Validation and Multicenter Clinical Experience of the Afirma® Gene Expression Classifier

Trevor E Angell, MD[1] and Erik K Alexander, MD[2]

1. Associate Physician, Brigham and Women's Hospital; Instructor in Medicine, Harvard Medical School, US;
2. Chief, Thyroid Section, Brigham & Women's Hospital; Associate Professor of Medicine, Harvard Medical School, US

Abstract

The current assessment of clinically relevant thyroid nodules is imprecise and nonspecific, primarily due to the presence of indeterminate cytology. This cytologic category implies a low but clinically important risk for thyroid cancer, historically prompting consideration for diagnostic thyroidectomy in most patients. Over the last 5 years, novel molecular testing options have become available and appear able to modify preoperative thyroid cancer risk and better inform clinical decisions. The Afirma® gene expression classifier (GEC) measures messenger RNA (mRNA) expression patterns in fine-needle aspiration (FNA) tissue and seeks to identify a benign signature despite indeterminate cytology. When detected, the cancer risk is low and allows for prevention of unnecessary surgeries. The results of a large, multicenter, blinded validation trial of the Afirma GEC demonstrated the test's high negative predictive value. Subsequent independent studies of the Afirma GEC in clinical use have documented the influence these results have on preoperative management recommendations. This review summarizes the current data regarding the Afirma GEC, including the original validation trial and recent multicenter clinical experience.

Keywords

Thyroid nodule, molecular tests, Afirma® gene expression classifier

Thyroid nodules are a common clinical illness. Neck ultrasonography identifies at least one nodule in 20–76 % of adults, the majority of which are benign.[1–3] Cytopathologic evaluation by ultrasound-guided fine needle aspiration (FNA) has emerged as the principal means to evaluate malignancy risk, and is recommended for most nodules greater than 1–1.5 cm.[4–6] However, in 15–25 % of cases, FNA yields an 'indeterminate' result, which implies a 20–30 % risk for cancer.[7–9] The Bethesda system for reporting thyroid cytology aids evaluation by stratifying indeterminate findings into groups of escalating malignancy risk: atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) with 5–25 % cancer risk; follicular neoplasm (FN) with 15–30 % risk; and suspicious for malignancy (SUSP) with 60–75 % risk.[10] Ultrasound findings may modify risk further. Despite gradation of cancer risk, none of the indeterminate classes imply sufficiently low risk for malignancy to allow monitoring in lieu of surgery. Therefore, diagnostic surgical resection has been generally recommended,[4,11] although the majority of indeterminate nodule have proved benign. For these patients, surgery was unnecessary, yet carried inherent risk, morbidity, and cost.[12,13] Improved preoperative diagnostic tools have sought to address this matter.

The diagnostic use of molecular markers has proved valuable toward modifying cancer risk assessment and addressing the shortfalls of visual microscopy. Initially, the discovery of several oncogenic gene mutations associated with well-differentiated thyroid cancer led to their use as diagnostic molecular markers. When applied to cytologically indeterminate nodules, the identification of mutations/translocation in *BRAF*, *RAS*, *RET/PTC*, and *PAX8/PPARG* increased the predictive value for malignancy.[14] Initial retrospective, nonblinded analysis of this mutation panel suggested its utility as a 'rule in' test for thyroid cancer that could influence the decision to proceed with surgery or modify its extent.[5,15,16] More recent prospective, blinded data confirm the utility of detecting *BRAF* mutations, though question the performance and predictive value of the entire panel, and most specifically the detection of *RAS* mutations.[17] Further, confirmation of malignancy in indeterminate nodules already destined for surgery may not substantially change the clinical management of patients.

Since the majority of indeterminate nodules are ultimately benign, a different testing approach sought to identify molecular profiles associated with benignity. Such a test, if validated, effectively 'rules out' malignancy, and identifies patients in whom surgery can be avoided. This novel diagnostic paradigm was utilized in the design of the Afirma® gene expression classifier (GEC) (Veractye, Inc.). The Afirma GEC employs measurement of 167 expressed messenger RNAs (mRNAs) from fresh aspiration tissue to

classify indeterminate nodules as 'benign' or 'suspicious.'[18] Importantly, the microarray platform for this test demonstrated excellent reproducibility and stability in different temperatures and shipping environments.[19] Here, we discuss the multicenter validation trial and subsequent published experiences with the Afirma GEC.

Validation and Clinical Use of the Afirma Gene Expression Classifier

In 2012, the *New England Journal of Medicine* (NEJM) published the results of a prospective, multicenter, blinded validation of the Afirma GEC in 3,789 patients enrolled from 49 academic and community sites.[20] Of 265 indeterminate nodules greater than 1 cm for which surgery was performed and gold-standard histopathology available for blinded review, 85 (32 %) nodules were malignant and 180 benign. The Afirma GEC demonstrated a sensitivity of 92 % (95 % confidence interval [CI] 84–97 %) and specificity of 52 % (95 % CI 44–59 %). In a population harboring a malignancy prevalence of 32 %, the negative predictive value (NPV) and positive predictive value (PPV) were 93 % and 47 %, respectively. More specifically, when Bethesda categories were AUS/FLUS or FN, the NPV for the Afirma GEC was 95 % and 94 %, respectively—similar to that of a benign cytology diagnosis for which conservative management is often favored. The NPV was 85 % for SUSP nodules and, at present, the Afirma GEC is not recommended for routine use in this population.

In spite of the demonstrated accuracy of this testing approach to identify benign nodules preoperatively, it remained unclear how implementation of this knowledge could affect clinical practice and vary in a real-world environment.

To address this concern, a multicenter experience with the Afirma GEC was performed and published by Alexander and colleagues in the *Journal of Clinical Endocrinology and Metabolism* in 2014.[21] From five academic centers, 346 consecutive samples were obtained from patients who underwent FNA biopsy for an ultrasound-confirmed thyroid nodule greater than 1 cm, determined to be of indeterminate cytology by local evaluation and subsequently sent for Afirma GEC at the discretion of the treating physician. After exclusion of seven samples, 339 samples were evaluated. These included 165 (49 %) cases cytologically classified as AUS/FLUS, 161 (47 %) classified as FN, and 13 (4 %) classified as SUSP.

Of nodules with AUS/FLUS cytology, the Afirma GEC result was 'benign' in 91/165 (55 %), 'suspicious' in 66/165 (40 %), and 'nondiagnostic' in 8/165 (5%). Nodules categorized as FN were 'benign,' 'suspicious,' or 'nondiagnostic' in 79 (49 %), 73 (45 %), and 9 (6 %) of cases, respectively. These distributions paralleled those of the initial NEJM validation.

For patients with an indeterminate nodule whose Afirma GEC results were 'suspicious,' surgery was recommended in 141 of 148 (95 %) cases, and was completed in 121 (85 %). Surgical pathology confirmed malignancy in 53 of 121 (44 %) cases. By contrast, only four of 174 patients with a 'benign' Afirma GEC result were recommended for surgery, all of whom complied. Documented follow-up monitoring was available in 71 patients with 'benign' Afirma GEC results. Ultimately, surgery was performed in 11 of these 71 patients, with only one proving malignant (1.0 cm sonographic nodule, confirmed to be a 0.6 cm papillary carcinoma).

This study showed a dramatic change in clinical management with Afirma GEC assessment. Using an 'intention-to-treat' model, Afirma GEC results modified the care recommendations of 171 of 339 (50 %) patients, and reduced surgery by 47 %.

Discussion

Data now confirm that molecular testing has the ability to inform and influence the clinical management of cytologically indeterminate thyroid nodules. Results from two large multicenter studies using the Afirma GEC have demonstrated high NPV for thyroid malignancy and the ability to alter the surgical recommendations among affected patients. The National Cancer Cooperative Network now proposes ultrasound monitoring rather than surgery 'if molecular testing predicts a risk for malignancy comparable to the risk for malignancy seen with a benign FNA cytology (approximately 5 % or less).'[22] Given the very low post-test malignancy risk demonstrated in the Afirma GEC validation trial, it appears reasonable to employ a strategy of watchful waiting in patients with an indeterminate nodule and Afirma GEC 'benign' result when no other mitigating clinical or radiographic factors prompt surgical intervention.

The impact of this change in management has been evaluated in two cost-effectiveness analyses. Cost-effectiveness was demonstrated with statistical modeling by Li et al.,[23] whereas Duick et al.[24] made cost calculations based on the retrospective finding that only 7.6 % of patients with an Afirma GEC 'benign' result underwent surgery. Estimated cost savings are between $1,453 and $2,600 per patient tested, and are attributed to a reduction in diagnostic surgery, surgical complications, and the need for life-long thyroid hormone replacement.

Central to the interpretation of Afirma GEC performance is the relationship between the prevalence of thyroid cancer in the population studied (i.e. pretest probability), and the NPV and PPV of a test. As the prevalence of malignancy increases, the NPV falls since there is an increasing probability that a negative test result represents a false negative. As the malignancy rate declines, NPV improves as the test result is increasingly likely to be a true negative. Differences in the underlying rate of malignant nodules within a population and the specific malignancy risk within nodules classified as indeterminate will influence both NPV and PPVs. Prior studies have confirmed substantial inter- and intra-rater variability for indeterminate cytology, even among experts[25–27] using the Bethesda system. Thus, the interpretation of an Afirma GEC result must be performed in the context of the thyroid cancer prevalence within a given practice setting.

Separate retrospective experiences using the Afirma GEC in the clinical setting have recently been published. While these are not independent validation trials and are thus susceptible to selection bias and variation in cytologic/histologic assessment, these trials nonetheless show consistent performance of the Afirma GEC in varied practice populations. In a study by Harrell et al.,[28] of 58 cytologically indeterminate nodules, 20 were Afirma GEC 'benign.' Though two malignancies were ultimately detected in this group, both false negative cases were attributed to sampling error (and not assay failure). No malignancies were detected in the remaining nodules with 'benign' Afirma GEC results, suggesting a very high NPV consistent with the 95 % CIs of the initial validation study.[29] A separate study by McIver et al.,[30] evaluated 90 indeterminate nodules primarily composed of lesions categorized as follicular (or Hurthle cell) neoplasm (i.e. Bethesda class IV).

The authors reported one malignancy among 16 nodules (6 %) 'benign' on Afirma GEC testing, again consistent with the initial validation findings. The PPV in this study was calculated to be 16 % and the authors question if their data suggest that the PPV of the Afirma GEC may be lower than initially reported. However, the rate of histologically proved malignancy in GEC 'suspicious' nodules was only 17 %, in comparison to 56 % in the original validation trial. This suggests that marked variation in the use and meaning of Bethesda categorizations at the time of initial cytologic review is the most likely explanation for such differences.

Other potential markers have been evaluated for use in indeterminate thyroid nodules, with some yielded encouraging preliminary results. Assessment of the cell surface glycoprotein galectin-3 by immunohistochemistry in two multicenter studies has shown the potential for clinical utility.[31,32] In the most recent, Bartolazzi et al.[32] evaluated 465 nodules with indeterminate cytology, finding a PPV and NPV for galectin-3 expression of 82 % and 91 %, respectively. Evaluation of a 4-microRNA prolife differentially expressed in thyroid cancers compared with benign nodules showed a diagnostic accuracy as high as 90 % in the research setting,[33] but demonstration of diagnostic accuracy and utility in large, prospective, validation studies has not yet been performed.[34] The 17-mutation panel (provided by some academic institutions or commercially marketed as miRInform™ Thyroid by Asuragen, Inc.) has more recently demonstrated inferior performance compared with original reports, although next-generation sequencing for

a larger set of mutations has been proposed as a next step.[17,35] Several have reasoned that combining mutational testing with Afirma GEC would provide ideal NPV and PPV performance. However, Kloos et al.[36] demonstrated that the additional determination of *BRAF* mutational status did not improve test sensitivity or specificity when added to the Afirma GEC. This was primarily because all *BRAF* mutation positive specimens were already classified by Afirma GEC as 'suspicious'.

Conclusions

In a short amount of time, molecular testing of thyroid nodule aspirates has dramatically changed the management of indeterminate thyroid nodules. Though sonographic, cytologic, and clinical assessment must still be performed, their limitations and imprecisions have made molecular testing particularly attractive. Results thus far indicate that molecular testing with the Afirma GEC significantly effects clinical decision-making and limits unnecessary surgical intervention in many patients. The application of the Afirma GEC to more diverse practice settings, along with increased experience and longer patient follow-up, will continue to document its clinical impact. Given the unique strengths of different molecular testing methods, combination testing with different and multiple molecular markers warrants further study. Clinician judgment remains paramount, though molecular diagnostic tests have now become a key adjunct to the existing clinical, biochemical, radiologic, and cytologic assessments for most indeterminate thyroid nodules. ∎

1. Mazzaferri EL, Management of a solitary thyroid nodule, *N Engl J Med*, 1993;328:553–9.
2. Popoveniuc G, Jonklaas J, Thyroid nodules, *Med Clin North Am*, 2012;96:329–49.
3. Tan GH, Gharib H, Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging, *Ann Intern Med*, 1997;126:226–31.
4. American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, 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;19:1167–214.
5. Bernet V, Hupart KH, Parangi S, et al., AACE/ACE Disease State Commentary: Molecular diagnostic testing of thyroid nodules with indeterminate cytopathology, *Endocr Pract*, 2014;20:360–3.
6. Sosa JA, Hanna J, Lanman RB, et al., Increases in thyroid nodule fine need aspirations, surgeries and diagnoses of thyroid cancer in the Unites States. Presented at: 34th Annual Meeting of the American Association of Endocrine Surgeons, April 14–16, 2013, Chicago, IL.
7. Baloch ZW, Livolsi VA, Follicular-patterned lesions of the thyroid: the bane of the pathologist, *Am J Clin Pathol*, 2002;117:143–50.
8. Nayar R, Ivanovic M, The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference, *Cancer*, 2009;117:195–202.
9. Bongiovanni M, Spitale A, Faquin WC, et al., The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis, *Acta Cytol*, 2012;56:333–9.
10. Cibas ES, Ali SZ, The Bethesda System for Reporting Thyroid Cytopathology, *Thyroid*, 2009;19:1159–65.
11. Gharib H, Papini E, Paschke R, et al., AACE/AME/ETA Task Force on Thyroid Nodules. 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: executive summary of recommendations, *J Endocrinol Invest*, 2010;33:51–6.
12. Sosa JA, Bowman HM, Tielsch JM, et al., The importance of surgeon experience for clinical and economic outcomes from thyroidectomy, *Ann Surg*, 1998;228:320–30.
13. Bergenfelz A, Jansson S, Kristoffersson A, et al., Complications to thyroid surgery: results as reported in a database from a multicenter audit comprising 3,660 patients, *Langenbecks Arch Surg*, 2008;393:667–73.
14. 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;94:2092–8.
15. Nikiforov YE, Yip L, Nikiforova MN. New strategies in diagnosing cancer in thyroid nodules: impact of molecular markers, *Clin Cancer Res*, 2013;19:2283–8.
16. 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;96:3390–7.
17. Beaudenon-Huibregtse S, Alexander EK, Guttler RB, et al., Centralized molecular testing for oncogenic gene mutations complements the local cytopathological diagnosis of thyroid nodules, *Thyroid*, 2014;24:1479–87.
18. Chudova D, Wilde JI, Wang ET, et al., Molecular classification of thyroid nodules using high-dimensionality genomic data, *J Clin Endocrinol Metab*, 2010;95:5296–304.
19. Walsh PS, Wilde JI, Tom EY, et al., Analytical performance verification of a molecular diagnostic for cytology-indeterminate thyroid nodules, *J Clin Endocrinol Metab*, 2012;97:E2297–306.
20. Alexander EK, Kennedy GC, Baloch ZW, et al., Preoperative diagnosis of benign thyroid nodules with indeterminate cytology, *N Engl J Med*, 2012;367:705–15.
21. Alexander EK, Schorr M, Klopper J, et al., Multicenter clinical experience with the Afirma gene expression classifier, *J Clin Endocrinol Metab*, 2014;99:119–25.
22. National Comprehensive Cancer Network. Thyroid Carcinoma Version 2.2013. Available at: http://www.nccn.org/professionals/ physician_gls/pdf/thyroid.pdf (accessed June 1, 2014).
23. Li H, Robinson KA, Anton B, et al., Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules, *J Clin Endocrinol Metab*, 2011;96:E1719–26.
24. 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.
25. Kocjan G, Chandra A, Cross PA, et al., The interobserver reproducibility of thyroid fine-needle aspiration using the UK Royal College of Pathologists' classification system, *Am J Clin Pathol*, 2011;135:852–9.
26. Gerhard R, da Cunha Santos G. Inter- and intraobserver reproducibility of thyroid fine needle aspiration cytology: an analysis of discrepant cases, *Cytopathology*, 2007;18:105–11.
27. Cibas ES, Baloch ZW, Fellegara G, et al., A prospective assessment defining the limitations of thyroid nodule pathologic evaluation, *Ann Intern Med*, 2013;159:325–32.
28. 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;20:364–9.
29. McIver B. Evaluating thyroid nodules in the information age: new technologies, new opportunities, new challenges, *Endocr Pract*, 2014;20:370–3.
30. McIver B, Castro MR, Morris JC, et al., An independent study of a gene expression classifier (Afirma™) in the evaluation of cytologically indeterminate thyroid nodules, *J Clin Endocrinol Metab*, 2014 [Epub ahead of print].
31. Bartolazzi A, Gasbarri A, Papotti M, et al., Thyroid Cancer Study Group. Application of an immunodiagnostic method for improving preoperative diagnosis of nodular thyroid lesions, *Lancet*, 2001;357:1644–50.
32. Bartolazzi A, Orlandi F, Saggiorato E, et al.; Italian Thyroid Cancer Study Group (ITCSG), Galectin-3-expression analysis in the surgical selection of follicular thyroid nodules with indeterminate fine-needle aspiration cytology: a prospective multicentre study, *Lancet Oncol*, 2008;9:543–9.
33. Keutgen XM, Filicori F, Crowley MJ, et al., A panel of four miRNAs accurately differentiates malignant from benign indeterminate thyroid lesions on fine needle aspiration, *Clin Cancer Res*, 2012;18:2032–8.
34. Kim MI, Alexander EK. Diagnostic use of molecular markers in the evaluation of thyroid nodules, *Endocr Pract*, 2012;18:796–802.
35. Nikiforova MN, Wald AI, Roy S, et al., Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer, *J Clin Endocrinol Metab*, 2013;98:1852–60.
36. Kloos RT, Reynolds JD, Walsh PS, et al., Does addition of BRAF V600E mutation testing modify sensitivity or specificity of the Afirma Gene Expression Classifier in cytologically indeterminate thyroid nodules?, *J Clin Endocrinol Metab*, 2013;98:E761–8.