Identification of genes and molecular pathways involved in the progression of premalignant oral epithelia

Abhijit G. Banerjee,1 Indraneel Bhattacharyya,2 and Jamboor K. Vishwanatha3

1Department of Oral Biology, Faculty of Dentistry, Health Sciences Center, University of Manitoba, Winnipeg, Manitoba, Canada; 2Department of Oral and Maxillofacial Surgery and Oral Diagnostic Sciences, College of Dentistry, University of Florida, Gainesville, Florida; and 3Department of Molecular Biology and Immunology, University of North Texas Health Science Center, Fort Worth, Texas

Abstract
An early interventional effort in oral premalignancy requires novel molecular targets and diagnostic biomarkers to delay or reverse incidences of malignant progression. Microarray-based transcriptional profiling in disease states provides global insight into the causal biomolecular processes and novel pathways involved. In this study, we investigated transcript profiles in precancerous oral lesions to identify nearly 1,700 genes as significantly overexpressed or underexpressed and a primarily affected metabolic pathway that may be responsible for irreversible transition to progressive stages of oral cancer. For the first time, we show a convergence of several genes and pathways known for their oncogenic capabilities, in progression of premalignant oral epithelial tissues. This study consequently provides a molecular basis for persistent proinflammatory conditions in oral premalignant tissues. We found that lipocalin-type prostaglandin D2 synthase (PTGDS), a key enzyme in the arachidonic acid metabolism pathway, as repressed in premalignant stages. We show the protective role of these enzyme-derived metabolites in inhibiting cell proliferation using an in vitro oral cancer progression model. We have also confirmed the overexpression of two invasion-related biomarkers, psoriasin (PSOR1) and versican (CSPG2), in oral premalignant and malignant archival tissues. Our results clearly indicate that pharmacologic intervention with anti-inflammatory prostaglandin D2-like analogues may help prevent or delay oral epithelial carcinogenesis because of metabolic restoration of a negative feedback regulatory loop through its several cognate receptors or target molecules. Further studies directed toward a multitude of possible protective mechanisms of this lipocalin-type enzyme or its products in oral cancer progression are warranted.

Introduction
Oral precancer is an ideal stage for preventive measures because morbidity associated with malignant transformation is high. The survival rate for patients with oral cancer at 5 years postdiagnosis is very low (~55%) despite improved methods of diagnosis, accessibility, and treatment (1). Intraepithelial neoplasia of the oral cavity, such as oral leukoplakia or erythroleukoplakia, occurs in 1% to 10% of the adult population in the Western world. Almost 17.5% of these progress to malignancy by 8 years, but the rate of malignant conversion for dysplasia, an advanced premalignant stage, at initial biopsy jumps up to 36.4% (2, 3). The current optimal treatment, including surgical removal of the lesion, is inadequate due to wide exposure of carcinogens over a large mucosal surface, resulting in “field carcinization” (4). Hence, there is a basic unmet medical need to develop better diagnostic measures and therapeutic targets.

Transcriptional profiling in premalignant stages is crucial for revealing molecular and biological changes that result because of oral epithelial transformation. We investigated gene expression profiles of the human oral precancerous tissue specimens and compared it to a known less aggressive variant of premalignant (verrucous hyperplasia) and malignant oral tissue (verrucous carcinoma) that seldom metastasizes. We hypothesized that this strategy would potentially identify differentially modulated genes that might be responsible for more aggressive behavior and imminent malignant potential. Such study may also provide novel diagnostic or therapeutic targets at premalignant stages that can potentially reverse, reduce, or at least delay the rates of malignant conversion in future. The development of a molecular diagnostic test based on the identified signature expression profile will help screen and identify a population or individuals at risk through minimal invasive procedures, such as periodic brush biopsies followed by multiplex reverse transcription-PCR analysis. Early intervention strategies can use such expression signatures as surrogate end point biomarkers of efficacy. In our study, we have also used cellular assays on an in vitro oral cancer progression model to validate a novel molecular target in a prominently perturbed pathway.
involved in protection from early carcinogenic events. Therefore, this study may provide a molecular basis and innovative approaches for clinical intervention to prevent oral cancer progression in the near future.

Materials and Methods

Tissue Specimens and Cell Lines

The oral biopsy specimens (n = 12) from the human subjects were collected as per the approved guidelines of the institutional review board (University of Nebraska Medical Center, Omaha, NE) with written informed consent of the patients attending regular oral health clinics and described in Results. Similarly, paraffin-embedded oral tissues (n = 20), previously fixed in 10% neutral buffered formalin, were obtained for immunohistochemical analysis. Cell lines of an oral cancer progression model (n = 4) were maintained as previously described (5).

Experimental Design

To investigate gene expression profiles, paired premalignant and normal adjacent tissues (n = 6) of oral cavity were compared within or outside similar diagnostic and pathologic description provided by an oral pathologist (I. Bhattacharyya) involved in this study. As per known clinical behavior of oral carcinomas, the less invasive sample of a paired premalignant oral tissue—verrucous hyperplasia and a seldom metastatic, yet malignant, variant called verrucous carcinoma (n = 2)—was included as a biological variant control. This control pair of tissues is hypothesized to help identify sets of differentially expressed genes in oral dysplasia that relates to a rather more invasive phenotype and imminent progression to malignancy. The identified set of genes were further confirmed with similar gene expression profiles of additional unpaired oral premalignant biopsy specimens (n = 2) to arrive at a list of significantly overexpressed or underexpressed genes.

Transcriptional Profiling by Genechip Microarray

Total RNA was isolated from the frozen biopsy specimens by embedding in OCT compound (VWR Scientific, San Diego, CA), cryosectioning into thin slices of 5 μm, followed by extraction of the tissues in TRIzol reagent (Invitrogen, Carlsbad, CA) and secondary cleanup using the RNAeasy mini kit (Qiagen, Valencia, CA). The quality and quantity of total RNA were characterized using the RNA NanoChip 6000 kit on a Bioanalyzer 2100 (Agilent Technologies, Foster City, CA). Total RNA (100 ng), meeting desired criteria of purity and integrity from each specimen, was used as a starting material for biotin-labeled cRNA probe synthesis using the Enzo BioArray high-yield RNA transcript labeling kit (Enzo, New York, NY) and modified small-sample amplification protocol (6). Fragmented cRNA (15 μg) was hybridized to human 133A Genechip Oligonucleotide microarray (Affymetrix, Inc., Santa Clara, CA). Posthybridization, the arrays were washed, stained with streptavidin-phycocerythrin fluorescent conjugates using Affymetrix Micro-Fluidics Station 400, and scanned for digital output intensity files (*.cel) using a Hewlett-Packard Gene Array Scanner.

Data Analysis and Statistical Validation

Each intensity output file (*.cel) for a hybridized tissue specimen was initially analyzed for probe hybridization quality control parameters, such as average background, target intensity, and raw noise (Q) values, to ascertain the appropriate gene expression profile analysis using the Microarray Suite (MAS 5.0) software (Affymetrix). To identify a differential gene expression pattern, we used the normal tissue expression profile as baseline for each paired premalignant tissue specimens in a comparison analysis using the Data Mining Tool (DMT-3.0) software (Affymetrix). The MAS 5.0 software uses the one-sided Wilcoxon’s signed rank test as the statistical method to generate a detection P value based on a discrimination score (R) derived using the equation $R = (PM - MM) / (PM + MM)$ for each probe pair, where PM denotes perfect match and MM denotes mismatch. Sixteen such probe pairs in the array represent the total length of each gene transcript. The discrimination score is a basic property of a probe pair that describes its ability to detect its intended target. It measures the target-specific intensity difference of the probe pair (PM − MM) relative to its overall hybridization intensity (PM + MM). Only if the detection P value is equal or below the default threshold value of 0.04 it is called present (P). The data analysis manual (PDF file) downloadable from the Affymetrix website provides further details of the statistical algorithms and is only mentioned here in brief. In a comparison analysis, two samples, hybridized to two Genechip probe arrays of the same type, are compared against each other to detect and quantify changes in gene expression. A change algorithm generates a Change P value again based on the Wilcoxon’s signed rank test and an associated Change. A second algorithm produces a quantitative estimate of the change in gene expression in the form of signal log ratio. The increase (I), decrease (D), or No (N) Change calls between two array sets (such as normal versus diseased sample) is based on such comparison analysis depending on the change P value. The P value ranges in scale from 0.0 to 1.0 and provides a measure of the likelihood of change and direction. Values close to 0.0 indicate likelihood for an increase in transcript expression level in the experiment array compared with the baseline, whereas values close to 1.0 indicate likelihood for a decrease in transcript expression level. The signal log ratio algorithm estimates the magnitude and direction of change of a transcript when two arrays are compared (experiment versus baseline). It is calculated by comparing each probe pair on the experiment array to the corresponding probe pair on the baseline array. This strategy cancels out differences due to different probe binding coefficients and is, therefore, more accurate than a single-array analysis. The log scale used is base 2, making it intuitive to interpret the signal log ratios in terms of multiples of 2. Thus, a signal log ratio of 1 indicates an increase of the transcript level by 2-fold, and −1.0 indicates a decrease by 2-fold. As with signal, computation of this number uses a one-step Tukey’s biweight method by taking a mean of the log ratios
of of probe pair intensities across the two arrays. The Tukey's biweight method gives an estimate of the amount of variation in the data exactly as SD measures the amount of variation for an average. The fold changes in gene expression can be calculated using signal log ratios using the following equation: fold change = \( \frac{2^{signal \ log \ ratio}}{C0} \) for signal log ratio >0 and \( (-1)^{signal \ log \ ratio} \) for signal log ratio <0. A signal log ratio of zero means no change.

Gene Clustering and Significant Gene Profiles
Hierarchical clustering analysis was done on the expression data set obtained upon MAS 5.0 analysis, with the Gene Cluster program, and visualized using the TreeView program\(^4\) to group genes according to their similarities in the expression levels thought to be linked in the disease process using an unsupervised clustering algorithm-based software (7). The set of probe IDs in a particular cluster node depicts values for coefficient of correlation (\( r^2 \)) and provides significant gene expression profiles. Such tight cluster of altered genes were queried on the NetAffx analysis center available at the Affymetrix website\(^5\) (8) and a downloadable program from the NIH, called EASE\(^6\), which provides Bonferroni’s correction and Fisher’s exact probability values for the microarray expression data sets besides annotations for the significant gene lists.

Quantitative Real-time PCR
The total RNA from all of the paired and unpaired frozen biopsy specimens (\( n = 12 \)) and cell lines of an oral progression model (\( n = 4 \)) were used in the real-time PCR analysis. We analyzed a subset of four genes by quantitative real-time PCR using SYBR Green I technology–based Protein Analysis PCR System (Bio-Rad, Hercules, CA). We did a melt curve analysis to determine the specificity of each amplification reaction based on GC content of an amplified DNA fragment and specific for a single amplified target. Appropriate gene-specific primers designed for the analysis and amplification conditions were as mentioned in Table 1. Relative fold changes in real-time PCR experiments were calculated using a comparative \( C_T \) (threshold cycle) equation (9) in reference to a housekeeping gene. We used statistical analysis tools of Microsoft Excel software to determine the significance values.

Protein Analysis
Immunohistochemistry was done on 5-μm-thick sections of paraffin-embedded oral premalignant biopsy specimens (\( n = 20 \)) using psoriasin (10) and versican (11) specific primary antibodies and Vectastain ABC peroxidase kits (Vector Labs, Burlingame, CA) as per protocols standardized earlier by us (12). We ran appropriate negative controls, including nonimmune serum (devoid of specific primary antibodies), in parallel to ascertain positive staining (>10% cells) in the specimens by light microscopy. We confirmed non–cross-reactivity of the immune serum on uninvolved and unrelated tissues before using it for dysplastic specimens as a standard control. We also confirmed negative staining with nonimmune serum as standard immunohistochemistry procedure on the same dysplastic specimens to rule out any nonspecific reactivity. We recorded the serial progressive specimens of dysplasia at various photographic magnifications that ascertain positive staining and localization.

Pathway Prediction Analysis
We obtained the annotations regarding involved bioprocesses, molecular function, and cellular localization using the Gene Ontology database as links from NetAffx-derived annotation tables. The significant gene clusters queried with the known components of biological pathways on the KEGG database\(^7\) and a web-based software “Ingenuity pathway analysis”\(^8\) helped identify genes within known networks. These gene networks were mapped with GenMAPP program\(^9\) to identify significant pathways involved in disease progression from premalignant stages to malignancy.

Biological Assays
We tested a protective anti-inflammatory gene as a plausible biological target in the process [i.e., lipocalin-type

---

\(^4\) Softwares available at http://rana.stanford.edu/software
\(^5\) http://www.affymetrix.com/analysis/index.affx
\(^6\) http://david.niaid.nih.gov/david/ease.htm
\(^7\) http://www.genome.ad.jp/kegg/kegg2.html
\(^8\) http://analysis.ingenuity.com
\(^9\) http://www.GenMAPP.org
prostaglandin D\(_2\) (PGD\(_2\)) synthase] using cell-based assays on the cell lines of oral cancer progression model. We observed this enzyme as repressed in dysplasia at the mRNA level. Consequently, the effect of substitution of the natural metabolite PGD\(_2\), a product of the target enzyme, was examined using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide dye reduction–based cell proliferation assays (13). The substitution of the enzyme itself by recombinant method calls for rather long-term experimental methods and may have additional implications to study the disease process. Briefly, \(1 \times 10^6\) cells/well were treated with an optimal final concentration (6.82 \(\mu\)mol/L) of PGD\(_2\) (Cayman Chemicals, Ann Arbor, MI) in the culture media and the cell growth was assessed every 24 to 96 hours in parallel with a vehicle-only (0.1% DMSO) negative control. The growth curve standardized to cell numbers (based on absorbance at 590 nm values) was plotted at each time point to assess the effect of PGD\(_2\) on the oral cell lines. We did colorimetric substrate-based assay for caspase-3 activation on similarly treated or untreated cells (Apoalert assay, BD Biosciences, Palo Alto, CA). Annexin V-FITC binding assay and cell cycle analysis of PGD\(_2\)-treated cells were done using the Telford assay method (14) followed by flow cytometry analysis on FACSCalibur instrument (BD Biosciences). We analyzed the flow cytometry data using the CellQuest software package (BD Biosciences).

**Results**

**Differential Gene Expression Profiles in Oral Pre-cancer**

In this study, we used paired premalignant and normal adjacent oral tissue biopsies \((n = 6)\) along with unpaired premalignant tissues \((n = 2)\) from patients attending oral health clinics for various premalignant oral conditions. The premalignant tissues represented various grades of oral epithelial dysplasia with hyperkeratosis. Additionally, a paired \((n = 2)\) but variant oral premalignant condition, namely verrucous hyperplasia, and its malignant but seldom metastatic counterpart, namely verrucous carcinoma, were included in the Genechip microarray comparison analysis as a control set for biological behavior. We collected the biopsy tissues under approval of the institutional review board as per their guidelines and patient informed written consent. We did human U133A, Genechip (Affymetrix) hybridization analysis experiments with amplified and labeled RNA (aRNA) from the above-mentioned biopsy samples, each starting with 100 ng of total RNA. Fluorescent scanning analysis of each Genechip array posthybridization resulted in output of raw intensity files (*.cel). The signal intensity for each probe pair data was further normalized based on signals from internal controls and background analysis values (raw noise, \(Q < 3.00\)) to a signal intensity of at least 150. The microarray data analysis shows a total differential gene expression of 5,264 up-regulated genes and 4,321 down-regulated genes based on the statistical analysis algorithms of MAS 5.0 software (Affymetrix). These represent ~41.6% of genes present on the array. The gene expression data from our microarray experiments was clustered for similarities in expression patterns using a hierarchical clustering method (15) across various array samples. Gene cluster heat maps (Fig. 1) includes the representative clusters of similarly expressed genes among the paired samples \((n = 8)\). We did not use the expression data sets of unpaired samples \((n = 2)\) to avoid any bias in node correlation coefficient values. Rather, we confirmed the expression trends of significant genes in these specimens to arrive at significant gene expression profiles for premalignant tissues. We have used Affymetrix probe set IDs to enable retrieval of specific gene identities and annotations directly using the NetAffx analysis tool available on the Affymetrix website and EASE software tool available from the NIH website. The expression profiles definitively classify sets of genes expressed differentially between normal and premalignant samples as red and green clusters. The intensity of expression levels is shown to vary from the mean of green \((-)\) or red \((+)\) spectrum for derived fold change values between ~9.053 and +12.042 (see Table 2). The correlation coefficient value \((r^2)\) for each representative node provides a measure of similarity in gene expression between the respective samples.

The scatter correlation graphs of signal intensities for each paired normal versus premalignant and verrucous hyperplasia versus verrucous carcinoma represents normal distribution around the line of unity for majority of the genes represented on the array (Fig. 2). The altered gene expression is represented in color (red or green) for each such comparison. A partial list of significant \((> \pm 1.5\)-fold in at least three of four such comparisons, i.e., 75%) differentially expressed genes based on comparison analysis and corresponding to positive and negative average fold change values calculated from average signal log ratios are presented (Table 2). We have identified almost 1,300 genes as significantly activated (increased) and 400 genes as significantly repressed (decreased), thus giving us a total of 1,700 genes as differentially expressed (see Supplementary Data)\(^{10}\) of initial 9,000 genes observed to be up- or down-regulated. This includes 52 expressed sequence tags or full-length clones of genes with unknown function and hypothetical proteins that could not be mapped to known gene networks. We expect these potentially novel and nonannotated genes or hypothetical proteins to be associated with malignant progression of oral dysplastic epithelia based on their similarity in expression within the cluster of genes known to be involved in essential cellular processes in normal or disease states. Functional validation for such novel genes identified through our study is necessary before appropriate biological annotation can be assigned to them.

We also analyzed the expression array data from each tissue sample for normal distribution independently using

\(^{10}\)Supplementary material for this article is available at Molecular Cancer Therapeutics Online (http://mct.aacrjournals.org/).
probability (P-P) plots and the resultant quantile derivatives (i.e., Q-Q plots) using statistical analysis tools of MATLAB software (see Supplementary Data) to ascertain validity before comparison analysis and found it to be similar to scatter correlation graphs (Fig. 2) from MAS 5.0 software. This step also helped us identify outlier populations.

**Validation of Microarray Data**

We compared the microarray-based expression data independently using quantitative real-time PCR technique based on SYBR green-1 technology, followed by amplicon identification using melting point dissociation curves. We did real-time PCR for four of the significant genes identified in our study with specifically designed primer pairs under indicated conditions (Table 1). The first-strand cDNA derived from total RNA for real-time PCR analysis were from the frozen normal and dysplastic biopsy specimens ($n = 12$), including paired normal, premalignant, verrucous hyperplasia, and verrucous carcinoma, as well as unpaired specimens previously used for microarray. In addition, we included cDNA from four cell lines of oral progression model ($n = 4$) and a prostate cancer cell line (PC3) as a positive control, whereas distilled water devoid of any cDNA served as negative control for PCR reactions. We are unaware of any cell lines that are null for expression of the four genes validated in these experiments and that could be used as additional negative controls. We calculated the comparative threshold cycle ($C_T$) values between the specific gene transcript message and a housekeeping gene message for the same unamplified RNA sample, termed as $\Delta C_T$. The $C_T$ value is similar to signal log ratios of microarray data after internal normalization (16). In our experiments, we have used the β-actin gene as a housekeeping gene control. The average fold expression change based on $2^{-\Delta \Delta C_T}$ values obtained in this experiment were plotted in comparison with $2^{signal \ log \ ratios}$ of microarray data that indicates similar expression trends and, therefore, ascertains the validity of our microarray results (Fig. 3). However, due to limitations of a PCR reaction (unequal efficiency and final yield based on starting cDNA concentration), the data between microarray and real-time PCR can never exactly

**Figure 1.** Gene clustering analysis reveals unique subset of genes differentially expressed in oral tissue specimens. A, total gene expression profile of paired oral tissue specimens represented in rows (for different genes) and in columns (for respective tissues). B, dendrograms classifying the gene expression pattern between the oral samples for a subset of increased or decreased genes. $N$, paired normal tissue; $D$, diseased premalignant oral tissue; $VH$, verrucous form of hyperplasia; $VCA$, verrucous form of carcinoma. The cluster nodes represent an increased or decreased subset of highly correlated genes identified by Affymetrix probe set IDs having a similar expression pattern with correlation coefficient ($r^2$) values close to 1.0. We used only paired specimens in clustering analysis and they appear in the dendrogram as per expression pattern—based similarity between specimens as well as genes.
be the same and, therefore, not directly comparable numerically as discussed previously (17, 18). For similar reasons, the negatively regulated genes, such as lipocalin-type PGD$_2$ synthase (PTGDS), in a real-time PCR show a negative correlation in our microarray data. We have shown a comparative analysis of real-time PCR and microarray (average fold change) data for four such significantly (at least in 75% of samples) altered genes, such as PTGDS, Table 2. Significantly regulated genes in oral premalignancy (partial list)

<table>
<thead>
<tr>
<th>Genbank accession</th>
<th>Gene ID</th>
<th>Average fold change</th>
<th>Biological function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_002260.2</td>
<td>KLRC2</td>
<td>+12.042</td>
<td>Natural killer cell immune response</td>
</tr>
<tr>
<td>AF_030514.1</td>
<td>CXCL11</td>
<td>+11.713</td>
<td>Calcium mobilization, chemotactant</td>
</tr>
<tr>
<td>NM_002963.2</td>
<td>S100A7</td>
<td>+11.713</td>
<td>Chemoattractant, cell invasion</td>
</tr>
<tr>
<td>AB046400.1</td>
<td>SCAA2</td>
<td>+8.574</td>
<td>Serine proteinase inhibitor</td>
</tr>
<tr>
<td>M21121</td>
<td>CCL5</td>
<td>+8.515</td>
<td>Chemokine, T-cell immune response</td>
</tr>
<tr>
<td>U34690.1</td>
<td>CORO1A</td>
<td>+8.225</td>
<td>Chemotaxis, cell motility</td>
</tr>
<tr>
<td>NM_002638.1</td>
<td>PI3</td>
<td>+6.635</td>
<td>Serine proteinase inhibitor</td>
</tr>
<tr>
<td>L24612.1</td>
<td>KRT6F</td>
<td>+5.352</td>
<td>Cellular hyperproliferation</td>
</tr>
<tr>
<td>AF061812.1</td>
<td>KRT16A</td>
<td>+5.098</td>
<td>Cellular hyperproliferation</td>
</tr>
<tr>
<td>AB000221.1</td>
<td>PARC</td>
<td>+4.347</td>
<td>Dendritic cell immune response</td>
</tr>
<tr>
<td>NM_002416.1</td>
<td>MIG</td>
<td>+4.199</td>
<td>T-cell trafficking, angiogenesis</td>
</tr>
<tr>
<td>NM_016184.1</td>
<td>CLEC6A</td>
<td>+4.112</td>
<td>Dendritic cell immunoreceptor with immunoreceptor tyrosine–based inhibitory motif</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_004319.1</td>
<td>TSCA6</td>
<td>+3.972</td>
<td>Dendritic cell maturation associated in cancer</td>
</tr>
<tr>
<td>NM_002160.1</td>
<td>TNC</td>
<td>+3.482</td>
<td>Growth, invasion and angiogenesis</td>
</tr>
<tr>
<td>NM_000450.1</td>
<td>SELE</td>
<td>+3.387</td>
<td>Endothelial cell adhesion, angiogenesis</td>
</tr>
<tr>
<td>U19556.1</td>
<td>SCAA2</td>
<td>+2.949</td>
<td>Serine proteinase inhibitor</td>
</tr>
<tr>
<td>NM_012128.2</td>
<td>CLCA4</td>
<td>+2.732</td>
<td>Voltage-gated chloride channel activity</td>
</tr>
<tr>
<td>D21524.1</td>
<td>OB-CDH1</td>
<td>+2.574</td>
<td>Cell-to-cell adhesion</td>
</tr>
<tr>
<td>NM_006536.2</td>
<td>CLCA2</td>
<td>+2.445</td>
<td>Voltage-gated chloride channel activity</td>
</tr>
<tr>
<td>NM_005863.1</td>
<td>NET1</td>
<td>+2.395</td>
<td>Guanyl-nucleotide exchange factor activity</td>
</tr>
<tr>
<td>NM_005627.1</td>
<td>SGK</td>
<td>+2.378</td>
<td>Protein kinase CK2 activity, apoptosis</td>
</tr>
<tr>
<td>NM_002851.1</td>
<td>PTPRN1</td>
<td>+2.362</td>
<td>Protein tyrosine phosphatase activity</td>
</tr>
<tr>
<td>NM_021960.1</td>
<td>MCL1</td>
<td>+2.313</td>
<td>Antiapoptosis, chemoresistance</td>
</tr>
<tr>
<td>M58664.1</td>
<td>CD24</td>
<td>+2.313</td>
<td>Invasion and metastasis biomarker</td>
</tr>
<tr>
<td>NM_006451.1</td>
<td>PAIP</td>
<td>+2.204</td>
<td>Translational initiation and EIF4A enhancer</td>
</tr>
<tr>
<td>NM_005850.1</td>
<td>MMP9</td>
<td>+2.173</td>
<td>Vascular endothelial growth factor release, angiogenesis</td>
</tr>
<tr>
<td>NM_005850.2</td>
<td>MMP2</td>
<td>+2.144</td>
<td>Angiogenesis, invasion, metastasis</td>
</tr>
<tr>
<td>NM_001953.2</td>
<td>EGF</td>
<td>+2.144</td>
<td>Angiogenesis, endothelial proliferation</td>
</tr>
<tr>
<td>NM_004385.1</td>
<td>CSPG2</td>
<td>+2.114</td>
<td>Heterophilic cell adhesion, migration</td>
</tr>
<tr>
<td>NM_004350.1</td>
<td>RUNX3</td>
<td>+2.099</td>
<td>Transcription factor, cell proliferation</td>
</tr>
<tr>
<td>NM_022357.1</td>
<td>RAB38</td>
<td>+2.085</td>
<td>Ras farnesylation, palmitoylation</td>
</tr>
<tr>
<td>NM_006101.1</td>
<td>HEC1</td>
<td>+2.085</td>
<td>Spindle checkpoint signaling, mitosis</td>
</tr>
<tr>
<td>NM_001920.1</td>
<td>DCN/4A</td>
<td>+1.932</td>
<td>Cell proliferation, angiogenesis, epithelial-mesenchymal transition</td>
</tr>
<tr>
<td>NM_000963.1</td>
<td>PTGS2</td>
<td>+1.866</td>
<td>Cell proliferation, angiogenesis, apoptosis</td>
</tr>
</tbody>
</table>

Decreased genes

<table>
<thead>
<tr>
<th>Genbank accession</th>
<th>Gene ID</th>
<th>Average fold change</th>
<th>Biological function</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_003281.1</td>
<td>TNN1</td>
<td>−9.060</td>
<td>Skeletal muscle development</td>
</tr>
<tr>
<td>NM_003837.1</td>
<td>FBP2</td>
<td>−8.515</td>
<td>Reversal of glycolysis</td>
</tr>
<tr>
<td>NM_004490.1</td>
<td>GBF14</td>
<td>−6.821</td>
<td>Platelet-derived growth factor–regulated serine kinase target</td>
</tr>
<tr>
<td>NM_003673.1</td>
<td>TCAP</td>
<td>−6.021</td>
<td>Muscle differentiation</td>
</tr>
<tr>
<td>NM_005963.2</td>
<td>MYH1</td>
<td>−5.776</td>
<td>Motility and vesicular transport</td>
</tr>
<tr>
<td>NM_000290.1</td>
<td>PGAM2</td>
<td>−5.618</td>
<td>Muscle-specific enzyme, muscle function</td>
</tr>
<tr>
<td>NM_000231.1</td>
<td>SGGC</td>
<td>−5.502</td>
<td>Muscle wasting disease, sarcolemma complex</td>
</tr>
<tr>
<td>NM_012278.1</td>
<td>ITGB1A</td>
<td>−5.134</td>
<td>Muscle-specific interactor for β3 integrin</td>
</tr>
<tr>
<td>NM_017533.1</td>
<td>MYH4</td>
<td>−4.112</td>
<td>Motor protein, cellular locomotion</td>
</tr>
<tr>
<td>NM_002469.1</td>
<td>MYF6</td>
<td>−3.837</td>
<td>Muscle regulatory factor</td>
</tr>
<tr>
<td>NM_002152.1</td>
<td>HRC</td>
<td>−3.458</td>
<td>Histidine-rich, calcium-binding protein</td>
</tr>
<tr>
<td>NM_003025.1</td>
<td>MYH7</td>
<td>−2.828</td>
<td>Skeletal muscle function</td>
</tr>
<tr>
<td>M98399.1</td>
<td>CD36</td>
<td>2.144</td>
<td>Fatty acid uptake, thrombospondin receptor</td>
</tr>
<tr>
<td>NM_005205.1</td>
<td>COX6A2</td>
<td>−2.129</td>
<td>Cytochrome oxidase isofrom</td>
</tr>
<tr>
<td>BC005939.1</td>
<td>PTGS1</td>
<td>−1.33</td>
<td>Retinoid transporter, prostaglandin biosynthesis</td>
</tr>
</tbody>
</table>
psoriasin (PSOR1 or S100A7), decorin (DCN), and killer cell C-type lectin receptor isoform 2 (KLRC2), to further validate our microarray results. Each of these genes is known to affect different mechanisms in malignant progression.

Additionally, the translated protein product of two of the up-regulated target genes, psoriasin and versican, were validated by immunohistochemistry in the paraffin sections (n = 20) of oral premalignant and malignant tissues using the appropriate primary antibodies followed by avidin-biotin complex–peroxidase–substrate staining. The representative data for versican and psoriasin expression (Fig. 4A and B) show that both proteins express at higher levels in dysplastic and carcinoma cells than in normal epithelial cells. Versican was found to be expressed along the leading edge of the motile cells consistent with its role in de-adhesion of transformed cells, whereas psoriasin expression was found to be higher in invasive epithelial cells. Future studies for detailed analysis may reveal localization differences if any. It is concluded here that expression of both of these markers is associated with the invasive potential of dysplastic epithelium and impending malignant transformation. Nonmalignant tissue specimens were included in the study as negative controls to rule out any false-positive reaction (data not shown).

Identification of a Malignant Progression-Related Pathway in Oral Precancer

We queried the tightly regulated genes of a cluster node obtained from the clustering analysis represented by their probe IDs on the NetAffx analysis center for Genbank annotations. Such annotations are linked to molecular and biological function databases, such as Gene Ontology, and pathway databases, such as KEGG. We predicted pathways involved in the disease process using analysis programs such as GenMAPP and Ingenuity pathway analysis. On summation of all pathways, we found that the genes differentially regulated between normal and dysplastic tissues are involved in cell signaling, cell proliferation, angiogenesis, apoptosis, matrix remodeling, motility, invasion, and innate immune surveillance mechanisms. After considering the most significantly altered gene expression profiles and its pathway representation analysis, the most perturbed pathway in dysplastic tissue was mapped relating to arachidonic acid metabolism and related inflammatory cascades. This pathway also represents appropriate crosstalk with the transforming growth factor-β signaling, Ras oncogenic signaling, prostanoid receptor signaling, and fatty acid degradation pathways as summarized in the progression pathway diagram (Fig. 5). Based on global analysis (see Supplementary Data) of the significant genes involved (30 cancer-related genes), we propose that it could be central in the malignant progression of oral dysplasia. This is because the resultant vector of the pathway points toward combined cell proliferation, anti-apoptotic, andangiogenesis mechanisms, with concomitant down-regulation of innate immune response–related genes that helps eliminate genetically aberrant cells from the system. Thus, our data supports a general notion that carcinogenesis is a result of multifactorial crosstalk between different molecular pathways.

Candidate Disease-Related Genes in the Pathway

Among the genes differentially expressed in premalignant versus paired normal tissues, the most significantly regulated genes are found associated with various malignant processes. These are the genes functionally known to be involved in angiogenic stimulation (MMP2, MMP9, ECGF, and SELE), apoptosis, cell proliferation, and cell cycle regulation (MCL1, CDC6, CDK4, MDM7, CCND2, CCNC, CCNB1, E2F, GADD45, RAB 38, and BNIP3), cell
However, the inflammation-related genes, such as cyclo-
A2, eIF6, and protein translation (MYOM2) of PGD 2 synthase in oral premalignant lesions could be a
appearance of the psoriasin and versican expression and loss
are warranted. Based on our study, we propose that the
mechanisms that promote invasiveness in oral dysplasia
specific studies directed toward the precise molecular
been reported in invasive breast carcinoma samples (26, 27),
contribution of psoriasin in the invasive process has recently
oral lesions remains to be elucidated. Although the
several expressed sequence tags and hypothetical proteins
involved in oral squamous carcinoma development by
several groups earlier (21–25). We have also identified
several genes from the list of 1,700
significant genes identified in our study have been found
in oral squamous carcinoma development by
several groups earlier (19, 20) in correlation with malignant potential of oral
premalignant tissues. Some of the genes from the list of 1,700
Figure 4. Versican and psoriasin protein expression in oral tissues. The immunohistochemistry-stained paraffin sections of various premalignant and
malignant stages of oral cancer progression specimens (n = 20) show representative staining for versican (in A) in sections of hyperplasia (a), and mild (b), moderate (c), or severe (d) dysplasia samples. B, representative staining for psoriasin in focal fibrous hyperplasia (a), mild (b), moderate (c), severe
dysplasia (d), well-differentiated carcinoma (e), and poorly differentiated carcinoma (f). All photographs of tissue sections are at ×40 magnification.
Hyperplasia specimens used in the experiments comprise as clinically nonmalignant, nonprogressive negative controls.

 signaling (PPARGC1, ITGB1BP2), DNA repair (XRCC5, ERCC1), epithelial to mesenchymal differentiation (TNNI1, MYF6, MYH1, MYH4, MYH7, MYLK, and MYOM2), oncogenesis (NET, ECT2, and SET), invasiveness (PSORI S100A7, CSPG2, and NRCAM), immunoregulation (CCL5, CD44v6, CXCL11, IGGH3, and KLRC2), and protein translation (eIF6, eIF3S10, and eIF4A1). However, the inflammation-related genes, such as cyclo-
oxygenase-2 (COX-2/PTGS2) and decorin (DCN) variant
A2, known to be involved in multiple bioprocesses, such as
in cell proliferation, angiogenesis, and immune surveillance,
seem most promising as molecular targets. Both of these
genes were studied individually in detail by our group
earlier (19, 20) in correlation with malignant potential of oral
premalignant tissues. Some of the genes from the list of 1,700
significant genes identified in our study have been found
involved in oral squamous carcinoma development by
several groups earlier (21–25). We have also identified
differentiated genes in the invasive process has recently
been reported in invasive breast carcinoma samples (26, 27),
specific studies directed toward the precise molecular
mechanisms that promote invasiveness in oral dysplasia
are warranted. Based on our study, we propose that the
appearance of the psoriasin and versican expression and loss
of PGD2 synthase in oral premalignant lesions could be a
definitive and novel molecular signature of impending
malignant potential.

Lipocalin-Type Prostaglandin D2 Synthase Acts as a
Protective Enzyme Target

Based on our microarray results, a biological assay-
based functional analysis was designed to test the
hypothesis of the central role of the COX pathway in
the oral malignant transformation. We identified, by
comparison analysis, the down-regulation of the PTGDS
message as an important variable in malignant pro-
gression with respect to nonaggressive verrucous forms
of oral tissue (verrucous hyperplasia). These enzyme-
derived metabolites (PGD2 and subsequent PGJ2 series
of compounds) are known to act as negative feedback
regulator of COX-2 gene transcription (28). Hence, we
hypothesized that supplementation of PGD2 in the oral
premalignant and malignant cell line culture media
directly should have an inhibitory effect on the cell lines.
Our hypothesis is also supported by studies done earlier
to see the effects of different prostaglandins or their
downstream metabolites on prostate and ovarian cancer
cells (29, 30).

Indeed, our experiments elucidated the effect of PGD2
substitution on the cell proliferation rates of oral cancer
progression model cell lines (Fig. 6A) and show that there
is significant inhibition at all stages of oral cancer
progression. We saw only minimal effect in cell prolif-
eration rates of the human papillomavirus-16 trans-
formed primary human oral keratinocyte cell line (HOK
16 B). The inhibition in dysplastic (DOK) to malignant
primary carcinoma (SCC25) to metastatic (OSC-2) stages
was measured utilizing 3-(4,5-dimethylthiazol-2-yl)-2,5-
diphenyltetrazolium bromide dye reduction assay, which
quantitates only metabolically active and viable cells.
Further study of inhibition mechanisms, done to analyze
the effect on apoptosis (caspase-3/8 activation assays,
Annexin V-FITC binding) or cell cycle distribution (Telford’s
assay) by flow cytemetry methods, confirmed this cell
proliferation inhibition as a result of G2-stage arrest in the
cell cycle (Fig. 6B). The results are significant (P < 0.01) when
compared with only vehicle-treated population of similar
cell lines. We did not observe any effects (data not shown) on
apoptotic mechanisms, and such data is explained by the down-regulation of gene expression profile related to the apoptotic effector molecules (cytochrome c oxidase release) observed in our microarray experiments. These effector molecules are known to initiate the classic caspase activation pathways. Thus, in our study, we have identified a route to natural cancer prevention process that effectively inhibits COX-2 at the transcriptional level and lowers the level of downstream metabolite (PGE2)–mediated cell proliferation. It seems from our study that in the absence or repression of the PTGDS enzyme levels in oral epithelial cells, dysplastic epithelial cells may progresses toward malignant stages and, therefore, could be termed as a protective target.

Discussion

The results of our study identifies, for the first time, a major change of expression levels in several genes of the arachidonic acid metabolism pathway in oral dysplasia. This central inflammatory pathway results in constitutive presence of several proinflammatory enzymes, including COX-2 (PTGS2) and cognate receptors (such as EP3a and EP3b) of its metabolic by-products, and thus is causal in malignant progression. The negative feedback regulation is not efficient enough due to the repression of a key enzyme in the arachidonic acid metabolism pathway (i.e., PTGDS), which negatively regulates COX-2 gene transcription. This pathway is supposedly initiated with a deregulated transforming growth factor-β pathway (31) and further induced by epidermal growth factor receptor activation (32). Stabilization of the COX-2 mRNA proceeds either because of altered splicing events described earlier by us (33) or through masking of AU-rich elements in the 3′ untranslated region of COX-2 message by Ras activation modulated proteins (34, 35), thus ensuring constant presence of COX-2 message in the system. In concert, metabolites of several differentially regulated genes coding for enzymes (e.g., PGE isomerase, PTGES, and lipoxygenases such as ALOX-5 and ALOX-12) in the pathway promote cell proliferation, antiapoptotic, angiogenic, and immunomodulatory processes. We found the gene expression profile changes associated with stage-dependent and biological behavior–consistent patterns. High correlation coefficients ($r^2 > 0.9$) depicting gene clusters resulted in a signature transcriptome profile for different premalignant and malignant stages. Based on such significant genes associated with disease progression, a multiplex gene expression diagnostic (reverse transcription-PCR or diagnostic microarray) can be developed for predicting the malignant transformation potential of premalignant oral tissues. In our earlier studies, we observed an early shift in cytokine profile (Th1/Th2 balance) due to PGE2 that may destabilize the innate immune surveillance mechanism (36). Due to this shift, natural killer and T-cell–mediated immunity against transformed oral epithelial cells may be lost. We have also characterized one such important aberrant gene expression of proteoglycan (decorin) and its localization that might be responsible for transforming growth factor-β pathway–dependent malignant progression, angiogenesis promotion, and depressed immunosurveillance (37). In this study, we further identify two invasion-related biomarkers, versican and psoriasin, with increased expression levels in oral premalignant and malignant tissues compared with paired normal control tissues that are functionally involved in tumor progression (38–40). Finally, using an in vitro oral cancer progression model, we observed efficient inhibition of cell proliferation rates through alteration in cell cycle events upon PGD2 treatment that substitute for PTGDS metabolic function. Interestingly, additional molecular mechanisms were proposed earlier. A nonenzymatically

![Diagram](MolCancerTherTherapeutics873.png)

**Figure 5.** Gene network and pathway prediction analysis identifies convergence of proinflammatory genes in malignant progression of oral cells. The key proinflammatory enzymes (PTGS2, ALOX5, ALOX12, and PTGES) and their receptors (PTGER 3β1, a2) are up-regulated (red) and, consequently, their metabolites propel disease progression, whereas anti-inflammatory and negative feedback regulatory enzyme (PTGDS, PTGIS) metabolites and receptors pathways (green) are down-regulated in the arachidonic acid metabolism pathway. Each node in the pathway represents significant alterations in expression compared with normal after Bonferroni’s correction and Fisher’s exact probability analysis.
PGD₂, has been reported to mediate the significant inhibition of cell proliferation rates (such as inducible nitric oxide synthase and COX-2. Transcription of nuclear factor-κB increasing its levels in the nuclei and inhibiting the two-tailed paired Student's t test.

The compound 15d-PGJ₂ also inhibits the degradation of IκBα, increasing its levels in the nuclei and inhibiting the transcription of nuclear factor-κB-dependent genes (45), such as inducible nitric oxide synthase and COX-2.

Therefore, it seems that all of these mechanisms may act synergistically to provide a natural prevention mechanism in oral malignant progression. Recently, PGD₂ was also reported as direct ligands of DP and CRTH₂ receptors. PGD₂ favors Th₂-type functions through CRTH₂ receptors known to be expressed mainly on T helper cells while restraining Th₁-like functions via DP receptors (46). DP1 receptor transcripts were reported to be present only in HT-29 cells by Hawcroft et al. (47), and none of the human colorectal epithelial cell lines tested by them expressed any DP2 receptor mRNA. Consequently, they concluded that PGD₂ is unlikely to have direct activity on neoplastic colorectal epithelial cells via cell surface DP receptors. We are in the process of deciphering the mechanistic effects of direct PGD₂ substitution on the cell lines of the oral cancer progression model and this includes possible receptor signaling. Based on our observations, we propose that combined early intervention strategies based on not only COX-2 inhibition at the enzymatic activity level but also at transcriptional regulation level using PGD₂ analogues may provide further therapeutic efficacy or might at least delay the malignant progression of oral epithelia.

**Acknowledgments**

We thank the microarray core facility of Johns Hopkins University School of Medicine and Public Health (Baltimore, MD) for extensive support; Munroe Meyer Institute microarray core facility at University of Nebraska Medical Center (Omaha, NE) for providing us the Affymetrix (MAS 5.0 suite and DMT) and Ingenuity analysis softwares; University of Nebraska Medical Center molecular biology and cell cycle analysis core facilities for assistance; Drs. Simon Sherman and Xiao Li (Nebraska Informatics Center in Life Sciences, Epiple University, University of Nebraska Medical Center, Omaha, NE) for their help in statistical validation and bioinformatics analysis of our microarray data; and Drs. Peter H. Watson (University of Manitoba, Winnipeg, Manitoba, Canada) and Larry Fisher (National Institute of Dental and Craniofacial Research, NIH, Bethesda, MD) for the gift of primary antibodies for psoriasin and versican, respectively.

**References**

6. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Emberley ED, Niu Y, Njue C, Kliewer EV, Murphy LC, Watson PH. Psoriasin (S100A7) expression is associated with poor outcome in


modulating prostate cancer angiogenesis: differential expression of 
Annexin II in prostate carcinomas from India and USA. Mol Cancer 
2003;2:34.


14. Telford WG, King LE, Fraker PJ. Evaluation of glucocorticoid-induced 


17. Ding C, Cantor CR. Quantitative analysis of nucleic acids—the last 


expression profiling assisted by laser capture microdissection and micro- 

carcinomas of the head and neck revealed by the use of laser 

23. Loercher A, Lee TL, Ricker JL, et al. Nuclear factor-κB is an important 


26. Emberley ED, Niu Y, Leygue E, et al. Psoriasis interacts with Jab1 and 

27. Emberley ED, Niu Y, Njue C, Kliever EV, Murphy LC, Watson PH. 

through hB kinase-independent and peroxisome proliferator-activated 

Campbell WB. Requirement of cyclooxygenase-2 expression and prosta- 
glandins for human prostate cancer cell invasion. Clin Exp Metastasis 

of human ovarian cancer cell growth in vitro and in nude mice by 


32. Dannenberg AJ, Lippman SM, Mann JR, Subbaramaiah K. 
DuBois RN. Cyclooxygenase-2 and epidermal growth factor receptor-
pharmacologic targets for chemoprevention. J Clin Oncol 2005;23: 
254 – 68.


34. Sheng H, Shao J, DuBois RN. K-Ras-mediated increase in cyclo-

inhibitor suppresses the growth of H-ras-transformed rat intestinal 


39. Emberley ED, Niu Y, Njue C, Kliever EV, Murphy LC, Watson PH. 
Psoriasis (S100A7) expression is associated with poor outcome in 


41. Itoh K, Mochizuki M, Ishi Y, et al. Transcription factor Nr2f2 regulates 
inflammation by mediating the effect of 15-deoxy-D12,14-prostaglandin 

42. Shibata T, Kondo M, Osawa T, Shibata N, Kobayashi M, Uchida K. 
15-Deoxy-12,14-prostaglandin J2. A prostaglandin D2 metabolite 
generated during inflammatory processes. J Biol Chem 2002;277: 
10459 – 66.

43. Cuzzocrea S, Wayman NS, Mazzon E, et al. The cyclopentenone 
prostaglandin 15-deoxy-D12,14-prostaglandin J(2) attenuates the de-
velopment of acute and chronic inflammation. Mol Pharmacol 2002;61: 
997 – 1007.

44. Thimmulappa RK, Mai KH, Srisuma S, Kensler TW, Yamamoto M. 
Biswal S. Identification of Nr2f2-regulated genes induced by the chemo-
preventive agent sulforaphane by oligonucleotide microarray. Cancer 

Inhibition of hB kinase and hB phosphorylation by 15-deoxy-D12,14-

46. Tanaka K, Hirai H, Takano S, Nakamura M, Nagata K. Effects of 
prostaglandin D2 on helper T cell functions. Biochem Biophys Res 

47. Havcroft G, Gardner SH, Hull MA. Expression of prostaglandin D2 
receptors DP1 and DP2 by human colorectal cancer cells. Cancer Lett 
Molecular Cancer Therapeutics

Identification of genes and molecular pathways involved in the progression of premalignant oral epithelia

Abhijit G. Banerjee, Indraneel Bhattacharyya and Jamboor K. Vishwanatha


Updated version
Access the most recent version of this article at:
http://mct.aacrjournals.org/content/4/6/865

Supplementary Material
Access the most recent supplemental material at:
http://mct.aacrjournals.org/content/suppl/2005/09/20/4.6.865.DC1

Cited articles
This article cites 37 articles, 19 of which you can access for free at:
http://mct.aacrjournals.org/content/4/6/865.full.html#ref-list-1

Citing articles
This article has been cited by 11 HighWire-hosted articles. Access the articles at:
/content/4/6/865.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.