Impact of the Afirma Gene Expression Classifier Result on the Surgical Management of Thyroid Nodules with Category III/IV Cytology and Its Correlation with Surgical Outcome

Shweta Chaudhary a Yanjun Hou a Rulong Shen a Shveta Hooda b Zaibo Li a

a Department of Pathology, Wexner Medical Center, Ohio State University, Columbus, Ohio, and b Department of Pathology, Ohio Valley Medical Center, Wheeling, W.Va., USA

Introduction

Thyroid nodules are very common, with clinically palpable nodules in 5% of adults and ultrasound-detected nodules in up to 68% of randomly selected individuals [1–4]. However, only 5–15% of thyroid nodules are malignant [5]. Although clinical assessment, thyroid-stimulating hormone measurement and ultrasound evaluation are able to enrich the yield of malignant nodules, fine-needle aspiration (FNA) plays a vital role in evaluating thyroid nodules with assessing cellular morphologic features [6, 7]. FNA can classify most thyroid nodules with a definitive diagnosis of either benign or malignant, but was 50% in cases suspicious for follicular neoplasm (SFN). The surgical excisional rate was significantly decreased in SFN cases after the Afirma test. Conclusions: The Afirma GEC is useful for further risk stratifying SFN cases.
Hürthle cell neoplasm (HCN) [6, 7]. Current guidelines recommend a repeat FNA for category III lesions, and lobectomy for category IV lesions and repeated category III lesions. However, only 15–35% prove to be malignant on follow-up surgical specimens, indicating that unnecessary surgeries are being performed for the majority of patients, leading to increased medical cost and surgical complications [8, 9]. Although molecular tests with specific gene mutations/rearrangements (including BRAF, RAS, RET/PTC and PAX8-PPARr, etc.) have a high specificity in detecting malignant cases, they have limited sensitivity and negative predictive value (NPV) and fail to detect as many as 30% of malignant cases [10–13]. Therefore, a preoperative test with a high sensitivity and NPV is needed to accurately identify benign nodules with indeterminate cytology and avoid unnecessary diagnostic surgery.

The Afirma gene expression classifier (GEC) developed by Veracyte Inc. (South San Francisco, Calif., USA) measures the expression of 142 genes to determine if an FNA sample with indeterminate cytology is benign or suspicious for malignancy. The Afirma GEC test is reported to have an NPV of 94–95% and a positive predictive value (PPV) of 37–38% in thyroid nodules with FLUS/AUS or FN/HCN [14, 15]. Beside the initial study, only limited studies with the Afirma GEC have been reported, with variable NPVs and PPVs [16, 17]. In this study, we aimed to review the experience of the Afirma GEC test in our institution and investigate the impact of its result on the surgical management of thyroid nodules with category III/IV cytology and its correlation with surgical outcome and cytomorphologic features.

Materials and Methods

Patient Selection

After institutional review board approval at Ohio State University, a pathology archive database search was performed for two periods of 3 years (the post-Afirma period, July 2012 to June 2015, and the pre-Afirma period, July 2009 to June 2012) to retrieve thyroid FNA cases. The Afirma GEC test was implemented during July 2012 for thyroid nodules with FLUS/AUS, SFN or HCN in our institution. The surgical pathology diagnosis was matched to the original FNA nodule by location and size. Incidental papillary thyroid microcarcinomas that did not match the original FNA nodule were excluded. All initial FNA cases with an interpretation of FLUS/AUS, but with repeat FNAs which were interpreted as benign, FLUS/AUS, SFN, suspicious for malignancy or malignancy, were also excluded from the current study cohort. However, the final repeat FNAs with an interpretation of FLUS/AUS were included. The majority of FLUS/AUS cases sent for Afirma test in our cohort were repeat FNAs (81%, 72/89), with a small portion being initial FNAs (19%, 17/89).

Table 1. Afirma GEC results in 158 thyroid nodules with category III/IV cytology

<table>
<thead>
<tr>
<th>Afirma results</th>
<th>FLUS/AUS</th>
<th>SFN</th>
<th>HCN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>41 (46)</td>
<td>22 (39)</td>
<td>0 (0)</td>
<td>63 (40)</td>
</tr>
<tr>
<td>Suspicious</td>
<td>41 (46)</td>
<td>31 (55)</td>
<td>13 (100)</td>
<td>85 (54)</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>7 (8)</td>
<td>3 (5)</td>
<td>0 (0)</td>
<td>10 (6)</td>
</tr>
<tr>
<td>Total</td>
<td>89</td>
<td>56</td>
<td>13</td>
<td>158</td>
</tr>
</tbody>
</table>

Values in parentheses are percentages.

Thyroid FNA Procedure

FNA biopsies were performed by radiologists under ultrasound guidance. The specimens were submitted for Romanowsky staining and/or Papanicolaou staining and additional material was submitted for cytospin preparations with Papanicolaou staining. The FNA cytology diagnoses were rendered by 6 cytopathologists using the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) criteria. All cases with Afirma GEC testing were further reviewed by Z.L. or R.S. to evaluate the cytomorphologic features.

Surgical Follow-Up Results and Histologic Pathology

The Afirma GEC results and surgical follow-up results were collected. Surgical specimens were received in 10% formalin, embedded in paraffin and stained with standard hematoxylin and eosin. Histologic diagnoses were rendered by subspecialized head/neck surgical pathologists.

Afirma GEC Testing

Besides passes for cytology analysis during the FNA procedure, one extra needle pass was also collected separately for the Afirma GEC test in our institution. Thyroid nodules with a cytology interpretation of FLUS/AUS, SFN or HCN were sent to Veracyte Inc. for the Afirma GEC test at the request of the submitting clinician and the results were reported back to our institution and recorded in the patient’s electronic medical record. Veracyte Inc. did not perform the cytology evaluation.

Statistical Analysis

Data were recorded using Microsoft Excel spreadsheet software (Microsoft, Redmond, Wash., USA). Fisher’s exact test was used to compare the difference with a two-sided analysis. All the analyses were done using the SAS 9.3 system (SAS Institute, Cary, N.C., USA), and a p value <0.05 was considered statistically significant.

Results

Afirma GEC Results and Correlation with Surgical Outcomes and Cytomorphologic Features

During a 3-year period after Afirma’s implementation, 376 (15%) FNAs were identified from a total of 2,560 as having a Bethesda System interpretation of category III/IV cytology (FLUS/AUS, SFN or HCN). In total, 158 out

Chaudhary/Hou/Shen/Hooda/Li
of 376 (42%) cases were sent for Afirma GEC analysis; these included a benign result in 63 (40%), suspicious result in 85 (54%) and unsatisfactory result in 10 (6%). When stratified by different categories, Afirma GEC was able to reclassify 46% of the FLUS/AUS cases and 39% of SFN cases as benign, but all 13 HCN FNAs showed suspicious results by Afirma GEC test (table 1).

Seventy-three (86%) out of 85 suspicious Afirma cases had surgery; 45 (62%; including 28 FLUS/AUS, 13 SFN and 4 HCN cases) were histologically benign and 28 (38%; including 7 FLUS/AUS, 13 SFN and 8 HCN cases) showed carcinoma [including 11 classic papillary thyroid carcinomas, 11 follicular variant papillary thyroid carcinomas (FVPTC; 3 invasive types and 8 noninvasive types) and 6 follicular carcinomas]. In contrast, only 8 (13%) out of 63 benign Afirma cases had surgery and all of them were histologically benign (table 2). Also, 5 (50%) out of 10 cases with an unsatisfactory Afirma result had surgery, with 4 cases as benign and 1 case as malignant histologically. Twelve HCN cases with suspicious Afirma results had surgical follow-up, with 4 cases as histologically benign and 8 cases as histologically malignant.

Afirma’s overall sensitivity, specificity, NPV and PPV in Bethesda category III/IV thyroid nodules was 100, 15, 100 and 38%, respectively. Among the different FNA categories, PPV was lowest (20%) in the FLUS/AUS group, 50% in the SFN group and highest (67%) in the HCN group (table 3).

Comparison between Pre-Afirma Data and Post-Afirma Data: The Distribution of FNA Interpretation Categories and Surgical Excision Rates of Thyroid Nodules with Bethesda Category III/IV Cytology

We also examined whether the implementation of the Afirma GEC test had any effect on the interpretation of thyroid FNA by cytopathologists. The distribution of each FNA category except suspicious for malignancy did not show a significant change after the implementation of the Afirma GEC test (data not shown). The decrease of FNAs with an interpretation of suspicious for malignancy in post-Afirma data was offset by the slight increase of malignant FNAs, indicating increased confidence among the cytopathologists. Each category’s percentage was also within the reported ranges. Similarly, no significant difference in patient age and female:male ratio was identified between the pre-Afirma data and post-Afirma data (data not shown).

Next, we examined whether the implementation of the Afirma GEC test had any effect on the surgical excision rates of thyroid nodules with category III/IV cytology. Since more than half of the category III/IV cases were not sent for an Afirma test, the comparison was analyzed between 158 cases with an Afirma test and pre-Afirma cases. The comparison data showed that the surgical excision rate was significantly decreased in the SFN group (76 vs. 52%, p = 0.001) after Afirma test, but not in the FLUS/AUS group (51 vs. 51%) or HCN group (69 vs. 92%; table 3).

---

Afirma Test on Thyroid Nodules

<table>
<thead>
<tr>
<th>Afirma results</th>
<th>FLUS/AUS with surgical follow-up</th>
<th>SFN with surgical follow-up</th>
<th>HCN with surgical follow-up</th>
<th>Total with surgical follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>6 (100)</td>
<td>2 (100)</td>
<td>4 (33)</td>
<td>8 (100)</td>
</tr>
<tr>
<td>Suspicious</td>
<td>28 (80)</td>
<td>13 (50)</td>
<td>8 (67)</td>
<td>45 (62)</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>3 (75)</td>
<td>1 (100)</td>
<td>0 (0)</td>
<td>4 (80)</td>
</tr>
<tr>
<td>Total</td>
<td>37 (82)</td>
<td>16 (55)</td>
<td>13 (45)</td>
<td>57 (66)</td>
</tr>
</tbody>
</table>

Values in parentheses are percentages.

Table 3. Sensitivity, specificity, NPV and PPV of the GEC in FLUS, SFN and HCN categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>NPV, %</th>
<th>PPV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLUS/AUS</td>
<td>100</td>
<td>18</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>SFN</td>
<td>100</td>
<td>13</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>HCN</td>
<td>100</td>
<td>0</td>
<td>n.a.</td>
<td>67</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>15</td>
<td>100</td>
<td>38</td>
</tr>
</tbody>
</table>

n.a. = Not statistically significant.
Since cases with a benign Afirma result should not have had surgery, we excluded the cases with a benign Afirma test and surgery from the post-Afirma data and then recompared the surgical excision rates. As shown in Table 4, the surgical excision rate was significantly decreased in the total cases (61 vs. 49%, p = 0.009) and SFN cases (76 vs. 48%, p = 0.0002) after Afirma test, but still not in the FLUS/AUS cases (51 vs. 44%) and HCN group (69 vs. 92%).

The overall malignancy rates in all category III/IV cases did not significantly change after Afirma GEC implementation (Table 5). There was an increased tendency in malignancy rates in SFN cases with surgical excision, but this was not statistically significant due to the small sample size.

### Table 4. Surgical follow-up rates for patients with category III/IV cytology between the pre- and post-Afirma periods

<table>
<thead>
<tr>
<th>FNA categories</th>
<th>Pre-Afirma</th>
<th>Post-Afirma cases with Afirma test</th>
<th>Post-Afirma cases with suspicious Afirma results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with surgery, n</td>
<td>total cases, n</td>
<td>%</td>
</tr>
<tr>
<td>FLUS/AUS</td>
<td>84</td>
<td>165</td>
<td>51</td>
</tr>
<tr>
<td>SFN</td>
<td>78</td>
<td>102</td>
<td>76</td>
</tr>
<tr>
<td>HCN</td>
<td>9</td>
<td>13</td>
<td>69</td>
</tr>
<tr>
<td>Total</td>
<td>171</td>
<td>280</td>
<td>61</td>
</tr>
</tbody>
</table>

n.s. = Not statistically significant.

1 These percentages are ratios of the suspicious Afirma nodules that had undergone surgery to all Afirma-tested nodules.

### Table 5. Malignant surgical follow-up rates for category III/IV cases with surgery between the pre- and post-Afirma periods

<table>
<thead>
<tr>
<th>FNA categories</th>
<th>Malignant/category III/IV cases with surgery</th>
<th>Malignant/category III/IV cases with surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pre-Afirma cases</td>
<td>post-Afirma cases with Afirma test</td>
</tr>
<tr>
<td>FLUS/AUS</td>
<td>23/84 (27)</td>
<td>8/45 (18)</td>
</tr>
<tr>
<td>SFN</td>
<td>27/78 (35)</td>
<td>13/29 (45)</td>
</tr>
<tr>
<td>HCN</td>
<td>5/9 (56)</td>
<td>8/12 (67)</td>
</tr>
<tr>
<td>Total</td>
<td>55/171 (32)</td>
<td>29/86 (34)</td>
</tr>
</tbody>
</table>

Values in parentheses are percentages. n.s. = Not statistically significant.

### Discussion

Consistent with the findings from Afirma’s initial studies, our data demonstrated that the Afirma GEC test was able to reclassify 40% of thyroid nodules with category III/IV cytology as benign, theoretically preventing them from undergoing unnecessary surgical excision [15, 18]. In our study, although the surgical excision rate in all cases with category III/IV cytology seemed to decrease from 61 to 54% after the implementation of Afirma GEC, the statistical analysis did not show a significant difference, which may have been caused by the small sample size. However, the surgical excision rate was significantly decreased in the SFN group (76 vs. 52%, p = 0.001) after the Afirma test. Although Afirma’s overall sensitivity and NPV in thyroid nodules with category III/IV cytology

---

4 Acta Cytologica
DOI: 10.1159/000446797

Chaudhary/Hou/Shen/Hooda/Li
were both 100% due to the absence of false negative cases, its specificity and PPV were only 15 and 38%, respectively. These analytical performance characteristics are consistent with the observations of another study [19]. However, the fact that only 8 cases with a benign Afirma result were surgically confirmed limited the accuracy of the sensitivity, specificity and NPV of the Afirma GEC test obtained in the current study, the results of which should therefore be interpreted with caution. One recent study investigated 90 thyroid nodules with indeterminate cytology and benign Afirma during a median follow-up period of 13 months and found that these nodules demonstrated a similar growth on sonogram to nodules with benign cytology, suggesting that the assessment of thyroid nodules with indeterminate cytology and benign Afirma may be performed similarly to those with benign cytology [20]. Two other recent studies evaluated the long-term management patterns and thyroid surgery rates of Afirma-benign patients compared to cytopathology-benign patients and found that patients with benign Afirma and cytopathology diagnoses were managed similarly and a nonoperative approach to follow-up was considered to be a safe alternative to diagnostic surgery by the majority of physicians [21, 22]. When stratified by different categories, the PPV was significantly improved to 50% in the SFN group, while both sensitivity and NPV remained at 100%, indicating a better performance of Afirma GEC in SFN cases in predicting the surgical outcome.

Indeed, when stratified by different cytology categories, a decreased surgical excision rate was only observed in the SFN group, but not in the FLUS/AUS and HCN groups. The similarity in surgical excisional rate in the FLUS/AUS group during the pre-Afirma versus post-Afirma period may be due to the low surgical excision rate during the pre-Afirma period (51%) and the low PPV of the Afirma GEC test (20%). In other words, Afirma GEC only reclassified about half of the FLUS/AUS or SFN cases as benign, and almost all of the other cases with suspicious Afirma GEC results underwent surgical excision. With a surgical excision rate of 51% before Afirma GEC in the FLUS/AUS group, the implementation of Afirma did not significantly decrease the surgical excision rate in the FLUS/AUS group. However, with a higher surgical excisional rate of 76% for the SFN group during the pre-Afirma period, the implementation of Afirma did significantly decrease the surgical excision rate. The findings are consistent with the study by Alexander et al. [15], which demonstrated that the most significant impact of the Afirma GEC test was noted among SFN thyroid nodules. However, in contrast to that study, we did not observe any impact of the Afirma GEC test on FLUS/AUS thyroid nodules. Repeat FNA was demonstrated to be able to redefine 75–80% of category III/IV cases as either benign or malignant, thus leaving only 20–25% of nodules as repeatedly indeterminate [23, 24]. In light of the results from the current study, it may be better to repeat FNA rather than to perform the Afirma GEC test for FLUS/AUS cases, even for repeated FLUS/AUS cases. However, further prospective studies are warranted to perform a point to point analysis between repeat FNA and the Afirma test for FLUS/AUS cases and an endpoint needs to be set for repeating FNA, at which stage the Afirma GEC test should be performed.

In our study, all HCN cases showed suspicious Afirma GEC results and the surgical excision rate was very high. Previous studies have demonstrated that the presence of Hürthle cells causes false positive Afirma GEC results [17, 25, 26]. Two recent studies have demonstrated that a suspicious Afirma GEC result does not increase the probability of malignancy in an HCN lesion [27, 28]. However, one of these studies [27] showed 63% (45/72) of patients had suspicious Afirma GEC results, which is much lower than the suspicious rate (100%) in our results. The discrepancy may be caused by the small sample in our study (13 cases). Therefore, the Afirma GEC test for HCNs should be considered with caution and the Afirma GEC result in such cases should be interpreted accordingly.

There was an increased tendency of malignancy rates in SFN cases with surgical excision after the Afirma test, but this was not statistically significant due to the small sample size. A significant portion of malignant cases from both pre- and post-Afirma were FVPTC (40% in pre-Afirma and 42% in post-Afirma), indicating the difficulty in making a diagnosis of FVPTC on cytology because the characteristic nuclear features are less frequently appreciated. Similarly, a diagnosis of follicular carcinoma requires capsular invasion or angioinvasion; therefore, it is almost impossible to make that diagnosis on cytology. With the presence of both FVPTC and follicular carcinoma cases in our cohort, the Afirma GEC test seems to be able to depict all thyroid carcinomas.

One of the limitations of the current study is that no clinical presentation/follow-up or imaging information was available for the study cohort, making it difficult to analyze why half of the FLUS/AUS cases were not sent for Afirma GEC test and analyze why some of the cases without Afirma GEC test did not undergo surgical excision. The other aspect our current study did not investigate is the cost analysis of Afirma GEC implementation in our institution. Even though a previous study has suggested

Afirma Test on Thyroid Nodules

DOI: 10.1159/000446797

Acta Cytologica

5
that the use of Afirma GEC testing in category III/IV thyroid nodules may be cost effective because it prevents unnecessary surgeries [29], whether the saving from reduced surgeries in SFN thyroid nodules is able to justify the cost of implementation of the Afirma GEC test in all category III/IV FNA cases in our institution warrants future studies with a detailed cost analysis.

In conclusion, our data have demonstrated that the use of the Afirma GEC test is able to reduce the number of unnecessary surgical excisions in thyroid nodules with SFN, but not in thyroid nodules with FLUS/AUS or HCN. Therefore, it may be better to repeat FNA rather than perform the Afirma GEC test for FLUS/AUS, and ordering an Afirma GEC test on HCN cases needs to be cautioned and the results should be interpreted accordingly.

Disclosure Statement

The authors have no financial conflicts of interest to disclose.

References