Salivary metabolite signatures of oral cancer and leukoplakia

Jie Wei1*, Guoxiang Xie2*, Zengtong Zhou3*, Peng Shi1, Yunping Qiu2, Xiaojiao Zheng3, Tianlu Chen4, Mingming Su5, Aihua Zhao3 and Wei Jia2

1 Department of Oral Mucosal Diseases, Ninth People’s Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Key Laboratory of Stomatology, Shanghai, 200011, China
2 Department of Nutrition, University of North Carolina at Greensboro, North Carolina Research Campus, Kannapolis, NC 28081
3 School of Pharmacy, Shanghai Jiao Tong University, Shanghai, 200240, People’s Republic of China
4 Ministry of Education Key Laboratory of Systems Biomedicine, Shanghai Center for System Biomedicine, Shanghai jiao Tong University, Shanghai, 200240, People’s Republic of China
5 David H. Murdock Research Institute, North Carolina Research Campus, Kannapolis, NC 28081

Oral cancer, one of the six most common human cancers with an overall 5-year survival rate of <50%, is often not diagnosed until it has reached an advanced stage. The aim of the current study is to explore salivary metabolomics as a disease diagnostic and stratification tool for oral cancer and leukoplakia and evaluate the potential of salivary metabolome for detection of oral squamous cell carcinoma (OSCC). Saliva metabolite profiling for a group of 37 OSCC patients, 32 oral leukoplakia (OLK) patients and 34 healthy subjects was performed using ultraperformance liquid chromatography coupled with quadrupole/time-of-flight mass spectrometry in conjunction with multivariate statistical analysis. The OSCC, OLK and healthy control groups demonstrate characteristic salivary metabolic signatures. A panel of five salivary metabolites including γ-aminobutyric acid, phenylalanine, valine, n-eicosanoic acid and lactic acid were selected using OPLS-DA model with S-plot. The predictive power of each of the five salivary metabolites was evaluated by receiver operating characteristic curves for OSCC. Valine, lactic acid and phenylalanine in combination yielded satisfactory accuracy (0.89, 0.97), sensitivity (86.5% and 94.6%), specificity (82.4% and 84.4%) and positive predictive value (81.6% and 87.5%) in distinguishing OSCC from the controls or OLK, respectively. The utility of salivary metabolome diagnostics for oral cancer is successfully demonstrated in this study and these results suggest that metabolomics approach complements the clinical detection of OSCC and stratifies the two types of lesions, leading to an improved disease diagnosis and prognosis.

About 1.5 million new cancer cases are expected to be diagnosed in 2009 in the United States and >0.5 million Americans are expected to die of cancer this year, averaging about 1,500 deaths per day.1 These numbers have been steadily increasing over the past 15 years, despite significant progress in cancer treatment. Oral cancer represents one of the six most common human cancers with a high morbidity rate and an overall 5-year survival rate of <50%.2-3 Reports indicate an increasing worldwide incidence of oral cancer at an earlier age.4-6 Over 90% of oral cancer is oral squamous cell carcinoma (OSCC) which arises from the oral mucosal lining.7 A critical factor in the lack of prognostic improvement is the fact that a significant proportion of cancers initially are asymptomatic lesions and are not diagnosed or treated until...
they reach an advanced stage. In OSCC, if the cancer is detected at T1 stage (T means the tumor size and invasion level), the 5-year survival rate is over 80%, compared to 20–40% if the cancer is diagnosed at later stages (T3 and T4). Oral leukoplakia (OLK) is the most frequent oral precancerous lesion with considerable malignant transformation ranging from 1.58 to 27.27%. Early detection of OSCC and OLK, as well as screening of high risk populations, is promising strategies for reducing the incidence of OSCC. Scientists expect to find high-throughput, low cost, more efficient and rapid diagnostic and screening approaches. Currently, the most definitive procedure for oral cancer diagnosis and screening involves a visit to the physician’s office, a scalpel biopsy usually on the tongue or gums, followed by a histopathological evaluation. In recent years, salivary transcriptome and proteome analyses for oral cancer has been reported and a number of specific genes and proteins have been proposed as biomarkers for clinical diagnostics. Recently, Nagler et al. conducted an analysis of OSCC patient’s saliva, in which significant increases in salivary concentrations of Cyfra 21-1, tissue polypeptide antigen and cancer antigen (CA125) were observed.

Human saliva is a mixture of secretions from multiple salivary glands, including the parotid, submandibular, sublingual and other minor glands lying beneath the oral mucosa. It is increasingly being viewed as a way to screen for diseases and was recently referred to as “the mirror of the body,” in the sense that it is the good biological medium for health and disease surveillance. Saliva may contain specific biomarkers associated with certain diseases, it has been used in diagnostics for >2000 years by many traditional medical systems such as traditional Chinese medicine.

It is now generally agreed that OSCC and precancerous lesions involve not only specially expressed genes and proteins (changes in the concentration of saliva composition, including, median total protein, sodium, potassium, calcium, inorganic phosphate, magnesium, albumin, etc.) but also changes in the concentration of endogenous metabolites. Metabolic profile of biofluids (urine, serum and saliva) can be altered by a variety of physiological processes following pathophysiological stimuli, therefore, global perturbation in these profiles may demonstrate the presence of a particular disease. Metabolomics or metabonomics is a complementary approach for early detection of oral cancer (OSCC), which utilizes a novel and unique strategy that provides a coherent perspective of the complete metabolic response of organisms to pathophysiological stimuli or genetic modification.

Salivary metabonomic analysis of OSCC and OLK has been performed by high performance liquid chromatography-mass spectrometry (HPLC-MS) and a total of 14 OSCC-related and 11 OLK-related biomarkers were discovered, although these discriminant metabolites were not identified. Recently, a capillary electrophoresis time-of-flight mass spectrometry based salivary metabolomics analysis of oral, breast and pancreatic cancer has been performed and a number of metabolites related were identified. Metabonomic analysis of serum samples of OSCC patients was also performed by nuclear magnetic resonance (NMR) spectroscopy, with a good discrimination between OSCC and the healthy controls. A panel of 23 discriminatory serum metabolites including valine, lactic acid and phenylalanine were identified. In the current work, we used an advanced version of LC-MS platform, ultraperformance liquid chromatography coupled with quadrupole/time-of-flight mass spectrometry (UPLC-QTOFMS), to study salivary metabolomics of the three groups, OSCC, OLK and healthy subjects. The purpose of this study is to identify salivary metabolites as potential biomarkers to diagnose and stratify OSCC and precancerous lesions using this high resolution and high sensitivity analytical platform.

**Material and Methods**

**Patient selection**

Saliva samples were collected from a group of 37 OSCC patients (26 men and 11 women, 9 of stage I, 12 of stage II, 6 of stage III and 10 of stage IV), whose mean age was 56 ± 11 years (34–77), 32 OLK patients (13 men and 19 women), whose mean age was 60 ± 13 years (34–80). They were compared to a control group of 34 healthy individuals (13 men and 21 women), whose mean age was 43 ± 14 years (21–73). Clinical information of the participants was provided in Supporting Information Table 1. They were all recruited from the Department of Oral Medicine and Surgery, School of Stomatology, Shanghai Jiao Tong University. There was no history of receiving medication and none had been treated with topical or systemical steroids. Diagnosis for all cases was based on clinical and histopathologic criteria. The Ethical Committee of the School of Stomatology, Shanghai Jiao Tong University, approved the protocol and all of the subjects signed an Ethical Committee consent form agreeing to serve as saliva donors for the experiments.

**Sample collection and preparation**

Saliva was collected between 9:00 and 10:00 a.m. with previously established protocols described by Navazesh. Briefly, subjects were refrained from eating, drinking, smoking or oral hygiene procedures for at least 1.5 hr before the collection. Saliva samples were centrifuged at 3,500g for 20 min at 4°C and the resulting supernatants were immediately stored at −80°C pending UPLC-QTOFMS analysis. Before analysis, the samples were thawed. Ultrapure water (200 µL) and acetonitrile (200 µL) were added to 200 µL of saliva and vortexed for 1 min, the mixture was allowed to stand for 10 min and centrifuged at 15,400g for 15 min. The supernatant was filtered through a syringe filter (0.22 µm) for UPLC-QTOFMS analysis.

**Salivary metabolomics analysis**

Saliva metabolite profiling was performed using a Waters ACQUITY UPLC system (Waters Corporation, Milford,
Table 1. Summary of the discriminant metabolites from VIP values of three component PLS-DA models accountable for variations among OSCC, OLK and control groups

<table>
<thead>
<tr>
<th>Retention time (min)</th>
<th>Metabolites identified</th>
<th>OSCC–control</th>
<th>OSCC–OLK</th>
<th>OLK–control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Rank</td>
<td>VIP&lt;sup&gt;²&lt;/sup&gt;</td>
<td>p-value&lt;sup&gt;³&lt;/sup&gt;</td>
</tr>
<tr>
<td>0.63</td>
<td>Alanine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>Lactic acid</td>
<td>↑(2)</td>
<td>7.62</td>
<td>2.13E-2</td>
</tr>
<tr>
<td>0.79</td>
<td>Valine</td>
<td>↓(1)</td>
<td>12.55</td>
<td>2.59E-4</td>
</tr>
<tr>
<td>0.99</td>
<td>γ-Aminobutyric acid</td>
<td>↓(5)</td>
<td>2.07</td>
<td>5.84E-4</td>
</tr>
<tr>
<td>1.22</td>
<td>Isoleucine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.91</td>
<td>Leucine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.20</td>
<td>n-Tetradecanoic acid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.21</td>
<td>Proline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.22</td>
<td>Phenylalanine</td>
<td>↓(4)</td>
<td>3.12</td>
<td>5.32E-5</td>
</tr>
<tr>
<td>2.61</td>
<td>Threonine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.95</td>
<td>n-Dodecanoic acid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.18</td>
<td>3-Indolepropionic acid</td>
<td>↑(5)</td>
<td>7.51</td>
<td>6.52E-04</td>
</tr>
<tr>
<td>15.18</td>
<td>n-Eicosanoic acid</td>
<td>↑(3)</td>
<td>3.64</td>
<td>9.19E-2</td>
</tr>
<tr>
<td>1.01</td>
<td>Homocysteine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.59</td>
<td>4-Methoxyphenylacetic acid</td>
<td>↑(6)</td>
<td>3.46</td>
<td>7.22E-03</td>
</tr>
</tbody>
</table>

† represents significantly elevated concentration, whereas ↓ represents significantly lowered concentration (OSCC, OLK group compared to healthy control group, and OSCC compared to OLK group).<sup>1</sup>
Metabolites are identified using available standards.<sup>2</sup>
VIP was obtained from OPLS-DA with a threshold of 1.0.<sup>3</sup>
<p-value> and fold change (FC) are calculated from nonparametric Wilcoxon-Mann-Whitney test (one-way ANOVA).<sup>4</sup>
All the metabolites were discriminant (Wilcoxon <i>p</i> < 0.05), with an FDR of 5%.<sup>5</sup>
FC with a value larger than 1 indicates a relatively higher concentration present in OSCC patients while a value lower than 1 means a relatively lower concentration as compared to the OLK patients and healthy controls. A fold change (>1) also indicates a relatively higher concentration present in OLK patients as compared to the healthy controls, while a value (<1) means a relatively lower concentration.
MA), which was equipped with a binary solvent delivery manager, and a sample manager coupled to Micromass Q-TOF Premier mass spectrometry (Waters Corporation, Milford, MA) equipped with an electrospray interface. Chromatographic separations were performed on a 2.1 × 100 mm² 1.7-µm ACQUITY BEH C18 chromatography column. The column was maintained at 45°C and eluted with a 5–95% acetonitrile (0.1 % (v/v) formic acid)—aqueous formic acid (0.1 % (v/v) formic acid) gradient over 18 min at a flow rate of 0.35 mL/min. A 5 µL aliquot was injected onto the column. The mass accuracy analysis and detailed MS parameters were optimized according to our previous work. During metabolite profiling experiments, centroid data were acquired for each sample from 50 to 1000 Da with a 0.10-sec scan time and a 0.01-sec interscan delay over an 18-min analysis time.

Data processing and statistical analysis
The UPLC-QTOFMS data of OSCC, OLK and control saliva samples were analyzed to identify potential discriminant variables. The positive mode electrospray raw data were analyzed by the MarkerLynx applications manager version 4.1 (Waters, Manchester, UK) and the parameters used were set according to our previous work. The parameters used were RT range 0–18 min, mass range 50–1000 Da, mass tolerance 0.02 Da, isotopic peaks were excluded for analysis, noise elimination level was set at 10.00, minimum intensity was set to 10% of base peak intensity, maximum masses per RT was set at 6, and, finally, RT tolerance was set at 0.01 min. After creating a suitable processing method, the next step is to process the dataset through the Create Dataset window. Selecting the method, we just created and creating our dataset, we would like to (i) detect Peaks and (ii) collect markers. Select all of these options from the processing options panel of the create dataset display, and at this point, it is also possible to automatically print reports and export data into a text file for use in third party software such as SIMCA-P. A list of the intensities of the peaks detected was generated for the first sample, using retention time and m/z data pairs as the identifier of each peak. The resulting three-dimensional matrix containing arbitrarily assigned peak index (retention time (RT)-m/z pairs), sample names (observations) and normalized peak areas (variables) were further exported to SIMCA-P software 11.5 (Umetrics, Umeå, Sweden) for multivariate statistical analysis. After data normalization, multivariate statistical methods such as principal component analysis (PCA) were performed to differentiate distinct salivary metabolite signatures of all groups and to identify the major discriminant metabolites between the three groups, the data were subsequently interrogated using an orthogonal partial least squares-discriminant analysis (OPLS-DA) model to identify the discriminant metabolites in OSCC group relative to healthy control subjects or OLK. The data sets were mean-centered, pareto-scaled in a columnwise manner before PCA and OPLS-DA modeling. As compared to UV-scaling (scaling to unit variance) method, the advantage of using this technique lies in the fact that it enhances the contribution of lower concentration metabolites without amplifying noise and artifacts commonly present in the metabolomic data sets. Additionally, to maximize the variations among the three groups, sophisticated supervised methods including OPLS-DA was further applied. PLS is a generalized multiple regression method that can deal with multiple collinear X and Y variables and has two objectives: one is to well approximate X and Y, and the other is to model the relationship between them. R²X and R²Y represent the fraction of the variance of X matrix and Y matrix, respectively, while Q²Y suggests the predictive accuracy of the model. The cumulative values of R²X, R²Y and Q²Y close to 1 indicate an excellent model. Variable importance in the projection (VIP) ranks the overall contribution of each variable to the OPLS-DA model, and those variables with VIP > 1.0 are considered relevant for group discrimination. Herein, VIP statistics and S-plot were applied to obtain the significant variables for subsequent metabolic pathway analysis. In SIMCA-P package, a typical crossvalidation procedure was conducted by leaving 1/7th samples out in each round so as to validate the OPLS-DA model against overfitting. To compare the combined MS dataset from OSCC to control and from OSCC to OLK, on a basis of the threshold of p values (<0.05) and fold change values from nonparametric Wilcoxon–Mann–Whitney test implemented in the Matlab statistical toolbox, the discriminant metabolites derived from the correlation coefficients of a crossvalidated OPLS-DA model were validated at a univariate level. False discovery rate (FDR), a statistical approach to the problem of multiple comparisons, was used in this study to verify the discriminant metabolites chosen by MW-p values (<0.05).

Receiver operating characteristic curve analysis and prediction models
Using the results obtained from the UPLC-QTOFMS and OPLS-DA analysis, we conducted receiver operating characteristic (ROC) curve analysis by PASW Statistics 18 (SPSS) to evaluate the predictive power of each of the discriminant metabolites. The cutpoint was determined for each biomarker by searching for those that yielded both high sensitivity and specificity. ROC curves were then plotted on the basis of the set of optimal sensitivity and specificity values. Area under the curve (AUC) was computed via numerical integration of the ROC curves. The metabolite signature that has the largest area under the ROC curve was identified as having the strongest predictive power for detecting OSCC.

A conventional logistic regression (LR) prediction model was developed to determine the best combination of salivary markers for cancer prediction. LR model was constructed using the binary outcome of the disease (OSCC) and healthy control (or OLK) as dependent variables and validated by using leave-one-out crossvalidation. The backward stepwise
regression was used to find the best final model. The leave-one-out cross validation is to predict the property value for a compound from the data set, which is in turn predicted from the regression equation calculated from the data for all other compounds. The backward stepwise LR method, available in both binary and multinomial regression in SPSS, determines automatically which variables to add or drop from the model. The marker values for the case that was left out were used to compute a predicted class for that observation. The crossvalidation error rate is then the number of samples predicted incorrectly divided by the number of samples. ROC curves for the logistic model were plotted with the fitted probabilities from the model as possible cut-points for computation of sensitivity and specificity.

Results

Salivary metabolomics analysis

Typical UPLC-QTOFMS chromatograms of saliva samples from the healthy control group, the OLK group and the OSCC group are shown in Figure 1a, where marked variations can be observed among the three salivary chromatograms. After data normalization, PCA was performed on the

Figure 1. (a) Typical UPLC-QTOFMS positive mode base peak intensity (BPI) chromatograms of OSCC, OLK and healthy control saliva samples (18-min separation on a 2.1 × 100 mm² 1.7-µm ACQUITY BEH C₁₈ UPLC column). (b) The scores plot of the OPLS-DA model of the UPLC-QTOF-MS (positive mode) spectral data from the OSCC group, OLK group and the healthy control group. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
Early Detection and Diagnosis

dataset, which showed a trend of intergroup separation on the scores plot (figures not provided). A supervised method,
OPLS-DA, was applied in the data analysis and the scores
plot showed three clusters (OSCC, OLK and healthy controls)
separating from each other (Fig. 1b).

The data were interrogated using an OPLS-DA model
which established intergroup separation between OSCC
group and healthy controls (Fig. 2a, left panel). An S-plot
model was used to select the significant metabolites that were
differentially expressed in OSCC group relative to the control
group (Fig. 2a, right panel). A total of 41 discriminant
metabolites were measured in OSCC relative to the control
group, 61 discriminant metabolites measured in OSCC rela-
tive to OLK, and 27 measured in OLK relative to the control
group. Table 1 provided a list of identified discriminant
metabolites, we used a smaller sample set with age (t-
test p>0.19) or gender (t-test p=0.28) matched OSCC
patients and healthy controls and constructed a new OPLS-
DA model. An S-plot model was used to select the significant

Receiver operating characteristic curve analysis

The detailed statistics of the area under the ROC curves
(AUC), and the corresponding sensitivities and specificities
for each of the five salivary metabolites for OSCC prediction
from healthy control are listed in Table 2. The calculated
area under the ROC curve for lactic acid and valine was 0.80
and 0.81, respectively. Lactic acid was shown at a signifi-
cant coefficient value while valine and phenylalanine
have negative coefficient values.

Discussion

The OPLS-DA models derived from our current UPLC-
QTOFMS metabolic analysis demonstrated good separation
among OSCC, OLK and healthy control; highlighting the
diagnostic potential of this non-invasive analytical approach.
Discriminant metabolites were identified in OSCC relative to
OLK by comparison the accurate mass of molecular weight (<2 ppm) with reference standards avail-
able (Table 1 and Fig. 2). However, there are significant dif-
fferences of age distribution and genders between OSCC and
OLK patients and between OSCC patients and healthy con-
trols, respectively. This may produce false positive results in
metabolite marker selection. To evaluate the impact of age
and gender imbalance on the identified discriminant metabo-
lites, we used a smaller sample set with age (n = 27, t-test
p=0.19) or gender (n = 27, t-test p=0.28) matched OSCC
patients and healthy controls and constructed a new OPLS-
DA model. An S-plot model was used to select the significant

Concentration in saliva decreased the probability that the
sample was obtained from an OSCC subject. The leave-one-
out crossvalidation error rate based on LR models was 18.9%
(7 of 37). The ROC curve was computed for the LR model.
Using a cutoff probability of 50%, we obtained a sensitivity
of 86.5% and a specificity of 82.4%. The positive predictive
value (PPV) was 81.6%. The calculated area under the ROC
curve was 0.89 (95% confidence intervals, 0.813–0.972) for
the LR model (Fig. 3a). Lactic acid was shown at a signifi-
cantly higher level in OSCC (2.26-fold to healthy control),
while valine (1.80-fold to healthy control) was lower in
OSCC (Fig. 3c). These were in accordance with the result
obtained from the LR model.

Analogously, we constructed the same LR model based on
the five validated biomarkers and consequently, lactic acid,
valine and phenylalanine, in combination, provided the best
prediction (Table 3). The coefficient value are positive for
lactic acid, indicating that the rise in its concentration in sa-
lica increased the probability that the sample was obtained
from an OSCC subject, while for valine and phenylalanine,
the coefficient value was negative, indicating that the rise in
their concentration in saliva decreased the probability that
the sample was obtained from an OSCC subject. The leave-one-
out crossvalidation error rate based on LR models was 13.5%
(5 of 37). The ROC curve was computed for the LR
model. Using a cutoff probability of 50%, we obtained a sen-
tivity of 94.6% and a specificity of 84.4%. The PPV was
87.5%. The calculated area under the ROC curve was 0.97
(95% confidence intervals, 0.932–1.000) for the LR model
(Fig. 3b). Lactic acid level was significantly elevated in OSCC
(2.97-fold to OLK), while valine (1.60-fold to OLK) and
phenylalanine (2.33-fold to OLK) were shown at lower levels
in OSCC (Fig. 3d). These findings were consistent with the
results obtained from the LR model, in which lactic acid has
a positive coefficient value while valine and phenylalanine
have negative coefficient values.
Figure 2. OPLS-DA scores plots and S-plots of metabonomic comparison between (a) OSCC group and the healthy control group, two-component model ($R^2_X = 0.149, R^2_Y = 0.818, Q^2_{(cum)} = 0.534$); (b) OSCC group and the OLK group, three-component model ($R^2_X = 0.253, R^2_Y = 0.908, Q^2_{(cum)} = 0.466$); and (c) OLK group and the healthy control group, three-component model ($R^2_X = 0.23, R^2_Y = 0.914, Q^2_{(cum)} = 0.721$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
metabolites that were differentially expressed in OSCC group relative to the control group. We obtained a very similar number of total discriminants in which the top 15 discriminant metabolites are the same as those identified using the whole sample set, see Supporting Information Tables 2–5. Therefore, we believe that the impact of age and gender on the identified discriminant metabolites in the diseased group is insignificant. In the same way, a smaller sample set with age (n = 25, t-test p =0.31) or gender (n = 25, t-test p =0.58) matched OSCC and OLK patients was selected and a new OPLS-DA and S-plot model were constructed. A very similar number of total discriminants were obtained and the top ten discriminant metabolites are the same as those identified using the whole sample set (Supporting Information Tables 6–9). Therefore, we believe that the impact of age and gender on the identified discriminant metabolites contributing for the separation of OSCC from OLK is insignificant.

The identified discriminant metabolites with MW-\(p\) values lower than 0.05 were validated with a FDR (threshold set at 0.05), which means there is a 5% false rate in the discovered

Table 2. Receiver operating characteristic (ROC) curve analysis of OSCC-associated salivary metabolomic signatures

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>OSCC–Healthy control</th>
<th>OSCC–OLK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC (95% CIs)</td>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>0.80 (0.700–0.904)</td>
<td>73.0</td>
</tr>
<tr>
<td>(\gamma)-Aminobutyric acid</td>
<td>0.56 (0.423–0.698)</td>
<td>61.8</td>
</tr>
<tr>
<td>Valine</td>
<td>0.81 (0.706–0.911)</td>
<td>82.4</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0.64 (0.508–0.765)</td>
<td>52.9</td>
</tr>
<tr>
<td>(n)-Eicosadienoic acid</td>
<td>0.67 (0.549–0.800)</td>
<td>51.4</td>
</tr>
</tbody>
</table>
discriminant metabolites. From Table 1, we can see that the number of discriminant metabolites in distinguishing OSCC from control is less than those between OLK and control. The two panels of discriminant metabolites identified from OSCC relative to OLK and OSCC relative to healthy control overlapped on five metabolites, which share the same direction of perturbation. The salivary levels of GABA, phenylalanine and valine were significantly lower, while n-eicosanoic acid and lactic acid were significantly higher, in the OSCC group than those in OLK group and the control group. Of these five metabolites, valine, lactic acid and phenylalanine were also identified as discriminatory serum metabolites in OSCC relative to healthy subjects.24 However, the levels of phenylalanine, valine and lactate changed differently (in opposite direction) is probably due to the different metabolic rates (for maintaining a homeostatic state) in the different biological compartments (serum and saliva). Each metabolite will have a unique abundance level in a given compartment, which alters differently in response to pathophysiological stimuli. An example is that the levels of phenylalanine and valine are reported to be about 233 and 57 µM in blood, but only 10 and 10 µM in saliva, respectively, in healthy adults older than 18 years old (www.hmdb.ca). We also found a depleted level of tryptophan in serum but an elevated level in urine of the same group of colorectal cancer patients in one of our recently published metabolomic studies.19,35 The reason that these two panels of markers shared five identical metabolites with the same direction of perturbation is presumably due to the similar metabolic alterations associated with the development of cancer either from a healthy state or an OLK state.

Lactic acid, an end product of glycolysis, was observed at a higher level in OSCC. Increased glycolysis is associated with many tumors or cancer cells, even in the presence of oxygen, which is known as the Warburg effect.36 Tumor cells are in a hypoxic condition and tend to utilize energy via glycolysis, and as a result, produce a higher level of lactate in body fluids, as evidenced by a higher level of lactic acid observed in many types of tumors.37 Increased lactic acid is associated with the decreased pyruvate entering into tricarboxylic acid (TCA) cycle. The impaired TCA production of energy due to the insufficient pyruvate supply is therefore supplemented by adjacent metabolic pathways with several intermediates infused into the TCA cycle. For example, the intermediates of branched chain amino acids (BCAAs) become involved in the TCA cycle when there is a shortage in energy supply.28 The catabolism of BCAAs, valine, leucine and isoleucine, using the same transaminase in the first two steps, would produce three α-keto acids which are further oxidized using a common branched-chain α-keto acid dehydrogenase (BCKD), yielding three CoA derivatives. Subsequently, the metabolic pathway diverge, producing many intermediates to be consumed in TCA cycle.16,39 We observed that valine, leucine and isoleucine were all at relatively decreased levels in saliva (an extracellular fluid) of OSCC, presumably due to the increased metabolic utilization by the TCA cycle in cancer cells.40 The decreased level of valine in saliva of OSCC was of statistical significance, while the change of leucine and isoleucine was not significant (p > 0.05). Decreased levels of BCAAs have been observed in patients with advanced tumors,16,41 which may be due to cancer cachexia and enhanced protein synthesis in tumor tissue.42 Additionally, the decreased level of BCAA may also be associated with increased glycolysis during the cell proliferation in cancer tissues.16

Similar to the increased metabolic utilization of BCAA intermediates by TCA, glutamine is degraded to glutamate and other TCA intermediates to reinforce the energy production through TCA cycle, a process known as glutaminolysis. Here we conceive that, along with the decrease of α-ketoglutarate, one of the TCA intermediates, more glutamate is transformed to 2-oxo-glutarate, leading to an increased utilization of GABA and thus decreased level of GABA in extracellular fluid such as saliva. Previous studies also suggest that GABA should be regarded as an important marker, such as a tumor signaling molecule in the periphery to control the proliferation of tumor cells,43 and a potential tumor suppressor for small airway-derived lung adenocarcinoma44 and colon cancer cell migration.45 Thus, decreased GABA level in the saliva of OSCC patients may be associated with the occurrence of OSCC. Our observation of a decreased salivary level of phenylalanine in the OSCC group is in accordance with the serum metabolomics analysis of human colorectal cancer19 but differs from other serum and tissue studies, which observed an increased level of phenylalanine.46,47 The increased level of n-eicosanoic acid may also be associated

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>OSCC–Control</th>
<th>OSCC–OLK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactic acid</td>
<td>0.1105</td>
<td>0.1232</td>
</tr>
<tr>
<td>Valine</td>
<td>−0.0462</td>
<td>0.0138</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0.0462</td>
<td>0.0176</td>
</tr>
<tr>
<td>Constant</td>
<td>0.9231</td>
<td>5.3701</td>
</tr>
</tbody>
</table>

The results of LR model were shown that lactic acid and valine are the best combination of salivary biomarkers for the discrimination of OSCC from healthy control, while lactic acid, valine and phenylalanine are the best combination of salivary biomarkers for the discrimination of OSCC from OLK.
with the occurrence of OSCC, as the risk of oral cancer was reported to be correlated with the ingestion of saturated fatty acids. 48

OSCC is a complex disease, resulting from an interdependent series of biochemical alterations, rather than a single disruptive event. Therefore, a panel of several metabolome markers will improve the sensitivity and specificity for OSCC detection. From the result of LR analysis (Tables 2 and 3), it showed that lactic acid and valine are the best predictors for distinguishing OSCC from healthy control, and lactic acid, valine and phenylalanine for OSCC from OLK.

It is unlikely that all cancer of a particular organ can be detected when using a single biomarker with high specificity and sensitivity because of the multifactorial nature of oncogenesis and the heterogeneity in oncogenic pathways. Therefore, multiple statistical strategies were used for our prediction model to identify combinations of biomarkers that can identify OSCC patients among all the samples. Although promising, the sensitivity (86.5%) and specificity (82.4%) cannot meet the demands for being a clinical tool for disease screening. Efforts should be made to validate and investigate other candidate biomarkers and to combine them to generate a higher power for oral cancer discrimination and prediction.

Three candidate biomarkers were identified using the salivary metabolomics approach. However, the discovered candidate biomarkers need to be extensively validated before they can be translated into real world diagnostic and screening application. If appropriately validated in large patient cohorts, the discovered candidates will be measured and verified with multiple complementary analytical technologies (e.g., a combined LC-MS and GC-MS approach). At present, OSCC is not detected until it reaches an advanced stage, which would generally result in a poor prognosis and survival rate. Therefore, early detection of OSCC as well as the screening of high risk populations with precancerous lesions remains to be an unmet need. The integration of various types of biomarkers including salivary metabolite signatures, coupled with a conventional oral examination (e.g., a scalpel biopsy) may become an applicable strategy for early detection of oral precancerous lesions and cancer.

References

15. Dr. David T. Wong and the UCLA Dental Research Institute. Saliva-Based Translational Research & Clinical Applications (StarCA), Los Angeles, 2007.
43. Young SZ, Bordey A. GABA’s control of stem and cancer cell proliferation in adult neural and peripheral niches. Physiology 2009;24:171–85.