Whole genome methylation analysis of nondysplastic Barrett esophagus that progresses to invasive cancer

Dilworth, Mark P; Nieto, Tom; Stockton, Jo D; Whalley, Celina M; Tee, Louise; James, Jonathan D; Noble, Fergus; Underwood, Tim J; Hallissey, Michael T; Hejmadi, Rahul; Trudgill, Nigel; Tucker, Olga; Beggs, Andrew D

DOI: 10.1097/SLA.0000000000002658

License: Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.
Whole Genome Methylation Analysis of Nondysplastic Barrett Esophagus that Progresses to Invasive Cancer

Mark P. Dilworth, FRCS(Gen),* Tom Nieto, MRCS,† Jo D. Stockton, PhD,* Celina M. Whalley, MSc,* Louise Tee, MSc,* Jonathan D. James, PhD,* Fergus Noble, MRCS,§ Tim J. Underwood, FRCS(Gen),¶ Michael T. Hallissey, FRCS,† Rahul Hejmadi, FRCPath,* Nigel Trudgill, FRCPath,‡ Olga Tucker, FRCS,§ and Andrew D. Beggs, FRCS(Gen) PhD*†

Objective: To investigate differences in methylation between patients with nondysplastic Barrett esophagus who progress to invasive adenocarcinoma and those who do not.

Background: Identifying patients with nondysplastic Barrett esophagus who progress to invasive adenocarcinoma remains a challenge. Previous studies have demonstrated the potential utility of epigenetic markers for identifying this group.

Methods: A whole genome methylation interrogation using the Illumina HumanMethylation 450 array of patients with nondysplastic Barrett esophagus who either develop adenocarcinoma or remain static, with validation of findings by bisulfite pyrosequencing.

Results: In all, 12 patients with “progressive” versus 12 with “nonprogressive” nondysplastic Barrett esophagus were analyzed via methylation array. Forty-four methylation markers were identified that may be able to discriminate between nondysplastic Barrett esophagus that either progress to adenocarcinoma or remain static. Hypomethylation of the recently identified tumor suppressor OR3A4 (probe cg08890332) validated in a separate cohort of samples (median methylation in progressors 67.8% vs 96.7% in nonprogressors; \( P = 0.0001, c = 3.85 \), Wilcoxon rank-sum test) and was associated with the progression to adenocarcinoma. There were no differences in copy number between the 2 groups, but a global trend towards hypomethylation in the progressor group was observed.

Conclusion: Hypomethylation of OR3A4 has the ability to risk stratify the patient with nondysplastic Barrett esophagus and may form the basis of a future surveillance program.

Keywords: Barrett esophagus, cancer genetics, esophageal cancer, methylation

From the *Institute of Cancer and Genomic Science, University of Birmingham, UK; †Queen Elizabeth Hospital, Birmingham, UK; ‡Sandwell and West Birmingham NHS Trust, Birmingham, UK; §Heart of England NHS Trust, Birmingham, UK; ¶Cancer Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, UK.

MPO, TN, OT, and ADB contributed equally to this study.

Funding: ADB acknowledges funding from the Wellcome Trust (102732/Z/13/Z), Cancer Research UK (C31641/A23923) and the Medical Research Council (MR/M016587/1). ADB also acknowledges support from the University of Southampton CRUK/ECMC Tissue Bank.

The authors declare no conflict of interests.

Supplemental digital content is available for this article. Direct URL citations appear in the HTML and PDF versions of this article on the journal’s Web site (www.annalsofsurgery.com).

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Reprints: Dr Andrew D. Beggs, FRCS(Gen) PhD, Institute of Cancer & Genomic Science, University of Birmingham, Vincent Drive, Birmingham B15 2TT, UK. E-mail: a.beggs@bham.ac.uk.

Copyright © 2018 The Author(s). Published by Wolters Kluwer Health, Inc. ISSN: 0003-4921/18/26903-0479 DOI: 10.1097/SLA.0000000000002658

Eosophageal adenocarcinoma (EADC) incidence is increasing and currently represents 5% of the digestive tract cancers in the UK. Overall disease survival is poor, but correlates with stage of cancer at presentation, demonstrating significant survival advantages with detection of early-stage disease. Barrett esophagus (BE), in which normal squamous mucosa is replaced with a metaplastic columnar phenotype, results from prolonged exposure to stomach acids and bile salts, which reflux into the esophagus causing chronic inflammation and tissue damage. The incidence of BE is increasing, largely thought to be a consequence of obesity-induced reflux disease. BE is associated with an increased risk of EADC, but for the majority of patients, BE will never progress beyond simple benign metaplasia. However, in a small number of patients, dysplasia will develop, with some progressing to EADC. The incidence of EADC in the BE population is up to 150 times greater than unaffected individuals.

Although the pathological changes seen in Barrett adenocarcinoma are understood as part of a well-established metaplasia-dysplasia-carcinoma sequence, the molecular drivers are less clear. The current dilemma is that for patients with nondysplastic BE, there are no accurate methods for identifying the small number of patients at high risk of progression to cancer.

Surveillance practice of the Barrett patients varies widely, between some who endoscope patients each year in contrast to others who will never repeat the investigation. The ongoing UK Medical Research Council funded BOSS study aims to understanding the optimum surveillance strategy, randomizing between prospective monitoring BE patients with frequent endoscopic assessment or a “watch and wait” policy.

Clearly, there is need for a method of risk stratification in these patients to facilitate a streamlined surveillance program by identifying high-risk nondysplastic BE patients. Attempts at biomarker development for stratification of high-risk BE have focused on mutational change, specifically around the role of TP53 mutation in predicting “high-risk” disease, given its role as a driver in esophageal cancer. However, Ross-Innes et al17 have convincingly demonstrated the presence of pathogenic TP53 mutations in apparently normal squamous esophageal mucosa, thus making its role in progression to invasive adenocarcinoma unclear. However, the role of epigenetic change in the pathogenesis of BE and esophageal cancer is less well-understood, but may well happen much earlier in the cancer development pathway, and, as a direct result, provide a more appropriate target for both predicting its development and potentially arresting tumorigenesis, should a suitable epigenetic modulator be identified.

Multiple methylation markers have been identified which can discriminate between high-risk and low-risk BE including APC, MGMT, PKP-1, TIMP3/TERT, RUNX3/HPP1, and AKA12. Agarwal et al25 performed a MeCIP array-based approach to compare the methylation of progressor (n = 5) versus nonprogressing patients (n = 4). In patients who progressed to invasive adenocarcinoma, their original biopsies began either with...
no dysplasia, indefinite dysplasia, or low-grade dysplasia, making comparison difficult. However, subsequent analysis of the top 25 differential methylation patterns found 3 gene regions with hypermethylation amongst the progression group (Pro_MMD2, Pro_ZNF358, and Intra_F10), with a trend towards global hypomethylation, in keeping with other epithelial premalignant conditions.30 Kaz et al31 also found significant differences in methylation in patients with BE due to factors such as obesity, smoking, and sex, which may be responsible for some of the observed risk.

While the studies reviewed do show variation in methylation between progressive versus nonprogressive BE, the methodology has been heterogeneous, and few have conducted the study with a group of the same patients tracked over time.

The study aims to determine whether there are differences in methylation in patients between high-risk nondysplastic BE, which will progress to cancer, versus low-risk BE.

METHODS

Patients and Samples

Two sample cohorts were identified containing patients who either progressed to EADC from nondysplastic BE or remained with nondysplastic BE (NDBE) identified from a prospectively maintained database of patients with BE at a large district general hospital. Inclusion criteria for the study were progressing patients with nondysplastic BE who, when observed over the study period, developed EADC. Samples were only included where there was a NDBE biopsy and then histology evidence that the patient developed adenocarcinoma. Nonprogressing patients were identified from a biopsy of NDBE, which, when followed over time, never progressed beyond NDBE. To be included in this group, the patient must have been in a surveillance program for a minimum of 15 years and have serial biopsies over that period. Patients within the surveillance program had endoscopy and biopsy every 2 years.32 Biopsies were taken at the time of the initial surveillance endoscopy and at all subsequent endoscopies including immediately before treatment as part of their staging. They were also required to have still been alive and to have had a NDBE biopsy within 2 years of this study.

Patients were excluded from the study if BE material was only available as part of tumor-associated BE or if dysplasia was identified in any BE biopsies.

All tissue used was formalin-fixed paraffin-embedded (FFPE) samples obtained from pathology laboratories and prepared by the University of Birmingham Human Biomatertials Resource Centre (ethical approval 09/H1010/75). H&E stained slides were reviewed by a consultant pathologist to ensure that the samples were BE and had no dysplasia throughout their extent. Cut 5-μm sections were mounted onto frosted slides, and macrodissection for BE was carried out. DNA extraction was then performed using Qiagen DNeasy Blood and Tissue kit following the manufacturer’s protocol. Each sample of extracted DNA was then quantified and qualified by Nanodrop spectrophotometry and Qubit fluorimetry. Bisulfite conversion was performed using a Zymo DNA methylation bisulfite conversion kit following the modified Illumina Infinium protocol on 500 ng of extracted DNA.

Methylation Arrays

The Illumina HumanMethylation 450 array, in which the methylation status of more than 485,000 individual CpG sites are examined,33 was used to compare sample groups (progressing NDBE vs nonprogressing NDBE). Once bisulfite converted, 1 ng of DNA was quality-controlled (Illumina FFPE QC kit) with only samples with ΔCt <5 being taken forward to array analysis. The resulting samples underwent repair suitable for array hybridization using the Illumina FFPE restore kit,34 followed by hybridization to Illumina HumanMethylation450 arrays using manufacturer’s protocols and scanned on an Illumina iScan. Normalized intensity files (iDAT) were exported using GenomeStudio for downstream analysis.

Immunohistochemistry

Immunohistochemistry (IHC) was carried out on a Leica Bond RX system using a mouse polyclonal anti-OR3A4 antibody (Abcam ab67107) at a dilution of 1:100 with a primary incubation time of 15 minutes.

Immunohistochemistry was scored on epithelial and stromal components and a composite score consisting of the sum of expression within membranous, nuclear and cytoplasmic compartments on a score of 1 to 4 was made, giving a combined maximum possible score of 12 for each compartment. Scoring was carried out by 2 independent observers blinded to progressor/nonprogressor status.

Bioinformatics Analysis of Array Data

Bioinformatics analysis of the methylation microarrays was performed using the ChAMP package35 via Bioconductor/R. In brief, red/green intensity values were captured from Illumina iDAT files, background-corrected and SWAN-normalized to produce M values (further details given in supplementary methods, http://links.lww.com/SLA/B376).

M values were analyzed using a logistic regression model using Empirical Bayesian shrinkage of moderated t-statistics to correct for small sample size. Small sample size was controlled for by setting stringent false discovery rate Q values of <0.05. Identification of variable methylated sites allowed the CpG site markers to be highlighted. DNA copy number analysis was carried out using the DNAcopy module of ChAMP.

Pyrosequencing Validation of Hits

Methylation-insensitive primers were designed and sourced, using Qiagen PyroMark Primer Design software v2.0. Primers were designed to flank CpG sites of interest. Illumina CG methylation probe locations were retrieved from the UCSC genome browser,36 and Fast-All sequence retrieved for −200 bp to +200 bp of the target CG dinucleotide. Primer design settings were optimized to design amplicons suitable for FFPE pyrosequencing, with the optimum amplicon size set to between 80 and 150 bp. All other settings were as per the standard Qiagen design parameters. Primers were ordered from Sigma-Aldrich, with the biotinylated pyrosequencing primer being purified by high-performance liquid chromatography and the remainder by desalting. Pyrosequencing PCR was performed using Qiagen PyroMark PCR Gold kit, consisting of 2 μL of bisulfite-converted DNA, 25 μL of PCR master mix, 5 μL of CoralLoad dye, 3 μL MgSO₄, 10 μL of Q reagent, and 2.5 μL each of forward (20 mM) and reverse (20 mM) primer. Reaction conditions were determined experimentally by use of a gradient PCR for each primer pair. A typical reaction consisted of activation at 95°C for 15 minutes, followed by 45 cycles of denaturation at 94°C for 30 seconds, annealing at 56°C for 30 seconds, and extension at 72°C for 30 seconds, followed by a final extension step at 72°C for 10 minutes. In addition to experimental DNA, each PCR was performed with 100% methylated DNA, 100% unmethylated DNA, and ddH₂O as controls. Methylated and unmethylated DNA was generated in house by means of M.SSI conversion (methylated DNA) and whole genome amplification using the Qiagen Repli-G kit (unmethylated DNA). Primer sequences for validation pyrosequencing were as follows: for FGFR2 cg17337672 these were

© 2018 The Author(s). Published by Wolters Kluwer Health, Inc.
endoscopic mucosal resection and the remainder underwent
3%), T2N0 (2/30, 7%), T2N1 (3/30, 10%), T3N0 (6/30, 20%), and
high-grade dysplasia (4/30, 13%), T1N0 (10/30, 33%), T1N1 (1/30,
0.0045
NMUR1
SNORD116-22
ZNF511
SNORD115-14
TRIB3
ZAP70
RBP3
ZPP2
ATP8B1
CUX2
TMEM151B

TCAATCTTCCCCCAAACAACACT, and sequencing
GGATTATAGGTGTGAGTTAT.

forward = AGGGGAGGAAGTATAGTT; reverse = [Btn]-
TCAAATCCCCCAACCAACACT, and sequencing = GTTTA-
GAGTTTTTTTTGGATTAGT; for OR3A4 cg09890332 these
were forward = GTGGTAGAAGTAGTGGAGGTGTGATAAT,
reverse = [Btn]CTTCAACTTCCTTCCCTTACATT, and sequenc-
ing = GGTTAGGGATGGAAGA; for OR3A4 cg09890332 these
were forward = TTTAAAGTGTAGTATAGGTTAGGTGAGTTA,
reverse = [Btn]TTTCCCCAACCTAATCTACTAATAAAAT, and sequencing = GAGTTATAGGTGTGAGTTAT.

RESULTS

Patient Selection

In all, 67 patients were recruited—37 from Sandwell and West
Birmingham NHS Trust (SWBNT) and 30 from University Hospital
Birmingham NHS Trust (UHBFT). Of these, 2067 progressed from
nondysplastic BE and 47 did not. The age range was between 42 and
60 years, with a median age of 56 years, of which 60/67 (89.6%) were
male. The median time to diagnosis of EADC in “progressor”
patients was 114 months, with a range of 14–162 months. Of the
patients recruited, in the SWBNT group, 6 progressors and 6 non-
progressors were taken forward to methylation array analysis, giving a total of 12
progressors, and in the UHBFT, 6 progressors and 6 non-
progressors. Patients recruited had validation pyrosequencing
performed.

Methylation Microarray Analysis

Twenty-four samples in total were hybridized successfully to
Illumina HumanMethylation450 microarrays. All arrays
passed manufacturers QC as specified by metrics in Illumina
GenomeStudio. Differential methylation analysis at the probe level
(Table 1) revealed significant differences in methylation between
progressor and nonprogressor in nondysplastic BE (Fig. 1). In all,
44 significantly [defined as Bayes factor (BF) >5, chosen as it is
equivalent to a genome, wide P value significance of 1 × 10–4] dif-
erentially methylated targets were identified, the bulk being hypo-
methylated, with a trend towards global hypomethylation in progressor
samples as demonstrated by left-shift of the Volcano plot (Fig. 1).

Differential Methylation at the Probe Level

The top ranked differentially methylated probe was
cg09890332 (chr17:3212495–3212495, hg19 coordinates), which
tags a CpG dinucleotide −1044 bp upstream of the transcription
start site of the long noncoding RNA, OR3A4 (NRR_024128:1).
The second highest ranked differentially methylated probe was
cg24007926 (chr2:206842761–206842761, hg19 coordinates).
This CpG dinucleotide is within a large, intragenic region, with the nearest
gene being INOSD (INOS complex subunit D, NM_017759.4),
15.684 bp downstream of this CpG. The third highest ranked differ-
entially methylated probe was cg17337672 (chr10:123354172–
123354172), which tags a CpG dinucleotide within intron 2 of
FGFR2 (fibroblast growth factor receptor 2, NM_000141.4).

Differentially methylated regions (DMRs) were called
between progressors and nonprogressors via the
dmrLasso function of the CHAMP software package (Table 2). Significant DMRs
were found from chr2:503065–503193 (which tags an intragenic region,
DNM P = 7.69 × 10−4), chr5:8217236–8217322 (which also tags an
intragenic region, DMR P = 1.27 × 10−5), and chr10: 123354172–
123354172.

### TABLE 1. Top 20 Array-identified CpG Sites With Methylation Variation Between Nondysplastic Samples Which Progress to
EADC Versus Those Which Remain Static

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>t</th>
<th>P</th>
<th>Adj. P</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg08990332</td>
<td>−12.26</td>
<td>2.02E-08</td>
<td>0.0031</td>
<td>OR3A4</td>
</tr>
<tr>
<td>cg24007926</td>
<td>−11.05</td>
<td>6.84E-08</td>
<td>0.0032</td>
<td>NA</td>
</tr>
<tr>
<td>cg17337672</td>
<td>−10.95</td>
<td>7.54E-08</td>
<td>0.0032</td>
<td>FGF2</td>
</tr>
<tr>
<td>cg02226469</td>
<td>−10.88</td>
<td>1.17E-08</td>
<td>0.0038</td>
<td>NA</td>
</tr>
<tr>
<td>cg17433294</td>
<td>−10.51</td>
<td>1.22E-07</td>
<td>0.0039</td>
<td>NMUR2</td>
</tr>
<tr>
<td>cg18479711</td>
<td>−10.31</td>
<td>1.52E-07</td>
<td>0.0040</td>
<td>HDAC4</td>
</tr>
<tr>
<td>cg00911162</td>
<td>−10.15</td>
<td>1.82E-07</td>
<td>0.0045</td>
<td>LMF1</td>
</tr>
<tr>
<td>cg19733294</td>
<td>−9.83</td>
<td>2.61E-07</td>
<td>0.0045</td>
<td>NMUR1</td>
</tr>
<tr>
<td>cg16150571</td>
<td>−9.64</td>
<td>3.26E-07</td>
<td>0.0045</td>
<td>SNORD116-22</td>
</tr>
<tr>
<td>cg24422417</td>
<td>−9.62</td>
<td>3.35E-07</td>
<td>0.0045</td>
<td>ZNF511</td>
</tr>
<tr>
<td>cg13161493</td>
<td>−9.57</td>
<td>3.52E-07</td>
<td>0.0045</td>
<td>RBP3</td>
</tr>
<tr>
<td>cg14010464</td>
<td>−9.53</td>
<td>3.70E-07</td>
<td>0.0045</td>
<td>TRIB3</td>
</tr>
<tr>
<td>cg24581378</td>
<td>−9.53</td>
<td>3.71E-07</td>
<td>0.0045</td>
<td>ZAP70</td>
</tr>
<tr>
<td>cg05233064</td>
<td>−9.36</td>
<td>4.54E-07</td>
<td>0.0051</td>
<td>SNORD115-14</td>
</tr>
<tr>
<td>cg12297814</td>
<td>−9.26</td>
<td>5.08E-07</td>
<td>0.0052</td>
<td>IGFN1</td>
</tr>
<tr>
<td>cg11212140</td>
<td>−9.23</td>
<td>5.33E-07</td>
<td>0.0052</td>
<td>NA</td>
</tr>
<tr>
<td>cg11864327</td>
<td>−9.16</td>
<td>7.40E-07</td>
<td>0.0063</td>
<td>ZPP2</td>
</tr>
<tr>
<td>cg16714147</td>
<td>−9.15</td>
<td>7.78E-07</td>
<td>0.0063</td>
<td>ATP8B1</td>
</tr>
<tr>
<td>cg17304276</td>
<td>−8.88</td>
<td>8.14E-07</td>
<td>0.0063</td>
<td>CUX2</td>
</tr>
<tr>
<td>cg11443888</td>
<td>−8.86</td>
<td>8.36E-07</td>
<td>0.0063</td>
<td>TMEM151B</td>
</tr>
</tbody>
</table>

Probe ID: the Illumina cg probe ID from the Illumina manifest; T: the t value (the size of the difference relative to the variation in the sample).
P = the raw P value, not corrected for multiple testing; adjusted P = the P value corrected for multiple testing.

TABLE 1. Top 20 Array-identified CpG Sites With Methylation Variation Between Nondysplastic Samples Which Progress to EADC Versus Those Which Remain Static

© 2018 The Author(s). Published by Wolters Kluwer Health, Inc.
www.annalsofsurgery.com | 481
12335576, which spans a region from the 5'-UTR of FGFR2 to the first exon within FGFR2 (DMR \( P = 4.79 \times 10^{-5} \)).

We took advantage of the information provided by the 2-color Illumina Infinium chemistry to call copy number aberrations (CNAs) within the regions targeted by the methylation probes using the CNA calling function of CHAMP. This did not demonstrate any recurrent copy number alterations between progressors and nonprogressors. There were no significant differences in the numbers of CNA between the 2 groups, with a median of 43 CNAs (range 24–92) in the progressors versus 44 CNAs (range 30–52) in the nonprogressors \( (P = 1.0, \text{ Wilcoxon rank-sum}) \).

Pathway methylation analysis was carried out using Database for Annotation, Visualisation and Integrated Discovery. Initially Kyoto Encyclopedia of Genes & Genomes pathway analysis showed that genes associated with mitogen-activated protein kinase signaling were enriched in the dataset \( (P = 0.012) \). Gene ontology analysis using the UP_KEYWORDS feature showed significant enrichment for the disease mutation \( (P = 9.6 \times 10^{-6}) \), polymorphism \( (P = 1.7 \times 10^{-5}) \), glycoprotein \( (P = 3.4 \times 10^{-5}) \), and alternate splicing \( (P = 1.1 \times 10^{-4}) \) terms.

### Validation Pyrosequencing

Because of the likely biological relevance of FGFR2, and the data demonstrating that OR3A4 was the top differentially methylated CpG, validation pyrosequencing was carried out on all 67 patients. Normality of distribution of methylation values was ascertained by histogram plots, in which it was found that methylation was non-normally distributed; therefore nonparametric testing was carried out.

For OR3A4 cg09890332, median methylation was 67.8\% [interquartile range (IQR) 12.1] in progressors versus 96.7\% (IQR 16.1) in nonprogressors \( (P = 0.0001, z = 5.158; \text{ Wilcoxon rank-sum test}) \). The pyrosequencing assay design used covered 2 additional CpG +4 bp and +10 bp downstream of cg09890332. Median methylation in these was 66.8\% and 59.7\% in progressors versus 75.0\% and 68.1\% in nonprogressors \( (P = 0.0280, z = 2.197, 2.088; \text{ Wilcoxon rank-sum test}) \). To investigate whether this phenomenon was localized to this region or was a gene-wide phenomenon, an additional pyrosequencing assay was designed based on probe ID cg07863524 (chr17:3213471–3213471), which is +976 bp downstream from cg09890332 and −68 bp from the transcription start site of OR4A4. This demonstrated that median methylation was 62.2\% in progressors and 56.7\% in nonprogressors \( (P = 0.600, z = −0.524; \text{ Wilcoxon rank-sum test}) \). A temporal analysis of change in methylation of cg09890332 over time is shown in Fig. 3, showing that the difference between methylation levels at initial biopsy is static between progressors and nonprogressors, and that the difference is maintained over time and is detectable for an extended period of time before diagnosis of EADC.

### Expression of OR3A4 in Progressors Versus Nonprogressors

We then carried out immunohistochemical assessment of expression of OR3A4 (Fig. 4), which, although is labeled as long noncoding RNA, is actually expressed in tissues (see Supplementary results, http://links.lww.com/SLA/B376), in a subset of 12 patients. For the stromal compartment, a median expression of 6 (IQR 4–7) was seen in progressors and 5.5 (IQR 4–6) in nonprogressors \( (P = 0.0308, z = 2.160) \). For the epithelial compartment, a median expression of 8 (IQR 8–10) was seen in progressors and 5.5 (IQR 3–10) in nonprogressors (Wilcoxon rank-sum \( P = 0.4587, z = −0.741 \)). Percentage methylation at OR3A4 and stromal expression were -0.741.

### Table 2. Table of Differentially Methylated Regions (DMRs)

<table>
<thead>
<tr>
<th>DMR ID</th>
<th>Probe ID</th>
<th>Probe Level Adjusted ( P )</th>
<th>Chromosome</th>
<th>Gene</th>
<th>Start of DMR (bp)</th>
<th>End of DMR (bp)</th>
<th>Size in bp</th>
<th>Change in Methylation</th>
<th>( P ) for DMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cg21273584</td>
<td>0.014</td>
<td>2</td>
<td>NA</td>
<td>502999</td>
<td>503195</td>
<td>197</td>
<td>−34%</td>
<td>7.69E-04</td>
</tr>
<tr>
<td>1</td>
<td>cg08854591</td>
<td>0.045</td>
<td>2</td>
<td>NA</td>
<td>502999</td>
<td>503195</td>
<td>197</td>
<td>−16%</td>
<td>7.69E-04</td>
</tr>
<tr>
<td>1</td>
<td>cg11573608</td>
<td>0.014</td>
<td>2</td>
<td>NA</td>
<td>502999</td>
<td>503195</td>
<td>197</td>
<td>−27%</td>
<td>7.69E-04</td>
</tr>
<tr>
<td>2</td>
<td>cg25568703</td>
<td>0.047</td>
<td>5</td>
<td>NA</td>
<td>8216903</td>
<td>8217655</td>
<td>753</td>
<td>−23%</td>
<td>1.27E-03</td>
</tr>
<tr>
<td>2</td>
<td>cg25016964</td>
<td>0.039</td>
<td>5</td>
<td>NA</td>
<td>8216903</td>
<td>8217655</td>
<td>753</td>
<td>−22%</td>
<td>1.27E-03</td>
</tr>
<tr>
<td>2</td>
<td>cg17642708</td>
<td>0.021</td>
<td>5</td>
<td>NA</td>
<td>8216903</td>
<td>8217655</td>
<td>753</td>
<td>−33%</td>
<td>1.27E-03</td>
</tr>
<tr>
<td>3</td>
<td>cg10788901</td>
<td>0.172</td>
<td>10</td>
<td>FGFR2</td>
<td>123352704</td>
<td>123355661</td>
<td>2958</td>
<td>−27%</td>
<td>4.79E-03</td>
</tr>
<tr>
<td>3</td>
<td>cg14856220</td>
<td>0.044</td>
<td>10</td>
<td>FGFR2</td>
<td>123352704</td>
<td>123355661</td>
<td>2958</td>
<td>−24%</td>
<td>4.79E-03</td>
</tr>
<tr>
<td>3</td>
<td>cg17337672</td>
<td>0.003</td>
<td>10</td>
<td>FGFR2</td>
<td>123352704</td>
<td>123355661</td>
<td>2958</td>
<td>−47%</td>
<td>4.79E-03</td>
</tr>
<tr>
<td>3</td>
<td>cg02412684</td>
<td>0.031</td>
<td>10</td>
<td>FGFR2</td>
<td>123352704</td>
<td>123355661</td>
<td>2958</td>
<td>−38%</td>
<td>4.79E-03</td>
</tr>
<tr>
<td>3</td>
<td>cg06791446</td>
<td>0.058</td>
<td>10</td>
<td>FGFR2</td>
<td>123352704</td>
<td>123355661</td>
<td>2958</td>
<td>−13%</td>
<td>4.79E-03</td>
</tr>
<tr>
<td>3</td>
<td>cg22633036</td>
<td>0.581</td>
<td>10</td>
<td>FGFR2</td>
<td>123352704</td>
<td>123355661</td>
<td>2958</td>
<td>−4%</td>
<td>4.79E-03</td>
</tr>
</tbody>
</table>
was strongly negatively correlated (Pearson correlation coefficient $-0.85$, $P = 0.014$), and a similar, but nonsignificant correlation was observed with epithelial expression and methylation at OR3A4 (Pearson correlation coefficient $-0.40$, $P = 0.373$).

**Ability of OR3A4 Methylation to Act as a Discriminator in BE**

To understand the accuracy of using methylation within cg09890332 of OR3A4 as a biomarker for high-risk BE, we carried out a multivariable reverse stepwise logistic regression analysis of methylation at the 3 tagged CpG dinucleotides within the pyrosequencing assay as the independent variables and progressor versus nonprogressor status as the dependent variables. In this model, CpGs 2 and 3 became nonsignificant ($P = 0.3325$ and $P = 0.4764$) and were removed from the model, leaving the first CpG in cg09890332 as being significant [coef $-0.0563$, SE $0.016$, $z = -3.40$, $P = 0.001$, 95% confidence interval (CI) $-0.089$ to $-0.024$]. Using receiver-operator curve modeling, the area under curve of this model was 0.82 (95% CI 0.80–0.83) in the cohort where the marker was originally generated (Supplementary Fig. 2, http://links.lww.com/SLA/B376).

We then used the `diagt` function of Stata 11.2 to model a set methylation threshold effect on sensitivity and specificity of the test, aiming for maximum negative predictive value and correcting for an incidence rate within the cohort of 0.7%. Modeling at a threshold of below 89% being significant showed that hypomethylation at OR3A4 can predict progression to invasive carcinoma with a sensitivity of 70.8%, specificity of 86%, positive predictive value of 85% and negative predictive value of 72.5%.

We then carried out validation bisulfite pyrosequencing on a cohort of progressors ($n = 18$) versus nonprogressors ($n = 14$, Southampton cohort), finding that there were significant differences ($P = 0.0477$, unpaired $t$ test) in methylation, with an average methylation of 59.2% (95% CI 56.2%–62.1%) in progressors versus
63.5% (95% CI 60.2%–66.7%) in the nonprogressors. Regression model demonstrated area under curve 0.70, and adjustment for a prevalence of 0.7% using a threshold of 58% demonstrated a sensitivity of 33.3%, specificity of 78.6%, positive predictive value of 10.5%, and negative predictive value of 94%.

CONCLUSIONS
We have identified that hypomethylation at cg09890332 corresponding to the CG nucleotide at position (CHR) of OR3A4 can discriminate between patients who progress from nondysplastic BE and those who did not. This association is maintained across independent cohorts, and seems to be related temporarily (ie, the association is maintained in the earliest set of samples from a time series of follow-up biopsies in patients with BE), and also by case status. Gastroesophageal reflux is a key risk factor in the development of BE, and also obesity and cigarette smoking.

The effect of hypomethylation on the OR3A4 gene seems to be functional, in that immunohistochemistry reveals an increase in OR3A4 expression in samples with hypomethylation. The finding that the stromal expression, in particular, is increased is of interest, given the known effect of “pathological” stroma in the pathogenesis of esophageal cancer. Our observed region coincides within 225 bp of a CTCF and RAD21 transcription factor binding site, further suggesting that methylation there has a functional effect to prevent transcription factor binding and alter gene expression. Guo et al performed a genome-wide screen of long noncoding RNAs in gastric adenocarcinoma, finding that OR3A4 was significantly (55.9-fold) overexpressed in these patients. They also observed that levels of OR3A4 were correlated with metastatic potential and prognosis. Furthermore, they utilized OR3A4 overexpression vectors and performed siRNA knockdown to demonstrate that OR3A4 seems to regulate cellular proliferation in gastric cancer cell lines. Finally, they utilized their overexpressing cell line models and implanted them into nude mice, finding that OR3A4 overexpressing gastric cancer cell lines grew significantly faster and more aggressively than with knockdown of OR3A4. Downstream analysis of target genes demonstrated that OR3A4 targets PDLIM2, a putative tumor suppressor that regulates cell cycle and adhesion; PCLL1, a transcriptional silencer; and DLX4 which induces epithelial-mesenchymal transition via TWIST1.

We found both at the individual probe level and as part of a DMR that there is hypomethylation in the CpG island associated with FGFR2; however, this did not validate at the single probe level when examined with bisulfite pyrosequencing. FGFR2 has been observed to undergo recurrent alteration in both esophageal adenocarcinoma and squamous cell carcinoma, with the latter demonstrating recurrent amplification. The disparity between our microarray results and validation by pyrosequencing may be due to probe inflation caused by small sample size, and is a significant weakness of our study; however, given its biological associations with esophageal adenocarcinoma, further work is needed.

In common with premalignant lesions in cancer, such as colorectal adenomatous polyps, we observed a trend towards genome-wide hypomethylation as demonstrated by a leftward shift of our genome-wide volcano plot, suggesting a widespread overexpression of genes as part of the development towards malignancy. We also found no difference in chromosomal instability between progressors and nonprogressors, although there was widespread instability within both sets of samples, in common with what has previously been observed in BE.

Another weakness of our study was the inability to carry out a more comprehensive validation of all observed markers as part of a larger panel of markers. Our study made use of extremely small tissue biopsies from endoscopic surveillance programs, which limited the quantity of usable DNA that could be extracted from these samples and used for downstream validation, and thus validation of the observed DMR and other differentially methylated position regions could not be carried out. We were also limited in the number of samples that could be tested, because of the rarity of biopsy samples before the diagnosis of esophageal cancer, as we took advantage of a local screening program to obtain samples. However, in both the genome-wide and in the validation phase, we believe we have sufficient power to detect methylation changes in this marker. In the genome-wide phase, our sample size of 24 patients would allow us to detect a methylation difference of 2%, with a statistical power of 90%. Similarly, in the validation phase, our sample size would allow us to detect a minimum methylation change of 10% in the sample set, given the previously observed median methylation and standard deviation in these samples.

We observed a median time to diagnosis of esophageal adenocarcinoma on commencement of the surveillance program of 114 months, which we believe is a reflection of the early identification of these patients and their enrolment into a screening program, and the known slow progression of esophageal adenocarcinoma. A further problem with molecular genetic analysis is heterogeneity, due to the low proportion of cells within a biopsy specimen that contain changes compatible with BE, which leads to less clear methylation changes.

Hypomethylation of OR3A4, although seemingly accurate for the detection of progression of Barrett to invasive adenocarcinoma, is likely to be of more utility as a multimodal stratifier in BE, taking account of previous findings at the mutational and copy number level, and also epigenetic change. However, for the purpose of designing a surveillance program with the ability to risk-stratify the nondysplastic BE population, this marker has significant potential utility.

Development of a streamlined surveillance program could lead to cost savings through the avoidance of unnecessary upper gastrointestinal endoscopy or via a less invasive technology such as the Cytosponge, in a low-risk cohort identified by a molecular marker panel. More frequent endoscopy in the high-risk cohort could lead to earlier diagnosis of EADC or initiate management of BE to arrest further progression. In conclusion, development of a stratified marker panel in the context of a clinical trial is now needed to improve diagnosis of high-risk BE.

REFERENCES

© 2018 The Author(s). Published by Wolters Kluwer Health, Inc.