**Supplemental Information**

**Supplemental Background and Methods**

**TissueCypher Barrett’s Esophagus Assay: Previous Development and Validation Studies**

The risk prediction test was developed and validated in a nested case-control study (n=366) with patients who progressed from ND, indefinite for dysplasia (IND) or LGD to HGD/EAC in ≥1 year, and patients with ND, IND or LGD with surveillance data showing no progression (1). A multivariable risk classifier was developed in a training set, all of the test parameters were locked, and then the test was evaluated in a separate validation set. The classifier stratified patients into low-, intermediate- and high-risk classes (hazard ratio (HR), 9.42; 95% confidence interval (C.I.), 4.6–19.24 (high-risk vs low-risk); p<0.0001), and also provided independent predictive information that outperformed predictions based on clinical variables, including subspecialist pathologist diagnosis. This first validation study demonstrated the ability of the test to predict incident progression in BE. In addition to predicting future risk of progression in BE, there is also a clinical need to identify prevalent HGD/EAC that may be missed by random sampling in BE. In a second study, the same risk prediction test (with all parameters locked) was evaluated in ND, IND and LGD biopsies from BE patients diagnosed with HGD/EAC on repeat endoscopy (prevalent cases, n=30, median time to HGD/EAC diagnosis 140.5 days) and non-progressors (n=145) (2). The test stratified prevalent cases and non-progressor patients into risk classes (odds ratio (OR), 46.0; 95% C.I., 14.86-169 (high-risk vs low-risk); p<0.0001), and also provided independent predictive information that outperformed the subspecialist and generalist diagnosis. The results of this second validation study indicated that the risk prediction test may provide an objective method to identify prevalent HGD/EAC and facilitate earlier identification of BE patients requiring therapeutic intervention.

The continuous risk score (ranges from 0 to 10) and categorical risk class (low, intermediate, or high) are calculated from the scaled and coefficient-weighted sum of 15 quantitative image analysis measurements (features) derived from 9 protein-based biomarkers and morphology as follows:

1. **Multiplexed immunofluorescence slide labeling** - serial sections of FFPE Barrett's esophagus biopsies are fluorescently immunolabeled for p16, AMACR, p53, HER2, K20, CD68, COX-2, HIF-1, and CD45RO, plus Hoechst according to a standardized operating procedure, as previously described (3);
2. **Whole slide fluorescence digital imaging** - labeled slides are imaged by whole slide fluorescence scanning that generates image data on each biomarker and nuclei according to a standardized operating procedure, as previously described (3);
3. **Automated image analysis** - Tissue images are analyzed by the TissueCypher® Image Analysis Platform (Cernostics, Pittsburgh, PA), which utilizes automated image analysis software to identify tissue fragments, then identify cell and subcellular objects and tissue structure objects (as described in detail previously (3)) from which 15 features are extracted. Features are quantitative measures of biomarkers, including protein-based biomarkers and also nuclear morphology. The 15 features are p53 nuclear sum intensity, p53 nuclear mean intensity, ratio of mean HER2/neu intensity:mean K20 intensity in nuclei clusters, ratio of 95th quantile HER2/neu intensity:95th quantile K20 intensity in nuclei clusters, coexpression of cellular COX2 mean intensity and cellular CD68 mean intensity, p53 mean intensity in nuclei clusters, nuclear solidity in p53+ p16- cells, CD45RO plasma membrane sum intensity, AMACR microenvironment standard deviation, COX2 texture in cytoplasm, HIF-1 microenvironment cell mean intensity, HIF-1 microenvironment cell moment (product of mean and standard deviation), p16 cytoplasm mean intensity, nuclear area in p53+ p16- cells, Hoechst nuclear intensity, as previously described in detail (1-3).
4. **Risk classification -** The 15 features are scaled using center and scale parameters for each feature defined in a training study, then weighted by coefficients derived from univariate Cox regression analysis of each feature and progression outcomes in the training study (1). The weighted sum of the 15 scaled features produces an unscaled risk score, which is scaled as follows:


Cutoff values are applied to the scaled risk score to classify patients for risk of progression: Risk class low if scaled score 0 to <5.5, risk class intermediate if scaled score ≥5.5 to <6.4, and risk class high if scaled score ≥6.4 to 10. These cutoffs were selected in the training study (1), and were chosen in a sequential manner. The first cut-point for separating the low risk class was selected to produce a negative predictive value of 95% (unadjusted for disease prevalence) to ensure very low rate of progression to HGD/EAC within 5 years in patients classified as low risk. The second cut-point for stratifying the high risk class was selected to produce a positive predictive value of 65% (unadjusted for disease prevalence) with the goal of identifying the majority of the patients who progressed to HGD/EAC within 5 years as high risk.

**Prevalence Adjustments:**

Positive predictive values (PPV), negative predictive value (NPV) and proportions of patients scoring low, intermediate and high risk by the risk prediction test were adjusted for prevalence due to the case-control nature of this study. It was estimated that 5.07% of patients with ND, IND or LGD at baseline would progress to HGD/EAC within 5 years in the studied patient population from which the nested case-control set was selected. The proportions of patients who progressed by 5 years were taken from the Kaplan-Meier curves for the high, intermediate and low risk groups (0.591, 0.240 and 0.153, respectively), and used to calculate the numbers of progressors and non-progressors in each of the risk groups at 5 years, i.e. 16.0 progressors and 11.0 non-progressors in the high risk group, 6.2 progressors and 19.8 non-progressors in the intermediate risk group, and 32.9 progressors and 182.1 non-progressors in the low risk group at 5 years. There were 55.1 progression events by 5 years, and 212.9 patients who had not progressed by that timepoint, which allows calculation of sensitivity and specificity of the test for 5 year risk prediction; sensitivity=29% and specificity=86%, the proportions of patients scoring low, intermediate and high risk (10%, 10%, and 80%, respectively), and unadjusted PPV=59.1%, and unadjusted NPV=84.7%. At a 5.07% 5-year prevalence of progression to HGD/EAC, there would be an estimated 13.6 progression events by 5 years within a group of 268 patients. Assuming that sensitivity and specificity of the test are not affected by prevalence, the adjusted PPV=23%, adjusted NPV=96.4%, and adjusted populations of patients scoring low, intermediate and high risk are 84.2%, 9.4% and 6.4%, respectively.

**Supplemental References**

1. Critchley-Thorne RJ, Duits LC, Prichard JW, et al. A Tissue Systems Pathology Assay for High-Risk Barrett's Esophagus. Cancer Epidemiol Biomarkers Prev 2016;25:958-68.

2. Critchley-Thorne RJ, Davison JM, Prichard JW, et al. A Tissue Systems Pathology Test Detects Abnormalities Associated with Prevalent High-Grade Dysplasia and Esophageal Cancer in Barrett's Esophagus. Cancer Epidemiol Biomarkers Prev 2017;26:240-248.

3. Prichard JW, Davison JM, Campbell BB, et al. TissueCypher: A Systems Biology Approach to Anatomic Pathology. Journal of Pathology Informatics 2015;6:48.

**Figure S1**

****

**Figure S1. Risk Prediction Test Compared to Quantitative Analysis of p53**. **Panel A:** KM analysis of probability of progression to HGD/EAC in BE patients with expert diagnosis ND, IND or LGD stratified into low-, intermediate-, and high-risk by the risk prediction test (58 IP and 210 NP); **Panel B:** KM analysis of probability of progression to HGD/EAC in BE patients stratified into low and high groups according to p53 nuclear sum intensity (cut-point was selected to score 27 patients p53-high, i.e. same number as scored high-risk by the risk prediction test); **Panel C:** KM analysis of probability of progression to HGD/EAC in BE patients with expert diagnosis ND stratified into low-, intermediate-, and high-risk by the risk prediction test (43 IP and 184 NP); **Panel D:** KM analysis of probability of progression to HGD/EAC in BE patients with expert diagnosis ND stratified into low and high groups according to p53 nuclear sum intensity.

**Figure S2**



**Figure S2. Evaluation of Risk Prediction Test Performance in Multi-Level Cases with Low Risk Results on One Biopsy Level.** One biopsy level was randomly selected from patients with risk prediction test results from multiple, separate biopsy levels. For patients whose randomly-selected biopsy level scored low-risk, the highest scoring additional biopsy level was assessed, and the risk stratification performance of the test was analyzed by KM to assess the probability of progression to HGD/EAC in BE patients scored low-, intermediate-, and high-risk by the test. **Panel A:** Results from the first round of random selection of a biopsy level, where patients with a randomly-selected biopsy scoring low-risk are scored based on the highest score from other spatial biopsies taken at the same endoscopy. **Panel B:** Results from the second round of random selection using the process described above. n=268; 58 incident progressors and 210 non-progressors in both panels.

|  |
| --- |
| **Table S1. Predictive Performance of Risk Prediction Test versus Clinical Variables** |
| *Multivariable Analysis of Clinical Variables (with diagnosis from expert pathologist)* |
| **Variable** | **HR (95% C.I.)** | **p value** |
| Age | 1.03 (1.00,1.05) | 0.0482 |
| Sex (M vs. F) | 3.00 (1.19, 7.58) | 0.0204 |
| Expert Dx IND vs. ND | 1.14 (0.40, 3.21) | 0.8046 |
| Expert Dx LGD vs. ND | 3.56 (1.75, 7.25) | 0.0005 |
| Segment length (L vs. S) | 1.80 (1.01, 3.22) | 0.0470 |
| Hiatal hernia (Present vs. Absent) | 0.87 (0.43, 1.75) | 0.6907 |
|  |
| *Multivariable Analysis of Clinical Variables and Risk Classes Produced by the Risk Prediction Test*  |
| **Variable** | **HR (95% C.I.)** | **p value** |
| Age  | 1.02 (1.00,1.05) | 0.0793 |
| Sex (M vs. F) | 3.04 (1.19, 7.74) | 0.0198 |
| Expert Dx IND vs. ND | 0.93 (0.33, 2.66) | 0.8959 |
| Expert Dx LGD vs. ND | 1.52 (0.62, 3.76) | 0.3636 |
| Segment length (L vs. S) | 1.64 (0.91, 2.97) | 0.0981 |
| Hiatal hernia (Present vs. Absent) | 0.79 (0.39, 1.61) | 0.5161 |
| Predicted Risk Classes: Int-Risk vs. Low-Risk | 1.58 (0.71, 3.52) | 0.2597 |
| Predicted Risk Classes: High-Risk vs. Low-Risk | 3.74 (1.64, 8.54) | 0.0017 |
| Note: Multivariable Cox models were run in which progression to HGD/EAC was evaluated first in relation to clinical variables alone, then in relation to the risk classes predicted by the test and clinical variables, as described in Materials and Methods. |

|  |
| --- |
| **Table S2. Predictive Performance of Risk Prediction Test versus Clinical Variables and p53** |
| *Multivariable Analysis of Clinical Variables, p53 (continuous variable)* *and Risk Classes Produced by the Risk Prediction Test* |
| **Variable** | **HR (95% C.I.)** | **p value** |
| Age | 1.0249 (0.9987, 1.052) | 0.0622 |
| Sex (M vs. F) | 3.1711 (1.241, 8.104) | 0.0159 |
| Segment length (S vs. L) | 0.6109 (0.3364, 1.109) | 0.1055 |
| Hiatal hernia (Present vs. Absent) | 0.7824 (0.3848, 1.591) | 0.4978 |
| p53 sum intensity in nuclei1 | 2.5425 (2.3510E-05, 2.7490E+05) | 0.8746 |
| Predicted Risk Classes: Int-Risk vs. Low-Risk | 1.6125 (0.7293, 3.565) | 0.2379 |
| Predicted Risk Classes: High-Risk vs. Low-Risk | 4.6255 (2.418, 8.85) | <0.0001 |
|  |
| *Multivariable Analysis of Clinical Variables, p53 (categorical variable)* *and Risk Classes Produced by the Risk Prediction Test* |
| **Variable** | **HR (95% C.I.)** | **p value** |
| Age | 1.0267  (1.0009 ,  1.053) | 0.0428 |
| Sex (M vs. F) |  3.5223  (1.3704 ,  9.053) | 0.00895 |
| Segment length (S vs. L) |  0.5608  (0.3076 ,  1.022) | 0.05899 |
| Hiatal hernia (Present vs. Absent) | 0.7746  (0.3811 ,  1.574) | 0.48031  |
| p53 sum intensity in nuclei2 | 2.1749  (0.9931 ,  4.763) | 0.05207 |
| Predicted Risk Classes: Int-Risk vs. Low-Risk |  1.5974  (0.7210 ,  3.539) | 0.24854  |
| Predicted Risk Classes: High-Risk vs. Low-Risk | 4.4284  (2.3191 ,  8.456) | <0.0001 |
|  |  |  |
| Note: Multivariable Cox models were run in which progression to HGD/EAC was evaluated first in relation to clinical variables alone, then in relation to the risk classes predicted by the test and clinical variables, as described in Materials and Methods. p53 sum intensity in nuclei was calculated by the TissueCypher Image Analysis Platform. 1p53 feature assessed as a continuous variable; 2p53 feature assessed as a categorical variable.  |

|  |
| --- |
| **Table S3. Predictive Performance of Continuous Risk Score versus GI Subspecialist/Expert Pathologist Review Diagnosis** |
| *Multivariable Analysis of Subspecialist Dx* |
| **Variable** | **HR (95% C.I.)** | **p value** |
| Expert IND vs. ND | 1.20 (0.47, 3.02) | 0.7034 |
| Expert LGD vs. ND | 3.79 (1.90, 7.56) | 0.0002 |
|  |
| *Multivariable Analysis of Subspecialist Dx and Continuous Risk Score Produced by the Risk Prediction Test* |
| **Variable** | **HR (95% C.I.)** | **p value** |
| Expert Dx IND vs. ND | 1.02 (0.40, 2.59) | 0.9642 |
| Expert Dx LGD vs. ND | 2.16 (1.00, 4.68) | 0.0503 |
| TissueCypher Risk Score | 1.28 (1.09, 1.49) | 0.0020 |

|  |
| --- |
| Note: Multivariable Cox models were run in which progression to HGD/EAC was evaluated first in relation to GI subspecialist (expert) diagnosis (Dx) alone, then in relation to the continuous risk score produced by the test and expert Dx, as described in Materials and Methods. |

|  |
| --- |
| **Table S4. Risk Scores in Biopsies from Separate Endoscopic Levels** |
| **Site** | **Mean standard deviation of risk scores** | **Concordance in risk classes** | **Patients with biopsies from** **multiple endoscopic levels** |
| **All****patients** | **Incident Progressors** | **Non-****Progressors** |
| **Cleveland Clinic** | 1.32 | 72% score same risk class26% score across 2 risk classes2% score across 3 risk classes | 47% | 56% | 45% |
| **University of Pittsburgh** | 1.51 | 75% score same risk class22% score across 2 risk classes3% score across 3 risk classes | 27% | 29% | 26% |