RESEARCH



Analytical validation of the Percepta Nasal Swab classifier; an RNA next-generation sequencing assay for the assessment of lung cancer risk in pulmonary nodules



Shuyang Wu¹, Ruochen Jiang¹, Grazyna Fedorowicz¹, Mei Wong¹, Janna S. Chamberlin¹, Lori Lofaro¹, P. Sean Walsh¹, Giulia C. Kennedy¹, Yangyang Hao¹, Jing Huang¹ and Bill Bulman^{1*}

Abstract

Background A novel molecular diagnostic test, Percepta Nasal Swab (PNS), was developed as a noninvasive lung cancer biomarker to aid in risk assessment for indeterminate pulmonary nodules in individuals who smoke or have previously smoked. Prior research has shown that exposure of the airway epithelium to cigarette smoke results in epithelial gene expression alterations throughout the respiratory tree that reflect the risk of lung cancer in a pulmonary nodule. The PNS classifier leverages this concept using whole transcriptome sequencing (RNASeq) of cells collected from the nasal epithelium and provides "high", "intermediate" and "low risk" classification calls to help guide clinical management decisions. The clinical validity of the PNS test was established on an independent validation set and demonstrated favorable sensitivity and specificity. This study aims to evaluate the analytical validity of the PNS test performance in our CLIA (Clinical Laboratory Improvement Amendments) laboratory.

Methods The reproducibility between RNASeq runs within a laboratory and the accuracy between laboratories were estimated and compared against the performance-based acceptance criterion. The impacts from varying RNA input amount, genomic DNA and blood RNA interference were evaluated to demonstrate the analytical sensitivity and specificity of the PNS test results to known conditions that may occur in routine laboratory processing.

Results Based on modeling the impact on clinical sensitivity/specificity, PNS test classifier scores can allow up to 0.776 score units of added noise/variability before any performance metrics drop below the pre-specified requirements. This allowable variability is six-fold higher than the observed variability estimated between runs and between laboratories under routine testing conditions, which are each less than 2% of the 98th percentile score range. In addition, PNS test results are shown to be robust against RNA input variation from 50 ng to 15 ng, up to 30% of genomic DNA by nucleic mass interference, and up to 14% of blood RNA interference.

Conclusions This study provided sufficient evidence for the accuracy, reproducibility, sensitivity, and specificity of the PNS molecular test and supported its utilization in clinical testing.

*Correspondence: Bill Bulman bill.bulman@veracyte.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Keywords Nasal Swab classifier, Analytical validation, Molecular diagnostic test, Lung cancer risk, Pulmonary nodules, Genomics

Background

Lung cancer is the second most common cancer in the U.S. and remains the number one cause of cancer-related mortality for both men and women [1]. The U.S. Preventative Services Task Force (USPSTF) recommends that certain individuals at high risk for lung cancer undergo annual screening with low dose CT [2], which has been shown to reduce the lung cancer mortality by 20% [3]. While highly sensitive for lung cancer, screening has a high false discovery rate, commonly identifying pulmonary nodules which will ultimately be proven to be benign. Pulmonary nodules are also common incidental findings on CT done for other reasons, and these too are most often benign [4]. When a pulmonary nodule is identified, current guidelines recommend that physicians base management decisions on the estimated probability of malignancy [5]. Currently, this estimation relies on known risk factors, radiographic features, and, in some cases, the use of validated clinical risk model calculators [5]. The use of a lung cancer biomarker has been suggested as a way to improve the accuracy of risk assessment [6], and genomic information has been shown to have clinical utility in this regard [7]. Extensive prior research has shown that a history of cigarette smoking is associated with temporary and permanent changes in gene expression in the airway epithelium in what has been referred to as "the field of injury", and that patterns of gene expression detectable throughout the respiratory tree correlate with the likelihood of lung cancer in a pulmonary nodule in individuals with a history of smoking [8-10]. This effect is manifested in individuals with a minimum exposure of only 100 cigarettes smoked in a lifetime. Classifiers leveraging gene expression profiling of benign-appearing bronchial epithelial cells collected during bronchoscopy have been developed and shown to successfully aid in the risk stratification for lung cancer management [11, 12]. The field of injury principle was subsequently shown to apply to cells from the nasal epithelium [13]. Leveraging that concept, a test [14, 15] utilizing patterns of gene expression in cells collected from the nasal epithelium along with clinical factors was developed to accurately predict the risk of lung cancer in a screen or incidentally-detected pulmonary nodule in individuals with a history of smoking.

The Percepta Nasal Swab (PNS) classifier was developed using nasal epithelial genomic information from 1120 patients with a pulmonary nodule (\leq 30 mm). The PNS classifier consists of two machine learning models ensembled in a hierarchical structure. The upstream model is a logistic regression model that mainly relies on clinical features to up-classify patients to high risk. Patients not classified as high risk in the upstream model are fed into the downstream model: a supporting vector machine that relies heavily on genomic features as well as interactions between genomic and clinical covariates to stratify the remaining patients to high/intermediate/low risk groups. The upstream and downstream models use a total of five clinical features (age, pack years, years since quit (cigarette smoking), nodule length and nodule spiculation) and 502 gene features as inputs [14, 15].

The PNS classifier was clinically validated in a set of 249 subjects and its associated performance is used as the foundation for this study. The clinical validation of the PNS classifier demonstrated robust performance on the primary set, with 96% sensitivity and 42% specificity in classifying patients to low-risk and 58% sensitivity and 90% specificity in classifying patients as having high risk of malignancy. More details regarding the clinical validity have been presented elsewhere [14, 15]. Taken together, the PNS test has the potential to provide accurate assessment of malignancy for ever-smokers with a newly detected lung nodule.

As a novel molecular diagnostic test, analytical validation of the PNS test is required to show that the assay results are reproducible and robust to technical variations that can be expected in routine laboratory processing and clinical testing. Specifically, analytical validation establishes the test performance in real-life settings where reagent lots, equipment, and operators can vary from run to run, and contaminants may be present in the sample [16]. Criteria for analytical validity of novel molecular diagnostic tests are established by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group and the Centers for Disease Control's ACCE Project (Analytic validity, Clinical validity, Clinical utility and associated Ethical, legal, and social implications) [17, 18]. Following the established criteria, this study aims to evaluate PNS for its variability and reproducibility of test results with variation in input RNA quantity, potential contaminants including blood and genomic DNA, and intra-run, inter-run intra-lab reproducibility, and inter-laboratory test accuracy. Variability and reproducibility of the pre-analytical phase, including specimen collection and RNA isolation steps, are beyond the scope of this analytical validation study and are not discussed in this paper.

Methods

Specimens

Nasal specimens utilized for analytical validation studies were collected using a Cyto-pak Cyto-Soft brush (CP-5B, Medical Packaging Corp., Camarillo, CA, USA). After sample collection, nasal specimens were stored in a nucleic acid preservative (RNAprotect, QIAGEN, Hilden, Germany) and either shipped chilled to a contract research lab for RNA extraction (AEGIS) or frozen at -80°C prior to RNA extraction (DECAMP-1, Lahey).

RNA extraction, amplification, and sequencing

Thawed nasal specimens in RNAprotect were agitated to remove cells from the CytoSoft brush either by vortexing or using a TissueLyser without bead (QIAGEN, Hilden, Germany) and then cells were pelleted by centrifugation (5000-10000 g, 5 min). Following removal of RNAprotect, the cell pellet was lysed using the QIAzol reagent and total RNA extracted using the miRNeasy Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. RNA quantification was performed using the QuantiFluor RNA System (Promega, Madison, WI), and 50 ng of RNA were used as input to the TruSeq RNA Access Library Prep procedure (Illumina, San Diego, CA), which enriches for the coding transcriptome. Libraries meeting quality control criteria for amplification yields were sequenced using NextSeq 500/550 instruments $(2 \times 75$ bp paired-end reads) with the High Output Kit (Illumina, San Diego, CA).

Raw sequencing (FASTQ) files were aligned to the Human Reference assembly 37 (Genome Reference Consortium) using the STAR RNA-seq aligner software. Uniquely mapped and non-duplicate reads were summarized for 63,677 annotated Ensembl genes using HTSeq. Data quality metrics were generated using RNA-SeQC. Samples were excluded when their library sequence data did not achieve minimum criteria for total reads, uniquely mapped reads, mean per-base coverage, base duplication rate, percentage of bases aligned

Table 1 Pre-specified performance thresholds used in flip rate analysis

Classification	Performance Metric	Clinical Validation Performance	Pre-spec- ified Per- formance Threshold
High Risk	Sensitivity	58.2%	50%
	Specificity	90.4%	85%
	PPV	67%	60%
	(prevalence = 25%)		
Low Risk	Sensitivity	96.3%	90%
	Specificity	41.7%	35%
	NPV	97.1%	90%
	(prevalence=25%)		

to coding regions, base mismatch rate and uniformity of coverage within each gene.

Study design overview

A simulation study was first performed to establish acceptance criterion of technical noise in the assay, then the inter-laboratory accuracy and intra-lab inter-/intrarun reproducibility were assessed against the established acceptance criterion. Next, the sensitivity of the PNS test to RNA input amount and specificity to interference from genomic DNA and blood RNA were established. These assessments utilized similar methods and study designs as previously employed and described in the analytical validation of a related molecular test, the Percepta Genomic Sequencing Classifier [19]. In addition, samples selected in each study showed comparable RNA quality measured by RIN value when compared with the clinical validation sample set of N= 249 patients, as shown in Supplement File 1 Figure S1.

Flip rate simulation

To establish the acceptance range for the technical noise that the PNS classifier can tolerate, a simulation study was performed to evaluate the amount of technical noise that could be added to classifier scores before the performance of the classifier was significantly impacted. Thresholds or allowable ranges were pre-specified for each of the clinical validation performance metrics to indicate when the classifier performance becomes significantly impacted (Table 1). The simulation was performed by generating random score noise from a normal distribution with a mean of 0 and standard deviation (SD) varying between 0.01 and 10. The generated random noise was then added to the scores from the 249 subjects in the independent validation set of the PNS test in the primary analysis [14, 15]. Simulated scores with added random noise were generated for both the upstream and downstream models of the PNS test. The amounts of random noise added to the upstream and downstream models were proportional to the expected variability observed in technical replicates during development. This is because the upstream model mainly assesses the patient's clinical risk for lung cancer, which is less likely to be impacted by the technical noise found in the assay, thus the same amount of technical noise would affect the upstream model with less magnitude compared to that of the downstream model. After generating the simulated validation set scores at each level of random noise, performance metrics including sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) were calculated based on the simulated scores. Note that for PPV and NPV calculation, data was extrapolated to 25% cancer prevalence, a prevalence reported by Tanner and colleagues in an observational

study of 18 geographically diverse community pulmonary practices [15]. In addition, the percentage of classifier calls that changed due to added noise, which is referred to as the flip-rate, was calculated at each noise level. The simulation was repeated 1000 times at each noise level, and median performance metrics were computed across all simulations at each noise level to quantify the impact from such noise. The maximum allowable score variability was determined as the largest added noise or variability for which all the performance metrics continued to meet the pre-specified thresholds which are chosen to be the lower end of the 95% confidence interval for the performance of the clinical validation of the classifier [15] (Table 1). This maximum allowable score variability was used as an acceptance criterion for the score variability estimated in the following experiments.

Assay reproducibility and accuracy

To assess the intra-run and inter-run reproducibility of the PNS classifier, 30 different samples, including six control samples, were processed in triplicate across three different experimental runs in a single laboratory. A single nasal brushing per patient, from which RNA was extracted and subsequently processed into aliquots, was used for analysis. The reagent lots, operators, and equipment were varied for each experimental run to represent the normal variation anticipated in routine processing. The 24 non-control samples were chosen to represent the entire classification score space of PNS, evenly divided between high, intermediate, and low scoring samples. The assay reproducibility was quantified by a linear mixed-effects model (Eq. 1), with S_{ijk} representing classifier scores of either upstream or downstream model for sample *i*, run *j* and replicate *k*, μ_i representing the fixed effect from different samples, r_i representing the random effect from different runs, and $\mu_i : r_j$ representing the random effect from interaction between sample iand run *j*.

$$S_{ijk} = \mu_i + r_j + \mu_i : r_j + \epsilon_{ijk} \tag{1}$$

To evaluate the inter-laboratory accuracy of PNS from the R&D laboratory where the test was developed to the CLIA laboratory where future testing will be performed, a panel of 85 patient samples and 6 controls were processed in the two laboratories to compare results. The two sets of libraries were prepared by different operators using different equipment from the two laboratories. Each sequencing set was performed by either the R&D or CLIA operators in the CLIA laboratory, on non-overlapping sequencers. The patient samples were selected to cover the entire score range of PNS. Linear modeling was used to evaluate inter-laboratory accuracy and variability (Eq. 2).

$$S_{ik} = \mu_i + \epsilon_{ik} \tag{2}$$

In this model, S_{ik} represents the classifier scores of either upstream or downstream models for sample *i* and replicate *k*, μ_i represents the sample effect, *k* represents technical replicates, and ϵ_{ik} is the residual. All 95% confidence intervals (CI) for SDs were obtained by bootstrapping.

Analytical Sensitivity and specificity

The PNS test specifies 50 ng of total RNA to be used as input to the library preparation procedure, however, the actual input amount can vary due to the nominal quantitation measurement error or due to pipetting accuracy. Lowered RNA input level can result in less diverse library populations which may affect test results. Therefore, the sensitivity of the PNS test to lower RNA input amounts was evaluated using four nasal swab RNA samples, two representing "High Risk" scores (positive score with malignant classification) and two representing "Low Risk" scores (negative score with benign classification). The four samples were plated in triplicate using input amounts of 15, 20, 36, and 50 ng RNA.

Genomic DNA can be co-extracted with RNA in the clinical nasal swab samples, representing a potential contaminant in the PNS assay. To evaluate the impact of genomic DNA on PNS scores and results, genomic DNA was added to four nasal swab RNA samples at 0%, 3%, 10%, and 30% contamination by mass of nucleic acid. Note that the total amount of RNA in each sample was constant at 50 ng. All four RNA samples were chosen to have low gDNA contamination level (must be <1%), with two representing "High Risk" scores (positive score with malignant classification) and two representing "Low Risk" scores (negative score with benign classification). The four samples were plated in triplicate with 3%, 10%, and 30% of genomic DNA added, and in duplicate with 0% genomic DNA addition.

Linear mixed-effect models (Eq. 3) were used to evaluate the impact of total RNA input amount and interfering genomic DNA on classifier scores (S_{ijk}) of either the upstream or downstream models. In Eq. 3, μ_i represents the random effect from different samples, b_j represents the fixed effect from different experimental conditions, i.e., varying RNA input amount or genomic DNA contamination, k represents technical replicates, and ϵ_{ijk} denotes the residual.

$$S_{ijk} = \mu_i + b_j + \epsilon_{ijk} \tag{3}$$

For each model, analysis of variance (ANOVA) was used to test if the experimental conditions introduce significant difference in the PNS test scores. P-values were considered significant at 5%. The number of samples and replicates used for the input amount and gDNA effect evaluation has been determined to be sufficient, providing greater than 80% statistical power when analyzed using a linear mixed-effects model to detect an effect size at the upper bound of the 95% confidence interval for the intra-run standard deviation (SD).

In addition to genomic DNA, the collection of the nasal swab sample may introduce blood, which could potentially lead to altered gene expression [20]. To evaluate the potential of blood as an interfering substance in the PNS test, 5%, 10%, 20%, 50%, and 75% of whole blood total RNA from two donors was added to total RNA purified from three nasal swab samples, while maintaining the total RNA input into the test constant at 50 ng. All selected blood and nasal samples were from male donors or patients to control for the potential interference from sex. Each of the three nasal swab samples was chosen to represent either the low, intermediate, and high cancer risk group, as determined by a preliminary PNS test result. Each of the two whole blood RNA samples was extracted by the same RNA isolation procedure as the nasal samples. Each nasal swab RNA sample with 0%, 10%, 20% and 100% whole blood RNA sample added was run in triplicate, whereas each sample with 5%, 50%, 75% blood RNA addition was run in duplicate. PNS scores were generated for all nasal and blood RNA mixture samples and the best fit curve was plotted through the scores. For each nasal swab sample, the minimum percentage of blood addition that would alter its original PNS classification call was estimated from the curve and reported as the blood interference tolerance limit.

Results

All aspects of the experiments were evaluated for both the upstream and downstream models of the PNS test. All data analysis was done in R version 3.2.3. The results, including variability and impact from interference for the upstream model alone, were better in all cases than the results of the downstream model alone. This is because the upstream model mainly relies on clinical features, while the downstream model relies heavily on genomic features. The downstream model is therefore subject to greater impact from assay variability and interference from potential contaminants. The impacts are being thoroughly evaluated in the analytical validation and summarized in the subsequent sections.

Flip rate simulation

As increasing amounts of technical variability were added to the validation set PNS scores in this simulation, the first observed impact was an increased probability of false up-classification, potentially reducing the accuracy of positive test calls. Specifically, when the amount of score variability exceeded 0.776 classifier score units, the PPV of the simulated validation set PNS scores dropped from 67 to 60% (Fig. 1), which was below the pre-specified acceptable range for this metric (Table 1). All other performance metrics, including Sensitivity, Specificity, NPV and Flip Rate, allowed higher amount of score variability. Therefore, the maximum acceptable score variability of the PNS test is limited by PPV and is established to be 0.776 classifier score units, which is 11.4% of the 98 percentile of the PNS score range.

Assay reproducibility

Between-run and within-run score SDs were estimated using a linear mixed-effects model (Eq. 1). The betweenrun SD was estimated to be 0.117 (95% CI 0.086-0.153; Fig. 2), and the within-run SD estimated to be 0.114 (95% CI 0.082-0.147; Fig. 2). This can be compared to the inter-class score variability between benign and malignant samples, which is almost four-fold higher (Fig. 2). The total SD was compared to the 98-percentile score range in the training set using cross-validation, which was calculated to be 6.82 classifier score units. The between-run SD ascribed to technical variation therefore represents 1.72% of the 98 percentile PNS score range, six-fold lower than the acceptable score variability, 0.776, as established in the Flip Rate simulation study. Thus, the technical SD within and between runs is far lower than the amount of variability that can affect PNS performance, as well as lower than the inherent biological signal on which the test operates. The conclusion remains valid when evaluated among samples with lower RNA quality, defined as those below the median of the clinical validation set, as shown in Supplementary File 1 Table S1.

Inter-laboratory accuracy

The total score variability between the R&D laboratory where the test was developed and the CLIA laboratory where future testing will be performed was modeled using a linear model (Eq. 2). The model estimated the between-lab SD to be 0.117 (95% CI 0.094 to 0.135; Fig. 2), which is in line with the between-run, within-lab SD observed in the reproducibility study. The conclusion remains valid when evaluated among samples with lower RNA quality, defined as those below the median of the clinical validation set, as shown in Supplementary File 1 Table S1. This suggests that technical variation associated with lab, equipment and operator variation is within the range of normal processing run variation in the same lab. The PNS scores generated between the two laboratories also showed a very high correlation ($R^2 = 0.98$). Between the two laboratories, discordant calls were made in 5 out of 162 (3.1%) classifier calls. The discordant calls are due to normal between-run score variability in samples with scores within 2 SDs of the classifier score decision boundary.



Fig. 1 Simulated high risk classification PPV in relation to increasing amount of noise added to the independent validation set scores (n = 249). Each box represents the PPV based on the simulated scores at each tested variability level (x-axis). The horizontal dashed line denotes the pre-specified performance requirement of 60% PPV. The vertical line denotes the maximum acceptable SD, 0.776, which is when PPV drops below the pre-specified requirement of 60%



Fig. 2 Comparison of PNS classifier score variability based on the downstream model scores where the dots indicate the point estimates, and two bars indicate the lower and upper bound of the 95% confidence interval. The Between Class score SD includes biological variation between cancer and benign samples and was computed from samples in the independent clinical validation set (n = 249) [15]. The between lab score variability is calculated from 87 samples that were sequenced both in R&D lab and CLIA lab (n = 174). The Between and within run score variability is estimated from 29 samples, each with 9 replicates across 3 runs (n = 261). The dashed line denotes the SD of 0.776, which was determined by flip-rate analysis to be the maximum acceptable SD at which performance of PNS still maintains all pre-specified requirements

Analytical sensitivity-total RNA input quantity

The sensitivity of the PNS test to varying levels of total RNA input amount was evaluated by linear mixed-effects model (Eq. 3). The result indicates that the RNA input amount does not have a significant impact on the PNS test scores (p-value = 0.25 for the downstream model). Also, samples with each RNA input amount tested (15 ng, 20 ng, 36 ng) were not significantly different from the scores of samples with the nominal input of 50 ng RNA (Fig. 3). This suggests that the PNS test results are robust over a range beyond what would be expected under routine test conditions, specifically, input amount has no significant impact on test results even when input is 70% lower than the nominal level. The conclusion remains valid when evaluated among samples with lower RNA quality, defined as those below the median of the clinical validation set, as shown in Supplementary File 1 Table S1.

Analytical specificity-genomic DNA

The impact on the PNS test scores from the potential contaminant, genomic DNA, was evaluated using a linear mixed-effects model (Eq. 3). Results indicate that the PNS scores of samples with gDNA additions were not significantly different from the corresponding pure RNA samples, when evaluated with a linear mixed-effects model (p-value = 0.93 for the downstream model), with no consistent trend in scores observed (Fig. 4). In addition, the amount of genomic DNA found in clinical samples collected as part of the AEGIS-1 and AEGIS-2

studies were consistently $\leq 1\%$ [18]. This study demonstrated that PNS results were robust in the presence of DNA contamination 10-fold higher than what has been observed in clinical nasal swab samples. Hence, genomic DNA contamination of test RNA has no meaningful impact to the PNS test results.

Analytical specificity - blood interference

The impact on the PNS test scores from another potential contaminant, blood RNA, was evaluated by regression analysis to estimate the maximum amount of blood RNA that could be tolerated before any classification call of the nasal swab sample would be altered. Two blood samples were used to spike in RNA in this study, with the first one having a lower PNS classifier score (more benign) than the second, thus the same nasal swab sample can have a different tolerance level to the two blood samples. With 31% of the first blood RNA or 14% of the second blood RNA added to the low-risk nasal sample, the PNS classification call changed from low to intermediate risk. This was due to the increase in the downstream model score of the nasal sample after the addition of blood RNA, which eventually caused the classification call to change. With 49% of the first blood RNA or 25% of the second blood RNA added to the intermediate risk nasal sample, the PNS classification call changed from intermediate risk to high risk due to the increase of the upstream model score with the blood addition. The high-risk nasal sample would be expected to be classified



Fig. 3 Analytical sensitivity of PNS to RNA input. The y-axis is a relative scale, with 0 representing the mean score of each sample across all input levels. Scores of the downstream model is shown here in the boxplot with the horizontal line in each boxplot indicating the median value. Each RNA input amount was run in triplicate for each sample. The region between the two dashed blue lines (0.4) represents 5.87% of the 98 percentiles of the PNS score range



Fig. 4 Analytical specificity of PNS against genomic DNA. The y-axis is a relative scale, with 0 representing the mean score of each sample across all input levels. Scores of the downstream model is shown here in the boxplot with the horizontal line in each boxplot indicating the median value. Each gDNA amount was run in triplicate for each sample and 0pcDNA was run in duplicate. The region between the two dashed blue lines (0.4) represents 5.87% of the 98 percentiles of the PNS score range

Table 2	Percentage of i	nterfering	blood	required	to	alte
Percepta	Nasal Swab cal					

PNS Post-Test Risk with-	Relevant Deci- sion Boundary	PNS Post-Test Risk with blood interference		
out blood interference		% Blood 1	% Blood 2	
Low	Downstream model Low Risk	> 31% Low changed to Intermediate	> 14% Low changed to Intermediate	
Intermediate	Downstream model Low Risk Downstream model High Risk Upstream model High Risk	100% Intermediate 100% Intermediate > 49% Intermediate changed to High	100% Intermediate 100% Intermediate > 25% Intermedi- ate changed to High	
High	Downstream model High Risk	> 95% High changed to Intermediate	100% High	
	Upstream model High Risk	100% High	100% High	

as high risk regardless of the blood contamination level, due to the upstream model score, and the classification call remained unchanged under all blood mixture conditions tested. Among all combinations of nasal and blood mixture samples tested (Table 2), the lowest percentage of blood tolerated without causing a change in the PNS result is 14%, suggesting that the PNS classifier can tolerate interference from $\,\leq 14\%$ of whole blood RNA.

To alleviate the effect of blood contamination on PNS classifier scores, a sample exclusion criterion was set to exclude nasal samples with higher than 10% of blood from clinical and analytical validation studies, as well as all future samples. Hemoglobin Subunit Beta (HBB) gene expression level, which is highly correlated with the percent blood spiked-in as shown in Supplementary File 1 Figure S2, was used as the metric for quantifying the blood content in the sample. This blood exclusion criterion was implemented as part of the PNS algorithm. 0% of the nasal RNA samples from the clinical validation and analytical validation sample sets were subject to the blood exclusion criterion. Further, analysis of the blood content in all available nasal swab samples revealed that less than 1% of the nasal swab RNA samples contained more than 5% of blood-derived RNA, with the most extreme samples having around 10% of blood-derived RNA. In other words, this blood-based sample exclusion criterion was developed to proactively safeguard future patient sample testing quality. Given that the maximum blood content tolerated by the PNS classifier is 14%, excluding samples with >10% of blood ensures that future patient samples' PNS results will be highly unlikely impacted by blood contamination.

Discussion

Newly developed molecular diagnostic tests are subject to analytical and clinical validation of performance. Clinical validition of the PNS test has shown it to be an accurate risk assessment tool for patients who have an indeterminate pulmonary nodule and have smoked a minimum of 100 cigarettes in their lifetime [14, 15]. The various studies described in this manuscript evaluated the analytical validity of the PNS test results through examination of technical variation and potential interference that could occur in the full workflow of sample collection, storage, shipping, laboratory processing, and generation of classification scores and calls. In-silico simulation using the clinical validation set scores established the acceptable range of technical variation that would significantly reduce the reported clinical test performance. Then, any impact from experimental conditions on the PNS test was compared against this acceptable range to demonstrate robustness regarding clinical performance.

Following the technical assessment criteria provided by EGAPP and ACCE, analytical reproducibility was established using patient samples with classifier scores covering the entire PNS score range [21]. The same panel of patient samples was processed and tested in two different laboratories: the research laboratory where the test was developed and the CLIA laboratory where future patient samples will be processed, to establish accuracy. Another panel of patient samples was processed in the CLIA laboratory using different lots of reagents, and different equipment used by different operators to establish routine run-to-run variability. PNS achieved 100% accuracy of the test result for patient samples with scores that were 2SD away from the decision boundary. The between-laboratory score variability was estimated to be 0.117 classifier score units, which is 1.72% of the 98 percentiles of the PNS score range. The within-laboratory score variability across multiple runs was also estimated to be 0.117, suggesting the technical variability associated with different laboratories are in line with the routine run-to-run variability of operating in the same laboratory. All evaluated score variabilities were much lower than the acceptance criterion, 0.776, established based on in-silico simulation of the level of variability that can be accepted before clinical test performance is significantly impacted. This further provides evidence that the PNS test can maintain its reported clinical validation performance in the CLIA laboratory under routine operations.

Moreover, analytical sensitivity evaluation demonstrated that the PNS scores did not differ significantly with varying levels of RNA input amount to as low as 15 ng when the nominal input amount is 50 ng. Analytical specificity evaluation demonstrated that PNS test results are robust to potential contaminants including genomic DNA and blood RNA. Specifically, PNS scores showed no impact from up to 30% of genomic DNA in nucleic acid mass while the PNS test tends to yield < 1% genomic DNA content in the standard and routine RNA extraction processes. The PNS test results also remained stable and produced consistent results with interference from up to 14% blood-derived RNA, while > 99.9% of nasal samples had < 5% of blood in the entire training and validation cohort. A blood-based exclusion criterion was established to ensure that only patient samples with less than 10% of blood are tested in the PNS classifier. The potential effects of bacterial or viral contamination were not evaluated, as the sequencing assay employed humanspecific probes designed to target only human sequences. Additionally, bacterial RNA was not detected in the Bio-Analyzer profiles of the nasal samples.

With this study and the clinical validation study [14, 15], the PNS test successfully passed all EGAPP level I analytic validity criteria and demonstrated robust performance.

Conclusions

This study provided evidence that Percepta Nasal Swab test is robust against various technical variabilities and potential contaminants that could be expected in a clinical setting. Percepta Nasal Swab for lung cancer risk assessment of pulmonary nodules in individuals with a history of smoking can be performed in clinical testing with high confidence of accuracy and reproducibility.

Abbreviations

PNS	Percepta Nasal Swab
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
EGAPP	Evaluation of Genomic Applications in Practice and Prevention
ACCE	Analytic validity, Clinical validity, Clinical utility and associated
	Ethical, legal and social implications
SD	Standard Deviation
CI	Confidence Interval
PPV	Positive Predictive Value
NPV	Negative Predictive Value
CV	Coefficients of Variation
CLIA	Clinical Laboratory Improvement Amendments

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12885-025-13683-2.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Author contributions

S.W., R.J., B.B., L.L., P.S.W., G.K., Y.H. and J.H. wrote the main manuscript text. S.W. prepared Table 1, and 2. S.W. and R.J. prepared Figs. 1, 2, 3 and 4. Y.H. and R.J. prepared supplementary files. G.F., M.W. and J.C. performed the experiment. All authors reviewed the manuscript.

Funding

Funding for this study and the publication of this article was provided by Veracyte, Inc. Veracyte, Inc. drafted the study design, executed the studies, oversaw the data analysis and manuscript preparation, and approved the decision to publish.

Data availability

The datasets generated and/or analyzed during the current study are available in the SRA repository, with BioProject number: PRJNA1072245 and link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA1072245?reviewer=ivrrts91vfqnj4t jmjmv561am5.

Declarations

Ethics approval and consent to participate

Samples used in this study were collected as part of the AEGIS study, a multi-center prospective study that included patients with lung nodules who underwent clinically indicated diagnostic bronchoscopy at participating medical centers across the US. Additional samples were collected from Lahey Hospital and Medical Center. Ethics review and IRB approval was obtained by each institution before enrollment and informed consent was obtained from all patients. AEGIS study IRBs are included in the Supplementary file 2. Lahey study samples were approved by Lahey Hospital and Medical Center IRB.

Consent for publication

Not applicable.

Disclosures

RJ, GF, MW, JSC, BB, LL, YH and JH are employees and equity owners of Veracyte, Inc.

Competing interests

All authors were employed by Veracyte, Inc. during the course of this study.

Author details

¹Veracyte, Inc., South San Francisco, CA 94080, USA

Received: 20 December 2023 / Accepted: 7 February 2025 Published online: 31 March 2025

References

- 1. American Cancer Society. Cancer Facts & Fig. 2021. Atlanta: American Cancer Society; 2020. Available at: https://www.cancer.org/research/cancer-facts-stat istics/all-cancer-facts-figures/cancer-facts-figures-2021.html
- Jonas DE, Reuland DS, Reddy SM, et al. Screening for Lung Cancer with Low-Dose Computed Tomography: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA. 2021;325(10):971–87. https://doi.org/10.1001/jama.2021.0377.
- National Lung Screening Trial, Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395–409.
- Gould MK, Tang T, Liu IL, Lee J, Zheng C, Danforth KN, et al. Recent trends in the identification of incidental pulmonary nodules. Am J Respir Crit Care Med. 2015;192(10):1208–14.
- Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, Wiener RS. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e93S-e120S. https://doi.org/10.1378/chest.12-2351. PMID: 23649456; PMCID: PMC3749714.

- Waterfield Price N et al. Prediction of Adenocarcinoma among Other Subtypes of Lung Cancer from CT Using Deep Learning. Journal of Clinical Oncology, vol. 39, no. 15_suppl, 2021, p. 3057. Crossref, https://doi.org/10.120 0/jco.2021.39.15_suppl.3057
- Lee HJ et al. Impact of the Percepta Genomic Classifier on Clinical Management Decisions in a Multicenter Prospective Study. Chest, vol. 159, no. 1, 2021, pp. 401–12. Crossref, https://doi.org/10.1016/j.chest.2020.07.067
- Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelia cell transcriptome. Proc Natl Acad Sci U S A. 2004;101(27):10143–8. https://doi.org/10.1073/pnas.0401422101.
- Wistuba II, Mao L, Gazdar AF. Smoking molecular damage in bronchial epithelium. Oncogene. 2002;21(48):7298–306. https://doi.org/10.1038/sj.onc.120 5806.
- Spira A, Beane JE, Shah V, Sterling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med. 2007;13(3):361–6. https://doi.org/10.1038/nm1556.
- Silvestri GA, Vachani A, Whitney D, Elashoff M, Porta Smith K, Ferguson JS, et al. A bronchial genomic classifier for the diagnostic evaluation of lung cancer. N Engl J Med. 2015;373(3):243–51. https://doi.org/10.1056/NEJMoa1504601.
- Choi Y, Qu J, Wu S, Hao Y, Zhang J, Ning J, et al. Improving Lung Cancer risk stratification leveraging whole transcriptome RNA sequencing and machine learning across multiple cohorts. BMC Med Genet. 2020;13(Suppl 10):151. htt ps://doi.org/10.1186/s12920-020-00782-1.
- Joseph F, Perez-Rogers J, Gerrein C, Anderlind G, Liu S, Zhang Y, Alekseyev KP, Smith D, Whitney WE, Johnson DA, Elashoff SM, Dubinett. Jerome Brody, Avrum Spira, Marc E. Lenburg, for the AEGIS Study Team, Shared Gene expression alterations in nasal and bronchial epithelium for Lung Cancer Detection. JNCI: J Natl Cancer Inst. July 2017;109:djw327. https://doi.org/10.1093/jnci/dj w327.
- Lamb C, Rieger-Christ K, Reddy C, Ding J, Qu J, Wu S, Johnson M, Whitney D, Walsh P, Wilde J, Bhorade S. A nasal clinical-genomic classifier for assessing risk of malignancy in lung nodules demonstrates accurate performance independent of nodule size or cancer stage. Chest. 2021;160(4):A2518. https:/ /doi.org/10.1016/j.chest.2021.08.026.
- Lamb C, Rieger-Christ et al. A nasal genomic classifier to assess Lung Cancer risk in pulmonary nodules, chest 2023 Nov 27:S0012-3692(23)05828-2. https:/ /doi.org/10.1016/j.chest.2023.11.036
- Choi Y, Huang J. Validation of Genomic-Based Assay in Fang L, Su, Cheng, editors Statistical methods in Biomarker and Early Clinical Development, Springer Nature, Switzerland, p. 117–36.
- Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Peper M, Calonge N, et al. The evaluation of genomic applications in practice and prevation (EGAPP) initiative: methods of the EGAPP working group. Genet Med. 2009;11(1):3–14. https://doi.org/10.1097/GIM.0b013e318184137c.
- Sun F, Bruening W, Uhl S, Ballard R, Tipton R, Schoelles K. Quality, regulation, and clinical utility of laboratory-developed molecular tests. Agency for Healthcare Research and Quality, Technology Assessment Program; 2010.
- Johnson MK, Wu S, Pankratz DG, et al. Analytical validation of the Percepta genomic sequencing classifier; an RNA next generation sequencing assay for the assessment of Lung Cancer risk of suspicious pulmonary nodules. BMC Cancer. 2021;21(400). https://doi.org/10.1186/s12885-021-08130-x.
- Hu Z, Whitney D, Anderson JR, Cao M, Ho C, Choi Y, et al. Analytical performance of a bronchial genomic classifier. BMC Cancer. 2016;16:61.
- Dimech W, Bowden DS, Brestovac B, Byron K, James G, Jardine D, et al. Validation of assembled nucleic acid-based tests in diagnostic microbiology laboratories. Pathology. 2004;36:45–50.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.