Gene Expression Levels In Different Stages of Progression in Oral Squamous Cell Carcinoma

Winston Patrick Kuo, DDS, MS[1,2,3,4], Tor-Kristian Jenssen, MSc[5,6], Peter J. Park, PhD[2], Mark W. Lingen, DDS, PhD[7], Rifat Hasina, DDS, PhD[7], Lucila Ohno-Machado, MD,PhD[1,4]

[1]Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School
[2]Children's Hospital Informatics Program, Harvard Medical School
[3]Harvard School of Dental Medicine
[4]Division of Health Sciences and Technology, Massachusetts Institute of Technology
[5]Department of Computer and Information Science, Norwegian University of Science and Technology
[6]Department of Tumor Biology, The Norwegian Radium Hospital
[7]Head and Neck Cancer Research Program, Cardinal Bernardin Cancer Center, Loyola University School of Medicine

ABSTRACT

Oral squamous cell carcinoma (OSCC) is one of the most common cancer types worldwide. The prognosis for patients with this disease is generally poor and little is known about its progression. Gene expression studies may provide important insights to the molecular mechanisms of this disease. We analyzed gene expression data from a small panel of patients diagnosed with OSCC. Even with only 13 patient samples we were able to find genes with significant differences in expression levels between normal, dysplasia, and cancer samples. The largest differences in expression were generally found between normal and cancer samples, but significant differences were also found for several genes between dysplasia and the other two sample types. We also represent the significance levels of differentially expressed genes on the chromosome domain. The genes and genetic features we examine are potentially important factors on the molecular level in the progression of OSCC.

INTRODUCTION

Oral squamous cell carcinoma (OSCC) represents approximately 3% of all cancers [1]. Pathologists and clinicians in charge of the management of OSCC are facing two major problems, namely the heterogeneity of the disease and the lack of conventional histological and clinical features that reliably predict the progression of the disease. The prognosis of OSCC patients is still poor, as little is known about the molecular mechanisms responsible for this malignancy. New prognostic and predictive factors are desired to allow for individualized treatment based on the characteristics of each patient.

Recent advances in technology have made it possible to develop molecular portraits of diseases on a global scale. Combined with laser-capture-microdissection (LCM) [2], DNA microarrays [3] allow for accurate measurement of the mRNA expression levels of thousands of genes simultaneously, and several studies using complementary DNA (cDNA) microarrays have already identified important genes comparing normal and head and neck squamous cell carcinomas [4, 5]. This technology should be equally useful for a comprehensive molecular characterization of OSCC. In particular, comprehensive analyses of gene expression can be used for classification of different stages of this disease, such as discrimination between normal, dysplasia, and malignant cells.

Incorporating microarray data with other types of genetic features is also an important task. For example, malignant tumors have been shown to contain chromosomal aberrations, and the pattern of abnormalities varies greatly among malignancies. However, as a rule, the total number of chromosomal aberrations is roughly proportional to the risk of metastasis [6]. Genetic aberrations in OSCC have been investigated as markers of disease progression including amplifications of 3q, 8q, 9q, 20q, 5p, 7p, and 11q13 [7].

Here we report the use of cDNA microarray filters to investigate differentially expressed genes through OSCC progression. After identifying those genes and expressed sequence tags (ESTs) that best discriminate between normal, dysplasia, and cancer samples, we map them to their respective chromosomal positions, and plot the p-values representing their significance in differentiating the groups. Our results demonstrate that gene expression profiling combined with such genetic features as chromosomal locations can be used to provide valuable information concerning the disease.

MATERIALS AND METHODS

The samples for analysis were classified according to histology from biopsy specimens into normal, dysplasia, and cancer. A total of 13 samples were used for this study. Normal (5 samples), dysplasia (3 samples), and cancer (5 samples) cells were specifically targeted and picked up from snap-frozen surgical tissue using LCM. Total RNA was extracted, amplified and reverse-transcribed into anti-sense RNA, labeled with p^{33} and hybridized on complementary DNA microarrays filters containing 4,132 genes.

To determine those genes that were differentially expressed in the three groups, we used the Kruskal-Wallis test. This is a nonparametric method for testing the null hypothesis that the mean is the same across all the groups. We chose a nonparametric approach since estimating the within group variances with such a small number of samples is unreliable and since the normality (or log-normality) assumption is questionable. Once we find that a gene is differentially expressed across three groups, we also used the Wilcoxon rank-sum test for pairwise comparisons. (The Kruskal-Wallis test reduces to the Wilcoxon test for two groups) With this nonparametric approach, there may be some loss of information, but we gain a valid testing procedure. We note that due to the small sample size, the p-value for a gene cannot be too low, even when it discriminates the groups perfectly. As identifiers for the genes, we used gene symbols or GenBank accession numbers when the former was not available.

Chromosome positions for chromosome-mapped ESTs from the Human Genome Working Draft (August 6, 2001 freeze; http://genome.ucsc.edu) were downloaded. The p-values from the Kruskal-Wallis test were obtained and mapped onto their respective positions on all 23 chromosomes. Chromosome positions for all mapped cytobands from the same version of the Human Genome Draft were also downloaded. Plots with chromosome-mapped negative logarithms (base 10) of p-values and cytobands were created in MATLAB (The Math Works, Natick, MA).

The negative logarithms of the p-values of all the scored genes were entered into a literature cluster analysis program called PubGene (www.pubgene.org), developed by Jenssen at al. [8]. The software returned clusters of genes with prognostic value, pointing to possible pathways (consisting of co-regulated genes) associated with OSCC.

RESULTS

The results of the Kruskal-Wallis tests were obtained and ranked. This procedure identified the genes with most significant differences in expression when all three groups were considered. The vast majority of the genes investigated showed lower expression levels in normal samples than in cancer samples, possibly reflecting a composition of the gene set biased towards oncogenes. Expression levels in dysplasia were generally intermediate and mostly closer to expression in normal samples than cancer samples. Log-transformed p-values from the Kruskal-Wallis tests were subsequently mapped to chromosomal positions (Figure 1). We have examined all of the chromosomes but we only showed one as an example. In the plot of chromosome 1, cytoband region 1q21-q23.2 appears to be a possible amplified region because many genes and ESTs were significantly up-regulated in cancer.

Table 1. List of significant genes found in Chromosome 1. (p <0.05)

Accession number	Gene name	Cytoband location
AA488073	MUC1	1q21
AA002086	CD1C	1q21-q23
AA479058	THPO	1q31.1
AA678021	SNRPE	1q32
AA485748	FMOD	1q32.1
AA458779	HMGCL	1pter-p33
N95761	FUCA1	1p34
AA424344	UROD	1p34
AA150416	TNFR2	1p36.3-p36.2

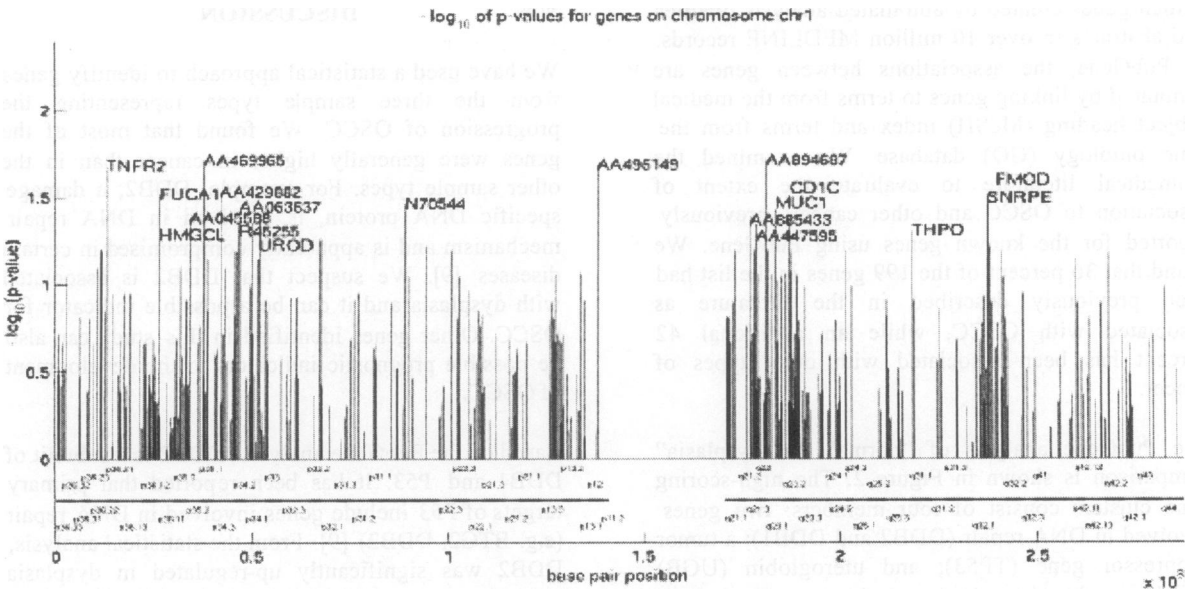

Figure 1. Chromosome plot of chromosome 1 (log-transformed *p*-values from the Kruskal-Wallis test). Each bar is located at the start of the EST used to detect expression of a given gene. The height of the bar is proportional to the log-transformed *p*-value. At the bottom cytobands (to-scale) are shown. Cytobands with positive staining are shown in red and bands with negative staining are colored in black.

For example, MUC1 has been reported to be associated with OSCC. Other significant up-regulated known genes on chromosome 1 are listed in Table 1.

As the first part of investigating changes in expression levels that could be related to cancer progression, we compared the normal and dysplasia groups. We identified seven genes and ESTs from the normal and dysplasia comparison that were significant in the rank-sum test ($p < 0.05$). All of these genes displayed higher expression in dysplasia than normal (Table 2). DDB2 [9], CYP21 [10] and CPS1 [11] have been reported to be associated with cancer.

Table 2. Normal and Dysplasia sample comparisons. List of significant ESTs and genes using the Rank-Sum Test. DDB2, CYP21 and CPS1 have been associated with cancer ($p < 0.05$).

Accession number	Gene name
AA469965	-
AA400187	-
H77597	-
AA410404	DDB2
T58430	CYP21
T61078	CPS1
H88329	-

Table 3 lists the ESTs and genes from the dysplasia and cancer comparison where we found gene expression levels higher in cancer than dysplasia: significant according to the rank-sum test ($p < 0.05$). Two of the five genes listed, RPL35A [12] and PAX6 [13], have been reported to be associated with cancer.

Table 3. Dysplasia and cancer sample comparisons. List of significant ESTs and genes using the Rank-Sum Test. Of the five known genes, RPL35A and PAX6 have been associated with cancer ($p < 0.05$).

Accession number	Gene name
AA878899	GLB1
AA634360	-
AA888182	RPS4X
AA863383	-
AA400893	-
AA873351	RPL35A
H73914	-
AA481543	PEPD
R95962	PAX6

The normal and cancer comparison had 365 ESTs and 199 known genes whose expressions were significantly up-regulated in cancer samples. We used PubGene to determine whether significant genes co-occurred in the literature. PubGene is a gene-to-gene co-citation network for the 13,712 named

human genes created by automated analysis of titles and abstracts in over 10 million MEDLINE records. In PubGene, the associations between genes are annotated by linking genes to terms from the medical subject heading (MeSH) index and terms from the gene ontology (GO) database. We examined the biomedical literature to evaluate the extent of association to OSCC and other cancers previously reported for the known genes using PubGene. We found that 36 percent of the 199 genes in the list had been previously described in the literature as associated with OSCC, while an additional 42 percent has been associated with other types of cancer.

The PubGene analysis of "normal and dysplasia" comparison is shown in Figure 2. The high-scoring gene clusters consist of four members: two genes involved in DNA repair (DDB2 and DDB1); a tumor suppressor gene (TP53); and uteroglobin (UGB) which was found to play a role in cancer [14]. The literature gene network confirms the relationship of DDB2 to DDB1 and TP53 as well as TTD and ERCC2 and ERCC3, all of which are involved in cancer.

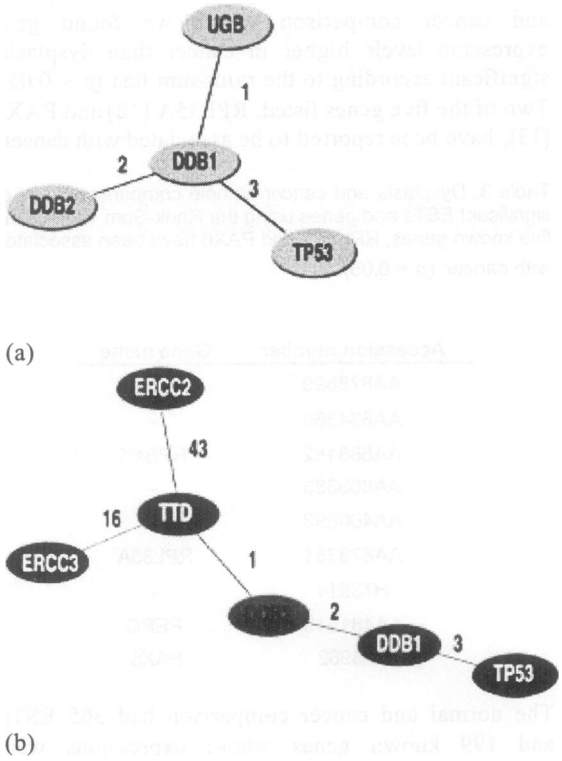

Figure 2. PubGene analysis of "normal and dysplasia" comparison. The top-scoring networks from search by gene-values contained four genes shown in (a). Literature gene network is shown in (b).

DISCUSSION

We have used a statistical approach to identify genes from the three sample types representing the progression of OSCC. We found that most of the genes were generally higher in cancer than in the other sample types. For example, DDB2, a damage specific DNA protein, is involved in DNA repair mechanism and is apparently compromised in certain diseases [9]. We suspect that DDB2 is associated with dysplasia and it can be a possible indicator for OSCC. Other genes identified in this study can also be possible prognostic indicators to the development of OSCC.

Based on the PubGene analysis, DDB2 is a subunit of DDB1 and P53. It has been reported that primary targets of P53 include genes involved in DNA repair (e.g. BTG2, DDB2) [9]. From the statistical analysis, DDB2 was significantly up-regulated in dysplasia compared to normal and in cancer compared to dysplasia. In contrast, DDB1 was significantly up-regulated only in cancer and dysplasia comparison. From this, we can speculate that DDB2 may be part of the P53 pathway.

Plotting significant genes on the chromosomes was useful for visualization of possible amplified regions. In chromosome 1, for example, the region near MUC1 has low p-values that may be associated with a possible amplified area. This area, 1q21-q23.1 has been associated with a gene that has been reported to be a possible diagnostic marker for prediction of premalignant and malignant lesions in the oral cavity [15].

Constructing a signaling pathway through the identification of the specific genes and the sequence in which they appear in the transformation of a normal to cancer can be beneficial in understanding OSCC. The results from this study are preliminary and further studies are necessary, but we believe these findings are promising. Future work will include more tissue samples and biological validation by RT-PCR.

REFERENCES

[1] Greenlee RT, Hill-Harmon MB, Murray T, Thun M: Cancer statistics, 2001. CA Cancer J Clin 2001;*51*:15-36.

[2] Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA: Laser capture

microdissection. Science 1996;*274*:998-1001.

[3] Schena M: Genome analysis with gene expression microarrays. Bioessays 1996;*18*:427-431.

[4] Leethanakul C, Patel V, Gillespie J, Pallente M, Ensley JF, Koontongkaew S, Liotta LA, Emmert-Buck M, Gutkind JS: Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. Oncogene 2000;*19*:3220-3224.

[5] Villaret DB, Wang T, Dillon D, Xu J, Sivam D, Cheever MA, Reed SG: Identification of genes overexpressed in head and neck squamous cell carcinoma using a combination of complementary DNA subtraction and microarray analysis. Laryngoscope 2000;*110*:374-381.

[6] Mitelman F, Johansson B, Mandahl N, Mertens F: Clinical significance of cytogenetic findings in solid tumors. Cancer Genet Cytogenet 1997;*95*:1-8.

[7] Gollin SM: Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease. Head Neck 2001;*23*:238-253.

[8] Jenssen TK, Laegreid A, Komorowski J, Hovig E: A literature network of human genes for high-throughput analysis of gene expression. Nat Genet 2001;*28*:21-28.

[9] Kannan K, Amariglio N, Rechavi G, Jakob-Hirsch J, Kela I, Kaminski N, Getz G, Domany E, Givol D: DNA microarrays identification of primary and secondary target genes regulated by p53. Oncogene 2001;*20*:2225-2234.

[10] Enberg U, Farnebo LO, Wedell A, Grondal S, Thoren M, Grimelius L, Kjellman M, Backdahl M, Hamberger B: In vitro release of aldosterone and cortisol in human adrenal adenomas correlates to mRNA expression of steroidogenic enzymes for genes CYP11B2 and CYP17. World J Surg 2001;*25*:957-966.

[11] Liu TH, Li DC, Gu CF, Ye SF: Carbamyl phosphate synthetase I. A novel marker for gastric carcinoma. Chin Med J (Engl) 1989;*102*:630-638.

[12] Pappas IS, Vizirianakis IS, Tsiftsoglou AS: Cloning, sequencing and expression of a cDNA encoding the mouse L35a ribosomal protein during differentiation of murine erythroleukemia (MEL) cells. Cell Biol Int 2001;*25*:629-634.

[13] Salem CE, Markl ID, Bender CM, Gonzales FA, Jones PA, Liang G: PAX6 methylation and ectopic expression in human tumor cells. Int J Cancer 2000;*87*:179-185.

[14] Kundu GC, Zhang Z, Mantile-Selvaggi G, Mandal A, Yuan CJ, Mukherjee AB: Uteroglobin binding proteins: regulation of cellular motility and invasion in normal and cancer cells. Ann N Y Acad Sci 2000;*923*:234-248.

[15] Nitta T, Sugihara K, Tsuyama S, Murata F: Immunohistochemical study of MUC1 mucin in premalignant oral lesions and oral squamous cell carcinoma: association with disease progression, mode of invasion, and lymph node metastasis. Cancer 2000;*88*:245-254.

Acknowledgments

Winston Patrick Kuo was supported by the National Library of Medicine Medical Informatics Training Grant number 5T15 LM07092–07. The HST Taplin faculty development award also supported this project. This work was supported in part by the NIH grants: DE12322 (MWL), DE00470 (MWL).