Metabolomics (2016) 12:116
DOI 10.1007/s11306-016-1050-5

ORIGINAL ARTICLE

Serum metabolomics for early diagnosis of esophageal squamous cell carcinoma by UHPLC-QTOF/MS

Jialin Wang[1,4] · Tao Zhang[2] · Xiaotao Shen[3] · Jia Liu[2,5] · Deli Zhao[6] · Yawen Sun[1,4] · Lu Wang[2] · Yingjun Liu[2] · Xiaoyun Gong[2] · Yanxun Liu[2] · Zheng-Jiang Zhu[3] · Fuzhong Xue[2]

Received: 12 March 2016 / Accepted: 27 May 2016 / Published online: 23 June 2016

Abstract
Introduction Previous metabolomics studies have revealed perturbed metabolic signatures in esophageal squamous cell carcinoma (ESCC) patients, however, most of these studies included mainly late-staged ESCC patients due to the difficulties of collecting the early-staged samples from asymptotic ESCC subjects.
Objectives This study aims to explore the early-staged ESCC metabolic signatures and potential of serum metabolomics to diagnose ESCC at early stages.
Methods Serum samples of 97 ESCC patients (stage 0, 39 cases; stage I, 17 cases; stage II, 11 cases, stage III, 30 cases) and 105 healthy controls (HC) were enrolled and randomly separated into training data (77 ESCCs, 84 HCs) and validation data (20 ESCCs, 21 HCs). Untargeted metabolomics was performed to identify ESCC-related metabolic signatures.
Results The global metabolomics profiles could clearly distinguish ESCC from HC in training data. 16 ascertained metabolites were found to be disturbed in the metabolic pathways characterized by dysregulated fatty acid biosynthesis, glycerophospholipid metabolism, choline metabolism in cancer and linoleic acid metabolism. The AUC value in validation data was 0.895, with sensitivity 85.0 % and specificity 90.5 %. Good diagnostic performances were also achieved for early stage ESCC, with the values of area under the curve (AUC) 0.881 for the ESCC patients in both stage 0 and I–II. In addition, six metabolites were found to discriminate ESCC stages. Among them, three biomarkers, dodecanoic acid, LysoPA(18:1), and LysoPC(14:0), exhibited clear trend for ESCC progression.
Conclusion These findings suggest serum metabolomics, performed in a minimally noninvasive and convenient manner, may possess great potential for early diagnosis of ESCC patients.

Keywords Esophageal squamous cell carcinoma · Metabolomics · Early diagnosis · UHPLC-QTOF/MS · Biomarker

Fuzhong Xue is the primary corresponding author.

J. Wang, T. Zhang and X. Shen made equally contributions.

Electronic supplementary material The online version of this article (doi:10.1007/s11306-016-1050-5) contains supplementary material, which is available to authorized users.

✉ Jialin Wang
wangjialin6681@sina.com

✉ Zheng-Jiang Zhu
jiangzhu@sioc.ac.cn

✉ Fuzhong Xue
xuefzh@sdu.edu.cn

[1] The Shandong Cancer Hospital Affiliated to Shandong University, Jinan 250117, China

[2] Department of Epidemiology and Biostatistics, School of Public Health, Shandong University, PO Box 100, Jinan 250012, China

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

[4] Shandong Cancer Hospital and Institute, Shandong Academy of Medical Sciences, Jinan 250117, China

[5] Yanjing Medical College, Capital Medical University, Beijing 101300, China

[6] Tumor Preventative and Therapeutic Base of Shandong Province, Feicheng People's Hospital, Feicheng 271600, China

1 Introduction

Esophageal squamous cell carcinoma (ESCC) remains the most predominant type of esophageal cancer in the developing world and an important health problem in high-risk areas, such as the Asian belt including Turkey, northeastern Iran, Kazakhstan, and northern and central China (Zhang et al. 2012a, b; Pennathur et al. 2013). Diagnosis of ESCC at early stages (0/I/II) is difficult because it is often asymptotic. Most of ESCC cases are diagnosed in late stages (III/IV), resulting in poor prognosis with the overall 5-year survival rate of only 9 %. If detected in early stages (such as stage 0, I and II), the 5-year survival rates are considerably improved to 47–83 % (Morimoto et al. 2010). Currently, early diagnosis of ESCC is primarily based on endoscopy, biopsy and pathological examination (Zhao et al. 2012). For example, endoscopy with iodine staining has been accepted as a population-level ESCC screening program in the high-risk areas of China (Guanrei and Songliang 1987; Roshandel et al. 2013). However, this method is expensive for the general population, and the screening is invasive that patients' compliance is relatively low (Dong et al. 2002; Yang et al. 2012). New cost-effective and non-invasive biomarkers are in great need for early diagnosis of ESCC.

Altered metabolism has recently been acknowledged as a key hallmark of cancer and metabolism-focused research has received renewed attention recently (Hanahan and Weinberg 2011; Vermeersch and Styczynski 2013). Metabolomics, the global semi-quantitative assessment of endogenous small molecule metabolites within a biological system, has been successfully utilized in cancer biomarker discovery (Spratlin et al. 2009; Zhang et al. 2013, b; Ke et al. 2015). Up to now, a few ESCC metabolomics studies has been reported using specimen of serum (Djukovic et al. 2010; Zhang et al. 2011; Ikeda et al. 2012; Zhang et al. 2012a, b, 2013, b; Jin et al. 2014; Mir et al. 2015), plasma (Hasim et al. 2012; Liu et al. 2013; Xu et al. 2013; Ma et al. 2014), urine (Davis et al. 2012; Hasim et al. 2012), tissue (Wu et al. 2009; Yakoub et al. 2010; Wang et al. 2013; Yang et al. 2013; Lynam-Lennon et al. 2014) or gastric content (Kumar et al. 2012), and conducted on mass spectrometry (MS) and nuclear magnetic resonance (NMR) platforms. These studies revealed significantly perturbed metabolic expression in ESCC patients compared with healthy controls. However, most of these studies had limited sample size of early-staged ESCC patients, due to the difficulties of collecting the samples from asymptotic ESCC subjects. Therefore, the disturbance of metabolic pathways in the early stages of ESCC remains unclear, and the potential of metabolomics for early diagnosis of ESCC still need further investigation.

In this study, serum samples of 97 ESCC cases and 105 healthy controls were collected at the Esophageal Cancer Screening Base in high-risk area of China, notably including 67 precious serum samples of early stages of ESCC (stage 0, 39 patients; stage I, 17 patients; stage II, 11 patients). An untargeted metabolomics study using ultra high performance liquid chromatography quadruple time-of-flight mass spectrometry (UHPLC-QTOF/MS) technique was performed to identify ESCC-related metabolic signatures. Metabolic pathways and biological relevance of potential biomarkers were intensively studied to gain insights into disturbed metabolism of ESCC at early stages. The potential metabolic biomarkers were evaluated on the validation data. In addition, the biomarkers related ESCC clinical staging and progression were investigated as well.

2 Materials and methods

2.1 Reagents and materials

LC–MS grade water (H_2O), acetonitrile (ACN), methanol (MeOH), 0.1 % formic acid (FA) in water and 0.1 % FA in ACN were purchased from Honeywell (Muskegon, MI, USA). Ammonium fluoride (NH_4F) was purchased from Sigma-Aldrich (St. Louis, USA) and dissolved in LC–MS grade water prior to use.

2.2 Serum sample collection

The study was approved by the Ethics Committee of the Shandong Tumor Hospital and written informed consent was obtained from all participants involved in this study. Serum samples from 97 ESCC cases and 105 healthy controls were collected at the Esophageal Cancer Screening Base of Shandong Province (City of Feicheng, Shandong, China) between June, 2013 and September, 2014. All the participants enrolled in this base with ages 40–69 years were screened for esophageal cancer using endoscopy with mucosal iodine staining (Dawsey et al. 1994, 1998). In this study, the participants with normal esophageal mucosa (iodine-positive) were regarded as the health controls. Meanwhile, biopsies of the iodine-negative participants were taken from the non-staining area of the mucosa, which were then underwent pathological evaluation to confirm and stage by two pathologists. Each ESCC patient was diagnosed and staged according to the American Joint Committee on Cancer (AJCC) TNM Classification of Carcinoma of the Esophagus and Esophagogastric Junction (7th edition, 2010) (Rice et al. 2010): stage 0 (tumor in situ, Tis), 39 patients; stage I, 17 patients; stage II, 11 patients; stage III, 30 patients. The enrolled subjects were randomly

separated into training subject group (77 ESCC cases, 84 healthy controls) and validation subject group (20 ESCC cases, 21 healthy controls). Detailed baseline and histopathologic characteristics for these patients are listed in Table S2.

Participants involved in this study did not take any medications, surgery, radiotherapy or chemotherapy, and those suffering from metabolic diseases, liver diseases, kidney diseases or any other cancers were excluded. All of the participants were in an overnight fasting state and 5 mL of peripheral venous blood was taken in the morning. The blood was then allowed to clot for 30 min at 37 °C water batch and followed by centrifugation at 3000 rpm for 15 min. Then the serum supernatant was taken, immediately freezed in liquid nitrogen, and stored at −80 °C until further analyses.

2.3 Sample preparation

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.

2.4 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 NH_4F 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 Supplement materials).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.

2.5 Data preprocessing and annotation

MS raw data (.d) files were converted to the mzXML format using ProteoWizard, and processed by R package XCMS (version 3.2). The preprocessing results generated a data matrix that consisted of the retention time (RT), mass-to-charge ratio (*m/z*) values, and peak intensity. R package CAMERA was used for peak annotation after XCMS data processing (Kuhl et al. 2012). 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 (Dunn et al. 2011). In briefly, a LOESS regression model was built based on the intensity drift of each metabolite in the QC samples and was used to predict and correct intensities of the same metabolite in subject samples. In addition, the unbalanced age and gender between ESCC and healthy controls are potential confounding factors for statistical analyses, therefore, the relative intensity of each metabolite was further adjusted for age, gender by regression residual analyses and then standardized with Z-transformation (mean = 0, SD = 1).

2.6 Statistical analyses

Principal component analysis (PCA) was first used to reduce the dimensionality of the multidimensional dataset,

while giving a comprehensive view of the clustering trend for the multidimensional data. Partial least-squares discriminant analysis (PLS-DA) was then used to understand global metabolic changes between ESCC and healthy controls, and corresponding variable importance in the projection (VIP values) were calculated in PLS-DA model as well. Validation plot was used to assess the validity of PLS-DA model by comparing the goodness of fit (R2 and Q2) of the PLS-DA models with the goodness of fit of 100 Y-permutated models. Meanwhile, the nonparametric Kruskal–Wallis rank sum test 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. Potential metabolic biomarkers were selected with VIP value more than 1 and FDR value less than 0.05. To evaluate the classification performance, the area under the receiver operating characteristic curve (AUC) was computed using the pROC package in the R platform. All of the statistical analyses were performed on the R platform (version 3.1.3), with the exception that PLS-DA was performed using SIMCA 14.0 (Umetrics AB, Umea, Sweden).

3 Results

The typical UHPLC-QTOF/MS chromatograms of serum samples for ESCC and healthy controls in positive (ESI+) and negative (ESI−) modes are shown in Supplement Fig. S1. In total, 1466 variables (metabolite peaks) were selected in the final data table for subsequent analyses (981 peaks in ESI+ mode, and 485 peaks in ESI− mode).

3.1 Metabolic profiles of ESCC and healthy controls

The PCA performed on the whole samples reveals that the QC samples are tightly clustered in PCA score plots (shown in Supplement Fig. S2), indicating the good analytical reproducibility of this metabolomics study. The PCA score plot also demonstrated a tendency of difference in metabolic profiles between ESCC and healthy controls.

PLS-DA was then used to understand global metabolic changes between ESCC and healthy controls in the training data. As shown in Fig. 1a, the PLS-DA score plot reveals a clear separation between ESCC subjects and healthy controls, with good fitting and predictive performances ($R2Y = 0.569$, $Q2Y = 0.523$). And the validation plot presented in Fig. 1b strongly supports the validity of the PLS-DA model, since the Q2 regression line (blue color) has a negative intercept and all permuted R2-values (green color in the left) are lower than the original point of the R2-value (green color in the right). Remarkably, excellent separations are also achieved using PLS-DA analyses for different cancer stages versus healthy controls, such as stage 0/Tis vs. healthy controls, stage I-II vs. healthy controls, and stage III vs. healthy controls (as shown in Fig. 1c–e).

3.2 Discovery and identification of potential biomarkers

A total of 439 differential peaks were selected with VIP value more than 1 and FDR value less than 0.05, including 343 peaks in ESI+ and 96 peaks in ESI−. According to the identity check by raw data and the peaks annotation by CAMERA, MS/MS experiments and standard compounds, 16 metabolites were ascertained as potential biomarkers to discriminate ESCC from healthy controls (11 and 5 biomarkers in ESI+ and ESI− modes respectively), as show in Table 1 and Supplement Table S2. The RSD of these 16 metabolites varied from 1.4 to 36.2 % with a median of 6.2 %, indicating the robustness of the metabolomics platform. In addition, the fold-change values, varied from 1.28 to 3.18, are apparently higher their values of RSD, indicating the magnitude of the changes of these metabolites is significant for future clinical analysis.

These metabolite biomarkers were identified through matching accurate mass and tandem MS/MS spectra with in-house metabolite tandem MS/MS databases. Procedures for metabolite identification and the detailed information of 16 serum metabolites are shown in the Supplement materials, Table S2, Fig. S3 and S4. Among then, eight biomarkers, including dodecanoic acid, cis-9-Palmitoleic acid, palmitic acid, oleic Acid, cortisol, L-Tyrosine, L-Tryptophan, and linoleic acid, were further confirmed using purchased standard references (shown in Supplement Table S3 and Fig. S3).

These relative intensity of 16 differential metabolites in ESCC patients versus healthy controls are summarized in Supplement Fig. S5, and mapped into their biochemical pathways through metabolic enrichment and pathway analysis based on database search (KEGG, http://www.genome.jp/kegg/) and MetaboAnalyst. The overview of pathway analysis was shown in Supplement Fig. S6. It revealed evident disorders in fatty acid biosynthesis, glycerophospholipid metabolism, choline metabolism in cancer, and linoleic acid metabolism (shown in Table 1).

3.3 Differential metabolites to discriminate ESCC stages

Figure 2 demonstrates a clear separation trend among different stages of ESCC (stage 0/Tis, I–II, III). To further determine the significant metabolites discriminating ESCC stages, the nonparametric Kruskal–Wallis rank sum test was performed and a value of FDR less than 0.05 was

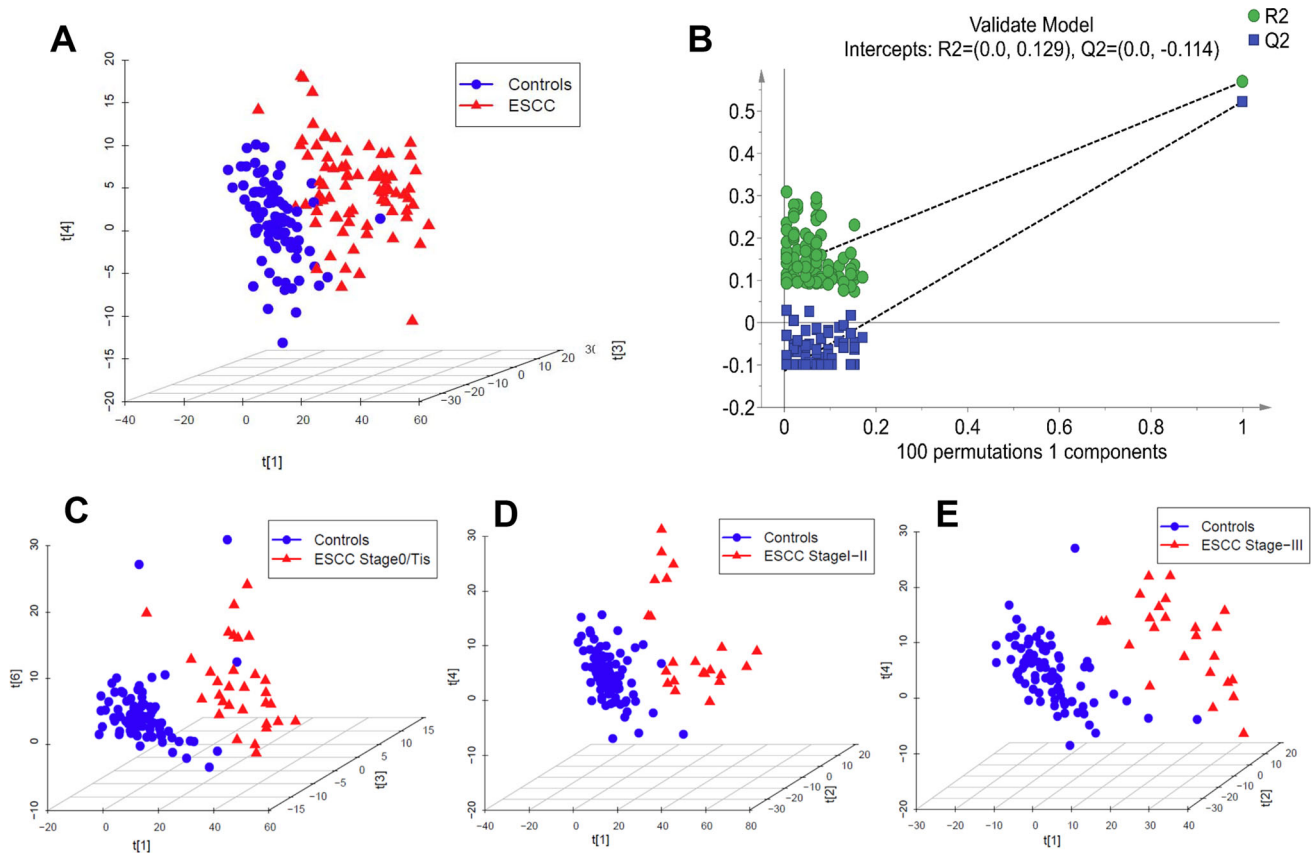

Fig. 1 **a** PLS-DA three-dimensional scores plot in the training data discriminates ESCC and healthy controls, with fitting and predictive performance (1 latent variables, R2Y = 0.569, Q2Y = 0.523); **b** validation plot obtained from 100 permutation tests; and the score plots of PLS-DA to discriminate. **c** ESCC stage 0/Tis vs healthy controls (1 latent variables, R2Y = 0.599, Q2Y = 0.542); **d** ESCC stage I–II vs healthy controls (6 latent variables, R2Y = 0.976, Q2Y = 0.607); **e** ESCC stage III vs healthy controls (5 latent variables, R2Y = 0.960, Q2Y = 0.746)

considered to be statistically significant. A total of 6 metabolites out of the above 16 potential biomarkers were identified as differential metabolites to discriminate ESCC stages (shown in Fig. 3). Notably, three biomarkers, dodecanoic acid, LPA(18:1/0:0), and LysoPC(14:0/0:0), exhibited clear trend for ESCC progression. In addition, compared with the healthy controls, the scaled intensity of palmitic acid, oleic Acid, and linoleic acid shown up-regulated expressions in stage 0/Tis other than stage I–II and III ESCC patients.

3.4 Validation and ROC analysis of potential biomarkers

The diagnostic potential of these metabolic biomarkers for ESCC, especially at early stages, was evaluated in the validation data. As shown in Fig. 4a, the heat map of the 16 biomarkers demonstrated clear differential metabolic profiles between ESCC and healthy controls in the validation data.

Furthermore, random forest model (RF), was constructed in R package randomForest composed of 16 biomarkers based on the data of training group. Then the RF was used to predict the class of subjects in the validation group. The AUC was 0.895 (95 % confidence interval, CI 0.783–1.000), with sensitivity of 85.0 % (95 % CI 70.0–100.0 %) and specificity of 90.5 % (95 % CI 76.2–100.0 %). Notably, the RF model based on 16 biomarkers exhibited good diagnostic performance in early stage ESCC (shown in Fig. 4b). The values of AUC were 0.881, 0.881 and 0.929 for the ESCC in the stage 0, I–II and III, respectively.

4 Discussion

In this study, an untargeted metabolomics study based on UHPLC-QTOF/MS technique was performed to investigate dysregulated metabolic signatures in serum sample of ESCC patients. In accordance with previous ESCC

Table 1 Serum metabolomics biomarkers discriminate ESCC from healthy controls in the training data

Pathway	Identity name	ESCC vs healthy controls					RSD%[b]
		VIP	FDR	AUC	Direction[c]	FC[c]	
Fatty acid biosynthesis	Dodecanoic acid[a]	1.00	6.7E−5	0.69	Down	1.28	3.2
	cis-9-Palmitoleic acid[a]	1.01	6.3E−4	0.67	Up	1.49	10.6
	Palmitic acid[a]	1.00	0.012	0.63	Up	1.63	7.0
	Oleic acid[a]	1.23	9.7E−5	0.69	Up	1.98	6.3
–	Cortisol[a]	1.44	3.5E−8	0.76	Up	2.57	3.3
–	L-Tyrosine[a]	1.00	0.020	0.62	Down	1.54	6.3
	L-Tryptophan[a]	1.05	5.1E−6	0.72	Down	1.85	1.4
Glycerophospholipid	LPA(18:1/0:0)	1.26	5.9E−7	0.74	Down	2.38	6.2
Metabolism, and choline	LysoPC(14:0/0:0)	1.13	8.1E−7	0.74	Down	2.04	5.2
Metabolism in cancer	LysoPC(18:2)	1.72	8.8E−13	0.84	Up	3.11	1.6
	LysoPC(24:0)	1.46	6.1E−13	0.85	Up	2.20	8.6
	LysoPC(18:4)	1.63	3.1E−12	0.83	Up	2.11	1.8
Glycerophospholipid	PC(14:1/P-18:1)	1.71	7.5E−13	0.85	Up	3.18	5.2
Metabolism	PC(16:0/18:2)	1.06	1.9E−6	0.73	Down	2.17	36.2
Linoleic acid metabolism	PC(24:1/22:6)	1.55	4.2E−13	0.85	Up	2.73	7.4
Linoleic acid metabolism	Linoleic acid[a]	1.06	3.9E−4	0.67	Up	1.50	10.2

[a] These compounds were confirmed using standard references

[b] RSD % for QC samples

[c] Direction "up" indicates a relative high concentration present in ESCC patients, and FC (fold change) was calculated from the ratio of the mean values of ESCCs relative to healthy controls; while "down" means a relative low concentration compared to the healthy controls, and FC was calculated from the ratio of the mean values of healthy controls relative to ESCCs

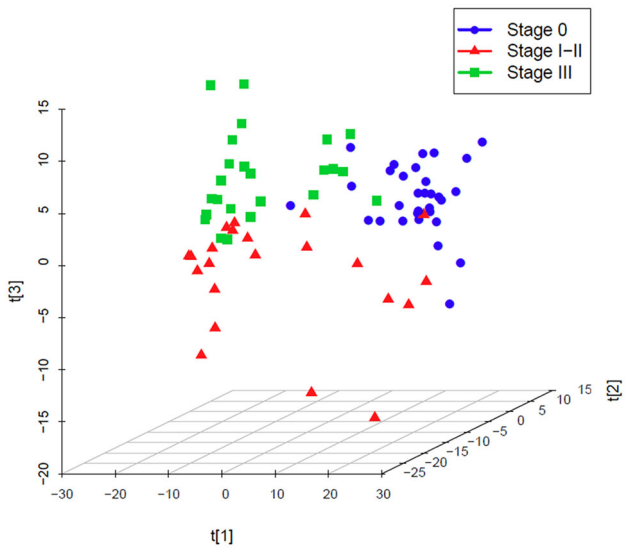

Fig. 2 PLS-DA three-dimensional scores plot in the training data discriminates different ESCC stages (3 latent variables, R2Y = 0.351, Q2Y = 0.185)

metabolomics studies, our finding demonstrated good discrimination between ESCC and healthy controls. A total of 16 serum metabolites were identified as potential biomarkers for early diagnosis of ESCC, while 6 biomarkers were further screened and selected to discriminate ESCC stages and 3 metabolites shown clear trend of ESCC progression. Among them, metabolites (oleic Acid, LysoPC(14:0/0:0), PC(24:1/22:6), PC(16:0/18:2), L-Tyrosine, L-Tryptophan, palmitic acid, and linoleic acid) have been previously reported to be disturbed in ESCC patients (Xu et al. 2013; Mir et al. 2015). Our study further confirmed these metabolite biomarkers in the early stages of ESCC. This is the first time to discover that three biomarkers, dodecanoic acid, LPA(18:1/0:0), and LysoPC(14:0/0:0), have decreased trend with ESCC progression, and may potentially be used for the prediction of cancer stage and progression.

A series of metabolites were found dysregulated in the pathways of glycerophospholipid metabolism and choline metabolism in cancer, including 4 lysophosphatidylcholines (LysoPC), 3 phosphatidylcholines (PC) and 1 lysophosphatidic acid (LPA), as shown in Fig. 5a. Among them, 3 LysoPCs and 2 PCs were up-regulated in ESCC patients compared with healthy controls (as shown in Fig. S5). Abnormal choline metabolism has been well-documented as a metabolic hallmark associated with oncogenesis and tumor progression (Glunde et al. 2006, 2011). In cancer, the choline metabolite profile is characterized by an elevation of PCs and total choline-containing compounds. The increased PC levels have been

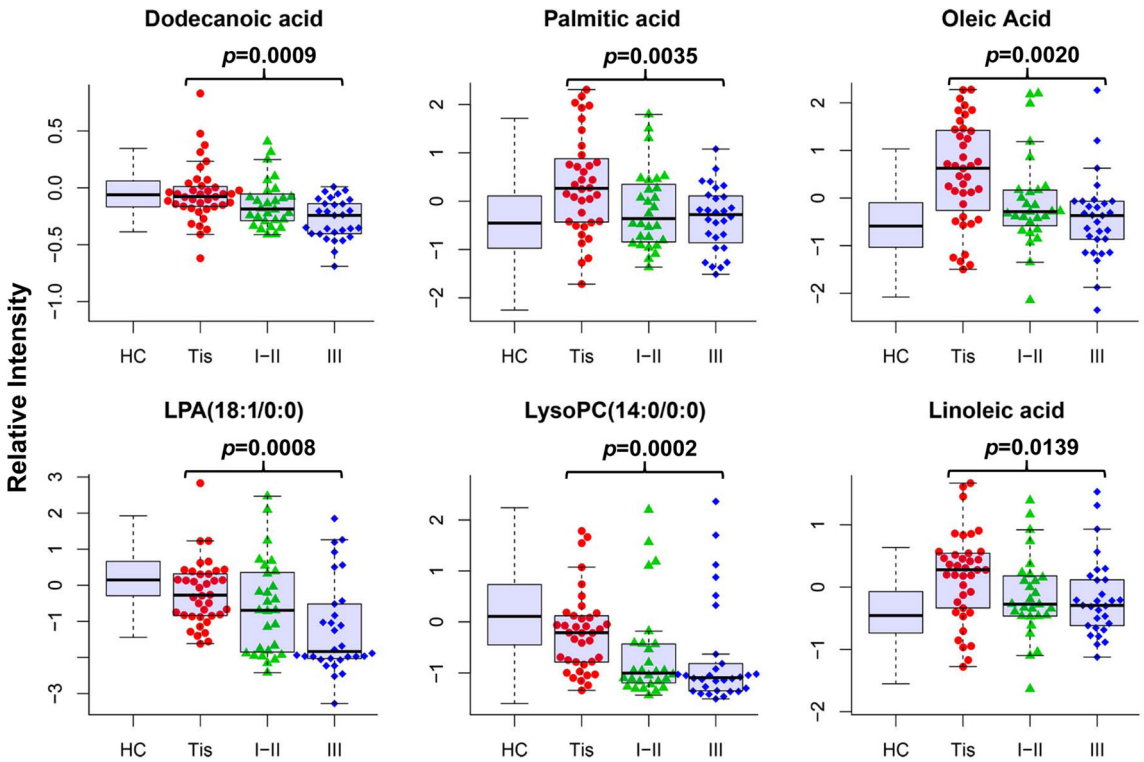

Fig. 3 Changing patterns of differential metabolites from healthy controls (HC) to different stages of ESCC (stage 0/Tis, I–II, III)

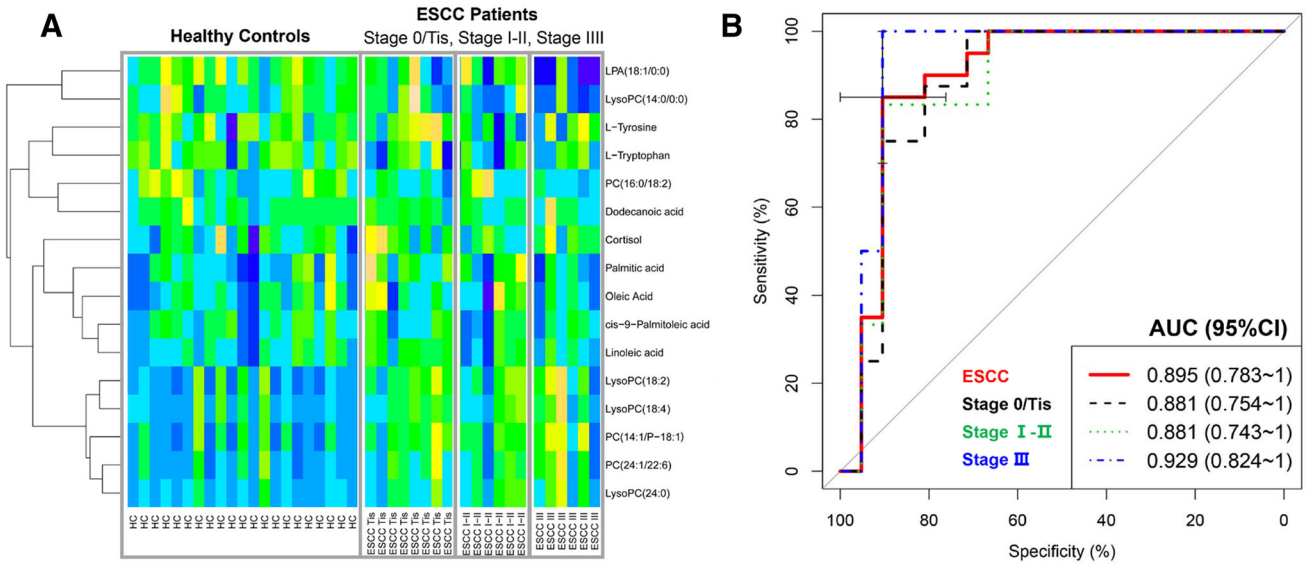

Fig. 4 **a** Heat map in the validation data. Each *row* represents a metabolite feature and each column represents a sample. **b** ROC analysis of random forest model combing 16 biomarkers to diagnose ESCC (stage 0, I–II and III) in the validation data

reported in ovarian cancer (Iorio et al. 2005), breast cancer (Eliyahu et al. 2007), and in ESCC from a recent study (Mir et al. 2015), which might be interpreted as a requirement for the high rate of cell proliferation. Notably, gradient decreases in the concentration of LysoPC(14:0/0:0) and LPA(18:1/0:0) in the pathway of choline metabolism in cancer exhibited clear trends for ESCC progression. In line with our study, Xu et al. also reported a decreased LysoPC(14:0/0:0) in the plasma of ESCC patients (Xu et al. 2013). LysoPCs can be transformed into LPA in a reaction catalyzed by lysophospholipase D (Lyso-PLD), and significant diminishment of LysoPCs might be

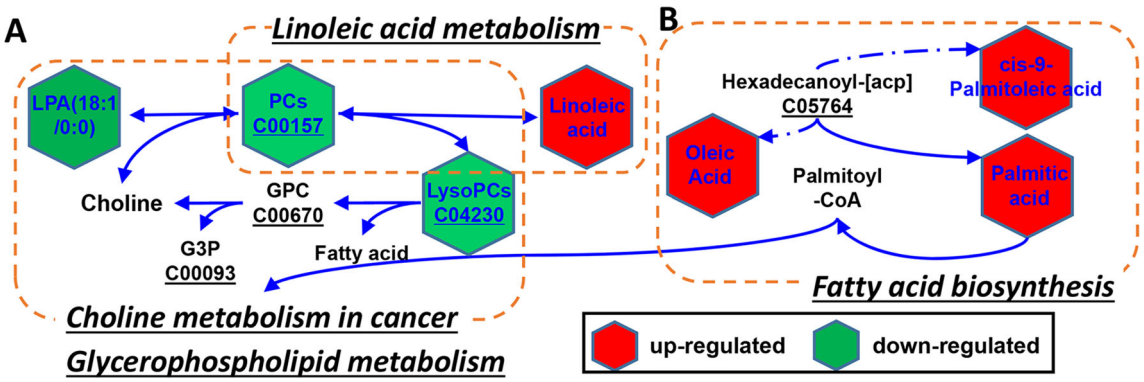

Fig. 5 Deregulated pathway in ESCC patients. **a** Dysregulated choline metabolism in cancer and Linoleic acid metabolism; **b** abnormal Fatty acid biosynthesis

correlated with the overexpression of Lyso-PLD. Elevated PCs are accompanied by a decrease in glycerophosphocholine (GPC), resulting in a decreased GPC/PC ratio compared with normal tissue (Aboagye and Bhujwalla 1999; Iorio et al. 2005).

Elevated linoleic acid metabolism was observed in ESCC patients, including linoleic acid and 3 PCs. Xu et al. also reported an increased linoleic acid in the plasma of ESCC patients (Xu et al. 2013). Linoleic acid is an essential polyunsaturated fatty acid in human nutrition and used in the biosynthesis of cell membranes, and replacement of saturated fat with linoleic acid has been advocated to improve serum lipid profiles (Zock and Katan 1998). Linoleic acid has been linked to the development of cancer in animals (Banni et al. 1999). Linoleic acid is prone to oxidation, which may play a role in carcinogenesis (Fang et al. 1996) and may increase the susceptibility of lipid particles to oxidative modification (Steinberg et al. 1989). Meta-analyses of a series of epidemiologic and experimental literatures suggest that high intake of linoleic acid substantially raises the risks of breast, colorectal, and prostate cancer (Zock and Katan 1998).

Significantly increased fatty acid biosynthesis was detected in the serum of ESCC patients compared with healthy controls (as shown in Fig. 5b). This agrees with the previous studies that found elevated level of palmitic acid and oleic acid in blood of ESCC (Xu et al. 2013; Jin et al. 2014), ovarian (Shen et al. 2001) and prostate cancer (Crowe et al. 2008; Chow 2009). Palmitic acid can be synthesized from other fatty acids and is the major fatty acid produced by de novo lipogenesis from acetyl CoA and malonyl CoA by the enzyme fatty acid synthase. Palmitic acid can be converted to palmitoyl-CoA, then incorporated mainly into glycerophospholipid metabolism.

The detection of the early, asymptomatic noninvasive stage ESCC has a profound impact on clinical outcome. However, the samples of early stages ESCC is very precious and difficult to collect because it is often asymptotic. Base on the Esophageal Cancer Screening Base of Feicheng (one of ESCC high-risk area of China), we are able to collect serum samples in the early ESCC stages (stage 0/Tis, 39 patients; stage I, 17 patients; stage II, 11 patients). Compared with previous studies, the current metabolomics study focused on early stage ESCC diagnosis. Our study supports that serum metabolomics signatures could be used to detect ESCC at early stages. The biomarkers identified in this study exhibited satisfactory diagnostic performance, especially in early stages of ESCC, with the values of AUC 0.881, 0.881, and 0.929 for stage 0/Tis, I–II and III, respectively. Nevertheless, considering the ESCC prevalence as low as less than 1/1000 in city of Feicheng, the positive predictive value (PPV) may be not acceptable for the ESCC screening in the general population (shown in Table S4) (Pearl 2002; Lin et al. 2013). If ESCC prevalence is assumed to be 1/100 in high-risk population or in clinical practice, the PPV is 8.27 % corresponding to sensitivity 85.00 % and specificity 90.48 %. Therefore, these serum metabolomics biomarkers may be applicable for ESCC screening in a high-risk population or early diagnosis in clinical practice.

The advantages of this study include: This study included 67 precious serum samples of early stages of ESCC (stage 0, 39 patients; stage I, 17 patients; stage II, 11 patients). In addition, all the potential biomarkers were identified and matched by MS/MS spectra, and eight biomarkers were confirmed by metabolite standard references. This study also has some limitations. First, the age distribution between ESCC and healthy controls was slightly unbalanced, and we don't have information on other potential confounding factor besides age and gender. We also showed the comparison of VIP, FDR, AUC and fold-change values for the selected biomarkers between age and gender-adjusted and unadjusted data in Supplement Fig. S7. Although these statistical values in adjusted

data were slightly lower than that in unadjusted data, the results in adjusted and non-adjusted data are quite consistent with each other. Second, potential biomarkers were preliminarily validated in validation data, however, the sample size of validation data was relatively small, especially for the stage specific analyses. Further investigations consisting of large-scale external cohort are needed to validate the clinical utility of biomarkers discovered in this study.

5 Concluding remarks

This serum metabolomics study was performed to identify ESCC early diagnostic metabolic signatures by including 67 precious serum samples from early stages ESCC patients. The findings suggest serum metabolomics may possess great potential for diagnosis of asymptomatic noninvasive early stages of ESCC patients. Metabolic pathways in ESCC patients, mainly characterized by dysregulated fatty acid biosynthesis, glycerophospholipid metabolism and linoleic acid metabolism, are found to be related to ESCC early stage development and progression.

Acknowledgments This work was funded by National Natural Science Foundation of China (Grant Number 81573246, 81302514, 81573259, 21575151), Science and Technology Research Projects of Shandong Province (2014GSF118022), and National Natural Science Foundation of Shandong Province (RZ2014HP044). Z.-J. Z. is supported by Thousand Youth Talents Program from Chinese government and Agilent Technologies Thought Leader Award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Human and animal rights All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Aboagye, E. O., & Bhujwalla, Z. M. (1999). Malignant transformation alters membrane choline phospholipid metabolism of human mammary epithelial cells. *Cancer Research, 59*, 80–84.

Banni, S., Angioni, E., Casu, V., et al. (1999). Decrease in linoleic acid metabolites as a potential mechanism in cancer risk reduction by conjugated linoleic acid. *Carcinogenesis, 20*, 1019–1024.

Chow, C. K. (2009). Fatty acid composition of plasma phospholipids and risk of prostate cancer. *Am J Clin Nutr, 89*, 1946. **author reply 1946–7**.

Crowe, F. L., Allen, N. E., Appleby, P. N., et al. (2008). Fatty acid composition of plasma phospholipids and risk of prostate cancer in a case-control analysis nested within the European Prospective Investigation into Cancer and Nutrition. *American Journal of Clinical Nutrition, 88*, 1353–1363.

Davis, V. W., Schiller, D. E., Eurich, D., & Sawyer, M. B. (2012). Urinary metabolomic signature of esophageal cancer and Barrett's esophagus. *World J Surg Oncol, 10*, 271.

Dawsey, S. M., Fleischer, D. E., Wang, G. Q., et al. (1998). Mucosal iodine staining improves endoscopic visualization of squamous dysplasia and squamous cell carcinoma of the esophagus in Linxian, China. *Cancer, 83*, 220–231.

Dawsey, S. M., Lewin, K. J., Wang, G. Q., et al. (1994). Squamous esophageal histology and subsequent risk of squamous cell carcinoma of the esophagus. A prospective follow-up study from Linxian, China. *Cancer, 74*, 1686–1692.

Djukovic, D., Baniasadi, H. R., Kc, R., Hammoud, Z., & Raftery, D. (2010). Targeted serum metabolite profiling of nucleosides in esophageal adenocarcinoma. *Rapid Communications in Mass Spectrometry, 24*, 3057–3062.

Dong, Z., Tang, P., Li, L., & Wang, G. (2002). The strategy for esophageal cancer control in high-risk areas of China. *Japanese Journal of Clinical Oncology, 32*(Suppl), S10–S12.

Dunn, W. B., Broadhurst, D., Begley, P., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols, 6*, 1060–1083.

Eliyahu, G., Kreizman, T., & Degani, H. (2007). Phosphocholine as a biomarker of breast cancer: molecular and biochemical studies. *International Journal of Cancer, 120*, 1721–1730.

Fang, J. L., Vaca, C. E., Valsta, L. M., & Mutanen, M. (1996). Determination of DNA adducts of malonaldehyde in humans: effects of dietary fatty acid composition. *Carcinogenesis, 17*, 1035–1040.

Glunde, K., Bhujwalla, Z. M., & Ronen, S. M. (2011). Choline metabolism in malignant transformation. *Nature Reviews Cancer, 11*, 835–848.

Glunde, K., Jacobs, M. A., & Bhujwalla, Z. M. (2006). Choline metabolism in cancer: implications for diagnosis and therapy. *Expert Review Molecular Diagnostics, 6*, 821–829.

Guanrei, Y., & Songliang, Q. (1987). Endoscopic surveys in high-risk and low-risk populations for esophageal cancer in China with special reference to precursors of esophageal cancer. *Endoscopy, 19*, 91–95.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*, 646–674.

Hasim, A., Ma, H., Mamtimin, B., et al. (2012). Revealing the metabonomic variation of EC using (1)H-NMR spectroscopy and its association with the clinicopathological characteristics. *Molecular Biology Reports, 39*, 8955–8964.

Ikeda, A., Nishiumi, S., Shinohara, M., et al. (2012). Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. *Biomedical Chromatography, 26*, 548–558.

Iorio, E., Mezzanzanica, D., Alberti, P., et al. (2005). Alterations of choline phospholipid metabolism in ovarian tumor progression. *Cancer Research, 65*, 9369–9376.

Jin, H., Qiao, F., Chen, L., Lu, C., Xu, L., & Gao, X. (2014). Serum metabolomic signatures of lymph node metastasis of esophageal squamous cell carcinoma. *Journal of Proteome Research, 13*, 4091–4103.

Ke, C., Hou, Y., Zhang, H., et al. (2015). Large-scale profiling of metabolic dysregulation in ovarian cancer. *International Journal of Cancer, 136*, 516–526.

Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry, 84*, 283–289.

Kumar, S., Huang, J., Cushnir, J. R., Spanel, P., Smith, D., & Hanna, G. B. (2012). Selected ion flow tube-MS analysis of headspace vapor from gastric content for the diagnosis of gastro-esophageal cancer. *Analytical Chemistry, 84*, 9550–9557.

Lin, Y., Totsuka, Y., He, Y., et al. (2013). Epidemiology of esophageal cancer in Japan and China. *J Epidemiol, 23*, 233–242.

Liu, R., Peng, Y., Li, X., et al. (2013). Identification of plasma metabolomic profiling for diagnosis of esophageal squamous-cell carcinoma using an UPLC/TOF/MS platform. *International Journal of Molecular Sciences, 14*, 8899–8911.

Lynam-Lennon, N., Connaughton, R., Carr, E., et al. (2014). Excess visceral adiposity induces alterations in mitochondrial function and energy metabolism in esophageal adenocarcinoma. *BMC Cancer, 14*, 907.

Ma, H., Hasim, A., Mamtimin, B., Kong, B., Zhang, H. P., & Sheyhidin, I. (2014). Plasma free amino acid profiling of esophageal cancer using high-performance liquid chromatography spectroscopy. *World Journal of Gastroenterology, 20*, 8653–8659.

Mir, S. A., Rajagopalan, P., Jain, A. P., et al. (2015). LC–MS-based serum metabolomic analysis reveals dysregulation of phosphatidylcholines in esophageal squamous cell carcinoma. *Journal of Proteomics, 127*, 96–102.

Morimoto, M., Nishiyama, K., Nakamura, S., et al. (2010). Significance of endoscopic screening and endoscopic resection for esophageal cancer in patients with hypopharyngeal cancer. *Japanese Journal of Clinical Oncology, 40*, 938–943.

Pearl, D. C. (2002). Proteomic patterns in serum and identification of ovarian cancer. *Lancet, 360*, 169–170. **author reply 170–1**.

Pennathur, A., Gibson, M. K., Jobe, B. A., & Luketich, J. D. (2013). Oesophageal carcinoma. *Lancet, 381*, 400–412.

Rice, T. W., Rusch, V. W., Ishwaran, H., & Blackstone, E. H. (2010). Cancer of the esophagus and esophagogastric junction: data-driven staging for the seventh edition of the American Joint Committee on Cancer/International Union Against Cancer Cancer Staging Manuals. *Cancer, 116*, 3763–3773.

Roshandel, G., Nourouzi, A., Pourshams, A., Semnani, S., Merat, S., & Khoshnia, M. (2013). Endoscopic screening for esophageal squamous cell carcinoma. *Arch Iran Med, 16*, 351–357.

Shen, Z., Wu, M., Elson, P., et al. (2001). Fatty acid composition of lysophosphatidic acid and lysophosphatidylinositol in plasma from patients with ovarian cancer and other gynecological diseases. *Gynecologic Oncology, 83*, 25–30.

Spratlin, J. L., Serkova, N. J., & Eckhardt, S. G. (2009). Clinical applications of metabolomics in oncology: a review. *Clinical Cancer Research, 15*, 431–440.

Steinberg, D., Parthasarathy, S., Carew, T. E., Khoo, J. C., & Witztum, J. L. (1989). Beyond cholesterol. Modifications of low-density lipoprotein that increase its atherogenicity. *New England Journal of Medicine, 320*, 915–924.

Vermeersch, K. A., & Styczynski, M. P. (2013). Applications of metabolomics in cancer research. *Journal of Carcinogenesis, 12*, 9.

Wang, L., Chen, J., Chen, L., et al. (2013). 1H-NMR based metabonomic profiling of human esophageal cancer tissue. *Mol Cancer, 12*, 25.

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

Xu, J., Chen, Y., Zhang, R., et al. (2013). Global and targeted metabolomics of esophageal squamous cell carcinoma discovers potential diagnostic and therapeutic biomarkers. *Molecular and Cellular Proteomics, 12*, 1306–1318.

Yakoub, D., Keun, H. C., Goldin, R., & Hanna, G. B. (2010). Metabolic profiling detects field effects in nondysplastic tissue from esophageal cancer patients. *Cancer Research, 70*, 9129–9136.

Yang, Y., Wang, L., Wang, S., et al. (2013). Study of metabonomic profiles of human esophageal carcinoma by use of high-resolution magic-angle spinning 1H NMR spectroscopy and multivariate data analysis. *Analytical and Bioanalytical Chemistry, 405*, 3381–3389.

Yang, J., Wei, W. Q., Niu, J., Liu, Z. C., Yang, C. X., & Qiao, Y. L. (2012). Cost-benefit analysis of esophageal cancer endoscopic screening in high-risk areas of China. *World Journal of Gastroenterology, 18*, 2493–2501.

Zhang, J., Bowers, J., Liu, L., et al. (2012a). Esophageal cancer metabolite biomarkers detected by LC–MS and NMR methods. *PLoS ONE, 7*, e30181.

Zhang, H. Z., Jin, G. F., & Shen, H. B. (2012b). Epidemiologic differences in esophageal cancer between Asian and Western populations. *Chinese Journal of Cancer, 31*, 281–286.

Zhang, J., Liu, L., Wei, S., et al. (2011). Metabolomics study of esophageal adenocarcinoma. *The Journal Thoracic and Cardiovascular Surgery, 141*, 469–475. **475.e1–4**.

Zhang, T., Wu, X., Ke, C., et al. (2013a). Identification of potential biomarkers for ovarian cancer by urinary metabolomic profiling. *Journal of Proteome Research, 12*, 505–512.

Zhang, X., Xu, L., Shen, J., et al. (2013b). Metabolic signatures of esophageal cancer: NMR-based metabolomics and UHPLC-based focused metabolomics of blood serum. *Biochimica et Biophysica Acta, 1832*, 1207–1216.

Zhao, L., Wei, W. Q., Zhao, D. L., et al. (2012). Population-based study of DNA image cytometry as screening method for esophageal cancer. *World Journal of Gastroenterology, 18*, 375–382.

Zock, P. L., & Katan, M. B. (1998). Linoleic acid intake and cancer risk: a review and meta-analysis. *American Journal of Clinical Nutrition, 68*, 142–153.