HHS Public Access

Author manuscript

Endocr Relat Cancer. Author manuscript; available in PMC 2020 December 29.

Published in final edited form as:

Endocr Relat Cancer. 2017 March ; 24(3): 127–136. doi:10.1530/ERC-16-0512.

Evaluation of ThyroSeq v2 performance in thyroid nodules with indeterminate cytology

Pablo Valderrabano[1], **Laila Khazai**[2], **Marino E Leon**[2,†], **Zachary J Thompson**[3], **Zhenjun Ma**[3], **Christine H Chung**[1], **Julie E Hallanger-Johnson**[1], **Kristen J Otto**[1], **Kara D Rogers**[1], **Barbara A Centeno**[2], **Bryan McIver**[1]

[1]Department of Head and Neck-Endocrine Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

[2]Department of Anatomic Pathology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

[3]Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

Abstract

ThyroSeq v2 claims high positive (PPV) and negative (NPV) predictive values in a wide range of pretest risks of malignancy in indeterminate thyroid nodules (ITNs) (categories B-III and B-IV of the Bethesda system). We evaluated ThyroSeq v2 performance in a cohort of patients with ITNs seen at our Academic Cancer Center from September 2014 to April 2016, in light of the new diagnostic criteria for non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP). Our study included 182 patients (76% female) with 190 ITNs consecutively tested with ThyroSeq v2. Patient treatment followed our institutional thyroid nodule clinical pathway. Histologies of nodules with follicular variant papillary thyroid carcinoma or NIFTP diagnoses were reviewed, with reviewers blinded to molecular results. ThyroSeq v2 performance was calculated in nodules with histological confirmation. We identified a mutation in 24% ($n = 45$) of the nodules. Mutations in *RAS* were the most prevalent ($n = 21$), but the positive predictive value of this mutation was much lower (31%) than that in prior reports. In 102 resected ITNs, ThyroSeq v2 performance was as follows: sensitivity 70% (46–88), specificity 77% (66–85), PPV 42% (25–61) and NPV 91% (82–97). The performance in B-IV nodules was significantly better than that in B-III nodules (area under the curve 0.84 vs 0.57, respectively; $P = 0.03$), where it was uninformative. Further studies evaluating ThyroSeq v2 performance are needed, particularly in B-III.

Correspondence should be addressed to P Valderrabano, pablo.valderrabano@moffitt.org.

†(M E Leon is now at Department of Pathology, College of Medicine, University of Florida, Gainesville, Florida, USA)

Declaration of interest

P V, L K, Z J T, Z M, C H C, K J O and K D R have nothing to disclose. B M and B A C receive grant support from GeneproDx. M E L received a sponsored research grant from Rosetta Genomics and holds equity in Rosetta Genomics.

Keywords

thyroid nodules; thyroid cancer; thyroid genetics; thyroid cytology; molecular diagnosis; non-invasive follicular thyroid neoplasms with papillary-like nuclear features

Introduction

Several molecular marker tests are available to refine the diagnosis of thyroid nodules with indeterminate cytology. Initially, these tests were classified as either 'rule-in' or 'rule-out' tests, depending on their ability to identify either cancerous or benign nodules before surgery, and their performance was highly dependent on the prevalence of malignancy (Ferris *et al.* 2015). A more recent generation of molecular tests, including ThyroSeq v2 (University of Pittsburgh Medical Center, Pittsburgh, PA. USA), claims to identify both cancerous and benign nodules with a very high level of confidence and to have a consistent performance over a wider range of prevalence of malignancy than previously available tests (Nikiforov *et al.* 2014, Labourier *et al.* 2015, Nikiforov *et al.* 2015). Despite a lack of independent validation studies, these newer tests are already being marketed and are actively changing the clinical management of ITNs. Moreover, it has recently been proposed to reclassify the encapsulated, non-invasive, follicular variant papillary thyroid carcinoma (FVPTC) as non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) (Nikiforov *et al.* 2016). This change in the nomenclature intentionally removes the word 'carcinoma' from the diagnosis to acknowledge the indolent behavior of these tumors. If NIFTPs are considered 'benign', the prevalence of malignancy among ITNs will decrease significantly, influencing the performance of molecular marker tests (Strickland *et al.* 2015, Faquin *et al.* 2016). A decrease in the pretest prevalence of malignancy will decrease the positive predictive value (PPV) and increase the negative predictive value (NPV) of all molecular tests. However, the specific impact on these tests will depend on the proportion of NIFTPs identified as 'benign' or 'malignant' by each test. In light of this new classification, we evaluated ThyroSeq v2 performance in ITNs (Bethesda category III (B-III), atypia/follicular lesion of undetermined significance or Bethesda category IV (B-IV), follicular/Hürthle cell neoplasm).

Materials and methods

Study cohort

Between September 2014 and April 2016, we evaluated 192 consecutive ITNs in 184 patients with ThyroSeq v2. Two specimens (1%) had complete test failure due to low cellularity and were excluded from the study. Demographic, clinical and pathological information was collected prospectively for 40 patients (22%) who consented to participate in an institutional review board–approved observational study initiated in July 2015 and was retrieved retrospectively under another institutional review board–approved study with waiver of consent for the remaining samples. We cataloged the following variables that might have an impact on molecular marker test results and/or on resection rates: age at cytological diagnosis; gender; presence of multinodular goiter (defined as the presence of at least 2 nodules >5 mm) or significant bilateral nodules (defined as the presence of at least

one contralateral nodule >1 cm); thyroid function test (normal = thyroid-stimulating hormone within reference range; hypothyroidism = thyroid-stimulating hormone above reference range, or patient on LT4 treatment; hyperthyroidism = thyroid-stimulating hormone below reference range or patient on anti-thyroidal drugs); nodule size (as determined by the largest diameter on ultrasound) and cytological diagnosis.

Pathology

All cytological specimens were interpreted by board-certified cytopathologists in the Department of Anatomic Pathology at Moffitt Cancer Center following the Bethesda system for reporting thyroid cytopathology, and each had a diagnosis of B-III or B-IV. Of the 190 patients with valid test results, 102 (54%) have been resected as of April 30th 2016. All but 2 (2%) of the histological diagnoses were rendered by head and neck pathology faculty at our institution. Although these pathologists were not blinded to the cytological diagnosis or molecular results, standard World Health Organization criteria were utilized in all cases. We began using the NIFTP designation in 2015 for encapsulated FVPTC after the 2015 United States and Canadian Academy of Pathology annual meeting (March 21–27, Boston, MA, USA), when the terminology was proposed in a companion meeting of the Endocrine Pathology Society (Tallini 2015). Because the definitive diagnostic criteria for NIFTP were not published until August 2016, the histology of all nodules originally diagnosed as FVPTC or NIFTP for which tissue was available, were reviewed by 2 pathologists (L K and B A C) blinded to the molecular results to ensure compliance with current diagnostic criteria. Slides were available for review in 7 of 8 tumors originally classified as NIFTP and 5 were reclassified as benign neoplasms upon review, as they were considered not to display sufficient nuclear atypia (score 1) to make this diagnosis (cytology was B-IV in 1 and B-III in 4; molecular markers were positive (*RAS* mutation) in 2 and negative in 3). All 6 tumors originally classified as FVPTC were reviewed and 2 were reclassified as NIFTP. Cytology–histology correlation was conducted by matching the location and size of the tumor from ultrasonographic scans and in the macroscopic description in the histology result. Only the histology of the biopsied nodule was used to calculate test performance.

Sample collection for ThyroSeq v2 and interpretation of results

Molecular marker tests for ITNs have been incorporated in our institutional thyroid nodule evaluation pathway since February 2014. Since that time, oncogene panel evaluation has been offered to all patients with ITNs who underwent biopsy at our institution. ThyroSeq v2 has been our chosen test since September 2014, due to its reported high sensitivity and specificity, yielding a theoretically superior performance in our institution compared to other tests, as previously reported (Valderrabano *et al.* 2016). The necessary specimens were collected in an identical fashion either at the radiology department or at the endocrine thyroid nodule clinic using 25 or 27 gauge needles, and the needle hub lavage of each aspirate (usually 5) was included in the media supplied by CBL Path (Rye Brook, NY, USA) for specimen preservation at the time of the ultrasound-guided biopsy. The specimen was stored in a freezer at −22°C until cytological diagnosis was available. If the cytology was indeterminate (B-III or B-IV), the specimen was shipped for analysis. Otherwise, it was discarded. The cytological and molecular specimens of 5 nodules (2.6%) were collected at an outside institution. ThyroSeq v2 results were considered positive when a mutation in

BRAF, *TP53*, *AKT1*, *CTNB1*, *PIK3CA*, *TERT* or *RET* genes was identified at an allelic frequency ≥5%; or ≥10% if the mutation was identified in *HRAS*, *KRAS*, *NRAS*, *PTEN*, *TSHR* or *EIF1AX* genes (Nikiforov *et al*. 2014, Nikiforov *et al*. 2015). Gene fusions were considered positive as previously described (Nikiforov *et al*. 2014, Nikiforov *et al*. 2015). Nodules with *MET* overexpression in which the estimated cancer risk was ~40% ($n = 2$); and nodules 'suspicious for ALK rearrangement' ($n = 1$) (due to disproportionately strong expression of the *ALK* tyrosine kinase domain over the extracellular domain), were also considered test-positive. Table 1 summarizes the molecular alterations identified by ThyroSeq v2 in our cohort. Specimens with detected somatic mutations below these thresholds or without molecular alterations identified were considered negative, as in prior publications (Nikiforov *et al*. 2014, Nikiforov *et al*. 2015). Parathyroid hormone (*PTH*) gene overexpression ($n = 1$), sodium–iodine symporter overexpression ($n = 2$) and specimens with *PTEN* mutation in patients with documented germ-line mutation (Cowden syndrome ($n = 2$)) were also considered negative.

Statistical analyses

Analysis cohort comparisons were made with Fisher's exact test for categorical variables and the Van der Waerden test for continuous variables. Logistic regression was used to test the association between the features and test results, and between the features and resection status. Odds ratios and 95% confidence intervals are shown for variables with a significant association. All *P* values are 2-sided unless otherwise stated and considered statistically significant at the 0.05 level. Sensitivity, specificity, NPV and PPV with exact binomial 95% confidence intervals were calculated for resected nodules only in 3 different cohorts: (1) all indeterminate thyroid nodules, (2) B-III specimens only and (3) B-IV specimens only. As NITFP are likely tumors in transformation on an adenoma-to-carcinoma sequence, these tumors do not clearly fit into a binary benign or malignant classification system. Recognizing this unsolved dilemma, we assessed the performance of ThyroSeq v2 in two different scenarios: (A) considering NIFTP 'malignant' (classical approach used in the validation studies for rule-in tests) and (B) considering NIFTP 'benign'. For ThyroSeq v2 performance comparisons between B-III and B-IV aspirates, ROC curves were calculated and the areas under the curves (AUC) were compared. The test statistic followed a chi-squared distribution: chi-squared = $(AUC_1 - AUC_2)^2/(s_1^2 + s_2^2)$, where AUC_1 and AUC_2 are the areas under the 2 independent ROC curves and s_1 and s_2 are their respective standard errors. All statistical analyses were performed using SAS (version 9.4, SAS Institute, Cary, NC, USA).

Results

The baseline characteristics of the 182 patients and 190 nodules tested with ThyroSeq v2 are shown in Table 2. The test was positive in 45 (23.7%) and negative in 145 (76.3%) nodules. There were no significant differences in any of the clinical characteristics between the test-positive and test-negative groups. Test-positive nodules were more frequently resected than test-negative nodules (73.3% vs 47.6%, $P = 0.0004$). Twelve patients with an identified mutation have not yet undergone thyroid surgery. Five patients with mutations less specific for cancer (3 with *TSHR* (1 B-IV and 2 B-III) and 2 with *EIF1AX* (1 B-IV and 1 B-III)

mutations) have elected follow-up with observation. Of the 7 patients with unresected nodules exhibiting more cancer-specific mutations (2 with *NRAS*, 1 with *NRAS* and *EIF1AX*, 2 with *HRAS*, 1 with *TERT* and 1 with *PAX8/PPARG*): 2 are scheduled for surgery; 1 had a concomitant unresectable lung cancer with decision not to operate on the thyroid and 4 have been lost to follow-up after having recommended surgery. Larger nodules were also more often resected ($P = 0.048$; OR 1.26 (1.00–1.58)), but the association between size and resection rates was seen only in the test-negative nodules ($P = 0.02$; OR 1.36 (1.05–1.76)), not in the test-positive group. The association between all other analyzed variables and rate of resections did not reach significance in any of the cohorts.

Twenty percent of the 102 resected nodules were malignant: 5 NIFTP (25%), 4 FVPTC (20%), 5 conventional variant of papillary thyroid carcinoma (CVPTC) (25%), 1 oncocytic variant of papillary thyroid carcinoma (5%), 4 minimally invasive follicular thyroid carcinomas (FTC) (20%), 1 of them oncocytic variant and one medullary thyroid carcinoma (5%) with an atypical *BRAF L597V* mutation and calcitonin gene overexpression. No mutations were identified in 5 of the malignancies (25%) (2 NIFTP, 2 CVPTC and 1 minimally invasive FTC), and a *KRAS* mutation was identified at a low (8%) allelic frequency in one FVPTC that was neither encapsulated nor well circumscribed, therefore, not meeting the criteria for NIFTP diagnosis (Table 3). *RAS* mutations were the most common genetic alteration, identified in 21 (47%) of the 45 patients with a positive result (*NRAS* in 14, *HRAS* in 4 and *KRAS* in 3). Of the 16 *RAS*-mutant nodules resected, 5 (31%) were malignant (1 NIFTP, 2 FVPTC, 1 CVPTC and 1 minimally invasive FTC), falling to 4 (25%) if we consider the NIFTP ‘benign’. Table 4 shows the distribution of positive and negative molecular results by histopathologic diagnosis.

The overall performance of ThyroSeq v2 is described in the Figure 1 (cohort 1). The sensitivity, specificity, PPV and NPV rates considering NIFTP malignant (scenario A) are 70%, 77%, 42% and 91%, respectively. If NIFTP was considered ‘benign’ (scenario B), the prevalence of malignancy fell from 20% to 15%, and the sensitivity, specificity, PPV and NPV are 73, 75, 33 and 94%, respectively.

The performance of ThyroSeq v2 was significantly better ($P = 0.03$) in B-IV than that in B-III specimens in both scenarios (A and B), where the AUC of the ROC curves were 0.84 vs 0.57 and 0.85 vs 0.55. ThyroSeq v2 was essentially uninformative for nodules with a B-III diagnosis (Fig. 1, cohort 2). A negative result in the B-III category did not reduce the prevalence of malignancy significantly in either of the scenarios (false-negative rate similar to prevalence of malignancy). Similarly, a positive result did not significantly increase the prevalence of malignancy, which was 19% in the best-case scenario (scenario A). The test was more informative in B-IV nodules (Fig. 1, cohort 3). The NPV was higher in B-IV than that in B-III specimens despite having a prevalence of malignancy twice that of B-III, achieving a false-negative rate of ~5% or less in both scenarios (Fig. 1, cohort 3). The PPV in B-IV was lower than that previously reported (65% and 53% in scenarios A and B, respectively) though a positive result increased the risk of malignancy 2.5-fold.

Discussion

We evaluated ThyroSeq v2 performance in ITNs and found that in our B-III cohort, the test was clinically uninformative, with no significant difference between the pretest and post-test rates of malignancy. In contrast, ThyroSeq v2 achieved a PPV of 65% and a NPV of 94% in B-IV specimens. In this subset of ITNs, the NPV may be high enough to consider observation in lieu of surgery (Nikiforov *et al.* 2014, NCCN 2015). These differences in ThyroSeq v2 performance with previous studies highlight the need for further studies across a range of institutions, to bring the performance of ThyroSeq v2 into a sharper focus.

Strengths and limitations of the study

This study is partly retrospective in nature, and the original histological diagnoses were not blinded to the cytological or molecular results, introducing the potential for ascertainment bias. Our conclusions are therefore somewhat limited by this design. However, this limitation applies equally to the original clinical validation studies for ThyroSeq v2 (Nikiforov *et al.* 2014, Nikiforov *et al.* 2015). In our study, all cytological reports and 98% of the histological diagnoses were issued by highly experienced endocrine pathologists. Nonetheless, we acknowledge that the expertise of the pathologists cannot eliminate the limitations of light microscopy, especially among follicular pattern lesions (Hirokawa *et al.* 2002, Lloyd *et al.* 2004, Elsheikh *et al.* 2008, Cibas *et al.* 2013). All cases included in this study were consecutively resected, and all molecular studies were performed prospectively before surgery.

Molecular markers have been incorporated in our institutional thyroid nodule evaluation pathway for several years. We have elected to use the results primarily to guide the extent of thyroidectomy, rather than to promote observation for negative results for 3 reasons: (1) the lack of independent validation studies confirming the NPV of ThyroSeq v2, (2) our prior experience with a 7-oncogene panel that revealed worse than predicted performance and (3) our relatively high institutional prevalence of malignancy in ITNs previously reported (~30%) (Valderrabano *et al.* 2016). Negative predictive values have been consistently under evaluated in independent validation studies of molecular marker tests due to their much lower rates of resection (typically ~10–20%) (Nishino 2016). In the original ThyroSeq v2 validation studies, the rate of resection of test-negative ITNs was 16% in the B-III study, and not reported in the B-IV study (Nikiforov *et al.* 2014, Nikiforov *et al.* 2015). We frequently offered diagnostic surgery to our patients with mutation-negative nodules, but treatment options were always discussed with the patient and resection rates were significantly lower when ThyroSeq v2 results were negative. However, almost 50% of ITNs with negative molecular test results were resected in our study, providing a better approximation to the true NPV of the test in our population.

The small sample size of our study (102 resected nodules) results in broad 95% confidence intervals, particularly when the B-III and B-IV specimens are analyzed individually. Nevertheless, our sample size is similar to that of the original validation studies, which included 143 and 98 resected nodules, respectively (Nikiforov *et al.* 2014, Nikiforov *et al.* 2015). Indeed, broad confidence intervals are seen in all of the extant studies of molecular

markers in ITNs, stressing the ongoing need for larger, independent, multicenter, clinical validation studies.

We blindly reviewed all the original FVPTC/NIFTP diagnoses, with the exception of one NIFTP, for which the specimen was not available. This review was performed for 2 reasons: (1) to ensure that the NIFTP diagnosis was made according to the recently published criteria (Nikiforov *et al.* 2016) and (2) to minimize any possible bias that knowledge of the molecular result might have had on the interpretation of follicular pattern lesions.

ThyroSeq v2 was uninformative in B-III specimens and the PPV lower than previously reported

ThyroSeq v2 was uninformative in our B-III cohort as the pretest prevalence of malignancy (13%) was not significantly modified by either a positive (19%) or negative (11%) result. Although we cannot conclude that the NPV in our study is significantly different from that in the previous publication (the 95% confidence intervals overlap: 74–97% in our B-III cohort vs 79–100% in the previous publication), the relatively high false-negative rate may preclude the use of observation in lieu of surgery if the test is negative (NCCN 2015, Nikiforov *et al.* 2015). In contrast, the NPV achieved in B-IV specimens (94%) was similar to that previously reported and high enough to justify observation in lieu of surgery in test-negative nodules (Nikiforov *et al.* 2014).

Our PPV was lower than reported in previous validation studies, particularly in the B-III cohort, where the difference was significant (95% confidence intervals do not overlap: 4–46% in our B-III cohort vs 61–93% in the previous publication) (Nikiforov *et al.* 2014, Nikiforov *et al.* 2015). These results are unlikely to change significantly with higher rates of resection of test-positive nodules. Five of 12 unresected ITNs with positive molecular results had mutations with low specificity for cancer (*TSHR* and *EIF1AX*), whereas the other 7 had more specific mutations. Even if we consider these 7 unresected nodules with more specific mutations to be truly malignant, the PPV would remain lower than that reported in prior validation studies (~53% overall in best-case scenario, scenario A) (Nikiforov *et al.* 2014, Nikiforov *et al.* 2015).

Differences between our study and the previous study in B-III nodules are likely to reflect differences in the cytological and molecular characteristics of the B-III cohorts. The B-III rate of malignancy in the current study (13%) was lower than that in our earlier publication (30%) and closer to reported national rates of malignancy for this category (Bongiovanni *et al.* 2012, Valderrabano *et al.* 2016). The reclassification of 5 NIFTPs as benign in this study and the implementation of a consensus review of all B-III diagnoses in 2014 to increase consistency and reduce the diagnostic rate of this category are likely responsible for these differences. However, the lower rate of malignancy alone is unlikely to explain the poor performance of ThyroSeq v2 among B-III specimens. In fact, a lower pretest prevalence of malignancy would be expected to improve the NPV, whereas in our study, the NPV was lower in the B-III than that in the B-IV category, despite the prevalence of malignancy being half that in B-IV.

The B-III group is a heterogeneous category by design, encompassing several scenarios that are associated with various risks of malignancy and likely different prevalence of mutations (Nishino & Wang 2014). We found a significantly higher rate of test-positive nodules among B-III specimens than that in the prior validation study (21% vs 7%) (Nikiforov *et al.* 2015). We hypothesize that this reflects differences in the cytological characteristics of the B-III cohorts that contribute to the discrepancy in ThyroSeq v2 performance. Some scenarios included in the B-III category, such as Hürthle cell predominance, are known to impair the performance of other molecular tests; and this or other scenarios might also affect ThyroSeq v2 performance (Harrell & Bimston 2014, Brauner *et al.* 2015). Future studies are required to clarify these questions.

Impact of NIFTP reclassification on the results interpretation

Upon review of histological specimens blinded to the molecular results, 5/7 (71%) NIFTPs were reclassified as benign neoplasms when a qualitative scale (scoring system) was applied (Nikiforov *et al.* 2016). This scoring system is still subject to personal interpretation but seems to improve the detection of mutation-positive tumors and standardizes the degree of nuclear atypia needed to make the diagnosis (Nikiforov *et al.* 2016). Had we relied on the original diagnoses, the prevalence of malignancy would be higher, but the test performance would not be significantly different. Under these circumstances, the sensitivity, specificity, PPV and NPV in scenario A (in which NIFTP is considered malignant) would be 64, 78, 48 and 87%, respectively, in the entire cohort; 45, 73, 31 and 83% in the B-III specimens and 79, 83, 65 and 91% in the B-IV specimens.

As reported by others, *RAS* mutations were the most common genetic abnormalities found in our series of ITNs (Nikiforov *et al.* 2014, Nikiforov *et al.* 2015). Mutations in *RAS* genes are traditionally associated with ~80% risk of malignancy (Gupta *et al.* 2013, Radkay *et al.* 2014). However, the most common malignancy among *RAS*-mutant tumors is the former encapsulated non-invasive FVPTC, recently reclassified as NIFTP (Gupta *et al.* 2013). If we only considered malignant tumors with invasive features, the rate of malignancy would drop from 83% to 33% in a previously reported series of 63 resected *RAS*-mutant tumors, similar to the rate observed in our series (25%) (Gupta *et al.* 2013). Only 1 (6%) of 16 resected *RAS* mutants had a sufficient degree of atypia to merit the NIFTP diagnosis in our series. It is unknown what proportion of the reported FVPTCs among resected *RAS* mutants in prior series would fulfill the recently proposed criteria for ‘significant atypia’ (Nikiforov *et al.* 2016). Altogether, it seems likely that differing NIFTP (nuclear atypia) diagnostic thresholds are one of the main factors in the test performance variability among different studies, given that the rate of invasive neoplasms is fairly consistent, at least among *RAS*-mutant tumors.

Of 33 resected mutation-positive nodules, 7 (26%) were hyperplastic/adenomatous nodules, a finding consistent with other series (Radkay *et al.* 2014). Although histopathological diagnosis of benign lesions depends on morphometric evaluation, oncogene panels offer information on clonality. Although the natural history of benign clonal lesions if left *in situ* is presently unknown, these may progress over time and acquire malignant potential. At one extreme of this spectrum are NIFTPs, which are considered surgical disease at this time, because invasiveness can only be excluded after adequate evaluation of the entire capsule

interface (Baloch *et al*. 2016). Although NIFTPs are considered as surgical disease, it might be reasonable to consider molecular marker test results as true positives or false negatives.

We identified 6 false-negative results in our series. In these nodules, cytohistological correlation is reliable due to concordance of size and location in 5 cases and to concordant location and absence of other nodules in the smallest one. In all cases, the epithelial cell control reported by ThyroSeq v2 was adequate, although it was reported borderline in 2, including one CVPTC with extrathyroidal extension that was biopsied and resected elsewhere. Two of these false-negative tumors were NIFTPs. Although the prognosis of these tumors is excellent once resected, their natural history if left *in situ* is as unknown as for mutation-positive NIFTPs, and therefore, we believe the same considerations regarding surgical resection and molecular marker test results interpretation should apply.

Conclusion

Our study suggests that ThyroSeq v2 is informative for nodules in the B-IV category, achieving a NPV robust enough to consider observation in lieu of surgery. However, the PPV of the test was lower than expected; and its performance particularly in B-III specimens may be variable among centers. In our institution, ThyroSeq v2 is unlikely to be informative in this category. These differing findings are likely to be explained by several factors including heterogeneity of the B-III category and differences in the diagnostic threshold for NIFTP between pathologists. These factors seem to affect the prevalence of mutation as well as the prevalence of malignancy within the population tested, therefore influencing the performance of the molecular tests and on the interpretation of their results. Validation studies at an institutional level may be necessary to adequately assess the predictive values of each test. Further studies evaluating the performance of ThyroSeq v2 in indeterminate thyroid nodules, particularly in B-III specimens, are needed.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References

Baloch ZW, Seethala RR, Faquin WC, Papotti MG, Basolo F, Fadda G, Randolph GW, Hodak SP, Nikiforov YE & Mandel SJ 2016 Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP): a changing paradigm in thyroid surgical pathology and implications for thyroid cytopathology. Cancer Cytopathology 124 616–620. (doi:10.1002/cncy.21744) [PubMed: 27203786]

Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L & Baloch ZW 2012 The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta Cytologica 56 333–339. (doi:10.1159/000339959) [PubMed: 22846422]

Brauner E, Holmes BJ, Krane JF, Nishino M, Zurakowski D, Hennessey JV, Faquin WC & Parangi S 2015 Performance of the Afirma gene expression classifier in Hurthle cell thyroid nodules differs from other indeterminate thyroid nodules. Thyroid 25 789–796. (doi:10.1089/thy.2015.0049) [PubMed: 25962906]

Cibas ES, Baloch ZW, Fellegara G, LiVolsi VA, Raab SS, Rosai J, Diggans J, Friedman L, Kennedy GC, Kloos RT, et al. 2013 A prospective assessment defining the limitations of thyroid nodule

pathologic evaluation. Annals of Internal Medicine 159 325–332. (doi:10.7326/0003-4819-159-5-201309030-00006) [PubMed: 24026318]

Elsheikh TM, Asa SL, Chan JK, DeLellis RA, Heffess CS, LiVolsi VA & Wenig BM 2008 Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. American Journal of Clinical Pathology 130 736–744. (doi:10.1309/AJCPKP2QUVN4RCCP) [PubMed: 18854266]

Faquin WC, Wong LQ, Afrogheh AH, Ali SZ, Bishop JA, Bongiovanni M, Pusztaszeri MP, VandenBussche CJ, Gourmaud J, Vaickus LJ, et al. 2016 Impact of reclassifying noninvasive follicular variant of papillary thyroid carcinoma on the risk of malignancy in the Bethesda system for reporting thyroid cytopathology. Cancer Cytopathology 124 181–187. (doi:10.1002/cncy.21631) [PubMed: 26457584]

Ferris RL, Baloch Z, Bernet V, Chen A, Fahey TJ 3rd, Ganly I, Hodak SP, Kebebew E, Patel KN, Shaha A, et al. 2015 American Thyroid Association Statement on Surgical Application of Molecular Profiling for Thyroid Nodules: current impact on perioperative decision making. Thyroid 25 760–768. (doi:10.1089/thy.2014.0502) [PubMed: 26058403]

Gupta N, Dasyam AK, Carty SE, Nikiforova MN, Ohori NP, Armstrong M, Yip L, LeBeau SO, McCoy KL, Coyne C, et al. 2013 RAS mutations in thyroid FNA specimens are highly predictive of predominantly low-risk follicular-pattern cancers. Journal of Clinical Endocrinology & Metabolism 98 E914–E922. (doi:10.1210/jc.2012-3396) [PubMed: 23539734]

Harrell RM & Bimston DN 2014 Surgical utility of Afirma: effects of high cancer prevalence and oncocytic cell types in patients with indeterminate thyroid cytology. Endocrine Practice 20 364–369. (doi:10.4158/EP13330.OR) [PubMed: 24246351]

Hirokawa M, Carney JA, Goellner JR, DeLellis RA, Heffess CS, Katoh R, Tsujimoto M & Kakudo K 2002 Observer variation of encapsulated follicular lesions of the thyroid gland. American Journal of Surgical Pathology 26 1508–1514. (doi:10.1097/00000478-200211000-00014)

Labourier E, Shifrin A, Busseniers AE, Lupo MA, Manganelli ML, Andruss B, Wylie D & Beaudenon-Huibregtse S 2015 Molecular testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. Journal of Clinical Endocrinology & Metabolism 100 2743–2750. (doi:10.1210/jc.2015-1158) [PubMed: 25965083]

Lloyd RV, Erickson LA, Casey MB, Lam KY, Lohse CM, Asa SL, Chan JK, DeLellis RA, Harach HR, Kakudo K, et al. 2004 Observer variation in the diagnosis of follicular variant of papillary thyroid carcinoma. American Journal of Surgical Pathology 28 1336–1340. (doi:10.1097/01.pas.0000135519.34847.f6)

National Comprehensive Cancer Network (NCCN) 2015 NCCN Clinical Practice Guidelines in Oncology: Thyroid carcinoma (version 2.2015). Fort Washington, PA, USA: NCCN (available at: https://www.nccn.org/professionals/physician_gls/pdf/thyroid.pdf)

Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, Gooding WE, Hodak SP, LeBeau SO, Ohori NP, et al. 2014 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 120 3627–3634. (doi:10.1002/cncr.29038) [PubMed: 25209362]

Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, Gooding WE, LeBeau SO, Ohori NP, Seethala RR, et al. 2015 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 25 1217–1223. (doi:10.1089/thy.2015.0305) [PubMed: 26356635]

Nikiforov YE, Seethala RR, Tallini G, Baloch ZW, Basolo F, Thompson LD, Barletta JA, Wenig BM, Al Ghuzlan A, Kakudo K, et al. 2016 Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. JAMA Oncology 2 1023–1029. (doi:10.1001/jamaoncol.2016.0386) [PubMed: 27078145]

Nishino M 2016 Molecular cytopathology for thyroid nodules: a review of methodology and test performance. Cancer Cytopathology 124 14–27. (doi:10.1002/cncy.21612) [PubMed: 26348024]

Nishino M & Wang HH 2014 Should the thyroid AUS/FLUS category be further stratified by malignancy risk? Cancer Cytopathology 122 481–483. (doi:10.1002/cncy.21412) [PubMed: 24634420]

Radkay LA, Chiosea SI, Seethala RR, Hodak SP, LeBeau SO, Yip L, McCoy KL, Carty SE, Schoedel KE, Nikiforova MN, et al. 2014 Thyroid nodules with KRAS mutations are different from nodules with NRAS and HRAS mutations with regard to cytopathologic and histopathologic outcome characteristics. Cancer Cytopathology 122 873–882. (doi:10.1002/cncy.21474) [PubMed: 25132659]

Strickland KC, Howitt BE, Marqusee E, Alexander EK, Cibas ES, Krane JF & Barletta JA 2015 The impact of noninvasive follicular variant of papillary thyroid carcinoma on rates of malignancy for fine-needle aspiration diagnostic categories. Thyroid 25 987–992. (doi:10.1089/thy.2014.0612) [PubMed: 26114752]

Tallini G 2015 Discussion of 'The Endocrine Pathology Society Conference for Re-examination of the Encapsulated Follicular Variant of Papillary Thyroid Cancer'. Presented at: United States & Canadian Academy of Pathology 104th Annual Meeting, Boston, MA, USA (available at: http://www.uscap.org/meetings/detail/2015-annual-meeting/sessions/1296).

Valderrabano P, Leon ME, Centeno BA, Otto KJ, Khazai L, McCaffrey JC, Russell JS & McIver B 2016 Institutional prevalence of malignancy of indeterminate thyroid cytology is necessary but insufficient to accurately interpret molecular marker tests. European Journal of Endocrinology 174 621–629. (doi:10.1530/EJE-15-1163) [PubMed: 26903551]

Cohort 1 (all ITN: B-III+B-IV)

A	Malignant (*n* = 20)	Benign (*n* = 82)
Positive (*n* = 33)	14	19
Negative (*n* = 69)	6	63

Sn 70% (46–88); Sp77% (66–85); PPV 42% (25–61); NPV 91% (82–97); Prevalence of malignancy 20%

B	Malignant (*n* = 15)	Benign (*n* = 87)
Positive (*n* = 33)	11	22
Negative (*n* = 69)	4	65

Sn73% (45–92); Sp75% (64–83); PPV 33% (18–52); NPV 94% (86–98); Prevalence of malignancy 15%

Cohort 2 (B-III category)

A	Malignant (*n* = 7)	Benign (*n* = 45)
Positive (*n* = 16)	3	13
Negative (*n* = 36)	4	32

Sn43% (10–82); Sp71% (56–84); PPV 19% (4–46); NPV 89% (74–97); Prevalence of malignancy 13%

B	Malignant (*n* = 5)	Benign (*n* = 47)
Positive (*n* = 16)	2	14
Negative (*n* = 36)	3	33

Sn40% (5–85); Sp70% (55–83); PPV 13% (2–38); NPV 92% (77–98); Prevalence of malignancy 10%

Cohort 3 (B-IV category)

A	Malignant (*n* = 13)	Benign (*n* = 37)
Positive (*n* = 17)	11	6
Negative (*n* = 33)	2	31

Sn85% (55–98); Sp84% (68–94); PPV 65% (38–86); NPV 94% (80–99); Prevalence of malignancy 26%

B	Malignant (*n* = 10)	Benign (*n* = 40)
Positive (*n* = 17)	9	8
Negative (*n* = 33)	1	32

Sn90% (56–100); Sp80% (64–91); PPV 53% (28–77); NPV 97% (84–100); Prevalence of malignancy 20%

Figure 1.

ThyroSeq v2 performance results. ITN indicates indeterminate thyroid nodules; B-III, atypia/follicular lesion of undetermined significance; B-IV, follicular/Hürthle cell neoplasm; Sn, sensitivity; Sp, specificity; NPV, negative predictive value; and PPV, positive predictive value. Test metrics were calculated on resected nodules only with exact binomial 95% confidence intervals in 2 different scenarios: NIFTP considered ‘malignant’ (A) and NIFTP considered ‘benign’ (B).

Table 1

Molecular alterations in test-positive nodules and associated risk of malignancy.

Molecular alteration	*n*	Cancer risk, % (n malignant/n resected)	Histological diagnosis
Point mutations			
BRAF K601E	2	50 (1/2)	H/AN, FVPTC
BRAF L597V	1	100 (1/1)	MTC
BRAF V600E	2	100 (2/2)	2 CVPTC
EIF1AX	5	0 (0/3)	H/AN, 2 FA
EIF1AX + NRAS Q61R	1	0 (0)	NA
EIF1AX + NRAS Q61K + TERT	1	100 (1/1)	FVPTC
EIF1AX + TSHR	1	0 (0/1)	H/AN
HRAS Q61K	1	100 (1/1)	FTC-MI
HRAS Q61R	3	0 (0/1)	H/AN
KRAS G12D	2	0 (0/2)	FA, HCA
KRAS G12V	1	0 (0/1)	H/AN
NRAS G13R	1	0 (0/1)	FA
NRAS Q61K	2	50 (1/2)	H/AN, CVPTC
NRAS Q61L	1	0 (0)	NA
NRAS Q61R	8	29 (2[a]/7)	H/AN, 4 FA, NIFTP, FVPTC
TERT	1	0 (0)	NA
TSHR	4	0 (0/1)	HCA
Gene fusions			
THADA/IGF2BP3	3	100 (3[a]/3)	2 NIFTP, FTC-MI
PAX8/PPARG	2	100 (1/1)	HCC-MI
Gene expression alterations			
MET overexpression	2	50 (1/2)	FA, OVPTC
ALK TK domain overexpression[b]	1	0 (0/1)	FA

[a]NIFTP considered ‘malignant’.

[b]For *ALK* TK domain overexpression: disproportional expression of the *ALK* tyrosine kinase domain as compared to the extracellular domain.

ALK TK, *ALK* tyrosine kinase; CVPTC, conventional variant of papillary thyroid carcinoma; FA, follicular adenoma; FTC-MI, follicular thyroid carcinoma minimally invasive; FVPTC, follicular variant of papillary thyroid carcinoma; HCA, Hürthle cell adenoma; H/AN, Hyperplastic/adenomatous nodule; HCC-MI, Hürthle cell carcinoma minimally invasive; MTC, medullary thyroid carcinoma; NA, not applicable; NIFTP, non-invasive follicular thyroid neoplasm with papillary-like nuclear features; OVPTC, oncocytic variant of papillary thyroid carcinoma.

Table 2

Baseline characteristics and clinical factors that could influence molecular marker tests results and rates of resection.

	All ITNs		Test positive		Test negative	
	Cohort, % (n)	**Resected, % (n)**	**Cohort, % (n)**	**Resected, % (n)**	**Cohort, % (n)**	**Resected, % (n)**
Patient level						
Total	100 (182)	53 (97)	25 (45)	71 (33)	75 (137)	47 (64)
Age (mean)	56.9	55.7	56.4	54.9	57.0	56.1
Male gender	24 (44)	45 (20)	16 (7)	43 (3)	27 (37)	46 (17)
Thyroid function						
Normal	80 (146)	53 (77)	76 (34)	73 (25)	82 (112)	46 (52)
Hypothyroidism	17 (31)	58 (18)	20 (9)	78 (7)	16 (22)	50 (11)
Hyperthyroidism	2 (3)	33 (1)	2 (1)	100 (1)	1 (2)	−(0)
N/A	1 (2)	50 (1)	2 (1)	−(0)	1 (1)	100 (1)
MNG[a]	66 (120)	48 (58)	69 (31)	65 (20)	65 (89)	43 (38)
Significant bilateral MNG[b]	32 (59)	54 (32)	27 (12)	67 (8)	34 (47)	51 (24)
Nodule level	100 (190)	54 (102)	24 (45)	73[c] (33)	76 (145)	48[c] (69)
Bethesda category						
III (AUS/FLUS)	55 (104)	50 (52)	49 (22)	73 (16)	57 (82)	44 (36)
IV (FN/HCN)	45 (86)	58 (50)	51 (23)	74 (17)	43 (63)	52 (33)
Size by ultrasound, cm	2.6	2.8[d]	2.5	2.5	2.6	2.9[d]
<2	39 (75)	47 (35)	33 (15)	73 (11)	41 (60)	40 (24)
2–3.9	49 (94)	57 (54)	58 (26)	77 (20)	47 (68)	50 (34)
≥4	11 (21)	62 (13)	9 (4)	50 (2)	12 (17)	65 (11)

Two patients had 1 nodule with mutation and ≥1 nodule without mutation; both are included in the test-positive cohort. Three nodules in this series measured ≤1 cm in largest diameter, and none had surgical follow-up. One had a *TSHR* mutation, another one had a BRAF-V600E mutation at low allelic frequency (4% and considered test negative), and no molecular alterations were identified in the other one. All other nodules were >1 cm. No association was found between any of the variables evaluated and the test result.

[a] MNG = at least 2 nodules >5 mm in the thyroid gland.

[b] Significant bilateral MNG = at least one nodule >1 cm in the contralateral lobe to the biopsied nodule.

[c] Nodules with a positive test result were more likely to be resected ($P = 0.0004$).

[d] Larger nodules were more likely resected ($P = 0.048$, OR 1.26 (1.00–1.58)), but only among test-negative nodules ($P = 0.02$, OR 1.36 (1.05–1.76)). Other variables were not significantly associated with the resection rates in any of the cohorts.

ITN, indeterminate thyroid nodule; MNG, multinodular goiter; N/A, not available.

Table 3

Characteristics of the false-negative specimens.

#	Age, years	Size-US, cm	MNG	ATA sonographic pattern	Bethesda category	Epithelial cell control	Molecular alteration identified	Histopathology	Size-path, cm
1	73	3.3	Yes	Heteroechoic with lobulated margins	IV	Good	None	NIFTP	3
2	71	3.7	Yes	Heteroechoic	III	Borderline	None	NIFTP	3
3	66	2.3	No	N/A	III	Borderline	None	CVPTC	2.3
4	36	6	Yes	Low-suspicion	IV	Good	None	FTC-MI (with vascular invasion)	5.4
5	42	1.1	No	High-suspicion	III	Good	None	CVPTC	0.6
6	62	3.2	No	Heteroechoic with lobulated margins	III	Good	*KRAS* (G12D)[a]	FVPTC	2.1

[a]Mutation was identified at 8% allelic frequency, putting it below the discriminative threshold.

ATA, American Thyroid Association; MNG, multinodular goiter; CVPTC, conventional variant papillary thyroid carcinoma; FTC-MI, minimally invasive follicular thyroid carcinoma; FVPTC, follicular variant papillary thyroid carcinoma; N/A, not available; NIFTP, non-invasive follicular thyroid neoplasm with papillary-like nuclear features; Size-path, size of the nodule measured on resected specimen; Size-US, largest dimension on ultrasound measurement.

Table 4

Distribution of positive and negative molecular results by histopathologic diagnosis.

Diagnosis (n) (n = 102)	Test positive, % (n) (n=33)	Test negative, % (n) (n = 69)
H/AN (27)	26 (7)	74 (20)
Adenoma (55)	22 (12)	78 (43)
NIFTP (5)	60 (3)	40 (2)
FVPTC (4)	75 (3)	25 (1)
Other PTC (6)	67 (4)	33 (2)
FTC (4)	75 (3)	25 (1)
MTC (1)	100 (1)	(0)

FTC, follicular thyroid carcinoma; FVPTC, follicular variant of papillary thyroid carcinoma; H/AN, hyperplastic/adenomatous nodule; MTC, medullary thyroid carcinoma; NIFTP, non-invasive follicular thyroid neoplasm with papillary-like nuclear features; PTC, papillary thyroid carcinoma.