

Geisinger Clinic

Annual Progress Report: 2011 Nonformula Grant

Reporting Period

July 1, 2013 – June 30, 2014

Nonformula Grant Overview

The Geisinger Clinic received \$1,000,000 in nonformula funds for the grant award period June 1, 2012 through August 29, 2014. Accomplishments for the reporting period are described below.

Research Project: Project Title and Purpose

Diagnostic-Prognostic Testing in Patients at High Risk for Esophageal Cancer – The purpose of this project is to clinically validate a diagnostic-prognostic test for esophageal cancer, which will accurately diagnose at a premalignant stage and predict which patients are at high risk for esophageal cancer to enable early, preventative therapy. A prototype test has been developed and proof-of-concept of the testing technology has been established in collaborative work by Geisinger and Cernostics. The project aims to perform clinical validation studies in a training cohort and two independent validation cohorts of esophageal biopsies with clinical outcome data from Geisinger, University of Pittsburgh and University of Pennsylvania to select diagnostic and prognostic classifiers and to establish the sensitivity, specificity and positive and negative predictive values of the diagnostic-prognostic test for patients at high risk for esophageal cancer.

Anticipated Duration of Project

6/1/2012 – 8/29/2014

Project Overview

The broad objective of the research is to clinically validate a diagnostic and prognostic test that accurately assigns diagnosis and predicts risk of developing esophageal cancer. The test is a spatial systems biology-based approach to anatomic pathologic testing. The test employs multiplexed fluorescence labeling of tumor system biomarkers, including malignant, immune and stromal processes in anatomic pathology specimens with digital imaging and image analysis to quantify biomarker expression and spatial relationships between biomarkers in the context of tissue morphology. This is coupled to classifier software to integrate biomarker data with morphology data and clinical data to produce diagnostic and prognostic scores. These scores will be used to accurately diagnose and predict the risk of developing esophageal cancer in individual patients to enable early treatment. A prototype test has been collaboratively developed by Geisinger (lead applicant) and Cernostics, Inc. (small business collaborator) as a proof-of-concept. As a next step, a consortium of investigators will perform retrospective clinical

validation studies of the test towards the long term goal of commercializing the test via a CLIA-certified laboratory. The test will be performed first in a training cohort of formalin-fixed paraffin-embedded esophageal biopsies with clinical data from Geisinger using Cernostics' spatial systems biology technology, and diagnostic and prognostic classifiers will be developed. The test, including the classifiers, will then be performed in two independent validation patient cohorts from the University of Pittsburgh and the University of Pennsylvania to determine specificity, sensitivity and positive and negative predictive values of the diagnostic-prognostic test. The specific research aims are, 1) Determine the performance of the prototype test in stratifying patients according to diagnosis and predicting risk for esophageal cancer in a retrospective training patient cohort; and 2) Validate the diagnostic and prognostic performance of the optimized diagnostic-prognostic test in two independent retrospective patient cohorts. The training and validation cohorts represent both urban and rural populations and are designed to reach the maximum number of the underserved and will ensure a significant statewide impact on the health of Pennsylvanians. Paralleling the proposed project, Cernostics and Geisinger will perform further analytical validation studies on the test. The test will be commercialized by Cernostics and will be offered as a service to pathologists and gastroenterologists to guide individualized patient management to help prevent the development of esophageal cancer.

Principal Investigator

Jeffrey W. Prichard, DO
Director of Surgical Pathology
Geisinger Medical Laboratory
100 North Academy Avenue
Danville, PA 17822

Other Participating Researchers

Jinhong Li, MD, PhD; David L. Diehl, MD – employed by Geisinger Clinic
Rebecca J. Critchley-Thorne, PhD; Bruce Campbell, MS – employed by Cernostics, Inc.
Gary W. Falk, MD, MSc; Anil K. Rustgi, MD; Nirag Jhala, MD, PhD – employed by the University of Pennsylvania
Jon M. Davison, MD; Chakra Chennubhotla, Ph.D. – employed by the University of Pittsburgh
Blair A. Jobe, MD; Ali H. Zaidi, MD – employed by West Penn Allegheny Health System
Yi Zhang, Ph.D. – consultant statistician

Expected Research Outcomes and Benefits

The project employs a testing technology for which Geisinger Health System and Cernostics have demonstrated proof-of-concept. The investigators have selected a comprehensive panel of diagnostic and prognostic biomarkers, many of which have established significance in diagnosing the stages of Barrett's esophagus and in predicting risk for esophageal cancer. Therefore, the expected research outcomes of the project are classifiers based on optimal sets of biomarker, morphology and clinical data that can accurately assign diagnosis and predict whether a patient will develop high grade dysplasia or esophageal cancer and also estimate the

sensitivity, specificity and overall accuracy of the diagnostic-prognostic test. It is expected that the test will have high sensitivity and specificity and high positive and negative predictive values based on the known diagnostic and prognostic significance of the panel of biomarkers and based on the high stringency of feature selection for the classifiers. It is also expected that the research will identify a key set of biomarkers and related molecular pathways involved in the progression of Barrett's esophagus to esophageal cancer, which will lead to a better understanding of the biology and behavior of esophageal cancer and aid in the design of new therapeutic agents to prevent and treat esophageal cancer.

The diagnostic utility of the test will improve health status by increasing the accuracy of pathological diagnosis, thus reducing the number of repeat endoscopies and biopsies that patients with Barrett's esophagus must currently undergo, particularly for patients who are initially diagnosed as "indefinite/indeterminate" for dysplasia. The prognostic utility of the test will improve health status by identifying patients at high risk for developing esophageal cancer early in the disease progression when treatments such as endoscopic mucosal resection and radiofrequency ablation can be applied to effectively prevent development of cancer. The prognostic utility will also identify low risk patients, who will not develop esophageal cancer, and who can be spared unnecessary endoscopies, biopsies and treatments.

The expected benefits of the project include; significant improvements in diagnostic and prognostic accuracy to prevent delays in treatment of patients at high risk for esophageal cancer, and a reduction in unnecessary and costly endoscopies and biopsies. This individualized approach will benefit patients by reducing the incidence and mortality associated with esophageal adenocarcinoma and will benefit health care systems by targeting treatments and screenings to the high-risk patients who need them.

Summary of Research Completed

Specific Aim 1: Determine the performance of the prototype test in stratifying patients according to diagnosis and predicting risk for esophageal cancer in a retrospective training patient cohort.

Progress during this project period towards achieving the milestones for specific aim 1:
Cohort for Retrospective Training Study

Patient Cases: The training cohort design was improved by inclusion of patient cases from multiple institutions. This enabled training of multivariate prognostic classifiers on a diverse set of patients, which increases the likelihood that the classifiers will generalize to patient cases outside the training population. To further increase diversity patient cases from Academic Medical Center (AMC), Amsterdam, Netherlands, have been included. The use of AMC's patient biospecimens and de-identified patient data for this study was approved by AMC's IRB. Cernostics has a separate collaboration (funded by Cernostics) with Jacques Bergman, M.D., Ph.D., at AMC to utilize Barrett's cohorts for validation studies of Cernostics' tests. The retrospective training cohort now includes patients from 4 different institutions; a summary is provided in Table 1. 315 cases from 200 patients were retrieved, reviewed and sectioned for the study during the project period 7-1-2013 to 6-30-2014. Patient cases were assigned to either the

training or independent validation cohort. Investigators at University of Pennsylvania provided tissue sections from 126 cases from 72 patients, University of Pittsburgh 98 cases from 78 patients and Geisinger provided 91 cases from 50 patients. A case is a set of biopsies in a single paraffin block. Some patients in the cohort have multiple cases from different parts of the esophagus taken at the same endoscopy. Table 2 summarizes the total number of patient cases acquired by each institution for the study thus far.

Patient Metadata: Each institution has provided de-identified clinical and pathological data for each patient case included in the study. The data collection has continued during this project period and now includes the following data elements: patient key, case key, time-shifted case collection date, original diagnosis, expert review diagnosis, outcome (progressor/non-progressor), progression endpoint diagnosis, time to progression, high grade dysplasia (HGD)/esophageal adenocarcinoma (EAC)-free surveillance time (for non-progressors), total surveillance time, time-shifted date and diagnosis of every surveillance biopsy, age, gender, race, Barrett's segment length (cm), Barrett's segment length class (short/long), hiatal hernia.

Development and Evaluation of Multivariable Prognostic Classification Algorithm

TissueCypher Data Generation: Cernostics investigators continued to develop their image analysis software for use in the project. During this project period Cernostics performed its 15-marker TissueCypher assay on Barrett's biopsies from an additional 224 patients from the four institutions. The TissueCypher assay has been performed on 416 patients to date. The patients were assigned to either the training or validation cohort (see Table 1). Data from the validation cohort was quarantined for later use in independent validation. Sections from the biopsies were labeled with Cernostics' assay and imaged by whole slide 4-channel fluorescence scanning at 20x magnification, resulting in 1,631 whole slide images for analysis. Slides were analyzed using Cernostics' TissueCypher image analysis software. 1,970 image analysis features were extracted from the 15 biomarkers, including morphology. Each feature was summarized as multiple measures, resulting in 26,939 feature/measures per patient case. The feature/measure data from biopsies in the training cohort was transferred to Yi Zhang, Ph.D. (consultant biostatistician).

Statistical Analyses: The goal of the statistical analyses performed by Dr Zhang was to build multivariable prognostic classifiers based on a subset of the 26,939 feature/measures that can predict risk of progression to HGD/EAC in individual Barrett's patients. There were 2 patient risk groups in the training cohort: 1. Progressors, i.e. progressed from no dysplasia (ND), indefinite for dysplasia (IND) or low grade dysplasia (LGD) to HGD or EAC, and 2. Non-Progressors (did not progress to HGD/EAC). Two feature selection methods were evaluated: 1. Univariate Conditional Logistic Regression (CLR) with all 26,939 feature/measures in progressors vs non-progressors (using 84 case-control sets); 2. Univariate Cox Proportional Hazard (Cox) with all 26,939 feature/measures in progressors vs non-progressors (using progressor group as the event, and non-progressors as censoring). The univariate results from both the CLR and Cox were reviewed and two subsets of 50 feature/measures (CLR-selected set and a Cox-selected set) were manually selected based on the performance in the univariate ranking (AUC and p-value) and on ability to capture known mechanisms of neoplastic progression while minimizing feature redundancy.

The top 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 features/measures for the CLR-selected set and the Cox-selected set were combined into prognostic classifiers by multivariate Cox model. Survival time was defined as the time between the tested and the diagnosis of HGD/ EAC for progressors or the last HGD/EAC-free follow-up for non-progressors. Leave-one-out cross validation was performed by setting 1 testing biopsy aside and using all other biopsies as the training set to select features/measures. The final prediction model was built using the Cox model that uses the selected features/measures on the training set of patients, then this model was applied on the 1 testing biopsies to calculate the probability of remaining high risk free at 2, 3, 4, and 5 years. This process was repeated until each biopsy was treated as the testing biopsy once. At the end of this process, each biopsy had a calculated probability of remaining HGD/EAC-free at 2, 3, 4, and 5 years. C-indices were calculated and Receiver Operating Characteristic (ROC) curves were plotted. Kaplan-Meier curves were plotted with 2 probability cutoffs. Area under the ROC curve (AUROC), C-indices, hazard ratios, positive predictive values (PPV) and negative predictive values (NPV) were used to select the top performing classifier from the CLR-selected features (classifier 1) and from the Cox-selected features (classifier 2) (see Table 3). The probability cutoffs were optimized to maximize NPV, PPV and hazard ratios for each classifier. Both classifiers stratified Barrett's patients into low, intermediate and high risk groups, identified progressor patients who are missed by the current standard pathology and showed similar performance across the 4 institutions. Results from the two prognostic classifiers are shown in Figures 1 and 2.

Classifier 1: Classifier 1 is based on 30 image analysis feature/measures derived from 12 biomarkers and nuclear morphology. The optimal probability cutoffs were 0.78 (low-int risk), 0.32 (int-high risk). Classifier 1 predicts probability of remaining HGD/EAC-free at 5 years with AUROC of 0.875 and hazard ratios of 6.3 (intermediate vs. low risk) and 21.9 (high vs. low risk), p value <0.0001 (Figure 1A-B). The PPV and NPV were 0.911 and 0.903, respectively, i.e. 91.1% of the high risk group are progressors and 90.3% of the low risk group are non-progressors. Classifier 1 showed similar performance across the diagnostic categories (Figure 1C, D, E). Multivariate Cox analysis of Classifier 1 versus the pathologist's diagnosis and Barrett's segment length showed that Classifier 1 adds independent prognostic information and has stronger prognostic power than either diagnosis or segment length (Figure 1F-G).

Classifier 2: Classifier 2 is based on 40 image analysis feature/measures derived from 8 biomarkers and nuclear morphology. Classifier 2 predicts probability of remaining HGD/EAC-free at 5 years with AUROC of 0.83 and hazard ratios of 3.7 (intermediate vs. low risk) and 17.4 (high vs. low risk), p value <0.0001 (Figure 2 A-B). The PPV and NPV were 0.803 and 0.909, respectively. Classifier 2 showed similar performance across the diagnostic categories (Figure 2C, D, E). Multivariate Cox analysis of Classifier 2 versus the pathologist's diagnosis and Barrett's segment length showed that Classifier 2 adds independent prognostic information and has stronger prognostic power than either diagnosis or Barrett's segment length (Figure 2F-G).

Classifier 1 has higher overall performance than classifier 2 but requires 12 biomarkers vs. only 8 biomarkers for classifier 2. Both classifiers will be tested on the independent validation set. The performance versus standard clinical variables and ability to generalize to independent patient populations will be considered when selecting which classifier to commercialize.

Milestones 1-3 associated with specific aim 1 were achieved during this project period.

Specific Aim 2: Validate the diagnostic and prognostic performance of the optimized diagnostic-prognostic test in two independent retrospective patient cohorts. Progress during this project period towards achieving the milestones for specific aim 2:

Patient Cases:

An independent validation cohort has been built during this project period that includes cases from 274 Barrett’s patients from the four institutions (Table 1). The 15-marker TissueCypher assay has been performed on 193 of the validation patient cases to date. The research plan aimed to acquire 400 patients for the validation study, however, this sample size was based on estimates prior to the training study. Dr Zhang calculated a sample size based on actual data from the training set. A total of 153 patients are needed to detect the survival differences we observed with the classifiers developed on the training set at 0.01 significance level with 90% power. The 193 patient validation set that the test has been performed on exceeds the required sample size.

TissueCypher Data Generation:

The TissueCypher assay has been performed on all 193 patients in the independent validation cohort. Sections from the biopsies were labeled with Cernostics’ assay and imaged by whole slide 4-channel fluorescence scanning at 20x magnification, resulting in 1,365 whole slide Barrett’s tissue images for analysis. Slides were analyzed using Cernostics’ TissueCypher image analysis software. 26,939 feature/measures per patient case were extracted from 15 biomarkers using Cernostics’ quantitative image analysis (same feature/measures as in the training set). This dataset from the validation cohort was quarantined for later use in independent validation and was not transferred to Dr Zhang or analyzed by the study investigators during the training phase. The remaining time on the project will be used to test classifiers 1 and 2 on the independent validation set. A manuscript will be submitted that describes both the development of the test in the training cohort and testing on the independent validation cohort.

Milestones 1-2 associated with Specific Aim 2 were achieved during this project period.

Cohort	Total # Patients	Progressors	Non-Progressors	Institution			
				Geisinger	UPitt	UPenn	AMC
Training	223	80	143	99	27	24	73
Validation	193	67	126	72	36	20	65

UPitt: University of Pittsburgh, UPenn: University of Pennsylvania, AMC: Academic Medical Center, Netherlands.

Institution	Patients	Non-Progressor Patients	Progressor Patients	HGD/EAC Patients
Geisinger Clinic	204	139	40	25
University of Pittsburgh	134	76	39	19
University of Pennsylvania	88	48	18	22

HGD: High grade dysplasia, EAC: Esophageal Adenocarcinoma

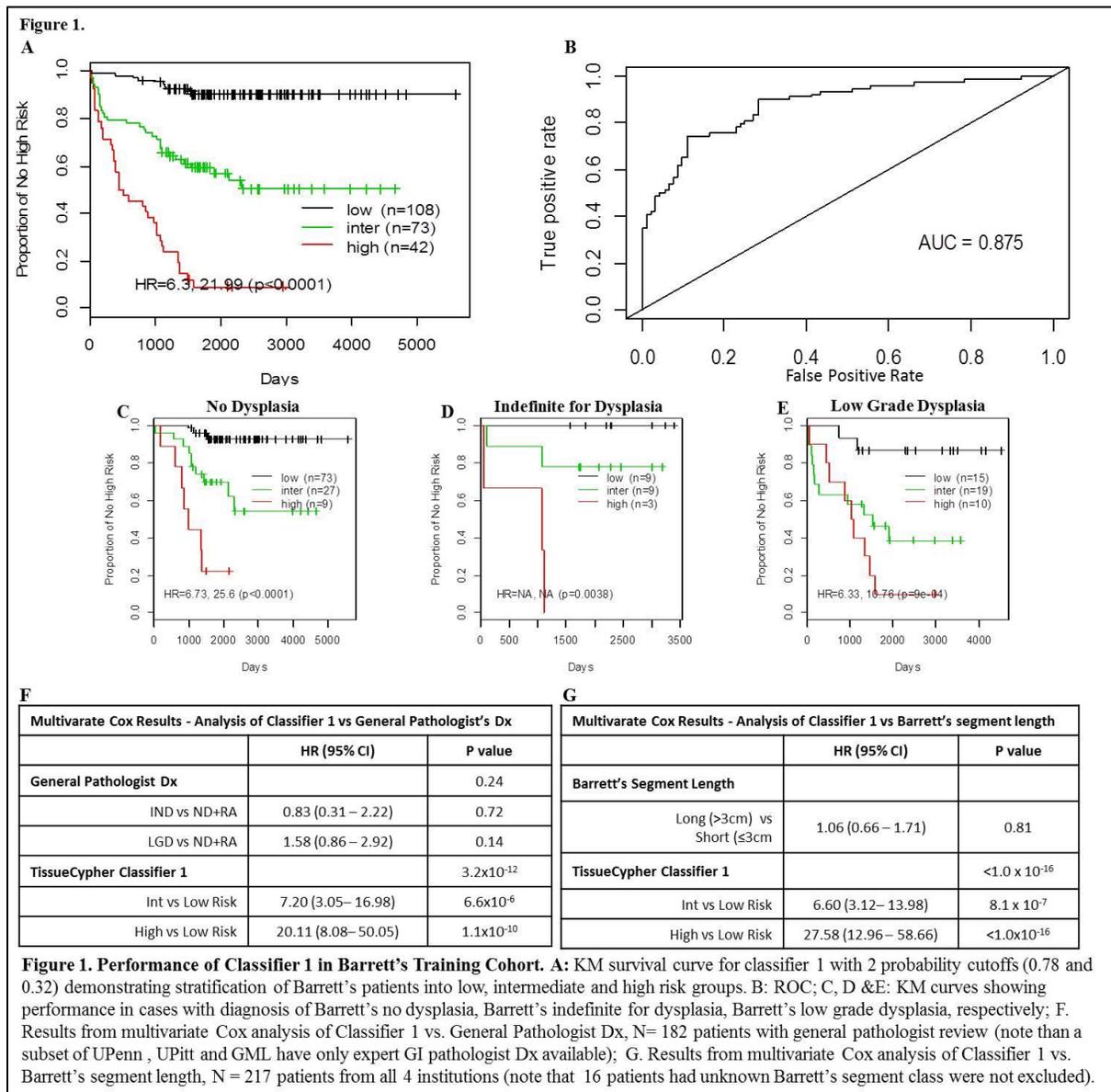


Table 3. Statistical Methods and Biomarker/Morphology Features used in the Development of TissueCypher Classifiers 1 and 2		
	TissueCypher Classifier 1	TissueCypher Classifier 2
Feature Selection	Univariate Conditional Logistic Regression	Univariate Cox Proportional Hazard
Model Building & Prediction	Cox Proportional Hazard	Cox Proportional Hazard
Probability Cutoffs in 3-Tier Classification	0.78 (low-int), 0.32 (int-high)	0.89 (low-int), 0.38 (int-high)
# Feature/measures	30	40
# Biomarkers	12	8
Biomarkers	p53, p16, HER2, CK-20, Ki-67, NF-kappaB, HIF1 alpha, CD45RO, Beta-catenin, COX2, CD68, AMACR (plus morphology)	p53, HER2, CK-20, Ki-67, HIF1 alpha, CD45RO, COX2, CD68 (plus morphology)

Figure 2.

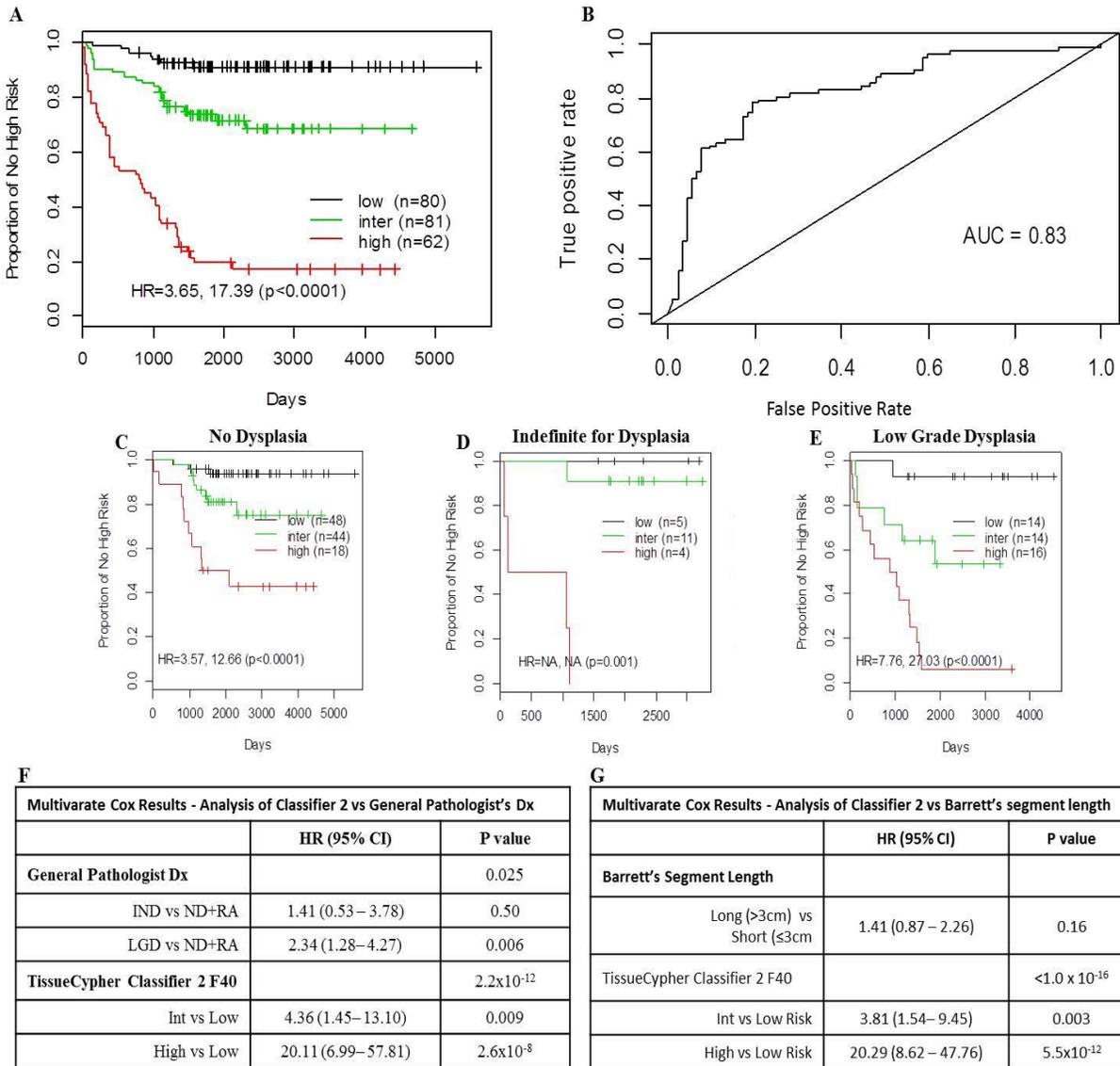


Figure 2. Performance of Classifier 2 in Barrett's Training Cohort. A: KM survival curve for classifier 2 with 2 probability cutoffs (0.89 and 0.38) demonstrating stratification of Barrett's patients into low, intermediate and high risk groups. B: ROC curve; C, D & E: KM curves showing performance in cases with diagnosis of Barrett's no dysplasia, Barrett's indefinite for dysplasia, Barrett's low grade dysplasia, respectively. F. Results from multivariate Cox analysis of Classifier 2 vs. General Pathologist Dx, N= 182 patients with general pathologist review (note that a subset of UPenn, UPitt and GML have only expert GI pathologist Dx available); G. Results from multivariate Cox analysis of Classifier 2 vs. Barrett's segment length, N = 217 patients from all 4 institutions (note that 16 patients had unknown Barrett's segment class were not excluded).