Published in final edited form as:

Proteomics. 2016 January ; 16(1): 159–173. doi:10.1002/pmic.201500224.

A targeted proteomic strategy for the measurement of oral cancer candidate biomarkers in human saliva

Rebeca Kawahara¹, James G. Bollinger², César Rivera¹, Ana Carolina P. Ribeiro³, Thaís Bianca Brandão³, Adriana F. Paes Leme^{1,*}, and Michael J. MacCoss^{2,*}

¹Laboratório de Espectrometria de Massas, Laboratório Nacional de Biociências, LNBio, CNPEM, Campinas, Brazil

²Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA

³Instituto do Câncer do Estado de São Paulo, Octavio Frias de Oliveira, São Paulo, Brazil

Abstract

Head and neck cancers, including oral squamous cell carcinoma (OSCC), are the sixth most common malignancy in the world and are characterized by poor prognosis and a low survival rate. Saliva is oral fluid with intimate contact with OSCC. Besides non-invasive, simple, and rapid to collect, saliva is a potential source of biomarkers. In this study, we build an SRM assay that targets fourteen OSCC candidate biomarker proteins, which were evaluated in a set of clinically-derived saliva samples. Using Skyline software package, we demonstrated a statistically significant higher abundance of the C1R, LCN2, SLPI, FAM49B, TAGLN2, CFB, C3, C4B, LRG1, SERPINA1 candidate biomarkers in the saliva of OSCC patients. Furthermore, our study also demonstrated that CFB, C3, C4B, SERPINA1 and LRG1 are associated with the risk of developing OSCC. Overall, this study successfully used targeted proteomics to measure in saliva a panel of biomarker candidates for OSCC.

Keywords

Selected reaction monitoring; saliva; oral cancer; skyline

1. Introduction

Head and neck cancers are the sixth most common malignant tumors worldwide [1]. Oral cancer is the most frequent subtype among these with oral squamous cell carcinoma (OSCC) annually affecting over 300,000 people worldwide [2]. Despite advancements in oral cancer prevention and multimodality treatments, the 5-year survival rate of OSCC patients is less

*To whom correspondence should be addressed: Adriana F. Paes Leme, Laboratório de Espectrometria de Massas, Laboratório Nacional de Biociências, LNBio, CNPEM, 13083-970, Campinas, Brazil. Phone: +55 19 3512-1118; Fax: +55 19 3512-1006, adriana.paesleme@lnbio.cnpem.br. *To whom correspondence should be addressed: Michael J. MacCoss, Ph.D., Department of Genome Sciences, University of Washington, Seattle, WA, Phone: 206.616.7451, maccoss@uw.edu.

Conflict of interest statement

The authors declare no competing financial interest.

than 50% and the prognosis of advanced cases has not changed considerably over the past 20 years [3, 4]. Currently, oral cancer diagnosis depends on a thorough oral examination and the most important prognostic factor for OSCC patients is still the clinical stage of disease (TNM stage) at the time of intervention [2, 5].

Scientists are continually searching for novel OSCC biomarker panels. Saliva is a great choice matrix for this task as it is an easy-to-obtain body fluid that can be collected noninvasively [6]. Previous studies have identified a large number of putative biomarker candidates in human saliva [7–11]. Unfortunately, none of these candidates have passed through the development and validation stage to gain utility in clinical practice [12]. The bottleneck in this pipeline is attributed, in part to a lack of robust methods to validate a majority of the candidates generated by discovery-phase studies [13, 14]. ELISA is currently the most commonly used analytical platform for the validation of biomarker candidates derived from bodily fluids. However, antibody availability restricts analysis to a small subset of potential candidates and the cost and low throughput of these methods ultimately renders a large majority of candidates untestable [15].

Selected reaction monitoring (SRM) tandem mass spectrometry has emerged as a technique capable of precisely and quantitatively measuring peptides in complex biological samples [16]. Due to its high specificity, sensitivity, precision, and multiplexing capabilities, SRM provides an attractive alternative for the validation of candidate biomarkers [17]. However, there are many challenges still associated with the development of a quantitative SRM assay. Among these is the challenge of identifying the optimal set of proteotypic peptides for each protein target. Proteotypic is defined here as a peptide that is unique to a particular gene product, devoid of single nucleotide polymorphisms, post-translational modifications and missed cleavages, besides readily detectable by the mass spectrometers, and has salient structural features with characteristic MS/MS fragmentation patterns. Furthermore, there are computational challenges related to assay development, analytical validation, data management, peak integration, data quality assessment and biostatistical analysis. Collectively, these challenges have hindered widespread use of the SRM methodology.

The current work presents a successful pipeline for the measurement of a panel of candidate biomarkers for oral cancer in human saliva via an SRM assay. The candidate biomarkers were selected from a discovery-based proteomics study, which was performed using the secretome from three classes of carcinoma (including OSCC originated from a human tongue), melanoma and non-cancerous cell lines (Kawahara et al., 2015, under review). For the selection of proteotypic peptides, we used an empirical approach that is based on the use of analytical standards (*in vitro*-synthesized fusion proteins, recombinant proteins or native proteins isolated from plasma). For each protein standard, we considered every doubly-charged monoisotopic tryptic peptide from 7 to 23 amino acids in length. For each peptide, we considered every singly-charged monoisotopic y-ion from y_n to y_{n-3} [18]. Skyline, a vendor-neutral tool [19], was used extensively for multiple reaction monitoring (MRM) assay development, method export, peak detection and peak integration as well as all subsequent statistical analysis. Overall, this study used targeted proteomics to study a panel of candidate biomarkers in human saliva collected from healthy patients and patients with OSCC.

2. Materials and Methods

2.1 Proteins and peptides

Human complement C4 (Cat# P151-0) and C3c (Cat# pro-555) derived from human plasma were purchased from BBI Solutions and Prospec, respectively. Recombinant SAA1 (Cat# cyt-657) and SERPINA1 (Cat# pro-529) were purchased from Prospec and recombinant CFB (Cat# C329) was purchased from Novo Protein.

Heavy [$^{13}\text{C}_6$ $^{15}\text{N}_2$]-lysine-labeled LLLEYLEEK and IEAIPQIDK peptides from *Schistosoma japonicum* (SJ) GST (Thermo Scientific) and VSEADSSNADWVTK and YGLVTYATYPK peptides from CFB (SynPeptide) were obtained to use as internal standards. Similarly, unlabeled LLLEYLEEK and IEAIPQIDK peptides were obtained to use as calibration standards. A mixture containing 15 heavy isotope-labeled peptides was used as retention time standards (Pierce[®] Retention Time Calibration Mixture, Thermo), quality control and normalization.

2.2 Recombinant proteins production using *in vitro* expression system

The recombinant proteins, native proteins isolated from human plasma and *in vitro*-synthesized proteins were used to empirically derive optimal proteotypic peptides and fragment ions as previously described [18] with some modifications described below.

For the nine *in-vitro* synthesized proteins (C1R, TINAG, SLPI, SERPINE1, LRG1, LCN2, TAGLN2, ANXA1 and FAM49B), clones containing an in-frame fused C-terminal SJ GST tag were obtained from the Arizona State University Biodesign Institute plasmid repository. Each bacterial plasmid was grown overnight in 5 ml of Luria broth supplemented with 100 µg/ml-1 ampicillin (LB-amp). Plasmid DNA was extracted using the manufacture's mini-prep protocol with the exception of an additional wash with PE buffer (Qiagen). All plasmid stocks were Sanger sequenced (Genewiz) using an M13 priming site upstream of the T7 promoter to confirm the identity of the insert and to check for contamination of the plasmid stocks. For *in vitro* protein synthesis, 1 µg of mini-prepped plasmid DNA was used directly in the 1-Step Human Coupled IVT Kit (Thermo Scientific) according to manufacturer's protocols with a few minor modifications. Briefly, a mix containing HeLa lysate, Accessory Proteins, and a proprietary Reaction Mix was added to 1 µg of plasmid DNA in a 25-µl transcription reaction supplemented with 0.3 µl RNase inhibitors (Thermo Scientific) and the reaction was incubated at 30 °C for 3 h/200 rpm.

2.3 Recombinant Protein Purification and digestion

To enrich GST-fusion protein, we used 200 µl of glutathione sepharose 4B beads (GE), washed 3 times with 1 ml 1× Dulbecco's phosphate-buffered saline (DPBS; Gibco) and resuspended in 1.25 ml of 1× DPBS. A 125-µl aliquot of the washed bead slurry was added to each 1.5 ml-tubes containing the HeLa lysate reaction such that each tube received 20 µl of packed beads. Completed translation reactions were added to the beads and the bead-protein mixture was rocked end-over-end for 16 h at 4 °C.

The bead-protein mixture was washed twice with 150 μ l of wash buffer (1 $\times$ DPBS supplemented with 863 mM NaCl) and twice with 150 μ l of 50 mM ammonium bicarbonate (pH 7.8) each. After the last wash, the beads were resuspended in 50 μ l Elution Buffer (0.05% PPS silent surfactant (Protein Discovery), 5 fmol/ μ l heavy isotope-labeled SJ GST peptides LLLEYLEEK and IEAIPQIDK (Thermo Scientific), and 50 mM ammonium bicarbonate (pH 7.8)). Ten microliters of each enriched protein sample was added to 4 μ l 4 $\times$ LDS buffer (Invitrogen) and saved for immunoblotting.

Bead bound protein samples (25 μ l) were diluted 1:1 in 0.05% PPS silent surfactant/50 mM ammonium bicarbonate (pH 7.8), boiled for 5 min at 95 $^{\circ}$ C, reduced with 5 mM dithiothreitol for 30 min at 60 $^{\circ}$ C, and alkylated with 15 mM iodoacetamide for 30 min at 25 $^{\circ}$ C in the dark. Alkylation reactions were quenched by bringing the DTT concentration up to 15 mM and proteins were subsequently digested with 1 μ g trypsin (Promega) at 37 $^{\circ}$ C for 2 h while shaking at 1200 rpm. Digestion progress was terminated via the addition of formic acid to 1.6% and samples were incubated for 1 h at room temperature to facilitate hydrolysis of the PPS silent surfactant.

2.4 Immunoblotting of *in-vitro* synthesized proteins

Undigested protein extracts from each of the *in vitro* translation reactions were boiled in 1 $\times$ LDS buffer (Invitrogen), resolved on a 4–12% Bis-Tris denaturing/reducing SDS-PAGE (Invitrogen) and transferred onto a nitrocellulose membrane (Bio-Rad) for immunoblotting. Membranes were then blocked with 5% non-fat dry milk (Safeway) in PBS-tween buffer and probed with a primary antibody targeting SJ GST (GE 27–4577–01) as an epitope. Incubations with the primary antibody were done at 4 $^{\circ}$ C overnight using a 1:1000 dilution. Secondary incubations were performed in 5% non-fat dry milk in PBS-tween using 1:10000 diluted peroxidase-conjugated rabbit anti-goat IgG (H+L) (Pierce). Membranes were visualized using an ECL plus western blotting kit (Amersham) and detected with radiographic film (Thermo).

2.5 SRM assay development and refinement

All samples were analyzed with a TSQ-Vantage triple-quadrupole instrument (Thermo Scientific) using either a nanoLC separation system (Eksigent) or a nanoACQUITY UPLC (Waters). A 3- μ l aliquot of each sample was separated on a self-packed 15-cm $\times$ 75 μ m inner diameter column (Polymicro Technologies) with ReproSil-Pur C18-AQ (3 μ m, particle size and 120 $\text{\AA}$ pore size). Peptides were separated using a 7-min gradient from 5% acetonitrile in 0.1% formic acid to 40% acetonitrile in 0.1% formic acid at a flow of 750 nl/min. The gradient was followed by 2 min linear gradient from 40% acetonitrile in 0.1% formic acid to 68% acetonitrile in 0.1 % formic acid also at 750 nl/min and washed for 3 min at 95% acetonitrile in 0.1% formic acid and finally re-equilibrated to initial conditions at 750 nl/min in 5% acetonitrile in 0.1% formic acid for 6 min. Ions were isolated in both Q1 and Q3 using 0.7 FWHM resolution. Peptide fragmentation was performed using Argon at 1.5 mTorr in Q2 using calculated peptide specific collision energies as described previously [20]. Data were acquired using a dwell time of 10 ms.

For each analytical standard, we initially monitored all monoisotopic, +2 charge states for every fully tryptic peptide ranging from 7 to 23 amino acids in length. In addition, we monitored the heavy and light forms of the SJ GST peptides LLEYLEEK and IEAIPQIDK. For each precursor, the y_3 to y_{n-1} fragment ions were considered in their monoisotopic, +1 charge state. All cysteines were monitored as carbamidomethyl cysteines.

After mass spectrometry analysis, the resulted raw files were imported into Skyline (<http://skyline.maccosslab.org>) [19]. Chromatographic data from each peptide were manually analyzed to determine the quality of the peptide signal and peak shape and the four best transitions for each peptide were selected based on the intensity and fragment m/z referentially higher than precursor m/z .

The samples were incubated in the 4 oC auto-sampler for 48 h to evaluate the peptide stability. The selected transitions refined in the first step were then submitted to another mass spectrometry run and the data were similarly analyzed in Skyline. The peptides that demonstrated more than 50% decrease in the peak area in integrated signal intensity after 48 h were deemed insufficiently unstable and excluded from further consideration.

2.6 Creating a chromatogram Library using PANORAMA

All the results from the method development using the recombinant, native or *in vitro*-synthesized proteins were uploaded to PANORAMA daily (<https://daily.panoramaweb.org>) and re-imported back into Skyline as a chromatogram library. Chromatogram libraries provide a way to store targeted assays that have been curated in Skyline and reuse them in the future for measuring proteins and peptides in other samples. Chromatogram libraries capture physiochemical and experimental properties such as the relative ion abundances, fragmentation patterns, measured chromatograms, and retention time information (iRT), and optimal instrument parameters such as collision energy.

2.7 Clinical Samples

The saliva was collected from healthy individuals (n=8), OSCC patients who underwent surgical resection (named as no lesion, n=9), and OSCC patients with active oral malignant lesions (named as lesion, n=13). All enrolled patients and volunteers signed a formulary stating their awareness and consent for the study as it was approved by the ethics review board of Instituto do Câncer do Estado de São Paulo, Octavio Frias de Oliveira, ICESP, São Paulo, SP, Brazil and Plataforma Brasil. Individuals were initially asked to rinse their mouth with 5 ml of drinking water. Saliva was then harvested into a clean glass receptacle, aliquoted in 2 ml eppendorf tubes and immediately stored at -80°C . Clinic pathological data, such as sex, anatomical site of the primary tumor, TNM classification of malignant tumors (TNM), WHO differentiation degree, and treatment, were collected from patient's charts and listed in Table 2.

2.8 Human saliva sample preparation and digestion

Protein extraction was performed by homogenizing the 100 μl of whole saliva with 100 μl of a solution containing 100 mM Tris-HCl, pH 7.5, 8 M urea, 2 M thiourea containing Protease Inhibitor Cocktail Complete Mini Tablets (Roche, Auckland New Zealand), 5 mM EDTA, 1

mM PMSF and 1 mM DTT. Samples were sonicated for 10 min and then centrifuged at 10,000 x g for 5 min. Protein 1 concentration was determined for each sample with a Bradford assay (Bio-Rad, Hercules, CA, USA) and a 10 µg aliquot of each was submitted to reduction (5 mM dithiothreitol, 25 min at 56°C), followed by alkylation (14 mM iodoacetamide, 30 min at room temperature in the dark) and digestion with trypsin (1:50, w/w). The reaction was stopped with 1% TFA and desalted with stage-tips C18 resin. The samples were dried in a vacuum concentrator and reconstituted in 0.1% formic acid.

2.9 Method refinement in the human saliva matrix

The saliva samples were resuspended to a final concentration of 0.33 µg/µl and three different samples from OSCC patients were pooled. The refined transitions for the 14 proteins (C1R, TINAG, SLPI, SERPINE1, LRG1, LCN2, TAGLN2, ANXA1, FAM49B, CFB, C4B, C3, SERPINA1 and SAA1) were exported from Skyline as transition list and imported into Xcalibur as multiple methods. Multiple runs were performed as described before with the pool of human saliva, using a gradient of 30 min from 5% to 40% acetonitrile in 0.1% formic acid at a flow of 250 nl/min. The gradient was followed by steeper 5-min linear gradient from 40% to 60% acetonitrile in 0.1% formic acid also at 250 nl/min. The column was then washed for 5 min at 95% acetonitrile in 0.1% formic acid and finally re-equilibrated to initial conditions at 500 nl/min. Peptide fragmentation was performed at 1.5 mTorr in Q2. Data were acquired using a scan width of 0.002 m/z and a dwell time of 10 ms.

2.10 Standard curve of synthetic peptides

A dilution curve of the unlabeled SJ GST peptides, LLEYLEEK and IEAIPQIDK, was spiked with 5 fmol/µl of each heavy isotope-labeled LLEYLEEK and IEAIPQIDK peptides. The calibration points used were 0, 0.1, 0.25, 5, 12.5 and 40 fmol/µl each of the light LLEYLEEK and IEAIPQIDK peptides (**the amount of injection volume was 1µl**). Each calibration point was diluted into 25 fmol/µl Bovine Quality Control (QC) standard mix (Michrome) in 2% acetonitrile and 0.1% formic acid in water. Peptide standards were measured in a single replicate and a linear regression of the data points was used to calibrate the SJ GST peptide light-to-heavy ratio of all other samples.

Similarly, an eight-point calibration curve was prepared using the aforementioned CFB recombinant protein digest in a theoretical concentration range of 0, 0.1, 1, 6.25, 12.5, 25, 50 and 100 fmol/µl with heavy isotope-labeled YGLVTYATYPK (4 fmol/µl) and VSEADSSNADWVTK (18 fmol/µl) peptides from CFB, heavy isotope-labeled LLEYLEEK and IEAIPQIDK (5 fmol/µl each) peptides from SJ GST as well as the Thermo Peptide Retention Time standards (6.7 fmol/µl). The curve was measured in triplicate and the results were analyzed using Skyline External Tool Quasar Quantitative Statistics to determine limits of detection and quantification, calculate mean and coefficient of variation for all transitions of each CFB heavy peptides.

2.11 SRM in the human saliva

An aliquot containing 1.2 µg of digested proteins from saliva spiked, at the same time, with 18 fmol/µl of VSEADSSNADWVT[¹³C₆]K from CFB, 4 fmol/µl of

YGLVTYATYP[¹³C₆]K, 5 fmol/μl of SJ GST heavy IEAIPQIDK peptide and 6.7 fmol/μl of Pierce Retention Time Calibration Mixture. Samples were analyzed on a TSQ-Vantage triple-quadrupole instrument (Thermo Scientific) as previously described. Briefly, the peptides were separated on a 20-cm column in a gradient of 30 min from 5 to 40% acetonitrile in 0.1% formic acid at a flow of 250 nl/min. The gradient was followed by steeper 5-min linear gradient from 40% to 60 % acetonitrile in 0.1 % formic acid also at 250 nl/min. The column was then washed for 5 min at 95 % acetonitrile in 0.1 % formic acid and finally re-equilibrated to initial conditions at 500 nl/min. Two technical replicates from each sample were performed. Peptide fragmentation was performed at 1.5 mTorr in Q2. Data was acquired using a scan width of 0.002 m/z and a dwell time of 10 ms.

2.12 SRM data analysis

Data sets were imported into Skyline, and peaks were automatically integrated and manually inspected.

The confidence of the peak identity was assessed by 1) calculating a dot-product correlation between the peak intensities of the transitions for the peptide of interest and the library spectrum for the same peptide and 2) calculating the retention time prediction obtained from the regression of RT by iRT values of five standards peptides (SSAAPPPPPR, IGDYAGIK, SAAGAFGPESLR, SFANQPLEVVYSK, LTILEELR) out of the 15 **Pierce** Retention Time Calibration Mixture peptides spiked and monitored in all samples. The following equation was used to predict RTs for the targeted peptides: $RT = 0.09 \times iRT + 18.51$ ($R = 0.9998$).

Statistical analysis was performed using MSstats [21–23], an external tool implemented directly in Skyline [24]. Samples were annotated into three respective conditions (Control, No lesion and Lesion) and individual biological replicates. The number of samples of each condition was: 8 for control, 9 for no lesion and 13 for lesion. Each sample was analyzed in two technical replicates. For more details see Supplemental Material and Methods.

3. Results and Discussion

In this work we demonstrated the use of human saliva collected from healthy patients or patients with oral squamous cell carcinoma to measure candidate biomarkers using selected reaction monitoring (SRM) assay. The candidate biomarkers were selected from a previous discovery-based proteomic study that used support vector machine-recursive features elimination (SVM-RFE), nearest shrunken centroids (NSC) and beta-binomial model tested in a well-controlled data obtained from the secretome of three classes of carcinoma (including OSCC originated from a human tongue) melanoma and non-cancerous cell lines. Among 2,574 proteins identified in the secretome, the statistical models indicated a panel of 137 candidate biomarkers for carcinoma and 271 for melanoma, which were differentially abundant between the tumor classes. Among the candidate biomarkers for carcinoma, fourteen proteins (C1R, TINAG, SLPI, SERPINE1, LRG1, LCN2, TAGLN2, ANXA1, FAM49B, CFB, C4B, C3, SERPINA1 and SAA1) were selected based on the availability of protein analytical standards for method development as well as cellular localization in the extracellular milieu (Kawahara et al., 2015, under review) and used for verification in human saliva using targeted proteomics.

3.1 Selection of proteotypic peptides

Our first task was to select an optimal set of proteotypic peptides of each protein to develop an SRM assays for human saliva. We utilized a previously described empirical approach for this task that is based on the use of analytical standards [18]. In the current report, we utilized *in-vitro* synthesized proteins, commercial recombinant proteins and commercial native proteins that were isolated from plasma.

Figure 1 illustrates our workflow for SRM method development. Using a fast gradient (7 minutes) at a relative high flow rate for nano-flow UPLC (750 nl/min), we were able to expedite method development during the arduous step of monitoring every transition for every tryptic peptide. In this step, 2,908 transitions, corresponding to 391 peptides were monitored using multiples methods, with 50 maximum transitions per sample injection, resulting in approximately 58 runs.

For quantification of the *in-vitro* synthesized proteins, we used the ratio of the SJ GST peptides from the C-terminal fusion protein to the spiked-in heavy isotope-labeled peptides. Observed ratios were fit into a calibration curve of the light and heavy isotope-labeled peptides (Supplemental Figure 1A). Each of the tested proteins produced at least 500 fmoles per 25 μ l of the *in vitro* reaction (Supplemental Figure 1A and B).

For each protein, we selected peptides and their respective fragment ions using Skyline. Peptides with non-existent or ambiguous chromatograms were excluded. Three to four transitions were chosen from the remaining peptides based on SRM intensity and fragment m/z higher than precursor m/z , preferentially. Stability of the peptides was evaluated after 48 h in the auto-sampler and the peptides that showed 50% decrease in the peak area were excluded. The final transitions obtaining after monitoring the standard proteins in these two steps were 650 that correspond to 169 peptides.

Obtained SRM results were exported to Panorama for the purpose of building a chromatogram library. This library was used as a reference to assign confidence in the identification of targeted peptides in the human saliva matrix. Skyline utilizes said library to calculate the dot-product (dotp) correlation for the ratio of the observed SRM peak intensities of a peptide in a specific biological matrix versus the library spectrum.

The filtered set of peptides and transitions were then monitored in a pool of digested saliva matrix. Assay refinement at this stage was based on those peptides and transitions that demonstrated the optimal signal-to-noise ratios. Ultimately, we selected three peptides per protein which showed good signal and good dotp (>0.8).

3.2 Analytical assay performance using CFB heavy peptides and global standards

The refinement process for optimal SRM peptides also guided our selection of peptides to be synthesized in a heavy isotope labeled analog. Figure 2A shows the peak area comparison of all tryptic peptides from 7 to 23 amino acids in length measured using the CFB recombinant protein. The first, fourth and fifth peptides, indicated by the colored arrows, were the peptides observed in highest intensity within the saliva matrix. Heavy isotope analogs for the peptides VSEADSSNADWVTCK and YGLVITYATYPK (red arrows) were purchased

from a commercial source. The relative signal intensities of the fragment ions generated from endogenous peptides in saliva were compared to those from the heavy labeled peptides in Figure 2B and C.

Calibration curves were constructed by dilution of digested recombinant CFB in known concentration range from 0.1 to 100 fmol/μl into constant amount of CFB heavy peptides VSEADSSNADWVTCK and YGLVTYATYPK were added at a final concentration of 18 and 4 fmol/μl, respectively, together with 6.7 fmol/μl of Pearce Retention Time Calibration Mixture and 5 fmol/μl of SJ GST heavy peptides.

The limit of detection (LOD) and the limit of quantification (LOQ) for both CFB peptides normalized by the respective labeled peptide were calculated using Quasar Quantitative Statistics [24], an external tool implementable in Skyline. The complete data used for calculating the LOD, LOQ and CVs are presented in Supplemental Tables 1 and 2. The calculated LOD and LOQ for VSEADSSNADWVTCK peptide were 0.90 fmol/μl and 2.7 fmol/μl, respectively. The calculated LOD and LOQ for YGLVTYATYPK were 1.8 fmol/μl and 5.5 fmol/μl, respectively.

Using the same calibration curve of recombinant CFB, we compared the performance of our SRM method when the peptide VSEADSSNADWVTCK and YGLVTYATYPK were normalized by the respective heavy peptides or by the global standards. Linear regression for the two peptide sequences was higher than 0.99 (R^2 values) within the investigated concentration range when normalizing by either the heavy peptides or by the global standards (Supplemental Figure 2).

The use of internal reference peptides from a targeted protein of interest has been previously described [25]. Our study demonstrates that external reference peptides can be used as an alternative method to internal reference peptides for the normalization of signal intensities. Besides, assay performance was compared with stable isotope labeling strategies, considering linearity and reproducibility. It is important to highlight that the use of retention time standards, as external reference peptides, is easy to implement, cost effective, and retains the potential to normalize for ion suppression across multiple different points in the chromatogram.

3.3 SRM analysis of OSCC candidate biomarkers in human saliva

For the optimized SRM assay, three peptides were monitored per protein with three to four transitions monitored per peptide. In total, 14 proteins, 48 peptides, 188 transitions (cycle time approximately 2s in a non-scheduled method) were considered and monitored in each 65-min run. Among the 48 peptides, six peptides were used as global standards, including five peptides from the Retention Time Calibration Mixture (SSAAPPPPPR, IGDYAGIK, SAAGAFGPESLR, SFANQPLEVVYSK, LTILEELR) as well as the SJ GST heavy isotope-labeled peptide IEAIPQIDK. The global standards were chosen to represent different points in the chromatography run. They were used to assess the quality of data over multiple replicates and sample injections. We used several features in Skyline to view the SRM data and evaluate the reproducibility and quality of data based on the retention time

and peak areas of the global standards and labeled peptides. Using these peptides as global standards, we observed CVs ranging from 7.5–25% (Table 1, Supplemental Table SM4).

All targeted peptides were manually analyzed to ensure correct peak picking. For that, we used the dot-product (dotp) correlation as well as the retention time prediction based on a linear regression of the retention times from the five peptides from the Retention Time Calibration Mixture.

We also used the advanced refinement tool in Skyline to automatically excluded peptides with dotp scores lower than 0.8. With the exception of SAA1, SLPI and TINAG, at least two peptides with dotp scores > 0.8 were detected for each protein that was within the retention time prediction window (2 min) in all samples and in both technical replicates. For SAA1 and TINAG no peptides were detected above the minimum dotp threshold of 0.8 and were thus excluded from statistical analysis. For the SLPI protein, only the peptide CLDPVDTPNPTR remained after the advanced refinement. A representative result showing one of the three peptides selected for each protein and the relative contribution of each fragment ion both in the library as well in the saliva matrix is displayed in Figure 3. The final complete list of targeted proteins, peptides and transitions used in this study is shown in Supplemental Table 3. Furthermore, a detailed report showing the dotp and retention time results for all peptides is available in Supplemental Table 4.

3.4 Statistical analysis using MSstats

MSstats is tool for quantitative proteomics that is based on a family of linear mixed-effects model. It represents all the quantitative measurements for targeted protein analysis to facilitate detection of systematic changes in protein abundance between conditions with high sensitivity and specificity all at a controlled false discovery rate [23].

After peak identification and quantitation in Skyline, the results were exported as a .csv file containing the following variables: ProteinName, PeptideSequence, PrecursorCharge, FragmentIon, ProductCharge, IsotopeLabelType, Condition, BioReplicate, Run, and Intensity.

Using the annotation feature present in Skyline, the experiment was designed as a “GroupComparison” experiment. We initially configured the conditions present in the study: Control (n=8, healthy patients), No lesion (n=9, patients who undergone surgical resection) and Lesion (n=13, patients who had active oral malignant lesion at the time of saliva collection). A unique identifier for each biological replicate in the experiment was annotated as “Bioreplicate.” The two technical replicates of each biological replicate were assigned with the same unique identifier. The intensity value was considered the transition area without any transformation. The complete MSstats input used in this analysis at a level of both protein and peptide and the results are shown in Supplemental Tables 5 and 6, respectively.

Before importing the file into R studio, rows containing missing values were excluded and all peptides were labeled as light. We performed statistical analysis in two steps. Initially, we summarized all features into protein-level conclusions. Secondly, we considered each

feature separately at the peptide level. For protein-level modeling and analysis, the protein id was assigned in the ProteinName column and was the same for all peptides derived from the same protein. For peptide-level comparisons, we alternatively used a unique peptide id in the Protein Name column.

Data were imported into R studio and pre-processed using Log2 followed by normalization to the global standard peptides. This normalization first equalizes endogenous intensities of global standard proteins across runs and applies the same between-run shifts to the remaining endogenous proteins in the experiment [23].

All data were then visualized for exploratory data analysis using dataProcessplots. It was observed from the quality control (QC) plot that the median intensities of all reference transitions from the global standards across all proteins were equal between runs (Figure 4A and Supplemental Figure 3).

Using the profile plot, we observed the profile of each peptide from each protein across all runs. Profile plot helps to identify potential sources of variation (both variation of interest and nuisance variation) for each protein. Most of the proteins studied demonstrated a very similar profile plots for all peptides considered in the analysis (Figure 4B and Supplemental Figure 4).

Condition plot was used to visualize potential systematic differences in protein intensities between conditions (Figure 4C and Supplemental Figure 5). Clear differences among the conditions are observed for the C3, C4B, CFB, FAM49B, LCN2, LRG1, SERPINA1 and TAGLN2 proteins, although a more refined model-based estimation was further conducted.

The statistical model was used to evaluate each protein for evidence of differential abundance between the conditions. This model takes into account the experimental design and considers the available sources of variation. The output for a differential abundance test is a table with the following columns: Protein, Label (of the comparison), log2 fold change (log2FC), standard error of the log2 fold change (SE), Student's *t*-test (Tvalue), degree of freedom of the Student's *t*-test (DF), raw p-values (p-value), and p-values adjusted for comparisons across multiple proteins using the approach by Benjamini and Hochberg (adjusted p-value). The cutoff of the adjusted p-value corresponds to the cutoff of the False Discovery Rate [26].

We performed three comparisons of our three conditions simultaneously: control versus OSCC lesion, control versus OSCC no lesion, and OSCC no lesion versus OSCC lesion. The complete MSstats statistical results of each comparison at the level of protein and peptide are shown in Supplemental Tables 7 and 8, respectively.

3.5 Volcano plot at a protein and peptide level

Volcano plots visualize the outcome of one comparison between conditions for all the proteins and combine the information on statistical and practical significance.

At a protein-level, the comparison between control vs OSCC lesion showed statistical difference with a FCcutoff>1.5 for all proteins (C1R, LCN2, SLPI, FAM49B, TAGLN2,

CFB, C3, C4B, LRG1, SERPINA1) except for ANXA1 and SERPINE1 (Figure 5A, Supplemental Table SM7). Alternatively, at the peptide level, only one of the peptides from ANXA1 and SERPINE1 failed to demonstrate a statistical difference. One peptide from C1R showed no statistical difference at the peptide level as well (Figure 5B).

Among the proteins that were found up-regulated in the saliva from OSCC patients, Hiromoto et al., 2011 also showed that the LCN2 expression was strongly up-regulated in well-differentiated OSCC tissues [27] and correlated with poor differentiation and adverse prognosis in oral squamous cell carcinoma [28]. Noorlag et al. demonstrated that SLPI expression correlates with lymph node metastases in OSCC [29]. LRG1 was reported to be upregulated in sera and tumors of ovarian cancer patients [30] and with a potential to become a useful biomarker for the diagnosis of early stage Epithelial Ovarian Cancer (EOC) [31] and for prediagnostic in colorectal cancer (CRC) plasmas [32]. Little is known about FAM48B function and expression, but recently this protein was detected in the secretome of five OSCC cell lines (OC3, OEC-M1, SAS, SCC4, and SCC2) [33]. Nohata et al. [34] showed that TAGLN2 is a target of post-transcriptional repression by miR-1 and silencing of TAGLN2 significantly inhibited cell proliferation and invasion in HNSCC cells. Overexpression of TAGLN2 was also observed in hepatocellular carcinoma [35], lung adenocarcinoma [36], and pancreatic cancer [37]. Proteins C1R, C4B, C3, CFB and SERPINA1 are proteins belong to the complement system that will be discussed further.

In the comparison of control versus OSCC no lesion, ANXA1, C3, LCN2 and SERPINE1 did not demonstrate a statistical difference at a protein level (Supplemental Table SM7). At a peptide level, the result was similar for the peptides from C3 (2 peptides), SERPINE1 (2 peptides), LNC2 (2 peptides) and ANXA1 (2 peptides). However, other peptides from TAGLN2 (2), C1R, SERPINA1 (1) CFB (2) also did not demonstrate a difference. One peptide from ANXA1 and one peptide from LCN2 were down-regulated and fourteen additional peptides were up-regulated: LRG1 (3), CFB (1), FAM49B (2), TAGLN2 (1), SLPI (1), C4B (3), C3 (1), C1R (1), SERPINA1 (2) (Supplemental Figure 6).

In the OSCC no lesion versus OSCC lesion conditions, we observed that four proteins did not have significant changes in abundance: SLPI, SERPINE1, C1R and FAM49B. The remaining eight proteins were up-regulated in the OSCC lesion condition (Supplemental Table SM7). Similar results were observed at a peptide level. Peptides from FAM49B (2), SERPINE1 (2), C1R (1) and SLPI (1) and TAGLN2 (2) did not demonstrate a difference at the peptide level (Supplemental Figure 7).

3.6 Comparison between relative quantification of CFB using labeled peptide or global standards

The ratio of the YGLVITYATYPK and VSEADSSNADWVTK peptides to their heavy isotope labeled analogs was compared to the ratio of these same peptides to the set of global standards. Integrated signal intensities were exported using Skyline (Supplemental tables 4 and 9) and compared by utilizing statistical significance among the three conditions (control, OSCC no lesion, and OSCC lesion) with a Student's *t*-test.

Both normalization approaches showed statistical significance between control and lesion condition for the two CFB peptides (Student's *t*-test, $p < 0.05$, Supplemental Figure 8). Both normalization approaches resulted in no statistically significant differences between the control versus OSCC no lesion and OSCC no lesion versus OSCC lesion comparison.

Although SRM with stable isotope-labeled standards is considered to be the “gold standard” MS-based quantitation method [38, 39], our results demonstrate that comparative analytical assay performance can be obtained by utilizing external reference standards.

However, it is important to remember that there are different “fit-for-purpose” approaches for targeted proteomics [40]. For comparative analysis, as in the case of this study, using semiquantitative measurements of proteins, protein forms, or peptides in biological systems (Tier 3), label-free measurement can be used with additional chromatographic and mass spectrometric information to establish confidence in peptide assignment. If the purpose is clinical bioanalysis/ diagnostic, laboratory test (Tier 1) or research use assays for quantifying proteins, peptides, and posttranslational modifications (Tier 2), best practice dictates the use of stable isotope-labeled internal standards for each targeted analyte.

3.7 Association analysis of OSCC presence, using the targeted peptides measured by SRM

To explore the relevance of the protein expression in clinical application, we used the peptide levels for each protein evaluated in this study to perform risk association to OSCC analysis. The normalized ratios for targeted peptides to global standards from 31 peptides, considering only those filtered by $\text{dotp} > 0.8$, in 30 samples are shown in the Supplemental Table 4.

The clinicopathological variables revealed that patients were mainly diagnosed at advanced clinical stage (most have lymph nodes involved) with a predominantly moderate/poor WHO differentiation degree and active lesions at the time of saliva collection (Table 2). Using univariate factor binary logistic regression, for a categorical variable OSCC presence/absence we observed an association between the CFB_1 (YGLVITYATYPK), C3_3 (TFISPIK), C4B_1 (TTNIQGINLLFSSR), C4B_3 (GSFEFPVGDAVSK), LRG1_1 (DLLLPQPDRL), LRG1_2 (YLFLNGNK) and SERPINA1_3 (LSITGTYDLK) peptides with OSCC **presence**. For all these peptides, the odds ratio for OSCC was 12.50 (95% of confidence interval (CI) and $p = 0.03$) (Table 3, Supplemental Material and Methods).

Interestingly, four out of five of these proteins (C3, CFB, C4B and SERPINA1) are involved in the complement and coagulation cascade, which was previously demonstrated to be enriched by network analysis using the carcinoma candidate biomarkers (KEGG pathway, $p < 0.05$) (Kawahara et al., 2015, under review).

The complement system has recently received increased attention in biomedical research. In light of evidence that complement proteins can facilitate cellular proliferation and contribute to the inflammatory state of the cancer [41, 42], there is increasing speculation that there is a relationship between complement proteins and cancer. In previous work, we demonstrated the role of CFB in tumor cell migration and macrophage chemotaxis (Kawahara et al., 2015, under review). Using serum proteomics, previous studies demonstrated the involvement of

complement proteins in cancer, for example C3a and C3f in colorectal cancer [43–45] and C4a and C3a in esophageal cancer [46]. Recently, Cho et al. 2014 [47] also described a role for the complement system in enhancing cancer growth and C3 was implicated as a key player for the activation of ovarian cancer growth and progression. Furthermore, Kim et al. (2014) suggested C3 as a plasma biomarker for the measurement disease progression in neuroblastoma patients [48] and Lee et al. (2014) showed CFB as biomarker candidate for pancreatic ductal adenocarcinoma [49]. Thus, a panel of complement proteins has emerged as potential candidates associated with cancer development. Our study suggests that these candidates are present in high abundance in saliva of OSCC patients and that they were also associated with OSCC risk. Collectively, our results open the potential of novel avenues for OSCC follow-up studies.

4. Concluding Remarks

In summary, this study presented a targeted proteomic approach for the measurement of OSCC candidate biomarkers in human saliva. We demonstrated a statistically significant higher abundance of ten candidate biomarkers in saliva of OSCC patients and among them, CFB, C3, C4B, SERPINA1 and LRG1 are associated with the risk of developing OSCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge FAPESP: 2009/54067-3, 2010/19278-0 and 2013/21603-5. We thank Dr. Michael MacCoss for supporting Rebeca Kawahara in the Research Internships Abroad (BEPE).

Abbreviations

ANXA1	Annexin A1
C1R	Complement C1r subcomponent
C3	Complement C3
C4B	Complement C4-B
CFB	Complement Factor B
DOTP	Dot-product
DPBS	Dulbecco's phosphate-buffered saline
FAM49B	Protein FAM49B
GST	Glutathione S-transferase
iRT	Retention time information
LCN2	Neutrophil gelatinase-associated lipocalin
LOD	Limit of detection

LOQ	Limit of quantification
LRG1	Leucine-rich alpha-2-glycoprotein
MRM	Multiple reaction monitoring
NSC	Nearest shrunken centroids
OSCC	Oral squamous cell carcinoma
SERPINA1	Alpha-1-antitrypsin
SERPINE1	Plasminogen activator inhibitor 1
SJ	Schistosoma japonicum
SLPI	Antileukoproteinase
SRM	Selected reaction monitoring
SVM-RFE	Support vector machine-recursive features elimination
TAGLN2	Transgelin-2

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: a cancer journal for clinicians. 2013; 63:11–30. [PubMed: 23335087]
2. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. Lancet. 2008; 371:1695–1709. [PubMed: 18486742]
3. Mydlarz WK, Hennessey PT, Califano JA. Advances and Perspectives in the Molecular Diagnosis of Head and Neck Cancer. Expert opinion on medical diagnostics. 2010; 4:53–65. [PubMed: 20161611]
4. da Silva SD, Ferlito A, Takes RP, Brakenhoff RH, Valentin MD, Woolgar JA, et al. Advances and applications of oral cancer basic research. Oral oncology. 2011; 47:783–791. [PubMed: 21802978]
5. Hall SF, Groome PA, Rothwell D, Dixon PF. Using TNM staging to predict survival in patients with squamous cell carcinoma of head & neck. Head & neck. 1999; 21:30–38. [PubMed: 9890348]
6. Bonne NJ, Wong DT. Salivary biomarker development using genomic, proteomic and metabolomic approaches. Genome medicine. 2012; 4:82. [PubMed: 23114182]
7. Hu S, Arellano M, Boontheung P, Wang J, Zhou H, Jiang J, et al. Salivary proteomics for oral cancer biomarker discovery. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008; 14:6246–6252. [PubMed: 18829504]
8. de Jong EP, Xie H, Onsongo G, Stone MD, Chen XB, Kooren JA, et al. Quantitative proteomics reveals myosin and actin as promising saliva biomarkers for distinguishing pre-malignant and malignant oral lesions. PloS one. 2010; 5:11148.
9. Chang KP, Yu JS, Chien KY, Lee CW, Liang Y, Liao CT, et al. Identification of PRDX4 and P4HA2 as metastasis-associated proteins in oral cavity squamous cell carcinoma by comparative tissue proteomics of microdissected specimens using iTRAQ technology. Journal of proteome research. 2011; 10:4935–4947. [PubMed: 21859152]
10. Sepiashvili L, Hui A, Ignatchenko V, Shi W, Su S, Xu W, et al. Potentially novel candidate biomarkers for head and neck squamous cell carcinoma identified using an integrated cell line-based discovery strategy. Molecular & cellular proteomics : MCP. 2012; 11:1404–1415. [PubMed: 22918226]
11. Hsu CW, Yu JS, Peng PH, Liu SC, Chang YS, Chang KP, et al. Secretome Profiling of Primary Cells Reveals That THBS2 Is a Salivary Biomarker of Oral Cavity Squamous Cell Carcinoma. Journal of proteome research. 2014

12. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies Nature clinical practice. Oncology. 2008; 5:588–599. [PubMed: 18695711]
13. Makawita S, Diamandis EP. The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification. Clinical chemistry. 2010; 56:212–222. [PubMed: 20007861]
14. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nature biotechnology. 2006; 24:971–983.
15. Sturgeon C, Hill R, Hortin GL, Thompson D. Taking a new biomarker into routine use--a perspective from the routine clinical biochemistry laboratory Proteomics. Clinical applications. 2010; 4:892–903. [PubMed: 21137030]
16. Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, et al. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nature methods. 2010; 7:43–46. [PubMed: 19966807]
17. Whiteaker JR, Lin C, Kennedy J, Hou L, Trute M, Sokal I, et al. A targeted proteomics-based pipeline for verification of biomarkers in plasma. Nature biotechnology. 2011; 29:625–634.
18. Stergachis AB, MacLean B, Lee K, Stamatoyannopoulos JA, MacCoss MJ. Rapid empirical discovery of optimal peptides for targeted proteomics. Nature methods. 2011; 8:1041–1043. [PubMed: 22056677]
19. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010; 26:966–968. [PubMed: 20147306]
20. Maclean B, Tomazela DM, Abbatiello SE, Zhang S, Whiteaker JR, Paulovich AG, et al. Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. Analytical chemistry. 2010; 82:10116–10124. [PubMed: 21090646]
21. Surinova S, Huttenhain R, Chang CY, Espona L, Vitek O, Aebersold R. Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. Nature protocols. 2013; 8:1602–1619. [PubMed: 23887179]
22. Chang CY, Picotti P, Huttenhain R, Heinzelmann-Schwarz V, Jovanovic M, Aebersold R, et al. Protein significance analysis in selected reaction monitoring (SRM) measurements. Molecular & cellular proteomics : MCP. 2012; 11:M111 014662.
23. Choi M, Chang CY, Clough T, Broudy D, Killeen T, Maclean B, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics. 2014
24. Broudy D, Killeen T, Choi M, Shulman N, Mani DR, Abbatiello SE, et al. A framework for installable external tools in Skyline. Bioinformatics. 2014; 30:2521–2523. [PubMed: 24813211]
25. Sherrod SD, Myers MV, Li M, Myers JS, Carpenter KL, Maclean B, et al. Label-free quantitation of protein modifications by pseudo selected reaction monitoring with internal reference peptides. Journal of proteome research. 2012; 11:3467–3479. [PubMed: 22559222]
26. Benjamini YHY. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B. 1995; 57:289–300.
27. Hiromoto T, Noguchi K, Yamamura M, Zushi Y, Segawa E, Takaoka K, et al. Up-regulation of neutrophil gelatinase-associated lipocalin in oral squamous cell carcinoma: relation to cell differentiation. Oncology reports. 2011; 26:1415–1421. [PubMed: 21874256]
28. Shinriki S, Jono H, Ueda M, Obayashi K, Nakamura T, Ota K, et al. Stromal expression of neutrophil gelatinase-associated lipocalin correlates with poor differentiation and adverse prognosis in oral squamous cell carcinoma. Histopathology. 2014; 64:356–364. [PubMed: 24118080]
29. Noorlag R, van der Groep P, Leusink FK, van Hooff SR, Frank MH, Willems SM, et al. Nodal metastasis and survival in oral cancer: Association with protein expression of SLPI, not with LCN2, TACSTD2, or THBS2. Head & neck. 2015; 37:1130–1136. [PubMed: 24764155]
30. Andersen JD, Boylan KL, Jemmerson R, Geller MA, Misemer B, Harrington KM, et al. Leucine-rich alpha-2-glycoprotein-1 is upregulated in sera and tumors of ovarian cancer patients. Journal of ovarian research. 2010; 3:21. [PubMed: 20831812]

31. Wu J, Yin H, Zhu J, Buckanovich RJ, Thorpe JD, Dai J, et al. Validation of LRG1 as a potential biomarker for detection of epithelial ovarian cancer by a blinded study. *PloS one*. 2015; 10:0121112.
32. Ladd JJ, Busald T, Johnson MM, Zhang Q, Pitteri SJ, Wang H, et al. Increased plasma levels of the APC-interacting protein MAPRE1, LRG1, and IGFBP2 preceding a diagnosis of colorectal cancer in women. *Cancer prevention research*. 2012; 5:655–664. [PubMed: 22277732]
33. Chang KP, Lin SJ, Liu SC, Yi JS, Chien KY, Chi LM, et al. Low-molecular-mass secretome profiling identifies HMGA2 and MIF as prognostic biomarkers for oral cavity squamous cell carcinoma. *Scientific reports*. 2015; 5:11689. [PubMed: 26138061]
34. Nohata N, Sone Y, Hanazawa T, Fuse M, Kikkawa N, Yoshino H, et al. miR-1 as a tumor suppressive microRNA targeting TAGLN2 in head and neck squamous cell carcinoma. *Oncotarget*. 2011; 2:29–42. [PubMed: 21378409]
35. Shi YY, Wang HC, Yin YH, Sun WS, Li Y, Zhang CQ, et al. Identification and analysis of tumour-associated antigens in hepatocellular carcinoma. *British journal of cancer*. 2005; 92:929–934. [PubMed: 15756260]
36. Rho JH, Roehrl MH, Wang JY. Tissue proteomics reveals differential and compartment-specific expression of the homologs transgelin and transgelin-2 in lung adenocarcinoma and its stroma. *Journal of proteome research*. 2009; 8:5610–5618. [PubMed: 19848416]
37. Chen R, Yi EC, Donohoe S, Pan S, Eng J, Cooke K, et al. Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape. *Gastroenterology*. 2005; 129:1187–1197. [PubMed: 16230073]
38. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology*. 2008; 4:222. [PubMed: 18854821]
39. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Molecular & cellular proteomics : MCP*. 2007; 6:2212–2229. [PubMed: 17939991]
40. Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Molecular & cellular proteomics : MCP*. 2014; 13:907–917. [PubMed: 24443746]
41. Rutkowski MJ, Sughrue ME, Kane AJ, Mills SA, Parsa AT. Cancer and the complement cascade. *Molecular cancer research : MCR*. 2010; 8:1453–1465. [PubMed: 20870736]
42. Rao SK, Pavicevic Z, Du Z, Kim JG, Fan M, Jiao Y, et al. Pro-inflammatory genes as biomarkers and therapeutic targets in oral squamous cell carcinoma. *The Journal of biological chemistry*. 2010; 285:32512–32521. [PubMed: 20702412]
43. Zhu D, Wang J, Ren L, Li Y, Xu B, Wei Y, et al. Serum proteomic profiling for the early diagnosis of colorectal cancer. *Journal of cellular biochemistry*. 2013; 114:448–455. [PubMed: 22961748]
44. Habermann JK, Roblick UJ, Luke BT, Prieto DA, Finlay WJ, Podust VN, et al. Increased serum levels of complement C3a anaphylatoxin indicate the presence of colorectal tumors. *Gastroenterology*. 2006; 131:1020–1029. quiz 284. [PubMed: 17030172]
45. Ward DG, Suggett N, Cheng Y, Wei W, Johnson H, Billingham LJ, et al. Identification of serum biomarkers for colon cancer by proteomic analysis. *British journal of cancer*. 2006; 94:1898–1905. [PubMed: 16755300]
46. Maher SG, McDowell DT, Collins BC, Muldoon C, Gallagher WM, Reynolds JV. Serum proteomic profiling reveals that pretreatment complement protein levels are predictive of esophageal cancer patient response to neoadjuvant chemoradiation. *Annals of surgery*. 2011; 254:809–816. discussion 16–7. [PubMed: 22005152]
47. Cho MS, Vasquez HG, Rupaimoole R, Pradeep S, Wu S, Zand B, et al. Autocrine effects of tumor-derived complement. *Cell reports*. 2014; 6:1085–1095. [PubMed: 24613353]
48. Kim PY, Tan O, Diakiw SM, Carter D, Sekerye EO, Wasinger VC, et al. Identification of plasma complement C3 as a potential biomarker for neuroblastoma using a quantitative proteomic approach. *Journal of proteomics*. 2014; 96:1–12. [PubMed: 24200836]

49. Lee MJ, Na K, Jeong SK, Lim JS, Kim SA, Lee MJ, et al. Identification of human complement factor B as a novel biomarker candidate for pancreatic ductal adenocarcinoma. *Journal of proteome research*. 2014; 13:4878–4888. [PubMed: 25057901]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Statement of significance of the study

Herein we present a feasible and successful pipeline to build an SRM assay that targets oral OSCC candidate biomarkers in human saliva. Saliva is a great choice matrix for this task as it is an easy-to-obtain body fluid that can be collected noninvasively. We demonstrated a statistically significant higher abundance of the C1R, LCN2, SLPI, FAM49B, TAGLN2, CFB, C3, C4B, LRG1, SERPINA1 candidate biomarkers in OSCC patients. Furthermore, our study also demonstrated that CFB, C3, C4B, SERPINA1 and LRG1 are associated with the risk of developing OSCC. Overall, this study successfully used targeted proteomics to measure a panel of novel biomarker candidates in OSCC that can be further correlated with prognosis and evaluated in larger cohort studies.

Figure 1. Experimental workflow of SRM assay development using empirical measurement for selection of proteotypic peptides.

Figure 2. Analytical assay performance using CFB heavy peptides and global standards
 (A) Recombinant CFB tryptic peptide intensities from 7 to 23 amino acids in length. The arrows show the proteotypic peptides measured in the saliva. The red arrows show peptides synthesized in the labeled form. B) Resulting SRM mass spectra (left, endogenous specific transitions for the different peptides corresponding to the protein used for quantification and right, corresponding heavy peptides) of VSEADSSNADWVTK and C) YGLVTYATYPK.

Figure 3. Proteotypic peptides identified using *in vitro*-synthesized protein monitored in human saliva

The relative contribution of each fragment ion to each peptide peak is displayed as different colors. The library contains the result from the recombinant, *in vitro*-synthesized or isolated from plasma proteins. The graph represents endogenous specific transitions for the different peptides measured in the human saliva.

Figure 4. Visualization for explanatory data analysis using MSstats

(A) Examples of quality control (QC) plots for Global Standards, SERPINA1 and CFB proteins. X-axis: run. Y-axis: log-intensities of transitions. B) Examples of profile plots for Global Standards, SERPINA1 and CFB proteins. Line colors indicate peptides and line types indicate transitions. C) Examples of condition plots for Global Standards, SERPINA1 and CFB proteins. X-axis: condition. Y-axis: log ratio of endogenous over reference intensities of each transition in a run. Dots indicate the mean of log ratio for each condition. Error bars have confidence intervals with 0.95 significant level for each condition. The plots visualize the differences between conditions, which are of the main biological interest.

Figure 5. Volcano plot of the comparison Control vs Lesion condition

X-axis: practical significance, model-based estimate of log-fold change. Y-axis: statistical significance, FDR-adjusted p-values on the negative log2 scale. The dashed line represents the FDR cutoff (default sig=0.05) and fold-change FCcutoff=1.5. A) Results at a protein level. Up regulated in OSCC (red): C1R, LCN2, SLPI, FAM49B, TAGLN2, CFB, C3, C4B, LRG1 and SERPINA1. No regulation (black): ANXA1, SERPINE1 and Global Standards. B) Results at a peptide level. Up-regulated in OSCC (red): ANXA1 (peptides TPA, NAL); CFB (peptides YGL, VSE, DIS); C3 (peptides LMN, SSL, TFI); C4B (peptides TTN, VGD, GSF); C1R (peptide DYF); LCN2 (peptides MYA, TFV, SYP); LRG1 (peptides DLL, YLF, VAA); FAM49B (peptides MSL, VML); SERPINA1 (peptides LSI, SVL, AVL); SERPINE1 (peptide FII), SLPI (peptide CLD); TAGLN2 (peptides IQA, TLM, NVI). No regulation (black): ANXA1 (peptide ALY); SERPINE1 (peptide GMI) and C1R (peptide NLP).

Table 1

Peptides used for normalization and the respective CV calculated for all runs.

Protein Name	Peptide	CV%
CFB (heavy)	YGLVTYATYPK	15.5%
CFB (heavy)	VSEADSSNADWVTK	13.8%
GST	IEAIPQIDK	19.50%
Pierce RT	SSAAPPPPPR	7.50%
Pierce RT	IGDYAGIK	13.40%
Pierce RT	SAAGAFGPESLR	21.00%
Pierce RT	SFANQPLEVVYSK	19.80%
Pierce RT	LTILEELR	25.00%

Table 2

Clinical and pathological characteristics of OSCC patients (n=22)

Features	OSCC n (%)
Sex	
Female	3 (13.6)
Male	19 (86.4)
TNM ¹	
Primary tumor (T)	
1	2 (15.4)
2	4 (30.8)
3	2 (15.4)
4	5 (38.5)
Regional lymph nodes (N)	
0	3 (23.1)
1	4 (30.8)
2	6 (46.2)
3	0
Metastasis (M)	
M0	2 (15.4)
Mx	11 (84.6)
WHO differentiation degree ²	
Well	5 (29.4)
Moderate	7 (41.2)
Poor	5 (29.4)
Active lesion ¹	
No	9 (40.9)
Yes	13 (59.1)
Survival state	
Alive	14 (63.6)
Deceased	8 (36.4)

¹
n=22,²
n=17

Table 3

Association analysis of OSCC presence, using the peptides intensities normalized by global standards, measured by SRM.

Peptides ^a	Peptide Sequence	Odds ratio ^b	IC 95%	p-value
ANXA1_1	TPAQFDADELRL	0.231	0.38–1.413	0.113
ANXA1_2	NALLSLAK	0.231	0.38–1.413	0.113
ANXA1_3	ALYEAGER	1	0.198–5.045	1
CFB_1	YGLVTYATYPK	12.250	1.268–118.361	0.030 *
CFB_2	VSEADSSNADWVTK	4.333	0.708–26.531	0.113
CFB_3	DISEVVTPLR	4.333	0.708–26.531	0.113
C1R_1	DYFIATCK	4.333	0.708–26.531	0.113
C1R_2	NLPNGDFR	0.231	0.038–1.413	0.113
C3_1	LMNIFLK	4.333	0.708–26.531	0.113
C3_2	SSLSVPYVIVPLK	4.333	0.708–26.531	0.113
C3_3	TFISPIK	12.250	1.268–118.361	0.030 *
C4B_1	TTNIQGINLLFSSR	12.250	1.268–118.361	0.030 *
C4B_2	VGDTLNLNLR	1.846E9	0.0–iv	0.998
C4B_3	GSFEFPVGDAVSK	12.250	1.268–118.361	0.030 *
LCN2_1	MYATIELK	2	0.381–10.511	0.413
LCN2_2	TFVPGCQPGFTLGNIK	1	0.198–5.045	1
LCN2_3	SYPLTSLVLR	1	0.198–5.045	1
LRG1_1	DLLLPQDLR	12.250	1.268–118.361	0.030 *
LRG1_2	YLFLNGNK	12.250	1.268–118.361	0.030 *
LRG1_3	VAAGAFQGLR	1.846E9	0.0–iv	0.998
FAM49B_1	MSLFYAEATPMLK	2	0.381–10.511	0.413
FAM49B_2	VMLETPEYR	1	0.198–5.045	1
SERPINA1_1	LSITGTYDLK	1.612E9	0.0–iv	0.998
SERPINA1_2	SVLGQLGITK	8.4	0.879–80.265	0.065
SERPINA1_3	AVLTIDEK	12.250	1.268–118.361	0.030 *
SERPINE1_1	FIINDWVK	0.5	0.095–2.628	0.413
SERPINE1_2	GMISNLLGK	0.867	0.164–4.579	0.866
SLPI	CLDPVDTNPTR	0.952	0.179–5.081	0.954
TAGLN2_1	IQASTMAFK	1.0	0.198–5.045	1.000
TAGLN2_2	TLMNLGGLAVAR	1.0	0.198–5.045	1.000
TAGLN2_3	NVIGLQMGTR	1.0	0.198–5.045	1.000

(a) Categorized as 'low' and 'high'.

(b) Odds ratio for Low/High.

(iv) Indeterminate value.

*
p<0.05

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript