Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2016 June ; 25(6): 958–968. doi:10.1158/1055-9965.EPI-15-1164.

A Tissue Systems Pathology Assay for High-Risk Barrett's Esophagus

Rebecca J. Critchley-Thorne^{1,*}, Lucas C. Duits², Jeffrey W. Prichard³, Jon M. Davison⁴, Blair A. Jobe⁵, Bruce B. Campbell¹, Yi Zhang¹, Kathleen A. Repa¹, Lia M. Reese¹, Jinhong Li³, David L. Diehl³, Nirag C. Jhala^{6,‡}, Gregory Ginsberg⁶, Maureen DeMarshall⁶, Tyler Foxwell⁴, Ali H. Zaidi⁵, D. Lansing Taylor⁷, Anil K. Rustgi⁶, Jacques J.G.H.M. Bergman², and Gary W. Falk⁶

¹Cernostics, Inc., 235 William Pitt Way, Pittsburgh, PA ²Department of Gastroenterology and Hepatology, Academic Medical Centre, Amsterdam, The Netherlands ³Department of Laboratory Medicine, Geisinger Medical Center, Danville, PA ⁴Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA ⁵Esophageal and Lung Institute, Allegheny Health Network, Pittsburgh, Pennsylvania, PA ⁶Division of Gastroenterology, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA ⁷Drug Discovery Institute and Department of Computational & Systems Biology, University of Pittsburgh, Pittsburgh, PA

Abstract

Background—Better methods are needed to predict risk of progression for Barrett's esophagus (BE). We aimed to determine whether a tissue systems pathology approach could predict progression in patients with non-dysplastic BE, indefinite for dysplasia, or low-grade dysplasia.

Methods—We performed a nested case–control study to develop and validate a test that predicts progression of BE to high-grade dysplasia (HGD) or esophageal adenocarcinoma (EAC), based upon quantification of epithelial and stromal variables in baseline biopsies. Data were collected from BE patients at 4 institutions. Patients who progressed to HGD or EAC in 1 year (n=79) were matched with patients who did not progress (n=287). Biopsies were assigned randomly to training or validation sets. Immunofluorescence analyses were performed for 14 biomarkers and quantitative biomarker and morphometric features were analyzed. Prognostic features were selected in the training set and combined into classifiers. The top-performing classifier was assessed in the validation set.

*Corresponding Author's Full Contact Information: Name: Rebecca J. Critchley-Thorne, Ph.D. Address: Cernostics, Inc., 235 William Pitt Way, Pittsburgh, PA 15238. Tel: 412-828-0900 Fax: 412-828-0900; Email: rthorne@cernostics.com

‡Currently at: Department of Pathology and Laboratory Medicine, Temple University School of Medicine, Philadelphia, PA

Disclosures: Lia Reese, Anil Rustgi, D. Lansing Taylor, and Rebecca Critchley-Thorne hold equity ownership or stock options in Cernostics, Inc., the commercial entity that developed the proprietary TissueCypher technology used in this study. Bruce Campbell, Lia Reese, Kathleen Repa and Rebecca Critchley-Thorne are current or former full-time employees of Cernostics, Inc. Yi Zhang is a consultant to Cernostics, Inc. Jeffrey Prichard, Jinhong Li and David Diehl are employees of Geisinger Health System, which has a financial interest in Cernostics, Inc. Bruce Campbell and Rebecca Critchley-Thorne report being inventors on a patent application for the TissueCypher technology used in this study.

Results—A 3-tier, 15-feature classifier was selected in the training set and tested in the validation set. The classifier stratified patients into low-, intermediate- and high-risk classes (hazard ratio, 9.42; 95% confidence interval, 4.6–19.24 (high-risk vs low-risk); $P<0.0001$). It also provided independent prognostic information that outperformed predictions based on pathology analysis, segment length, age, sex, or p53 overexpression.

Conclusion—We developed a tissue systems pathology test that better predicts risk of progression in BE than clinicopathologic variables.

Impact—The test has the potential to improve upon histologic analysis as an objective method to risk stratify BE patients.

Keywords

Barrett's esophagus; risk assessment; systems biology

Introduction

Barrett's esophagus (BE) is a precursor to esophageal adenocarcinoma (EAC). Although the risk of progression of BE to EAC is very low (¹⁻³), treatment options for advanced EAC are limited and early detection is critical for optimal patient management. EAC can be prevented if dysplasia is detected and treated early with endoscopic therapies such as radiofrequency ablation (RFA) and/or endoscopic mucosal resection (EMR) (⁴⁻⁶). Despite endoscopic surveillance programs, the increasing incidence of EAC continues to remain a health concern (⁷). Accurate tests are needed to identify BE patients who are at high risk for progression and require therapeutic intervention as well as to recognize low risk BE patients who can potentially reduce the frequency of surveillance.

Current practice guidelines recommend endoscopic surveillance with biopsies at frequencies determined by the histologic diagnosis (⁸). However, the histologic evaluation is limited by inter-observer variation and random sampling (⁹⁻¹²). Furthermore, the molecular and cellular changes associated with malignant transformation can precede the morphologic changes that pathologists can evaluate by histology (¹³). Efforts have long been underway to identify risk prediction biomarkers in BE. This concept has become more important with the advent of highly effective endoscopic therapies such as RFA and EMR. Many biomarkers have been evaluated in BE (¹⁴⁻¹⁷) but risk prediction biomarkers have been difficult to identify. The British Society of Gastroenterology (BSG) recommends use of p53 immunohistochemistry (IHC) to aid diagnosis (¹⁸), however, biomarkers for accurate risk prediction have not been validated to date. The complex structure of tissues highlights the need for an alternative systems biology approach (¹⁹). Assessment of tissues as a “system” has the potential to improve upon current tools by quantifying genetic, immunologic, vascular and morphologic features relevant to patient outcomes. This tissue systems pathology approach has been demonstrated to have potential diagnostic applications in BE (²⁰). This approach may also have prognostic applications by objectively quantifying multiple molecular and cellular features that precede definitive morphologic changes. The aim of this study was to develop and validate a tissue systems pathology test that predicts future risk of progressing to HGD/EAC in patients with BE.

Materials and Methods

Study Design and Patients

A nested case-control study was constructed that utilized a multi-center cohort of BE patients in surveillance programs with clinical outcome data from four institutions. BE cases with a diagnosis of non-dysplasia (ND), indefinite for dysplasia (IND) or low-grade dysplasia (LGD) were retrieved from Geisinger Health System, University of Pittsburgh, University of Pennsylvania and Academic Medical Center (AMC), Amsterdam, Netherlands. The diagnoses were confirmed by a single gastrointestinal (GI) subspecialist pathologist at each US institution (J.M.D., N.C.J., J.L). Inclusion criteria were availability of tissue and clinicopathologic data and confirmation of BE. Exclusion criteria were history of HGD/EAC, diagnosis of HGD/EAC in less than 1 year, insufficient tissue quality as assessed by a pathologist, and preparation of tissue with Bouin's fixative or methylene blue. The earliest surveillance biopsy that satisfied inclusion/exclusion criteria was selected for each patient. Additional information can be found in Supplementary Methods. Patients who progressed to HGD/EAC in 1 year (incident progressors/cases, n=41 training, n=38 validation) were matched to non-progressor controls with median HGD/EAC-free surveillance time of 5.6 years (n=142 training, n=145 validation) based on gender, segment length and age where possible, and also by location (i.e. US cases matched to US controls and Netherlands cases matched to Netherlands controls). Case-control sets from the US and Netherlands institutions were randomly assigned to training or validation sets (Table 1). In the training set 33/41 progressor patients progressed to HGD and 8/41 to EAC and in the separate validation set, 29/38 progressor patients progressed to HGD and 9/38 to EAC (Supplementary Table S1). Data elements collected were: case collection date, original pathologic diagnosis provided by a generalist pathologist and GI subspecialist review diagnosis for the case tested, date and original diagnosis of every surveillance biopsy, progression endpoint (HGD/EAC), HGD/EAC-free surveillance time (time between case tested and HGD/EAC diagnosis or last follow-up), age, sex, race and segment length (short 3cm, long >3cm). The study was approved by the institutional review boards at each institution.

Candidate Biomarker Selection

The following candidate panel of 14 protein biomarkers was selected and examined in the study: K20, Ki-67, β -catenin, p16INK4a, AMACR, p53, HER2/neu, CDX-2, CD68, NF- κ B-p65, COX-2, HIF-1 α , CD45RO, and CD1a. The biomarkers included markers of epithelial cell abnormalities described in the progression of BE and also stromal biomarkers known to play a role in carcinogenesis (^{14, 21-33}).

Fluorescence Immunolabeling

5 μ m sections of formalin-fixed paraffin-embedded (FFPE) BE biopsies were stained with H&E by standard methods. Additional sections were labeled by multiplexed immunofluorescence for the candidate biomarker panel, plus Hoechst, according to previously described methods (²⁰). The biomarkers were multiplexed in sub-panels consisting of Hoechst and 3 biomarkers/slide detected via Alexa Fluor-488, -555 and -647-conjugated secondary antibodies (Life Technologies, Carlsbad, CA).

Whole Slide Imaging

H&E-stained slides were imaged at 20× magnification on a NanoZoomer Digital Pathology scanner (Hamamatsu Photonics, K.K., Japan). Fluorescently-immunolabeled slides were imaged at 20× on a ScanScope FL (Aperio Technologies/Leica BioSystems, Vista, CA) with a standardized imaging procedure as previously described (²⁰).

Image Analysis

Whole slide fluorescence images were analyzed using the TissueCypher™ Image Analysis Platform (Cernostics, Inc., Pittsburgh, PA), which utilizes algorithms for collection of quantitative biomarker and morphology feature data at the cellular and sub-cellular level, and within tissue compartments such as epithelium, metaplasia and lamina propria. Features (continuous, quantitative measurements of biomarkers and/or morphology) included biomarker intensities and coexpression within subcellular and tissue compartments, morphometrics and microenvironment-based measurements, as described previously (²⁰). 1,184 features/biopsy were extracted from the biomarkers and morphology by the software, and summarized as multiple measures, (percentiles, IQR, percent positive, spatial summary statistics) resulting in 13,538 feature/measures per biopsy. The image analysis software was blind to case-control status.

Statistical Analyses

A risk prediction classifier was developed within the training set and prospectively defined prior to testing in the validation set. We tested the hypotheses that patients in the predicted low-risk class have significantly lower risk for progression to HGD/EAC than patients in the predicted high-risk class, and that the risk classes would add independent prognostic information beyond that of the pathologic diagnosis and segment length.

Development of Risk Prediction Model—Univariate conditional logistic regression (CLR) was performed in the training set with the 13,538 feature/measures to compare non-progressors to progressors and enable feature selection for multivariable model building. Selected features were combined into classifiers and leave-one-out cross validation (LOOCV) was performed to estimate prognostic performance of the classifiers. In each iteration of LOOCV, 1 case-control group (progressor and matched non-progressors) was set aside and the remaining case-control groups were used as the training set. The feature/measures were selected and the prediction model was built in the training set by the sum of the features weighted by the univariate Cox coefficients, and then this model was applied on the testing cases to calculate a score. A univariate model was utilized since a multivariate model with many independent variables can suffer from overfitting and thus often capture spurious relationships among independent variables, which may lead to decreased performance when the prediction model is tested in separate cohorts (³⁴). The LOOCV process was repeated until all case-control groups were treated as the testing case once. The end result of LOOCV was a risk score for each patient ranging from 0-10. Survival time for Cox proportional hazards regression was defined as the time between the case tested and the diagnosis of HGD/EAC for progressors or last follow-up for non-progressors. Cox regression was only used after the features had been selected (by CLR) in order to derive the

weights for the selected features to compute a risk score for the prediction model. Concordance-indices (C-indices) were calculated and Receiver Operating Characteristic (ROC) curves based on the binary outcome of low/high for 5-year risk of progression were plotted. Cutoffs were determined to stratify patients into low-, intermediate-and high-risk classes. Two cutoffs were chosen in a sequential manner. The first cut-point for separating the low-risk group from the rest was chosen, corresponding to a 95% negative predictive value (NPV) unadjusted for disease prevalence to ensure that the patients scored low-risk had very low risk for progression. The second cut-point for separating the high-risk group from the rest was chosen, corresponding to a 65% positive predictive value (PPV), unadjusted for disease prevalence, to ensure that the majority of the progressors would be captured by the high-risk group. Kaplan-Meier (KM) curves were used to represent the probability of progression in the 3 risk classes. Hazard ratios (HRs) with 95% confidence intervals (C.I.) were calculated from Cox regression and odds ratios with 95% C.I. were calculated from the CLR. Log-rank test was used to assess the equality of probability of progression curves of the risk groups from KM analysis, while score test was performed with CLR to examine the significance of association of the risk groups with incidence of HGD/EAC.

Validation of Risk Prediction Model—The separate validation set was quarantined during the training phase. Sample size calculations indicated that a total of 43 patients were required in the validation set to ensure 80% power to detect a significant difference of 50% in the 5-year risk of progression to HGD/EAC between those classified as high-risk vs low-risk (as was observed in the training set), at 0.05 significance level. All assay parameters were pre-specified during the training phase and locked down prior to testing in the validation set. The risk score for each patient was calculated and risk classes were assigned based on the pre-specified cutoffs. Prevalence-adjusted NPV and PPV were calculated for the low- and high-risk groups based on previously reported progression rates⁽³⁵⁾. The intermediate-risk group was not included in calculations of NPV and PPV.

Comparison of Risk Prediction Model versus Clinical Variables—Multivariate Cox models were performed to assess whether the test would add independent prognostic information beyond traditional clinical factors. Diagnosis ((LGD versus ND/IND combined, sex (0=F, 1=M), segment length (0= short, 1=long) were dichotomized. Percent cells overexpressing p53 (determined by image analysis software, as described previously⁽²⁰⁾) and age were evaluated as continuous variables. Patients with missing segment length and/or original generalist diagnosis were excluded from the multivariate Cox models.

Results

Patients

The nested case-control cohort included pre-progression samples from 79 (n=41 in training, n=38 in validation) BE patients with ND, IND or LGD who progressed to HGD or EAC at least 1 year later and 287 samples from matched control patients who did not show progression (n=142 in training, n=145 in validation), as summarized in Table 1. Case-control sets from the US and European institutions were randomly assigned to the training or

validation set. The non-progressor patients had median HGD/EAC-free surveillance time of 5.9 (IQR 4.5, 8.2) and 5.5 years (IQR 4.1, 8.5) in the training and validation sets, respectively. The median time-to-progression was 2.9 (IQR 2.3, 3.7) and 2.8 (IQR 2.0, 4.2) years in the training and validation sets, respectively.

Development of 15-Feature Classifier in the Training Set

Slide images of multiplexed biomarker labeling were analyzed by the image analysis software to generate 13,538 feature/measures per biopsy in the training set. A set of 17 image analysis features derived from p53, HER2, Ki-67, K20, COX2, CD68, HIF-1 α , p16INK4A, AMACR, CD45RO and nuclear morphology were selected based on p-values from univariate CLR comparing cases versus controls (Table 2). Features derived from CDX2, β -catenin, CD1a and NF- κ B-p65 were not selected due to low ranking based on p-values. The false discovery rate for the 17 selected features was 0.025%. Therefore, the likelihood that a feature was selected by random chance was negligible. The top 3, 6, 9, 12, 15 and 17 feature/measures based on p-values from CLR were scaled using center and scale parameters derived from the training set. The weighted (by univariate Cox model coefficients) sum was calculated to produce a risk score. 98% of the raw scores ranged from -10 to 10. Scaling was performed as follows:

$$\text{score} = \begin{cases} 0 & \text{if raw score} < -10 \\ \frac{\text{raw score} + 10}{2} & \text{if } -10 < \text{raw score} < 10 \\ 10 & \text{if raw score} > 10 \end{cases}$$

LOOCV from feature/measure selection to predictive model building was conducted to evaluate the performance of the risk classifier in the training set. Using risk scores generated by these classifiers through LOOCV, C-indices for the top 3, 6, 9, 12, 15 and 17 features were 0.674, 0.672, 0.716, 0.755, 0.797 and 0.792, respectively, demonstrating that the top performing model was based on 15 feature/measures. The 15 feature/measures were derived from p53, HER2, K20, COX2, CD68, HIF-1 α , p16INK4A, AMACR, CD45RO and nuclear morphology and included multiple features derived from individual biomarkers (Table 2). The correlation among features derived from the same candidate biomarker is summarized in Supplementary Table S2, which demonstrates weak to moderate correlations except for two p53-related features showing relatively larger correlation (cor=0.82), but not to the extent of being collinear. A flowchart detailing the classifier development is shown in Supplementary Figure S1. Expression patterns of the biomarkers on which the classifier is based are shown in Supplementary Figure S2. AUROC for the classifier was 0.872 in patients from all four institutions (Fig. 1A), 0.842 in US patients and 0.870 in AMC patients, indicating high prognostic accuracy.

Two cutoffs were chosen to produce a 3-tier classifier. KM plots of the 5-year probability of progression to HGD/EAC in patients scored as low-, intermediate- and high-risk demonstrated that the classifier stratified progressors from non-progressors in all institutions combined and in US and AMC patients separately (Fig. 1B-D). HRs were 4.19 (95% C.I. 1.52, 11.57) for intermediate- vs. low-risk and 14.73 (95% C.I. 6.55, 33.16) for high- vs.

low-risk. Both the log-rank and score tests showed that the 3 predicted risk classes had different risk for progression to HGD/EAC ($p<0.0001$).

The molecular and cellular changes that are captured by the classifier are illustrated in Figure 2, which compares endoscopy images, H&E-stained biopsy images and images of multiplexed fluorescence labeling of the 9 protein candidate biomarkers from which the 15 features are derived in a progressor (Fig. 2A-C) and a non-progressor (Fig. 2D-F). Endoscopy images for both patients showed BE with no apparent lesions (Fig. 2A (progressor) and 2D (nonprogressor)). Biopsies from both patients were confirmed as ND by a GI subspecialist (Fig. 2B (progressor) and 2E (non-progressor)). The classifier scored the progressor high-risk due to multiple molecular and cellular changes (Fig. 2C). The non-progressor was scored low-risk due to absence of high-risk features (Fig. 2F).

In multivariate Cox models in which progression to HGD/EAC was evaluated first in relation to clinical variables alone, then in relation to the predicted risk classes added to the clinical variables, the intermediate-risk and high-risk classes provided prognostic power that was independent of the pathologist's diagnosis (generalist and subspecialist), segment length, age, sex and percent cells overexpressing p53 (Supplementary Table S3A-B). The magnitude of HRs indicated that the predicted risk classes provided stronger prognostic power than the clinical variables (Supplementary Table S3). Similar results were observed when the 15-feature risk score was evaluated as a continuous variable (Supplementary Table S4).

Performance of 15-Feature Classifier in the Separate Validation Set—All

parameters of the test were locked down prior to testing in the separate validation set. The prospectively defined test was then evaluated in the separate validation set of patients (Table 1). ROC analysis showed that the pre-specified 15-feature classifier predicted 5-year risk of progression to HGD/EAC with AUROCs of 0.804 in patients from all four institutions (Fig. 3A), 0.860 in US patients and 0.717 in AMC patients. In a post-hoc analysis in the validation set, the c-indices for the top 3-, 6-, 9-, 12- and 15-feature classifiers were 0.69, 0.659, 0.7, 0.758 and 0.772, respectively, confirming that the 15-feature classifier was the top performing risk prediction model in the separate validation set. KM analysis demonstrated that the classifier could distinguish progressors from non-progressors in the full validation set and in US and AMC patients separately (Fig. 3B-D), validating the classifier performance that was observed in the training set. HRs were 2.45 (95% C.I. 0.99, 6.07) for the comparison of the intermediate-risk versus low-risk group and 9.42 (95% C.I. 4.61, 19.24) (Fig. 3E), for high-risk versus low-risk ($p<0.0001$ for log-rank and score tests). 17 of the progressors scored high-risk, 7 scored intermediate-risk and 14 scored low risk in the separate validation set. 9 of the non-progressors scored high-risk, 21 scored intermediate-risk and 115 scored low-risk. The probability of progression to HGD/EAC by 5 years increased continuously as the 15-feature risk score increased (Fig. 3F). Prevalence-adjusted NPV and PPV were 0.98 and 0.26 using reported progression rates⁽³⁵⁾. The prevalence-adjusted proportions of patients scored low-, intermediate- and high-risk were 77%, 15% and 8%, respectively.

Multivariate Cox models evaluating a reduced model with clinical variables only and a full model with the 15-feature classifier added showed that the high-risk class provided prognostic power that was independent of the general and GI subspecialist pathologist's diagnosis, segment length, age, gender and percentage of cells overexpressing p53 in the validation set (Table 3A). The GI subspecialist diagnosis showed prognostic power when evaluated alone; however, it was no longer statistically significant when the predicted risk classes were added to the Cox model (Table 3B). The risk classifier and the GI subspecialist diagnosis were correlated (chi-square $p=0.002$), indicating that both were prognostic. However, attenuation of the prognostic power of the GI subspecialist diagnosis whereas the risk classifier remained significant in the multivariate Cox model indicated that the risk classifier had stronger prognostic power. There was no strong correlation between the risk classifier and the other evaluated variables including segment length, age, percentage of cells overexpressing p53 (correlations = 0.22, -0.06 and 0.16, respectively), and gender (chi-square $p=0.36$). Similar results were observed when the risk score was evaluated as a continuous variable (Supplementary Table S5).

Discussion

Using a nested case-control study design we developed and validated a novel multivariable test that predicts future risk of progression to HGD/EAC in BE patients. The test produces a risk score that can be used as a predictor to estimate 5-year risk for progression to HGD/EAC. The test incorporates 3-tier stratification to classify patients as low-, intermediate- or high-risk for progression. The predicted high-risk group was at 9.4-fold increased risk of developing HGD/EAC compared to the low-risk group. Furthermore, the risk classes provided independent predictive information that outperformed traditional risk factors in this study, including the general pathologist diagnosis and the expert GI diagnosis. Importantly, the test demonstrated risk stratification that was independent of traditional clinical variables in a separate validation cohort of BE patients.

The test has the potential to complement histologic analysis and offer providers and patients an individualized risk score for progression. The 3-tier classifier identifies patients at very low risk of progression within 5 years, as demonstrated by the prevalence-adjusted NPV of 0.98 in the validation set. If further validated, this finding suggests that the frequency of surveillance in this patient group can potentially be extended to 5 years. The classifier also identifies patients at very high risk of progression, with prevalence-adjusted PPV estimated at 0.26, demonstrating high predictive performance considering the very low frequency of progression in BE. Current clinical guidelines recommend endoscopic ablative therapy for confirmed HGD⁽⁸⁾ and there is growing evidence to support ablative therapy for confirmed LGD^(36, 37). Validation of this risk prediction approach provides potential support to extend ablative therapy to BE patients with IND and ND with high-risk scores. As with any clinical test the predictive accuracy will depend on disease prevalence, which is likely to vary between different clinical settings. The approach described here is quantitative, objective and in this study outperformed the generalist and subspecialist diagnosis, which are both prone to inter-observer variation. While our approach would initially add cost, there is the compensatory potential to lower future costs due to reduced surveillance frequency in low-risk patients, and early treatment in high-risk patients.

The limitations of this study include the retrospective nature and the limited sample size. A larger prospective study would not have been feasible due to the very low prevalence of progression in BE. The study lacked central pathology review, although all cases were reviewed by a single GI subspecialist at each US institution. The cohort included patients in surveillance at multiple centers, which prevented standardization of biopsy fixation and storage protocols. However, the biopsies reflect routine BE samples. While digital pathology has gained traction, the use of imaging and image analysis in pathology laboratories remains limited. The approach described here could be deployed in a central reference clinical laboratory equipped with the necessary imaging instrumentation, software, and staff skilled in digital pathology. The study design included training on a set of patients from multiple institutions with the rationale that a multi-institutional set would be more representative of the BE patient population, and therefore a predictive model built in the training set would be more likely to validate in a separate testing set and more likely to be of utility in the general population. An alternative analysis approach would be to train the classifier in a set of patients from a single institution and validate it in an independent set from a different institution. Future studies will include independent validation in BE patient cohorts from institutions not included in the development and initial validation of the 15-feature risk score.

Many biomarkers have shown promise for risk prediction in BE (¹⁴⁻¹⁷), including the biomarkers evaluated in this study, however, no biomarkers for risk prediction have been validated or translated into practice to date. The challenges to risk prediction include genetic and non-genetic heterogeneity (³⁸) and the resulting need to assess multiple pathways of carcinogenesis. The role of both epithelial and stromal components in carcinogenesis suggest that a systems biology approach may overcome prior study limitations(¹⁹). While p53 IHC has diagnostic value (¹⁸), it is not sufficient as a single biomarker for risk prediction (^{15, 24}). The methods used to evaluate biomarkers have also hindered implementation. Traditional pathology methods are limited by the difficulties in managing multiple IHC tests on limited biopsies and the challenges of manually integrating multivariable data into a prognosis. The testing approach validated in this study aids pathology by objectively measuring multiple molecular and cellular abnormalities that can precede the morphologic changes. The risk score identifies high-risk BE patients as having loss of tumor suppression and cell cycle control, stromal angiogenesis, altered patterns of infiltrating lymphocytes, increased inflammation and morphology abnormalities, which are early indicators of progression. The classifier utilizes multiple image analysis features extracted from the same biomarker to capture multiple expression patterns (Table 2). For example, p53 is frequently mutated in BE and while some mutations lead to p53 protein accumulation, others lead to loss (³⁹). By assessing multiple different features, the classifier aims to quantify multiple patterns of biomarker abnormalities in a standardized, objective manner. The classifier also incorporates microenvironment-based features (²⁰) that capture localized abnormalities such as focal AMACR-overexpression and clusters of HIF-1 α -overexpressing cells. Various molecular approaches have also been applied to diagnostic and prognostic testing in BE (⁴⁰⁻⁴²). While these technologies have aided biomarker discovery and show promise in risk prediction, they have the disadvantage of requiring tissues to be digested, resulting in loss of morphology and spatial relationships, which may be relevant to

patient outcomes. Furthermore, specific tests using these genomic technologies have not been independently validated in BE.

In addition to the advantages of the technology platform, this study was strengthened by use of a diverse patient cohort from four institutions. An additional strength was the exclusion of patients with prevalent HGD/EAC, enabling development of a test that predicts incident progression. Furthermore, the test was validated on a separate set of BE patients. The assay can be performed on sections from FFPE blocks.

In summary, the tissue systems pathology approach validated in this study quantifies multiple epithelial and stromal processes and better predicted risk of progression to HGD/EAC in BE patients than clinical variables, including pathologic diagnosis in this study. This approach provides opportunity to improve upon current qualitative histology as a quantitative method to risk stratify patients with BE.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Georgiann Noll and Matthew Barley (Geisinger) and Xuan Mai Nguyen for technical assistance (Cernostics) and David Rimm, M.D., Ph.D. and Janet Warrington, Ph.D. for many helpful discussions and advice on the research study.

Financial Support: This work was supported by a grant from the Pennsylvania Department of Health Cure Program Grant, Research on Cancer Diagnostics or Therapeutics with Commercialization Potential RFA#10-07-03 (R.J. Critchley-Thorne, J.M. Davison, G.W. Falk, J.W. Prichard, Y. Zhang), a Qualifying Therapeutic Discovery Project Grant, Internal Revenue Service/Affordable Care Act 2010, (R.J. Critchley-Thorne) and by Cernostics, Inc. Partial support by NIH/NCI U54-CA163004 (A.K. Rustgi, G.W. Falk) and the Molecular Pathology and Imaging and Molecular Biology Core Facilities at University of Pennsylvania.

References

1. Wani S, Falk G, Hall M, Gaddam S, Wang A, Gupta N, et al. Patients with nondysplastic Barrett's esophagus have low risks for developing dysplasia or esophageal adenocarcinoma. *Clin Gastroenterol Hepatol*. 2011; 9:220–7. quiz e26. [PubMed: 21115133]
2. Wani S, Falk GW, Post J, Yerian L, Hall M, Wang A, et al. Risk factors for progression of low-grade dysplasia in patients with Barrett's esophagus. *Gastroenterology*. 2011; 141:1179–86. 1186 e1. [PubMed: 21723218]
3. Thota PN, Lee HJ, Goldblum JR, Liu X, Sanaka MR, Gohel T, et al. Risk stratification of patients with barrett's esophagus and low-grade dysplasia or indefinite for dysplasia. *Clin Gastroenterol Hepatol*. 2014; 13:459–465 e1. [PubMed: 25102445]
4. Shaheen NJ, Overholt BF, Sampliner RE, Wolfsen HC, Wang KK, Fleischer DE, et al. Durability of radiofrequency ablation in Barrett's esophagus with dysplasia. *Gastroenterology*. 2011; 141:460–8. [PubMed: 21679712]
5. Bulsiewicz WJ, Kim HP, Dellon ES, Cotton CC, Pasricha S, Madanick RD, et al. Safety and efficacy of endoscopic mucosal therapy with radiofrequency ablation for patients with neoplastic Barrett's esophagus. *Clin Gastroenterol Hepatol*. 2013; 11:636–42. [PubMed: 23103824]
6. Pasricha S, Bulsiewicz WJ, Hathorn KE, Komanduri S, Muthusamy VR, Rothstein RI, et al. Durability and Predictors of Successful Radiofrequency Ablation for Barrett's Esophagus. *Clin Gastroenterol Hepatol*. 2014
7. ACS. Cancer Facts & Figures 2014. American Cancer Society; 2014.

8. Spechler SJ, Sharma P, Souza RF, Inadomi JM, Shaheen NJ. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology*. 2011; 140:1084–91. [PubMed: 21376940]
9. Reid BJ, Haggitt RC, Rubin CE, Roth G, Surawicz CM, Van Belle G, et al. Observer variation in the diagnosis of dysplasia in Barrett's esophagus. *Hum Pathol*. 1988; 19:166–78. [PubMed: 3343032]
10. Montgomery E, Bronner MP, Goldblum JR, Greenson JK, Haber MM, Hart J, et al. Reproducibility of the diagnosis of dysplasia in Barrett esophagus: a reaffirmation. *Hum Pathol*. 2001; 32:368–78. [PubMed: 11331953]
11. Odze RD. Diagnosis and grading of dysplasia in Barrett's oesophagus. *J Clin Pathol*. 2006; 59:1029–38. [PubMed: 17021130]
12. Goldblum JR. Controversies in the diagnosis of Barrett esophagus and Barrett-related dysplasia: one pathologist's perspective. *Arch Pathol Lab Med*. 2010; 134:1479–84. [PubMed: 20923304]
13. Stachler MD, Taylor-Weiner A, Peng S, McKenna A, Agoston AT, Odze RD, et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet*. 2015; 47:1047–55. [PubMed: 26192918]
14. Barrett MT, Sanchez CA, Galipeau PC, Neshat K, Emond M, Reid BJ. Allelic loss of 9p21 and mutation of the CDKN2/p16 gene develop as early lesions during neoplastic progression in Barrett's esophagus. *Oncogene*. 1996; 13:1867–73. [PubMed: 8934532]
15. Bird-Lieberman EL, Dunn JM, Coleman HG, Lao-Sirieix P, Oukrif D, Moore CE, et al. Population-based study reveals new risk-stratification biomarker panel for Barrett's esophagus. *Gastroenterology*. 2012; 143:927–35 e3. [PubMed: 22771507]
16. Kastelein F, Biermann K, Steyerberg EW, Verheij J, Kalisvaart M, Looijenga LH, et al. Aberrant p53 protein expression is associated with an increased risk of neoplastic progression in patients with Barrett's oesophagus. *Gut*. 2012; 62:1676–83. [PubMed: 23256952]
17. Kastelein F, Biermann K, Steyerberg EW, Verheij J, Kalisvaart M, Looijenga LH, et al. Value of alpha-methylacyl-CoA racemase immunohistochemistry for predicting neoplastic progression in Barrett's oesophagus. *Histopathology*. 2013; 63:630–9. [PubMed: 24004067]
18. Fitzgerald RC, di Pietro M, Ragunath K, Ang Y, Kang JY, Watson P, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut*. 2014
19. Gough, A.; Lezon, T.; Faeder, J.; Chennubhotla, C.; Murphy, R.; Critchley-Thorne, R., et al. Content Analysis And Cellular And Tissue Systems Biology: A Bridge Between Cancer Cell Biology And Tissue-Based Diagnostics. In: Mendelsohn, J.; Howley, PM.; Israel, MA.; Gray, JM.; Thompson, CB., editors. *The Molecular Basis of Cancer*. 4th. New York: Elsevier; 2014.
20. Prichard JW, Davison JM, Campbell BB, Repa KA, Reese LM, Nguyen XM, et al. A Systems Biology Approach to Anatomic Pathology. *Journal of Pathology Informatics*. 2015; 6:48. [PubMed: 26430536]
21. Ormsby AH, Goldblum JR, Rice TW, Richter JE, Falk GW, Vaezi MF, et al. Cytokeratin subsets can reliably distinguish Barrett's esophagus from intestinal metaplasia of the stomach. *Hum Pathol*. 1999; 30:288–94. [PubMed: 10088547]
22. Moons LM, Bax DA, Kuipers EJ, Van Dekken H, Haringsma J, Van Vliet AH, et al. The homeodomain protein CDX2 is an early marker of Barrett's oesophagus. *J Clin Pathol*. 2004; 57:1063–8. [PubMed: 15452161]
23. Weston AP, Banerjee SK, Sharma P, Tran TM, Richards R, Cherian R. p53 protein overexpression in low grade dysplasia (LGD) in Barrett's esophagus: immunohistochemical marker predictive of progression. *Am J Gastroenterol*. 2001; 96:1355–62. [PubMed: 11374668]
24. Sikkema M, Kerkhof M, Steyerberg EW, Kusters JG, van Strien PM, Looman CW, et al. Aneuploidy and overexpression of Ki67 and p53 as markers for neoplastic progression in Barrett's esophagus: a case-control study. *Am J Gastroenterol*. 2009; 104:2673–80. [PubMed: 19638963]
25. van Dekken H, Hop WC, Tilanus HW, Haringsma J, van der Valk H, Wink JC, et al. Immunohistochemical evaluation of a panel of tumor cell markers during malignant progression in Barrett esophagus. *Am J Clin Pathol*. 2008; 130:745–53. [PubMed: 18854267]
26. Dorer R, Odze RD. AMACR immunostaining is useful in detecting dysplastic epithelium in Barrett's esophagus, ulcerative colitis, and Crohn's disease. *Am J Surg Pathol*. 2006; 30:871–7. [PubMed: 16819330]

27. Rossi E, Grisanti S, Villanacci V, Della Casa D, Cengia P, Missale G, et al. HER-2 overexpression/amplification in Barrett's oesophagus predicts early transition from dysplasia to adenocarcinoma: a clinico-pathologic study. *J Cell Mol Med*. 2009; 13:3826–33. [PubMed: 19292734]
28. Miyashita T, Tajima H, Shah FA, Oshima M, Makino I, Nakagawara H, et al. Impact of inflammation-metaplasia-adenocarcinoma sequence and inflammatory microenvironment in esophageal carcinogenesis using surgical rat models. *Ann Surg Oncol*. 2014; 21:2012–9. [PubMed: 24526548]
29. Rauser S, Langer R, Tschernitz S, Gais P, Jütting U, Feith M, et al. High number of CD45RO+ tumor infiltrating lymphocytes is an independent prognostic factor in nonmetastasized (stage I-IIA). esophageal adenocarcinoma *BMC Cancer*. 2010; 10
30. Cappello F, Rappa F, Anzalone R, La Rocca G, Zummo G. CD1a expression by Barrett's metaplasia of gastric type may help to predict its evolution towards cancer. *Br J Cancer*. 2005; 92:888–90. [PubMed: 15756258]
31. Griffiths EA, Pritchard SA, McGrath SM, Valentine HR, Price PM, Welch IM, et al. Increasing expression of hypoxia-inducible proteins in the Barrett's metaplasia-dysplasia-adenocarcinoma sequence. *Br J Cancer*. 2007; 96:1377–83. [PubMed: 17437013]
32. O'Riordan JM, Abdel-latif MM, Ravi N, McNamara D, Byrne PJ, McDonald GS, et al. Proinflammatory cytokine and nuclear factor kappa-B expression along the inflammation-metaplasia-dysplasia-adenocarcinoma sequence in the esophagus. *Am J Gastroenterol*. 2005; 100:1257–64. [PubMed: 15929754]
33. Lagorce C, Paraf F, Vidaud D, Couvelard A, Wendum D, Martin A, et al. Cyclooxygenase-2 is expressed frequently and early in Barrett's oesophagus and associated adenocarcinoma. *Histopathology*. 2003; 42:457–65. [PubMed: 12713622]
34. Harrell, F. *Multivariable Modeling Strategies Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer Series in Statistics. Switzerland: Springer International Publishing; 2015.
35. Duits LC, Phoa KN, Curvers WL, Ten Kate FJ, Meijer GA, Seldenrijk CA, 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; 64:700–6. [PubMed: 25034523]
36. Phoa KN, van Vilsteren FG, Weusten BL, Bisschops R, Schoon EJ, Ragunath K, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA*. 2014; 311:1209–17. [PubMed: 24668102]
37. Small AJ, Araujo JL, Leggett CL, Mendelson AH, Agarwalla A, Abrams JA, et al. Radiofrequency Ablation is Associated with Decreased Neoplastic Progression in Patients with Barrett's Esophagus and Confirmed Low-Grade Dysplasia. *Gastroenterology*. 2015
38. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013; 501:338–45. [PubMed: 24048066]
39. Kaye PV, Haider SA, James PD, Soomro I, Catton J, Parsons SL, et al. Novel staining pattern of p53 in Barrett's dysplasia--the absent pattern. *Histopathology*. 2010; 57:933–5. [PubMed: 21166706]
40. Jin Z, Cheng Y, Gu W, Zheng Y, Sato F, Mori Y, et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. *Cancer Res*. 2009; 69:4112–5. [PubMed: 19435894]
41. Eluri S, Brugge WR, Daglilar ES, Jackson SA, Styn MA, Callenberg KM, et al. The Presence of Genetic Mutations at Key Loci Predicts Progression to Esophageal Adenocarcinoma in Barrett's Esophagus. *Am J Gastroenterol*. 2015; 110:828–34. [PubMed: 26010308]
42. Del Portillo A, Lagana SM, Yao Y, Uehara T, Jhala N, Ganguly T, et al. Evaluation of Mutational Testing of Preneoplastic Barrett's Mucosa by Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Endoscopic Samples for Detection of Concurrent Dysplasia and Adenocarcinoma in Barrett's Esophagus. *J Mol Diagn*. 2015
43. Haralick RM, Shanmugam K, Dinstein I. Textural Feature for Image Classification. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, SMC-3*. 1973:610–621.

Abbreviations

AMACR	alpha-methylacyl-CoA racemase
AUROC	area under receiver operating characteristic curve
BE	Barrett's esophagus
C-index	concordance index
COX-2	cyclo-oxygenase-2
EAC	esophageal adenocarcinoma
EMR	endoscopic mucosal resection
HGD	high grade dysplasia
HIF-1α	hypoxia-inducible factor 1-alpha subunit
HR	hazard ratio
IND	indefinite for dysplasia
K20	cytokeratin-20
KM	Kaplan-Meier
LGD	low grade dysplasia
LOOCV	leave-one-out-cross-validation
ND	non-dysplastic
OR	odds ratio
RFA	radiofrequency ablation
ROC	receiver operating characteristic

Figure 1. Development and Performance of 15-Feature Risk Score in Training Set of BE Patients
A ROC curve for 15-feature risk score in training set of incident progressor and nonprogressor patients. **B, C and D:** KM analysis of probability of progression to HGD/EAC in patients scored low-, intermediate- and high-risk by the 15-feature risk classifier from all four institutions, the three US institutions and AMC, respectively. **E:** Univariate HRs and ORs with 95% C.I. for comparisons between risk groups.

Figure 2. Detection of High Risk Features that Precede Morphologic Changes in BE

Endoscopy, H&E and multiplexed fluorescence biomarker images are shown for an incident progressor (IP) (panels A-C) and a non-progressor (NP) (panels D-F) with GI subspecialist diagnosis of BE ND. The IP patient progressed to HGD 6.3 years later and was scored high-risk by the 15-feature classifier. The NP patient had 7.8 years of endoscopic surveillance showing no progression to HGD/EAC and was scored low-risk. **A:** Endoscopy image from IP showing BE without visible lesions; **B:** H&E-stained biopsy from IP showing ND; **C:** Biomarker patterns in ND biopsy from IP (upper-left fragment: p53-yellow, p16-green, AMACR-red, upper-right: HER2/neu-green, K20-red, lower-left: CD68-green, COX-2-red, lower-right: HIF-1 α -green, CD45RO-red; **D:** Endoscopy image from NP showing BE without visible lesions; **E:** H&E-stained biopsy from NP showing ND; **F:** Biomarker patterns in ND biopsy from NP showing absence of high-risk changes (upper-left: p53-yellow, p16-green, AMACR-red, upper-right: HER2/neu-green, K20-red, lower-left: CD68-green, COX-2-red, lower-right: HIF-1 α -green, CD45RO-red. Hoechst labeling shown in blue in panels C and F.

Figure 3. Validation of 15-Feature Risk Classifier in Separate Set of BE Patients
A ROC curve for 15-feature risk classifier in validation set. **B, C and D:** KM analysis of probability of progression to HGD/EAC in validation set patients scored low-, intermediate- and high-risk by the 15-feature risk classifier in patients from all four institutions, US institutions and AMC, respectively. **E:** HRs and ORs (95% C.I.) for comparisons between risk groups. **F:** 5-year progression rate as a continuous function of the risk score.

Table 1

trols

Training Set			Separate Validation Set									
Non-Progressors		Incident Progressors	Non-Progressors		Incident Progressors							
	142	41	145	38								
5.9 (4.5, 8.2)		2.9 (2.3, 3.7)	5.5 (4.1, 8.5)	2.8 (2.0, 4.2)								
56.5 ±11.6		60.9 ±12.2	61.0 ±12.1	60.1 ±11.3								
63 (44.4)		16 (39.0)	58 (40.0)	10 (26.3)								
71 (50)		24 (58.5)	73 (50.3)	27 (71.1)								
8 (5.6)		1 (2.4)	14 (9.7)	1 (2.6)								
119 (83.8)		32 (78)	114 (78.6)	33 (86.8)								
23 (16.2)		9 (22)	31 (21.4)	5 (13.2)								
IND	LGD	IND	ND	LGD	IND	ND	LGD	IND	ND	LGD	IND	LGD
3 (2.1)	5 (3.5)	1 (2.4)	26 (63.4)	14 (34.1)	138 (95.2)	2 (1.4)	5 (3.4)	31 (81.6)	2 (5.3)	5 (13.2)		

Gei	UPenn	UPitt	AMC	Gei	UPenn	UPitt	AMC	Gei	UPenn	UPitt	AMC	Gei	UPenn	UPitt
50	16 (11.3)	9 (6.3)	25 (61.0)	9 (22.0)	1 (2.4)	6 (14.6)	46 (31.7)	63 (43.5)	15 (10.3)	21 (14.5)	28 (73.7)	4 (10.5)	3 (7.9)	3 (7.9)
9, 8, 9	4.6 (3.2, 5.5)	11.6 (7.3, 12.0)	3.2 (2.9, 4.0)	2.2 (1.4, 3.0)	2.0 (N/A)	2.5 (1.6, 3.3)	5.0 (4.0, 7.8)	6.5 (4.2, 10.1)	4.1 (3.5, 4.8)	6.3 (5.6, 8.6)	2.8 (2.3, 4.3)	2.1 (1.4, 4.7)	1.7 (1.0, 3.9)	3.1 (1.5, 5.5)

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2017 June 01.

tested and last endoscopy with ND, IND or LGD (non-progressors) or endoscopy with diagnosis HGD or EAC (incident progressors). Diagnosis
ents. Original diagnosis provided by a generalist pathologist was available for 160 patients in the training set and 144 patients in the validation set. AMC:
nger Health System; UPenn: University of Pennsylvania; UPitt, University of Pittsburgh. S.D.: standard deviation, IQR: interquartile range.

Table 2
17 Candidate Image Analysis Features (Including Top 15 Features Utilized by Risk Classifier)

Biomarker	Image Analysis Feature	P value	P value Adjusted by Diagnosis	Coefficient	Utilized by 15-Feature Risk Classifier
p53	p53 nuclear sum intensity	3.81E-05	5.81E-05	-8.04439	
p53	p53 nuclear mean intensity	7.48E-05	0.000116779	6.358257	
HER2/neu and K20	Ratio of mean HER2/neu intensity:mean K20 intensity in nuclei clusters	0.000155084	0.000312461	4.547325	
HER2/neu and K20	Ratio of 95th quantile HER2/neu intensity:95th quantile K20 intensity in nuclei clusters	0.000318651	0.00064884	4.286031	
COX2 and CD68	Coexpression cellular COX2 mean intensity and cellular CD68 mean intensity	0.00046706	0.000911804	-0.02203	
p53	p53 mean intensity in nuclei clusters	0.000501393	0.000498535	3.099642	
p53, p16 and nuclear morphology (solidity)	Nuclear solidity in p53+ p16- cells	0.000866951	0.00126554	15.62477	
CD45RO	CD45RO plasma membrane sum intensity	0.000873155	0.002141764	-3.76449	
AMACR	AMACR microenvironment standard deviation	0.000924663	0.001502611	0.000789	
COX2	COX2 texture in cytoplasm*	0.001318862	0.001909437	10.39816	
HIF1alpha	HIF-1α microenvironment cell mean intensity	0.001758646	0.001040583	0.000349	
HIF1alpha	HIF-1α microenvironment cell moment (product of mean and standard deviation)	0.002189596	0.002420038	1.02E-06	
p16	p16 cytoplasm mean intensity	0.003522243	0.003240205	-4.98699	
p53, p16 and nuclear morphology (area)	Nuclear area in p53+ p16- cells	0.00386791	0.005517011	0.014368	
Nuclear morphology	Hoechst nuclear 95th quantile intensity	0.022822291	0.03429373	10.78732	
Ki-67 and K20	Ratio of 95th quantile Ki-67 intensity:95th quantile K20 intensity in nuclei clusters	0.279157	0.340456	0.95634	x
Ki-67 and K20	95th quantile Ki-67 intensity in nuclei of metaplastic cells	0.10297	0.053479	1.73172	x

Univariate conditional logistic regression was performed with the 13,538 feature/measures extracted by the image analysis to compare non-progressors (controls) versus incident progressors (cases) in the training set of BE patients. The table lists the selected subset of 17 features derived from 10 biomarkers and nuclear morphology that showed significant differences in incident progressors versus non-progressors. This set of 17 features was entered into the multivariable model building. P-values shown are estimated from the conditional logistic regression. Coefficients were derived from Cox proportional hazards regression of each feature/measure.

* Contrast textural feature is extracted from a co-occurrence matrix and is a measure of the COX2 intensity contrast between a pixel and its neighbor over the whole tissue image, as described by Haralick et al (43).

Table 3
Prognostic Performance of Risk Classes vs. Clinical Variables in Validation Set of BE Patients

A. Prognostic Performance of Risk Classes vs. Clinical Variables*		
Variable	Hazard Ratio (95% CI)	P Value
Analysis without Risk Prediction Test		
General Pathologist's Dx (LGD vs. ND/IND)	1.55 (0.67-3.58)	0.31
Segment length (Long vs. Short)	2.53 (1.00-6.42)	0.05
Age	0.99 (0.96-1.02)	0.38
Gender	1.47 (0.51-4.29)	0.48
p53 (% cells overexpressing) [‡]	6.87 (0.01-4755.13)	0.56
Analysis with Risk Prediction Test		
General Pathologist's Dx (LGD vs. ND/IND)	1.27 (0.53-3.01)	0.59
Segment length (Long vs. Short)	1.91 (0.75-4.87)	0.17
Age	0.99 (0.96-1.02)	0.4
Gender	1.01 (0.34-3.05)	0.98
p53 (% cells overexpressing)	0.6 (0-728.87)	0.89
Risk Classes (predicted by the test)		
<i>Intermediate vs. Low Risk</i>	2.11 (0.66-6.7)	0.21
<i>High vs. Low Risk</i>	7.27 (3.2-16.49)	<0.0001
B. Prognostic Performance of Risk Classes vs. GI Subspecialist**		
Variable	Hazard Ratio (95% CI)	P Value
Analysis without Risk Prediction Test		
GI Subspecialist Pathologist's Dx (LGD vs. ND/IND)	3.19 (1.24-8.2)	0.02
Analysis with Risk Prediction Test		
GI Subspecialist Pathologist's Dx (LGD vs. ND/IND)	1.33 (0.5-3.53)	0.57
Risk Classes (predicted by the test)		
<i>Intermediate vs. Low Risk</i>	2.37 (0.95-5.93)	0.07
<i>High vs. Low Risk</i>	8.95 (4.27-18.77)	<0.0001

Multivariate Cox models were run in which progression to HGD/EAC was evaluated first in relation to clinical variables alone, then in relation to risk classes and clinical variables in non-progressor patients and incident progressor patients in the validation set. Pathologist diagnosis, gender, segment length, and Risk Classes were dichotomized as described in Methods. Age and p53 were evaluated as continuous variables. Mann-Whitney tests showed no statistically significant difference between age, gender or segment length in progressors versus non-progressors (p=0.72, 0.36, 0.09, respectively).

* n=30 incident progressor patients and n=103 non-progressor patients with complete data for all evaluated variables.

** n=38 incident progressor patients and n=145 non-progressor patients (all validation set patients) for analysis in part B.

[‡]calculated by the image analysis software as described in Methods.