Title: Genetic Testing- Molecular Markers in Fine Needle Aspirates (FNA) of the Thyroid

Description/Background

Fine needle aspiration (FNA) of a thyroid lesion to identify which patients need to undergo surgery has diagnostic limitations and has led to the development of molecular markers in an attempt to improve the accuracy of patient selection.

FNA of the thyroid

Thyroid nodules are common, present in 5% to 7% of the U.S. adult population. Most are benign, and most cases of thyroid cancer are curable by surgery when detected early. FNA of the thyroid is currently the most accurate procedure to distinguish benign thyroid lesions and malignant ones, reducing the rate of unnecessary thyroid surgery for patients with benign nodules and triaging patients with thyroid cancer to appropriate surgery.

About 60% to 70% of thyroid nodules are classified cytologically as benign, and 4% to 10% of nodules are cytologically deemed malignant. (1) However, the remaining 20% to 30% have equivocal findings (inclusive, indeterminate, atypical, suspicious), usually due to overlapping cytologic features between benign and malignant nodules; these nodules usually require surgery for a final diagnosis. Thyroid FNA cytology is classified by according to Bethesda System criteria into the following groups: nondiagnostic; benign; follicular lesion (FLUS) or atypia of undetermined significance (AUS); follicular neoplasm (or suspicious for follicular neoplasm); suspicious for malignancy; and malignant. Lesions with FNA cytology in the FLUS/AUS or follicular neoplasm category are often considered indeterminate.

There is some individualization of management for patients with FNA-indeterminate nodules, but many patients will ultimately require a surgical biopsy, typically thyroid lobectomy, with intraoperative pathology consultation would typically be the next step in diagnosis. Approximately 80% of patients with indeterminate cytology undergo surgical resection; postoperative evaluation reveals a malignancy rate ranging from 6% to 30%, making this a
clinical process with very low specificity. (2) Thus, if an analysis of FNA samples could reliably identify the risk of malignancy as low, there is potential for patients to avoid surgical biopsy.

Preoperative planning of optimal surgical management in patients with equivocal cytologic results is challenging, as different thyroid malignancies may require different surgical procedures (e.g., unilateral lobectomy versus total or subtotal thyroidectomy with or without lymph node dissection) depending on several factors, including histologic subtype and risk-stratification strategies (tumor size, patient age, etc.) If a diagnosis cannot be made intraoperatively, a lobectomy is typically performed, and if on postoperative histology the lesion is malignant, a second surgical intervention may be necessary for completion thyroidectomy.

**Thyroid cancer**

Most thyroid cancers originate from thyroid follicular cells and include well-differentiated papillary thyroid carcinoma (PTC) (80% of all thyroid cancers) and follicular carcinoma (15%). Poorly differentiated and anaplastic thyroid carcinomas are uncommon and can arise de novo or from preexisting well-differentiated papillary or follicular carcinomas. Medullary thyroid carcinoma originates from parafollicular or C cells and accounts for about 3% of all thyroid cancers.

The diagnosis of malignancy in the case of PTC is primarily based on cytologic features. If an FNA in a case of PTC is indeterminate, intraoperative consultation is most often diagnostic, although its efficacy and therefore use will vary between institutions, surgeons, and pathologists.

For follicular carcinoma, the presence of invasion of the tumor capsule or of blood vessels is diagnostic and cannot be determined by cytology, as tissue sampling is necessary to observe these histologic characteristics. Intraoperative diagnosis of follicular carcinoma is challenging and often not feasible, because extensive sampling of the tumor and capsule is usually necessary and performed on postoperative permanent sections.

New approaches for improving the diagnostic accuracy of thyroid FNA include mutation analysis for somatic genetic alterations, to more accurately classify which patients need to proceed to surgery (and may include the extent of surgery necessary) and a gene expression classifier to identify patients who do not need surgery and can be safely followed.

**Mutations associated with thyroid cancer**

Various mutations have been discovered in thyroid cancer. The 4 gene mutations that are the most common and carry the highest impact on tumor diagnosis and prognosis are *BRAF* and *RAS* point mutations and *RET/PTC* and *PAX8/PPARγ* rearrangements.

Papillary carcinomas carry point mutations of the *BRAF* and *RAS* genes, as well as *RET/PTC* and *TRK* rearrangements, all of which are able to activate the mitogen-activated protein kinase (MAPK) pathway. These mutually exclusive mutations are found in more than 70% of papillary carcinomas. *BRAF* mutations are highly specific for PTC. Follicular carcinomas harbor either *RAS* mutations or *PAX8/PPARγ* rearrangement. These mutations are also mutually exclusive and identified in 70% to 75% of follicular carcinomas. Genetic alterations involving the *PI3K/AKT* signaling pathway also occur in thyroid tumors, although they are rare in well-differentiated thyroid cancer and have higher prevalence in less differentiated thyroid carcinomas. (3) Additional mutations known to occur in poorly differentiated and anaplastic
carcinomas involve the \textit{TP53} and \textit{CTNNB1} genes. Medullary carcinomas, which can be familial or sporadic, frequently possess point mutations located in the \textit{RET} gene.

\textbf{Available Molecular Diagnostic Testing}

\textbf{Mutation/Rearrangement Testing}
Point mutations in specific genes, including \textit{BRAF}, \textit{RAS}, and \textit{RET}, and evaluation for rearrangements associated with thyroid cancers can be accomplished by gene sequencing with Sanger sequencing or pyrosequencing or by real-time polymerase chain reaction (rtPCR). Panels of tests for mutations associated with thyroid cancer are also available. For example, Quest Diagnostics offers a Thyroid Cancer Mutation Panel, which includes \textit{BRAF} and \textit{RAS} mutation analysis and testing for \textit{RET/PTC} and \textit{PAX8/PPARγ} rearrangements.

The ThyroSeq® v.2 Next Generation Sequencing panel (CBLPath, Ocala, FL) is a NGS sequencing panel of more than 60 genes. According to the ThyroSeq’s manufacturer’s website, the test is indicated when FNA cytology indicates atypia of uncertain significance or follicular lesion of undetermined significance, follicular neoplasm or suspicious for follicular neoplasm, or suspicious for malignancy. (7) In particular, it has been evaluated in patients with follicular neoplasm/suspicious for follicular neoplasm on FNA as a test to increase both sensitivity and specificity for cancer diagnosis.

The ThyGenX™ Thyroid Oncogene Panel (formerly miRInform® Thyroid; Interpace Diagnostics, Parsippany, NJ; testing done at Asuragen Clinical Laboratory) is another NGS sequencing panel designed to be used in patients with indeterminate thyroid FNA results. It includes sequencing of eight genes associated with papillary thyroid carcinoma and follicular carcinomas.

\textbf{Gene Expression Profiling}
Genetic alterations associated with thyroid cancer can be assessed through the use of gene expression profiling, which refers to analysis of messenger RNA (mRNA) expression levels of many genes simultaneously. Several gene expression profiling tests are now available to biologically stratify tissue from thyroid nodules.

The Afirma® Gene Expression Classifier (Afirma GEC; Veracyte, South San Francisco, CA) analyzes the expression of 142 different genes to determine patterns associated with benign findings on surgical biopsy. It is designed to be used for thyroid nodules that have an “indeterminate” classification on FNA as a method to select patients who are at low risk for cancer (“rule out”).

Veracyte also markets 2 “malignancy classifiers” that use mRNA expression-based classification to evaluate for \textit{BRAF} mutations or mutations associated with medullary thyroid carcinoma (Afirma BRAF and Afirma MTC, respectively). In a description of the generation of the Afirma BRAF test, the authors outline the following proposed benefits of the mRNA-based expression test for \textit{BRAF} mutations:
(1) PCR based methods may have low sensitivity, requiring that a large proportion of the nodule have a relevant mutation;
(2) testing for only 1 mutation may not detect patients with low-frequency mutations that result in the same pattern of pathway activation; and
(3) PCR-based approaches with high analytic sensitivity may require a large of amount of DNA that is difficult to isolate from small FNA samples. (6) The Afirma MTC is an option when the Afirma GEC is ordered for thyroid nodules with an “intermediate” classification on FNA, and can also be used for thyroid nodules with “malignant” or “suspicious” results on Afirma GEC. The Afirma BRAF is designed to be used for nodules with “suspicious” results on Afirma GEC.

ThyraMIR™ (Interpace Diagnostics, Parsippany, NJ) is a microRNA expression based classifier that is intended for use in thyroid nodules with indeterminate cytology on FNA.

Other gene expression profiles have been reported in investigational settings, but have not been widely validated or in commercial use (e.g., Barros-Filho et al [2015], (10) Zheng et al [2015] (9); these are not addressed in this review.

**Regulatory Status**

Testing for mutations associated with thyroid cancer via sequencing or rtPCR are laboratory-developed tests (LDTs). Clinical laboratories may develop and validate tests in-house and market them as a laboratory service; LDTs must meet the general regulatory standards of the Clinical Improvement Act (CLIA). Laboratories that offer LDTs must be licensed by CLIA for high-complexity testing.

In 2013, the U.S. Food and Drug Administration (FDA) approved through the premarket approval process the THxID™-BRAF kit, which is an in vitro diagnostic device to assess specific BRAF mutations in melanoma tissue via rtPCR. However, there are currently no diagnostic tests for thyroid cancer mutation analysis with approval from FDA.

**Medical Policy Statement**

Mutation analysis in fine needle aspirates of the thyroid is considered to be experimental/investigational.

The use of the Afirma gene expression classifier in fine needle aspirates of the thyroid that are cytologically considered to be indeterminate (follicular lesion of undetermined significance or follicular neoplasm) may be established in patients who have the following characteristics:

- Thyroid nodules without strong clinical or radiological findings suggestive of malignancy
- In whom surgical decision-making would be affected by test results

Gene expression classifiers in fine needle aspirates of the thyroid not meeting above outlined criteria are considered experimental/investigational.

**Inclusionary and Exclusionary Guidelines** (Clinically based guidelines that may support individual consideration and pre-authorization decisions)

N/A
CPT/HCPCS Level II Codes (Note: The inclusion of a code in this list is not a guarantee of coverage. Please refer to the medical policy statement to determine the status of a given procedure.)

Established codes:
81545

Other codes (investigational, not medically necessary, etc.):
N/A

Rationale

Molecular Markers to Predict Benignancy

Clinical Context and Proposed Clinical Utility
Molecular markers to predict benignancy are tests designed to have a high negative predictive value (NPV). The focus of this section is the Afirma Gene Expression Classifier (GEC), which is proposed as a risk stratifying test for patients who have indeterminate findings on fine needle aspirate (FNA). These patients presently proceed to surgical resection. The purpose of the test is to select patients at low risk of malignancy who could avoid unnecessary surgery.

Analytic Validity
Walsh et al verified the analytic performance of the Afirma gene expression classifier (GEC) in the classification of cytologically indeterminate (FNAs from thyroid nodules. (10) The analytic performance studies were designed to characterize the stability of the RNA in the aspirates during collection and shipment, analytical sensitivity and specificity, and assay performance studies including intranodule, intra-assay, interassay, and interlaboratory reproducibility. The authors concluded that the analytic sensitivity and specificity, robustness and quality control of the GEC were successfully verified.

Clinical Validity
Chudova et al developed a molecular test to distinguish between benign and malignant thyroid nodules using FNAs. (2) The authors used mRNA analysis to measure more than 247,000 transcripts in 315 thyroid nodules. The data set consisted of 178 retrospective surgical specimens, representing the most common benign and malignant histologic subtypes, and 137 prospectively collected aspirate specimens. Two classifiers were trained separately on surgical samples and aspirates. The performance was evaluated using an independent test set of 48 prospective FNA samples, which had known surgical pathology diagnoses, and included 50% with indeterminate cytopathology. The performance of the classifier was markedly lower in the FNAs than in tissue, likely due to differences in cellular heterogeneity between the 2 types of specimens. On the test set, NPV and specificity were estimated to be 96% and 84%, respectively.

Prospective Clinical Validation
Alexander et al reported on a 19-month, prospective, multicenter validation study of the Afirma GEC, which involved 49 clinical sites (both academic and community centers), 3789 patients and 4812 FNAs from thyroid nodules that were at least 1 cm in size. (11) Local pathology reports of the cytologic diagnosis were collected for all patients, and reports without a definitive benign or malignant diagnosis at the local site were reviewed by 3 expert cytopathologists, who reclassified them as atypical, follicular neoplasm or suspicious for a follicular neoplasm, or suspicious for malignancy. Of all nodules screened, 577 (12%) were considered indeterminate after central review, and 413 of those had tissue pathology available.

The GEC used in the Alexander et al study was retrained on a set of 468 samples, comprised of 220 banked tissue samples, 14 ex vivo operative FNA samples, and 234 prospective clinical FNA samples. The authors noted that 25 of those prospective clinical FNA samples were derived from the 413 samples described above.

After exclusion of the 25 used for test validation and those that did not have a valid GEC result, 265 FNA samples were evaluated with the Afirma GEC. Of the 265 samples, 85 were malignant; the GEC correctly identified 78 of the 85 as suspicious (92% sensitivity; 95% CI, 84% to 97%). Specificity was 52% (95% CI, 44%-59%). NPV ranged from 85% for “suspicious cytologic findings” to 95% for “atypia of undetermined clinical significance.” There were seven FNAs with false-negative results, six of which were thought to be due to hypocellular aspirate specimens.

Retrospective Clinical Validation
In 2014, Alexander et al reported results from a retrospective analysis of 339 thyroid nodules which underwent Afirma GEC testing for indeterminate cytology on FNA (follicular lesion of undetermined significance/atypia of undetermined significance, follicular neoplasm, or suspicious for malignancy) at five academic medical centers. (12) Most of the nodules sent for GEC testing were follicular lesions of undetermined significance/atypia of undetermined significance or follicular neoplasm. The distribution of GEC testing results for each cytologic classification is shown in Table 1.

A subset of patients whose nodules underwent GEC testing had a subsequent thyroid resection. Among 148 cases with suspicious Afirma GEC findings, surgery (thyroid resection) was recommended for 141 (95%). For the 174 cases with benign Afirma GEC findings, surgery was recommended for 4 (2%; p<0.01). Using the assumption that, absent the GEC results, thyroid surgery would be recommended for patients with cytologically indeterminate FNA results, the authors reported that the GEC results altered management in 50% of patients. Table 2 shows thyroidectomy biopsy results for the subset of patients shown in Table 1 who underwent surgery.
Seventeen patients who had indeterminate cytology, benign Afirma GEC results, and did not undergo surgery had follow-up beyond 1 year. Of those, 3 patients underwent surgical removal of the nodule because of compressive symptoms (n=2) or nodule growth (n=1); all nodules were benign on final histology. The remaining 14 patients had ongoing follow-up with ultrasound with no ongoing evidence of malignancy. The study demonstrated site-to-site variation in the proportion of samples that were GEC benign. A benign GEC result did not completely rule out malignant pathology. Long-term follow-up was available for only a small proportion of patients with benign GEC findings who did not undergo surgery.

In 2016, Santhanam et al reported results of a meta-analysis of studies reporting on the performance of the Afirma GEC in cytologically indeterminate nodules. (13) Seven studies met the inclusion criteria, which required that studies reported on the use of the Afirma GEC in nodules that were indeterminate on FNA (including atypia of undetermined significance or follicular lesion of undetermined significance; suspicious for follicular/Hürthle cell neoplasm; suspicious for malignancy), and thyroidectomy was performed as a reference standard in at least the cases where the index test was suspicious. All studies were judged to be at low risk of bias for patient selection, and most for GEC test selection, whereas the risk of bias in the final histopathology was low in 3 studies, unclear in 3 studies, and high in 1 study. In the pooled cohort, the prevalence of malignancy was 37.1%. The main results of the analysis are summarized in Table 3.

Retrospective single-center studies, including Harrell and Bimston (2014), (14) Lastra et al (2014), (15) McIver et al (2014), (16) Yang et al (2016) (17) have reported the diagnostic accuracy of the Afirma GEC, are summarized in Table 4. These studies are all subject to ascertainment bias, because a large proportion of individuals with Afirma benign reports did not undergo surgery, which makes determining the sensitivity and specificity of the GEC assay impossible. However, the rates of malignancy among patients with Afirma benign results who did undergo surgery are consistently low. One exception is the study by Harrell and Bimston (2014); it may be reflective of a higher-than-usual overall rate of malignancy in patients with indeterminate FNA results. One additional publication (Celik et al, 2015) reported on Afirma GEC testing, but included in its sample population individuals with benign and suspicious cytology on FNA, which is not the targeted use of the test. (18)
There are limited data on the true negative rate of individuals with indeterminate FNA cytology and Afirma GEC benign results. Supportive information on the accuracy Afirma GEC benign results can be obtained from studies that report on long-term follow-up of individuals with indeterminate FNA cytology and Afirma GEC benign results. Angell et al reported on a retrospective analysis comparing clinical outcomes for individuals with indeterminate FNA cytology and Afirma GEC benign results with individuals with cytologically benign nodules. (20) A total of 95 cytologically indeterminate/Afirma GEC benign nodules in 90 patients were compared with 1224 cytologically benign nodules identified from a single center, prospectively collected database. Five nodules in the cytologically indeterminate were resected; of the remaining 90 nodules, 58 (64.4%) had follow-up ultrasound available at a median of 13 months postdiagnosis. When nodule growth was defined by a volume increase of 50% or more, 17.2% cytologically indeterminate/Afirma GEC benign were considered to have grown, compared with 13.8% of cytologically benign nodules (p=0.44). Surgical resection was more common in cytologically indeterminate/Afirma GEC benign nodules (13.8% vs. 0.9%, p<0.001).

Clinical Utility
No evidence directly demonstrating improved outcomes in patients managed with the Afirma GEC was identified. Therefore, a chain of indirect evidence was developed, which addresses 2 key questions:

1. Does the use of the Afirma GEC in individuals with cytologically indeterminate thyroid nodules lead to changes in management (in this case, reduced thyroid resections)?

The clinical setting in which the Afirma GEC is meant to be used is well-defined: individuals with atypia of undetermined significance/follicular lesion of undetermined significance or follicular neoplasm/suspicious for follicular neoplasm on FNA, who do not have other
indications for thyroid resection (i.e., in whom the GEC results would play a role in surgical decision making).

Decision impact studies, most often reporting on clinical management changes but not on outcomes after surgical decisions were made, suggest that in at least some cases, surgical decision-making is changed. These studies are described briefly.

Duick et al reported on the impact of Afirma GEC test results on physician and patient decision making to operate on thyroid nodules with indeterminate cytology and Afirma GEC benign results in a sample of 395 nodules from 368 patients. (21) Data were collected on 368 patients with 395 nodules. Surgery was performed in 7.6% of the patients with indeterminate cytology and a benign GEC. The authors compared this surgical excision rate of the study population (7.6%) with a historical rate of surgical excision of 74% previously reported for patients with an indeterminate cytologic diagnosis (but no GEC test).

The 2014 study by Alexander et al provides some evidence about clinical management changes for patients with indeterminate thyroid nodules with the use of the Afirma GEC. While the treating physicians presumably elected to obtain the GEC testing with the intent of altering management recommendations, the magnitude of the difference in surgical recommendations for patients with GEC suspicious or benign results was large.

Two studies, Aragon Han et al [2014], (22) Noureldine et al [2015] (23) were identified that evaluated the potential for the Afirma GEC to change surgical decision making by comparing actual surgical decision making when the Afirma GEC was used to predict surgical decision making based on a management algorithm. In both, surgical decision making was estimated to change in at least some proportion of patients (approximately 10%-15%).

2. Do those management changes improve outcomes?

A simplified decision model was developed for use with the Afirma GEC in individuals with cytologically indeterminate FNA samples. It is shown in Appendix 2. It is assumed that when the Afirma GEC is not used, patients with cytologically indeterminate FNA results undergo thyroid resection. When the Afirma GEC is used, those with Afirma suspicious lesions undergo resection, while those who have Afirma benign lesions do not. In this case, compared to the standard care plan, some patients without cancer will have avoided a biopsy, which is weighed against the small increase in missed cancers in patients who had cancer but tested as Afirma benign.

Assuming that the rate of cancer in cytologically indeterminate thyroid nodules is approximately 20%, (24) in the standard care plan, 80% of patients with cytologically indeterminate FNA samples will undergo an unnecessary biopsy. Applying the test characteristic values from Alexander et al (2012), (11) it is estimated that approximately 1.6% of individuals with a true cancer would be missed, but approximately 38%, instead of 80%, would undergo unneeded surgery.

Whether the tradeoff between avoiding unneeded surgeries and the potential for missed cancer is worthwhile depends, in part, on patient and physician preferences. However, some general statements may be made by considering the consequences of a missed malignancy and the consequences of unnecessary surgery. Most missed malignancies will be papillary
thyroid carcinomas (PTCs), which have an indolent course. Thyroid nodules are amenable to ongoing surveillance (clinical, ultrasound, and with repeat FNAs), with minimal morbidity.

Thyroid resection is a relatively low-risk surgery. However, consequences of surgery can be profound. Patients who undergo a hemi- or subtotal thyroidectomy have a risk of recurrent laryngeal nerve damage and parathyroid gland loss.

At present, the existing standard of care for thyroid nodules is based on intervention that is stratified by FNA cytology results, which are grouped into categories with differing prognosis. Avoiding an invasive surgery in situations where patients are at very low likelihood of having an invasive tumor is likely beneficial, given the small but potentially significant adverse effects associated with thyroidectomy or hemithyroidectomy. The alternative to surgical biopsy in the low-risk population is ongoing active surveillance.

Section Summary: Molecular Markers to Predict Benignancy

In one multicenter validation study, the Afirma GEC has been reported to have a high NPV (range, 90%-95%). These results are supported by an earlier development and clinical validation study (Chudova et al), but the classifiers used in the two studies do not appear to be identical. In an additional multicenter and multiple single-center studies, there is suggestive evidence that rates of malignancy are low in Afirma benign patients, but the exact NPV is unknown. The available evidence suggests that physician decision making about surgery is altered by GEC results, although long-term follow-up of patients with thyroid nodules who avoided surgery based on GEC results is limited. An indirect chain of evidence can be constructed to establish the potential for clinical utility with GEC testing in cytologically indeterminate lesions, but with only 1 study with the marketed test reporting a true NPV, the clinical validity is uncertain.

Molecular Markers to Predict Malignancy

Clinical Context

Molecular markers associated with malignancy in thyroid nodules are generally used as “rule-in” tests to identify cancer or tumors with more aggressive behavior. The purpose of the test is in patients with cytologically indeterminate FNA results when knowing the presence of a certain mutation or having a high enough pretest probability of cancer would change the surgical approach or some other aspect of management.

For thyroid nodules that have indeterminate findings on FNA cytology, a surgical biopsy with intraoperative pathology consultation would typically be the next step. Following a diagnosis of a thyroid malignancy, preoperative surgical planning with regard to the extent of thyroid resection and lymph node dissection is an important consideration. Conventional factors determining biopsy strategy and surgical resection strategy include histologic subtype and risk stratification based on factors such as tumor size and patient age.

Point mutation/rearrangement testing is discussed separately from GECs to predict malignancy in the discussion of their analytic and clinical validity, because the development and validation of the GECs is unique to each specific marketed test. However, the proposed clinical utility is similar.
Point Mutation/Rearrangement Testing

Analytic Validity
Point mutations in specific genes associated with thyroid cancer, such as the *BRAF* V600E gene, and the detection of genetic rearrangements associated with thyroid cancer, such as the *RET/PTC* rearrangement, are typically detected with Sanger sequencing or next-generation sequencing (NGS) methods. In the case of mutation testing for genes associated with thyroid cancer malignancy, analytic validity refers to a test’s technical accuracy in detecting a mutation that is present or in excluding a mutation that is absent. The rtPCR-based methods are generally considered to have high accuracy. For example, Smith et al reported technical performance characteristics for *BRAF* mutation detection by qualitative PCR in thyroid FNA samples with high within- and between-run reproducibility. (25)

Next-generation sequencing (NGS) is expected to have high accuracy for detecting a mutation that is present. However, with increasing numbers of tested mutations, there is increased risk of detection of variants of uncertain significance (VOUS). The VOUS rate for currently available NGS panels for thyroid cancer is not well-characterized. Nikiforova et al described the development and validation of a multi-gene NGS panel for thyroid cancer, the ThyroSeq panel. (26) The authors developed a custom library of gene sequence variants based on mutations previously reported in the literature. The assay demonstrated 100% accuracy in evaluating samples of 15 thyroid tumors and 3 cell lines with known genetic alterations and 15 DNA samples with no mutations. In analysis of 229 DNA samples from frozen tissues, formalin-fixed, paraffin-embedded tissues, and FNAs (n=105, 72, and 52, respectively), the panel identified mutations in 19 of 27 (70%) of classic papillary thyroid carcinomas (PTCs), 25 of 30 (83%) follicular variant PTCs, 14 of 18 (78%) conventional and 7/18 (39%) Hürthle cell carcinomas, 3 of 10 (30%) poorly differentiated carcinomas, 20 of 27 (74%) anaplastic thyroid carcinomas, and 11 of 15 (73%) medullary thyroid carcinomas. Of 83 benign nodules, 5 (6%) were positive for mutations.

Clinical Validity
A number of studies have evaluated whether testing for point mutations or gene fusions (either single mutation or panels of mutations) can be used to improve the sensitivity and specificity for diagnosing indeterminate FNA of the thyroid, with the goal of identifying mutations that predict malignancy in FNA samples.

Mutations Association with Malignancy
In 2015, Fnais et al reported on a systematic review and meta-analysis of studies reporting on the test accuracy of *BRAF* mutation testing in the diagnosis of PTC. (27) The review included 47 studies with 9924 FNA samples. For all cytologically indeterminate nodules, the pooled sensitivity estimate for *BRAF* mutation testing was 31% (95% CI, 6% to 56%). Among nodules suspicious for malignancy on FNA, the pooled sensitivity estimate for *BRAF* mutation testing was 52% (95% CI, 39% to 64%; I²=77%).

Ferraz et al evaluated 20 publications that reported on the type and number of mutations in cases of FNA of the thyroid diagnosed as indeterminate and compared the results with final histology after surgical resection. (28) Sixteen studies analyzed one mutation (e.g., *BRAF* or *RET/PTC*) and 4 studies analyzed a panel of several mutations (*BRAF*, *RAS*, *RET/PTC*, *PAX8/PPARγ*). The detection of a mutation in a histologically (surgically resected) benign thyroid lesion was categorized as a false positive case, detecting no mutation in an FNA
sample from a histologically benign surgical sample was considered a true negative, and finding no mutation in a histologically malignant lesion was categorized as a false negative. Based on 4 studies that examined a panel of mutations, there was a broad sensitivity range of 38% to 85.7% (mean 63.7%), a mean specificity of 98% (range, 95%-100%), mean false positive rate of 1.25% (0%-4%) and mean false negative rate of 9% (1%-21%). Based on 2 studies that examined RET/PTC rearrangements, mean sensitivity was 55% (50%-60%), specificity 100%, false positive rate of 0% and mean false negative rate 3.5% (91%-6%). Based on 3 studies that examined BRAF mutations, mean sensitivity was 13% (0%-37.5%), mean specificity, 92.3% (75%-100%), mean false positive rate, 0.5% (0%-1%) and mean false negative rate of 6% (3%-12%). The authors concluded that testing for a panel of mutations leads to an improvement in the sensitivity and specificity for indeterminate FNA of the thyroid but that further standardizations and further molecular markers are needed before broad application of molecular FNA cytology for the diagnosis of thyroid nodules.

The largest body of literature on mutation testing for prediction of malignancy in indeterminate thyroid nodules is related to the development an evaluation of a NGS panel that includes BRAF, RAS, RET/PTC, or PAX8/PPARγ, the ThyroSeq. Studies that address these panels are described in more length; studies that include subsets of these mutations or additional mutations are summarized in the following section.

Nikiforov et al prospectively tested a panel of mutations (BRAF, RAS, RET/PTC, PAX8/PPARγ) in 470 FNA samples of thyroid nodules from 328 consecutive patients. (29) Mutational status was correlated with cytology and either surgical pathology diagnosis or follow-up (mean 34 months). A total of 40 patients were excluded for poor quality of specimen or loss to follow-up (mean, 34 months). A total of 40 patients were excluded for poor quality of specimen or loss to follow-up. Sixty-nine patients (with 86 thyroid FNA samples) underwent surgery soon after completion of the cytologic evaluation; preoperative cytologic diagnosis was: positive for malignancy in 22 samples, indeterminate (including atypical and suspicious for malignancy) in 52 samples, and negative for malignancy in 12 samples. By FNA, 32 mutations were found (18 BRAF, 8 RAS, 5 RET/PTC, 1 PAX8/PPARγ); after surgery, 31 mutation-positive nodules (97%) were diagnosed as malignant on pathologic examination, and 1 was a benign tumor (3%). Thirteen of the 32 mutation-positive FNA samples had a definitive cytologic diagnosis of malignancy, whereas the rest were either indeterminate or negative for malignancy.

Of the remaining 219 patients, 147 (229 FNAs) who did not undergo surgery were followed using serial ultrasound with no change in the nodule status (124 patients) or using repeated FNA with cytology negative for malignancy (23 patients) and no mutation found in the FNA material. These nodules were considered negative for malignancy. The remaining 72 patients who were initially in the follow-up group underwent subsequent surgery. Combining all 3 groups, the specificity for malignancy was high (99.7%), but the sensitivity of the molecular test alone was not (62%).

Ohori et al performed mutation screening in 117 FNA samples classified as a follicular lesion of indeterminate significance/atypia of indeterminate significance. (30) BRAF, RAS, RET/PTC, or PAX8/PPARγ mutations were detected in 10% of this category. They demonstrated that the probability of having a malignancy in this cytology category together with a detection of one of the somatic mutations investigated was 100%, whereas the probability of having a thyroid malignancy without a mutation detected was 7.6%.
In 2011, Nikiforov et al reported results of a prospective study to assess the clinical utility of a panel of mutations to predict the likelihood of malignancy in thyroid nodules that were indeterminate on FNA. (31) The authors included 1056 consecutive FNA samples with indeterminate cytology on FNA that underwent mutation testing, with 967 of those adequate for molecular analysis (653 follicular lesion of undetermined significance/atypia of undetermined significance; 247 follicular or Hurthle cell neoplasm or suspicious for follicular neoplasm; 67 suspicious for malignant cells). One hundred seventeen of the samples were included in the Ohori et al analysis previously described. Eighty-seven BRAF, RAS, RET/PTC, or PAX8/PPARγ mutations were detected. At the time of analysis, 479 patients had undergone thyroidectomy for further evaluation, providing a histopathologic diagnosis for 513 samples. The presence of a mutation had low sensitivity for predicting malignant histology (63%, 57%, 68% for samples with follicular lesion of undetermined significance/atypia of undetermined significance, follicular or Hurthle cell neoplasm/suspicious for follicular neoplasm, and suspicious for malignant cells on cytology, respectively), but high specificity (99%, 97%, and 96%, respectively). The NPV for the mutation analysis results was 94%, 86%, and 72% for samples with follicular lesion of undetermined significance/atypia of undetermined significance, follicular or Hurthle cell neoplasm/suspicious for follicular neoplasm, and suspicious for malignant cells on cytology, respectively. The authors conclude that mutation analysis may be useful in surgical planning, such as determining whether patients should undergo a thyroid lobectomy or a total thyroidectomy as a first surgery.

In a subsequent study, Nikiforov et al evaluated the accuracy of an NGS panel that included tests for point mutations in 13 genes and for 42 types of gene fusions (ThyroSeq v.2 NGS panel) in a series of 143 consecutive thyroid FNA samples with a cytologic diagnosis of follicular or Hürthle cell neoplasm/suspicious for follicular or Hürthle cell neoplasm. (32) Molecular testing was retrospectively performed for 91 samples and prospectively performed for the remaining 52. The prevalence of cancer on histology was 27.5% and 26.9% in the retrospective and prospective cohorts, respectively. In the retrospective cohort, of the 25 malignant nodules, 22 were PTCs, and 3 were follicular thyroid carcinomas (FTCs). In the prospective cohort, of the 14 malignant nodules, 11 were PTCs and 3 were FTCs. The performance of the ThyroSeq in both cohorts is shown in Table 5.

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>Retrospective (n=91)</td>
</tr>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Prospective (n=52)</td>
</tr>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Overall (N=143)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
</tr>
<tr>
<td>PPV (95% CI)</td>
</tr>
<tr>
<td>NPV (95% CI)</td>
</tr>
<tr>
<td>CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.</td>
</tr>
</tbody>
</table>

The authors noted that, compared with the panel of mutations used in their 2011 study, the NGS panel was associated with marked increase in NPV, with a similar PPV. In this case, the authors proposed that the panel could be used to both “rule in” and “rule out” invasive cancers.
The same group (Nikiforov et al) reported the performance of a subsequent generation ThyroSeq panel (ThyroSeq v2.1) with an expanded gene panel in a series of 465 thyroid FNA samples with a diagnosis of atypia of undetermined significance/follicular lesion of undetermined significance. (33) Molecular analysis was performed prospectively in all patients. Ninety patients (96 nodules) eventually underwent thyroid surgery, based on either patient preference, the presence of another nodule with a diagnosis of suspicious for malignancy or malignant on FNA, or positive molecular testing. An additional 2 patients were considered to have a definitive nonsurgical diagnosis of primary hyperparathyroidism based on biochemical testing. The performance of the test in patients with a known clinical outcome is summarized in Table 5.

In addition to studies that describe the clinical validity of the mutations that comprise the ThyroSeq panel, studies have reported on the diagnostic performance of individual mutations and combinations of mutations to predict malignancy in thyroid nodules that are indeterminate on FNA. The results that pertain to the use of mutation testing in indeterminate thyroid nodules are summarized in Table 6. (In some cases, measures of agreement were calculated from data provided in the manuscript.)

Table 6. Studies of Clinical Validity of Molecular Markers to Predict Malignancy in Indeterminate Thyroid FNA Samples

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Population</th>
<th>Genes Tested</th>
<th>Insufficient /Inadequate for Analysis</th>
<th>Measures of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moses et al (2010)</td>
<td>110 indeterminate thyroid nodules</td>
<td>BRAF, KRAS, NRAS, RET/PTC1, RET/PTC3, NTRK1</td>
<td>2</td>
<td>Sen 80  Spec 38  PPV 95  NPV 67  ACC 79 77</td>
</tr>
<tr>
<td>Ohori et al (2010)</td>
<td>100 patients with 117 follicular lesions of undetermined significance/atypia of undetermined significance</td>
<td>BRAF, NRAS, HRAS, KRAS, RET/PTC1, RET/PTC3, PAX8/PPARγ</td>
<td>NR</td>
<td>60 100 100 92 93</td>
</tr>
<tr>
<td>Cantara et al (2010)</td>
<td>41 indeterminate and 54 suspicious thyroid nodules</td>
<td>BRAF, H-K-NRAS, RET/PTC, TRK, PAX8-PPARγ</td>
<td>53</td>
<td>86 80a  97a  86a  97a  95a</td>
</tr>
<tr>
<td>Xing et al (2004)</td>
<td>25 indeterminate, dominant nodules</td>
<td>BRAF</td>
<td>NR</td>
<td>14 100 100 48 52</td>
</tr>
<tr>
<td>Rossi et al (2015)</td>
<td>140 indeterminate or suspicious for malignancy or malignant nodules</td>
<td>BRAF</td>
<td>NR</td>
<td>90 50d  100d 100d 93d 96c</td>
</tr>
<tr>
<td>Beaudenon-Huibregtset et al (2014)</td>
<td>53 nodules with indeterminate/ nondiagnostic FNA</td>
<td>BRAF, HRAS, KRAS, NRAS, PAX8-PPARγ, RET-PTC1, RET-PTC3</td>
<td>48</td>
<td>89 81 64</td>
</tr>
</tbody>
</table>
Additional studies report on differences in mutation frequency in malignant versus benign tumors, and may report on the sensitivity and specificity of mutation testing in unselected populations (i.e., all patients with nodules, rather than just those with indeterminate cytology). These studies are summarized next.

Mathur et al collected thyroid FNA samples, thyroid tissue, clinical and histopathology data, and tumor genotyping for mutations \(\text{BRAF V600E, NRAS, KRAS, RET/PTC1, RET/PTC3, and NTRK1}\) for 341 patients with 423 dominant thyroid nodules. (40) A cytologic examination of the samples showed that 51% were benign (one quarter of these were surgically resected), 21% were malignant, 11% were atypical lesions, 12% were follicular or Hurthle cell neoplasms, and 4% were suspicious for malignancy. On final analysis, 165 nodules were benign and 123 were malignant. Of the 423 FNA samples, 24 \(\text{BRAF V600E}\) mutations, 7 \(\text{KRAS}\), 21 \(\text{NRAS}\), 4 \(\text{PAX8-PPARγ}\) rearrangements, 3 \(\text{RET/PTC1}\), and 2 \(\text{RET/PTC3}\) rearrangements were detected. In all, 17 of 165 (10.3%) benign thyroid nodules had a mutation compared with 26% (32 of 123) malignant tumors (p<0.05).

Eszlinger et al retrospectively analyzed a panel of mutations (\(\text{BRAF}\) and \(\text{RAS}\) point mutations and \(\text{PAX8-PPARγ}\) and \(\text{RET/PTC}\) rearrangements) in a sample of 310 thyroid air-dried FNA specimens with available corresponding formalin-fixed, paraffin embedded thyroid biopsy samples (164 indeterminate, 57 malignant, and 89 benign on FNA). (41) A total of 47 mutations were detected on FNA: 22 \(\text{BRAF}\) mutations, 13 \(\text{NRAS}\) mutations, 3 \(\text{HRAS}\) mutations, 8 \(\text{PAX8-PPARγ}\) rearrangements, and 1 \(\text{RET/PTC}\) rearrangement. The addition of mutation analysis to cytology results was associated with a sensitivity and specificity of 75.3% and 90.4% for the detection of malignancy, respectively, with a PPV and NPV of 77.2% and 89.4%, respectively. The presence of a \(\text{BRAF}\) or \(\text{RET/PTC}\) mutation was associated with cancer in 100% of samples.

The association between \(\text{BRAF}\) mutations and PTC is supported by a report by Park et al (2015) on 294 patients with thyroid nodules whose FNA samples were evaluated with \(\text{BRAF}\) mutation testing by two methods, real-time PCR with Taq-Man minor groove-binding probes and allele-specific PCR using dual-priming oligonucleotides. (42) The detection rate of PTC by \(\text{BRAF}\) mutation testing by real-time PCR and allele-specific PCR was 80.2% (95% CI, 71.9% to 86.9%) and 76.9% (95% CI, 68.3% to 84.0%), respectively.

**Mutations Association with Tumor Behavior**

As reported in the studies previously described, the presence of \(\text{BRAF}\) mutations is strongly associated with malignancy in thyroid nodule FNA samples. \(\text{BRAF}\) mutations have also been associated with more aggressive clinicopathologic features in individuals who are diagnosed with PTC.

Adeniran et al conducted a study of 157 cases with equivocal thyroid FNA readings (indeterminate and suspicious for papillary thyroid carcinoma [PTC]) or a positive diagnosis for
PTC and concomitant BRAF mutation analysis. (1) The results of histopathologic follow-up were correlated with the cytologic interpretations and BRAF status. Based on the follow-up diagnosis after surgical resection, the sensitivity for diagnosing PTC was 63.3% with cytology alone and 80.0% with the combination of cytology and BRAF testing. No false positives were noted with either cytology or BRAF mutation analysis. All PTCs with extrathyroidal extension or aggressive histologic features were positive for BRAF mutation. The authors concluded that patients with an equivocal cytologic diagnosis and BRAF V600E mutation could be candidates for total thyroidectomy and central lymph node dissection.

Xing et al investigated the utility of BRAF mutation testing of thyroid FNA specimens for preoperative risk stratification of PTC in 190 patients. (36) A BRAF mutation in preoperative FNA specimens was associated with poorer clinicopathologic outcomes of PTC. In comparison with the wild-type allele, a BRAF mutation strongly predicted extrathyroidal extension (23% vs. 11%; p=0.039), thyroid capsular invasion (29% vs. 16%; p=0.045), and lymph node metastasis (38% vs. 18%; p=0.002). During a median follow-up of 3 years (range, 0.6-10 years), PTC persistence/recurrence was seen in 36% of BRAF mutation-positive patients versus 12% of BRAF mutation-negative patients, with an odds ratio of 4.16 (95% confidence interval [CI], 1.70 to 10.17; p=0.002). The positive and NPVs for preoperative FNA-detected BRAF mutation to predict PTC persistence/recurrence were 36% and 88%, respectively, for all histologic subtypes of PTC. The authors concluded that preoperative BRAF mutation testing of FNA specimens may provide a novel tool to preoperatively identify PTC patients at higher risk for extensive disease (extrathyroidal extension and lymph node metastases) and those who are more likely to manifest disease persistence/recurrence.

Gene Expression Classifiers to Predict Malignancy

Analytic Validity
In 2015, Diggans et al described the development and validation Afirma BRAF malignancy classifier. (8) The study included FNA biopsies from 716 thyroid nodules. Biopsies were evaluated with quantitative PCR (qPCR) for the BRAF V600E gene, with 181 used as a training sample and 535 used as a validation sample. The Afirma BRAF malignancy classifier was generated using robust multichip average-normalized gene expression summaries, and the classifiers were evaluated for positive percent agreement (PPA) and negative percent agreement (NPA) with the PCR-derived gene classification. The highest scoring classification method and gene set were then used in a final round of model building. The maximum PPA and NPA for all cytology categories were observed when the threshold for BRAF-positive status was 5% or more BRAF mutations. At 5% analytic sensitivity, Afirma BRAF demonstrated a PPA with PCR results of 90.4% (95% exact binomial CI, 83.5% to 95.1%) and an NPA of 99.0% (95% CI, 97.6% to 99.7%). Two samples in the training set and 4 samples in the validation set that were Afirma BRAF-positive but negative (0% mutation) on PCR, which the authors attributed to either technical variability in either assay or to mutations other than the BRAF V600E mutation that cause similar gene expression changes.

Intra- and interrun reproducibility of the classifier were evaluated using 9 FNA biopsies (FNABs) and 3 tissue controls selected from among training samples with high (BRAF-positive) or low (BRAF-negative) classifier scores and scores near the classifier decision boundary. Each FNAB and tissue was processed from total RNA in triplicate in each of 3 different runs across days, operators, and reagent lots. The intraassay standard deviation (SD) of Afirma BRAF scores was 0.171 (95% CI, 0.146 to 0.204). Of the 106 Afirma BRAF calls
produced (2 arrays failed quality control requirements), 106 resulted in concordant calls across all 3 runs (100% concordance). The interassay SD of scores was 0.204 (95% CI, 0.178 to 0.237) for scores measured on a 6-point scale. These results suggest low intra- and interrun variability.

In 2016, Kloos et al described the development of the Afirma MTC classifier in a study that also described the clinical validity of the MTC classifier. (43)

**Clinical Validity**

Less evidence exists on the validity of gene expression profiling (specifically, the Afirma BRAF and Afirma MTC tests). Mutations can be used to improve the sensitivity and specificity for diagnosing indeterminate FNA of the thyroid, with the goal of identifying mutations that predict malignancy in FNA samples.

In the Diggans et al study, describing the development and validation of the Afirma BRAF test (previously described), for a subset of 213 thyroid nodule FNA samples for which histopathology was available, Afirma BRAF test results were compared with pathologic findings. (8) The Afirma BRAF classified all histopathologically benign samples as *BRAF* V600E-negative (specificity, 100%; 95% CI, 97.4% to 100%). Of the 73 histopathologically malignant samples, the Afirma BRAF test identified 32 as *BRAF*-positive (sensitivity, 43.8%; 95% CI, 32.2% to 55.9%).

In the Kloos study describing the development and validation of the Afirma MTC classifier, the MTC classifier was evaluated in a sample of 10,488 thyroid nodule FNA samples referred for GEC testing (the Afirma GEC described below). (43) In this sample, 43 cases were Afirma MTC-positive, of which 42 were considered to be clinically consistent with medullary thyroid carcinoma on pathology or biochemical testing, for a PPV of 97.7% (95% CI, 86.2% to 99.9%).

Labourier et al reported on the sensitivity and specificity of a test algorithm combining micro-RNA measurements from 17 genes (miRInforme: Asuragen Laboratory, Austin, TX) with a 10-gene GEC in 109 FNA samples with atypia of undetermined significance/follicular lesion of undetermined significance or follicular neoplasm/suspicious for follicular neoplasm on cytology evaluated at the Asuragen Laboratory with known final pathology. (44) Seventy-four nodules were diagnosed as benign and 35 as malignant. The Performance of the combined test (micro-RNA measurements and the 10-gene GEC) is summarized in Table 7.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>All Patients</th>
<th>Atypia of Undetermined Significance/Follicular Lesion of Undetermined Significance</th>
<th>Follicular Neoplasm/Suspicious for Follicular Neoplasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>109</td>
<td>59</td>
<td>51</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>89% (73% to 97%)</td>
<td>94% (73% to 100%)</td>
<td>82% (57% to 96%)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>85% (75% to 92%)</td>
<td>80% (64% to 91%)</td>
<td>91% (76% to 98%)</td>
</tr>
<tr>
<td>PPV (95% CI)</td>
<td>74% (58% to 86%)</td>
<td>68% (46% to 85%)</td>
<td>82% (57% to 96%)</td>
</tr>
<tr>
<td>NPV (95% CI)</td>
<td>94% (85% to 98%)</td>
<td>97% (84% to 100%)</td>
<td>91% (76% to 98%)</td>
</tr>
</tbody>
</table>

CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

**Clinical Utility**

Testing for specific mutations associated with thyroid cancer (e.g., *BRAF* V600E mutations, *RET* mutations, and *RET*/PTC and *PAX8/PPARγ* rearrangements) are generally designed to
“rule in” cancer in nodules that have indeterminate cytology on FNA. (45) (Of note, some mutation panels, such as the ThyroSeq panel, may have a high enough NPV that their clinical use could also be considered as a molecular marker to predict benignancy; see next section). A potential area for clinical utility for this type of mutation testing would be in informing preoperative planning for thyroid surgery following initial thyroid FNA, such as planning for a hemi- versus a total thyroidectomy or performance of a central neck dissection.

In a retrospective analysis, Yip et al reported outcomes after implementation of an algorithm incorporating molecular testing of thyroid FNA samples to guide the extent initial thyroid resection. (46) The study included a cohort of patients treated at a single academic center at which molecular testing (BRAF V600E, BRAF K601E, NRAS codon 61, HRAS codon 61, and KRAS codon 12 and 13 point mutations; RET/PTC1, RET/PTC3, and PAX8/PPARγ rearrangements) was prospectively obtained for all FNAs with indeterminate cytology (follicular lesion of undetermined significance, follicular neoplasm, and suspicious for malignancy), and for selective FNAs at the request of the managing physician for selected nodules with either benign or nondiagnostic cytology. The study also included a second cohort of patients who did not have molecular testing results available. For the patients treated with molecular diagnosis, a positive molecular diagnostic test was considered to be an indication for an initial total thyroidectomy. Patients with follicular lesion of undetermined significance and negative molecular diagnostic results were followed with repeat FNA, followed by a lobectomy or total thyroidectomy if indeterminate pathology persisted. Patients with follicular neoplasm or suspicious for malignancy results on cytology and a negative molecular diagnostic result were managed with lobectomy or total thyroidectomy.

The sample included 671 patients, 322 and 349 managed with and without molecular diagnostics, respectively. Positive molecular testing results were obtained in 56 patients (17% of those managed with molecular diagnostics), most commonly RAS mutations (42/56 [75%]), followed by BRAF V600E (10/56 [18%]), BRAF K601E (2/56 [4%]), and PAX8/PPARγ rearrangements (2/56 [4%]). Compared with those managed without molecular diagnostics, patients managed with molecular diagnostics were nonsignificantly less likely to undergo total thyroidectomy as an initial procedure (63% vs. 69%, p=0.08). However, they had nonsignificantly higher rates of central compartment lymph node dissection (21% vs. 15%, p=0.06). Across both cohorts, 25% of patients (170/671) were found to have clinically significant thyroid cancer, with no difference in thyroid cancer rates based on the type of initial operation (26% for total thyroidectomy vs. 22% for lobectomy, p=0.3). The incidence of clinically significant thyroid cancer after initial lobectomy (i.e., requiring a 2-stage surgery) was significantly lower for patients managed with molecular diagnostics (17% vs. 43%, p<0.001). An indeterminate FNA result had sensitivity and specificity for the diagnostic of thyroid cancer of 89% and 27%, respectively, with PPV and NPV of 29% and 88%, respectively. The addition of molecular diagnostics to FNA results increased the specificity for a cancer diagnosis to 95% and the PPV to 82%.

In 2015, a task force from the American Thyroid Association (ATA) reported on a review with recommendations for the surgical management of FNA-indeterminate nodules with various molecular genetic tests. (47) This review reported on the estimated likelihood of malignancy in an FNA-indeterminate nodule depending on results of the Afirma GEC (described above) and other panels designed to rule in malignancy. Depending on the estimated prebiopsy likelihood of malignancy, recommendations for surgery include observation, active surveillance, repeat FNA, diagnostic lobectomy, or oncologic thyroidectomy.
Section Summary: Molecular Markers to Predict Malignancy
The available evidence suggests that the use of mutation testing in thyroid FNA samples is generally associated with a high specificity and PPV for clinically significant thyroid cancer. The most direct evidence related to the clinical utility of mutation testing for genes associated with malignancy in thyroid cancer comes from one single-center retrospective study, which reported surgical decisions and pathology findings in patients managed with and without molecular diagnostics. There is potential clinical utility for identifying malignancy with higher certainty on FNA if such testing allows better preoperative planning at the time of thyroid biopsy, potentially avoiding the need for a separate surgery. An ATA statement provides some guidelines for surgeons managing patients with indeterminate nodules. However, the adoption of these guidelines in practice and the outcomes associated with them are uncertain.

SUMMARY OF EVIDENCE
For individuals with thyroid nodule(s) and indeterminate findings on fine needle aspiration (FNA) who receive FNA sample testing with the Afirma Gene Expression Classifier (GEC), the evidence includes 1 prospective clinical validity study with the marketed test, and an indirect chain to support clinical utility. Relevant outcomes are disease-specific survival, test accuracy and validity, morbidity events, and resource utilization. In 1 multicenter validation study, the Afirma GEC has been reported to have a high negative predictive value (NPV; range, 90%-95%). These results are supported by an earlier development and clinical validation study (Chudova et al), but the classifiers used in the 2 studies do not appear to be identical. In an additional multicenter and multiple single center studies, there is suggestive evidence that rates of malignancy are low in Afirma benign patients, but the exact NPV is unknown. The available evidence suggests that physician decision making about surgery is altered by GEC results, although long-term follow-up of patients with thyroid nodules who avoided surgery based on GEC results is limited. An indirect chain of evidence can be constructed to establish the potential for clinical utility with GEC testing in cytologically indeterminate lesions, but with only 1 study with the marketed test reporting a true NPV, the clinical validity is uncertain. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive FNA sample testing with molecular markers to predict malignancy, the evidence includes prospective and retrospective studies of clinical validity. Relevant outcomes are disease-specific survival, test accuracy and validity, morbidity events, and resource utilization. Mutation analysis has the potential to improve the accuracy of an equivocal FNA of the thyroid and may play a role in preoperative risk stratification and surgical planning. Single-center studies suggest that testing for a panel of mutations associated with thyroid cancer may allow the appropriate selection of patients for surgical management with an initial total thyroidectomy. Prospective studies in additional populations are needed to validate these results. Mutation analysis does not achieve a high enough NPV to identify which patients can undergo watchful waiting over thyroid surgery. Although the presence of certain mutations may predict more aggressive malignancies, the management changes that would occur as a result of identifying higher risk tumors are not well-established. The evidence is insufficient to determine the effects of the technology on health outcomes.

Ongoing and Unpublished Clinical Trials
Some currently unpublished trials that might influence this policy are listed in Table 8.
SUPPLEMENTAL INFORMATION

Clinical Input Received Through Physician Specialty Society and Academic Medical Center
While the various physician specialty societies and academic medical centers may collaborate with and make recommendations during this process, through the provision of appropriate reviewers, input received does not represent an endorsement or position statement by the physician specialty societies or academic medical centers, unless otherwise noted.

2016 Input
In response to requests, input was received from two physician specialty societies, 1 of which provided 3 responses, and 1 academic medical center while this policy was under review in 2016. Input focused on the use of gene expression classifiers (GEC) designed to with a high negative predictive (NPV) value in nodules indeterminate on fine needle aspirate (FNA). Although individual uses of GEC with NPV in these situations varied, there was general agreement that these tests are considered standard in the evaluation of some indeterminate cases of FNA.

2013 Input
In response to requests, input was received from one physician specialty society (4 reviewers) and 6 academic medical centers, for a total of 10 reviewers, while this policy was under review in 2013. There was general agreement with the policy statements that mutation analysis and use of the gene expression classifier is investigational. Input was mixed as to whether either test changes patient management and whether prospective randomized trials are necessary to establish the clinical utility of these tests.

SUPPLEMENTAL INFORMATION

Practice Guidelines and Position Statements

American Thyroid Association
In 2015, the American Thyroid Association (ATA) issued updated guidelines on the management of thyroid nodules and differentiated thyroid cancer in adults. (48) These guidelines make the following statements on molecular diagnostics in thyroid nodules that are atypia of undetermined significance (AUS)/follicular lesion of undetermined significance (FLUS) on cytology:

“For nodules with AUS/FLUS cytology, after consideration of worrisome clinical and sonographic features, investigations such as repeat FNA or molecular testing may be used to supplement...
malignancy risk assessment in lieu of proceeding directly with a strategy of either surveillance or diagnostic surgery. Informed patient preference and feasibility should be considered in clinical decision-making.” (Weak recommendation, Moderate-quality evidence).

“If repeat FNA cytology, molecular testing, or both are not performed or inconclusive, either surveillance or diagnostic surgical excision may be performed for an AUS/FLUS thyroid nodule, depending on clinical risk factors, sonographic pattern, and patient preference.” (Strong recommendation, Low-quality evidence).

The guidelines make the following statements on molecular diagnostics in thyroid nodules that are follicular neoplasm (FN)/suspicious for follicular neoplasm (SFN) on cytology:

“Diagnostic surgical excision is the long-established standard of care for the management of FN/SFN cytology nodules. However, after consideration of clinical and sonographic features, molecular testing may be used to supplement malignancy risk assessment data in lieu of proceeding directly with surgery. Informed patient preference and feasibility should be considered in clinical decision-making.” (Weak recommendation, Moderate-quality evidence).

The guidelines also state: “there is currently no single optimal molecular test that can definitively rule in or rule out malignancy in all cases of indeterminate cytology, and long-term outcome data proving clinical utility are needed.”

National Comprehensive Cancer Network

National Comprehensive Cancer Network (NCCN) guidelines on the treatment of thyroid cancer (v2.2015) make the following comments about the use of molecular diagnostics in thyroid cancer (49):

For thyroid nodules evaluated with FNA, molecular diagnostics may be employed in the following cases (category 2B recommendation):

- Follicular or Hürthle cell neoplasm
- Atypia of undetermined significance/Follicular lesion of undetermined significance

The guidelines state: “Molecular testing (both the Gene Expression Classifier and individual mutation analysis) was available in the majority of NCCN Member Institutions (>75%). About 70% of the panelists would recommend using a gene expression classifier in the evaluation of follicular lesions.”

Government Regulations

National/Local:

Palmetto GBA determines coverage and reimbursement for laboratories that perform molecular diagnostic testing and submit claims to Medicare in Medicare Jurisdiction E (California, Nevada, and Hawaii). Palmetto GBA’s decisions apply for all molecular diagnostic tests for Medicare.
Palmetto GBA has completed an assessment of the Afirma GEC and determined that the test meets criteria for analytic and clinical validity, and clinical utility as a reasonable and necessary Medicare benefit. (50) Effective January 1, 2012, Palmetto GBA will reimburse Afirma services for patients with the following conditions:

- Patients with one or more thyroid nodules with a history or characteristics suggesting malignancy such as:
  - Nodule growth over time
  - Family history of thyroid cancer
  - Hoarseness, difficulty swallowing or breathing
  - History of exposure to ionizing radiation
  - Hard nodule compared with rest of gland consistency
  - Presence of cervical adenopathy

OR

- Patients who have an indeterminate follicular pathology on fine needle aspiration

Medicare will cover the Afirma GEC test for all Medicare members meeting inclusionary guidelines, no matter which jurisdiction the Medicare member is are enrolled in.

MolDX expects this test will be performed once per patient lifetime. Should the unlikely situation of a second, unrelated thyroid nodule with indeterminate pathology occur, coverage may be considered upon appeal with support documentation.

**Michigan Department of Health and Human Services:**
**Michigan Medicaid Provider Manual, section 5.5, Genetic and Molecular Testing: Updated October 2016**

Whenever possible, Michigan Medicaid follows Medicare guidelines. Medicare does not cover a genetic test for a clinically affected individual for purposes of medical research, family planning, disease risk assessment of other family members or when the treatment and surveillance of the beneficiary will not be affected, or in any other circumstance that does not directly affect the diagnosis or treatment of the beneficiary.

Genetic testing is considered a covered benefit when it is medically necessary to establish a molecular diagnosis and treatment of a genetic disease and all of the following are met:

- The testing must be ordered by a physician (MD or DO) who is an enrolled provider.
- The beneficiary has documented clinical features symptomatic of a condition or disease, or is at risk of inheriting the disease based upon personal history, family history, documentation of a genetic mutation and/or ethnic background.
- Following history, physical examination, pedigree analysis, and completion of conventional diagnostic testing, a definitive diagnosis remains uncertain and a genetic diagnosis is suspected.
- The test results will be used to significantly alter the management or treatment of the disease.
- If applicable, the testing method is an FDA-approved method for the identification of a specific genetically-linked inheritable disease as evidenced by the following measures:
  - The genotypes to be detected by a genetic test must be shown, by scientifically valid methods, to be associated with the occurrence of the disease;
  - The analytical and clinical validity of the test must be established;
− The observations must be independently replicated and subject to peer review; and
− The clinical testing laboratory must be an enrolled provider who is properly certified by CLIA.
Testing is allowed once during the beneficiary’s lifetime per disease for diagnostic purposes. If medically necessary, and on a case-by-case basis, prior authorization may be requested to allow for exceptions to this restriction.

Providers must follow state law (Public Act 368 of 1978, Section 333.17020 Genetic test; informed consent) regarding informed consent for predictive genetic testing. This includes any statutory requirements for pre- or post-testing genetic counseling. There must be made available, upon request, documentation of pre-testing informed consent provided before testing. This documentation must include the limitations of the test, possible outcomes, and methods for communicating and maintaining confidentiality of results.

Genetic testing is not considered a covered benefit for:

- Criteria other than those outlined above.
- Testing to confirm a diagnosis or disorder that can be diagnosed by conventional diagnostic methods.
- Testing for conditions or purposes where the test results would not directly influence the management or treatment of the disease or condition (e.g., a disease without known treatment).
- Testing for informational purposes or management of a beneficiary’s family member.
- Confirmatory testing for validation of laboratory results.
- Screening for investigational or research purposes.
- Minors under the age of 18 for adult onset conditions that have no preventative or therapeutic treatments.
- Testing that has not been performed in a CLIA-certified laboratory.

The sole purpose of family planning counseling and infertility services.

5.5.B. PRIOR AUTHORIZATION REQUIREMENTS AND DOCUMENTATION For genetic testing that requires prior authorization, the following documentation must be submitted prior to the testing being performed:

- Indication for the test.
- Clinical notes that clearly detail the beneficiary’s related signs and symptoms, including relevant family history. A family pedigree analysis must be made available upon request.
- Other related testing or clinical findings of the beneficiary or family member.
- Documentation supporting that the test results will be used to significantly alter the management or treatment of the disease.
- The name and NPI number of the laboratory performing the test. (added per bulletin MSA 13-42)

(The above Medicare information is current as of the review date for this policy. However, the coverage issues and policies maintained by the Centers for Medicare & Medicare Services [CMS, formerly HCFA] are updated and/or revised periodically. Therefore, the most current CMS information may not be contained in this document. For the most current information, the reader should contact an official Medicare source.)
Related Policies

- Genetic Cancer Susceptibility Panels Using Next Generation Sequencing
- Genetic Testing to Determine the Prognosis of Breast Cancer Patients
- Genetic Testing: Mass Spectrometry Based Proteomic Profiling to Determine Treatment for Non Small-Cell Lung Cancer (NSCLC), e.g., VeriStrat®

References


44. Labourier E, Shifrin A, Busseniers AE, et al. Molecular testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with
indeterminate cytology. J Clin Endocrinol Metab. Jul 2015;100(7):2743-2750. PMID 25965083


The articles reviewed in this research include those obtained in an Internet based literature search for relevant medical references through November 2016, the date the research was completed.
<table>
<thead>
<tr>
<th>Policy Effective Date</th>
<th>BCBSM Signature Date</th>
<th>BCN Signature Date</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/1/14</td>
<td>6/17/14</td>
<td>6/23/14</td>
<td>Joint policy established</td>
</tr>
<tr>
<td>1/1/16</td>
<td>10/13/15</td>
<td>11/5/15</td>
<td>Routine maintenance</td>
</tr>
</tbody>
</table>
| 3/1/17               | 12/13/16             | 12/13/16           | Added code 81545, deleted NOC code. Updated Hayes information, references and rationale section. Medical Policy Statement changed: The use of the Afirma gene expression classifier in fine needle aspirates of the thyroid that are cytologically considered to be indeterminate (follicular lesion of undetermined significance or follicular neoplasm) may be established in patients who have the following characteristics:  
  - Thyroid nodules without strong clinical or radiological findings suggestive of malignancy  
  - In whom surgical decision-making would be affected by test results |

Next Review Date: 4th Qtr, 2017
BLUE CARE NETWORK BENEFIT COVERAGE
POLICY: GENETIC TESTING- MOLECULAR MARKERS IN FINE NEEDLE ASPIRATES (FNA)
OF THE THYROID

I. Coverage Determination:

<table>
<thead>
<tr>
<th>Plan Type</th>
<th>Coverage Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial HMO (includes Self-Funded groups unless otherwise specified)</td>
<td>Covered, criteria apply</td>
</tr>
<tr>
<td>BCNA (Medicare Advantage)</td>
<td>Covered; criteria apply; See Government Regulations section of policy.</td>
</tr>
<tr>
<td>BCN65 (Medicare Complementary)</td>
<td>Coinsurance covered if primary Medicare covers the service.</td>
</tr>
</tbody>
</table>

II. Administrative Guidelines:

- The member’s contract must be active at the time the service is rendered.
- Coverage is based on each member’s certificate and is not guaranteed. Please consult the individual member’s certificate for details. Additional information regarding coverage or benefits may also be obtained through customer or provider inquiry services at BCN.
- The service must be authorized by the member’s PCP except for Self-Referral Option (SRO) members seeking Tier 2 coverage.
- Services must be performed by a BCN-contracted provider, if available, except for Self-Referral Option (SRO) members seeking Tier 2 coverage.
- Payment is based on BCN payment rules, individual certificate and certificate riders.
- Appropriate copayments will apply. Refer to certificate and applicable riders for detailed information.
- CPT – HCPCS codes are used for descriptive purposes only and are not a guarantee of coverage.