Original Article

Analytical Validity of a MicroRNA-Based Assay for Diagnosing Indeterminate Thyroid FNA Smears From Routinely Prepared Cytology Slides

Hila Benjamin, MSc[1]; Temima Schnitzer-Perlman, MSc[2]; Alexander Shtabsky, MD, PhD[3,4]; Christopher J. VandenBussche, MD[5]; Syed Z. Ali, MD[5]; Zdenek Kolar, MD, PhD[6]; Fabio Pagni, MD[7]; Rosetta Genomics Group[1,2]; Dganit Bar, PhD[2]; and Eti Meiri, PhD[2]

BACKGROUND: The majority of thyroid nodules are diagnosed using fine-needle aspiration (FNA) biopsies. The authors recently described the clinical validation of a molecular microRNA-based assay, RosettaGX Reveal, which can diagnose thyroid nodules as benign or suspicious using a single stained FNA smear. This paper describes the analytical validation of the assay. **METHODS:** More than 800 FNA slides were tested, including slides stained with Romanowsky-type and Papanicolaou stains. The assay was examined for the following features: intranodule concordance, effect of stain type, minimal acceptable RNA amounts, performance on low numbers of thyroid cells, effect of time since sampling, and analytical sensitivity, specificity, and reproducibility. **RESULTS:** The assay can be run on FNA slides for which as little as 1% of the cells are thyroid epithelial cells or from which only 5 ng of RNA have been extracted. Samples composed entirely of blood failed quality control and were not classified. Stain type did not affect performance. All slides were stored at room temperature. However, the length of time between FNA sampling and processing did not affect assay performance. There was a high level of concordance between laboratories (96%), and the concordance for slides created from the same FNA pass was 93%. **CONCLUSIONS:** The microRNA-based assay was robust to various physical processing conditions and to differing sample characteristics. Given the assay's performance, robustness, and use of routinely prepared FNA slides, it has the potential to provide valuable aid for physicians in the diagnosis of thyroid nodules. *Cancer Cytopathol* 2016;124:711-21. © 2016 Rosetta Genomics. Cancer Cytopathology published by Wiley Periodicals, Inc. on behalf of American Cancer Society.

KEY WORDS: fine-needle aspiration (FNA); indeterminate; microRNA; molecular test; nodules; smears; thyroid.

INTRODUCTION

Fine-needle aspiration (FNA) biopsy is the most widely used method for the diagnosis of thyroid nodules. The evaluation of stained FNA smears leads to a definitive benign or malignant diagnosis in the majority of cases. However, 10% to 40% of FNAs are classified as cytologically indeterminate according to the Bethesda reporting

system.[1] Indeterminate classification includes 3 categories: atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) (Bethesda category III), follicular neoplasm or suspicious for a follicular neoplasm (FN/SFN) (Bethesda category IV), and suspicious for malignancy (SM) (Bethesda category V). The majority of patients with indeterminate cytology are referred to surgery regardless of their risk of malignancy, which is 5% to 15%, 15% to 30%, and 60% to 75% for Bethesda categories III, IV, and V, respectively.[1,2] Therefore, improving diagnostic accuracy for patients with indeterminate cytology would benefit many of these patients, reducing the rates of surgery and the subsequent risk for metabolic and anatomic complications (such as hypoparathyroidism and laryngeal nerve injury), as well as the resulting life-long requirement of thyroid hormone supplementation.

To overcome the diagnostic limitations of FNA cytology and to spare patients with benign nodules unnecessary surgeries, several molecular tests have been developed. Some of these tests use gene point mutations (eg, BRAF and RAS) and gene rearrangements (eg, RET/PTC, PAX8/PPARG) associated with thyroid cancer.[3,4] Another test uses the expression profiles of 167 genes to classify indeterminate nodules.[5] More recently, a molecular test combining DNA, messenger RNA and microRNA (miRNA) biomarkers was developed for the diagnosis of nodules classified as Bethesda categories III and IV.[6] A major disadvantage of these tests is that they all require fresh FNA tissue or special collection and shipment conditions. To overcome these limitations, we developed a molecular test that can diagnose indeterminate thyroid nodules using stained FNA smears prepared for initial cytopathologic evaluation. In this test, expression levels of a set of miRNA biomarkers are measured and used for classification of thyroid nodules as benign or suspicious.[7] miRNAs are an important class of noncoding RNA implicated in gene expression regulation, and their expression profiles have been associated with the pathophysiology of cancer, including thyroid cancer.[8–16] It has been demonstrated that miRNA expression profiles can differentiate between malignant and benign thyroid lesions in resected nodules as well as FNA samples.[14] miRNAs have also been shown to be extremely stable and remain intact in various clinical samples.[17] This, as well as their important role in thyroid cancer, makes miRNAs excellent biomarkers for the diagnosis of indeterminate thyroid nodules.

Air-dried Romanowsky-type (eg, Giemsa, Diff-Quik) stained slides are commonly used for the evaluation of cytoplasm or extracellular material; and alcohol-fixed, Papanicolaou-stained slides are used for the optimal assessment of nuclear details.[18] The test was developed and validated on FNA smears that were stained with both of these stain types.

Test performance was determined in a previously completed, multinational and multicenter, blinded validation study.[7] In the current study, the test's robustness and compatibility with a variety of FNA smear processing and staining techniques were demonstrated. The ability of the assay to work on minute amounts of thyroid cells originating in a single FNA smear slide was tested, no effect on correct classification rates was found. Finally, the test was able to yield accurate results in the presence of blood in the sample.

MATERIALS AND METHODS

Patients and Samples

De-identified, retrospective FNA smear samples with corresponding histologic diagnoses from 576 nodules were obtained for the assay's training and clinical validation studies and were profiled as previously described.[7] A single slide from each nodule was used for the assay training and clinical validation studies. The analytical validation utilized these as well as additional slides from the above-mentioned nodules (for a description of all slides used in this study, see Table 1). Whole blood was collected from 17 de-identified donors and was used to create 48 nonstained blood smear slides (10 µL per slide). To study the effect of the time to processing on miRNA expression, 5 additional Bethesda II thyroid FNAs were cut in half; 1 half of each slide was immediately extracted and assayed, and the other half was stored at room temperature for 54 days and then extracted and assayed. The samples used in this study were collected under approval of the Institutional Review Board Committee, and a waiver of consent was obtained from the Institutional Review Board.

Test Quality Control

Negative controls consisting of double-distilled water were processed alongside clinical samples. Two negative controls were used: One was processed starting at the RNA extraction stage, and the other was processed from the complimentary DNA (cDNA) synthesis stage. These negative controls ensured that samples were not contaminated during either of these stages.

TABLE 1. Analytical Validation Studies and Samples

| Section	Objective	Experiment	No. of Samples
1	Assay concordance on slides taken from the same nodule	Assay protocol was run on 139 samples taken from 59 FNA passes on 49 nodules; assay classification was compared within each pass	139
2A	Assay performance using different RNA amounts	Assay protocol was repeated on the same RNA using 4 different initial RNA amounts for cDNA synthesis	12
2B	Sensitivity of the PCR platform	Linearity of miRNA expression was examined on a series of 8 dilutions of positive control RNA	1
3A	Effect of FNA sample time to processing on assay results	Slides were split in half, with the first half processed within 12 d and the second half processed after 54 d; miRNA expression profiles and assay classifications were compared	5
3B		Slides processed within 1 year of sampling were compared with slides that had a longer time to processing; classification correctness was compared in clinical validation set samples that passed assay QC	189
4A	Assay performance on blood and slides with low number of thyroid cells	Assay was run on whole blood smears	48
4B		Assay performance was examined on 1 benign and 1 malignant thyroid sample, which were mixed with different amounts of blood	2
4C		Assay classification correctness was compared between samples with many lymphocytes and samples with few lymphocytes in slides representing 36 nodules from the assay training and clinical validation studies	188
4D		Assay sensitivity and specificity were evaluated on samples that had 30-500 thyroid cells	106
5A	Intralaboratory reproducibility	Technician work-flow reproducibility was tested for RNA extraction, cDNA synthesis, and qRT-PCR quantification	2
5B		qRT-PCR reproducibility was tested on repeats of a positive control sample	1
6A	Interlaboratory reproducibility	Overall process concordance was examined for RNA extraction, cDNA preparation, and qRT-PCR	27
6B		Concordance for preparation of cDNA followed by qRT-PCR was tested	93
6C		qRT-PCR concordance was tested	48

Abbreviations: cDNA, complimentary DNA; FNA, fine-needle aspiration; miRNA, microRNA; PCR, polymerase chain reaction; QC, quality control; qRT-PCR, quantitative reverse transcriptase-polymerase chain reaction.

The assay's positive control consisted of a mix of RNAs extracted from several benign and malignant samples of thyroid formalin-fixed, paraffin-embedded (FFPE) resections, which were augmented with RNA extracted from whole blood to ensure detectable expression levels of all miRNAs measured by the assay. FFPE samples were used as positive controls to ensure high expression levels of all 24 assay miRNAs in these samples. In addition, resection tissue provides a high miRNA yield, facilitating use of the same positive control sample for extended periods of time. A positive control sample was run alongside each batch of clinical samples to ensure that the cDNA and polymerase chain reaction (PCR) reagents yielded consistent results. Each sample was profiled in duplicate. If repeats differed beyond an allowed range, then PCR was repeated on the sample.

The assay analyzed thyroid FNA smears to detect miRNA blood markers, thyroid epithelial markers, and malignant/benign markers. Samples that displayed high blood marker levels and low epithelial marker levels were deemed as not having enough thyroid tissue, were not classified, and were reported as failed. In addition, samples in which malignant/benign marker levels were below an allowed threshold were failed for insufficient RNA levels and were not classified.

Classifier

The assay classifier measured 24 miRNAs and combined several linear discriminant analysis (linear DA) steps and a K-nearest neighbor classifier step to differentiate between benign samples and samples that were "suspicious for
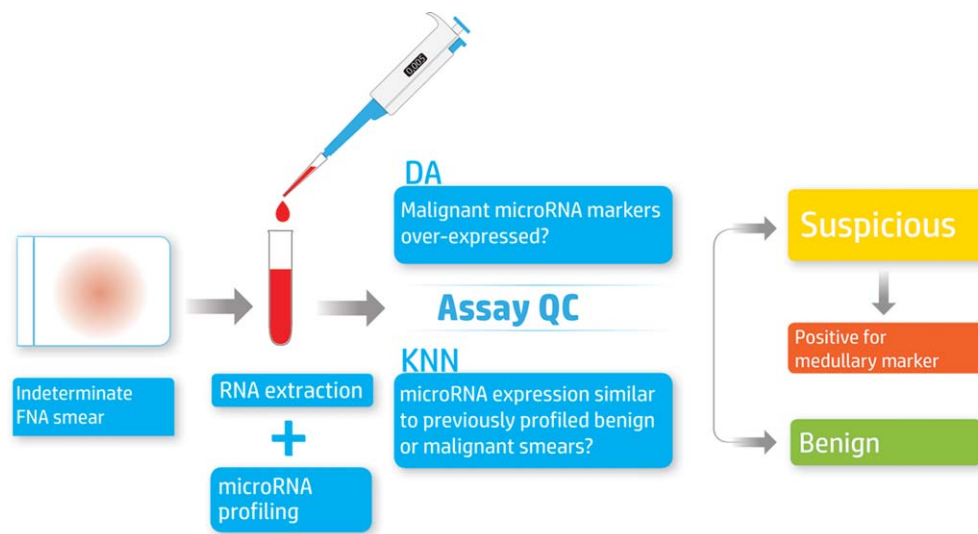
Figure 1. overview of the molecular microRNA (miRNA)-based assay. RNA is extracted from a single stained diagnostic smear and is used to synthesize complimentary DNA (cDNA). Expression levels of 24 miRNAs are quantified using quantitative reverse transcriptase-polymerase chain reaction. The sample is processed along with 2 negative controls and 1 positive control. If either the positive or negative controls fail, the process is repeated. Assay quality control (QC) ensures that samples with low epithelial and malignant/benign markers fail and are not classified. Sample classification combines several linear discriminant analysis (DA) steps and a K-nearest neighbor (KNN) classifier step (K = 9). Possible classifications are: suspicious for malignancy by miRNA markers, positive for medullary marker carcinoma, and benign.

malignancy by miRNA profiling". Medullary carcinoma samples were classified by a linear discriminant analysis step based on hsa-miR-375 expression. Samples classified in this step received a final classification of "positive for medullary marker". The assay's performance was based on miRNA correlations and thus did not require miRNA expression level normalization.

Intra-laboratory and inter-laboratory reproducibility studies

The assay was developed in Rosetta Genomics' research and development laboratory (Rosetta Genomics Ltd, Rehovot, Israel-RG-IL) and was then transferred to the Rosetta Genomics' Philadelphia laboratory (Rosetta Genomics, Inc, Philadelphia, Pa-RGL-US), which is certified under the Clinical Laboratory Improvement Amendments (CLIA) Act of 1988 to perform high-complexity testing and is accredited by the College of American Pathologists. Both laboratories participated in the inter-laboratory reproducibility studies.

Interlaboratory overall process concordance

The study was run on 2 types of tissue samples: FNA-like samples ($1/4$ of a 5-micron FFPE resection sample diluted with whole blood; n = 8) and FNA smear slides that were

cut in half, with half of the slide extracted and quantified in each laboratory (total = 27 samples).

Statistical Analysis

Pearson correlation was used to compare miRNA expression profiles between samples. The chi-square test was used for comparisons of correctness of classification under different slide characteristics, and the coefficient of determination (R^2) metric was used to assess marker linearity between different RNA concentrations.

miRNA Marker Linearity

A series of 8 dilutions of positive control RNA (dilution factor = 2.5) was used to reach a final RNA quantity range of 2×10^{-3} ng to 1.25 ng. Assay markers were quantified in triplicate. Linearity of assay markers in the relevant cycle threshold (C_T) range was evaluated by calculating efficiency and R^2 values. PCR efficiency was calculated using the following formula: efficiency $= (2^{[-1/slope]} - 1) \times 100$, where slope denotes the function slope of the linear fit between cycle threshold (C_T) values and $\log_2$ (RNA quantity).

RESULTS

The assay discussed in this study was designed to diagnose indeterminate thyroid FNA smears as benign or suspicious

TABLE 2. Assay Results at Different Levels of Initial RNA

Sample	Initial RNA Level				Histologic Type
	5 ng	10 ng	20 ng	40 ng	
FNA-1	Failed assay QC	Failed assay QC	Benign	Benign	Nodular hyperplasia
FNA-2	Benign	Malignant	Malignant	Malignant	Papillary carcinoma
FNA-3	Malignant	Malignant	Malignant	Malignant	Papillary carcinoma
FNA-4	Benign	Benign	Benign	Benign	Papillary carcinoma
FNA-5	Malignant	Malignant	Malignant	Malignant	Papillary carcinoma
FNA-6	Benign	Benign	Benign	Benign	Follicular adenoma
FNA-7	Failed assay QC	Malignant	Malignant	Benign	Follicular adenoma
FNA-8	Failed assay QC	Benign	Benign	Benign	Nodular hyperplasia
FNA-9	Malignant, medullary	Malignant, medullary	Malignant, medullary	Malignant, medullary	Medullary carcinoma
FNA-10	Benign	Benign	Benign	Benign	Nodular hyperplasia
FNA-11	Malignant	Malignant	Malignant	Malignant	Papillary carcinoma
FNA-12	Benign	Benign	Benign	Failed assay QC	Nodular hyperplasia (nodular goiter)

Abbreviations: FNA, fine-needle aspiration; QC, quality control.

for malignancy by miRNA profiling.[7] For the latter, the test also indicates whether or not a medullary marker is highly expressed. The testing process (described in Fig. 1) requires a single, routinely prepared, diagnostic stained smear stored at room temperature. In this report, we describe the analytical validation of the assay, its reproducibility and robustness. A list of all studies performed is provided in Table 1.

Intranodule Concordance

To assess intranodule concordance, multiple slides originating from the same FNA pass were examined. There were 59 passes in which 2 slides or more were defined as diagnostic by the referring cytologist and that also passed assay quality control (QC). Of these, 32 passes represented 2 stain types or more. In this analysis, each slide's classification was compared with the majority of classifications given to slides within the same pass. In passes that had only 2 samples and disagreeing classifications, 1 classification was chosen randomly as the majority classification, and the other classification was considered a disagreement. On the basis of this analysis, 129 of 139 slides agreed with the majority of classifications within their pass (classification agreement, 92.81%). When analyzing the 32 passes comprised of two or more stain types, 73 of 80 slides agreed with the majority classification (classification agreement, 91.25%). These results indicate the assay's reproducibility between different slides taken from the same pass, regardless of stain type.

Assay Performance With Minute RNA Amounts

To determine the amount of RNA recommended for use by the assay protocol, 12 FNA smears were extracted (6 benign nodules and 6 malignant nodules, including 1 medullary carcinoma) and were processed with 4 different initial RNA amounts for the cDNA synthesis reaction: 5 ng, 10 ng, 20 ng, and 40 ng (Table 1, section 2A; Table 2). At 5 ng of initial RNA material, assay QC failed in only three of 12 samples, whereas 1 sample or none failed QC criteria for amounts of 10, 20, and 40 ng. Based on discovery studies (data not shown), we observed that the majority of FNA slides contained at least 20 ng of RNA; therefore, 20 ng was set as the default RNA amount for processing in the assay protocol. Since assay classification was consistent in all tested initial RNA amounts (see below and Table 2), it was concluded that slides with <20 ng initial RNA could also be processed by the assay. Within the assay's consecutive clinical validation set, only 11 of 201 slides had <20 ng of RNA, and only one of those 11 slides failed because of insufficient thyroid material (2 ng). The rate of QC failure for samples with <20 ng in the clinical validation was 9%, whereas the overall failure rate for all samples was 6%.

Median miRNA expression levels, as measured by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR), in the 12 FNA samples described above, displayed linearity between different RNA amounts (minimum $R^2 = 0.96$), as illustrated in Figure 2. This linear association between initial RNA amount and median miRNA expression indicates that the assay miRNA profiles
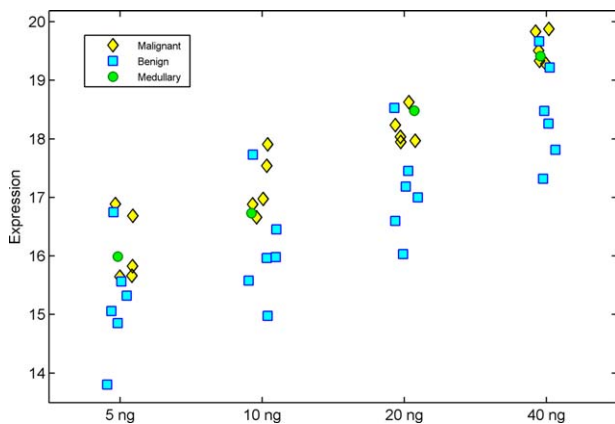
Figure 2. Median Expression levels over different amounts of initial RNA are illustrated. Expression levels from 12 fine-needle aspiration samples were measured at 4 initial RNA concentrations: 5 ng, 10 ng, 20 ng, and 40 ng. Median miRNA levels were calculated for each sample and at each initial RNA amount over all assay miRNAs. The coefficient of determination (R^2) was used to assess the linearity of median miRNA expression levels for each sample over all initial RNA amounts. The minimum R^2 was 0.96. In order to visualize the median expression of miRNAs, inverted cycle threshold (C_T) values are shown: The C_T values were subtracted from 50, so that high values represent high median expression. Yellow diamonds denote malignant samples; blue squares, benign samples; green circles, medullary samples.

for both benign and suspicious samples are consistent over this range of RNA amounts. Assay classification for slides that passed assay QC was also consistent in 11 of 12 samples.

To test the sensitivity of the PCR platform and the linear range of markers, a series of eight dilutions of positive control RNA was used (Table 1, section 2B). Assay sensitivity was determined by the lowest RNA amount that gave a detectable signal. All 24 miRNAs were detectable at 1.28×10^{-2} ng. The efficiency for all markers ranged between 95.4% and 110.1%. R^2 for all markers was >0.982.

Assay Stability Over Time

The assay was designed to process routinely prepared FNA slides. To account for possible delays in sample receipt, assay robustness was tested on recently prepared slides as well as on slides that were processed several weeks after preparation. Five recently prepared cytologic slides were cut in half. One half of each slide was extracted and processed immediately upon receipt (within 12 days of sampling), and the second half was kept at room temperature for 54 days and then processed. Analysis demonstrated that

miRNA expression profiles were highly correlated between slide halves that were processed immediately and slide halves that were processed after nearly 8 weeks. The minimal Pearson correlation between samples was 0.988. Assay classification for both half slides was identical in all 5 samples (Fig. 3; see Table 1, section 3A).

Furthermore, an analysis conducted on 189 clinical slides which were stored at room temperature from the time of preparation, revealed no significant effect of slide age on assay performance ($P = 0.158$) in slides where the time to processing was <1 year (n = 48) relative to slides where time to processing was between 1 year and 11 years (n = 141). The results of this analysis highlight the assay's capability to give correct results for both recently processed and older FNA smears (see Table 1, section 3B).

Analytical Sensitivity and Specificity

Assay performance on samples with low numbers of thyroid cells

Since FNA smear samples are rich in blood and contain very few thyroid cells, assay sensitivity and specificity were tested in studies that simulated various levels of blood and thyroid tissue (see Table 1, sections 4A-4D). First, to evaluate assay QC on samples with no thyroid cells, the assay was tested on 48 smears of whole blood. Due to lack of thyroid material, all 48 samples were given a final status of failed assay QC for epithelial and malignant/benign markers and were not given a classification (Fig. 4).

Next, test performance was evaluated on simulated concentrations of blood and thyroid tissue. Two FFPE resection samples originating in thyroid nodules (one benign and one malignant) were diluted with various amounts of blood. Each sample was classified by the assay at several levels of dilution: 100% thyroid-derived RNA (0% whole blood-derived RNA), 50% thyroid-derived RNA (50% whole blood-derived RNA), and 10% thyroid-derived RNA (90% whole blood-derived RNA). The assay correctly classified both samples at all 3 dilution points.

To assess the influence of lymphocytes on assay performance, we examined a set of 188 FNA slides on which lymphocyte levels had been quantified. Correct classification by the assay was tested in slides with many lymphocytes relative to slides with few lymphocytes. The results of a chi-square test indicated no statistically
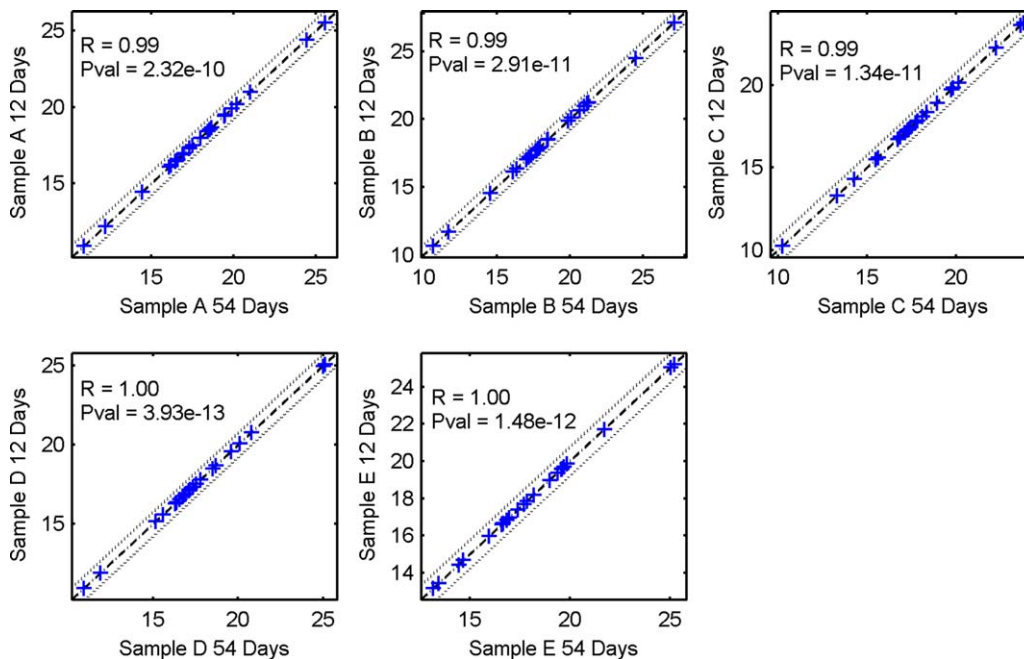
Figure 3. The correlation in expression profiles of recent and older samples. Expression profiles for 5 recently prepared slides that were cut in half are illustrated. One half of each slide was processed immediately, and the other half was stored at room temperature for 54 days and then processed. Each blue cross represents the expression level of a single microRNA (miRNA) measured by the assay. The x-axis represents slide halves that had a longer time to processing, and the y-axis represents recently processed slide halves. Dashed lines represent a 1.5-fold change in expression. In order to visualize the differential expression of miRNAs, inverted cycle threshold (C_T) values are shown: The C_T values were subtracted from 50, so that high values represent high expression. These charts indicate that expression levels were identical in both halves of each slide. Pval indicates *P* value; R, correlation coefficient.

significant link between the amount of lymphocytes present in the sample and the correctness of assay diagnosis ($P = .87$).

Finally, to evaluate assay performance on clinical FNA samples with low numbers of thyroid cells, the assay was applied to FNA smears that had minute amounts of thyroid tissue. Figure 5 illustrates the sensitivity and specificity of the assay on 106 slides that had up to 60, 100, and 500 thyroid cells (as determined by the referring cytologists). Each of these samples consisted of 98% to 99% blood and 1% to 2% thyroid cells. Assay classification was compared to a post-surgical diagnosis, which was agreed upon by 3 independent pathologists. When examining slides with up to 60 cells, 33 of 36 samples passed assay QC for sufficient epithelial and malignant/benign markers. The assay's performance on these samples (7 malignant and 26 benign) exhibited high levels of sensitivity and specificity, demonstrating that, while the assay ensured QC failure of samples that had no thyroid material, it correctly classified FNA smears that had very small amounts of thyroid material.

Analytical reproducibility

To demonstrate the reproducibility of the assay, a series of intra-laboratory and inter-laboratory studies was conducted.

Intra-laboratory reproducibility studies. To evaluate the level of reproducibility between technicians, PCR instruments, and positive control cDNA, intralaboratory assay reproducibility was assessed in 3 studies. An outline of the study design is described in Table 1 (sections 5A and 5B).

The objective of the first study was to evaluate the reproducibility of the entire laboratory process (see Fig. 6). Two FFPE samples were quantified 4 times by different technicians on different days using different PCR instruments. Reference quantifications of both samples were processed from start to finish by a single technician.

To demonstrate the reproducibility of RNA extraction between technicians, RNA was extracted from each of the 2 samples by 2 different technicians on different days, and cDNA from each extraction was then synthesized and quantified by a single technician. Expression levels were
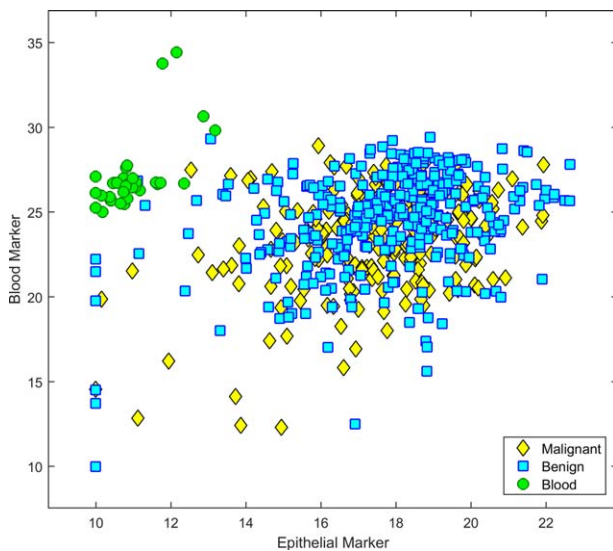
Figure 4. Assay separation of blood smears from thyroid fine-needle aspiration (FNA) smears is illustrated. Expression profiles of thyroid epithelial markers and blood markers are shown in whole blood samples and in thyroid FNA samples from the assay training and clinical validation sets.[7] The data sets included 229 malignant samples and 330 benign samples. In order to visualize the differential expression of microRNAs, inverted cycle threshold (C_T) values are shown: The C_T values were subtracted from 50, so that high values represent high expression. Yellow diamonds denote malignant samples; blue squares, benign samples; green circles, whole blood smears. Blood and thyroid epithelial markers clearly separate whole blood smears from thyroid samples.

Figure 5. Assay performance on samples that had various amounts of thyroid cells. The assay's performance was tested on a set of 106 samples in which the number of thyroid cells ranged from 30 to 500 (as evaluated by the referring cytologists). The assay sensitivity and specificity are described for samples in which the number of thyroid cells was $\leq$60 (n = 36), $\leq$100 (n = 64), and $\leq$500 (n = 106). The y-axis represents sensitivity and specificity values. The x-axis represents the cumulative number of cells. Within the range tested, assay sensitivity and specificity were not affected by the number of thyroid cells on the slide.
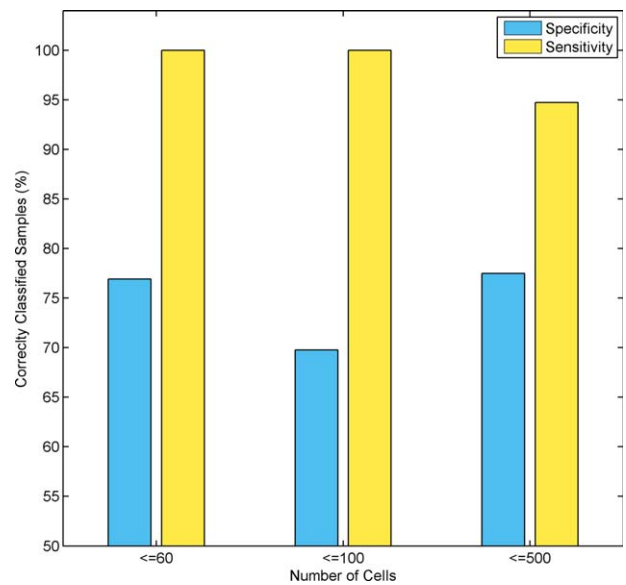
compared within each sample, and the minimum Pearson correlation was 0.99.

To demonstrate cDNA synthesis reproducibility, cDNA was synthesized by 2 technicians from RNA extracted by a single technician. The 2 cDNA preparations from each sample were profiled by qRT-PCR (by a single technician), expression profiles were compared within each sample, and the minimum Pearson correlation was 0.98.

To demonstrate qRT-PCR and cDNA synthesis reproducibility, RNA was extracted from each sample by a single technician. cDNA synthesis and miRNA were quantified by qRT-PCR performed by 2 technicians, expression profiles were compared within each sample, and the minimum Pearson correlation was 0.99.

The assay classifier was then run on the resulting 4 repeats of each sample. Each sample was correctly classified all 4 times, showing the robustness of the assay for all stages of miRNA preparation and quantification.

The second study was conducted using positive control samples. Eleven positive control cDNA samples were

synthesized from assay positive control RNA and quantified by three technicians. Each cDNA sample was quantified twice, resulting in 22 positive control expression values (C_T) for each of the assay miRNA. The minimal Pearson correlation between all pairs of positive control reference samples was 0.99 ($P < .001$), demonstrating complete intra-laboratory reproducibility of cDNA synthesis and qRT-PCR under assay conditions.

Interlaboratory reproducibility study. To assess the interlaboratory reproducibility of the final classification given by the assay, samples were run in 2 laboratories (RGL-US and RG-IL), classified by the assay, and compared. Three laboratory-to-laboratory concordance studies were run on a total of 168 samples (Table 1, sections 6A-6C).

The first study assessed concordance of the entire process on 27 FNA and FNA-like samples. RNA extraction, cDNA synthesis, and qRT-PCR quantification were performed in each laboratory according to the assay protocol, and samples were classified by the assay. Twenty-five
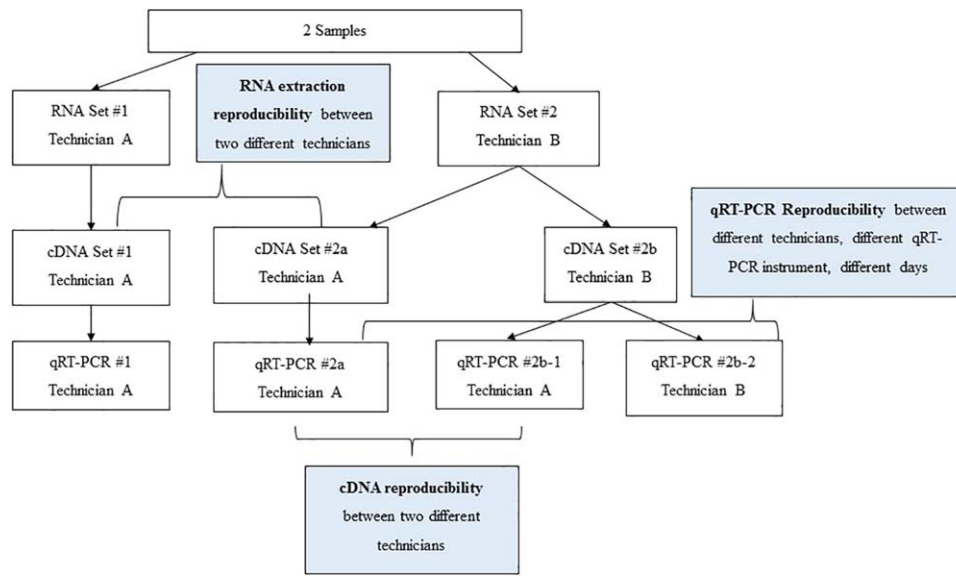
Figure 6. The scheme of intra-laboratory validation is illustrated. For RNA extraction reproducibility, 2 formalin-fixed, paraffin-embedded samples were extracted by 2 technicians on different days. Complimentary DNA (cDNA) was synthesized from each sample and quantified by a single technician. Expression levels were compared to demonstrate the reproducibility of RNA extraction between technicians. For cDNA reproducibility, cDNA was synthesized by 2 technicians from the same RNA sample. The 2 cDNA preparations were then profiled by a single technician using quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR), and expression profiles were compared to demonstrate the reproducibility of cDNA synthesis between technicians. For qRT-PCR reproducibility, RNA was extracted and cDNA was synthesized by a single technician. MicroRNA was quantified by 2 technicians, and expression profiles were compared to demonstrate reproducibility between PCR plates, PCR instruments, and technicians on different days.

samples passed assay QC; and, of these, 24 received an identical classification in both laboratories, resulting in 96% concordance and a Pearson correlation of 0.90 for assay miRNA expression ($P < .001$).

In the second study, concordance of cDNA preparation followed by qRT-PCR was tested for 93 samples, representing 7 histologic types: follicular carcinoma, follicular adenoma, Hashimoto thyroiditis, medullary carcinoma, nodular hyperplasia (nodular goiter), papillary carcinoma, and Graves disease. In this study, RNA was extracted in the RG-IL laboratory, and subsequent steps were performed independently in both laboratories (RG-IL and RGL-US).

Eleven samples failed assay QC. Of the remaining 82 samples, 76 received the same classification in the 2 laboratories, representing 92.68% agreement. A minimal Pearson correlation of 0.974 over all assay miRNAs was observed ($P < .001$) after removing 1 sample from analysis based on a single outlier miRNA.

In the third study, the concordance of qRT-PCR between RG-IL and RGL-US was tested on 48 cDNA samples, representing 6 histologic types: follicular carcinoma, follicular adenoma, Hashimoto thyroiditis, medul-

lary carcinoma, nodular hyperplasia (nodular goiter), and papillary carcinoma. Classification of 40 of the 44 samples that passed assay QC was identical between laboratories (90.9% agreement). The minimal Pearson correlation for assay miRNAs in these 48 samples was 0.95 ($P < .001$). Because the current study was performed at a time when the laboratory technicians were at the beginning of their qRT-PCR training, this level of concordance may be an underestimation of qRT-PCR concordance. That possibility is supported by the high concordance measured in the entire assay process study, despite the greater number of factors in the entire process study relative to this study.

DISCUSSION

Diagnosis of thyroid nodules as benign or malignant is an important and challenging task. Molecular diagnostic tests have been in use to aid this diagnostic challenge, but they require non-routine collection or shipment procedures. In the current study, we tested over 200 clinical FNA samples to examine the technical characteristics of RosettaGX Reveal[TM], a molecular test that works on routinely used, stained thyroid FNA smears, and its ability to process such samples and return accurate diagnostic results. We

demonstrated that the assay is able to work with various routinely used stains, even in particularly challenging cases, such as when minute amounts of RNA are present or when only blood is present, using appropriate QC measures.

Analytical validation studies establishing the robustness of a diagnostic test over a variety of technical conditions are of utmost importance. In a previous study, we demonstrated the clinical validity of the assay.[7] In the current study, we demonstrated robustness of the assay for all stages of sample processing, beginning with different types of routinely stained FNA slides, through handling conditions, and ending with laboratory processing and assay classification. We have shown that assay results are not affected by various physical processing conditions, including laboratories, reagents, technicians, time to processing, and PCR instruments. We have also shown that the assay is stable across multiple samples taken from the same FNA pass, regardless of staining method.

Throughout the analytical validation of the assay, we aimed to achieve a high level of certainty regarding the reproducibility and stability of assay results. To this end, we processed a total of 168 FNA samples for inter-laboratory validation. This validation study demonstrated high concordance between slide halves processed from start to finish in the research and development laboratory and slide halves run from start to finish in the CLIA approved CL laboratory (classification agreement was 96% between laboratories). The assay also exhibited high stability for samples that were extracted in 1 laboratory and processed in both laboratories and between samples for which qRT-PCR was run in 2 laboratories after extraction and cDNA synthesis was run in a single laboratory (classification agreement, 92.68% and 90.9%, respectively). These results demonstrate a high level of reproducibility.

The assay was designed to diagnose FNA smear slides and does not require an additional needle pass or special transport tube. Any representative stained slide from an FNA pass can be used for assay classification. The use of an existing slide for molecular testing has the additional advantage of running the assay on the same slide that was evaluated by the referring cytopathologist. This ensures that the assay's results correspond to the relevant thyroid material in question. Use of the same slide for cytologic evaluation (cell composition and quantity) and assay processing further substantiates the analytical results pertaining to slide characteristics observed in this study.

The FNAs used in this study and in the assay's clinical validation set were comprised of approximately 98% to 99% blood and approximately 1% to 2% thyroid cells. A major issue when working with FNAs is the small amount of diagnostic material. In this study, we have demonstrated that the assay gives correct classifications for FNAs with minute amounts of thyroid material. A designated experiment was conducted on 12 thyroid samples using different initial RNA amounts. MiRNA signals exhibited linearity across the RNA quantity range, and classification consistency was also similar for all RNA amounts. We have also demonstrated that, when the assay was run on samples that were adequate for diagnosis, the QC failure rate was very low (6%).[7] These data suggest that the assay is suitable for the clinical setting, with low failure rates and the ability to work with extremely low RNA amounts while maintaining high accuracy.

Other molecular tests have been shown to mistakenly diagnose whole blood samples as suspicious for malignancy or as benign.[19] Although the assay described in the current report can correctly classify samples with low numbers of thyroid cells, it does not mistakenly diagnose samples that do not contain any thyroid material.

The samples used for clinical validation of the assay were FNA slides stored at room temperature that ranged in age from a few months to approximately 11 years. To determine whether the time to processing was a factor in the correctness of assay classification, we compared older slides to more recently prepared slides and observed no significant difference in assay performance. These results align with published data regarding the stability of miRNAs in clinical samples.[17]

In this study, we have demonstrated the stability of the RosettaGX Reveal assay to conditions often encountered during clinical sample processing. We have also demonstrated that the assay is robust to the qualities and characteristics of samples used in assay development and validation. The use of a validated diagnostic assay that works with routinely used thyroid FNA smears can aid physicians in diagnostic decisions pertaining to course of treatment.

FUNDING SUPPORT

CONFLICT OF INTEREST DISCLOSURES

AUTHOR CONTRIBUTIONS

Hila Benjamin: Conceptualization, methodology, software, validation, formal analysis, data curation, writing–original draft, and visualization. **Temima Schnitzer-Perlman**: Software, validation, formal analysis, data curation, writing–original draft, and visualization. **Alexander Shtabsky**: Resources, writing–review and editing, and investigation. **Syed Z. Ali**: Resources, writing–review and editing, and investigation. **Christopher J. VandenBussche**: Resources, writing–review and editing, and investigation. **Zdenek Kolar**: Resources, writing–review and editing, and investigation. **Fabio Pagni**: Resources, writing–review and editing, and investigation. **Rosetta Genomics Group**: Validation, investigation, data curation, writing–review and editing, and project administration. **Dganit Bar**: Conceptualization, methodology, writing–original draft, supervision, and project administration. **Eti Meiri**: Conceptualization, methodology, writing–review and editing, supervision, and project administration.

REFERENCES

1. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol.* 2012;56:333-339.

2. Cibas ES, Ali SZ; NCI Thyroid FNA. State of the Science Conference. The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol.* 2009;132:658-665.

3. Nikiforov YE, Carty SE, Chiosea SI, et al. Impact of the multi-gene ThyroSeq next-generation sequencing assay on cancer diagnosis in thyroid nodules with atypia of undetermined significance/follicular lesion of undetermined significance cytology. *Thyroid.* 2015;25:1217-1223.

4. Beaudenon-Huibregtse S, Alexander EK, Guttler RB, et al. Centralized molecular testing for oncogenic gene mutations complements the local cytopathologic diagnosis of thyroid nodules. *Thyroid.* 2014;24:1479-1487.

5. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med.* 2012;367:705-715.

6. Labourier E, Shifrin A, Busseniers AE, et al. Molecular Testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. *J Clin Endocrinol Metab.* 2015;100:2743-2750.

7. Bar D, Lithwick-Yanai G, Goren Y, et al. A first-of-its-kind, microRNA-based diagnostic assay for accurate thyroid nodule classification. Poster presented at: 15th International Thyroid Congress and 85th Annual Meeting of the American Thyroid Association; October 19, 2015; Lake Buena Vista, FL.

8. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer.* 2006;6:857-866.

9. Farazi TA, Spitzer JI, Morozov P, Tuschl T. miRNAs in human cancer. *J Pathol.* 2011;223:102-115.

10. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature.* 2005;435:834-838.

11. Pallante P, Visone R, Ferracin M, et al. MicroRNA deregulation in human thyroid papillary carcinomas. *Endocr Relat Cancer.* 2006;13: 497-508.

12. He H, Jazdzewski K, Li W, et al. The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci U S A.* 2005;102: 19075-19080.

13. Weber F, Teresi RE, Broelsch CE, Frilling A, Eng C. A limited set of human MicroRNA is deregulated in follicular thyroid carcinoma. *J Clin Endocrinol Metab.* 2006;91:3584-3591.

14. Nikiforova MN, Tseng GC, Steward D, Diorio D, Nikiforov YE. MicroRNA expression profiling of thyroid tumors: biological significance and diagnostic utility. *J Clin Endocrinol Metab.* 2008;93: 1600-1608.

15. Chou CK, Chen RF, Chou FF, et al. miR-146b is highly expressed in adult papillary thyroid carcinomas with high risk features including extrathyroidal invasion and the BRAF(V600E) mutation. *Thyroid.* 2010;20:489-494.

16. Cahill S, Smyth P, Finn SP, et al. Effect of ret/PTC 1 rearrangement on transcription and post-transcriptional regulation in a papillary thyroid carcinoma model [serial online]. *Mol Cancer.* 2006;5:70.

17. Xi Y, Nakajima G, Gavin E, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA.* 2007;13:1668-1674.

18. Bibbo M, Wilbur D. Comprehensive Cytopathology. 3rd ed. Philadelphia, PA: Saunders; 2008.

19. Walsh PS, Wilde JI, Tom EY, et al. Analytical performance verification of a molecular diagnostic for cytology-indeterminate thyroid nodules. *J Clin Endocrinol Metab.* 2012;97:E2297-E2306