Blinded Prospective Validation Study of a Whole Blood Gene-Expression Classifier for the Diagnosis of Benign Versus Malignant Pulmonary Nodules

ABSTRACT 7450
POSTER #421

Philip McQuary[1] , Mona Shahbazian[1], Lyssa Friedman[2], Nadia Sheibani[1], Mary Del Rosario[1], Esther Nith[1], Karen Copeland[3], B. Ring[4], Doug Harrington[5], Lyndal Hesterberg[1]

[1]OncoCyte Corporation, Alameda, CA, United States, [2]Lyssa Friedman Consulting, LLC, Mill Valley, CA, United States, [3]Boulder Statistics, LLC, Steamboat Springs, CO, United States, [4]College of Life Science, Huazhong University of Science and Technology, Wuhan, China, [5]Predictive Health Diagnostics, Inc., Irvine, CA, United States

BACKGROUND

- Pulmonary nodules are frequently identified incidentally by thoracic imaging and via low-dose CT (LDCT) for lung cancer screening
- The risk of lung cancer for any nodule is largely determined by its nodule size (NS) and appearance, patient age, and smoking history
- Patients with pulmonary nodules frequently undergo invasive diagnostic testing, such as transthoracic or bronchoscopic biopsy however, these approaches may result in non-diagnostic results and/or complications
- Although increased NS is associated with an increase in cancer risk, a significant number of invasive procedures yield benign results and are associated with significant morbidity (20+%)[1]
- Blood-based biomarkers may identify patients with higher-risk lesions that warrant invasive procedures while allowing patients at lower risk to undergo surveillance
- The multivariate gene expression classifier was trained on a set of 700 samples from 57 clinical sites
- Random decision forests and bootstrapped p-values from a generalized regression modeling using a lasso fitting algorithm were used for gene selection and for selecting the final classifier

METHODS

- 250 blinded PAXgene whole blood samples were collected from current or former smokers who had pulmonary nodules 5-30mm identified by imaging
- Validation set samples were from 44 U.S. clinical sites (Figure 1)
- RNA was extracted using the Qiagen PAXgene Blood miRNA kit
- Samples were analyzed using the ThermoFisher Ion GeneStudio S5 Next-Generation Sequencing System (one sample with no result) to validate the classifier performance
- The overall performance of the multivariate gene expression classifier was evaluated and compared to the Mayo Model[2] (Figure 3)
- All shown data was generated from 249 specimens

RESULTS

- The validation sample set reflected a robust clinical trial population collected from 44 sites distributed across the U.S. with overall cancer prevalence of 32%
- The classifier using only gene-expression biomarkers from whole blood – no clinical parameters – yielded an overall AUC of 0.89 with Sensitivity ~90% (95%CI 82%-95%) and Specificity ~75% (95%CI 68%-81%) (Figure 2)
- In the LungRADS 4 study sub-population (8 - 30mm NS) the biomarker test had an AUC of 0.88 while the Mayo Model, which includes NS, yielded an AUC of 0.72 (p=0.0001) (Figure 3)
- There was no difference in classifier performance observed within the validation group (249 patients) based on age, presence of fungal infections, comorbidities (e.g., COPD) or gender

Table 1: Demographics, N=249

	Benign (%)	Malignant (%)	Total
Gender			
Female	46 (27%)	41 (51%)	87
Male	123 (73%)	39 (49%)	162
Age			
Mean	65.9	69.1	
Range	24 - 92	42 - 89	
Nodule size			
0.5-0.7	59	0	59
0.8-1.0	42	9	51
1.1-2.0	50	45	95
2.1-3.0	18	26	44
Total	**169 (68%)**	**80 (32%)**	**249**

Malignant subtypes included in analysis: Adenocarcinoma (n=50); Squamous (n=18); Small cell (n=5); Carcinoid (n=1); Adenosquamous (n=1); Subtype unknown (n=5)

Table 2: Classifier Performance

		Final Diagnosis		
		Malignant	**Benign**	**Total**
Test call	Malignant	72	43	115
	Benign	8	126	134
	Total	**80**	**169**	**249**

Figure 1: Robust Sample Distribution in Both Studies

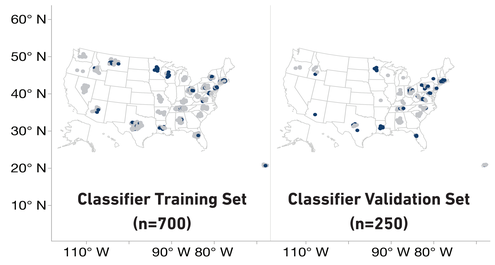
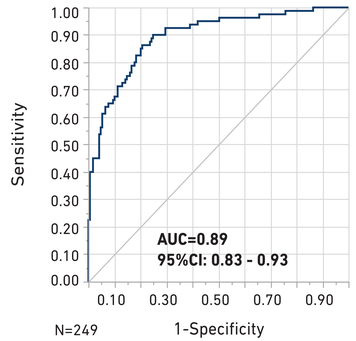


Figure 2: Classifier Performance



The classifier is a measure of biological risk: biomarkers alone

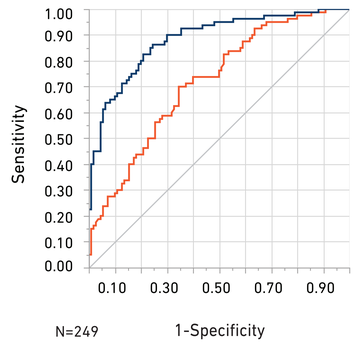
No clinical factors are used

Table 3: Classifier NPV across Cancer Prevalence Ranges

Cancer Prevalence	10%	20%	30%
NPV	98.0%	95.6%	92.7%

Negative Predictive Value (NPV), important to rule out unnecessary biopsies, remains high in different cancer prevalence ranges

Figure 3: Performance Comparison in LungRADs 4



Classifier
AUC=0.88
95%CI: 0.82 - 0.93

Mayo Model Score
AUC=0.72
95%CI: 0.64 - 0.78

LungRADs 4 = 8-30 mm nodules

SUMMARY

- A multivariate gene expression classifier identifies benign from malignant nodules between 5-30mm with high accuracy in a diverse population of current and former smokers
- This classifier, which is based on biomarkers only, significantly outperformed the Mayo Model for cancer risk that utilizes nodule size as one of the variables[2]
- A larger, blinded clinical validation study of the classifier with an independent specimen set is underway

DISCUSSION

- The blood-based classifier demonstrates high sensitivity (90%) with a specificity of 75% in a prospective blinded validation set of 249 samples
- This non-invasive gene expression classifier is a biological assessment of the risk of cancer with predictive value independent of nodule size or clinical factors
- This method supports improved care for patients for whom clinical management is not straightforward (pulmonary nodules, 8 to 30 mm) by identifying the subset of patients at low risk of malignancy

REFERENCE:

1. Huo J, Xu Y, Sheu T, Volk RJ, Shih Y-CT. Complication Rates and Downstream Medical Costs Associated with Invasive Diagnostic Procedures for Lung Abnormalities in the Community Setting. JAMA Intern Med. 2019 Jan 14;179:324–332
2. Swensen SJ, Silverstein MD, Ilstrup DM, et. al. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. Arch Intern Med. 1997 Apr 28;157(8):849-55

Study funded by OncoCyte Corporation

OncoCyte™