UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Advanced endoscopic imaging of esophageal neoplasia; old looks and new visions

Boerwinkel, D.F.

Publication date
2014

Citation for published version (APA):
Boerwinkel, D. F. (2014). *Advanced endoscopic imaging of esophageal neoplasia; old looks and new visions*. [Thesis, fully internal, Universiteit van Amsterdam].

8

THE COMBINATION OF AUTOFLUORESCENCE ENDOSCOPY AND MOLECULAR BIOMARKERS IS A NOVEL DIAGNOSTIC TOOL FOR DYSPLASIA IN BARRETT'S OESOPHAGUS

Massimiliano di Pietro; David F Boerwinkel;
Mohammed Kareem Shariff; Xinxue Liu;
Emmanouil Telakis; Pierre Lao-Sirieix; Elaine Walker;
George Couch; Leanne Mills; Susan Slininger;
Maria O'Donovan; Mike Visser; Sybren Meijer;
Philip V Kaye; Lorenz Wernisch; Krish Ragunath;
Jacques JGHM Bergman; Rebecca C Fitzgerald

ABSTRACT

OBJECTIVE: Endoscopic surveillance for Barrett's oesophagus (BO) has major limitations including multiple biopsies, sampling error and subjectivity of pathological diagnosis of dysplasia. We aimed to compare a biomarker panel on minimal biopsies directed by autofluorescence imaging (AFI) with the standard surveillance protocol for the diagnosis of prevalent dysplasia.

DESIGN: We performed a cross-sectional prospective study on 175 patients with BO in three tertiary referral centers. All patients underwent high resolution endoscopy followed by AFI and 157 completed the biopsy protocol. Aneuploidy/tetraploidy; 9p and 17p loss of heterozygosity; *RUNX3*, *HPP1* and *p16* methylation; p53 and cyclin A immunohistochemistry (IHC) were assessed on biopsies targeted by AFI. Bootstrap resampling was used to select the best diagnostic biomarker panel for high grade dysplasia (HGD) and early cancer (EC). This panel was validated in an independent cohort of 39 patients.

RESULTS: Aneuploidy, p53 IHC and cyclin A had the strongest association with dysplasia in the per-biopsy analysis and, as a panel, had an area under the receiver operating characteristic curve of 0.97 (95%CI 0.95-0.99) for the diagnosis of HGD/EC. The diagnostic accuracy for HGD/EC of the 3-biomarker panel form AFI+ areas was superior to AFI- areas ($p<0.001$). Furthermore, compared to the standard protocol this 3-biomarker panel had equal sensitivity for HGD/EC, with a 4.5-fold reduction in the number of biopsies. In an independent cohort of patients, the panel had sensitivity and specificity for HGD/EC of 100% and 90% respectively.

CONCLUSIONS: A 3-biomarker panel on a small number of AFI-targeted biopsies provides an accurate and objective diagnosis of dysplasia in BO.

INTRODUCTION

The incidence of oesophageal adenocarcinoma (OAC) has dramatically increased over the last 40 years in the Western world, while survival rates remain dismal. (1-3) Barrett's oesophagus (BO) is the only recognized precursor to OAC and progresses to invasive cancer through low grade dysplasia (LGD), high grade dysplasia (HGD) and early intramucosal cancer (EC). (4) Despite the contradictory evidence on the benefit of endoscopic surveillance, (5, 6) all specialist societies recommend surveillance with protocol entailing four-quadrant random biopsies every two centimeters plus targeted biopsies of visible lesions (Seattle protocol). (7, 8) Diagnosis of HGD or EC leads to treatment recommendations generally involving endoscopic therapy to prevent development of advanced disease. (9) However, endoscopic surveillance has major limitations. The Seattle protocol is time consuming and invasive with subsequent poor adherence by endoscopists. (10, 11) In addition, dysplasia and EC can be inconspicuous leading to sampling error by random biopsies. (12) Hence, clinical justification and cost-effectiveness of endoscopic surveillance has been questioned, especially following recent evidence that the cancer risk in BO may be lower than previously thought. (13, 14)

To improve recognition of inconspicuous dysplasia, autofluorescence imaging (AFI) has been investigated. (15-17) AFI utilizes short wave-lengths of light to excite certain endogenous molecules (fluorophores) in the (GI) mucosa to emit a fluorescent signal. Dysplasia and EC can be associated with loss of autofluorescence, due to distortion of tissue architecture and molecular changes. (17) Two previous cross-over studies have shown that, although AFI can improve detection of inconspicuous dysplasia, it suffers from a high false positive rate (>60%) and its accuracy was not sufficient to replace the Seattle protocol. (15, 16) The nature of AFI false positivity is still unclear. We hypothesized that the AFI positivity may correlate with molecular abnormalities of the glandular tissue, not yet manifest as histological dysplasia.

An additional limitation of endoscopic surveillance is the significant inter-observer variability in the histopathological diagnosis of dysplasia, even among expert pathologists. (18, 19) To improve the objectivity of disease stage assessment, researchers have investigated molecular biomarkers. Some of these, such as over-expression of p53 (20-22) and cyclin A, (23) methylation of specific genes, (24) loss of heterozygosity (LOH) at the 17p and 9p loci and DNA ploidy abnormalities (25) associate with dysplasia. However, with the exception of p53 expression, these biomarkers have only been tested by single groups in single, retrospective cohorts of patients and this has hampered their translation into clinical practice. In addition, there is a lack of prospective studies testing multiple biomarkers in the same prospective cohort in order to identify the smallest, clinically applicable panel with the highest diagnostic accuracy. A further problem is that biomarkers performed on random biopsies are potentially subject to sampling error, due to the non-uniform distribution of molecular changes in BO. (25, 26) This limitation could be overcome using advanced imaging modalities to help target biopsies for biomarkers. We assessed the feasibility of this novel, combined approach in a recent retrospective pilot study. (27)

We thus hypothesised that combining AFI with molecular biomarkers could provide a synergistic approach to elucidating the overall dysplasia status of the patient and overcome limitations of conventional histology performed on random biopsies. Therefore, the primary aim of this

prospective study was to compare the accuracy of a panel of molecular biomarkers on AFI-directed biopsies with the Seattle protocol for the diagnosis of HGD/EC in BO. The secondary aims were: (*i*) assessment of diagnostic accuracy for the biomarkers for any grade of dysplasia; (*ii*) validation of a large panel of biomarkers in an independent prospective study by an independent laboratory.

METHODS

Patients and setting

The study was approved by the Cambridgeshire 2 Research Ethics Committee (09/H0308/118). For the generation of the biomarker panel, a training cohort of 175 patients was recruited prospectively between April 2009 and October 2011 from three centers (Amsterdam, Cambridge, and Nottingham). For the validation of the panel an independent cohort of 39 patients was prospectively recruited at a single institution (Cambridge) between March 2012 and April 2013. Inclusion criteria were: age >18 years, known BE with minimum length of C≥2, M≥2 or C<2, M≥4 according to the Prague classification,(28) referral for evaluation of dysplastic BE or follow-up post endoscopic resection for HGD/EC. Short segments of <2cm were excluded due to the excess of AFI-false positivity at the gastro-oesophageal junction (GOJ) and since sampling error is not such an issue for these cases. (29) Exclusion criteria were: oesophagitis (Los Angeles grade ≥B); previous upper GI surgery (with exception of Nissen fundoplication) or known upper-GI tract abnormality (e.g. pharyngeal pouch); coagulopathy or anticoagulant/antiplatelet therapy for high risk conditions; active or severe cardiopulmonary disease or liver disease; dysphagia; special communication needs.

Endoscopic procedure

After written informed consent, patients were endoscoped with FQ260Z endoscopes (Olympus inc, Tokyo, Japan) as previously described. (15) Four endoscopists performed the study procedures, but in order to uniform the interpretation of the AFI signal, the two endoscopists with less experience in AFI had to perform at least 30 AFI procedures prior to the study under supervision of endoscopists with extensive prior experience in AFI (JJGHMB and KR). The oesophagus was inspected first with a white light high resolution endoscope (HRE) to record the presence of any macroscopically visible lesions (Supplementary Figure 1A). Then, using the AFI mode, AFI positive (AFI+) areas (violet-purple in color) were carefully mapped. A representative AFI negative (AFI-) area (green color) was then selected as a negative control (Supplementary Figure 1B). In patients who had diffuse patchy AFI positivity throughout the BO (n=8) an AFI- control area could not be selected, but they were included in the per-biopsy (Figure 2) and final per-patient analysis (figure 4).

Biopsy and histology

Each AFI+ area and one AFI- area were biopsied for biomarker analysis and histopathology (Supplementary Figure 1A). All AFI+ areas within the tubular oesophagus, up to a maximum of four, were included in the research biopsy protocol, however small AFI+ areas (<1cm) within 1

cm from the gastro-oesophageal junction were excluded, due to the well known false positivity in close proximity to the gastric folds (30). Where possible (depending on the size of the AFI+ area), a maximum of three biopsies were taken in the following order of priority: one biopsy in formalin, one snap frozen biopsy in 10% dimethysulfoxide (DMSO), one biopsy snap frozen dry. For AFI+ positive areas larger than 1 cm two or more biopsies were taken for histology to minimize sampling error, whereas AFI+ with a complex shape were considered as multiple AFI+ areas (Supplementary Figure 1B). Histopathological assessment of each AFI-targeted area relied on biopsies stored in formalin. Random biopsies were then taken according to the Seattle protocol. The histology was assessed by an expert GI pathologist at the respective participating center according to the Vienna classification. (31) All dysplastic cases, including indefinite for dysplasia (ID), were then reviewed by a second pathologist, with random pairing of the initial pathologist with one of the pathologists from the other two centres. Four pathologists took part to initial assessment, but only three participated to the reviewing process. In case of disagreement, consensus among three pathologists was reached through a review process, and this diagnosis was considered as the final histological outcome for the per-biopsy and per-patient analysis. Cases with ID without definite dysplasia after consensus review were regarded as non-dysplastic (NDBO). In accordance with the Vienna classification cases of HGD, carcinoma in situ and intramucosal adenocarcinoma were grouped together (HGD/EC), as they represent a common endpoint for endoscopic therapeutic intervention as per clinical guidelines (7) and expert consensus statements. (32) For the purpose of the biomarker analysis, all biopsies taken using the AFI mode were regarded as AFI-targeted (AFI+ and AFI-). With regards to the overall per-patient histopathological diagnosis, we considered diagnoses made using: a) the current clinical standard (Seattle protocol, *i.e.*, biopsies on areas visible on HRE + quadrantic random biopsies); b) the overall histology (a + histology on AFI-targeted biopsies).

Biomarker analyses

All molecular analyses were done at the Medical Research Council, Cancer Cell Unit (Cambridge, UK). The panel of molecular biomarkers analyzed in the training cohort (9-biomarker panel) comprised: aneuploidy, G2/tetraploidy, LOH at 9p and 17p loci, hypermethylation of *p16*, *RUNX3*, *HPP1* and immunohistochemistry for p53 and cyclin A. Only the 3 biomarkers included in the final panel (aneuploidy and IHC for p53 and cyclin A) were performed in the validation cohort.

Flow cytometry

The snap frozen biopsies in DMSO were used for analysis of DNA content abnormalities and processed as described previously. (33) The isolated nuclei were analyzed by flow cytometry using a MoFlow (Beckman Coulter, Miami, FL, USA) or BD Influx™ (Becton, Dickenson biosciences, New Jersey, USA) by an expert cytometrist. The cell cycle histogram was analyzed using ModFit LT (Verity Software House, Topsham, ME, USA) by two investigators, who were blind to the histological outcome, and discordant results were reviewed jointly to reach agreement.

Quantitative methylation-specific PCR (Methylight)

DNA was extracted from snap frozen esophageal biopsies using the DNeasy® Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA was bisulfite modified using an EpiTect® Bisulfite Kit (Qiagen) and the degree of methylation was analyzed by Methylight using a LightCycler® 480 (Roche diagnostics Ltd, Rotkreuz, Switzerland) with previously published Taqman primers and probes. (24) Normalized methylation values were calculated as previously described. (24)

Immunohistochemistry (IHC)

Staining was performed with the BOND™ System (Leica Microsystems, Ltd, Milton Keynes, UK) with an H1 antigen retrieval program. Antibodies used were: p53 clone DO7 (Dakocytomation, 1:50) and cyclin A (Novocastra, 1:40). The scoring was performed by a single author without prior knowledge of the clinical diagnosis. Surface cyclin A was scored as previously described, using a cut-off for positivity of 1% of positive surface cells. (23) P53 expression was scored positive when there were areas of strong staining or complete loss of staining, compared to the background levels, as previously described. (34)

Loss of heterozygosity (LOH)

LOH at 17p and 9p loci (*p53* and *p16* genes, respectively) was assessed on DNA extracted by phenol/chloroform. Microsatellite markers were used as previously published. (35) The microsatellite loci were amplified by multiplex PCR and the PCR products were combined with GeneScan™ LIZ® Size Standard and electrophoresed on an automated sequencing system (ABI 3130 xl Genetic Analyzer, Applied Biosystems, California, USA) according to the manufacturer's protocol. GeneMapper® Software (Applied Biosystems) was used to analyse the peak height ratios of the alleles of a duodenal control and oesophageal biopsy and highlight LOH candidates. LOH was assessed on the basis of the following allelic imbalance ratios:

$$\text{Allelic ratio of sample} = \frac{\text{Peak height of Allele 1}}{\text{Peak height of Allele 2}}$$

$$\text{Allelic Imbalance} = \frac{\text{Allele Ratio of Control Sample}}{\text{Allele Ratio of AFI Sample}}$$

Ratio of > 1.35 or < 0.067 was considered as suggestive of LOH as instructed by the software. The sample was considered positive for LOH if ≥2 loci were suggestive for LOH for a particular gene.

Sample size and power

Previous data suggest that single biomarkers have a sensitivity for HGD/EC varying between 60 and 87%, however we assumed that, when combined into a panel, biomarkers could have 90%-95% sensitivity and 85-95% specificity. For a sensitivity of 90%, we calculated that to have an accuracy of 10% (95% CI: ±10%) we needed to recruit at least 35 patients with HGD/EC.

Statistics

For the purpose of the per-biopsy analysis, each AFI-targeted area was classified with: *i*. patient research number; *ii*. AFI status (AFI+ or AFI-); *iii*. binary outcome of the 9 biomarkers (positive or negative); *iv*. histological diagnosis (NDBO, LGD, HGD, EC). Chi-square tests were used to compare differences between groups. A p-value <0.05 was considered statistically significant. We used κ -statistics to quantify interobserver agreement. For the reliability of pathological diagnosis, agreement among three expert pathologists for a diagnosis of any grade of dysplasia was calculated and results were expressed as weighted κ-value. For the reliability of endoscopic location and level of subjective suspicion for dysplasia of AFI targeted areas, the agreement was calculated and expressed as κ-value. In particular for the location of the AFI positive areas, agreement among endoscopists was defined as their specification of the epicenter of the AFI positive area within 30 degrees.

In some AFI-targeted areas some biomarkers were recorded as missing, due to the fact that some AFI+ areas were not large enough to permit the assessment of the full 9-biomaker panel, which relied on 3 biopsies. To ensure enough power and avoid the possibility that missing information biased the outcome of the analysis, missing data were imputed by multiple imputation (MI), as this method controls the influence of both randomly missing and non-randomly missing data. (36, 37) The imputation model included the 9 biomarkers and the histological diagnosis as variables. Five independent datasets were imputed and they were included in the analysis together with the original database, after removing samples containing missing values in variables of interest.

To avoid inflating the diagnostic performance of biomarkers by applying the model to the same dataset from which it was developed (over-fitting), and prevent random selection of biomarker panels, bootstrap resampling was applied. In this analysis multiple rounds of a random selection of sub-samples were performed and the calculated accuracy of the parameter of interest validated in the remaining sub-sample. Details of the statistical methodology used for the generation and the validation of the biomarker panel are described in Supplementary material All the analyses were performed using SPSS version 19.0 (SPSS Inc, Chicago, IL, USA) and R version 2.15.2.

RESULTS

Of the 175 participants recruited as part of the training cohort, 18 were excluded because of either a lack of biopsies for laboratory tests (n=4) or breach of biopsy protocol (e.g. biopsy mislabeling, no research biopsy in AFI+ areas taken, no random biopsies taken) (n=14). From the remaining 157 participants, we obtained 373 AFI-targeted biopsies (AFI+: AFI- = 230:143), which were processed for biomarkers and included in the per-biopsy analysis. The validation cohort consisted of 39 patients, from which 130 AFI-targeted biopsies were taken (AFI+:AFI- = 89:41). There were no significant differences between the two cohorts in the demographics and histologic outcomes as well as characteristics of AFI+ areas, however the validation cohort had a slightly higher number of AFI+ areas per patient (Table 1). The overall interobserver agreement among three expert pathologists for a diagnosis of any grade of dysplasia was

Table 1. Demographics, histological stage and endoscopic characteristics of AFI+ areas for patients included in training and validation cohorts.

Variables	Training cohort	Validation cohort	p
Number of patients	157	39	N/A
Male:female (%)	79:21	92:8	0.06
Mean age (range)	66.4 (35-87)	68.7 (35-84)	0.23
Mean length of BO in cm (range)	7.3 (2-17)	7.6 (3-18)	0.85
Histological diagnosis			
NDBO	99 (63%)	22 (56.4%)	0.24
LGD	21 (13.4%)	8 (20.5%)	
HGD	24 (15.3%)	3 (7.7%)	
EC	13 (8.3%)	6 (15.4%)	
Endoscopic features			
Number of AFI+ areas	229	89	N/A
AFI+ areas visible on HRE	28.4%	23.6%	0.16
AFI+ areas with HGD/EC	21.1%	15.7%	0.34

'moderate' (weighted κ-value 0.56, standard error 0.055). This is in keeping with previous studies comparing dysplasia rating between pathologists, underscoring the subjectivity of the diagnosis. With reference to the interpretation of the AFI status of selected areas for targeted biopsies, the agreement between observers was 'good' for the location of the area of interest (κ-value 0.68, standard error 0.104) and 'moderate' for the level of suspicion for dysplasia (κ-value 0.59, standard error 0.104).

Association of biomarkers with dysplasia (per-biopsy analysis)

We first sought to identify a small biomarker panel with the best diagnostic accuracy for prevalent dysplasia. To this end, we identified nine biomarkers that had the most robust correlation with dysplasia in BO based on published phase III and IV clinical studies. (20-25) Figure 1 shows the strategy used for the generation of the panel. The 9 biomarkers were tested in the AFI-targeted biopsies from the training cohort to assess their association with dysplasia on a per-biopsy basis. As shown in Table 2, all of the biomarkers associated with the presence of confirmed dysplasia (LGD, HGD or EC), with the exception of 9p LOH (p16). When restricting the analysis to the association between biomarkers and HGD/EC, we found similar results, in that only tetraploidy and 9p LOH lacked statistical significance.

Since large numbers of biomarkers are difficult to apply in the clinical setting, we used a strict statistical methodology to identify a small and clinically feasible biomarker panel. Since some AFI+ areas were not sufficiently large to allow three biopsies as per protocol, the dataset had 20% of missing values, which ranged between 13.4% and 28.5% for different biomarkers (Table 2). We used bootstrap resampling of the original database as well as on 5 imputed databases to
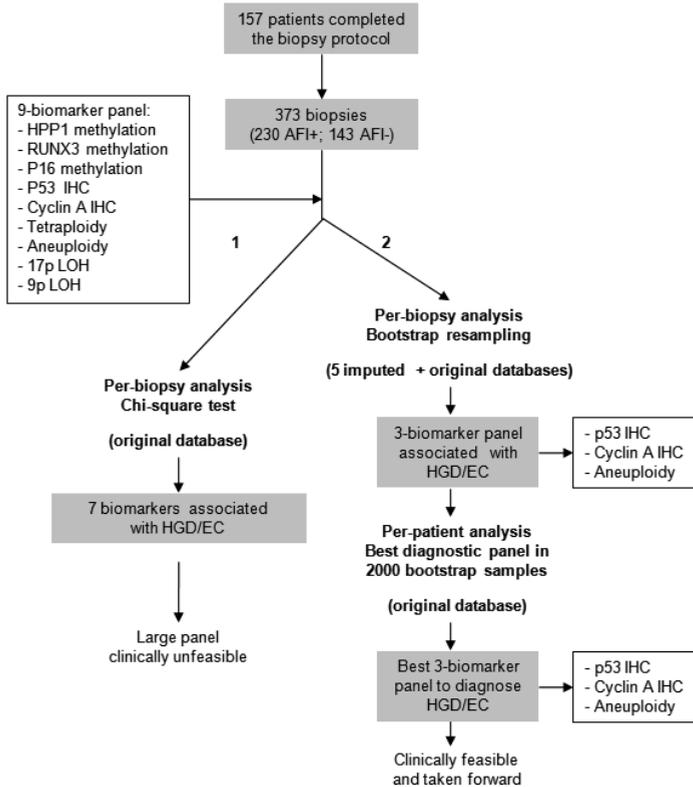
Figure 1. Strategy for the generation of the biomarker panel. Chi-square test showed that form the initial panel of 9 biomarkers seven significantly associated with a diagnosis of HGD/EC in the corresponding biopsy (arm 1 – left side). To identify a small biomarker panel bootstrap resampling on 5 imputed and 1 original database was applied (arm 2 – right side). Both per-biopsy and per-patient analyses identified the same 3-biomaker panel as best diagnostic panel for HGD-EC.

account for missing values. This statistical methodology has been previously developed to handle datasets with up to 50% of missing data in clinical studies. (38) We concentrated on the highest histological outcome for this analysis (HGD/EC), as this is what currently triggers therapeutic decisions according to clinical guidelines. We found that two IHC markers (p53 and cyclin A) and aneuploidy had the highest rate of inclusion in best bootstrap models in all the 5 imputed databases, as well as in the original database (Figure 2A) and for this reason they were selected to form a 3-biomaker panel. On a per-biopsy analysis this panel had AUCs of 0.93 (95% CI: 0.89-0.98) and 0.97 (95%CI: 0.95-0.99) for a diagnosis of any grade of dysplasia and HGD/EC, respectively (Figure 2B-C). These data showed that these 3 biomarkers have the strongest association with dysplasia and could be combined into a panel to aid clinical diagnosis. Association between molecular biomarkers and AFI statusIn order to translate the diagnostic accuracy for dysplasia of the biomarker panel from a single biopsy to a per-patient level, we determined that the biopsies should be targeted using an imaging tool rather than relying on

Table 2. Association of biomarkers with dysplasia in the per-biopsy analysis.

Biomarker	Missing values (%)	Biomarker outcome	HGD/EC			Any dysplasia		
			No	Yes	P	No	Yes	P
HPP1 methylation	19.1	Negative	59 (22.9%)	3 (7.1%)	0.02	56 (24.0%)	6 (9.0%)	<0.01
		Positive	199 (77.1%)	39 (92.9%)		177 (76.0%)	61 (91.0%)	
RUNX3 methylation	19.1	Negative	102 (39.5%)	6 (14.3%)	<0.01	97 (41.6%)	11 (16.4%)	<0.01
		Positive	156 (60.5%)	36 (85.7%)		136 (58.4%)	56 (83.6%)	
P16 methylation	19.1	Negative	143 (55.4%)	12 (28.6%)	<0.01	132 (56.7%)	23 (34.3%)	<0.01
		Positive	115 (44.6%)	30 (71.4%)		101 (43.3%)	44 (65.7%)	
P53 IHC	13.2	Negative	196 (70.5%)	3 (6.8%)	<0.01	189 (76.8%)	10 (13.2%)	<0.01
		Positive	82 (29.5%)	41 (93.2%)		57 (23.2%)	66 (86.8%)	
Cyclin A IHC	14.0	Negative	238 (85.0%)	8 (20.5%)	<0.01	223 (88.8%)	23 (33.8%)	<0.01
		Positive	42 (15.0%)	31 (79.5%)		28 (11.2%)	45 (66.2%)	
Tetraploidy	19.7	Negative	182 (69.2%)	20 (57.1%)	0.15	171 (72.2%)	31 (50.8%)	<0.01
		Positive	81 (30.8%)	15 (42.9%)		66 (27.8%)	30 (48.2%)	
Aneuploidy	19.7	Negative	232 (88.2%)	10 (28.6%)	<0.01	216 (91.1%)	26 (42.6%)	<0.01
		Positive	31 (11.8%)	25 (71.4%)		21 (8.9%)	35 (57.4%)	
17p LOH	27.4	Negative	112 (49.3%)	7 (16.7%)	<0.01	107 (51.9%)	12 (19.0%)	<0.01
		Positive	115 (50.7%)	35 (83.3%)		99 (48.1%)	51 (81.0%)	
9p LOH	28.5	Negative	43 (19.2%)	5 (12.2%)	0.29	40 (19.5%)	8 (13.3%)	0.27
		Positive	181 (80.8%)	36 (87.8%)		165 (80.5%)	52 (86.7%)	

a random biopsy. As discussed in the introduction we hypothesized that the AFI positive areas may be enriched for molecular abnormalities due to an abnormal genetic field effect. Hence, we then asked whether AFI positivity correlates with molecular abnormalities of the mucosa. We first analyzed the association between individual biomarker outcome and AFI status of the corresponding endoscopic area. Aneuploidy, 17p LOH, p53 IHC and cyclin A significantly associated with AFI positivity (p<0.05) (Supplementary Table 1). Since dysplasia is enriched in AFI positive areas and could represent a confounding factor, we looked at this association after exclusion of areas harboring dysplasia and found that aneuploidy and p53 IHC retained a significant association with AFI positivity (Supplementary Table 1). To confirm that this association translates into a clinical advantage, we looked at whether the 3-biomarker panel (aneuploidy, p53 IHC and cyclin A) had a better accuracy for an overall diagnosis of HGD/EC if assessed on AFI+ areas compared with AFI- areas. We performed this comparative analysis in 5 imputed databases as well as in the original database (Figure 3). To avoid over-fitting, we used the bootstrap samples as a training set and patients not selected in bootstrap sampling as a validation set. Notably, average AUCs in all 6 databases were significantly higher for the 3-biomarker panel assessed on AFI+ areas compared to AFI- areas. We therefore concluded that AFI positivity is associated with an enrichment of biomarkers and is a suitable tool to guide biopsies for analysis of biomarkers.
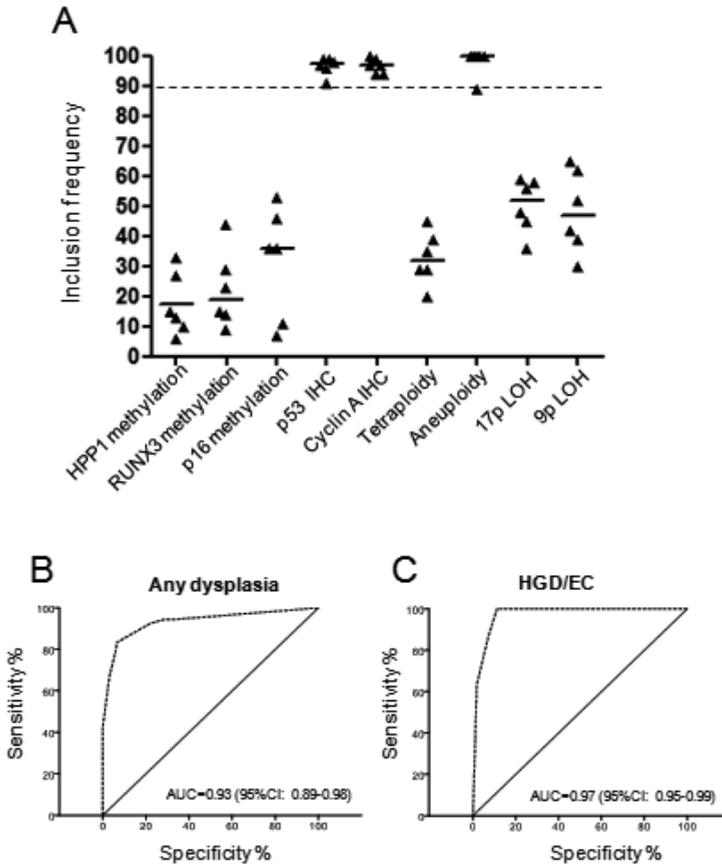
Figure 2. A 3-biomaker panel including p53 IHC, cyclin A IHC and aneuploidy has high diagnostic accuracy for dysplasia. A. Inclusion frequencies of the nine biomarkers in 100 bootstrap samples for each MI database (n=5) and the original database (n=1). A stringent cut-off of 90 for the median over all 6 databases was used to select the best biomarkers. P53 IHC, cyclin A IHC and aneuploidy had median inclusion frequency above the threshold. B-C Area under the curve (AUC) for the diagnosis of any grade of dysplasia (B) and HGD/EC only (C) was calculated using the panel of biomarker selected in A.

Internal validation of the 3-biomaker panel (per-patient analysis)

We sought to confirm that the biomarker panel selected in the per-biopsy analysis by the logistic regression model could correctly classify patients based on their highest grade of dysplasia assessed histologically. From a pool of 9 biomarkers, there are 84 different combinations of 3 biomarkers and 36 combinations of 2 biomarkers. We calculated the diagnostic accuracy of these 120 biomarker panels with different cut-offs of positive biomarkers in the 2000 bootstrap samples generated from the original database. Strikingly, the best diagnostic accuracy was achieved by the same panel identified in the per-biopsy analysis (p53 IHC, cyclin A IHC and aneuploidy), with a cut-off of 2 positive biomarkers to diagnose HGD/EC (Supplementary Table 2). We therefore took this combination of biomarkers forward as the final diagnostic panel to determine
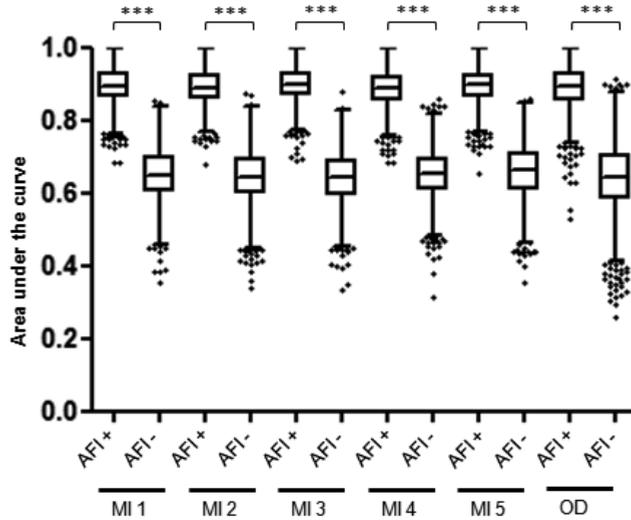
Figure 3. Diagnostic accuracy for HGD/EC of the 3-biomarker panel assessed on AFI+ areas and AFI- areas. This analysis was performed on the 114 patients with biopsies available on both AFI+ and AFI- areas, after exclusion of patients without AFI positivity (n=35) and those with diffuse AFI positivity (n=8). The AUC for a diagnosis of overall HGD/EC was calculated using the 3-biomaker panel from AFI+ areas and AFI- areas in 2000 bootstrap samples from 5 imputed databases (MI) and the original database (OD). In this plot, each box represents the median AUC with first and third quartiles for the bootstrap samples of each group and the whiskers include data within 1.5 interquartile range of the upper and lower quartile. Outliers are depicted separately*** indicates p value <0.001.

histological outcome in the patients from the training cohort. Hence, we included patients with the 3-biomarker full dataset on AFI-targeted biopsies (n=112) (Figure 1B). In keeping with the results from bootstrap resampling, a cut-off of 2 biomarkers positive had the best accuracy (Supplementary Table 3). Using this cut-off to identify patients with high histological grade, only one patient with HGD was mis-classified with a low-risk biomarker signature (Figure 4 and Table 3). The biomarker panel had a sensitivity and a specificity of 95.8% (95%CI: 76.9%-99.8%) and 88.6% (95%CI: 79.7%-94.1%), respectively, for a diagnosis of HGD/EC, and 74.4% (95%CI: 57.6%-86.4%) and 94.5% (95%CI: 85.8%-98.2%), respectively, for a diagnosis of any grade of dysplasia. This translated into a similar diagnostic accuracy to the Seattle protocol, which also missed one patient with HGD/EC (Table 3). By comparison the Seattle protocol had a sensitivity of 95.8% (95%CI: 76.9%-99.8%) for a diagnosis of HGD/EC (p=1.0 when compared with 3-biomarker panel) and 84.6% (95%CI: 68.8%-93.6%) for a diagnosis of any grade of dysplasia (p=0.26). Relying on the histology of AFI+ areas, the sensitivity for HGD/EC was lower at 91.7%, although this difference was not statistically significant. Importantly, using this novel approach 2.8 biopsies per patient were taken on average compared with 12.8 for the standard biopsy protocol (p<0.001) (Table 3).

External validation of the 3-biomaker panel

We then tested the 3 biomarkers in an independent group of 39 patients. The panel had a sensitivity and a specificity of 100% and 90% (95%CI: 79.3%-100.0%), respectively, for a diagnosis

Table 3. Comparison among Seattle protocol, AFI-targeted histology and 3-biomarker panel on AFI+ areas. Seattle protocol includes all biopsies taken on HRE white light endoscopy (random + targeted on macroscopically visible abnormal areas). AFI+ histology includes biopsies taken on AFI+ areas, regardless of their appearance on white light. AFI + biomarkers column includes only the biopsies from AFI+ areas processed for the 3-biomaker panel.

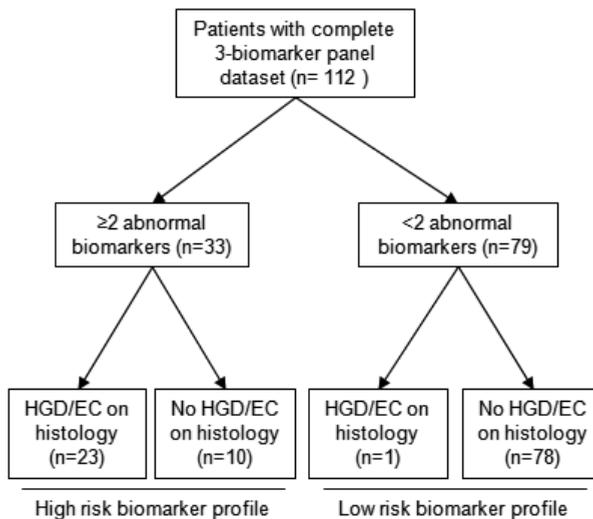|	Seattle protocol + histology	AFI + histology	AFI + Biomarkers	p value
No. of HGD/EC missed	1	2	1	N/A
Sensitivity for HGD/EC	95.8%	91.7%	95.8%	ns
Total no. of biopsies	1385	169	310	N/A
No of biopsies per patient	12.4	1.5	2.8	<0.001
No of biopsies for every HGD/EC case diagnosed	60.2	7.7	13.5	N/A

Figure 4. Flow-chart of biomarker outcome in patient from the training cohort with full 3-biomarker dataset. Using a cut-off of 2 abnormal biomarkers to diagnose patients with prevalent HGD/EC, the 3-biomaker panel only missed one patient with HGD/EC. Ten patients with NDBO or LGD were classified as high risk for HGD/EC. The sensitivity and specificity of 3-biomaker panel for a diagnosis of HGD/EC were 95.8% (95%CI: 76.9%-99.8%) and 88.6% (95%CI: 79.7%-94.1%), respectively.

of HGD/EC, and 70.6% (95%CI: 48.9%-92.3%) and 100%, respectively, for diagnosis of any grade of dysplasia (Supplementary Table 4). The sensitivity of the panel for HGD/EC was not significantly different from the Seattle protocol, which however missed one case of HGD. This current clinical standard protocol had a sensitivity of 88.9% for a diagnosis of HGD/EC (p=1.0 when compared with 3-biomarker panel) and 94.1% for a diagnosis of any grade of dysplasia (p=0.17), but relied on a significantly higher number of biopsies (p<0.001). We conclude that this 3-biomarker panel assessed on AFI-targeted biopsies has a similar sensitivity for HGD/EC compared to the Seattle protocol and allows a significant reduction in the number of biopsies required.

DISCUSSION

In the present study, we devised a novel strategy in which advanced endoscopic imaging can be used to direct biopsy sampling for biomarkers, which can provide a more objective read-out than dysplasia. Single biomarkers or panels of biomarkers have been shown to correlate with dysplasia in BO (21, 23, 24, 35) and in some cases to estimate the individual risk of progression. (23, 24, 35, 39, 40) This prospective multi-centre phase-4 study evaluated nine biomarkers with the most robust published data, and confirmed that eight of them were significantly associated with any grade of dysplasia and seven with HGD/EC. To our knowledge this is the first study that has validated such a large panel of biomarkers in a single prospective patient cohort. Furthermore, it is the first time that the previously published biomarkers have been validated in an independent laboratory. The rationale for using a biomarker panel stems from evidence that a combination of biomarkers performs better than single biomarkers on their own. (35) Clinical applicability of biomarkers relies on the feasibility of the specific laboratory assays, low enough costs and a manageable number of assays if combined into a panel. A stringent statistical methodology, which took into account missing data and assessed the strength of different combinations of biomarkers, indicated that p53 IHC, cyclin A IHC and aneuploidy constitute the panel with the best diagnostic accuracy for HGD/EC. This panel satisfies criteria for clinical applicability. P53 and cyclin A rely on IHC, which is a routine technique in pathology laboratories. Of these, p53 is currently already employed by some pathologists to aid in the diagnosis of dysplasia (20, 21) and has been recommended in the revised BSG guidelines. (41) For the evaluation of aneuploidy, we used a methodology based on flow cytometry analysis from frozen samples preserved in DMSO, as this is the method used in previously published results. (35) For future clinical applications image cytometry (IC) is an alternative technique applicable to paraffin sections and has been recently shown to be comparable to flow cytometry in BO. (42) Hence in the future, the 3-biomaker panel could be applied from a single diagnostic biopsy stored in formalin.

Since molecular events and mutations occurring in BO are not distributed uniformly within BO, (25, 26) we postulated that use of one or a few random biopsies to assess biomarkers could lead to sampling error. In recent years, a large effort has been invested in novel imaging modalities for the detection of inconspicuous neoplasia. AFI is promising, but its applicability in clinical practice has been limited by a high false positive rate. We hypothesized that AFI positivity might correlate with a field of molecular abnormality surrounding or preceding subtle histological changes. Indeed, we found that aneuploidy and p53 abnormalities significantly correlated with AFI positivity independently of dysplasia. DNA is known to be a weak fluorophore, therefore the loss of autofluorescence may be explained by the increased DNA content observed in aneuploidy. Similarly, *p53* is a key gene in the control of DNA amplification and proliferation of aneuploid cells. (43) Even though the other biomarker in the panel (cyclin A) was not associated with AFI positivity in non-dysplastic areas, cyclin A can be overexpressed as result of cell cycle deregulation leading to aneuploidy. (44) Inflammation has previously been shown to associate with AFI false positivity. (15) Since inflammatory pathways can drive genetic and epigenetic changes during carcinogenesis in BO, (45) it is possible that the AFI signal, inflammation and biomarkers are all closely related.

In this study, assessment of dysplasia in the AFI-targeted areas relied only on diagnostic biopsies stored in formalin, therefore there is a possibility of sampling error for very focal dysplasia. To reduce this possibility, large AFI+ areas had two or more diagnostic biopsies taken. Since the molecular changes normally precede development of dysplasia and expand over large areas of mucosa,(25) it is reasonable to consider the field of molecular abnormalities to be larger than the focus of dysplasia, and therefore less affected by sampling error.

Notably, the new diagnostic tool had the same sensitivity for dysplasia as the Seattle protocol with significantly fewer biopsies. This could be further reduced using image cytometry to assess aneuploidy on the same diagnostic biopsy used for immunohistochemical markers. Therefore, this novel risk stratification approach, which is based on more objectively measurable outcomes, has the potential to overcome several of the major limitations of BO surveillance endoscopy, including the sampling error and the subjectivity of a dysplasia diagnosis. In our study, four pathologists with special interest in BO from three centres were involved. Yet, the agreement for a diagnosis of any grade of dysplasia was moderate (κ-value 0.56). Hence, a binary outcome of a molecular test can prove advantageous in this regard. In addition, the small number of biopsies may lead to shorter endoscopic procedures, lower risk of complications and improved cost-effectiveness. From an economic perspective, we estimated that the costs required to assess the 3-biomaker panel (aneuploidy by IC + IHC for p53 and cyclin A) on biopsies targeted by AFI were similar to the costs of the Seattle protocol and these costs would come down with high throughput assay techniques (data not shown).

This is a cross-sectional study designed to assess the ability of biomarkers to diagnose prevalent dysplasia and therefore we did not set out to analyze follow up data and look at prediction of future cancer risk by biomarkers. However, it is notable that among the ten patients in the training cohort with a high risk biomarker profile ($\geq$2 abnormal biomarkers) and no evidence of HGD/EC at the time of the endoscopy (Figure 4), six of them (60%) had pathological progression within six months (four from LGD to HGD/EC, one from ID to EC and one from ID to LGD). Assuming this could have been prevalent disease missed by sampling error, this confirms that biomarkers overcome this limitation of the Seattle protocol.

This study has several strengths. It is the first study designed prospectively to assess a pre-defined group of multiple biomarkers in a large cohort of patients with BE. A single laboratory assessed a large panel of biomarkers that had previously been studied by independent groups in different patient cohorts. To our knowledge, this is the largest imaging study in BO and it has a sample size comparable to other biomarker studies. (24, 35) Finally the endoscopic protocol was carefully monitored to guarantee proper training and strict adherence.

However, this study has some limitations. Only approximately 70% of patients could enter the per-patient analysis in the original database (112 out of 157). This is due to the fact that at the start of the study we lacked information about which of the 9 biomarkers would outperform the others in the statistical analysis. Therefore, since not all the AFI+ areas were large enough to allow for three research biopsies, we evenly distributed different biomarker analyses among the available biopsies and then used a combination of MI and bootstrap resampling to deal with missing data. (38) However, all patients included in the validation cohort could be fully assessed

with the 3 biomarkers, suggesting that a smaller panel is clinically feasible. Furthermore, the patient cohorts were enriched for dysplasia, therefore they do not reflect the general BO population. However, enrichment for high risk patients was necessary to guarantee an adequate number of pathologic outcomes to correlate with biomarker positivity. Moreover, we cannot exclude that the improved diagnostic accuracy of biomarkers on AFI+ areas (Figure 3) is influenced by the higher number of AFI+ areas sampled compared to AFI- regions. In order to maximize the molecular information obtained, without significantly biasing the performance of the Seattle protocol and exceeding the overall number of research biopsies permitted by the ethical committee, we aimed to sample at most 4 AFI+ and one AFI- area per patient and. This notwithstanding, the ratio between AFI+ and AFI- areas was 1.6:1 in the training cohort and 2.1:1 in the validation cohort, hence we believe that this had a small impact on the final results. An additional limitation of this study is that AFI is not compatible with all the endoscopic technologies available across the world. We believe that this study has paved the way for a new diagnostic approach based on the integration of endoscopic and molecular data, and in the future other advanced imaging modalities could be used to target biopsies for biomarker assessment. The advantage of AFI is that, like chromoendoscopy, the assessment of a colour-coded images is relatively easy and our data on the reliability of AFI signal interpretation confirmed a good level of agreement among endoscopists. Finally, although the exclusion of patients with short segments makes this study mostly relevant to patients with long segment of BO, we believe that the latter group of patients is the one that particularly benefits from red-flag endoscopic techniques and biomarkers due to the higher rate of sampling error and number of biopsies compared to short segments of BO.

In conclusion, this study provides evidence that a 3-biomarker panel on AFI-targeted biopsies has equivalent diagnostic accuracy for dysplasia compared to the current gold standard with a significant reduction in the number of biopsies required.

ACKNOWLEDGMENTS

REFERENCES

1. Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. Ann Oncol 2012, Jul 30 [Epub ahead of print].

2. Cancer Research UK. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/oesophagus/?script=true . Sept 2012.

3. Eloubeidi MA, Mason AC, Desmond RA, et al. Temporal trends (1973-1997) in survival of patients with esophageal adenocarcinoma in the United States: a glimmer of hope? Am J Gastroenterol 2003;98:1627-33.

4. Reid BJ, Li X, Galipeau PC, et al. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat Rev Cancer 2010;10:87-101.

5. Cooper GS, Kou TD, Chak A. Receipt of previous diagnoses and endoscopy and outcome from esophageal adenocarcinoma: a population-based study with temporal trends. Am J Gastroenterol 2009;104:1356-62.

6. Corley DA, Mehtani K, Quesenberry C, et al. Impact of Endoscopic Surveillance on Mortality from Barrett's Esophagus-Associated Esophageal Adenocarcinomas. Gastroenterology 2013.

7. Spechler SJ, Sharma P, Souza RF, et al. American Gastroenterological Association technical review on the management of Barrett's esophagus. Gastroenterology 2011;140:e18-52.

8. Watson A, Heading RC, Shepherd NA, et al. Guidelines for the diagnosis and management of Barrett's columnar-lined oesophagus. http://wwwbsgorguk/images/stories/docs/clinical/guidelines/oesophageal/Barretts_Oespdf 2005.

9. Shaheen NJ, Sharma P, Overholt BF, et al. Radiofrequency ablation in Barrett's esophagus with dysplasia. N Engl J Med 2009;360:2277-88.

10. Abrams JA, Kapel RC, Lindberg GM, et al. Adherence to biopsy guidelines for Barrett's esophagus surveillance in the community setting in the United States. Clin Gastroenterol Hepatol 2009;7:736-42.

11. Peters FP, Curvers WL, Rosmolen WD, et al. Surveillance history of endoscopically treated patients with early Barrett's neoplasia: nonadherence to the Seattle biopsy protocol leads to sampling error. Dis Esophagus 2008;21:475-9.

12. Cameron AJ, Carpenter HA. Barrett's esophagus, high-grade dysplasia, and early adenocarcinoma: a pathological study. Am J Gastroenterol 1997;92:586-91.

13. Desai TK, Krishnan K, Samala N, et al. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. Gut 2012;61:970-6.

14. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. N Engl J Med 2011;365:1375-83.

15. Curvers WL, Herrero LA, Wallace MB, et al. Endoscopic tri-modal imaging is more effective than standard endoscopy in identifying early-stage neoplasia in Barrett's esophagus. Gastroenterology 2010;139:1106-14.

16. Curvers WL, van Vilsteren FG, Baak LC, et al. Endoscopic trimodal imaging versus standard video endoscopy for detection of early Barrett's neoplasia: a multicenter, randomized, crossover study in general practice. Gastrointest Endosc 2011;73:195-203.

17. Kara M, DaCosta RS, Wilson BC, et al. Autofluorescence-based detection of early neoplasia in patients with Barrett's esophagus. Dig Dis 2004;22:134-41.

18. Curvers WL, ten Kate FJ, Krishnadath KK, et al. Low-grade dysplasia in Barrett's esophagus: overdiagnosed and underestimated. Am J Gastroenterol 2010;105:1523-30.

19. Wani S, Falk GW, Post J, et al. Risk factors for progression of low-grade dysplasia in patients with Barrett's esophagus. Gastroenterology 2011;141:1179-86, 86 e1.

20. Jouret-Mourin A, Sempoux C, Duc KH, et al. Usefulness of histopathological markers in diagnosing Barrett's intraepithelial neoplasia (dysplasia). Acta Gastroenterol Belg 2009;72:425-32.

21. Kaye PV, Haider SA, Ilyas M, et al. Barrett's dysplasia and the Vienna classification: reproducibility, prediction of progression and impact of consensus reporting and p53 immunohistochemistry. Histopathology 2009;54:699-712.

22. Skacel M, Petras RE, Gramlich TL, et al. The diagnosis of low-grade dysplasia in Barrett's esophagus and its implications for disease progression. Am J Gastroenterol 2000;95:3383-7.

23. Lao-Sirieix P, Lovat L, Fitzgerald RC. Cyclin A immunocytology as a risk stratification tool for Barrett's esophagus surveillance. Clin Cancer Res 2007;13:659-65.

24. 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;24:4138-48.

25. Galipeau PC, Prevo LJ, Sanchez CA, et al. Clonal expansion and loss of heterozygosity at chromosomes 9p and 17p in premalignant esophageal (Barrett's) tissue. J Natl Cancer Inst 1999;91:2087-95.

26. Leedham SJ, Preston SL, McDonald SA, et al. Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. Gut 2008;57:1041-8.

27. Boerwinkel DF, Di Pietro M, Liu X, et al. Endoscopic TriModal imaging and biomarkers for neoplasia conjoined: a feasibility study in Barrett's esophagus. Dis Esophagus 2012;Oct 15, Epub ahead of print.

28. Sharma P, Dent J, Armstrong D, et al. The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. Gastroenterology 2006;131:1392-9.

29. Curvers WL, Singh R, Wallace MB, et al. Identification of predictive factors for early neoplasia in Barrett's esophagus after autofluorescence imaging: a stepwise multicenter structured assessment. Gastrointest Endosc 2009;70:9-17.

30. Kara MA, Peters FP, Ten Kate FJ, et al. Endoscopic video autofluorescence imaging may improve the detection of early neoplasia in patients with Barrett's esophagus. Gastrointest Endosc 2005;61:679-85.

31. Dixon MF. Gastrointestinal epithelial neoplasia: Vienna revisited. Gut 2002;51:130-1.

32. Bennett C, Vakil N, Bergman J, et al. Consensus Statements for Management of Barrett's Dysplasia and Early-Stage Esophageal Adenocarcinoma, Based on a Delphi Process. Gastroenterology 2012;143:336-46.

33. Reid BJ, Haggitt RC, Rubin CE, et al. Barrett's esophagus. Correlation between flow cytometry and histology in detection of patients at risk for adenocarcinoma. Gastroenterology 1987;93:1-11.

34. Depledge DP, Evans KJ, Ivens AC, et al. Comparative expression profiling of Leishmania: modulation in gene expression between species and in different host genetic backgrounds. PLoS Negl Trop Dis 2009;3:e476.

35. Galipeau PC, Li X, Blount PL, et al. NSAIDs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma. PLoS Med 2007;4:e67.

36. Su YS, Gelman A, Hill J, et al. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. Journal of Statistical Software 2011;10.

37. Graham JW, Hofer SM, Piccinin AM. Analysis with missing data in drug prevention research. NIDA Res Monogr 1994;142:13-63.

38. Heymans MW, van Buuren S, Knol DL, et al. Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC Med Res Methodol 2007;7:33.

39. Murray L, Sedo A, Scott M, et al. TP53 and progression from Barrett's metaplasia to oesophageal adenocarcinoma in a UK population cohort. Gut 2006;55:1390-7.

40. Kastelein F, Biermann K, Steyerberg EW, et al. Aberrant p53 protein expression is associated with an increased risk of neoplastic progression in patients with Barrett's oesophagus. Gut 2012;Dec 20, Epub ahead of print.

41. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. Gut 2013.

42. Dunn JM, Mackenzie GD, Oukrif D, et al. Image cytometry accurately detects DNA ploidy abnormalities and predicts late relapse to high-grade dysplasia and adenocarcinoma in Barrett's oesophagus following photodynamic therapy. Br J Cancer 2010;102:1608-17.

43. Fukasawa K, Choi T, Kuriyama R, et al. Abnormal centrosome amplification in the absence of p53. Science 1996;271:1744-7.

44. Gardner L, Malik R, Shimizu Y, et al. Geminin overexpression prevents the completion of topoisomerase IIalpha chromosome decatenation, leading to aneuploidy in human mammary epithelial cells. Breast Cancer Res 2011;13:R53.

45. Poehlmann A, Kuester D, Malfertheiner P, et al. Inflammation and Barrett's carcinogenesis. Pathol Res Pract 2012;208:269-80.

SUPPLEMENTARY TEXT AND TABLES
Statistical methods

In order to identify a biomarker panel that could predict histological outcome, we used a bootstrap resampling method (1) and took 100 bootstrap samples for each MI dataset and the original database. For each bootstrap sample, we tested individually the 512 different combinations of 9 biomarkers for their association with HGD/EC. Out of the 512 possible models, the best one was selected according to the Akaike information criterion (AIC) of a logistic regression model. In order to correct for potential patient effects when multiple AFI-targeted areas derived from the same patient, they were included as random effects in the model. For each biomarker we calculated the bootstrap inclusion frequency, which was defined as the number of times the biomarker was selected in the 100 best models. The biomarkers with a median inclusion frequency of at least 90 over the imputed datasets and original database were selected for the biomarker panel. (2) The predicted probability of dysplasia for each endoscopic area obtained from a logistic regression with the selected biomarker panel as predictor was used to generate the receiver operating characteristic (ROC) curves to evaluate the diagnostic accuracy of the panel and the area under the ROC curve (AUC) was reported.

For the purpose of the per-patient analysis, in patients with at least one AFI+ area and one AFI- area two datasets, called AFI+ and AFI- dataset, were created. In the AFI+ dataset, each biomarker was assigned a positive outcome, if the biomarker was positive in at least one AFI+ area, and negative if negative in all AFI+ areas. An identical process was followed for the AFI-dataset. The histological diagnosis for a patient was the overall histology (see biopsy and histology section). A logistic regression model, with the selected biomarker panel as predictor, was fitted using patients from a bootstrap sample and was then validated in the patients that were not selected in the bootstrap sample. This process was repeated 2000 times separately for the AFI+ and the AFI- datasets, resulting in 2000 AUC values for each dataset, which were compared in a paired t-test pairing bootstrap samples from the AFI+ and AFI- datasets. This comparison was done in all the 5 imputed databases and the original database.

To validate the biomarker panel selected in the per-biopsy analysis in the per-patient analysis, we applied a simple counting approach to the original database. For each possible panel of two or three biomarkers, a patient was predicted as HGD or EC if the number of positive biomarkers in the panel reached a certain cut-off point. We calculated sensitivity and specificity of each panel for a diagnosis for HGD/EC based on different cut-off points. The best biomarker panel was defined as the one with the highest accuracy (average of sensitivity and specificity). This process was repeated in 2000 bootstrap samples from the original database and the mean and standard deviation of the accuracy was reported for each panel at different cut-off points.

BIBLIOGRAPHY

1.   Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. Appl Statist 1999;48(3):313-29.

2.   Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. Journal of the Royal Statistical Society: Series C (Applied Statistics) 1998;48(3):313-29.
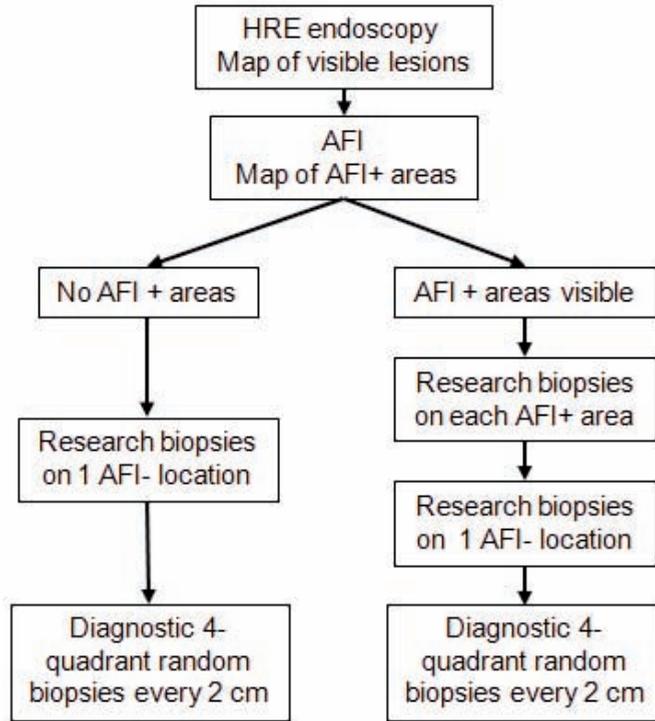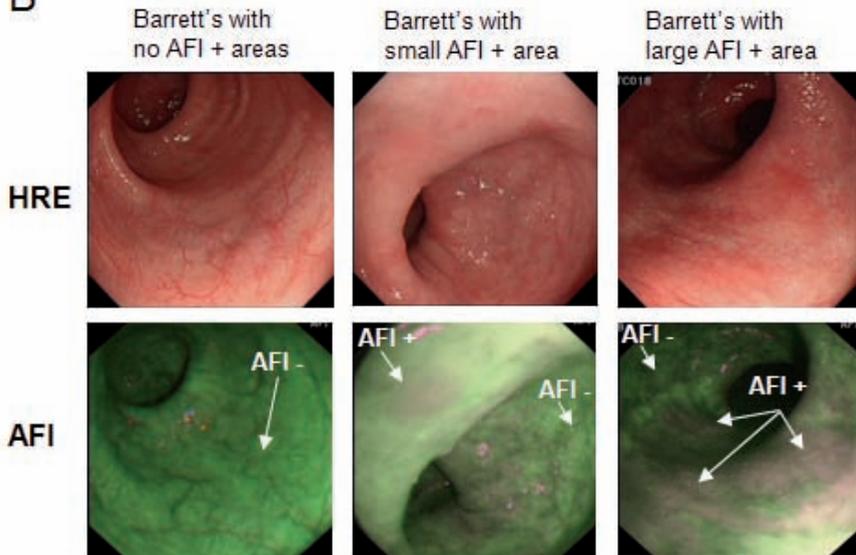
Supplementary Figure 1. Endoscopic imaging and biopsy protocols. A. Flowchart of endoscopic imaging and biopsy protocol. B. Examples of endoscopic view on high resolution endoscopy (HRE) and autofluorescence imaging (AFI) in patients with no AFI+ areas (left), small AFI+ areas (middle) and large/complex AFI+ areas (right).

Supplementary Table 1. Association of molecular biomarkers with AFI status.

Biomarker	Biomarker outcome	All areas			Non dysplastic areas		
		AFI-	AFI+	P	AFI-	AFI+	P
HPP1 methylation	Negative	30 (23.8%)	32 (18.3%)	0.25	29 (25.2%)	27 (22.9%)	0.76
	Positive	96 (76.2%)	143 (81.7%)		86 (74.8%)	91 (77.1%)	
RUNX3 methylation	Negative	50 (39.7%)	58 (33.1%)	0.27	49 (42.6%)	48 (40.7%)	0.79
	Positive	76 (60.3%)	117 (66.9%)		66 (57.4%)	70 (59.3%)	
P16 methylation	Negative	71 (56.3%)	84 (48.0%)	0.16	67 (58.3%)	65 (55.1%)	0.69
	Positive	55 (43.7%)	91 (52.0%)		48 (41.7%)	53 (44.9%)	
P53 IHC	Negative	96 (76.8%)	103 (52.0%)	<0.01	93 (83.0%)	96 (71.6%)	0.05
	Positive	29 (23.2%)	95 (48.0%)		19 (17.0%)	38 (28.4%)	
Cyclin A IHC	Negative	111 (84.1%)	135 (72.2%)	0.02	106 (89.1%)	117 (88.6%)	1.00
	Positive	21 (15.9%)	52 (27.8%)		13 (10.9%)	15 (11.4%)	
Tetraploidy	Negative	89 (70.6%)	114 (65.9%)	0.45	85 (74.6%)	86 (69.9%)	0.47
	Positive	37 (29.4%)	59 (34.1%)		29 (25.4%)	37 (30.1%)	
Aneuploidy	Negative	115 (91.3%)	128 (74.0%)	<0.01	109 (95.6%)	107 (87.0%)	0.02
	Positive	11 (8.7%)	45 (28%)		5 (4.4%)	16 (13.0%)	
17p LOH	Negative	60 (53.6%)	60 (38.0%)	0.01	57 (55.9%)	50 (48.1%)	0.27
	Positive	52 (46.4%)	98 (62.0%)		45 (44.1%)	54 (51.9%)	
9p LOH	Negative	19 (17.0%)	30 (19.5%)	0.63	17 (16.7%)	23 (22.3%)	0.38
	Positive	93 (83.0%)	124 (80.5%)		85 (83.3%)	80 (77.7%)	

Supplementary Table 2. Twenty best panels of biomarkers for the diagnosis of HGD/EC from a total pool of 84 different combinations of 3 biomarkers and 36 combinations of 2 biomarkers with different cut-off values. Cut-off refers to the number of abnormal biomarkers required to yield a positive outcome. For each cut off the means of the diagnostic accuracy form 2000 bootstrap samples and the standard deviation are represented. Highlighted in bold the best cut off for each panel.

Biomarker 1	Biomarker 2	Biomarker 3	Mean			Standard deviation		
			cut-off 1	cut-off 2	cut-off 3	cut-off 1	cut-off 2	cut-off 3
Cyclin A IHC	p53 IHC	aneuploidy	0.765272	0.866055	0.780642	0.025394	0.037145	0.048938
Cyclin A IHC	p53 IHC	tetraploidy	0.682272	0.859118	0.709647	0.024396	0.033878	0.051788
Cyclin A IHC	p53 IHC	P53 LOH	0.683017	0.856624	0.798002	0.024707	0.034406	0.047969
Cyclin A IHC	p53 IHC	p16 methyl	0.683315	0.839828	0.76563	0.024397	0.034727	0.048326
p53 IHC	tetraploidy	P53 LOH	0.635249	0.836269	0.65821	0.024351	0.036256	0.053326
Cyclin A IHC	p53 IHC	N/A	0.778035	0.83559	N/A	0.023653	0.040751	N/A
Cyclin A IHC	tetraploidy	aneuploidy	0.707605	0.835231	0.658476	0.030779	0.042051	0.047343
Cyclin A IHC	p53 IHC	RUNX3 methyl	0.632646	0.799535	0.834621	0.022021	0.035843	0.044334
Cyclin A IHC	aneuploidy	RUNX3 methyl	0.641786	0.832034	0.752845	0.023248	0.040472	0.050709
Cyclin A IHC	p53 IHC	p16 LOH	0.599525	0.827687	0.793979	0.020432	0.025124	0.047418
Cyclin A IHC	aneuploidy	P53 LOH	0.679156	0.826015	0.712769	0.038479	0.045307	0.050269
p53 IHC	RUNX3 methyl	N/A	0.632093	0.811182	N/A	0.028636	0.040024	N/A
Cyclin A IHC	RUNX3 methyl	N/A	0.646782	0.808857	N/A	0.028894	0.044359	N/A
p53 IHC	HPP1 methyl	P53 LOH	0.537662	0.706159	0.806901	0.013877	0.034944	0.046313
p53 IHC	p53 LOH	N/A	0.71064	0.801271	N/A	0.030795	0.045302	N/A
Cyclin A IHC	p53 IHC	HPP1 methyl	0.540834	0.815344	0.805547	0.013671	0.024707	0.046789
Cyclin A IHC	p53 IHC	P53 LOH	0.683017	0.856624	0.798002	0.024707	0.034406	0.047969
p53 IHC	RUNX3 methyl	P53 LOH	0.609351	0.778615	0.795858	0.021577	0.03136	0.046959
Cyclin A IHC	p53 IHC	P53 LOH	0.599525	0.827687	0.793979	0.020432	0.025124	0.047418
p53 IHC	aneuploidy	N/A	0.78885	0.79205	N/A	0.031008	0.049292	N/A

Supplementary Table 3. Detailed biomarker outcome in patients with a complete 3-biomaker panel dataset.

Number of positive biomarkers	Number of patients with NDBO/LGD (%)	Number of patients with HGD/EC (%)	Total of patients
0	50 (100.0%)	0 (0)	50
1	28 (96.6%)	1 (3.4%)	29
2	8 (53.3%)	7 (46.7%)	15
3	2 (11.1%)	16 (88.9%)	18
Total of patients	88	24	112

Supplementary Table 4. Detailed biomarker outcome in patients with a complete 3-biomaker panel dataset.

Outcome of the 3-biomarker panel	Number of patients with NDBO/LGD (%)	Number of patients with HGD/EC (%)	Total of patients
Negative (<2 positive biomarkers)	27 (90.0%)	0 (0%)	27
Positive (≥2 positive biomarkers)	3 (10.0%)	9 (100.0%)	12
Total of patients	30	9	39