Proteomics Analysis of Cells in Whole Saliva from Oral Cancer Patients via Value-added Three-dimensional Peptide Fractionation and Tandem Mass Spectrometry*[S]

Hongwei Xie‡, Getiria Onsongo§, Jonathan Popko‡, Ebbing P. de Jong‡, Jing Cao‡, John V. Carlis§, Robert J. Griffin¶, Nelson L. Rhodus‖, and Timothy J. Griffin‡**

Whole human saliva possesses tremendous potential in clinical diagnostics, particularly for conditions within the oral cavity such as oral cancer. Although many have studied the soluble fraction of whole saliva, few have taken advantage of the diagnostic potential of the cells present in saliva, and none have taken advantage of proteomics capabilities for their study. We report on a novel proteomics method with which we characterized for the first time cells contained in whole saliva from patients diagnosed with oral squamous cell carcinoma. Our method uses three dimensions of peptide fractionation, combining the following steps: preparative IEF using free flow electrophoresis, strong cation exchange step gradient chromatography, and microcapillary reverse-phase liquid chromatography. We determined that the whole saliva samples contained enough cells, mostly exfoliated epithelial cells, providing adequate amounts of total protein for proteomics analysis. From a mixture of four oral cancer patient samples, the analysis resulted in a catalogue of over 1000 human proteins, each identified from at least two peptides, including numerous proteins with a role in oral squamous cell carcinoma signaling and tumorigenesis pathways. Additionally proteins from over 30 different bacteria were identified, some of which putatively contribute to cancer development. The combination of preparative IEF followed by strong cation exchange chromatography effectively fractionated the complex peptide mixtures despite the closely related physiochemical peptide properties of these separations (pI and solution phase charge, respectively). Furthermore compared with our two-step method combining preparative IEF and reverse-phase liquid chromatography, our three-step method identified significantly more cellular proteins while retaining higher confidence protein identification enabled by peptide pI information gained through IEF. Thus, for detecting salivary markers of oral cancer and possibly other conditions of the oral cavity, the results confirm both the potential of analyzing the cells in whole saliva and doing so with our proteomics method. ***Molecular & Cellular Proteomics 7:486–498, 2008.***

Whole human saliva is easily collected in the clinic in a non-invasive, on-demand manner and in relatively large, easily stored quantities, making it an optimal bodily fluid for clinical diagnostics (1, 2). Diagnosis of oral cancer could benefit greatly from the development of whole saliva-based clinical tests given the physical proximity of the site of cancer development with the diagnostic fluid. In fact, oral cancer, usually diagnosed in the form of oral squamous cell carcinoma (OSCC),[1] has not seen a drop in its 50% mortality rate over the last 30 years (3), motivating the development of new and reliable, saliva-based clinical diagnostic tests for the early detection of OSCC. Such tests would lead to more informed treatment of patients and a reduction in the suffering and death caused by this cancer.

To develop these tests, molecular markers that are predictive of cancer development need first to be identified within whole saliva. Many researchers have studied the soluble fraction of whole saliva as a source of these markers (4–8) with some of these studies using large scale mass spectrometry-based proteomics technologies to catalogue its protein components (9–14). Fewer researchers, however, have investigated using the cells found in whole human saliva for oral cancer diagnostics, and none have used proteomics methods to identify cellular protein markers for potential OSCC detection. Instead these studies have concentrated on detecting genetic or pathological changes associated with OSCC in these cells (15–17).

To catalogue proteins from complicated biological samples

From the Departments of ‡Biochemistry, Molecular Biology, and Biophysics, §Computer Science and Engineering, and ‖Oral Medicine, Diagnosis, and Radiology, School of Dentistry, University of Minnesota, Minneapolis, Minnesota 55455 and ¶Department of Radiation Oncology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205

Received, April 2, 2007, and in revised form, October 24, 2007
Published, MCP Papers in Press, November 28, 2007, DOI 10.1074/mcp.M700146-MCP200

[1] The abbreviations used are: OSCC, oral squamous cell carcinoma; FFE, free flow electrophoresis; μLC, microcapillary LC; SCX, strong cation exchange; DAPI, 4′,6-diamidino-2-phenylindole; iTRAQ, isobaric tags for relative and absolute quantitation.

using mass spectrometry, peptide fractionation methods are commonly used. Complex proteolytic protein digests are divided into less complex subsets of peptide fractions from which lower abundance proteins can be detected by the mass spectrometer. Generally these methods combine multiple different chromatography or electrophoresis steps that fractionate peptide sequences using different physiochemical properties (18, 19). For historical reasons (20), the different fractionation steps used are commonly thought of as being "orthogonal" to each other, leading to the generally used description of these methods as "multidimensional" separations. However, because for the case of modern proteomics applications, spatial dimensionality no longer holds for the fractionation steps comprising these methods, these are more accurately termed "multistep" peptide fractionation methods. Two-step peptide fractionation using a first step of strong cation exchange (SCX) chromatography followed by a second step of microcapillary reverse-phase liquid chromatography (μLC) on line with ESI-MS/MS has become a standard for proteomics analysis (18, 21, 22). With increasing popularity, an alternative two-step method using preparative peptide IEF followed by μLC-ESI-MS/MS for sensitive analysis of complex protein mixtures is being used as described by our group (11, 23) and others (9, 24–29). This alternative gets valuable peptide pI information from the IEF fractionation that aids in the high confidence identification of peptide sequence matches determined by sequence database searching of the MS/MS data. One group has also demonstrated the effectiveness of a three-step, LC-based peptide fractionation method for increasing the number of proteins identified by MS/MS (30), although none have described a three-step method using preparative IEF as one of the steps.

Here we introduce a novel three-step peptide fractionation method and demonstrate its effectiveness for proteomics cataloguing of cells found in whole saliva from patients with diagnosed OSCC lesions. Furthermore we show that whole saliva contains reliable amounts of cells, mainly exfoliated from the oral epithelium, which provide adequate amounts of protein for proteomics studies using mass spectrometry. The steps of our three-step method are: 1) preparative IEF using free flow electrophoresis (FFE), 2) SCX chromatography, and 3) μLC on line with ESI-MS/MS. We show that our method effectively fractionates complex peptide mixtures and increases significantly the number of proteins identified compared with our earlier described two-step method while retaining the benefits of peptide pI for high confidence protein identification (25, 26, 31). We identified over 1000 human proteins (each matched by at least two peptides) from the cells in whole saliva from OSCC patients, with a number of these being low abundance proteins having a previously described association in oral cancer progression, and identified for the first time directly from whole saliva. We also identified proteins from over 30 different bacteria, many of which have not been described previously in whole human saliva and several of which have possible links to cancer. These results provide the first description of the potential for using cells in whole human saliva to identify protein markers of oral cancer progression and the effectiveness of our proteomics method for such studies.

EXPERIMENTAL PROCEDURES

Patient Description and Whole Saliva Collection—The study was conducted with approval by the Institutional Review Board of the School of Dentistry, University of Minnesota. Saliva was collected from four subjects with clinically diagnosed and histopathologically confirmed primary OSCC ulcerative lesions located on the tongue. Unstimulated whole saliva was obtained using a standard, controlled protocol (6) by first having each subject swallow and then expectorate continuously into a 50-ml sterile, polypropylene conical tube for a period of 5 min. This resulted in ~4 ml or more of total saliva from each subject. Following collection, the samples were immediately placed on ice and stored at −70 °C until further processing.

Measurement of Cell Numbers and Cellular Protein Concentrations in Whole Saliva—1 ml of whole saliva from each of the four different subjects was separately centrifuged at 1500 × *g* at 4 °C for 10 min, and each was processed identically in parallel. After carefully discarding the supernatant, the insoluble cell pellet was washed twice using ice-cold PBS buffer (Invitrogen), and the cells were resuspended into 1 ml of ice-cold PBS. Next 100 μl was removed, and the cells were stained with trypan blue followed by counting using a light microscope (Nikon). The remaining 900 μl of solution containing cells was again pelleted by centrifugation, the supernatant was removed, and the cells were lysed using radioimmune precipitation assay buffer containing Triton X-100 (Boston BioProducts, Worcester, MA) and a protease inhibitor mixture (Roche Applied Science). The protein concentration was determined by the BCA protein assay (Pierce).

Cytological Characterization of Cells in Whole Saliva—For these experiments, 500 μl of whole saliva from each of the four subject saliva samples was combined together, and the cell pellet was isolated by centrifugation as described above. Immunostaining of the cells was carried out following a previously described protocol (32). After washing the cell pellet with ice-cold PBS buffer, it was redissolved in 1 ml of PBA (0.1% BSA in 1× PBS buffer), and an equal volume of 0.2% Triton X-100 in PBS was added. The solution was incubated for 3 min on ice and then centrifuged at 500 × *g* for 10 min, and the supernatant was removed. The cells were redissolved with 1 ml of PBA and transferred to a 1.5-ml microcentrifuge tube with protection from light, and 5 μl of FITC-labeled anti-cytokeratin mouse monoclonal antibody was added (Abcam, Cambridge, MA) followed by gentle shaking for 30 min at room temperature. The cells were then centrifuged, washed with ice-cold PBS, redissolved in 1 ml PBS, and then stained with DAPI DNA stain (Fisher Scientific) by adding 3 μl of a 3 mg/ml DAPI stock solution in water with gentle shaking for 12 min at room temperature. The cells were again centrifuged, washed with ice-cold PBS, redissolved in 300 μl of PBS, and counted using a fluorescence microscope (BX60, Olympus, New York, NY).

Cellular Protein and Peptide Preparation—For proteomics studies, new 2-ml aliquots of whole saliva from each of the four subjects were combined, the cell pellet was isolated by centrifugation, the cells were lysed as described above, and total protein was measured using the BCA method. From this mixture, 200 μg of total protein was exchanged into a buffer containing 50 mM Tris-HCl, 100 mM NaCl, and 1% SDS using an Amicon ultracentrifugal filter device (5-kDa molecular mass cutoff, Millipore, Billerica, MA). The sample was boiled briefly to denature proteins, and additional Tris/NaCl buffer was added to achieve a final buffer consisting of 50 mM Tris-HCl, 100 mM NaCl, 0.1% SDS, pH 7.5. Tris(2-carboxyethyl)phosphine reducing agent (Pierce) was added to reach a final concentration of 5 mM, 20

μg of modified trypsin (Promega, Madison, WI) was added, and the mixture was incubated overnight at 37 °C. The resulting peptides were desalted using a mixed mode cation exchange cartridge (Waters, Milford, MA) and concentrated by vacuum centrifugation. The peptides were labeled with the iTRAQ reagent (33) (Applied Biosystems) using the manufacturer's protocol. Although the objective of this study was not to gain quantitative information, these samples were labeled with the iTRAQ reagent to test the amenability of labeled peptides to our three-step peptide fractionation method for its potential use in future quantitative studies. After labeling, the peptides were desalted again by mixed mode cation exchange cartridge chromatography and dried by vacuum centrifugation.

FFE Fractionation—The overall proteomics method used for peptide fractionation and protein identification is summarized in Fig. 2 and described under "Results." Here we provide the relevant experimental details that go with this figure and description. The peptides (200 μg) were redissolved in 250 μl of FFE buffer (pH ~8.5) and fractionated using an FFE system (BD Biosciences) enabling preparative IEF of peptides and collection into a 96-deepwell microtiter plate as we have described previously (11, 23). Initially a 50-μl aliquot (~10% of total) was taken from each of the FFE fractions and processed using an Amicon Ultrafree-MC centrifugal filter device (5-kDa molecular mass cutoff, Millipore) to remove high molecular weight hydroxypropylmethylcellulose polymer as described previously (23). The flow-through from the centrifugal device, containing purified peptides, was dried by vacuum centrifugation and reconstituted in 30 μl of HPLC load buffer (0.1% formic acid, 2% acetonitrile in water). Each sample was analyzed by μLC-MS/MS analysis (as described below) to obtain an initial profile of the peptide distribution across the microtiter plate fractions and to identify those fractions containing the most complex peptide mixtures, necessitating their further fractionation by SCX as described below.

SCX Fractionation—For those FFE fractions containing relatively large numbers of peptides, a second step of fractionation was performed using a PolySULFOETHYL SCX guard column (Javelin guard column, 1.0-mm inner diameter × 10 mm, 5 μm, 300 Å, PolyLC, Inc.) using an automated syringe pump capable of highly accurate sub-μl/min flow rates (Harvard Apparatus Inc.). For SCX fractionation, a 250-μl aliquot was removed from each acidic FFE fraction (pH range 3.5–6, 10 FFE fractions total) as these contained the most abundant numbers of peptides, consistent with our previous descriptions of peptide fractionation by FFE (23). For the other neutral and basic fractions, two 125-μl aliquots were combined from each adjacent FFE fraction as these were determined to contain lower numbers of peptides in initial screening experiments. Combining these fractions resulted in an additional 10 fractions. Each of the 250-μl aliquots was subjected to ultrafiltration as described above to remove the high molecular weight hydroxypropylmethylcellulose polymer, and the filtrate was then desalted using a Sep-Pak C_{18} cartridge (Waters) followed by concentration by vacuum centrifugation. Each peptide fraction was then redissolved in 200 μl of SCX loading buffer (10 mM KH_2PO_3 containing 20% acetonitrile, pH = 3.0) and loaded onto a preconditioned SCX column at a flow rate of 50 μl/min. After washing with loading buffer and the loading buffer containing 15 mM KCl (to remove loosely retained contaminants), peptides were eluted using step gradient chromatography, using steps with increasing KCl concentration, at a flow rate of 50 μl/min. For the acidic fractions, steps of 20, 25, 50, and 200 mM KCl in loading buffer were collected (200-μl total volume); for neutral and basic samples, slightly adjusted steps of 20, 25, 100, and 200 mM KCl were collected, accounting for the expected presence of more highly charged peptides with more basic peptide pI values compared with the acidic peptides. Each collected fraction was concentrated by vacuum centrifugation and reconstituted in 30 μl of HPLC load buffer.

μLC-MS/MS Analysis—All on-line μLC separations were done on an automated Paradigm MS4 system (Michrom Bioresources, Inc., Auburn, CA). Each processed FFE/SCX fraction was automatically loaded across a Paradigm Platinum Peptide Nanotrap (Michrom Bioresources, Inc.) precolumn (0.15 × 50 mm, 400-μl volume) for sample concentrating and desalting at a flow rate of 50 μl/min in HPLC buffer A (0.1% formic acid in a solution of 5% acetonitrile and 95% water). The in-line analytical capillary column (75 μm × 12 cm) was home-packed using C_{18} resin (5 μm, 200-Å Magic C18AG, Michrom Bioresources, Inc.) with the exception that the electrospray tip was made with a hand-held torch. Peptides were eluted using a linear gradient of 10–35% HPLC buffer B (0.1% formic acid in a solution of 95% acetonitrile and 5% water) over 60 min followed by isocratic elution at 80% buffer B for 5 min with a flow rate of 0.25 μl/min across the capillary column.

All mass spectrometry was done on an LTQ linear ion trap mass spectrometer (Thermo Fisher, Inc.) using Xcalibur version 2.0 operating software. Ionized peptides eluting from the capillary column were automatically selected for MS/MS using a data-dependent procedure that alternated between one MS scan followed by four MS/MS scans for the four most abundant precursor ions in the MS survey scan. Those *m/z* values selected for MS/MS were dynamically excluded for 30 s using a repeat count of 2. The electrospray voltage applied was 2.0 kV. MS and MS/MS spectra were acquired with a maximum fill time of 50 and 100 ms for MS and MS/MS analysis, respectively. MS spectra were acquired from a single microscan, whereas MS/MS spectra were acquired using two microscans. For MS scans, the *m/z* scan range was set from 400 to 1800 daltons.

Sequence Database Searching and Data Analysis—To identify the human proteins from the cellular samples, obtained MS/MS spectra were searched using SEQUEST (34) (Bioworks version 3.2, Thermo Finnigan, San Jose, CA) against a non-redundant human protein sequence database from the European Bioinformatics Institute (ipi.HUMAN.v3.18.fasta, containing 62,000 entries). This database was chosen because it is manually curated to minimize protein redundancy due to identical protein sequences appearing under different accession codes. A reversed-sequence version of the same database was appended to the end of the forward version for the purpose of false positive rate estimation (22). Search parameters included static mass shift (+144.0 Da) for the N terminus and lysine due to modification with the iTRAQ reagent. Differential amino acid mass shifts for oxidized methionine (+16 Da) were also included. Precursor peptide mass tolerance was ±2.0 Da with no tryptic specificity. Fragment ion tolerance was set at 1.0. To each matched peptide sequence a predicted pI using the Shimura *et al.* (35) algorithm was automatically assigned using a script developed in house. The pK_a values for the N terminus and lysine ε amino group were set to a value of 8.6 to account for the addition of the *N*-methylpiperazine ring of the iTRAQ reagent (33). The search results were validated using the peptide validation program PeptideProphet (36), which assigns a comprehensive probability (*p*) score from 0 to 1 to each peptide sequence match based on its SEQUEST scores (Xcorr, ΔCn, Sp, and RSp) and additional information, including mass difference between the precursor ion and the assigned peptide and the number of tryptic termini. The peptide sequence match results were organized and interpreted using the software tool Interact (37), allowing up to two missed cleavage for identified peptides.

For each FFE fraction, an average pI value was calculated from peptide matches with $p \geq 0.9$. Our previous studies (11, 23) have shown that the average pI of the FFE fraction approximates the pH of the fraction and is useful for filtering peptide sequence matches. Using this average pI assigned to each FFE fraction, only peptide matches (regardless of assigned *p* score) were kept for further consideration if their predicted pI was within ±0.5 units, based on the FFE

resolution (23), of the average pI value for the FFE fraction from which they were identified and if the peptide sequence was at least partially tryptic to maximize the high confidence matches (31). Those peptides passing this first phase of filtering were then further filtered using a two-step procedure similar to that we have described previously (11). In the first step, we took the entire dataset of filtered peptide matches and determined the lowest p score at which the estimated false positive rate was maintained at 1%, which was $p = 0.75$ (by comparison, without the initial pI filtering step the p score threshold needed to be 0.98 to maintain this low false positive rate). The second step filtered the peptide sequence matches first using a low stringency p score of 0.2 with the added stipulation that a protein would be accepted only if it was matched by two or more unique peptides. This filtering step also resulted in a false positive rate below 1% (again by comparison, without the initial pI filtering step, the p score using this method must be raised to 0.5 to maintain the same false positive rate). The protein matches resulting from both of these filtering steps were combined, and only proteins matched from at least two or more unique peptides were included in our final catalogue of proteins. The estimated false positive rate for our protein catalogue was 0.24%.

To identify proteins derived from bacteria within the cell pellet of whole saliva, the obtained MS/MS spectra were also searched against the updated, comprehensive UniProt/Swiss-Prot protein sequence database release 6.0 (38) containing 194,317 entries from over 10,000 different organisms as well as over 135,000 protein sequences from bacteria and viruses. A reversed-sequence version of the same database was appended to the end of the forward version to enable false positive rate estimation. Average FFE pI values (derived from the human peptide sequences identified in each fraction) and calculated peptide pI values were used to filter the resulting bacterial peptide sequence matches as described above for the human protein sequence database results, and the same two-step filtering method was used to obtain a final catalogue of proteins with an estimated false positive rate of 1%. Protein identifications matching to non-human organisms (*i.e.* bacteria and viruses) were manually extracted and further inspected. To ensure high confidence, bacteria were only considered to be present in the sample if peptides matched to two or more proteins derived from the bacteria were observed, or if only a single protein from the bacteria was identified, two or more unique peptides from the protein were matched. Additionally at least one of the peptide sequences identified from each of the bacteria had to be expressed only in that bacterium (*i.e.* not a redundant peptide sequence found in proteins known to be expressed in multiple bacteria types).

Immunoblotting Confirmation Experiments—For these studies, the washed cell pellet was lysed by suspending the pellet in 2 volumes of 50 mM Tris-HCl (pH 6.8) with 2% (w/v) SDS, 5% (v/v) β-mercaptoethanol, and a protein inhibitor mixture (Roche Applied Science 13457200) and boiling for 10 min. After cooling, the solution was centrifuged at 16,000 $\times$ g for 15 min at 4 °C to remove cellular debris. Proteins in the supernatant were concentrated by acetone precipitation and redissolved in ~200 μl of 50 mM Tris-HCl (pH 7.4) containing 2% SDS.

Western blots were performed for three proteins, whose genes are denoted in Table II as *STAT3*, *PRDX3*, and *SCCA1*. SDS-PAGE was performed on a Bio-Rad Tris-HCl minigel (10, 12, and 12%) with Tris-glycine-SDS buffer at 144 V. The proteins were transferred to a PVDF membrane for 60 min at 100 V in a cold room. The membrane was blocked with 1$\times$ PBS containing 2.5 mg/ml BSA and 0.25% (v/v) Tween 20 for 15, 80, and 120 min. This was followed by incubation with the primary antibody (mouse monoclonal (23G5) to *STAT3* (phospho-Ser-727), Abcam ab24922, at 1:200 dilution for P40736; anti-AOP1, Sigma-Aldrich A7674, at 1:10,000 dilution for P30048; *SCCA1* (8H11), Santa Cruz Biotechnology, Inc. sc-21767, at 1:400 dilution for P29508) in 1$\times$ PBS containing 2.5 mg/ml BSA and 0.25% (v/v) Tween 20. After three washes, the membranes were incubated with secondary antibody solution (horseradish peroxidase-conjugated goat anti-mouse IgG1, Immunology Consultants Laboratory GG1-90P) at 1:20,000, 1:10,000, and 1:30,000 dilution in 1$\times$ PBS containing 5 mg/ml BSA and 0.25% (v/v) Tween 20) for 60, 45, and 50 min. Finally the blot was visualized using the SuperSignal West Pico ECL Substrate (Thermo Scientific 34078) and autoradiography film.

TABLE I
Cell and total protein concentrations obtained from whole saliva samples collected from OSCC-diagnosed individuals

Individual	Cells/ml ($\times 10^6$)	Protein amount
		mg/ml
1	1.42	0.11
2	1.56	0.16
3	0.84	0.06
4	2.52	0.22
Average ± S.D.	1.58 ± 0.7	0.14 ± 0.068

RESULTS

Because only a few groups have described the use of the cells in whole saliva for diagnostic studies and none have described its analysis using proteomics methods, we first investigated the dependability of whole saliva to provide adequate amounts of starting material for proteomics studies. Four whole saliva samples, collected using non-invasive methods (see "Experimental Methods") from patients displaying oral lesions diagnosed as OSCC, were centrifuged to obtain a cell pellet. Cell counting revealed that on average the whole saliva samples contained over 1 million cells/ml as shown in Table I. Total protein measurements after cell lysis revealed that each whole saliva sample provided more than 100 μg of total protein/ml on average (Table I). These results demonstrate that whole saliva samples contain dependable numbers of cells, providing more than enough protein for proteomics analysis.

Next to gain further insight into their potential for the detection of oral cancer, we characterized the type and condition of the cells present in the whole saliva samples. For these studies, an equal volume of whole saliva from each of the four oral cancer patients was combined, and the cell pellet was isolated by centrifugation. The cells were stained with a monoclonal antibody with broad affinity for human cytokeratin proteins, a standard marker of epithelial cells (39). The cells were also stained with the DNA-binding DAPI dye, which stains all cells containing intact DNA regardless of cell type. Fig. 1 shows a representative picture of the cells from whole saliva visualized by fluorescence microscopy. Although heterogeneous in their staining patterns, almost all of the cells detected showed positive staining for keratin (*yellow fluorescence* in Fig. 1), indicating exfoliated cells from the oral epithelium are the major constituent of the cells in whole saliva. The majority of these cells also showed intact cell morphologies, presumably still containing intact intracellular protein components and structures.

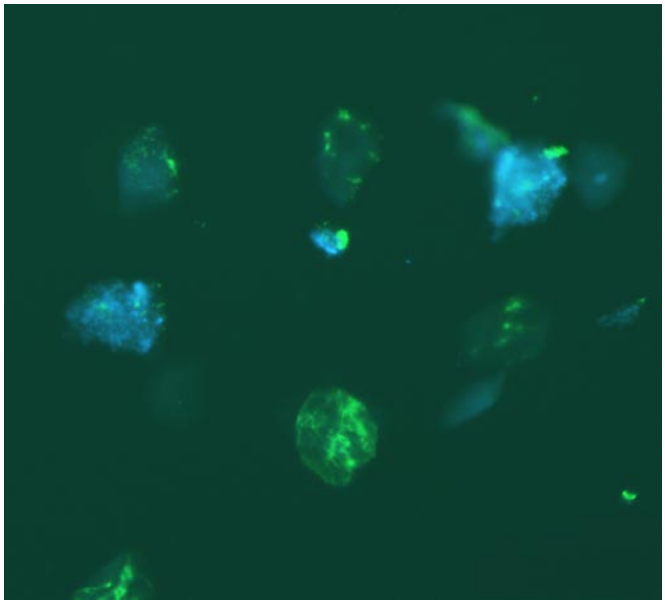

FIG. 1. **Immunostaining results of cells from whole saliva.** Cells were stained with a monoclonal antibody recognizing cytokeratin (*yellow fluorescence*) and DAPI DNA stain (*blue fluorescence*).

With the suitability of the samples established, we next sought to catalogue the proteins present within these samples. Fig. 2 outlines the overall strategy used in our proteomics method. To begin, we pooled equal amounts of saliva from the four OSCC patients to conserve sample from each patient for future studies. The cell pellet was isolated by centrifugation of whole saliva, and the proteins were extracted by cell lysis and then subjected to digestion with trypsin, resulting in a complex peptide mixture.

To sensitively identify proteins from these highly complex cellular samples, we developed a novel, three-step peptide fractionation method prior to MS/MS analysis. This three-step method is a modification to our previously described, two-step peptide fractionation method (23) combing FFE and μLC wherein we have added the intervening step of SCX to more extensively fractionate the complex peptide mixtures with the objective of increasing the coverage of proteins and our sensitivity for low abundance proteins. Each of these steps are *numbered* and in *bold text* in the flowchart shown in Fig. 2 and shown in a simplified schematic on the *right side* of the figure. After a first step of preparative IEF fractionation using FFE, single or combined FFE fractions are each subjected to a

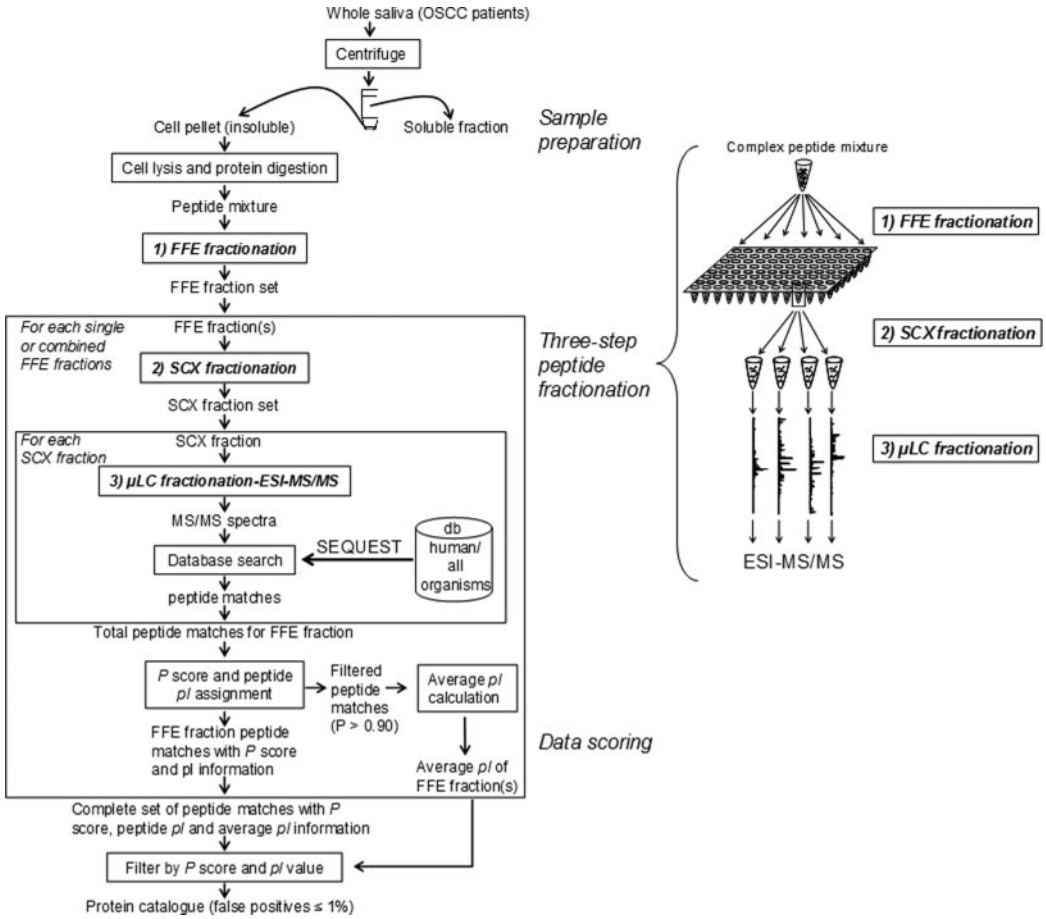

FIG. 2. **Sample preparation and proteomics method used.** The entire process consisting of sample preparation, three-step peptide fractionation, and data scoring is summarized, ultimately leading to a high confidence catalogue of proteins from cells in whole saliva. A simplified schematic of our three-step peptide fractionation method is also shown on the *right side* of the figure.

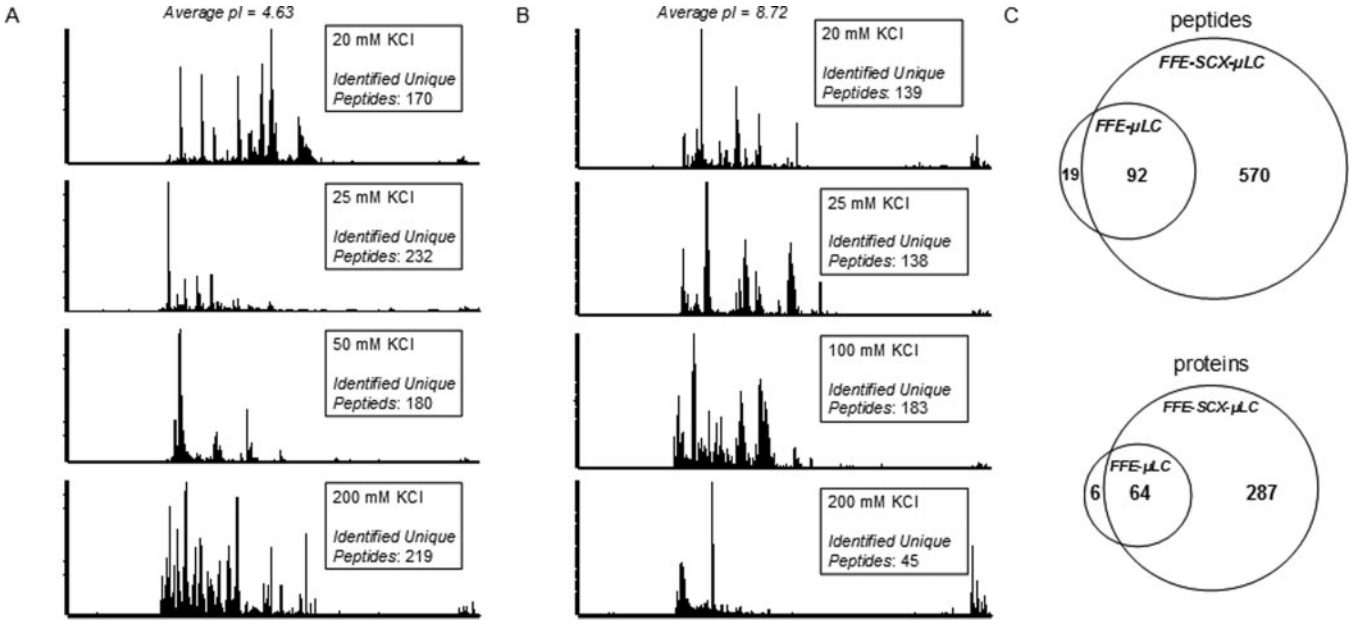

FIG. 3. **Testing the effectiveness of the three-step peptide fractionation method.** *A*, μLC mass chromatograms of SCX fractions for an acidic FFE fraction. *B*, μLC mass chromatograms of SCX fractions for a basic FFE fraction. *C*, number of peptides (*top*) and resulting proteins (*bottom*) identified from the same FFE fraction analyzed using either the two-step (*FFE-μLC*) or three-step fractionation (*FFE-SCX-μLC*) method.

second step of fractionation by SCX step elution chromatography, and each SCX fraction is analyzed by μLC-ESI-MS/MS using an LTQ linear ion trap mass spectrometer. Identical to our two-step fractionation method, peptide pI information added during the first stage of FFE fraction is used for filtering of peptide matches determined by sequence database searching to provide a high confidence catalogue of proteins (see description of this process below).

Given the novelty of our method, before proceeding to the large scale proteomics analysis of cells in whole saliva, we initially undertook studies to test its effectiveness. We first investigated the compatibility of coupling IEF peptide fractionation with a subsequent step of SCX chromatography. Because peptide fractionation using IEF and SCX have closely related physiochemical properties of peptide pI and solution phase charge, respectively, it was necessary to confirm that SCX effectively fractionated peptides of similar pI values contained in each FFE fraction. We selected an FFE fraction containing acidic peptide sequences (average predicted peptide pI, ~4.6) and a fraction containing basic peptide sequences (average predicted peptide pI, ~8.7) to test the compatibility of these different FFE fractions to SCX fractionation. Fig. 3, *A* and *B*, shows the results of each of these FFE fractions where each was separately loaded to the SCX cartridge and eluted by step gradient chromatography using increasing concentrations of salt at each elution step. Each elution fraction was then analyzed by μLC-ESI-MS/MS. We used a slightly higher salt concentration (100 mM) for the third SCX elution step for the basic peptide fraction compared with the acidic fraction (50 mM) because of the expectation that more of the basic peptides would have stronger affinity for the SCX column and elute at higher salt concentrations. For both the acidic and basic FFE fractions shown in Fig. 3, *A* and *B*, qualitative differences between the chromatograms of each different SCX fraction were observed, and the number of unique peptides identified in each fraction were evenly distributed, indicating resolution of distinct subsets of peptides with each SCX elution step.

We next measured the degree to which our three-step method can increase the number of peptides identified compared with our established two-step method. Our objective was to determine whether the increase in peptide sequence matches was significant enough to justify the extra labor involved with the SCX fractionation step. To measure the increase, we selected a single FFE fraction containing peptide sequences with acidic pI values, as these FFE fractions tend to contain the most peptides, and selected a portion of this fraction (about 10% of the total, and the maximum loading amount for the μLC system) for direct analysis by μLC-ESI-MS/MS, the same procedure used in our two-step method (23). Another portion of the FFE fraction (about 50% of the total) was subjected first to SCX step gradient chromatography, resulting in four fractions, each of which was analyzed by μLC-ESI-MS/MS. Fig. 3*C* compares the peptide sequence matches obtained and the number of proteins identified from these peptides for the two-step fractionation and the three-step fractionation. For each dataset of peptide matches with assigned PeptideProphet (36) probability scores (*p* scores), thresholds were used such that the false positive rate was estimated to be 1%. The results show that the three-step

method increases the number of peptides identified by almost 6-fold compared with the two-step method (111 peptides compared with 662 peptides, respectively) and increases the number of identified proteins by 5-fold (70 proteins compared with 351 proteins, respectively).

We next applied our method to the remaining FFE fractions to obtain a catalogue of proteins from cells in whole saliva. Single FFE fractions from the acidic region of the pH gradient were subjected to SCX fractionation followed by μLC-ESI-MS/MS analysis. For basic pH fractions, adjacent pairs of FFE fractions were combined prior to SCX fractionation because these fractions are less abundant in peptides than the acidic fractions (23) and combining these fractions helped to minimize the number of μLC-MS/MS analyses. After subjecting the resulting MS/MS spectra to sequence database searching using SEQUEST (34), we used the pI information, determined in the first FFE fractionation step, to filter the peptide sequence matches and obtain high confidence protein matches. As shown in the *bottom portion* of Fig. 2, for each FFE fraction or pair of adjacent fractions, we calculated an average peptide pI value using only high confidence peptide sequence matches (*p* score of 0.90 or greater) for this calculation and the algorithm described by Shimura *et al.* (35) to predict the pI value for each sequence. Although a quantitative comparison was not the objective of this study, the peptide mixture was labeled with the iTRAQ reagent to test the amenability of the labeled peptides with our fractionation method for possible future quantitative studies. Therefore, we accounted for iTRAQ reagent labeling at the N terminus and lysine residues in our pI calculations (see "Experimental Procedures"). The calculated average pI values for each FFE fraction were compared with their measured pH, and these values were within 0.5 pH unit (data not shown), similar to our past descriptions of preparative peptide fractionation by FFE (23). Based upon this finding, only peptide sequence matches with predicted pI values within 0.5 unit of the average pI for the FFE fraction from which they were derived were accepted as we have done in our past studies using FFE (23). *p* score thresholds were then applied to these peptide sequence matches at a level such that the estimated false positive rate was 1%.

Our proteomics method generated a catalogue of over 1000 human proteins identified by two or more unique peptides (5884 total peptides were identified from these proteins). Supplemental Table 1 shows all of these proteins along with annotations and relevant information on the peptide sequence matches determined by SEQUEST. Although not displayed in Supplemental Table 1, we identified another 847 proteins from matches to single peptides (*i.e.*"single hits"). The use of pI filtering increases the confidence of these single hit proteins despite the limited sequence coverage, and the catalogue of these proteins is available upon request. The human proteins were identified by searching the MS/MS data against the non-redundant International Protein Index (IPI) human protein database (40), minimizing the potential for matching to identical protein sequences that may be listed under multiple accession numbers. Use of pI information in the filtering of peptide sequence matches (as described under "Experimental Methods") enabled a lowering of the *p* score threshold for peptide matches while still maintaining a low estimated false positive rate of 0.24%. Similar to our past descriptions using peptide pI information (11, 23), the decreased *p* score thresholds resulted in significantly more proteins identified with high confidence, identifying an additional 1027 peptide sequences and adding 144 proteins to our catalogue when considering only those proteins identified from two or more peptides. The additional peptide sequences identified as a consequence of the pI filtering process are shown in red text in Supplemental Table 1.

A number of identified human proteins play a role in cell signaling and tumorigenesis pathways of OSCC. We searched our catalogue for proteins with known association to OSCC whose identification may be a result of our analysis of samples derived from patients with OSCC lesions. Table II shows a selection of these proteins from the catalogue, all of which are known to be expressed in epithelial cells. Three of these proteins (*STAT3*, *IKBKB*, and *LY6D*) were identified from a single peptide sequence (matched to multiple MS/MS spectra), and Supplemental Fig. 1 shows annotated MS/MS spectra supporting these sequence matches. Proteins involved in OSCC signaling pathways include the transcription factor *STAT3*, which putatively plays an early role in the development of OSCC (41–43), and also several proteins known to regulate the activity of the transcription factor NF-κB, which has a well described role in early events of cancer onset including OSCC (43, 44). The proteins *SCCA1* and *SCCA2* are tumor markers of OSCC and may play a role in its progression (45). Also identified were numerous membrane-spanning proteins with either confirmed or putative roles in cell adhesion and growth mechanisms important to OSCC tumor growth.

We also independently confirmed the presence of some of the proteins shown in Table II in cells from whole saliva collected from three additional patients diagnosed with OSCC. These were separate samples from the four patient samples that were combined and used for our proteomics analysis. We selected the proteins *STAT3*, *PRDX3*, and *SCCA1* from Table II for confirmation studies by immunoblotting. We detected each of these proteins in each of the three patient samples, further confirming the presence of these protein in cells from whole saliva and validating the results from our proteomics analysis. Results from our immunoblotting experiments are shown in Supplemental Fig. 2.

Because bacteria are known to be present within the oral environment (46) and play an important role in oral health and possibly even oral cancer development (47), we also interrogated our MS/MS dataset for the presence of proteins derived from bacterial sources. For this investigation, the entire MS/MS dataset was subjected to sequence database searching against the comprehensive UniProt protein sequence da-

TABLE II
Identified epithelial cellular proteins with described involvement in OSCC progression
GPI, glycosylphosphatidylinositol; TCR, T-cell receptor.

UniProt no.	Gene name	Protein name (N_{up}/N_{sm}[a])	Subcellular localization	Description	Ref.
OSCC signaling pathways					
P40763	*STAT3*	Signal transducer and activator of transcription 3 (1/8)	Cytoplasm/nucleus	Transcription factor that binds to the interleukin-6-responsive elements.	41–43
O43353	*RIPK2*	Receptor-interacting serine/threonine-protein kinase 2 (2/2)	Cytoplasm	Activates NF-κB.	43, 57
O14920	*IKBKB*	Inhibitor of nuclear factor κB kinase subunit β (1/3)	Cytoplasm	Activates NF-κB by phosphorylating its inhibitors, leading to their degradation.	58, 59
P30048	*PRDX3*	Thioredoxin-dependent peroxide reductase, mitochondrial (3/14)	Mitochondrion	Involved in redox regulation of the cell. Acts synergistically with MAP3K13 to regulate the activation of NF-κB in the cytosol.	60, 61
OSCC tumor markers					
P29508	*SCCA1*	Serpin B3 (3/5)	Cytoplasm	May act as a protease inhibitor to modulate the host immune response against tumor cells.	45
P48594	*SCCA2*	Serpin B4 (8/23)	Cytoplasm	May act as a protease inhibitor to modulate the host immune response against tumor cells.	45
OSCC cell adhesion and growth					
P27701	*CD82*	CD82 antigen (3/9)	Multipass membrane protein	Associates with CD4 or CD8 and delivers costimulatory signals for the TCR/CD3 pathway.	62
P08195	*SLC3A2*	4F2 cell surface antigen heavy chain (4/9)	Single pass type II membrane protein	Involved in sodium-independent, high affinity transport of large neutral amino acids. Required for normal and neoplastic cell growth.	63
Q14210	*LY6D*	Lymphocyte antigen Ly-6D (1/19)	Cell membrane; lipid anchor; GPI anchor	May be involved in cell-cell adhesion and signal transduction.	64
P21926	*CD9*	CD9 antigen (4/56)	Multipass membrane protein	Involved in platelet activation and aggregation. Involved in cell adhesion, cell motility, and tumor metastasis.	65
P09758	*TACSTD2*	Tumor-associated calcium signal transducer 2 (6/47)	Single pass type I membrane protein	GA733 tumor-associated antigen gene family may function as growth factor receptors.	66

[a] N_{up}, number of unique peptides identified from the protein. Supporting MS/MS spectra for the proteins STAT3, IKBKB, and LY6D, identified from single peptide sequences, are shown in Supplemental Table 1. N_{sm}, total number of MS/MS spectrum matches for the protein.

tabase (38), which contains proteins expressed in a variety of organisms including bacteria. As with the human proteins, the resulting peptide sequence matches, regardless of the organism from which they were derived, were filtered using peptide pI information and *p* score thresholds to obtain high confidence protein matches. Next those proteins identified from bacteria were extracted from the larger dataset for further analysis. The high confidence protein matches were grouped based upon the distinct bacteria in which they are expressed, and only bacteria with multiple peptide matches or matches to multiple different proteins were considered for acceptance. Of these, at least one of the peptide sequence matches also had to be specific to that bacterium (*i.e.* not a redundant peptide sequence found in proteins expressed in other organisms). Using these strict criteria, we identified proteins from 34 different bacteria. These bacteria are summarized in Table III. The organisms in bold text in Table III have not been reported previously in the literature as appearing in saliva.

TABLE III
Bacteria from which proteins were identified in the cell pellet of whole saliva

Organism name (proteins[a])	Description[b]
***Bacillus halodurans* (3)**	Produces many industrially useful alkaliphilic enzymes
Bacillus stearothermophilus (2)	A thermophile widely distributed in soil, hot springs, ocean sediment and a cause of spoilage in food products
Bacillus subtilis (3)	Metabolizes a wide variety of carbon sources and secretes large quantities of industrially important enzymes
B. fragilis (6)	A human pathogen with the potential to severely limit the success of gastrointestinal surgery
Bdellovibrio bacteriovorus (2)	Burrows into another microbe, killing it and using the remains to make offspring
***Bifidobacterium longum* (4)**	Keeps the human digestive system running smoothly
Borrelia burgdorferi (2)	A spirochete that is the causative agent of Lyme disease
***Buchnera aphidicola* (2)**	Bacterial symbiont with aphids
Campylobacter jejuni (2)	A food-borne pathogen that is the leading cause of food poisoning
***Caulobacter crescentus* (2)**	A model organism that is used in comparative analysis in the pathogenicity of other organisms
***Clostridium acetobutylicum* (3)**	Has the ability to convert starch into the organic solvents acetone and butanol
Clostridium tetani (3)	Causes tetanus disease
***Corynebacterium glutamicum* (2)**	A natural producer of glutamic acid
E. corrodens (2)	Anaerobic bacterium normally in the adult human oral cavity but that may become an opportunistic pathogen, especially in immunocompromised individuals
Escherichia coli (5)	One of the foremost causes of food poisoning, also found in the human digestive tract
Haemophilus influenzae (13)	An opportunistic pathogen involved in numerous diseases
Lactococcus lactis subsp. *lactis* (11)	Used in the dairy industry for cheese, yogurt, and buttermilk production
***Leifsonia xyli* subsp. *xyli* (3)**	Causes ratoon stunting disease in sugarcane
***Methanococcus jannaschii* (3)**	Cannot tolerate oxygen and is one of the methane-producing microbes
Mycobacterium leprae (2)	The bacterium that causes leprosy
Mycobacterium tuberculosis (2)	The major cause of tuberculosis
Mycoplasma genitalium (2)	Causes reproductive tract infections
P. multocida (5)	A pathogenic bacteria causing many serious diseases in humans and animals
Porphyromonas gingivalis (8)	One of the leading causes of periodontal disease
***Rhizobium loti* (2)**	A new species of legume root nodule bacteria
***Vibrio parahaemolyticus* (3)**	Causes food-borne gastroenteritis
Streptococcus mutans (5)	A dental pathogen, is the leading cause of tooth decay worldwide
Streptococcus pneumoniae (21)	Causes pneumonia, blood infections, and meningitis
***Streptomyces coelicolor* (3)**	Used to produce most of the natural antibiotics in use today
Thermoanaerobacter tengcongensis (2)	A thermophile
Treponema pallidum (3)	Causes syphilis
***Streptococcus equisimilis* [3]**	Can cause infections in wounds
***Rhodospirillum molischianum* [2]**	A non-sulfur purple bacteria with lamellar chromatophores
Streptococcus gordonii [2]	Has been found in small numbers on the oropharyngeal mucosa and in mature supragingival plaque

[a] Number of proteins identified from the organism. Organisms shown in bold have not been reported in the literature as being detected in whole saliva. For the last three bacteria shown underlined, only a single protein was identified but from multiple different peptide sequences with the number of different peptide sequences shown in brackets.

[b] Descriptions are from the European Molecular Biology Laboratory-European Bioinformatics Institute where applicable.

DISCUSSION

An essential feature of any sample with potential use in clinical diagnostics is the dependable availability of sample material collected from individuals in the clinic in amounts adequate for the methods used in its analysis. Given that few have investigated the cells in whole saliva as a diagnostic sample and none have studied its utility as a sample for proteomics analysis, determining the amount of cells and resulting protein material in individual patient saliva samples was critical for determining its potential for clinical diagnostics. Although of course there was some variability between the four individual whole saliva samples, each provided several hundred micrograms of total protein from typical volumes (~3 ml or more) of saliva collected in the clinic using non-invasive methods. These amounts of protein provide more than enough starting material for large scale proteomics studies, which usually require 100 μg or more of total starting protein. Therefore, our finding that the cells from non-invasively collected whole saliva dependably provide adequate protein material for proteomics analysis demonstrates its potential in studies discovering proteins relevant to the diagnosis of oral cancer and other conditions.

We also pursued cytological studies to characterize the makeup of the cells in whole saliva using immunostaining for cytokeratin expression, a well described marker for epithelial cells (32, 48). Because many researchers are interested in

using whole saliva as a bodily fluid for the early detection and diagnosis of OSCC, which predominantly affects oral epithelial cells (49, 50), these experiments provided an assessment of the suitability of exfoliated cells present in non-invasively collected whole saliva for studies of OSCC progression. The results showing the presence of mainly intact, keratin-expressing cells in whole saliva (Fig. 1) support their potential as samples for diagnostic tests of diseases of the oral epithelium such as OSCC. Our findings are consistent with a previous study by members of our research team that also detected a high proportion of intact cells present in saliva with the majority of these being dead cells based upon trypan blue staining (15). Our observation of different cellular keratin staining patterns is consistent with other descriptions using an antibody with broad affinity for different types of keratin, which have also shown a heterogeneous mixture of staining patterns indicating epithelial cells at different stages of differentiation (48, 51, 52).

Our findings answer two questions critical to evaluating our novel three-step peptide fractionation method. First, is the peptide IEF step orthogonal enough to the SCX step to be effectively coupled together? Because the effective fractionation of peptide mixtures in a multistep method is dependent upon each step resolving peptides based upon different (*i.e.* orthogonal) physiochemical properties, the closely related properties of peptide pI (IEF) and solution phase charge (SCX) could potentially limit the effectiveness of coupling these two steps in our method. Particularly acidic peptides often contain fewer basic amino acids, which might limit their retention and chromatographic resolution by SCX due to its dependence on the affinity of positively charged basic amino acid side chains for the negatively charged SCX packing material. However, the qualitatively different mass chromatograms shown in Fig. 3 demonstrate that both acidic peptides (Fig. 3*A*) and basic peptides (Fig. 3*B*) are effectively resolved by a second SCX fractionation step. Furthermore the roughly equal number of unique peptides identified from each different SCX fraction demonstrates the resolution of the peptide mixtures by SCX. Additionally although iTRAQ reagent-labeled peptides were used in the described studies, we have observed similar results on complex mixtures of non-labeled peptides. Collectively these results provide a strong "yes" to the question of orthogonality between the IEF and SCX steps.

Second, does the inclusion of a third peptide fractionation step significantly increase peptide sequence matches relative to a two-step method combining preparative IEF and μLC? Fig. 3*C* clearly answers this question, showing a greater than 6-fold increase in peptide sequence matches using the three-step fractionation method compared with a two-step fractionation on the same FFE peptide fraction. Our three-step fractionation method should therefore prove an effective and general method for large scale cataloguing of complex protein mixtures. Given the increasing use of preparative IEF in proteomics studies using a variety of different instrumental platforms (9, 11, 23–29), our findings should be useful for researchers wishing to further fractionate complex peptide mixtures post-IEF.

One slight drawback to our three-step fractionation method is decreased throughput due to multiple sample handling steps and the generation of the large number of fractions generated by the first two fractionation steps. From each FFE fraction (20 total) we generated four SCX fractions, resulting in 80 fractions needing to be analyzed by μLC-MS/MS. Using automated sample loading, μLC-MS/MS analysis of this set of fractions took over 4 days of instrument time. Although the instrument time needed to analyze such a large number of fractions cannot be easily decreased, automation of the peptide fractionation steps could decrease time spent on manual sample processing and help to increase throughput. Notably the second and third steps of our method (SCX and μLC-MS/MS) could be combined in an automated platform, like the popular multidimensional protein identification technology (MudPIT) method (18, 21), increasing the throughput of our method.

Our method retains peptide pI information, enabling the identification of more proteins with high confidence. Preparative peptide IEF using FFE adds information on the pI of fractionated peptides shown by our group (31) and others (24–26) to provide a valuable constraint for high confidence peptide sequence matching from MS/MS data. Because preparative IEF is the first step in our three-step method, peptide pI information is retained for all subsequent fractionation steps and can still be used for filtering the peptide sequence matches after sequence database searching. As shown at the *bottom* of Fig. 2, for each FFE fraction, after SCX fractionation and μLC-MS/MS analysis, an average peptide pI, adjusted for the presence of the iTRAQ reagent label, was calculated for the FFE fraction based upon high confidence peptide matches (p score >0.9) from all of the SCX fractions, and this average pI value was used to further filter the peptide matches. Our results showed that the average peptide pI calculated for these iTRAQ reagent-labeled peptides approximates the pH of the FFE fraction with accuracy similar to our previous descriptions of peptide fractionation by FFE (11, 23). Consequently only considering peptide sequence matches that have predicted pI values within 0.5 pH units of the average pI value to be correct increases the confidence of these matches. Others have similarly demonstrated the utility of using average peptide pI values to filter peptide sequence matches (24–26). This filtering enables the lowering of the p score threshold while still maintaining a low false positive rate ($\leq$1%) as estimated by reverse database searching. This decrease in the p score threshold increased the number of peptides identified by over 1000 and increased the size of our protein catalogue by almost 14%, further supporting the value of preparative IEF fractionation of peptides for MS/MS-based proteomics studies. The overall estimated false positive rate for our protein catalogue was 0.24%.

Among the proteins in our catalogue, the identification of proteins involved in cellular pathways linked to OSCC has 2-fold importance. First, it demonstrates for the first time that these epithelial cellular proteins can be detected in the cells contained in non-invasively collected whole saliva using proteomics methods or otherwise. All work in the literature describing proteins and biochemical pathways associated with OSCC development have been identified either from studies on model cell lines or from invasively collected patient tissue biopsies. For example, the role of the transcription factor *STAT3*, shown in Table II, has been characterized from a head and neck cancer cell line (43) and oral tissue biopsies (42); likewise the tumor inhibitor role of the *SCCA1* and *SCCA2* proteins was characterized in a model head and neck cancer cell line (45). The presence of both of these proteins in the cells from whole saliva was also confirmed by immunoblotting experiments in independent patient samples (see Supplemental Fig. 2). Our detection of these proteins with ties to OSCC directly from cells in whole saliva bolsters its potential as a sample for clinical diagnostics of oral cancer.

Second, the identification of proteins from signaling and tumorigenesis pathways associated with OSCC demonstrates the ability of our proteomics method for discovering proteins with potential diagnostic value in oral cancer detection directly from the cells in whole saliva. In particular, the identification of proteins involved in transcriptional control, such as *STAT3*, and kinases involved in signal transduction, such as *RIPK2* and *IKBKB*, indicate the ability of our proteomics method to detect relatively low abundance proteins with roles in OSCC development. The identification of numerous transmembrane-bound proteins with roles in cell adhesion and proliferation and ties to OSCC development demonstrates the ability of our method to detect this class of hydrophobic proteins, which challenge other proteomics methods such as those based upon two-dimensional gel electrophoresis (53). Although the identification of these proteins of potential diagnostic importance directly from cells in whole saliva is a significant first step, quantitative studies will be necessary to confirm their diagnostic power and, importantly, discover new protein markers of OSCC. Our demonstration here of the effectiveness of our proteomics method provides a foundation for these quantitative proteomics studies of the cells in whole saliva from patients at different stages of OSCC progression.

We also identified proteins derived from dozens of interesting bacteria within the cell pellet of whole saliva. Many of the bacteria have not been detected previously in whole saliva as determined by a search of the literature, which included results from a recent study of the soluble fraction of whole saliva using proteomics methods (9). Interestingly several of the bacteria detected have known associations with cancer, and their detection may be a consequence of our analysis of OSCC patient samples in this study, consistent with another recent study suggesting a role for bacteria in diagnosing OSCC (47). For example, *Eikenella corrodens* has been found in elevated levels of infection in head and neck cancer patients (54), and *Bacteroides fragilis* and *Pasteurella multocida* both produce toxins that are thought to promote cell proliferation in cancer (55, 56). Although further study is necessary to elucidate the role, if any, of these bacteria in OSCC development, their detection demonstrates the potential of our proteomics method for detecting bacteria within the oral environment and investigating their possible role in cancer development using quantitative proteomics methods.

CONCLUSIONS

Collectively our results presented here have confirmed the suitability of the cells in whole saliva as a source of potential protein markers of OSCC and established the effectiveness of our proteomics method for their discovery. We have demonstrated the effectiveness of a three-step peptide fractionation method for complex mixture analysis, showing for the first time the amenability of a first step of preparative peptide IEF with a second step of SCX chromatography as part of this method. Our results open the door to using our method for quantitative proteomics profiling via iTRAQ reagent labeling of protein samples derived from the cells in whole saliva collected from patients at different points in the OSCC progression process to identify protein changes associated with cancer development. We are currently initiating such studies, which should advance us toward the long term goal of developing reliable and inexpensive, protein-based clinical assays in whole saliva for the detection of oral cancer, leading to improved treatment and a drop in the mortality rate from this cancer.

Acknowledgments—We thank the Center for Mass Spectrometry and Proteomics facility at the University of Minnesota for instrumentation support and maintenance and the Minnesota Supercomputing Institute for maintenance of the SEQUEST cluster and general hardware and software support.

*This work was supported by NIDCR, National Institutes of Health Grant 1 R01 DE17734 and in part by grants from the Minnesota Medical Foundation, American Cancer Society Institutional Research Grant IRG-58-001-46, and a research award from Eli Lilly and Co. (to T. J. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

** To whom correspondence should be addressed: University of Minnesota, 321 Church St. SE, 6-155 Jackson Hall, Minneapolis, MN 55455. Phone: 612-624-5249; Fax: 612-624-0432; E-mail: tgriffin@umn.edu.

REFERENCES

1. Hofman, L. F. (2001) Human saliva as a diagnostic specimen. *J. Nutr.* **131,** 1621S–1625S
2. Lawrence, H. P. (2002) Salivary markers of systemic disease: noninvasive diagnosis of disease and monitoring of general health. *J. Can. Dent. Assoc.* **68,** 170–174
3. Landis, S. H., Murray, T., Bolden, S., and Wingo, P. A. (1998) Cancer

statistics, 1998. *CA Cancer J. Clin.* **48,** 6–29
4. Wong, D. T. (2006) Towards a simple, saliva-based test for the detection of oral cancer. 'Oral fluid (saliva), which is the mirror of the body, is a perfect medium to be explored for health and disease surveillance'. *Expert Rev. Mol. Diagn.* **6,** 267–272
5. Rhodus, N. L., Cheng, B., Myers, S., Miller, L., Ho, V., and Ondrey, F. (2005) The feasibility of monitoring NF-κB associated cytokines: TNF-α, IL-1α, IL-6, and IL-8 in whole saliva for the malignant transformation of oral lichen planus. *Mol. Carcinog.* **44,** 77–82
6. Rhodus, N. L., Ho, V., Miller, C. S., Myers, S., and Ondrey, F. (2005) NF-κB dependent cytokine levels in saliva of patients with oral preneoplastic lesions and oral squamous cell carcinoma. *Cancer Detect. Prev.* **29,** 42–45
7. Warnakulasuriya, S., Soussi, T., Maher, R., Johnson, N., and Tavassoli, M. (2000) Expression of p53 in oral squamous cell carcinoma is associated with the presence of IgG and IgA p53 autoantibodies in sera and saliva of the patients. *J. Pathol.* **192,** 52–57
8. Park, N. J., Zhou, X., Yu, T., Brinkman, B. M., Zimmermann, B. G., Palanisamy, V., and Wong, D. T. (2007) Characterization of salivary RNA by cDNA library analysis. *Arch. Oral Biol.* **52,** 30–35
9. Guo, T., Rudnick, P. A., Wang, W., Lee, C. S., Devoe, D. L., and Balgley, B. M. (2006) Characterization of the human salivary proteome by capillary isoelectric focusing/nanoreversed-phase liquid chromatography coupled with ESI-tandem MS. *J. Proteome Res.* **5,** 1469–1478
10. Hu, S., Xie, Y., Ramachandran, P., Ogorzalek Loo, R. R., Li, Y., Loo, J. A., and Wong, D. T. (2005) Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. *Proteomics* **5,** 1714–1728
11. Xie, H., Rhodus, N. L., Griffin, R. J., Carlis, J. V., and Griffin, T. J. (2005) Catalogue of human saliva proteins identified by free flow electrophoresis-based peptide separation and tandem mass spectrometry. *Mol. Cell. Proteomics* **4,** 1826–1830
12. Wilmarth, P. A., Riviere, M. A., Rustvold, D. L., Lauten, J. D., Madden, T. E., and David, L. L. (2004) Two-dimensional liquid chromatography study of the human whole saliva proteome. *J. Proteome Res.* **3,** 1017–1023
13. Ghafouri, B., Tagesson, C., and Lindahl, M. (2003) Mapping of proteins in human saliva using two-dimensional gel electrophoresis and peptide mass fingerprinting. *Proteomics* **3,** 1003–1015
14. Vitorino, R., Lobo, M. J., Ferrer-Correira, A. J., Dubin, J. R., Tomer, K. B., Domingues, P. M., and Amado, F. M. (2004) Identification of human whole saliva protein components using proteomics. *Proteomics* **4,** 1109–1115
15. Cheng, B., Rhodus, N. L., Williams, B., and Griffin, R. J. (2004) Detection of apoptotic cells in whole saliva of patients with oral premalignant and malignant lesions: a preliminary study. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* **97,** 465–470
16. Casartelli, G., Bonatti, S., De Ferrari, M., Scala, M., Mereu, P., Margarino, G., and Abbondandolo, A. (2000) Micronucleus frequencies in exfoliated buccal cells in normal mucosa, precancerous lesions and squamous cell carcinoma. *Anal. Quant. Cytol. Histol.* **22,** 486–492
17. Nunes, D. N., Kowalski, L. P., and Simpson, A. J. (2000) Detection of oral and oropharyngeal cancer by microsatellite analysis in mouth washes and lesion brushings. *Oral Oncol.* **36,** 525–528
18. Link, A. J. (2002) Multidimensional peptide separations in proteomics. *Trends Biotechnol.* **20,** S8–13
19. Issaq, H. J., Chan, K. C., Janini, G. M., Conrads, T. P., and Veenstra, T. D. (2005) Multidimensional separation of peptides for effective proteomic analysis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **817,** 35–47
20. Giddings, J. C. (1984) Two-dimensional separations: concept and promise. *Anal. Chem.* **56,** 1258A–1260A, 1262A, 1264A
21. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., III (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17,** 676–682
22. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2,** 43–50
23. Xie, H., Bandhakavi, S., and Griffin, T. J. (2005) Evaluating preparative isoelectric focusing of complex peptide mixtures for tandem mass spectrometry-based proteomics: a case study in profiling chromatin-enriched subcellular fractions in Saccharomyces cerevisiae. *Anal. Chem.* **77,** 3198–3207
24. Malmstrom, J., Lee, H., Nesvizhskii, A. I., Shteynberg, D., Mohanty, S., Brunner, E., Ye, M., Weber, G., Eckerskorn, C., and Aebersold, R. (2006) Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. *J. Proteome Res.* **5,** 2241–2249
25. Cargile, B. J., Bundy, J. L., Freeman, T. W., and Stephenson, J. L., Jr. (2004) Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. *J. Proteome Res.* **3,** 112–119
26. Krijgsveld, J., Gauci, S., Dormeyer, W., and Heck, A. J. (2006) In-gel isoelectric focusing of peptides as a tool for improved protein identification. *J. Proteome Res.* **5,** 1721–1730
27. Heller, M., Ye, M., Michel, P. E., Morier, P., Stalder, D., Junger, M. A., Aebersold, R., Reymond, F., and Rossier, J. S. (2005) Added value for tandem mass spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. *J. Proteome Res.* **4,** 2273–2282
28. Horth, P., Miller, C. A., Preckel, T., and Wenz, C. (2006) Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol. Cell. Proteomics* **5,** 1968–1974
29. An, Y., Fu, Z., Gutierrez, P., and Fenselau, C. (2005) Solution isoelectric focusing for peptide analysis: comparative investigation of an insoluble nuclear protein fraction. *J. Proteome Res.* **4,** 2126–2132
30. Wei, J., Sun, J., Yu, W., Jones, A., Oeller, P., Keller, M., Woodnutt, G., and Short, J. M. (2005) Global proteome discovery using an online three-dimensional LC-MS/MS. *J. Proteome Res.* **4,** 801–808
31. Xie, H., and Griffin, T. J. (2006) Trade-off between high sensitivity and increased potential for false positive peptide sequence matches using a two-dimensional linear ion trap for tandem mass spectrometry-based proteomics. *J. Proteome Res.* **5,** 1003–1009
32. Glogovac, J. K., Porter, P. L., Banker, D. E., and Rabinovitch, P. S. (1996) Cytokeratin labeling of breast cancer cells extracted from paraffin-embedded tissue for bivariate flow cytometric analysis. *Cytometry* **24,** 260–267
33. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3,** 1154–1169
34. Eng, J., McCormack, A. L., and Yates, J. R., III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989
35. Shimura, K., Kamiya, K., Matsumoto, H., and Kasai, K. (2002) Fluorescence-labeled peptide pI markers for capillary isoelectric focusing. *Anal. Chem.* **74,** 1046–1053
36. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392
37. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19,** 946–951
38. Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **33,** D154–D159
39. Bartek, J., Vojtesek, B., Staskova, Z., Bartkova, J., Kerekes, Z., Rejthar, A., and Kovarik, J. (1991) A series of 14 new monoclonal antibodies to keratins: characterization and value in diagnostic histopathology. *J. Pathol.* **164,** 215–224
40. Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4,** 1985–1988
41. Jewett, A., Head, C., and Cacalano, N. A. (2006) Emerging mechanisms of immunosuppression in oral cancers. *J. Dent. Res.* **85,** 1061–1073
42. Shah, N. G., Trivedi, T. I., Tankshali, R. A., Goswami, J. A., Jetly, D. H., Kobawala, T. P., Shukla, S. N., Shah, P. M., and Verma, R. J. (2006) Stat3 expression in oral squamous cell carcinoma: association with clinicopathological parameters and survival. *Int. J. Biol. Markers* **21,** 175–183
43. Squarize, C. H., Castilho, R. M., Sriuranpong, V., Pinto, D. S., Jr., and Gutkind, J. S. (2006) Molecular cross-talk between the NFκB and STAT3

signaling pathways in head and neck squamous cell carcinoma. *Neoplasia* **8,** 733–746
44. Bindhu, O. S., Ramadas, K., Sebastian, P., and Pillai, M. R. (2006) High expression levels of nuclear factor κB and gelatinases in the tumorigenesis of oral squamous cell carcinoma. *Head Neck* **28,** 916–925
45. Nakashima, T., Yasumatsu, R., Kuratomi, Y., Masuda, M., Kuwano, T., Toh, S., Umezaki, T., Cataltepe, S., Silverman, G. A., and Komune, S. (2006) Role of squamous cell carcinoma antigen 1 expression in the invasive potential of head and neck squamous cell carcinoma. *Head Neck* **28,** 24–30
46. Kolenbrander, P. E. (2000) Oral microbial communities: biofilms, interactions, and genetic systems. *Annu. Rev. Microbiol.* **54,** 413–437
47. Mager, D. L., Haffajee, A. D., Devlin, P. M., Norris, C. M., Posner, M. R., and Goodson, J. M. (2005) he salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *J. Transl. Med.* **3,** 27
48. Moll, R. (1998) Cytokeratins as markers of differentiation in the diagnosis of epithelial tumors. *Subcell. Biochem.* **31,** 205–262
49. Rhodus, N. L. (2005) Oral cancer: leukoplakia and squamous cell carcinoma. *Dent. Clin. N. Am.* **49,** 143–165, ix
50. Forastiere, A., Koch, W., Trotti, A., and Sidransky, D. (2001) Head and neck cancer. *N. Engl. J. Med.* **345,** 1890–1900
51. Moll, R., Franke, W. W., Schiller, D. L., Geiger, B., and Krepler, R. (1982) The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell* **31,** 11–24
52. Cooper, D., Schermer, A., and Sun, T. T. (1985) Classification of human epithelia and their neoplasms using monoclonal antibodies to keratins: strategies, applications, and limitations. *Lab. Investig.* **52,** 243–256
53. Roe, M. R., and Griffin, T. J. (2006) Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. *Proteomics* **6,** 4678–4687
54. Sheng, W. S., Hsueh, P. R., Hung, C. C., Teng, L. J., Chen, Y. C., and Luh, K. T. (2001) Clinical features of patients with invasive Eikenella corrodens infections and microbiological characteristics of the causative isolates. *Eur. J. Clin. Microbiol. Infect. Dis.* **20,** 231–236
55. Lax, A. J. (2005) Opinion: bacterial toxins and cancer—a case to answer? *Nat. Rev. Microbiol.* **3,** 343–349
56. Oswald, E., Nougayrede, J. P., Taieb, F., and Sugai, M. (2005) Bacterial toxins that modulate host cell-cycle progression. *Curr. Opin. Microbiol.* **8,** 83–91
57. Kobayashi, K., Inohara, N., Hernandez, L. D., Galan, J. E., Nunez, G., Janeway, C. A., Medzhitov, R., and Flavell, R. A. (2002) RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. *Nature* **416,** 194–199
58. Basseres, D. S., and Baldwin, A. S. (2006) Nuclear factor-κB and inhibitor of κB kinase pathways in oncogenic initiation and progression. *Oncogene* **25,** 6817–6830
59. Tamatani, T., Azuma, M., Aota, K., Yamashita, T., Bando, T., and Sato, M. (2001) Enhanced IκB kinase activity is responsible for the augmented activity of NF-κB in human head and neck carcinoma cells. *Cancer Lett.* **171,** 165–172
60. Jin, D. Y., Chae, H. Z., Rhee, S. G., and Jeang, K. T. (1997) Regulatory role for a novel human thioredoxin peroxidase in NF-κB activation. *J. Biol. Chem.* **272,** 30952–30961
61. Lo, W. Y., Tsai, M. H., Tsai, Y., Hua, C. H., Tsai, F. J., Huang, S. Y., Tsai, C. H., and Lai, C. C. (2007) Identification of over-expressed proteins in oral squamous cell carcinoma (OSCC) patients by clinical proteomic analysis. *Clin. Chim. Acta* **376,** 101–107
62. Imai, Y., Sasaki, T., Shinagawa, Y., Akimoto, K., and Fujibayashi, T. (2002) Expression of metastasis suppressor gene (KAI1/CD82) in oral squamous cell carcinoma and its clinico-pathological significance. *Oral Oncol.* **38,** 557–561
63. Kim, D. K., Ahn, S. G., Park, J. C., Kanai, Y., Endou, H., and Yoon, J. H. (2004) Expression of L-type amino acid transporter 1 (LAT1) and 4F2 heavy chain (4F2hc) in oral squamous cell carcinoma and its precusor lesions. *Anticancer Res.* **24,** 1671–1675
64. Schrijvers, A. H., Gerretsen, M., Fritz, J. M., van Walsum, M., Quak, J. J., Snow, G. B., and van Dongen, G. A. (1991) Evidence for a role of the monoclonal antibody E48 defined antigen in cell-cell adhesion in squamous epithelia and head and neck squamous cell carcinoma. *Exp. Cell Res.* **196,** 264–269
65. Kusukawa, J., Ryu, F., Kameyama, T., and Mekada, E. (2001) Reduced expression of CD9 in oral squamous cell carcinoma: CD9 expression inversely related to high prevalence of lymph node metastasis. *J. Oral Pathol. Med.* **30,** 73–79
66. Nakashima, K., Shimada, H., Ochiai, T., Kuboshima, M., Kuroiwa, N., Okazumi, S., Matsubara, H., Nomura, F., Takiguchi, M., and Hiwasa, T. (2004) Serological identification of TROP2 by recombinant cDNA expression cloning using sera of patients with esophageal squamous cell carcinoma. *Int. J. Cancer* **112,** 1029–1035