Urine Metabolite Profiling Offers Potential Early Diagnosis of Oral Cancer

By: Guo X. Xie, Tian L. Chen, Yun P. Qiu, Peng Shi, Xiao J. Zheng, Ming M. Su, Ai H. Zhao, Zeng T. Zhou, and Wei Jia

Abstract:

Oral cancer is the sixth most common human cancer, with a high morbidity rate and an overall 5-year survival rate of less than 50%. It is often not diagnosed until it has reached an advanced stage. Therefore, an early diagnostic and stratification strategy is of great importance for oral cancer. In the current study, urine samples of patients with oral squamous cell carcinoma (OSCC, n = 37), oral leukoplakia (OLK, n = 32) and healthy subjects (n = 34) were analyzed by gas chromatography-mass spectrometry (GC–MS). Using multivariate statistical analysis, the urinary metabolite profiles of OSCC, OLK and healthy control samples can be clearly discriminated and a panel of differentially expressed metabolites was obtained. Metabolites, valine and 6-hydroxynicotic acid, in combination yielded an accuracy of 98.9%, sensitivity of 94.4%, specificity of 91.4%, and positive predictive value of 91.9% in distinguishing OSCC from the controls. The combination of three differential metabolites, 6-hydroxynicotic acid, cysteine, and tyrosine, was able to discriminate between OSCC and OLK with an accuracy of 92.7%, sensitivity of 85.0%, specificity of 89.7%, and positive predictive value of 91.9%. This study demonstrated that the metabolite markers derived from this urinary metabolite profiling approach may hold promise as a diagnostic tool for early stage OSCC and its differentiation from other oral conditions.

Article:

INTRODUCTION

Oral cancer is the sixth most common human cancer with a high morbidity rate and an overall 5-year survival rate of less than 50% (Epstein et al. 2002; Mao et al. 2004). Over 90% of oral cancers are oral squamous cell carcinoma (OSCC) which arise from the oral mucosal lining (Scully and Felix 2006). Reports indicate an increasing worldwide incidence of oral cancer at an earlier age (Kantola et al. 2000; Myers et al. 2000). A critical factor in the lack of prognostic improvement is the fact that a significant proportion of cancers are initially asymptomatic lesions and are not diagnosed or treated until they reach an advanced stage. In OSCC, if the cancer is detected at T1 (T means the tumor size invasion level) stage, the 5 year survival rate is over 80%, compared to 20–40% if diagnosed at later stages (T3 and T4) (Wong 2006). Oral leukoplakia (OLK) is the most frequently occurring oral precancerous lesion and it has considerable

malignant transformation in clinic (ranging from 1.58 to 27.27%) (Xu et al. 1992). Early detection and screening of high risk populations of OSCC and OLK are 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 (Wong 2006).

It is now generally agreed that OSCC and precancerous lesions involve not only specially expressed genes and proteins, but also changes in the concentration of endogenous metabolites (Griffin and Shockcor 2004; Sugimoto et al. 2010; Tiziani et al. 2009). The metabolic profile of biofluid (urine, serum, and saliva) can be altered by a variety of physiological processes following pathophysiological stimuli. Therefore, a global perturbation in one or more of these profiles may demonstrate the presence of a particular disease (Dunn et al. *2005*; Qiu et al. 2009). There is an urgent need to discover more biomarkers for diagnosis, prognosis, therapeutic response prediction and population screening of human cancers, which can hopefully improve treatment and reduce cancer mortality. Furthermore, an ideal cancer screening method should be accurate, non-invasive, and low-cost. Metabolomics or metabonomics (Sugimoto et al. 2010; Tiziani et al. 2009; Yan et al. 2008; Zhou et al. 2009) 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 pathophysiologic stimuli or genetic modification. Recent urinary metabonomic analysis demonstrates its applicability for the diagnosis and prognosis of disease (Kind et al. 2007; Qiu et al. 2010; Wu et al. 2009a).

Metabolomic analysis of OSCC patient serum samples (Tiziani et al. 2009) was 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, alanine, tyrosine and serine were identified. Salivary (Yan et al. 2008) metabonomics analysis of OSCC and OLK was also performed by high performance liquid chromatography–mass spectrometry (HPLC–MS) and a total of 14 OSCC-related and 11 OLK-related metabolites were discovered; although, these differential metabolites were not identified. Recently, an UPLC-QTOFMS based salivary metabolomics approach has been successfully demonstrated as a complementary tool in the early detection and diagnosis of oral cancer by our group (Wei et al. 2010). We get an accuracy of 89.3% in distinguishing OSCC from healthy control with a sensitivity of 86.5% and specificity of 82.4% using a combination of two metabolites. In the current work, we investigated the ability of gas chromatography-mass spectrometry (GC–MS) based urinary metabonomics study to completely separate patients with OSCC from the OLK and healthy controls and to obtain data-rich information for oral cancer diagnosing. The purpose of this study is to identify urinary metabolites as potential biomarkers to diagnose and stratify OSCC and precancerous lesions. These biomarkers can be used as a complementary diagnosis method by using saliva and blood.

MATERIALS AND METHODS
Patient selection
Urine was collected from a group of 37 OSCC patients (26 males and 11 females, 9 of stage I, 12 of stage II, 6 of stage III, 10 of stage IV), whose mean age was 56 ± 11 years (34–77), 32 OLK

patients (13 males and 19 females), whose mean age was 60 ± 13 years (34–80). They were compared to a control group of 34 healthy individuals (13 males and 21 females), whose mean age was 43 ± 14 years (21–73). Clinical information on participants was provided in Supplementary 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 was based in all cases 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 the Ethical Committee approved consent form agreeing to serve as urine donors for the experiments.

Table 1: Summary of the differential metabolites from VIP values of two component OPLS-DA models accountable for variations among OSCC, OLK and control groups

RT (min)	Metabolites[a]	OSCC-Ccontrol					OLK-control					OCSS-OLK				
		Rank	VIP[b]	P^c	FDR[d]	FC[e]	Rank	VIP[b]	P^c	FDR[d]	FC[e]	Rank	VIP[b]	P^c	FDR[d]	FC[e]
8.69	Alanine	(3)	2.01	6.19E-05	1.40E-3	1.09										
9.31	Hexanedioic acid						(1)	2.2	2.43E-05	6.67E-4	1.94					
10.48	3-Methylglutakonic acid						(4)	2.06	1.79E-04	1.33E-3	1.95					
10.71	Valine	(4)	1.94	1.29E-05	1.00E-3	1.39										
12.10	Leucine											(5)	2.01	5.79E-04	1.70E-3	1.51
12.41	Serine	(7)	1.78	3.15E-04	2.20E-3	1.53										
13.35	Cysteine											(1)	2.19	2.45E-05	2.00E-4	1.74
13.54	6-Hydroxynicotinic acid	(1)	3.1	2.87E-11	1.00E-4	0.17						(2)	2.15	9.36E-04	2.1E-3	0.30
19.23	Hippurate	(2)	2.57	3.85E-07	3.00E-4	0.49										
20.75	Phenylalanine											(6)	1.96	5.06E-4	1.60E-3	1.37
27.21	Histidine						(3)	2.12	3.44E-04	2.17E-3	0.39					
30.22	Histamine											(7)	1.87	1.26E-03	2.40E-3	1.51
30.77	Glutamine,N2-phenlyalacetyl						(3)	2.12	1.82E-05	3.33E-4	0.36					
32.68	Tyrosine	(5)	1.92	1.43E-05	1.10E-3	1.65						(4)	2.1	2.54E-05	3.00E-4	1.7
34.49	Tryptophan						(5)	1.95	2.80E-03	5.00E-3	0.70	(3)	2.12	1.32E-06	1.00E-4	1.99
37.89	Cystine	(6)	1.88	9.17E-04	2.90E-3	3.22	(2)	2.18	6.29E-06	1.67E-4	3.62					

[a]Metabolites are identified using available library databases; [b]Variable importance in the projection (VIP) was obtained from OPLS-DA with a threshold of 1.0. [c] P-value and fold change (FC) are calculated from nonparametric Wilcoxon–Mann–Whitney test (one-way ANOVA). [d] All the metabolites were discriminant (Wilcoxon $P < 0.05$), with an FDR of 5%. [e] FC with a value larger than 1.0 indicates a relatively higher concentration present in OSCC patients while a value lower than 1.0 means a relatively lower concentration as compared to the OLK patients and the 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

Urine sample collection and preparation

Urine samples (300 ml) were collected from participants between 9:00 and 10:00 a.m. and there was no diet or other restrictions in the sample collection. Urine samples were centrifuged at $15,400 \times g$ for 10 min and the resulting supernatants were immediately stored at $-80°C$ pending GC–MS analysis. Before analysis, the urine samples were processed according to our previous work (Qiu et al. 2007).

Urinary metabonomics analysis

The urine samples were derivatized with ethyl chloroformate (ECF) and analyzed with a PerkinElmer gas chromatography coupled with a TurboMass-Autosystem XL mass spectrometer (PerkinElmer Inc., USA) with minor modifications to our previous work (Qiu et al. 2007). Briefly, A typical 600 µl aliquot of supernatant of urine sample (300 µl urine + 300 µl of ultrapure water) spiked with internal standards (100 µl of L-2-chlorophenylalanine in water, 0.1 mg/ml) was prepared for ECF derivatization. After adding 400 µl of anhydrous ethanol, 100 µl of pyridine, and 50 µl of ECF to the urine sample, the derivatization was conducted by ultrasonication for 60 s. The extraction was carried out using 300 µl of chloroform, with the aqueous layer pH was carefully adjusted to 9–10 using 100 µl of NaOH (7 mol/l). The derivatization procedure was repeated with another 50 µl of ECF into the aforementioned products. After the overall mixtures were vortexed for 30 s and centrifuged for 10 min, the

aqueous layer was aspirated off, while the remaining chloroform layer containing derivatives was obtained and dried with anhydrous sodium sulfate for subsequent GC–MS analysis. A 1 μl extract aliquot of the extracts was injected onto a DB-5 MS capillary column coated with 5% diphenyl cross-linked 95% dimethylpolysiloxane (30 m × 250 μm i.d., 0.25 μm film thickness; Agilent J&W Scientific, Folsom, CA) in the split mode (3:1). Either the injection temperature or the interface temperature was set to 260°C and the ion source temperature was adjusted to 200°C. Initial GC oven temperature was 80°C; 2 min after injection, the GC oven temperature was raised to 140°C at a rate of 10°C/min, to 240°C at a rate of 4°C/min, then to 280°C at a rate of 10°C/min, and finally held at 280°C for 3 min. Helium was the carrier gas with a flow rate of 1 ml/min. The measurements were made with electron impact ionization (70 eV) in the full scan mode (m/z 30–550).

Data processing and statistical analysis

The data from the GC–MS analysis were converted to NetCDF format via the data analysis interface of the PE Instrument (PerkinElmer Inc., USA). Each file was extracted subsequently using custom scripts in the MATLAB 7.0 (The MathWorks, Inc., USA) for data pretreatment procedures such as baseline correction, de-noising, smoothing and alignment, time-window splitting, and peak deconvolution (based on multivariate curve resolution algorithm) (Ni et al. 2007). The exclusion of internal standard (IS) and solvent peaks, and normalization to a total chromatogram based on the total urine collected were carried out. The output data was organized in the form of an arbitrary peak index (retention time-m/z pairs), sample names (observations), and peak intensity information (variables). IS and any known artificial peaks, such as peaks caused by noise, column bleed and ethyl chloroformate derivatization procedure, were removed from the data set. Multivariate statistical analyses, including principal component analysis (PCA) and orthogonal partial least squares projection to latent structures-discriminant analysis (OPLS-DA) were performed by SIMCA-P 11.5 software (Umetrics, Umeå, Sweden) (Ni et al. 2007). Based on a variable importance in projection (VIP) threshold (VIP > 1.5) from the 7-fold cross-validated OPLS-DA model together with the V-plot (plot constructed with the VIP value versus p (corr) value of each metabolite), which calculated with VIP and p (corr) (Ni et al. 2008), a number of metabolites responsible for the differentiation of metabolic profiles of diseased individuals and healthy controls could be obtained. In parallel, the metabolites identified by the OPLS-DA model were validated at a univariate level using nonparametric Wilcoxon–Mann–Whitney test from the Matlab statistical toolbox with the critical P-value set to 0.05. False discovery rate (FDR) (Benjamini and Hochberg 1995; Storey 2002), 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). The corresponding fold change shows how these selected differential metabolites varied in the OSCC patients from those of healthy controls. Additionally, compound identification was performed by comparing the mass fragments with NIST 05 Standard mass spectral databases in NIST MS search 2.0 (NIST, Gaithersburg, MD) software with a similarity of more than 70% and finally verified by available reference compounds.

Receiver operating characteristic curve analysis and prediction models

Using the results obtained from the GC-TOFMS and OPLS-DA analysis, Receiver Operating Characteristic (ROC) curve analysis was further conducted by PASW Statistics 18 (SPSS Inc.) to evaluate the predictive power of each of the differential metabolites as we done with the saliva metabolomics (Wei et al. 2010). The ROC curve is the graph of points defined by sensitivity and

(1-specificity) and was produced with the whole GC–MS data. ROC curves were plotted on the basis of the set of optimal sensitivity and specificity values of a diagnostic test at different cutpoint. The cutpoint was determined for each biomarker by searching for those that yielded both high sensitivity and specificity. Area under the curve (AUC) was computed via numerical integration of the ROC curves. AUC is often presented along with its 95% confidence interval (CI). The 95% CI gives the range of values in which the true value lies and the associated degree of confidence. The area under the ROC curves (AUC) and its 95% confidence interval (CI) were obtained by nonparametric estimation. AUC is the value of the Wilcoxon–Mann–Whitney statistic, an estimate of the probability Pxy that the variable classify a randomly chosen sample to the right group. CI is the interval of confidence for the AUC value (Qin and Hotilovac 2008). 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 urinary markers for cancer prediction (Li et al. 2004). Logistic regression, aims to correctly predict the category of outcome for individual cases using the most parsimonious model which created using all predictor variables that are useful in predicting the response variable. Backward stepwise regression appears to be the preferred method of exploratory analyses, where the analysis begins with a full or saturated model and variables are eliminated from the model in an iterative process. The fit of the model is tested after the elimination of each variable to ensure that the model still adequately fits the data. When no more variables can be eliminated from the model, the analysis has been completed (http://userwww.sfsu.edu/~efc/classes/biol710/logistic/logisticreg.htm). In this study, LR model was constructed using the binary outcome of the disease (OSCC) and healthy control (or OLK) as dependent variables. The backward stepwise method 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 logistic regression 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 cross-validation 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 cutpoints for computation of sensitivity and specificity.

RESULTS AND DISCUSSION
Urinary metabonomics analysis
Typical GC–MS total ion current (TIC) chromatograms of urine samples from the OSCC group, OLK group and healthy control group are shown in Fig. 1a; obvious differences can be observed among the three urinary chromatograms. After data normalization, PCA was performed on the dataset, which showed a trend of inter-group separation on the scores plot (Figure not provided). The OPLS-DA model was applied to the data analysis and the scores plot showed three clusters (OSCC, OLK and healthy controls) separating from each other (Fig. 1b).
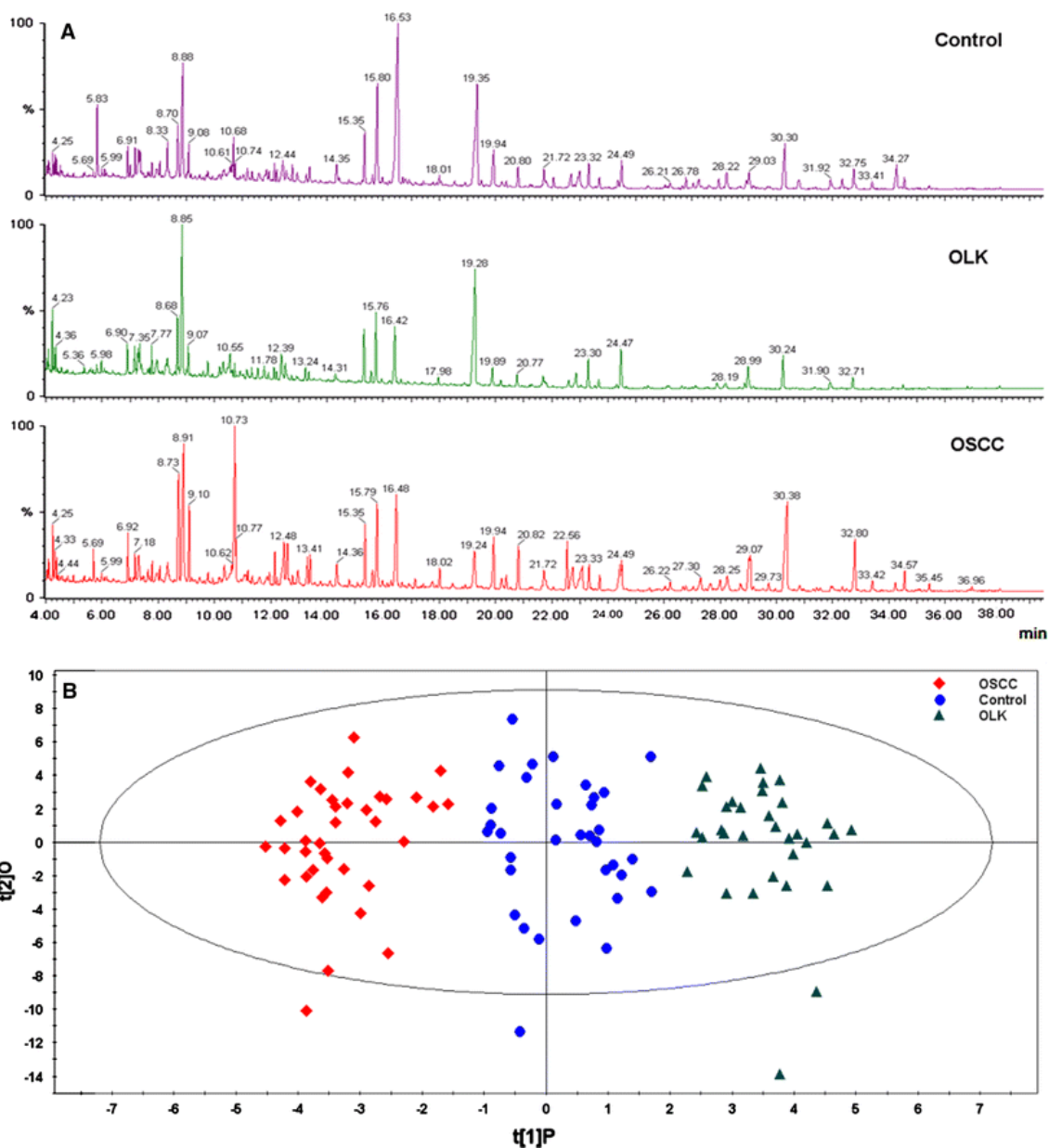
Figure 1: a Typical GC–MS total ion current (TIC) chromatograms of OSCC, OLK and healthy control urine samples; **b** the scores plot of the OPLS-DA model of the GC–MS spectral data from the OSCC group, OLK group and the healthy control group

The OPLS-DA model demonstrated satisfactory modeling and predictive abilities using one predictive component and two orthogonal components (R2X = 0.166, R2Y = 0.836, Q2(cum) = 0.706), achieving a distinct separation between the metabolite profiles of the OSCC group and healthy controls (Fig. 2a, left panel). A V–plot model was used to select the significant metabolites that were differentially produced between the healthy control group and OSCC group (Fig. 2b, right panel). A total of 65 differentially expressed metabolites were measured. The metabolites most strongly influencing the differentiation are listed in Table 1 with the VIP, a measure of their relative influence on the model, along with the fold change. All

metabolites listed in Table 1 were discriminant (Wilcoxon $P < 0.05$), with an FDR of 5%. The most significant differential metabolites were decreased urinary excretion of hippurate, and 6-hydroxynicotic acid, and increased urinary excretion of alanine, tyrosine, valine, serine and cystine (Fig. 2 and Table 1).
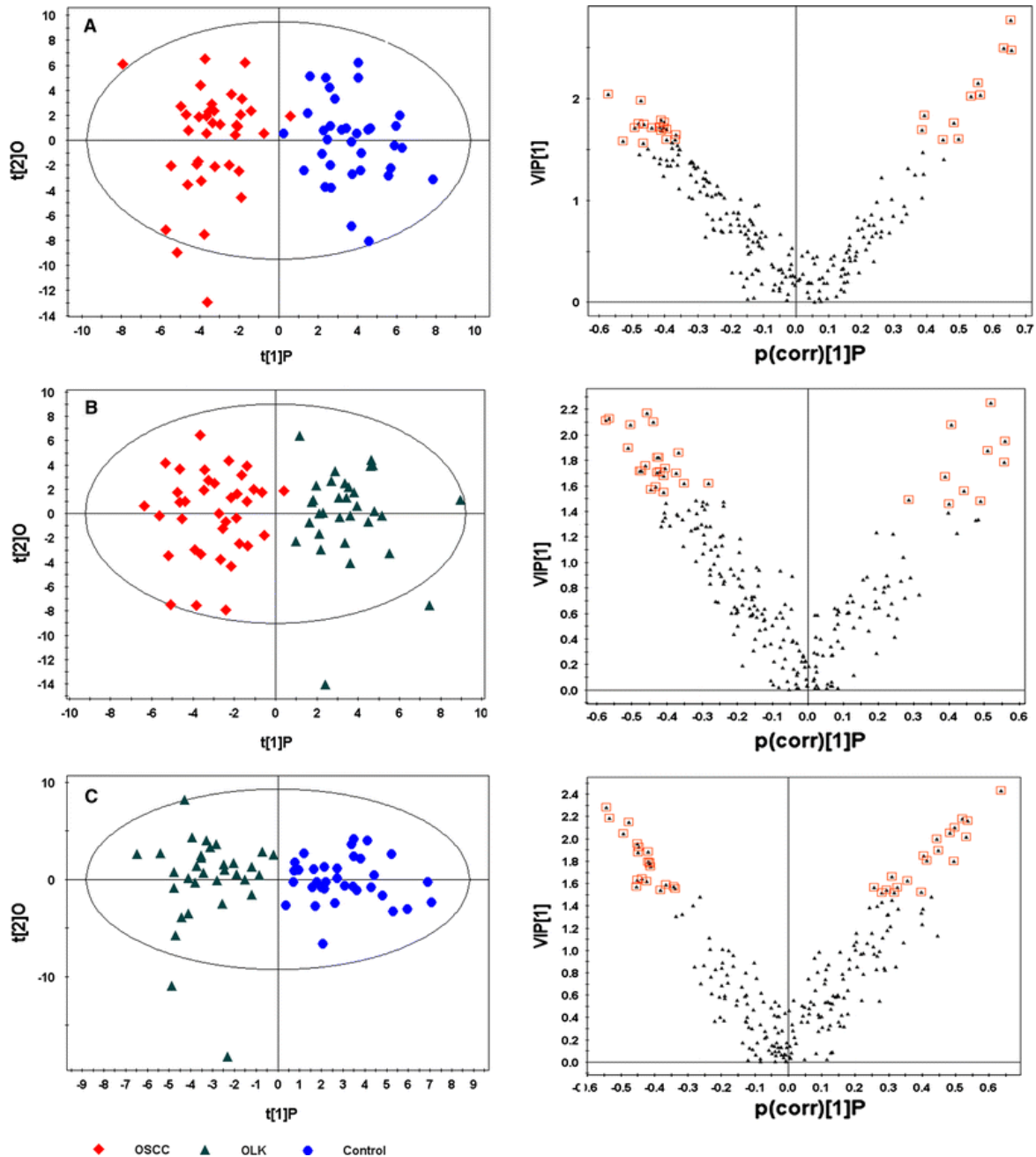

Figure 2: OPLS-DA scores plots and V-plots of metabonomic comparison between **a** OSCC group and the healthy control group, two-component model (R2X = 0.166, R2Y = 0.836, Q2(cum) = 0.706); **b** OSCC group and the OLK group, two-component model (R2X = 0.153, R2Y = 0.797, Q2(cum) = 0.581); and **c** OLK group and the healthy control group, two-component model (R2X = 0.154, R2Y = 0.785, Q2(cum) = 0.521)

The differential metabolites which accounted for the inter-group separation between the OSCC group and the OLK group, and between the OLK group and the healthy control group, were also identified by two two-component OPLS-DA model (R2X = 0.153, R2Y = 0.797, Q2(cum) = 0.581), and R2X = 0.154, R2Y = 0.785, Q2(cum) = 0.521), respectively. The most significantly altered metabolites are listed in Table 1 using a V-plot model.

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 total differential urinary metabolites for OSCC prediction from healthy control are listed in Table 2.

Table 2: Receiver operating characteristic (ROC) curve analysis of OSCC-associated urinary metabonomic signatures

Metabolites	OSCC-healthy control			OSCC-OLK		
	AUC (95% CIs)	Sensitivity (%)	Specificity (%)	AUC (95% CIs)	Sensitivity (%)	Specificity (%)
Alanine	0.777 (0.670, 0.883)	67.6	64.1			
Leucine				0.74 (0.622, 0.861)	72.7	61.7
Cystine	0.729 (0.605, 0.853)	72.2	68.6			
Valine	0.801 (0.697, 0.906)	79.4	73.0			
Serine	0.749 (0.634, 0.863)	66.7	63.6			
Cysteine				0.80 (0.690, 0.902)	78.8	69.4
6-Hydroxynicotic acid	0.959 (0.919, 1.000)	92.1	94.0	0.73 (0.604, 0.861)	68.9	78.0
Hippurate	0.851 (0.757, 0.944)	75.0	85.2			
Phenylalanine				0.74 (0.626, 0.862)	69.4	63.6
Histamine				0.73 (0.607, 0.845)	65.7	58.8
Tyrosine	0.80 (0.699, 0.900)	73.3	63.4	0.80 (0.691, 0.900)	78.6	63.4
Tryptophan				0.84 (0.743, 0.936)	82.4	74.3

To demonstrate the utility of urinary metabolites for the discrimination between OSCC and healthy control, a logistic regression model was built based on seven validated biomarkers. As a result, 6-hydroxynicotic acid and valine in combination provided the best prediction (Table 3). The coefficient value are positive for 6-hydroxynicotic acid (control to OSCC), indicating that the rise in its concentration in urine decreased the probability that the sample was obtained from an OSCC subject, while for valine, the coefficient value was negative (control to OSCC), indicating that the rise in its concentration in urine increased the probability that the sample was obtained from an OSCC subject. The leave-one-out cross-validation error rate based on logistic regression models was 2.8% (2 of 71). The ROC curve was computed for the logistic regression model. Using a cutoff probability of 50%, we obtained a sensitivity of 94.4% and a specificity of 91.4%. The positive predictive value (PPV) was 91.9%. The calculated area under the ROC curve was 0.99 (95% confidence intervals, 0.979, 1.000) for the logistic regression model (Fig. 3a). 6-Hydroxynicotic acid (5.66-fold to healthy control, decreased) was shown lower levels in OSCC, while valine (1.39-fold to healthy control, increased) was shown higher levels in OSCC (Fig. 3c, d).

Table 3: Urinary metabolite signatures selected by logistic regression model for the discrimination of OSCC from healthy control and OLK

Metabolite	OSCC-control			OSCC-OLK		
	Coefficient value	S.E.	*P* value	Coefficient value	S.E.	*P* value
Valine	−3.393	1.484	0.022			
Cysteine				−0.057	0.026	0.027
6-Hydroxynicotic acid	8.224	2.620	0.002	1.028	0.442	0.020
Tyrosine				−0.019	0.007	0.007
Constant	1.529	0.946	0.106	3.442	1.237	0.005

Note: The results of logistic regression model show that 6-hydroxynicotic acid and valine are the best combination of urinary biomarkers for the discrimination of OSCC from healthy controls, while 6-hydroxynicotic acid, cysteine, and tyrosine are the best combination of urinary biomarkers for the discrimination of OSCC from OLK
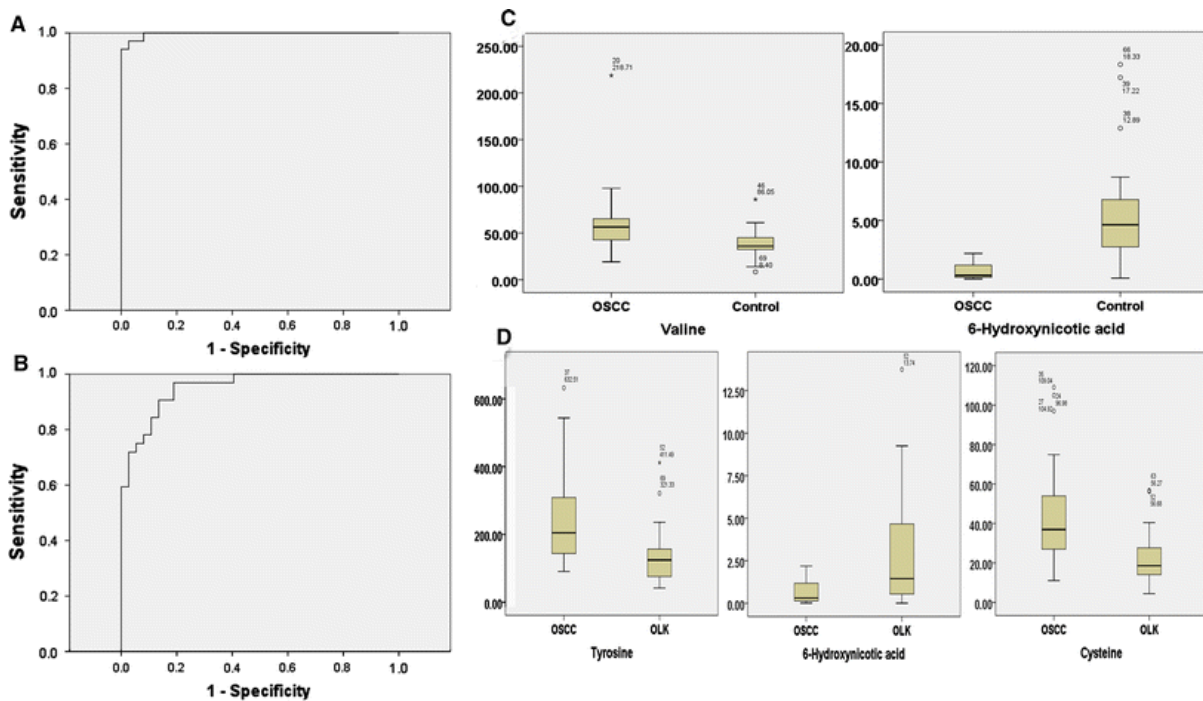

Figure 3: **a** ROC curve analysis for the predictive power of combined urinary biomarkers for distinguishing OSCC from healthy control. The final logistic model included two urinary biomarkers, 6-hydroxynicotic acid and valine. Using a cutoff probability of 50%, we obtained sensitivity of 94.4% and specificity of 91.4% by ROC. The calculated area under the ROC curve was 0.989 (95% confidence intervals, 0.979, 1.000); **b** ROC curve analysis for the predictive power of combined urinary biomarkers for distinguishing OSCC from OLK. The final logistic model included three urinary biomarkers, 6-hydroxynicotic acid, cysteine, and tyrosine. Using a cutoff probability of 50%, we obtained sensitivity of 85.0% and specificity of 89.7% by ROC. The calculated area under the ROC curve was 0.927 (95% confidence intervals, 0.869, 0.985); **c** Box plots of two differential metabolites in distinguishing OSCC from healthy control; **d** Box plots of three differential metabolites in distinguishing OSCC from OLK

Analogously, we constructed the same logistic regression model based on seven validated biomarkers including leucine, cysteine, 6-hydroxynicotic acid, phenylalanine, histamine, tyrosine, and tryptophan for the discrimination between OSCC and OLK. The combination of

cysteine, 6-hydroxynicotic acid, and tyrosine provided the best prediction (Table 3). The coefficient value is positive for 6-hydroxynicotic acid (OLK to OSCC), indicating that the rise in its concentration in urine decreased the probability that the sample was obtained from an OSCC subject, while for cysteine and tyrosine, the coefficient value was negative (OLK to OSCC), indicating that the rise in their concentration in urine increased the probability that the sample was obtained from an OSCC subject. The leave-one-out cross-validation error rate based on logistic regression models was 4.3% (3 of 69). The ROC curve was computed for the logistic regression model. Using a cutoff probability of 50%, we obtained a sensitivity of 85.0% and a specificity of 89.7%. The PPV was 91.9%. The calculated area under the ROC curve was 0.93 (95% confidence intervals, 0.869, 0.985) for the logistic regression model (Fig. 3b). 6-Hydroxynicotic acid (3.33-fold to OLK, downregulated) was shown lower level in OSCC, while cysteine (1.74-fold to OLK, upregulated) and tyrosine (1.70-fold to OLK, upregulated) were shown higher levels in OSCC (Fig. 3c, d).

Metabolite annotation and biochemical hypothesis

The OPLS-DA models derived from our current GC–MS metabonomics analysis demonstrated good separation among OSCC and OLK patients and healthy controls, highlighting the diagnostic potential of this non-invasive analytical approach. Differential metabolites were identified in OSCC relative to OLK or the control group by comparing the mass fragments with NIST 05 standard mass spectral databases and available reference compounds (Fig. 2 and Table 1). However, there are significant differences of age distribution and genders between OSCC and OLK patients and between OSCC patients and healthy controls, respectively. This may produce false positive results in metabolite marker selection. According the method used to evaluate the impact of age and gender imbalance on the identified discriminant metabolites in our previous saliva study (Wei et al. 2010), a smaller sample set with age or gender matched OSCC patients and healthy controls was used and constructed a new OPLS-DA model. A V–plot model was used to select the significant metabolites that were differentially expressed in OSCC group relative to the control group. As a result, a very similar number of total discriminants and the top 10 discriminant metabolites are the same as those identified using the whole sample set (data was not provided). Therefore, we believe that the impact of age and gender on the identified discriminant metabolites in the diseased group is insignificant. The impact of age and gender on the identified discriminant metabolites contributing for the separation of OSCC from OLK was studied in the same way.

The corresponding V-plot (Fig. 2, right panel) indicated that the differential metabolites contributing to the separation of the metabolic patterns of the control group and OSCC group are decreased urinary levels of hippurate and 6-hydroxynicotic acid, and increased levels of alanine, serine, cystine, tyrosine, and valine. However, of these seven metabolites, valine, tyrosine and serine were shown lower levels in serum metabolites (Tiziani et al. 2009) and valine was shown lower level in saliva metabolites (Wei et al. 2010) in OSCC, relative to healthy subjects. The levels of valine, tyrosine and serine in urine and serum and levels of valine in urine and saliva changed differently (in opposite direction) in OSCC are probably due to the different metabolic rates (for maintaining a homeostatic state) in the different biological compartments (serum, urine, 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 level of valine and alanine are reported to be about 233 and 333 μM in blood, but only 10 and 10 μM in

saliva, and 41 ± 1 and 4.23 ± 0.64 μM in urine, 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 metabonomic studies (Qiu et al. 2009; Qiu et al. 2010). A list of the most significant differential metabolites ($P < 0.05$) in the OSCC group and in the OLK group (compared to the control group) and between the OSCC group and the OLK group is summarized in Table 1. 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 discriminant metabolites.

The branched chain amino acid (BCAA), valine, was shown higher level in OSCC patients relative to the healthy controls. It is worth noting that increased valine level was also found in hepatocellular carcinoma, due to an increased glycolysis (Chan et al. 2009), and a direct inhibitory effect on tumor growth was confirmed after the depletion of valine (Ishigure et al. 2001). Leucine, the other BCAA, was also significantly higher in the OSCC group compared to the OLK group. Valine and leucine share the same enzyme systems (branched-chain alpha-ketoacid dehydrogenase for their initial degradative steps), which are considered as a group in terms of their roles in amino acid homeostasis (Wu et al. 2009b). Another BCAA, isoleucine, though was also higher in the OSCC group compared to the healthy control group but with no statistical significance. In this study, an increased level of alanine and serine was also found in oral cancer samples, suggesting involvement of glycolysis (Wu et al. 2009b). The level of tyrosine was higher in the OSCC than that in the control group in our study. Tyrosine can be synthesized in the body from phenylalanine and the conversion of phenylalanine to tyrosine is catalyzed by the enzyme phenylalanine hydroxylase. As reported, levels of most amino acids and their primary derivatives were significantly higher in tumors than in normal colon and stomach tissues (Hirayama et al. 2009) and all free amino acids are essential metabolic substrates for tumor cells. Solid tumors exhibit relatively specific amino acid dependency that functionally regulates their survival, proliferation, and metastasis. Tyrosine/phenylalanine restriction also inhibits invasion and metastasis of other cancers (Pelayo et al. 2001). Specific amino acid dependency is one of the metabolic abnormalities of cancer cells and can be regarded as the metabolic basis for their malignant behavior (Wu et al. 2009b).

It was observed that patients with OSCC have altered urinary excretion levels of hippurate, which is believed to be the product of the gut microflora co-metabolization (Claus et al. 2008). This suggests that there is a significant involvement of gut flora in the disease-induced metabolic alteration. Hippurate had antitumor effects both in vitro and in vivo by killing various tumor strains (Spustova and Oravec 1989) and the decreased levels of hippurate in the urine of OSCC patients may be evidence on the occurrence of OSCC. Cystine is an essential precursor for the biosynthesis of glutathione, a major redox regulatory molecule that protects cells from endogenously produced reactive oxygen species. Inhibition of cystine uptake disrupts the growth of primary brain tumors (Chung et al. 2005) and increased cystine uptake promotes the malignant progression of Nb2 lymphoma cells (Gout et al. 1997). Cystine/cysteine redox cycle is a discrete major regulator of cell survival (Banjac et al. 2008) and as proposed by Ye and Sontheimer (Ye and Sontheimer 1999), in the central nervous system, deregulated activity of the cystine/cysteine cycle, as observed in human malignant glioma cells, leads to a detrimental increase in secreted glutamate associated with increased glutamate toxicity, neuronal death and

seizures. The elevation of cysteine and tyrosine were consistent with enhanced glycolysis and effects on the TCA cycle (Wu et al. 2010). The reason that decreased level of 6-hydroxynicotic acid in OSCC remains to be unknown.

OSCC is a complex disease, resulting from an interdependent series of biochemical alterations, rather than a single disruptive event. Therefore, a panel of several metabolite markers will improve the sensitivity and specificity for OSCC detection. A logistic regression model was applied to evaluate the predictive power of each of the differential metabolite and then get the optimal combination of metabolites that can be used for cancer prediction. From the result of logistic regression analysis (Table 3), it showed that 6-hydroxynicotic acid and valine are the best predictors for distinguishing OSCC from healthy control, and 6-hydroxynicotic acid, cysteine, and tyrosine for OSCC from OLK. The PPV for OSCC to healthy control was 91.9, and 91.9% for OSCC to OLK, which indicates that the above metabolite biomarkers may be promising for screening. However, although promising in distinguishing OSCC from the controls, the sensitivity (94.4%) and specificity (91.4%) cannot meet the demands for being a clinical tool for disease screening. Efforts should be made to investigate and validate other candidate biomarkers and to combine them to generate a higher power for oral cancer discrimination and prediction. The discovered candidate biomarkers need to be extensively validated before they can be translated into real world diagnostic and screen application.

CONCLUDING REMARKS
In summary, the urine metabolic profiling approach identified distinct metabolic signatures of OSCC, OLK and healthy controls, which are characterized by a number of differentially expressed urine metabolites. The study demonstrates that this sufficiently robust and non-invasive profiling approach can be a promising screening tool for the early diagnosis of oral cancer. The results of this study also highlight the applicability of urinary metabolite markers that can be used as a stratification tool in the diagnosis of different oral conditions, complementary to the existing clinical procedures.

ABBREVIATIONS
OSCC	Oral squamous cell carcinoma
OLK	Oral leukoplakia
GC–MS	Gas chromatography–mass spectrometry
PCA	Principal component analysis
OPLS-DA	Orthogonal partial least squares-discriminant analysis
VIP	Variable importance in the projection
ROC	Receiver operating characteristic
LR	Logistic regression
FDR	False discovery rate
R2X	Fraction of sum of squares (SS) of X explained by each component
R2Y	Fraction of sum of squares (SS) of Y explained by each component
Q2cum	The cumulative Q2 for the extracted components
V-plot	Plot constructed with the VIP value versus p(corr) value of each metabolite
p(corr)	P scaled as correlation coefficient between X and T
t[2]O	Score of the orthogonal component
t[1]P	Score of the first non-orthogonal component

REFERENCES

Banjac, A., Perisic, T., Sato, H., et al. (2008). The cystine/cysteine cycle: a redox cycle regulating susceptibility versus resistance to cell death. *Oncogene, 27*(11), 1618–1628.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate—a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological, 57*(1), 289–300.

Chan, E. C. Y., Koh, P. K., Mal, M., et al. (2009). Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *Journal of Proteome Research, 8*(1), 352–361.

Chung, W. J., Lyons, S. A., Nelson, G. M., et al. (2005). Inhibition of cystine uptake disrupts the growth of primary brain tumors. *Journal of Neuroscience, 25*(31), 7101–7110.

Claus, S. P., Tsang, T. M., Wang, Y. L., et al. (2008). Systemic multicompartmental effects of the gut microbiome on mouse metabolic phenotypes. *Molecular Systems Biology, 4*, 219.

Dunn, W. B., Bailey, N. J., & Johnson, H. E. (2005). Measuring the metabolome: Current analytical technologies. *Analyst, 130*(5), 606–625.

Epstein, J. B., Zhang, L., & Rosin, M. (2002). Advances in the diagnosis of oral premalignant and malignant lesions. *Journal of the Canadian Dental Association, 68*, 617–621.

Gout, P. W., Kang, Y. J., Buckley, D. J., Bruchovsky, N., & Buckley, A. R. (1997). Increased cystine uptake capability associated with malignant progression of Nb_2 lymphoma cells. *Leukemia, 11*(8), 1329–1337.

Griffin, J. L., & Shockcor, J. P. (2004). Metabolic profiles of cancer cells. *Nature Reviews Cancer, 4*(7), 551–561.

Hirayama, A., Kami, K., Sugimoto, M., et al. (2009). Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Research, 69*(11), 4918–4925.

Ishigure, K., Shimomura, Y., Murakami, T., et al. (2001). Human liver disease decreases methacrylyl-CoA hydratase and beta-hydroxyisobutyryl-CoA hydrolase activities in valine catabolism. *Clinica Chimica Acta, 312*(1–2), 115–121.

Kantola, S. P. M., Jokinen, K., Hyrynkangs, K., Soini, Y., Alho, O. P., & Salo, T. (2000). Prognostic factors in tongue cancer-relative importance of demographic, clinical and histopathological factors. *British Journal of Cancer, 83*, 614–619.

Kind, T., Tolstikov, V., Fiehn, O., & Weiss, R. H. (2007). A comprehensive urinary metabolomic approach for identifying kidney cancer. *Analytical Biochemistry, 363*(2), 185–195.

Li, Y., St John, M. A. R., Zhou, X. F., et al. (2004). Salivary transcriptome diagnostics for oral cancer detection. *Clinical Cancer Research, 10*(24), 8442–8450.

Mao, L., Hong, W. K., & Papadimitrakopoulou, V. A. (2004). Focus on head and neck cancer. *Cancer Cell, 5*, 311–316.

Myers, J. N., Elkins, T., Roberts, D., & Byers, R. M. (2000). Squamous cell carcinoma of the tongue in young adults: increasing incidence and factors that predict treatment outcomes. *Otolaryngology: Head and Neck Surgery, 122*, 44–51.

Ni, Y., Su, M. M., Lin, J. C., et al. (2008). Metabolic profiling reveals disorder of amino acid metabolism in four brain regions from a rat model of chronic unpredictable mild stress. *FEBS Letters, 582*(17), 2627–2636.

Ni, Y., Su, M., Qiu, Y., et al. (2007). Metabolic profiling using combined GC-MS and LC-MS provides a systems understanding of aristolochic acid-induced nephrotoxicity in rat. *FEBS Letters, 581*(4), 707–711.

Pelayo, B. A., Fu, Y. M., & Meadows, G. G. (2001). Decreased tissue plasminogen activator and increased plasminogen activator inhibitors and increased activator protein-1 and specific promoter 1 are associated with inhibition of invasion in human A375 melanoma deprived of tyrosine and phenylalanine. *International Journal of Oncology, 18*(4), 877–883.

Qin, G. S., & Hotilovac, L. (2008). Comparison of non-parametric confidence intervals for the area under the ROC curve of a continuous-scale diagnostic test. *Statistical Methods in Medical Research, 17*(2), 207–221.

Qiu, Y., Cai, G., Su, M., et al. (2009). Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. *Journal of Proteome Research, 8*, 4844–4850.

Qiu, Y. P., Cai, G. X., Su, M. M., et al. (2010). Urinary metabonomic study on colorectal cancer. *Journal of Proteome Research, 9*(3), 1627–1634.

Qiu, Y. P., Su, M. M., Liu, Y. M., et al. (2007). Application of ethyl chloroformate derivatization for gas chromatography–mass spectrometry based metabonomic profiling. *Analytica Chimica Acta, 583*, 277–283.

Scully, C., & Felix, D. H. (2006). Oral medicine—update for the dental practitioner—oral cancer. *British Dental Journal, 200*(1), 13–17.

Spustova, V., & Oravec, C. (1989). Antitumor effect of hippurate—an experimental-study using various mouse-tumor strains. *Neoplasma, 36*(3), 317–320.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 64*, 479–498.

Sugimoto, M., Wong, D. T., Hirayama, A., Soga, T., & Tomita, M. (2010). Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics, 6*(1), 78–95.

Tiziani, S., Lopes, V., & Gunther, U. L. (2009). Early stage diagnosis of oral cancer using H-1 NMR-based metabolomics. *Neoplasia, 11*(3), 269–276.

Wei, J., Xie, G. X., Zhou, Z. T., et al. (2010). Salivary metabolite signatures of oral cancer and leukoplakia. *International Journal of Cancer*. doi:10.1002/ijc.25881.

Wong, D. T. (2006). Towards a simple, saliva-based test for the detection of oral cancer. *Expert Review of Molecular Diagnostics, 6*, 267–272.

Wu, H., Xue, R. Y., Dong, L., et al. (2009a). Metabolomic profiling of human urine in hepatocellular carcinoma patients using gas chromatography/mass spectrometry. *Analytica Chimica Acta, 648*(1), 98–104.

Wu, H., Xue, R. Y., Lu, C. L., et al. (2009b). Metabolomic study for diagnostic model of oesophageal cancer using gas chromatography/mass spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 877*(27), 3111–3117.

Wu, H., Xue, R. Y., Tang, Z. Q., et al. (2010). Metabolomic investigation of gastric cancer tissue using gas chromatography/mass spectrometry. *Analytical and Bioanalytical Chemistry, 396*(4), 1385–1395.

Xu, G. Q., Li, B. Q., & Li, H. H. (1992). *Oral precancerous pathological changes-oral leukoplakia and oral lichen planus*. Beijing: Chinese Medical Science and Technology Publishing House.

Yan, S. K., Wei, B. J., Lin, Z. Y., Yang, Y., Zhou, Z. T., & Zhang, W. D. (2008). A metabonomic approach to the diagnosis of oral squamous cell carcinoma, oral lichen planus and oral leukoplakia. *Oral Oncology, 44*(5), 477–483.

Ye, Z. C., & Sontheimer, H. (1999). Glioma cells release excitotoxic concentrations of glutamate. *Cancer Research, 59*(17), 4383–4391.

Zhou, J. L., Xu, B., Huang, J., et al. (2009). H-1 NMR-based metabonomic and pattern recognition analysis for detection of oral squamous cell carcinoma. *Clinica Chimica Acta, 401*(1–2), 8–13.