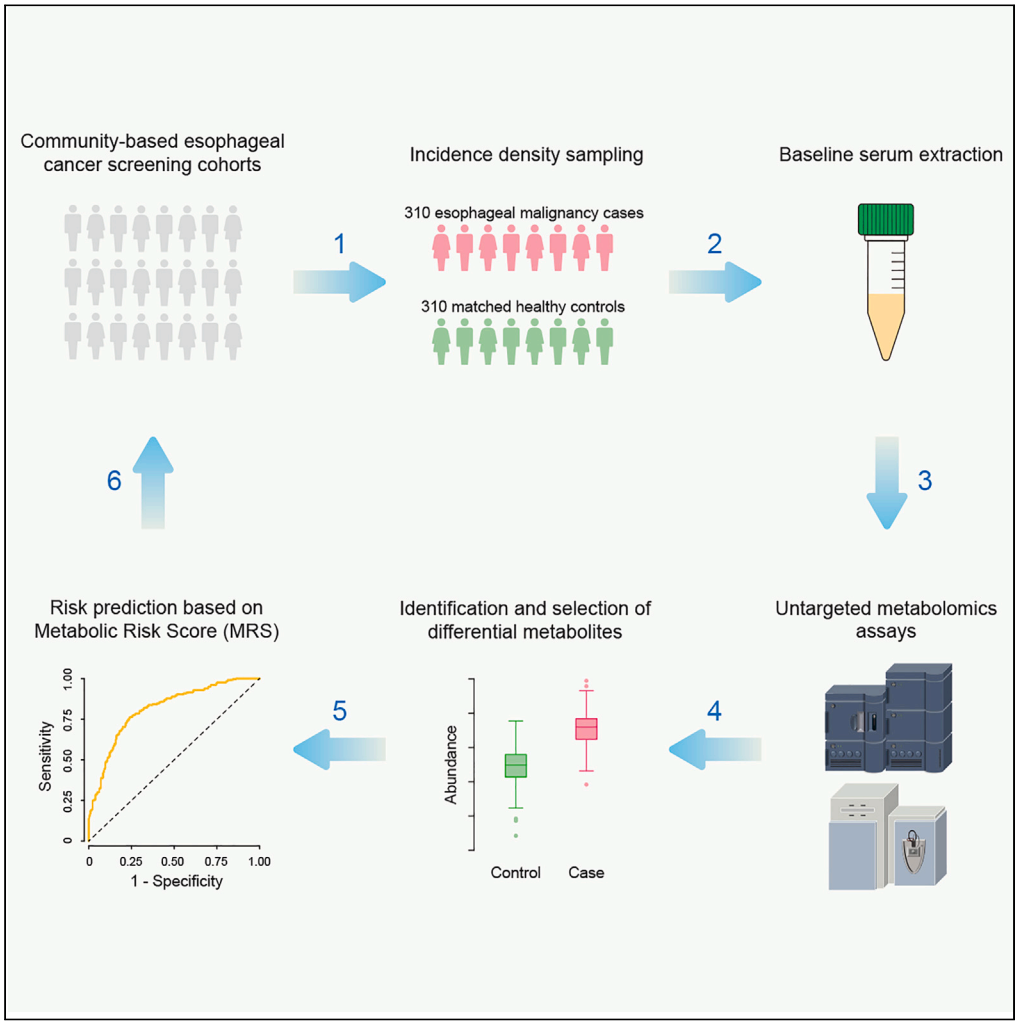
iScience

CellPress
OPEN ACCESS

Article

Construction and validation of serum Metabolic Risk Score for early warning of malignancy in esophagus



Mengfei Liu, Hongrui Tian, Minmin Wang, ..., Hong Cai, Zhonghu He, Yang Ke

zhonghuhe@foxmail.com (Z.H.)
keyang@bjmu.edu.cn (Y.K.)

Highlights

Alterations of serum metabolites are associated with esophageal malignancy

Metabolic Risk Score (MRS) could identify prevalent and short-term esophageal cancer

MRS could be applied to population-level risk-based esophageal cancer screening

Liu et al., iScience *27*, 109965
June 21, 2024
https://doi.org/10.1016/j.isci.2024.109965

iScience

CellPress
OPEN ACCESS

Article

Construction and validation of serum Metabolic Risk Score for early warning of malignancy in esophagus

Mengfei Liu,[1,11] Hongrui Tian,[1,11] Minmin Wang,[1,2] Chuanhai Guo,[1] Ruiping Xu,[3] Fenglei Li,[4] Anxiang Liu,[5] Haijun Yang,[6] Liping Duan,[7] Lin Shen,[8] Qi Wu,[9] Zhen Liu,[1] Ying Liu,[1] Fangfang Liu,[1] Yaqi Pan,[1] Zhe Hu,[1] Huanyu Chen,[1] Hong Cai,[1] Zhonghu He,[10,12,*] and Yang Ke[10,*]

SUMMARY

Using noninvasive biomarkers to identify high-risk individuals prior to endoscopic examination is crucial for optimization of screening strategies for esophageal squamous cell carcinoma (ESCC). We conducted a nested case-control study based on two community-based screening cohorts to evaluate the warning value of serum metabolites for esophageal malignancy. The serum samples were collected at enrollment when the cases had not been diagnosed. We identified 74 differential metabolites and two prominent perturbed metabolic pathways, and constructed Metabolic Risk Score (MRS) based on 22 selected metabolic predictors. The MRS generated an area under the receiver operating characteristics curve (AUC) of 0.815. The model performed well for the within-1-year interval (AUC: 0.868) and 1-to-5-year interval (AUC: 0.845) from blood draw to diagnosis, but showed limited ability in predicting long-term cases (>5 years). In summary, the MRS could serve as a potential early warning and risk stratification tool for establishing a precision strategy of ESCC screening.

INTRODUCTION

Esophageal cancer (EC) is one of the most common and lethal cancers worldwide, with 604,100 new cases and 544,076 deaths in 2020.[1] In China, EC ranks as the sixth most frequent cancer and fifth leading cause of cancer death.[2] Esophageal squamous cell carcinoma (ESCC) is the predominant subtype. As the precursor lesions for ESCC, esophageal squamous dysplasia (mild/moderate/severe dysplasia) and carcinoma *in situ* (CIS) have been reported to have a significantly increased risk of progressing into invasive cancer.[3] Since the main cause of EC has not yet been determined, secondary prevention, which can also be referred to as screening, early diagnosis, and early treatment, serves as the mainstay for ESCC prevention and control in China.

Currently, chromoendoscopy with iodine staining and indicative biopsy is recognized as the gold standard technique for the diagnosis of early ESCC and its precursor lesions. However, traditional screening strategies, which theoretically cover the entire population of a given age group in an area, may impose a heavy economic burden on both individuals and government in view of the extensive requirement for resources. Moreover, the majority of the participants would not benefit immediately from endoscopic screening due to the relatively low detection rate of esophageal malignant lesions even in high-incidence areas.[4] Because of the invasive nature of endoscopic examination, ESCC screening is also accompanied with non-negligible risk of perforation, bleeding, cardiopulmonary events, and psychological harm.

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Genetics, Peking University Cancer Hospital & Institute, Beijing 100142, China
[2]Department of Global Health, School of Public Health, Peking University, Beijing 100191, China
[3]Anyang Cancer Hospital, Anyang 455000, China
[4]Hua County People's Hospital, Anyang 456400, China
[5]Endoscopy Center, Anyang Cancer Hospital, Anyang 455000, China
[6]Department of Pathology, Anyang Cancer Hospital, Anyang 455000, China
[7]Department of Gastroenterology, Peking University Third Hospital, Beijing 100191, China
[8]State Key Laboratory of Holistic Integrative Management of Gastrointestinal Cancers, Beijing Key Laboratory of Carcinogenesis and Translational Research, Department of Gastrointestinal Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China
[9]State Key Laboratory of Holistic Integrative Management of Gastrointestinal Cancers, Beijing Key Laboratory of Carcinogenesis and Translational Research, Endoscopy Center, Peking University Cancer Hospital & Institute, Beijing 100142, China
[10]State Key Laboratory of Molecular Oncology, Beijing Key Laboratory of Carcinogenesis and Translational Research, Department of Genetics, Peking University Cancer Hospital & Institute, Beijing 100142, China
[11]These authors contributed equally
[12]Lead contact
*Correspondence: zhonghuhe@foxmail.com (Z.H.), keyang@bjmu.edu.cn (Y.K.)
https://doi.org/10.1016/j.isci.2024.109965

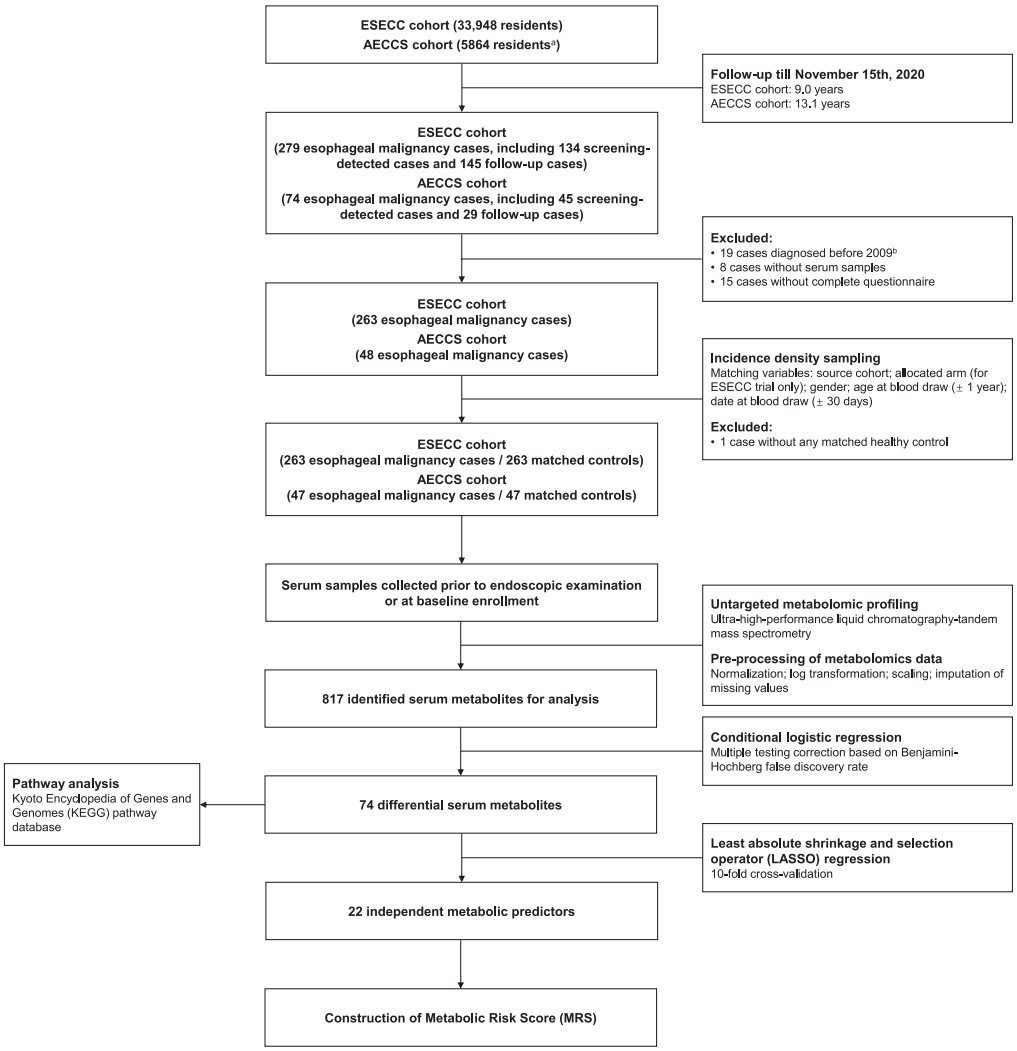


Figure 1. Flowchart of the study design and analysis

[a] Among the ten target villages (9208 enrolled residents) in the AECCS cohort, four villages where blood samples were collected only at the first round of endoscopic screening were excluded, considering the potential influence of storage time of serum samples on metabolomics detection.

[b] For the AECCS cohort, cancer cases detected at the first round of endoscopic screening (n = 19) were excluded. Abbreviations: AECCS, Anyang Esophageal Cancer Cohort Study; ESECC, Endoscopic Screening for Esophageal Cancer in China.

Use of non-invasive early warning biomarkers to identify individuals at high-risk or in high-risk subgroups prior to endoscopic examination should be the main direction of ESCC screening projects. It has increasingly been recognized that cancer is a metabolic disease, where the disturbance of cellular energy metabolism leads to dysregulated cellular proliferation.[5] Metabolomics is a systematic approach for characterizing hundreds to thousands of small molecule constituents of a biologic system (<1 kDa) in a high-throughput manner. Several studies have been conducted to assess the diagnostic value of blood metabolites for ESCC (Table S1).[6–14] However, the majority of previous studies were hospital-based, and lacked representativeness due to selection bias both from cases (cases mainly at an advanced stage) and controls (non-ESCC volunteers). Furthermore, these studies were unable to determine whether the alteration of metabolite levels preceded the occurrence of ESCC. Hence, in contrast to a hospital-based study design which is more suitable for prognostic research, prospective investigation with community-based screening cohorts is the optimal design for discovering metabolites with diagnostic and early warning value for ESCC.

Here, we conducted a nested case-control study based on two large-scale prospective screening cohorts in a region which is high-risk for ESCC in rural China. We aim to provide compelling evidence for the early warning value of serum metabolites for esophageal malignancy and to construct a Metabolic Risk Score (MRS) to identify high-risk individuals in real-world screening scenarios. The flowchart of the study design is illustrated in Figure 1.

Table 1. Baseline characteristics of cases of esophageal malignancy and 1:1 matched controls

Variables	Control n (%)	Case n (%)	*p* value[a]
N	310	310	NA
Age at blood draw			
Mean (SD)	61.8 (4.8)	61.8 (4.8)	0.625
Gender			
Female	126 (40.7)	126 (40.7)	NA
Male	184 (59.3)	184 (59.3)	
Cohort			
ESECC	263 (84.8)	263 (84.8)	NA
AECCS	47 (15.2)	47 (15.2)	
Stage			
Severe dysplasia	–	86 (27.8)	NA
Carcinoma *in situ*	–	59 (19.0)	
Squamous cell carcinoma	–	165 (53.2)	
Duration of time between blood draw and analysis (year)			
Mean (SD)	8.2 (1.7)	8.2 (1.7)	0.665
Family history of esophageal cancer			
No	282 (91.0)	249 (80.3)	<0.001
Yes	28 (9.0)	61 (19.7)	
Body mass index			
>22 kg/m^2	258 (83.2)	233 (75.2)	0.013
≤22 kg/m^2	52 (16.8)	77 (24.8)	
Cigarette smoking			
No	215 (69.4)	214 (69.0)	0.922
Yes	95 (30.6)	96 (31.0)	
Alcohol consumption			
No	264 (85.2)	254 (81.9)	0.245
Yes	46 (14.8)	56 (18.1)	

Abbreviations: AECCS, Anyang Esophageal Cancer Cohort Study; ESECC, Endoscopic Screening for Esophageal Cancer in China; NA, not applicable; SD, standard deviation.

[a]*p* values were derived using the McNemar test (categorical variables) or the Wilcoxon signed-rank test (continuous variables).

RESULTS

Participant baseline characteristics

The participant baseline characteristics are shown in Table 1. Among the esophageal malignancy cases, 86 (27.8%) were diagnosed with severe dysplasia, 59 (19.0%) with CIS, and 165 (53.2%) with squamous cell carcinoma. The distribution of age at blood draw, gender, source cohort, and duration of time between blood draw and analysis (matching variables) were similar between 310 cases and 310 controls. As compared with controls, cases were more likely to have a family history of EC (19.7% vs. 9.0%, $p < 0.001$) and lower body mass index (BMI) (≤22) (24.8% vs. 16.8%, $p = 0.013$). Among the 310 cases, 149 cases were diagnosed within the first year after blood draw (attributable to baseline screening), 112 cases were diagnosed after one year and before five years, and 49 cases were diagnosed after five years (Figure S1).

Identification of differential metabolites

A total of 74 differential metabolites were identified (false discovery rate (FDR) < 0.2 in univariate conditional logistic regression) and these were sorted into the following metabolic classes: lipid (40, 54.1%), peptide (11, 14.9%), amino acid (10, 13.5%), carbohydrate (3, 4.1%), cofactors and vitamins (3, 4.1%), nucleotide (3, 4.1%), and other classes (4, 5.4%) (Tables 2 and S2). The differential metabolites included 58 positive associations (odds ratios (ORs) ranged from 1.2 to 1.7) and 16 negative associations (ORs ranged from 0.7 to 0.8).

All 74 metabolites were mapped based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. At the FDR threshold of 0.05 for enrichment analysis and the pathway impact threshold of 0.1, two metabolic pathways were significantly associated with esophageal

Table 2. Associations of 74 differential serum metabolites with esophageal malignancy risk

Metabolite	Class of metabolism	OR per SD (95% CI)	BH-FDR
5-oxoproline	Amino Acid	1.5 (1.2,1.9)	0.019
glutamate	Amino Acid	1.4 (1.1,1.8)	0.093
methionine sulfoxide	Amino Acid	1.4 (1.1,1.8)	0.094
isovalerate (i5:0)	Amino Acid	1.4 (1.1,1.7)	0.098
beta-citrylglutamate	Amino Acid	1.3 (1.1,1.6)	0.120
N-methylproline	Amino Acid	0.8 (0.7,0.9)	0.157
S-methylmethionine	Amino Acid	0.8 (0.7,0.9)	0.164
2-oxoarginine	Amino Acid	0.7 (0.6,0.9)	0.037
N-formylmethionine	Amino Acid	0.7 (0.6,0.9)	0.114
4-methyl-2-oxopentanoate	Amino Acid	0.7 (0.5,0.9)	0.171
ribose	Carbohydrate	1.4 (1.1,1.7)	0.082
galactonate	Carbohydrate	0.8 (0.7,0.9)	0.130
pyruvate	Carbohydrate	0.8 (0.6,1.0)	0.190
4-oxo-retinoic acid	Cofactors and Vitamins	1.3 (1.1,1.6)	0.086
bilirubin (E,E)	Cofactors and Vitamins	0.8 (0.6,0.9)	0.164
alpha-tocopherol	Cofactors and Vitamins	0.7 (0.6,0.9)	0.070
alpha-ketoglutarate	Energy	0.8 (0.6,0.9)	0.179
4-cholesten-3-one	Lipid	1.7 (1.4,2.2)	0.002
glycerophosphoethanolamine	Lipid	1.7 (1.3,2.1)	0.002
heptanoate (7:0)	Lipid	1.7 (1.4,2.2)	0.002
1-(1-enyl-palmitoyl)-GPE (P-16:0)	Lipid	1.7 (1.3,2.1)	0.011
1-(1-enyl-stearoyl)-GPE (P-18:0)	Lipid	1.6 (1.3,2.0)	0.011
2-hydroxyheptanoate	Lipid	1.6 (1.3,2.1)	0.011
13-HODE + 9-HODE	Lipid	1.6 (1.2,2.0)	0.016
caproate (6:0)	Lipid	1.6 (1.2,2.0)	0.019
glycerophosphoinositol	Lipid	1.5 (1.2,1.8)	0.016
1-oleoylglycerophosphate (18:1)	Lipid	1.5 (1.2,2.0)	0.027
1-palmitoyl-GPA (16:0)	Lipid	1.5 (1.2,1.9)	0.065
pelargonate (9:0)	Lipid	1.4 (1.1,1.7)	0.049
1-stearoyl-GPI (18:0)	Lipid	1.4 (1.1,1.7)	0.098
glycerophosphoserine	Lipid	1.4 (1.1,1.7)	0.120
butyrate (4:0)	Lipid	1.4 (1.1,1.7)	0.125
1-(1-enyl-oleoyl)-GPE (P-18:1)	Lipid	1.4 (1.1,1.7)	0.130
1-stearoyl-2-oleoyl-GPS (18:0/18:1)	Lipid	1.3 (1.1,1.6)	0.048
mead acid (20:3n9)	Lipid	1.3 (1.1,1.5)	0.077
1,2-dipalmitoyl-GPC (16:0/16:0)	Lipid	1.3 (1.1,1.5)	0.078
1-oleoyl-GPC (18:1)	Lipid	1.3 (1.1,1.5)	0.098
chenodeoxycholate	Lipid	1.3 (1.1,1.5)	0.098
ximenoylcarnitine (C26:1)	Lipid	1.3 (1.1,1.6)	0.098
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	Lipid	1.3 (1.1,1.5)	0.125
1-palmitoyl-GPE (16:0)	Lipid	1.3 (1.1,1.5)	0.130
glycosyl-N-nervonoyl-sphingosine (d18:1/24:1)	Lipid	1.3 (1.1,1.5)	0.130
undecanoate (11:0)	Lipid	1.3 (1.1,1.5)	0.130
1-(1-enyl-palmitoyl)-GPC (P-16:0)	Lipid	1.3 (1.1,1.7)	0.145

(*Continued on next page*)

Table 2. *Continued*

Metabolite	Class of metabolism	OR per SD (95% CI)	BH-FDR
malonate	Lipid	1.3 (1.0,1.6)	0.188
3-hydroxystearate	Lipid	1.2 (1.0,1.5)	0.164
glycosyl-N-behenoyl-sphingadienine (d18:2/22:0)	Lipid	1.2 (1.0,1.5)	0.169
2-oleoylglycerol (18:1)	Lipid	1.2 (1.0,1.4)	0.183
glycosyl ceramide (d18:1/20:0, d16:1/22:0)	Lipid	1.2 (1.0,1.5)	0.183
palmitoyl dihydrosphingomyelin (d18:0/16:0)	Lipid	1.2 (1.0,1.5)	0.188
1-stearoyl-GPE (18:0)	Lipid	1.2 (1.0,1.5)	0.191
palmitoyl sphingomyelin (d18:1/16:0)	Lipid	1.2 (1.0,1.4)	0.191
glycerophosphorylcholine (GPC)	Lipid	1.2 (1.0,1.5)	0.199
palmitoyl-sphingosine-phosphoethanolamine (d18:1/16:0)	Lipid	1.2 (1.0,1.5)	0.199
sebacate (C10-DC)	Lipid	0.8 (0.6,0.9)	0.098
azelate (C9-DC)	Lipid	0.8 (0.6,0.9)	0.125
linoleoyl-linolenoyl-glycerol (18:2/18:3) [2]	Lipid	0.8 (0.7,1.0)	0.171
N6-methyladenosine	Nucleotide	1.4 (1.2,1.7)	0.016
2′-O-methyluridine	Nucleotide	1.3 (1.1,1.6)	0.125
dihydroorotate	Nucleotide	0.8 (0.6,0.9)	0.094
pentose acid	Partially Characterized Molecules	0.8 (0.7,0.9)	0.100
gamma-glutamylglutamate	Peptide	1.6 (1.2,1.9)	0.012
gamma-glutamyl-epsilon-lysine	Peptide	1.5 (1.2,1.9)	0.014
leucylglycine	Peptide	1.5 (1.2,1.8)	0.016
gamma-glutamylleucine	Peptide	1.4 (1.1,1.7)	0.086
gamma-glutamylisoleucine	Peptide	1.4 (1.1,1.7)	0.094
gamma-glutamylphenylalanine	Peptide	1.4 (1.1,1.7)	0.098
leucylglutamine	Peptide	1.3 (1.1,1.6)	0.060
histidylalanine	Peptide	1.3 (1.1,1.6)	0.130
gamma-glutamyl-alpha-lysine	Peptide	1.3 (1.1,1.6)	0.169
HWESASXX	Peptide	1.3 (1.1,1.7)	0.183
gamma-glutamylthreonine	Peptide	1.3 (1.1,1.6)	0.183
histidine betaine (hercynine)	Xenobiotics	1.3 (1.1,1.6)	0.184
ergothioneine	Xenobiotics	0.7 (0.6,0.9)	0.125

Abbreviations: BH-FDR, Benjamini-Hochberg false discovery rate; CI, confidence interval; OR, odds ratio; SD, standard deviation.
Note: Differential metabolites were determined using univariable conditional logistic regression with BH-FDR <0.2.

malignancy, namely, glycerophospholipid metabolism (FDR = 0.002, pathway impact = 0.25) and D-glutamine and D-glutamate metabolism (FDR = 0.046, pathway impact = 0.50) (Figure 2).

Construction of MRS

Among the 74 differential metabolites, 22 metabolites were further selected as independent predictors based on least absolute shrinkage and selection operator (LASSO) regression using the largest value of lambda (0.03170988) where binomial deviance was within one standard error of the minimum (Figure S2), and were sorted into the following metabolic classes: lipid (14, 63.6%), carbohydrate (2, 9.1%), nucleotide (2, 9.1%), amino acid (1, 4.5%), cofactors and vitamins (1, 4.5%), partially characterized molecules (1, 4.5%), and peptide (1, 4.5%). A subset of the 74 differential metabolites exhibited relatively high correlations (Spearman correlation coefficient: ≥ 0.6 or ≤ -0.6); however, the subsequent selection through LASSO regression ensured that there were no high correlations among the 22 metabolic predictors (Figure S3).

Following the algorithm described in the STAR Methods section, MRS was established based on the 22 independent metabolites (Figure 3). The distribution of MRS for esophageal malignancy cases was significantly distinct from that for healthy controls (Student's *t* test

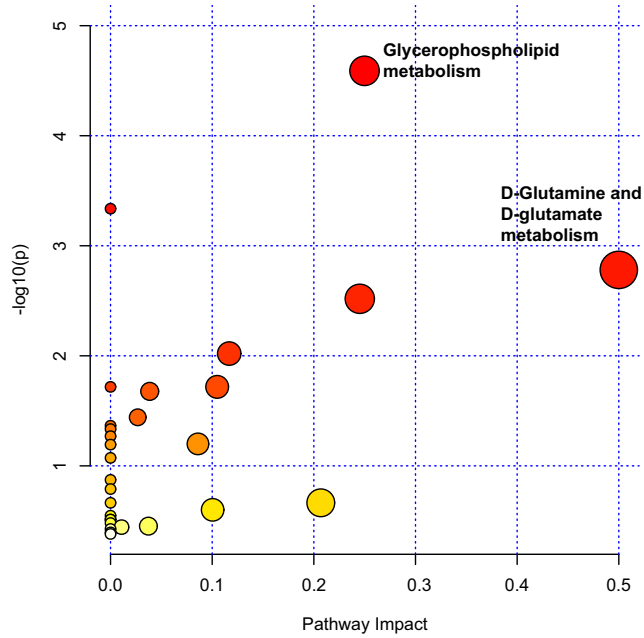


Figure 2. Pathway analysis of differential serum metabolites for esophageal malignancy based on the Kyoto Encyclopedia of Genes and Genomes pathway database

Note: Glycerophospholipid metabolism and D-glutamine and D-glutamate metabolism were identified as significant metabolic pathways associated with esophageal malignancy, at the false discovery rate threshold of 0.05 for enrichment analysis and a pathway impact threshold of 0.1.

p value = 1.38E-19), reflecting the discriminative ability of the MRS-based risk assessment tool (Figure S4). The MRS was then categorized into tertiles: "low-MRS group" (60 cases and 146 controls), "intermediate-MRS group" (100 cases and 107 controls), and "high-MRS group" (150 cases and 57 controls). Notably, 48.4% of the cases were classified as having a high MRS, while only 18.4% of the healthy controls fell into the high-MRS category (Table 3). The risk of esophageal malignancy for the intermediate-MRS group and the risk for the high-MRS group were three times (OR: 3.1 (95% confidence interval (CI): 1.9–5.0)) and 12 times (OR: 12.1 (95% CI: 6.6–22.3)) as high as that for the low-MRS group, respectively.

MRS = 0.267 × ximenoylcarnitine (C26:1) –
0.264 × sebacate (C10-DC) –
0.270 × azelate (C9-DC) +
0.228 × glycosyl-N-nervonoyl-sphingosine (d18:1/24:1) +
0.257 × mead acid (20:3n9) +
0.252 × 1-oleoyl-GPC (18:1) +
0.554 × heptanoate (7:0) +
0.233 × undecanoate (11:0) +
0.263 × 1,2-dipalmitoyl-GPC (16:0/16:0) +
0.289 × 1-stearoyl-2-oleoyl-GPS (18:0/18:1) +
0.398 × glycerophosphoinositol +
0.317 × glycerophosphoserine +
0.237 × chenodeoxycholate +
0.559 × 4-cholesten-3-one –
0.224 × galactonate +
0.320 × ribose +
0.361 × N6-methyladenosine –
0.271 × dihydroorotate –
0.236 × N-methylproline +
0.288 × 4-oxo-retinoic acid –
0.242 × pentose acid +
0.300 × leucylglutamine

Figure 3. The formula of Metabolic Risk Score for esophageal malignancy based on 22 independent metabolic predictors

Abbreviations: MRS, Metabolic Risk Score. Note: The MRS was constructed by summation of the metabolite level of each independent predictor multiplied by the respective effect size (β-coefficient in univariate conditional logistic regression).

Table 3. The distribution of the MRS for cases of esophageal malignancy and controls

			Cases		
Group	MRS	Controls	Screening-detected cases	Follow-up cases	Total
Low MRS	≤ −1.0936490	146 (47.1%)	23 (14.4%)	37 (24.7%)	60 (19.4%)
Intermediate MRS	(-1.0936490, 0.8867129]	107 (34.5%)	44 (27.5%)	56 (37.3%)	100 (32.3%)
High MRS	>0.8867129	57 (18.4%)	93 (58.1%)	57 (38.0%)	150 (48.4%)

Abbreviations: MRS, Metabolic Risk Score.
Note: The cutoff values of MRS groups were determined based on the tertiles of the MRS.

As shown in Figure 4A, the MRS model generated an area under the receiver operating characteristics curve (AUC) of 0.815 (95% CI: 0.782–0.848). Leave-one-out cross-validation yielded an AUC of 0.813 (95% CI: 0.779–0.846) (Figure 4B). The performance of the MRS model was superior to that of the questionnaire-based model (AUC: 0.648 (95% CI: 0.606–0.690)), with Delong test p = 3.12E-10 and a net reclassification improvement (NRI) of 0.329 (p = 3.44E-10). The integrated model combining MRS with questionnaire-based risk factors generated an AUC of 0.830 (95% CI: 0.798–0.861), significantly improving the performance of the questionnaire-based model (Delong test p = 5.39E-17; NRI: 0.329 (p = 9.42E-12)).

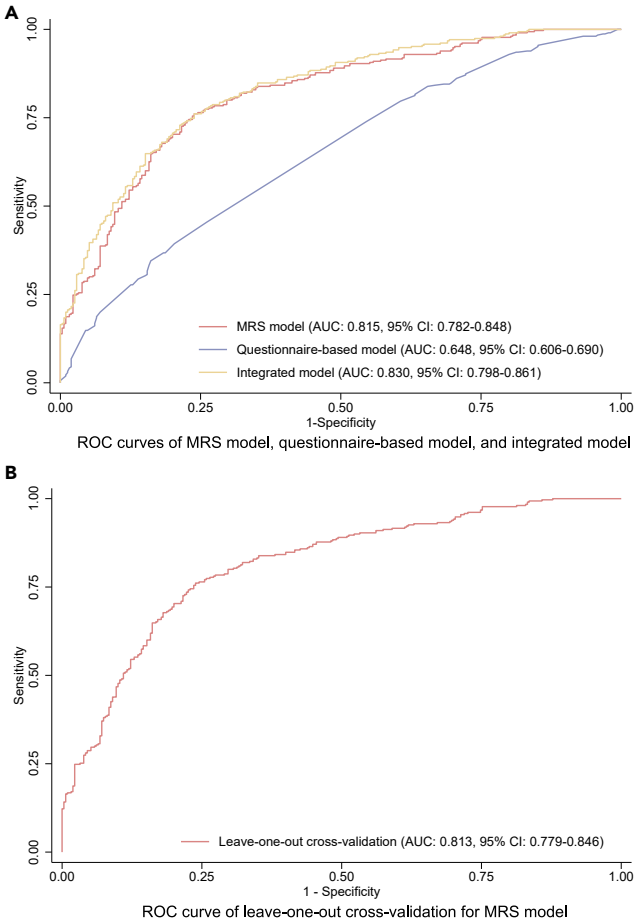


Figure 4. ROC curves of the MRS model (including leave-one-out cross-validation), the questionnaire-based model, and the integrated model

(A) ROC curves of MRS model, questionnaire-based model, and integrated model.

(B) ROC curve of leave-one-out cross-validation for MRS model.

Abbreviations: AUC, area under the receiver operating characteristics curve; CI, confidence interval; MRS, Metabolic Risk Score; ROC, receiver operating characteristics. Note: The MRS model contained MRS only; the questionnaire-based model contained risk factors for esophageal cancer which are well-recognized including body mass index, family history of esophageal cancer, cigarette smoking, and alcohol consumption; the integrated model contained MRS and aforementioned risk factors.

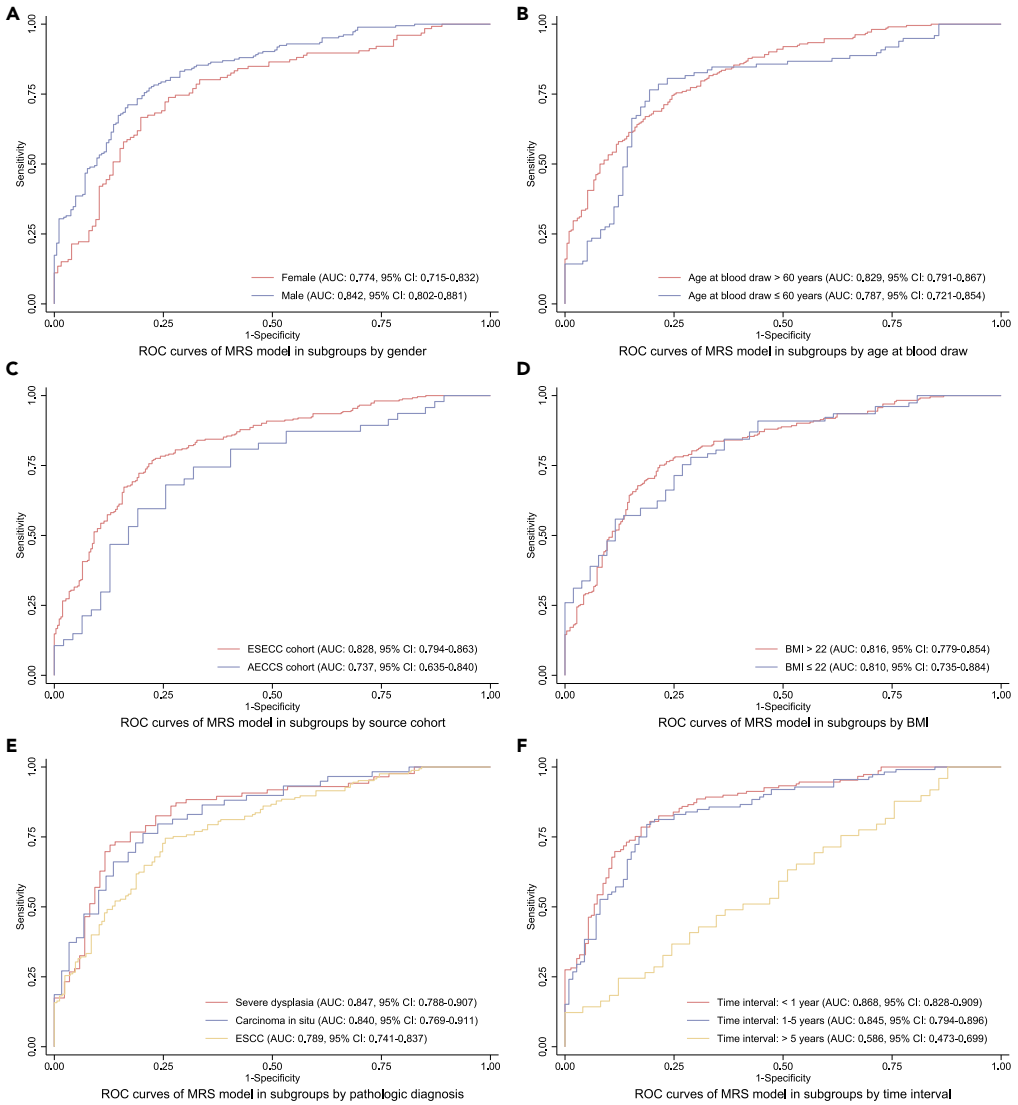


Figure 5. ROC curves of the MRS model in subgroups

(A) ROC curves by gender.

(B) ROC curves by age at blood draw.

(C) ROC curves by source cohort.

(D) ROC curves by BMI.

(E) ROC curves by pathologic diagnosis.

(F) ROC curves by time interval from blood draw to the diagnosis of esophageal malignancy.

Abbreviations: AECCS, Anyang Esophageal Cancer Cohort Study; AUC, area under the receiver operating characteristics curve; BMI, body mass index; CI, confidence interval; ESCC, esophageal squamous cell carcinoma; ESECC, Endoscopic Screening for Esophageal Cancer in China; MRS, Metabolic Risk Score; ROC, receiver operating characteristics.

Subgroup evaluation of the MRS model

As shown in Figure 5, no statistically significant differences were detected among the subgroups of female and male (AUC: 0.774 vs. 0.842, Delong test $p = 0.058$) (Figure 5A), >60-year and ≤60-year age groups (AUC: 0.829 vs. 0.787, Delong test $p = 0.287$) (Figure 5B), Endoscopic Screening for Esophageal Cancer in China (ESECC) and Anyang Esophageal Cancer Cohort Study (AECCS) cohorts (AUC: 0.828 vs. 0.737, Delong test $p = 0.100$) (Figure 5C), BMI >22 and BMI ≤22 groups (AUC: 0.816 vs. 0.810, Delong test $p = 0.877$) (Figure 5D), and severe dysplasia, CIS, and ESCC (AUC: 0.847 vs. 0.840 vs. 0.789, Delong test $p = 0.265$) (Figure 5E). Stratified by the time interval from blood draw to diagnosis, the MRS model generated similar AUCs for the within-1-year interval (AUC: 0.868 (95% CI: 0.828–0.909)) and 1-to-5-year interval (AUC: 0.845 (95% CI: 0.794–0.896)). However, the performance decreased when the time interval was over five years (AUC: 0.586 (95% CI: 0.473–0.699)) (Delong test $p = 2.41E\text{-}05$) (Figure 5F).

Additionally, stratified by the storage time from serum collection to metabolomics analysis, the discriminative ability of MRS in "storage time ≤8.2 years" group (AUC: 0.808) was similar to that in "storage time >8.2 years" group (AUC: 0.819) (Delong test $p = 0.735$), supporting the robustness of the MRS performance across serum samples with different storage times (Figure S5).

Sensitivity analyses

For the MRS model, sensitivity analysis was carried out by changing the coding forms of metabolite levels (tertiles, quartiles, and quintiles) and the methods for selecting independent metabolic predictors (conditional logistic regression). The AUCs of the re-established MRS models ranged from 0.713 to 0.819 (Figure S6).

DISCUSSION

Universal screening for ESCC, which covers the entire population of certain age groups in a given area, has been adopted in China and used worldwide for decades. Previous studies have shown that the majority of subjects would not immediately benefit from endoscopic examination due to low detection rate, high cost, and collateral harm of universal screening, emphasizing the necessity of identifying high-risk subgroups in advance, using low-invasive and efficient risk assessment tools. Early-warning biomarkers play an increasingly important role in risk stratification of ESCC,[15] and serum metabolomics has great potential due to low invasiveness and convenience of collection, transportation, and conservation. However, previous metabolomics research on ESCC mainly employed traditional case-control designs based on samples from clinical settings, where a considerable proportion of the cases were already at an advanced stage. Moreover, there might be heterogeneity between the selected controls and the general population. To overcome these limitations and provide a more robust evaluation of the early-warning value of serum metabolites for ESCC, we conducted this nested case-control study based on large community-based ESCC screening cohorts. We constructed a Metabolic Risk Score (MRS) with 22 selected serum metabolites and demonstrated their good performance in diagnosing prevalent and short-term malignancy of the esophagus, thus providing a promising tool for distinguishing high-risk individuals for ESCC screening.

In the first step of analysis, we identified 74 differential serum metabolites, and the majority of these belonged to lipid metabolism (54.1%), peptide metabolism (14.9%), and amino acid metabolism (13.5%). Dysregulation of some of our differential metabolites detected in this study were in line with other ESCC research. For instance, increasing levels of glutamate and malonate, and decreasing of alpha-tocopherol and pyruvate had been observed in patients with esophageal malignancy.[7,13,16] But we also found inverse associations of differential metabolites (such as ribose[17] and 2-oxoarginine[18]) and ESCC as compared to other studies. The discrepancy of these findings might be explained by hospital-based sampling methods adopted in prior studies. Due to absence of special symptoms (for example dysphagia) for early-stage malignant lesions in the esophagus, ESCC patients clinically diagnosed in hospitals are usually at an advanced stage. Such individuals may have had abnormal dietary habits for long periods of time. Thus, the possibility that alteration in metabolism was influenced by the disease per se cannot be excluded when exploring differential metabolites using a hospital-based design.

We further observed two significant metabolic pathways associated with the risk of esophageal malignancy, including glycerophospholipid metabolism and D-glutamine and D-glutamate metabolism. For energy metabolism in normal nonproliferating cells, glucose is metabolized to pyruvate through glycolysis, and pyruvate then tends to enter the process of oxidative phosphorylation in the mitochondria in the presence of oxygen, or is completely converted to lactate in the absence of oxygen (anaerobic glycolysis). Instead, the energy metabolism of cancer cells is characterized as aerobic glycolysis or "Warburg effect", where the majority of glucose is converted to lactate regardless of whether oxygen is sufficient. Emerging evidence has shown that to meet cellular requirements for synthesis of new biomass (phospholipids, nucleotides, amino acids, and so on), cancer cells preferentially utilize glycolysis to generate a series of metabolic intermediates. As depicted in Figure S7, dihydroxyacetone phosphate is synthesized during the process of glycolysis, and sustains the formation of glycerol-3-phosphate, which can be utilized in the synthesis of various glycerophospholipids (structural components of cellular membranes).[19] In addition, since glucose is mostly converted to lactate, cancer cells rely heavily on glutamine which serves as an alternative source of energy generation. Glutamine is converted by glutaminase into glutamate, which is further converted by glutamate dehydrogenase into α-ketoglutarate, replenishing the tricarboxylic acid cycle so that growing demand for energy can be met. Perturbation in glycerophospholipid metabolism and D-glutamine and D-glutamate metabolism have also been reported in previous ESCC-related studies.[13,20] Furthermore, Lugol's chromoendoscopy in ESCC screening is based on the reaction of iodine and intracellular glycogen, the latter of which is diminished or absent in the dysplastic tissues with rapidly dividing cells. Thus, the evidence noted above implies that the occurrence of esophageal malignancy may be related to abnormal energy metabolism. Moreover, this study has revealed over half of the differential serum metabolites belong to lipid metabolic pathways, and consistent with this finding, our team has also previously observed that dyslipidemia is associated with malignant esophageal lesions.[4,21] Further basic research is warranted to comprehensively explore the biologic mechanism of lipid metabolism and ESCC.

Based on these 74 differential metabolites, 22 independent metabolic predictors of esophageal malignancy were carefully selected and integrated as an MRS. Compared with the low-MRS group, the risk of esophageal malignancy for intermediate- and high-MRS groups increased two times and 11 times respectively, reflecting the strong capacity of the MRS for risk stratification. Overall, the MRS model showed a good performance in distinguishing cases of esophageal malignancy from healthy controls (AUC: 0.815), which is significantly superior to the performance of the questionnaire-based model. We further employed multiple strategies for validation of the MRS model. Similar performance was observed in leave-one-out cross-validation (AUC: 0.813) as well as among different genders, age groups, source cohorts, BMI, and pathologic diagnoses, suggesting the robustness of our MRS model.

Further stratified analysis showed that the MRS model can efficiently identify prevalent cases and the individuals at high risk of being diagnosed within five years. In contrast, we observed poor performance of the MRS model for distinguishing long-term risk, i.e., being diagnosed with ESCC >5 years later. There may be two assumptions explaining the observed time-sensitive associations. First, metabolic alterations may precede esophageal malignancy within a relatively short time window from exposure to malignancy onset. Second, metabolic alterations may arise as consequences of esophageal malignancy, presenting along with the carcinogenic process of esophageal squamous epithelium. In this case, metabolic alteration is a tumor burden indicator, thus not able to predict long-term risk of developing esophageal malignancy. Further molecular epidemiologic studies and laboratory studies are needed to reveal the underlying mechanism. All in all, the MRS model demonstrates its suitability for identifying prevalent esophageal malignancy cases as a diagnostic tool and for predicting future cases that may occur within five years after screening.

Our previous studies have demonstrated that Lugol-unstained lesions (LULs) of larger size are predictive of higher risk of progression to esophageal malignancy.[22] To confirm that the metabolic perturbations observed in this study were originated from the malignant lesions in the esophagus, we further evaluated the association of the MRS and the size of LULs in severe dysplasia and above (SDA) cases from the screening arm of the ESECC trial (size was determined by the smaller one between diameter and length of a LUL). MRS was found to be significantly higher in cases with large LULs (>10 mm) with an OR of 1.2 (95% CI: 1.0–1.4, $p = 0.035$), supporting the concept that alteration of serum metabolites can be derived from lesions in the esophagus.

In terms of application, we recommended the MRS as an initial risk-stratification tool to identify high-risk individuals prior to endoscopic screening. Given the complexity of metabolomic profiling, we suggest that blood samples be collected at community health centers and subsequently sent to high-quality central laboratories equipped with metabolomics platforms for further analysis and estimation of MRS. In settings where resources are abundant and increasing the proportion of detected cancer cases among all the cases (sensitivity) is the priority, we would recommend referring individuals with high MRS or even intermediate MRS to endoscopic examination, achieving a sensitivity of 80.6% (MRS cutoff value: −1.0936490). Whereas in settings where resources available for screening are limited, detecting as many malignancies as possible becomes the priority. In such scenarios, we recommend endoscopic examinations only for individuals with high MRS (MRS cutoff value: 0.8867129). Even if no malignant lesions are found during the endoscopy, those individuals are still recommended to undergo regular endoscopic surveillance over the next five years to identify any incident cases arising in the short-term future. For individuals not reaching the MRS cutoff value, we recommend consultation with professional physicians to make informed decisions regarding the necessity of undergoing endoscopic examinations. Besides, given the observed 5-year time window of predictive ability of the MRS model, continuous risk monitoring is also encouraged through repeated MRS assessments every five years, which allows for timely detection of any changes in metabolic profiles that might indicate an increased risk of esophageal malignancy.

Considering the importance of metabolite concentration stability, in this study, we have taken relevant measures in multiple aspects. Firstly, the collection, processing, and storage of all blood samples followed a unified implementation plan. Secondly, to ensure the comparability of serum metabolites between the case group and the control group, we took the baseline blood collection date as one of the matching variables when selecting controls. Thirdly, for metabolomics analysis, rigorous quality control measures were implemented, and the results indicated that the instrument had met the stability requirements (the median relative standard deviation of the internal standard compounds was lower than the allowable upper limit). Fourthly, in the data pre-processing stage, normalization was used to eliminate the influence of detection batch effects on the comparability of metabolite concentrations between samples. In addition, to further provide supporting evidence on analyte stability, we tested the association between smoking habit and tobacco-related metabolites (cotinine and hydroxycotinine) (Figure S8). We observed significantly higher levels of both cotinine ($OR_{per\ standard\ deviation} = 11.4$, $p = 1.98E\text{-}31$) and hydroxycotinine ($OR_{per\ standard\ deviation} = 10.6$, $p = 7.08E\text{-}45$) in the smoking group compared to the non-smoking group. Further, no significant heterogeneity of the association was detected between subgroups stratified by the median time interval from blood draw to metabolomics analysis (8 years) (cotinine: $p_{interaction} = 0.533$; hydroxycotinine: $p_{interaction} = 0.189$). These results also indicated the stability of the metabolite levels in our serum samples used in this study.

This study has the following strengths: 1) serum samples were collected before the onset of a symptomatic clinically diagnosed tumor, ensuring that the identified metabolic biomarkers could be used for early detection and early warning of esophageal malignancy as a pre-endoscopy test; 2) both cases and controls were recruited from the community general population, ensuring ideal population representativeness and generalizability of our results (external validity); 3) through incidence density sampling from the same cohort, cases and controls were meticulously matched based on a series of factors, which guaranteed comparability between cases and controls (internal validity).

In summary, this study has revealed a series of serum metabolites associated with esophageal malignancy, and has constructed an MRS which can efficiently predict the risk of current and short-term esophageal cancer. The MRS proposed in this study could be an ideal risk-stratification tool used prior to invasive endoscopic examination, which is essential for establishing new precision strategies for both organized and opportunistic screening for ESCC in China and worldwide.

Limitations of the study

The limitation of this study should also be noted. Despite adopting nested case-control design on the basis of two prospective community-based cohorts with relatively large sample sizes (>600 individuals), this is nevertheless a single-center investigation carried out in a high-risk region for ESCC. Multicenter studies are warranted to validate the performance of the MRS in different populations with various ESCC disease burden, further enhancing the robustness and applicability of our findings.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Study participants and sample collection
  - Case definition and control selection
- METHOD DETAILS
  - Metabolomic profiling
  - Pre-processing of metabolomics data
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109965.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2021YFC2500405), the National Natural Science Foundation of China (82073626), the Beijing Natural Science Foundation (J200016, 7222243), and the Beijing Nova Program (Z201100006820093). The funders of the study had no role in study design, collection, analysis, and interpretation of data, or writing of the report.

We would like to thank Prof. Dong Hang from Nanjing Medical University for assistance in conception and design of the study and data analysis, Dr. Michael A. McNutt for editing and correction of this manuscript, and Calibra Lab at DIAN Diagnostics and Key Laboratory of Digital Technology in Medical Diagnostics of Zhejiang Province for the support with untargeted metabolomics analysis.

AUTHOR CONTRIBUTIONS

Y.K., Z.He, and M.L. contributed to the conception and design of the study. Z.He, M.L., H.T., M.W., C.G., R.X., F.Li., A.L., H.Y., L.D., L.S., Q.W., Z.L., Y.L., F.Liu, Y.P., Z.Hu., H.Chen, and H.Cai contributed to the acquisition of data. Z.He., M.L., and H.T. contributed to data analysis. Y.K., Z.He, M.L., and H.T. contributed to interpretation of data and checking of the results. H.T., M.L., Z.He, and Y.K. contributed to drafting of the manuscript, which was reviewed and approved by all coauthors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 11, 2023
Revised: March 20, 2024
Accepted: May 9, 2024
Published: May 11, 2024

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. *71*, 209–249. https://doi.org/10.3322/caac.21660.
2. Zheng, R., Zhang, S., Zeng, H., Wang, S., Sun, K., Chen, R., Li, L., Wei, W., and He, J. (2022). Cancer incidence and mortality in China, 2016. J. Natl. Canc. Center *2*, 1–9. https://doi.org/10.1016/j.jncc.2022.02.002.
3. Wang, G.Q., Abnet, C.C., Shen, Q., Lewin, K.J., Sun, X.D., Roth, M.J., Qiao, Y.L., Mark, S.D., Dong, Z.W., Taylor, P.R., and Dawsey, S.M. (2005). Histological precursors of oesophageal squamous cell carcinoma: results from a 13 year prospective follow up study in a high risk population. Gut *54*, 187–192. https://doi.org/10.1136/gut.2004.046631.
4. He, Z., Liu, Z., Liu, M., Guo, C., Xu, R., Li, F., Liu, A., Yang, H., Shen, L., Wu, Q., et al. (2019). Efficacy of endoscopic screening for esophageal cancer in China (ESECC): design and preliminary results of a population-based randomised controlled trial. Gut *68*, 198–206. https://doi.org/10.1136/gutjnl-2017-315520.
5. Seyfried, T.N., Flores, R.E., Poff, A.M., and D'Agostino, D.P. (2014). Cancer as a metabolic disease: implications for novel therapeutics. Carcinogenesis *35*, 515–527. https://doi.org/10.1093/carcin/bgt480.
6. Wang, P.P., Song, X., Zhao, X.K., Wei, M.X., Gao, S.G., Zhou, F.Y., Han, X.N., Xu, R.H., Wang, R., Fan, Z.M., et al. (2022). Serum Metabolomic Profiling Reveals Biomarkers for Early Detection and Prognosis of Esophageal Squamous Cell Carcinoma. Front. Oncol. *12*, 790933. https://doi.org/10.3389/fonc.2022.790933.
7. Yu, M., Wen, W., Yi, X., Zhu, W., Aa, J., and Wang, G. (2022). Plasma Metabolomics Reveals Diagnostic Biomarkers and Risk Factors for Esophageal Squamous Cell Carcinoma. Front. Oncol. *12*, 829350. https://doi.org/10.3389/fonc.2022.829350.

8. Zhao, J., Zhao, X., Yu, J., Gao, S., Zhang, M., Yang, T., and Liu, L. (2022). A multi-platform metabolomics reveals possible biomarkers for the early-stage esophageal squamous cell carcinoma. Anal. Chim. Acta *1220*, 340038. https://doi.org/10.1016/j.aca.2022.340038.
9. Chen, Z., Huang, X., Gao, Y., Zeng, S., and Mao, W. (2021). Plasma-metabolite-based machine learning is a promising diagnostic approach for esophageal squamous cell carcinoma investigation. J. Pharm. Anal. *11*, 505–514. https://doi.org/10.1016/j.jpha.2020.11.009.
10. Yuan, Y., Zhao, Z., Xue, L., Wang, G., Song, H., Pang, R., Zhou, J., Luo, J., Song, Y., and Yin, Y. (2021). Identification of diagnostic markers and lipid dysregulation in oesophageal squamous cell carcinoma through lipidomic analysis and machine learning. Br. J. Cancer *125*, 351–357. https://doi.org/10.1038/s41416-021-01395-w.
11. Zhang, S., Lu, X., Hu, C., Li, Y., Yang, H., Yan, H., Fan, J., Xu, G., Abnet, C.C., and Qiao, Y. (2020). Serum Metabolomics for Biomarker Screening of Esophageal Squamous Cell Carcinoma and Esophageal Squamous Dysplasia Using Gas Chromatography-Mass Spectrometry. ACS Omega *5*, 26402–26412. https://doi.org/10.1021/acsomega.0c02600.
12. Zhu, Z.J., Qi, Z., Zhang, J., Xue, W.H., Li, L.F., Shen, Z.B., Li, Z.Y., Yuan, Y.L., Wang, W.B., and Zhao, J. (2020). Untargeted Metabolomics Analysis of Esophageal Squamous Cell Carcinoma Discovers Dysregulated Metabolic Pathways and Potential Diagnostic Biomarkers. J. Cancer *11*, 3944–3954. https://doi.org/10.7150/jca.41733.
13. Liu, Y.Y., Yang, Z.X., Ma, L.M., Wen, X.Q., Ji, H.L., and Li, K. (2019). (1)H-NMR spectroscopy identifies potential biomarkers in serum metabolomic signatures for early stage esophageal squamous cell carcinoma. PeerJ *7*, e8151. https://doi.org/10.7717/peerj.8151.
14. Xu, J., Chen, Y., Zhang, R., Song, Y., Cao, J., Bi, N., Wang, J., He, J., Bai, J., Dong, L., et al. (2013). Global and targeted metabolomics of esophageal squamous cell carcinoma discovers potential diagnostic and therapeutic biomarkers. Mol. Cell. Proteomics *12*, 1306–1318. https://doi.org/10.1074/mcp.M112.022830.
15. He, Z., and Ke, Y. (2023). Challenge and future of cancer screening in China: Insights from esophageal cancer screening practice. Chin. J. Cancer Res. *35*, 584–594. https://doi.org/10.21147/j.issn.1000-9604.2023.06.03.
16. Ye, W., Lin, Y., Bezabeh, T., Ma, C., Liang, J., Zhao, J., Ouyang, T., Tang, W., and Wu, R. (2021). (1) H NMR-based metabolomics of paired esophageal tumor tissues and serum samples identifies specific serum biomarkers for esophageal cancer. NMR Biomed. *34*, e4505. https://doi.org/10.1002/nbm.4505.
17. Liu, L., Wu, J., Shi, M., Wang, F., Lu, H., Liu, J., Chen, W., Yu, G., Liu, D., Yang, J., et al. (2022). New Metabolic Alterations and A Predictive Marker Pipecolic Acid in Sera for Esophageal Squamous Cell Carcinoma. Dev. Reprod. Biol. *20*, 670–687. https://doi.org/10.1016/j.gpb.2021.08.016.
18. Chen, Z., Gao, Y., Huang, X., Yao, Y., Chen, K., Zeng, S., and Mao, W. (2021). Tissue-based metabolomics reveals metabolic biomarkers and potential therapeutic targets for esophageal squamous cell carcinoma. J. Pharm. Biomed. Anal. *197*, 113937. https://doi.org/10.1016/j.jpba.2021.113937.
19. Pavlova, N.N., and Thompson, C.B. (2016). The Emerging Hallmarks of Cancer Metabolism. Cell Metab. *23*, 27–47. https://doi.org/10.1016/j.cmet.2015.12.006.
20. Yang, T., Hui, R., Nouws, J., Sauler, M., Zeng, T., and Wu, Q. (2022). Untargeted metabolomics analysis of esophageal squamous cell cancer progression. J. Transl. Med. *20*, 127. https://doi.org/10.1186/s12967-022-03311-z.
21. Wang, M.M., Guo, C.H., Li, F.L., Xu, R.P., Liu, Z., Pan, Y.Q., Liu, F.F., Liu, Y., Cai, H., Liu, M.F., et al. (2021). Family history of esophageal cancer modifies the association of serum lipids and malignant esophageal lesions: a nested case-control study from the "Endoscopic Screening for Esophageal Cancer in China" trial. Chin. Med. J. (Engl.) *134*, 1079–1086. https://doi.org/10.1097/cm9.0000000000001432.
22. Liu, M., Zhou, R., Guo, C., Xu, R., Liu, A., Yang, H., Li, F., Duan, L., Shen, L., Wu, Q., et al. (2021). Size of Lugol-unstained lesions as a predictor for risk of progression in premalignant lesions of the esophagus. Gastrointest. Endosc. *93*, 1065–1073.e3. https://doi.org/10.1016/j.gie.2020.09.020.
23. Pang, Z., Chong, J., Zhou, G., de Lima Morais, D.A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.É., Li, S., and Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res. *49*, W388–W396. https://doi.org/10.1093/nar/gkab382.
24. Tian, H., Yang, W., Hu, Y., Liu, Z., Chen, L., Lei, L., Zhang, F., Cai, F., Xu, H., Liu, M., et al. (2020). Estimating cancer incidence based on claims data from medical insurance systems in two areas lacking cancer registries in China. EClinicalMedicine *20*, 100312. https://doi.org/10.1016/j.eclinm.2020.100312.
25. Liu, F., Guo, F., Zhou, Y., He, Z., Tian, X., Guo, C., Ning, T., Pan, Y., Cai, H., and Ke, Y. (2012). The Anyang Esophageal Cancer Cohort Study: study design, implementation of fieldwork, and use of computer-aided survey system. PLoS One *7*, e31602. https://doi.org/10.1371/journal.pone.0031602.
26. Liu, M., Zhou, R., Liu, Z., Guo, C., Xu, R., Zhou, F., Liu, A., Yang, H., Li, F., Duan, L., et al. (2022). Update and validation of a diagnostic model to identify prevalent malignant lesions in esophagus in general population. EClinicalMedicine *47*, 101394. https://doi.org/10.1016/j.eclinm.2022.101394.
27. Wang, M., Liu, F., Pan, Y., Xu, R., Li, F., Liu, A., Yang, H., Duan, L., Shen, L., Wu, Q., et al. (2021). Tumor-associated autoantibodies in ESCC screening: Detecting prevalent early-stage malignancy or predicting future cancer risk? EBioMedicine *73*, 103674. https://doi.org/10.1016/j.ebiom.2021.103674.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human serum samples	The ESECC trial	NCT01688908; https://doi.org/10.1136/gutjnl-2017-315520
Human serum samples	The AECCS cohort	https://doi.org/10.1371/journal.pone.0031602
Deposited data		
Raw metabolomic data	This paper	https://ngdc.cncb.ac.cn/omix (OMIX database: OMIX004101)
Software and algorithms		
STATA version 15.0	StataCorp LLC	https://www.stata.com/
R version 4.1.3	R Project	https://www.r-project.org
MetaboAnalyst version 5.0	Pang et al.[23]	www.metaboanalyst.ca
Other		
ACQUITY BEH C18 column, 2.1 × 100 mm, 1.7 μm	Waters	Cat#186008316
ACQUITY BEH Amide column, 2.1 × 150 mm, 1.7 μm	Waters	Cat#186008315

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Zhonghu He (zhonghuhe@foxmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The raw metabolomic data have been deposited in the OMIX, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix, OMIX database: OMIX004101) and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report the original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study participants and sample collection

This case-control study is nested in two independent prospective community-based cohorts, namely the Endoscopic Screening for Esophageal Cancer in China (ESECC, ClinicalTrials.gov identifier: NCT01688908) and the Anyang Esophageal Cancer Cohort Study (AECCS), in the Taihang Mountain area, which is a region of risk for ESCC in China.[24] These two cohorts employed different methods for target village selection and eligibility criteria for participation.[4,25] Briefly, for the ESECC cohort, in 2012–2016 a total of 668 villages with population sizes of 500–3000 from Hua County in rural Anyang were randomly selected and allocated at 1:1 to the screening arm or the non-screening arm based on blocked randomization. 33,948 residents aged 45–69 were enrolled (sampling proportion: ~20%). For the AECCS cohort, in 2007–2014 9208 residents aged 25–65 were enrolled from ten target villages in rural Anyang (sampling proportion: ~70%), which were selected according to location, population size, administrative capabilities, etc., and underwent three rounds of endoscopic examination (2007-2009, 2009-2011, and 2012-2014). Among the ten target villages in the AECCS cohort, four villages where blood samples were collected only at the first round of

endoscopic screening were excluded in this study, considering the potential influence of storage time of serum samples on metabolomics detection.

For the screening arm of the ESECC trial and the AECCS cohort, standard upper gastrointestinal endoscopy with iodine staining was performed by experienced physicians. Biopsies were taken if focal lesions were observed and were fixed in 10% formaldehyde, embedded in paraffin, sectioned at 5 μm, and stained with hematoxylin and eosin. The biopsy specimens were reviewed by two experienced pathologists without knowledge of endoscopic findings, and discrepancies in histologic diagnoses were adjudicated by consultation.

At baseline recruitment of both these cohorts, blood sample collection, physical examination and a questionnaire interview were conducted for each participant by well-trained investigators prior to endoscopic examination. Each participant provided a fasting blood sample. These blood samples were placed overnight in a 4°C refrigerator and, on the following day, centrifuged to extract serum. The collected serum was promptly transferred and temporarily stored at −20°C, and later stored at −80°C when it was transported to Beijing.

The occurrence of esophageal malignancy among the participants was either detected through endoscopic screening (including re-examination) or captured by using annual follow-up, which involved active door-to-door interview and passive linkage with claims data from medical insurance system.[24]

Case definition and control selection

Esophageal malignancy in this study was defined as lesions of severe dysplasia and above (SDA) (severe dysplasia, CIS, and ESCC) detected through endoscopic screening or follow-up, and these cases were further categorized into prevalent cases and incident cases. Prevalent cases in this study referred to both esophageal malignancies detected at baseline screening and interval cancers diagnosed within one year after the initial screening, in order to capture cases that might have been "overlooked" during endoscopic examination.[26,27] Incident cases in this study included esophageal malignancies diagnosed at least one year after the initial screening in both the ESECC screening arm and AECCS cohort, as well as all the clinically diagnosed ESCC cases in the ESECC non-screening arm.

Up to November 15th, 2020 (the longest follow-up time was 9.0 years for the ESECC cohort and 13.1 years for the AECCS cohort), a total of 353 SDA cases from ESECC (279 cases, including 134 screening-detected cases and 145 follow-up cases) and AECCS (74 cases, including 45 screening-detected cases and 29 follow-up cases) cohorts were detected. For the AECCS cohort, considering the potential impact of relatively long serum storage time on metabolomics analysis, cancer cases detected at the first round of endoscopic screening (n = 19) were excluded. Besides, those cases without serum samples (n = 8), complete questionnaire information (n = 15), or any matched healthy control (n = 1) were also excluded. Finally, a total of 310 cases of esophageal malignancy (ESECC cohort: 263 cases; AECCS cohort: 47 cases) were included in this study.

For each case, one control was randomly selected through incidence density sampling, with the matching variables including source cohort, allocated arm (for ESECC trial only), gender, age at blood draw (±1 year), and date at blood draw (±30 days). Given the nested case-control study design and the utilization of 1:1 incidence density sampling, two cases of esophageal malignancy were also selected as matched controls, before their diagnosis, for two other cases. Thus, a total of 618 blood samples at enrollment for both cases and controls were used in this study.

The study participants were all Han Chinese. The mean age at blood draw was 61.8 years and 40.7% were female. This study was approved by the Institutional Review Board of the Peking University School of Oncology, China (Approval number: 2011101110, 2006020). All participants provided written informed consent.

METHOD DETAILS

Metabolomic profiling

Untargeted ultra-high-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) was used to detect and characterize metabolomic profiles from the de-identified serum samples (Waters ACQUITY 2D UPLC system; Thermo Fisher Scientific Q-Exactive (QE) high-resolution/accurate mass orbitrap spectrometer). The metabolite identification process strictly adhered to the Metabolomics Standards Initiative (MSI) tier 1 standard. More detailed information regarding sample preparation, UPLC and MS methods, compound identification, and quality control are as follows.

Sample preparation was carried out on the automated MicroLabSTAR® system (Hamilton). Metabolites were extracted by precipitating protein with methanol and centrifugation after vigorous shaking for two minutes (GenoGrinder). To guarantee the quantity and reliability of metabolite detection, the extract was divided into four fractions for analysis: three were used for UPLC methods using reverse phase liquid chromatography based on C18 columns (UPLC BEH C18-2.1 × 100 mm, 1.7 μm; Waters) (Method A and Method B with positive ion mode electrospray ionization (ESI), and Method C with negative ion mode ESI), and one was used for UPLC method using a hydrophilic interaction liquid chromatography column (UPLC BEH Amide 2.1 × 150 mm, 1.7 μm; Waters) for metabolites with strong polarity (Method D with negative ion mode ESI). The mobile solutions used for gradient elution were as follows: (1) water, methanol, 0.05% perfluoropentanoic acid (PFPA), and 0.1% formic acid (FA) (Method A); (2) methanol, acetonitrile, water, 0.05% PFPA, and 0.01% FA (Method B); (3) methanol and water in 6.5 mM ammonium bicarbonate at pH 8 (Method C); (4) water and acetonitrile with 10 mM ammonium formate at pH 10.8 (Method D). All the four fractions were dried using nitrogen gas (removing the organic solvent) and were later re-dissolved in compatible reconstitution solvents. The QE mass spectrometer analysis was alternated between MS and data-dependent MS2 scans using dynamic exclusion at 35,000 mass resolution, and the scan range was 70–1,000 m/z.

Data acquired from the mass spectrometer were converted to be processed using proprietary in-house software. Raw data pre-processing, peak finding/alignment, and peak annotation were orderly performed, and metabolites were identified by searching the in-house library (more than 3300 purified standard compounds had been registered). Identification of metabolites was based on three criteria: (1) narrow window retention index; (2) accurate mass match to the library with variation <5 ppm; (3) MS/MS spectra with high forward and reverse scores (comparing the ions in the experimental spectrum to that in the library spectrum entries).

Quality control (QC) samples were generated by taking a small volume of each serum sample, and these samples were spaced evenly among the randomized experimental samples. Internal standards, which were carefully selected and did not interfere with the measurement of metabolites, were added into each sample to aid chromatographic peak alignment and instrument stability monitoring. The stability of the equipment was evaluated by calculating the median relative standard deviation (RSD) for the standards among the QC samples. In this study, median RSD was reported as 4.3% which was below the limit (5%), denoting a stable LC-MS system performance during the metabolomics assay (Figure S9).

Pre-processing of metabolomics data

To balance batch variability, the level of each metabolite was divided by the median value in a given batch to obtain normalized data. Natural log transformation was applied to improve the normality of distribution. To facilitate comparison of metabolites, a scaling method (standardization) was also used according to the mean and standard deviation (SD) of all serum samples. Missing values were imputed with the minimum non-missing value, and only those metabolites with missingness less than 20% were retained. Finally, a total of 817 metabolites were included in the analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Baseline characteristics of the study participants were presented as means (SDs) for continuous variables and numbers (percentages) for categorical variables. Differences between cases and matched controls were tested using the Wilcoxon signed-rank test (for continuous variables) and the McNemar test (for categorical variables). Univariate conditional logistic regression was used to calculate odds ratio (OR) per one SD increase in metabolite concentrations. The Benjamini-Hochberg procedure was applied for multiple testing correction using false discovery rate (FDR), and those metabolites with FDR <0.2 were identified as differential metabolites. Pathway analysis was then performed using MetaboAnalyst 5.0 online software (www.metaboanalyst.ca) based on these differential metabolites and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database.[23] Pathway impact value was calculated from pathway topology analysis. In consideration of multiple testing of pathways, we used FDR to adjust the *p* value calculated from the enrichment analysis. Based on the differential metabolites, least absolute shrinkage and selection operator (LASSO) with a 10-fold cross-validation was performed to select the optimal combination of independent metabolic predictors. Specifically, metabolites that could independently predict outcomes were initially identified based on the value of lambda that gave a minimum mean cross-validated error (binomial deviance), and then a simpler model with the largest value of lambda where error was within one standard error of the minimum was finally retained.

A Metabolic Risk Score (MRS) for esophageal malignancy was constructed by summation of the metabolite level of each independent predictor multiplied by the respective effect size (β-coefficient in univariate conditional logistic regression). We constructed the following prediction models using conditional logistic regression for comparison: (1) MRS model: containing MRS only; (2) questionnaire-based model: containing traditional risk factors including body mass index (BMI), family history of EC, cigarette smoking, and alcohol consumption; (3) integrated model: containing both MRS and aforementioned risk factors. We evaluated the performance of the prediction model by calculating the area under the receiver operating characteristics curve (AUC), and compared the ability for discrimination among these models using the Delong test and net reclassification improvement (NRI).

Multiple methods were used for validation of the robustness of the MRS model. First, leave-one-out cross-validation was performed based on the predicted probabilities of each matched pair generated from models built on all the remaining pairs. Then, the MRS model was further validated within each of the subgroups by gender, age at blood draw, source cohort, BMI, pathologic diagnosis, and time interval from blood draw to diagnosis of esophageal malignancy.

Sensitivity analysis for the performance of the MRS model was also performed using the following strategies: (1) metabolite levels were grouped into tertiles, quartiles, and quintiles; (2) independent metabolic predictors were selected using a conditional logistic regression.

Statistical analysis was conducted using STATA version 15.0 and R version 4.1.3 ("glmnet" package). All tests were 2-sided and had a significance level of 0.05 unless otherwise specified.

ADDITIONAL RESOURCES

84.8% of the samples in this study were derived from ESECC, which is a clinical trial (NCT01688908): https://classic.clinicaltrials.gov/ct2/show/NCT01688908?term=01688908&draw=2&rank=1.