OPEN ACCESS

Multicentre validation of a microRNA-based assay for diagnosing indeterminate thyroid nodules utilising fine needle aspirate smears

Gila Lithwick-Yanai,[1] Nir Dromi,[1] Alexander Shtabsky,[2,3] Sara Morgenstern,[3,4] Yulia Strenov,[3,4] Meora Feinmesser,[3,4] Vladimir Kravtsov,[3,5] Marino E Leon,[6] Marián Hajdúch,[7] Syed Z Ali,[8] Christopher J VandenBussche,[8] Xinmin Zhang,[9,10] Leonor Leider-Trejo,[2,3] Asia Zubkov,[2] Sergey Vorobyov,[11] Michal Kushnir,[1] Yaron Goren,[1,12] Sarit Tabak,[1] Etti Kadosh,[1] Hila Benjamin,[13] Temima Schnitzer-Perlman,[1] Hagai Marmor,[1] Maria Motin,[1] Danit Lebanony,[1] Sharon Kredo-Russo,[1] Heather Mitchell,[13] Melissa Noller,[13] Alexis Smith,[13] Olivia Dattner,[13] Karin Ashkenazi,[13] Mats Sanden,[13] Kenneth A Berlin,[13] Dganit Bar,[1] Eti Meiri[1]

ABSTRACT

Aims The distinction between benign and malignant thyroid nodules has important therapeutic implications. Our objective was to develop an assay that could classify indeterminate thyroid nodules as benign or suspicious, using routinely prepared fine needle aspirate (FNA) cytology smears.

Methods A training set of 375 FNA smears was used to develop the microRNA-based assay, which was validated using a blinded, multicentre, retrospective cohort of 201 smears. Final diagnosis of the validation samples was determined based on corresponding surgical specimens, reviewed by the contributing institute pathologist and two independent pathologists. Validation samples were from adult patients (≥18 years) with nodule size >0.5 cm, and a final diagnosis confirmed by at least one of the two blinded, independent pathologists. The developed assay, RosettaGX Reveal, differentiates benign from malignant thyroid nodules, using quantitative RT-PCR.

Results Test performance on the 189 samples that passed quality control: negative predictive value: 91% (95% CI 84% to 96%); sensitivity: 85% (CI 74% to 93%); specificity: 72% (CI 63% to 79%). Performance for cases in which all three reviewing pathologists were in agreement regarding the final diagnosis (n=150): negative predictive value: 99% (CI 94% to 100%); sensitivity: 98% (CI 87% to 100%); specificity: 78% (CI 69% to 85%).

Conclusions A novel assay utilising microRNA expression in cytology smears was developed. The assay distinguishes benign from malignant thyroid nodules using a single FNA stained smear, and does not require fresh tissue or special collection and shipment conditions. This assay offers a valuable tool for the preoperative classification of thyroid samples with indeterminate cytology.

INTRODUCTION

Thyroid cancer has been increasing worldwide over the past few decades and is the most rapidly increasing cancer in the US.[1] More than 64 000 new cases are expected to be diagnosed in the US in 2016, with 1980 associated deaths.[2] Thyroid cancer usually presents as a palpable thyroid nodule identified on physical exam or incidentally when imaging studies are performed.

Fine needle aspiration (FNA) is currently the recommended method for sampling thyroid tissue in order to diagnose thyroid nodules. FNA cytology results in a definitive benign or malignant diagnosis in the majority of cases. However, depending on the institution, approximately 10–40% of FNAs are not conclusively diagnosed by cytology and are categorised as indeterminate.[3 4] In the Bethesda System for Reporting Thyroid Cytopathology, indeterminate categories include: 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). Most patients with cytologically indeterminate nodules are referred for a diagnostic lobectomy or complete thyroidectomy, however as many as 70% of these nodules prove to be benign on final surgical pathology.[3 4] To overcome this limitation of FNA cytology, several molecular tests have been developed, offering a refined diagnosis for cytologically indeterminate thyroid nodules and leading to a reduction in unnecessary surgeries.[5–9]

MicroRNAs (miRNAs) comprise a class of short (~21–23 nucleotides), non-coding endogenous RNAs that regulate gene expression by directing their target mRNAs for degradation or translational repression.[10–12] miRNA expression profiling has identified signatures associated with cancer diagnosis, prognosis and response to treatment.[10 13–16] In addition, miRNA expression profiles have been shown to differentiate histological types[17 18] and are currently used in several commercially available tests.[9 19 20] Numerous studies have described the role of miRNAs in the pathogenesis of thyroid cancer.[21–24]

miRNAs are extremely stable and remain intact in tissues, whether fresh, frozen or formalin-fixed paraffin-embedded (FFPE).[25] This property of miRNAs has been exploited for the development of several commercially available miRNA-based molecular tests.[19 20] It has also allowed us, as described here, to develop a miRNA-based diagnostic test, RosettaGX Reveal. Unlike other commercially available molecular tests, this test does not require fresh FNA tissue or special collection and shipment conditions, and can be performed on a single, routinely prepared FNA smear, stained with Papanicolaou stain or Romanowsky-type stains (Diff-Quik and Giemsa).

We describe here the discovery, development and blinded validation of the above-referenced miRNA-based diagnostic test. The test measures a set of miRNAs by qRT-PCR to classify a nodule as benign or 'suspicious for malignancy by miRNA profiling'. The test also measures a miRNA specific to medullary carcinoma. The negative predictive value (NPV) in indeterminate nodules where all three reviewing pathologists were in agreement regarding final diagnosis is 99%; it is 91% for the entire validation set.

METHODS

Patients and samples

The study was composed of three stages: (1) discovery, (2) training and (3) validation (figure 1). Under Institutional Review Board (IRB)-approved protocols, archived, preoperative stained FNA smear samples were gathered from several sources, as detailed in the online supplementary materials and methods. In the discovery and training sets, we sought to enrich for the various histological types and subtypes and therefore collected non-consecutive samples of Bethesda categories II–VI. Samples for the independent, retrospectively collected, validation set were received with a corresponding H&E-stained slide, obtained from the excised nodule, along with its associated histological diagnosis (reference diagnosis). The validation set consisted of indeterminate samples from five sources, which were blinded both to the lab technicians and to the investigators performing the analyses. To approximate the true distribution in the population, the samples in the validation set were consecutive (ie, each institute gave all the indeterminate smears that had a matching resection sample, gathered within a defined period of time). A detailed description of the training and validation samples is presented in table 1.

In a separate evaluation study, 41 Bethesda II and Bethesda VI samples and 48 FNA cell block samples were tested with the final assay classifier.

Cytopathological assessment

All cytological slides were categorised according to the Bethesda system[4] by the contributing institute. Since some samples date back to before the establishment of the Bethesda system, all samples were assigned a Bethesda category by the cytopathologist of the medical centre of origin ('the original pathologist'), based on the entire set of cytological slides. The cytological samples were stained with either Papanicolaou stain or Romanowsky-type stains (Diff-Quik and Giemsa).

Histological diagnosis and inclusion criteria

For all FNA samples, the reference diagnosis was based on the pathological assessment of the H&E stained excised tumour. Samples were included in the training and validation cohorts if the patient was at least 18 years old and if the nodule size was greater than 0.5 cm.
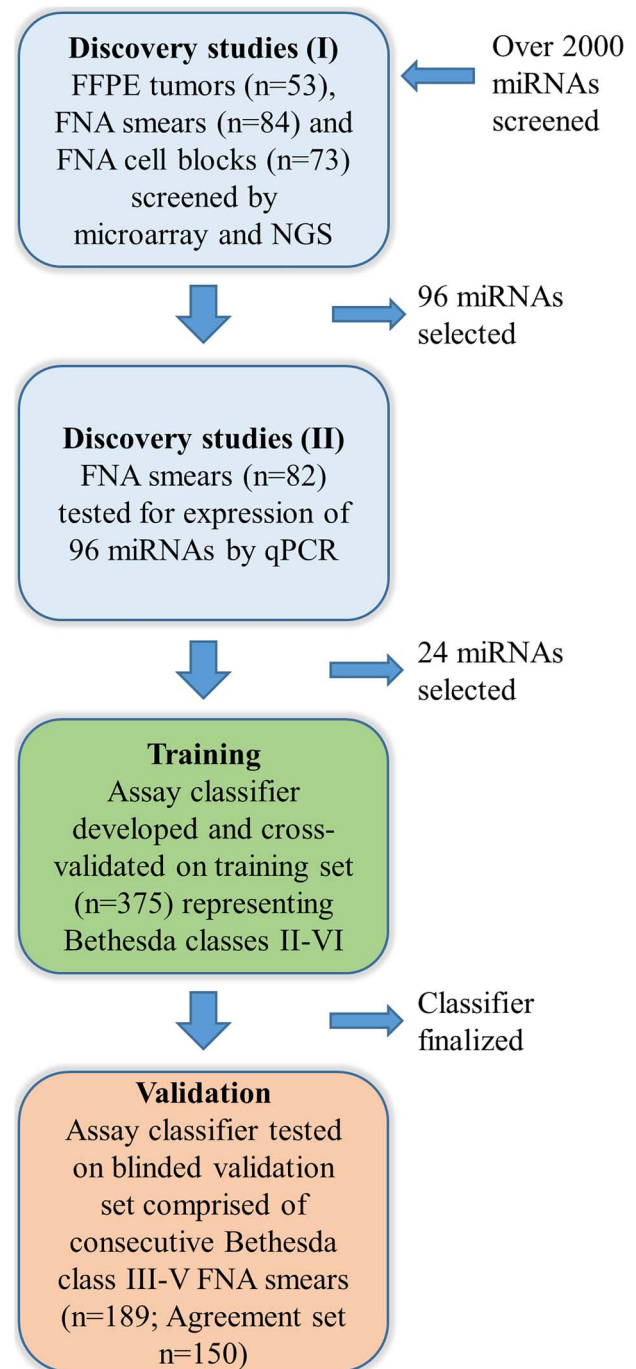

Figure 1 Assay development. The study was composed of three stages: (1) a discovery phase, (2) training, (3) validation. Discovery studies: (I) An initial set of 96 miRNAs was selected based on their differential expression in benign and malignant samples (53 formalin-fixed paraffin-embedded (FFPE), 73 cell block and 84 fine needle aspirate (FNA) samples) as seen on custom microarray and next generation sequencing (NGS) experiments. (II) This set of miRNAs was then evaluated on FNA smears (n=82) using qRT-PCR and 24 miRNAs were selected for the assay training and validation stages. Training: the final assay classifier was developed and cross validated on an FNA training set (n=375). Validation: the test was validated on a blinded set of 189 indeterminate samples from Bethesda classes III, IV and V for which at least one out of two independent pathologists agreed with the original pathologist regarding the final diagnosis (benign or malignant) of the sample. The results of the test on a subset of validation set samples for which all three pathologists agreed (n=150) were also assessed.

Table 1 Tumour samples used in the study

	Training*	Validation
Cohort		
#Samples	375	201
#Patients	357	201
% Malignant	49	30
Age (median)	54	53
% Females	73	80
Cytology		
#Giemsa	212	90
#Diff-Quik	95	21
#Papanicolaou	62	90
#Bethesda II	27	0
#Bethesda III	80	29
#Bethesda IV	142	131
#Bethesda V	77	41
#Bethesda VI	49	0

*Patient age was missing for 64 training samples and patient gender was missing for 10 training samples.
Three training samples were created by mixing more than one slide (with different stains), two were unstained, and for one the stain was unknown.

For the samples in the discovery and training sets, the original pathologist's review was the sole review and determined the final reference diagnosis. Samples in the validation set were also reviewed by two additional independent pathologists (ASh and LL-T). If at least one of the two independent pathologists agreed with the original pathologist's diagnosis regarding whether the resection sample was benign or malignant, then that sample was included (36 samples did not meet this criterion, and were therefore not included). All cases in which the reference diagnosis was medullary carcinoma were included, since only the original pathologist had information regarding calcitonin immunostaining. The histological type that was used for analyses was the one assigned by the original pathologist.

Data regarding the new diagnosis of 'non-invasive follicular thyroid neoplasm with papillary-like nuclear features' (NIFTP) were not collected, since this diagnosis was suggested after the study was concluded.[26]

The test was run after receiving the pathological reviews and defining the validation set, and thus, pathologists were blinded to the test results. Two smears were excluded from the validation set following unblinding, since it was discovered that there were two other smears from the same two samples. The duplicates had identical test results.

Classifier

The classifier combines several linear discriminant analysis (LDA) steps and a K-nearest neighbour (KNN) classifier step to differentiate between benign samples and samples that are 'suspicious for malignancy by miRNA profiling'. Samples classified by one of the LDA steps are marked as being positive for expression of the medullary marker. Several quality control (QC) steps accompany the test.[27] Further details regarding the assay protocol and classifier can be found in the online supplementary materials and methods.

RESULTS

We developed an assay which classifies indeterminate thyroid smears as benign or 'suspicious for malignancy by miRNA profiling'. In addition, the assay tests for the presence of a medullary carcinoma marker (hsa-miR-375). There were three phases in the development of the assay (figure 1): a discovery phase in which the set of miRNA biomarkers was selected; a training phase in which the final classifier was determined; and a validation study, in which the diagnostic protocol was tested in the CLIA-approved US laboratory, on a blinded independent validation cohort (table 1). The validation study was preceded by an inter-laboratory validation study and other analytical validation studies.[27] In addition, there was an evaluation study on Bethesda II/VI samples and cell blocks.

Discovery studies

To select the set of miRNAs for classification, several screening stages were performed (figure 1). In the first stage, 53 FFPE samples of resected tumours, 73 cell blocks of FNAs and a set of 156 stained FNA smears, corresponding to 84 unique samples, were profiled on Agilent custom-designed miRNA microarrays containing over 2000 miRNA probes. In addition, a subset of the follicular FFPE samples were profiled using next generation sequencing (data not shown). Next, a subset of 96 miRNAs that showed differential expression in benign and malignant tumours was selected. The selected miRNA set also included biomarkers described in the literature, biomarkers of epithelial cells and markers of various blood components discovered based on the profiling of smears that contained only blood.[27] These miRNAs were measured using qRT-PCR analysing 95 stained FNA smears, corresponding to 82 unique samples (71 of which were previously profiled on microarrays). Based on these experiments, a final set of 24 miRNAs was selected (table 2).

Training set and classifier

To establish the final sample reference set and classifier, the 24 miRNAs were quantified in 375 samples, according to the final assay protocol, in two laboratories (252 FNA smear samples profiled in the Rosetta Genomics Israel laboratory and 123 samples profiled in the Rosetta Genomics US laboratory in Philadelphia, Pennsylvania, USA). The type of cytological stain used did not affect the classification performance.[27]

The classification method used for this miRNA-based assay, named RosettaGX Reveal, combines several LDA steps along with a KNN-based classifier. The performance of the training set is summarised in table 3. Based on the results from this training set (as estimated using cross validation), the sensitivity of the classifier on indeterminate samples (Bethesda categories III, IV and V) was estimated to be 86%, and the specificity was estimated to be 75%.

Validation set

An independent set of 201 consecutive, indeterminate FNA samples (table 1) from five sources was classified blindly, in the US CLIA-approved laboratory, by the assay. This set of 201 samples included only samples for which at least one of the two independent pathologists agreed with the original pathologist on the final diagnosis (benign or malignant) of the excised H&E stained nodule.

Only 12 of the 201 samples (6%) failed during processing or QC steps, with the most common reason being low miRNA expression. All of these 12 samples were histologically benign based on the resections. Of the remaining 189 samples, 101 (53.4%) were classified as benign. The performance of the validation set was found to be very similar to the performance estimates of the training set, as can be seen in tables 3 and 4 (NPV: 91%, sensitivity: 85%, specificity: 72%; and positive predictive

Table 2 MicroRNAs profiled in the assay

MicroRNA*	Sequence†	Forward primer sequence‡
hsa-miR-31-5p	AGGCAAGATGCTGGCATAGCT	AGGCAAGATGCTGGCATAGCT
hsa-miR-5701	TTATTGTCACGTTCTGATT	AGTCATTTGGCTTATTGTCACGTTCTGATT
hsa-miR-424-3p	CAAAACGTGAGGCGCTGCTAT	CAAAACGTGAGGCGCTGCTAT
MID-50971	ATACTCTGGTTTCTTTTC	CAGTCATTTGGCATACTCTGGTTTCTTTTC
MID-20094	TAAGCCAGTTTCTGTCTGATA	CATTTGGCTAAGCCAGTTTCTGTCTGATA
MID-50976	CTGTCTGAGCGCCGCTC	CCTGTCTGAGCGCCGCTC
hsa-miR-3074-5p	GTTCCTGCTGAACTGAGCCAG	CGTTCCTGCTGAACTGAGCCAG
hsa-miR-222-3p	AGCTACATCTGGCTACTGGGT	GCAGCTACATCTGGCTACTGGGT
MID-50969	ATGACAGATTGACATGGACAATT	TGGCATGACAGATTGACATGGACAATT
hsa-miR-146b-5p	TGAGAACTGAATTCCATAGGCT	TGGCTGAGAACTGAATTCCATAGGCT
hsa-miR-346	TGTCTGCCCGCATGCCTGCCTCT	TGTCTGCCCGCATGCCTGCCTCT
MID-16582	AGTGAAGCATTGGACTGTA	TTGGCAGTGAAGCATTGGACTGTA
hsa-miR-342-3p	TCTCACACAGAAATCGCACCCGT	CAGTCATTTGGGTCTCACACAGAAATCG
hsa-miR-181c-5p	AACATTCAACCTGTCGGTGAGT	CAGTCATTTGGCAACATTCAACCTGTCG
hsa-miR-125b-5p	TCCCTGAGACCCTAACTTGTGA	CAGTCATTTGGGTCCCTGAGACCCTAAC
hsa-miR-375	TTTGTTCGTTCGGCTCGCGTGA	CAGTCATTTGGGTTTGTTCGTTCGGCTC
hsa-miR-486-5p	TCCTGTACTGAGCTGCCCCGAG	CAGTCATTTGGCTCCTGTACTGAGCTGC
hsa-miR-551b-3p	GCGACCCATACTTGGTTTCAG	CAGTCATTTGGCGCGACCCATACTTGGT
hsa-miR-23a-3p	ATCACATTGCCAGGGATTTCC	CAGTCATTTGGCATCACATTGCCAGGGA
hsa-miR-574-3p	CACGCTCATGCACACACCCACA	CAGTCATTTGGCCACGCTCATGCACACA
hsa-miR-152-3p	TCAGTGCATGACAGAACTTGG	CAGTCATTTGGCTCAGTGCATGACAGAA
hsa-miR-200c-3p	TAATACTGCCGGGTAATGATGGA	CAGTCATTTGGGTAATACTGCCGGGTAA
hsa-miR-138-5p	AGCTGGTGTTGTGAATCAGGCCG	CAGTCATTTGGCAGCTGGTGTTGTGAAT
hsa-miR-345-5p	GCTGACTCCTAGTCCAGGGCTC	CAGTCATTTGGCGCTGACTCCTAGTCCA

*microRNA names that begin with 'hsa' are in miRBase, those that begin with 'MID' were sequenced and/or predicted at Rosetta Genomics.
†microRNA sequence is from miRBase[28] V.20.
‡Reverse primer sequence: GCGAGCACAGAATTAATACGAC.

Table 3 Performance of the assay

		Indeterminate (Bethesda III, IV,V)*	Indeterminate (Bethesda III, IV)*	Bethesda II and Bethesda VI*
Training†	#Malignant	115	59	40
	#Benign	147	137	26
	Sensitivity	86 [78–92]	78 [65–88]	96 [85–100]
	Specificity	75 [67–81]	76 [68–83]	82 [62–94]
Validation‡, entire set	#Malignant	61	31	0
	#Benign	128	119	0
	Sensitivity	85 [74–93]	74 [55–88]	NA
	Specificity	72 [63–79]	74 [65–82]	NA
	NPV	91 [84–96]	92 [84–96]	NA
	PPV	59 [48–69]	43 [29–57]	NA
Validation‡, agreement set	#Malignant	40	14	0
	#Benign	110	102	0
	Sensitivity	98 [87–100]	100 [77–100]	NA
	Specificity	78 [69–85]	80 [71–88]	NA
	NPV	99 [94–100]	100 [96–100]	NA
	PPV	62 [49–74]	41 [25–59]	NA
Evaluation study on Bethesda II/VI samples§	#Malignant	NA	NA	9
	#Benign	NA	NA	32
	Sensitivity	NA	NA	89 [52–100]
	Specificity	NA	NA	63 [44–79]

*95% CIs are in square brackets.
†For training, estimates are based on the mean of 10 10-fold cross-validation runs. Samples with very low expression in any of the classification steps, as well as medullary samples, are not included.
‡Samples that failed quality control are not included.
§Additional blinded study.
NA, not applicable.

value (PPV): 59%). When focusing on nodules of size ≥1 cm (n=166), the sensitivity was 84% and the specificity was 72%. The sensitivity and specificity of the subset of Bethesda III and IV samples are both 74%, with an NPV of 92% and a PPV of 43% (table 3). The accuracy of oncocytic follicular adenoma (FA) samples was slightly lower than that of non-oncocytic FA samples, however this difference was not statistically significant (see online supplementary results).

The nine malignant samples misclassified as benign (table 5) included samples from all three indeterminate Bethesda categories; included both Giemsa and Papanicolaou stained samples; and came from three different sources. The follicular carcinoma (FC) sample misclassified as benign by the assay was described as having minimal capsular invasion, according to the original pathologist, as were the other two FC samples that were correctly classified by the assay. The samples from patients with chronic lymphocytic thyroiditis (CLT) showed a lower correct classification rate (ie, relatively more were misclassified as 'suspicious for malignancy by miRNA profiling'), relative to the training performance and to the other benign samples (table 4). However, this difference may be due to the small number of CLT samples in the validation set.

Validation agreement set

To test the assay on a set of samples with a higher degree of certainty in the final diagnosis, a subset of the validation samples ('agreement set') was compiled post hoc. This set was composed of 160 samples (80% of the validation set) for which all three

Table 4 Performance of the assay for different histological types

Histological type	Training*		Validation, entire set		Validation, agreement set	
	#Samples†	% Correct‡	#Samples†	% Correct‡	#Samples†§	% Correct‡
Medullary	5	100 [48–100]	3	100 [29–100]	1 (33.3%)	100 [3–100]
PTC classic	48	94 [83–100]	17	88 [64–99]	15 (88.2%)	100 [78–100]
FVPTC	40	81 [65–92]	37	84 [68–94]	23 (62.2%)	96 [78–100]
FC	16	56 [30–80]	3	67 [9–99]	1 (33.3%)	100 [3–100]
PDC	5	100 [48–100]	1	100 [3–100]	0 (0%)	NA
Papillary, other	6	100 [54–100]	0	NA	0	NA
FA	90	76 [66–84]	95	76 [66–84]	82 (86.3%)	82 [72–89]
Nodular hyperplasia	48	75 [60–86]	28	64 [44–81]	23 (82.1%)	74 [52–90]
CLT	9	82 [44–99]	5	40 [5–85]	5 (100.0%)	40 [5–85]
Total	267		189		150	

*Only indeterminate training samples are listed in the table. Estimates for the training performance are based on the mean of 10 10-fold cross-validation runs. Samples with very low expression in any of the classification steps, as well as medullary samples, are not included.
†Number of samples includes only those that passed quality control steps.
‡95% CIs are in square brackets.
§Numbers in parentheses signify the percentage of test samples in the agreement set.
CLT, chronic lymphocytic thyroiditis; FA, follicular adenoma; FC, follicular carcinoma; FVPTC, follicular variant of papillary carcinoma; NA, not applicable; PDC, poorly differentiated carcinoma.

pathologists were in agreement on the final diagnosis of benign or malignant; 150 of these samples passed QC steps. This set demonstrated very high performance (table 3). The NPV of the agreement set was 99% (only one malignant sample was misclassified as benign), with a sensitivity of 98%, a specificity of 78% and a PPV of 62%. The NPV and PPV for both the entire set and the agreement set are plotted in figure 2.

As expected, the samples in the agreement set (table 4) had a much higher correct classification rate compared with the remainder of the validation set samples (ie, where only one of the independent pathologists agreed with the diagnosis made by the original pathologist): 125/150 (83%) samples in the agreement set were correctly classified, whereas 19/39 (49%) of the remaining samples were correctly classified (p=6.14e-06, χ^2 test).

Concordance between pathologists

The assay performance is influenced by the accuracy of the diagnosis. Therefore, we examined the level of agreement between the pathologists for the different histological types (table 6). There was a large number of encapsulated follicular variant of papillary carcinomas (FVPTCs) in the entire validation set that were not in the agreement set. This higher proportion of encapsulated FVPTCs in the subset of samples for which only one of the two independent pathologists agreed with the original

pathologist, was statistically significant compared with the proportion of non-encapsulated FVPTCs (p=0.0029, Fisher's exact test).

Medullary carcinoma

Medullary carcinoma is a rare form of thyroid cancer which often demonstrates overexpression of hsa-miR-375.[29] To identify medullary carcinoma, the assay tests for the upregulation of hsa-miR-375 in one of the LDA steps (figure 3). Elevated expression of this medullary marker is provided as part of the assay results. In the training set, there were 14 medullary samples, including five indeterminate medullary samples, and all of these presented high expression of hsa-miR-375. In the validation set, there were three medullary carcinoma samples. All were correctly classified as suspicious. However, one (assigned Bethesda V) did not demonstrate overexpression of hsa-miR-375 and was therefore not denoted as medullary carcinoma (this sample was confirmed to be medullary carcinoma, with positive calcitonin immunostaining).

Evaluation study on FNA cell blocks and Bethesda II/VI samples

The assay was also tested on cell blocks, and in benign (Bethesda II) and malignant (Bethesda VI) smears. The sensitivity and specificity of the cell block indeterminate samples were 72% and

Table 5 The malignant validation set samples misclassified as benign

| Bethesda	Stain	Extracted amount (ng)	Gender	Histological type	Histological subtype	In agreement set?
V	Giemsa	294	Female	Papillary carcinoma	Follicular variant, non-encapsulated	Yes
IV	Giemsa	4716	Female	Papillary carcinoma	Classic variant	No
IV	PAP	138	Male	Papillary carcinoma	Follicular variant, encapsulated	No
III	PAP	115	Female	Papillary carcinoma	Follicular variant, encapsulated	No
IV	PAP	103	Female	Papillary carcinoma	Follicular variant, encapsulated	No
IV	Giemsa	51	Female	Papillary carcinoma	Follicular variant, encapsulated	No
IV	PAP	1242	Female	Papillary carcinoma	Follicular variant, encapsulated	No
IV	Giemsa	249	Female	Follicular carcinoma	Minimal capsular invasion	No
IV	Giemsa	451	Male	Papillary carcinoma	Classic variant	No

PAP, Papanicolaou.
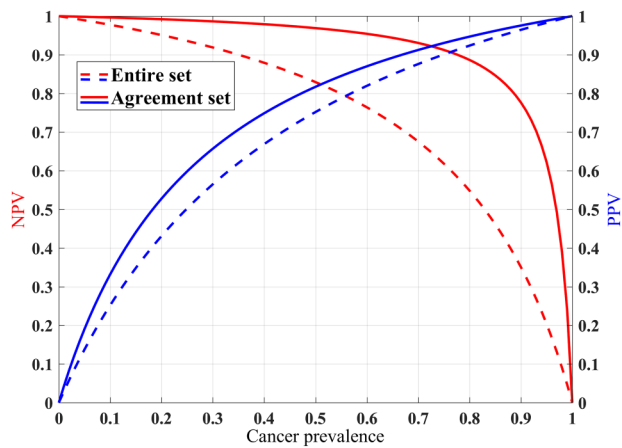
Figure 2 Negative predictive value (NPV) and positive predictive value (PPV) for varying prevalence values. NPV and PPV were calculated, based on the observed sensitivity and specificity in the blinded validation set, for varying prevalence values. Dashed lines: the entire validation set (sensitivity: 85.2%, specificity: 71.9%), solid lines: the agreement subset (sensitivity: 97.5%, specificity: 78.2%). Red line: calculated NPV. Blue line: calculated PPV.

Figure 3 Medullary carcinoma linear discriminant analysis (LDA) step. An LDA classifier based on the expression of hsa-miR-375 is used to differentiate medullary carcinoma samples. All the training medullary carcinoma stained smears and two of the three medullary smears in the test set demonstrate overexpression of hsa-miR-375. Yellow diamonds: malignant non-medullary training samples. Blue squares: benign training samples. Green circles: medullary carcinoma training samples. Red stars: medullary carcinoma validation samples.

79%. The sensitivity of the malignant Bethesda VI smears was 89% and the specificity of the benign Bethesda II samples was 63% (table 3). More details can be seen in the online supplementary results.

DISCUSSION

We present here a first-of-its-kind assay by which miRNA material is successfully extracted from routinely stained FNA cytology smears and classified as 'suspicious for malignancy by miRNA profiling' or 'benign'. In contrast to currently available tests,[6 8 9 30] the test presented here does not require an additional FNA biopsy and can be performed on the same specimen as that initially used to categorise the sample as indeterminate. In addition, this test does not require specially designated preservation material, or unique shipment conditions. Instead, a single routinely prepared cytological slide, stained with Papanicolaou stain or Romanowsky type stains (Diff-Quik and Giemsa), can be used. The test does not require a large amount
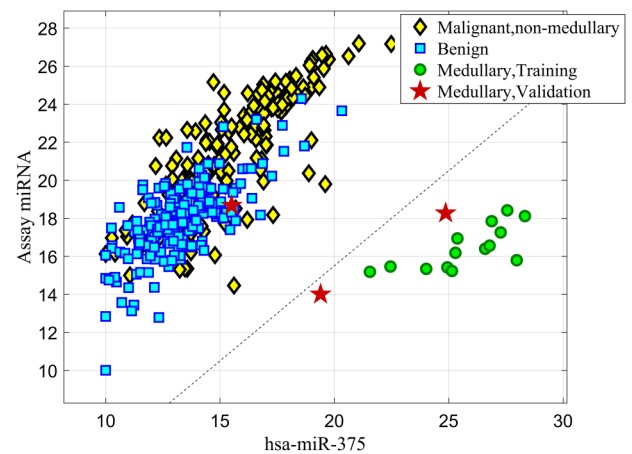
of cytological material, and the failure rate is quite low if there is minimal adequate cellularity, with 94% of the samples in the validation set being successfully processed.

The assay's performance was evaluated based on a validation set composed of blinded, indeterminate, consecutive samples gathered from five sources in the USA, Europe and Israel. Since the test is run on cytology slides routinely prepared for examination, and does not require any special preservation conditions, it was possible to perform the study on a retrospective cohort.

The development of a molecular test requires a reliable gold-standard reference diagnosis with which to compare the test results. This leads to two inherent biases in the tested set of samples. The first bias is that only samples with a corresponding surgically excised histopathological specimen were gathered. The second bias is that only samples for which the reference resection-based diagnosis was confirmed by an independent pathologist were included. Since there is a high level of disagreement between pathologists regarding the diagnosis of such specimens,[31–34] relying on a single pathologist may lead to the inclusion of samples with an inaccurate final diagnosis, which could lead to an incorrect estimation of the performance of the assay. However, we cannot rule out the possibility that the exclusion of these samples alters the true sample distribution and, as a result, affects the performance estimates.

The majority of malignant samples that were not included in the agreement subset were encapsulated FVPTCs. This is in accordance with previous reports that there is a relatively high level of inter-observer variability between pathologists with regard to FVPTC diagnoses,[31 33] in particular for non-invasive, encapsulated FVPTCs versus FA.[35] There is current evidence that encapsulated FVPTC is a neoplasm of relatively low malignant potential, particularly if there is no capsular or vascular invasion.[36] Additional evidence supporting a reclassification is suggested by the findings in their molecular profile.[37] It has been suggested that cases of encapsulated FVPTC that cannot be unequivocally diagnosed as benign or malignant should be reclassified as 'follicular tumour of uncertain malignant potential' by some authors[35 38] or, as proposed by the Endocrine Pathology Society, as NIFTP.[26] We are actively collecting data on

Table 6 The malignant histological types in the validation set

	Agreement set		Not in agreement set	
	Total	#Misclassified*	Total	#Misclassified*
Medullary†	1	0	2	0
Papillary classic	15	0	2	2
FVPTC, encapsulated	12	0	14	5
FVPTC, non-encapsulated	10	1	0	0
FC	1	0	2	1
PDC	0	0	1	0
Total‡	39	1	21	8

*Misclassified as benign.
†The two independent pathologists did not have information regarding immunostaining.
‡One FVPTC sample (in the agreement set and correctly classified) is not included in the table, since there was no information available regarding the encapsulation status.
FC, follicular carcinoma; FVPTC, follicular variant of papillary carcinoma; PDC, poorly differentiated carcinoma.

this new diagnosis for future studies of the classifier. It has also been suggested that papillary thyroid cancer should be reclassified according to its molecular profile.[37] Our study offers additional evidence supporting the need for reclassification of encapsulated FVPTC.

The expression levels of several documented thyroid malignant markers are measured in our assay. For example, the miRNAs used in the assay include hsa-miR-146b-5p and hsa-miR-222-3p, which have both been found to be upregulated in papillary thyroid cancer and involved in tumour progression and aggressiveness.[39–41] In contrast, hsa-miR-152-3p and hsa-miR-138-5p have been shown to be downregulated in papillary thyroid cancer.[42] miRNAs, including several of those used in the assay, have been previously found to differentiate malignant and benign thyroid FNA samples,[9] [43–46] even in FNA smears.[47–49]

In conclusion, we presented a new diagnostic assay and evaluated its performance on a blinded set of 189 samples from several sources. Additional cohorts, both academic and non-academic, could help to further validate the performance of the assay. The test described in this paper is a novel, multicentre, clinically evaluated, commercially available assay that can accurately differentiate between malignant and benign thyroid nodules using routinely prepared FNA-stained smears.

Take home messages

- 10–40% of thyroid fine needle aspirates (FNAs) are not conclusively diagnosed by cytology and are categorised as indeterminate.
- The RosettaGX Reveal assay, which was blindly validated, differentiates benign from malignant thyroid nodules in indeterminate smears.
- The smear used for the assay can be a routinely prepared smear, which was used to make the indeterminate diagnosis, and does not require a repeat FNA.
- In contrast with currently available tests, the assay does not require fresh tissue or special collection and shipment conditions.

Author affiliations
[1]Rosetta Genomics Ltd, Rehovot, Israel
[2]Pathology Institute, Tel-Aviv Sourasky Medical Center, Tel Aviv, Israel
[3]The Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel
[4]Pathology Institute, Rabin Medical Center, Petach Tikva, Israel
[5]Pathology Institute, Meir Medical Center, Kfar Saba, Israel
[6]H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA
[7]Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic
[8]The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[9]Temple University Hospital, Philadelphia, Pennsylvania, USA
[10]Cooper University Hospital, Cooper Medical School of Rowan University at Camden, New Jersey, USA
[11]National Centre of Clinical and Morphological Diagnostics, St Petersburg, Russia
[12]Geha Mental Health Center, Petach Tikva, Israel
[13]Rosetta Genomics Inc, Philadelphia, Pennsylvania, USA

REFERENCES

1. Zevallos JP, Hartman CM, Kramer JR, et al. Increased thyroid cancer incidence corresponds to increased use of thyroid ultrasound and fine-needle aspiration: a study of the Veterans Affairs health care system. *Cancer* 2015;121:741–6.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7–30.
3. Bongiovanni M, Spitale A, Faquin WC, et al. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol* 2012;56:333–9.
4. Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol* 2009;132:658–65.
5. Duick DS, Klopper JP, Diggans JC, et al. The impact of benign gene expression classifier test results on the endocrinologist-patient decision to operate on patients with thyroid nodules with indeterminate fine-needle aspiration cytopathology. *Thyroid* 2012;22:996–1001.
6. 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;25:1217–23.
7. 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;24:1479–87.
8. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med* 2012;367:705–15.
9. 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;100:2743–50.
10. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6:857–66.
11. Farazi TA, Spitzer JI, Morozov P, et al. miRNAs in human cancer. *J Pathol* 2011;223:102–15.
12. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature* 2005;435:834–8.
13. Li MH, Fu SB, Xiao HS. Genome-wide analysis of microRNA and mRNA expression signatures in cancer. *Acta Pharmacol Sin* 2015;36:1200–11.
14. Rosenfeld N, Aharonov R, Meiri E, et al. MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 2008;26:462–9.
15. Weissmann-Brenner A, Kushnir M, Lithwick Yanai G, et al. Tumor microRNA-29a expression and the risk of recurrence in stage II colon cancer. *Int J Oncol* 2012;40:2097–103.
16. Shepshelovich D, Ram R, Uziel O, et al. MicroRNA signature is indicative of long term prognosis in diffuse large B-cell lymphoma. *Leuk Res* 2015;39:632–7.
17. Landgraf P, Rusu M, Sheridan R, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 2007;129:1401–14.

18 Bentwich I. A postulated role for microRNA in cellular differentiation. *Faseb J* 2005;19:875–9.

19 Gilad S, Lithwick-Yanai G, Barshack I, *et al*. Classification of the four main types of lung cancer using a microRNA-based diagnostic assay. *J Mol Diagn* 2012;14:510–17.

20 Meiri E, Mueller WC, Rosenwald S, *et al*. A second-generation microRNA-based assay for diagnosing tumor tissue origin. *Oncologist* 2012;17:801–12.

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

22 He H, Jazdzewski K, Li W, *et al*. The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci USA* 2005;102:19075–80.

23 Weber F, Teresi RE, Broelsch CE, *et al*. A limited set of human MicroRNA is deregulated in follicular thyroid carcinoma. *J Clin Endocrinol Metab* 2006;91:3584–91.

24 Nikiforova MN, Tseng GC, Steward D, *et al*. MicroRNA expression profiling of thyroid tumors: biological significance and diagnostic utility. *J Clin Endocrinol Metab* 2008;93:1600–8.

25 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;13:1668–74.

26 Nikiforov YE, Seethala RR, Tallini G, *et al*. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol* 2016;2:1023–9.

27 Benjamin H, Schnitzer-Perlman T, Shtabsky A, *et al*. Analytical validity of a microRNA-based assay for diagnosing indeterminate thyroid FNA smears from routinely prepared cytology slides. *Cancer Cytopathol* 2016. http://dx.doi.org/10.1002/cncy.21731

28 Griffiths-Jones S, Grocock RJ, van Dongen S, *et al*. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006;34(Database issue):D140–4.

29 Hudson J, Duncavage E, Tamburrino A, *et al*. Overexpression of miR-10a and miR-375 and downregulation of YAP1 in medullary thyroid carcinoma. *Exp Mol Pathol* 2013;95:62–7.

30 Nikiforov YE, Carty SE, Chiosea SI, *et al*. Highly accurate diagnosis of cancer in thyroid nodules with follicular neoplasm/suspicious for a follicular neoplasm cytology by ThyroSeq v2 next-generation sequencing assay. *Cancer* 2014;120:3627–34.

31 Elsheikh TM, Asa SL, Chan JK, *et al*. Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. *Am J Clin Pathol* 2008;130:736–44.

32 Wang CC, Friedman L, Kennedy GC, *et al*. A large multicenter correlation study of thyroid nodule cytopathology and histopathology. *Thyroid* 2011;21:243–51.

33 Lloyd RV, Erickson LA, Casey MB, *et al*. Observer variation in the diagnosis of follicular variant of papillary thyroid carcinoma. *Am J Surg Pathol* 2004;28:1336–40.

34 Hirokawa M, Carney JA, Goellner JR, *et al*. Observer variation of encapsulated follicular lesions of the thyroid gland. *Am J Surg Pathol* 2002;26:1508–14.

35 Ganly I, Wang L, Tuttle RM, *et al*. Invasion rather than nuclear features correlates with outcome in encapsulated follicular tumors: further evidence for the reclassification of the encapsulated papillary thyroid carcinoma follicular variant. *Hum Pathol* 2015;46:657–64.

36 Haugen BR, Alexander EK, Bible KC, *et al*. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* 2016;26:1–133.

37 Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 2014;159:676–90.

38 Williams ED. Guest editorial: two proposals regarding the terminology of thyroid tumors. *Int J Surg Pathol* 2000;8:181–3.

39 Benvenga S, Koch CA. Molecular pathways associated with aggressiveness of papillary thyroid cancer. *Curr Genomics* 2014;15:162–70.

40 Yang Z, Yuan Z, Fan Y, *et al*. Integrated analyses of microRNA and mRNA expression profiles in aggressive papillary thyroid carcinoma. *Mol Med Rep* 2013;8:1353–8.

41 Chen YT, Kitabayashi N, Zhou XK, *et al*. MicroRNA analysis as a potential diagnostic tool for papillary thyroid carcinoma. *Mod Pathol* 2008;21:1139–46.

42 Cong D, He M, Chen S, *et al*. Expression profiles of pivotal microRNAs and targets in thyroid papillary carcinoma: an analysis of The Cancer Genome Atlas. *Onco Targets Ther* 2015;8:2271–7.

43 Dettmer M, Vogetseder A, Durso MB, *et al*. MicroRNA expression array identifies novel diagnostic markers for conventional and oncocytic follicular thyroid carcinomas. *J Clin Endocrinol Metab* 2013;98:E1–7.

44 Agretti P, Ferrarini E, Rago T, *et al*. MicroRNA expression profile helps to distinguish benign nodules from papillary thyroid carcinomas starting from cells of fine-needle aspiration. *Eur J Endocrinol* 2012;167:393–400.

45 Keutgen XM, Filicori F, Crowley MJ, *et al*. A panel of four miRNAs accurately differentiates malignant from benign indeterminate thyroid lesions on fine needle aspiration. *Clin Cancer Res* 2012;18:2032–8.

46 Kitano M, Rahbari R, Patterson EE, *et al*. Evaluation of candidate diagnostic microRNAs in thyroid fine-needle aspiration biopsy samples. *Thyroid* 2012;22:285–91.

47 Shen R, Liyanarachchi S, Li W, *et al*. MicroRNA signature in thyroid fine needle aspiration cytology applied to 'atypia of undetermined significance' cases. *Thyroid* 2012;22:9–16.

48 Fassina A, Cappellesso R, Simonato F, *et al*. A 4-MicroRNA signature can discriminate primary lymphomas from anaplastic carcinomas in thyroid cytology smears. *Cancer Cytopathol* 2014;122:274–81.

49 Panebianco F, Mazzanti C, Tomei S, *et al*. The combination of four molecular markers improves thyroid cancer cytologic diagnosis and patient management. *BMC Cancer* 2015;15:918.