An Independent Study of a Gene Expression Classifier (Afirma) in the Evaluation of Cytologically Indeterminate Thyroid Nodules

Bryan McIver, M. Regina Castro, John C. Morris, Victor Bernet, Robert Smallridge, Michael Henry, Laura Kosok, and Honey Reddi

Department of Head and Neck, and Endocrine Oncology (B.M.), Moffitt Cancer Center, Tampa, Florida 33612; Division of Endocrinology (M.R.C., J.C.M., L.K.), Mayo Clinic, Rochester, Minnesota 55905; Division of Endocrinology (V.B., R.S.), Mayo Clinic, Jacksonville, Florida 32216; Department of Surgical Pathology (M.H.), Mayo Clinic, Rochester, Minnesota 55905; and Prevention Genetics (H.R.), Marshfield, Wisconsin 54449

Context: Molecular markers hold the promise of improved diagnostic yield in thyroid fine-needle biopsy. The Afirma gene expression classifier (GEC), available commercially, reports a negative predictive value of 94% in the diagnosis of benign nodules after indeterminate cytology. However, there are currently no independent studies of the performance of this assay.

Objective: The aim was to assess the performance of the Afirma GEC in an academic medical center.

Design: Samples for the GEC were collected according to the manufacturer's recommended protocol from patients undergoing thyroid fine-needle aspiration. We requested GEC analysis on nodules reported cytologically as follicular neoplasm or atypia or follicular lesion of undetermined significance from patients willing to defer surgery.

Patients: All patients undergoing thyroid fine-needle aspiration during the study period, whose cytology was reported as follicular neoplasm or atypia of undetermined significance/follicular lesion of undetermined significance, were offered access to the test and recruited to this study.

Intervention: Patients whose GEC was "benign" were offered ultrasound follow-up in lieu of surgery. Those with a "suspicious" GEC were advised to undergo diagnostic lobectomy.

Setting: The study was conducted at a large academic medical center.

Main Outcome Measure: We measured the rate of benign and suspicious calls from the Afirma GEC and histological diagnosis after surgery.

Results: A total of 72 nucleic acid samples were sent for GEC analysis. In 12 (17%) of these samples, there was insufficient mRNA, leaving 60 Afirma results for analysis. Of these, 16 (27%) were benign, whereas 44 (73%) were suspicious. The rate of confirmed malignancy in GEC-suspicious nodules was only 17%.

Conclusion: The Afirma GEC demonstrates a lower than expected rate of benign reports in follicular or Hürthle cell neoplasm and a lower than anticipated malignancy rate within GEC-suspicious nodules. These data suggest that the positive predictive value of the GEC is lower than previously reported and call into question the performance of the test when applied in the context of specialized academic cytopathology. **(*J Clin Endocrinol Metab* 99: 4069–4077, 2014)**

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

Received September 24, 2013. Accepted April 22, 2014.
First Published Online April 29, 2014

For article see page 4062

Abbreviations: AUS, atypia of undetermined significance; FLUS, follicular lesion of undetermined significance; FN, follicular neoplasm; FNA, fine-needle aspiration; GEC, gene expression classifier; HCN, Hürthle cell neoplasm; NPV, negative predictive value; PPV, positive predictive value; PTC, papillary thyroid carcinoma.

Thyroid nodules are the most common endocrine tumor, with population-based screening studies identifying clinically palpable nodules in approximately 5% of adults, whereas ultrasound and autopsy studies demonstrate nodules in more than 50% of women and 20% of men over age 50 (1). Thyroid cancer, which usually presents as a nodule, is uncommon but is increasing in incidence more rapidly than any other cancer type (2). However, only a small minority of nodules, whether palpable or incidentally discovered, proves to be malignant (3). Selection of nodules for biopsy based on suspicious ultrasound features enriches the yield of malignant nodules, but the proportion that ultimately proves malignant remains a mere 10–15% (4).

Both the American Thyroid Association (ATA) and the American Association of Clinical Endocrinologists (AACE), working in collaboration with the Associazione Medicie Endocrinologi and the European Thyroid Association, have published guidelines for the evaluation of thyroid nodules that recommend a multistep strategy: clinical assessment, measurement of TSH, ultrasound evaluation, and biopsy of nodules selected according to size and ultrasound characteristics (5, 6). For those nodules that require biopsy, fine-needle aspiration (FNA) cytology provides sufficient information to classify most nodules as benign (72%; range, 62–85%), whereas approximately 5% (range, 1–8%) of nodules are cytologically malignant, most often papillary thyroid carcinoma (PTC) (7).

However, 10–30% of biopsied nodules exhibit "indeterminate" cytology, including the subtypes of atypia of undetermined significance (AUS) or follicular lesion of undetermined significance (FLUS); (suspicious for) follicular neoplasm (FN) or Hürthle cell neoplasm (HCN); and suspicious for malignancy (8). Current guidelines recommend surgical resection for most of these nodules to permit adequate pathological evaluation, although repeat biopsy is supported for AUS/FLUS when the risk of malignancy is felt to be sufficiently low, which occurs particularly when the categorization of AUS/FLUS is driven primarily by features other than nuclear atypia (<10% vs >50% risk of malignancy) (9).

Overall, only 15–35% of nodules with indeterminate cytology prove to be malignant on histological evaluation, usually either follicular variant of PTC or follicular carcinoma (7, 8). For these cancers, a lobectomy—the diagnostic procedure of choice—is regarded as inadequate therapeutic surgery in most cases, so these patients will require a second surgical procedure to complete the initial treatment of their malignancy. A more specific test for the preoperative diagnosis of malignancy might aid in determining the extent of initial surgery. For the 65–85% of indeterminate nodules that prove benign, however, a lobectomy is arguably too much surgery, required only as a diagnostic rather than a therapeutic procedure, in the absence of local compressive symptoms or cosmetic concerns. A more sensitive preoperative test that would allow safe identification of clearly benign nodules with indeterminate cytology raises the prospect of avoiding a purely diagnostic surgery. It is in these areas that molecular markers have been developed and are being marketed to improve the preoperative evaluation of thyroid nodules.

The Afirma gene expression classifier (GEC) is a proprietary diagnostic test developed by Veracyte, Inc, for the preoperative identification of benign thyroid nodules with indeterminate cytology. Testing is offered through a single Clinical Laboratory Improvement Amendments (CLIA)-certified reference laboratory. The assay classifies nodules as either benign or suspicious for malignancy, with reported post-test probability of malignancy of 5–6% for a "benign" result and 37–38% for a "suspicious" result in AUS/FLUS and FN/HCN, usually described as a negative predictive value (NPV) of 94–95% and a positive predictive value (PPV) of 37–38% (10, 11). Because the risk of malignancy for a thyroid nodule with AUS/FLUS or FN/HCN and a benign GEC diagnosis is reported to be comparable to that of an operated nodule with a benign cytopathology diagnosis (11, 12), observation or ultrasound follow-up has been recommended in lieu of thyroid surgery in a recent revision of guidelines for the management of thyroid cancer, issued by the National Cancer Cooperative Network (NCCN) (13). The data that support this approach, however, are limited to a single confirmation study, which encompassed only 81 FN/HCN and 129 AUS/FLUS nodules, with wide confidence intervals for sensitivity (68–99%) and specificity (36–63%) and significant consequent uncertainty in both the NPV (79–99%) and PPV (23–52%) of the test, even as applied within this trial setting, with a known pretest probability of malignancy (11). Two subsequent studies have been reported, demonstrating higher PPV (54 and 57%, respectively) and implying a lower NPV (estimated at 92 and 90%), although neither of these studies was able directly to assess NPV because most patients with benign GEC results have not undergone surgery (14, 15). To increase the reported experience of the GEC, we therefore set out to assess the performance of the test as applied in a large academic, multispecialty clinic setting, using both highly specialized cytopathology services and a clinical assessment to determine which samples should be reflexed to the GEC assay.

Patients and Methods

Mayo Clinic was the first center that was not part of the industry-sponsored confirmatory study (11) to receive the Veracyte designation of "Enabled Center." This designation allowed our group, and subsequently a limited number of other academic centers, to send appropriately collected nucleic acid specimens directly for Afirma GEC analysis, based on our in-house specialized cytopathology services, rather than being required to use Thyroid Cytopathology Partners (TCP, Austin, Texas), the cytopathology group chosen by Veracyte to act as gatekeepers to the commercial assay. Starting in May 2011, we began to offer access to the GEC to adult patients over 21 years of age, with nodules >1.0 cm, whose cytology was reported as FN/HCN or AUS/FLUS. After obtaining written informed consent, an ultrasound-guided FNA biopsy was conducted in the usual way, using four to six passes with a 27-gauge needle placed into the index nodule under direct ultrasound control. Either capillary action or gentle aspiration was used to obtain cytological material in a minimally traumatic fashion, and the aspirated material was spread onto labeled glass slides and fixed in ethanol preservative for cytological evaluation. At the end of the procedure, either one or two additional needle passes were collected from the index nodule and washed into the GEC collection tube containing a proprietary nucleic acid preservative, following the protocol recommended by Veracyte.

Patients were offered access to the test if they attended Mayo Clinic for assessment of a thyroid nodule and underwent thyroid FNA within the Endocrine Clinic in Rochester, Minnesota, between May 2011 and December 2012; in Jacksonville, Florida, between October 2011 and December 2012; or in the Radiology department of Mayo Clinic, Rochester, between December 2011 and December 2012. Afirma GEC samples were labeled and stored at 4°C, pending receipt of the cytology report, typically within 2–4 hours. After a further discussion with the patient, the samples from nodules that were reported either as FN/HCN or AUS/FLUS, for patients who were willing to defer surgery and who were deemed not to be at particularly high risk for malignancy (no history of head or neck irradiation; negative family history of thyroid cancer; no prior history of thyroid cancer; no worrisome imaging characteristics) were shipped to Veracyte for GEC analysis. Patients with symptomatic nodules, with worrisome imaging features, or with significant risk factors for thyroid cancer, in whom a decision had already been made to take the patient to surgery, were not recommended for the GEC. Their nucleic acid samples were disposed of, along with those from patients with cytology that was unsatisfactory, definitively benign, malignant, or suspicious for malignancy.

Samples were shipped on ice to Veracyte's Clinical Laboratory Improvement Amendments laboratory in South San Francisco, California, by overnight courier using the cold-packs and shipping containers provided by Veracyte. Samples that would arrive on Saturday morning were held over the weekend at 4°C and shipped on Monday morning for a Tuesday morning delivery. Afirma-GEC reports were received, typically within 7–10 days, and scanned into the patients' electronic medical record.

We maintained a prospective register of patients undergoing the GEC assay, including demographic, ultrasonographic, cytological, surgical, histological, and follow-up findings. Patients with AUS/FLUS or FN/HCN and a GEC result of suspicious were advised to undergo diagnostic lobectomy, intraoperative frozen section, and immediate completion thyroidectomy if malignancy was diagnosed. Patients whose GEC was reported to be benign were offered the option of continued ultrasound follow-up, with plans to intervene surgically only if the nodule enlarged or changed over time. A structured follow-up plan was implemented to ensure that these patients were not lost to follow-up, with repeat ultrasound recommended after 3–6 months, and then annually for at least 5 years.

Statistical analysis was performed using the SPSS Statistical Package (SPSS Inc). Comparisons were made using binomial distribution statistics, and confidence intervals were estimated using the normal distribution approximations.

Results

As illustrated in Figure 1, we performed FNA biopsies on 1207 nodules from 984 patients within the "enabled system," which permitted collection of a nucleic acid sample. Of these, 12 (1.0%) were categorized as AUS/FLUS on cytology, and 93 (7.7%) were FN/HCN, rates that are in keeping with previously published data from this institution, although lower than those reported from several other academic and community centers (14). The necessary nucleic acid sample was not collected in 15 of these patients, at the discretion of the physician performing the biopsy, leaving a total of 90 GECs available for analysis from cytologically indeterminate nodules in eligible patients (10 with a cytological diagnosis of AUS/FLUS, 17 HCN, and 63 FN).

Of these 90 patients, 18 (20%) chose to move ahead with a diagnostic or therapeutic lobectomy, rather than consider use of the GEC. In six of these cases, the decision was prompted by concerns about a highly suspicious appearance on ultrasound (n = 3), the synchronous discovery of a contralateral malignancy (n = 2), or (in one case) the finding that the nodule was positron emission tomography-positive, leading to increased concern about a malignant condition. In an additional four cases, the nodule was large and/or causing compressive symptoms. The remaining eight patients underwent surgery because either the patient or their treating physician was sufficiently concerned about the nodule that a decision not to operate would have been an unacceptable choice. Consequently, the GEC was not performed on any of these 18 patients. Thyroid lobectomy was performed for diagnostic purposes on all of these cases, except the two in whom a contralateral malignancy was identified on preoperative cytology.

A total of 72 nucleic acid samples were sent to Veracyte for GEC analysis. In 12 (17%) of these samples, there was insufficient quantity or quality of mRNA to obtain a result, leaving a total of 60 Afirma results available for analysis. Of these, 16 (27%) were reported as benign, while 44 (73%) were reported as suspicious. Patients with benign GEC results were offered the opportunity to defer diagnostic surgery and undergo ultrasound-based follow-up as an alternative. Fourteen of the 16 patients with benign

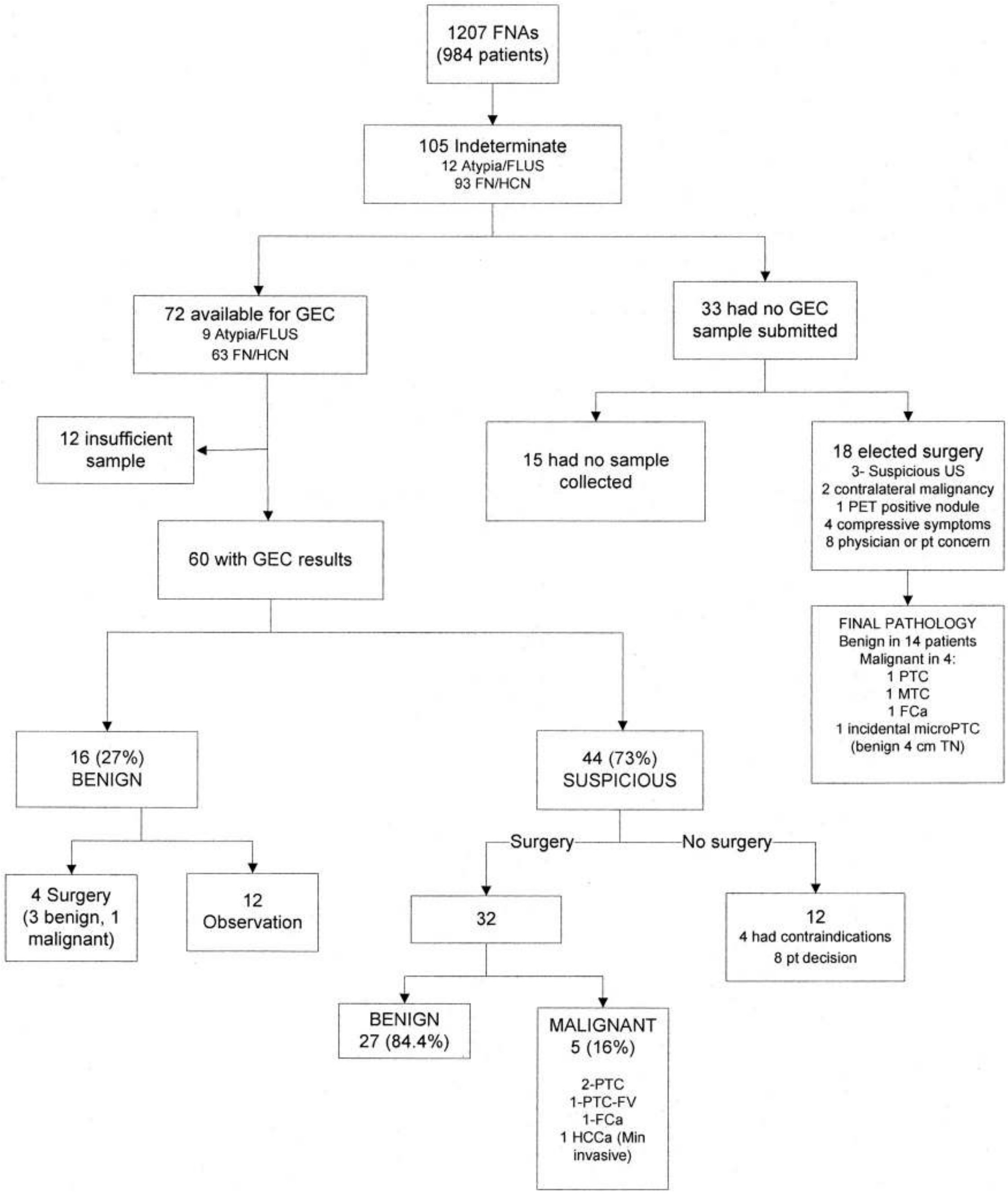

Figure 1. Flow diagram of patient disposition and outcomes in the study.

GEC reports chose this conservative management, whereas two patients chose surgery, a rate of conservative management that closely correlates with the findings of our previous clinical utility study (12). Two patients with a GEC-benign classifier (one HCN, one FN) underwent later surgery: in one case because of the growth of the index nodule after 11 months of follow-up, and in the other case because of the patient's preference 6 months later, despite lack of growth. Of these four GEC-benign nodules that were resected, three proved to be histologically benign at surgery, whereas the fourth (which demonstrated no growth on follow-up) proved to be a 3.2-cm follicular carcinoma, exhibiting focal capsular and vascular invasion. Twelve patients with benign-GEC results remain under active follow-up after a median of 9.5 months.

Of the 44 patients with suspicious GEC results, 32 (73%) have undergone surgery to date. Of the remaining 12 patients, four have contraindications to surgery, and eight have chosen to avoid surgery despite medical recommendations to proceed. A total of five cancers have been identified in the 32 operated thyroids, giving a posttest probability of cancer of 15.6%. These malignancies encompassed two classic PTC, one follicular variant PTC, and one follicular thyroid cancer. The remaining lesion

Table 1. Performance of GEC According to Final Histopathology Diagnosis for Cytologically Indeterminate (Suspicious) Nodules That Underwent Surgery: Suspicious for Follicular Neoplasm (SFN) (n = 18)

	Histopathology		
	Malignant	**Benign**	**Total**
Suspicious	2[a]	13[b]	15
Benign	1[c]	2[d]	3
Total	3	15	18

[a] One papillary carcinoma, one follicular carcinoma; [b] nine follicular adenomas, one nodular hyperplasia, three Hurthle cell adenomas; [c] one follicular carcinoma; [d] two follicular adenomas.

was described as either an "atypical adenoma with Hürthle cell features" or a minimally invasive Hürthle cell cancer. For the purpose of this analysis, we have considered this nodule to be a malignancy. A total of 31 patients with cytology suspicious for neoplasm (18 FN and 13 HCN) underwent surgery. Tables 1 to 5 summarize the results according to GEC category and final histopathology. Five patients with cytology read as AUS/FLUS had samples submitted for GEC. All of these were read as GEC suspicious and underwent surgery. Of these, one lesion proved to be malignant (microPTC arising within a larger adenoma); the others showed one follicular adenoma, one adenomatous nodule, and two hyperplastic nodules. Because most of our patients with a benign GEC report have not undergone surgery, we are unable to directly assess the probability of cancer in GEC-benign nodules (the NPV). These patients remain under careful long-term follow-up.

Discussion

Remarkable advances have been made over the last two decades in our understanding of the genetic and molecular changes that drive thyroid neoplasia (16). Chromosomal rearrangements involving the RET proto-oncogene or the V600E point mutation in the BRAF gene underpin most PTC, whereas mutations of RAS and rearrangements of the PPARγ gene have been implicated in a significant proportion of follicular thyroid cancer. Nikiforov et al (17, 18) have been instrumental in developing the concept of using a panel of oncogenes to more accurately diagnose thyroid cancers in the preoperative phase. The high specificity of these oncogenes in predicting cancer—particularly BRAF and RET/PTC, with specificities close to 100%—allows the surgeon to plan an appropriate cancer surgery when a mutation is present, even when cytology alone is indeterminate. However, our understanding of oncogenic triggers is incomplete, particularly for Hürthle cell cancer and follicular variant PTC, in which only a minority exhibits RAS mutations (16, 19). Consequently, the currently available oncogene panels lack sensitivity in the detection of these malignant subtypes, resulting in a number of false-negative results.

An alternative approach has been developed by Veracyte, Inc, and marketed under the trade name Afirma. Based on a commercially available gene expression chip (Affymetrix Inc), Veracyte has developed a "gene expression classifier" (GEC), using a proprietary algorithm to distinguish benign from suspicious nodules, based on the expression pattern of mRNA extracted from one or two dedicated FNA needle passes. The algorithm utilizes a screening "cassette" of 25 genes designed to identify medullary thyroid cancer and certain metastatic malignancies. Thereafter, the expression levels of 142 genes are processed through a "support vector machine" that classifies the expression pattern as either benign or suspicious (10). The support vector machine is a supervised machine-learning algorithm that identifies patterns of gene expression in a recursive learning process, which was trained on

Table 2. GEC Results: Suspicious for HCN (SHCN) (n = 13)

	Histopathology		
	Malignant	**Benign**	**Total**
Suspicious	2[a]	10[b]	12
Benign	0	1[c]	1
Total	2	11	13

[a] One papillary carcinoma, one Hürthle cell carcinoma; [b] one benign (per report), one follicular adenoma with HC change, seven Hürthle cell adenomas, one Hashimoto's thyroiditis; [c] one Hürthle cell adenoma.

Table 3. GEC Results: Suspicious for Neoplasm (SFN + SHCN) (n = 31)

	Histopathology		
	Malignant	**Benign**	**Total**
Suspicious	4	23	27
Benign	1	3	4
Total	5	26	31

Sensitivity (4/5), 80%; specificity (3/26), 12%; PPV (4/27), 15%; NPV(3/4), 75%; prevalence of malignancy = 16%.

Table 4. GEC Results: ATYPIA/FLUS (n = 5)

	Histopathology		
	Malignant	**Benign**	**Total**
Suspicious	1[a]	4[b]	5
Benign	0	0	0
Total	1	4	5

[a] One microPTC arising in larger adenoma; [b] one follicular adenoma, one adenomatous nodule, two hyperplastic nodules.

Table 5. GEC Results: Performance Across the Entire Dataset of Indeterminate Samples (SFN + SHCN + Atypia/FLUS)

	Histopathology		
	Malignant	**Benign**	**Total**
Suspicious	5	27	32
Benign	1	3	4
Total	6	30	36

Sensitivity (5/6), 83%; specificity (3/30), 10%; PPV (5/32), 16%; NPV (3/4), 75%; prevalence of malignancy, 17%.

a set of mRNA samples derived from nodules with known histology. In the only clinical validation study yet published, the performance of the GEC was assessed in a set of 256 FNA samples with indeterminate cytology, including 81 FN/HCN and 129 AUS/FLUS (11). Within the AUS/FLUS group, sensitivity of the GEC was 90%, specificity was 53%, and NPV was 95%, whereas the figures for FN/HCN were 90% sensitivity, 49% specificity, and 94% NPV. The post-test probability of malignancy in GEC-benign, cytologically AUS/FLUS or FN/HCN nodules in this study was therefore 5–6%, a risk deemed acceptable by the NCCN to consider "observation in lieu of surgery." In their reported experience to date, approximately 50% of samples submitted to the GEC are reported as having a benign profile, leading Veracyte to claim that up to 50 000 patients could avoid surgery each year if the GEC was implemented across the country, leading to a substantial net savings to the health care system as a whole (12). Similar rates of benign GEC results have been reported in a recently published study from five academic medical centers, all of which were included in the original clinical validation study (14). However, a community practice-based study, utilizing Veracyte's chosen cytopathology group, TCP, reported that only 34% of indeterminate samples proved benign on the GEC, closer to our data, with only 27% of samples receiving a benign GEC result. Li et al (20), in a theoretical cost-effectiveness modeling analysis, assumed that surgery could be avoided in "almost three-fourths of currently performed surgeries in patients with benign nodules," concluding that the test would be cost-effective, saving money for the health care system as a whole, despite the high direct cost of the assay. The much lower rate of surgery avoidance that could have been achieved by Alexander et al (11) (41% of calls in AUS/FLUS and FN/HCN categories were benign), by Harrell et al (15) (34% benign GEC reports), and in our experience (27% benign) call that analysis into question and suggest that the costs of widespread implementation of this molecular test may be substantially higher than initially reported because of the greater number of tests needed to be run in order to avoid one surgery. In our study, surgery might have been avoided in one patient for every four tests run, whereas one surgery was avoided for every two tests run in a study based on TCP cytopathology (12), in which approximately 50% of tested samples were reported benign on the Afirma GEC ($P < .05$). Furthermore, because of the uncertainty about the true NPV achieved by the GEC, we have implemented a structured follow-up plan to ensure that these patients were not lost to follow-up, with repeat ultrasound after 3–6 months and then annually for at least 5 years. This more intensive follow-up is, we believe, clinically appropriate, but it may add to the overall costs of the assay, compared to the analysis by Li et al (20).

In our study, only 15.6% of the nodules that were GEC suspicious, for which final histology is known, have proven malignant, a PPV for malignancy substantially and significantly lower than the 38% reported by Alexander et al (11) ($P < .05$), which was itself rather lower than the 48% PPV in early reports (10, 11) and dramatically lower than the 57% PPV reported recently by Harrell et al (15).

The most obvious explanation for this discrepancy might be a significantly lower than reported specificity of the Afirma GEC, when applied to samples from nodules reported as AUS/FLUS or FN/HCN by the specialized thyroid cytopathologists at Mayo Clinic, compared to either Veracyte's cytopathology partners, TCP, or to the "local cytopathologists" who provided the data for the Veracyte confirmatory study (11). Because we have elected not to operate on all patients with an Afirma-GEC benign result, we are unable to directly calculate the precise sensitivity and specificity of the assay in our hands, a limitation of all the subsequent reports (14, 15). If we assume that the single false-negative result we have identified (one GEC-benign patient had a follicular carcinoma) is the only such error, the assay would have a sensitivity in our hands of 83% (Table 6), well within the 95% confidence interval of 68–99% reported by Alexander et al (11). However, the

Table 6. Estimates for Sensitivity, Specificity, NPV, and PPV of Afirma GEC

	Histopathology		
Afirma GEC Results	**Malignant**	**Benign**	**Total**
Suspicious	5	27	32
Benign	1	15	16
Total	6	42	48

For the purpose of this analysis we have included all of the GEC-benign nodules (n = 16), but only those GEC-suspicious nodules that have undergone surgery (n = 32). One of 16 GEC-benign nodules has proven to be malignant at surgery; this analysis is based on the presumption that this was the only false-negative result in this group. Sensitivity = 83%; specificity = 36%; prevalence = 12.5%; NPV of benign GEC = 94% (assumes no additional false-negatives); PPV of suspicious GEC = 15.6%.

lower incidence of cancer we have identified in GEC-suspicious nodules yields a specificity of Afirma in our study of only 36%, significantly lower than the 49 and 53% specificity reported by Alexander et al (11) ($P < .05$) in FN/HCN and AUS/FLUS, respectively. The second contributor to the lower PPV for Afirma in our experience is likely to be the low pretest probability of malignancy in the group of nodules we have selected for GEC analysis. Only 6 cancers have been identified in the 36 nodules that have gone to surgery so far, for a histological malignancy rate of 16.7%. Including the one false-negative within the GEC-benign group, the overall malignancy rate in our study is only approximately 12.5%. Although low, this probability of malignancy is close to the 14% risk of malignancy reported for FN/HCN at Mayo Clinic in previous studies (21) that were completed without the use of genetic markers and is similar to the rates of malignancy reported by several other large academic medical centers, including MD Anderson Cancer Center (16%) and Northwestern University (15%), although it stands in contrast to some other centers, including Yale University (48%) and Washington Hospital Medical Center (49%) (for a detailed meta-analysis, see Wang et al, Ref. 22). The pretest risk of malignancy in our study may have been further lowered (from 14% to ~12.5%) by our decision not to send for GEC analysis samples from patients we deemed to be at high risk for cancer. Patients with compressive symptoms, large nodules, highly suspicious ultrasound features, or other risk factors for cancer were advised to undergo surgery, and the GEC was not requested on these patients (Figure 1), a decision that may have depleted the study pool of some cancers but which, in our view, was medically appropriate. The rate of malignancy among this group of patients was, indeed, slightly higher than among the patients for whom the GEC was sent. Four of the 18 (22%) patients for whom the GEC was drawn, but not analyzed, had malignancy diagnosed at the time of surgery, of which three (17%) were cancers within the index nodule. However, because the decision to operate had already been made in these patients, we believe that the GEC would have represented an unnecessary medical expenditure, which would not have altered the management of the patient.

The consequence of this low pretest probability of malignancy is to lower the post-test probability of malignancy for both benign and suspicious results. The relationship between the pretest probability of malignancy and the achieved NPV and PPV of the Afirma GEC is shown in Figure 2A, assuming a sensitivity of 90% and a specificity of 49% for the test as reported by Alexander et al (11). As the pretest probability of malignancy increases,

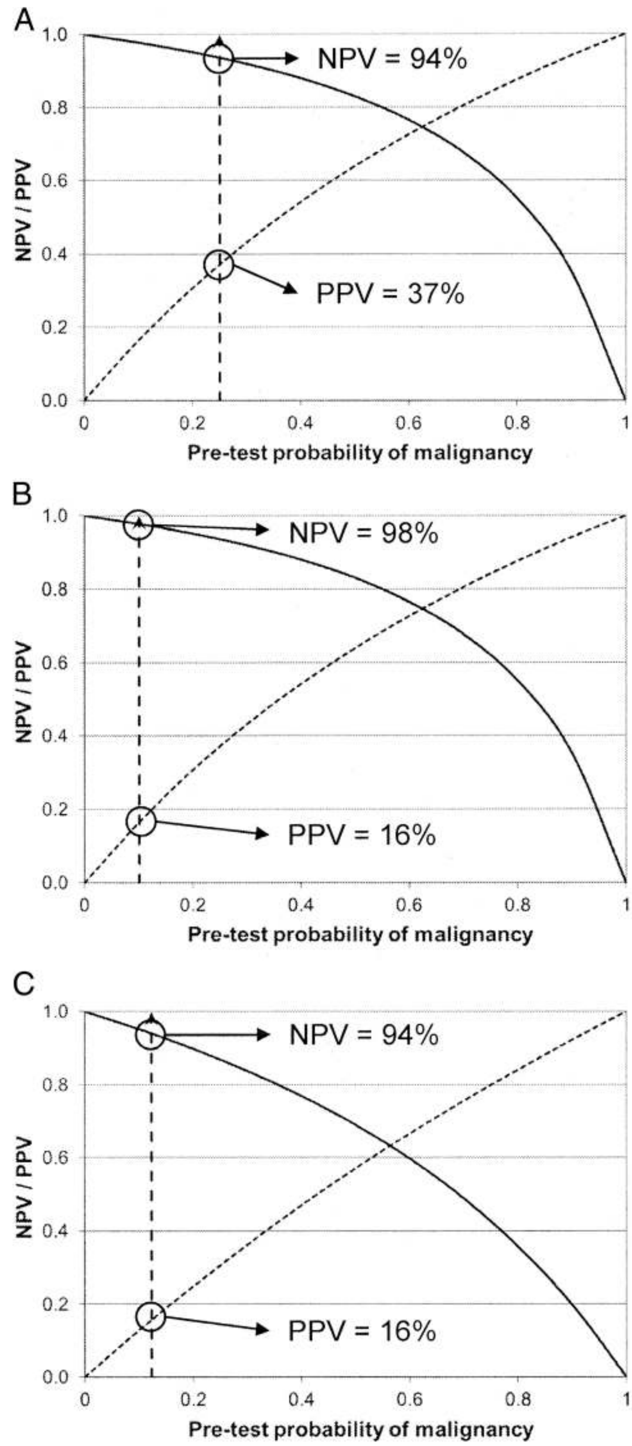

Figure 2. A, NPV and PPV for Afirma GEC for FN/HCN, as a function of the pretest probability of malignancy, based on sensitivity and specificity reported by Alexander et al (14). The Afirma GEC has a reported sensitivity of 90% and specificity of 48% in FN/HCN, which results in a PPV of 38% and NPV of 94%. B, At the same sensitivity and specificity, but with a lower pretest probability of malignancy, estimated at 10% in our study, the PPV falls to 16% whereas NPV is expected to increase to 98%. C, Our study data imply a specificity of 39% and sensitivity of 83% if there are no additional false-negatives among the Afirma GEC-benign nodules. This changes the shape of the curves relating PPV and NPV to pretest probability of malignancy changes, as shown. A sensitivity of 38%, specificity of 83%, and a 12% pretest probability of malignancy yields the observed PPV of 16% and an NPV of 94%.

the achieved NPV falls while the PPV rises. Consequently, the PPV and NPV achieved by the Afirma GEC assay depend critically on the pretest probability of malignancy, and indeed can be defined only when that risk of malignancy is known. An NPV of 94% and PPV of 37%, as reported by Alexander et al (11) in FN/HCN, are achieved with a pretest risk of malignancy of 25%. However, based on the same test sensitivity and specificity, with a pretest risk of malignancy of 12%, the GEC is predicted to offer a PPV of 17%, close to the PPV we have demonstrated in this study (Figure 2B). Under these circumstances, we would predict that the NPV in our study should be higher, at 98%, but caution that this assumes that the point estimates for sensitivity and specificity reported by Alexander et al (11) are accurate, an assumption that awaits independent confirmation and that is not entirely supported by our study. As a consequence of these uncertainties, our patients with benign GEC results who have not yet undergone surgery remain under close surveillance, pending the publication of additional confirmatory studies.

In summary, our experience of the Afirma GEC has demonstrated a lower than expected rate of benign Afirma GEC reports in AUS/FLUS and FN/HCN, increasing the number of tests needed to avoid one surgery from two to four and raising questions about the costs of widespread application of this assay. In addition, we found the PPV of a suspicious classifier result to be lower than previously reported (16 vs 38%), so that more than 80% of GEC-suspicious nodules proved to be benign at surgery. This disappointing result, however, is consistent with the performance of the classifier as reported by Alexander et al (11), when applied to a group of patients at low risk for malignancy, and it is a reminder that the performance of this or any other molecular test depends critically on the input cytopathology. Unless pretest probability of malignancy is known, claims of reproducible NPV and PPV are spurious and should be treated with caution. Additional confirmatory studies are necessary to assess the performance characteristics of the Afirma GEC before widespread adoption of this technology can be recommended.

Acknowledgments

The authors are grateful to Ms Desirae Howe-Clayton and Mr Russell Ward for their valuable contributions to sample collection and handling and management of the clinical database. We are also grateful to Ms Heidi Gautier, of Veracyte Inc, who developed the "Enabled Model" and assisted in writing the standard operating procedures for use of the Afirma GEC.

Address all correspondence and requests for reprints to: Bryan McIver, MBPhD, Chair, Department of Head and Neck, and Endocrine Oncology, Leader, Endocrine Tumor Program, Moffitt Cancer Center, 12902 Magnolia Drive, University of South Florida, Tampa, FL 33612. E-mail: Bryan.McIver@moffitt.org.

Disclosure Summary: B.M. has received speaker honoraria and scientific consulting fees from Veracyte Inc and Asuragen Inc. No other author has any relevant financial conflict of interest.

References

1. **Mazzaferri EL.** Management of a solitary thyroid nodule. *N Engl J Med*. 1993;328:553–559.
2. **American Cancer Society.** Cancer Facts & Figures 2012. http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-031941.pdf. Accessed July 2, 2013.
3. **Caruso DR, Mazzaferri EL.** Fine needle aspiration biopsy in the management of thyroid nodules. *Endocrinologist*. 1991;1:194–202.
4. **Yassa L, Cibas ES, Benson CB, et al.** Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer*. 2007;111:508–516.
5. **Cooper DS, Doherty GM, Haugen BR, et al.** Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19:1167–1214.
6. **Gharib H, Papini E, Paschke R, et al.** American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *J Endocrinol Invest*. 2010;33:1–50.
7. **Gharib H, Goellner JR.** Fine-needle aspiration biopsy of the thyroid: an appraisal. *Ann Intern Med*. 1993;118:282–289.
8. **Cibas ES, Ali SZ.** The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol*. 2009;132:658–665.
9. **Horne MJ, Chhieng DC, Theoharis C, et al.** Thyroid follicular lesion of undetermined significance: evaluation of the risk of malignancy using the two-tier sub-classification. *Diagn Cytopathol*. 2012;40:410–415.
10. **Chudova D, Wilde JI, Wang ET, et al.** Molecular classification of thyroid nodules using high-dimensionality genomic data. *J Clin Endocrinol Metab*. 2010;95:5296–5304.
11. **Alexander EK, Kennedy GC, Baloch ZW, et al.** Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med*. 2012;367:705–715.
12. **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.
13. **National Comprehensive Cancer Network.** Thyroid Carcinoma Version 2.2013. http://www.nccn.org/professionals/physician_gls/pdf/thyroid.pdf. Accessed July 15, 2013.
14. **Alexander EK, Schorr M, Klopper J, et al.** Multicenter clinical experience with the Afirma gene expression classifier. *J Clin Endocrinol Metab*. 2014;99(1):119–125.
15. **Harrell RM, Bimston DN.** Surgical utility of Afirma: effects of high cancer prevalence and oncocytic cell types in patients with indeterminate thyroid cytology. *Endocr Pract*. 2014;20:364–369.
16. **Mitsiades N, Fagin JA.** Molecular genetics of thyroid cancer: pathogenetic significance and clinical applications. In: Weiss RE, Refetoff S, eds. *Genetic Diagnosis of Endocrine Disorders*. Chap 11. Waltham, MA: Academic Press; 2010:117–138.

17. **Nikiforov YE, Steward DL, Robinson-Smith TM, et al.** Molecular testing for mutations in improving the fine-needle aspiration diagnosis of thyroid nodules. *J Clin Endocrinol Metab*. 2009;94:2092–2098.
18. **Nikiforov YE, Ohori NP, Hodak SP, et al.** Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1056 FNA samples. *J Clin Endocrinol Metab*. 2011;96:3390–3397.
19. **Ganly I, Ricarte Filho J, Eng S, et al.** Genomic dissection of Hurthle cell carcinoma reveals a unique class of thyroid malignancy. *J Clin Endocrinol Metab*. 2013;98:E962–E972.
20. **Li H, Robinson KA, Anton B, Saldanha IJ, Ladenson PW.** Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules. *J Clin Endocrinol Metab*. 2011;96:E1719–E1726.
21. **Castro MR, Espiritu RP, Bahn RS, et al.** Predictors of malignancy in patients with cytologically suspicious thyroid nodules. *Thyroid*. 2011;21:1191–1198.
22. **Wang CC, Friedman L, Kennedy GC, et al.** A large multicenter correlation study of thyroid nodule cytopathology and histopathology. *Thyroid*. 2011;21:243–251.