Title: Whole genome methylation analysis of non-dysplastic Barrett's oesophagus that progresses to invasive cancer

Authors: Dilworth MP FRCS(Gen) *,[1], Nieto T MRCS*,[1], Stockton JD PhD[1], Whalley C MSc[1], Tee L MSc[1], James JD PhD[1], Noble F MRCS[5], Underwood TJ FRCS(Gen)[5], Hallissey MT FRCS[2], Hejmadi R FRCPath[1], Trudgill N FRCP[3], Tucker O FRCS*,[4], Beggs AD FRCS(Gen) PhD*,[1,2]

1. Institute of Cancer and Genomic Science, University of Birmingham, U.K.
2. Queen Elizabeth Hospital, Birmingham, U.K.
3. Sandwell and West Birmingham NHS Trust, Birmingham, U.K.
4. Heart of England NHS Trust, Birmingham, U.K.
5. Cancer Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, U.K.

* - these authors contributed equally to this study.

Corresponding author:

Dr Andrew D Beggs,

Institute of Cancer & Genomic Science

University of Birmingham

Vincent Drive

Birmingham

B15 2TT

Email: a.beggs@bham.ac.uk

Telephone: +44 (0)121 414 7458

Reprints will not be available

Funding support: AB acknowledges funding from the Wellcome Trust (102732/Z/13/Z), Cancer Research UK (C31641/A23923) and the Medical Research Council (MR/M016587/1). AB also acknowledges support from the University of Birmingham Human Biomaterials Resource Centre. MPD and TN acknowledge funding from the QE Hospital Charities.TJU acknowledges funding by the Medical Research Council (G1002565) and Cancer Research UK (C10104/A23924). TJU thanks the University of Southampton CRUK/ECMC Tissue Bank.

Running head: Epigenetics of Barrett's oesophagus

MINI-ABSTRACT:

Identifying patients with Barrett's oesophagus (BO) who progress to adenocarcinoma amenable to surgical treatment remains a challenge. In order to investigate this cohort, we carried out analysis of the epigenome of "progressive" vs. "non-progressive" BO finding significant epigenetic variation between these groups that identified new pathways for therapy & diagnosis.

ABSTRACT

Objective: To investigate differences in methylation between patients with non-dysplastic Barrett's' oesophagus who progress to invasive adenocarcinoma and those who do not.

Summary background data: Identifying patients with non-dysplastic Barrett's oesophagus 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 non-dysplastic Barrett's Oesophagus who either develop adenocarcinoma or remain static, with validation of findings by bisulfite pyrosequencing

Results: In total, 12 patients with "progressive" vs. 12 with "non-progressive" non-dysplastic Barrett's oesophagus were analysed via methylation array. Forty-four methylation markers were identified that may be able to discriminate between non-

dysplastic Barrett's Oesophagus that either progress to adenocarcinoma or remain static. Hypomethylation of the recently identified tumour suppressor *OR3A4* (probe cg09890332) validated in a separate cohort of samples (median methylation in progressors = 67.8% vs. 96.7% in non-progressors, p=0.0001, z = 3.85, Wilcoxon rank sum test) and was associated with the progression to adenocarcinoma. There were no differences in copy number between the two 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 non-dysplastic Barrett's Oesophagus and may form the basis of a future surveillance program.

Keywords: Barrett's oesophagus; Oesophageal cancer; Methylation; Cancer Genetics

INTRODUCTION

Oesophageal adenocarcinoma (OADC) incidence is increasing [1] and currently represents 5% of the digestive tract cancers in the UK [2]. Overall disease survival is poor [3] but correlates with stage of cancer at presentation, demonstrating significant survival advantages with detection of early stage disease [4, 5].

Barrett's Oesophagus (BO), 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 oesophagus causing chronic inflammation and tissue damage [6].

The incidence of BO is increasing, largely thought to be a consequence of obesity induced reflux disease [7-10]. Barrett's oesophagus is associated with an increased risk of OADC [11], but for the majority of patients, BO will never progress beyond simple benign metaplasia [12, 13]. However, in a small number of patients, dysplasia will develop with some progressing to OADC [14]. The incidence of OADC in the BO population is up to 150 times greater than unaffected individuals [12].

Although the pathological changes seen in Barrett's adenocarcinoma are understood as part of a well-established metaplasia-dysplasia-carcinoma sequence [15], the molecular drivers are less clear[16-18]. The current dilemma is that for patients with non-dysplastic BO, there are no accurate methods for identifying the small number of patients at high risk of progression to cancer.

Surveillance practice of the Barrett's patients varies widely, between some who endoscope patients each year in contrast to others who will never repeat the investigation[19]. The on-going UK Medical Research Council funded BOSS study

aims to understanding the optimum surveillance strategy, randomising between prospective monitoring BO patients with frequent endoscopic assessment or a “watch and wait” policy [20].

Clearly, there is need for a method of risk stratification in these patients in order to facilitate a streamlined surveillance programme by identifying high risk non-dysplastic BO patients. Attempts at biomarker development for stratification of high risk Barrett’s oesophagus have focused on mutational change, specifically around the role of *TP53* mutation in predicting “high risk” disease [21] given its role as a driver in oesophageal cancer. However Ross-Innes et al [17] have convincingly demonstrated the presence of pathogenic *TP53* mutations in apparently normal squamous oesophageal mucosa thus making its role in progression to invasive adenocarcinoma unclear. However, the role of epigenetic change in the pathogenesis of Barrett’s oesophagus and oesophageal 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 tumourigenesis, should a suitable epigenetic modulator be identified.

Multiple methylation markers have been identified which can discriminate between high risk and low risk BO including *APC/ p16* [22], *MGMT* [23], *PKP-1* [24], *TIMP3/TERT* [25], *RUNX3/HPP1* [26 27] and *AKAP12* [28]. Agarwal *et al* [29] performed a MeCIP array based approach to compare the methylomes of progressor (n=5) vs. non-progressing 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 three 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 pre-malignant conditions [30]. Kaz et al [31] also found significant differences in methylation in patients with BO due to factors such as obesity, smoking and gender which may be responsible for some of the observed risk.

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

AIMS

To determine whether there are differences in methylation in patients between high risk non-dysplastic Barrett's Oesophagus which will progress to cancer vs low risk Barrett's

METHODS

Patients and Samples:

Two sample cohorts were identified containing patients who either progressed to OADC from non-dysplastic BO or remained with non-dysplastic BO (NDBO) identified from a prospectively maintain database of patients with BO at a large district general hospital. Inclusion criteria for the study were progressing patients with non-dysplastic Barrett's oesophagus who, when observed over the study period, developed OADC. Samples were only included where there was a NDBO biopsy and then histology evidence that the patient developed adenocarcinoma. Non-progressing patients were identified from a biopsy of NDBO which when followed over time never progressed beyond NDBO. To be included in this group, the patient must have been in a surveillance programme for a minimum of 15 years and have serial biopsies over that period. Patients within the surveillance programme 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 prior to treatment as part of their staging. They were also required to have still been alive and to have had a NDBO biopsy within 2 years of this study.

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

All tissue used was formalin fixed paraffin embedded (FFPE) samples obtained from pathology libraries and prepared by the University of Birmingham Human Biomaterials Resource Centre (ethical approval 09/H1010/75). H&E stained slides were reviewed by a consultant pathologist to ensure that the samples were BO and

had no dysplasia throughout their extent. Cut 5μM paraffin sections were mounted onto frosted slides, and macrodissection for BO 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. Bisulphite conversion was performed using a Zymo DNA Methylation bisulphite conversion kit following the modified Illumina Infinium protocol on 500ng 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 NDBO vs non-progressing NDBO). Once bisulphite converted, 1ng of DNA was quality controlled (Illumina FFPE QC kit) with only samples with dCt <5 being taken forward to array analysis. The resulting samples underwent repair suitable for array hybridisation using the Illumina FPPE restore kit. [34] followed by hybridisation to Illumina HumanMethylation450 arrays using manufacturers protocols and scanned on an Illumina iScan. Normalised intensity files (iDAT) were exported using GenomeStudio for downstream analysis.

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.

IHC 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 1-4 was made, giving a combined maximum possible score of 12 for each compartment. Scoring was carried out by two independent observers blinded to progressor/non-progressor status.

Bioinformatics Analysis of Array Data

Bioinformatics analysis of the methylation microarrays was carried using the ChAMP package [35] via Bioconductor/R. In brief, red/green intensity values were captured from Illumina iDAT files, background corrected and SWAN normalised to produce M values (further details given in supplementary methods).

M values were analysed 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 FDR 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 FASTA sequence retrieved for -200bp to + 200bp of the target CG dinucleotide. Primer design settings were optimised to design amplicons suitable for

FFPE pyrosequencing, with the optimum amplicon size set to between 80-150bp. 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 2uL of bisulphite converted DNA, 25uL of PCR master mix, 5uL of CoralLoad dye, 3uL $MgSO_4$, 10uL of Q reagent and 2.5uL each of forward (20mM) and reverse (20mM) primer. Reaction conditions were determined experimentally by use of a gradient PCR for each primer pair. A typical reaction consisted of activation at 95C for 15 minutes, followed by 45 cycles of denaturation at 94C for 30s, annealing at 56C for 30s and extension at 72C for 30 seconds, followed by a final extension step at 72C for 10 minutes. In addition to experimental DNA, each PCR was performed with 100% methylated DNA, 100% un-methylated 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 Forward = AGGGGAAGGGAATTTAGGTT; Reverse = [Btn]TCAATCTTCCCCCAAACAACCACT and sequencing = GTTTAGAAGTTTTTTTTGGATTAGT. For ORA3A4 cg07863524 these were forward = GTGGTAGAAGTAGGATGAGGTGTTGATAAT; Reverse =[Btn]CTTCAACTTCCTTCCCCTTACATTT and sequencing = GGGTAGGGATGGAAGA . For OR3A4 cg09890332 these were Forward = TTAAAGTGTTAGGATTATAGGTGTGAGTTA, reverse =

[Btn]TTTCCCAACCCTAATCACTACTAATAAAAT and sequencing = GGATTATAGGTGTGAGTTAT.

RESULTS:

Patient selection:

In total, 67 patients were recruited, 37 from Sandwell and West Birmingham NHS Trust (SWBNT) and 30 from University Hospital Birmingham NHS Trust (UHBFT). Of these, 20/67 progressed from non-dysplastic Barrett's oesophagus and 47 did not. The age range was between 42-60 years with a median age of 56 years of which 60/67 (89.6%) were of male gender. The median time to diagnosis of OADC 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 and in the UHBFT 6 progressors and 6 non-progressors were taken forward to methylation array analysis, giving a total of 12 progressors and 12 non-progressors. This samples size was chosen because of our previous experience with biomarker discovery in methylation arrays as a suitable size for biomarker discovery. A validation cohort of 32 patients (progressors=18, non-progressors=14) were obtained from University Hospital Southampton. All patients included in the study had symptoms of reflux disease as a presenting symptom. For all patients in the progressor cohort (n=30) the observed pathological disease stages at the time of resection were high grade dysplasia (4/30, 13%), T1N0 (10/30, 33%), T1N1 (1/30, 3%), T2N0 (2/30, 7%), T2N1 (3/30,10%), T3N0 (6/30, 20%) and T3N1 (4/30, 13%). All patients with high grade dysplasia underwent endoscopic mucosal resection and the remainder underwent oesophagectomy. All patients recruited had validation pyrosequencing performed.

Methylation microarray analysis

Twenty four samples in total were hybridised 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 non-progressors in non-dysplastic BO (Figure 1). In total, 44 significantly (defined as Bayes Factor, BF>5, chosen as it is equivalent to a genome wide p-value significance of $1x10^{-6}$) differentially methylated targets were identified, the bulk being hypomethylated, with a trend towards global hypomethylation in progressor samples as demonstrated by left-shift of the Volcano plot (Figure 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 -1044bp upstream of the transcription start site of the long non-coding 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 *INO80D* (INO80 complex subunit D, NM_017759.4), 15,684bp downstream of this CpG. The third highest ranked differentially 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 (DMR) were called between progressors and non-progressors via the *dmrLasso* function of the CHAMP software package (Table 2).

Significant DMRs were found from chr2:503065-503193 (which tags an intragenic region, DMR p= 7.69x10-4), chr5:8217236-8217322 (which also tags an intragenic region, DMR p=1.27x10-3) and chr10: 123353418- 123355576 which spans a region from the 5' UTR of *FGFR2* to the 1st exon within *FGFR2* (DMR p=4.79x10^{-3}).

We took advantage of the information provided by the two colour Illumina Infinium chemistry to call copy number aberrations (CAN) 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 non-progressors. There were no significant differences in the numbers of CNA between the two groups, with a median of 43 CNA (range 24-92) in the progressors vs. 44 CAN (range 30-52) in the non-progressors (p=1.0, Wilcoxon rank sum).

Pathway methylation analysis was carried out using DAVID. Initially KEGG pathway analysis showed that genes associated with MAPK signalling were enriched in the dataset (p=0.012). Gene ontology analysis using the UP_KEYWORDS feature showed significant enrichment for the disease mutation (p=9.6x10-6), polymorphism (p=1.7x10-5), glycoprotein (p=3.4x10-5), and alternate splicing (p=1.1x10-4) 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 non-parametric testing was carried out

For OR3A4 cg09890332, median methylation was 67.8% (IQR = 12.1) in progressors vs. 96.7% (IQR 16.1) in non-progressors (p=0.0001, z = 5.158, Wilcoxon rank sum test) (Figure 2). The pyrosequencing assay design used covered two additional CpG +4bp and +10bp downstream of cg09890332. Median methylation in these was 66.8% and 59.7% in progressors vs. 75.0% and 68.1% in non-progressors (p=0.0280 and 0.0368, z=2.197 and 2.088, Wilcoxon rank sum test). In order to investigate whether this phenomenon was localised 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 +976bp downstream from cg09890332 and -68bp from the TSS of OR4A4. This demonstrated that median methylation was 62.2% in progressors and 56.7% in non-progressors (p=0.600, z=-0.524, Wilcoxon rank sum test). A temporal analysis of change in methylation of cg09890332 over time is shown in Figure 3, showing that the difference between methylation levels at initial biopsy is static between progressors and non-progressors and that the difference is maintained over time and is detectable for an extended period of time before diagnosis of OADC.

We then validated cg17337672 within FGFR2, finding that median methylation was 83.4% in progressors vs. 82.2% in non-progressors (p=0.51,z=0.653, Wilcoxon rank sum test).

Expression of OR3A4 in progressors vs. non-progressors

We then carried out Immunohistochemical assessment of expression of OR3A4 (Figure 4), which, although is labelled as long non-coding RNA is actually expressed in tissues (see supplementary results), in a subset of 12 patients. For the stromal

compartment, a median expression of 6 (IQR 4-7) was seen in progressors and 2 (IQR 1-3) in non-progressors (Wilcoxon rank sum 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 non-progressors (Wilcoxon rank sum p=0.4587, z=-0.741). Percentage methylation at OR3A4 and stromal expression was strongly negatively correlated (Pearson correlation coefficient = -0.85, p=0.014) and a similar, but non-significant 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 BO

In order to understand the accuracy of using methylation within cg09890332 of OR3A4 as a biomarker for high-risk Barrett's oesophagus we carried out a multivariable reverse stepwise logistic regression analysis of methylation at the three tagged CpG dinucleotides within the pyrosequencing assay as the independent variables and progressor vs. non-progressor status as the dependent variable. In this model, CpGs two and three became non-significant (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% CI -0.089- -0.024). Using ROC modelling the AUC of this model was 0.82 (95% CI 0.80-0.83) in the cohort where the marker was originally generated (Supplementary figure 2).

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%. Modelling 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 bisulphite pyrosequencing on a cohort of progressors (n=18) vs. non-progressors (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 vs. 63.5% (95% CI 60.2-66.7%) in the non-progressors. Regression model demonstrated AUC = 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 non-dysplastic Barrett's oesophagus and those who did not. This association is maintained across independent cohorts, and seems to be related temporally (i.e. the association is maintained in the earliest set of samples from a time series of follow-up biopsies in patients with Barretts oesophagus) as well as by case status. Gastro-oesophageal reflux is a key risk factor in the development of Barretts oesophagus [19] as well as obesity and cigarettes 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 oesophageal cancer[37]. Our observed region coincides within 225bp 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 [38] performed a genome wide screen of long non-coding RNAs in gastric adenocarcinoma, finding that *OR3A4* was significantly (55.9 fold,) over-expressed in these patients. They also observed that levels of *OR3A4* were correlated with metastatic potential and prognosis. Furthermore, they utilised *OR3A4* over-expression vectors and performed siRNA knockdown to demonstrate that *OR3A4* seems to regulate cellular proliferation in gastric cancer cell lines. Finally, they utilised their over-expressing cell line models and implanted them into nude mice, finding that *OR3A4* over-expressing 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 tumour suppressor than regulates cell cycle and adhesion; *PIWIL1*, 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 differentially methylated region that there is hypomethylation in the CpG island associated with *FGFR2*, however this did not validate at the single probe level when examined with bisulphite pyrosequencing. *FGFR2* has been observed to undergo recurrent alteration in both oesophageal adenocarcinoma [39] and squamous cell carcinoma [40] 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 oesophageal adenocarcinoma further work is needed.

In common with pre-malignant lesions in cancer, such as colorectal adenomatous polyps [30], we observed a trend towards genome wide hypomethylation as demonstrated by a leftward shift of our genome wide volcano plot, suggesting a widespread over-expression of genes as part of the development towards malignancy. We also found no difference in chromosomal instability between progressors and non-progressors, although there was widespread instability within both sets of samples, in common with what has previously been observed [41] in Barrett's oesophagus.

Another weakness of our study is 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 programmes, that 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 DMP 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 prior to the diagnosis of oesophageal cancer, as we took advantage of a local screening programme 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 [42] 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 an median time to diagnosis of oesophageal 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 programme and the known slow progression of oesophageal 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 Barrett's oesophagus, which leads to less clear methylation changes.

Hypomethylation of *OR3A4*, although seemingly accurate for the detection of progression of Barrett's to invasive adenocarcinoma, is likely to be of more utility as a multi-modal stratifier in Barrett's oesophagus, taking account of previous findings at the mutational and copy number level, as well as epigenetic change. However, for the purpose of designing a surveillance programme with the ability to risk stratify the non-dysplastic Barrett's oesophagus patient, this marker has significant potential utility.

Development of a streamlined surveillance programme could lead to cost savings through the avoidance of un-necessary upper gastro-intestinal endoscopy or via a less invasive technology such as the CytoSponge [43], 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 OADC or initiate management of BO 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 BO.

REFERENCES

1. Lepage C, Rachet B, Jooste V, et al. Continuing rapid increase in esophageal adenocarcinoma in England and Wales. *Am J Gastroenterol* 2008; 103(11):2694-9.
2. Lepage C, Bouvier AM, Manfredi S, et al. Trends in incidence and management of esophageal adenocarcinoma in a well-defined population. *Gastroenterol Clin Biol* 2005; 29(12):1258-63.
3. Pennathur A, Gibson MK, Jobe BA, et al. Oesophageal carcinoma. *Lancet* 2013; 381(9864):400-12.
4. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med* 2003; 349(23):2241-52.
5. Pennathur A, Farkas A, Krasinskas AM, et al. Esophagectomy for T1 esophageal cancer: outcomes in 100 patients and implications for endoscopic therapy. *Ann Thorac Surg* 2009; 87(4):1048-54; discussion 1054-5.
6. Sun D, Wang X, Gai Z, et al. Bile acids but not acidic acids induce Barrett's esophagus. *Int J Clin Exp Pathol* 2015; 8(2):1384-92.
7. Spechler SJ, Goyal RK. The columnar-lined esophagus, intestinal metaplasia, and Norman Barrett. *Gastroenterology* 1996; 110(2):614-21.
8. McDougall NI, Johnston BT, Collins JS, et al. Three- to 4.5-year prospective study of prognostic indicators in gastro-oesophageal reflux disease. *Scand J Gastroenterol* 1998; 33(10):1016-22.
9. Lagergren J, Bergstrom R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med* 1999; 340(11):825-31.
10. Cook MB, Corley DA, Murray LJ, et al. Gastroesophageal reflux in relation to adenocarcinomas of the esophagus: a pooled analysis from the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON). *PLoS One* 2014; 9(7):e103508.
11. Phillips WA, Lord RV, Nancarrow DJ, et al. Barrett's esophagus. *J Gastroenterol Hepatol* 2011; 26(4):639-48.
12. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med* 2011; 365(15):1375-83.
13. Bhat S, Coleman HG, Yousef F, et al. Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. *J Natl Cancer Inst* 2011; 103(13):1049-57.
14. Wani S, Falk G, Hall M, et al. Patients with nondysplastic Barrett's esophagus have low risks for developing dysplasia or esophageal adenocarcinoma. *Clin Gastroenterol Hepatol* 2011; 9(3):220-7; quiz e26.
15. Brown CS, Ujiki MB. Risk factors affecting the Barrett's metaplasia-dysplasia-neoplasia sequence. *World J Gastrointest Endosc* 2015; 7(5):438-45.
16. Cancer Genome Atlas Research N, Analysis Working Group: Asan U, Agency BCC, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017; 541(7636):169-175.
17. Ross-Innes CS, Becq J, Warren A, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet* 2015; 47(9):1038-46.
18. Secrier M, Li X, de Silva N, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat Genet* 2016; 48(10):1131-41.
19. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 2014; 63(1):7-42.
20. Jankowski J, Barr H. Improving surveillance for Barrett's oesophagus: AspECT and BOSS trials provide an evidence base. *BMJ* 2006; 332(7556):1512.
21. Dolan K, Walker SJ, Gosney J, et al. TP53 mutations in malignant and premalignant Barrett's esophagus. *Dis Esophagus* 2003; 16(2):83-9.

22. Wang JS, Guo M, Montgomery EA, et al. DNA promoter hypermethylation of p16 and APC predicts neoplastic progression in Barrett's esophagus. *Am J Gastroenterol* 2009; 104(9):2153-60.
23. Kuester D, El-Rifai W, Peng D, et al. Silencing of MGMT expression by promoter hypermethylation in the metaplasia-dysplasia-carcinoma sequence of Barrett's esophagus. *Cancer Lett* 2009; 275(1):117-26.
24. Kaz AM, Luo Y, Dzieciatkowski S, et al. Aberrantly methylated PKP1 in the progression of Barrett's esophagus to esophageal adenocarcinoma. *Genes Chromosomes Cancer* 2012; 51(4):384-93.
25. Clement G, Braunschweig R, Pasquier N, et al. Methylation of APC, TIMP3, and TERT: a new predictive marker to distinguish Barrett's oesophagus patients at risk for malignant transformation. *J Pathol* 2006; 208(1):100-7.
26. Schulmann K, Sterian A, Berki A, et al. Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. *Oncogene* 2005; 24(25):4138-48.
27. Jin Z, Cheng Y, Gu W, et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. *Cancer Res* 2009; 69(10):4112-5.
28. Jin Z, Hamilton JP, Yang J, et al. Hypermethylation of the AKAP12 promoter is a biomarker of Barrett's-associated esophageal neoplastic progression. *Cancer Epidemiol Biomarkers Prev* 2008; 17(1):111-7.
29. Agarwal R, Jin Z, Yang J, et al. Epigenomic program of Barrett's-associated neoplastic progression reveals possible involvement of insulin signaling pathways. *Endocr Relat Cancer* 2012; 19(1):L5-9.
30. Beggs AD, Jones A, El-Bahrawy M, et al. Whole-genome methylation analysis of benign and malignant colorectal tumours. *J Pathol* 2013; 229(5):697-704.
31. Kaz AM, Wong CJ, Varadan V, et al. Global DNA methylation patterns in Barrett's esophagus, dysplastic Barrett's, and esophageal adenocarcinoma are associated with BMI, gender, and tobacco use. *Clin Epigenetics* 2016; 8:111.
32. Theron BT, Padmanabhan H, Aladin H, et al. The risk of oesophageal adenocarcinoma in a prospectively recruited Barrett's oesophagus cohort. *United European Gastroenterol J* 2016; 4(6):754-761.
33. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011; 98(4):288-95.
34. Illumina FFPE restore manual 2015. Available at: http://support.illumina.com/content/dam/illumina-support/documents/myillumina/5c3d90a3-793c-4a8b-932b-0434590f98ef/infinium_ffpe_sample_restore_booklet_15014614_c.pdf.
35. ChAMP package 2015. Available at: http://www.bioconductor.org/packages/release/bioc/html/ChAMP.html.
36. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002; 12(6):996-1006.
37. Underwood TJ, Hayden AL, Derouet M, et al. Cancer-associated fibroblasts predict poor outcome and promote periostin-dependent invasion in oesophageal adenocarcinoma. *J Pathol* 2015; 235(3):466-77.
38. Guo X, Yang Z, Zhi Q, et al. Long noncoding RNA OR3A4 promotes metastasis and tumorigenicity in gastric cancer. *Oncotarget* 2016; 7(21):30276-94.
39. Gallo LH, Nelson KN, Meyer AN, et al. Functions of Fibroblast Growth Factor Receptors in cancer defined by novel translocations and mutations. *Cytokine Growth Factor Rev* 2015; 26(4):425-49.

40. Kato H, Arao T, Matsumoto K, et al. Gene amplification of EGFR, HER2, FGFR2 and MET in esophageal squamous cell carcinoma. *Int J Oncol* 2013; 42(4):1151-8.
41. Paulson TG, Maley CC, Li X, et al. Chromosomal instability and copy number alterations in Barrett's esophagus and esophageal adenocarcinoma. *Clin Cancer Res* 2009; 15(10):3305-14.
42. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* 2015; 44(4):1429-1441.
43. Offman J, Fitzgerald RC. Alternatives to Traditional Per-Oral Endoscopy for Screening. *Gastrointest Endosc Clin N Am* 2017; 27(3):379-396.

Figure legends:

Figure 1: Volcano plot of probe-level methylation in progressors vs. non-progressors. Blue points = Bayes factor < 5; Red points = Bayes factor > 5. The plot shows a leftward shift of probes towards the left, suggesting global hypomethylation.

Figure 2: Boxplot of OR3A4 methylation differentiating high risk non-dysplastic BO

Figure 3: Change in methylation across temporally acquired samples for OR3A4. The y-axis shows percentage methylation at cg09890332 as observed by pyrosequencing. Timepoints on the x-axis refer to the sampling points, with 1 representing the initial baseline endoscopy and subsequent visits referred by increasing numbers (surveillance intervals are variable).

Figure 4: Light micrographs of representative examples of expression of OR3A4 via IHC of non-progressor (A=10X view, B=40X view) and progressor (C=10X view, D=40X) view

Table 1: Top20 array identified CpG sites with methylation variation between non-dysplastic samples which progress to OADC vs those which remain static

Key: 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-value = the raw p-value, not corrected for multiple testing, Adjusted p-value = the p-value corrected for multiple testing

Probe ID	t	P.Value	Adj.P.Value	Gene Name
cg09890332	-12.26	2.02E-08	0.0031	OR3A4
cg24007926	-11.05	6.84E-08	0.0032	NA
cg17337672	-10.95	7.54E-08	0.0032	FGFR2
cg02226469	-10.88	8.17E-08	0.0032	NA
cg17433294	-10.51	1.22E-07	0.0038	NMUR2
cg18479711	-10.31	1.52E-07	0.0039	HDAC4
cg09011162	-10.15	1.82E-07	0.0040	LMF1
cg19733463	-9.83	2.61E-07	0.0045	NMUR1
cg16150571	-9.64	3.26E-07	0.0045	SNORD116-22
cg24424217	-9.62	3.35E-07	0.0045	ZNF511
cg13164993	-9.57	3.52E-07	0.0045	RBP3
cg14019464	-9.53	3.70E-07	0.0045	TRIB3
cg24581378	-9.53	3.71E-07	0.0045	ZAP70
cg05230642	-9.36	4.54E-07	0.0051	SNORD115-14
cg12297814	-9.26	5.08E-07	0.0052	IGFN1
cg11231240	-9.23	5.33E-07	0.0052	NA
cg11864327	-8.96	7.40E-07	0.0063	ZFP2

cg16771467	-8.92	7.78E-07	0.0063	ATP8B1
cg17304276	-8.88	8.14E-07	0.0063	CUX2
cg11443888	-8.86	8.36E-07	0.0063	TMEM151B

Table 2: Table of differentially methylated regions (DMR)

DMR ID	Probe ID	Probe level adjusted p-value	Chromosome	Gene	Start of DMR (bp)	End of DMR (bp)	Size in bp	Change in methylation	P-value for DMR
1	cg21273584	0.014	2	NA	502999	503195	197	-34%	7.69E-04
1	cg00854591	0.045	2	NA	502999	503195	197	-16%	7.69E-04
1	cg11573608	0.014	2	NA	502999	503195	197	-27%	7.69E-04
2	cg25568703	0.047	5	NA	8216903	8217655	753	-25%	1.27E-03
2	cg25016964	0.039	5	NA	8216903	8217655	753	-22%	1.27E-03
2	cg17642708	0.021	5	NA	8216903	8217655	753	-33%	1.27E-03
3	cg10788901	0.172	10	FGFR2	123352704	123355661	2958	-27%	4.79E-03
3	cg14856220	0.044	10	FGFR2	123352704	123355661	2958	-24%	4.79E-03
3	cg17337672	0.003	10	FGFR2	123352704	123355661	2958	-47%	4.79E-03
3	cg02412684	0.031	10	FGFR2	123352704	123355661	2958	-38%	4.79E-03
3	cg06791446	0.058	10	FGFR2	123352704	123355661	2958	-13%	4.79E-03

3	cg22633036	0.581	10	FGFR2	123352704	123355661	2958	-4%	4.79E-03