CANCER GENOMICS & PROTEOMICS *4*: 55-64 (2007)

Discovery of Oral Fluid Biomarkers for Human Oral Cancer by Mass Spectrometry

SHEN HU[1,2], TIANWEI YU[1], YONGMING XIE[3], YANAN YANG[3], YANG LI[1], XIAOFENG ZHOU[1,5], STEPHANIE TSUNG[1], RACHEL R. LOO[2,4,6], JOSEPH A. LOO[2-6] and DAVID T. WONG[1,5-8]

[1]School of Dentistry and Dental Research Institute, [2]Mass Spectrometry and Proteomics Center, [3]Department of Chemistry and Biochemistry, [4]Department of Biological Chemistry, [5]Jonsson Comprehensive Cancer Center, [6]Molecular Biology Institute, [7]Division of Head and Neck Surgery/Otolaryngology, School of Medicine and [8]Henry Samueli School of Engineering and Applied Science, University of California Los Angeles, Los Angeles, CA, U.S.A.

Abstract. *Proteomic analysis of human oral fluid (whole saliva) holds promise as a non-invasive method to identify biomarkers for human oral cancer, a high impact local disease in the oral cavity affecting 38,000 Americans and with 350,000 cases worldwide annually. In this study, matrix-assisted laser desorption/ionization – mass spectrometry (MALDI-MS) was used to profile oral fluid samples from oral cancer and control subjects, and 46 peptides/proteins were found at significantly different levels between the two groups. To identify a candidate protein biomarker, oral fluid samples were separated by liquid chromatography (LC) using a C4 reversed-phase column. The collected LC fractions were monitored by MALDI-MS and the fraction containing the candidate biomarker was digested for LC-MS/MS analysis to identify it. The use of nanospray MS/MS for the identification of candidate peptide biomarkers was also demonstrated. This approach can be useful for the identification of protein or peptide biomarkers following MALDI-MS or surface-enhanced laser desorption/ionization MS profiling of clinical samples. This study clearly demonstrated that oral fluid contains proteomic signatures that may serve as biomarkers for human diseases such as oral cancer. Once discovered and validated on a large and independent clinical cohort, oral fluid proteomic biomarkers may be extensively used for future disease diagnosis.*

Correspondence to: Dr. David Wong, School of Dentistry and Dental Research Institute, University of California Los Angeles, Los Angeles, CA, U.S.A. / Dr. Joseph Loo, Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA, U.S.A.

Key Words: Cancer biomarker, oral fluid diagnostics, oral squamous cell carcinoma, QqTOF MS.

Biomarkers are measurable and quantifiable biological parameters that can serve as indices for health-related assessments, such as disease diagnosis, environmental exposure and its effects, metabolic processes and epidemiological studies. As proteomic tools mature, protein biomarker discovery becomes a central application of proteomics. Profiling proteins in tissues or body fluids over the course of disease progression could reveal potential biomarkers indicative of specific disease status, which may be used extensively for future medical diagnostics. Similarly, analyses of protein profiles before and after pharmacological treatment may provide vital clues regarding drug efficacy and toxicity.

Proteomics is a powerful approach for global study of the structure and function of all proteins expressed in a biological system. Mapping proteomes, the protein complements to genomes, from tissues, cells and organisms is being used to identify new protein targets, to explore mechanisms of action or toxicology of pharmaceuticals, and to discover new disease biomarkers for clinical and diagnostic applications. Due to its particular sensitivity and highly accurate mass measurements, mass spectrometry (MS) has become one of the core technologies for proteomics. A very common proteomics approach is to map out proteins using 2-D gel electrophoresis (2-DE) followed by in-gel digestion and MS measurement of the resulting peptide fragments, either by mass fingerprinting or tandem MS (MS/MS), to identify proteins from each gel spot. Such global analysis can also be performed by using a "shotgun" proteomics approach, in which 2-D liquid chromatography (LC) and tandem MS are typically used to analyze a fully digested proteomic sample. Both approaches rely on searching protein or DNA sequence databases to identify the proteins.

Because proteomics aims at global profiling and the characterization of proteins, it is a powerful approach to

discover protein biomarkers in human diseases for clinical and diagnostic applications. A variety of MS techniques, including surface-enhanced laser desorption/ionization (SELDI), matrix-assisted laser desorption/ionization (MALDI), MALDI-TOF/TOF and LC-MS/MS have been extensively applied to biomarker study in human cancers. SELDI provides rapid profiling of peptides and small proteins and has been applied to biomarker studies in ovarian, breast and prostate cancer (1-4). Offline the LC-MALDI TOF/TOF-based platform offers several advantages including high throughput, mass range and accuracy, selective MS/MS analysis based on MS information, and an easy to interpret data structure (5). Peptide identification can be performed on the same MALDI plate afterwards by MS/MS. The ability to re-examine and verify the same sample set can be very beneficial for quantitative applications. MS with isotope-coded tagging has the potential for high sensitivity, coverage and throughput (6). This gel-free proteomic strategy also holds significant promise for the rapid identification of protein markers/targets for human diseases (7, 8).

Such a cancer biomarker discovery study also requires a robust statistical methodology for data analysis and building prediction models in the task of forecasting malignant potential. A major biological difficulty confronting statistical prediction model building is the multifactorial nature of oncogenesis. Most tumors result from an inter-dependent series of genetic alterations, rather than a single decisive event. To overcome the difficulty, multiple statistical strategies are often used for prediction model development to identify a combination of individual biomarkers that may indicate increased risk above and beyond the presence of each individual abnormality (9, 10). A robust statistical strategy will not only reveal signatures between disease and normal control samples, but will also build prediction models and, therefore, identify a combination of protein biomarkers representing the best sensitivity and specificity for a specific human cancer.

One might think tumor tissue is the ideal sample for protein biomarker studies. However, in terms of diagnostic applications, human body fluids (*e.g.*, blood, oral fluid, urine, *etc.*) are attractive because they are simple to collect and process. Human oral fluid, the product of multiple salivary glands (parotid, submandibular, sublingual and other minor glands) lying beneath the oral mucosa, contains a large array of proteins such as enzymes and enzyme inhibitors, growth factors and cytokines, immunoglobulins, mucins and other glycoproteins. We have recently conducted a human saliva proteome analysis using 2-D gel electrophoresis/MS and shotgun proteomics (11-13), to provide a "dictionary" for the discovery of salivary protein biomarkers for human diseases. Human oral fluid is especially attractive for disease diagnosis because: i) its collection is totally non-invasive as compared to blood for serum/plasma analyses; ii) many, if not all blood components, are reflected in oral fluid. This bio-fluid has been proven to be very valuable for the diagnosis of HIV, periodontal diseases and hepatitis (14-16). An oral fluid assay has been on the market for HIV, and may soon be marketed for the protein c-erbB-2, which is a prognostic breast cancer marker, assayed in tissue biopsies of women diagnosed with malignant tumors (17). The soluble fragments of the c-erbB-2 oncogene and the cancer antigen 15-3 (CA 15-3) were significantly higher in the oral fluid samples of women who had cancer than in those of healthy controls and patients with benign tumors. Pilot studies have indicated that the oral fluid test for this oncogene is sensitive, reliable and potentially useful in the initial detection of and/or follow-up screening for breast cancer (18).

Squamous cell carcinoma of the oral cavity and oropharynx is the sixth most common cancer worldwide, with approximately 38,000 new cases in the U.S. and 350,000 new cases worldwide each year (19). The overall 5-year survival rates for oral cancer have remained low at approximately 30-40% for the past decades, being considerably lower than that for colorectal, cervical and breast cancer (20, 21). Delayed detection is one of the main reasons for the high morbidity rate of oral cancer, which suggests an imperative need to develop biological markers in order to improve its detection. In this work, we described a proteomic approach, by combining MALDI-MS profiling with LC fractionation and QqTOF MS/MS, to discover and identify oral fluid biomarkers for human oral cancer. The study clearly demonstrated that oral fluid is a valuable body fluid for the diagnosis of human diseases such as oral cancer.

Materials and Methods

Sample collection and processing. All the oral squamous cell carcinoma (OSCC) patients involved in this study had not received any prior treatment in the form of chemotherapy, radiotherapy, surgery, or alternative remedies prior to sample collection. An equal number of age- and gender-matched subjects with comparable smoking histories were selected as a control group. Among the two subject groups, there were no significant differences in terms of mean age, gender or smoking history. No subjects had a history of prior malignancy, immunodeficiency, autoimmune disorders, hepatitis, or HIV infection. All of the subjects signed the institutional review board-approved consent form. Unstimulated oral fluid samples were collected between 9 a.m. and 10 a.m. with prior mouth rinsing with water (22). The donors were asked to abstain from eating, drinking, smoking, or using oral hygiene products for at least 1 hour prior to collection. A well-defined and standardized protocol was used for the collection, storage and processing of the oral fluid specimens. The oral fluid samples were centrifuged at 2,600 xg for 15 minutes at 4˚C to remove debris and cells. The supernatant was removed from the pellet and protease inhibitors were included in the collected samples to ensure preservation of the protein integrity (Sigma, 2 μL per mL oral fluid). The samples were divided into 1-mL aliquots and stored at –80˚C.

MALDI-TOF MS profiling and data analysis. MALD-TOF MS (Applied Biosystems, Foster City, CA, USA) profiling of oral fluid samples from 20 oral cancer and 20 control subjects was performed using either α-cyano-4-hydroxycinnamic acid (CHCA, 10 mg/mL in 50% ACN containing 0.1% TFA) or sinapinic acid (SA, 12.5 mg/mL in 50% ACN containing 0.1% TFA) as the matrix. The oral fluid sample was diluted 10 times with the matrix solution and 1 μL of the mixture was spotted on the MALDI plate, in duplicate, for MALDI-TOF MS analysis.

The initial processing of MALDI data included baseline correction, noise filtering and peak calibration. The MALDI signals were averaged and normalized for subsequent statistical analysis to reveal peptides/proteins at significant difference between the cancer and control populations. All statistical analyses were performed by R 1.9.1 and S-plus 6.2. The Student's *t*-test was used to identify peptides that were differentially-expressed in cancer patients. To examine the ability of the selected peptides to differentiate between normal and cancer, Principal Component Analysis (PCA) was first used to reduce the dimension and visually examine the separation between the two groups. Then, classification models were conducted using diagonal linear discriminant analysis (DLDA), which is a form of "weighted voting" scheme (23), and support vector machine (SVM) with linear kernel (24).

LC fractionation and off-line MALDI-MS measurement of LC fractions. LC separation of human oral fluid samples (114 mg total proteins) was performed using an Agilent (Santa Clara, CA, USA) 1100 LC with a Vydac C4 reverse-phase column (2.1 mm x 250 mm, 5 mm; Grace Vydac, San Diego, CA, USA). The eluents used for the LC were (A) 0.1% TFA and (B) 95% ACN/0.1% TFA. The flow rate was 0.25 mL/min and the gradient was from 5% B to 85% B within 40 minutes. Thirty fractions (one fraction per minute) were automatically collected, equally divided and then completely dried using a speed vacuum. Subsequent MALDI-MS measurement of the fractions was performed using both CHCA (10 mg/mL in 50% ACN containing 0.1% TFA) and SA (12.5 mg/mL in 50% ACN containing 0.1% TFA) matrices. Each fraction was dissolved in 5 μL matrix solution and 1 μL of each sample was then spotted on the MALDI plate for measurement. Therefore, two sets of MALDI spectra were produced and used to match with the original MALDI-MS spectra to determine which LC fractions contained the candidate peptide or protein.

LC-MS/MS. Tryptic digestion was performed on the LC fractions containing the potential protein/peptide biomarkers. Each fraction was reduced with 10 mM DTT for 1 hour, alkylated with 50 mM iodoacetamide for 1 hour and then digested with 0.2 μg sequence-grade trypsin (in 50 mM NH_4HCO_3) at 37˚C for overnight. LC-MS/MS of peptide digest was conducted on a nano-LC system (LC Packings, Sunnyvale, CA, USA) with a nanoelectrospray (nano-ESI) interface (Protana, Odense, Denmark) and an Applied Biosystems/Sciex QSTAR® XL (QqTOF) mass spectrometer (Foster City, CA, USA). The samples were first loaded onto an LC Packings PepMap C18 precolumn (300 μm x 1 mm; particle size 5 μm) and washed for five minutes with the loading solvent, 0.1% formic acid. The samples were then eluted onto a LC Packings PepMap C18 column (75 μm x 150 mm; particle size 5 μm) for nano-LC separation at a flow rate of 200 nL/min. The eluents used for the LC were (A) 0.1% formic acid and (B) 95% ACN/5% H_2O/0.1% formic acid, and the gradient was from 5% B to 85% B within 40 minutes. For online LC-MS/MS analyses, a Proxeon (Protana, Odense, Denmark) nano-bore stainless steel online emitter (30 μm i.d.) was used for electrospray ionization at 1900 V. Argon was used as the collision gas. Peptide product ion spectra were automatically recorded during the LC-MS runs by information-dependent analysis on the QqTOF mass spectrometer.

Nanospray MS. Prior to nanospray MS analysis, a control oral fluid sample was pretreated with Millipore ultracentrifuge filter Microcon YM-3K (Billerica, MA, USA) to remove salts and small molecules. Nanospray MS/MS analysis was performed using a QSTAR® XL (QqTOF) mass spectrometer with a nano-ESI interface. Four μL of filtered sample was loaded into a nanospray emitter (New Objective, Woburn, MA, USA) and ionization of the selected intact peptides was conducted at 1000 V.

Database searching. LC-MS/MS and nanospray MS/MS data were processed to search the Swissprot protein database using the MASCOT search engine (Matrix Science, London, UK). Standard search parameters were used, including mass tolerance of 0.3 Da for the precursor ion and 0.2 Da for fragment ions as well as single missed cleavage.

Western blotting. Western blotting was used to validate the thioredoxin levels on a new independent set of oral fluid samples (24 oral cancer/24 controls). The proteins were separated on 12% SDS-PAGE gels (15 lanes, Bio-Rad Laboratories, Hercules, CA, USA) at 150 V for about 1 hour and then transferred to polyvinylidene difluoride membrane using a Mini Trans-Blot electrophoretic transfer cell (Bio-Rad). After saturating with 5% milk in TBST buffer overnight at 4˚C, the blots were sequentially incubated with primary mouse monoclonal thioredoxin antibody (Abcam, Cambridge, MA, USA; 1:1000 dilution) and horseradish peroxidase-conjugated anti-mouse IgG secondary antibody (Amersham, Piscataway, NJ, USA; 1:5000 dilution). Finally, the bands were visualized by enhanced chemiluminescence (Amersham).

Results and Discussion

We have demonstrated a MS-based approach to identify candidate proteins and peptides in oral fluid as biomarkers for oral cancer detection. The overall experimental strategy is summarized in Figure 1. Initially, MALDI-TOF MS was used to profile oral fluid samples from 20 oral cancer and 20 control subjects. The MALDI-TOF MS data were processed and normalized for subsequent statistical analysis to reveal a panel of proteins and peptides present at significantly different levels between the cancer and control populations. To identify the candidate markers, oral fluid samples were separated by HPLC using a C4 reversed-phase column with fraction collector. MALDI-TOF MS was then used to monitor those fractions and the fractions containing candidate masses were tryptically digested and analyzed by LC-MS/MS. Subsequently MASCOT database searching was conducted to identify the candidate biomarkers. Nanospray MS/MS, followed by database searching, was

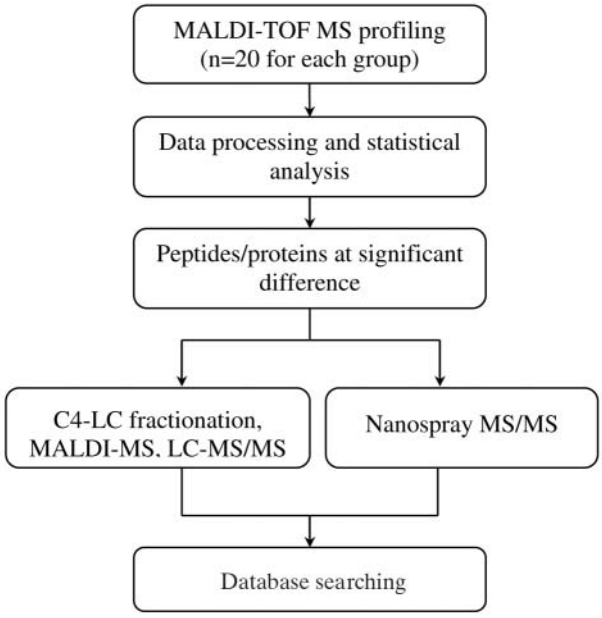

Figure 1. *Outline of the strategy for identification of potential biomarkers using mass spectrometry. MALDI-TOF MS was used for profiling of oral fluid samples from 20 oral cancer and 20 control subjects. The obtained data were processed and statistical analysis was conducted to reveal peptides/proteins at significant difference between the two groups. In order to identify potential biomarkers, oral fluid samples were fractionated by HPLC. Subsequently, MALDI-MS was used to monitor LC fractions, and the fractions containing candidate biomarkers were processed for biomarker identification using LC-MS/MS analysis and database searching. Intact peptide biomarkers were also analyzed by nanospray MS/MS and identified by database searching.*

also employed for the identification. Alternatively, the initial step by MALDI-TOF MS measurement can be performed by other MS profiling techniques such as SELDI-TOF MS, MALDI-Fourier Transform Ion Cyclotron Resonance (FTICR) MS or LC-ESI-MS. Depending on whether the candidate is a peptide or protein, an appropriate reversed-phase column (C18, C8 or C4) can be chosen for optimum separation and fraction collection. Once the fraction containing a candidate biomarker has been confirmed, it is straightforward to identify the candidate using tandem MS and database searching.

MALDI-TOF MS profiling and data analysis. MALDI-TOF MS profiling was performed on the oral fluid samples from 20 oral cancer patients and 20 matched control subjects, using CHCA (10 mg/ml) and SA (12.5 mg/ml) as the matrix. Without any pre-treatment, the samples were directly diluted (1:10) with the matrix solution for MALDI-MS measurement. Figure 2 presents the MALDI-MS spectra of oral fluid samples from an oral cancer patient and a control subject, using two different matrices. The MALDI peaks are well aligned, however, their abundance varies between the cancer and control samples. The two MALDI matrices revealed very different sets of peaks. In total, 1623 features were observed from 20 OSCC and 20 control subjects by using two MALDI matrices.

Analysis of the acquired MALDI-MS data included two steps: data pre-processing and statistical analysis. The data pre-processing was performed according to a uniform set of procedures. Baseline correction was initially conducted to eliminate broad artifacts from the data set followed by noise filtering to remove noisy spikes. Afterwards, the MALDI spectra were calibrated using a set of commonly shared peaks (1471, 1867, 2520, 4368 for CHCA; 4439, 5378, 5796, 11159 for SA) to ensure accurate alignment of peaks across the whole set of spectra. Finally, peak normalization was performed against Total Peak Intensity (multiplied by a scaling factor of 10000), which is similar to the normalization method used in SELDI (25, 26). These steps ensured comparability of MALDI spectra among different subjects (cancer and controls). Statistical analysis (t-test) of the processed MALDI-MS data (1623 features in total) indicated that 46 peptides/proteins were at significantly higher or lower levels ($p<0.05$) between cancer and control populations (Table I). Interestingly, a small protein with m/z of 5584 was shown at significantly lower levels in oral cancer by both MALDI matrices. Figure 3A presents the expression levels of those peptides/proteins as a heatmap, and more down-regulated peptides are observed than up-regulated. After dimension reduction by PCA analysis, we observed that the peptides contained sufficient information to stratify the control/cancer groups (Figure 3B). It is of interest to note that, the panel of peptides/proteins significant for oral cancer is completely distinct from the one we observed for primary Sjögren's syndrome, which is a chronic autoimmune disorder clinically characterized by xerostomia (dry mouth) and xerophthalmia (dry eyes) (data not shown). This suggests that the panels of discriminatory salivary peptides/proteins are likely to be different for different diseases.

For the discovered candidates, validation can be realized if their antibodies are available from either commercially or from other researchers. However, validating peptide biomarkers is challenging due to the general lack of highly specific means to target the specific peptide. To further analyze the ability of the candidate peptide markers to discriminate cancer from normal, we fitted the DLDA and SVM models and evaluated the generalization error rate by leave-one-out cross-validation. Considering the sample size, we only used the top 13 peptides that had t-test p-values <0.02. The DLDA model fitted 87.5% (35/40) of the data correctly. In cross-validation, the success rate was 85% (34/40). The SVM with linear kernel fitted the data with 95% (39/40) accuracy, while showing a cross-validation success rate of 90% (36/40).

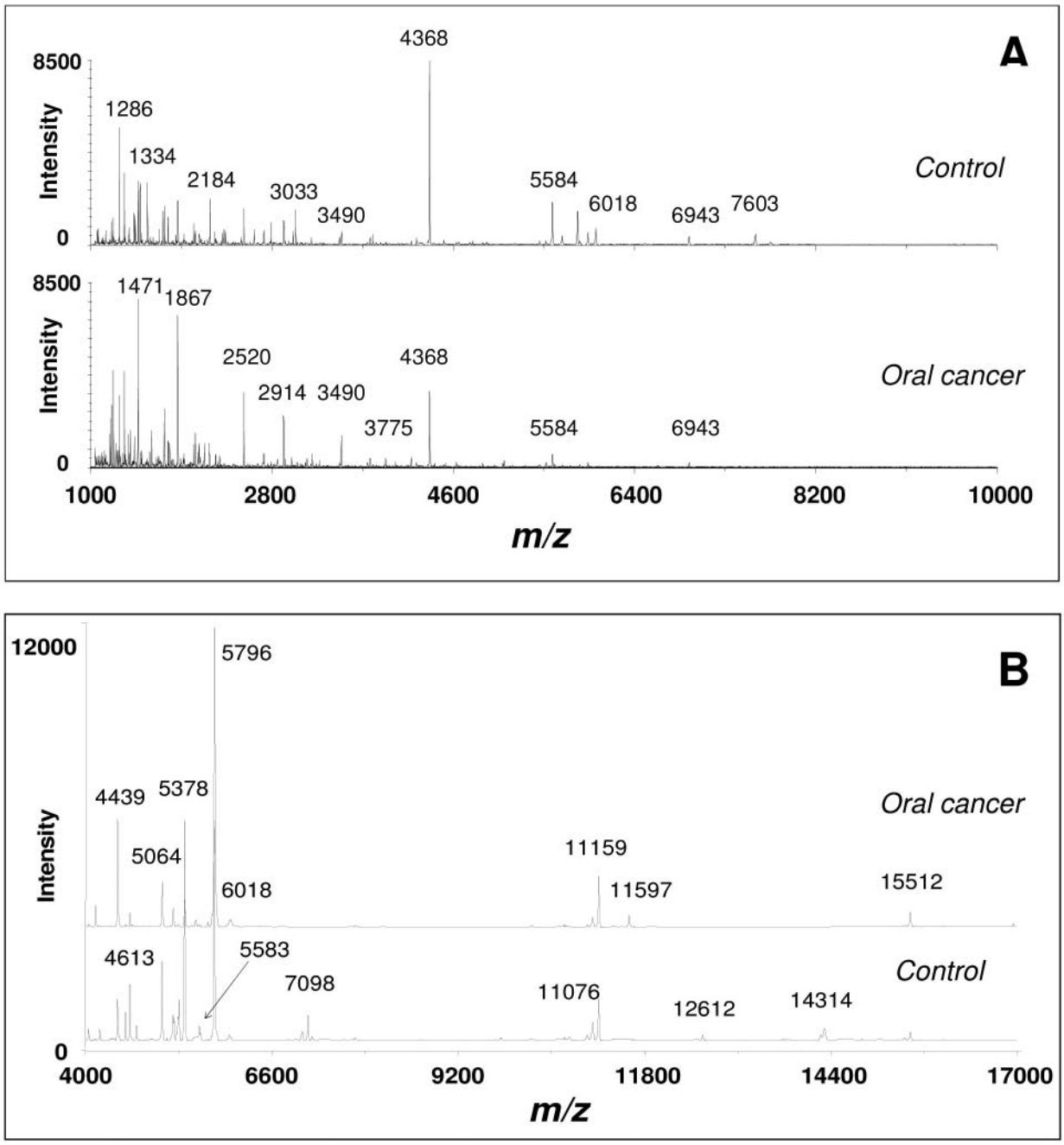

Figure 2. *MALDI-TOF MS profiling of oral fluid samples from oral cancer and control subjects. MALDI-MS experiment was performed using either (A) α-cyano-4-hydroxycinnamic acid (10 mg/ml) or (B) sinapinic acid (12.5 mg/ml) as the matrix. The saliva sample was diluted 10 times with the matrix solution and 1 μl of the mixture was used for MALDI profiling.*

LC fractionation and MALDI measurement of LC fractions. Previous reports of cancer biomarkers identified by SELDI or MALDI typically show panels of candidates with only their masses listed. Few studies provided the sequence information of those candidates. In this work, we have established a robust method to sequence candidate protein/peptide biomarkers. By performing MALDI-MS profiling and data analysis of 20 cancer/20 control samples, we have shown that a protein with a molecular weight of 11,597 Da is present at significantly higher levels in oral cancer patients' oral fluid. In order to identify this 11,597-Da protein, the oral fluid sample from an oral cancer

Table I. *Peptides and proteins at significant difference between oral cancer and control groups ($p<0.05$). MALDI-TOF MS was used for profiling of oral fluid samples from 20 oral cancer and 20 control subjects.*

m/z	Control (mean, n=20)	Cancer (mean, n=20)	*p*
1070	2.300	4.364	0.031
1073	1.204	0.279	0.045
1150	1.703	0.481	0.011
1156	0.687	0.076	0.032
1214	0.598	0.143	0.045
1225	19.400	41.833	0.013
1240	1.644	0.638	0.035
1262	0.851	2.745	0.046
1300	0.577	0.033	0.035
1323	0.138	1.183	0.026
1381	3.925	0.335	0.035
1394	0.000	0.558	0.020
1426	0.755	0.184	0.019
1430	0.567	0.000	0.024
1480	0.000	1.147	0.031
1491	4.554	0.946	0.033
1517	0.957	0.151	0.021
1605	1.973	5.644	0.015
1619	2.487	0.381	0.007
1674	1.072	5.527	0.040
1732	16.945	40.302	0.019
1819	4.817	10.956	0.032
1905	3.822	1.735	0.022
1927	0.400	0.040	0.043
1936	0.409	0.000	0.027
1942	0.469	0.000	0.033
2081	1.141	0.113	0.029
2183	4.606	0.795	0.008
2558	1.657	0.485	0.017
2623	3.453	0.085	0.014
3016	0.203	0.000	0.038
3034	9.214	0.711	0.007
3123	0.378	0.000	0.047
3800	0.932	0.000	0.006
4185	4.685	2.214	0.026
4192	18.468	7.816	0.038
4368	43.258	18.900	0.003
5568	2.217	0.216	0.022
5584	9.900	3.401	0.015
5601	1.998	0.295	0.038
6000	1.506	5.479	0.028
6931	2.517	0.472	0.043
7100	22.334	6.352	0.040
7308	0.879	0.000	0.046
11597	1.527	3.882	0.032
12612	3.417	0.261	0.004

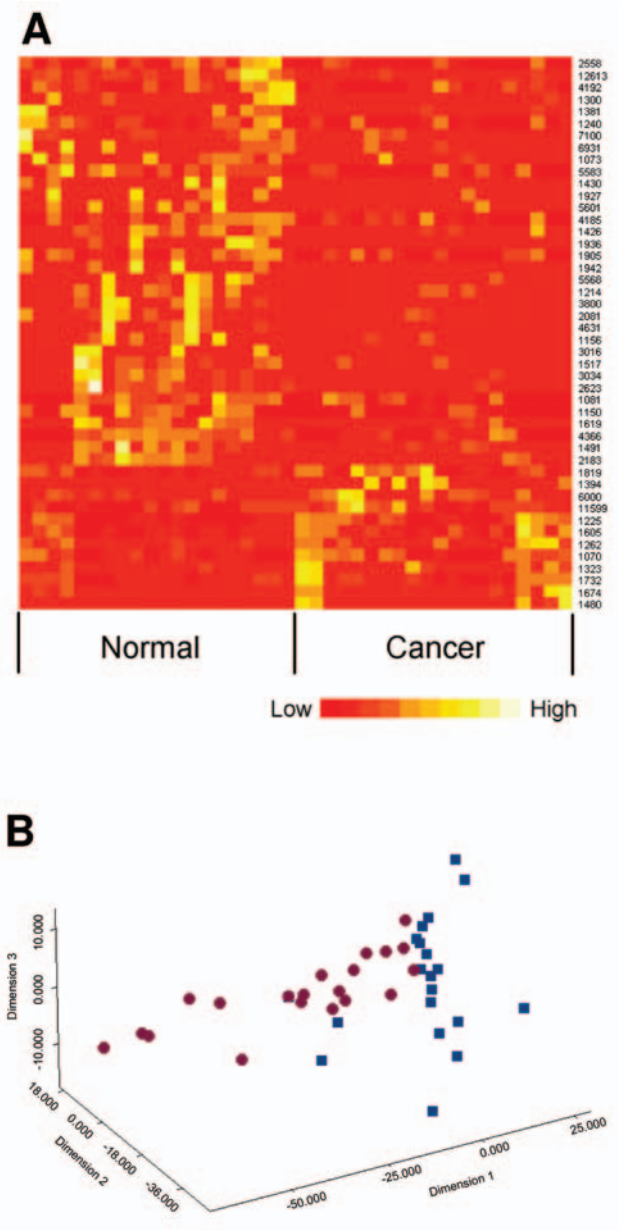

Figure 3. *Heatmap (A) of 46 peptides/proteins showing significantly differential levels between two populations. 3-D multidimensional scaling (B) shows that the peptides contain sufficient information to separate the control/cancer groups.*

patient was separated by HPLC using a C4 reversed-phase column and 30 fractions were collected on a 1-min basis. These fractions were then monitored with MALDI-MS, using both CHCA and SA as matrices, and the fractions containing the candidate markers were determined for further identification experiments. The MALDI-MS spectrum of an LC fraction is shown in Figure 4 A. This fraction was found to contain the 11,597-Da protein (arrow pointed), which was at significantly higher levels in cancer samples (Figure 2B).

LC-MS/MS and nanospray MS/MS identification. Figure 4 B depicts the tandem MS spectrum of a double-charged, tryptic peptide (*m/z*, 697.4). The precursor ion was well

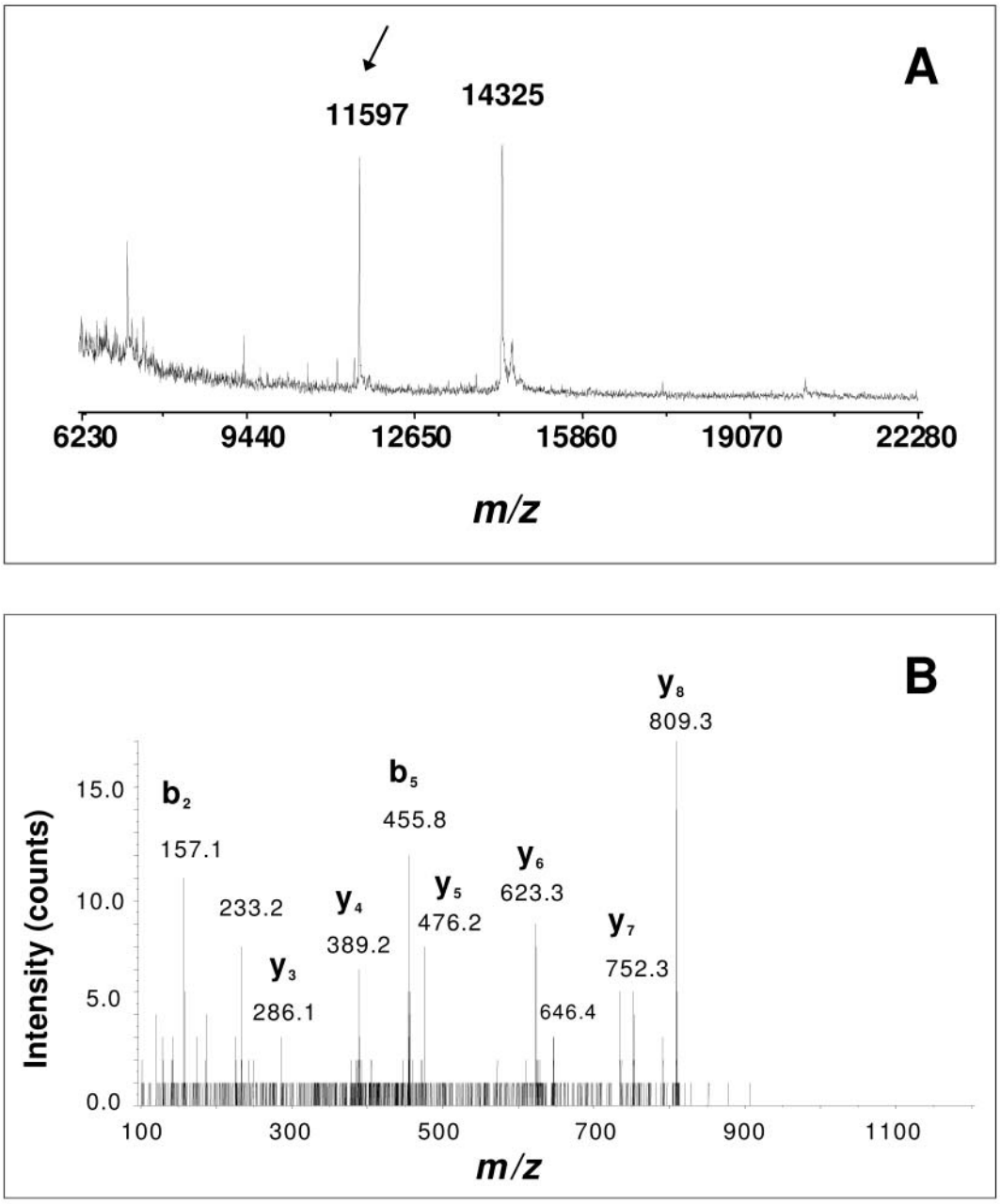

Figure 4. *Identification of a candidate oral fluid protein biomarker for oral cancer. Oral fluid sample was fractionated by C4-HPLC. A) The MALDI-MS spectrum of a LC fraction containing the protein of 11597 Da. This fraction was digested for LC-MS/MS analysis. B) The MS/MS spectrum of a tryptic peptide, VGEFSGANK, originated from thioredoxin.*

fragmented to yield sufficient structural information for identification of the peptide sequence, VGEFSGANK, originated from thioredoxin. Mascot database searching indicated that 4 peptides were matched to this protein, resulting in a sequence-coverage of 31%.

The feasibility for identification of candidate peptide biomarkers by nanospray MS was also demonstrated. Without *in vitro* enzymatic digestion, intact peptide ions were selected for nanospray QqTOF MS/MS analysis and subsequent database searching. Figure 5 shows the ESI mass spectrum (Figure 5A) of a 5^{+}-charged precursor ion (*m*/*z*, 874.6, monoisotopic) and its collision-induced fragments (Figure 5B). The peptide was determined to be a fragment of salivary proline-rich phosphoprotein precursor PRH1/PRH2.

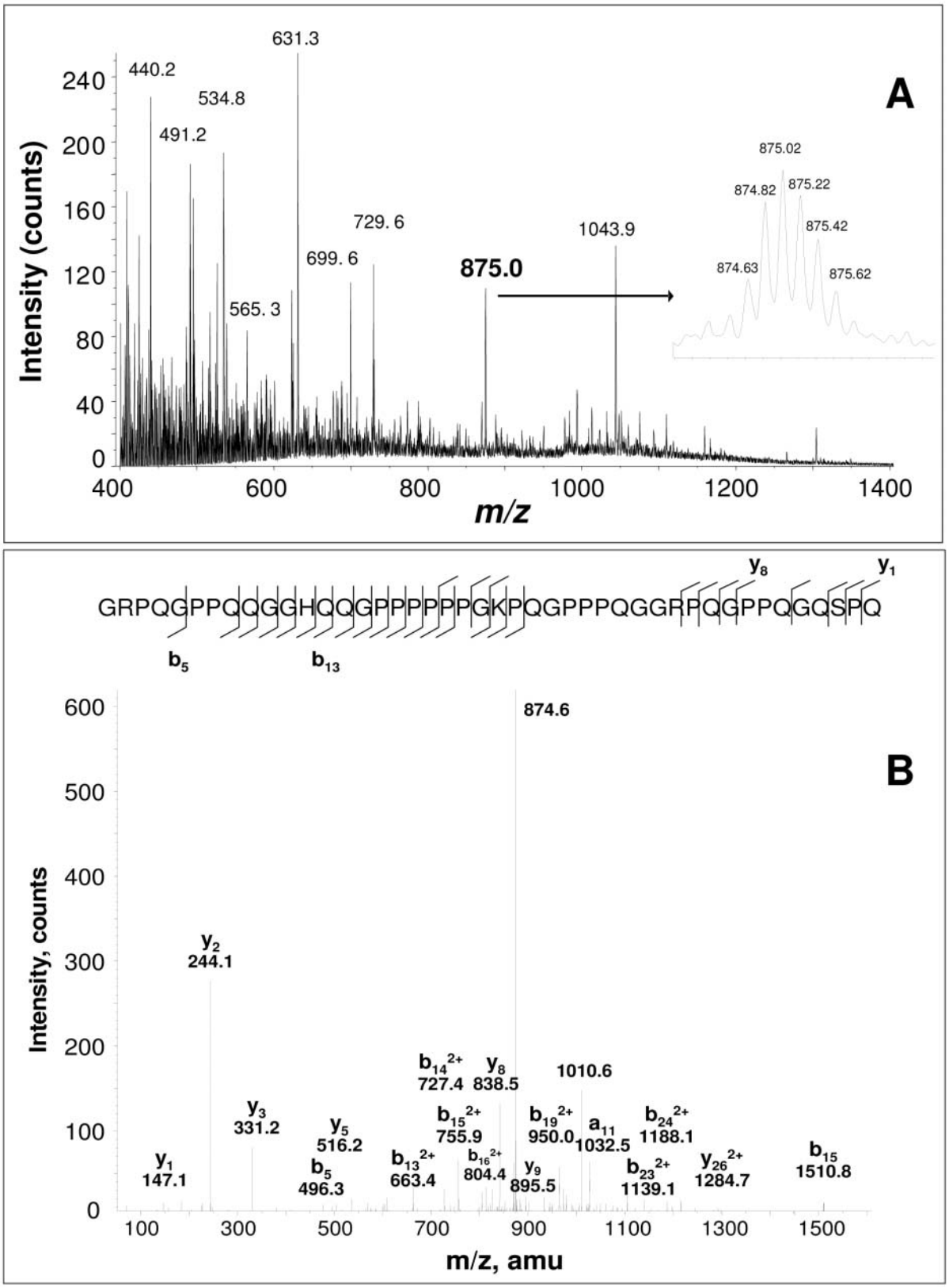

Figure 5. *Nanospray MS/MS identification of a candidate peptide biomarker for oral cancer. A 5^+-charged precursor ion at m/z 875 yielded sufficient product ions to identify a proline-rich peptide. The insert shows the origins of the 1^+- and 2^+-charged product ions.*

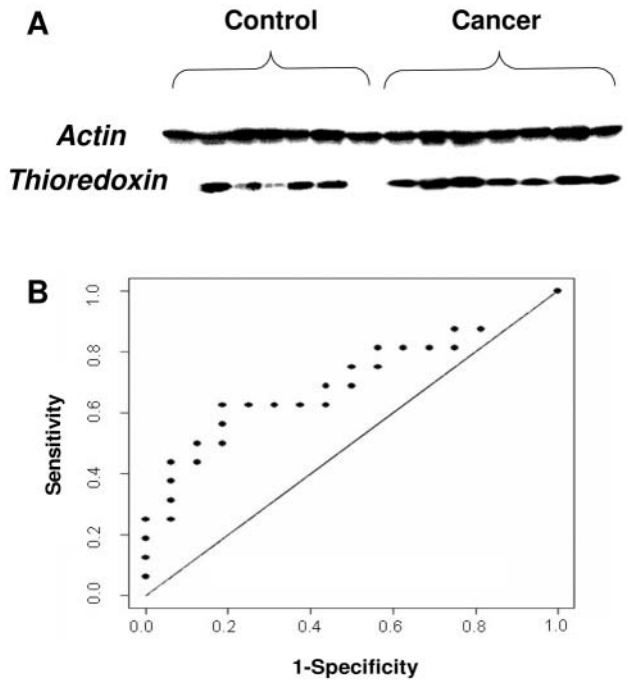

Figure 6. *Western blotting of thioredoxin in oral fluid samples from oral cancer and control subjects (A). The chart (B) indicates the ROC curve based on the validation data of thioredoxin between oral cancer and control saliva samples (n=24 for each group).*

The measured molecular mass (4367.96 Da) showed good agreement to the theoretical mass (4368.18 Da). This peptide was found to be significantly lower in oral cancer patients' oral fluid. The C4-LC fraction mainly containing this peptide was also analyzed using LC-MS/MS. A 4^+-charged precursor ion (*m*/*z*, 1093.10) produced well-defined fragment ions and database searching resulted in the positive identification of the same peptide (data not shown).

Western blotting validation and receiver operating characteristic (ROC) analysis. In order to validate the MS identification of thioredoxin, western blotting was used to compare the levels of thioredoxin in a new independent set of cancer and control samples, Figure 6A indicates western blot analysis of thioredoxin in 7 oral cancer samples versus 7 control samples. By using densitometric analysis and actin for normalization, the average level of thioredoxin in cancer samples was found to be about three times higher than that in normal samples ($p<0.01$), which confirms the result obtained by MALDI-MS.

A conventional method for evaluating the predictive power of biomarkers for disease discrimination uses receiver operating characteristic (ROC) curve (10), which is a plot of the true positive rate versus false positive rate at different classification threshold settings. The area under the ROC curve (ROC value) is used as a summary measure of diagnostic accuracy. By performing ROC analysis of the validation data on 24 OSCC and 24 control subjects, the ROC value for thioredoxin was determined to be 0.71 (Figure 6B), and the specificity and sensitivity were determined to be 70.8% and 70.8% respectively using a cut-off probability of 50.1%.

It is interesting to note that expression microarray analysis of oral fluids from oral cancer and control subjects (n=20, 10 OSCC, 10 control) revealed that saliva thioredoxin mRNA level was concordantly up-regulated in OSCC subjects ($p<0.02$) (27). In addition, thioredoxin was found over-expressed in human cancers such as non-small cell lung, gastric, cervical and hepatocellular carcinomas (28-31). Serum levels of thioredoxin were also reported to be elevated ~2-fold in patients with hepatocellular carcinoma (32). These published results support our finding and suggest an interesting role of thioredoxin in human cancers.

Conclusion

We have proven that human oral fluid from OSCC patients contains signature peptides and proteins potentially useful for OSCC detection. These discovered candidates, once validated on new and independent patient cohorts, may lead to extensive use of oral fluid for oral cancer diagnosis in the future dental clinics. Similar approach can also be used to monitor oral fluids from patients with oral precancer lesions and discover oral fluid biomarkers for early oral cancer detection. In addition, we have established a straightforward method by combining LC fractionation with MALDI-MS and tandem MS for biomarker identification. This approach is practical towards identifying oral fluid biomarkers considering oral fluid is simpler than other body fluids such as serum. Although not explored, the candidate peptide biomarkers can be separated and fractionated using C18 RP-HPLC and subsequently identified by using tandem MS and database searching. The described approach mainly focused on discovery of low-molecular-weight candidates for OSCC. In the future, a 2-DE/MS or quantitative proteomics approach can be used to further discover high-molecular-weight protein signatures for OSCC detection.

References

1 Petricoin EF 3rd, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, Velassco A, Trucco C, Wiegand L, Wood K, Simone CB, Levine PJ, Linehan WM, Emmert-Buck MR, Steinberg SM, Kohn EC and Liotta LA: Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst *94*: 1576-1578, 2002.

2 Jacobs IJ and Menon U: Progress and challenges in screening for early detection of ovarian cancer. Mol Cell Proteomics *3*: 355-366, 2004.

3 Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF, Yasui Y, Feng Z and Wright GL Jr: Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res *62*: 3609-3614, 2002.
4 Li J, Zhang Z, Rosenzweig J, Wang YY and Chan DW: Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem *48*: 1296-1304, 2002.
5 Pan S, Zhang H, Rush J, Eng J, Zhang N, Patterson D, Comb MJ and Aebersold R: High throughput proteome screening for biomarker detection. Mol Cell Proteomics *4*: 182-190, 2005.
6 Tao WA and Aebersold R: Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. Curr Opin Biotechnol *14*: 110-118, 2003.
7 Moseley MA: Current trends in differential expression proteomics: isotopically coded tags. Trends Biotechnol *19*: S10-S16, 2001.
8 Martin DB, Gifford DR, Wright ME, Keller A, Yi E, Goodlett DR, Aebersold R and Nelson PS: Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. Cancer Res *64*: 347-355, 2004.
9 Terrin N, Schmid CH, Griffith JL, D'Agostino RB and Selker HP: External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. J Clin Epidemiol *56*: 721-729, 2003.
10 Grunkemeier GL and Jin R: Receiver operating characteristic curve analysis of clinical risk models. Ann Thorac Surg *72*: 323-326, 2001.
11 Hu S, Xie Y, Ramachandran P, Loo RR, Li Y, Loo JA and Wong D: 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, 2005.
12 Hu S, Denny P, Denny P, Xie Y, Loo JA, Wolinsky LE, Li Y, McBride J, Loo RR, Navazesh M and Wong DT: Differentially expressed protein markers in human submandibular and sublingual secretions. Int J Oncol *25*: 1423-1430, 2004.
13 Hu S, Li Y, Xie Y, Tjon K, Loo RR, Loo JA and Wong DT: Human salivary proteome and transcriptome. J Dent Res *85*: 1129-1133, 2006.
14 Malamud D: Oral diagnostic testing for detecting human immunodeficiency virus-1 antibodies: a technology whose time has come. Am J Med *102(4A)*: 9-14, 1997.
15 Beeley JA and Khoo KS: Salivary proteins in rheumatoid arthritis and Sjogren's syndrome: one-dimensional and two-dimensional electrophoretic studies. Electrophoresis *20*: 1652-1660, 1999.
16 Kaufman E and Lamster IB: Analysis of saliva for periodontal diagnosis–a review. J Clin Periodontol *27*: 453-465, 2000.
17 Streckfus C, Bigler L, Dellinger T, Dai X, Kingman A and Thigpen JT: The presence of soluble c-erbB-2 in saliva and serum among women with breast carcinoma: a preliminary study. Clin Cancer Res *6*: 2363-2370, 2000.
18 Streckfus C, Bigler L, Tucci M and Thigpen JT: A preliminary study of CA15-3, c-erbB-2, epidermal growth factor receptor, cathepsin-D, and p53 in saliva among women with breast carcinoma. Cancer Invest *18*: 101-109, 2000.
19 Vokes EE, Weichselbaum RR, Lippman SM and Hong WK: Head and neck cancer. New Engl J Med *328*: 184-194, 1993.
20 Sidransky D: Molecular genetics of head and neck cancer. Curr Opin Oncol *7*: 229-233, 1995.
21 Schantz SP: Biologic markers, cellular differentiation, and metastatic head and neck cancer. Eur Arch Otorhinolaryngol *250*: 424-428, 1993.
22 Navazesh M: Methods for collecting saliva. Ann NY Acad Sci *694*: 72-77, 1993.
23 Dudoit S, Fridlyand J and Speed T: Comparison of discrimination methods for the classification of tumors using gene expression data. JASA *97(457)*: 77-87, 2002.
24 Duda R, Hart P and Stork D: Pattern Classification Wiley-Interscience, 2000.
25 Vlahou A, Giannopoulos A, Gregory BW, Manousakas T, Kondylis FI, Wilson LL, Schellhammer PF, Wright GL Jr and Semmes OJ: Protein profiling in urine for the diagnosis of bladder cancer. Clin Chem *50*: 1438-1441, 2004.
26 Wang Z, Yip C, Ying Y, Wang J, Meng XY, Lomas L, Yip TT and Fung ET: Mass spectrometric analysis of protein markers for ovarian cancer. Clin Chem *50*: 1939-1942, 2004.
27 Li Y, St John MAR, Zhou X, Kim Y, Sinha U, Jordan RCK, Eisele D, Abemayor E, Elashoff D, Park NH and Wong DT: Salivary transcriptome diagnostics for oral cancer detection. Clin Cancer Res *10*: 8442-8450, 2004.
28 Soini Y, Kahlos K, Napankangas U, Kaarteenaho-Wiik R, Saily M, Koistinen P, Paaakko P, Holmgren A and Kinnula VL: Widespread expression of thioredoxin and thioredoxin reductase in non-small cell lung carcinoma. Clin Cancer Res *7*: 1750-1757, 2001.
29 Hedley D, Pintilie M, Woo J, Nicklee T, Morrison A, Birle D, Fyles A, Milosevic M and Hill R: p-regulation of the redox mediators thioredoxin and apurinic/apyrimidinic excision (APE)/Ref-1 in hypoxic microregions of invasive cervical carcinomas, mapped using multispectral, wide-field fluorescence image analysis. Am J Pathol *164*: 557-565, 2004.
30 Ding SJ, Li Y, Shao XX, Zhou H, Zeng R, Tang ZY and Xia QC: Proteome analysis of hepatocellular carcinoma cell strains, MHCC97-H and MHCC97-L, with different metastasis potentials. Proteomics *4*: 982-994, 2004.
31 Powis G, Mustacich D and Coon A: The role of the redox protein thioredoxin in cell growth and cancer. Free Radic Biol Med *29*: 312-322, 2000.
32 Miyazaki K, Noda N, Okada S, Hagiwara Y, Miyata M, Sakurabayashi I, Yamaguchi N, Sugimura T, Terada M and Wakasugi H: Elevated serum level of thioredoxin in patients with hepatocellular carcinoma. Biotherapy *11*: 277-288, 1998.

Received June 7, 2006
Accepted June 20, 2006