMA3C*	3	1559400_s_at	Pregnancy-associated plasma protein A, pappalysin 1	*PAPPA*	3
215715_at	Solute carrier family 6, member 2	*SLC6A2*	2	213577_at	Squalene epoxidase	*SQLE*	2
243834_at	Trinucleotide repeat containing 6A	*TNRC6A*	3	243834_at	Trinucleotide repeat containing 6A	*TNRC6A*	3
216450_x_at	Tumor rejection antigen 1	*TRA1*	2	242554_at	Two pore segment channel 2	*TPCN2*	2
229574_at	Transformer-2 α	*TRA2A*	2				
244871_s_at	Ubiquitin-specific protease 32	*USP32*	2				
226176_s_at	Ubiquitin-specific protease 42	*USP42*	2				
Genes downregulated in pN^+ cases				**Genes downregulated in ECS^+ cases**			
208848_at	Alcohol dehydrogenase 5, chi polypeptide	*ADH5*	2	209864_at	Frequently rearranged in advanced T-cell lymphomas 2	*FRAT2*	2
224655_at	Adenylate kinase 3	*AK3*	2	1555419_a_at	*N*-acylsphingosine amidohydrolase 1	*ASAH1*	2
210980_s_at	*N*-acylsphingosine amidohydrolase 1	*ASAH1*	2	213902_at	*N*-acylsphingosine amidohydrolase 1	*ASAH1*	3
209366_x_at	Cytochrome *b*-5	*CYB5*	2	212096_s_at	Mitochondrial tumor suppressor 1	*MTUS1*	2
225864_at	Family with sequence similarity 84B	*FAM84B*	2	226622_at	Mucin 20	*MUC20*	2
211569_s_at	L-3-hydroxyacyl-coenzyme A dehydrogenase, short chain	*HADHSC*	2	202174_s_at	Pericentriolar material 1	*PCM1*	2
1555037_a_at	Isocitrate dehydrogenase 1	*IDH1*	3	210145_at	Phospholipase A2, group IVA	*PLA2G4A*	2
227297_at	Integrin, α9	*ITGA9*	2	215772_x_at	Succinate-coenzyme A ligase, GDP-forming, β subunit	*SUCLG2*	2
212956_at	KIAA0882 protein	*KIAA0882*	2	200883_at	Ubiquinol–cytochrome *c* reductase core protein II	*UQCRC2*	2
222603_at	KIAA1815	*KIAA1815*	2				
204249_s_at	LIM domain only 2	*LMO2*	2				
202174_s_at	Pericentriolar material 1	*PCM1*	2				
212096_s_at	Mitochondrial tumor suppressor 1	*MTUS1*	3				
223130_s_at	Myosin light chain interacting protein	*MYLIP*	2				
209120_at	Nuclear receptor subfamily 2, group F, member 2	*NR2F2*	2				
223315_at	Netrin 4	*NTN4*	2				

*Number of times presented in the dChip-, MAS 5.0–, and RMA-based significant gene lists.

$ECS^{+/-}$ (Figure 2). The hierarchical clustering misclassified one for each of two groups (Figure 2*A*). Apparent separation between ECS^+ and ECS^- groups was observed on MDS (Figure 2*B*).

qRT-PCR was performed to validate the expressional differences of eight candidate genes in the cases we used for global expression profiling study. As shown in Table 4, differences in RNA levels are either statistically significant ($P < .05$) or suggestive ($P < .1$) for six of eight genes tested for pN^- *versus* pN^+ (including *BMP2*, *CTTN*, *EEF1A1*, *GTSE1*, *MMP9*, and *EGFR*). For the ECS^- and ECS^+ groups, five of eight genes tested show either statistically significant ($P < .05$) or suggestive ($P < .1$) differences in RNA levels (including *BMP2*, *CTTN*, *EEF1A1*, *MMP9*, and *EGFR*).

To test the prediction/classification power of these markers, a logistic model and ROC curve analysis were performed based on qRT-PCR results (Table 5). *CTTN* and *MMP9* show the best prediction powers (0.94 and 0.88 AUC, respectively, for pN; 0.88 and 0.95 AUC, respectively, for ECS). Specific combinations of markers (*CTTN*+*MMP9*+*EGFR* for $pN^{+/-}$; *CTTN*+*EEF1A1*+*MMP9* for $ECS^{+/-}$) that fit the data best achieved perfect specificity and sensitivity. As evaluated by leave-one-out cross-validation, 85% overall prediction accuracy rates were achieved for both specific marker combinations for $pN^{+/-}$ and $ECS^{+/-}$. Our results demonstrated that pN metastasis and ECS can be predicted by gene expression analyses. This provides a foundation for further validation of the prediction/classification

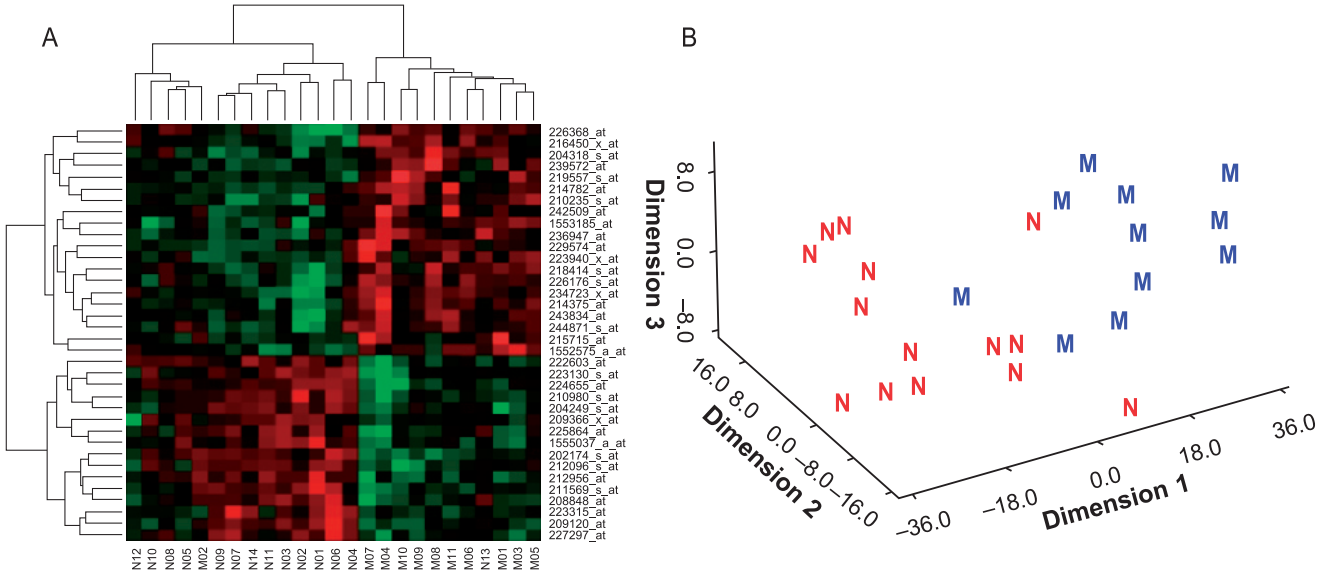

Figure 1. *Classification of OTSCC lymph node metastasis using global gene expression analysis. Global gene expression profiling on 11 lymph node metastasis–positive (pN^+) and 14 lymph node metastasis–negative (pN^-) SCC samples of the tongue was carried out using U133+2.0 array. A signature gene set of 33 genes was created based on the integration of three microarray methods (dChip, MAS 5.0, and RMA), as described in the Materials and Methods section. Hierarchical clustering (A) and MDS (B) were performed based on this signature gene set. The metastasis group (labeled M; n = 11) and the nonmetastasis group (labeled N; n = 14).*

power of identified molecular markers in a large independent sample set.

Discussion

Lymph Node Metastasis and ECS of HNSCCs

ECS of lymph node metastasis is one of the most important negative prognostic factors of HNSCCs [14,15]. However, the underlying biology determining these aggressive features is largely unknown. Improvement in patient survival requires an increased understanding of tumor metastasis so that aggressive tumors can be detected early in the disease process and targeted therapeutic interventions can be developed. We used a global approach to uncover gene expression signatures associated with these aggressive features. Because our method covers most of the human transcriptome, the differentially expressed genes consistently seen in tumors with lymph node metastasis and ECS of the node are likely to be biologically important. In this study, we identified

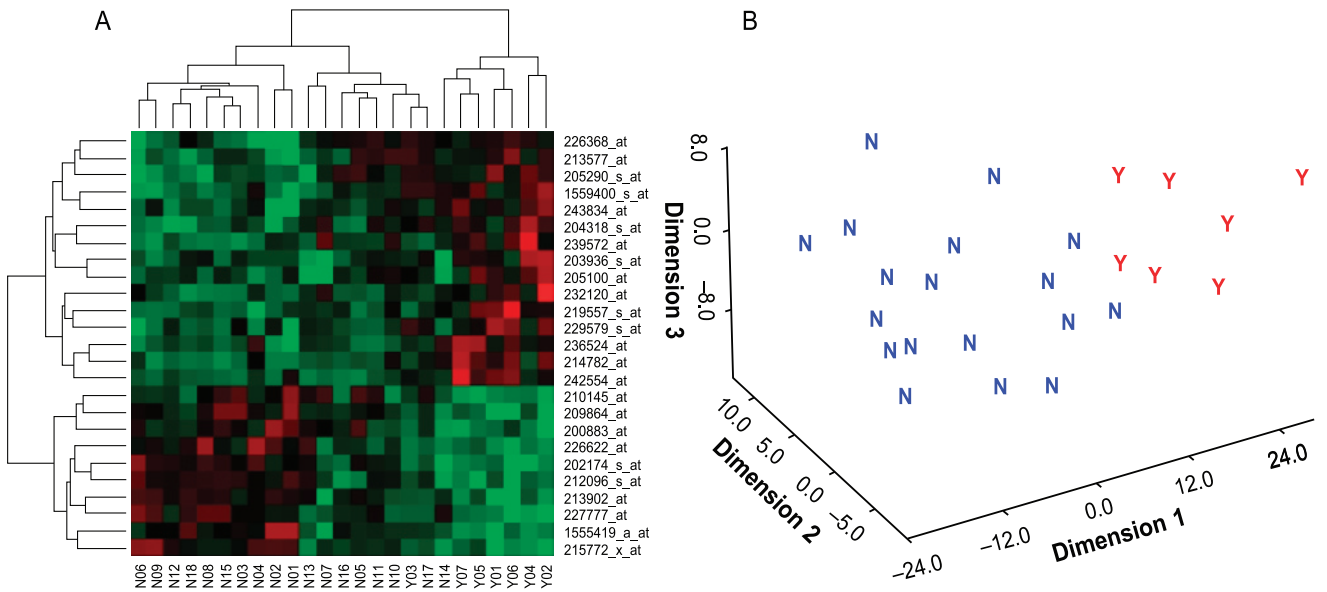

Figure 2. *Classification of OTSCC node ECS using global gene expression analysis. A second signature gene set was created for the ECS^+ group (labeled Y; n = 7) and the ECS^- group (labeled N; n = 18). Hierarchical clustering (A) and MDS (B) were performed based on this signature gene set.*

Table 4. Real-Time qRT-PCR Validation of Selected Candidate Genes.

	pN^{-}*	pN^{+}**	Wilcoxon *P*	ECS^{-}*	ECS^{+}**	Wilcoxon *P*
BMP2	1.19 (0.42–3.45)	4.76 (4.1–9.57)	.043	1.57 (0.71–4.59)	7.21 (4.68–10.56)	.069
CTTN	1.52 (1.22–3.45)	7.21 (5.58–13.49)	<.0001	2.93 (1.37–4.95)	7.21 (6.15–21.73)	.004
EEF1A1	0.56 (0.26–1.07)	0.33 (0.21–0.41)	.078	0.5 (0.26–0.89)	0.26 (0.16–0.34)	.064
ASAH1	1.04 (0.45–1.35)	0.27 (0.14–1.2)	.259	0.97 (0.28–1.35)	0.26 (0.12–0.87)	.209
MTUS1	0.81 (0.25–1.77)	0.42 (0.09–0.66)	.364	0.76 (0.25–1.73)	0.42 (0.09–0.66)	.544
GTSE1	1.68 (0.60–2.51)	3.48 (2.48–7.17)	.013	2.3 (0.77–3.2)	3.48 (2.22–6.73)	.107
MMP9	2.38 (1.48–3.43)	13.0 (5.67–16.30)	.001	2.46 (1.52–4.26)	13.93 (9.64–38.43)	.001
EGFR	0.71 (0.41–1.72)	3.03 (1.91–5.21)	.004	1.11 (0.45–2.69)	3.03 (1.93–7.09)	.055

*The median value (25th to 75th percentiles) of the RNA level was computed with the $2^{-\Delta\Delta C_t}$ method, as described previously [26], where *β-actin* was used as reference gene.

CTTN and *MMP9* as consistently overexpressed genes in tumors with lymph node metastasis that best predict lymph node metastasis and ECS in our models.

It is important to note that prediction works better when expressions of several candidate genes were considered at the same time, suggesting a compounded effect of the genes on HNSCC progression. Although the precise role of classification models in predicting patients' clinical outcomes needs to be determined in future studies, our results suggest that these genes may be therapeutic targets for patients with HNSCC. More importantly, it suggests that targeting a combination of several of these genes may achieve therapeutic synergy to treat or prevent tumor cell spread and metastasis.

Rationale of Focusing on Tongue Cancer

It has been realized that head and neck cancers are groups of diverse cancers that develop from many different anatomic sites and are associated with different risk factors [29] and genetic characteristics [30]. In this study, we focused on OTSCC—one of the most common sites for HNSCC. OTSCC is significantly more aggressive than other forms of HNSCCs, with a propensity for rapid local invasion and spread [31]. It is possible that signature gene sets for tumor metastases of HNSCC originating from different anatomic sites maybe different. More studies will be needed to address this important question.

Discussion of Validated Candidate Genes

A set of candidate genes, including genes that are known to be involved in metastasis, such as *MMP9* and *CTTN*, has been identified and validated in this study. *MMP9* is a member of a group of secreted zinc metalloproteases, which, in mammals, degrade the collagen of the extracellular matrix. An elevated expression of *MMP9* has been linked to metastasis in many different cancer types [32,33]. *CTTN* has been shown to be an oncogene residing in the 11q13 region that is frequently amplified in HNSCC and breast cancer [34,35]. Interestingly, cyclin D1, which also resides in the 11q13 and is often coupregulated with *CTTN* [36], does not appear to be one of the most predictive markers for either pN metastasis or ECS. Our signature gene set also contains several candidate genes that are involved in tumorigenesis, such as *BMP2* and *EGFR*. *BMP2* is a member of the transforming growth factor-β superfamily, which controls proliferation, differentiation, and other functions in many cell types. *EGFR* is one of the most frequently amplified and mutated genes in many different type of cancers, including HNSCC [37,38]. Other identified candidate genes, whose roles in the metastatic process have not been clearly defined, include *GTSE1* and *EEF1A1*. *GTSE1* is a microtubule-localized protein. Its expression is cell cycle–regulated and can induce G2/M–phase accumulation when overexpressed [39]. It has been demonstrated that *GTSE1* downregulates the levels and activity of the p53

Table 5. Classification Power for Selected Candidate Genes Based on Real-Time qRT-PCR.

	Prediction Power for pN^{-}/pN^{+}				Prediction Power for ECS^{-}/ECS^{+}			
	Wilcoxon *P*	ROC (AUC)	Sensitivity	Specificity	Wilcoxon *P*	ROC (AUC)	Sensitivity	Specificity
BMP2	.043	74	82	73	.069	74	86	68
CTTN	<.0001	94	91	80	.004	88	86	74
EEF1A1	.078	71	55	33	.064	74	57	26
ASAH1	.259	64	27	73	.209	67	29	74
MTUS1	.364	61	64	33	.544	58	57	37
GTSE1	.013	79	73	80	.107	71	71	74
MMP9	.001	88	82	93	.001	95	100	89
EGFR	.004	82	73	80	.055	75	71	79
	The best model*							
	Model for pN^{-}/pN^{+}	Sensitivity	Specificity	Overall accuracy rate (%)	Model for ECS^{-}/ECS^{+}	Sensitivity	Specificity	Overall accuracy rate (%)
	CTTN+MMP9+EGFR	100	100	84.6**	*CTTN+EEF1A1+MMP9*	100	100	84.6**

*The best model is generated based on stepwise logistic model selection.
**The overall accuracy rate is estimated by the leave-one-out cross validation approach.

tumor-suppressor protein and represses its ability to induce apoptosis after DNA damage [40]. The *EEF1A1* gene codes for the α subunit of elongation factor-1, which is involved in the binding of aminoacyl-tRNA to 80*S* ribosome. The involvement of this gene with tumorigenesis is not clear.

Rationale of Using Multiple Microarray Analysis Methods for Identification of Candidate Genes

Several approaches for defining a measure of expression representing the amount of corresponding mRNA for each microarray probe set have been proposed. These include the Affymetrix MAS 5.0 (which is now implemented in GCOS) [22], the MBEI from Li and Wong [23] (implemented in dChip), and the RMA from Irizarry et al. [24] (implemented in R). Each approach uses a different data preparatory procedure for normalization and background correction, and uses its own model to define the expression index. Consequently, expression indices will differ when applied to the same probe set data. These differences reflect varying biologic attributes that each mathematical model highlights. Previous studies (David Elashoff, Myungshin Oh, Nik Brown, Yang Li, David T. Wong, Steve Horvath, ''Empirical Study of the Influence of Expression Index on the Standard Statistical Analysis of Oligonucleotide Microarray Data,'' manuscript in preparation), as well as our own results here, have found a low level of overlap between gene lists produced by various methods, suggesting that a combination of methods can be used in the statistical analysis of oligonucleotide microarray data. This may be a reflection of the high false-positive rate in microarray-based studies. A stratagem will be needed to reduce these false-positives. In addition, it is important to narrow down the candidate gene list for future validation and eventual implementation of these biomarkers in a clinical setting.

In this study, to narrow down candidate genes, a final signature gene set was compiled by selecting genes present in at least two of three initial candidate gene lists generated using dChip, MAS 5.0, and RMA. The rationale is that using a combination of several methods/gene lists will increase the statistical power to identify true biomarkers, reduce the chances of false-positives, and improve reproducibility in downstream validation experiments.

Requirement for Future Validation Studies with Independent Sample Sets

Our data (ROC analyses) showed the potential of using these candidate genes as biomarkers for the prediction/classification of lymph node metastasis and the ECS of OTSCC. Specific combinations of several markers provide further enhancement of prediction/classification power. Interestingly, several markers (*CTTN* and *MMP9*) provide superb classification power for both pN metastasis and ECS. This suggests that the same (or similar) functional/biologic events account for both pN metastasis and ECS phenotypes, and these phenotypes are a reflection of continuous progression of metastasis potential from pN metastasis to ECS. The specific thresholds for these markers and the roles used to combine these markers will most likely be different for the two distinct phenotypes. Additional studies on a larger sample sets are needed.

In summary, our results demonstrate the feasibility of using biomarkers discovered by global expression profiling analyses as potential biomarkers for the prediction/classification of OTSCC metastasis and ECS. These results, together with those from recent studies, will set the stage for translating this molecular genetic-based finding into clinically valid markers for the prediction and classification of HNSCC metastasis, pending future validation of these markers with large independent sample sets.

Acknowledgement

Microarray hybridization and scanning were performed at the University of California at Los Angeles DNA microarray facility. We thank Francois Janot at Institut Gustave-Roussy for providing the samples.

References

[1] Pantel K and Brakenhoff RH (2004). Dissecting the metastatic cascade. *Nat Rev Cancer* **4** (6), 448–456.
[2] Woolgar JA (1999). Pathology of the N_0 neck. *Br J Oral Maxillofac Surg* **37** (3), 205–209.
[3] Jones AS, Phillips DE, Helliwell TR, and Roland NJ (1993). Occult node metastases in head and neck squamous carcinoma. *Eur Arch Otorhinolaryngol* **250** (8), 446–449.
[4] Pillsbury HC III and Clark M (1997). A rationale for therapy of the N_0 neck. *Laryngoscope* **107** (10), 1294–1315.
[5] Robbins KT, Clayman G, Levine PA, Medina J, Sessions R, Shaha A, Som P, and Wolf GT (2002). Neck dissection classification update: revisions proposed by the American Head and Neck Society and the American Academy of Otolaryngology—Head and Neck Surgery. *Arch Otolaryngol Head Neck Surg* **128** (7), 751–758.
[6] Short SO, Kaplan JN, Laramore GE, and Cummings CW (1984). Shoulder pain and function after neck dissection with or without preservation of the spinal accessory nerve. *Am J Surg* **148** (4), 478–482.
[7] van Wilgen CP, Dijkstra PU, van der Laan BF, Plukker JT, and Roodenburg JL (2004). Shoulder and neck morbidity in quality of life after surgery for head and neck cancer. *Head Neck* **26** (10), 839–844.
[8] van Wilgen CP, Dijkstra PU, van der Laan BF, Plukker JT, and Roodenburg JL (2004). Morbidity of the neck after head and neck cancer therapy. *Head Neck* **26** (9), 785–791.
[9] van Wilgen CP, Dijkstra PU, Nauta JM, Vermey A, and Roodenburg JL (2003). Shoulder pain and disability in daily life, following supraomohyoid neck dissection: a pilot study. *J Craniomaxillofac Surg* **31** (3), 183–186.
[10] Nakamura K, Ozaki N, Yamada T, Hata T, Sugimoto S, Hikino H, Kanazawa A, Tokuka A, and Nagaoka S (2005). Evaluation of prognostic significance in extracapsular spread of lymph node metastasis in patients with gastric cancer. *Surgery* **137** (5), 511–517.
[11] Yamashita H, Noguchi S, Murakami N, Kawamoto H, and Watanabe S (1997). Extracapsular invasion of lymph node metastasis is an indicator of distant metastasis and poor prognosis in patients with thyroid papillary carcinoma. *Cancer* **80** (12), 2268–2272.
[12] van der Velden J, van Lindert AC, Lammes FB, ten Kate FJ, Sie-Go DM, Oosting H, and Heintz AP (1995). Extracapsular growth of lymph node metastases in squamous cell carcinoma of the vulva. The impact on recurrence and survival. *Cancer* **75** (12), 2885–2890.
[13] Mambo NC and Gallager HS (1977). Carcinoma of the breast: the prognostic significance of extranodal extension of axillary disease. *Cancer* **39** (5), 2280–2285.
[14] Greenberg JS, Fowler R, Gomez J, Mo V, Roberts D, El Naggar AK, and Myers JN (2003). Extent of extracapsular spread: a critical prognosticator in oral tongue cancer. *Cancer* **97** (6), 1464–1470.
[15] Myers JN, Greenberg JS, Mo V, and Roberts D (2001). Extracapsular spread. A significant predictor of treatment failure in patients with squamous cell carcinoma of the tongue. *Cancer* **92** (12), 3030–3036.

[16] Ramaswamy S, Ross KN, Lander ES, and Golub TR (2003). A molecular signature of metastasis in primary solid tumors. *Nat Genet* **33** (1), 49–54.

[17] van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415** (6871), 530–536.

[18] Schmalbach CE, Chepeha DB, Giordano TJ, Rubin MA, Teknos TN, Bradford CR, Wolf GT, Kuick R, Misek DE, Trask DK, et al. (2004). Molecular profiling and the identification of genes associated with metastatic oral cavity/pharynx squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg* **130** (3), 295–302.

[19] O'Donnell RK, Kupferman M, Wei SJ, Singhal S, Weber R, O'Malley B, Cheng Y, Putt M, Feldman M, Ziober B, et al. (2005). Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene* **24** (7), 1244–1251.

[20] Roepman P, Wessels LF, Kettelarij N, Kemmeren P, Miles AJ, Lijnzaad P, Tilanus MG, Koole R, Hordijk GJ, van der Vliet PC, et al. (2005). An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet* **37** (2), 182–186.

[21] Ohyama H, Zhang X, Kohno Y, Alevizos I, Posner M, Wong DT, and Todd R (2000). Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. *Biotechniques* **29** (3), 530–536.

[22] Affymetrix (2002). Statistical Algorithms Description Document. Affymetrix, Santa Clara, CA, USA.

[23] Li C and Wong WH (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci USA* **98** (1), 31–36.

[24] Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, and Speed TP (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31** (4), e15.

[25] Li Y, St John MA, Zhou X, Kim Y, Sinha U, Jordan RC, Eisele D, Abemayor E, Elashoff D, Park NH, et al. (2004). Salivary transcriptome diagnostics for oral cancer detection. *Clin Cancer Res* **10** (24), 8442–8450.

[26] Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. *Methods* **25** (4), 402–408.

[27] Renger R and Meadows LM (1994). Use of stepwise regression in medical education research. *Acad Med* **69** (9), 738.

[28] Li Y, Elashoff D, Oh M, Sinha U, St John MA, Zhou X, Abemayor E, and Wong DT (2006). Serum circulating human mRNA profiling and its utility for oral cancer detection. *J Clin Oncol* **24** (11), 1754–1760.

[29] Dobrossy L (2005). Epidemiology of head and neck cancer: magnitude of the problem. *Cancer Metastasis Rev* **24** (1), 9–17.

[30] Timar J, Csuka O, Remenar E, Repassy G, and Kasler M (2005). Progression of head and neck squamous cell cancer. *Cancer Metastasis Rev* **24** (1), 107–127.

[31] Franceschi D, Gupta R, Spiro RH, and Shah JP (1993). Improved survival in the treatment of squamous carcinoma of the oral tongue. *Am J Surg* **166** (4), 360–365.

[32] Osman M, Tortorella M, Londei M, and Quaratino S (2002). Expression of matrix metalloproteinases and tissue inhibitors of metalloproteinases define the migratory characteristics of human monocyte–derived dendritic cells. *Immunology* **105** (1), 73–82.

[33] Turner HE, Nagy Z, Esiri MM, Harris AL, and Wass JA (2000). Role of matrix metalloproteinase 9 in pituitary tumor behavior. *J Clin Endocrinol Metab* **85** (8), 2931–2935.

[34] Schuuring E, Verhoeven E, Mooi WJ, and Michalides RJ (1992). Identification and cloning of two overexpressed genes, *U21B31/PRAD1* and *EMS1*, within the amplified chromosome 11q13 region in human carcinomas. *Oncogene* **7** (2), 355–361.

[35] Schuuring E, van Damme H, Schuuring-Scholtes E, Verhoeven E, Michalides R, Geelen E, de Boer C, Brok H, van Buuren V, and Kluin P (1998). Characterization of the *EMS1* gene and its product, human Cortactin. *Cell Adhes Commun* **6** (2–3), 185–209.

[36] Schuuring E (1995). The involvement of the chromosome 11q13 region in human malignancies: *cyclin D1* and *EMS1* are two new candidate oncogenes—a review. *Gene* **159** (1), 83–96.

[37] Grandis JR and Tweardy DJ (1993). Elevated levels of transforming growth factor a and epidermal growth factor receptor messenger RNA are early markers of carcinogenesis in head and neck cancer. *Cancer Res* **53,** 3579–3584.

[38] Dassonville O, Formento JL, Francoual M, Santini J, Schneider M, Demard F, and Milano G (1993). Expression of epidermal growth factor receptor and survival in upper aerodigestive tract cancer. *J Clin Oncol* **11,** 1873–1878.

[39] Monte M, Collavin L, Lazarevic D, Utrera R, Dragani TA, and Schneider C (2000). Cloning, chromosome mapping and functional characterization of a human homologue of murine *gtse-1* (B99) gene. *Gene* **254** (1–2), 229–236.

[40] Monte M, Benetti R, Collavin L, Marchionni L, Del Sal G, and Schneider C (2004). hGTSE-1 expression stimulates cytoplasmic localization of p53. *J Biol Chem* **279** (12), 11744–11752.