Multi-Omic Analysis of Esophageal Adenocarcinoma Uncovers Candidate Therapeutic Targets and Cancer-Selective Posttranscriptional Regulation

O'Neill, J. Robert; Mayordomo, Marcos Yébenes; Mitulović, Goran; Al Shboul, Sofian; Bedran, Georges; Faktor, Jakub

Published in:
Molecular and Cellular Proteomics

DOI:
10.1016/j.mcpro.2024.100764

Publication date:
2024

Multi-Omic Analysis of Esophageal Adenocarcinoma Uncovers Candidate Therapeutic Targets and Cancer-Selective Posttranscriptional Regulation

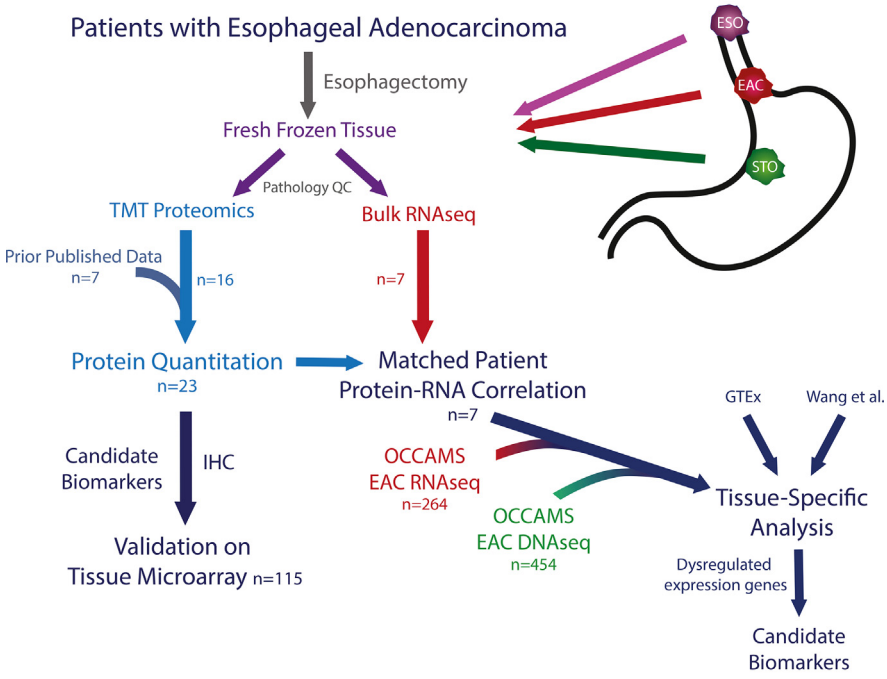
Authors

J. Robert O'Neill, Marcos Yébenes Mayordomo, Goran Mitulović, Sofian Al Shboul, Georges Bedran, Jakub Faktor, Lenka Hernychova, Lukas Uhrik, Maria Gómez-Herranz, Mikołaj Kocikowski, Vicki Save, Bořivoj Vojtěšek, Mark J. Arends, OCCAMS Consortium, Ted Hupp, and Javier Antonio Alfaro

Correspondence

robertoneill@nhs.net; marcos.yebenes@gmail.com; Javier.alfaro@proteogenomics.ca

Graphical Abstract



In Brief

This study provides the largest proteomic characterization in esophageal adenocarcinoma (EAC), analyzing proteins from 5879 genes across 23 patients. GPA33 is identified as a key EAC-selective biomarker overexpressed in a significant portion of EAC patients. Potential biomarkers characterized by RNA to protein dysregulation are uncovered by proteogenomic analysis. The EAC-selective overexpression of genes including AGMAT and RBM3 highlights their possible roles in disease pathogenesis and opens new possibilities for early diagnosis and therapy development.

Highlights

- Largest multi-omic characterization in EAC.
- Protein quantitation from 5,879 genes across 23 EAC patients tissues.
- GPA33, significantly over-expressed in EAC.
- Tumor-selective, post-transcriptional protein regulation in EAC.

2024, Mol Cell Proteomics *23*(*6*), 100764

https://doi.org/10.1016/j.mcpro.2024.100764

Multi-Omic Analysis of Esophageal Adenocarcinoma Uncovers Candidate Therapeutic Targets and Cancer-Selective Posttranscriptional Regulation

J. Robert O'Neill[1,2,‡*], Marcos Yébenes Mayordomo[2,3,‡*], Goran Mitulović[4,5], Sofian Al Shboul[6], Georges Bedran[3], Jakub Faktor[3], Lenka Hernychova[7], Lukas Uhrik[7], Maria Gómez-Herranz[3], Mikołaj Kocikowski[3], Vicki Save[8], Bořivoj Vojtěšek[7], Mark J. Arends [9], OCCAMS Consortium, Ted Hupp[2,3], and Javier Antonio Alfaro[3,10,11,12,13,*]

Efforts to address the poor prognosis associated with esophageal adenocarcinoma (EAC) have been hampered by a lack of biomarkers to identify early disease and therapeutic targets. Despite extensive efforts to understand the somatic mutations associated with EAC over the past decade, a gap remains in understanding how the atlas of genomic aberrations in this cancer impacts the proteome and which somatic variants are of importance for the disease phenotype. We performed a quantitative proteomic analysis of 23 EACs and matched adjacent normal esophageal and gastric tissues. We explored the correlation of transcript and protein abundance using tissue-matched RNA-seq and proteomic data from seven patients and further integrated these data with a cohort of EAC RNA-seq data (n = 264 patients), EAC whole-genome sequencing (n = 454 patients), and external published datasets. We quantified protein expression from 5879 genes in EAC and patient-matched normal tissues. Several biomarker candidates with EAC-selective expression were identified, including the transmembrane protein GPA33. We further verified the EAC-enriched expression of GPA33 in an external cohort of 115 patients and confirm this as an attractive diagnostic and therapeutic target. To further extend the insights gained from our proteomic data, an integrated analysis of protein and RNA expression in EAC and normal tissues revealed several genes with poorly correlated protein and RNA abundance, suggesting posttranscriptional regulation of protein expression. These outlier genes, including SLC25A30, TAOK2, and AGMAT, only rarely demonstrated somatic mutation, suggesting post-transcriptional drivers for this EAC-specific phenotype. AGMAT was demonstrated to be overexpressed at the protein level in EAC compared to adjacent normal tissues with an EAC-selective, post-transcriptional mechanism of regulation of protein abundance proposed. Integrated analysis of proteome, transcriptome, and genome in EAC has revealed several genes with tumor-selective, posttranscriptional regulation of protein expression, which may be an exploitable vulnerability.

Esophageal cancer is the seventh most common cancer worldwide and the sixth leading cause of cancer death (1). Esophageal adenocarcinoma (EAC) is now the prevalent histological subtype in Western countries and continues to increase in incidence (2, 3). Gastro-esophageal reflux is thought to be the main driver of the development of EAC as refluxed bile and gastric acid combine to create a carcinogenic environment in the distal esophagus. In this environment, the normally squamous lined esophagus undergoes a metaplastic change to columnar mucosa, known as Barrett's esophagus, which can progress

From the [1]Cambridge Oesophagogastric Centre, Addenbrooke's Hospital, Cambridge, United Kingdom; [2]Institute of Genetics and Cancer (IGC), University of Edinburgh, Edinburgh, Scotland; [3]International Center for Cancer Vaccine Science (ICCVS), University of Gdansk, Gdansk, Poland; [4]Clinical Department of Laboratory Medicine Proteomics Core Facility, Medical University Vienna, Vienna, Austria; [5]Bruker Austria, Wien, Austria; [6]Department of Pharmacology and Public Health, Faculty of Medicine, The Hashemite University, Zarqa, Jordan; [7]Research Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, Brno, Czech Republic; [8]Department of Pathology, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom; [9]Edinburgh Pathology, Institute of Genetics and Cancer (IGC), University of Edinburgh, Edinburgh, Scotland; [10]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, UK; [11]International Centre for Cancer Vaccine Science, University of Gdańsk, Gdańsk, Poland; [12]Department of Biochemistry and Microbiology, University of Victoria, Victoria, Canada; [13]The Canadian Association for Responsible AI in Medicine, Victoria, BC, Canada

[‡]Co-first authors contributed equally and can prioritize their name when adding this paper's reference to résumés.

*For correspondence: Javier Antonio Alfaro, Javier.alfaro@proteogenomics.ca; J. Robert O'Neill, robertoneill@nhs.net; Marcos Yébenes Mayordomo, marcos.yebenes@gmail.com.

https://doi.org/10.1016/j.mcpro.2024.100764

over time through dysplasia to EAC (4). If treated by resection at a mucosa-confined, often asymptomatic, early stage, the 5-year overall survival approaches 90% (5). However, despite recent advances in the detection and surveillance of Barrett's esophagus, the majority of patients with EAC still present with symptomatic, advanced disease, and the prognosis remains grave with a 5-year overall survival of less than 20% (6, 7).

Encouraging advances have been made in systemic chemotherapy regimens for EAC (8), yet most remain resistant to chemotherapy. The lack of validated early detection methods, the late presentation of symptomatic cancers, and the aggressiveness of this tumor type combine to lead to a highly lethal disease (9). Identifying an EAC-specific biomarker could improve the detection of invasive disease within a premalignant Barrett's esophagus segment using either advanced imaging techniques during endoscopy or on *ex vivo* cytology specimens or have application in imaging for cancer staging or, in selected cases, as a therapeutic target (10).

Despite knowledge of the genetic changes in EAC developed during large-scale sequencing efforts, no dominant driver oncogene or EAC-specific biomarker has been identified (11) and this impairs efforts to develop screening tools. Many imaging tools are directed at protein targets, yet few studies of global protein abundances have been performed in EAC. Shotgun proteomics is an ideal technique to explore tissue protein abundances in a hypothesis-free context (12). Our previous work remains one of the most extensive proteomic studies to date in EAC and explored protein abundances by shotgun proteomics in seven patients (13). In this current study, we significantly extend this to 23 patients but also improve proteome coverage by leveraging newer generation mass spectrometry (MS) instruments.

Previous studies have explored EAC pathogenesis by genomic or transcriptomic methods (14, 15) or proteomics in isolation (16, 17). With developments in sequencing and proteomic techniques, combining DNA, RNA-seq, and proteomic analysis from the same tissue has become a realistic prospect and can illuminate mechanisms of protein abundance regulation and potentially reveal EAC-specific biomarkers obscured in previous single-omic studies.

In this biomarker discovery study, we characterize changes in protein and RNA abundance in EAC and generalize these results to a broader cohort. We identified EAC-specific changes in the ratio of protein to RNA abundance, which we propose originate through posttranscriptional regulatory mechanisms. These new insights may allow the development of EAC-specific diagnostic or therapeutic tools.

EXPERIMENTAL PROCEDURES

Ethical and Institutional Approvals

All patients gave prospective, written, informed consent to the use of tissue, clinical data, and publication of results arising. Ethical approvals were provided by the Lothian Local Research Ethics Committee, the Tayside Committee on Medical Research Ethics, and the National Research Ethics Service (NRES) Committee East of England, Cambridge South (references 06/S1101/16, 10/S1402/33, 10/H0305/1). Institutional approval was provided by the National Health Service (NHS) Lothian Research and Development (references 2006/W/PA/01 and 2011/W/ON/27). Patients were deidentified at the time of consent to the use of tissue and clinical data and therefore no patient-identifiable data have been included. The studies in this work abide by the Declaration of Helsinki principles.

Experimental Design and Statistical Rationale

Sample Size—Proteomic data have been included in this work from samples from a total of 23 patients with EAC analyzed by tandem mass tag (TMT) LC-MS/MS. Raw peptide reporter ion data were obtained for seven patients from our previous publication (13), and we generated new data for 16 patients. All subsequent analysis steps in the proteomic informatic analysis included all 23 patients. Due to the very limited available published data on the EAC proteome and as this was a hypothesis-generating study, a formal power calculation was not undertaken to determine the sample number for proteomic analysis. This was determined based on similar discovery-phase proteomic studies using human tumor tissue and the availability of high-quality clinical material.

Due to the risk of false-positives because of the relatively small sample size in the discovery cohort, verification of proposed proteomic biomarkers was performed by immunohistochemistry (IHC) using a tissue microarray containing formalin-fixed, paraffin-embedded cores from multiple tissue types from an independent cohort of 115 patients with EAC.

Our previous proteomic data suggested high technical reproducibility with the majority of sample variance coming from between patient variability (13). The number of biological, rather than technical, replicates was maximized for the available reporter ion channels to address this (supplemental Fig. S1, *B–F*).

Tissue Sampling—Fresh biopsies surplus to diagnostic requirements representing tumor, proximal normal squamous esophagus, and distal normal gastric tissue were collected from resection specimens from patients undergoing surgery for locally advanced esophageal or esophagogastric junctional adenocarcinoma within 30 min of specimen resection. The use of patient-matched, tumor adjacent normal tissues provided a patient level control to identify tumor-enriched protein expression. Biopsies were snap-frozen in liquid nitrogen and maintained at −80 °C until processing. All biopsies were deidentified at the time of collection, subjected to frozen section and histopathological review by a specialist upper gastrointestinal pathologist to confirm tissue of origin, and only tumor samples with >50% tumor cellularity proceeded to analysis with the remaining biopsy used for proteomic analysis and, in cases with sufficient material, DNA and RNA extraction and sequencing.

Tissue Processing for Proteomics—Tissue samples (median sample mass = 17 mg) were processed to tryptic peptides as previously described (13) (detailed protocol in Supplemental Methods). Briefly, samples were homogenized using a bead-mill, lysed in 4% SDS lysis buffer at 20:1 buffer to biopsy mass, sonicated, and buffer-exchanged using 30 kDa cut-off spin columns (Millipore). Buffer-exchanged lysates were reduced and alkylated before protein concentration determination using the RC-DC protein assay (Bio-Rad) according to the manufacturer's instructions. Lysates were digested overnight at 37 °C using trypsin (sequencing grade, Promega) at a 1:100 enzyme-to-protein ratio (*w/w*). Tryptic peptides solutions were lyophilized using a SpeedVac concentrator (Savant SPD121P, Thermo Fisher Scientific) and resuspended in 50 µl 100 mM triethylammonium bicarbonate prior to tandem mass tag (TMT) labeling.

TMT Labeling—TMT label tags were individually resuspended in 41 μl of anhydrous acetonitrile at room temperature (RT). Tryptic peptide samples were labeled with 20.5 μl of the corresponding TMT label reagent (supplemental Fig. S1*A*) at RT for 1 h before reaction quenching using 4 μl of 5% hydroxylamine for 15 min. TMT-labeled samples were pooled in low peptide retention tubes (Pierce), lyophilized, reconstituted in 0.1% (*v/v*) formic acid in MS-grade water (Thermo Fisher Scientific), desalted using Micro SpinColumns C18 (Harvard Apparatus), and lyophilized again.

Fractionation—The samples from seven initial patients described in our prior published work were fractionated by OFF-GEL electrophoresis as previously described (13). Subsequent pooled TMT-labeled samples for this work were fractionated using the Pierce high-pH reversed-phase peptide fractionation kit (Thermo Fisher Scientific) into eight peptide fractions according to the manufacturer's instructions (Supplemental Methods). All fractions and the subsequent flow-through were then subjected to LC-MS/MS.

Mass Spectrometry—Individual peptide fractions were subjected to mass-spectrometry (MS) using a Q-Exactive Orbitrap Plus equipped with the Flex nano-electrospray ionization source and stainless steel needle (20 μm ID × 10 μm tip ID). The needle voltage was set to 3.1 kV, and the scan range was 200 to 2000 *m/z*. Full MS resolution was set to 70,000, automated gain control (AGC) target to 3×10^6, and maximum injection time was set to 50 ms. For MS/MS analysis, the mass resolution was set to 35,000, the AGC target to 1×10^5, and the maximum injection time to 120 ms. The isolation width for tandem mass spectrometry (MS/MS) was set to *m/z* 1.5, and the top 15 ions were selected for fragmentation, singly charged ions, and ions bearing a charge higher than +7 were excluded from MS/MS. Dynamic exclusion time was set to 20 s.

Database Searching and Peptide Identification—For the 16 patients with new MS data generated in this study, Thermo Fisher Scientific RAW files were converted using msconvert (18) to mzML files. Database searches were performed using MS-GF+ (19) (v2019.04.18) against a custom database of Ensemble v94 proteins concatenated with reverse decoy sequences and common contaminants. Peptide-spectrum matching (PSM) was conducted with the following parameters: 10 parts per million precursor tolerance, tryptic cleavage termini with up to two missed cleavages, peptide length of 6 to 40 amino acids, fixed cysteine carbamidomethylation, fixed TMT 6 plex modification (229.1629 Da) on any lysine terminus, variable modification (–187.152366 Da) on any lysine residue to account for acetylation rather than TMT modification, and variable methionine oxidation. Independent false discovery rate (FDR) calculations were made at each stage of identification using Scavenger (19, 20) (v0.1.29) with a threshold of 1% FDR at the PSM peptide and protein level, and two unique peptides required for protein identification.

Quantitative Analysis of Proteomes—For quantitative analysis, reporter ion intensities were extracted from mzML files using the pyOpenMS (21) package with a 0.01 Da tolerance. Summary peptide reporter ion intensities were derived from the sum of all corresponding PSM reporter ion intensity values.

To allow integration with the paired RNA expression data and avoid confounding from the uncertainty in transcript and protein isoform identification, all peptides mapping to a unique gene (Ensembl gene identifier; ENSG) were grouped under that gene identifier. To obtain relative expression between patient-matched tissues for a given gene, the ratio of peptide reporter ion intensities was calculated and then the geometric mean of ratios was calculated for each gene as previously reported (13). Only peptides unique to an ENSG were used for quantification. A summary relative expression value between tumor and normal esophagus (TvE) and tumor and normal gastric tissue (TvG) was calculated for each gene across patients and technical replicates using a meta-analysis approach with a fixed-effect model as previously described (13). Welch's modified *t* test was used to test the hypothesis that relative protein abundances in the TvE and TvG ratios were not different from the protein abundances of any technical replicates (TvT, EvE, and GvG). The Benjamini–Yekutieli method (22) was used to correct for multiple hypothesis testing, and the FDR- corrected significance threshold was set to $p < 0.05$.

For the correlation of protein and RNA abundance, the geometric mean of all peptide reporter ion intensities for tumor tissues only was calculated for each ENSG for each patient. The meta-analysis with a fixed-effect model approach was then applied to combine patients, weighting the contribution of each patient according to the inverse of the variance of the geometric mean intensity.

External Proteomic Data—Raw peptide abundances from normal tissues were obtained from external published MS data (23, 24). Peptide intensities were processed using the same principles as the EAC samples, pooling all unique peptides that map for each ENSG identifier, and deriving a geometric mean for each ENSG.

To account for differing MS instruments, fractionation methods, and possible experimental batch effects, three sequential steps of normalization were applied to the combined dataset of our study and previous publications. The normalization consisted of a sample loading normalization, followed by the trimmed mean of M-values and quantile normalization (code available on request).

Immunohistochemistry—Tissue microarrays (TMAs) were constructed as previously reported (13). Cores on the array included patient-matched normal esophagus, normal stomach, ECA, normal lymph nodes, and lymph node metastases (where present) (supplemental Fig. S5). TMA blocks were sectioned at a thickness of 5 μm and placed on positively charged slides (Thermo Fisher Scientific) to maximize core adherence. IHC was performed using BOND III autostainer with antibodies to GPA33 (Abcam, ab108938, 1:250 dilution), (Sigma, HPA018858, 1:100 dilution) or insulin-like growth factor–binding protein 1 (IGF2BP1) (Sigma, HPA002037, 1:500 dilution) incubated for 20 min at RT and detection using the Leica Bond Polymer Refine Detection kit (DS9800; Leica Biosystems), following the manufacturer's instructions. Sections representing normal colonic epithelium and normal tonsillar tissue were stained in parallel as positive and negative controls.

Assessment of the IHC intensity staining of TMA cores was performed by two expert histopathologists reaching consensus scores, and these samples were graded from 0 to 3 (0 = nil, 1 = weak, 2 = moderate, and 3 = strong). Representative IHC images and corresponding intensity scores are shown in (supplemental Fig. S2*A*). Cores with significant artifacts (*i.e.*, folded tissue) or loss of tissue material were excluded from the analysis.

EAC RNA-Seq Data—RNA expression values were derived by RNA-seq for 264 EAC samples as part of the ICGC ESAD project as previously described (11). A total of seven tumors evaluated by shotgun proteomics in this study also had paired RNA-seq data generated. For these seven patients, unprocessed FASTQ files were mapped to the hg38 reference genome using STAR aligner (v2.6.1). The resulting binary alignment map (BAM) files were then used to quantify gene expression with RSEM (v1.3.3, https://deweylab.github.io/RSEM/), obtaining normalized transcript per million (TPM) values for each unique ENSG.

RNA expression results from a further 257 EAC samples were derived as described in Frankell *et al.* (11), and expression values in TPM reads were obtained for each unique ENSG quantified for each additional patient.

External RNA-Seq Data—To complement our study of EAC, transcript expression values (obtained as TPM values) from normal esophageal samples and a wide range of other normal tissues were incorporated from two external studies - the Wang *et al.* (23) dataset and the genotype-tissue expression consortium dataset (GTEx consortium) analyses of normal human tissues (25). Transcript expression from normal tissues was pooled with EAC samples and normalized together at the ENSG level using the trimmed mean of M-values and quantile normalization (26). The GTEx data were downloaded from the GTEx Portal on 12/13/2018.

EAC Whole-Genome Sequencing—Whole-genome sequencing data from 454 ECAs generated during the International Cancer Genome Consortium ECA cancer genome project (ICGC ESAD) were included in the study. Tumors were sequenced at a mean depth of 70× coverage in EAC (minimum depth 52×) and a mean depth of 42× coverage in patient-matched germline tissue (minimum depth 29×). Data from 420 tumors were previously published in Frankell *et al.* (11), while data from 34 newly processed samples were included in this study. Raw unprocessed FASTQ files were aligned to the human reference genome (hg38) using bwa (v0.7.17), and duplicate reads were marked using Picard tools (v2.18.23). Genomic variants were called using the genome analysis toolkit best practices Mutect2 (v.4.1.3) and filtered against a germline sample for each patient. Germline variants were also labeled in the same step using data from the 1000 genomes project (27). The resulting variant files were annotated using VCF2MAF (28), combined with Ensembl variant effect predictor, and then processed using custom scripts in R (version 4.1.0, https://www.r-project.org/).

RESULTS

Aberrant Expression in the EAC Proteins

To identify proteins with enriched or EAC-specific expression, patient-matched samples representing the primary tumor and adjacent normal esophagus and stomach were collected from resection specimens from 23 patients undergoing esophagectomy for EAC (Fig. 1, cohort clinical and pathological characteristics in Tables 1 and 2). A total of 5879 gene products were quantified in at least one patient across the three different tissues (1% FDR at the PSM, peptide, and protein levels) (supplemental Tables S1 and S2). Exploring the expression of proteins between adjacent patient-matched tissues revealed relative overexpression of several proteins in EAC compared to surrounding normal tissues (Fig. 2). Although data were collected using various experimental procedures, no batch effects were observed following normalization and technical replicate ratios clustered together (supplemental Fig. S3). Although our ratiometric expression analysis cannot uncover protein abundances that are specific to EAC, we found that none of the proteins detected in the raw data had unique peptide reporter ion signatures that were exclusively present in EAC tissue, indicating that there were no proteins detected with entirely EAC-specific expression. Proteins ubiquitously identified across patients were generally from “housekeeping” or cytoskeletal genes and few proteins with tissue-enhanced expression profiles were detected in more than 75% of patients.

We once again identified epithelial cell adhesion molecule (EPCAM) as being highly expressed in EAC compared to surrounding normal tissues and have previously validated this expression pattern (13). Glycoprotein A33 (GPA33) is a cell surface glycoprotein that has been identified as expressed in intestinal tissues and colorectal and gastric cancer and has a putative role in cell adhesion (29, 30). We quantified GPA33 in eight patients’ tissues and demonstrated 4-fold higher expression in EAC than both patient-matched normal squamous esophagus and normal stomach. Proteins with known expression in intestinal tissues such as intelectin-1 (31) and villin (32) were found to be enriched in EAC, likely reflecting the origin of this cancer from intestinal metaplasia of the esophagus (33). Other proteins previously demonstrated to be overexpressed in EAC were also identified (DSG2 (34), GATA6 (11, 35), or REG4 (36)), as well as many other novel genes presenting enriched expression in EAC.

We focused on EAC-enriched proteins, defining those pragmatically as quantified in the majority of our samples (at least 60%) and significantly overexpressed in EAC compared to both normal esophagus and stomach (FDR-adjusted $p < 0.05$). Nineteen members of the RNA-binding motif (RBM) protein family were identified in this study and four (RBM3, RBM6, RBM25, and RBMX) demonstrated EAC-enriched expression. RBM3 was at least 2-fold overexpressed in EAC compared to both normal esophageal and normal gastric tissues.

Cancer-testis antigens have long been held as compelling targets for specific cancer therapies or biomarkers for cancer screening due to their highly restricted tissue expression (37). We identified several cancer-testis antigens as EAC-enriched, including melanoma-associated antigen family members MAGEA4, MAGEA10, MAGED2, and MAGEB2. This group of genes has a well-established role in other cancers and MAGEA4 has been previously demonstrated to be overexpressed in esophageal cancer (38). Due to its EAC-enriched expression, MAGEA4 is a compelling target for immunotherapeutic approaches. Clinical trials are already underway for MAGEA4-directed adoptive T-cell therapies for patients with MAGEA4-enriched expression in esophageal cancer (39, 40) (https://clinicaltrials.gov/ct2/show/NCT04752358). For this reason, we evaluated another cancer-testis antigen; insulin-like growth factor 2 mRNA binding protein (IGF2BP1) which was expressed in 10% of patients in this cohort. This gene was over-expressed in EAC compared to patient-matched adjacent normal tissues. There is little published literature on the role of this protein in EAC and we, therefore, sought to validate its expression by IHC.

Validation of GPA33 and IGF2BP1 as EAC-Enriched Proteins

IGF2BP1 is an RNA binding protein with expression in a wide range of fetal tissues and several cancers but the expression in adult tissues is limited to the testis, ovary, prostate, and kidney (41). A recent study reported higher

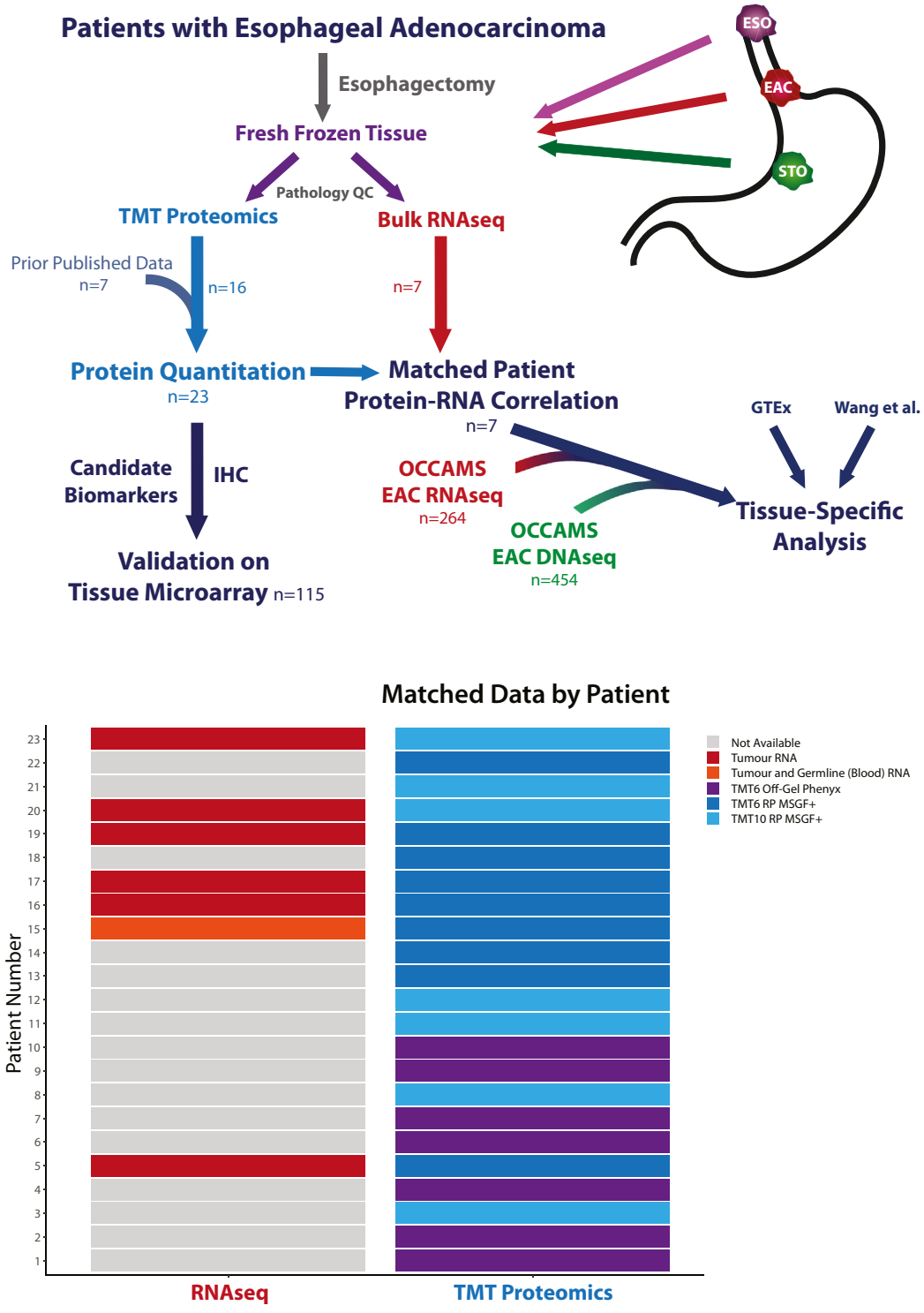


FIG. 1. **Flow diagram of the proteomic and transcriptomic analysis performed.** Global workflow of the proteomic analysis performed using EAC and matched adjacent normal esophagus and normal stomach samples. EAC, esophageal adenocarcinoma.

expression of IGF2BP1 in colonic adenocarcinoma relative to normal colonic mucosa in a set of 13 paired samples (42).

The expression of IGF2BP1 in EAC was evaluated by IHC using a tissue microarray comprising patient-matched cores from formalin-fixed, paraffin-embedded archival tissues representing primary ECA, lymph node metastases (where present), uninvolved lymph nodes, normal gastric mucosa, and normal squamous esophageal samples. A total of 115

TABLE 1
Cohort clinical characteristics

Clinical covariable	Total
Gender	
Male	18 (78.3)
Female	5 (21.7)
Age at diagnosis	
Median (IQR)	66.6 (13.1)
ASA Grade	
I	11 (47.8)
II	11 (47.8)
III	1 (4.3)
Histology	
Adenocarcinoma	23 (100)
Tumour location	
Eso Mid 1/3	1 (4.3)
Eso Low 1/3	12 (52.2)
EGJ I	3 (13.0)
EGJ II	3 (13.0)
EGJ III	4 (17.4)
cT	
T2	1 (4.3)
T3	22 (95.7)
cN	
N0	0 (0)
N1	8 (34.8)
N2	11 (47.8)
N3	4 (17.4)
cM	
M0	23 (100.0)
Preoperative chemotherapy	
2xCF	16 (69.6)
3xECX	2 (8.7)
6xECF	1 (4.3)
None	4 (17.4)

Abbreviations: ASA, American Society of Anesthesiologists; ECF, epirubicin, cisplatin, 5-fluorouracil; ECX, epirubicin, cisplatin, capecitabine; EGJ, esophagogastric junctional; Eso, esophageal; IQR, interquartile range.

TABLE 2
Cohort pathological characteristics

Pathological covariable	Total
pT	
T1b	3 (13.0)
T2	3 (13.0)
T3	15 (65.2)
T4a	2 (8.7)
pN	
N0	6 (26.1)
N1	6 (26.1)
N2	5 (21.7)
N3	6 (26.1)
pM	
M0	23 (100.0)
R	
R0	13 (56.5)
R1	10 (43.5)
TRG	
TRG I-III	0 (0)
TRG IV	6 (26.1)
TRG V	8 (34.8)
Missing	5 (21.7)
N/A	4 (17.4)

Abbreviations: pM, pathological metastasis stage from eighth edition of the UICC TNM cancer staging manual (74); pN, pathological nodal stage; pT, pathological tumor invasion stage; R, resection margin status; TRG, mandard tumor regression grade.

patients' tissues were included in the array, all of whom had undergone esophagectomy for ECA and 75% had no oncological treatment prior to surgery. The clinicopathological characteristics of these patients have been previously reported (13).

Across the TMA, IGF2BP1 was highly expressed in 12% of the EAC samples, a similar prevalence to the discovery cohort undergoing shotgun proteomic analysis (identified in 2/23 patients). Expression was absent or very low in normal esophageal squamous tissue (Fig. 3*A*). On analysis of patient-matched cores, there were five EAC cases that showed expression of IGF2BP1 with absent expression in the patient-matched normal squamous tissue (Figs. 3*B*, C-50, C-25, C-26, C-33, and C-86). These observations further support the quantitative proteomic data. High IGF2BP1 protein abundance was identified in 19% of EAC lymph node metastases (Fig. 3*A*). Although around a quarter of uninvolved lymph nodes (5/20 cases) and normal gastric mucosa (7/31 cases) were classified as high IGF2BP1 protein abundances, the staining was confined mainly to a few scattered lymphocytes or potentially nonspecific staining within gastric glands (supplemental Fig. S2*B*), suggesting that IGF2BP1 tends to be expressed in epithelial-derived tumor cells in EAC. However, more than two-thirds (72%) of the EAC cases showed no protein expression of IGF2BP1 by IHC. Overall, the data suggest that IGF2BP1 protein can be considered to be a moderately tumor-enriched biomarker in EAC, but it may only be useful as a therapeutic target or diagnostic biomarker in a limited proportion of patients. The role of IGF2BP1 expression in lymphocytes remains to be determined but may impact the utility of this biomarker in a clinical context for identifying EAC lymph node metastases.

GPA33 was highly expressed in around a third of the EAC samples (33/110 cases) in the TMA (Fig. 3*C*), which is again consistent with the prevalence in our proteomic cohort (EAC-enriched expression in 8/23 patients). In contrast, only two cases across all of the normal tissues (normal lymph node, normal gastric mucosa, and normal esophageal squamous epithelium) showed similarly high GPA33 expression (Fig. 3*C* uninvolved node). Of the 43 patients with multiple tissue types assessed by IHC, 14 (33%) showed high GPA33 protein levels in primary ECA. While sample numbers are low, there are representative examples of agreement with primary and metastatic lymph nodes with either high primary and lymph node expression (1/1 patients, Fig. 3*D*, C13) or low primary and either low or high lymph node expression (4/4

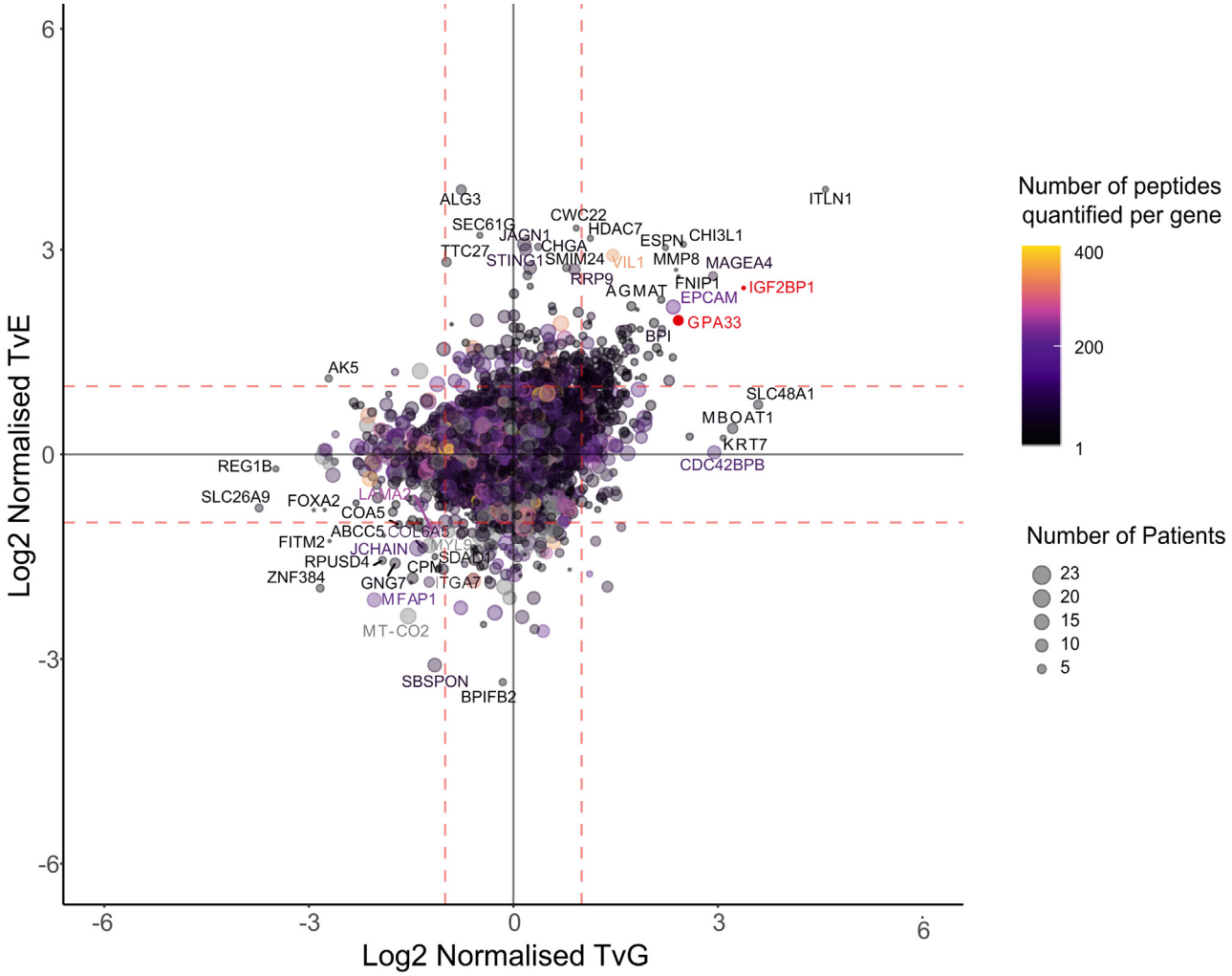


FIG. 2. **The landscape of protein abundance in EAC relative to patient-matched normal esophageal and normal gastric tissue in 23 patients.** Relative expression of 5879 genes across EAC, normal squamous esophagus, and normal stomach in more than one patient. The size of each point indicates the number of patients in which the protein has been quantified, and the color represents the number of peptides quantified per protein. All values correspond to a weighted mean derived across patients with a geometric mean of the ratio of peptide reporter ion intensities across all unique peptides mapping to a gene calculated for each patient. Symbols represent ENSG identifiers.

patients, Fig. 3*D*, C41, C44, C103, C113). For the patients with high GPA33-expressing primary tumors, no normal tissues had high expression. Similarly in all assessed normal tissues (normal lymph node, normal gastric mucosa, and normal esophageal squamous epithelium), there were only two cases of high GPA33 staining (2/37 cores, Fig. 3*D*, C-100, C-107). In both of these uninvolved lymph nodes, the staining appeared to arise in a few lymphocytes (supplemental Fig. S4). We further verified the staining pattern with a second anti-GPA33 antibody (ab108938, Fig. 3*E*) on a subset of the cohort (n = 62 patients) and found a highly significant correlation in staining scores (Rho = 0.822, $p < 0.001$). Overall, these data demonstrate that GPA33 has high EAC specificity and is worthy of further exploration as an EAC biomarker and potential therapeutic target.

Correlation of RNA and Protein Abundances in EAC

The correlation between RNA abundance and protein abundance has been explored in mammalian cell systems and using global methods in normal human tissues and cancer (23–25, 43). Uncovering tissue-specific patterns of the regulation of protein abundance independent of RNA abundance implies a posttranscriptional regulatory mechanism, which could represent a disease biomarker or drug target. The control of protein abundance in normal and diseased tissues is still largely unexplored for many genes and if occurring at the posttranscriptional level would limit the relevance of transcriptome analysis to understanding disease biology. Conversely, the greater genome coverage of RNA-seq provides a distinct advantage over shotgun proteomics, and this technique may be better suited to future translational studies

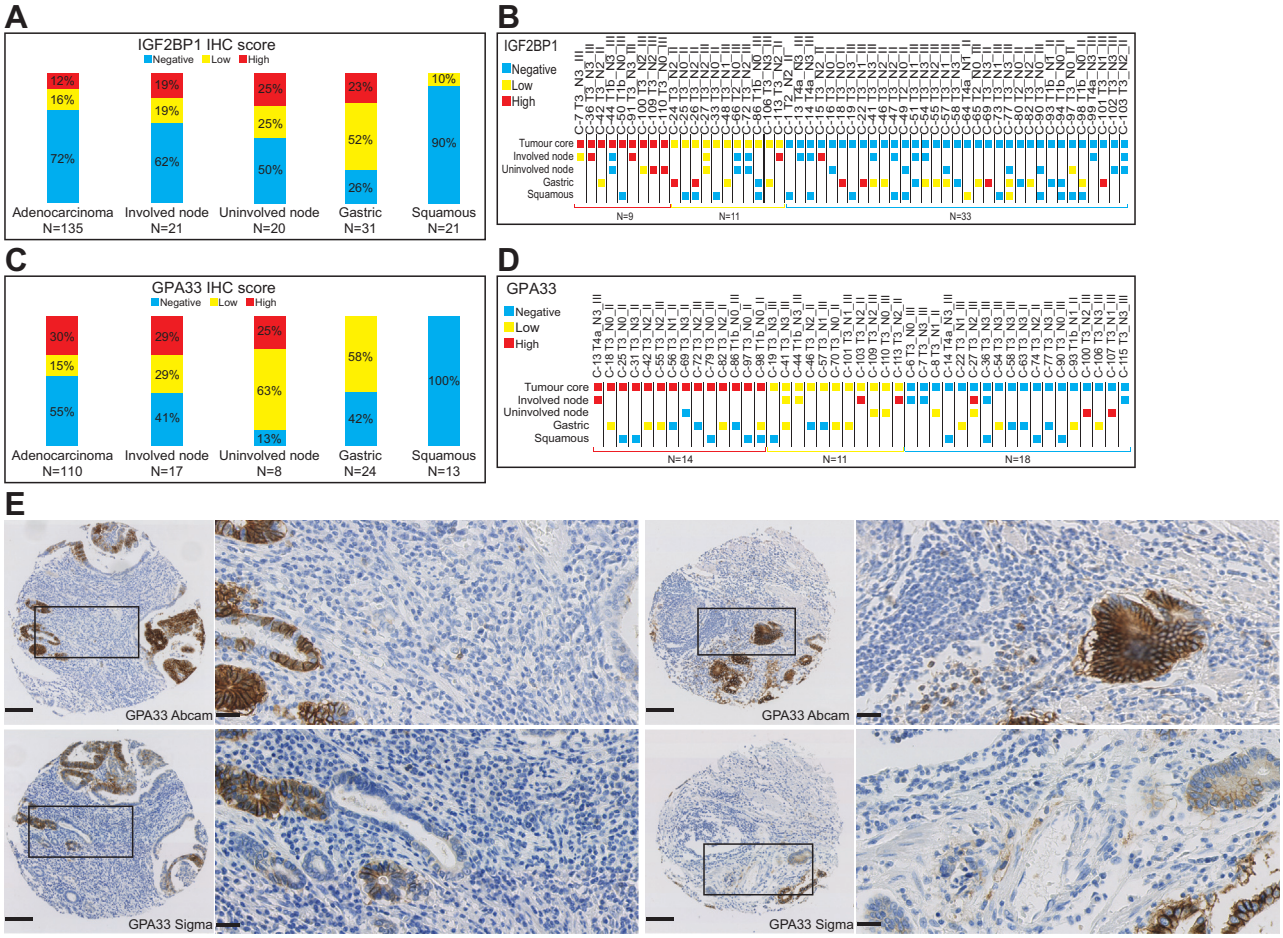


FIG. 3. **Expression of IGF2BP1 and GPA33 across a tissue microarray containing patient-matched primary esophageal adenocarcinoma, involved (metastatic) lymph nodes, uninvolved lymph nodes, normal gastric, and normal squamous esophageal squamous.** *A*, immunohistochemistry (IHC) scores for IGF2BP1 protein abundances according to tissue type. *B*, scoring according to patient-matched tissue samples, including tumor cores and/or involved and uninvolved lymph nodes along with normal gastric and normal esophageal squamous tissues (total n = 53). *C*, IHC scores for GPA33 protein abundances according to tissue type. *D*, scoring according to patient-matched tissue samples, including tumor cores and/or involved and uninvolved lymph nodes along with normal gastric and normal esophageal squamous tissues (total n = 43). *E*, representative GPA33 IHC images showing similar staining patterns of two different anti-GPA33 antibodies. IHC staining was evaluated according to 3,3′-diaminobenzidine (DAB) intensity. In (*B*) and (*D*), each core is identified by a core number, followed by pathological T-stage, N-stage, and tumor grade. GPA33, glycoprotein A33; IGF2BP1, insulin-like growth factor–binding protein 1.

in the case that protein and RNA abundance are well correlated. To explore this in EAC, we correlated the expression of protein and mRNA in seven patients in our proteomic discovery cohort who had sufficient tumor tissue to undergo both RNA-seq and shotgun proteomics (supplemental Fig. S6, *A*–*H*). These tissue-matched datasets present a unique opportunity to independently explore the decoupling of RNA and protein abundance.

We quantified both RNA and protein abundance in EAC tissues in at least one patient across 5531 genes. Although genome coverage and the strength of correlation varied across patients, there was a moderate correlation between RNA and protein abundance overall (Spearman's value of 0.46), in keeping with previous reports from other human tissues (23).

By plotting genes according to both protein and RNA expression, we identify a group of 15 outlier genes with low protein-to-RNA expression ratios (Log_2 protein intensity < 26, Log_2 RNA TPM > 10; Fig. 4*A*, highlighted blue). Although these genes may have been identified in a transcriptome analysis as highly expressed, they are less likely to be of interest from a biomarker perspective due to low relative protein expression. In contrast, we also identified a group of 30 outlier genes with high protein-to-RNA expression ratios (Log_2 protein intensity > 28, Log_2 RNA TPM < 8; Fig. 4*A*, highlighted red). These may not have been identified from a transcriptome-only analysis and may well be biomarker candidates or therapeutic targets.

Several of these genes with high protein-to-RNA ratios are of interest, including RHNO1, CHFR, and CENPE (supplemental Fig. S6*H*). RHNO1 has been recently reported

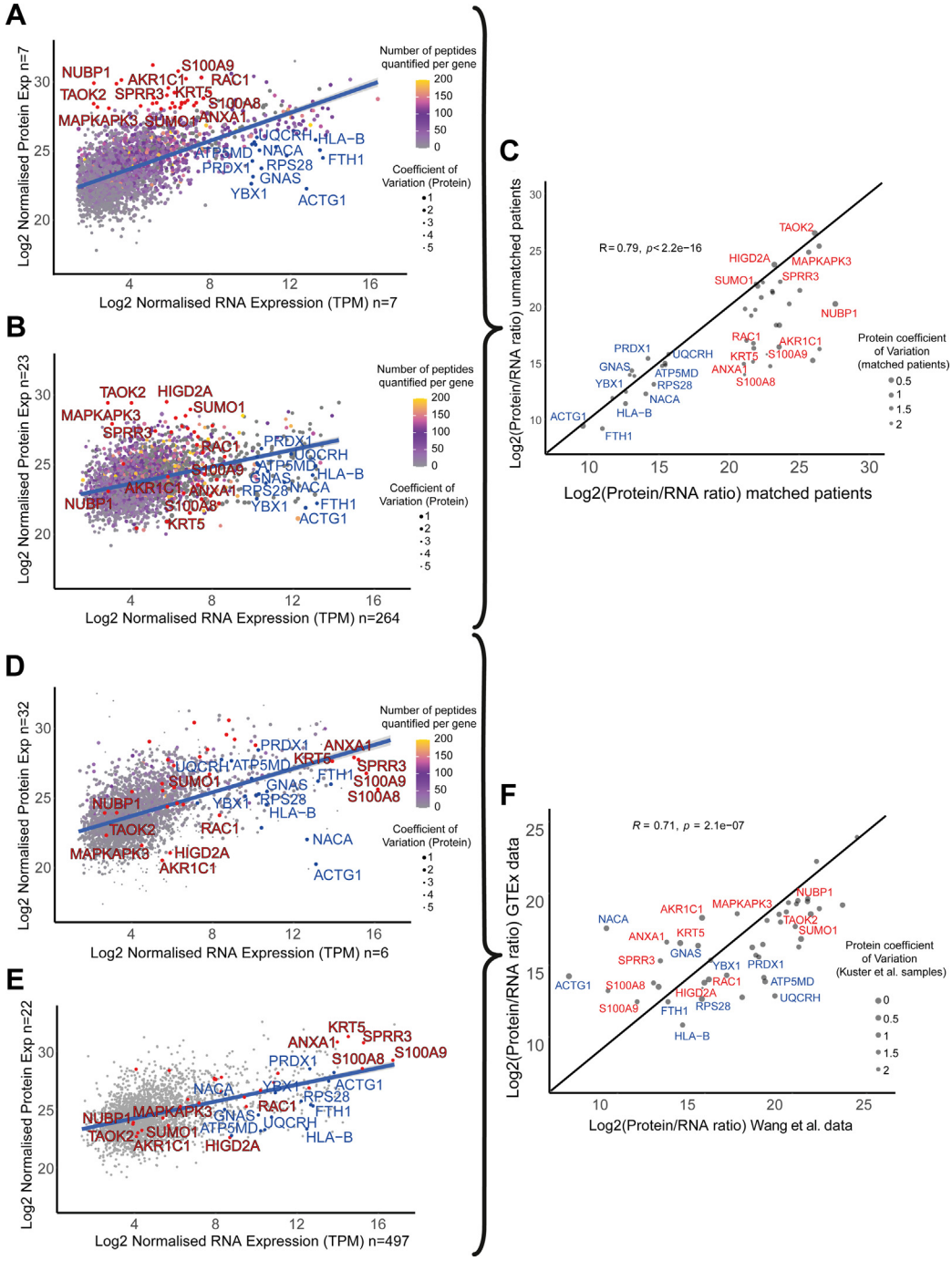


FIG. 4. **Direct correlation of protein and RNA expression in EAC and normal squamous esophagus.** *A*, patient-matched expression of RNA and protein in EAC (n = 7 patients). Protein abundance represents the weighted mean across patients with a geometric mean of the tumor reporter ion abundance calculated for each unique peptide mapping to a gene. RNA abundance has been calculated as the geometric mean across patients of the normalized transcript per million (TPM) values for each unique ENSG. *B*, correlation of RNA and protein abundances in unmatched EAC samples (n = 23 proteomic data, n = 264 transcriptomic data). Abundances have been calculated as in (*A*). *C*, correlation of outlier protein to RNA ratios in matched and unmatched EAC cohorts. *D*, correlation of RNA and protein abundances in the normal squamous esophagus using Wang *et al.* data. Outliers from (*A*) are *highlighted*. *E*, correlation of RNA and protein abundances in the normal squamous esophagus using GTEx consortium data. Outliers from (*A*) are *highlighted*. *F*, correlation of outlier protein-to-RNA ratios in Wang *et al.* and GTEx datasets. EAC, esophageal adenocarcinoma; GTEx , genotype-tissue expression.

as a prognostic marker in colorectal cancer (44), renal, and liver cancer, where the high expression is associated with a poor prognosis (45). CHFR is an E3 ubiquitin ligase with a role in cell cycle checkpoints and is known to be downregulated EAC at the RNA level (46, 47). This was again demonstrated in this analysis yet CHFR was highly expressed at the protein level, highlighting the need to corroborate RNA expression results with data at the protein level. Another gene following this pattern of expression is CENPE, which is associated with lung adenocarcinoma cell proliferation (48), and the expression is correlated with poor prognosis in EAC (49). It is possible that protein abundance is not transcriptionally controlled for these genes and therefore in future work their expression must be studied directly at the protein level.

Tissue-Matched Protein-to-RNA Expression Ratios are Representative of Global EAC Protein-to-RNA Expression Ratios

We have directly assessed protein-to-RNA expression ratios by undertaking both proteomic and RNA-seq analysis from the same tumors from seven patients, yet it is not clear if this small sample is representative of protein-to-RNA expression ratios in EAC in general. To assess this, we used RNA expression data derived from RNA-seq analysis of 264 EACs from patients in the OCCAMS study taking part in the ICGC ESAD project. Protein expression in EAC alone derived from the pooled analysis of the 23 patients presented in Figure 2 was plotted against RNA expression derived from the OCCAMS cohort on a gene-wise basis (Fig. 4*B*).

Most outlier genes with either a high or low protein-to-RNA ratio identified on the matched tissue analysis presented a consistent ratio in the unmatched analysis (Fig. 4*C*, Spearman's r = 0.79, $p < 0.001$). Where differences were observed, this was usually down to a difference in the quantified protein abundances level accompanied by a high coefficient of variation between samples, *e.g.*, S100A8, S100A9 or AKR1C1. This may reflect challenges associated with protein quantitation using a shotgun proteomic strategy and our stringent approach using only uniquely identified peptides for protein quantification. We, therefore, conclude that for most of our outlier genes, the protein-to-RNA ratios identified in EAC are robust.

Exploring Tissue Differences in Protein-to-RNA Expression Ratio

Previous reports from analysis across a range of normal human tissues suggest that protein-to-RNA ratio is both a surrogate of translation rate and is gene intrinsic. It has been proposed that protein-to-RNA abundance ratio, although widely variable between genes, is preserved for a specific gene across multiple tissue types (50–52). We used two external datasets of tissue-matched RNA and proteomic analysis, the GTEx consortium dataset, and the Wang *et al.* dataset to evaluate the protein-to-RNA ratio of outlier genes identified in our analysis of EAC in normal squamous esophageal tissue.

Outlier genes with consistent protein-to-RNA ratios across both matched and unmatched EAC cohorts demonstrated significantly different protein-to-RNA ratios in normal squamous esophageal tissue in both the Wang *et al.* data (Fig. 4*D*) and the GTEx data (Fig. 4*E*). Despite disparate experimental strategies and quantitative proteomic methods between these two large-scale studies, the protein-to-RNA ratios for outlier genes were well correlated between studies (Fig. 4*F*, Spearman's r = 0.71, $p < 0.001$). This raises the possibility that there are EAC-specific mechanisms of posttranscriptional control of protein abundance at least for these outlier genes.

Several outlier genes (KRT5, ANXA1, SPRR3, S100A8, and S100A9) demonstrated lower RNA expression in EAC when than the normal esophageal tissue while still being highly expressed at the protein level in both tissues. RHNO1 was also demonstrated to have very low RNA expression in EAC, leading to very high protein-to-RNA ratios in comparison with normal tissues.

In contrast, when comparing the EAC data with the external normal squamous esophageal data, other dysregulated genes, including TAOK2, MAPKAPK3, and HIGD2A, appear to have a higher protein abundance in EAC than normal squamous esophageal tissue while presenting similar transcript expression levels.

Exploring Changes of Protein to RNA Expressions Across Tissues

We next compared protein-to-RNA ratios for outlier genes in matched and unmatched EAC cohorts and the wider range of normal tissues assessed in the GTEx and Wang *et al.* studies to determine if these findings were generalizable to other normal human tissues.

To shortlist 45 genes, we employed two distinct thresholding criteria based on the combination of RNA intensities and protein abundances in tumor samples. These thresholds allowed us to capture a comprehensive set of genes, demonstrating significant differences in RNA and protein expression levels in tumor tissues.

The protein-to-RNA ratios across EAC and multiple tissues for several outlier genes are summarized in Figure 5. We focus on genes with extreme protein/RNA ratios or that display significant differences in expression between tumor and healthy tissues. To further explore the potential deregulation of protein from RNA abundance during the development of EAC from Barrett's, genes associated with intestinal differentiation or Barrett's esophagus (AGMAT, HMGA1, EPHB2, OLFM4) (33) were also included in Figure 5.

The striking finding is that the protein-to-RNA ratio is similar for each gene across a wide range of normal tissues, in keeping with prior reports that this is a gene-intrinsic property. In contrast, however, in our analysis of EAC, there are widely different protein-to-RNA ratios that are consistent between

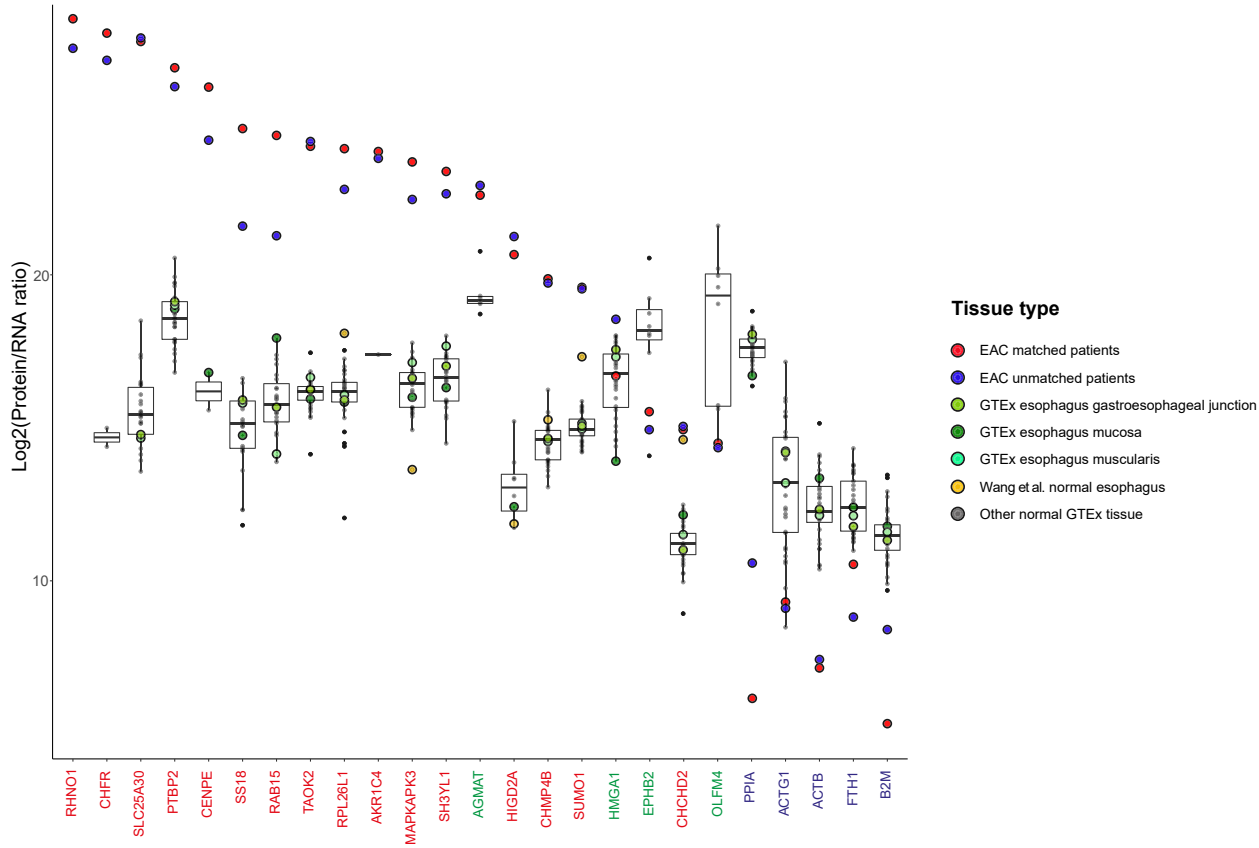


FIG. 5. **Distribution of protein-to-RNA ratios for outlier genes across normal tissues and EAC.** The genes included in this figure are the outliers in Figure 4*A*. Each *colored point* summarizes an esophageal tissue type. Normal tissues from other anatomical regions have been summarized as a *boxplot* with median, box limits as 25th and 75th centile, and tails representing maximum and minimum values as well as outliers. Dysregulated genes with high protein abundances and low RNA expression were colored in *red* in the *x*-axis, as well as genes associated with intestinal differentiation or Barrett's esophagus (*green*) and genes with low protein abundances and high RNA expression (*blue*). EAC, esophageal adenocarcinoma.

matched and unmatched cohorts for several outlier genes, suggesting this property is not gene-intrinsic for these genes in EAC.

For the genes previously reported to be associated with intestinal differentiation or Barrett's esophagus (HMGA1, EPHB2, and OLFM4) the protein-to-RNA ratios were similar between normal tissues and EAC, suggesting regulation of protein abundance is constant for these genes during the development of EAC from Barrett's esophagus. AGMAT, however, has a higher protein-to-RNA ratio in EAC than a wide range of normal tissues, suggesting this regulation may be altered during Barrett's carcinogenesis. We previously demonstrated that AGMAT was significantly overexpressed at the protein level in EAC compared to the patient-matched normal squamous esophagus and gastric tissue (Fig. 2), and AGMAT has recently been described with a role in cancer by promoting tumorigenesis *via* MAPK signaling (53). Together, these data confirm AGMAT is overexpressed in EAC, and we propose this overexpression is post-transcriptionally controlled. Further work is required to understand the mechanism and the role of AGMAT in EAC development.

Dysregulated Protein-to-RNA Ratio is not Driven by Somatic Mutation

EAC is associated with a significant tumor mutational burden (54), and so we next sought to determine if the somatic mutation was correlated with posttranscriptional changes in protein abundance for the outlier genes. Somatic variants were only detected in 15 of the 45 outlier genes, with only 13 of them presenting amino acid changes and the mutant allele frequency ranged from <1 to 3% (supplemental Fig. S7). This low rate of somatic mutation, affecting less than 3% of the total population despite high tumor sequencing coverage, suggests somatic variants are not the driver of the protein-to-RNA expression ratio changes for these outlier genes. We, therefore, conclude the EAC-specific posttranscriptional change in protein abundance revealed for the outlier genes is driven by an alternative mechanism.

DISCUSSION

In this study, we describe the expressed protein landscape in EAC and matched adjacent normal tissues from 23 patients undergoing surgical resection for locally advanced disease and describe the relative expression of at least 5879 proteins in these tissues. To our knowledge, this represents the largest characterization of relative protein abundances in EAC in the literature to date. In keeping with other published shotgun proteomic studies, the majority of proteins were only identified in a subset of the 23 patients. This may reflect a combination of low proteome coverage, an intrinsic limitation in the shotgun proteomic strategy, and interpatient heterogeneity. Despite this, we gain valuable new insights into the expressed protein landscape in this cancer.

We once again confirm EPCAM as highly expressed relative to surrounding normal tissues and have previously validated the EAC-specific expression of EPCAM (13). We also identified several RBM protein family members as demonstrating EAC-enriched expression, including RBM3. In support of this finding, RBM3 has been previously implicated in cancer and reported to be overexpressed in esophageal and gastric adenocarcinoma with particularly high expression in those cancers with a background of intestinal metaplasia (55, 56).

We report several further proteins with EAC-enriched expression, some of which are novel but others (INTL1, VIL1, OLFM4, REG4, and ANXA13) have been previously described as expressed selectively in intestinal tissue or goblet cells, a histological marker of intestinal differentiation. We propose this may reflect the origin of EAC from glandular metaplasia of the esophagus. We further identify and validate the EAC-enriched expression of GPA33 in a third of patients with EAC and the overexpression of IGF2BP1 in 10% of patients with EAC. The limited expression of GPA33 in patient-matched normal tissues provides a compelling basis for further development of this as an EAC biomarker for imaging and cancer detection or development as a therapeutic target. GPA33 has previously been demonstrated to be highly expressed in gastric and colorectal cancer and human anti-GPA33 antibodies are being evaluated as immuno-oncological treatments (57–59). Early phase clinical trials of monoclonal anti-GPA33 showed good safety and tolerability (60), while other clinical trials investigating novel anti-GPA33 antibodies in colorectal cancer are ongoing (NCT02248805).

Although we did not identify any background normal squamous expression and only limited nonspecific gastric mucosal expression of GPA33, two cores from normal lymph nodes across our validation TMA cohort demonstrated some scattered GPA33-expressing cells, presumed to be lymphocytes (supplemental Fig. S4). A recent study has confirmed GPA33 expression in a subset of CD4-positive T-regulatory cells with an immunosuppressive phenotype, which would support our findings (61). The role of these cells in cancer remains to be determined.

We further extended our analysis to seven patients in this cohort by quantifying RNA expression by RNA-seq in the same EAC tissues subjected to quantitative shotgun proteomics. We initially explored the correlation between protein and RNA abundance in this tissue. The changes in expression between protein and RNA abundances and how they are regulated as part of normal tissue homeostasis are still being delineated (62) and are only recently being evaluated in malignancy (63). However, it is clear that both protein degradation and synthesis pathways can be tightly regulated (64). The former through post-translational modifications of proteins targeting them toward proteasomal degradation and the latter arises through post-transcriptional regulation decorrelating RNA and protein abundances. Mutation, copy number variation, and epigenetic modification are well established to modify protein abundances, and regulation of protein abundance *via* posttranslational or posttranscriptional pathways is likely another exploitable mechanism.

The combined analysis of protein and RNA expression in tissue-matched EAC patients revealed two groups of outlier genes with either high or low protein-to-RNA expression ratios. We validated these findings by comparison of protein abundance in EAC from our full proteomic cohort of 23 patients with RNA abundance from 264 EACs. We demonstrate a moderate to good correlation (above 0.5 and below 0.9) in protein-to-RNA ratios between the matched and unmatched cohorts for these outlier genes, suggesting our cohort of seven patients was a representative sample of EAC.

We next sought to determine if the protein-to-RNA ratios observed for the outlier genes in EAC were consistent across other tissue types, and therefore a gene-intrinsic phenomenon, or if these findings were solely observable in EAC, a tissue-specific phenomenon.

There have been two significant efforts to study the correlation of RNA and protein abundances in normal tissue: the GTEx project (24, 25) and the Wang *et al.* (23) dataset. We used the available raw data from these studies to evaluate protein-to-RNA ratios across normal tissues for our outlier genes and found protein-to-RNA ratios to be well correlated between these studies. Protein-to-RNA ratios for the EAC outlier genes were similar across a wide range of normal human tissues in support of the previous reports that protein-to-RNA ratio is gene intrinsic.

To our surprise, several outlier genes had very consistent protein-to-RNA ratios across both matched and unmatched EAC cohorts but demonstrated significantly different protein-to-RNA ratios across all the normal tissues assessed in GTEx and Wang *et al.* studies. We conclude that for these genes, protein abundance is controlled by a post-transcriptional mechanism, which is altered in EAC. This offers a potentially cancer-specific targetable mechanism and is worthy of further exploration.

Outlier genes were associated with several pathways including those involved in mitochondrial function and the p38/MAPK signaling cascade. Both SLC25A30 and PTBP2 are described to influence mitochondrial function and promote cancer cell proliferation (65, 66). Prior work has suggested S100A9 is downregulated in the progression from Barrett's metaplasia to EAC but this only assessed expression at the mRNA level (67), in contrast in a further study, the protein has been found to be overexpressed in EAC in comparison to Barrett's tissue samples and upregulated in serum in patients with EAC (68). Our findings further support these data and emphasize the importance of evaluating protein abundances directly.

Both MAPKAPK3 and TAOK are members of the p38 MAPK signaling pathway, which has a well-established role in oncogenesis and is a potential therapeutic target (69, 70). HIGD2A is upregulated in response to hypoxia (71), a likely feature of the tumor microenvironment. It is possible, for these outlier genes, post-transcriptional regulation of protein abundance differs in EAC. It is not clear if these effects are unique to EAC or more broadly represent a property of malignant or diseased tissue. For selected candidates, altered post-transcriptional regulation may offer a tumor-specific property that could be exploited for diagnosis or therapy.

AGMAT was demonstrated to have a high protein-to-RNA ratio in EAC and a significantly lower protein-to-RNA ratio in other normal tissues. AGMAT was also overexpressed at the protein level in EAC compared to both patient-matched normal squamous esophagus and normal stomach. Based on these data, we propose that the mechanism of overexpression of AGMAT in EAC is posttranscriptional. The role of AGMAT in EAC and the exact mechanism leading to protein overexpression are worthy of further exploration in follow-up studies.

We have directly assessed global protein abundances in EAC and patient-matched adjacent normal tissues in 23 patients but have only directly assessed RNA expression in the same EAC tissue from seven of these patients. It is possible our findings regarding protein-to-RNA expression ratio in EAC were limited by statistical power. For this reason, we evaluated RNA expression in a large external cohort of 264 EACs and compared it with protein abundances from our cohort of 23 patients. In support of our findings, there was an excellent correlation in protein to RNA expression between patient-matched and unmatched cohorts.

We have also iterated our proteomic methods across time in collecting these data. From our previous proteomic study, the largest source of biological variation was found to be inter-patient heterogeneity rather than technical variation. To address this, we used data from the maximum number of patients rather than selecting those with a more homogenous experimental strategy. In support of our quantitative proteomic findings, many of our EAC-enriched proteins have been previously published as upregulated in EAC and both GPA33 and IGF2BP1 were overexpressed in a very similar proportion of our discovery cohort and the external TMA validation cohort.

The use of diverse proteomic and transcriptomic methods in both GTEx and Wang *et al.* studies has the potential to confound our conclusions. However, we have applied robust normalization to deal with batch effects and we have only explored relative abundances to further limit the impact of variable quantitative dynamic ranges. A good correlation was achieved for these two studies when comparing outlier protein-to-RNA ratios in normal squamous esophageal tissue in support of these methods. The use of these large external datasets provided the capacity to explore protein to RNA in a wide range of tissues beyond that achievable using our dataset. We caution that detailed validation of our findings is still required for candidate genes, where we propose EAC-specific posttranscriptional regulation of protein abundances. If confirmed, however, this novel finding may provide a tumor-specific targetable mechanism with potential applications in diagnosis and therapy. The identification of protein products arising from the candidate genes examined in this study can be accomplished using blood samples. According to the human protein atlas, RBM3, TAOK, and AGMAT can be detected using MS, GPA33 can be identified through proximity extension assay, and S100A can be quantified using immunoassay techniques.

We hold the belief that for this cancer, early detection is likely to be best achieved *via* a luminal endoscopic or cytological technique. In the future, it is possible that the plasma proteome may hold value once the luminal environment is understood, biomarker candidates identified and verified as tractable for the early disease detection, and these robust biomarkers confirmed to be detectable in plasma.

In this study, we use proteomic methods to present the landscape of protein abundances in EAC and matched adjacent tissues. We identify several EAC-enriched proteins and externally validate GPA33 as overexpressed in a third of patients with EAC. Therapeutic trials are already underway using GPA33-directed therapies in other cancers, and our finding of EAC-specific expression of this cell surface protein in a significant proportion of patients provides evidence to extend trials to consider patients with EAC.

We provide the first combined global analysis of protein and RNA expression in EAC and identify several novel genes with high or low protein-to-RNA expression ratios, suggesting the decorrelation of RNA and protein abundance. We further confirm these findings are specific to EAC for several candidate genes by undertaking a combined analysis of protein and RNA expression across multiple normal tissue types using external datasets and confirm these genes are not dysregulated by somatic mutation. We finally identify AGMAT as overexpressed at the protein level in EAC compared to adjacent normal tissues and identify the mechanism of overexpression as post-transcriptional which, due to its EAC specificity, may offer a targetable vulnerability.

DATA AVAILABILITY

The MS data have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD042792 and 10.6019/PXD042792.

The raw DNA sequencing data used in this study are deposited at the European Genome-Phenome Archive under accession codes: EGAD00001007785 (whole-genome sequencing of primary tumors and matched normal). The raw sequencing data are available under restricted access due to data privacy laws; access can be requested to the ICGC data access compliance office as described here: https://docs.icgc-argo.org/docs/data-access/daco/applying. The mutation data for the primary EACs are also available at the ICGC data portal (https://dcc.icgc.org/), under accession code ESAD-UK."

Raw RNA-seq data are available to collaborators on application to the OCCAMS consortium (https://www.occams.org.uk/contact.html).

Supplemental Data—This article contains supplemental data (72, 73).

Acknowledgments—Additional infrastructure support and assistance with tissue collection, storage, and tissue microarray construction were provided from the Cancer Research UK–funded Experimental Cancer Medicine Centers in Edinburgh and Cambridge. The authors would like to thank the PL-Grid Infrastructure and CI-TASK, Poland for providing the hardware and software resources and the Human Research Tissue Bank from Addenbrookes Hospital, which is supported by the UK National Institute for Health Research (NIHR) Cambridge Biomedical Research Center.

The genotype-tissue expression (GTEx) project was supported by the common Fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Funding and additional information—OCCAMS was funded by a program Grant from Cancer Research UK (RG66287). S. A. S. is supported by the Deanship of Scientific Research, The Hashemite University: grants no. 785/48/2022 and 738/54/2022. B. V. and L. H. were supported by the European Regional Development Fund, Project ENOCH (CZ.02.1.01/0.0/0.0/16_019/0000868) and the Ministry of Health, Czech Republic, conceptual development of research organization (MMCI, 00209805). The study was supported by the project "International Center for Cancer Vaccine Science", which is carried out within the International Agendas Programme of the Foundation for Polish Science cofinanced by the European Union under the European Regional Development Fund. J. R. O. received support from the CRUK Cambridge Centre Thoracic Cancer Programme (CTRQQR-2021\100012). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health Research.

Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium Authors: Rebecca C Fitzgerald, Paul A W Edwards, Nicola Grehan, Barbara Nutzinger, Christine Loreno, Aisling M Redmond, Sujath Abbas, Adam Freeman, Elizabeth C Smyth, Maria O'Donovan, Ahmad Miremadi, Shalini Malhotra, Monika Tripathi, Calvin Cheah, Hannah Coles, Curtis Millington, Ginny Devonshire, Matthew Eldridge, Maria Secrier, Sriganesh Jammula, Jim Davies, Charles Crichton, Nick Carroll, Richard H Hardwick, Peter Safranek, Andrew Hindmarsh, Vijayendran Sujendran, Stephen J Hayes, Yeng Ang, Andrew Sharrocks, Shaun R Preston, Izhar Bagwan, Vicki Save, Richard J E Skipworth, Ted R Hupp, J Robert O'Neill, Olga Tucker, Andrew Beggs, Philippe Taniere, Sonia Puig, Gianmarco Contino, Timothy J Underwood, Robert C Walker, Ben L Grace, Jesper Lagergren, James Gossage, Andrew Davies, Fuju Chang, Ula Mahadeva, Vicky Goh, Francesca D Ciccarelli, Grant Sanders, Richard Berrisford, David Chan, Ed Cheong, Bhaskar Kumar, L Sreedharan, Simon L Parsons, Irshad Soomro, Philip Kaye, John Saunders, Laurence Lovat, Rehan Haidry, Michael Scott, Sharmila Sothi, Suzy Lishman, George B Hanna, Christopher J Peters, Krishna Moorthy, Anna Grabowska, Richard Turkington, Damian McManus, Helen Coleman, Russell D Petty, Freddie Bartlet

Author contributions—J. R. O., S. A. S., G. M., J. F., L. H., L. U., M. G. H., V. S., M. Y. M., and J. A. A. investigation; J. R. O., B. V. and T. H. funding acquisition; J. R. O., M. Y. M., and J. A. A. conceptualization; J. R. O. and J. A. A. supervision; J. R. O. N., M. Y. M., G. B., and M. K. formal analysis; J. R. O., M. Y. M., S. A. G. B., and M. K. visualization; J. R. O., M. Y. M., S. A. S., G. B., G. M., J. F., L. H., L. U., M. G. H., M. K., V. S., B. V., M. J. A., R. C. F. (OCCAMS), T. H., and J. A. A. writing-review and editing.

Conflict of interest—The authors declare that they have no competing interests.

Abbreviations—The abbreviations used are: EAC, esophageal adenocarcinoma; ENSG, Ensembl gene identifier; FDR, false discovery rate; GPA33, glycoprotein A33; GTEx, genotype-tissue expression; IGF2BP1, insulin-like growth factor–binding protein 1; MS, mass spectrometry; RBM, RNA-binding motif; TMA, tissue microarray; TMT, tandem mass tag; TPM, transcript per million; SDS, sodium dodecyl sulfate.

Received September 19, 2023, and in revised form, March 8, 2024
Published, MCPRO Papers in Press, April 9, 2024, https://doi.org/10.1016/j.mcpro.2024.100764

REFERENCES

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018) Global cancer statistics 2018: GLOBOCAN estimates of

incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424
2. Then, E. O., Lopez, M., Saleem, S., Gayam, V., Sunkara, T., Culliford, A., *et al*. (2020) Esophageal cancer: an updated surveillance epidemiology and end results database analysis. *World J. Oncol.* **11**, 55–64
3. Edgren, G., Adami, H.-O., Weiderpass, E., Nyrén, O., and Nyrén, O. (2013) A global assessment of the oesophageal adenocarcinoma epidemic. *Gut* **62**, 1406–1414
4. Killcoyne, S., and Fitzgerald, R. C. (2021) Evolution and progression of Barrett's oesophagus to oesophageal cancer. *Nat. Rev. Cancer* **21**, 731–741
5. Rahman, S. A., Walker, R. C., Maynard, N., Trudgill, N., Crosby, T., Cromwell, D. A., *et al*. (2021) The AUGIS survival predictor: prediction of long-term and conditional survival after esophagectomy using random survival forests. *Ann. Surg.* **277**, 267–274
6. Thrift, A. P. (2016) The epidemic of oesophageal carcinoma: where are we now? *Cancer Epidemiol.* **41**, 88–95
7. Pilonis, N. D., Killcoyne, S., Tan, W. K., O'Donovan, M., Malhotra, S., Tripathi, M., *et al*. (2022) Use of a Cytosponge biomarker panel to prioritise endoscopic Barrett's oesophagus surveillance: a cross-sectional study followed by a real-world prospective pilot. *Lancet Oncol.* **23**, 270–278
8. Al-Batran, S.-E., Homann, N., Pauligk, C., Goetze, T. O., Meiler, J., Kasper, S., *et al*. (2019) Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. *Lancet* **393**, 1948–1957
9. Coleman, H. G., Xie, S.-H., and Lagergren, J. (2018) The epidemiology of esophageal adenocarcinoma. *Gastroenterology* **154**, 390–405
10. Chen, J., Jiang, Y., Chang, T. S., Rubenstein, J. H., Kwon, R. S., Wamsteker, E. J., *et al*. (2022) Detection of Barrett's neoplasia with a near-infrared fluorescent heterodimeric peptide. *Endoscopy* **54**, 1198–1204
11. Frankell, A. M., Jammula, S., Li, X., Contino, G., Killcoyne, S., Abbas, S., *et al*. (2019) The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat. Genet.* **51**, 506–516
12. O'Neill, J. R. (2019) An overview of mass spectrometry-based methods for functional proteomics. *Methods Mol. Biol.* **1871**, 179–196
13. O'Neill, J. R., Pak, H. S., Pairo-Castineira, E., Save, V., Paterson-Brown, S., Nenutil, R., *et al*. (2017) Quantitative shotgun proteomics unveils candidate novel esophageal adenocarcinoma (EAC)-specific proteins. *Mol. Cell. Proteomics* **16**, 1138–1150
14. Kim, J. (2017) Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175
15. Dai, J. Y., Wang, X., Buas, M. F., Zhang, C., Ma, J., Wei, B., *et al*. (2018) Whole-genome sequencing of esophageal adenocarcinoma in Chinese patients reveals distinct mutational signatures and genomic alterations. *Commun. Biol.* **1**, 1–9
16. Bratlie, S. O., Wallenius, V., Edebo, A., Fändriks, L., and Casselbrant, A. (2019) Proteomic approach to the potential role of angiotensin II in Barrett dysplasia. *Proteomics Clin. Appl.* **13**, e1800102
17. Weke, K., Singh, A., Uwugiaren, N., Alfaro, J. A., Wang, T., Hupp, T. R., *et al*. (2021) MicroPOTS analysis of Barrett's esophageal cell line models identifies proteomic changes after physiologic and radiation stress. *J. Proteome Res.* **20**, 2195–2205
18. Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536
19. Kim, S., and Pevzner, P. A. (2014) MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5**, 5277
20. Ivanov, M. V., Levitsky, L. I., Bubis, J. A., and Gorshkov, M. V. (2019) Scavager: a versatile postsearch validation algorithm for shotgun proteomics based on gradient boosting. *Proteomics* **19**, e1800280
21. Röst, H. L., Schmitt, U., Aebersold, R., and Malmström, L. (2014) pyOpenMS: a Python-based interface to the OpenMS mass-spectrometry algorithm library. *Proteomics* **14**, 74–77
22. Benjamini, Y., and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188
23. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., *et al*. (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503
24. Jiang, L., Wang, M., Lin, S., Jian, R., Li, X., Chan, J., *et al*. (2020) A quantitative proteome map of the human body. *Cell* **183**, 269–283.e19
25. Consortium, T. Gte (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330
26. Robinson, M. D., and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25
27. Auton, A., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., *et al*. (2015) A global reference for human genetic variation. *Nature* **526**, 68–74
28. Kandoth, C., Gao, J., qwangmsk, Mattioni, M., Struck, A., Boursin, Y., *et al*. (2018) mskcc/vcf2maf: vcf2maf v1.6.16 (v1.6.16). *Zenodo*. https://doi.org/10.5281/zenodo.1185418
29. Heath, J. K., White, S. J., Johnstone, C. N., Catimel, B., Simpson, R. J., Moritz, R. L., *et al*. (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 469–474
30. van Niel, G., Raposo, G., Candalh, C., Boussac, M., Hershberg, R., Cerf-Bensussan, N., *et al*. (2001) Intestinal epithelial cells secrete exosome-like vesicles. *Gastroenterology* **121**, 337–349
31. Owen, R. P., White, M. J., Severson, D. T., Braden, B., Bailey, A., Goldin, R., *et al*. (2018) Single cell RNA-seq reveals profound transcriptional similarity between Barrett's oesophagus and oesophageal submucosal glands. *Nat. Commun.* **9**, 4261
32. Maunoury, R., Robine, S., Pringault, E., Léonard, N., Gaillard, J. A., and Louvard, D. (1992) Developmental regulation of villin gene expression in the epithelial cell lineages of mouse digestive and urogenital tracts. *Development* **115**, 717–728
33. Nowicki-Osuch, K., Zhuang, L., Jammula, S., Bleaney, C. W., Mahbubani, K. T., Devonshire, G., *et al*. (2021) Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. *Science* **373**, 760–767
34. Liu, Y.-Q., Chu, L. Y., Yang, T., Zhang, B., Zheng, Z. T., Xie, J. J., *et al*. (2022) Serum DSG2 as a potential biomarker for diagnosis of esophageal squamous cell carcinoma and esophagogastric junction adenocarcinoma. *Biosci. Rep.* **42**, BSR20212612
35. Pavlov, K., Honing, J., Meijer, C., Boersma-van Ek, W., Peters, F. T. M., van den Berg, A., *et al*. (2015) GATA6 expression in Barrett's oesophagus and oesophageal adenocarcinoma. *Dig. Liver Dis.* **47**, 73–80
36. Dai, Y., Wang, Q., Gonzalez Lopez, A., Anders, M., Malfertheiner, P., Vieth, M., *et al*. (2018) Genome-wide analysis of Barrett's adenocarcinoma. A first step towards identifying patients at risk and developing therapeutic paths. *Transl. Oncol.* **11**, 116–124
37. Scanlan, M. J., Simpson, A. J. G., and Old, L. J. (2004) The cancer/testis genes: review, standardization, and commentary. *Cancer Immun.* **4**, 1
38. Tang, W.-W., Liu, Z.-H., Yang, T.-X., Wang, H.-J., and Cao, X.-F. (2016) Upregulation of MAGEA4 correlates with poor prognosis in patients with early stage of esophageal squamous cell carcinoma. *Oncotargets Ther.* **9**, 4289–4293
39. Zhang, Y., Zhang, Y., and Zhang, L. (2019) Expression of cancer–testis antigens in esophageal cancer and their progress in immunotherapy. *J. Cancer Res. Clin. Oncol.* **145**, 281–291
40. Hong, D. S., Jalal, S. I., Elimova, E., Ajani, J. A., Murphy, M. A. B., Cervantes, A., Evans, T. R. J., *et al*. (2022) SURPASS-2 trial design: a phase 2, open-label study of ADP-A2M4CD8 SPEAR T cells in advanced esophageal or esophagogastric junction cancers. *J Clin Oncol* **40**, TPS363
41. Hammer, N. A., Hansen, T. V. O., Byskov, A. G., Rajpert-De Meyts, E., Grøndahl, M. L., Bredkjaer, H. E., *et al*. (2005) Expression of IGF-II mRNA-binding proteins (IMPs) in gonads and testicular cancer. *Reproduction* **130**, 203–212
42. Chen, H.-M., Lin, C. C., Chen, W. S., Jiang, J. K., Yang, S. H., Chang, S. C., *et al*. (2021) Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) is a prognostic biomarker and associated with chemotherapy responsiveness in colorectal. *Cancer Int. J. Mol. Sci.* **22**, 6940
43. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., *et al*. (2011) Global quantification of mammalian gene expression control. *Nature* **473**, 337–342
44. Yang, F., Cai, S., Ling, L., Zhang, H., Tao, L., and Wang, Q. (2020) Identification of a five-gene prognostic model and its potential drug repurposing in colorectal cancer based on TCGA, GTEx and GEO databases. *Front. Genet.* **11**, 622659

45. Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., *et al*. (2017) A subcellular map of the human proteome. *Science* **356**, eaal3321
46. Shibata, Y., Haruki, N., Kuwabara, Y., Ishiguro, H., Shinoda, N., Sato, A., *et al*. (2002) Chfr expression is downregulated by CpG island hypermethylation in esophageal cancer. *Carcinogenesis* **23**, 1695–1699
47. Soutto, M., Peng, D., Razvi, M., Ruemmele, P., Hartmann, A., Roessner, A., *et al*. (2010) Epigenetic and genetic silencing of CHFR in esophageal adenocarcinomas. *Cancer* **116**, 4033–4042
48. Shan, L., Zhao, M., Lu, Y., Ning, H., Yang, S., Song, Y., *et al*. (2019) CENPE promotes lung adenocarcinoma proliferation and is directly regulated by FOXM1. *Int. J. Oncol.* **55**, 257–266
49. Zhu, X., Luo, X., Feng, G., Huang, H., He, Y., Ma, W., *et al*. (2019) CENPE expression is associated with its DNA methylation status in esophageal adenocarcinoma and independently predicts unfavorable overall survival. *PLoS One* **14**, e0207341
50. Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., *et al*. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587
51. Eraslan, B., Wang, D., Gusic, M., Prokisch, H., Hallström, B. M., Uhlén, M., *et al*. (2019) Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. *Mol. Syst. Biol.* **15**, e8513
52. Edfors, F., Danielsson, F., Hallström, B. M., Käll, L., Lundberg, E., Pontén, F., *et al*. (2016) Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.* **12**, 883
53. Zhu, H., Yin, J., Chen, D., He, S., and Chen, H. (2019) Agmatinase promotes the lung adenocarcinoma tumorigenesis by activating the NO-MAPKs-PI3K/Akt pathway. *Cell Death Dis.* **10**, 1–15
54. Secrier, M., Li, X., de Silva, N., Eldridge, M. D., Contino, G., Bornschein, J., *et al*. (2016) Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat. Genet.* **48**, 1131–1141
55. Zhou, R.-B., Lu, X.-L., Zhang, C.-Y., and Yin, D.-C. (2017) RNA binding motif protein 3: a potential biomarker in cancer and therapeutic target in neuroprotection. *Oncotarget* **8**, 22235–22250
56. Jonsson, L., Hedner, C., Gaber, A., Korkocic, D., Nodin, B., Uhlén, M., *et al*. (2014) High expression of RNA-binding motif protein 3 in esophageal and gastric adenocarcinoma correlates with intestinal metaplasia-associated tumours and independently predicts a reduced risk of recurrence and death. *Biomark. Res.* **2**, 11
57. Garinchesa, P., Sakamoto, J., Welt, S., Real, F., Rettig, W., and Old, L. (1996) Organ-specific expression of the colon cancer antigen A33, a cell surface target for antibody-based therapy. *Int. J. Oncol.* **9**, 465–471
58. Wu, Z., Guo, H.-F., Xu, H., and Cheung, N.-K. V. (2018) Development of a tetravalent anti-GPA33/anti-CD3 bispecific antibody for colorectal cancers. *Mol. Cancer Ther.* **17**, 2164–2175
59. Moore, P. A., Shah, K., Yang, Y., Alderson, R., Roberts, P., Long, V., *et al*. (2018) Development of MGD007, a gpA33 x CD3-bispecific DART protein for T-cell immunotherapy of metastatic colorectal cancer. *Mol. Cancer Ther.* **17**, 1761–1772
60. Infante, J. R., Bendell, J. C., Goff, L. W., Jones, S. F., Chan, E., Sudo, T., *et al*. (2013) Safety, pharmacokinetics and pharmacodynamics of the anti-A33 fully-human monoclonal antibody, KRN330, in patients with advanced colorectal cancer. *Eur. J. Cancer* **49**, 1169–1175
61. Opstelten, R., de Kivit, S., Slot, M. C., van den Biggelaar, M., Iwaszkiewicz-Grześ, D., Gliwiński, M., *et al*. (2020) GPA33: a marker to identify stable human regulatory T cells. *J. Immunol.* **204**, 3139–3148
62. Buccitelli, C., and Selbach, M. (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644
63. Kosti, I., Jain, N., Aran, D., Butte, A. J., and Sirota, M. (2016) Cross-tissue analysis of gene and protein expression in normal and cancer tissues. *Sci. Rep.* **6**, 24799
64. Liu, Y., Beyer, A., and Aebersold, R. (2016) On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550
65. Cheung, H. C., Hai, T., Zhu, W., Baggerly, K. A., Tsavachidis, S., Krahe, R., *et al*. (2009) Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. *Brain* **132**, 2277–2288
66. Rochette, L., Meloux, A., Zeller, M., Malka, G., Cottin, Y., and Vergely, C. (2020) Mitochondrial SLC25 carriers: novel targets for cancer therapy. *Molecules* **25**, 2417
67. Maru, D. M., Singh, R. R., Hannah, C., Albarracin, C. T., Li, Y. X., Abraham, R., *et al*. (2009) MicroRNA-196a is a potential marker of progression during Barrett's metaplasia-dysplasia-invasive adenocarcinoma sequence in esophagus. *Am. J. Pathol.* **174**, 1940–1948
68. Zaidi, A. H., Gopalakrishnan, V., Kasi, P. M., Zeng, X., Malhotra, U., Balasubramanian, J., *et al*. (2014) Evaluation of a 4-protein serum biomarker panel-biglycan, annexin-A6, myeloperoxidase, and protein S100-A9 (B-AMP)-for the detection of esophageal adenocarcinoma. *Cancer* **120**, 3902–3913
69. Koul, H. K., Pal, M., and Koul, S. (2013) Role of p38 MAP kinase signal transduction in solid tumors. *Genes Cancer* **4**, 342–359
70. Zou, X., and Blank, M. (2017) Targeting p38 MAP kinase signaling in cancer through post-translational modifications. *Cancer Lett.* **384**, 19–26
71. Salazar, C., Barros, M., Elorza, A. A., and Ruiz, L. M. (2021) Dynamic distribution of HIG2A between the mitochondria and the nucleus in response to hypoxia and oxidative stress. *Int. J. Mol. Sci.* **23**, 389
72. Schöbinger, M., Klein, O.-J., and Mitulović, G. (2016) Low-temperature mobile phase for peptide trapping at elevated separation temperature prior to nano RP-HPLC-MS/MS. *Separations* **3**, 6
73. Tóth, G., Panić-Janković, T., and Mitulović, G. (2019) Pillar array columns for peptide separations in nanoscale reversed-phase chromatography. *J Chromatogr A* **1603**, 426–432
74. Rice, T. W., Gress, D. M., Patil, D. T., Hofstetter, W. L., Kelsen, D. P., and Blackstone, E. H. (2017) Cancer of the esophagus and esophagogastric junction-major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* **67**, 304–317