Autofluorescence-directed confocal endomicroscopy in combination with a 3-biomarker panel can inform management decisions in Barrett's esophagus

Running Title: confocal endomicroscopy and biomarkers in Barrett's

Authors: Massimiliano di Pietro[1], Elizabeth Bird-Lieberman[1,2], Xinxue Liu[1], Tara Nuckcheddy-Grant[1], Helga Bertani[3], Maria O'Donovan[4], Rebecca C Fitzgerald[1]

[1] MRC Cancer Unit at the University of Cambridge, Cambridge UK

[2] Translational Gastroenterology Unit, Experimental Medicine Division, John Radcliffe Hospital, Oxford, UK

[3] Department of Digestive Endoscopy, Nuovo Ospedale Civile S.Agostino Modena, Italy

[4] Department of Histopathology, Cambridge University Hospitals, Cambridge UK

Corresponding author: Massimiliano di Pietro

MRC Cancer Unit at the University of Cambridge

Cambridge Biomedical Campus, Box 197

Email: md460@mrc-cu.cam.ac.uk

Tel +44 1223 763349; fax: +44 1223 763296

Word count: 5,455

Keywords: Barrett's esophagus, esophageal adenocarcinoma, confocal laser endomicroscopy, molecular biomarkers, autofluorescence imaging

ABSTRACT

OBJECTIVES: Barrett's esophagus (BE) surveillance with white-light endoscopy and quadrantic biopsies (Seattle protocol) is resource intensive and limited by sampling error. Previous work suggests that autofluorescence imaging (AFI) in combination with a molecular panel might reduce the number of biopsies, but this was not sufficiently sensitive for low grade dysplasia, which is now a point for endoscopic intervention. Here we used AFI to direct narrow-field imaging tools for real-time optical assessment of dysplasia and biopsies for a biomarker panel. We compared the new diagnostic algorithm with the current standard

METHODS: 55 patients with BE were recruited at a single tertiary referral center. Patients underwent high-resolution endoscopy followed by autofluorescence imaging (AFI). AFI-targeted areas (n=194) were examined in turn by narrow-band imaging with magnification (NBIz) and probe-based confocal laser endomicroscopy (pCLE). Biopsies were taken from AFI-targeted areas and tested using an established molecular panel comprising aneuploidy plus Cyclin A and p53 immunohistochemistry.

RESULTS: In the per-patient analysis the overall sensitivity and specificity of AFI-targeted pCLE were 100% and 53.6% for high grade dysplasia/intramucosal cancer and 96.4% and 74.1% for any grade of dysplasia, respectively. NBIz had equal specificity for dysplasia detection (74.1%), but significantly lower sensitivity (57.1%) than pCLE. The time required to perform AFI-targeted pCLE was shorter that that taken by the Seattle protocol (p=0.0004). We found enrichment of molecular abnormalities in areas with optical dysplasia by pCLE (p<0.001), regardless of histologic dysplasia. The addition of the 3-biomarker panel reduced the false positive rate of pCLE by 50%, leading to sensitivity and specificity for any grade of dysplasia of 89.2% and 88.9%, respectively.

CONCLUSION: The combination of pCLE on AFI-targeted areas and a 3-biomarker panel identifies patients with dysplasia.

STUDY HIGHLIGHTS

1. WHAT IS CURRENT KNOWLEDGE

- Endoscopic surveillance for Barrett's esophagus (BE) is problematic due to sampling error by random biopsies.

- Based on the current evidence, advanced imaging cannot replace random biopsies

- We showed previously that a 3-biomarker panel directed by autofluorescence imaging (AFI) can accurately diagnose high grade dysplasia, but sensitivity for low-grade dysplasia was low.

2. WHAT IS NEW HERE

- Probe-based confocal laser endomicroscopy (pCLE) directed by AFI has high sensitivity for both high-grade and low-grade dysplasia in BE

- Optical dysplasia by pCLE correlates with the field of molecular change regardless of histologic dysplasia

- The addition to AFI-directed pCLE of a 3-biomarker panel on minimal biopsy sampling further improves the overall diagnostic accuracy for any grade of dysplasia

INTRODUCTION

Barrett's esophagus (BE) is the only known precursor to esophageal adenocarcinoma (EA), with a

conversion rate to cancer estimated to be around 0.3% per year [1, 2] . All major gastroenterology

and endoscopy societies recommend endoscopic surveillance with a protocol entailing 4 random

biopsies every 2 centimetres in addition to targeted sampling of visible lesions to optimise detection

of dysplasia (Seattle protocol)[3-5]. The Seattle protocol is limited by sampling error due to the

heterogeneous and frequent inconspicuous nature of dysplasia. In addition, inter-observer

agreement for a pathological diagnosis of dysplasia is poor [6, 7].

In order to improve detection of dysplasia, several enhanced imaging modalities have been

investigated, which include both conventional dye and electronic (virtual) chromoendoscopic

techniques [8]. Autofluorescence imaging (AFI) is a red-flag endoscopic technique, which exploits the

intrinsic fluorescent properties of the gastro-intestinal (GI) mucosa [9]. Architectural tissue changes

associated with dysplasia and early cancer can cause loss of autofluorescence and can be detected in

AFI mode as purple-red signal within green background. Despite high sensitivity for intra-epithelial

neoplasia, AFI suffers from a high false positive rate (up to 80%) [10]. To obviate to this, AFI has been

incorporated into trimodal endoscopes with high resolution endoscopy (HRE) and narrow band

imaging with magnification (NBIz). However, two cross-over studies showed that this technology

lacks sufficient accuracy to replace the Seattle protocol since 7 to 21% of patients with HGD can be

missed if relying on targeted biopsies only [11, 12]. Possible explanations for this are: *i.* low

sensitivity of NBIz for focal HGD and LGD; *ii.* sampling error by single biopsies within large areas of

AFI positivity; *iii.* true occurrence of AFI-negative dysplasia. More recently, confocal laser

endomicroscopy (CLE) has allowed real-time microscopic analysis of the GI mucosa to detect

histologic dysplasia [13]. Two CLE systems are currently available, *i.e.* the probe-based CLE (pCLE)

and the endoscopy-integrated one (eCLE). In a cross-sectional study on 101 patients, Sharma and

colleagues found that the combination of pCLE and HRE had a sensitivity and a specificity for

HGD/IMC of 93.5 and 67.1% respectively [14]. In a more recent study, patients were randomized to have either HRE (n=98) or HRE + eCLE (n=94) [15]. The addition of eCLE reduced the total number of biopsies and led to a 2.7-fold higher diagnostic yield for HGD/IMC in the per-patient analysis with an overall 95% sensitivity for BE neoplasia. A disadvantage of CLE is the narrow field of view that can lead to sampling error despite multiple optical biopsies, similar to random sampling for conventional histology. However, pCLE is compatible with all commercially available endoscopes, hence it can be used in combination with other a red-flag imaging techniques to reduce the number of locations required for assessment of optical dysplasia. The accuracy of pCLE for diagnosing dysplasia has never been compared to that of other magnified imaging techniques, such as NBIz, particularly when used in combination with a red-flag technique such as AFI. Furthermore, these magnified endoscopic tools have not been investigated for the detection of low grade dysplasia (LGD).

The advent of endoscopic therapy and the evidence of cancer prevention by radiofrequency ablation (RFA) in patients with a consensus diagnosis of LGD have reinforced the importance of improving dysplasia detection during BE surveillance [16, 17]. Hence, the issues of poor inter-observer agreement for a pathological diagnosis of LGD and lack of validation of advanced imaging techniques for detection of this stage of disease need urgent consideration to improve patient management. Tissue biomarkers offer the advantage of measuring molecular changes that correlate with dysplasia as well as disease progression [18, 19] and could be used as adjunct to guide patient management. In a recent multi-center study, we tested a large panel of 9 molecular biomarkers in combination with AFI. We found that AFI positive [AFI(+)] areas had enrichment of biomarkers and we showed that 3 of these biomarkers (aneuploidy, cyclin A and p53 immunohistochemistry) outperformed the others in their association with prevalent dysplasia. This 3-biomarker panel assessed on biopsies directed by AFI had a very high per-patient diagnostic accuracy for HGD and IMC (96% sensitivity and 89% specificity). However, the sensitivity for any grade of dysplasia was lower (75%) [20].

Having established previously the role of AFI in enhancing detection of molecular changes in BE, in the present study we took a step further and used AFI to guide the use of pCLE for a real-time optical diagnosis of dysplasia without the need for histology. We thus compared the diagnostic accuracy of AFI-targeted pCLE with that of NBIz, which has previously been investigated in combination with AFI [11, 12]. We then hypothesized that a 3-biomarker panel on biopsies form the same AFI-targeted areas could further refine patient prognostication. The primary objectives were: *i*. the diagnostic accuracy for HGD/IMC of AFI-targeted optical dysplasia by NBIz or pCLE compared to the current gold standard Seattle protocol; and *ii*. the added value of molecular biomarkers to optical dysplasia to diagnose any grade of dysplasia. The secondary aims were: *i*. the direct comparison of pCLE and NBIz for the prediction of targeted histology; *ii*. the overall diagnostic accuracy for any grade dysplasia of AFI-targeted NBIz or pCLE compared to the Seattle protocol; *iii*. the interobserver agreement for optical dysplasia by pCLE.

METHODS

Patients and setting

The study was approved by the Cambridgeshire 2 Research Ethics Committee (09/H0308/118). Patients were recruited prospectively between March 2012 and April 2014 at a single tertiary referral center for BE. Inclusion criteria were: age >18 years, known BE with minimum length of C ≥2cm or M ≥4cm (if C<2) according to the Prague classification [21], referral for evaluation of dysplastic BE or follow-up post endoscopic resection for HGD/IMC. Short segments of BE were excluded due to the excess of AFI false positivity at the gastro-esophageal junction and since sampling error and time taken to perform biopsies is not such an issue for these cases [22]. Exclusion criteria were: esophagitis (Los Angeles grade ≥B); previous upper-GI surgery (with exception of Nissen fundoplication) or known upper-GI tract abnormality (e.g. pharyngeal pouch); known allergy to fluorescein; severe or uncontrolled asthma, which increases the risk of adverse reactions to

fluorescein; coagulopathy or anticoagulant/antiplatelet therapy for high risk conditions; active or severe cardiopulmonary disease or liver disease; dysphagia; special communication needs.

Endoscopic procedure

Prior to the procedure patients received either local anaesthesia with lignocaine spray or conscious sedation with Midazolam +/- Fentanyl. Patients were endoscoped with FQ260Z endoscopes (Olympus inc, Tokyo, Japan) as previously described.[11] Two endoscopists performed the procedures without randomization. Both endoscopists had extensive experience in AFI and NBIz, with more than 100 procedures each. Prior to patient recruitment, the endoscopists underwent online pCLE training (www.cellvizio.net) until they reached at least 90% correct scoring in 5 consecutive sets of 10 videos. The esophagus was first inspected by white light, high resolution endoscopy (HRE) to detect visible lesions (Figure 1). Then, imaging was switched to the AFI mode, and AFI was used as a wide-field detector to identify and carefully map AFI(+) areas (red-purple in colour). For each patient, a control AFI negative [AFI(-)] area (green colour) was randomly chosen. Only these AFI-targeted areas [AFI(+) and one AFI(-) control] were further studied with the narrow-field optical techniques (NBIz and pCLE) and with biomarkers on biopsies. Up to a maximum of four AFI(+) areas were permitted in the research protocol, however small AFI(+) areas within 1 cm from the gastro-oesophageal junction were excluded, due to the well-known false positivity in close proximity to the gastric folds [23]. First, the mucosal pit pattern and the vascular pattern of AFI-targeted areas were evaluated by NBIz and classified as regular and irregular as previously described [24]. In the absence of clear endoscopic landmarks to help locate AFI-targeted areas on HRE or NBI, these were marked with APC (ERBE, VIO200, 2.0 litres, 35W), as fluorescein injection required for the confocal examination causes loss of AFI positivity (Figure 1). Second, pCLE was performed on AFI-targeted areas after injection of fluorescein (2.5ml of 10% solution). pCLE sequences were generated and recorded. pCLE patterns were classified in "live-mode" as dysplastic and non-dysplastic based on previously published criteria [13]. Immediately after the end of the procedure, the recorded pCLE

videos were reviewed "off-line" by the endoscopist. The post-procedure assessment was recorded as the final outcome of the optical biopsy and was matched with the histological outcome.

Biopsy and histology

Each AFI(+) area and one AFI(-) area were biopsied for histopathology and biomarker analysis. Where possible, depending on the size of the AFI(+) area, two biopsies were taken in the following order of priority: one formalin-fixed and paraffin-embedded (FFPE) and one snap frozen in 10% dimethysulfoxide (DMSO). For AFI(+) positive areas larger than 1 cm two or more biopsies were taken for histology to minimize sampling error. Histopathological assessment of each AFI-targeted area relied on the FFPE biopsies. Random biopsies were then taken according to the Seattle protocol for histopathological analysis only, including any visible lesions, which had not been flagged by AFI. The histology on AFI-targeted and random biopsies was assessed by one of the hospital GI pathologists according to the Vienna classification [25]. All dysplastic cases, including indefinite for dysplasia (ID), were then reviewed by a second GI pathologist (MOD), who has extensive expertise in BE and satisfactory level of inter-observer agreement with external pathologists in previous studies [18, 20]. For the purpose of the data analysis, cases with ID were regarded as non-dysplastic, as they lacked definite dysplasia. In accordance with the Vienna classification cases of HGD, carcinoma in situ and intramucosal adenocarcinoma (IMC) were grouped together (HGD/IMC), as they represent a common endpoint for endoscopic therapeutic intervention [4, 26]. 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); and b) the overall histology (criterion a + histology on AFI targeted biopsies), which for the purpose of the analyses was regarded as the gold standard.

Procedure time

pCLE time was defined as the time between i.v. injection of fluorescein and the end of the last pCLE sequence recorded. Seattle protocol time was defined as the time taken to perform the Seattle protocol including biopsies of visible lesions.

Biomarker analyses

The biomarker panel included cyclin A, p53 and aneuploidy selected on the basis of previously published data [20]. Cyclin A and p53 expression were assessed by immunohistochemistry (IHC) with automated staining (BOND™ System, Leica Microsystems, Ltd, Milton Keynes, UK). Surface cyclin A was scored as previously described by a single investigator (MdP). Cases with a percentage of positive surface cells ≥1% were regarded as positive [27]; dubious cases were reviewed by an independent investigator (PL-S) and consensus reached. P53 expression was scored by a single investigator (MOD); staining was reported positive in the presence of areas of strong staining or complete loss of staining, compared to the background levels, as previously described [28]. Aneuploidy was assessed by flow-cytometry on cell isolated from frozen biopsies. The cell cycle histogram was analyzed using ModFit LT (Verity Software House, Topsham, ME, USA) by a single investigator (MdP), who was blind to the histological outcome; dubious cases were reviewed by an independent investigator (CR-I) and consensus was reached. When combined into a panel, we used a cut-off of 2 biomarkers positive out of 3 for a diagnosis of HGD/IMC and 1 out of three for a diagnosis of any grade of dysplasia.

Sample size calculation

The study was powered to address the first primary objective (diagnostic accuracy for HGD/IMC) with particular reference to AFI-targeted pCLE which represented the novel imaging approach. From previous studies we expected that the combination of AFI and pCLE would have a sensitivity and a specificity of at least 90% and 80%, respectively [13, 14]. With this level of diagnostic accuracy we calculated that we would have needed at least 14 patients with HGD/IMC and 31 with a lower stage

of disease to show a sensitivity of at least 0.70 or better and a specificity of at least 0.65 or better at a significance level of 0.05.

Statistical analysis

For the purpose of the per-location analysis, each AFI-targeted area was classified as follows: *i*. NBI status (NBI irregular or NBI regular); *ii*. pCLE status (pCLE irregular or pCLE regular); *iii*. binary outcome of the 3 biomarkers; *iv*. histological diagnosis. For the purpose of the per-patient analysis, we considered the overall outcome of the NBIz and pCLE (irregular if at least 1 area was NBIz or pCLE irregular) and the pathological outcome from the biopsies taken according to the Seattle protocol, and compared these with the overall histology. Chi-square tests were used to compare differences between groups. A p-value <0.05 was considered statistically significant. Intra-observer agreement was assessed by κ–statistics on all good quality sequences (n=292) between the "live mode" diagnosis made during the procedure and the "off-line" diagnosis made at the end of the procedure. Inter-observer agreement for pCLE pattern recognition was assessed by κ–statistics on 120 randomly selected sequences after exclusion of 15 with sub-optimal quality. The agreement was assessed between the off-line post-procedure diagnosis by MdP or EB-L and a third independent endoscopist with extensive expertise in pCLE (HB). According to previously published criteria [13], the endoscopist defined each sequence as non-dysplastic or dysplastic, and were allowed to use a call "dubious for dysplasia" if undecided (n=10). Agreement was expressed as a weighted κ-value. The "dubious for dysplasia" category was allowed since these sequences were randomly selected for off-line evaluation, whereas when performing pCLE live the endoscopist had the possibility to perform additional sequences on the same location if in doubt.

RESULTS

Fifty-five patients with BE were recruited and underwent trimodal imaging (HRE+AFI+NBIz) followed by AFI-targeted pCLE. The demographics are shown in Table 1. Overall, in this enriched population 51% of patients had definite dysplasia, with an equal prevalence of LGD and HGD/IMC cases. In total 194 AFI-targeted locations were identified at endoscopy [AFI(+):AFI(-) 138:56] with the following distribution of histological diagnoses: NDBE n=122, ID n=20; LGD n=24, HGD n=21 and IMC n=7. These endoscopic areas were subjected to optical analysis by NBIz and pCLE in turn and finally to biopsies for histology and biomarkers. Patients had a median number of 3 AFI(+) areas analyzed (range 0-5), in addition to one AFI(-) area. In two patients an AFI(-) area could not be selected due to diffuse low intensity AFI(+) signal and two additional patients had more than 1 AFI(-) area biopsied due to AFI(-) lesions visible on HRE. The majority of AFI(+) locations (81%) were inconspicuous at HRE, and on histopathology 17% of them contained HGD or IMC and 36% corresponded to BE with any degree of dysplasia (LGD + HGD/IMC) (Table 1). We first examined the diagnostic accuracy of the AFI-targeted biopsies for dysplasia, in comparison with the Seattle protocol. Conventional histopathology on AFI-targeted biopsies correctly staged all the patients with HGD/IMC and identified 2 additional patients with focal and inconspicuous HGD, who would have been missed if only Seattle protocol biopsies had been taken. However, AFI-targeted histology missed 7 patients found to have LGD elsewhere on random biopsies. Hence, histology on AFI-targeted biopsies had a sensitivity of 100% for HGD/IMC and 75% for any grade of dysplasia, whereas the Seattle protocol had a sensitivity of 85.7% for HGD/IMC and 92.8% for any grade of dysplasia (Table 1).

In keeping with previously published data, although highly sensitive, in this study AFI alone had a high false positive rate for HGD/IMC (82.7%) and any grade of dysplasia (69.5%) (Table 2). To assess whether additional advanced imaging modalities could predict histologic dysplasia with higher specificity, we used NBIz and pCLE to study areas flagged by AFI in more detail. In the per-location analysis, pCLE had significantly better sensitivity than NBIz for all grades of dysplasia (Table 2). pCLE reduced the false positive rate of AFI for HGD/IMC and any grade of dysplasia to 69.7% and 48.7%

11

respectively. Similarly NBIz decreased the false positive rate of AFI for HGD/IMC and any grade of

dysplasia to 53.5% and 39.3%, respectively. The intra-observer agreement on optical histology

assessed by pCLE between live-mode and off-line diagnosis was 'good' with a κ-value of 0.751

(standard error 0.04). The pCLE agreement with an independent endoscopist was 'moderate' with a

weighted κ-value of 0.575 (standard error 0.079), which is not dissimilar from that reported by

Sharma and colleagues in a previous multicenter study [14].

We next performed a per-patient analysis to assess the accuracy of pCLE or NBIz in predicting the

overall pathological outcome based on the imaging patterns only. The basis for this is that the

presence of optical dysplasia can not only reflect dysplasia in any given location but also be

informative about the overall status of the patient. pCLE correctly identified all the patients with

HGD/IMC, while NBIz missed 2 patients with HGD (Table 1), however the differences in sensitivity

and specificity for HGD/IMC between the two modalities did not reach statistical significance (Table

3). When we looked at diagnosis of any degree of dysplasia within the BE segment by optical biopsy,

pCLE had significantly better sensitivity than NBIz (96.4% vs 57.1%), with equal specificity (74.1%)

(Table 3). Real time histology by pCLE down-graded 1 patient with LGD, whereas it overestimated 3

patients with non-dysplastic BE. As a result, AFI-targeted pCLE had equal sensitivity for any grade of

dysplasia when compared to the Seattle protocol (96.4% vs 92.8%, respectively) (Table 3). However,

AFI-targeted pCLE relied on an average of 3.5 optical biopsies, which compared favourably with 12.9

histological biopsies per patient required by the Seattle protocol. We compared the time taken to

perform AFI-targeted pCLE and the multiple biopsies according to the Seattle protocol and found

that the optical biopsies required significantly shorter time (p=0.0004) (Figure 2).

Despite high diagnostic accuracy for any grade of dysplasia, pCLE was affected by a 49% false

positive rate in the per-location analysis when matched with histology within the corresponding

area. It is possible that this discrepancy could be explained by sampling error for focal dysplasia,

which was detected by detailed examination of AFI(+) areas by pCLE, but missed by a single targeted

biopsy for histology. Hence, this would explain why AFI-targeted pCLE had better sensitivity for any

grade of dysplasia than AFI-targeted histology (96.4% vs 75%). In particular, if this was true, one

would expect a correlation between optical dysplasia by pCLE and the field of molecular change,

which normally has a larger spatial distribution than histological dysplasia [29]. To test this

hypothesis we looked at the correlation between pCLE patterns and molecular biomarkers. We

concentrated on a 3-biomaker panel (aneuploidy + IHC for p53 and cyclin A) that was shown in a

recent phase IV prospective cross-sectional study to have a high degree of association with prevalent

dysplasia [20]. As shown in Figure 3a, we found a significant correlation between the number of

positive biomarkers and the presence of optical dysplasia within the corresponding targeted area. As

this could be biased by the high enrichment of dysplasia in our cohort of patients, we repeated this

analysis after exclusion of areas with neoplasia visible on HRE (Figure 3b) or all areas with any degree

of dysplasia (Figure 3c). The significant correlation between positive biomarkers and pCLE

irregularity was retained in both analyses, suggesting optical dysplasia associates with molecular

changes regardless of histologic dysplasia found on the biopsy sample. We then looked at the

correlation between molecular biomarkers and histologic dysplasia (figure 3d). As expected the areas with

HGD/IMC were highly molecularly abnormal. Interestingly there was a striking difference between ID and LGD,

in that all the areas with ID, except one, had no positive biomarkers as opposed to 85% of LGD areas having at

least 1 positive biomarker. The correlation between molecular biomarkers and histologic dysplasia

occurred regardless of the presence of optical dysplasia in the corresponding area (Figure 3e and 3f,

respectively).

In this study pCLE yielded a high diagnostic accuracy for any degree of dysplasia, however by using

the published criteria we could not distinguish between different grades of dysplasia. In particular, 4

out of 5 patients with indefinite for dysplasia (ID) had at least one area with optical dysplasia on

pCLE, suggesting that inflammatory changes often underlying this pathological diagnosis could

confound the optical assessment. Therefore, we looked at whether the biomarkers could be used to

integrate the information obtained with real-time pCLE to allow more accurate patient stratification.

As showed in Figure 4, all the patients with HGD/IMC had optical dysplasia as well as at least 2 abnormal biomarkers, but a cut-off of 2 positive biomarkers was too stringent for a diagnosis of LGD. By lowering the biomarker threshold, we found that 79% of patients with LGD presented with optical dysplasia and at least one positive biomarker. On the other hand, 86% of patients with NDBE had no optical dysplasia and half of them had no biomarker positivity and no optical dysplasia. Conversely, among the 5 patients with ID as overall diagnosis, 4 had optical dysplasia but only 2 had at most 1 positive biomarker. Therefore, using a cut-off of 1 biomarker or more for the diagnosis of any grade of dysplasia in combination with optical dysplasia by pCLE, the panel improved the specificity from 74.1% to 88.9%, with a minimal drop in the sensitivity, compared to AFI-targeted pCLE alone (Table 3). Of note, the group of patients with LGD was almost equally split between patients having ≥2 and <2 positive biomarkers. We looked at the correlation between clinical factors (age, length of the BE, spatial distribution of dysplasia) and the molecular outcome and found that in patients with LGD the presence of ≥2 positive biomarkers was associated with multifocal dysplasia (p=0.03). Overall, these results indicate that for long BE segments the combination of AFI-targeted pCLE and a small biomarker panel has high diagnostic accuracy for dysplasia in BE and that biomarkers help provide an objective measure of the dysplasia status of the patient in a more time-efficient manner than the Seattle protocol.

DISCUSSION

This study has two main novel findings: *i*. Real-time assessment by pCLE performed on AFI-targeted locations has a good diagnostic accuracy for any grade of dysplasia in BE (sensitivity 96.4% and specificity 74.1%); *ii*. the integration of pCLE and molecular biomarkers further improves specificity by reducing the false positive rate and has the potential for further patient stratification compared to the standard clinical protocol.

Recently two large multicenter studies have evaluated the accuracy of CLE to diagnose neoplasia in BE based on assessment of random as well as HRE-targeted locations. Even though the diagnostic accuracy for HGD/IMC reported by these studies is similar to ours, these trials relied on a high

number of areas assessed by CLE. For example, Sharma and colleagues analysed 8 locations per patient in a cohort that had a much shorter average BE length compared to ours (3.6cm vs 7.4cm). Our study is the first to use CLE exclusively directed by a flagging imaging technique. Of note, in our cohort on average only 3-4 locations per patient were used for the optical biopsies. When analysing the time required to perform pCLE, this was significantly shorter than that taken to carry out the Seattle protocol. Even though the use of AFI as a flagging technique for pCLE has the drawback of requiring mucosal marking due to the loss of AFI positivity after I.V. fluorescein, our results suggest that this imaging protocol is unlikely to prolong the procedure, if used as an alternative to the Seattle protocol.

One of the particularly complex diagnostic categories in BE is LGD, since there is a high inter-observer variability among GI pathologists and increasingly treatment is recommended for this patient group. Although pCLE has been validated for the diagnosis of HGD and IMC [13, 30], previous studies have not evaluated the accuracy for a diagnosis of LGD. Hence, although the primary endpoint of our study was to assess diagnostic accuracy for HGD/IMC we also extended the analysis to include LGD. In keeping with previous studies [13, 30], in our cohort of patients the sensitivity of pCLE for HGD/IMC in the per-patient analysis was high (100%), but this was accompanied by a low specificity (53%). Notably, we found that, extending the analysis to LGD as a histological endpoint, the specificity increased considerably (74%), without a significant drop in the sensitivity (96.4%). This is interesting since in the per-location analysis we did not find an optimal correspondence between pCLE and conventional histopathology, due to a false positive rate of 49%. Explanations for this could be that *i.* focal dysplasia was detected by pCLE by detailed examination of each AFI positive area, but it was missed by a single biopsy; *ii.* pCLE detected as dysplastic inflammatory changes that are often associated with a diagnosis of indefinite for dysplasia (ID). If the first hypothesis was true, we would expect a correlation between optical dysplasia and the field of molecular change, which is known to precede definite dysplasia and extend over larger mucosal surface [19, 27, 29, 31]. Indeed, we found that biomarker positivity correlated not only with histologic dysplasia, but also with optical dysplasia,

even in the analysis that excluded areas with a pathological evidence of dysplasia (Fig. 3C). This indicates that there is a tight relationship between the field of molecular change and dysplasia, whether it is detected by conventional histology or optical imaging. Furthermore it suggests that at least in a proportion of AFI-targeted areas, regarded as false pCLE positive on the basis negative histology, focal dysplasia might have been missed by conventional histology.

In this study, the combination of optical dysplasia and biomarkers with a cut-off of 1 had 89.2% sensitivity and 88.9% specificity for an overall per-patient diagnosis of confirmed dysplasia. Our data indicate that the presence of optical dysplasia and at least one molecular biomarker defines a high risk group for prevalent dysplasia which should be managed with early endoscopic treatment also in absence of visible neoplasia (Figure 4). It was clinically useful to distinguish between patients with 1 vs 2 positive biomarkers since we found a definite association between a diagnosis of HGD/IMC and positivity of two or more markers. Additionally in our cohort, patients with LGD and 2 or more positive biomarkers were more likely to have multifocal dysplasia. Previously published studies demonstrated a wide variety of cancer progression rates in patients with LGD [7, 32] suggesting that this category embraces a rather heterogeneous group of patients, which is in keeping with the findings from our optical/molecular stratification model. As there is evidence from previous studies that the cancer risk in BE relates to the number of positive biomarkers [18, 19], we feel that this stratification based on our panel is clinically relevant. On the other hand the absence of optical dysplasia and molecular changes clearly defines a low risk group, suitable for a more prolonged surveillance-interval, whereas an intermediate risk group with either optical dysplasia or one positive biomarker could have closer follow-up. Hence, the added value of the novel integrated protocol is that while it correctly diagnosed all the patients with HGD/IMC and 11 out of 14 patients with LGD, it also detected a high risk profile in a small fraction of patients with ID or non dysplastic BE, which could perhaps benefit from early endoscopic treatment. Larger studies with follow up data are required to confirm these findings. For the purpose of the statistical analysis we excluded cases with ID from the dysplastic group. There is controversy about the increased cancer risk in this group of patients [32, 33], hence ID is not

currently an indication to endoscopic intervention. It was interesting that the small group of patients with a diagnosis of ID had a different risk profile from patients with non dysplastic BE, since none of the ID patients had a low risk signature (no optical dysplasia and no positive biomarkers). This is in keeping with the accepted clinical practice that these patients should have closer follow up. However areas with ID had a much lower yield of positive biomarkers than LGD (10% vs 85%, Figure 3d), providing molecular support to the distinction between this category and cases with definite dysplasia.

While the optical diagnosis based on pCLE could allow avoidance of biopsies altogether, the addition of biomarkers relied on a small number of AFI targeted biopsies. On average 3 to 4 AFI targeted areas per patient were biopsied in this study. Our protocol included 2 biopsies for each AFI targeted area, one FFPE and one frozen for aneuploidy by flow cytometry. There is evidence, that image cytometry on FFPE samples is comparable to flow cytometry for assessment aneuploidy [34], hence the panel can be effectively applied to a single FFPE biopsy. These biomarker assays are straightforward and could be performed in a clinical laboratory, with the added advantage that they give a more objective result that histopathological analysis of dysplasia. Considering that the Seattle protocol in our cohort relied on an average of 12.9 biopsies per patient, the biomarkers would lead to a nearly 4-fold reduction in the biopsies required compared to the current standard.

This study has several limitations. Firstly, this a single-center study performed in a cohort of patients that was highly enriched for dysplasia. We chose this design and patient number in order to give an early indication concerning whether a larger, multicenter study is warranted. It is possible that in an unselected population the specificity may be lower, but it is promising that in this study the vast majority of patients with non-dysplastic BE (21 out 22) were correctly classified by the optical/molecular algorithm. Secondly, the endoscopists were not blinded to the prior histological findings and this could have influenced the optical diagnosis. We have looked at the pathological outcome preceding the study endoscopy and found that for 20% of patients histology was not available and for another 20% the outcome of the pCLE was discordant from the previous

histological findings comparing dysplasia with no dysplasia so this does not seem to have been a problem. Furthermore, all the laboratory scoring of biomarkers was done without knowledge of the histological diagnosis. Thirdly, the pathological finding of HDG/IMC was associated with visible lesions with the exception of two patients with focal HGD, indicating that within this group the proportion of patients where pCLE contributed to this diagnosis is small. However, all the patients with LGD (n=14) had no macroscopically visible lesions on HRE and 13 of them had optical dysplasia, indicating that pCLE can efficiently recognize dysplasia even within inconspicuous segment of BE. Fourth, the time taken to perform pCLE in this study is a slight under-estimate since did not take into account the selection of the locations by AFI, because it was difficult to extrapolate this time from that taken to perform trimodal imaging (HRE+AFI+NBIz).

In conclusion, this study provides evidence that the integration of multimodal imaging based on real time assessment of dysplasia by pCLE and a small panel of molecular biomarker can provide effective patient stratification to direct clinical management. Future studies on larger and less-selected cohorts of patients are required to confirm our findings.

References

1. Reid, B.J., et al., *Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis.* Nat Rev Cancer, 2010. **10**(2): p. 87-101.
2. Desai, T.K., et al., *The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis.* Gut, 2012. **61**(7): p. 970-6.
3. Fitzgerald, R.C., et al., *British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus.* Gut, 2013. **63**(1): p. 7-42.
4. Spechler, S.J., et al., *American Gastroenterological Association medical position statement on the management of Barrett's esophagus.* Gastroenterology, 2011. **140**(3): p. 1084-91.
5. Wang, K.K. and R.E. Sampliner, *Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus.* Am J Gastroenterol, 2008. **103**(3): p. 788-97.
6. Curvers, W.L., et al., *Low-grade dysplasia in Barrett's esophagus: overdiagnosed and underestimated.* Am J Gastroenterol, 2010. **105**(7): p. 1523-30.
7. Wani, S., et al., *Risk factors for progression of low-grade dysplasia in patients with Barrett's esophagus.* Gastroenterology, 2011. **141**(4): p. 1179-86, 1186 e1.
8. Coda, S. and A.V. Thillainayagam, *State of the art in advanced endoscopic imaging for the detection and evaluation of dysplasia and early cancer of the gastrointestinal tract.* Clin Exp Gastroenterol, 2014. **7**: p. 133-50.

9. Boerwinkel, D.F., et al., *Effects of Autofluorescence Imaging on Detection and Treatment of Early Neoplasia in Patients With Barrett's Esophagus.* Clin Gastroenterol Hepatol, 2013.

10. Boerwinkel, D.F., et al., *Fluorescence imaging for the detection of early neoplasia in Barrett's esophagus: old looks or new vision?* Eur J Gastroenterol Hepatol, 2014. **26**(7): p. 691-8.

11. Curvers, W.L., et al., *Endoscopic tri-modal imaging is more effective than standard endoscopy in identifying early-stage neoplasia in Barrett's esophagus.* Gastroenterology, 2010. **139**(4): p. 1106-14.

12. Curvers, W.L., 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**(2): p. 195-203.

13. Gaddam, S., et al., *Novel probe-based confocal laser endomicroscopy criteria and interobserver agreement for the detection of dysplasia in Barrett's esophagus.* Am J Gastroenterol, 2011. **106**(11): p. 1961-9.

14. Sharma, P., et al., *Real-time increased detection of neoplastic tissue in Barrett's esophagus with probe-based confocal laser endomicroscopy: final results of an international multicenter, prospective, randomized, controlled trial.* Gastrointest Endosc, 2011. **74**(3): p. 465-72.

15. Canto, M.I., et al., *In vivo endomicroscopy improves detection of Barrett's esophagus-related neoplasia: a multicenter international randomized controlled trial (with video).* Gastrointest Endosc, 2014. **79**(2): p. 211-21.

16. Shaheen, N.J., et al., *Radiofrequency ablation in Barrett's esophagus with dysplasia.* N Engl J Med, 2009. **360**(22): p. 2277-88.

17. Phoa, K.N., et al., *Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial.* JAMA, 2014. **311**(12): p. 1209-17.

18. Bird-Lieberman, E.L., et al., *Population-based study reveals new risk-stratification biomarker panel for Barrett's esophagus.* Gastroenterology, 2012. **143**: p. 927-35.

19. Galipeau, P.C., et al., *NSAIDs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma.* PLoS Med, 2007. **4**(2): p. e67.

20. di Pietro, M., et al., *The combination of autofluorescence endoscopy and molecular biomarkers is a novel diagnostic tool for dysplasia in Barrett's oesophagus.* Gut, 2014: p. Epub ahead of print.

21. Sharma, P., et al., *The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria.* Gastroenterology, 2006. **131**(5): p. 1392-9.

22. Curvers, W.L., 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**(1): p. 9-17.

23. Kara, M.A., et al., *Endoscopic video autofluorescence imaging may improve the detection of early neoplasia in patients with Barrett's esophagus.* Gastrointest Endosc, 2005. **61**(6): p. 679-85.

24. Kara, M.A., et al., *Detection and classification of the mucosal and vascular patterns (mucosal morphology) in Barrett's esophagus by using narrow band imaging.* Gastrointest Endosc, 2006. **64**(2): p. 155-66.

25. Dixon, M.F., *Gastrointestinal epithelial neoplasia: Vienna revisited.* Gut, 2002. **51**(1): p. 130-1.

26. Bennett, C., et al., *Consensus Statements for Management of Barrett's Dysplasia and Early-Stage Esophageal Adenocarcinoma, Based on a Delphi Process.* Gastroenterology, 2012. **143**(2): p. 336-46.

27. Lao-Sirieix, P., L. Lovat, and R.C. Fitzgerald, *Cyclin A immunocytology as a risk stratification tool for Barrett's esophagus surveillance.* Clin Cancer Res, 2007. **13**(2 Pt 1): p. 659-65.

28.   Depledge, D.P., 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**(7): p. e476.

29.   Galipeau, P.C., 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**(24): p. 2087-95.

30.   Wallace, M.B., et al., *Preliminary accuracy and interobserver agreement for the detection of intraepithelial neoplasia in Barrett's esophagus with probe-based confocal laser endomicroscopy.* Gastrointest Endosc, 2010. **72**(1): p. 19-24.

31.   Kastelein, F., et al., *Aberrant p53 protein expression is associated with an increased risk of neoplastic progression in patients with Barrett's oesophagus.* Gut, 2013. **62**(12): p. 1676-83.

32.   Duits, L.C., et al., *Barrett's oesophagus patients with low-grade dysplasia can be accurately risk-stratified after histological review by an expert pathology panel.* Gut, 2015. **64**(5): p. 700-6.

33.   Kestens, C., et al., *Endoscopic follow-up intervals of indefinite for dysplasia in Barrett's esophagus should be similar to those recommended for low-grade dysplasia: a dutch nationwide cohort study.* United European Gastroenterology Journal, 2013. **1**(1): p. A39.

34.   Dunn, J.M., 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**(11): p. 1608-17.

Table 1. Patient demographics

Variables	Patients n=55
Sex (M:F)	13:1
Mean age (range)	69.5 (61.7-77.3)
Length of Barrett's (mean M value)	7.4
Histological diagnosis	n (%)
NDBO	22 (40%)
Indefinite	5 (9%)
LGD	14 (25.5%)
HGD/IMC	14 (25.5%)
Endoscopic features	
Number of AFI location (AFI+:AFI-)	138:56
Median number of AFI+ areas per patient (range)	3 (0-5)
AFI+ areas visible on HRE	19%
AFI+ areas with HGD/IMC	17%
AFI+ areas with any grade of dysplasia	36%
Dysplastic patients missed	HGD/LGD
Seattle protocol	2/0
AFI-targeted histology	0/7
AFI-targeted pCLE (optical biopsy)	0/1
AFI-targeted NBIz (optical biopsy)	2/10

Table 2. Diagnostic performance of different imaging modalities in the per-location analysis.

	HGD/IMC				Any grade of dysplasia			
	AFI	NBIz[†]	pCLE[†]	p value[#]	AFI	NBIz[†]	pCLE[†]	p value[#]
Sensitivity	92.3%	61.9%	100%	0.001	86%	39.5%	83.3%	<0.0001
Specificity	31.1%	90.5%	67.1%	<0.0001	31.9%	92%	72.5%	<0.0001
FP rate	82.7%	53.5%	69.7%	0.12	69.5%	39.3%	48.7%	0.02

FP rate: false positive rate

[†]NBIz and pCLE were only performed on AFI-targeted locations

[#]comparison between NBIz and pCLE

Table 3. Diagnostic performance of different diagnostic modalities in the per-patient analysis

	HGD/IMC						
	Seattle^	AFI + NBI	AFI + pCLE	AFI + pCLE + biomarkers*	p value†	p value#	p value~
Sensitivity	86%	84.6%	100%	100%	0.14	0.13	1
Specificity	N/A	70.7%	53.6%	87.8%	N/A	0.09	0.0007
FP rate	N/A	52.1%	57.6%	26%	N/A	0.69	0.03
	LGD + HGD/IMC						
	Seattle^	AFI + NBI	AFI + pCLE	AFI + pCLE + biomarkers*	p value†	p value#	p value~
Sensitivity	92.8%	57.1%	96.4%	89.2%	0.89	0.002	0.3
Specificity	N/A	74.1%	74.1%	88.9%	N/A	0.78	0.16
FP rate	N/A	30.4%	20.1%	10.7%	N/A	0.33	0.29

FP rate: false positive rate

^ Specificity and FP rate do not apply as false positive cannot be found at histology

*cut offs: 2 biomarkers positive for HGD/IMC and 1 biomarker positive for LGD+HGD/IMC

† comparison between histology based on Seattle protocol biopsies and pCLE

comparison between NBIz and pCLE

~ comparison between pCLE and pCLE + biomarkers

Figure Legends

Figure 1 . Endoscopic imaging protocol. **A** Inconspicuous BE at high-resolution white light endoscopy. **B** Autofluorescence imaging (AFI) reveals an area of positive signal (white arrowheads). **C** NBI with optical magnification shows subtle changes in the mucosal pit (arrowhead) and vascular pattern (arrows) of difficult interpretation. **D** The AFI positive area is marked with APC. **E** After I.V. injection of 10% fluorescein the AFI positivity is lost due to widespread green fluorescent signal. **F** Analysis of the area by pCLE shows saw-tooth appearance of the glands surface, irregular inter-glandular space and loss of goblet cells in keeping with dysplasia. Pathological analysis showed low grade dysplasia.

Figure 2. Time taken to perform optical and conventional biopsies. pCLE indicates the time to perform endomicroscopic imaging of AFI targeted location. Seattle protocol refers to the time needed to take multiple biopsies according to the standard protocol (quadrantic random biopsies every 2 cm + targeted on visible lesions). The boxes represent the interquartile ranges with the median and the vertical line ranges from minimum to maximum values.

Figure 3. Association between molecular biomarkers and optical/histologic dysplasia. **A-C** Graphs represent the number of positive molecular biomarkers (x-axis) in endoscopic areas with no optical dysplasia (dark gray) and with optical dysplasia (light gray). Only areas with all three biomarkers available were included in this analysis. In **A** all endoscopic areas (n=125) are represented; in **B** only areas without neoplasia visible at high-resolution endoscopy (n=115) and in **C** only endoscopic areas without histologic dysplasia (n=85). **D** Percentage distribution of different numbers of positive biomarkers (0, 1 and 2/3) in endoscopic areas (n=125) with different histological stage of disease. **E-F** Graphs represent the number of positive molecular biomarkers (x-axis) in endoscopic areas with no histologic dysplasia (dark grey) and with histologic dysplasia (light grey). In **E** only endoscopic areas with optical dysplasia (n=56) are represented; in **F** only endoscopic areas without optical dysplasia (n=69) are shown.

Figure 4. Integration of optical dysplasia and molecular biomarker to inform patient management.

Patients are grouped based on their overall histology (NDBE: non-dysplastic Barrett's esophagus; ID: indefinite for dysplasia; LGD: low grade dysplasia; HGD/IMC: high grade dysplasia/intramucosal cancer). Groups are colour-coded and based on different combination of optical dysplasia and positive molecular biomarkers.