

Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity

Rebekah K O'Donnell^{1,7}, Michael Kupferman^{2,7}, S Jack Wei³, Sunil Singhal⁵, Randal Weber², Bert O'Malley Jr², Yi Cheng¹, Mary Putt⁶, Michael Feldman⁴, Barry Ziober² and Ruth J Muschel^{*,1}

¹Department of Pathology, Children's Hospital of Philadelphia, Hospital of the University of Pennsylvania, Philadelphia, PA, USA; ²Department of Otorhinolaryngology–Head and Neck Surgery, Hospital of the University of Pennsylvania, Philadelphia, PA, USA; ³Department of Radiation Oncology, Hospital of the University of Pennsylvania, Philadelphia, PA, USA; ⁴Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA, USA; ⁵Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA; ⁶Department of Biostatistics and Epidemiology, University of Pennsylvania, PA, USA

Metastasis via the lymphatics is a major risk factor in squamous cell carcinoma of the oral cavity (OSCC). We sought to determine whether the presence of metastasis in the regional lymph node could be predicted by a gene expression signature of the primary tumor. A total of 18 OSCCs were characterized for gene expression by hybridizing RNA to Affymetrix U133A gene chips. Genes with differential expression were identified using a permutation technique and verified by quantitative RT-PCR and immunohistochemistry. A predictive rule was built using a support vector machine, and the accuracy of the rule was evaluated using crossvalidation on the original data set and prediction of an independent set of four patients. Metastatic primary tumors could be differentiated from nonmetastatic primary tumors by a signature gene set of 116 genes. This signature gene set correctly predicted the four independent patients as well as associating five lymph node metastases from the original patient set with the metastatic primary tumor group. We concluded that lymph node metastasis could be predicted by gene expression profiles of primary oral cavity squamous cell carcinomas. The presence of a gene expression signature for lymph node metastasis indicates that clinical testing to assess risk for lymph node metastasis should be possible.

Oncogene (2005) **24**, 1244–1251. doi:10.1038/sj.onc.1208285
Published online 22 November 2004

Keywords: lymphatic metastasis; squamous cell carcinoma; microarray; gene expression pattern

Introduction

Metastasis, the dissemination of tumor cells that colonize new areas of the body, may progress via the

bloodstream or the lymphatics. Head and neck squamous cell carcinoma (HNSCC) characteristically metastasizes to the regional lymph nodes through the draining lymphatics. We focused on HNSCC originating in the oral cavity (OSCC). Typically, 50% of patients with OSCC have detectable lymph node involvement at presentation. Metastasis to distant sites is relatively uncommon, occurring in less than 5% at presentation. Patients with lymph node metastasis have a markedly worse prognosis than patients without metastasis. Only 25–40% of patients with lymph node metastasis at presentation will achieve 5-year survival, compared to approximately 90% of patients without metastasis (Hong and Weber, 1995; Greenberg *et al.*, 2003). Some of the patients without lymph node metastasis at presentation will subsequently manifest metastasis. Since removal of lymph nodes and/or radiation and chemotherapy can reduce emergence of occult metastasis, attempts to predict the patients in this category are made clinically (Kim *et al.*, 1993; Myers *et al.*, 1998). Currently, node-negative patients estimated to have a 20% or greater risk of metastasis often have surgical removal of draining lymph nodes and radiation. The ability to better predict lymph node metastasis could allow therapy better tailored to each patient.

The pioneering studies of metastasis by Fidler and Kripke (1977) established that a primary tumor is composed of cells with widely differing metastatic potentials. This concept of tumor heterogeneity might predict that studying the bulk tumor would not be useful in predicting metastasis, since a large number of poorly metastatic cells might obscure the properties of a small number of highly metastatic cells (Poste and Fidler, 1980). However, gene signatures have been identified in several tumor types which correlate primary tumor gene expression with increased risk of metastasis (van't Veer *et al.*, 2002; Kikuchi *et al.*, 2003; Ramaswamy *et al.*, 2003; Weiss *et al.*, 2003; Bertucci *et al.*, 2004; Nakamura *et al.*, 2004). These data are part of an increasing body of work indicating that the genetic signature of the majority of cells in a primary tumor holds significant

*Correspondence: RJ Muschel; E-mail: muschel@xrt.upenn.edu

⁷These authors contributed equally to this work

Received 4 June 2004; revised 11 October 2004; accepted 11 October 2004; published online 22 November 2004

		Initial Set, N+ (n=11)	Initial Set, N- (n=7)	Validation Set (n=4)	Histochemistry Only Set (n=3)
Age	Median	69	59	53	55
	(range)	(42-80)	(47-83)	(45-63)	(43-82)
	Mean	63	58	54	60
Gender	Male (%)	64	100	100	33
	Female (%)	36	0	0	67
Ethnicity	White (%)	73	100	50	67
	Black (%)	18	0	0	33
	Others (%)	9	0	50	0
Tobacco Use	Yes (%)	73	71	100	67
	No (%)	27	29	0	33
Alcohol Use	Heavy (%)	36	29	50	33
	Social (%)	45	43	25	33
	None (%)	18	29	25	33
Tumor Site	FOM/Buccal/Tonsil (%)	9	43	0	33
	Gingiva (%)	9	0	0	0
	Larynx (%)	0	0	25	0
	Mandible (%)	18	0	0	0
	Tongue (%)	64	57	75	67
Pathological T stage	1 (%)	9	14	50	67
	2 (%)	57	29	0	0
	3 (%)	18	14	25	33
	4 (%)	64	43	25	0
Pathological N stage	1 (%)	18	0	0	0
	2 (%)	82	0	75	100
Grade	1 (%)	9	0	0	0
	1-2 (%)	9	0	0	0
	2 (%)	36	71	25	33
	2-3 (%)	9	29	75	0
	3 (%)	36	0	0	67

Figure 1 Clinical features of patients in study. A total of 18 were included in the original test group, four patients compose the microarray validation set, and three compose the supplemental immunohistochemistry set

value. Differences in gene expression between metastatic and nonmetastatic primary tumors may expose patterns of metastatic potential and could serve as a basis for clinical testing for metastatic potential at biopsy.

Metastasis research has focused largely on hematogenous metastasis although spread through the lymphatics is common in many cancers. Recent data indicate that the processes of hematogenous and lymphatic metastasis are markedly different (He *et al.*, 2002). Studies in breast, lung, colon, and pancreatic cancer have begun to address the issue of lymphatic metastasis, but these diseases are also prone to the confounding factor of distant metastasis (West *et al.*, 2001; Kikuchi *et al.*, 2003; Bertucci *et al.*, 2004; Nakamura *et al.*, 2004). HNSCC may be a useful system for studying lymphatic metastasis because of the high probability of lymphatic spread and low probability of hematogenous spread. A recent study has identified a gene signature for recurrent disease in HNSCC and several studies have investigated the changes in gene expression from normal tissue to carcinoma, but studies involving lymphatic metastasis have been limited by the small number of genes assessed and/or lack of a rigorous efficacy test of the metastatic signature (Alevizos *et al.*, 2001; Belbin *et al.*, 2002; Mendez *et al.*, 2002; Hwang *et al.*, 2003; Gonzalez *et al.*, 2003; Leethanakul *et al.*, 2003; Nagata *et al.*, 2003; Chung *et al.*, 2004; Ginos *et al.*, 2004; Schmalbach *et al.*, 2004; Warner *et al.*, 2004).

We designed our experiment to investigate whether differences in gene expression between N+ and N- primary tumor groups could be used to predict the N status of an independent patient set, and whether such a discriminatory gene signature would be similar to those predicting hematogenous metastasis.

Results

We compared the gene expression profiles obtained from primary squamous cell carcinomas of the oral cavity (OSCC) that were metastatic to lymph nodes (N+) to those that were not (N-). Tumor samples were collected from 18 patients undergoing surgical treatment for OSCC. Samples from lymph node metastases were obtained from five patients. An additional four primary tumor samples were obtained independently for use as a microarray test set, and three primary and two lymph node samples were obtained for immunohistochemical validation. The clinical characteristics of these patients are outlined in Figure 1. Total RNA was isolated from tissue samples shown by frozen section to be devoid of normal tissue. Probes generated from this RNA were hybridized to Affymetrix U133A genechips.

Application of the permutation-based method Significance Analysis of Microarrays (SAM) to the signal data identified 147 differentially expressed transcripts with an estimated median false discovery rate of 6.1%. The differences in transcript levels found by hybridization to the Affymetrix chip were confirmed by quantitative real-time RT-PCR. We compared the average values of the N+ patients to the average values of the N- patients for four genes: hypothetical protein MGC3731 and keratin 13 from the list of genes generated by SAM, and BAG1 and RPB6, whose Affymetrix signals were significantly different between N+ and N- using a two-sided *t*-test but which were not present in the SAM-generated list. The direction of change for all four genes using quantitative real-time PCR corresponded with the direction of change from the microarrays (data not shown).

The set of genes selected by SAM was modified by removing genes whose expression levels were below background in both N+ and N- groups. The resulting signature set of 116 genes (Figure 2b and Supplementary Figure 1) was evaluated for its ability to discriminate between metastatic and nonmetastatic primary tumors. As expected, hierarchical clustering using the signature set showed that all metastatic primary tumors clustered together, and all nonmetastatic primary tumors with one exception (patient 16, which exhibits features of both groups) clustered together (Figure 2a).

The 116-gene signature set was also used to compare N+ to N- primary tumors using two-dimensional principal-components analysis. The principal-components analysis reduces the information in the individual genes into linear combinations of the gene transcript signals and is commonly used to visualize differences among groups of samples (Butte, 2002). This analysis readily distinguished the primary tumors with metastasis from those without, again with patient 16 positioned between the two groups. The signature gene set was compared with gene sets of an equal number of randomly selected genes that failed to discriminate between the tumor groups (Figure 3).

We hypothesized that lymph node metastases from the metastatic patients would associate with the N+

group, since the metastases were derived from N+ tumors. Using principal-components analysis with the signature gene set, all lymph node metastases clearly clustered with the N+ group (Figure 4).

We tested the ability of the principal-components analysis to predict lymph node metastasis by evaluating an independent test set of four patients. Three patients presented with OSCC (two metastatic, one nonmetastatic), and one presented with metastatic cancer of the larynx. Using principal-components analysis, all metastatic tumors from the test set clearly associated with the N+ from the initial data set, but the nonmetastatic tumor from the test set was not clearly classified with either group (Figure 4).

Unlike principal-components analysis, which reduces the dimensionality of the data, support vector machines (SVMs) use kernel functions to increase the number of dimensions in order to better separate two groups of data (Butte, 2002). The machine searches for a hyperplane in multidimensional space that is maximally distant from both groups, then reclassifies all samples based upon their orientation to the hyperplane. The vectors, which map the position of each sample in multidimensional space, are summed to yield the discriminant score, which is used to classify the sample into one of two groups. The leave-one-out

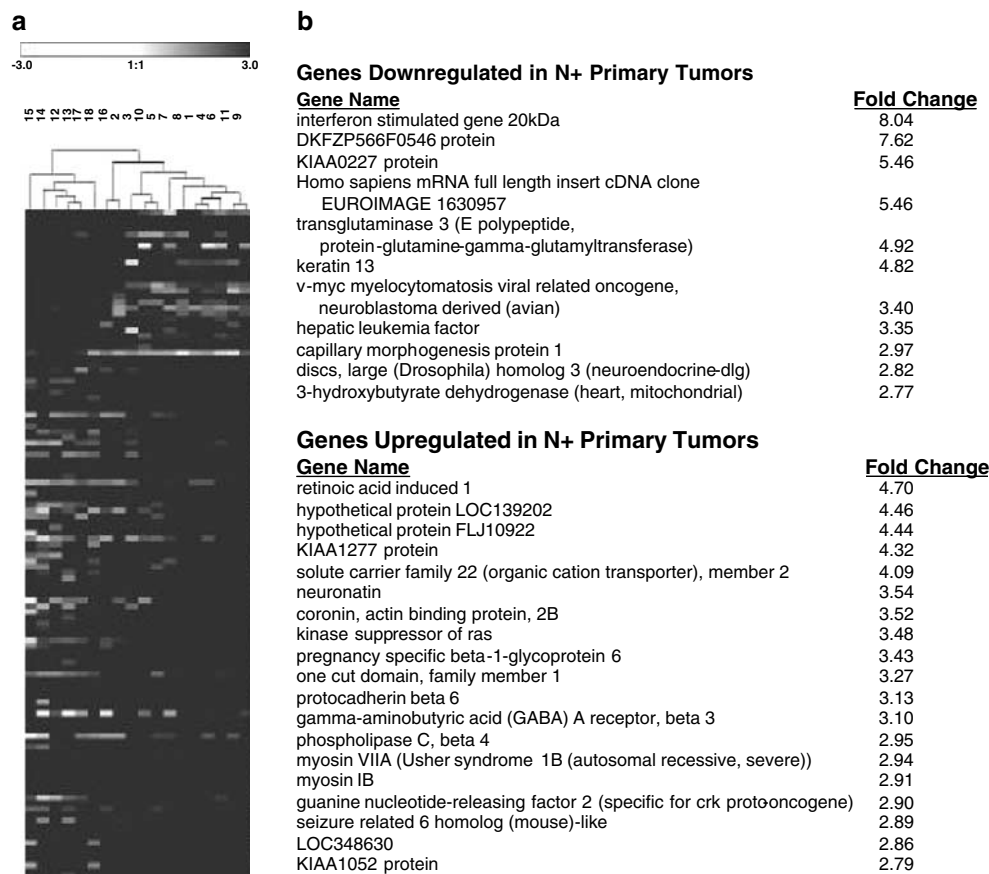


Figure 2 Signature gene set selected by SAM. (a) Hierarchical clustering of primary tumors using the 116-gene signature set classified the tumors into two groups (N-, left; N+, right). Patient 16, an N-, shares features of both groups. (b) List of genes with the highest fold-change values from the 116-gene signature set (for full list, see Supplementary Figure 1)

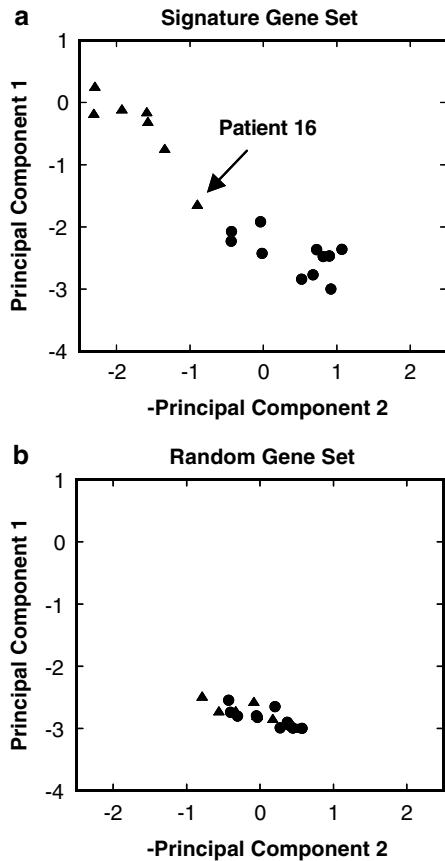


Figure 3 Two-dimensional principal-components analysis of original patient group using signature set or random gene set. (a) Principal-components analysis using the 116-gene signature set separated N+ (●) from N- (▲) patients, with patient 16 (N-) not clearly classified into either group. (b) Principal-components analysis using a representative random set of genes did not separate the groups

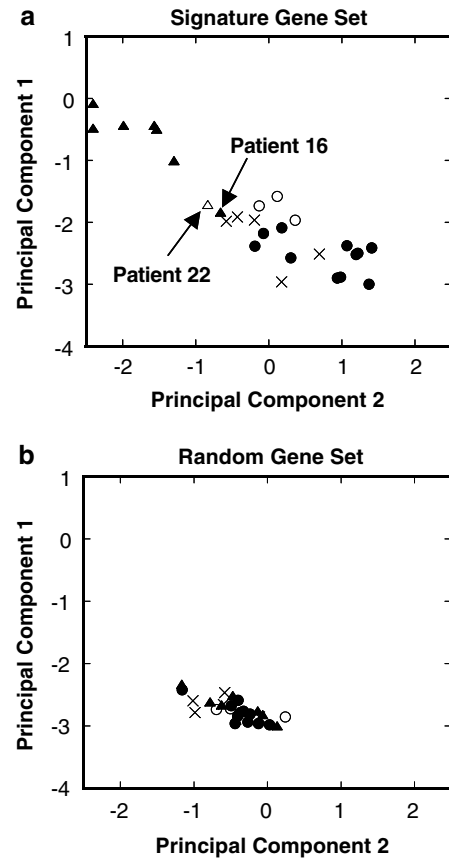


Figure 4 Two-dimensional principal-components analysis classifies lymph node metastases and independent validation set. (a) Principal-components analysis using the signature gene set clustered the lymph node metastases (X) and validation set N+ (○) with the original N+ (●). The validation set N- (△, patient 22) is intermediate between the groups, but closer to the original N- (▲) than patient 16 (N-). (b) Principal-components analysis using a representative random gene set did not separate the groups

analysis employs multiple permutations of the data to determine the hyperplane, at each iteration removing a different training (original patient set) sample. Using this approach, all lymph node metastases were classified with the N+ group, all patients from the test set were classified with the appropriate groups, and patient 16, who had not been clearly classified using hierarchical clustering or principal-components analysis, was classified correctly (Figure 5b). Because we used the 116-gene set chosen by SAM as our input to the SVM, our classification using the leave-one-out approach may be overly optimistic. To ensure that our results were not an artifact of the analysis, we used the SVM to predict the tumor classes using sets of an equal number of randomly chosen, above-threshold genes. These random gene sets were consistently unable to discriminate between N+ and N- groups (Figure 5a).

None of the genes in our signature set have been previously associated with lymph node metastasis in HNSCC, although different experimental methods have implicated some in squamous cell cancer (TDRD1 (Loriot *et al.*, 2003), transglutaminase 3 (Chen *et al.*,

2000), ERBB4 (Bei *et al.*, 2001), keratin 13 (Depondt *et al.*, 1999)) and correlated others with aggressiveness and invasiveness of a diverse group of primary tumors (ERBB4 and keratin 13, OSCC (Depondt *et al.*, 1999; Bei *et al.*, 2001); KLF12, melanoma (Karjalainen *et al.*, 1998; Roth *et al.*, 2000); ATP6V1C1, pancreatic (Ohta *et al.*, 1996); KCNJ5, breast (Kennedy *et al.*, 1999; Stringer *et al.*, 2001)). Unexpectedly CXCR4, which has been previously correlated with metastasis in HNSCC as well as other tumor types, did not emerge in our list of differentially expressed genes (Uchida *et al.*, 2003; Delilbasi *et al.*, 2004). To evaluate CXCR4 expression in these patients, we examined RNA and protein levels by quantitative RT-PCR and immunohistochemistry, but found no significant differences in CXCR4 cytoplasmic or nuclear expression among N+, N-, or lymph node metastases (Figure 6a-c), consistent with the Affymetrix data. In contrast with CXCR4, immunohistochemistry revealed the differential expression of Col2 between N+ and N- patient groups, consistent with its differential expression at the RNA level (Figure 6d).

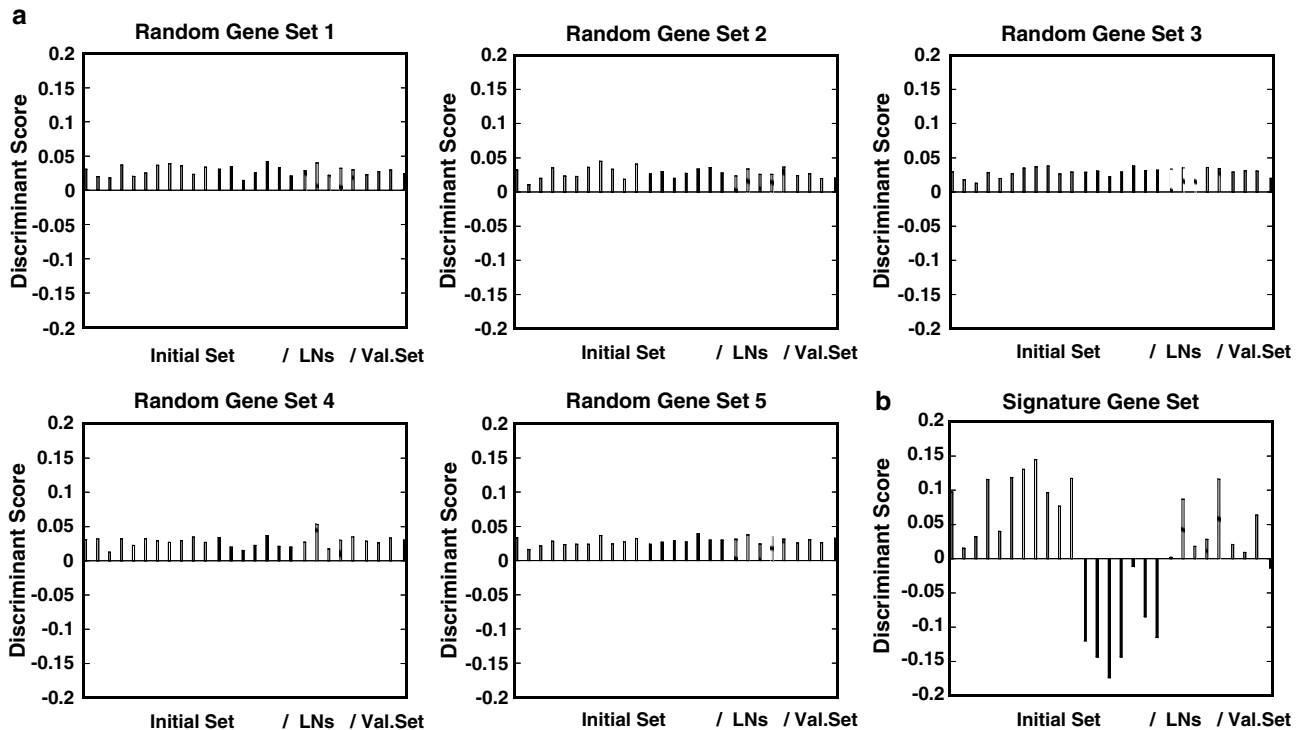


Figure 5 Prediction of lymph node metastases and independent validation set using SVM. (a) A representative sample of five random gene sets consistently classified all samples as a single group. (b) The signature gene set correctly discriminated between N+ (□) and N- (■) for 100% of patients in the original set (crossvalidation) and 100% of the validation set and lymph node metastases (⊗)

Discussion

These results demonstrate that measurable differences in gene expression exist between N+ and N- primary tumors and that these differences are sufficient to predict the N status of an independent set of patients using a leave-one-out crossvalidation approach. Despite the relatively small sample size, significance was high. The relevance of our results is supported by the correct prediction of an independent test set. Although the test set is small, it is extremely important since independent validation is very rare in microarray studies (Ntzani and Ioannidis, 2003). These results are consistent with the ability of gene signatures developed from studies in hematogenous metastasis systems to predict overall prognosis based on the gene expression of the primary tumor (van de Vijver *et al.*, 2002; van't Veer *et al.*, 2002; Ramaswamy *et al.*, 2003).

We evaluated several methods for discriminating between N+ and N- tumors. Hierarchical clustering, although commonly used, suffers from a lack of robustness, statistical instability, and variable biological relevance (Clare and King, 2002; Datta and Datta, 2003), and we used this method strictly as a visualization tool. We generated the signature gene set using a widely accepted statistical method, SAM, that selects genes that are differentially expressed between two groups and estimates a false discovery rate (Tusher *et al.*, 2001). Due to the wide variation in published gene signatures even pertaining to the relatively narrow question of

transformation in HNSCC, we considered the accurate prediction of an independent validation set essential to evaluating our results (Alevizos *et al.*, 2001; Mendez *et al.*, 2002; Gonzalez *et al.*, 2003; Hwang *et al.*, 2003; Leethanakul *et al.*, 2003; Ntzani and Ioannidis, 2003). Principal-components analysis is also a useful visual tool but is not considered a robust predictor, and in our hands showed a trend differentiating N- from N+. We chose to use an SVM as our primary prediction algorithm (Butte, 2002). SVMs, which predict test cases based upon the expression profiles of the original data set mapped into an expanded dimensional space, have been shown to classify accurately a high percentage of tumors of multiple types in microarray experiments (Furey *et al.*, 2000; Ramaswamy *et al.*, 2001; Butte, 2002).

Although the differences in gene expression were sufficient for the SVM to classify correctly all the tumors in our study, they do not appear to reveal absolute differences between metastatic and nonmetastatic primary tumors. The visual principal-components analysis suggests a continuum of gene expression similar to that previously shown in lymph node metastasis of breast cancer (West *et al.*, 2001). Clearly larger studies are warranted to determine boundaries for defining 'clearly metastatic,' 'clearly nonmetastatic,' and 'potentially metastatic' categories.

We chose to embrace the complexity of the tumor system in our experimental design by using non-microdissected tissue samples because interactions with

immune and stromal cells have been shown to play a role in tumor aggressiveness (van Kempen *et al.*, 2003). Although it is known that keratin 13 is expressed by tumor cells, many of the other genes in the signature could be expressed by tumor or stroma (Depondt *et al.*,

1999). The presence of immune cells in the primary tumors differed from N+ to N- patients, as indicated by differing expression of HLA II DO β and CD64. It is possible that the genetic make-up of the host, through immune or stromal cell interaction with the tumor, may provide a critical step in the progression of a nonmetastatic to a metastatic phenotype.

Some of these steps may have been previously identified, since three of the genes in our signature gene set have been previously associated with an aggressive phenotype in other types of cancers. *KCNJ5/GIRK4* is a subunit of a G-protein-controlled channel which regulates potassium flow into a cell. Overexpression of its binding partner, *GIRK1*, was found to be correlated with lymph node metastasis in breast carcinoma (Stringer *et al.*, 2001). Downregulation of the transcription factor *AP-2* is associated with metastasis in melanoma, and our signature gene set includes upregulated *KLF12*, an *AP-2* repressor gene (Karjalainen *et al.*, 1998; Roth *et al.*, 2000). Vacuolar ATP synthase subunit C, overexpressed in our N+ patients, has been correlated with invasiveness of pancreatic tumors (Ohta *et al.*, 1996). Our signature gene set also bears some similarities with the one defined for lymph node metastasis of pancreatic tumors by Nakamura *et al.* (2004) (*DLG3*, *Smad6*, and *AP1S1* compared with *DLG5*, *Smad3*, and *AP3D1*), perhaps indicating the presence of common pathways promoting lymphatic metastasis in different types of tumors.

One gene we were surprised to find missing from our signature set was *CXCR4*, a chemokine receptor that is being investigated as a regulator of metastasis. Although immunohistochemical studies in HNSCC, melanoma, and breast cancer have suggested a role for *CXCR4* in lymph node metastasis, our examination found no correlation between *CXCR4* expression and metastasis either in differentiating N+ from N- or N+ from lymph node metastasis (Robledo *et al.*, 2001; Kato *et al.*, 2003; Uchida *et al.*, 2003; Delilbasi *et al.*, 2004). Experiments in mice which inserted *CXCR4* into metastatic cells resulted in an increase in hematogenous metastasis only, and blockade of *CXCR4/SDF-1* interaction has a much greater inhibitory effect on hematogenous than lymphatic metastasis (Muller *et al.*, 2001; Murakami *et al.*, 2002). Recent data show that *CXCR4* is able to enhance T-cell entrance to the lymph node via the bloodstream but not the lymphatics (Scimone *et al.*, 2004). Therefore, we propose that *CXCR4*'s role in

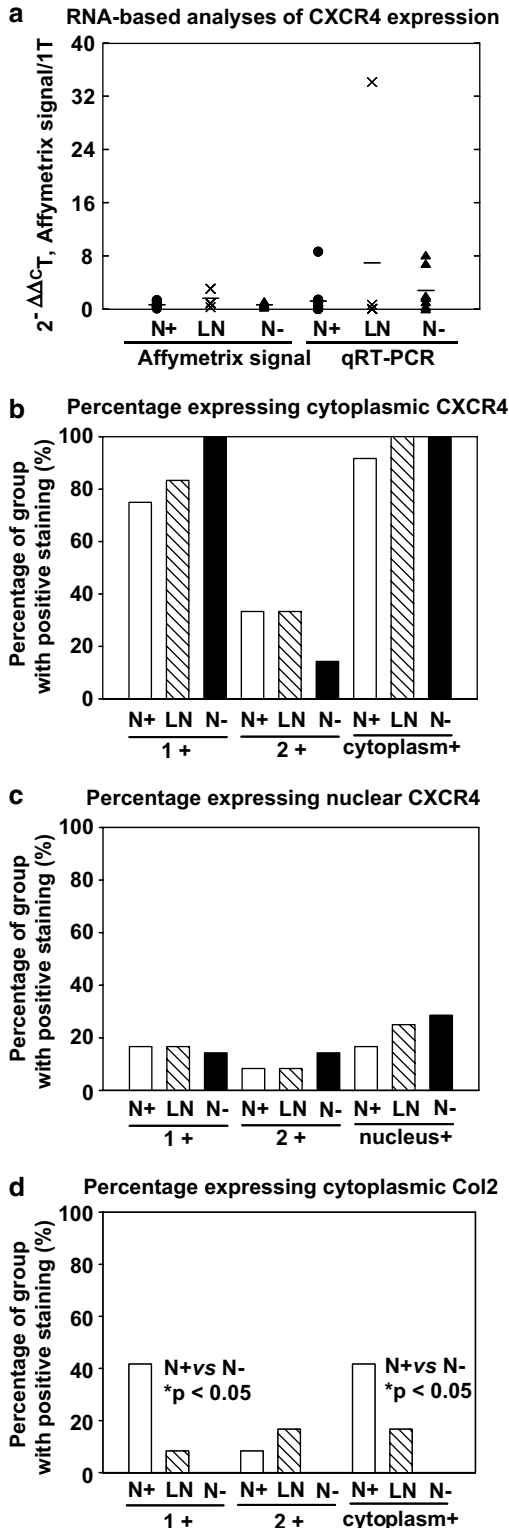


Figure 6 Col2 but not CXCR4 associated with lymph node metastasis. (a) CXCR4 RNA levels were assayed by microarray and quantitative real-time RT-PCR in 11 N+ (●), four lymph node metastases (X), and seven N- (▲). (b, c) CXCR4 protein levels were assayed by immunohistochemistry in nine N+ (□), 10 lymph node metastases (▨), and seven N- (■) from the initial patient set, and three N+ and two lymph node metastases from the supplemental immunohistochemistry set. There were no significant differences between N+ and N- groups, or between N+ and lymph node metastasis groups. (d) Col2 protein levels were assayed by immunohistochemistry in the same patients

metastasis to the lymph node is primarily through the hematogenous route and that this molecule is not a major participant in lymphatic tumor spread. The absence of CXCR4 and the lack of overlap of our gene signature with other gene signatures defined for hematogenous metastasis indicate that metastases through the lymphatic and hematogenous routes employ different pathways.

Materials and methods

Tumor procurement and RNA extraction

Samples from squamous cell carcinomas of the oral cavity were identified through the Head and Neck Tumor Database of the University of Pennsylvania. Institutional Review Board approval and informed consent was obtained for all tissue use. Tumors were pathologically staged according to AJCC guidelines. Within 30 min after surgical extirpation, tissues were frozen in liquid nitrogen. RNA was extracted using Trizol (Life Technologies Inc., Gaithersburg, MD, USA) from 20 to 110 mg of tissue. RNA quality and quantity were confirmed with agarose gel electrophoresis and spectrophotometry.

Array hybridization and data analysis

Probes were generated using the procedures described by Affymetrix (Santa Clara, CA, USA) by the University of Pennsylvania Microarray Facility and hybridized to an Affymetrix U133A Genechip (Affymetrix, CA, USA). The microarrays were evaluated as described by Affymetrix using a GeneArray 2500 confocal scanner (Affymetrix, CA, USA). The average signal from two sequential scans was calculated for each microarray feature.

Background subtraction was carried out using the algorithms provided by Affymetrix Microarray Suite 5.0 (Affymetrix, CA, USA). Total gene expression signal for each array was scaled to 150 signal units to allow comparison of arrays.

Scaled data were imported into the TIGR Multiexperiment Viewer version 2.2 from the Institute of Genomic Research (Rockville, MD, USA). SAM was used to identify genes differentially expressed between N+ and N- samples (Tusher

et al., 2001). Genes whose mean expression levels for both groups fell below background, estimated by the average of *Bacillus subtilis* gene signals, were discarded. The remaining signature set of genes was analysed using principal-components analysis (scaled for equal variance, results extracted by JScatter, a lab-based program, for visualization) and SVM. Crossvalidation of the original samples using a leave-one-out approach and a diagonal factor of 2 provided an estimate of the accuracy of the SVM algorithm, which was also used to predict the classes of four independent samples.

Quantitative RT-PCR

Real-time RT-PCR primers and probes were designed using ABI Primer Express (Applied Biosystems, Foster City, CA, USA), manufactured by Integrated DNA Technologies Inc. (Coralville, IA, USA), and analysed in the University of Pennsylvania Center for AIDS Research ABI Prism 7700 Sequence Detection System (Applied Biosystems, CA, USA). Samples were quantitated according to the $-\Delta\Delta C_T$ method (Livak and Schmittgen, 2001).

Immunohistochemistry

Formalin-fixed, paraffin-embedded sections of patient tissues were dewaxed according to standard procedures and blocked in 2% H₂O₂ in methanol at room temperature for 20 min. CXCR4 staining included incubation in 10 mM sodium citrate at pH 6.0 for 5 min at 95°C, blocking with 10% goat serum at 37°C for 1 h, and overnight incubation at 4°C with antibody (12G5, R&D Systems, Minneapolis, MN, USA) at 1:400 dilution. Col2 staining included incubation with pepsin (Abcam, Cambridge, MA, USA) at 37°C for 30 min and 1 h incubation at 37°C with antibody (6B3, Chemicon, Temecula, CA, USA) at 1:25 dilution.

Acknowledgements

We thank Amy Ziober for RNA preparation of the validation patient set, Jim O'Donnell for development of JScatter, the Greater Philadelphia Bioinformatics Alliance and the Institute for Genomics Research for presenting the Microarray Data Analysis Workshop, and the University of Pennsylvania Microarray Core Facility for chip processing and consultation.

References

- Alevizos I, Mahadevappa M, Zhang X, Ohyama H, Kohno Y, Posner M, Gallagher GT, Varvares M, Cohen D, Kim D, Kent R, Donoff RB, Todd R, Yung CM, Warrington JA and Wong DT. (2001). *Oncogene*, **20**, 6196–6204.
- Bei R, Pompa G, Vitolo D, Moriconi E, Ciocci L, Quaranta M, Frati L, Kraus MH and Muraro R. (2001). *J. Pathol.*, **195**, 343–348.
- Belbin TJ, Singh B, Barber I, Socci N, Wenig B, Smith R, Prystowsky MB and Childs G. (2002). *Cancer Res.*, **62**, 1184–1190.
- Bertucci F, Salas S, Eysteris S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Lloriod B, Bachelart L, Montfort J, Victorero G, Viret V, Ollendorff V, Fert V, Giovaninni M, Delpero J-R, Nguyen C, Viens P, Monges G, Birnbaum D and Houlgatte R. (2004). *Oncogene*, **23**, 1377–1391.
- Butte A. (2002). *Nat. Rev.*, **1**, 951–960.
- Chen B-S, Wang M-R, Xu X, Cai Y, Xu Z-X, Han Y-L and Wu M. (2000). *Int. J. Cancer*, **88**, 862–865.
- Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X, Shockley WW, Weissler MC, Dressler LG, Shores CG, Yarbrough WG and Perou CM. (2004). *Cancer Cell*, **5**, 489–500.
- Clare A and King RD. (2002). *In Silico Biol.*, **2**, 511–522.
- Datta S and Datta S. (2003). *Bioinformatics*, **19**, 459–466.
- Delilbasi CB, Okura M, Iida S and Kogo M. (2004). *Oral Oncol.*, **40**, 154–157.
- Depondt J, Shabana A-H, Sawaf H, Gehanno P and Forest N. (1999). *Eur. J. Oral Sci.*, **107**, 442–454.
- Fidler IJ and Kripke ML. (1977). *Science*, **197**, 893–895.
- Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M and Haussler D. (2000). *Bioinformatics*, **16**, 906–914.
- Ginos MA, Page GP, Michalowicz BS, Patel KJ, Volker SE, Pambuccian SE, Ondrey FG, Adams GL and Gaffney PM. (2004). *Cancer Res.*, **64**, 55–63.
- Gonzalez HE, Gujrati M, Frederick M, Henderson Y, Arumugam J, Spring PW, Mitsudo K, Kim H-W and Clayman GL. (2003). *Arch. Otolaryngol. Head Neck Surg.*, **129**, 754–759.

- Greenberg JS, Fowler R, Gomez J, Mo V, Roberts D, El Naggar AK and Myers JN. (2003). *Cancer*, **97**, 1464–1470.
- He Y, Kozaki K-I, Karpanen T, Koshikawa K, Yla-Herttuala S, Takahashi T and Alitalo K. (2002). *J. Natl. Cancer Inst.*, **94**, 819–825.
- Hong WK and Weber R. (1995). *Head, Neck Cancer: Basic, Clinical Aspects*. Kluwer Academic Publishers: Dordrecht, Boston.
- Hwang D, Alevizos I, Schmitt WA, Misra J, Ohyama H, Todd R, Mahadevappa M, Warrington JA, Stephanopoulos G, Wong DT and Stephanopoulos G. (2003). *Oral Oncol.*, **39**, 259–268.
- Karjalainen J, Kellokoski J, Eskelinen M, Alhava E and Kosma V. (1998). *J. Clin. Oncol.*, **16**, 3584–3591.
- Kato M, Kitayama J, Kazama S and Nagawa H. (2003). *Breast Cancer Res.*, **5**, 144–150.
- Kennedy ME, Nemej J, Corey S, Wickman K and Clapham DE. (1999). *J. Biol. Chem.*, **274**, 2571–2582.
- Kikuchi T, Daigo Y, Katagiri T, Tsunoda T, Okada K, Kakiuchi S, Zembutsu H, Furukawa Y, Kawamura M, Kobayashi K, Imai K and Nakamura Y. (2003). *Oncogene*, **22**, 2192–2205.
- Kim HC, Kusukawa J and Kameyama T. (1993). *Kurume Med. J.*, **40**, 183–192.
- Leethanakul C, Knezevic V, Patel V, Amornphimoltham P, Gillespie JW, Shillitoe EJ, Emko P, Park MH, Emmert-Buck MR, Strausberg RL, Krizman DB and Gutkind JS. (2003). *Oral Oncol.*, **39**, 248–258.
- Livak K and Schmittgen T. (2001). *Methods*, **25**, 402–408.
- Loriot A, Boon T and De Smet C. (2003). *Int. J. Cancer*, **105**, 371–376.
- Mendez E, Cheng C, Farwell DG, Ricks S, Agoff SN, Futran ND, Weymuller J, Ernest A Maronian NC, Zhao LP and Chen C. (2002). *Cancer*, **95**, 1482–1494.
- Muller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME, McClanahan T, Murphy E, Yuan W, Wagner SN, Barrera JL, Mohar A, Verastegui E and Zlotnik A. (2001). *Nature*, **410**, 50–56.
- Murakami T, Maki W, Cardones AR, Fang H, Kyi AT, Nestle FO and Hwang ST. (2002). *Cancer Res.*, **62**, 7328–7334.
- Myers LL, Wax MK, Nabi H, Simpson GT and Lamonica D. (1998). *Laryngoscope*, **108**, 232–236.
- Nagata M, Fujita H, Ida H, Hoshina H, Inoue T, Seki Y, Oshnishi M, Ohyama T, Shingaki S, Kaji M, Saku T and Takagi R. (2003). *Int. J. Cancer*, **106**, 683–689.
- Nakamura T, Furukawa Y, Nakagawa H, Tsunoda T, Ohigashi H, Murata K, Ishikawa O, Ohgaki K, Kashimura N, Miyamoto M, Hirano S, Kondo S, Katoh H, Nakamura Y and Katagiri T. (2004). *Oncogene*, **23**, 2385–2400.
- Ntzani EE and Ioannidis JPA. (2003). *Lancet*, **362**, 1439–1428.
- Ohta T, Numata M, Yagishita H, Futagami F, Tsukioka Y, Kitagawa H, Kayahara M, Nagakawa T, Miyazaki I, Yamamoto M, Iseki S and Ohkuma S. (1996). *Br. J. Cancer*, **73**, 1511–1517.
- Poste G and Fidler IJ. (1980). *Nature*, **283**, 139–145.
- Ramaswamy S, Ross KN, Lander ES and Golub TR. (2003). *Nat. Genet.*, **33**, 49–54.
- Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang C-H, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES and Golub TR. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 15149–15154.
- Robledo MM, Bartolome RA, Longo N, Rodriguez-Frade JM, Mellado M, Longo I, van Muijen GNP, Sanchez-Mateo P and Teixeira J. (2001). *J. Biol. Chem.*, **276**, 45098–45105.
- Roth C, Schuierer M, Gunther K and Buettner R. (2000). *Genomics*, **63**, 384–390.
- Schmalbach CE, Chepeha DB, Giordano TJ, Rubin MA, Teknos TN, Bradford CR, Wolf GT, Kuick R, Misek DE, Trask DK and Hanash S. (2004). *Arch. Otolaryngol. Head Neck Surg.*, **130**, 295–302.
- Scimone ML, Felbinger TW, Mazo IB, Stein JV, von Andrian UH and Weninger W. (2004). *J. Exp. Med.*, **199**, 1113–1120.
- Stringer BK, Cooper AG and Shephard SB. (2001). *Cancer Res.*, **61**, 582–588.
- Tusher VG, Tibshirani R and Chu G. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 5116–5121.
- Uchida D, Begum N-M, Almofti A, Nakashiro K-I, Kawamata H, Tateishi Y, Hamakawa H, Yoshida H and Sato M. (2003). *Exp. Cell Res.*, **290**, 298–302.
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil D, Schreiber GJ, Peterse J, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH and Bernards R. (2002). *N. Engl. J. Med.*, **347**, 1999–2009.
- van Kempen LC, Ruiter DJ, van Muijen GN and Coussens LM. (2003). *Eur. J. Cell Biol.*, **82**, 539–548.
- 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 A, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH. (2002). *Nature*, **415**, 530–536.
- Warner GC, Reis PP, Jurisica I, Sultan M, Arora S, MacMillan C, Makitie AA, Grenman R, Reid N, Sukhai M, Freeman J, Gullane P and Irish J. (2004). *Int. J. Cancer*, **110**, 857–868.
- Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, Westerga J, Meuwissen SG, Albertson DG and Meijer GA. (2003). *Oncogene*, **22**, 1872–1879.
- West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson Jr JA, Marks JR and Nevins JR. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 11462–11467.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>).