Medical Policy

BlueCross BlueShield of Kansas
An independent licensee of the Blue Cross Blue Shield Association

Title: Molecular Markers in Fine Needle Aspirates of the Thyroid

Professional
Original Effective Date: March 17, 2017
Revision Date(s): March 17, 2017; October 1, 2017; April 15, 2018; September 14, 2018; January 1, 2019
Current Effective Date: September 14, 2018

Institutional
Original Effective Date: March 17, 2017
Revision Date(s): March 17, 2017; October 1, 2017; April 15, 2018; September 14, 2018; January 1, 2019
Current Effective Date: September 14, 2018

Populations	Interventions	Comparators	Outcomes
Individuals: • With thyroid nodule(s) and indeterminate findings on fine needle aspirate	Interventions of interest are: • Fine needle aspirate sample testing with molecular markers to rule out malignancy and to avoid surgical biopsy or resection	Comparators of interest are: • Surgical biopsy	Relevant outcomes include: • Disease-specific survival • Test accuracy • Test validity • Morbid events • Resource utilization
Individuals: • With thyroid nodule(s) and indeterminate findings on fine needle aspirate	Interventions of interest are: • Fine needle aspirate sample testing with molecular markers to rule in malignancy and to guide surgical planning	Comparators of interest are: • Surgical management based on clinicopathologic risk factors	Relevant outcomes include: • Disease-specific survival • Test accuracy • Test validity • Morbid events • Resource utilization

Populations	Interventions	Comparators	Outcomes
Individuals: • With thyroid nodule(s) and indeterminate findings on fine needle aspirate	Interventions of interest are: • Fine needle aspirate sample testing with molecular markers to rule out or to rule in malignancy for surgical planning	Comparators of interest are: • Surgical management based on clinicopathologic risk factors and/or surgical biopsy	Relevant outcomes include: • Disease-specific survival • Test accuracy • Test validity • Morbid events • Resource utilization

DESCRIPTION

To determine which patients need thyroid resection, many physicians will perform a cytologic examination of fine needle aspirate (FNA) samples from a thyroid lesion; however, this method has diagnostic limitations. As a result, assays using molecular markers have been developed in an attempt to improve the accuracy of thyroid FNA biopsies.

OBJECTIVE

The objective of this policy is to evaluate whether testing for molecular markers in fine needle aspirates of the thyroid leads to improved health outcomes in individuals with thyroid nodule(s) with an indeterminate finding on the fine needle aspirate.

BACKGROUND

Thyroid Nodules

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.

Diagnosis

Fine needle aspirate (FNA) samples 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, 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 Bethesda System criteria into the following groups: nondiagnostic; benign; follicular lesion of undetermined significance (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 AUS or FLUS or follicular neoplasm categories are often considered indeterminate.

Management

There is some individualization of management for patients with FNA-indeterminate nodules, but many patients will 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 has revealed a malignancy rate ranging from 6% to 30%, making this a clinical process with very low specificity.[2] Thus, if 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, because different thyroid malignancies require different surgical procedures (eg, unilateral lobectomy vs 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). 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 FNA in a case of PTC is indeterminate, surgical biopsy with intraoperative pathology consultation is most often diagnostic, although its efficacy and therefore its use will vary across institutions, surgeons, and pathologists. In 2016, reclassification of encapsulated follicular-variant PTC as a noninvasive follicular tumor with papillary-like nuclei was proposed and largely adopted; this classification removes the word *carcinoma* from the diagnosis to acknowledge the indolent behavior of these tumors.[3]

For follicular carcinoma, the presence of invasion of the tumor capsule or of blood vessels is diagnostic and cannot be determined by cytology, because 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.

Genetic Variants Associated with Thyroid Cancer

Various genetic variants have been discovered in thyroid cancer. The most common 4 gene mutations are *BRAF* and *RAS* single nucleotide variants (SNVs), and *RET/PTC* and *PAX8/PPARγ* rearrangements.

Papillary carcinomas carry SNVs 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 pathway.[4] These mutually exclusive variants are found in more than 70% of papillary carcinomas.[4] *BRAF* SNVs are highly specific for PTC. Follicular carcinomas harbor either *RAS* SNVs or *PAX8/PPARγ* rearrangements. These variants have been identified in 70% to 75% of follicular carcinomas.[4] Genetic alterations involving the PI3K/AKT signaling pathway also occur in thyroid tumors, although they are rare in well-differentiated thyroid cancers and have a higher prevalence in less differentiated thyroid carcinomas.[4] Additional variants known to occur in poorly differentiated and anaplastic carcinomas involve the *TP53* and *CTNNB1* genes. Medullary carcinomas, which can be familial or sporadic, frequently possess SNVs located in the *RET* gene.

Studies have evaluated the association between various genes and cancer phenotype in individuals with diagnosed thyroid cancer.[5-7]

Telomerase reverse transcriptase (*TERT*) promoter variants occur with varying frequency in different thyroid cancer subtypes. Overall, *TERT* C228T or C250T variants have been reported in approximately 15% of thyroid cancers, with higher rates in the undifferentiated and anaplastic subtypes compared with the well-differentiated subtypes.[8] *TERT* variants are associated with several demographic and histopathologic features such as older age and advanced TNM stage. *TERT* promoter variants have been reported to be independent predictors of disease recurrence and cancer-related mortality in well-differentiated thyroid cancer.[9-11] Also, the co-occurrence of *BRAF* or *RAS* variants with *TERT* or *TP53* variants may identify a subset of thyroid cancers with unfavorable outcomes.[12-14]

Molecular Diagnostic Testing

Variant Detection and Rearrangement Testing

SNVs in specific genes, including *BRAF*, *RAS*, and *RET*, and evaluation for rearrangements associated with thyroid cancers can be accomplished with Sanger sequencing or pyrosequencing or with real-time polymerase chain reaction (PCR) of single or multiple genes or by next-generation sequencing (NGS) panels. Panel tests for genes associated with thyroid cancer, with varying compositions, are also available. For example, Quest Diagnostics offers a Thyroid Cancer Mutation Panel, which includes *BRAF* and *RAS* variant analysis and testing for *RET/PTC* and *PAX8/PPARγ* rearrangements.

The ThyroSeq v.2 Next-Generation Sequencing panel (CBLPath, Ocala, FL) is an NGS panel of more than 60 genes. According to the CBLPath'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.[15] In particular, it has been evaluated in patients with follicular neoplasm and/or suspicious for follicular neoplasm on FNA as a test to increase both sensitivity and specificity for cancer diagnosis.

ThyGenX is an NGS panel that sequences 8 genes and identifies specific gene variants and translocations associated with thyroid cancer. ThyGenX is intended to be used in conjunction with the ThyraMIR microRNA expression test when the initial ThyGenX test is negative.

Gene Expression Profiling

Genetic alterations associated with thyroid cancer can be assessed using gene expression profiling, which refers to the analysis of messenger RNA (mRNA) expression levels of many genes simultaneously. Several gene expression profiling tests are now available to stratify tissue from thyroid nodules biologically.

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 evaluate thyroid nodules that have an "indeterminate" classification on FNA as a method to select patients ("rule out") who are at low risk for cancer.

Other gene expression profiles have been reported in investigational settings, but have not been widely validated or used commercially (eg, Barros-Filho et al [2015],[16] Zheng et al [2015][17]); they are not addressed in this review.

ThyraMIR™ is a microRNA expression–based classifier intended for use in thyroid nodules with indeterminate cytology on FNA following a negative result from the ThyGenX™ Thyroid Oncogene Panel.

Algorithmic Testing

Algorithmic testing involves the use of two or more tests in a prespecified sequence, with a subsequent test automatically obtained depending on results of an earlier test.

Algorithmic Testing Using Afirma GEC with Afirma MTC and Afirma BRAF

In addition to Afirma GEC, Veracyte also markets 2 "malignancy classifiers" that use mRNA expression-based classification to evaluate for *BRAF* variants (Afirma BRAF) or variants associated with medullary thyroid carcinoma (Afirma MTC). Table 1 describes the testing algorithms for Afirma MTC and Afirma BRAF.

Table 1. Afirma MTC and Afirma BRAF Testing Algorithms

Test 1	Test 1 Result	Reflex to Test 2
Thyroid nodule on fine needle aspirate	"Indeterminate"	Afirma MTC
Afirma GEC	"Malignant" or "suspicious"	Afirma MTC
Afirma GEC	"Suspicious"	Afirma BRAF

In a description of the Afirma BRAF test, the following have been proposed as benefits of the mRNA-based expression test for *BRAF* variants: (1) PCR-based methods may have low sensitivity, requiring that a large proportion of the nodule have a relevant variant; (2) testing for only 1 variant may not detect patients with low-frequency variants 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.[18]

The testing strategy for both Afirma MTC and Afirma BRAF is to predict malignancy from an FNA sample with increased pretest probability for malignancy. A positive result with Afirma MTC or Afirma BRAF would inform preoperative planning such as planning for a hemi- vs a total thyroidectomy or performance of a central neck dissection.

Algorithmic Testing Using ThyGenX and ThyraMIR

The ThyGenX Thyroid Oncogene Panel (Interpace Diagnostics, Parsippany, NJ; testing done at Asuragen Clinical Laboratory) is an NGS panel designed to assess patients with indeterminate thyroid FNA results. It includes sequencing of 8 genes associated with papillary thyroid carcinoma and follicular carcinomas. ThyGenX has replaced the predicate miRInform Thyroid test that assesses for 17 validated gene alterations.

ThyraMIR (Interpace Diagnostics, Parsippany, NJ) is a microRNA expression–based classifier intended for use in thyroid nodules with indeterminate cytology on FNA following a negative result from the ThyGenX Thyroid Oncogene Panel.

The testing strategy for combined ThyGenX and ThyraMIR testing is first to predict malignancy. A positive result on ThyGenX would "rule in" patients for surgical resection. The specific testing results from a ThyGenX positive test would be used to inform preoperative planning when positive. For a ThyGenX negative result, the reflex testing involves the ThyraMIR microRNA expression test to "rule out" for a surgical biopsy procedure given the high negative predictive value of the second test. Patients with a negative result from the ThyraMIR test would be followed with active surveillance and avoid a surgical biopsy.

REGULATORY STATUS

Clinical laboratories may develop and validate tests in-house and market them as a laboratory service; laboratory-developed tests must meet the general regulatory standards of the Clinical Laboratory Improvement Amendments. Thyroid variant testing and gene expression classifiers are available under the auspices of the Clinical Laboratory Improvement Amendments. Laboratories that offer laboratory-developed tests must be

licensed by the Clinical Laboratory Improvement Amendments for high-complexity testing. To date, the U.S. Food and Drug Administration has chosen not to require any regulatory review of this test.

In 2013, the THxID™-BRAF kit (bioMérieux, Marcy l'Etoile, France), an in vitro diagnostic device, was approved by the Food and Drug Administration through the premarket approval process to assess specific *BRAF* variants in melanoma tissue via real-time PCR. However, there are currently no diagnostic tests for thyroid cancer mutation analysis with approval from the Food and Drug Administration.

Table 2 provides a summary of commercially available molecular diagnostic tests for indeterminate thyroid pathology.

Table 2. Summary of Molecular Tests for Indeterminate Thyroid Cytopathology FNA Specimens

Test	Methodology	Analyte(s)	Report
Afirma® GEC	mRNA gene expression	167 genes	Benign/suspicious
Afirma® BRAF	mRNA gene expression	1 gene	Negative/positive
Afirma® MTC	mRNA gene expression		Negative/positive
ThyroSeq v2, v3	Next-generation sequencing	60+ genes[b]	Specific gene variant/translocation
ThyGenX™[a]	Next-generation sequencing	8 genes	Specific gene variant/translocation
TERT single-gene test	Unclear for commercially available test[c]	1 gene	Specific gene variants
miR*Inform*®[a]	Multiplex PCR by sequence-specific probes	14 DNA variants, 3 RNA fusions	Specific gene variant/translocation
ThyraMIR™	microRNA expression	10 microRNAs	Negative/positive
RosettaGX™ Reveal	microRNA expression	24 microRNAs	• Benign • Suspicious for malignancy • High risk for medullary carcinoma

FNA: fine needle aspirate; NGS: next-generation sequencing; PCR: polymerase chain reaction.
[a] The miR*Inform*® test is the predicate test to ThyGenX™ and is not commercially available.
[b] Includes *TERT*.
[c] Available literature on *TERT* testing used PCR.

POLICY

A. The use of either Afirma Gene Expression Classifier or ThyroSeq v2 in fine needle aspirates of thyroid nodules with indeterminate cytologic findings (ie, Bethesda diagnostic category III [atypia / follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm / suspicion for a follicular neoplasm]) may be considered **medically necessary** in patients who have **ALL** of the following characteristics:

    1. Thyroid nodules without strong clinical or radiologic findings suggestive of malignancy, **AND**

2. In whom surgical decision making would be affected by test results.

B. The use of any of the following types of molecular marker testing or gene variant analysis in fine needle aspirates of thyroid nodules with indeterminate findings (Bethesda diagnostic category III [atypia / follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm / suspicion for a follicular neoplasm]) or suspicious findings (Bethesda diagnostic category V [suspicious for malignancy]) to rule in malignancy to guide surgical planning for initial resection rather than a 2-stage surgical biopsy followed by definitive surgery may be considered **medically necessary**:

1. ThyroSeq v2;

2. ThyraMIR microRNA / ThyGenX;

3. Afirma BRAF after Afirma Gene Expression Classifier; or

4. Afirma MTC after Afirma Gene Expression Classifier.

C. Gene expression classifiers, genetic variant analysis, and molecular marker testing in fine needle aspirates of the thyroid not meeting criteria outlined above, including, but not limited to, use of RosettaGX Reveal and single-gene *TERT* testing, are considered **experimental / investigational**.

Policy Guidelines

1. In patients who do not undergo surgical biopsy or thyroidectomy on the basis of gene expression classifier or molecular marker results, regular active surveillance is indicated.
2. Use of molecular marker testing based on fine needle aspirate of a thyroid nodule to rule in malignancy prior to surgical biopsy may guide surgical planning, particularly factors such as choice of surgical facility provider to ensure that the capability is available to conduct a frozen section pathologic reading during surgical biopsy so that surgical approach may be adjusted accordingly in 1 surgery.
3. Genetics Nomenclature Update
   Human Genome Variation Society (HGVS) nomenclature is used to report information on variants found in DNA and serves as an international standard in DNA diagnostics. It is being implemented for genetic testing medical evidence review updates starting in 2017 (see Table PG1). HGVS nomenclature is recommended by HGVS, the Human Variome Project, and the HUman Genome Organization.

   The American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) standards and guidelines for interpretation of

sequence variants represent expert opinion from ACMG, AMP, and the College of American Pathologists. These recommendations primarily apply to genetic tests used in clinical laboratories, including genotyping, single genes, panels, exomes, and genomes. Table PG2 shows the recommended standard terminology—"pathogenic," "likely pathogenic," "uncertain significance," "likely benign," and "benign"—to describe variants identified that cause Mendelian disorders.

Table PG1. Nomenclature to Report on Variants Found in DNA

Previous	Updated	Definition
Mutation	Disease-associated variant	Disease-associated change in the DNA sequence
	Variant	Change in the DNA sequence
	Familial variant	Disease-associated variant identified in a proband for use in subsequent targeted genetic testing in first-degree relatives

Table PG2. ACMG-AMP Standards and Guidelines for Variant Classification

Variant Classification	Definition
Pathogenic	Disease-causing change in the DNA sequence
Likely pathogenic	Likely disease-causing change in the DNA sequence
Variant of uncertain significance	Change in DNA sequence with uncertain effects on disease
Likely benign	Likely benign change in the DNA sequence
Benign	Benign change in the DNA sequence

ACMG: American College of Medical Genetics and Genomics; AMP: Association for Molecular Pathology.

RATIONALE

The most recent policy update includes a review of the literature through April 9, 2018.

Evidence reviews assess whether a medical test is clinically useful. A useful test provides information to make a clinical management decision that improves the net health outcome. That is, the balance of benefits and harms is better when the test is used to manage the condition than when another test or no test is used to manage the condition.

The first step in assessing a medical test is to formulate the clinical context and purpose of the test. The test must be technically reliable, clinically valid, and clinically useful for that purpose. Evidence reviews assess the evidence on whether a test is clinically valid and clinically useful. Technical reliability is outside the scope of these reviews, and credible information on technical reliability is available from other sources.

Molecular Markers to Rule Out Malignancy

Clinical Context and Test Purpose

The purpose of molecular testing in individuals with indeterminate findings on fine needle aspirate(s) (FNA) of thyroid nodules is to rule out malignancy and eliminate the need for surgical biopsy or resection.

The relevant question addressed in this evidence review is: Does molecular testing appropriately eliminate the need for surgical biopsy or resection and lead to improved health outcomes? The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest is individuals with indeterminate findings on FNAs of thyroid nodules who would be willing to undergo watchful waiting, depending on results of their molecular testing. Patients with indeterminate findings after FNA of thyroid nodule presently proceed to surgical biopsy or resection.

Interventions
The test being considered is molecular testing, which includes either Afirma Gene Expression Classifier (GEC), ThyroSeq v2, or RosettaGX Reveal.

Comparators
The following practice is currently being used: standard surgical management through surgical biopsy or resection for biopsy.

Outcomes
The potential beneficial outcome of primary interest would be avoiding an unneeded surgical biopsy or resection (eg, lobectomy or hemithyroidectomy) in a true-negative thyroid nodule that is benign.

Potential harmful outcomes are those resulting from false-negative test results, which may delay diagnosis and surgical resection for thyroid cancer. For small, slow-growing tumors, it is uncertain that a delay in diagnosis would necessarily a worsen health outcomes.

Timing
The time frame for evaluating the performance of the test varies from the initial FNA to surgical resection to weeks to months following an indeterminate result to years. Papillary thyroid cancer (PTC) is indolent, and a nodule could be observed for many years to ensure no clinical change.

Setting
The primary setting would be in endocrinology.

Afirma GEC
Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Chudova et al (2010) described the development and initial clinical validation of a version of the Afirma GEC.[2] The classifier was trained on 178 retrospectively identified surgical thyroid specimens, which represented a variety of malignant and benign disorders, and separately on a set of 137 FNA samples with known surgical pathology. The classifier was developed with the objective of achieving a negative predictive value (NPV) specificity of 95% and a specificity of 70%. The tissue-trained classifier was tested on an independent sample of 48 FNAs (24 with

indeterminate cytopathology, 24 with a mix of malignant and benign cytopathology). The FNA-trained classifier was tested separately on the same sample of 48 FNAs. In the 24 samples with indeterminate cytopathology, sensitivity and specificity were 100% (95% confidence interval [CI], 64% to 100%) and 73.3% (95% CI, 49% to 89%), respectively.

Prospective Clinical Validation: Alexander et al (2012) reported on a 19-month, prospective, multicenter (49 academic and community sites), study of the Afirma GEC.[19] A total of 4812 nodules were screened for inclusion with centralized cytopathology. 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 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 without 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% to 59%). The NPV ranged from 85% for "suspicious cytologic findings" to 95% for "atypia of undetermined clinical significance." Seven FNAs had false-negative results, six of which were thought to be due to hypocellular aspirate specimens.

Retrospective Clinical Validation: Santhanam et al (2016) conducted a meta-analysis of studies reporting on the performance of the Afirma GEC in cytologically indeterminate nodules.[20] Seven studies met inclusion criteria, which required that studies reported on the use of the Afirma GEC in nodules found indeterminate on FNA (including atypia of uncertain significance [AUS] or follicular lesion of undetermined significance [FLUS]; suspicious for follicular or 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. Although the authors reported pooled results, these results (particularly specificity) were likely biased given the lack of reference standard diagnosis for most lesions in the included studies (see the following section and Table 3).

Duh et al (2017) reported on an updated systematic review that included the publications in the Santhanam (2016) review along with additional publications up to mid-2016.[21] Twelve studies were selected. Reviewers did not report pooled results because most included studies did not meet minimal quality standards, primarily due to the lack of reference standard diagnoses for GEC-benign nodules. Several studies from the systematic review are included in Table 3. Retrospective multicenter and single-center studies, including Harrell and Bimston (2014),[22] Lastra et al (2014),[23] McIver et al (2014),[24] Yang et al (2016),[25] Witt (2016),[26] Baca et al (2017),[27] Harrison et al (2017),[28] Kay-Rivest et al (2017),[29] Hang et al (2017),[30] and Samulski et al (2016)[31] have reported on the diagnostic accuracy of the Afirma GEC (see Table 3). All studies

were subject to ascertainment bias because a large proportion of individuals, with Afirma benign reports did not undergo surgery, which made 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 were 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. An additional publication (Celik et al [2015]) reported on Afirma GEC testing; however, included in this publication were individuals with benign and suspicious cytology on FNA—and those individuals are not necessarily considered to be part of the "target population."[32]

Table 3. Studies Reporting Afirma GEC Results

Study	Indeterminate FNA Samples, n (%)	Afirma GEC Test Result	N	With Thyroidectomy, n	With Malignancy on Thyroidectomy, n
Alexander et al (2014)[33]	165 AUS/FLUS 174 FN/SFN	Suspicious Benign Nondiagnostic	148 174 17	121 11	53 1
Harrell and Bimston (2014)[22]	58 AUS/FLUS or FN	Suspicious Benign	36[a] 20	30 5	21 2
Lastra et al (2014)[23]	69 (51.5) AUS/FLUS 39 (29.5) FN 25 (19) FNOF	Suspicious Benign	62 70	48 2	22 0
McIver et al (2014)[24]	12 (11.4) AUS/FLUS 93 (88.6) FN/HCN	Suspicious Benign	44[b] 16	32 4	5 1
Yang et al (2016)[25]	165 (76) AUS/FLUS 24 (11) SFN/FN	Suspicious Benign	80 94	62 5	32 0
Witt (2016)[26]	47 AUS/FLUS or SFN/FN (32 with GEC attempted[c])	Suspicious Benign	15 14	15 0	6 NA; followed clinically
Samulski et al (2016)[31]	1 B/HN 166 AUS/FLUS 77 FN/SFN 45 FNOF 1 neoplasm present 4 nondiagnostic	Suspicious Benign QNS	136 145 13	107 16	42 3
Baca et al (2017)[27]	36 AUS-A 25 AUS-C 39 AUS-C/A	Suspicious Benign	105 122	90 10	37 0
Harrison et al (2017)[28]	38 AUS/FLUS 4 FN/SFN 2 B/HN 1 Susp malignancy 1 nondiagnostic	Suspicious Benign	57 52	46 3	16 0
Kay-Rivest et al (2017)[29]	105 AUS/FLUS 67 FN/SFN	Suspicious Benign	83 89	77 0	38 NA; followed clinically
Hang et al (2017)[30]	304 AUS 80 SFN	Suspicious Benign	202 182	151 24	56 1

AUS: atypia of undetermined significance; AUS-A, AUS with architectural atypia; AUS-C: AUS with cytologic atypia; AUS-C/A: AUS with cytologic and architectural atypia; B/HN: benign/hyperplastic; FLUS: follicular lesion of undetermined significance; FN: follicular neoplasm; FNA: fine needle aspirates; FNOF: follicular neoplasm with oncocytic features; HCN: Hürthle cell neoplasm; NA: not applicable; QNS: quality not sufficient; SFN: suspicious for follicular neoplasm.

[a] Two samples inadequate due to low mRNA content.

[b] GEC results were available for 60 subjects.

[c] Three samples were inadequate.

There are limited data on the true-negative rates 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 have reported on long-term follow-up of individuals with indeterminate FNA cytology and Afirma GEC benign results. Angell et al (2015) retrospectively compared clinical outcomes for individuals who had indeterminate FNA cytology and Afirma GEC benign results with individuals who had cytologically benign nodule results.[34] A total of 95 cytologically indeterminate and 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 and Afirma GEC benign nodules (13.8% vs 0.9%, p<0.001).

Clinically Useful

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials.

No evidence directly demonstrating improved outcomes in patients managed with the Afirma GEC was identified.

Chain of Evidence: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Because no direct evidence of utility was identified, a chain of evidence was developed, which addresses 2 key questions:

1. Does use of the Afirma GEC in individuals with cytologically indeterminate thyroid nodules change clinical management (in this case, reduced thyroid resections)?
2. Do those management changes improve outcomes?

Changes in Management: The clinical setting in which the Afirma GEC is meant to be used is well-defined: individuals with AUS or FLUS or follicular neoplasm or who are suspicious for follicular neoplasm (SFN) on FNA, who do not have other indications for thyroid resection (ie, 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, have suggested that, in at least some cases, surgical decision making changed. These studies are described briefly.

Duick et al (2012) reported on the impact of Afirma GEC test results on physician and patient decision making to resect thyroid nodules with indeterminate cytology and Afirma GEC benign results in a sample of 395 nodules from 368 patients.[35] Surgery was performed in 7.6% of the

patients with indeterminate cytology and a benign GEC result, less than the historical rate of thyroid resection (74%) in patients with indeterminate cytology.

Sipos et al (2016) retrospectively analyzed nonacademic medical practices using the Afirma GEC to determine the long-term nonoperative rate of thyroid nodules with benign results.[36] Of the patients with Afirma benign results during 36 months of follow-up, 17.3% underwent surgery. Eighty-eight percent of all surgeries were performed within the first 2 years after a benign Afirma GEC result.

The study by Alexander et al (2014) provided evidence on clinical management changes for patients with indeterminate thyroid nodules tested with Afirma GEC.[33] 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],[37] Noureldine et al [2015][38]) evaluated the potential for the Afirma GEC test to change surgical decision making by comparing actual surgical decision making when 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 (10%-15%).

Abeykoon et al (2016) studied the impact of implementing Afirma GEC at a single center.[39] Surgical recommendations for patients with indeterminate thyroid nodules decreased from 81.5% pre-Afirma GEC to 50% post-Afirma GEC. The rate of malignant surgical pathology diagnosis increased from 20% pre-Afirma GEC to 85.7% post-Afirma. The implementation of Afirma GEC decreased the number of surgical recommendations and increased the rate of malignancy detected for patients who received a surgical biopsy.

Chaudhary et al (2016) studied the impact on surgical outcomes pre- and postimplementation of Afirma GEC.[40] A total of 158 FNAs were sent for Afirma GEC with 73 suspicious and 8 benign Afirma cases going for surgeries. Compared with before implementation of Afirma GEC, the rate for surgical biopsy decreased from 61% to 54% but was not statistically significant. In the SFN, the rate of surgical biopsy significantly decreased from 76% to 52%.
Dhingra (2016) studied the effects of an FNA protocol combining expert thyroid cytopathology plus Afirma GEC in community practice.[41] Historical data were compared with data after implementation of the FNA protocol. Prior to protocol implementation, the rates of indeterminate cytology and diagnostic surgeries were 26% and 24%. After protocol implementation, the rates of indeterminate cytology and diagnostic surgeries decreased to 10% and 6%. The effect of Afirma GEC implementation could not be ascertained given the FNA protocol combined expert thyroid cytopathology and Afirma GEC.

Improved Outcomes: A simplified decision model was developed for use with Afirma GEC in individuals with cytologically indeterminate FNA samples. It is shown in Appendix Figure 1. It is assumed that when Afirma GEC is not used, patients with cytologically indeterminate FNA results undergo thyroid resection. When Afirma GEC is used, those with Afirma suspicious lesions undergo resection, while those who have Afirma benign lesions do not. In this case, compared with 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%,[42] 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),[19] it is estimated that approximately 1.6% of individuals with 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 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 events associated with thyroidectomy or hemithyroidectomy. Among the low-risk population, the alternative to surgical biopsy is ongoing active surveillance.

RosettaGX Reveal

Technically Reliable

Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist.

This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid

A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Lithwick-Yanai et al (2017) described the development and initial clinical validation of the RosettaGX Reveal quantitative real-time PCR assay for 24 microRNA samples in a multicenter, retrospective cohort study using 201 FNA smears.[43] The results of the clinical validation study are reported in Table 4.

Table 4. RosettaGX Reveal Performance

Outcomes	Value, %	95% Confidence Interval, %
Samples passing QC (n=189)		
Sensitivity	85	74 to 93
Specificity	72	63 to 79
Negative predictive value	91	84 to 96

Outcomes	Value, %	95% Confidence Interval, %
Samples with 3 pathologists agreeing on final diagnosis (n=150, subset of samples passing QC)		
Sensitivity	98	87 to 100
Specificity	78	69 to 85
Negative predictive value	99	94 to 100

Adapted from Lithwick-Yanai et al (2017).[43]
QC: quality control.

No prospective clinical studies for RosettaGX Reveal were identified.

Clinically Useful
Direct Evidence: No evidence directly demonstrating improved outcomes in patients managed with the RosettaGX Reveal was identified.

Chain of Evidence: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Section Summary: Molecular Markers to Rule Out Malignancy
In a single multicenter validation study, the Afirma GEC test 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 [2010]), 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 has suggested 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. A chain of evidence can be constructed to establish the potential for clinical utility with GEC testing in cytologically indeterminate lesions, but with only a single study with the marketed test reporting a true NPV, the clinical validity is uncertain. For the RosettaGX Reveal test, a retrospective clinical validation has been reported. No prospective studies for patients managed with the RosettaGX Reveal were identified, so the clinical validity remains uncertain.

Molecular Markers to Rule In Malignancy
Clinical Context and Test Purpose
The purpose of testing for molecular markers (eg, single nucleotide variants [SNVs] and gene rearrangements) in individuals with indeterminate findings on FNA of thyroid nodules is to rule in malignancy and to guide surgical approach or management.

The relevant question addressed in this evidence review is: Does testing for molecular markers predict malignancy and alter surgical approach or management and lead to improved health outcomes?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest is individuals with indeterminate findings on FNA(s) of thyroid nodules. Patients with indeterminate findings would presently proceed to surgical biopsy perhaps with intraoperative pathology consultation (ie, intraoperative frozen section) if available.

Interventions
The test being considered is testing for molecular markers (eg, SNVs and gene rearrangements) to rule in malignancy and to use molecular marker results that are positive for variants associated with malignancy to guide surgical planning to ensure the capability for intraoperative pathologic confirmation of malignancy to adjust to definitive surgery for initial resection if appropriate.

Comparators
The following practices are currently being used: standard surgical management through surgical resection, including a 2-stage surgical biopsy (ie, lobectomy) followed by definitive surgery (ie, hemithyroidectomy or thyroidectomy).

Outcomes
The potential beneficial outcome of primary interest is appropriate surgical planning in the preoperative period (eg, hemithyroidectomy or thyroidectomy when malignancy is predicted). This has the potential benefit of reducing the likelihood of having the patient repeating surgery if a diagnosis is not made on frozen pathology section during the initial surgery if lobectomy is done as a first procedure.

Potential harmful outcomes are those resulting from false-positive results. However, the use of intraoperative confirmation of malignancy through frozen pathology section in patients with positive molecular marker testing would mitigate any risk of inappropriately performing more extensive thyroidectomy in the absence of malignancy.

Timing
The time frame for evaluating the performance of the test varies from the initial FNA to surgical resection to weeks to months following an indeterminate result.

Setting
The primary setting would be in endocrinology.

Gene Expression Classifiers to Predict Malignancy
Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

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

Afirma BRAF and Afirma MTC: In the study by Diggans et al (2015), 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.[18] 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 a study describing the development and validation of the Afirma MTC classifier, Kloos et al (2016) evaluated the MTC classifier in a sample of 10,488 thyroid nodule FNA samples referred for GEC testing.[44] In this sample, 43 cases were Afirma MTC–positive, of which 42 were considered to be clinically consistent with MTC on pathology or biochemical testing, for a positive predictive value (PPV) of 97.7% (95% CI, 86.2% to 99.9%).

TERT Single-Gene Testing: Published literature describing performance characteristics of the marketed version (Interpace Diagnostics) of a *TERT* single-gene test was not identified. Three studies have reported performance characteristics of a *TERT* test; all 3 studies evaluated *TERT* promoter variants C228T and C250T. Study characteristics and results are shown in Tables 5 and 6. Study relevance, design, and conduct gaps are shown in Tables 7 and 8.

Nikiforov et al (2014) evaluated the accuracy of the ThyroSeq v2 NGS panel that included tests for SNVs in 13 genes (including *TERT*) and for 42 types of gene fusions in a series of 143 consecutive thyroid FNA samples with a cytologic diagnosis of follicular or Hürthle cell neoplasm or suspicious for follicular or Hürthle cell neoplasm.[45] Molecular testing was retrospectively performed for 91 samples and prospectively performed for the remaining 52. Results for performance characteristics of the *TERT* variant alone were reported. Four of 39 total cancers were identified as *TERT*-positive (2 were unique diagnostic events); there were no *TERT*-positive results in the benign samples.

Liu and Xing (2014) described the performance of *TERT* as a single-gene test. FNA biopsy specimens were obtained preoperatively from thyroid nodules in 308 patients who underwent thyroidectomy.[46] The percentage of samples that showed indeterminate cytologic findings on FNA biopsy was not described. The disposition of samples meeting eligibility criteria and a number of samples that did not produce results, were not described. Standard PCR was performed for direct genomic DNA sequencing to identify *TERT* promoter variants (C228T, C250T). One hundred twenty-nine (42%) of the samples were positive for thyroid cancer by pathology following surgery (111 PTC, 18 follicular thyroid carcinomas). *TERT* promoter variants C228T and C250T were found in 9 cases of thyroid cancer and no *TERT* variants were found in the 179 benign samples.

Decaussin-Petrucci et al (2017) described molecular testing for *BRAF*, *RAS*, and *TERT* variants in a prospective cohort of 326 cases, including 61 AUS, 124 follicular neoplasms, 72 suspicious for malignancy, and 69 malignant cases.[47] Diagnosis of malignancy was confirmed by histology on paired surgical specimens in 163 samples. The flow of samples meeting eligibility criteria and number of samples that did not produce results were not described. The results here focus on the analysis of *TERT* single-gene tests. Nine *TERT* variants were detected, and all were confirmed to be malignant.

Table 5. Clinical Validity Study Characteristics of *TERT* as a Single-Gene Test

Study	Study Population	Design	Reference Standard for Dx	Threshold Score for Positive Test	Timing of Reference and TERT Tests	Blinding of Assessors
Nikiforov et al (2014)[45]	Patients who underwent surgery with known surgical pathology outcome (FN/SFN)	• Retrospective (91 samples) and prospective (52 samples)	Pathologic diagnosis	Positive for either *TERT* promoter variants C228T or C250T	• Retrospective: samples collected before study and analyzed afterward • Prospective samples collected and analyzed in real-time	Yes
Liu and Xing (2014)[46]	Patients who underwent thyroidectomy for established thyroid cancer, cytologically indeterminate thyroid nodules or symptomatic goiter	• Retrospective • Not clear if samples were consecutive	Pathologic diagnosis	Positive for either *TERT* promoter variants C228T or C250T	FNAB samples obtained preoperatively; time between FNAB and surgery not specified	Not specified
Decaussin-Petrucci et al (2017)[47]	Patients with indeterminate or malignant cytology	• Prospective	Histology	Positive for either *TERT* promoter variants C228T or C250T	Samples collected before surgery	Yes

Dx: diagnosis; FN: follicular neoplasm; FNAB: fine needle aspiration biopsy; SFN: suspicious for a follicular neoplasm.

In summary, no studies of the validity of the marketed version of the *TERT* single-gene test were identified. Three studies reported information on *TERT* testing sufficient to calculate performance characteristics. The sample sizes of the included studies are approximately 150 to 350, with the prevalence of *TERT* variants between 3% and 7% and prevalence of cancer between 27% and 42%. Specificity was 100% in all studies (ie, there were no false-positives); however, the confidence intervals for PPV were extremely wide.

Table 6. Clinical Validity Results of *TERT* as a Single-Gene Test

Study	N		Excluded Samples	Prevalence, %		Sensitivity (95% CI), %	Specificity (95% CI), %	PPV (95% CI), %	NPV (95% CI), %
	Initial	Final		*TERT* Variants	Cancer				
Nikiforov et al (2014)[45]	144	143	Insufficient amount of isolated NAs	3	27	10 (3 to 24)[a]	100 95 to 100)[a]	100 (28 to 100)[a]	75 (67 to 82)[a]
Liu and Xing (2014)[46]	NR	308	NR	7	42	7 (3 to 13)[a]	100 (97 to 100)[a]	100 (55 to 100)[a]	60 (54 to 65)[a]
Decaussin-Petrucci et al (2017)[47]									
Overall	NR	326	Unclear	5.5	50	6 (3 to 10)[a]	100 (97 to 100)[a]	100 (55 to 100)[a]	51 (46 to 57)[a]

Study	N		Excluded Samples	Prevalence, %		Sensitivity (95% CI), %	Specificity (95% CI), %	PPV (95% CI), %	NPV (95% CI), %
IC (AUS/ FN/SM)	257	197	Insufficient sample	5.2	39	5 (1 to 13)[a]	100 (95 to 100) [a]	100 (28 to 100)[a]	62 (55 to 69)[a]

AUS: atypia of undetermined significance; CI: confidence interval; FN: follicular neoplasm; IC: indeterminate cytology; NPV: negative predictive value; NA: nucleic acid; NR: not reported; PPV: positive predictive value; SM: suspicious for malignancy.
[a] CIs calculated from data provided.

As mentioned, none of the 3 studies reported on the marketed version of the test, although one reported on *TERT* performance within the existing ThyroSeq marketed test. Liu and Xing (2014) was not limited to nodules that were indeterminate, and neither Liu (2014) nor Decaussin-Petrucci (2017) reported on the disposition of all eligible samples.

Table 7. Clinical Validity Study Relevance Gaps of *TERT* as a Single-Gene Test

Study	Population[a]	Intervention[b]	Comparator[c]	Outcomes[d]	Duration of Follow-Up[e]
Nikiforov et al (2014)[45]				3. Performance characteristics for *TERT* not reported but calculated from data provided	
Liu and Xing (2014)[46]	4. Study population is not limited to nodules that are indeterminate following FNA	3. Marketed version of *TERT* test not used	3. No comparison to other tests available for this purpose	3. NPV and PPV not reported but calculated from data provided	
Decaussin-Petrucci et al (2017)[47]		3. Marketed version of *TERT* test not used		3. Performance characteristics for *TERT* not reported but calculated from data provided	

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.
FNA: fine needle aspirate; NPV: negative predictive value; PPV: positive predictive value.
[a] Population key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.
[b] Intervention key: 1. Classification thresholds not defined; 2. Version used unclear; 3. Not intervention of interest.
[c] Comparator key: 1. Classification thresholds not defined; 2. Not compared to credible reference standard; 3. Not compared to other tests in use for same purpose.
[d] Outcomes key: 1. Study does not directly assess a key health outcome; 2. Evidence chain or decision model not explicated; 3. Key clinical validity outcomes not reported (sensitivity, specificity and predictive values); 4. Reclassification of diagnostic or risk categories not reported; 5. Adverse events of the test not described (excluding minor discomforts and inconvenience of venipuncture or noninvasive tests).
[e] Follow-Up key: 1. Follow-up duration not sufficient with respect to natural history of disease (true positives, true negatives, false positives, false negatives cannot be determined).

Table 8. Clinical Validity Study Design and Conduct Gaps of *TERT* as a Single-Gene Test

Study	Selection[a]	Blinding[b]	Delivery of Test[c]	Selective Reporting[d]	Data Completeness[e]	Statistical[f]
Nikiforov et al (2014)[45]				1. No registration reported		1. CIs for *TERT* performance characteristics not reported but calculated

Study	Selection[a]	Blinding[b]	Delivery of Test[c]	Selective Reporting[d]	Data Completeness[e]	Statistical[f]
						from data provided
Liu and Xing (2014)[46]	2. Not clear if samples consecutive or random	1. Blinding not specified but likely to be blinded	1. Unclear how much time elapsed between FNAB and surgery	1. No registration reported	1. No description of samples that were insufficient for processing or failed to produce result	1. CIs for performance characteristics not reported but calculated from data provided
Decaussin-Petrucci et al (2017)[47]			1. Unclear how much time elapsed between FNAB and surgery	1. No registration reported	1. No description of samples that were insufficient for processing or failed to produce result	1. CIs for performance characteristics not reported but calculated from data provided

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.
CI: confidence interval; FNAB: fine needle aspirate biopsy.
[a] Selection key: 1. Selection not described; 2. Selection not random or consecutive (ie, convenience).
[b] Blinding key: 1. Not blinded to results of reference or other comparator tests.
[c] Test Delivery key: 1. Timing of delivery of index or reference test not described; 2. Timing of index and comparator tests not same; 3. Procedure for interpreting tests not described; 4. Expertise of evaluators not described.
[d] Selective Reporting key: 1. Not registered; 2. Evidence of selective reporting; 3. Evidence of selective publication.
[e] Data Completeness key: 1. Inadequate description of indeterminate and missing samples; 2. High number of samples excluded; 3. High loss to follow-up or missing data.
[f] Statistical key: 1. Confidence intervals and/or p values not reported; 2. Comparison to other tests not reported.

Clinically Useful

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials.

Testing for specific variants associated with thyroid cancer (eg, *BRAF* V600E, *TERT,* and *RET* variants, *RET/PTC* and *PAX8/PPARγ* rearrangements) is generally designed to "rule in" cancer in nodules with indeterminate cytology on FNA.[48] (Of note, some gene 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 variant testing would be in informing preoperative planning for thyroid surgery following initial thyroid FNA, such as planning for a hemi- vs a total thyroidectomy or performance of a central neck dissection.

In a retrospective analysis, Yip et al (2014) reported on outcomes after implementation of an algorithm incorporating molecular testing of thyroid FNA samples to guide the extent of initial thyroid resection.[49] 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 SNVs; *RET/PTC1*, *RET/PTC3*, and *PAX8/PPARγ* rearrangements) was

prospectively obtained for all FNAs with indeterminate cytology (FLUS, follicular neoplasm, suspicious for malignancy), and for selective FNAs at the request of the managing physician for selected nodules with benign or nondiagnostic cytology. The study also included a second cohort of patients who did not have molecular testing results available. For patients treated with molecular diagnosis, a positive molecular diagnostic test was considered an indication for an initial total thyroidectomy. Patients with FLUS and negative molecular diagnostic results were followed with repeat FNA, followed by 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 managed with and 349 without molecular diagnostics. Positive molecular testing results were obtained in 56 (17% of those managed with molecular diagnostics) patients, most commonly *RAS* variants (42/56 [75%]), followed by *BRAF* V600E (10/56 [18%]) and *BRAF* K601E (2/56 [4%]) variants, and *PAX8/PPARγ* rearrangements (2/56 [4%]). Compared with those managed without molecular diagnostics (63%), patients managed with molecular diagnostics (69%) were nonsignificantly less likely to undergo total thyroidectomy as an initial procedure (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% (170/671) of patients had clinically significant thyroid cancer, with no difference in thyroid cancer rates based on the type of initial surgery (26% for total thyroidectomy vs 22% for lobectomy, p=0.3). The incidence of clinically significant thyroid cancer after initial lobectomy (ie, 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 a sensitivity and specificity for the diagnostic of thyroid cancer of 89% and 27%, respectively, with a PPV of 29% and an NPV of 88%. The addition of molecular diagnostics to FNA results increased the specificity for a cancer diagnosis to 95% and the PPV to 82%.

Chain of Evidence: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

In 2015, a task force from the American Thyroid Association published a review with recommendations for the surgical management of FNA-indeterminate nodules using various molecular genetic tests.[50] This review reported on the estimated likelihood of malignancy in an FNA-indeterminate nodule depending on results of the Afirma GEC test (described above) and other panels designed to rule in malignancy. Depending on the estimated prebiopsy likelihood of malignancy, recommendations for surgery included observation, active surveillance, repeat FNA, diagnostic lobectomy, or oncologic thyroidectomy.

Section Summary: Molecular Markers to Predict Malignancy
The available evidence has suggested that use of variant testing in thyroid FNA samples is generally associated with high specificity and PPV for clinically significant thyroid cancer. The most direct evidence related to the clinical utility of variant testing for genes associated with malignancy in thyroid cancer comes from a single-center retrospective study that 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 permits better preoperative planning at the time of thyroid biopsy, potentially avoiding the need for a separate surgery. A statement from the American Thyroid Association

provides some guidelines for surgeons managing patients with indeterminate nodules. However, adoption of these guidelines in practice and outcomes associated with them are uncertain.

Molecular Markers to Rule Out and Rule In Malignancy

Clinical Context and Test Purpose

The purpose of the ThyroSeq v2 test and the combined ThyGenX Thyroid Oncogene Panel plus ThyraMIR microRNA classifier in individuals with indeterminate findings on FNA(s) of thyroid nodules is to predict malignancy and inform surgical planning decisions with positive results using ThyroSeq v2 or the ThyGenX, and if negative, to predict benignancy using ThyraMIR microRNA classifier to eliminate or necessitate the need for surgical biopsy and guide surgical planning.

The relevant question addressed in this evidence review is: Does the ThyroSeq v2 test or the combined use of ThyGenX and ThyraMIR appropriately eliminate or necessitate the need for surgical resection or biopsy and lead to improved health outcomes?

The following PICOTS were used to select literature to inform this review.

Patients

The relevant population of interest includes individuals with indeterminate findings on FNA(s) of thyroid nodules. Patients with indeterminate findings presently proceed to surgical resection.

Interventions

The tests being considered are either: (a) the ThyroSeq v2 test or (b) the combined ThyGenX Thyroid Oncogene Panel and ThyraMIR microRNA classifier testing.

Comparators

The following practices are currently being used: surgical biopsy and/or standard surgical management through surgical resection.

Outcomes

The potential beneficial outcomes of primary interest are using a true-negative result to avoid an unneeded surgical biopsy or using a true-positive result to guide surgical resection (eg, hemithyroidectomy or thyroidectomy).

Potential harmful outcomes are those resulting from false-positive or false-negative test results. False-positive test results can lead to unnecessary surgical biopsy or resection and procedure-related complications. False-negative test results can lead to lack of surgical biopsy or resection for thyroid cancer and delay in diagnosis.

Timing

The time frame for evaluating the performance of the test varies from the initial FNA to surgical resection to weeks to months following an indeterminate result.

Setting

The primary setting would be in outpatient endocrinology.

ThyroSeq v2 Test

Technically Reliable

Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid

A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

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

Variants Association with Malignancy: Fnais et al (2015) conducted a systematic review and meta-analysis of studies reporting on the test accuracy of *BRAF* variant testing in the diagnosis of PTC.[51] Reviewers included 47 studies with 9924 FNA samples. For all cytologically indeterminate nodules, the pooled sensitivity estimate for *BRAF* variant testing was 31% (95% CI, 6% to 56%). Among nodules suspicious for malignancy on FNA, the pooled sensitivity estimate for *BRAF* variant testing was 52% (95% CI, 39% to 64%; I^2=77%).

Ferraz et al (2011) evaluated 20 publications that reported on the type and number of variants in cases of FNA of the thyroid diagnosed as indeterminate and compared the results with final histology after surgical resection.[52] Sixteen studies analyzed a single variant (eg, *BRAF* variant or *RET/PTC* rearrangement) and four analyzed a panel of variants (*BRAF* and *RAS* variants, *RET/PTC* and *PAX8/PPARγ* rearrangements). The detection of a variant in a histologically (surgically resected) benign thyroid lesion was categorized as a false-positive case, detecting no variant in an FNA sample from a histologically benign surgical sample was considered a true-negative, and finding no variant in a histologically malignant lesion was categorized as a false-negative. Based on 4 studies that examined a panel of variants, there was a broad sensitivity range (38%-85.7%; mean, 63.7%), a mean specificity of 98% (range, 95%-100%), a mean false-positive rate of 1.25% (range, 0%-4%), and a mean false-negative rate of 9% (range, 1%-21%). Based on 2 studies that examined *RET/PTC* rearrangements, mean sensitivity was 55% (range, 50%-60%), specificity 100%, a false-positive rate of 0%, and mean false-negative rate 3.5% (1%-6%). Based on 3 studies that examined *BRAF* variants, mean sensitivity was 13% (range, 0%-37.5%), mean specificity was 92.3% (range, 75%-100%), the mean false-positive rate was 0.5% (0%-1%), and the mean false-negative rate was 6% (range, 3%-12%). Authors concluded that testing for a panel of variants improved the sensitivity and specificity for indeterminate FNA of the thyroid but that further standardizations and further molecular markers would be needed before broad application of molecular FNA cytology for the diagnosis of thyroid nodules.

Additional studies describing the clinical validity of the genes that comprise the ThyroSeq panel or other individual variants and combinations of variants to predict malignancy in thyroid nodules that are indeterminate on FNA have been reported. The results that pertain to the use of gene testing in indeterminate thyroid nodules are summarized in Table 9. (In some cases, measures of agreement were calculated from data provided in the published article.)

Table 9. Clinical Validity of Molecular Markers to Predict Malignancy in Indeterminate Thyroid FNA Samples

Study	Population	Genes and Rearrangements Tested	Insufficient or Inadequate for Analysis	Measures of Agreement, %				
				Sen	Spec	PPV	NPV	Acc
Moses et al (2010)[53]	110 indeterminate thyroid nodules	*BRAF*, *KRAS*, *NRAS*, *RET/PTC1*, *RET/PTC3*, *NTRK1*	2	38	95	67	79	77
Ohori et al (2010)[54]	100 patients with 117 atypia or follicular lesions of uncertain significance	*BRAF*, *NRAS*, *HRAS*, *KRAS*, *RET/PTC1*, *RET/PTC3*, *PAX8/PPARγ*	NR	60	100	100	92	93
Cantara et al (2010)[55]	41 indeterminate and 54 suspicious thyroid nodules	*BRAF*, *H-K-NRAS*, *RET/PTC*, *TRK*, *PAX8-PPARγ*	53	86[a] 80[b]	97[a] 100[b]	86[a] 100[b]	97[a] 47[b]	95[a] 83[b]
Xing et al (2009)[56]	25 indeterminate, dominant nodules	*BRAF*	NR	14	100	100	48	52
Jara et al (2015)[57]	66 nodules suspicious for PTC	*BRAF*	NR	46	88	88	44	61
Rossi et al (2015)[58]	140 indeterminate or suspicious for malignancy or malignant nodules	*BRAF*	NR	90[c] 50[d] 40[e]	100[c] 100[d] 100[e]	100[c] 100[d] 100[e]	93[c] 69[d] 14[e]	96[c] 77[d] 46[e]
Beaudenon-Huibregtse et al (2014)[59]	53 nodules with indeterminate or nondiagnostic FNA	*BRAF*, *HRAS*, *KRAS*, *NRAS*, *PAX8-PPARγ*, *RET-PTC1*, *RET-PTC3*		48	89	81	64	

Acc: accuracy; FNA: fine needle aspiration; NPV: negative predictive value; NR: not reported; PPV: positive predictive value; PTC: papillary thyroid carcinoma; Sen: sensitivity; Spec: specificity.
[a] FNA-indeterminate nodules.
[b] FNA suspicious nodules.
[c] Atypia of indeterminate significance.
[d] Follicular neoplasm or suspicious for follicular neoplasm.
[e] Suspicious for malignancy.

NGS Panel (ThyroSeq): The largest body of literature on variant testing for prediction of malignancy in indeterminate thyroid nodules is related to the development an NGS panel (ThyroSeq) that includes *BRAF*, *RAS*, *RET/PTC*, or *PAX8/PPARγ.* Studies that address these panels are detailed below.

Nikiforov et al (2009) prospectively tested a panel of variants (*BRAF*, *RAS*) and rearrangements (*RET/PTC*, *PAX8/PPARγ*) in 470 FNA samples of thyroid nodules from 328 consecutive patients.[60] Variant status correlated with cytology and either surgical pathology diagnosis or follow-up (mean, 34 months). Forty patients were excluded for poor quality specimens, or loss to follow-up. Sixty-nine patients (with 86 thyroid FNA samples) underwent surgery soon after completing 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 variants and rearrangements were found (18 *BRAF*, 8 *RAS*, 5 *RET/PTC*, 1 *PAX8/PPARγ*); after surgery, 31 (97%) variant-positive nodules were diagnosed as malignant on pathologic examination and 1 (3%) as a benign tumor. Thirteen of the 32 variant-positive FNA samples had a definitive cytologic diagnosis of malignancy, whereas the rest were 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 variant found in the FNA. 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 (2010) performed variant screening in 117 FNA samples classified as AUS or FLUS.[54] *BRAF*, *RAS*, *RET/PTC*, or *PAX8/PPARγ* variants and rearrangements were detected in 10% of this category. The screening demonstrated that the probability of having a malignancy in this cytology category together with detection of one of the somatic variants investigated was 100%, whereas the probability of having a thyroid malignancy without a variant detected was 7.6%.

Nikiforov et al (2011) reported on results of a prospective study that assessed the clinical validity of a panel of variants to predict the likelihood of malignancy in thyroid nodules found indeterminate on FNA.[61] The authors included 1056 consecutive samples with indeterminate cytology on FNA that underwent variant testing, with 967 of those adequate for molecular analysis (653 AUS or FLUS; 247 follicular or Hürthle cell neoplasms or SFNs; 67 suspicious for malignant cells). Eighty-seven *BRAF*, *RAS*, *RET/PTC*, or *PAX8/PPARγ* variants and rearrangements were detected. At analysis, 479 patients had undergone thyroidectomy for further evaluation, providing a histopathologic diagnosis for 513 samples. The presence of a variant had a low sensitivity for predicting malignant histology (63%, 57%, and 68% for samples with AUS or FLUS, follicular or Hürthle cell neoplasms or SFNs, and suspicious for malignant cells on cytology, respectively), but a high specificity (99%, 97%, 96%, respectively). The NPVs for the variant analysis results were 94%, 86%, and 72% for samples with AUS or FLUS, follicular or Hürthle cell neoplasms or SFN, and suspicious for malignant cells on cytology, respectively. The authors concluded that variant analysis might be useful in surgical planning, such as determining whether patients should undergo a thyroid lobectomy or a complete thyroidectomy as a first surgery.

Subsequently, Nikiforov et al (2014) evaluated the accuracy of an NGS panel that included tests for SNVs in 13 genes (including *TERT*) and for 42 types of gene fusions (ThyroSeq v2 NGS panel) in a series of 143 consecutive thyroid FNA samples with a cytologic diagnosis of follicular or Hürthle cell neoplasm or suspicious for follicular or Hürthle cell neoplasm.[45] 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. In the prospective cohort, of the 14 malignant nodules, 11 were PTCs, and 3 were follicular thyroid carcinomas. Performance data for the ThyroSeq in both cohorts are shown in Table 10.

The authors noted that, compared with the gene panel used in their 2011 study, the NGS panel was associated with a marked increase in NPV, with a similar PPV. In this case, they proposed that the panel could be used to both "rule in" and "rule out" invasive cancers.

Nikiforov et al (2015) also reported on 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 AUS or FLUS (see Table 10).[62] Molecular analysis was performed prospectively in all patients. Ninety patients (96 nodules) 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. Two other patients were considered to have a definitive nonsurgical diagnosis of primary hyperparathyroidism based on biochemical testing.

Table 10. Performance of ThyroSeq Panel

	Nikiforov et al (2014)[45]			Nikiforov et al (2015)[62]
Variant Testing Outcomes	**Retrospective (n=91)**	**Prospective (n=52)**	**Overall (N=143)**	**Known Outcome (n=98)**
Negative	64 (2 cancer; 62 benign)	37 (2 cancer; 35 benign)		72 (2 cancer; 70 benign)
Positive	27 (23 cancer; 4 benign)	15 (12 cancer; 3 benign)		26 (20 cancer; 6 benign)
Sensitivity (95% CI), %	92	86	90 (80 to 99)	90.9 (78.8 to 100)
Specificity (95% CI), %	94	92	93 (88 to 98)	92.1 (86.0 to 98.2)
PPV (95% CI), %	85	80	83 (72 to 95)	76.9 (60.7 to 93.1)
NPV (95% CI), %	97	95	96 (92 to 95)	97.2 (78.8 to 100)

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

Nikiforov et al (2018) reported on the performance of ThyroSeq v3 with 112 genes.[63] The training sample included 238 surgically removed tissue samples consisting of 205 thyroid tissue samples representing all main types of benign and malignant tumors and nontumoral conditions. The validation sample included an independent set of 175 FNA samples of indeterminate cytology. Using the cutoff identified in the training set, the ThyroSeq v3 sensitivity was 98% (95% CI, 93% to 99%), specificity was 82% (95% CI, 72% to 89%), with accuracy of 91% (95% CI, 86% to 94%).

Additional studies describing the clinical validity of the ThyroSeq panel in external settings (outside of the institution where it was developed) have reported on the diagnostic performance to predict malignancy in thyroid nodules that are indeterminate on FNA have been reported (see Table 11). These studies differed from the previous studies in that noninvasive follicular thyroid neoplasm with papillary-like nuclear features was classified as not malignant for calculation of performance characteristics. (In some cases, measures of agreement were calculated from data provided in the published article.)

Table 11. Additional Clinical Validity Studies of ThyroSeq to Predict Malignancy in Indeterminate Thyroid FNA Samples

Study	**Population**	**Genes and Rearrangements Tested**	**Insufficient or Inadequate for Analysis**	**Measures of Agreement (95% CI), %**			
				Sen	**Spec**	**PPV**	**NPV**
Valderrabano et al (2017)[64]	190 indeterminate thyroid nodules	ThyroSeq v2 (60+ genes)	2	70 (46 to 88)	77 (66 to 85)	42 (25 to 61)	91 (82 to 97)
Taye et al (2018)[65]	156 indeterminate thyroid nodules	ThyroSeq v2 (60+ genes)	3	89 (52 to 100)	43 (29 to 58)	22 (10 to 38)	96 (78 to 99)

Acc: accuracy; FNA: fine needle aspiration; NPV: negative predictive value; PPV: positive predictive value; Sen: sensitivity; Spec: specificity.

Additional studies have reported on differences in variant frequency in malignant vs benign tumors, and reported on the sensitivity and specificity of gene testing in unselected populations (ie, all patients with nodules, rather than just those with indeterminate cytology).[66-68]

Genetic Variants Association with Tumor Behavior: As noted, the presence of *BRAF* or *TERT* variants is strongly associated with malignancy in thyroid nodule FNA samples. *BRAF* or *TERT* variants have also been associated with more aggressive clinicopathologic features in individuals diagnosed with PTC.

Adeniran et al (2011) assessed 157 cases with equivocal thyroid FNA readings (indeterminate and suspicious for PTC) or with a positive diagnosis for PTC and concomitant *BRAF* variant analysis.[1] The results of histopathologic follow-up 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* variant analysis. All PTCs with an extrathyroidal extension or aggressive histologic features were positive for a *BRAF* variant. The authors concluded that patients with an equivocal cytologic diagnosis and a *BRAF* V600E variant could be candidates for total thyroidectomy and central lymph node dissection.

Xing et al (2009) investigated the utility of *BRAF* variant testing of thyroid FNA specimens for preoperative risk stratification of PTC in 190 patients.[56] A *BRAF* variant in preoperative FNA specimens was associated with poorer clinicopathologic outcomes for PTC. Compared with the wild-type allele, a *BRAF* variant 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 or recurrence was seen in 36% of *BRAF* variant–positive patients and 12% of *BRAF* variant–negative patients, with an odds ratio (OR) of 4.16 (95% CI, 1.70 to 10.17; p=0.002). The PPV and NPV for preoperative FNA-detected *BRAF* variant to predict PTC persistence or recurrence were 36% and 88%, respectively, for all histologic subtypes of PTC. The authors concluded that preoperative *BRAF* variant testing of FNA specimens might provide a novel tool to preoperatively identify PTC patients at higher risk for extensive disease (extrathyroidal extension and lymph node metastases) and those more likely to manifest disease persistence or recurrence.

Yin et al (2016) reported on a systematic review and meta-analysis evaluating *TERT* promoter variants and aggressive clinical behaviors in PTC.[69] Eight eligible studies (total N=2035 patients; range, 30-507) were included. Compared with wild-type, *TERT* promoter variant status was associated with lymph node metastasis (OR=1.8; 95% CI, 1.3 to 2.5; p=0.001), extrathyroidal extension (OR=2.6; 95% CI, 1.1 to 5.9; p=0.03), distant metastasis (OR=6.1; 95% CI, 3.6 to 10.3; p<0.001), advanced TNM stages III or IV (OR=3.2; 95% CI, 2.3 to 4.5; p<0.001), poor clinical outcome (persistence or recurrence; OR=5.7; 95% CI, 3.6 to 9.3; p<0.001), and mortality (OR=8.3; 95% CI, 3.8 to 18.2; p<0.001).

ThyGenX Thyroid Oncogene Panel and ThyraMIR microRNA Classifier

Technically Reliable

Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid

Labourier et al (2015) evaluated the diagnostic algorithm combining a 17-variant panel with ThyraMIR on a cross-sectional cohort of thyroid nodules comprised of 109 FNA samples with AUS/FLUS or follicular neoplasm or SFN across 12 endocrinology centers.[70] A summary of the sensitivity and specificity of the combined test is listed in Table 12.

Table 12. Summary of Clinical Validity for 17-Variant Panel and ThyraMIR on FNA Samples

Groups	No. of Cases	Sensitivity	Specificity	PPV	NPV	Odds Ratio
Cohort (95% CI), %	109	89 (73 to 97)	85 (75 to 92)	74 (58 to 86)	94 (85 to 98)	44 (13 to 151)
AUS/FLUS (95% CI), %	58	94 (73 to 100)	80 (64 to 91)	68 (46 to 85)	97 (84 to 100)	68 (8 to 590)
FN/SFN (95% CI), %	51	82 (57 to 96)	91 (76 to 98)	82 (57 to 96)	91 (76 to 98)	48 (9 to 269)

Adapted from Labourier et al (2015).[70]

AUS: atypia of undetermined significance; CI: confidence interval; FLUS: follicular lesion of undetermined significance; FN: follicular neoplasm; FNA: fine needle aspiration; NPV: negative predictive value; PPV: positive predictive value; SFN: suspicious for a follicular neoplasm.

Clinically Useful

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials.

Direct evidence for the clinical utility for the ThyroSeq v2 test and the combined ThyGenX and ThyraMIR diagnostic testing algorithm is lacking.

Chain of Evidence: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

A chain of evidence may be constructed to infer potential clinical utility of the combined diagnostic testing algorithm. No studies using ThyGenX NGS panel in FNA samples were identified. However, available evidence has suggested that use of variant testing using NGS in thyroid FNA samples is generally associated with high specificity and PPV for clinically significant thyroid cancer. There is potential clinical utility for identifying malignancy with higher certainty on FNA if such testing permits better preoperative planning at the time of thyroid biopsy, potentially avoiding the need for a separate surgery. However, variant analysis does not achieve an NPV sufficiently high enough to identify which patients can undergo active surveillance over thyroid surgery. In the diagnostic algorithm that reflexes to the ThyraMIR after a negative ThyGenX

result, patients receiving reflex testing could identify who may undergo active surveillance over thyroid surgery. A single study using a 17-variant panel with ThyraMIR showed an NPV of 94%. Therefore, the high NPV of ThyraMIR has the potential to accurately predict benignancy and triage patients to active surveillance.

Section Summary: Molecular Markers to Rule Out and Rule in Malignancy
Evidence for the clinical validity of the ThyroSeq NGS panel comes from 3 studies at the institution where the study was developed (1 with both retrospective and prospective samples, 1 with prospective samples, 1 with retrospective samples) and 2 independent retrospective studies. In the 2 original clinical validity studies at the developing institution, the performance characteristics were similar for ThyroSeq v2 (sensitivity, 90%-91%; specificity, 92%-93%; PPV, 77%-83%; NPV, 96%-97%). In the 2 independent validation studies in which noninvasive follicular thyroid neoplasm with papillary-like nuclear features was categorized as not malignant, performance characteristics were lower and variable (sensitivity, 70%-89%; specificity, 43%-77%; PPV, 22%-42%; NPV, 91%-96%). One study has evaluated the clinical validity of the ThyroSeq v3 panel, reporting a sensitivity of 98% and specificity of 82%.

Evidence for the clinical validity of combined testing for miRNA gene expression using ThyraMIR and a targeted 17-variant panel comes from 2 retrospective studies using archived surgical specimens and FNA samples. One study combined a 17-variant panel with ThyraMIR testing on archived surgical specimens and resulted in a sensitivity of 85% and specificity of 95%. The second study combined a 17-variant panel (miR*Inform*) with ThyraMIR testing on FNA samples and resulted in a sensitivity of 89%, specificity of 85%, PPV of 74%, and NPV of 94%. No studies were identified that demonstrated the clinical validity of a combined ThyGenX and ThyraMIR test on FNA samples.

Direct evidence for the clinical utility for the ThyroSeq v2 test and the combined ThyGenX and ThyraMIR reflex testing is lacking. However, available evidence has suggested that testing for gene variants and rearrangements can predict malignancy and inform surgical planning decisions when the test is positive. Pooled retrospective and prospective clinical validation studies of ThyroSeq v2 have reported a combined NPV of 96% (95% CI, 92% to 95%) and PPV of 83% (95% CI, 72% to 95%) and might potentially assist in selecting patient to avoid surgical biopsy in negative and guide surgical planning if positive. The NPV of the ThyGenX to identify patients who should undergo active surveillance over thyroid surgery is unknown. In a reflex testing setting, the high NPV for a microRNA gene expression test used on the subset of patients with a negative result from a variant and gene rearrangement testing may provide incremental information in identifying patients appropriately for active surveillance, but improvements in health outcomes are still uncertain.

SUMMARY OF EVIDENCE

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive FNA sample testing with molecular markers to rule out malignancy and to avoid surgical biopsy or resection, the evidence includes a prospective clinical validity study with the Afirma GEC and a chain of evidence to support clinical utility. Relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. In a multicenter validation study, the Afirma GEC was reported to have a high NPV (range, 90%-95%). These results are supported by an earlier development and clinical validation study (Chudova et al [2010]), but the classifiers used in both studies do not appear to be identical. In other multicenter and multiple single-center

studies, there is suggestive evidence that rates of malignancy are low in Afirma patients who are benign, but the exact NPV is unknown. The available evidence suggests that the decisions a physician makes regarding surgery are altered by GEC results; however, it should be noted that long-term follow-up of patients with thyroid nodules who avoided surgery based on GEC results is limited. A chain of evidence can be constructed to establish the potential for clinical utility with GEC testing in cytologically indeterminate lesions, but with only a single study of the marketed test reporting a true NPV, the clinical validity is uncertain. For the RosettaGX Reveal test, no prospective clinical studies were identified. 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 rule in malignancy and to guide surgical planning, the evidence includes prospective and retrospective studies of clinical validity. Relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. Variant 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 have suggested that testing for a panel of genetic variants associated with thyroid cancer may allow for the appropriate selection of patients for surgical management with an initial complete thyroidectomy. Prospective studies in additional populations are needed to validate these results. The variant analysis does not achieve an NPV sufficiently high enough to identify which patients can undergo active surveillance over thyroid surgery. Although the presence of certain variants 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.

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive FNA sample testing with molecular markers to rule out malignancy and avoid surgical biopsy or to rule in malignancy for surgical planning, the evidence includes multiple retrospective and prospective clinical validation studies for the ThyroSeq v2 or v3 test and 2 retrospective clinical validation studies that used a predicate test 17-variant panel (miR*Inform*) test to the current ThyGenX and ThyraMIR. Relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. In a retrospective validation study on FNA samples, the 17-variant panel (miR*Inform*) test and ThyraMIR had a sensitivity of 89%, and an NPV of 94%. Pooled retrospective and prospective clinical validation studies of ThyroSeq v2 have reported a combined NPV of 96% and PPV of 83% in studies conducted at the institution developing the test but poorer performance at external institutions. No studies were identified demonstrating the diagnostic characteristics of the marketed ThyGenX. No studies were identified demonstrating evidence of direct outcome improvements. A chain of evidence for the ThyroSeq v2 test and combined ThyGenX and ThyraMIR testing would rely on establishing clinical validity. The evidence is insufficient to determine the effects of the technology on health outcomes.

CLINICAL INPUT FROM PHYSICIAN SPECIALTY SOCIETIES AND ACADEMIC MEDICAL CENTERS

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.

2017 Input
In response to requests, clinical input on 7 tests for molecular markers was received from 9 respondents, including 1 specialty society-level response, 1 physician from academic center, and 7 physicians from 2 health systems while this policy was under review in 2017. Based on the evidence and independent clinical input, the clinical input supports that the following indications provide a clinically meaningful improvement in the net health outcome and are consistent with generally accepted medical practice:

- Use of the following types of molecular marker testing in fine needle aspirate (FNA) of thyroid nodules with indeterminate cytologic findings (ie, Bethesda diagnostic category III [atypia/follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm/suspicion for a follicular neoplasm]) to rule out malignancy and to avoid surgical biopsy:
    - Afirma Gene Expression Classifier; or
    - ThyroSeq v2
- Use of the following type of molecular marker testing in FNA of thyroid nodules with indeterminate cytologic findings or Bethesda diagnostic category V (suspicious for malignancy) to rule in the presence of malignancy to guide surgical planning for the initial resection rather than a 2-stage surgical biopsy followed by definitive surgery:
    - ThyroSeq v2;
    - ThyraMIR microRNA/ThyGenX;
    - Afirma BRAF after Afirma Gene Expression Classifier; or
    - Afirma MTC after Afirma Gene Expression Classifier.

Based on the evidence and independent clinical input, the clinical input does not support whether the following indication provides a clinically meaningful improvement in the net health outcome or is consistent with generally accepted medical practice:

- Use of the following types of molecular marker testing in FNA of thyroid nodules:
    - RosettaGX Reveal.

2016 Input
In response to requests, input was received from 2 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 designed to with a high negative predictive value (NPV) in nodules indeterminate on fine needle aspirate (FNA). Although individual uses of a gene expression classifier with NPV in these situations varied, there was general agreement that the tests are considered standard in the evaluation of some indeterminate cases of FNA.

2013 Input
In response to requests, input was received from 1 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.

PRACTICE GUIDELINES AND POSITION STATEMENTS

American Association of Clinical Endocrinologists et al

In 2016, the American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazone Medici Endocrinologi updated their joint guidelines and made the following statements on molecular testing for cytologically indeterminate thyroid nodules[71]:

- "Cytopathology expertise, patient characteristics, and prevalence of malignancy within the population being tested impact the negative predictive values (NPVs) and positive predictive values (PPVs) for molecular testing."
- "Consider the detection of *BRAF* and *RET/PTC* and, possibly, *PAX8/PPARG* and *RAS* mutations if such detection is available."
- "*TERT* mutational analysis on FNA, when available, may improve the diagnostic sensitivity of molecular testing on cytologic samples."
- "Because of the insufficient evidence and the limited follow-up, we do not recommend either in favor of or against the use of gene expression classifiers (GECs) for cytologically indeterminate nodules."

For the role of molecular testing for deciding extent of surgery the following recommendations were made:

- "Currently, with the exception of mutations such as BRAFV600E that have a PPV approaching 100% for papillary thyroid carcinoma (PTC), evidence is insufficient to recommend in favor of or against the use of mutation testing as a guide to determine the extent of surgery."

American Thyroid Association

In 2016, the American Thyroid Association (ATA) updated its guidelines on the management of thyroid nodules and differentiated thyroid cancer in adults.[72] These guidelines made the following statements on molecular diagnostics in thyroid nodules that are atypia of undetermined significance (AUS) or follicular lesion of undetermined significance (FLUS) on cytology and follicular neoplasm (FN) or suspicious for follicular neoplasm (SFN) on cytology (see Table 11):

Table 11. Molecular Diagnostics in Thyroid Nodules That Are AUS or FLUS or FN or SFN on Cytology

Recommendation	SOR	QOE
AUS or FLUS		
"For nodules with AUS/FLUS cytology, after consideration of worrisome clinical and sonographic features, investigations such as repeat FNA [fine needle aspirate] 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	Moderate
"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	Low
FN or SFN		
"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	Moderate

AUS: atypia of undetermined significance; FLUS: follicular lesion of undetermined significance; FN: follicular neoplasm; FNA: fine needle aspirate; QOE: quality of evidence; SFN: suspicious for follicular neoplasm; SOR: strength of evidence.

The guidelines also stated: "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 (v.1.2018) comment on the use of molecular diagnostics in thyroid cancer.[73] For thyroid nodules evaluated with FNA, molecular diagnostics may be employed when lesions are suspicious for (category 2B recommendation):

- Follicular or Hürthle cell neoplasms.
- Atypia of undetermined significance or follicular lesion of undetermined significance.

The guidelines also state: "Molecular testing (both the Gene Expression Classifier and individual variant 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."

U.S. PREVENTIVE SERVICES TASK FORCE RECOMMENDATIONS

Not applicable.

ONGOING AND UNPUBLISHED CLINICAL TRIALS

Some currently unpublished trials that might influence this review are listed in Table 14.

Table 14. Summary of Key Trials

NCT No.	Trial Name	Planned Enrollment	Completion Date
Ongoing			
NCT02352766	Role of NGS-based ThyroSeq Panel in Cancer Diagnosis in Thyroid Nodules	300	Jun 2018
NCT03170804	Genomic Profiling of Nodular Thyroid Disease and Thyroid Cancer	200	Jan 2020

NCT: national clinical trial.

CODING

The following codes for treatment and procedures applicable to this policy are included below for informational purposes. Inclusion or exclusion of a procedure, diagnosis or device code(s) does not constitute or imply member coverage or provider reimbursement. Please refer to the member's contract benefits in effect at the time of service to determine coverage or non-coverage of these services as it applies to an individual member.

CPT/HCPCS

81345	TERT (telomerase reverse transcriptase) (eg, thyroid carcinoma, glioblastoma multiforme) gene analysis, targeted sequence analysis (eg, promoter region)
81445	Targeted genomic sequence analysis panel, solid organ neoplasm, DNA analysis, and RNA analysis when performed, 5-50 genes (eg, ALK, BRAF, CDKN2A, EGFR, ERBB2, KIT, KRAS, NRAS, MET, PDGFRA, PDGFRB, PGR, PIK3CA, PTEN, RET), interrogation for sequence variants and copy number variants or rearrangements, if performed
81479	Unlisted molecular pathology procedure

81545	Oncology (thyroid), gene expression analysis of 142 genes, utilizing fine needle aspirate, algorithm reported as a categorical result (eg, benign or suspicious)
0018U	Oncology (thyroid), microRNA profiling by RT-PCR of 10 microRNA sequences, utilizing fine needle aspirate, algorithm reported as a positive or negative result for moderate to high risk of malignancy
0026U	Oncology (thyroid), DNA and mRNA of 112 genes, next-generation sequencing, fine needle aspirate of thyroid nodule, algorithmic analysis reported as a categorical result ("Positive, high probability of malignancy" or "Negative, low probability of malignancy")

ICD-10 Diagnoses

C73	Malignant neoplasm of thyroid gland
D44.0	Neoplasm of uncertain behavior of thyroid gland

REVISIONS

03-17-2017	Policy added to the bcbsks.com web site on 02-15-2017 with an effective date of 03-17-2017.
10-01-2017	In Coding section: ▪ Added CPT code: 0018U.
04-15-2018	Updated Description section.
	In Policy section: ▪ In Item A, added "either", "or ThyroSeq v2", "nodules with", "cytologic findings (ie, Bethesda diagnostic category III [atypia", "or Bethesda diagnostic category IV", "suspicion for a follicular neoplasm", and removed "that are cytologically considered to be" to read, "The use of either the Afirma Gene Expression Classifier or ThyroSeq v2 in fine needle aspirates of thyroid nodules with indeterminate cytologic findings (ie, Bethesda diagnostic category III [atypia/follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm/suspicion for a follicular neoplasm]) may be considered medically necessary in patients who have ALL of the following characteristics." ▪ In Item B, removed "Mutation analysis in fine needle aspirates of the thyroid is experimental / investigational" and added "The use of any of the following types of molecular marker testing or gene variant analysis in fine needle aspirates of thyroid nodules with indeterminate findings (Bethesda diagnostic category III [atypia/follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm/suspicion for a follicular neoplasm]) or suspicious findings (Bethesda diagnostic category V [suspicious for malignancy]) to rule in malignancy to guide surgical planning for initial resection rather than a 2-stage surgical biopsy followed by definitive surgery may be considered medically necessary: 1. ThyroSeq v2; 2. ThyraMIR microRNA/ThyGenX; 3. Afirma BRAF after Afirma Gene Expression Classifier; or 4. Afirma MTC after Afirma Gene Expression Classifier. ▪ In Item C, added "genetic variant analysis, and molecular marker testing" and "including, but not limited to, use of RosettaGX Reveal" to read, "Gene expression classifiers, genetic variant analysis, and molecular marker testing in fine needle aspirates of the thyroid not meeting criteria outlined above, including, but not limited to, use of RosettaGX Reveal, are considered experimental / investigational." ▪ Updated Policy Guidelines.
	Updated Rationale section.
	In Coding section: ▪ Added CPT codes: 81445, 81479.

	Updated References section.
09-14-2018	Policy published on 08-15-2018 with an effective date of 09-14-2018.
	Updated Description section.
	In Policy section: ▪ In Item C, added "and single-gene *TERT* testing" to read, "Gene expression classifiers, genetic variant analysis, and molecular marker testing in fine needle aspirates of the thyroid not meeting criteria outlined above, including, but not limited to, use of RosettaGX Reveal and single-gene *TERT* testing, are considered experimental / investigational."
	Updated Rationale section.
	In Coding section: ▪ Added CPT code: 0026U.
	Updated References section.
	Updated Appendix section.
01-01-2019	In Coding section: ▪ Added new CPT code: 81345.

REFERENCES

1. Adeniran AJ, Theoharis C, Hui P, et al. Reflex BRAF testing in thyroid fine-needle aspiration biopsy with equivocal and positive interpretation: a prospective study. *Thyroid.* Jul 2011;21(7):717-723. PMID 21568726
2. Chudova D, Wilde JI, Wang ET, et al. Molecular classification of thyroid nodules using high-dimensionality genomic data. *J Clin Endocrinol Metab.* Dec 2010;95(12):5296-5304. PMID 20826580
3. Nikiforov YE, Seethala RR, Tallini G, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol.* Aug 1 2016;2(8):1023-1029. PMID 27078145
4. Nikiforov YE. Molecular diagnostics of thyroid tumors. *Arch Pathol Lab Med.* May 2011;135(5):569-577. PMID 21526955
5. Han PA, Kim HS, Cho S, et al. Association of BRAF mutation and microRNA expression with central lymph node metastases in papillary thyroid cancer: a prospective study from four endocrine surgery centers. *Thyroid.* Apr 2016;26(4):532-542. PMID 26950846
6. Yip L, Nikiforova MN, Yoo JY, et al. Tumor genotype determines phenotype and disease-related outcomes in thyroid cancer: a study of 1510 patients. *Ann Surg.* Sep 2015;262(3):519-525; discussion 524-515. PMID 26258321
7. Lin JD, Fu SS, Chen JY, et al. Clinical manifestations and gene expression in patients with conventional papillary thyroid carcinoma carrying the BRAF(V600E) mutation and BRAF pseudogene. *Thyroid.* May 2016;26(5):691-704. PMID 26914762
8. Alzahrani AS, Alsaadi R, Murugan AK, et al. TERT promoter mutations in thyroid cancer. *Horm Cancer.* Jun 2016;7(3):165-177. PMID 26902827
9. Landa I, Ganly I, Chan TA, et al. Frequent somatic TERT promoter mutations in thyroid cancer: higher prevalence in advanced forms of the disease. *J Clin Endocrinol Metab.* Sep 2013;98(9):E1562-1566. PMID 23833040
10. Liu X, Bishop J, Shan Y, et al. Highly prevalent TERT promoter mutations in aggressive thyroid cancers. *Endocr Relat Cancer.* Aug 2013;20(4):603-610. PMID 23766237
11. Liu T, Wang N, Cao J, et al. The age- and shorter telomere-dependent TERT promoter mutation in follicular thyroid cell-derived carcinomas. *Oncogene.* Oct 16 2014;33(42):4978-4984. PMID 24141777
12. Xing M, Liu R, Liu X, et al. BRAF V600E and TERT promoter mutations cooperatively identify the most aggressive papillary thyroid cancer with highest recurrence. *J Clin Oncol.* Sep 1 2014;32(25):2718-2726. PMID 25024077
13. Song YS, Lim JA, Choi H, et al. Prognostic effects of TERT promoter mutations are enhanced by coexistence with BRAF or RAS mutations and strengthen the risk prediction by the ATA or TNM

staging system in differentiated thyroid cancer patients. *Cancer.* May 1 2016;122(9):1370-1379. PMID 26969876
14. 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.* Nov 2013;98(11):E1852-1860. PMID 23979959
15. CBLPath. ThyroSeq® v.2 Next Generation Sequencing. n.d.; http://www.cblpath.com/products-and-services/test-menu-cblpath/item/1628-thyroseq-v2. Accessed May 25, 2018.
16. Barros-Filho MC, Marchi FA, Pinto CA, et al. High diagnostic accuracy based on CLDN10, HMGA2, and LAMB3 transcripts in papillary thyroid carcinoma. *J Clin Endocrinol Metab.* Jun 2015;100(6):E890-899. PMID 25867809
17. Zheng B, Liu J, Gu J, et al. A three-gene panel that distinguishes benign from malignant thyroid nodules. *Int J Cancer.* Apr 1 2015;136(7):1646-1654. PMID 25175491
18. Diggans J, Kim SY, Hu Z, et al. Machine learning from concept to clinic: reliable detection of braf v600e DNA mutations in thyroid nodules using high-dimensional RNA expression data. *Pac Symp Biocomput.* 2015;20:371-382. PMID 25592597
19. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med.* Aug 23 2012;367(8):705-715. PMID 22731672
20. Santhanam P, Khthir R, Gress T, et al. Gene expression classifier for the diagnosis of indeterminate thyroid nodules: a meta-analysis. *Med Oncol.* Feb 2016;33(2):14. PMID 26749587
21. Duh QY, Busaidy NL, Rahilly-Tierney C, et al. A systematic review of the methods of diagnostic accuracy studies of the Afirma Gene Expression Classifier. *Cytopathology.* Oct 2017;27(10):1215-1222. PMID 28741442
22. 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.* Apr 2014;20(4):364-369. PMID 24246351
23. Lastra RR, Pramick MR, Crammer CJ, et al. Implications of a suspicious afirma test result in thyroid fine-needle aspiration cytology: an institutional experience. *Cancer Cytopathol.* Oct 2014;122(10):737-744. PMID 25123499
24. 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.* Nov 2014;99(11):4069-4077. PMID 24780044
25. Yang SE, Sullivan PS, Zhang J, et al. Has Afirma gene expression classifier testing refined the indeterminate thyroid category in cytology? *Cancer Cytopathol.* Feb 2016;124(2):100-109. PMID 26422098
26. Witt RL. Outcome of thyroid gene expression classifier testing in clinical practice. *Laryngoscope.* Feb 2016;126(2):524-527. PMID 26343268
27. Baca SC, Wong KS, Strickland KC, et al. Qualifiers of atypia in the cytologic diagnosis of thyroid nodules are associated with different Afirma gene expression classifier results and clinical outcomes. *Cancer Cytopathol.* May 2017;125(5):313-322. PMID 28152275
28. Harrison G, Sosa JA, Jiang X. Evaluation of the Afirma Gene Expression Classifier in repeat indeterminate thyroid nodules. *Arch Pathol Lab Med.* Jul 2017;141(7):985-989. PMID 28467214
29. Kay-Rivest E, Tibbo J, Bouhabel S, et al. The first Canadian experience with the Afirma(R) gene expression classifier test. *J Otolaryngol Head Neck Surg.* Apr 4 2017;46(1):25. PMID 28372589
30. Hang JF, Westra WH, Cooper DS, et al. The impact of noninvasive follicular thyroid neoplasm with papillary-like nuclear features on the performance of the Afirma gene expression classifier. *Cancer Cytopathol.* Sep 2017;125(9):683-691. PMID 28544601
31. Samulski TD, LiVolsi VA, Wong LQ, et al. Usage trends and performance characteristics of a "gene expression classifier" in the management of thyroid nodules: An institutional experience. *Diagn Cytopathol.* Nov 2016;44(11):867-873. PMID 27534929
32. Celik B, Whetsell CR, Nassar A. Afirma GEC and thyroid lesions: An institutional experience. *Diagn Cytopathol.* Dec 2015;43(12):966-970. PMID 26466552
33. Alexander EK, Schorr M, Klopper J, et al. Multicenter clinical experience with the Afirma gene expression classifier. *J Clin Endocrinol Metab.* Jan 2014;99(1):119-125. PMID 24152684

34. Angell TE, Frates MC, Medici M, et al. Afirma Benign thyroid nodules show similar growth to cytologically benign nodules during follow-up. *J Clin Endocrinol Metab.* Nov 2015;100(11):E1477-1483. PMID 26353010
35. 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.* Oct 2012;22(10):996-1001. PMID 22873825
36. Sipos JA, Blevins TC, Shea HC, et al. Long-term nonoperative rate of thyroid nodules with benign results on the Afirma Gene Expression Classifier. *Endocr Pract.* Jun 2016;22(6):666-672. PMID 26789352
37. Aragon Han P, Olson MT, Fazeli R, et al. The impact of molecular testing on the surgical management of patients with thyroid nodules. *Ann Surg Oncol.* Jun 2014;21(6):1862-1869. PMID 24522987
38. Noureldine SI, Olson MT, Agrawal N, et al. Effect of gene expression classifier molecular testing on the surgical decision-making process for patients with thyroid nodules. *JAMA Otolaryngol Head Neck Surg.* Dec 1 2015;141(12):1082-1088. PMID 26606459
39. Abeykoon JP, Mueller L, Dong F, et al. The effect of implementing Gene Expression Classifier on outcomes of thyroid nodules with indeterminate cytology. *Horm Cancer.* Aug 2016;7(4):272-278. PMID 27102883
40. Chaudhary S, Hou Y, Shen R, et al. Impact of the Afirma Gene Expression Classifier result on the surgical management of thyroid nodules with category III/IV cytology and its correlation with surgical outcome. *Acta Cytol.* Jun 2016;60(3):205-210. PMID 27344463
41. Dhingra JK. Office-based ultrasound-guided FNA with molecular testing for thyroid nodules. *Otolaryngol Head Neck Surg.* Oct 2016;155(4):564-567. PMID 27246437
42. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. *Thyroid.* Nov 2009;19(11):1159-1165. PMID 19888858
43. Lithwick-Yanai G, Dromi N, Shtabsky A, et al. Multicentre validation of a microRNA-based assay for diagnosing indeterminate thyroid nodules utilising fine needle aspirate smears. *J Clin Pathol.* Jun 2017;70(6):500-507. PMID 27798083
44. Kloos RT, Monroe RJ, Traweek ST, et al. A genomic alternative to identify medullary thyroid cancer preoperatively in thyroid nodules with indeterminate cytology. *Thyroid.* Jun 2016;26(6):785-793. PMID 26992356
45. Nikiforov YE, Carty SE, Chiosea SI, et al. Highly accurate diagnosis of cancer in thyroid nodules with follicular neoplasm/suspicious for a follicular neoplasm cytology by ThyroSeq v2 next-generation sequencing assay. *Cancer.* Dec 1 2014;120(23):3627-3634. PMID 25209362
46. Liu R, Xing M. Diagnostic and prognostic TERT promoter mutations in thyroid fine-needle aspiration biopsy. *Endocr Relat Cancer.* Oct 2014;21(5):825-830. PMID 25121551
47. Decaussin-Petrucci M, Descotes F, Depaepe L, et al. Molecular testing of BRAF, RAS and TERT on thyroid FNAs with indeterminate cytology improves diagnostic accuracy. Dec 2017;28(6):482-487. PMID 29094776
48. Bernet V, Hupart KH, Parangi S, et al. AACE/ACE disease state commentary: molecular diagnostic testing of thyroid nodules with indeterminate cytopathology. *Endocr Pract.* Apr 2014;20(4):360-363. PMID 24727662
49. Yip L, Wharry LI, Armstrong MJ, et al. A clinical algorithm for fine-needle aspiration molecular testing effectively guides the appropriate extent of initial thyroidectomy. *Ann Surg.* Jul 2014;260(1):163-168. PMID 24901361
50. Ferris RL, Baloch Z, Bernet V, et al. American Thyroid Association statement on surgical application of molecular profiling for thyroid nodules: current impact on perioperative decision making. *Thyroid.* Jul 2015;25(7):760-768. PMID 26058403
51. Fnais N, Soobiah C, Al-Qahtani K, et al. Diagnostic value of fine needle aspiration BRAF(V600E) mutation analysis in papillary thyroid cancer: a systematic review and meta-analysis. *Hum Pathol.* Oct 2015;46(10):1443-1454. PMID 26232865

52. Ferraz C, Eszlinger M, Paschke R. Current state and future perspective of molecular diagnosis of fine-needle aspiration biopsy of thyroid nodules. *J Clin Endocrinol Metab.* Jul 2011;96(7):2016-2026. PMID 21593119
53. Moses W, Weng J, Sansano I, et al. Molecular testing for somatic mutations improves the accuracy of thyroid fine-needle aspiration biopsy. *World J Surg.* Nov 2010;34(11):2589-2594. PMID 20703476
54. Ohori NP, Nikiforova MN, Schoedel KE, et al. Contribution of molecular testing to thyroid fine-needle aspiration cytology of "follicular lesion of undetermined significance/atypia of undetermined significance". *Cancer Cytopathol.* Feb 25 2010;118(1):17-23. PMID 20099311
55. Cantara S, Capezzone M, Marchisotta S, et al. Impact of proto-oncogene mutation detection in cytological specimens from thyroid nodules improves the diagnostic accuracy of cytology. *J Clin Endocrinol Metab.* Mar 2010;95(3):1365-1369. PMID 20130073
56. Xing M, Clark D, Guan H, et al. BRAF mutation testing of thyroid fine-needle aspiration biopsy specimens for preoperative risk stratification in papillary thyroid cancer. *J Clin Oncol.* Jun 20 2009;27(18):2977-2982. PMID 19414674
57. Jara SM, Bhatnagar R, Guan H, et al. Utility of BRAF mutation detection in fine-needle aspiration biopsy samples read as "suspicious for papillary thyroid carcinoma". *Head Neck.* Dec 2015;37(12):1788-1793. PMID 24989827
58. Rossi M, Buratto M, Tagliati F, et al. Relevance of BRAF(V600E) mutation testing versus RAS point mutations and RET/PTC rearrangements evaluation in the diagnosis of thyroid cancer. *Thyroid.* Feb 2015;25(2):221-228. PMID 25333496
59. Beaudenon-Huibregtse S, Alexander EK, Guttler RB, et al. Centralized molecular testing for oncogenic gene mutations complements the local cytopathologic diagnosis of thyroid nodules. *Thyroid.* Oct 2014;24(10):1479-1487. PMID 24811481
60. 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.* Jun 2009;94(6):2092-2098. PMID 19318445
61. 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 sample. *J Clin Endocrinol Metab.* 2011;96(11):3390-3397. PMID 21880806
62. Nikiforov YE, Carty SE, Chiosea SI, et al. Impact of the multi-gene ThyroSeq Next-generation sequencing assay on cancer diagnosis in thyroid nodules with atypia of undetermined significance/follicular lesion of undetermined significance cytology. *Thyroid.* Nov 2015;25(11):1217-1223. PMID 26356635
63. Nikiforova MN, Mercurio S, Wald AI, et al. Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. *Cancer.* Apr 15 2018;124(8):1682-1690. PMID 29345728
64. Valderrabano P, Khazai L, Leon ME, et al. Evaluation of ThyroSeq v2 performance in thyroid nodules with indeterminate cytology. *Endocr Relat Cancer.* Mar 2017;24(3):127-136. PMID 28104680
65. Taye A, Gurciullo D, Miles BA, et al. Clinical performance of a next-generation sequencing assay (ThyroSeq v2) in the evaluation of indeterminate thyroid nodules. *Surgery.* Jan 2018;163(1):97-103. PMID 29154079
66. Mathur A, Weng J, Moses W, et al. A prospective study evaluating the accuracy of using combined clinical factors and candidate diagnostic markers to refine the accuracy of thyroid fine needle aspiration biopsy. *Surgery.* Dec 2010;148(6):1170-1176; discussion 1176-1177. PMID 21134548
67. Eszlinger M, Krogdahl A, Munz S, et al. Impact of molecular screening for point mutations and rearrangements in routine air-dried fine-needle aspiration samples of thyroid nodules. *Thyroid.* Feb 2014;24(2):305-313. PMID 23837487
68. Park KS, Oh YL, Ki CS, et al. Evaluation of the Real-Q BRAF V600E detection assay in fine-needle aspiration samples of thyroid nodules. *J Mol Diagn.* Jul 2015;17(4):431-437. PMID 25937618
69. Yin DT, Yu K, Lu RQ, et al. Clinicopathological significance of TERT promoter mutation in papillary thyroid carcinomas: a systematic review and meta-analysis. *Clin Endocrinol (Oxf).* Aug 2016;85(2):299-305. PMID 26732020

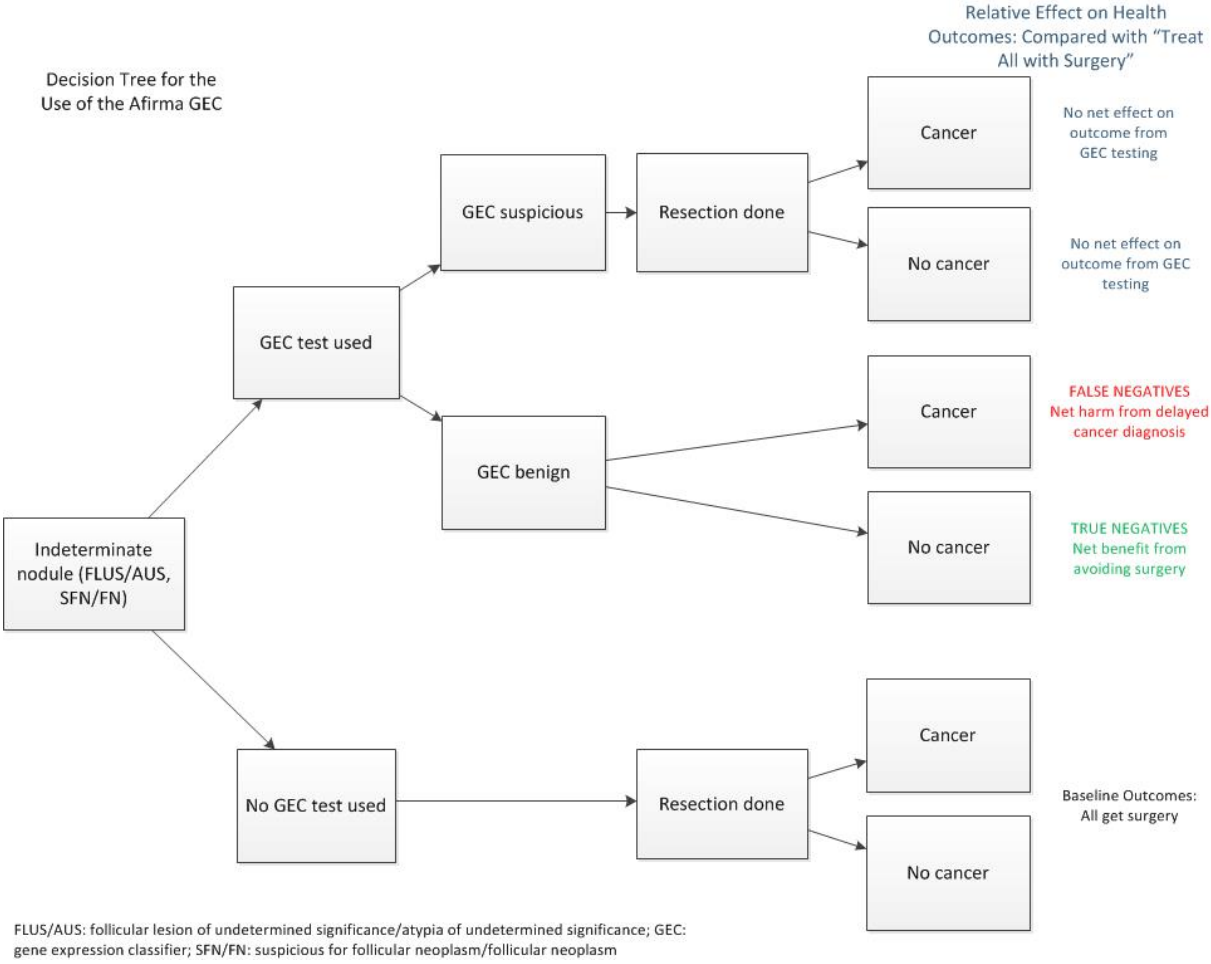
70. Labourier E, Shifrin A, Busseniers AE, et al. Molecular testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. *J Clin Endocrinol Metab.* Jul 2015;100(7):2743-2750. PMID 25965083
71. Gharib H, Papini E, Garber JR, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules--2016 Update. *Endocr Pract.* May 2016;22(5):622-639. PMID 27167915
72. Haugen BR, Alexander EK, Bible KC, et al. American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid.* Jan 2016;26(1):1-133. PMID 26462967
73. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology: Thyroid Carcinoma. Version 1.2018. https://www.nccn.org/professionals/physician_gls/pdf/thyroid.pdf. Accessed May 25, 2018.
74. Palmetto GBA. MolDX: Afirma Assay by Veracyte Coding and Billing Guidelines (M00015, V12). 2017; https://www.palmettogba.com/palmetto/MolDX.nsf/DocsCat/MolDx%20Website~MolDx~Browse%20By%20Topic~Covered%20Tests~8Q7MRU7038?open&navmenu=%7C%7C. Accessed May 25, 2018.

Other References

1. Blue Cross and Blue Shield of Kansas Pathology Liaison Committee, May 2018.

APPENDIX

Appendix Figure 1. Decision Model for the Afirma GEC Use



FLUS/AUS: follicular lesion of undetermined significance/atypia of undetermined significance; GEC: gene expression classifier; SFN/FN: suspicious for follicular neoplasm/follicular neoplasm