Oncogene (2001) 20, 6196–6204

www.nature.com/onc

Oral cancer *in vivo* gene expression profiling assisted by laser capture microdissection and microarray analysis

Ilias Alevizos[1,10], Mamatha Mahadevappa[2,10], Xue Zhang[3], Hiroe Ohyama[1], Yohko Kohno[1], Marshall Posner[4], George T Gallagher[1], Mark Varvares[5], Donald Cohen[6], Dae Kim[1], Ralph Kent[7], R Bruce Donoff[8], Randy Todd[8], Chou Ming Yung[9], Janet A Warrington[2] and David TW Wong*[,1]

[1]*Department of Oral Medicine and Diagnostic Sciences, Division of Oral Pathology, 188 Longwood Avenue, Harvard School of Dental Medicine, Boston, Massachusetts, MA 02115, USA;* [2]*Affymetrix Inc., 3380 Central Expressway, Santa Clara, California, CA 95051, USA;* [3]*Department of Biology, China Medical University, Shenyang 110001, China;* [4]*Department of Medicine, Division of Hematology Oncology, Dana Farber Cancer Institute, Boston, Massachusetts, MA 02115, USA;* [5]*Department of Otolaryngology, Massachusetts Eye & Ear Infirmary, Boston, Massachusetts, MA 02115, USA;* [6]*University of Nebraska Medical Center, College of Dentistry, Lincoln, Nebraska, NE 68583-0740, USA;* [7]*The Forsyth Institute, Boston, Massachusetts, MA 02115, USA;* [8]*Department of Oral and Maxillofacial Surgery, Massachusetts General Hospital, Boston, Massachusetts, MA 02114, USA;* [9]*Chung Shan Medical and Dental College, Taichung, Taiwan*

Large scale gene expression profiling was carried out on laser capture microdissected (LCM) tumor and normal oral epithelial cells and analysed on high-density oligonucleotide microarrays. About 600 genes were found to be oral cancer associated. These oral cancer associated genes include oncogenes, tumor suppressors, transcription factors, xenobiotic enzymes, metastatic proteins, differentiation markers, and genes that have not been implicated in oral cancer. The database created provides a verifiable global profile of gene expression during oral carcinogenesis, revealing the potential role of known genes as well as genes that have not been previously implicated in oral cancer. *Oncogene* (2001) **20,** 6196–6204.

Keywords: oral cancer; gene expression; laser capture microdissection; microarrays

Introduction

High-throughput technologies, such as DNA microarrays, comprehensively profile and monitor gene expression in biological processes, including cancer. While this approach has been applied to hematopoietic tumors (Alizadeh *et al*., 2000; Golub *et al*., 1999), its application to the major type of human cancer, solid tumors, has only been reported for tumor homogenates and cell lines (Alon *et al*., 1999; Perou *et al*., 2000; Sgroi *et al*., 1999). The cell-specific profiling of solid tumor gene expression has been hampered by the inability to procure specific pure cell populations. This obstacle was recently overcome through the development of laser capture microdissection (LCM), which allows the harvesting of specific cells from complex tissues such as solid tumors.

In this study we demonstrate the application of LCM to the harvesting of normal and tumor cells from a solid tumor site, oral cavity cancer. The mRNAs were linearly amplified by three rounds of T7 RNA polymerase reaction, biotinylated, and then hybridized to HuGeneFL® microarrays. Analysis of the hybridization outcome revealed that 26 to 40% of the genes queried were present in all samples analysed. Data analysis revealed intriguing clues to the biological pathways involved in oral malignancy. These include genes associated with metastatic/invasion, transcription factors, oncogenes/tumor suppressor genes, differentiation markers and members in the xenobiotic pathway. To validate the GeneChip® data, we selected three genes that are consistently altered (5/5) in the metastatic/invasion pathway (collagenase, urokinase plasminogen activator (UPA) and cathepsin L) and validated their differential expression by real time quantitative PCR. Our approach should be applicable to any solid tumor gene expression profiling study using currently available microarrays to reveal relevant mechanistic pathways and events in normal and pathological processes.

Results and discussion

Gene expression profiling of normal and tumor oral epithelia

The use of LCM to harvest cells from their native tissue environment and the use of high-density oligonucleotide probe arrays to identify gene expression

*Correspondence: DTW Wong;
E-mail: dwong@warren.med.harvard.edu
[10]Both investigators contributed equally and should be considered as first author
Received 28 February 2001; revised 23 May 2001; accepted 31 May 2001

differences between normal and malignant oral epithelial cells provide powerful means to decode the molecular events involved in the genesis and progression of head and neck/oral cancer. We isolated paired normal and malignant epithelium from five snap frozen biopsies (10 samples total). Most of these patients have a history of smoking and alcohol consumption, which are the major etiological causes of oral cancer (see Materials and methods). The quality and quantity of isolated RNA was examined by reverse transcription polymerase chain reaction (RT–PCR) of five cellular maintenance gene transcripts of high to low abundance (glyceraldehyde-3-phosphate dehydrogenase; tubulin-α; β-actin; ribosomal protein S9; uniquitin C) (Ohyama *et al*., 2000). The quantity of isolated RNA was also assessed with RiboGreen RNA Quantitation Reagent and kit (Molecular Probes, Eugene, OR, USA) using spectrofluorometry (Bio-Rad, Hercules, CA, USA). Only those samples exhibiting PCR products for all five cellular maintenance genes were used for subsequent analysis. The biotinylated cRNA from the 10 samples (normal and cancer) were further used to hybridize the Affymetrix Test-1 probe arrays to determine cRNA quality and integrity. The arrays contain probes representing a handful of maintenance genes and a number of controls (Ohyama *et al*., 2000). Analysis of the arrays confirmed the RT–PCR findings. cRNA linearly amplified from human oral cancer tissue produced no nonspecific or unusual hybridization patterns and the transcripts for the maintenance genes were detected. The 5′ region of the RNA was degraded but enough 3′ transcript was intact to proceed for hybridization using the HuGeneFL probe arrays. In addition, probes synthesized on the arrays are biased to the last 600 bp in the 3′ region of the transcripts. Yields from the LCM and amplification steps are shown in Table 1a. Linear amplification of the total RNA began with $\sim$100 ng of total RNA.

Table 1b summarizes the hybridization outcome of the five paired cases of oral cancers. The per cent transcript detected ranged from 26 to 40%, indicating satisfactory quality and representation of the harvested RNA. Note that the difference between the normal and cancer samples from each patient is very similar, indicating little variability among each pair, suggesting that the quality of the RNA isolated from the normal and tumor epithelium is similar.

Microarray hybridization results

Differential gene expression using GeneChip® analysis software revealed 404 probe sets changed in the majority of the cases (3/5). Among the 404,211 were increased in tumor and 193 were decreased in tumor, compared to normal. There were 39 probe sets that changed in all five cases (Table 2a). Sixteen of them were increased in tumor and 23 were decreased in tumor, compared to normal. Table 2b lists a subset of the differentially expressed genes grouped into biological pathways known to be relevant in carcinogenesis.

Our data revealed that many known genes involved in neoplasia are differentially expressed in the five paired cases of oral cancer. Our analysis also revealed members of known biological pathways whose expression are altered during oral carcinogenesis. These include metastatic and invasion pathways, transcription factors, oncogenes and tumor suppressor genes, and differentiation markers (Table 2b). Of particular importance are the differentially expressed genes that are not yet functionally characterized or genes that have not been studied by classic methods in head and neck/oral carcinogenesis. One such example is neuromedin U (Nmu), which is down-regulated in 5/5 tumors (Szekeres *et al*., 2000). Nmu is a poorly understood protein that manifests potent contractile activities on smooth muscle cells. Recently, two G-protein coupled receptors (NMU1 and NMU2) have been identified to interact with Nmu with nanomolar potency (Fujii *et al*., 2000; Raddatz *et al*., 2000). Our data provide strong evidence that Nmu is relevant in the development of oral malignancy and suggest the need for further study of the role of Nmu (down regulated expression in tumor) in carcinogenesis.

In order to validate our findings, three metastatic pathway genes whose expression are consistently altered in the five paired cases of oral cancer, were selected. Real-time quantitative PCR (RT–QPCR) in conjunction with the TaqMan specific probe system or

Table 1a LCM, RNA isolation and amount of cDNA after two rounds of T7 amplification

	Case 1		*Case 2*		*Case 3*		*Case 4*		*Case 5*	
	Normal	*Cancer*	*Normal*	*Cancer*	*Normal*	*Cancer*	*Normal*	*Cancer*	*Normal*	*Cancer*
Number of LCM shots	48 000	40 000	40 000	40 000	40 000	40 000	40 000	40 000	40 000	40 000
Approx # of cells	200 000	120 000	120 000	120 000	120 000	120 000	120 000	120 000	120 000	120 000
Approx amt of RNA used for T7 amp	100 ng	75 ng	100 ng	120 ng	100 ng	115 ng	100 ng	100 ng	100 ng	100 ng
ds-cDNA amt after two rounds of T7 amp	8.75 μg	5.88 μg	1.56 μg	6.88 μg	2.5 μg	2.97 μg	23.5 μg	6.0 μg	17.74 μg	23.13 μg

Amount of ds-cDNA after two rounds of T7 amplification is dependent on the quality of the LCM-generated RNA from the normal and tumor tissues

Table 1b Per cent transcripts detected in normal and tumor tissues

	Case 1		*Case 2*		*Case 3*		*Case 4*		*Case 5*	
	Normal	*Cancer*	*Normal*	*Cancer*	*Normal*	*Cancer*	*Normal*	*Cancer*	*Normal*	*Cancer*
% Genes detected	25.8	30.3	30.3	33.0	29.4	30.3	35.3	35.8	40.3	40.2

Table 2a Thirty-nine genes whose expression changed in 5/5 cases

Accession number	*Name*	*Average % fold changes*	*Chromosome localization*	*Oral cancer association*	*References*
Increase in tumor					
D86983	p53-responsive gene 2	1483	2pter-p25.1	Yes	(Yeudall *et al*., 1997)
J03040	Secreted protein, acidic, cysteine-rich, osteonectin	2620	5q31.3-q32	Yes	(Porte *et al*., 1998)
M11147	Ferritin, light polypeptide	432	19q13.3-q13.4	Yes	(Kimura *et al*., 2000)
X12451	Cathepsin L	330	9q21-q22	Yes	(Kawamata *et al*., 1997; Strojan *et al*., 2000)
X17042	Proteoglycan 1, secretory granule	320	10q22.1		
X54925	Matrix metalloproteinase 1 (interstitial collagenase)	17600	11q22.3	Yes	(Magary *et al*., 2000)
Z29083	5T4 oncofetal trophoblast glycoprotein	443	6		
D43968	Runt-related transcription factor 1 (acute myeloid leukemia 1)	500	21q22.23		
X02761	Fibronectin 1	5117	2q34	Yes	(Mighell *et al*., 1997)
Z74616	Collagen, type I, alpha 2	785	7q22.1		(Junien *et al*., 1983)
X57579	Inhibin, beta A (activin A, activin AB alpha polypeptide)	12984	7p15-p13		
M30257	Vascular cell adhesion molecule 1	324	1p32-p31	Yes	(Liu *et al*., 1999)
M55998	Human alpha-1 collagen type I gene,	1493	N/A		
X65965	Superoxide dismutase 2, mitochondrial	360	6q25.3	Yes	(Kurokawa *et al*., 1998; Muramatsu *et al*., 1995)
X54489	GRO1 oncogene (melanoma growth stimulating activity, alpha)	4652	4q21	Yes	(Loukinova *et al*., 2000)
L13923	Fibrillin 1 (Marfan syndrome)	616	15q21.1		
Decrease in tumor					
X78932	Zinc finger protein 273	485	N/A		
J04469	Creatine kinase, mitochondrial 1 (ubiquitous)	4526	15q15	Yes	(Lin and Chen, 1991)
L05779	Epoxide hydrolase 2, cytoplasmic	320	8p21-p12	Yes	(Janot *et al*., 1993)
M32402	Placental protein 11 (serine proteinase)	21707	12q13.1		
M69177	Monoamine oxidase B	757	Xp11.4-p11.3		
M98447	Transglutaminase 1 (K polypeptide epidermal type I, protein-glutamine-gamma-glutamyltransferase)	570	14q11.2	Yes	(Shin *et al*., 1994)
U17077	BENE protein	2810	2q13		
U34252	Aldehyde dehydrogenase 9 (gamma-aminobutyraldehyde dehydrogenase, E3 isozyme)	285	1q22-q23		
U46689	Aldehyde dehydrogenase 10 (fatty aldehyde dehydrogenase)	419	17p11.2		
U67963	Lysophospholipase-like	671	3		
U83115	Absent in melanoma 1	509	6q21		
U90902	Human clone 23612	315	N/A		
X07695	Keratin 4	4321	12q13	Yes	(Suo *et al*., 1993)
X07696	Keratin 15	9109	17q21	Yes	(Viaene and Baert, 1995)
X76029	Neuromedin U	1005	4q12		
X76180	Sodium channel, nonvoltage-gated 1 alpha	968	12p13		
X78549	Protein tryosine kinase 6	535	20q13.3		
Y09267	Flavin containing monooxygenase 2	3197	1q23-q251		
Y09616	Carboxylesterase 2 (intestine, liver)	331	16		
Y07909	Epithelial membrane protein 1	842	12p12.3		
X53296	Interleukin 1 receptor antagonist	1324	2q14.2	Yes	(von Biberstein *et al*., 1996)
M61855	Cytochrome P450, subfamily IIC (mephenytoin4-hydroxylase), polypeptide 9	461	10q24	Yes	(Murray *et al*., 1994)
U18934	TYRO3 protein tyrosine kinase	454	15q15.1-q21.1		

SYBR® Green system was used to validate the expression levels of interstitial collagenase (a member of the MMP's involved in metastasis), urokinase plasminogen activator (UPA, associated with metastasis) and cathepsin L (a member of the serine proteases). Comparison of the microarray and RT–QPCR data revealed that they approximate each other (Table 3). The actual comparative data is for collagenase is graphically shown in Figure 1a while Figure 1b shows the gel electrophoresis results of the collagenase RT–QPCR products. Similar data were obtained for UPA and cathepsin L (Table 3, gel electrophoresis data not shown). We have further validated a number of other high and low genes including Neuromedin U, GST, cytochrome P450, ALDH-9, ALDH-10 and Wilm's tumor-related protein (data not shown). Figure 2 is a schematic highlighting our findings in the proteolysis pathway that may contribute in the development of oral cancer.

SOM and hierarchical clustering analyses

The microarray data, though voluminous, can be analysed by pattern recognition (clustering) software to aid in deriving lists of genes that distinguish and characterize disease versus normal biopsies, thus

Table 2b Representative sample of differentially regulated genes

Accession number	*Description/ function*	*Chromosomal location*	*Altered regulation*	*Average % fold changes*
Metastatic/invasion pathway				
X12451	Cathepsin-L	9q21-22	Up 5/5	330
X54925	Collagenase/MMP I	11q22.3	Up 5/5	17600
X02419	UPA	10q24	Up 5/5	363
Z29083	5T4 Oncofetal protein	6q13-14	Up 5/5	443
Transcriptional factors				
D43968	Runt-related transcription factor/AML1	21q22	Up 5/5	499
U85658	ERF1 transcription factor	20q13.2	Down 4/5	300
Ocogenes/suppressors				
M57731	Gro-beta oncogene	4q12-13	Up 4/5	555
M16038	Lyn-tyrosine kinase (oncogene LYN)	8q13	Up 4/5	496
L13698	Gas-1	9q21.3-22.1	Up 4/5	216
	Tumor associated membrane protein (control cell-			
Y07909	Cekk interactions and cell proliferation)	12p12.3	Down 5/5	842
X98311	CEA2 (onco-suppressor)	19q13.2	Down 5/5	611
M16750	PIM-1-oncogene	6p21.2	Down 4/5	372
U83115	AIM1 (onco-suppressor)	6q21	Down 4/5	509
Differentiation markers				
X07695	Cytokeratin 4	12q13	Down 5/5	4321
X07696	Cytokeratin 15	17q21	Down 5/5	9110
M98447	Transglutaminase 1	14q11.2	Down 5/5	570
Others				
U37546	Apoptosis-inhibitor protein	11q22-23	Up 5/5	525
X76029	Neuromedin U	4q12	Down 5/5	1005
S45630	Alpha-beta-crystallin/Rosenthal protein	11q22.3-23.1	Down 4/5	668
X78932	ZFP HZF9 (potent repressor)	19p12	Down 5/5	485
U47414	Cyclin G2 (cell cycle regulators)	4q21	Down 4/5	1244

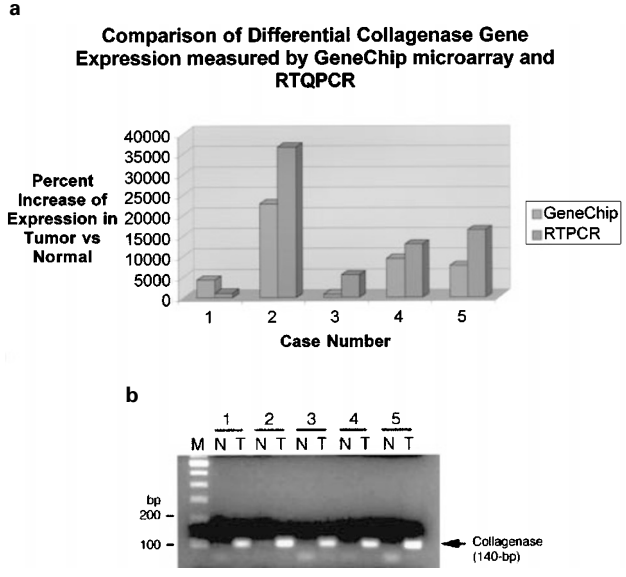

Figure 1 Comparison and validation of microarray data by RT–QPCR. (**a**) Comparison of gene expression data (from GeneChip©) and by RT–QPCR for collagenase. (**b**) Visualization of actual by RT–QPCR products by agarose gel electrophoresis

shedding light on molecular genetic profiles and ultimately the mechanism of the disease under study. Techniques used for clustering include self-organizing maps (SOM), Bayesian, hierarchical, and k-means. SOM was selected for our analysis because of advantages in initial exploration of the data allowing the operator to impose partial structure on the clusters (Tamayo *et al*., 1999). Other advantages of SOM include good computational properties, computational speed and easy implementation. SOM analysis was applied to the microarray data on the five paired cases of oral cancers. The clusters graphically represent gene expression patterns across all 10 samples (normal and tumor), each cluster differing in gene number and grouping. This method provides candidate set of genes whose differing expression activity can be used to distinguish normal and tumor cell behavior.

By SOM analysis, 178 transcripts were found to be differentially expressed between tumor and normal tissues. An important observation is that many of the differentially down-regulated genes are known to be important enzymes in the xenobiotic metabolic pathway (Jourenkova-Mironova *et al*., 1999; Katoh *et al*., 1999; Park *et al*., 1997; Sato *et al*., 1999). These include cytochrome *c* oxidase subunit Vb (coxVb), gamma-aminobutyraldehyde dehydrogenase, microsomal glutathione S-transferase (GST-II), aldehyde dehydrogenase 7 (ADH7), COX C VIII, ALDH8, EPH2 cytosolic epoxide hydrolase and ALDH10. Further data analysis revealed that other xenobiotic pathway genes, not included in this cluster, were also down-regulated in all five cases, suggesting perhaps a general downregulation of xenobiotic pathway genes during oral cancer development.

The xenobiotic pathway is of importance in the degradative metabolism of both foreign/native toxic and carcinogenic products. Phase I and II xenobiotic enzymes are two key sequential steps in the metabolism of toxic substances including alcohol and tobacco products. It is interesting to note that most of the five

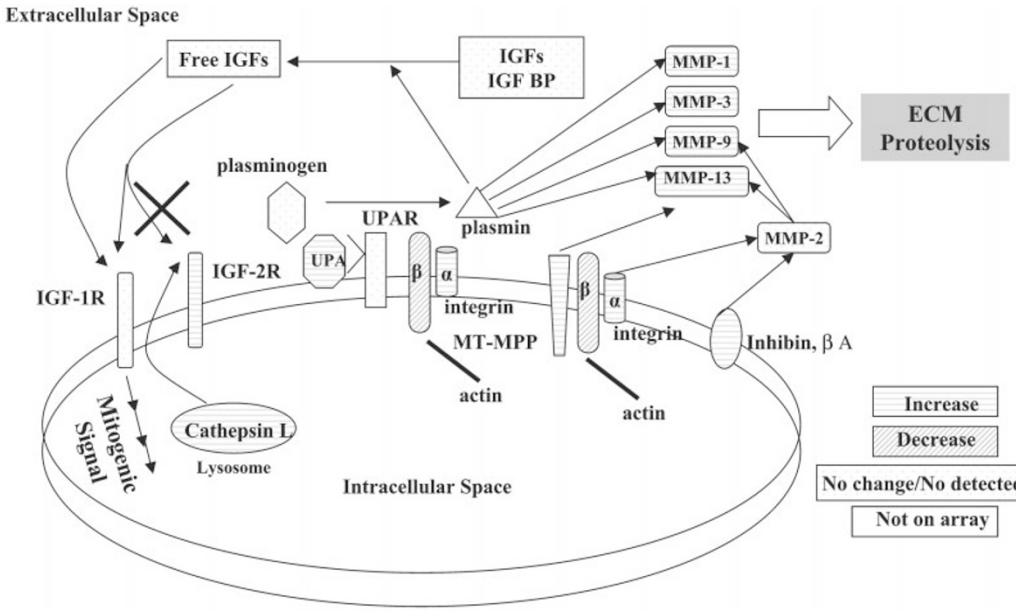

Figure 2 Proteolysis, remodeling of the extracellular matrix and cell migration pathway in oral cancer based on differential gene expression data. One of the central themes in this schematic is the activation of plasmin from plasminogen through the action of urokinase plasminogen activator (UPA). Plasmin is the main activator of matrix metalloproteinases (MMP) such as MMP-1, 3, 9 and 13. Activation of insulin growth factors (IGF) by plasmin and internalization of IGF-2 receptor by cathepsins are also of significance for continuous and uncontrolled cell division. Inhibin beta alpha is another molecule that has a function in activating MMP-2 which also activates MMP-9 and MMP-13. In our analysis, UPA as well as several metalloproteinases, cathepsin L, inhibin alpha beta and IGF-2 receptor are upregulated in the tumors. MT-MPP = Membrane Type Matrix Metalloproteinase

cases of oral cancer were from heavy smokers and drinkers. These data may suggest that key regulatory events were altered in the xenobiotic pathways during oral carcinogenesis that may contribute to the increased susceptibility towards carcinogens such as tobacco and alcohol, the two major etiological factors for oral carcinogenesis.

Using Matlab analysis, 117 transcripts were identified to be differentially expressed between normal and tumor cells. Hierarchical clustering is shown in Figure 3. The distinct clustering of the normal samples, from the tumor samples suggests that LCM procured pure, homogenous samples. It is interesting to note that the tumor samples 1B and 3B clustered more tightly together than the other three tumor samples. The sample size is too small to draw any conclusions for this however. A larger study is in progress to determine if this tight sub-clustering reflects a developmental stage of oral cancer. Based on the outcome of three analytical methods (GeneChip, SOM and Matlab), ~600 candidate oral cancer genes were identified. Of this comprehensive set, 27 of the differentially expressed genes were identified by all three methods (Table 4). Of the 600 candidate genes, 41% were detected at low levels, 1–5 copies per cell.

Shillitoe *et al*. (2000) and Leethanakul *et al*. (2000a,b) have created expression libraries of human oral cancer cell lines and LCM-generated oral cancer tissues. Their studies revealed 52 genes to be differentially expressed at more than twofold in at least three of the cancer tissue sets. Of these 52 genes,

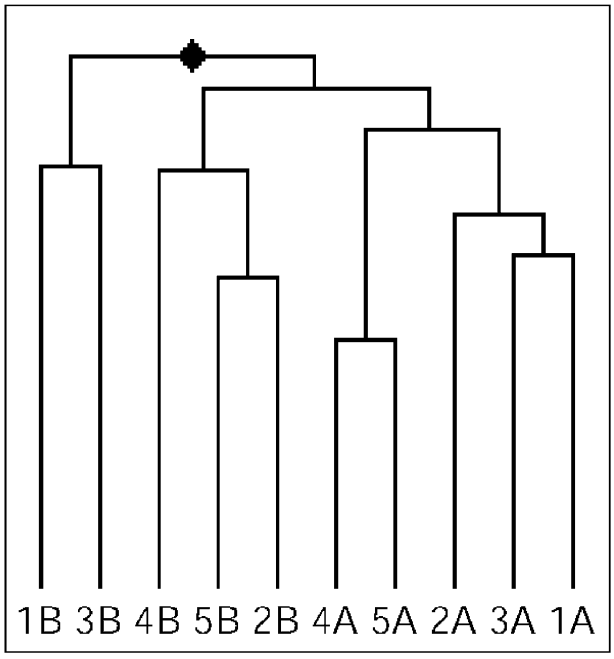

Figure 3 Hierarchical clustering was done on intensity values standardized by dividing by root mean square. Cosine correlation of similarity coefficient and complete linkage clustering classified the samples as shown. Normal 'A'; Tumor 'B'

26 were present on the Affymetrix GeneChip®. By our analysis of these 26 overlapping genes, 18 were called absent (not detectable) in both normal and tumor

Table 3 Comparison of per cent increases for three upregulated genes measured by GeneChip and real-time quantitative PCR data

		Samples				
		1	*2*	*3*	*4*	*5*
Accession	*Gene*	*GC/RT−QPCR*	*GC/RT−QPCR*	*GC/RT−QPCR*	*GC/RT−QPCR*	*GC/RT−QPCR*
X54925	Collagenase	4300/1100	22900/36800	800/5500	49500/130000	7800/16500
X12451	Pro-cathepsin L	300/800	570/700	180/300	300/2800	300/400
X02419	UPA	400/100	440/800	150/3200	400/980	425/900

GC = Gene expression data. RT−QPCR = real-time quantitive PCR. The observed discrepancy in the precise quantitation of the GeneChip® and the RT−QPCR reflects the fact that a minute amount (ng) of LCM-generated total RNA was used for amplification followed by biotinylation and hybridization to the GeneChip® microarrays. Using the same LCM-generated total RNA, we validated the GeneChip® data of three tumor metastatic genes by real-time quantitative PCR (RT−QPCR). These two independent approaches yielded data which indicated a similar trend (Figure 1). Both methods showed that genes are upregulated from undetectable levels in the control to moderate abundance in the tumor cells. Similar results of GeneChip® versus RT−QPCR correlation were previously used by Welsh *et al*. (2001) to validate candidates identified in an ovarian cancer study. The RT−QPCR data confirmed the upregulation and downregulation of selected candidates. So while there is discrepancy in the precise quantitation of GeneChip® and RT−QPCR data of each sample, the overall trend and correlation are similar. The array data produces information about relative abundance that is accurate to within 1.5–2-fold (Lockhart *et al*., 1996; Redfern *et al*., 2000) providing information that allows binning of the transcript levels by low, low-medium, medium high or high abundance (Warrington *et al*., 2000a,b; Lockhart *et al*., 1996; Redfern *et al*., 2000)

samples (DP-2/U18422; TIMP-4/U76456; VEGF-C/U43142; FGF3/X14445; FGF5/M37825; FGF6/X63454; IGFBP5/M65062; EGF cripto protein CR1 and 2/M96956; APC/M74088; ERK6/X79483; GDI dissociation protein/U82532; MAP kinase p38/L35253; MKK6/U39657; MEKK3/U78876; Frizzled/L37782; FZD3/U82169; Dishevelled homolog/U46461; Patched homolog/U43148;); one gene shows no difference between normal and tumor tissues (cyclin H/U11791); one gene was upregulated in 5/5 tumors (beta1-catenin/X87838); three genes were upregulated in 4/5 tumors (thrombospondin2 precursor/L12350; inhibitor of apoptosis protein/U45878; Caspase 5 precursor/U28014) and one gene was upregulated in 3/5 tumors (MMP-10/X07820); one gene was downregulated in 4/5 tumors (RhoA/L25080). Finally one gene was upregulated in two tumors, downregulated in two tumors and called absent in the fifth oral cancer (TRAF2/U78798).

Of the 52 genes, two genes were detected present only through our LCM/GeneChip® analysis. They are human SPARC/osteonectin (J03040) and 5T4 oncofetal antigen (Z29083), which are consistently altered in the same manner in all five oral cancers examined.

Of interest is that a number of genes were identified by either our LCM/oligonucleotide microarray approach or the LCM/cDNA library approach (Leethanakul *et al*., 2000a,b; Shillitoe *et al*., 2000) to be highly expressed/upregulated in oral cancer tissues. These include: ferritin heavy polypeptide I, urokinase plasminogen activator, ATP-binding cassette transporter, interleukin-1 receptor antagonist and keratin 4.

In addition, there are genes that were differentially expressed and detectable in the cell line study (Shillitoe *et al*., 2000), not in the Head and Neck CGAP (HNCGAP) libraries (Leethanakul *et al*., 2000a,b), but were detected present in our dataset. Good examples of these genes are the collagen type 1 alpha 2 genes and the heat shock protein 70 kD gene. An example of a gene that was not identified by either LCM approach (HNCGAP libraries or our method), but detected present in the cell line filtered cDNA microarray analysis is the transforming growth factor alpha gene, suggesting perhaps the elevated expression of this gene maybe associated with *in vitro* culturing.

The different outcome of the various studies are likely reflective of the experimental approaches and methods of analyses. First, by using LCM-generated RNA, contamination of heterogeneous cellular elements is avoided. Second, sample number and the type of microarray used in the respective studies may be relevant to the discrepancies. Third, the stage of the tumor, source and anatomical site of the oral cancers, and handling methods can further result in different gene expression levels. In our study, the detection of 39 cellular genes consistently altered in 5/5 different paired cases of human oral cancer lends strong support to the experimental approach using LCM-generated RNA, linearly amplified by T7 RNA polymerase and subsequently analysed by high-density oligonucleotide GeneChip® probe arrays. These oral cancer associated genes will now be tested to determine their usefulness as classifiers to predict the normal/malignant nature of oral epithelial tissues. The biology associated with these genes could also be explored to evaluate their role in oral cancer development. A number of these genes are secretory proteins that are upregulated in cancer tissues and could be evaluated as biomarkers of oral malignancy. These include osteonectin, ferritin, cathepsin L, proteoglycan (secretory granule) and oncofetal trophoblast glycoprotein. Our results also indicate that our approach is applicable to the molecular analysis of solid tumor, providing a means for obtaining information about consistent molecular alterations that advance both the understanding of the basic biology of this tumor as well as the clinically relevant aspects of the molecular epidemiology of oral cancer. Our data supports the use of LCM-derived RNA to be used on microarrays and that array hybridization coupled with hierarchical and non-hierarchical analysis methods provide powerful approaches for identifying candidate genes and molecular profiling. A larger study is underway to validate these findings and improve our

understanding of the molecular changes associated with oral cancer.

Materials and methods

Matched normal and malignant human oral cancer biopsies

See Table 5 which summarizes the key demographic data of the five cases of human oral cancers used in the study.

RNA isolation, linear amplification (aRNA) from laser capture microdissection (LCM)-generated cells and target sample preparations

These procedures were carried out as previously described (Ohyama *et al*., 2000). Normal mucosa was obtained from the contralateral side of the patient's oral cavity.

Hybridization of biotinylated cRNA to Test 1 and HuGeneFL® probe arrays

The cRNA was fragmented as described by Wodicka *et al*. (1997) All array washing, staining and scanning was carried out as described in the Gene Expression Manual (Affymetrix, Inc. 1999) (Tamayo *et al*., 1999).

GeneChip® probe arrays

The probe sets consist of oligonucleotides 25 bases in length. Probes are complementary to the published sequences (GeneBank) as previously described (Lockhart *et al*., 1996). The sensitivity and reproducibility of the GeneChip® probe arrays is such that RNAs present at a frequency of 1 : 100 000 are unambiguously detected, and detection is quantitative over more than three orders of magnitude (Redfern *et al*., 2000; Warrington *et al*., 2000). In this set of experiments with oral cancer samples, the bacterial transcript (BioB), spiked before the hybridization at concentration of 1.5 pM which translates to three copies per cell (based on the assumption that there are 300 000 transcripts per cell with an average transcript length of 1 kb), were called present in nine out of 10 experiments (Lockhart *et al*., 1996). Array controls, and performance with respect to specificity and sensitivity are the same as those previously described (Lockhart *et al*., 1996; Mahadevappa and Warrington, 1999; Wodicka *et al*., 1997). Information regarding the genes represented on the arrays used in this study can be found at www.netaffx.com.

GeneCluster/Self-Organizing Maps (SOM)

For GeneCluster analysis, we input gene expression levels and geometry of nodes. Before the computation of the SOM, two preprocessing steps took place. First, a filter was applied to exclude genes that did not change significantly across the pairs. Genes were eliminated if they did not show a relative change of $x=2$ and an absolute change of $y=35$, $(x,y)=(2,35)$. Second, normalization of expression levels across experiments was carried out, thus emphasizing the expression pattern rather than the absolute expression values. Data was normalized using GeneChip software. Description of normalization procedure can be found on pp. A5–14, GeneChip Expression Analysis Technical Manual, 1999 (Tamayo *et al*., 1999).

Table 4 Differentially expressed genes identified by all three methods

Accession number	*Name*	*Expression up/down in oral cancer*	*Chromosome localization*	*Oral cancer association*	*References*
D00408	Cytochrome P450, subfamily IIIA polypeptide 7	See comment 1	ND		
D13666	Osteoblast specific factor 2	Up in cancer	13		
D42047	KIAA0089	Down in cancer	3		
D43968	Runt-related transcription factor 1	Up in cancer	21q22.3		
D84276	CD38 antigen (p45)	Up in cancer	4p15		
HG3494–HT3688	Nuclear protein IL6	Up in cancer	ND	Yes	(Ondrey *et al*., 1999)
J03258	Vitamin D receptor	Down in cancer	12q12-q14	Yes	(Lotan, 1997)
J03473	ADP-ribosyltransferase	Up in cancer	1q41-q42		
J03909	Interferon, gamma inducible protein 30	Up in cancer	19p13.1		
J04080	Complement component 1, s subcomponent	Up in cancer	12p13		
M14200	Diazepam binding inhibitor	Down in cancer	2q12-q21		
M15661	Ribosomal protein L36a	Down in cancer	14		
M34309	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog3	Down in cancer	12q13	Yes	(Shintani *et al*., 1995)
M69177	Monoamine oxidase B	Down in cancer	Xp11.4-p11.3		
M81182	ATP-binding cassette, subfamily D	Down in cancer	1p22-p21		
S45630	Crystallin, alpha B	Down in cancer	11q22.3-q23.1		
U18934	TYRO3 protein tyrosine kinase	Down in cancer	15q15.1-q21.1		
U34252	Aldehyde dehydrogenase 9	Down in cancer	1q22-q23		
U46689	Aldehyde dehydrogenase 10	Down in cancer	17p11.2		
U56814	Deoxyribonuclease I-like 3	Down in cancer	3p21.1-p14.3		
X12451	Cathepsin L	Up in cancer	9q21-q22	Yes	(Kawamata *et al*., 1997; Strojan *et al*., 2000)
X54489	GRO1 oncogene (MGSA)	Up in cancer	4q21	Yes	(Loukinova *et al*., 2000)
X76029	Neuromedin U	Down in cancer	4q12		
X78932	Zinc finger protein 273	Down in cancer	ND		
X87241	FAT tumor suppressor	Up in cancer	4q34-q35		
Y00815	Protein tyrosine phosphatase, receptor, type F	Down in cancer	1q34	Yes	
Z29083	5T4 oncofetal trophoblast glycoprotein	Up in cancer	6		

Table 5 Summarization of the key demographic data of the five cases of human oral cancers used in the study

Identifier	*Gender*	*Histopathology*	*Age*	*Smoking history*	*Alcohol consumption*
1A, 1B	F	SCC, moderately differentiated	80	Not known	Not known
2A, 2B	M	SCC, well differentiated	61	2ppd/15 years	2 shots/day
3A, 3B	M	SCC, well differentiated	68	1-2ppd/40 years	3 beers/day
4A, 4B	M	SCC, moderately differentiated	75	1-2 cig./day	2 drinks/day
5A, 5B	F	SCC, moderate-poorly differentiated	60	40pk/year	Heavy

SCC = squamous cell carcinoma; ppd = pack per day; pk = pack; cig = (cigarette or cigar)

Real time quantitative PCR (RT–QPCR)

The cDNA product of the reverse transcription was used as the template for the RT–QPCR. For the RT–QPCR reaction we used iCycler IQ™ Real Time PCR detection system (Bio-Rad, Hercules, CA, USA) with TaqMan specific probes and primers for Cathepsin, and SYBR® Green buffer and reagents (Perkin Elmer/Applied Biosystems Foster City, CA, USA) for Urokinase Plasminogen Activator and Collagenase I (Heid *et al*., 1996). For designing the specific primers and probes we used PE/ABD Primer Express software as well as MacVector. Primer sequences used are: Collagenase forward: 5′-ACACGGAACCCCAAGGACA-3′; Collagenase Reverse: 5′-GTTTTGTTGCCGGTGGTTTT-3′; UPA forward: 5′-GCACCATCAAACAAACCCCCTTAC-3′; UPA reverse: 5′-CAGACAGAAAAACCCCTGCCTG-3′; Cathepsin L forward: 5′-CAGTGTGGTTCTTGTTGGG-CT-3′; Cathepsin L reverse: 5′-CTTGAGGCCCAGAGCA-GTCTA-3′. The final PCR products were run on 2% minigel to ensure single product amplification during the PCR assay.

Acknowledgments
The work is supported by the National Institute of Dental and Craniofacial Research (NIDCR) grants P01 DE12467 (DTW Wong), P30 DE11814 (DTW Wong), R29 DE11983 (R Todd), Harvard University William F. Milton Fund (R Todd), Oral & Maxillofacial Surgery Foundation (H Ohyama and R Todd), and China Site Key Basic Research Program Grant G199805123 (X Zhang). H Ohyama is supported by the Scholar in Medicine Fellowship, Harvard Medical School.

References

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM. (2000). *Nature,* **403,** 503–511.

Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D and Levine AJ. (1999). *Proc. Natl. Acad. Sci. USA,* **96,** 6745–6750.

Fujii R, Hosoya M, Fukusumi S, Kawamata Y, Habata Y, Hinuma S, Onda H, Nishimura O and Fujino M. (2000). *J. Biol. Chem.,* **275,** 21068–21074.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. (1999). *Science,* **286,** 531–537.

Heid CA, Stevens J, Livak KJ and Williams PM. (1996). *Genome Res.,* **6,** 986–994.

Janot F, Massaad L, Ribrag V, de Waziers I, Beaune PH, Luboinski B, Parise Jr O, Gouyette A and Chabot GG. (1993). *Carcinogenesis,* **14,** 1279–1283.

Jourenkova-Mironova N, Voho A, Bouchardy C, Wikman H, Dayer P, Benhamou S and Hirvonen A. (1999). *Int. J. Cancer,* **81,** 44–48.

Junien C, Huerre C and Rethore MO. (1983). *Am. J. Hum. Genet.,* **35,** 584–591.

Katoh T, Kaneko S, Kohshi K, Munaka M, Kitagawa K, Kunugita N, Ikemura K and Kawamoto T. (1999). *Int. J. Cancer,* **83,** 606–609.

Kawamata H, Nakashiro K, Uchida D, Harada K, Yoshida H and Sato M. (1997). *Int. J. Cancer,* **70,** 120–127.

Kimura Y, Fujieda S, Takabayashi T, Tanaka T, Sugimoto C and Saito H. (2000). *Cancer Lett.,* **155,** 163–168.

Kurokawa H, Sakimoto M, Yamashita Y, Murata T and Kajiyama M. (1998). *Fukuoka Igaku Zasshi,* **89,** 321–327.

Leethanakul C, Patel V, Gillespie J, Pallente M, Ensley JF, Koontongkaew S, Liotta LA, Emmert-Buck M and Gutkind JS. (2000a). *Oncogene,* **19,** 3220–3224.

Leethanakul C, Patel V, Gillespie J, Shillitoe E, Kellman RM, Ensley JF, Limwongse V, Emmert-Buck MR, Krizman DB and Gutkind JS. (2000b). *Oral Oncol.,* **36,** 474–483.

Lin LM and Chen YK. (1991). *J. Oral Pathol. Med.,* **20,** 479–485.

Liu CM, Sheen TS, Ko JY and Shun CT. (1999). *Br. J. Cancer,* **79,** 360–362.

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H and Brown EL. (1996). *Nat. Biotechnol.,* **14,** 1675–1680.

Lotan R. (1997). *Environ. Health Perspect,* **105** (Suppl. 4), 985–988.

Loukinova E, Dong G, Enamorado-Ayalya I, Thomas GR, Chen Z, Schreiber H and Van Waes C. (2000). *Oncogene,* **19,** 3477–3486.

Magary SP, Ryan MW, Tarnuzzer RW and Kornberg L. (2000). *Otolaryngol Head Neck Surg.,* **122,** 712–716.

Mahadevappa M and Warrington JA. (1999). *Nat. Biotechnol.,* **17,** 1134–1136.

Mighell AJ, Thompson J, Hume WJ, Markham AF and Robinson PA. (1997). *Oral Oncol.,* **33,** 155–162.

Muramatsu H, Kogawa K, Tanaka M, Okumura K, Nishihori Y, Koike K, Kuga T and Niitsu Y. (1995). *Cancer Res.,* **55,** 6210–6214.

Murray GI, Shaw D, Weaver RJ, McKay JA, Ewen SW, Melvin WT and Burke MD. (1994). *Gut,* **35,** 599–603.

Ohyama H, Zhang X, Kohno Y, Alevizos I, Posner M, Wong DT and Todd R. (2000). *Biotechniques,* **29,** 530–536.

Ondrey FG, Dong G, Sunwoo J, Chen Z, Wolf JS, Crowl-Bancroft CV, Mukaida N and Van Waes C. (1999). *Mol. Carcinog.,* **26,** 119–129.

Park JY, Muscat JE, Ren Q, Schantz SP, Harwick RD, Stern JC, Pike V, Richie Jr JP and Lazarus P. (1997). *Cancer Epidemiol. Biomarkers Prev.,* **6,** 791–797.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D. (2000). *Nature,* **406,** 747–752.

Porte H, Triboulet JP, Kotelevets L, Carrat F, Prevot S, Nordlinger B, DiGioia Y, Wurtz A, Comoglio P, Gespach C and Chastre E. (1998). *Clin. Cancer Res.,* **4,** 1375–1382.

Raddatz R, Wilson AE, Artymyshyn R, Bonini JA, Borowsky B, Boteju LW, Zhou S, Kouranova EV, Nagorny R, Guevarra MS, Dai M, Lerman GS, Vaysse PJ, Branchek TA, Gerald C, Forray C and Adham N. (2000). *J. Biol. Chem.,* **275,** 32452–32459.

Redfern CH, Degtyarev MY, Kwa AT, Salomonis N, Cotte N, Nanevicz T, Fidelman N, Desai K, Vranizan K, Lee EK, Coward P, Shah N, Warrington JA, Fishman GI, Bernstein D, Baker AJ and Conklin BR. (2000). *Proc. Natl. Acad. Sci. USA,* **97,** 4826–4831.

Sato M, Sato T, Izumo T and Amagasa T. (1999). *Carcinogenesis,* **20,** 1927–1931.

Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson Jr JR and Elkahloun AG. (1999). *Cancer Res.,* **59,** 5656–5661.

Shillitoe EJ, May M, Patel V, Lethanakul C, Ensley JF, Strausberg RL and Gutkind JS. (2000). *Oral. Oncol.,* **36,** 8–16.

Shin DM, Hittelman WN and Hong WK. (1994). *Cancer Epidemiol. Biomarkers Prev.,* **3,** 697–709.

Shintani S, Funayama T, Yoshihama Y, Alcalde RE and Matsumura T. (1995). *Cancer Lett.,* **95,** 79–83.

Strojan P, Budihna M, Smid L, Svetic B, Vrhovec I, Kos J and Skrk J. (2000). *Clin. Cancer Res.,* **6,** 1052–1062.

Suo Z, Holm R and Nesland JM. (1993). *Histopathology,* **23,** 45–54.

Szekeres PG, Muir AI, Spinage LD, Miller JE, Butler SI, Smith A, Rennie GI, Murdock PR, Fitzgerald LR, Wu H, McMillan LJ, Guerrera S, Vawter L, Elshourbagy NA, Mooney JL, Bergsma DJ, Wilson S and Chambers JK. (2000). *J. Biol. Chem.,* **275,** 20247–20250.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES and Golub TR. (1999). *Proc. Natl. Acad. Sci. USA,* **96,** 2907–2912.

Viaene AI and Baert JH. (1995). *Histochem. J.,* **27,** 69–78.

von Biberstein SE, Spiro JD, Lindquist R and Kreutzer DL. (1996). *Arch. Otolaryngol. Head Neck Surg.,* **122,** 751–759.

Warrington JA, Dee S and Trulson M. (2000a). *Microarray Biochip Technology, Vol. 6*. Schena, M. (ed.). Eaton Publishing, pp. 119–148.

Warrington JA, Nair A, Mahadevappa M and Tsyganskaya M. (2000b). *Physical Genomics,* **2,** 143–147.

Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA and Hampton GM. (2001). *Proc. Natl. Acad. Sci. USA,* **98,** 1176–1181.

Wodicka L, Dong H, Mittmann M, Ho MH and Lockhart DJ. (1997). *Nat. Biotechnol.,* **15,** 1359–1367.

Yeudall WA, Jakus J, Ensley JF and Robbins KC. (1997). *Mol. Carcinog.,* **18,** 89–96.