Serum metabolomics analysis for the progression of esophageal squamous cell carcinoma

Xia Li¹, Ihong Zhao², Mengke Wei¹, Jiali Lv¹, Yawen Sun³, Xiaotao Shen⁴, Deli Zhao², Fuzhong Xue¹, Tao Zhang¹#, Jialin Wang¹,³#

1 Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, China

2 Tumor Preventative and Therapeutic Base of Shandong Province, Feicheng People’s Hospital, Feicheng 271600, China

3 Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan 250117, China

4 Interdisciplinary Research Center on Biology and Chemistry, and Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China

Running Head: metabolites and the progression of ESCC

# Correspondence & Reprints:

Tao Zhang, MD, PhD.
Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University. PO Box 100, 44 Wenhua Xi Road, Jinan 250012, China.
Email: taozhang@sdu.edu.cn

Jialing Wang
Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences. 440 Jiyan Road, Jinan 250117, China.
Email: wangjialin6681@gmail.com
**Abstract**

**BACKGROUND** Previous metabolomics studies have found differences in metabolic characteristics between the healthy and ESCC patients. However, few of these studies concerned the whole process of the progression of ESCC. This study aims to explore serum metabolites associated with the progression of ESCC.

**METHODS** Serum samples from 653 participants (305 normal, 77 esophagitis, 228 LGD, and 43 HGD/ESCC) were examined by ultra high performance liquid chromatography quadruple time-of-flight mass spectrometry (UHPLC-QTOF/MS). Principal component analysis (PCA) was first applied to obtain an overview of the clustering trend for the multidimensional data. Fuzzy c-means (FCM) clustering was then used to screen metabolites with a changing tendency in the progression of ESCC. Univariate ordinal logistic regression analysis and multiple ordinal logistic regression analysis were applied to evaluate the association of metabolites with the risk of ESCC progression, and adjusted for age, gender, BMI, tobacco smoking, and alcohol drinking status.

**RESULTS** After FCM clustering analysis, a total of 38 metabolites exhibiting changing tendency among normal, esophagitis, LGD, and HGD/ESCC patients. Final results showed 15 metabolites associated with the progression of ESCC. Ten metabolites (dopamine, L-histidine, 5-hydroxyindoleacetate, L-tryptophan, 2'-O-methylcytidine, PC (14:0/0:0), PC (O-16:1/0:0), PE (18:0/0:0), PC (16:1/0:0), PC (18:2/0:0)) were associated with decreased risk of developing ESCC. Five metabolites (hypoxanthine, inosine, carnitine (14:1), glycochenodeoxycholate, PC (P-18:0/18:3)) were associated with increased risk of developing ESCC.

**CONCLUSIONS** These results demonstrated that serum metabolites are associated with the progression of ESCC. These metabolites are capable of potential biomarkers for the risk prediction and early detection of ESCC.
Keywords: esophageal squamous cell carcinoma; serum metabolites; progression; FCM; ordinal logistic regression
Introduction

Esophageal squamous cell carcinoma (ESCC) is the predominant subtype of esophageal carcinoma, accounts for about 90% of cases of esophageal carcinoma worldwide[1, 2]. Most global ESCC cases occurred in the Central and South-East Asian region, which called “Asian esophageal cancer belt”, such as Iran, Turkey, Kazakhstan, and China[3, 4]. ESCC has a long progressive stage, including esophagitis, basal cell hyperplasia (BCH), esophageal squamous dysplasia (ESD), cancer in situ, invasive cancer, and metastatic cancer [5]. Unfortunately, it is asymptomatic in the early stage, the majority of patients are diagnosed at an advanced stage, with a poor prognosis. So identifying patients before they progress to advanced-stage disease is very important. Currently, early detection and screening of ESCC is primarily based on endoscopy, biopsy and pathological examination[6]. Endoscopic examination with iodine staining is widely used in high-risk areas of China[7, 8]. However, this method is expensive and invasive, it is difficult to be accepted by general residents[9, 10]. New cost-effective and non-invasive methods for identifying these patients with high sensitivity and specificity are not available.

Metabolic reprogramming has been proposed to be a key hallmark of cancer[11]. Recently, the application of metabolomics towards cancer research has been paid more attention. Metabolomics is a comprehensive and quantitative analysis of metabolites in the biological system under study[12]. Currently, metabolomics has been a powerful tool in the identification of metabolic changes in cancer progression and the discovery of non-invasive biomarkers for cancer prediction and diagnosis[13-16]. However, previous metabolomics studies on ESCC mainly focused on exploring potential diagnostic biomarkers based on healthy controls and ESCC patients[17-21], few studies pay attention to the whole process of the progression of ESCC. Zhang et al.[22] used serum metabolomic strategy based on gas chromatography-mass spectrometry (GC-MS) to identify and validate potential
metabolic markers for the discrimination of normal control, ESCC and ESD patients, it showed reasonable performance. Nevertheless, they still chose the traditional two-group comparison method in screening differential metabolites and ignored the ordinal relationship about disease progression.

In this study, serum samples from 653 participants (305 normal, 77 esophagitis, 228 LGD, and 43 HGD/ESCC) were collected at the Esophageal Cancer Screening Base in the high-risk area of China, covering the whole progression of ESCC. An untargeted ultra high performance liquid chromatography quadruple time-of-flight mass spectrometry (UHPLC-QTOF/MS) metabolomics approach was applied to these serum samples. Fuzzy c-means (FCM) clustering analysis and ordinal logistic regression analysis were combined to identify serum metabolites associated with the progression of ESCC. This study could be helpful for discovering new biomarkers for risk prediction and early detection of ESCC, especially precancerous lesions.

Materials and Methods

Study population

This study included subjects aged 40 to 69 years who had screened for esophageal cancer at the Esophageal Cancer Screening Base of Shandong Province (City of Feicheng, Shandong, China) between June 2013 and September 2014. The study was approved by the Ethics Committee of the Shandong Cancer Hospital and Institute, and all participants signed written informed consent.

The inclusion criteria for all patients in this study were as follows: (1) Diagnosis of esophageal squamous epithelial lesions at the Esophageal Cancer Screening Base of Shandong Province between June 2013 and September 2014. (2) Aged 40-69 years.

The exclusion criteria were as follows: (1) Coexisting other malignant tumors. (2)
Having history of surgery, radiotherapy or chemotherapy for other malignant tumors. (3) Suffering from metabolic diseases, liver diseases or kidney diseases. (4) Taking any medications. (5) Patients with inadequate clinical information.

The questionnaire interview was conducted by trained investigators to obtain their information, including age, gender, body mass index (BMI), tobacco smoking, and alcohol drinking status. In this study, all participants underwent a simple physical examination and endoscopy with mucosal iodine staining. For the participants who had suspicious tissues in esophageal mucosa (iodine-negative), the non-staining tissues were taken for biopsy and underwent pathological evaluation by two pathologists.

ESD is subdivided into three levels of severity: mild, moderate, and severe. Mild and moderate ESD are combined to low-grade dysplasia (LGD), while severe ESD and cancer in situ are considered as high-grade dysplasia (HGD). In this study, a total of 305 healthy subjects, 77 patients with esophagitis, 188 patients with mild ESD, 40 patients with moderate ESD, 15 patients with severe ESD, 12 patients with cancer in situ, and 16 patients with ESCC were enrolled. We grouped the participants into 4 groups: 'normal', 'esophagitis', 'LGD', and 'HGD/ESCC'. For the participants with multiple diagnoses in one biopsy, we only care about the most severe one.

**Serum samples collection and UHPLC-QTOF/MS analysis**

In brief, blood samples were taken in the morning from the participants after an overnight fasting. Whole blood specimens were immediately processed to obtain the serum samples and were immediately stored at -80°C until further analysis. Before UHPLC-QTOF/MS analysis, serum samples were thawed at 4 °C on ice and underwent the preprocessing procedure. The serum samples were randomly injected for the UHPLC-QTOF/MS analysis. A more detailed description can be found in the previous paper[17]. Blank samples (75 % ACN in water) and QC samples were injected every eight samples
during acquisition. Then raw data obtained by UHPLC-QTOF/MS analysis was further preprocessed and annotated (see details in supplement materials). Finally, 341 metabolite features were obtained for statistical analysis.

**Statistical analysis**

Principal component analysis (PCA) was first applied to obtain an overview of the clustering trend for the multidimensional data and evaluate the overall stability of the metabolomics data. Fuzzy c-means (FCM) clustering was then used to screen metabolites with a changing tendency in the progression of ESCC.

FCM [23, 24] is a kind of data soft clustering techniques, it allows each data point to belong to multiple clusters. Each data point has a fuzzy degree of belonging to each cluster. The algorithm of FCM is based on minimize an objective function:

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ij}^{m} \| x_i - c_j \|^2, \quad 1 \leq m \leq \infty$$  \hspace{1cm} (1)

Where, $u_{ij}$ is the degree to which an observation $x_i$ belongs to a cluster $c_j$, $c_j$ is the center of the cluster $j$, $m$ is a hyper-parameter that controls the level of fuzzy cluster fuzziness. The whole algorithm is made up of three steps: (a) Specify a number of clusters. (b) Assign coefficients randomly to each data point for being in the clusters. (c) Repeat until the algorithm has converged: Compute the centroid for each cluster, using the formula (2). For each data point, compute its coefficients of being in the clusters, using the formula (3).

$$c_j = \frac{\sum_{i=1}^{n} u_{ij} x_i}{\sum_{i=1}^{n} u_{ij}}$$  \hspace{1cm} (2)

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{\| x_i - c_j \|}{\| x_i - c_k \|} \right)^{m-1}}$$  \hspace{1cm} (3)

Then the univariate ordinal logistic regression analysis was performed to determine the significance of each metabolite, and the relevant false discovery rates (FDR) based on the $p$ values were estimated in the context of multiple testing. Metabolites with FDR value less
than 0.05 were further adjusted for age, gender, BMI, tobacco smoking, and alcohol drinking status. Finally, all metabolites founded above and five covariates mentioned above were included in multiple ordinal logistic regression analysis. All of the statistical analyses were performed on the R platform (version 3.5.3).

**Results**

**General Characteristics of Study Participants**

Table 1 shows the baseline characteristics of study participants included in this study. The mean ages calculated for the normal, esophagitis, LGD, and HGD/ESCC groups were 52.91, 58.31, 58.24, and 62.47 years, respectively. The univariate ordinal logistic analysis showed that significant differences ($P < 0.05$) in age, BMI, tobacco smoking, and alcohol drinking status among these groups. Notably, the highest percentages for tobacco smoking and alcohol drinking status were recorded in the HGD/ESCC group.

**Metabolic Profiling of Serum Samples**

In the UHPLC-QTOF-MS dataset, a total of 8182 metabolite features were detected. First, the reproducibility of the polar metabolite features was evaluated by relative standard deviation (RSD); metabolite features showing large variations were removed (RSD > 30%) before conducting the statistical analysis. After annotation, we obtained 341 metabolite features. To determine whether the metabolic profiles of the four groups were different, a multivariate statistical analysis using the principal component analysis (PCA) model was conducted. The score plots obtained from the PCA model are presented in Fig. S1. We excluded 12 PCA-based-outlier serum samples (7 normal, 2 esophagitis, 3 LGD) from further analysis. The PCA scores plot shows a clear cluster of the QC sample, indicating the high stability and reproducibility of the instrument.
FCM clustering analysis of metabolites

According to changing tendency, metabolites in the progression of ESCC were divided into six categories by FCM clustering analysis (Fig. 1). First, we excluded metabolites in cluster3 and cluster6 because they don't change regularly. Then, metabolites in cluster4 and cluster5 did not change significantly in the esophagitis, LGD, and HGD/ESCC groups, they were also excluded. Finally, 38 metabolites with membership greater than 0.45 in cluster 1 and cluster 2 were selected for further analysis.

Metabolites associated with the risk of ESCC Progression

After univariate ordinal logistic regression analysis, 15 metabolites were significant (FDR < 0.05) among with 38 metabolites mentioned above (Table S1). Then, adjusted age, gender, BMI, tobacco smoking, and alcohol drinking status, their expression were remains associated with the risk of ESCC progression (Table 2). Ten metabolites (dopamine, L-histidine, 5-hydroxyindoleacetate, L-tryptophan, 2'-O-methylcytidine, PC (14:0/0:0), PC (O-16:1/0:0), PE (18:0/0:0), PC (16:1/0:0), PC (18:2/0:0)) were associated with decreased risk of developing ESCC. Their boxplots of relative intensity of expression in the esophagitis, LGD, and HGD/ESCC groups were depicted in Fig. 2. Five metabolites (hypoxanthine, inosine, carnitine (14:1), glycochenodeoxycholate, PC (P-18:0/18:3)) were associated with increased risk of developing ESCC. Their changes of relative intensity of expression in each group are shown in Fig. S2.

Pathway enrichment analysis for metabolites that associated with the progression of ESCC was performed, we can see these metabolites main involved tryptophan metabolism, purine metabolism, histidine metabolism et al pathway (Fig. 3, see details in Table S2).

Further multiple ordinal logistic regression analysis included all 15 metabolites founded by above analysis and age, gender, BMI, tobacco smoking, and alcohol drinking status. After adjustment, several metabolites (dopamine, L-Histidine, 5-Hydroxyindoleacetate,
PC(O-16:1/0:0), hypoxanthine, Inosine, glycochenodeoxycholate) were remain associated with the progress of ESCC (Table 3).

**Discussion**

In the process of normal esophageal squamous epithelium develops into cancer, body metabolism also changes. In this study, we used untargeted metabolomics study based on UHPLC-QTOF/MS technique to explore serum metabolites associated with the progression of ESCC. A total of 15 metabolites associated with the progression of ESCC. These metabolites related to purine metabolism and energy metabolism (fatty acid metabolism and amino acid metabolism).

A series of phospholipid metabolites were found dysregulated in lesions associated with ESCC, mainly phosphatidylcholines (PC). Phospholipids converted from choline, they can form the bilayer structures of all cell membranes and are an important part of all cells[25]. Ma et al.[26] reported the serum expression levels of PCs in ESCC patients were lower than healthy subjects. In another serum metabolomics study, PCs were a major class of dysregulated metabolites, both down-regulated and up-regulated, suggesting potential perturbation of phosphocholine metabolism in ESCC[27]. Reduced PC metabolism has been also observed in prostate cancer[28] and cervical cancer[29]. Most of the PCs found in this study down-regulated with the development of ESCC. This is consistent with previous studies. The reduced metabolism of PCs in patients may meet the needs of high proliferation of cancer cells.

Carnitine is an important metabolite in fatty acid β-oxidation, it transports acyl CoA across the mitochondrial membranes into the mitochondria for oxidative decomposition[30]. The present result indicated that fatty acid β-oxidation was disturbed in the process of developing ESCC. Xu et al.[31] also found that L- carnitine up-regulated in ESCC patients
compared to healthy controls. Fatty acid β-oxidation provides extra ATP and is an important source of NADPH in cancer cells metabolic reprogramming[32]. The increased carnitine promotes fatty acid β-oxidation, which helps the development of cancer.

We observed that metabolites from the purine metabolic pathway, hypoxanthine and inosine were associated with ESCC risk. In the purine degradation pathway, inosine can convert to hypoxanthine by purine nucleoside phosphorylase (PNP), then xanthine-oxidase (XO) catalyzes hypoxanthine oxidation to form uric acid finally[33]. Previous studies found an up-regulated uric acid in ESCC patients’ urine samples[18] and plasma samples[31]. This change in expression also showed in serum samples of esophageal cancer patients, gastric cancer patients and colorectal cancer patients[34]. Previous studies also showed that uric acid contributes to cancer risk, recurrence, and mortality[35].

Glycochenodeoxycholate is a primary bile acid, it showed an up-regulated trend in the process of developing ESCC in the present study. Nishioka et al.[36] reported the possibility of bile acid contributing to squamous carcinogenesis of the esophagus, it promotes the proliferation of ESCC.

In the present study, L-tryptophan and 5-hydroxyindoleacetate were down-regulated expression in lesions associated with ESCC. Zhu et al.[20] also reported the plasma levels of L-tryptophan were lower in ESCC patients. 5-hydroxyindoleacetate is a product of L-tryptophan metabolism in tryptophan metabolic pathway. Compared to the healthy controls, the concentration of tryptophan is down-regulated in both ESCC and metastatic ESCC patients, further analysis showed that disturbed tryptophan metabolism correlating to progression and metastasis of ESCC[37, 38]. The level of tryptophan decreased in the esophageal cancer (EC) and esophageal adenocarcinoma (EAC) patients compared with healthy controls, which indicates an increased demand for and overutilization of amino acids in the tumor tissue [39, 40].
Ma et al. [41] have reported that the expression of histidine is significantly down-regulated in ESCC patients compared to healthy people. Wang et al. [42] also found that histidine significantly decreased in EC tissues in comparison to normal mucosae. In this study, L-histidine also has a down-regulated trend in the development of ESCC. Histidine catabolism produces one carbon unit to participate in the synthesis of purines and pyrimidines. The lower levels of histidine in body mainly due to active nucleic acid metabolism in tumor cells, histidine consumption is markedly increased.

Dopamine is involved in the tyrosine metabolic pathway. We found it down-regulated in the development of ESCC, possibly because of a decrease in tyrosine. Tyrosine down-regulated has also been observed in EC and EAC patients [39, 40].

Although our study included 653 serum samples from the normal to ESCC progression stages (305 normal, 77 esophagitis, 228 LGD, and 43 HGD/ESCC) and used ordinal information of these four groups in statistical analysis to explore serum metabolites associated with the progression of ESCC. This study also has some limitations. First, the distribution of sample size in each group was unbalanced, which could affect the results of our statistical analysis. Second, this study found some metabolites related to the progress of ESCC, but because it is difficult to obtain more samples covering all stages of ESCC, the role of these metabolites in risk prediction and early detection of ESCC lacks further analysis and validation.

In conclusion, this serum metabolomics study demonstrated that serum metabolites associated with the progression of ESCC. The results obtained in the current study showed that 15 metabolites were significantly altered among normal, esophagitis, LGD and HGD/ESCC groups. These metabolites are capable of potential biomarkers for the risk prediction and early detection of ESCC. Further studies are necessary in order to fully understand the association of metabolic changes with ESCC progression and their risk.
prediction and early detection potentials.
**Abbreviations:** ESCC: Esophageal squamous cell carcinoma; LGD: low-grade dysplasia; HGD: High-grade dysplasia; BMI: body mass index; RSD: relative standard deviation; PCA: Principal component analysis; FCM: Fuzzy c-means clustering; FDR: false discovery rates; OR: odds ratio; CI: confidence interval.

**Acknowledgements:** We greatly appreciate those individuals who took part in our study. Without their dedication and cooperation, our study could not have begun. In addition, our study has been supported by many research groups and organizations. We are very grateful for all the supports giving by investigators, clinicians and pathologists from Shandong Cancer Hospital, Shandong Cancer Control and Prevention office, Department of Biostatistics of Shandong University, Cancer Center of Feicheng Hospital, Centers of Disease Control, and Community Health Centers.

**Author Contributions:** All of the authors contributed significantly to this work. Tao Zhang and Jialin Wang designed the study and enrolled the subjects; Xiaotao Shen carried out metabolomics experiments; Xia Li, Mengke Wei, and Jiali Lv analyzed the data and wrote the manuscript; Yawen Sun, Deli Zhao, Fuzhong Xue, and Jialin Wang supervised the field activities and data collection and edited the manuscript. All of the authors read and approved the final manuscript.

**Funding:** This study was supported by grants from National Natural Science Foundation of China (81573246, 81973147 and 81673271), Cheeloo Young Scholars Program of Shandong University, Shandong University multidisciplinary research and innovation team of young scholars (2020QNQT11 and IFYT18034). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.
References

**Figure Legends**

**Figure 1** FCM clustering analysis of 341 metabolites from normal participants to HGD/ESCC patients. LGD, low-grade dysplasia; HGD, high-grade dysplasia; ESCC, esophageal squamous cell carcinoma. Each of these lines represents a metabolite. The color of the line indicates the membership of the metabolite in cluster.

**Figure 2** Boxplots of relative intensity of expression of ten metabolites that associated with decreased risk of developing ESCC between normal and esophagitis, LGD and HGD/ESCC patients. LGD, low-grade dysplasia; HGD, high-grade dysplasia; ESCC, esophageal squamous cell carcinoma.

**Figure 3** Pathway enrichment analysis for metabolites that associated with the progression of esophageal squamous cell carcinoma.
Figure 1
Figure 2
Figure 3
Table 1. General characteristics of study subjects

<table>
<thead>
<tr>
<th>Variables</th>
<th>Normal</th>
<th>Esophagitis</th>
<th>LGD</th>
<th>HGD/ESCC</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>305</td>
<td>77</td>
<td>228</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Age (years) (SD)</td>
<td>52.9 (7.6)</td>
<td>58.3 (6.9)</td>
<td>58.2 (7.3)</td>
<td>62.5 (6.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>175 (57.4)</td>
<td>39 (50.7)</td>
<td>114 (50.0)</td>
<td>21 (48.8)</td>
<td>0.068</td>
</tr>
<tr>
<td>BMI (Kg/m²) (SD)</td>
<td>24.8 (3.4)</td>
<td>23.6 (3.7)</td>
<td>24.2 (3.2)</td>
<td>23.1 (3.6)</td>
<td>0.002</td>
</tr>
<tr>
<td>Tobacco smoking, n (%)</td>
<td>53 (17.4)</td>
<td>18 (23.4)</td>
<td>61 (26.8)</td>
<td>14 (32.6)</td>
<td>0.002</td>
</tr>
<tr>
<td>Alcohol drinking, n (%)</td>
<td>75 (24.6)</td>
<td>23 (29.9)</td>
<td>69 (30.2)</td>
<td>18 (41.9)</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; SD, standard deviation; LGD, low-grade dysplasia; HGD, high-grade dysplasia; ESCC, esophageal squamous cell carcinoma
Table 2. Metabolites associated with the progression of ESCC by ordinal logistic analysis adjusted for age, gender, BMI, tobacco smoking and alcohol drinking status.

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Cluster</th>
<th>Adjusted OR (95% CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dopamine</td>
<td>1</td>
<td>0.70 ( 0.59 ~ 0.84 )</td>
<td>1.67E-04</td>
</tr>
<tr>
<td>L-Histidine</td>
<td>1</td>
<td>0.60 ( 0.50 ~ 0.72 )</td>
<td>9.03E-08</td>
</tr>
<tr>
<td>5-Hydroxyindoleacetate</td>
<td>1</td>
<td>0.72 ( 0.61 ~ 0.84 )</td>
<td>5.08E-05</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>1</td>
<td>0.73 ( 0.62 ~ 0.85 )</td>
<td>8.65E-05</td>
</tr>
<tr>
<td>2'-O-methylcytidine</td>
<td>1</td>
<td>0.72 ( 0.53 ~ 0.97 )</td>
<td>3.02E-02</td>
</tr>
<tr>
<td>PC (14:0/0:0)</td>
<td>1</td>
<td>0.68 ( 0.58 ~ 0.81 )</td>
<td>9.13E-06</td>
</tr>
<tr>
<td>PC (O-16:1/0:0)</td>
<td>1</td>
<td>0.59 ( 0.50 ~ 0.69 )</td>
<td>2.06E-10</td>
</tr>
<tr>
<td>PE (18:0/0:0)</td>
<td>1</td>
<td>0.59 ( 0.49 ~ 0.71 )</td>
<td>2.26E-08</td>
</tr>
<tr>
<td>PC (16:1/0:0)</td>
<td>1</td>
<td>0.64 ( 0.54 ~ 0.76 )</td>
<td>2.64E-07</td>
</tr>
<tr>
<td>PC (18:2/0:0)</td>
<td>1</td>
<td>0.74 ( 0.63 ~ 0.87 )</td>
<td>2.32E-04</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>2</td>
<td>1.37 ( 1.17 ~ 1.59 )</td>
<td>7.21E-05</td>
</tr>
<tr>
<td>Inosine</td>
<td>2</td>
<td>1.35 ( 1.16 ~ 1.56 )</td>
<td>9.11E-05</td>
</tr>
<tr>
<td>Carnitine (14:1)</td>
<td>2</td>
<td>1.23 ( 1.05 ~ 1.43 )</td>
<td>9.52E-03</td>
</tr>
<tr>
<td>Glycochenodeoxycholate</td>
<td>2</td>
<td>1.22 ( 1.05 ~ 1.42 )</td>
<td>1.09E-02</td>
</tr>
<tr>
<td>PC (P-18:0/18:3)</td>
<td>2</td>
<td>1.20 ( 1.03 ~ 1.41 )</td>
<td>1.85E-02</td>
</tr>
</tbody>
</table>
Table 3 Multiple ordinal logistic regression analysis for association between the progression of ESCC and all metabolites

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>$\beta$</th>
<th>Std. Error</th>
<th>t value</th>
<th>p value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dopamine</td>
<td>-0.213</td>
<td>0.099</td>
<td>-2.157</td>
<td>3.10E-02</td>
<td>0.81 (0.67 ~ 0.98 )</td>
</tr>
<tr>
<td>L-Histidine</td>
<td>-0.365</td>
<td>0.101</td>
<td>-3.601</td>
<td>3.17E-04</td>
<td>0.69 (0.57 ~ 0.85 )</td>
</tr>
<tr>
<td>5-Hydroxyindoleacetate</td>
<td>-0.230</td>
<td>0.094</td>
<td>-2.463</td>
<td>1.38E-02</td>
<td>0.79 (0.66 ~ 0.95 )</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>-0.056</td>
<td>0.096</td>
<td>-0.582</td>
<td>5.60E-01</td>
<td>0.95 (0.78 ~ 1.14 )</td>
</tr>
<tr>
<td>PC(14:0/0:0)</td>
<td>0.179</td>
<td>0.136</td>
<td>1.322</td>
<td>1.86E-01</td>
<td>1.20 (0.92 ~ 1.56 )</td>
</tr>
<tr>
<td>PC(O-16:1/0:0)</td>
<td>-0.251</td>
<td>0.103</td>
<td>-2.425</td>
<td>1.53E-02</td>
<td>0.78 (0.64 ~ 0.95 )</td>
</tr>
<tr>
<td>PE(18:0/0:0)</td>
<td>-0.231</td>
<td>0.120</td>
<td>-1.920</td>
<td>5.49E-02</td>
<td>0.79 (0.63 ~ 1.00 )</td>
</tr>
<tr>
<td>PC(16:1/0:0)</td>
<td>-0.212</td>
<td>0.135</td>
<td>-1.573</td>
<td>1.16E-01</td>
<td>0.81 (0.62 ~ 1.05 )</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>0.366</td>
<td>0.084</td>
<td>4.372</td>
<td>1.23E-05</td>
<td>1.44 (1.22 ~ 1.70 )</td>
</tr>
<tr>
<td>PC(18:2/0:0)</td>
<td>-0.106</td>
<td>0.101</td>
<td>-1.053</td>
<td>2.92E-01</td>
<td>0.90 (0.74 ~ 1.10 )</td>
</tr>
<tr>
<td>Inosine</td>
<td>0.348</td>
<td>0.085</td>
<td>4.106</td>
<td>4.02E-05</td>
<td>1.42 (1.20 ~ 1.67 )</td>
</tr>
<tr>
<td>carnitine(14:1)</td>
<td>0.065</td>
<td>0.089</td>
<td>0.729</td>
<td>4.66E-01</td>
<td>1.07 (0.90 ~ 1.27 )</td>
</tr>
<tr>
<td>Glycochenodeoxycholate</td>
<td>0.258</td>
<td>0.085</td>
<td>3.037</td>
<td>2.39E-03</td>
<td>1.29 (1.10 ~ 1.53 )</td>
</tr>
<tr>
<td>2’-O-methylcytidine</td>
<td>-0.154</td>
<td>0.117</td>
<td>-1.312</td>
<td>1.90E-01</td>
<td>0.86 (0.68 ~ 1.08 )</td>
</tr>
<tr>
<td>PC(P-18:0/18:3)</td>
<td>0.144</td>
<td>0.081</td>
<td>1.771</td>
<td>7.66E-02</td>
<td>1.16 (0.98 ~ 1.36 )</td>
</tr>
</tbody>
</table>
Supplementary materials

Serum metabolomics analysis for the progression of esophageal squamous cell carcinoma

Serum samples collection and preparation

Blood samples were taken in the morning from the participants after an overnight fasting. The samples were clotted at 37 °C water bath for 30 min and followed by centrifugation for 15 min at 3000 rpm. Then the serum supernatant was taken, immediately stored at -80°C until further analysis.

Before UHPLC-QTOF/MS analysis, serum samples were thawed at 4 °C on ice. Then 50 μL of serum sample was taken and placed in a 96-well plate, then extracted with 150 μL of MeOH (which was kept at -20 °C before extraction) using Bravo liquid handling system (Agilent Technologies, USA), and followed by vortex for 30 s and incubation for 2 h at -20 °C to precipitate proteins. The 96-well plate was then centrifuged at 4000 rpm for 20 min at 4 °C. The resulting supernatants were transferred to LC–MS vials and stored at -80 °C until the UHPLC-QTOF/MS analysis.

UHPLC-QTOF/MS analysis

The serum samples were randomly injected for the UHPLC-QTOF/MS analysis. Quality control (QC) samples were prepared by pooling aliquots of all serum samples that were representative of the serum samples under analysis, and used for data normalization. Blank samples (75 % ACN in water) and QC samples were injected every eight samples during acquisition.

The UHPLC-QTOF/MS analyses were performed using a UHPLC system (1290 series, Agilent Technologies, USA) coupled to a quadruple time-of-flight (QTOF) mass spectrometer (Agilent 6550 iFunnel Q-TOF, Agilent Technologies, USA). Waters ACQUITY UHPLC HSS T3 columns [particle size, 1.8 μm; 100 mm (length) × 2.1 mm (i.d.)] were used
for the LC separation and the column temperature was kept as 25 °C. The flow rate was 0.5 mL/min and the sample injection volume was 6 μL. The mobile phases A was 0.1 % FA in water in positive mode (ESI+) or 0.5 mM NH4F in water in negative mode (ESI-), and B was 0.1 % FA in ACN in positive mode or 100 % ACN in negative mode. The linear gradient was set as follows: 0–1 min: 1 % B, 1–8 min: 1 % B to 100 % B, 8–10 min: 100 % B, 10–10.1 min: 100 % B to 1 % B, 10.1–12 min: 1 % B. The acquisition rate was set as 4 spectra/s and the TOF mass range was set as m/z 50–1200 Da. The parameters of MS data acquisition were set as follows: sheath gas temperature, 400 °C; dry gas temperature, 250 °C; sheath gas flow, 12 L/min; dry gas flow, 16 L/min; capillary voltage, 3000 V in positive mode or -3000 V in negative mode, respectively; nozzle voltage, 0 V; and nebulizer pressure, 20 psi in positive or 40 psi in negative mode, respectively.

Tandem mass spectrometry (MS/MS) data acquisition was performed using another quadruple time-of-flight mass spectrometer (Triple TOF 5600+, AB SCIEX, USA). QC samples were used for MS/MS data acquisition. To expand the coverage of MS/MS spectra, the mass range were divided into four segments: 50–300 Da, 290–600 Da, 590–900 Da, 890–1200 Da. The acquired MS/MS spectra were matched against in-house tandem MS spectral library for metabolite identification (see details in follow). The source parameters were set as follows: GAS1, 60; GAS2: 60; CUR: 30; TEM: 600 °C; ISVF: 5500 V and -4500 V in positive and negative modes, respectively.

**Metabolite identification:** Tandem mass spectrometry (MS/MS) spectra data for metabolites were acquired using AB Sciex TripleTOF 5600+. The acquired MS/MS spectra were matched against our in-house standard MS/MS spectral library and metabolite standards. The MS/MS spectra match score was calculated using dot-product algorithm ranging from 0-1. The cutoff for match score was set as 0.8. The MS/MS spectra match results were further manually checked to confirm the identification. All the m/z errors are
less than 25 ppm and all the RT errors are less than 18 seconds. Secondly, the biomarkers whose MS/MS spectra were not matched in our in-house databases or cannot be interpreted by MS/MS spectra, were searched against online databases of HMDB (www.hmdb.ca), METLIN (http://metlin.scripps.edu/) and KEGG (http://www.genome.jp/kegg/). The mass tolerance between the measured m/z values and the exact masses of the components of interest was set to within 30 ppm. The potential biomarkers were identified according to their molecular weights.

**Data preprocessing and annotation**

Raw data obtained by UHPLC-QTOF/MS analysis was firstly converted to the mzXML format using ProteoWizard, and processed by R package XCMS (version 3.2). Then a data matrix consisted of the retention time (RT), mass-to-charge ratio (m/z) values, and peak intensity was generated. R package CAMERA was used for peak annotation after XCMS data processing. Metabolic features detected less than 80 % in all the QC samples were discarded. Only monoisotopic peaks annotated by CAMERA were selected for the subsequent statistical analyses. To remove the unwanted analytical variations occurring intra- and inter- batches, each metabolite peak in all subject samples was normalized using the LOESS method based on QC samples. Then the relative intensity of each metabolite was standardized with Z-transformation (mean = 0, SD = 1).
**Supplementary figure legends:**

**Figure S1.** The PCA performed on the whole samples including normal participants, esophagitis, LGD and HGD/ESCC patients, and QC samples. LGD, low-grade dysplasia; HGD, high-grade dysplasia; ESCC, esophageal squamous cell carcinoma.

**Figure S2.** Boxplots of relative intensity of expression of five metabolites that associated with increased risk of developing ESCC between normal and esophagitis, LGD and HGD/ESCC patients. LGD, low-grade dysplasia; HGD, high-grade dysplasia; ESCC, esophageal squamous cell carcinoma.
Figure S1
Figure S2

Groups

- Normal
- Esophagitis
- LGD
- HGD/ESCC

Relative Intensity

Metabolites

- Hypoxanthine
- Inosine
- Camilin(14:1)
- Glycochenodeoxycholate
- PC(P-18:0/18:3)
<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Cluster</th>
<th>p value</th>
<th>FDR</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dopamine</td>
<td>1</td>
<td>2.63E-05</td>
<td>1.17E-04</td>
<td>0.69 (0.59 ~ 0.82)</td>
</tr>
<tr>
<td>L-Histidine</td>
<td>1</td>
<td>1.13E-09</td>
<td>2.26E-08</td>
<td>0.59 (0.49 ~ 0.70)</td>
</tr>
<tr>
<td>5-Hydroxyindoleacetate</td>
<td>1</td>
<td>8.73E-04</td>
<td>2.91E-03</td>
<td>0.78 (0.67 ~ 0.90)</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>1</td>
<td>1.50E-07</td>
<td>1.50E-06</td>
<td>0.67 (0.58 ~ 0.78)</td>
</tr>
<tr>
<td>2'-O-methylcytidine</td>
<td>1</td>
<td>1.99E-02</td>
<td>4.06E-02</td>
<td>0.70 (0.52 ~ 0.95)</td>
</tr>
<tr>
<td>PC (14:0/0:0)</td>
<td>1</td>
<td>1.70E-06</td>
<td>1.30E-05</td>
<td>0.68 (0.58 ~ 0.79)</td>
</tr>
<tr>
<td>PC (O-16:1/0:0)</td>
<td>1</td>
<td>6.42E-10</td>
<td>2.26E-08</td>
<td>0.61 (0.52 ~ 0.71)</td>
</tr>
<tr>
<td>PE (18:0/0:0)</td>
<td>1</td>
<td>5.54E-08</td>
<td>7.38E-07</td>
<td>0.63 (0.53 ~ 0.74)</td>
</tr>
<tr>
<td>PC (16:1/0:0)</td>
<td>1</td>
<td>1.95E-06</td>
<td>1.30E-05</td>
<td>0.68 (0.58 ~ 0.80)</td>
</tr>
<tr>
<td>PC (18:2/0:0)</td>
<td>1</td>
<td>4.42E-04</td>
<td>1.61E-03</td>
<td>0.76 (0.65 ~ 0.89)</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>2</td>
<td>1.23E-03</td>
<td>3.77E-03</td>
<td>1.27 (1.10 ~ 1.47)</td>
</tr>
<tr>
<td>Inosine</td>
<td>2</td>
<td>2.24E-05</td>
<td>1.12E-04</td>
<td>1.36 (1.18 ~ 1.57)</td>
</tr>
<tr>
<td>Carnitine (14:1)</td>
<td>2</td>
<td>9.73E-05</td>
<td>3.89E-04</td>
<td>1.34 (1.16 ~ 1.56)</td>
</tr>
<tr>
<td>Glycochenodeoxycholate</td>
<td>2</td>
<td>2.53E-03</td>
<td>7.22E-03</td>
<td>1.25 (1.08 ~ 1.44)</td>
</tr>
<tr>
<td>PC (P-18:0/18:3)</td>
<td>2</td>
<td>1.88E-02</td>
<td>4.06E-02</td>
<td>1.19 (1.03 ~ 1.38)</td>
</tr>
</tbody>
</table>
Table S2 Pathway enrichment analysis for metabolites associated with the progression of ESCC

<table>
<thead>
<tr>
<th>Pathway Name</th>
<th>Total</th>
<th>Expected</th>
<th>Hits</th>
<th>Raw p</th>
<th>-log10(p)</th>
<th>Holm P</th>
<th>FDR</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptophan metabolism</td>
<td>41</td>
<td>0.19</td>
<td>2</td>
<td>0.01</td>
<td>1.88</td>
<td>1</td>
<td>0.75</td>
<td>0.16</td>
</tr>
<tr>
<td>Aminoacyl-tRNA biosynthesis</td>
<td>48</td>
<td>0.22</td>
<td>2</td>
<td>0.02</td>
<td>1.75</td>
<td>1</td>
<td>0.75</td>
<td>0.00</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>65</td>
<td>0.29</td>
<td>2</td>
<td>0.03</td>
<td>1.50</td>
<td>1</td>
<td>0.89</td>
<td>0.02</td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td>16</td>
<td>0.07</td>
<td>1</td>
<td>0.07</td>
<td>1.15</td>
<td>1</td>
<td>1.00</td>
<td>0.22</td>
</tr>
<tr>
<td>beta-Alanine metabolism</td>
<td>21</td>
<td>0.09</td>
<td>1</td>
<td>0.09</td>
<td>1.04</td>
<td>1</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Tyrosine metabolism</td>
<td>42</td>
<td>0.19</td>
<td>1</td>
<td>0.18</td>
<td>0.76</td>
<td>1</td>
<td>1.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Primary bile acid biosynthesis</td>
<td>46</td>
<td>0.21</td>
<td>1</td>
<td>0.19</td>
<td>0.72</td>
<td>1</td>
<td>1.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>