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


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 differentiates 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. More than 64 000 new cases are expected to be diagnosed in the US in 2016, with 1980 associated deaths. 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.

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. 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.

MicroRNAs (miRNAs) comprise a class of short (~21–23 nucleotides), non-coding endogenous RNAs that regulate gene expression by directing their target miRNAs for degradation or translational repression. miRNA expression profiling has identified signatures associated with cancer diagnosis, prognosis and response to treatment. In addition, miRNA expression profiles have been shown to differentiate histological types and are currently used in several commercially available tests. Numerous studies have described the role of miRNAs in the pathogenesis of thyroid cancer.
miRNAs are extremely stable and remain intact in tissues, whether fresh, frozen or formalin-fixed paraffin-embedded (FFPE). This property of miRNAs has been exploited for the development of several commercially available miRNA-based molecular tests. 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 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.
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 analysis of 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 value 98%).

### Table 1 Tumour samples used in the study

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

*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.

---


Original article
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 ‘suspect’ 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
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, $\chi^2$ 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 VI) 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 4** Performance of the assay for different histological types

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

*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.
‡% Correct are in square brackets. 
§Numbers in parentheses signify the percentage of test samples in the agreement set.

**Table 5** The malignant validation set samples misclassified as benign

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

PAP, Papanicolaou.
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 ‘suspicous for malignancy by miRNA profiling’ or ‘benign’. In contrast to currently available tests, 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, 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, in particular for non-invasive, encapsulated FVPTCs versus FA. There is current evidence that encapsulated FVPTC is a neoplasm of relatively low malignant potential, particularly if there is no capsular or vascular invasion. Additional evidence supporting a reclassification of these samples alters the true sample distribution and, as a result, affects the performance estimates.

The assay was developed based on a blinded 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, 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, in particular for non-invasive, encapsulated FVPTCs versus FA. There is current evidence that encapsulated FVPTC is a neoplasm of relatively low malignant potential, particularly if there is no capsular or vascular invasion. Additional evidence supporting a reclassification 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, in particular for non-invasive, encapsulated FVPTCs versus FA. There is current evidence that encapsulated FVPTC is a neoplasm of relatively low malignant potential, particularly if there is no capsular or vascular invasion. Additional evidence supporting a reclassification of these samples alters the true sample distribution and, as a result, affects the performance estimates.

**Table 6** The malignant histological types in the validation set

<table>
<thead>
<tr>
<th>Agreement set</th>
<th>Not in agreement set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Medullary†</td>
<td>1</td>
</tr>
<tr>
<td>Papillary classic</td>
<td>15</td>
</tr>
<tr>
<td>FVPTC, encapsulated</td>
<td>12</td>
</tr>
<tr>
<td>FVPTC, non-encapsulated</td>
<td>10</td>
</tr>
<tr>
<td>FC</td>
<td>1</td>
</tr>
<tr>
<td>PDC</td>
<td>0</td>
</tr>
<tr>
<td>Total†</td>
<td>39</td>
</tr>
</tbody>
</table>

*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.

**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.
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. 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. In contrast, hsa-miR-152-3p and hsa-miR-138-5p have been shown to be downregulated in papillary thyroid cancer. miRNAs, including several of those used in the assay, have been previously found to differentiate malignant and benign thyroid FNA samples, even in FNA smears.

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, academic, could help to further validate the performance of the assay described in this paper is a novel, multicentre, academic.

Contributors GL-Y analysed the data, developed the algorithm and the software, and wrote the manuscript. ND, TS-P, HB analysed the data, developed the algorithm and the software, and reviewed and edited the manuscript. ASH collected data, reviewed the slides, was involved in interpretation of the data, and reviewed and edited the manuscript. SM, YS, MF, VK, MEL, MH, SZA, CIVB, XZ, AZ, SV identified clinical cases, reviewed the slides and collected data. LL-T and MS reviewed and analysed slides. MK contributed to experimental design, collected the data, was involved in interpretation of the data, and reviewed and edited the manuscript. YG contributed to experimental design and collected the data. ST, EK, HM, MM, DL, SK-R, HM, MN, ASM, OD, KA performed experiments and were involved in calibrations and protocol development. KAB was involved in conceptualisation and reviewed and edited the manuscript. DB was involved in conceptualisation, was involved in interpretation of the data, and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests Authors affiliated with Rosetta Genomics are full-time employees of the company and/or hold equity in the company, which stands to gain from the publication of this manuscript. One of the authors (A. Shatabsky) is a paid consultant for Rosetta Genomics. The authors from medical/clinical centres have received funding from the company as part of this and/or other collaborative projects.

Ethics approval IRB (for each institution).

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

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.

REFERENCES


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.


תקציר:

מטרות: לאבחנה בין קשריות שפירות לממאירות בלוטת התריס יש השלכות רפואיות חשובות. מטרתנו היתה לפתח בדיקה אשר תסווג קשריות, שלא ניתן לסווגן בבדיקה ציטולוגית, כಶפירות או כממאירות. זאת, באמצעות משטחים ציטולוגים בדגימו ותביופסיה מחט דקה.

שיטות: הבדיקה, המבוססת על קביעת ביטויים של מקרו-רנ"א, פותחה עבור 375 דגימות ועなかった戾 בדיקת ויליארמנס של 201 דגימה. נסח הלויזיאנה היה רטוספектив, ובמרכזי ההבתות באומות מ Sinai.romise והדברים נשלחו הביתה פיתולוגית על הגידול והשחיק. לכל ויליארמנס בדיקת "U" פיתלוג המרכז וbadge נשלחו בדיקה資源 של RosettaGX Reveal, המבנה בה וקריאת שם ב-RT-PCR.

תוצאות: ביצועי הבדיקה עבור 211 הדגימות שpassed בקרת האיכות היו ערך ניבוי שלילי (negative predictive value) 12% (CI: 84%-96%), רגישות 13% (CI: 74%-93%), ספציפיות 71% (CI: 63-79%). ביצועי הבדיקה עבור 230 דגימות עם הסכמת של שלושת הפתולוגים היו ערך ניבוי שלילי (negative predictive value) 11% (CI: 94%-100%), רגישות 11% (CI: 94%-100%), ספציפיות 71% (CI: 69-85%).

מסקנה: בדיקה חדשה המבוססת על ביטוי מקרו-רנ"א במשטחים ציטולוגים של בלוטת התריס פותחה. הבדיקה מחלולה בין קשריות שפירות לממאירות, ולא médecin_mid 처리ים שונים ונו軍ית. המסקנה היא שהבדיקה מספקת כל חשב ליאנות צרפיה, ולא רק בקירה של בלוטת התריס.