

Identification of Human Esophageal Cancer Subtypes by Serum Protein Analysis Using MALDI-TOF Mass Spectrometry

Timothy P. Millar, Yuliya Karpevitch*, Elizabeth G. Hill*, Charles E. Hammond, Brenda J. Hoffman, Adam J. Smolka

Departments of Medicine and *Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina, Charleston, SC 29425

Overview

Purpose To use mass spectrometric analysis of human sera to identify spectral fingerprints that discriminate between healthy controls and patients diagnosed with esophageal cancer.

Methods Sera were obtained from 38 patients, treated with C8 superparamagnetic beads, and analyzed by MALDI-TOF MS.

Results MALDI-TOF MS discriminates between sera from healthy controls and patients diagnosed with esophageal adenocarcinoma

Objectives

To generate proteomic spectra of biophysically defined subsets of serum proteins from patients diagnosed with esophageal squamous cell carcinoma or esophageal adenocarcinoma using matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS), and to correlate differential features of the mass spectra with patient clinical status.

Introduction

The prognosis of patients with esophageal cancer is largely dependent on the presence or absence of lymph node metastasis at the time of diagnosis (30%-40% survival for stages I-IIa, and 5%-10% survival for stage IIb at 5 years). Most patients presenting with symptoms of any sub-type of esophageal cancer will have advanced disease. In those with resectable disease, the cure rate is still extremely low due to local recurrence and metastatic disease. The best hope of improving survival is early detection and effective therapeutic intervention. Numerous studies have shown that molecular genetic markers can be detected in lymph nodes and blood samples of patients with breast and lung cancer by real-time RT-PCR. Analogous protein markers may be associated with development and progression of esophageal cancers. We hypothesize that molecular signatures of such biomarkers in patient sera may be acquired by proteomic analysis using MALDI-TOF MS. Confirmation of this hypothesis will have clear prognostic and treatment implications. Earlier detection and more accurate staging of disease will optimize treatment strategies, and may provide more objective and stringent criteria for measurement of treatment outcome.

Contact

Adam J. Smolka, Ph.D.
smolkaaj@musc.edu

Elizabeth G. Hill, Ph.D.
hille@musc.edu

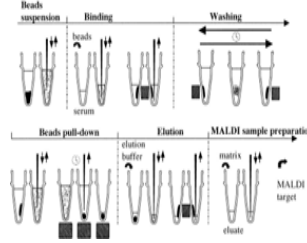
Timothy P. Millar
millart@musc.edu

Methods

SAMPLE PREPARATION. Sera were obtained from esophageal adenocarcinoma patients, esophageal squamous cell carcinoma patients, and from normal controls at Medical University Hospital, Charleston, SC. The study was approved by the MUSC Institutional Review Board and informed consent was obtained from all patients. Blood samples were collected into 9.5 ml BD Vacutainer SST Gel and Clot Activator tubes. After 30 min, blood samples were centrifuged at 3500 rpm for 15 min, and sera were aliquoted and stored at -80°C.

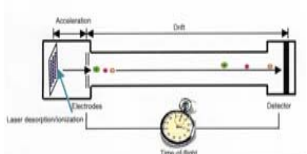
Sera were processed using reversed-phase ligand (C8)-derivatized superparamagnetic beads (MB-HIC8; Bruker Daltonics). Serum aliquots (5 µl) were mixed in a PCR tube with 10 µl binding buffer and 5 µl MB-HIC8 beads. The tube was placed in a magnetic separator, and supernatant was withdrawn. The beads were washed 3X with 100 µl wash buffer, and bound serum proteins were eluted with 5 µl 50% acetonitrile.

Serum Processing with Magnetic Beads



(Villanueva *et al* (2004) Anal Chem 76:1560-1569)

TARGET APPLICATION Gold-plated MALDI-TOF target plates were sonicated 15 min in 50% methanol/50% H₂O, rinsed with 100% ethanol and air-dried. Saturated matrix solution (10 µl α-cyano-hydroxy-cinnamic acid in 40% acetonitrile/50% methanol/10% H₂O) was added to 5 µl serum aliquots, mixed by gentle pipetting, and 1.5 µl was applied to randomly-selected target plate spots. Samples were air-dried for 5-10 min before analysis.



MALDI-TOF MS. Sera were analyzed in a bench-top laser desorption/ionization-time-of-flight mass spectrometer (Voyager STR) equipped with a 337 nm nitrogen laser operating at 3.0 Hz in linear mode geometry. Ions were accelerated at 25 kV with 500 nsec of pulsed ion extraction delay time. Spectra were externally calibrated with a mixture of protein standards between 2000 and 16,000 Da.

Methods

SIGNAL ACQUISITION To increase detection sensitivity, each sample spot was irradiated with 20 laser shots at an intensity of 4493, and *m/z* data were subsequently acquired from the same position with 100 laser shots at an intensity of 2496. Quadruplicate mass spectra spanning an *m/z* window of 800-20,000 Da were acquired for each sample on 3 different days.

DATA PROCESSING [using Matlab 7 (R14) SP 3]

Step 1 Intensity interpolation to impose a common time scale across all spectra.

Step 2 Removal of intensities <2000 Da from all spectra to eliminate 'matrix noise', a large-amplitude matrix signal that swamps the biological signal.

Step 3 Construction of a heat map to visually assess peak alignment (calibration).

Step 4 Identification and quantitation of peaks.

Peak detection

An undecimated discrete wavelet transform (UDWT; Rice Wavelet Toolbox) was used to smooth and denoise the mean spectrum by setting wavelet coefficients below a specified threshold to zero, leaving all other coefficients unchanged (Morris *et al*, 2005). Baseline artifact was removed from the denoised average spectrum by constructing a local minimum non-increasing curve. Local maxima with intensities exceeding a pre-specified signal/noise threshold δ were identified as *m/z* peaks. A total of 295 peaks was identified by estimating local noise using median absolute deviation based on the wavelet-based noise estimates in a window comprised of 41 *m/z* locations (20 to either side of the putative peak).

Peak quantification

Individual spectra were denoised using UDWT, and the estimated monotone minimum baseline was removed. Spectra were normalized to total ion current by dividing peak intensities by the sum of all intensities for a given spectrum. Each peak was quantified based on the maximum normalized intensity in a window defined by local minima used in identifying peak *m/z* locations (Morris *et al*, 2005). The resulting 295 normalized peak intensities for each of 507 spectra were used in subsequent statistical analyses.

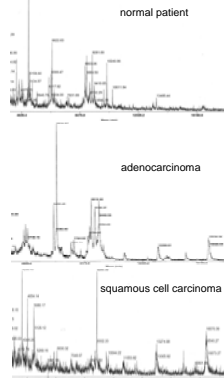
Statistical Analyses (SAS version 9.1.3)

Differentially-expressed peaks were identified based on false discovery rate (FDR) using Q-value software. Generalized estimating equations (GEEs) were used to model log_e peak intensity as a function of disease class. Differentially expressed peaks were identified on each fitted model using linear contrasts. Multiple comparisons (885 total) were adjusted by controlling the false discovery rate (FDR) at level 0.1 (Storey and Tibshirani, 2003).

Results

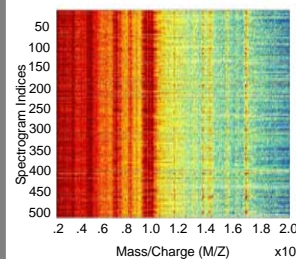
A total of 507 spectra were analyzed, comprising replicate spectra from 17 adenocarcinoma, 13 squamous cell carcinoma, and 8 control patients. The number of replicates per patient ranged from 7 to 24 (mean = 13.3 spectra)

Representative Serum MALDI-TOF Mass Spectra



These spectra are raw, unprocessed graphic files in which the x axis shows the *m/z* ratio and the y axis shows the relative peak intensities.

A Heat Map Showing 507 Individual Serum Spectra from All Patients



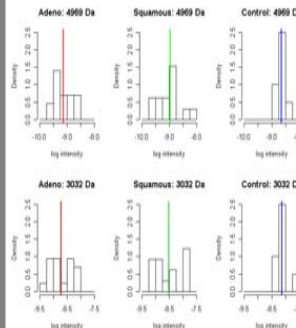
The map shows ranked peak intensities based on interpolated and truncated spectra

The predictive power of differentially-expressed peaks was assessed using simple logistic regression to model the log odds of adenocarcinoma vs. control and squamous cell carcinoma vs. control as a function of average log_e intensity. Trends in biomarker expression across disease states were assessed by modeling log_e peak intensity as a function of disease, modeled as a continuous covariate. Finally, the odds of classification to a given disease were measured as a function of average log_e intensity for differentially-expressed peaks.

Results

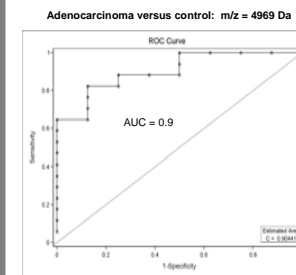
COMPARISON OF SPECTRA FROM ESOPHAGEAL CANCER AND CONTROL SERA

Histograms of Average Log Intensities for 4969 Da and 3032 Da Peaks

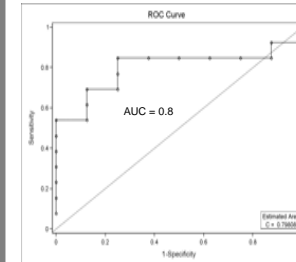


Peaks at 4969, 3032, 2982, and 3055 Da were significantly under-expressed in adenocarcinoma relative to control patients. Both 4969 and 3032 Da peaks were under-expressed in squamous cell carcinoma vs. controls (neither was significant with the FDR set at 0.1)

Receiver Operating Characteristic (ROC) Curves for the 4969 Da Peak



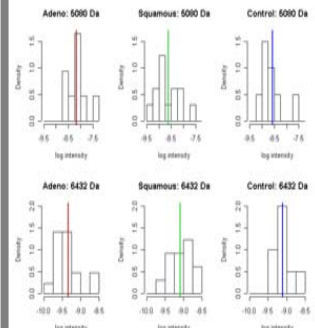
Squamous cell carcinoma versus control: m/z = 4969 Da



Results

COMPARISONS OF SPECTRA FROM ESOPHAGEAL ADENOCARCINOMA AND SQUAMOUS CELL CARCINOMA SERA

Histograms of Average Log Intensities for 5080 Da and 6432 Da Peaks



Comparing adenocarcinoma to squamous cell carcinoma, two peaks at 5080 Da and 6432 Da were differentially-expressed (neither was significant with FDR set at 0.1).

Conclusions

- Peaks at 4969, 3032, 2982, and 3055 Da were significantly under-expressed in adenocarcinoma relative to control patients.
- ROC curves indicate good classification for adenocarcinoma vs. controls and moderate classification of squamous cell carcinoma vs. controls
- A 10% increase in 5080 Da peak intensity reduces the odds of adenocarcinoma vs. squamous by 27% (OR = 1.27, 95% CI = 1.06 to 1.59, *p* = 0.020).
- A 10% increase in 6432 Da peak intensity reduces the odds of adenocarcinoma vs. squamous by 26% (OR = 0.74, 95% CI = 0.54 to 0.96, *p* = 0.034).
- MALDI-TOF MS discriminates between sera from healthy patients and those diagnosed with esophageal adenocarcinoma.

References

Morris JS *et al* (2005). Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* 21:1764 - 1775

Storey JD and Tibshirani R (2003). Statistical significance for genomewide studies. *PNAS* 100: 9440 - 9445

Acknowledgments

DOD/Hollings Cancer Center/ MUSC GC-3319-05-4498CM to A.J.S., E.G.H. and B.J.H.

NIDCR/NIH K25 DE016863 -01 to E.G.H.