Serum glycoprotein biomarker validation for esophageal adenocarcinoma and application to Barrett’s surveillance

Running title: Biomarkers for esophageal adenocarcinoma

Alok K. Shah,1,2 Gunter Hartel,1 Ian Brown,3 Clay Winterford,1 Renhua Na,1 Kim-Anh Lê Cao,2,4 Bradley A. Spicer,5 Michelle Dunstone,5 Wayne A. Phillips,6 Reginald V. Lord,7 Andrew P. Barbour,2 David I. Watson,8 Virendra Joshi,9 David C. Whiteman,1 and Michelle M. Hill1,2

1QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2The University of Queensland Diamantina Institute, Faculty of Medicine, The University of Queensland, Translational Research Institute, Brisbane, Queensland, Australia; 3Envoi Pathology, Brisbane, Queensland, Australia; 4Melbourne Integrative Genomics and School of Mathematics and Statistics, The University of Melbourne, Victoria, Australia; 5Monash University, Melbourne, Victoria, Australia; 6Peter MacCallum Cancer Centre, and Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria, Australia; 7St Vincent’s Centre for Applied Medical Research and University of Notre Dame School of Medicine, Sydney, Australia; 8Discipline of Surgery, Flinders University, Adelaide, South Australia, Australia; 9Ochsner Health System, Gastroenterology, New Orleans, LA, United States.

Correspondence: Michelle M. Hill, PhD, QIMR Berghofer Medical Research Institute, Locked Bag 2000, Royal Brisbane Hospital, 300 Herston Road, Herston QLD 4029 Australia. Tel: +617 3845 3020; e-mail: Michelle.Hill@qimrberghofer.edu.au
**Abbreviations**

| Abbreviation | Description |
|--------------|-------------|
| AAL | *Aleuria aurantia* lectin |
| % CV | % Co-efficient of variation |
| AUROC | Area under receiver operating characteristics curve |
| BE | Barrett's esophagus |
| BE-HGD | Barrett's esophagus with high-grade dysplasia |
| BE-ID | Barrett's esophagus which is indefinite for dysplasia |
| BE-LGD | Barrett's esophagus with low-grade dysplasia |
| BMI | Body mass index |
| C1QB | Complement C1q subcomponent subunit B |
| C2 | Complement C2 |
| C3 | Complement C3 |
| C4B | Complement C4-B |
| C4BPA | C4b-binding protein alpha chain |
| C4BPB | C4b-binding protein beta chain |
| C9 | Complement component C9 |
| CFB | Complement factor B |
| CFI | Complement factor I |
| CI | Confidence interval |
| CP | Ceruloplasmin |
| EAC | Esophageal adenocarcinoma |
| EPHA | Erythroagglutinin from *Phaseolus vulgaris* |
| FFPE | Formalin-fixed, paraffin-embedded |
| Abbreviation | Description |
|--------------|-------------|
| GERD         | Gastroesophageal reflux disease |
| GSN          | Gelsolin |
| JAC          | Jacalin from *Artocarpus integrifolia* |
| LeMBA        | Lectin magnetic bead array |
| MRM-MS       | Multiple reaction monitoring-mass spectrometry |
| NPL          | *Narcissus pseudonarcissus* lectin |
| NSE          | Non-specialized epithelium |
| OR           | Odds ratio |
| PGLYRP2      | N-acetylglucosaminidase |
| PON1         | Serum paraoxonase/arylesterase 1 |
| PON3         | Serum paraoxonase/lactonase 3 |
| RBP4         | Retinol-binding protein 4 |
| SERPINA4     | Kallistatin |
| SIS          | Stable isotope-labeled internal standard |
SUMMARY

BACKGROUND & AIMS: Esophageal adenocarcinoma (EAC) is thought to develop from asymptomatic Barrett’s esophagus (BE) with a low annual rate of conversion. Current endoscopy surveillance for BE patients is probably not cost-effective. Previously, we discovered serum glycoprotein biomarker candidates which could discriminate BE patients from EAC. Here, we aimed to validate candidate serum glycoprotein biomarkers in independent cohorts, and to develop a biomarker panel for BE surveillance.

METHODS: Serum glycoprotein biomarker candidates were measured in 301 serum samples collected from Australia (4 states) and USA (1 clinic) using lectin magnetic bead array (LeMBA) coupled multiple reaction monitoring mass spectrometry (MRM-MS). The area under receiver operating characteristic curve was calculated as a measure of discrimination, and multivariate recursive partitioning was used to formulate a multi-marker panel for BE surveillance.

RESULTS: Different glycoforms of complement C9 (C9), gelsolin (GSN), serum paraoxonase/arylesterase 1 (PON1) and serum paraoxonase/lactonase 3 (PON3) were validated as diagnostic glycoprotein biomarker candidates for EAC across both cohorts. A panel of 10 serum glycoproteins accurately discriminated BE patients not requiring intervention [BE+/ low grade dysplasia] from those requiring intervention [BE with high grade dysplasia (BE-HGD) or EAC]. Tissue expression of C9 was found to be induced in BE, dysplastic BE and EAC. In longitudinal samples from subjects that have progressed towards EAC, levels of serum C9 glycoforms were increased with disease progression.

CONCLUSIONS: Further prospective clinical validation of the confirmed biomarker candidates in a large cohort is warranted. A first-line BE surveillance blood test may be developed based on these findings.
KEYWORDS: esophageal adenocarcinoma; Barrett’s esophagus; serum glycoprotein biomarker; complement pathway; multiple reaction monitoring mass spectrometry; lectin
INTRODUCTION

Esophageal cancer is the sixth most common cause of cancer related mortality in men, with 3-fold higher rates in men than women (1, 2). Of the two main histological subtypes (adenocarcinoma and squamous cell carcinoma), the incidence of esophageal adenocarcinoma (EAC) has been rising continuously in Western countries, and accounts for the majority of cases (3-5). Despite aggressive treatment, EAC has a 5-year survival of less than 20% (6). EAC is thought to develop from the metaplastic condition Barrett’s esophagus (BE) as a consequence of gastroesophageal reflux disease (GERD) through a metaplasia–dysplasia–adenocarcinoma sequence (Figure 1A) (7-9).

Currently, BE patients usually undergo endoscopy-biopsy surveillance with the degree of dysplasia assessed by histopathology as a biomarker to monitor risk of neoplastic progression (10). Patients diagnosed with high grade dysplasia (BE-HGD) are treated with endoscopic mucosal resection, radiofrequency ablation or surgery, in an attempt to halt further disease progression (10-12). The significant cost of endoscopy plus the low annual progression rate to HGD or EAC means that the cost-effectiveness of endoscopic surveillance is questioned at the population level (13-16). Furthermore the evaluation of dysplasia in tissue biopsies by histopathology is prone to inter-observer variability and sampling error (17). A less costly and minimally invasive diagnostic procedure is needed for cost-effective screening and surveillance of at-risk populations (18, 19).

As the first step to developing blood-based EAC diagnostic test, we focused on differential glycosylation of serum glycoproteins during EAC pathogenesis. We established a new glycoprotein biomarker pipeline which couples lectin-based glycoprotein isolation with state-of-the-art discovery and targeted proteomics (20-23). We then applied it to identify and verify changes in lectin binding profile of serum glycoproteins between healthy, BE and EAC patients.
Here, we report results from validation in independent cohorts, and evaluation of biomarker panels for surveillance of BE patients.

**Figure 1.** (A) Pathogenesis of esophageal adenocarcinoma (EAC) and clinical management of patients during each stage. In response to exposure to gastric and bile acid, non-specialized esophageal epithelium (NSE) converts to Barrett’s esophagus (BE), and may progress through low grade dysplasia (LGD) and high grade dysplasia (HGD) stages to EAC. Patients at high risk of BE undergo endoscopic screening to detect asymptomatic metaplastic BE condition. Patients with BE or BE-LGD undergo endoscopy-biopsy surveillance to detect BE-HGD for endoscopic treatments. (B) Workflow of the study. A total of 301 serum samples from two different patient cohorts were subjected to lectin magnetic bead array-coupled multiple reaction monitoring mass spectrometry (LeMBA-MRM-MS). Biomarker candidates for EAC and surveillance were identified by statistical analysis. Tissue expression of top glycoprotein biomarker candidate complement C9 was evaluated by immunohistochemistry.
EXPERIMENTAL PROCEDURES

Clinical cohorts

Ethical approval was obtained from all participating institutions, and all patients provided informed consent to participate in the studies. We investigated two independent cohorts recruited in Australia and the United States of America, respectively. The Australian samples were selected from participants recruited into The Progression of Barrett's Esophagus to Cancer Network (PROBE-NET) study across 4 states (New South Wales, Queensland, South Australia and Victoria). A total of 252 serum samples collected from 242 patients were analyzed [Normal – 43, BE – 65, BE-LGD – 39, BE-HGD – 35, and EAC – 60, at baseline]. Of the 252 serum samples, 10 samples were from patients who progressed to the subsequent stage of the disease while 242 were baseline serum.

Independently, a cross-sectional cohort of 49 serum samples collected at Ochsner Health System, New Orleans, United States were also analyzed [Normal – 14, BE – 13, Barrett’s mucosa which is indefinite for dysplasia (BE-ID) – 3, BE-LGD – 2, BE-HGD – 7, and EAC – 10]. Some patients in the Normal group had a history of Barrett’s related pathology and received endoscopic treatments. These patients were confirmed to have no Barrett’s mucosa by histology at the time of serum collection, hence the normal groups will be called non-specialized epithelium [NSE, epithelium without Barrett’s mucosa].

The clinical diagnosis linked to the sample was based on histological examination of biopsies taken at the same endoscopy. The diagnoses were provided to the researcher performing the biomarker measurements to allow batch randomization design for the assay. Serum samples were stored at -80°C, aliquoted, and shipped to the Translational Research Institute, Brisbane on
dry ice for this study. Formalin-fixed, paraffin-embedded (FFPE) tissue sections from the Ochsner Health System were selected and shipped to Brisbane, Australia for immunohistochemistry.

Both PROBE-NET and the Ochsner cohort provided information on patients’ age, sex and body mass index (BMI, calculated as weight (kg) / [height (m)]²); whereas ethnicity was provided by the Ochsner cohort only and education, alcohol drinking and tobacco smoking was only available in the PROBE-NET cohort. Data on demographics and lifestyle factors were compared among different clinical and histological groups using Pearson’s chi-square test or Fisher’s exact test as appropriate. $P < 0.05$ was considered to be statistically significant. Analyses were performed using SAS 9.4 software.

**Serum glycoprotein biomarker measurement and analysis**

Serum glycoprotein biomarker candidates were measured using our previously reported lectin magnetic bead array (20, 21)-coupled multiple reaction monitoring (MRM) mass spectrometry assay (24). In this method, lectin binding is used to isolate glycoproteins with particular glycan structures. Based on our previous work for BE/EAC (22, 23), four lectins were selected for the independent validation cohorts, namely AAL, EPHA, JAC, and NPL (Vector Laboratories, Burlingame, CA). PROBE-NET and Ochsner cohorts were independently analyzed with block randomization design for each cohort. The PROBE-NET method measured 365 peptides belonging to 106 proteins while Ochsner method measured 381 peptides belonging to 115 proteins, inclusive of standard peptides. Detailed methods are provided in Supplemental Methods and Supplemental Table 1.

Skyline was used for inspecting and processing MRM data (25). The quality of acquired datasets (Supplemental Table 2A for PROBE-NET cohort and Supplemental Table 2B for Ochsner cohort) were evaluated by % coefficient of variation (% CV) of spiked-in stable isotope-labeled
internal standard (SIS) peptides and peptides derived from the spiked-in internal standard protein chicken ovalbumin. The data were normalized using the median intensity of 8 SIS peptides. Peptide intensities were converted into protein intensity with Pearson correlation coefficient cut-off set at 0.6 (22). As recently highlighted by others (26), this step serves as quality control for peptide level measurements resulting in a robust protein level quantitative dataset for down-stream statistical analysis. The normalized protein intensities were transformed using the natural logarithm and z-scores were calculated for downstream statistical analysis.

JMP Pro 13.2 (SAS Institute, Inc., Cary, NC, USA) was used for univariate and multivariate biomarker statistical analyses. Univariate logistic regressions were conducted on EAC vs NSE outcome, EAC vs BE outcome, and the surveillance outcome (BE-HGD or EAC vs BE or BE-ID or BE-LGD), against each of the glycoprotein_lectin biomarker candidates. Odds ratios (ORs) with 95% Wald confidence intervals (95% CIs), area under receiver operating characteristic curves (AUROCs) and Likelihood Ratio P values were calculated for both PROBE-NET and Ochsner cohorts. Recursive partitioning (also known as Classification and Regression Trees, CART) was used to identify a multivariate panel of markers that would discriminate between surveillance outcomes (BE-HGD+EAC vs BE+BE-ID+BE-LGD). The PROBE-NET dataset was used as the training set to develop predictive models, and the Ochsner dataset was used as the validation set. The set of 217 markers that were available in both PROBE-NET and Ochsner datasets was used in the training set, as well as baseline characteristics, including age, gender, and BMI. To avoid overfitting, models were limited to 6, 8, and 10 biomarkers. The prediction formulas derived from PROBE-NET were then applied to the Ochsner dataset to determine sensitivity and specificity in the validation set.
| Characteristics | All $^1$ | Histology diagnosis | $^1$ |
|-----------------|----------|---------------------|-----|
|                 | N (% )  | NSE N (%) | BE N (%) | BE-LGD N (%) | BE-HGD N (%) | EAC N (%) |
| Ochsner Cohort (N) | 49 (22) | 14 (29) | 13 (23) | 5 (40) | 7 (43) | 10 (14) |
| Age at baseline | <0.001 | 11 (4) | 3 (23) | 2 (46) | 0 (3) | 2 (20) |
| 60-69 | 22 (45) | 5 (26) | 4 (31) | 3 (66) | 3 (43) | 7 (70) |
| ≥70 | 16 (33) | 5 (36) | 6 (46) | 0 (3) | 3 (43) | 2 (20) |
| Sex | 0.83 | Female | 11 (22) | 3 (21) | 2 (15) | 2 (40) | 2 (29) | 2 (20) |
| Male | 38 (78) | 11 (79) | 11 (85) | 3 (60) | 5 (71) | 8 (80) |
| Ethnicity | 0.57 | Caucasian | 45 (92) | 12 (86) | 12 (92) | 5 (100) | 6 (86) | 10 (100) |
| Hispanic/Latino | 1 (2) | 0 (0) | 1 (8) | 0 (0) | 0 (0) | 0 (0) |
| African-American | 3 (6) | 2 (14) | 0 (0) | 0 (0) | 1 (14) | 0 (0) |
| BMI at baseline | 0.19 | <25 | 12 (24) | 2 (14) | 7 (54) | 0 (0) | 1 (14) | 2 (20) |
| 25-29.9 | 15 (31) | 3 (21) | 2 (15) | 2 (40) | 3 (43) | 5 (50) |
| ≥30 | 22 (45) | 9 (64) | 4 (31) | 3 (60) | 3 (43) | 3 (30) |
| PROBENET (N) | 242 | 43 | 65 | 39 | 35 | 60 |
| Age at baseline | 0.04 | <60 | 95 (39) | 24 (57) | 29 (45) | 11 (28) | 11 (31) | 20 (33) |
| 60-69 | 74 (31) | 14 (33) | 16 (25) | 17 (43) | 11 (31) | 16 (27) |
| ≥70 | 73 (30) | 5 (10) | 20 (31) | 11 (30) | 13 (37) | 24 (40) |
| Sex | <0.001 | Female | 58 (24) | 28 (64) | 10 (15) | 7 (18) | 4 (11) | 9 (15) |
| Male | 184 (76) | 15 (36) | 55 (85) | 32 (82) | 31 (89) | 51 (85) |
| Education | 0.57 | High School | 87 (40) | 13 (32) | 25 (40) | 17 (49) | 11 (37) | 21 (40) |
| College equivalent | 91 (41) | 16 (39) | 26 (42) | 12 (34) | 12 (40) | 25 (48) |
| University | 42 (19) | 12 (29) | 11 (18) | 6 (17) | 7 (23) | 6 (12) |
| BMI at baseline | 0.41 | <25 | 45 (22) | 13 (32) | 14 (24) | 3 (10) | 6 (20) | 9 (19) |
| 25-29.9 | 94 (45) | 13 (32) | 26 (44) | 16 (51) | 13 (43) | 26 (54) |
| ≥30 | 69 (33) | 14 (35) | 19 (32) | 12 (39) | 11 (37) | 13 (27) |
| Alcohol drinking | 0.45 | Never | 32 (14) | 9 (22) | 8 (13) | 5 (14) | 2 (6) | 8 (15) |
| Ever | 190 (86) | 32 (78) | 54 (87) | 30 (86) | 30 (94) | 44 (85) |
| Tobacco smoking | 0.18 | Never | 73 (33) | 15 (37) | 26 (42) | 12 (34) | 6 (19) | 14 (27) |
| Ever | 149 (67) | 26 (63) | 36 (58) | 23 (66) | 26 (81) | 38 (73) |
1 counts in subcategories may not add up to the total number because of missing data.

2 $\chi^2$ test or Fisher exact test as appropriate.

**Immunohistochemical analysis**

Immunohistochemical staining was performed on 42 tissue sections collected from 34 patients using anti-C9 primary antibody (Sigma Aldrich #SAB4503059). Detailed immunohistochemistry protocol is described in Supplemental Methods. To confirm the staining specificity, the anti-C9 antibody was neutralized with C9 recombinant protein before application to the slide, as reported in Supplemental Methods and Supplemental Figure 1. Staining intensities in squamous mucosa, columnar epithelium without intestinal metaplasia, Barrett’s esophagus mucosa (with intestinal metaplasia), dysplasia (low and high grade), EAC and inflammatory infiltrate were scored by a specialist gastrointestinal pathologist. Each component, if present in the tissue, was scored separately using a 4 grade assessment of intensity (0 no staining, 1+ weak staining, 2+ moderate staining, 3+ strong staining). Where the staining was non-uniform in a component, the maximum intensity of staining was used for the score, providing at least 10% of the cells of that component stained to this intensity. Due to limited numbers, for statistical analysis, we combined dysplasia with EAC group; similarly, those with 3+ staining in Barrett’s mucosa were combined with 2+ staining group. The relationship between histological features and the staining intensity was evaluated using Fisher exact test.

**RESULTS**

Workflow of this study is depicted in Figure 1B. Table 1 details the baseline demographic and clinical characteristics of the Australian (PROBE-NET) and Ochsner cohorts.

Univariate analysis of the PROBE-NET cohort dataset revealed 46 and 54 biomarker candidates with $P<0.05$ for EAC vs NSE and EAC vs BE comparisons respectively (Supplemental
Table 3. Data for the top 10 biomarkers that differentiate EAC from NSE, and EAC from BE in PROBE-NET cohort are shown in Figure 2A and 2B, respectively. Out of these candidates, 16 candidates for EAC vs NSE comparison and 9 candidates for EAC vs BE comparison were also significantly different in the Ochsner cohort, confirming these glycoproteins as validated biomarkers. As illustrated in Figure 2C, 8 validated biomarkers overlap between the two lists. These biomarkers are potentially most useful, being able to distinguish EAC from NSE and BE. Interestingly, the 8 biomarkers comprised of different glycosylated forms of 4 proteins, namely, complement C9 (C9), gelsolin (GSN), serum paraoxonase/arylesterase 1 (PON1), serum paraoxonase/lactonase 3 (PON3).

Next we considered BMI as a potential confounding factor for EAC biomarker validation. Correlation analysis between BMI and biomarker levels in all PROBE-NET samples revealed no substantial correlation ($|r|<0.6$) (Supplemental Table 4), although three proteins namely complement C3 (C3), C4b-binding protein alpha chain (C4BPA), and complement factor I (CFI) with multiple lectin pull-downs showed positive correlations with BMI in NSE patients only ($|r|\ 0.6234-0.7069$). Importantly, none of the top 10 biomarkers showed strong correlation with BMI.
Biomarkers for BE surveillance

Having confirmed univariate biomarkers for detection of EAC from NSE and BE in independent cohorts, we next evaluated the ability of serum glycoproteins to be used as a surveillance tool for BE and BE-LGD patients, i.e. to distinguish between patients who require treatment (BE-HGD and EAC) and those who do not (BE, BE-ID and BE-LGD). Eight biomarkers that showed AUROC > 0.6 in a BE surveillance setting for both PROBE-NET and Ochsner cohort
(Figure 3, Supplemental Table 3), comprising of 3 glycoforms of GSN, 2 glycoforms of C9, AAL-binding PON1, AAL-binding PON3, as well as EPHA-binding, Complement factor B (CFB).

Next we sought to generate a multimarker panel for BE surveillance, using the PROBE-NET cohort for modeling and Ochsner cohort for model validation. The minimal panel of six biomarkers showed 0.83 AUROC, 83% specificity and 67% sensitivity for the PROBE-NET cohort, and a moderate specificity of 61% and sensitivity of 63% for the Ochsner cohort (Table 2). Addition of 4 more biomarkers to the panel increased the AUROC to 0.93, and improved the specificity and sensitivity measures for PROBE-NET, as well as the sensitivity for Ochsner cohort (Table 2).

Figure 3. Serum glycoprotein biomarkers for BE surveillance in PROBE-NET and Ochsner cohorts. Glycoproteins that distinguish patients who require treatment (BE-HGD+EAC) from those who do not require treatment (BE+BE-ID+BE-LGD) with an area under receiver operating characteristic curve (AUROC) value of > 0.6 in both PROBE-NET and Ochsner cohorts are shown. Odds ratio (OR) with 95% Wald confidence intervals, likelihood ratio \(P\) values and AUROC values for (A) PROBE-NET cohort and (B) Ochsner cohort. AAL = Aleuria aurantia lectin; AUROC = area under receiver operating characteristic curve; BE = Barrett’s esophagus; BE-HGD = BE with high-grade dysplasia; BE-LGD = BE with low-grade dysplasia; C9 = complement C9; CFB = complement factor B; CI = confidence interval; EAC = esophageal adenocarcinoma; EPHA = erythroagglutinin Phaseolus vulgaris; GSN = gelsolin; JAC = jacalin from Artocarpus integrifolia; NPL = Narcissus pseudonarcissus lectin; PON1 = serum paraoxonase/arylesterase 1; PON3 = serum paraoxonase/lactonase 3.
Table 2. Glycoprotein biomarker panel performance for BE surveillance (BE-HGD+EAC vs BE+BE-ID+BE-LGD) in PROBE-NET and Ochsner cohorts.

| Multimarker panel | PROBE-NET cohort | Ochsner cohort |
|-------------------|------------------|----------------|
|                   | AUROC | Sensitivity | Specificity | Sensitivity | Specificity |
| Panel 1:          |       |             |             |             |
| C9_EPHA + PON1_AAL + SERPINA4_EPHA + RBP4_AAL + C1QB_JAC + PGLYRP2_NPL | 0.83  | 83%         | 69%         | 61%         | 65%         |
| Panel 2:          |       |             |             |             |
| Panel 1 + CP_NPL + C4BPB_JAC + C2_EPHA + C4B_AAL | 0.93  | 86%         | 85%         | 61%         | 71%         |

AAL = Aleuria aurantia lectin; AUROC = area under receiver operating characteristic curve; BE = Barrett’s esophagus; BE-HGD = BE with high-grade dysplasia; BE-LGD = BE with low-grade dysplasia; C1QB = Complement C1q subcomponent subunit B; C2 = Complement C2; C4B = Complement C4-B; C4BPB = C4b-binding protein beta chain; C9 = complement component C9; CP = Ceruloplasmin; EAC = esophageal adenocarcinoma; EPHA = erythroagglutinin Phaseolus vulgaris; GSN = gelsolin; JAC = jacalin from Artocarpus integrifolia; NPL = Narcissus pseudonarcissus lectin; PGLYRP2 = N-acetyl muramoyl-L-alanine amidase; PON1 = serum paraoxonase/arylesterase 1; PON3 = serum paraoxonase/lactonase 3; RBP4 = Retinol-binding protein 4; SERPINA4 = Kallistatin.

Complement C9 expression in BE and EAC tissue

In agreement with our previous finding of complement pathway dysregulation in EAC pathogenesis (22), 5 of the 10 glycoprotein biomarker candidates in the final surveillance biomarker panel (Table 2) belonged to the complement pathway. As a first step to evaluate alterations of the complement pathway in EAC at a tissue level, we optimized immunohistochemistry staining for the top candidate complement C9. Staining specificity of the method was confirmed by neutralization of the antibody with recombinant C9 protein prior to staining (Supplemental Figure 1). We then evaluated expression of C9 in esophageal tissue sections from a subset of the Ochsner cohort. As shown in Figure 4A, C9 was detected in BE and EAC, but not squamous epithelium. Dysplastic BE showed particularly strong staining in the plasma membrane and/or cytoplasm (Figure 4A). Strong staining in immune infiltrates served as an expected positive control. Quantitation of staining intensity score against the tissue
phenotype (Figure 4B) showed statistically significant association between C9 expression levels and histology groups of squamous epithelium, columnar epithelium, Barrett’s mucosa, dysplasia/EAC ($P<0.001$).

Figure 4. Immunohistochemical analysis of complement C9 (C9) in lower esophageal biopsies. (A) A panel of representative images of C9 staining in squamous epithelium, Barrett’s mucosa, dysplastic epithelium, and esophageal adenocarcinoma conditions. The left image shows 2+ staining in the non-dysplastic Barrett’s epithelium. The middle image shows 1+ and 3+ staining in the BE. The right image shows 2+ staining in cancer cells and 3+ staining in inflammatory cells. (B) Summary of C9 staining intensity according to histology groups. $P$ value from Fisher exact test. EAC = esophageal adenocarcinoma; IM = intestinal metaplasia.

Serum complement C9 glycoforms in progressor samples

As an additional evaluation, we examined C9 lectin pull-down levels in samples from PROBE-NET participants who had progressed from BE to BE-LGD ($N = 4$), BE-LGD to BE-
HGD (N = 3), BE to BE-HGD (N = 1), BE to EAC (N = 1), and NSE to gastric type mucosa (N = 1) phenotype during follow-up. The progression time varied from 110 days up to 5 years. Significant elevation of C9_EPHA and C9_NPL was observed following progression in this small patient cohort (Figure 5).

**DISCUSSION**

This study advances our previous serum glycoprotein research (22, 23) and provides a critical breakthrough towards developing cost-effective EAC surveillance. Over the years, several studies have been carried out to identify circulatory biomarker candidates to diagnose BE-dysplasia-EAC disease spectrum (19, 27). These studies have explored genetic alterations in cell free circulating DNA (28), serum miRNA changes (29), circulatory tumor cells (30), glycan profile alteration in serum (31, 32), circulatory autoantibodies (against cancer antigens) (33), volatile organic compounds found in breath analysis (34), metabolic changes in urine (35), and a panel of
serum proteins (36) as promising diagnostic biomarker candidates for BE and/or EAC. However, none of these biomarker candidates have progressed to from bench to bedside, likely due to the lack of subsequent validation studies in large independent cohort of patients. Here, we have addressed this gap by validating eight serum glycoprotein biomarkers for EAC in two independent patient cohorts including dysplastic samples.

In addition to demonstrating the robustness of our mass spectrometry based glycoprotein-centric proteomics workflow for biomarker validation, the current study confirmed our previous finding of complement activation in EAC (22). The complement system consists of a cascade of circulating proteases that are locally activated leading to the formation of the membrane attack complex on the immunogen and recruitment of phagocytes. Complement components are predominantly expressed and secreted into the plasma by the liver but are also found to be expressed in other tissues (37), but some complement proteins may be expressed by tumor cells (38). While complement components are primarily involved in mediating innate immune response, recent studies have revealed an apparently paradoxical tumor-promoting role of the complement system (38, 39). Complement C3 has been reported to play an autocrine role in ovarian and lung cancer tumor growth (40). C5a is elevated in serum of lung cancer patients (41) and increases the invasiveness of C5aR+ tumors (42). Fucosylated C9 was previously reported to be elevated in the serum of lung cancer patients (43).

Two previous publications reported on complement component changes in BE and EAC. Bobryshev et al. reported reduced C1q expression in dendritic cells and macrophages in the epithelium of BE and EAC, and suggest this to be an immune-escape mechanism (44). Song et al. identified complement pathway proteins, complement C3 and complement C1r subcomponent to be increased in serum collected from BE-HGD and EAC patients respectively as compared to
disease free individuals (45). The current study is the first to report C9 protein expression in BE and EAC cells, as demonstrated on tissue sections. In addition to validating our previous finding of elevated circulating C9 glycoforms in EAC, we further confirmed the utility of serum C9_JAC and C9_EPHA for BE surveillance in independent cohorts (Figure 3). The latter result suggests a pathological change in the secretion, glycosylation, or expression of C9 during the progression of BE to BE-HGD/EAC. Modulation of C9 expression is unlikely to be the mechanism, since we detected strong C9 staining in both BE and BE-HGD tissue. Interestingly, bile (deoxycholic acid) treatment was reported to altered glycosylation and Golgi structure in esophageal epithelial and BE cells, resulting in impaired protein secretion via the classical pathway (46). Hence, C9 may be sequentially regulated by expression level and glycosylation/secretion during BE-EAC progression. Further studies are required to determine the precise molecular mechanisms.

The strengths of this study include the validation of biomarkers in independent cohorts, evaluation of biomarker panel for BE surveillance, and the use of immunohistochemistry to determine the cellular origin of the top serum glycoprotein biomarker. There are several limitations to our study. While our glycoprotein biomarker pipeline allows high-throughput glycoprotein biomarker discovery, the glycosylation site and glycan structural changes are not determined at early stages. Furthermore, the functional consequences of these glycosylation changes in cancer progression also remain to be evaluated. The small number of progressor samples available is a limitation, not only for the current study, but for evaluation of BE/EAC biomarkers in general. International collaborations and longitudinal cohort sampling are therefore required to advance the search for EAC biomarkers. Finally, the tissue expression of C9 and other candidates also need to be confirmed in a larger number of samples.
In summary, we have validated a number of serum glycoprotein biomarkers that warrant further clinical testing in large independent cohorts including longitudinal patient samples. Further development of these biomarkers to a blood test can aid the current endoscopy-biopsy surveillance program of BE patients for early detection of dysplastic progression.

REFERENCES

1. Siegel, R. L., Miller, K. D., and Jemal, A. (2017) Cancer Statistics, 2017. CA Cancer J Clin 67, 7-30
2. Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136, E359-386
3. Chen, Z., Ren, Y., Du, X. L., Yang, J., Shen, Y., Li, S., Wu, Y., Lv, M., Dong, D., Li, E., Li, W., Liu, P., Yang, J., and Yi, M. (2017) Incidence and survival differences in esophageal cancer among ethnic groups in the United States. Oncotarget 8, 47037-47051
4. Thrift, A. P. (2016) The epidemic of oesophageal carcinoma: Where are we now? Cancer Epidemiol 41, 88-95
5. Hur, C., Miller, M., Kong, C. Y., Dowling, E. C., Nattinger, K. J., Dunn, M., and Feuer, E. J. (2013) Trends in esophageal adenocarcinoma incidence and mortality. Cancer 119, 1149-1158
6. Malhotra, G. K., Yanala, U., Ravipati, A., Follet, M., Vijayakumar, M., and Are, C. (2017) Global trends in esophageal cancer. J Surg Oncol 115, 564-579
7. Reid, B. J., Li, X., Galipeau, P. C., and Vaughan, T. L. (2010) Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat Rev Cancer 10, 87-101
8. Anaparthy, R., and Sharma, P. (2014) Progression of Barrett oesophagus: role of endoscopic and histological predictors. Nat Rev Gastroenterol Hepatol 11, 525-534
9. Rubenstein, J. H., and Shaheen, N. J. (2015) Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma. *Gastroenterology* 149, 302-317 e301

10. Fitzgerald, R. C., di Pietro, M., Ragunath, K., Ang, Y., Kang, J. Y., Watson, P., Trudgill, N., Patel, P., Kaye, P. V., Sanders, S., O'Donovan, M., Bird-Lieberman, E., Bhandari, P., Jankowski, J. A., Attwood, S., Parsons, S. L., Loft, D., Lagergren, J., Moayyedi, P., Lyrratzopoulos, G., de Caestecker, J., and British Society of, G. (2014) British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 63, 7-42

11. Shaheen, N. J., Falk, G. W., Iyer, P. G., Gerson, L. B., and American College of, G. (2016) ACG Clinical Guideline: Diagnosis and Management of Barrett's Esophagus. *Am J Gastroenterol* 111, 30-50; quiz 51

12. Whiteman, D. C., Appleyard, M., Bahin, F. F., Bobryshev, Y. V., Bourke, M. J., Brown, I., Chung, A., Clouston, A., Dickins, E., Emery, J., Eslick, G. D., Gordon, L. G., Grimpen, F., Hebbard, G., Holliday, L., Hourigan, L. F., Kendall, B. J., Lee, E. Y., Levert-Mignon, A., Lord, R. V., Lord, S. J., Maule, D., Moss, A., Norton, I., Olver, I., Pavey, D., Raftopoulos, S., Rajendra, S., Schoeman, M., Singh, R., Sitas, F., Smithers, B. M., Taylor, A. C., Thomas, M. L., Thomson, I., To, H., von Dincklage, J., Vuletich, C., Watson, D. I., and Yusoff, I. F. (2015) Australian clinical practice guidelines for the diagnosis and management of Barrett's esophagus and early esophageal adenocarcinoma. *J Gastroenterol Hepatol* 30, 804-820

13. Desai, T. K., Krishnan, K., Samala, N., Singh, J., Cluley, J., Perla, S., and Howden, C. W. (2012) The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut* 61, 970-976

14. Barbiere, J. M., and Lyrratzopoulos, G. (2009) Cost-effectiveness of endoscopic screening followed by surveillance for Barrett's esophagus: a review. *Gastroenterology* 137, 1869-1876
15. Gerson, L. B., Groeneveld, P. W., and Triadafilopoulos, G. (2004) Cost-effectiveness model of endoscopic screening and surveillance in patients with gastroesophageal reflux disease. *Clin Gastroenterol Hepatol* 2, 868-879

16. Yang, Y., Chen, H. N., Wang, R., Tang, Y. J., and Chen, X. Z. (2015) Cost-Effectiveness Analysis on Endoscopic Surveillance Among Western Patients With Barrett's Esophagus for Esophageal Adenocarcinoma Screening. *Medicine (Baltimore)* 94, e1105

17. Goldblum, J. R. (2003) Barrett's esophagus and Barrett's-related dysplasia. *Mod Pathol* 16, 316-324

18. Vaughan, T. L., and Fitzgerald, R. C. (2015) Precision prevention of oesophageal adenocarcinoma. *Nat Rev Gastroenterol Hepatol* 12, 243-248

19. Shah, A. K., Saunders, N. A., Barbour, A. P., and Hill, M. M. (2013) Early diagnostic biomarkers for esophageal adenocarcinoma--the current state of play. *Cancer Epidemiol Biomarkers Prev* 22, 1185-1209

20. Choi, E., Loo, D., Dennis, J. W., O'Leary, C. A., and Hill, M. M. (2011) High-throughput lectin magnetic bead array-coupled tandem mass spectrometry for glycoprotein biomarker discovery. *Electrophoresis* 32, 3564-3575

21. Loo, D., Jones, A., and Hill, M. M. (2010) Lectin magnetic bead array for biomarker discovery. *J Proteome Res* 9, 5496-5500

22. Shah, A. K., Le Cao, K. A., Choi, E., Chen, D., Gautier, B., Nancarrow, D., Whiteman, D. C., Saunders, N. A., Barbour, A. P., Joshi, V., and Hill, M. M. (2015) Serum Glycoprotein Biomarker Discovery and Qualification Pipeline Reveals Novel Diagnostic Biomarker Candidates for Esophageal Adenocarcinoma. *Mol Cell Proteomics* 14, 3023-3039
23. Shah, A. K., Le Cao, K. A., Choi, E., Chen, D., Gautier, B., Nancarrow, D., Whiteman, D. C., Baker, P. R., Clauser, K. R., Chalkley, R. J., Saunders, N. A., Barbour, A. P., Joshi, V., and Hill, M. M. (2016) Glyco-centric lectin magnetic bead array (LeMBA) - proteomics dataset of human serum samples from healthy, Barretts esophagus and esophageal adenocarcinoma individuals. Data Brief 7, 1058-1062

24. (2013) Method of the Year 2012. Nature methods 10, 1

25. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966-968

26. Zhang, B., Pirmoradian, M., Zubarev, R., and Kall, L. (2017) Covariation of Peptide Abundances Accurately Reflects Protein Concentration Differences. Mol Cell Proteomics 16, 936-948

27. Yazbeck, R., Jaenisch, S. E., and Watson, D. I. (2016) From blood to breath: New horizons for esophageal cancer biomarkers. World J Gastroenterol 22, 10077-10083

28. Openshaw, M. R., Richards, C. J., Guttery, D. S., Shaw, J. A., and Thomas, A. L. (2017) The genetics of gastroesophageal adenocarcinoma and the use of circulating cell free DNA for disease detection and monitoring. Expert Rev Mol Diagn 17, 459-470

29. Chiam, K., Wang, T., Watson, D. I., Mayne, G. C., Irvine, T. S., Bright, T., Smith, L., White, I. A., Bowen, J. M., Keefe, D., Thompson, S. K., Jones, M. E., and Hussey, D. J. (2015) Circulating Serum Exosomal miRNAs As Potential Biomarkers for Esophageal Adenocarcinoma. J Gastrointest Surg 19, 1208-1215

30. Gallerani, G., and Fabbri, F. (2016) Circulating Tumor Cells in the Adenocarcinoma of the Esophagus. Int J Mol Sci 17
31. Gaye, M. M., Valentine, S. J., Hu, Y., Mirjankar, N., Hammoud, Z. T., Mechref, Y., Lavine, B. K., and Clemmer, D. E. (2012) Ion mobility-mass spectrometry analysis of serum N-linked glycans from esophageal adenocarcinoma phenotypes. *J Proteome Res* 11, 6102-6110

32. Mitra, I., Zhuang, Z., Zhang, Y., Yu, C. Y., Hammoud, Z. T., Tang, H., Mechref, Y., and Jacobson, S. C. (2012) N-glycan profiling by microchip electrophoresis to differentiate disease states related to esophageal adenocarcinoma. *Anal Chem* 84, 3621-3627

33. Kilic, A., Schuchert, M. J., Luketich, J. D., Landreneau, R. J., Lokshin, A. E., Bigbee, W. L., and El-Hefnawy, T. (2008) Use of novel autoantibody and cancer-related protein arrays for the detection of esophageal adenocarcinoma in serum. *J Thorac Cardiovasc Surg* 136, 199-204

34. Kumar, S., Huang, J., Abbassi-Ghadi, N., Mackenzie, H. A., Veselkov, K. A., Hoare, J. M., Lovat, L. B., Spanel, P., Smith, D., and Hanna, G. B. (2015) Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. *Ann Surg* 262, 981-990

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

36. Zaidi, A. H., Gopalakrishnan, V., Kasi, P. M., Zeng, X., Malhotra, U., Balasubramanian, J., Visweswaran, S., Sun, M., Flint, M. S., Davison, J. M., Hood, B. L., Conrads, T. P., Bergman, J. J., Bigbee, W. L., and Jobe, B. A. (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

37. Lubbers, R., van Essen, M. F., van Kooten, C., and Trouw, L. A. (2017) Production of complement components by cells of the immune system. *Clin Exp Immunol* 188, 183-194
38. Pio, R., Corrales, L., and Lambris, J. D. (2014) The role of complement in tumor growth. *Advances in experimental medicine and biology* 772, 229-262

39. Mamidi, S., Hone, S., Teufel, C., Sellner, L., Zenz, T., and Kirschfink, M. (2015) Neutralization of membrane complement regulators improves complement-dependent effector functions of therapeutic anticancer antibodies targeting leukemic cells. *Oncoimmunology* 4, e979688

40. Cho, M. S., Vasquez, H. G., Rupaimoole, R., Pradeep, S., Wu, S., Zand, B., Han, H. D., Rodriguez-Aguayo, C., Bottsford-Miller, J., Huang, J., Miyake, T., Choi, H. J., Dalton, H. J., Ivan, C., Baggerly, K., Lopez-Berestein, G., Sood, A. K., and Afshar-Kharghan, V. (2014) Autocrine effects of tumor-derived complement. *Cell Rep* 6, 1085-1095

41. Corrales, L., Ajona, D., Rafail, S., Lasarte, J. J., Riezu-Boj, J. I., Lambris, J. D., Rouzaut, A., Pajares, M. J., Montuenga, L. M., and Pio, R. (2012) Anaphylatoxin C5a creates a favorable microenvironment for lung cancer progression. *J Immunol* 189, 4674-4683

42. Nitta, H., Wada, Y., Kawano, Y., Murakami, Y., Irie, A., Taniguchi, K., Kikuchi, K., Yamada, G., Suzuki, K., Honda, J., Wilson-Morifuji, M., Araki, N., Eto, M., Baba, H., and Imamura, T. (2013) Enhancement of human cancer cell motility and invasiveness by anaphylatoxin C5a via aberrantly expressed C5a receptor (CD88). *Clin Cancer Res* 19, 2004-2013

43. Narayanasamy, A., Ahn, J. M., Sung, H. J., Kong, D. H., Ha, K. S., Lee, S. Y., and Cho, J. Y. (2011) Fucosylated glycoproteomic approach to identify a complement component 9 associated with squamous cell lung cancer (SQLC). *Journal of proteomics* 74, 2948-2958

44. Bobryshev, Y. V., Lu, J., and Lord, R. V. (2010) Expression of C1q complement component in Barrett’s esophagus and esophageal adenocarcinoma. *J Gastrointest Surg* 14, 1207-1213
45. Song, E., Zhu, R., Hammoud, Z. T., and Mechref, Y. (2014) LC-MS/MS quantitation of esophagus disease blood serum glycoproteins by enrichment with hydrazide chemistry and lectin affinity chromatography. *J Proteome Res* 13, 4808-4820

46. Byrne, A. M., Sharma, R., Duggan, G., Kelleher, D., and Long, A. (2012) Deoxycholic acid impairs glycosylation and fucosylation processes in esophageal epithelial cells. *Glycobiology* 22, 638-648

**Author contributions**

MMH conceived and supervised the project. AKS designed and performed the proteomics experiments including data analysis. AKS, GH, RN, and KALC undertook statistical analysis. WAP, RVL, APB, DIW, DCW, and VJ contributed to sample collection. RN, DCW, and VJ selected the patient samples for the study. CW developed, optimized and performed the immunohistochemistry. IB evaluated and scored immunohistochemistry slides. BAS and MD contributed to recombinant C9 protein expression and purification. AKS, GH and MMH wrote the manuscript. DCW helped with critical revision of the manuscript. All authors edited and approved the manuscript.

We thank Dr Thomas Hennessey (Agilent Technologies) and Mr Elliot McElroy (Agilent Technologies) for technical assistance, and Mr Cris Molina (Ochsner) for clinical research management support.

**Grant support**

This work was supported by The University of Queensland-Ochsner Seed Fund for Collaborative Research Grant 2014, The University of Queensland Faculty of Medicine and Biomedical Science Cancer Bequest Grant 2014, and internal funding support from The University
of Queensland Diamantina Institute 2015. PROBE-NET was supported by a National Health and Medical Research Council (NHMRC) Centre of Research Excellence grant (APP1040947). KALC is a recipient of NHMRC Career Development Fellowship (APP1087415). DCW is supported by an NHMRC Research Fellowship (APP1058522). The sponsors had no influence on the study design, collection, analysis, and interpretation of data.

Disclosures

The authors have no conflicts of interest to disclose.