Metabolomics study of esophageal adenocarcinoma

Jian Zhang, MS,[a] Lingyan Liu, MS,[b] Siwei Wei, MS,[a] G. A. Nagana Gowda, PhD,[a] Zane Hammoud, MD,[c] Kenneth A. Kesler, MD,[d] and Daniel Raftery, PhD[a]

Objective: The objective of this study was to detect and evaluate reliable metabolite markers for screening and monitoring treatment of patients with esophageal adenocarcinoma (EAC) by studying metabolomics. The sensitivity and specificity of the study were evaluated not only for EAC but also for Barrett esophagus and high-grade dysplasia, which are widely regarded as precursors of EAC.

Methods: Profiles of metabolites in blood serum were constructed using nuclear magnetic resonance spectroscopy and statistical analysis methods. The metabolite biomarkers discovered were selected to build a predictive model that was then used to test the classifications accuracies.

Results: Eight metabolites showed significant differences in their levels in patients with cancer and in the control group on the basis of Student *t* test. A partial least-squares discriminant analysis model built on these metabolites provided excellent classifications of patients with cancer and the control group, with the area under the receiver operating in a characteristic curve of > 0.85 for both training and validation sample sets. Evaluated by the same model, the Barrett esophagus samples were of mixed classification, and the high-grade dysplasia samples were classified primarily as cancer samples. A pathway study indicated that altered energy metabolism and changes in the trochloroacetic acid cycle were the dominant factors in the biochemistry of EAC.

Conclusions: ^{1}H nuclear magnetic resonance–based metabolite profiling analysis was shown to be an effective approach to differentiating between patients with EAC and healthy subjects. Good sensitivity and selectivity were shown by using the 8 metabolite markers discovered to predict the classification of samples from the healthy control group and the patients with the disease. Serum metabolic profiling may have potential for early diagnosis of EAC and may enhance our understanding of its mechanisms. (J Thorac Cardiovasc Surg 2011;141:469-75)

GTS

Supplemental material is available online.

Esophageal cancer (EC) is a leading cause of death from cancer worldwide. The two principal types of EC, squamous cell carcinoma and adenocarcinoma, are relatively uncommon in the United States, composing approximately 1% of all cancers. However, the incidence of adenocarcinoma is rising at a rapid rate. According to a report from the American Cancer Society, 12,300 new cases and 12,100 deaths were reported in 2000,[1] and the corresponding numbers for 2009 were 16,470 and 14,530, respectively.[2] The 5-year survival rates of those with localized and all stages combined are 34% and 17%, respectively.[2] Moreover, currently there is no reliable method of early detection or of prediction of treatment outcome.

Barrett's esophagus (BE), high-grade dysplasia (HGD), and invasive cancer are thought to compose a multistep process in the development of esophageal adenocarcinoma (EAC).[3] HGD has been considered to be the immediate precursor of invasive adenocarcinoma.[4] Most patients with HGD are usually bearing or developing cancer,[5] so HGD has been regarded as a marker of progression to carcinoma.[6] However, no intervention currently in existence prevents the progression of BE or HGD to esophageal cancer.[5] The traditional methods of diagnosing EC include endoscopy and barium swallow,[7] but the poor specificity and sensitivity of these methods results in the detection of EC only at advanced stages. Recently, prognostic and predictive protein and genetic markers have been introduced to aid in the diagnosis of EC.[8,9] However, markers effective at a potentially curative stage are lacking.

Metabolomics, a growing field in systems biology, offers a powerful and promising approach to identifying valuable biomarkers. Metabolomics (or metabolite profiling) describes the study of concentrations and fluxes of low molecular weight metabolites present in biofluids or tissues that provide detailed information about biologic systems and their current status.[10,11] The information-rich analytic

From the Department of Chemistry,[a] Weldon School of Biomedical Engineering,[b] Purdue University, West Lafayette, Ind; Department of Cardiothoracic Surgery,[c] Henry Ford Hospital, Detroit, Mich; and Department of Surgery,[d] Thoracic Surgery Division, Indiana University School of Medicine, Indianapolis, Ind.

Financial support: Purdue University's Center for Cancer Research and the Oncological Sciences Center in Discovery Park at Purdue.

Disclosures: Authors have nothing to disclose with regard to commercial support.

Received for publication April 19, 2010; revisions received July 12, 2010; accepted for publication Aug 1, 2010; available ahead of print Sept 29, 2010.

Address for reprints: Zane Hammoud, MS, Department of Cardiothoracic Surgery, Henry Ford Hospital, K-14, 2799 West Grand Boulevard, Detroit, MI 48202 (E-mail: zhammou1@hfhs.org).

0022-5223/$36.00

doi:10.1016/j.jtcvs.2010.08.025

Abbreviations and Acronyms

AUC = area under the curve
BE = Barrett's esophagus
EAC = esophageal adenocarcinoma
EC = esophageal cancer
GC/MS = gas chromatography/mass spectrometry
HGD = high-grade dysplasia
LV = latent variable
MS = mass spectrometry
NMR = nuclear magnetic resonance
PLS = partial least-squares
PLSDA = partial least-squares discriminant analysis
ROC = receiver operating characteristic
TCA = citrate cycle
TSP = trimethylsilylpropionic acid-d4 sodium salt

GTS

techniques of nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) combined with multivariate statistical analyses are the premier methods for metabolomics-based studies.[12] Various diseases, including but not limited to cancers, diabetes, inborn errors of metabolism, and coronary heart disease, have been explored using metabolomics as a tool, and a number of putative biomarkers have been detected and evaluated with the goal of improved diagnoses, assessment of risk, and even prediction of therapy outcomes.[11] A recent example is the identification of sarcosine as a potential marker for prostate tumor aggressiveness; it was discovered using a metabolomics approach.[13] However, few metabolomics studies of EC have been reported to date. Recently, Shen and coworkers reported 20 metabolite markers that were detected in fresh tumor tissue and corresponding normal esophageal mucosa by using gas chromatography/mass spectrometry (GC/MS).[14] To decrease the mortality rate of patients with esophageal cancer, development of early diagnostic methods, especially the exploitation of biomarkers that offer high sensitivity and specificity, is still in great demand. In addition, information provided by metabolomics in the area of biology may be useful in further understanding the biology of the disease.

In this study, ^{1}H NMR and multivariate statistical analysis were employed to detect molecular changes in human blood serum samples by comparing the metabolic profiles of patients with BE, HGD, and EAC as well as of normal controls in an attempt to identify a metabolite profile of EAC. We also attempted to identify a set of putative markers that may be useful in understanding the pathogenesis of EAC.

MATERIALS AND METHODS

Chemicals

Deuterium oxide (D_2O, 99.9% D) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, Mass), and trimethylsilylpropionic acid-d4 sodium salt (TSP) was purchased from Sigma-Aldrich (Milwaukee, Wis). All chemical reagents were analytic grade.

Serum Sample Collection and Storage

All work was conducted under a protocol approved by the Indiana University School of Medicine and Purdue University Institutional Review Board. All subjects included in the study provided informed consent according to institutional guidelines. All samples were collected when subjects were in the fasting state. Whole blood samples were collected from patients with histologically documented BE (n = 5), BE with HGD (n = 11), and adenocarcinoma (n = 68). Blood samples from 34 healthy volunteers served as controls. Each blood sample was allowed to clot for 45 minutes and then was centrifuged at 2000 rpm for 10 minutes. The serum was collected, and an aliquote was put into in a separate vial and stored at –80°C until use.

^{1}H NMR and Statistical Analysis

Samples were prepared by mixing 200 μL serum with 330 μL D_2O. A 60-μL solution of TSP (0.12 mg/mL) sealed in a separate capillary was used as an internal standard, which acted as the frequency standard (δ = 0.00). A Bruker DRX 500-MHz spectrometer equipped with a room-temperature hydrogen cyanide probe was used to acquire 1-dimensional ^{1}H NMR spectra. The water signal was suppressed using a standard 1-dimensional Carr-Purcell-Meiboom-Gill pulse sequence coupled with water presaturation. For each spectrum, 64 transients were collected, and 16K data points were acquired using a spectral width of 6000 Hz. An exponential weighting function corresponding to 1 Hz line broadening was applied to the free-induced decay before Fourier transformation. Phasing and baseline correction were applied using Bruker TopSpin software.

To remove the errors resulting from the small fluctuations of chemical shifts due to pH or ion concentration variations, NMR spectral regions were binned to 4K buckets of equal width (1.5 Hz). Each spectrum was aligned to the methyl peak of alanine at 1.48 ppm, and was normalized using the integrated TSP signal. Spectral regions within the range of 0.3 to 10 ppm were used after deleting the region containing the water resonance and urea signal (4.5 to 6. ppm).

To visualize the differences between spectra better, partial least-squares (PLS), a robust supervised method to detect subtle changes between group variations, was employed. PLS fits data matrices X (which consists of NMR spectra) and Y (which is set to 1 for cancer and 0 for control) and recasts these data as score plots and loading plots. The NMR spectral signals, or variables, were autoscaled (by subtracting the mean value of each variable and dividing by its standard deviations) prior to all statistical analyses. The score plot shows the possible relationships (or clustering) among the samples to estimate the classification; each orthogonal axis is named a latent variable (LV). The corresponding loading plot of each LV contains the weight or contribution of each variable in the modeling. To explore potential biomarker candidates, univariate analysis was performed by calculating the *P* value (unpaired Student *t* test) and a Benjamini-Hochberg correction was followed in order to control false discovery errors originating from multiplicity.[15]

Subsequently, a partial least-squares discriminant analysis (PLS-DA) model was built to evaluate the biomarker candidates when combined as a metabolite profile. Predictions were made visually using a Y-predicted scatter plot with a cut-off value chosen for potential class membership. The NMR data were imported into Matlab (R2008a; MathWorks, Natick, Mass) and installed using a PLS toolbox (version 4.1; Eigenvector Research, Inc.) for PLS and PLS-DA analysis.