ABSTRACT 7412

A Multi-Gene Classifier for the Diagnosis of Benign Versus Malignant Pulmonary Nodules

Anil Vachani[1], Michael Atalay[2], Ross Bremner[3], Brad Broussard[4], Karen Copeland[5], Katerine von Lang Egressy[6], J. Scott Ferguson[7], Lyssa Friedman[8], Randall J. Harris[9], Jennifer Kidd[8], Joseph Leach[10], Philip McQuary[8], Liqun Qiu[8], Saiyad Sarkar[11], Mona Shahbazian[8], Nadia Sheibani[8], Jaime Shuff[12], Brad Vincent[13], Lyndal Hesterberg[8]

[1] University of Pennsylvania, Philadelphia, PA; [2] Rhode Island Hospital, Providence, RI; [3] St. Joseph's Hospital and Medical Center, Phoenix, AZ; [4] Lafayette General Health, Lafayette, LA; [5] Boulder Statistics, LLC, Boulder, CO; [6] Kalispell Regional Healthcare, Kalispell, MT; [7] University of Wisconsin, Madison, WI; [8] OncoCyte Corporation, Alameda, CA; [9] Mercy Medical Center, Canton, OH; [10] Minnesota Oncology, Saint Louis Park, MN; [11] Medstar Franklin Medical Center, Baltimore, MD; [12] St. Mary's Regional Medical Center, Reno, NV; [13] Mary Bird Perkins Cancer Center, Baton Rouge, LA

BACKGROUND

- Pulmonary nodules are commonly encountered in clinical practice as a result of increased use of CT imaging
- The risk of lung cancer for any nodule is largely determined by size and appearance, patient age, and smoking history
- Patients with pulmonary nodules frequently undergo invasive diagnostic testing, such as transthoracic or bronchoscopic biopsy, and these approaches are limited by non-diagnostic results and complications
- Blood-based biomarkers may identify patients with higher-risk lesions that warrant invasive testing while allowing patients at lower risk to undergo surveillance
- The initial classifier consisting of 8 genes used in 106 patients with pulmonary nodules (50 benign, 56 malignant, including 50 NSCLC, 4 SCLC, 1 carcinoid and 1 lymphoma) distinguished benign from malignant nodules with high accuracy (AUC 0.87, 95% CI: 0.80, 0.93); accuracy of the classifier improved when nodule size (NS) was added (AUC 0.99, 95% CI: 0.96, 1.00)

METHODS

- 299 PAXgene samples were collected from 29 U.S. centers from current or former smokers who had pulmonary nodules 5-30mm
- RNA was extracted using Qiagen PAXgene Blood miRNA kit
- Samples were run on NanoString nCounter® MAX Analysis System along with RNA reference controls
- Generalized regression with a lasso-fitting algorithm was used with all 299 samples to build a model with 15 markers and nodule size
- A six-fold cross validation was performed using random forest, boot-strap methods with lasso optimization for marker selection and model development
- Upon selection of top candidate markers, a repeat six-fold cross-validation was performed to optimize the model
- The overall performance of the optimized model was evaluated and compared to nodule size alone and the VA Model[1]

RESULTS

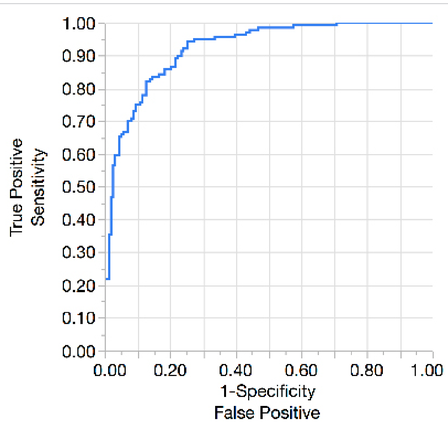
- In the first stage of classifier development, 18 markers were selected across a six-fold internal cross-validation
- The average performance of the classifier in the test sets across all six folds yielded a sensitivity of 0.90 (95% CI: 0.73-0.96), specificity of 0.62 (95% CI: 0.48-0.72), and AUC of 0.86
- In the optimization phase of classifier development, the top 18 markers selected from the six-fold models were then used to train a classifier on all 299 samples
- The optimized, final classifier used 15 markers and NS, yielding an overall AUC 0.92 (95% CI: 88.8-95%)
- In this study population, there was no significant difference between the performance of NS alone and the VA model[1], with AUCs of 0.85 (95% CI: 80.7-90.1%) and 0.82 (95% CI: 76.9-86.7%) respectively

Table 1: Demographics, N=299

	Benign (%)	Malignant (%)	Total
Gender			
Female	84 (50%)	85 (50%)	169
Male	74 (57%)	56 (43%)	130
Age			
Mean	65.0	68.3	
Range	36-89	46-89	
Nodule size			
0.5-0.7	63 (97%)	2 (3%)	65
0.8-1.0	38 (75%)	13 (25%)	51
1.1-2.0	45 (38%)	74 (62%)	119
2.1-3.0	12 (19%)	52 (81%)	64
Total	**158 (53%)**	**141 (47%)**	**299**

Malignant subtypes: 124 NSCLC, 8 SCLC, 4 carcinoid, 1 metastatic colorectal, 1 lymphoma and 3 cancer type unknown

Figure 1: Final Classifier Test Performance

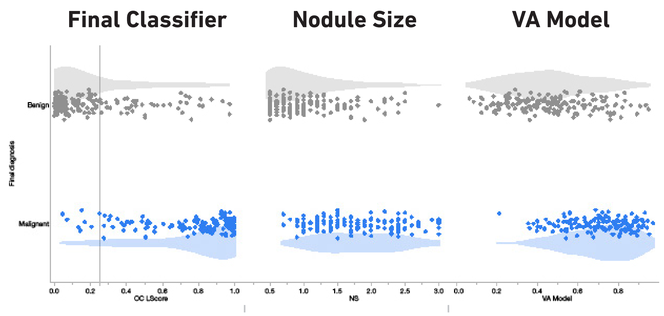


Observed sensitivity: 95% *(90-98%)*
Observed specificity: 73% *(65-79%)*
Cancer prevalence: 47%

Table 2: Final Classifier Test Performance

		Final Diagnosis		
		Malignant	Benign	Total
Test call	Malignant	134	43	177
	Benign	7	115	122
	Total	141	158	299

Figure 2: Performance Comparison



SUMMARY

- A blood-based gene expression classifier in current and former smokers with pulmonary nodules 5-30mm allows identification of benign from malignant nodules with high accuracy
- Classifier performance was superior to nodule size alone and to a cancer risk probability model[1]
- A clinical validation study with a large independent test set is planned

DISCUSSION

- The current classifier replicates initial results
- Although increased NS is associated with an increase in cancer risk, a significant number of invasive procedures yield benign results across nodule sizes, particularly in regions with endemic fungal infections
- These procedures are associated with significant morbidity in a patient population with co-morbid conditions
- A non-invasive gene expression classifier with high sensitivity can be used to identify a population of patients with pulmonary nodules at low risk of malignancy and can facilitate conservative patient management

REFERENCE:

1. Gould MK, Ananth L, Barnett PG, Veterans Affairs SNAP Cooperative Study Group. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *CHEST* 2007 131:383-388.

Study funded by OncoCyte Corporation

OncoCyte™