

# SCIENTIFIC REPORTS



OPEN

## Global metabolomics reveals potential urinary biomarkers of esophageal squamous cell carcinoma for diagnosis and staging

Received: 05 April 2016  
Accepted: 22 September 2016  
Published: 11 October 2016

Jing Xu<sup>1</sup>, Yanhua Chen<sup>1</sup>, Ruiping Zhang<sup>1</sup>, Jiuming He<sup>1</sup>, Yongmei Song<sup>2</sup>, Jingbo Wang<sup>3</sup>, Huiqing Wang<sup>1</sup>, Luhua Wang<sup>3</sup>, Qimin Zhan<sup>2</sup> & Zeper Abliz<sup>1,4</sup>

We performed a metabolomics study using liquid chromatography-mass spectrometry (LC-MS) combined with multivariate data analysis (MVDA) to discriminate global urine profiles in urine samples from esophageal squamous cell carcinoma (ESCC) patients and healthy controls (NC). Our work evaluated the feasibility of employing urine metabolomics for the diagnosis and staging of ESCC. The satisfactory classification between the healthy controls and ESCC patients was obtained using the MVDA model, and obvious classification of early-stage and advanced-stage patients was also observed. The results suggest that the combination of LC-MS analysis and MVDA may have potential applications for ESCC diagnosis and staging. We then conducted LC-MS/MS experiments to identify the potential biomarkers with large contributions to the discrimination. A total of 83 potential diagnostic biomarkers for ESCC were screened out, and 19 potential biomarkers were identified; the variations between the differences in staging using these potential biomarkers were further analyzed. These biomarkers may not be unique to ESCCs, but instead result from any malignant disease. To further elucidate the pathophysiology of ESCC, we studied related metabolic pathways and found that ESCC is associated with perturbations of fatty acid  $\beta$ -oxidation and the metabolism of amino acids, purines, and pyrimidines.

Esophageal cancer (EC) is a common cause of cancer-related death. In 2012, an estimated 455,800 new cases of EC were observed worldwide, and 400,200 deaths occurred as a result of EC<sup>1</sup>. China is a high-risk area for EC, and more than 90% of cases are esophageal squamous cell carcinoma (ESCC)<sup>2,3</sup>. Early-stage asymptomatic EC is usually curable with excellent long-term survival (90% or above at 5 years). Unfortunately, most patients exhibit locally advanced or metastatic EC at the time of diagnosis and have a poor prognosis (5-year survival rate, <20%)<sup>4-6</sup>. Even among patients with radical esophagectomies, the 5-year survival rate is only 10–20%<sup>7</sup>. Clearly, the discovery of potential biomarkers for early diagnosis of EC are urgently needed.

The advent of EC is accompanied by metabolic changes that are reflected by changes in gene expression, microRNA profiles, and the concentrations of circulating proteins and small metabolites<sup>8,9</sup>. Metabolomics involves the global and unbiased definition of the complement of small molecules in biofluids, tissues, organs, or organisms<sup>10-13</sup>. Therefore, this approach facilitates the screening and early detection of EC. Because metabolomics can provide information about disease processes, this method has been widely applied to the diagnosis of multiple pathologies<sup>14-17</sup>. Currently, nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are the most widely employed methods for metabolomics studies<sup>18</sup>. Compared with NMR, MS (particularly liquid chromatography [LC]-MS) offers several advantages, including high sensitivity and a wide dynamic range, and thus, LC-MS has become increasingly popular for metabolomics studies in recent years<sup>19-22</sup>.

<sup>1</sup>State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100050, P. R. China. <sup>2</sup>State Key Laboratory of Molecular Oncology, Cancer Institute & Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, P. R. China. <sup>3</sup>Department of Radiation Oncology, Cancer Institute & Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, P. R. China. <sup>4</sup>Centre for Bioimaging & Systems Biology, Minzu university of China, Beijing 100081, P. R. China. Correspondence and requests for materials should be addressed to Z.A. (email: zeper@imm.ac.cn)

Compounds	LC-(+)ESI-MS			
	Rt (min)	RSD (%)	Peak Area ( $\times 10^5$ )	RSD (%)
L-Carnitine	1.4	0.86	9.788	5.01
Phenylalanine	2.9	0.16	6.415	2.44
Hippuric acid	8.2	1.50	3.170	2.56
Hydrocortisone	15.0	0.51	18.375	4.20
Estrone	17.4	0.2	0.317	11.14
LC-(−)ESI-MS				
Phenylalanine	2.9	1.22	1.569	4.33
Hippuric acid	7.5	2.56	5.916	3.47
Tryptophan	5.1	0.86	1.674	4.02
Cholic acid	13.5	0.52	1.406	10.35
Linoleic acid	20.9	0.38	8.785	5.26

**Table 1.** Retention times of mixed standards peaks detected by LC-MS.

Several metabolomics studies of EC have been performed using various analytical platforms<sup>23–26</sup>. Previously, we also performed global and targeted metabolomics study of ESCC plasma to discover potential diagnostic and therapeutic biomarkers<sup>27</sup>. Compared to plasma, urine is more readily available and is noninvasively collected. Moreover, urine is not subject to homeostatic mechanisms, and greater varieties of endogenous metabolites could occur in urine, thereby better reflecting the changes in human metabolism.

Here, we performed LC-MS combined with multivariate data analysis (MVDA) to investigate the global urinary profiles of ESCC patients and normal controls. In addition, we evaluated the possibility of using urine metabolomics for the classification of ESCC and used an independent test set to examine the predictive ability of the analytical platform. Potential biomarkers were discovered, identified, and evaluated by receiver operating characteristic analysis (ROC). Moreover, we monitored the variation in these biomarkers during staging. Compare with the plasma metabolomics results, we investigated related metabolic pathways. The overall goals of this study were to (1) develop a LC-MS-based urine metabolomics method for ESCC diagnosis and staging, (2) discover potential biomarkers, and (3) illustrate the pathological changes associated with ESCC. The workflow of this study is shown in Figure S1.

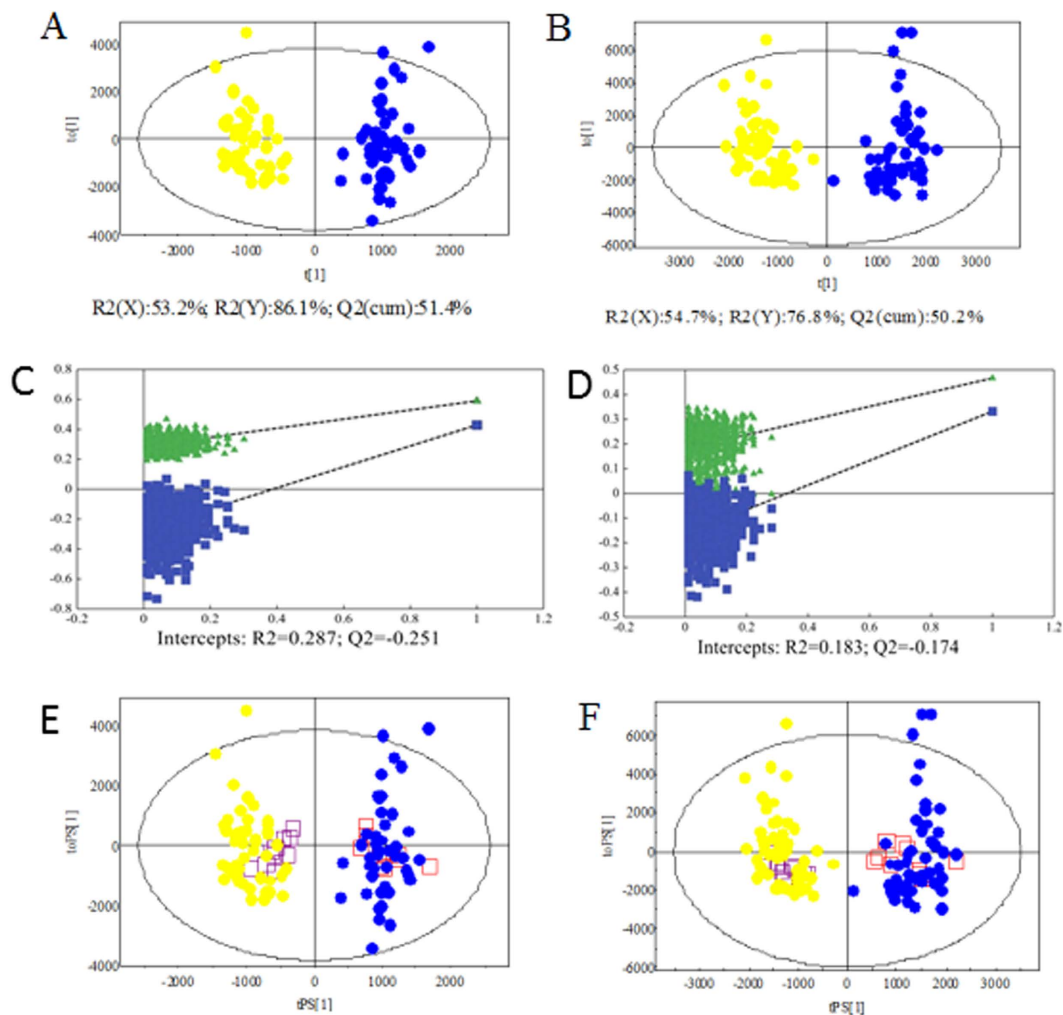
## Results

**Data quality assessment.** To obtain reliable data from metabolomics analysis, using a stable analytical method is important. Accordingly, unsupervised principal component analysis (PCA) was performed on all samples (including ESCC patients, NC and quality control (QC) samples) as part of an assessment of the stability of the process. Figure S2 shows that all QCs clustered in the center of the PCA. This results demonstrated no drift in retention time and chromatographic shape during the whole run-sequence, indicating the LC-MS results were statistically acceptable for analysis<sup>28,29</sup>. Furthermore, the mixed standard was also analyzed simultaneously with the acquired samples to evaluate the reproducibility of the method. The extracted ion chromatograms (XICs) of the mixed standard are shown in Figure S3. The relative standard deviations of the retention times of each standard compound in both ion modes were less than 5%, and the relative standard deviations of peak areas were below 15% (Table 1). In addition, the retention time deviation profiles of all urine samples resulted from R-software exhibited  $\pm 20$  s fluctuation in most of the LC-( $\pm$ ) electrospray ionization (ESI)-MS analyses (Figure S4). These results indicated that chromatographic separation and mass measurement were highly stable and reproducible throughout the sequence. The typical total ion chromatograms (TICs) produced from urine samples by LC-( $\pm$ ) ESI-MS are presented in Figure S5.

**Multivariate statistical analysis.** The LC-ESI-MS data sets obtained in both positive- and negative-ion modes contained 1463 and 2153 peaks, respectively, with retention times of 0.9–25 min. The program coded for LC-( $\pm$ )ESI-MS analysis is available in the Supporting Information. Unsupervised PCA by SIMCA-P was initially conducted to obtain an overview of the urinary LC-MS data from cancer patients and healthy controls. As shown in Figure S6, obvious separation trends between the two groups indicated that the ESCC patients exhibited metabolic changes relative to those of the controls.

To maximize the discrimination among the classes of observations and explore potential biomarker candidates in more detail, we applied orthogonal partial least squares discriminant analysis (OPLS-DA) as a stoichiometric analysis method to explore the difference between patients and controls. Ten ESCC patients and 10 controls were randomly selected to form an independent test set, and the training set consisted of the remaining subjects. The score scatter plots for the LC-( $\pm$ )ESI-MS data of the training set showed clear discrimination between the ESCC and NC groups (Fig. 1A,B). For LC-(+)ESI-MS data, the one predictive ( $t_p$ ) and two orthogonal ( $t_o$ ) (1 + 2) components were calculated with  $R^2(Y)$  and  $Q^2(\text{cum})$  values of 86.1% and 51.4%, respectively. The LC-(−)ESI-MS data set generated one predictive and three orthogonal components with  $R^2(Y)$  and  $Q^2(\text{cum})$  values of 76.8% and 50.2%, respectively.

To prevent original model overfitting, permutation tests with 999 iterations were performed (Fig. 1C,D). These permutation tests produced intercepts of  $R^2$  and  $Q^2$  with values of 0.287 and  $-0.251$  for the positive model data and 0.183 and  $-0.174$  for the negative model data (Fig. 1C,D). The criteria for validity were as follows:  $R^2$  less



**Figure 1.** Score plots of OPLS-DA models (A,B); Validation plots of the PLS-DA models (C,D) and T-predicted scatter plots (E,F) of OPLS-DA model based on the data from LC-(+)ESI-MS (A,C,E) and LC-(-)ESI-MS (B,D,F). (●, ESCC patients; ●, controls; □, ESCC patient prediction set; and □, control prediction set).

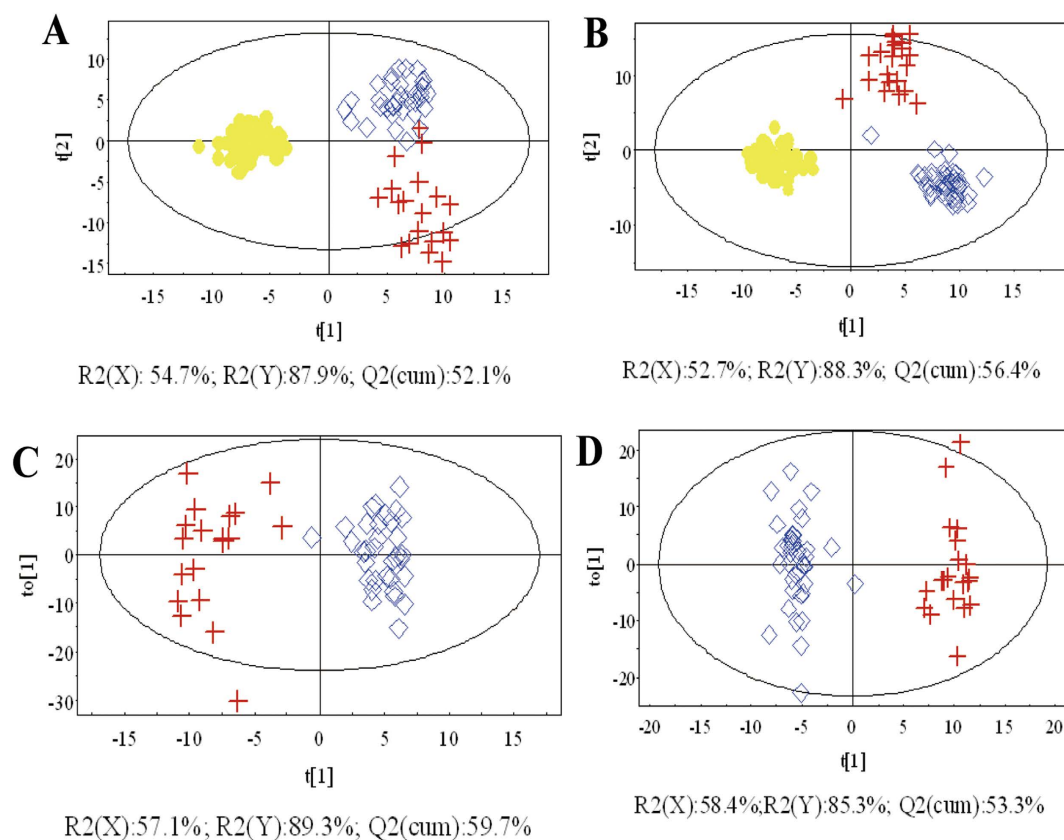
than 0.4, and  $Q^2$  less than 0.05<sup>30</sup>. Thus, the results indicate that the OPLS-DA models generated from the LC-(±)ESI-MS data were reliable.

To further evaluate the predictive ability of the established models, an external test using plasma samples from 10 patients and 10 controls was performed. As shown in Fig. 1E,F, satisfactory results were obtained. The OPLS-DA model correctly predicted all ESCC patients and healthy controls with 100% sensitivity and specificity while 1463 and 2153 ions of interest were applied. This result indicated that LC-MS-based urine metabolomics has potential applications for non-invasive ESCC diagnosis.

Furthermore, OPLS-DA was applied to differentiate early-stage (T1–2) and advanced-stage (T3–4) ESCC patients. The score scatter plots of OPLS-DA models from the LC-(±)ESI-MS data showed the clear differentiation of early-stage (T1–2) ESCC patients, advanced-stage (T3–4) ESCC patients, and NC groups (Fig. 2), indicating that changes in some endogenous metabolites were related to disease stage. The results of permutation tests with 999 iterations showed that the models were not overfitted (Figure S7).

**Discovery, identification and characterization of potential biomarkers.** Based on the OPLS-DA results, which facilitated a good group classification of ESCC patients and controls, we extracted potential markers of interest from the combined S-plot, variable importance in project (VIP), and raw data plots<sup>31,32</sup>. An independent t-test ( $P < 0.05$ ) was also performed to validate the significance of the discriminated variables selected by these methods. XICs were used to reduce the redundant variables originating from the same compound, such as adduct ions, fragments and isotopes. Ultimately, we selected 83 biomarker candidates for further identification.

The possible elemental compositions of the biomarkers were determined based on their exact masses, considering the relative intensities of the isotope peaks on high-resolution MS spectra. Furthermore, we elucidated the structures of the potential biomarkers based on high-resolution MS/MS spectra and searches in various databases. Subsequently, standard compounds were used to confirm the structures of these metabolites. Ultimately, 19 potential diagnostic biomarkers were identified, including nine carnitine derivatives (L-carnitine and eight acyl-carnitines), four amino acid derivatives (pyroglutamic acid, indoxyl, urocanic acid, and phenylacetylglutamine),



**Figure 2.** OPLS-DA score plots based on the data from (A,C) LC-(+)ESI-MS and (B,D) LC-(-)ESI-MS of ESCC patients and healthy controls. +, early-stage ESCC patients (T1–2);  $\diamond$ , advanced-stage ESCC patients (T3–4); and  $\bullet$ , controls.

three nucleosides (deoxycytidine, cyclic adenosine monophosphate [cAMP], and cyclic guanosine monophosphate [cGMP]), two purine derivatives (uric acid and paraxanthine), and L-Fucose. Detailed information regarding these compounds is listed in Table 2 and Figures S8–S25. Nine were further confirmed by comparison with authentic standards, including retention times and MS/MS fragmentation patterns. The identification score is also calculated by the scoring metric<sup>33</sup>. Hierarchical clustering analysis (HCA) of these potential biomarkers was conducted, and the results are shown in Fig. 3.

To further characterize the utility of these potential biomarkers for the prediction of ESCC, univariate ROC analysis and heat maps were carried out. The ROC curve could provide information regarding the sensitivity and specificity of the potential biomarkers. The metabolites were ranked according to the area under the ROC curve (AUC) values in heat maps, which were used to illustrate the discriminatory power of potential biomarkers (Fig. 4A). All AUC values were between 0.663 and 0.941. Because the ESCC is a complex disease that involves the systemic disorder of biochemical pathways, a biomarker panel containing a group of biomarkers rather than a single biomarker could be more powerful to discriminate and provide pathophysiology information. Therefore, metabolites with  $AUC > 0.85$  were analyzed by binary logistic regression combined with ROC curves to build the biomarker panel. The results (Fig. 4B) showed that the panel of five metabolites (decanoylcarnitine, cAMP, heptanoylcarnitine, cGMP, and phenylacetylglutamine) had an AUC of 0.981. The values of sensitivity and specificity reached 91.3% and 98.4%, respectively, at the best cut-off points. These results indicated that the biomarker panel could provide more reliable discrimination between ESCC patients and normal controls. In the future, larger urine samples will be acquired to validate these conclusions.

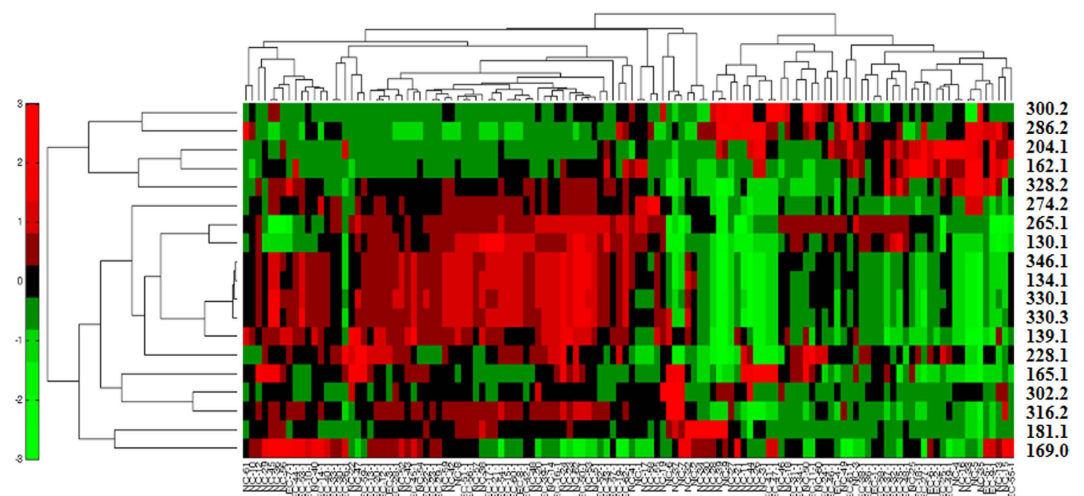
**Biological significance of biomarkers.** Among the 19 identified potential biomarkers, 11 were up-regulated, and eight down-regulated in patients (Table 2). The trends in the levels of these potential biomarkers in T1–2 and T3–4 ESCC patients relative to controls were further analyzed, and representative metabolites are shown in Fig. 5.

## Discussion

EC is an aggressive malignancy with poor prognosis due to the delayed diagnosis in part. Metabolomics, which is developed in recent years, offers a novel, convenient and sensitive approach to get the disturbed metabolic pathways and the tumor-associated biomarkers. Several metabolomics studies of EC have been carried out. Using NMR-based approach, Davis, V. W. *et al.* performed the urinary metabolomics of EC. The results showed clear distinctions between EC, Barrett's esophagus and healthy controls, and the related biomarkers were discovered.

No.	RT(min)	m/z	metabolite identification	P-value <sup>c</sup>	VIP <sup>d</sup>	trend <sup>e</sup>	related pathway	The identification score <sup>f</sup>
1	9.43	130.0501	Pyroglutamic acid <sup>ab</sup>	3.65E-06	6.59	↑	Glutamine metabolism	6
2	7.45	134.0591	Indoxyl <sup>b</sup>	0.006	1.52	↑	Tryptophan metabolism	6
3	2.27	139.0507	Urocanic acid <sup>ab</sup>	0.017	1.53	↑	Histidine metabolism	8
4	1.45	162.1097	L-Carnitine <sup>ab</sup>	0.021	2.79	↑	Fatty acid transportation	8
5	2.27	165.0554	L-Fucose <sup>ab</sup>	0.003	1.88	↑	Fructose and mannose degradation	8
6	2.89	169.0378	Uric acid <sup>ab</sup>	4.89E-05	7.12	↑	Purine metabolism	8
7	9.56	181.0719	Paraxanthine <sup>b</sup>	0.012	2.31	↓	Caffeine metabolism	6
8	2.07	204.1249	Acetylcarnitine <sup>b</sup>	0.025	6.65	↑	Fatty acid β oxidation	6
9	3.41	228.0801	Deoxycytidine <sup>ab</sup>	2.51E-09	4.36	↑	Pyrimidine metabolism	8
10	9.43	265.1169	Phenylacetylglutamine <sup>ab</sup>	7.8E-07	12.91	↑	Phenylalanine metabolism	8
11	11.93	274.2009	Heptanoylcarnitine (carnitine C 7:0) <sup>b</sup>	0.031	1.69	↓	Fatty acid β oxidation	6
12	12.3	286.2005	Octenoylcarnitine (carnitine C 8:1)	0.008	2.13	↓	Fatty acid β oxidation	4
13	13.39	300.2163	Nonenoylcarnitine (carnitine C 9:1)	0.025	1.39	↓	Fatty acid β oxidation	4
14	14.03	302.2319	Nonanoylcarnitine (carnitine C 9:0) <sup>b</sup>	0.0025	3.39	↓	Fatty acid β oxidation	6
15	14.92	316.2472	Decanoylcarnitine (carnitine C 10:0) <sup>ab</sup>	0.012	1.36	↓	Fatty acid β oxidation	8
16	15.14	328.2473	Undecenoylcarnitine (carnitine C 11:1)	0.019	1.78	↓	Fatty acid β oxidation	4
17	5.08	330.0588	cAMP <sup>ab</sup>	0.016	2.15	↑	Purine metabolism	8
18	11.70	330.2650	Undecanoylcarnitine (carnitine C 11:0) <sup>b</sup>	0.019	1.45	↓	Fatty acid β oxidation	6
19	4.94	346.0547	cGMP <sup>ab</sup>	2E-14	3.17	↑	Purine metabolism	8

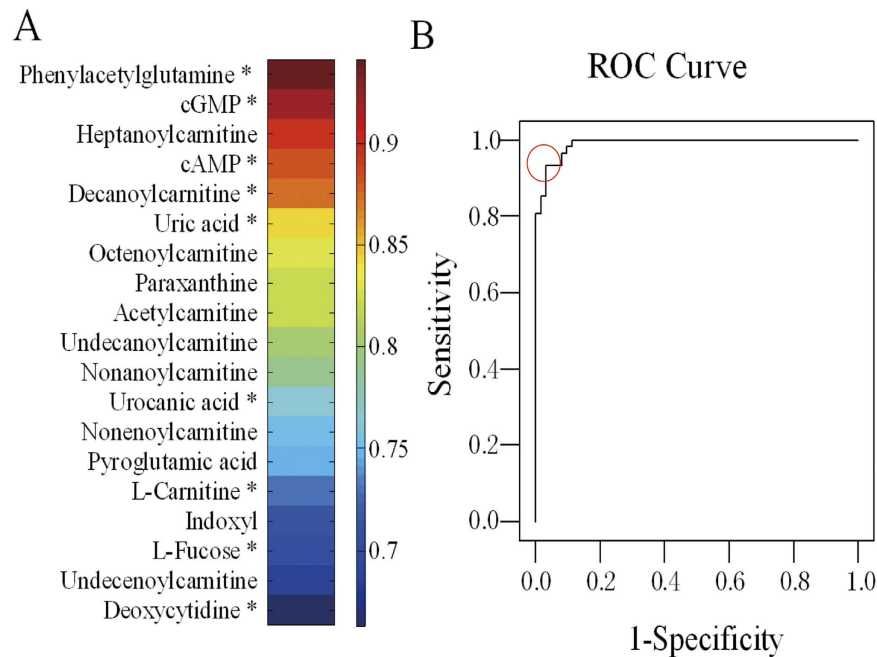
**Table 2. Urine potential biomarkers associated with ESCC.** <sup>a</sup>Metabolites confirmed by standard compounds. <sup>b</sup>Metabolites provisionally identified by database searches and MS fragmentation. Others, proposals based on MS fragmentation and exact mass data. <sup>c</sup>P value of independent *t*-test. <sup>d</sup>VIP is variable importance in the projection obtained from OPLS-DA with a threshold of 1.0. <sup>e</sup>Change trend compared with controls. (↑): up-regulated. (↓): down-regulated. <sup>f</sup>The identification score is calculated by the scoring metric.



**Figure 3. Hierarchical Clustering Analysis (HCA) of 19 potential biomarkers.**

However, the patients in this study were limited to esophageal adenocarcinoma (EAC)<sup>23</sup>. Jin, H. *et al.* performed the serum metabolomics signatures of lymph node metastasis of ESCC with gas chromatography (GC)-MS. A series of differential metabolites in serum for ESCC and lymph node metastatic ESCC patients were discovered and identified, and finally a potential biomarker panel (valine,  $\gamma$ -aminobutyric acid and pyrrole-2-carboxylic acid) were screened for ESCC diagnosis<sup>25</sup>. Previously, we also carried out the LC-MS-based plasma metabolomics of ESCC, and some potential biomarkers were discovered for diagnosis and therapeutic effect prediction<sup>27</sup>. In present study, the ESCC urinary metabolomics based on LC-MS approach was performed to discover the biomarkers for disease diagnosis and staging. The results might offer the supplement information for the previous studies, and be useful for the ESCC diagnosis.





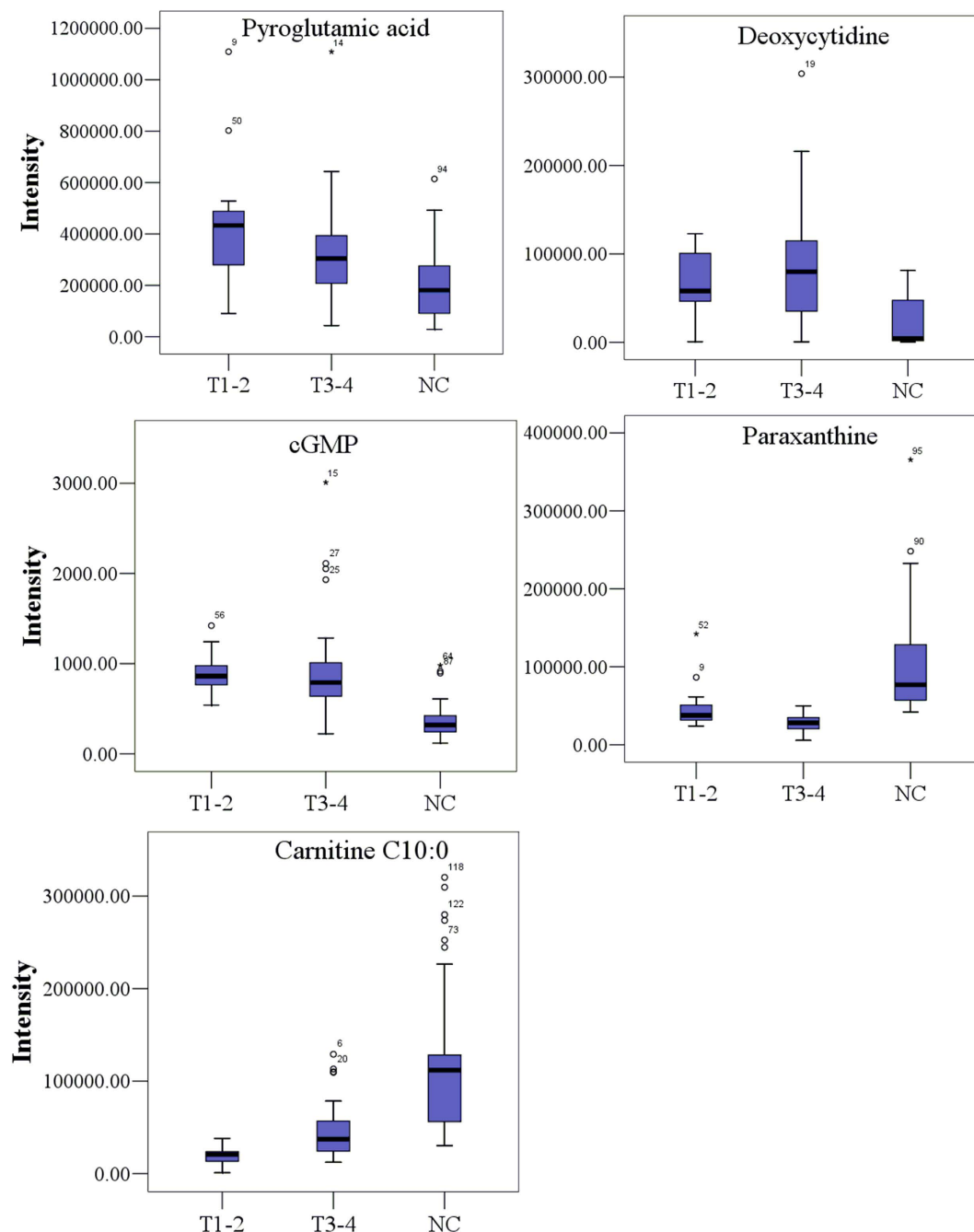
**Figure 4. Visualization of the discriminatory power of individual and combined potential diagnostic biomarkers.**

Carnitine and acylcarnitines, which are intermediates in the key energy metabolic pathways of fatty acid  $\beta$ -oxidation, are present at different concentrations in the urine of ESCC patients than in the urine of matched control patients (Table 2). Carnitines play an important role in transporting long-chain fatty acids across the mitochondrial membranes and short-chain fatty acids across mitochondria into the cytosol, and further participate in  $\beta$ -oxidation and energy metabolism<sup>34,35</sup>. The present results indicated that fatty acid  $\beta$ -oxidation was disturbed in ESCC patients. This finding is consistent with the Warburg effect, in which most cancer cells preferentially utilize glycolysis over other forms of energy production, including fatty-acid oxidation through acetyl-CoA<sup>36,37</sup>. The carnitine system in cancers has also been explored in previous studies examining other diseases<sup>38,39</sup>. Our previous results based on the ESCC plasma metabolomics also revealed that the levels of carnitines changed in patients<sup>27</sup>. This study provides the complementary information about the relationship between carnitines and cancers, especially for ESCC. The results may establish not only a new screening method, but also identify a new therapeutic target for this disease. For example, promoters of fatty acid  $\beta$ -oxidation might be evaluated to determine whether they have salutary effects on ESCC cells *in vitro*.

The dramatic increases in the levels of phenylacetylglutamine, pyroglutamic acid, urocanic acid, and indoxyl indicate that amino acid metabolism is disturbed in ESCC patients. Phenylacetylglutamine, a normal constituent of human urine, forms in the liver following the condensation of glutamine with phenylacetyl-CoA<sup>40</sup>. The urinary levels of phenylacetylglutamine have been used to monitor surrogate liver glutamate and to investigate a liver citric acid cycle intermediate<sup>41</sup>. Pyroglutamic acid, a cyclized derivative of L-glutamic acid, is formed nonenzymatically from glutamate, glutamine, and  $\gamma$ -glutamylated peptides, but it can also be produced through the action of  $\gamma$ -glutamylcyclotransferase on L-amino acid<sup>42</sup>. Elevated urine levels of pyroglutamic acid may be associated with glutamine metabolism problems. Urocanic acid is an intermediate in the conversion of histidine to glutamic acid. The up-regulation of urinary urocanic acid could result from a histidine metabolism disorder in tumor tissue and/or a glutamic acid metabolism disorder. Indoxyl is reported with tryptophan metabolism in Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis. The hydrolysis of tryptophan yields indole, and the oxidation of indole produces indoxyl.

Purine and pyrimidine metabolism were also abnormal in ESCC patients relative to controls. The cAMP and cGMP levels were significantly altered in ESCC patients. cAMP, the first second messenger to be identified, plays fundamental roles in cellular responses to many hormones and neurotransmitters<sup>43</sup>. cGMP, a ubiquitous second messenger, mediates several signal transduction pathways in mammalian cells<sup>44</sup>. Increasing evidence suggests that cGMP plays an important role in cellular proliferation, differentiation, and apoptosis<sup>45</sup>. Uric acid, produced by the enzyme xanthine oxidase during purine metabolism<sup>46</sup>, was found to be up-regulated in ESCC patients' urine samples, as observed in plasma<sup>27</sup>. Paraxanthine, the preferential product of caffeine metabolism in humans, is formed by the demethylation of caffeine by P450 1A2 in the liver. The reduction of paraxanthine may indicate that the metabolic activity of P450 1A2 declined<sup>47</sup>. Deoxycytidine is the intermediate or end-product of nucleotide and nucleic acid metabolism<sup>48</sup>. The increase in deoxycytidine might stem from abnormal cell proliferation in cancer tissue.

Cancer is considered to be a complex disease involving the systemic deregulation of cell proliferation, survival, apoptosis, and the cell cycle. Consequently, it can lead to disorders of some related metabolic pathways. Therefore,



**Figure 5.** Typical metabolite variations in urine samples from T1–2 and T3–4 ESCC patients relative to controls.

a biomarker panel might be more effective than a single biomarker for diagnosing cancer patients and elucidating the pathophysiology of cancer. It should be noted that the biomarker panel described here may not be unique to ESCC. Further validation of a highly ESCC-specific biomarker panel, including larger cohorts of different patients, will be conducted in the near future.

### Conclusions

Identifying metabolic biomarkers can contribute to improving diagnostics, prognostication, and therapy. Because the development and prognosis of ESCC varies significantly with genetic background, noninvasive ESCC biomarkers would significantly improve screening and diagnosis. Urinary metabolomics offers a novel and sensitive approach to simultaneously evaluating tumor-associated perturbations of multiple metabolic pathways and their downstream functional significance prior to the appearance of gross phenotypic changes.

In this study, we coupled LC-MS with MVDA to perform global urine metabolomics analysis of ESCC. The resulting data clearly demonstrated differences between patients and healthy controls. Furthermore, the data from

characteristics	ESCC patients	Healthy controls
No. of subjects	62	62
Age (mean, range)	62, 46–78	60, 45–74
BMI (mean, range)	22.1, 16.4–30.4	21.6, 16.7–29.6
Gender	male	male
Race	Chinese	Chinese
Cancer stage		
Early stage (T1–2, without metastases)	22 (T1: 7, T2: 15)	—
Advanced stage (T3–4, with metastases)	40 (T3: 19, T4: 21)	—

**Table 3. Clinicopathologic characteristics of the study samples.**

ESCC patients clustered according to the cancer stage. Finally, we identified 19 metabolites as potential diagnostic biomarkers and studied their related metabolic pathways. Significant differences in these biomarkers suggest that ESCC patients have disorders in fatty acid  $\beta$ -oxidation; amino acid, purine, pyrimidine metabolism; and fructose or mannose degradation. Importantly, metabolites are regulated by both intrinsic and extrinsic factors, and thus, the specificity of these endogenous markers must be further evaluated. Further studies will be conducted to validate these biomarkers in larger cohorts of different patients. This study confirmed the feasibility of using a LC-MS-based urine metabolomics platform to characterize ESCC.

## Methods

**Chemicals.** High-performance LC (HPLC)-grade acetonitrile (ACN) and formic acid (FA) were obtained from Merck (Germany). Standard compounds, including L-phenylalanine, hippuric acid, hydrocortisone, estrone, tryptophan, cholic acid, linoleic acid, urocanic acid, L-fucose, L-carnitine, uric acid, deoxycytidine, cAMP, and cGMP were purchased from Sigma-Aldrich (USA). Phenylacetylglutamine, trioxymethylantraquinone, and 2-hydroxybenzoic acid were purchased from the National Institute for the Control of Pharmaceutical and Biological Products (China).

**Sample Collection.** Sixty-two ESCC patients and 62 healthy volunteers from the Cancer Institute and Hospital of the Chinese Academy of Medical Sciences (Beijing, China) were enrolled in the study. All patients were diagnosed by histopathological examination. No patients had received chemotherapy or radiation, and they had not undergone surgical operation before sample collection. ESCC stage was built due to the 2009 Tumor Node Metastasis (TNM) staging system. The detailed demographic profiles of the participants are provided in Table 3. The study was approved by the Cancer Institute and Hospital of the Chinese Academy of Medical Sciences ethics committee and with the approval of corresponding regulatory agencies, and all the experiments were carried out in accordance with the approved guidelines. Meanwhile, all the patients involved in the study signed the informed consent form and agreed to participate. All urine samples were collected before breakfast following the informed consent guidelines and immediately stored at  $-80^{\circ}\text{C}$  prior to sample preparation and analysis.

**Sample preparation.** The urine samples were thawed at  $4^{\circ}\text{C}$  before analysis. Creatinine analysis was performed by the Inspection Department of the Cancer Institute and Hospital of the Chinese Academy of Medical Sciences using an enzymatic procedure. The samples were prepared by centrifugation at 10,000 relative centrifugal force (rcf) at  $4^{\circ}\text{C}$  for 5 min, followed by creatinine value-calibrated dilution. A pooled QC sample was prepared by mixing the same volume ( $10\mu\text{L}$ ) of each sample<sup>49</sup>. Moreover, a mixed standard including (1) L-carnitine, (2) L-phenylalanine, (3) hippuric acid, (4) hydrocortisone, (5) estrone, (6) tryptophan, (7) cholic acid, and (8) linoleic acid was also used to monitor the stability of the analytical system.

**LC-MS analysis.** The study was performed on a 1200 Series Rapid Resolution Liquid Chromatography system (Agilent Technologies, Germany) coupled to a quadrupole time-of-flight (Q-TOF) mass spectrometer (QSTAR Elite, AB Sciex, USA) equipped with ESI sources. The system was controlled by Analyst QS 2.0 (QSTAR Elite, AB Sciex, USA).

A  $10\text{ cm} \times 2.1\text{ mm}$  Zorbax Aq-C<sub>18</sub>  $1.8\mu\text{m}$  column was used and maintained at  $60^{\circ}\text{C}$ . The mobile phase was (A) 0.1% FA-water and (B) ACN, with multi-step gradient conditions as follows: initial 0% B maintained for 8 min, then increased to 10% B in 5 min; increased to 60% B over 5–15 min; 15–20 min to 100% B, and finally maintained at 100% B for 8 min, at a flow rate of  $200\mu\text{L}/\text{min}$ . The injection volume was  $5\mu\text{L}$  for each run. Healthy volunteers and ESCC patients samples were analyzed in random order. QC and mixed-standard samples were also analyzed repeatedly within the analytical run after every ten plasma samples to evaluate chromatographic reproducibility.

The LC-MS data were acquired in both positive and negative ion modes. The detailed parameters were as follows: spray voltage 5.5 kV or  $-4.5\text{ kV}$ , declustering voltage 50 V or  $-50\text{ V}$ , vaporizer temperature  $375^{\circ}\text{C}$ , turbo gas 75 psi, nebulizer gas 65 psi, curtain gas 40 psi. Full Scan analysis was performed in TOF mode with the scan range of  $m/z$  65–850, and the MS/MS analysis was accomplished with information-dependent acquisition (IDA) mode with collision energy (CE) 40/20 or  $-35/-20\text{ eV}$ . Acquired data were auto-calibrated by background ions in positive ion mode (phthalates:  $m/z$  149.0233 and  $m/z$  391.2843) and standard solutions (2-hydroxybenzoic acid:  $m/z$  137.0244 and trioxymethylantraquinone  $m/z$  269.0455) which were introduced by post-column mixing in negative ion mode.



**Data preprocessing and MVDA analysis.** Freely available software XCMS version 2.10.0 and commercial software SIMCA-P version 12.0 (Umetrics AB, Umeå, Sweden) were used in this study. Raw data obtained by LC-MS analysis was firstly converted to the mzData format by the Wiff-to-mzData translator and then imported into XCMS software<sup>50</sup> (<http://masspec.scripps.edu/xcms/xcms.phpUT>). Parameters for detailed data preprocessing in XCMS are available in the Supporting Information. After the data preprocessing, the SIMCA-P was further adopted for MVDA of the resultant 2D data matrices with mean centering and pareto scaling. Principal component analysis (PCA) was used to visualize the stability of the system. The cross-validation was used to test the model validity against overfitting. Potential biomarker candidates were selected based on variable importance in project (VIP > 1), S-plot, and the raw data plot in orthogonal partial least-square discriminant analysis (OPLS-DA) model, and independent t-test ( $P < 0.05$ ). Finally, the fragment, isotope and adduct ions were manually removed according to the corresponding extracted ion chromatograms (XICs) and the potential biomarkers were screened out.

**Metabolite identification and characterization.** The structure of potential biomarkers was identified as described<sup>27,33,51,52</sup>, firstly by searching the free databases such as HMDB (<http://hmdb.ca>), Massbank (<http://massbank.imm.ac.cn/MassBank>), and METLIN (<http://metlin.scripps.edu>) with exact molecular weights; and then using high-resolution LC-MS/MS spectra for further identification; applying standard compounds to verify the potential structures; and finally obtained the identification score by the scoring metric.

The discriminatory power of potential biomarkers was characterized by the area under the ROC curve (AUC) produced by SPSS (version 17.0)<sup>53</sup>, and visually displayed by heat maps<sup>54</sup>.

## References

- Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J. Clin.* **65**, 87–108 (2015).
- Vizcaino, A. P. *et al.* Time trends incidence of both major histologic types of esophageal carcinomas in selected countries, 1973–1995. *Int. J. Cancer* **99**, 860–868 (2002).
- Tran, G. D. *et al.* Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int. J. Cancer* **113**, 456–463 (2005).
- Zeng, H. *et al.* Cancer survival in China, 2003–2005: A population-based study. *Int. J. Cancer* **136**, 1921–1930 (2015).
- Besharat, S. *et al.* Inoperable esophageal cancer and outcome of palliative care. *World J Gastroenterol* **14**, 3725–3728 (2008).
- Tachibana, M. *et al.* Surgical treatment of superficial esophageal cancer. *Langenbecks Arch Surg* **391**, 304–321 (2006).
- Coleman, M. P. *et al.* EURO CARE-3 summary: cancer survival in Europe at the end of the 20<sup>th</sup> century. *Ann Oncol* **14**, Suppl. 5, v128–149 (2003).
- Hu, N. *et al.* Integrative genomics analysis of genes with biallelic loss and its relation to the expression of mRNA and micro-RNA in esophageal squamous cell carcinoma. *BMC Genomics* **16**, 732 (2015).
- Xu, Q. *et al.* Overexpression of the DEC1 protein induces senescence *in vitro* and is related to better survival in esophageal squamous cell carcinoma. *PLoS One* **7**, e41862 (2012).
- Nicholson, J. K. & Lindon, J. C. Metabonomics. *Nature* **455**, 1054–1056 (2008).
- Nicholson, J. K., Lindon, J. C. & Holmes, E. ‘Metabonomics’: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189 (1999).
- Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genom* **2**, 155–168 (2001).
- McMillan, A. *et al.* Multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. *Sci Rep* **5**, 14174 (2015).
- Spratlin, J. L., Serkova, N. J. & Eckhardt, S. G. Clinical applications of metabolomics in oncology: a review. *Clin Cancer Res* **15**, 431–440 (2009).
- Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–915 (2009).
- Hirayama, A. *et al.* Metabolic profiling reveals new serum biomarkers for differentiating diabetic nephropathy. *Anal Bioanal Chem* **404**, 3101–3109 (2012).
- Mörén, L. *et al.* Metabolomic Screening of Tumor Tissue and Serum in Glioma Patients Reveals Diagnostic and Prognostic Information. *Metabolites* **5**, 502–20 (2015).
- Zhang, A. *et al.* Modern analytical techniques in metabolomics analysis. *Analyst* **137**, 293–300 (2012).
- Luan, H. *et al.* Comprehensive urinary metabolomic profiling and identification of potential noninvasive marker for idiopathic Parkinson’s disease. *Sci Rep* **5**, 13888 (2015).
- Huan, T. *et al.* DnsID in MyCompoundID for Rapid Identification of Dansylated Amine- and Phenol-Containing Metabolites in LC-MS-Based Metabolomics. *Anal Chem* **87**, 9838–9845 (2015).
- Patterson, A. D. *et al.* Aberrant lipid metabolism in hepatocellular carcinoma revealed by plasma metabolomics and lipid profiling. *Cancer Res* **71**, 6590–6600 (2011).
- Liu, X. *et al.* Discovery and validation of plasma biomarkers for major depressive disorder classification based on liquid chromatography-mass spectrometry. *J Proteome Res* **14**, 2322–2330 (2015).
- Davis, V. W. *et al.* Urinary metabolomic signature of esophageal cancer and Barrett’s esophagus. *World J Surg Oncol* **10**, 271–296 (2012).
- Hasim, A. *et al.* Revealing the metabonomic variation of EC using <sup>1</sup>H-NMR spectroscopy and its association with the clinicopathological characteristics. *Mol Biol Rep* **39**, 8955–8964 (2012).
- Jin, H. *et al.* Serum metabolomic signatures of lymph node metastasis of esophageal squamous cell carcinoma. *J Proteome Res* **13**, 4091–4103 (2014).
- Sanchez-Espiridon, B. *et al.* Identification of serum markers of esophageal adenocarcinoma by global and targeted metabolic profiling. *Clin Gastroenterol Hepatol* **13**, 1730–1737 (2015).
- Xu, J. *et al.* Global and targeted metabolomics of esophageal squamous cell carcinoma discovers potential diagnostic and therapeutic biomarkers. *Mol Cell Proteomics* **12**, 1306–1318 (2013).
- Want, E. J. *et al.* Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc* **5**, 1005–18 (2010).
- Graham, S. F. *et al.* High resolution metabolomic analysis of ASD human brain uncovers novel biomarkers of disease. *Metabolomics* **12**, 1–10 (2016).
- Eriksson, L. *et al.* In: *Multivariate and Megavariate Data Analysis Part I: Basic Principles and Applications*. Umetrics AB, Swede (2006).

31. Wiklund, S. *et al.* Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* **80**, 115–122 (2008).
32. Wang, X. *et al.* Urine metabolomics analysis for biomarker discovery and detection of jaundice syndrome in patients with liver disease. *Mol Cell Proteomics* **11**, 370–380 (2012).
33. Sumner, L. *et al.* Proposed quantitative and alphanumeric metabolite identification metrics. *Metabolomics* **10**, 1047–9 (2014).
34. Borum, P. R. Changing perspective of carnitine function and the need for exogenous carnitine of patients treated with hemodialysis. *Am. J. Clin. Nutr* **64**, 976–977 (1996).
35. Bremer, J. Carnitine-metabolism and functions. *Physiological Reviews* **63**, 1420–1480 (1983).
36. Kim, J. W. & Dang, C. V. Cancer's molecular sweet tooth and the Warburg effect. *Cancer Res* **66**, 8927–8930 (2006).
37. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).
38. Qiu, Y. *et al.* Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. *J. Proteome Res.* **8**, 4844–4850 (1996).
39. Ganti, S. *et al.* Urinary acylcarnitines are altered in human kidney cancer. *Int J Cancer* **130**, 2791–2800 (2011).
40. Webster, L. T. *et al.* Identification of separate acyl-CoA: glycine and acyl-CoA: L-glutamine N-acyltransferase activities in mitochondrial fractions from liver of rhesus monkey and man. *J Biol Chem* **251**, 3352–3358 (1976).
41. Yang, D. *et al.* Assay of the human liver citric acid cycle probe phenylacetylglutamine and of phenylacetate in plasma by gas chromatography-mass spectrometry. *Anal Biochem.* **212**, 277–282 (1993).
42. Scott, I. R., Harding, C. R. & Barrett, J. G. Histidine-rich protein of the keratohyalin granules. Source of the free amino acids, urocanic acid and pyrrolidone carboxylic acid in the stratum corneum. *Biochim Biophys Acta* **719**, 110–117 (1982).
43. Sassone-Corsi, P. The cyclic AMP pathway. *Cold Spring Harb Perspect Biol* **4**, 1–3 (2012).
44. Fallahian, F. *et al.* Cyclic GMP induced apoptosis via protein kinase G in oestrogen receptor-positive and-negative breast cancer cell lines. *FEBS J* **278**, 3360–3369 (2011).
45. Loweth, A. C. *et al.* Evidence for the involvement of cGMP and protein kinase G in nitric oxide-induced apoptosis in the pancreatic B-cell line, HIT-T15. *FEBS Lett* **400**, 285–288 (1997).
46. Ames, B. N. *et al.* Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis. *Proc Natl Acad Sci USA* **78**, 6858–6862 (1981).
47. Furge, L. L. & Fletke, K. J. HPLC determination of caffeine and paraxanthine in urine: An assay for cytochrome P450 1A2 activity. *Biochem Mol Biol Educ* **35**, 138–144 (2007).
48. Mills, G. C., Schmalstieg, F. C. & Goldblum, R. M. Urinary purines and nucleosides in immune deficiencies. *Ann N Y Acad Sci* **451**, 66–77 (1985).
49. Gika, H. G. *et al.* Within-day reproducibility of an HPLC-MS-based method for metabolomic analysis: application to human urine. *J. Proteome Res.* **6**, 3291–3303 (2007).
50. Smith, C. A. *et al.* XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
51. An, Z. *et al.* Integrated ionization approach for RRLC-MS/MS-based metabolomics: finding potential biomarkers for lung cancer. *J. Proteome Res.* **9**, 4071–4081 (2010).
52. Chen, J. *et al.* Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. *Anal. Chim. Acta* **650**, 3–9 (2009).
53. Swets, J. A. Measuring the accuracy of diagnostic systems. *Science* **240**, 1285–1293 (1988).
54. Weinstein, J. N. Biochemistry. A postgenomic visual icon. *Science* **319**, 1772–1773 (2008).

## Acknowledgements

This study has been supported by the National Natural Science Foundation of China (Key Program, 21335007) and National High Technology Research and Development Program of China (863 Program, 2014AA021101).

## Author Contributions

Z.A. designed and supervised the entire study. Z.A. and J.X. planned the experiments; Y.S. and J.W. provided the clinical diagnosis information. Q.Z., L.W., J.W. and Y.S. collected the clinical samples. J.X. and H.W. performed experiments. J.X. wrote the manuscript. Z.A., Y.C. and R.Z. revised the manuscript. Z.A., Y.C., R.Z. and J.H. elucidated the implications of the results.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Xu, J. *et al.* Global metabolomics reveals potential urinary biomarkers of esophageal squamous cell carcinoma for diagnosis and staging. *Sci. Rep.* **6**, 35010; doi: 10.1038/srep35010 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016