CrossMark

REVIEW ARTICLE

Gene expression classifier for the diagnosis of indeterminate thyroid nodules: a meta-analysis

Prasanna Santhanam[1] · Rodhan Khthir[1] · Todd Gress[2] · Ayman Elkadry[1] · Omolola Olajide[1] · Abid Yaqub[3] · Henry Driscoll[1]

Abstract Prior studies demonstrate that a novel genomic test, the gene expression classifier (GEC), could identify a benign gene expression signature in those nodules with indeterminate cytology with a negative predictive value of greater than 95 %. Examine the performance of the AFIRMA gene expression classifier in predicting benign and malignant nodules in patients with cytologically indeterminate nodules. MEDLINE and EMBASE search for studies meeting eligibility criteria between January 1, 2005, and August 30, 2015. A total of 58 studies identified. After excluding duplicates, case reports, reviews, commentary, insufficient data, a total of seven studies selected for analysis. We combined individual patient data from seven studies that examined the GEC test for indeterminate thyroid nodules. The reference standard for determination of benign or malignant nodules was the histopathology of the thyroidectomy specimen. A QUADAS-2 report for all studies included in the final analysis was tabulated for risk of bias and applicability.

The pooled sensitivity of the GEC was 95.7 % (95 % CI 92.2–97.9, I^2 value 45.4 %, $p = 0.09$), and the pooled specificity was 30.5 % (95 % CI 26.0–35.3, I^2 value 92.1 %, $p < 0.01$). Overall, the diagnostic odds ratio was 7.9 (95 % CI 4.1–15.1). Patients with benign GEC were not followed long enough to ascertain the actual false-negative rates of the index test. Our meta-analysis revealed a high pooled sensitivity and a low specificity for the AFIRMA-GEC test for indeterminate thyroid nodules. This makes it an excellent tool to rule out malignancy.

Keywords Thyroid nodules · Gene expression classifier

✉ Prasanna Santhanam
prasanna@marshall.edu

1 Section of Endocrinology, Department of Internal Medicine, Byrd Clinical Center, Joan C Edwards School of Medicine, Marshall University, 1249 15th Street, Suite 3046, Huntington, WV 25701, USA

2 Department of Clinical and Translational Sciences and Department of Internal Medicine, Joan C Edwards School of Medicine, Marshall University, Huntington, WV 25701, USA

3 Division of Endocrinology, Department of Internal Medicine, University of Cincinnati School of Medicine, Cincinnati, OH, USA

Introduction

In a landmark research paper published in 2005, Jarzab et al. [1] identified more than 19 genes expressed in papillary thyroid cancer tissue and developed a potent molecular classifier that could discriminate between papillary thyroid cancer (PTC) and nonmalignant thyroid tissue in the majority of samples. Since then, a gene expression classifier (AFIRMA by Veracyte) was developed for commercial use and is considered the standard of care for the identification of indeterminate nodule(s) that are at higher risk of malignancy [2]. However, there are no large-scale studies examining the diagnostic accuracy of the test. This is due in large part to the fact that the rates of indeterminate biopsy result vary widely among institutions. Also, some tertiary care centers have reported contrasting results. We performed a systematic review and meta-analysis looking at the sensitivity and specificity of the gene expression classifier since no study has ever reported findings across different centers even though gene

expression classifier has now become the standard of care in the USA. We have also briefly looked at the clinical utility of the test in a brief review about alternative methods employed to distinguish benign and malignant results.

Methods

Data sources and search

We searched in MEDLINE and EMBASE for literature regarding gene expression classifier and indeterminate thyroid nodules. Different meSH terms for PubMed/MEDLINE as well as EMBASE were employed. A detailed method of our search meSH terms has been elucidated in the supplementary material (supplementary material table a). The details about the search in EMBASE are also outlined (supplementary material table b).

Study selection and eligibility criteria

The study search was conducted from January 1, 2005, to August 30, 2015. We included studies which had information on the performance of the AFIRMA gene expression classifier (AGEC; RNA method) for the diagnosis of indeterminate thyroid nodules.

Inclusion criteria for studies

- Indeterminate thyroid result on FNA that included Atypia of undetermined significance (AUS) or follicular lesion of undetermined significance (FLUS) Suspicious for follicular/Hurthle cell neoplasm Suspicious for malignancy
- Use of AFIRMA Veracyte gene expression classifier test as a screening tool
- Thyroidectomy performed as a reference standard in at least the cases where the index test was suspicious

Exclusion criteria for studies

- Opinions, reviews, case reports
- Lack of clear inclusion and exclusion criteria in the individual studies
- Absence of thyroidectomy as a reference standard
- Other methods of gene expression analysis for indeterminate nodule(s) like mRNA

We included both prospectively and retrospectively done studies. In the final meta-analysis, only studies which had information on the true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) rates were

included. The index test was the gene expression classifier test for indeterminate thyroid nodules. The reference standard was the histopathology of the thyroidectomy specimen. The Bethesda System for Reporting Thyroid Cytology: Atypia of undetermined significance (AUS) or follicular lesion of undetermined significance (FLUS), suspicious for follicular/Hurthle cell neoplasm, and suspicious for malignancy were all considered as indeterminate cytopathological result for the nodules [3].

Data extraction

Two authors (P.S. and R.K.) were involved in reviewing the literature from MEDLINE and screened the titles and abstract of the search results and retrieved all potentially relevant reports, and one author (A.Y.) reviewed literature from EMBASE. After selecting the studies that fulfilled the initial screening, the reviewers split into two different groups (P.S. and H.D. in group 1 and A.E., O.O., and RK in group 2) and independently reviewed the selected studies and screened the full texts to identify those that met the inclusion criteria.

Reviewers determined whether the index test and the reference test were performed according to the highest standards with regard to patient selection (random vs discrete), applicability (regional and/or tertiary centers involved), risk of bias (independent interpretation of the gene expression classifier as well as the thyroidectomy histopathological sample), and flow and timing of the assessment (GEC prior to thyroidectomy).

Quality review—QUADAS-2 reporting

Reviewers in groups 1 and 2 independently worked and analyzed the eligible articles to assess the reported quality of the methods using the QUADAS-2 reporting tool [4]. The QUADAS-2 tool is performed with 4 key steps: framing of the review question, creation of the tool and review-specific guidance, construction of the flow diagram, and evaluation of bias and applicability [4].

A detailed system of QUADAS-2 questionnaire was designed, and the responses on all the questions were obtained for each individual study by the group. The summary of the questionnaire is outlined below.

(1) *Patient selection* Risk of bias—Were the patients in the GEC group as well as the thyroidectomy group (for indeterminate thyroid nodules) both selected randomly? Applicability—Was the patient included in the studies from local and regional hospitals too or were they from higher centers where the cytopathology is excellent? (2) *Index test* Risk of bias—Was the GEC interpreted independently prior to performance of thyroidectomy? Was the result of the thyroid tissue cytopathology known prior to

GEC? Applicability—Are there differences in the way GEC tests were interpreted among the different tests used? (3) *Reference standard* Risk of bias—Was the reference standard interpreted without the knowledge of the GEC result? Applicability—Was the reference standard (cytopathology of specimen) performed at the highest standards? Were any small centers involved? (4) *Flow and timing* Risk of bias—Was there an appropriate interval between the GEC and thyroidectomy cytopathology? Did all patients receive thyroidectomy? Were patients without thyroidectomy included in the analysis?

All differences were resolved through consensus within the groups. If there was a difference in viewpoint between the groups, it was resolved by majority opinion.

A QUADAS-2 report for all the studies included in the final analysis was tabulated.

Data synthesis and statistical analysis

Using STATA (metandi command; Stata Corp., College Station, TX), we performed a meta-analysis of diagnostic test accuracy in which both the gene expression classifier and the thyroidectomy histopathological sample were considered as dichotomous variables (each having two outcomes—benign or suspicious/malignant). The degree of variability in study results was first evaluated by plotting the sensitivity and specificity from each study on a *forest plot*. The publication bias was evaluated using Eggers' test (funnel plot asymmetry that is measured by the intercept from standard normal regression) [5]. This method though useful has limited capacity to detect bias when meta-analysis is done with a small number of studies [5].

The method to find the averages of the logs of diagnostic odds ratio and the likelihood ratio has been described by DerSimonian–Laird [6]. This method also termed the random effects model (REM) which has been used to estimate the various parameters in different meta-analysis [7–9].

We used the meta-DISC software for estimating the pooled sensitivity, specificity, diagnostic odds ratio, and likelihood ratio using the methods described by Deeks et al. [10]. We tested for heterogeneity of data using the i-square as well as H which is the square root of the Chi-square heterogeneity statistic divided by its degrees of freedom [11].

Using the Moses–Shapiro–Littenberg to study *diagnostic odds ratio* variation according to threshold and fit symmetrical or asymmetrical curves, a single graph of sensitivity and specificity was created [summary receiver operating characteristic (SROC)] that was weighted by study size, and each data point represented a separate study [12].

A bivariate normal model (a two-level mixed logistic regression model, with separate binomial distributions for the true positives and true negatives depending on the sensitivity and specificity in each study) for the logit transforms of sensitivity and specificity between studies was created, using a generalized linear mixed model approach to meta-analysis of sensitivity and specificity [13, 14]. This method that has been described by Reitsma et al. [14] uses a sound and flexible approach in which explanatory variables may be added to the model and may lead to analysis of their independent effects on sensitivity and specificity. Prediction ellipses and confidence intervals within the ROC space were also created using the bivariate model [14].

Results

Literature search

A comprehensive PubMed/MEDLINE search was done using the following meSH terms -"AFIRMA and/or thyroid" (20 papers), "AFIRMA and/or indeterminate thyroid nodules" (18 papers), "gene expression classifier and/or thyroid" (48 papers), and "gene expression classifier and/or indeterminate thyroid nodules" (65 studies). EMBASE search session yielded the following results (in parenthesis)—Gene AND expression AND classifier AND thyroid AND nodules (58), AFIRMA AND indeterminate AND thyroid AND nodules (1), and Gene AND expression AND classifier AND indeterminate AND thyroid AND nodules (48). The details of search are in the flowchart (Fig. 1). The search terms and results are in supplementary table a and supplementary table b.

There were seven studies that qualified for our systematic review [15–21]. The details on the methodology and QUADAS-2 grading are outlined in Table 1. The characteristics of the studies included are shown in Table 2.

Meta-analysis results

The pooled sensitivity of the GEC (as estimated by the DerSimonian–Laird method) was 95.7 % (95 % CI 92.2–97.9, I^2 value 45.4 %, $p = 0.09$), and the pooled specificity was 30.5 % (95 % CI 26.0–35.3, I^2 value 92.1 %, $p < 0.01$). The pooled positive LR was 1.20 (95 % CI 0.996–1.44), and the pooled negative LR was 0.2 (95 % CI 0.11–0.36). Overall, the diagnostic odds ratio was 7.9 (95 % CI 4.1–15.1). The results of the meta-analysis are outlined in Table 3.

There was some degree of asymmetry in the funnel plot as estimated by Eggers' test, and this has been shown in Fig. 2. The coefficient of slope was −0.46 [standard error (SE) 3.68, $t = -0.12$, $p > (t) = 0.91$]. The bias was 0.65 [SE 1.22, $t = 0.53$, $p > (t) = 0.62$].

Fig. 1 Flowchart showing algorithm for screening and study selection
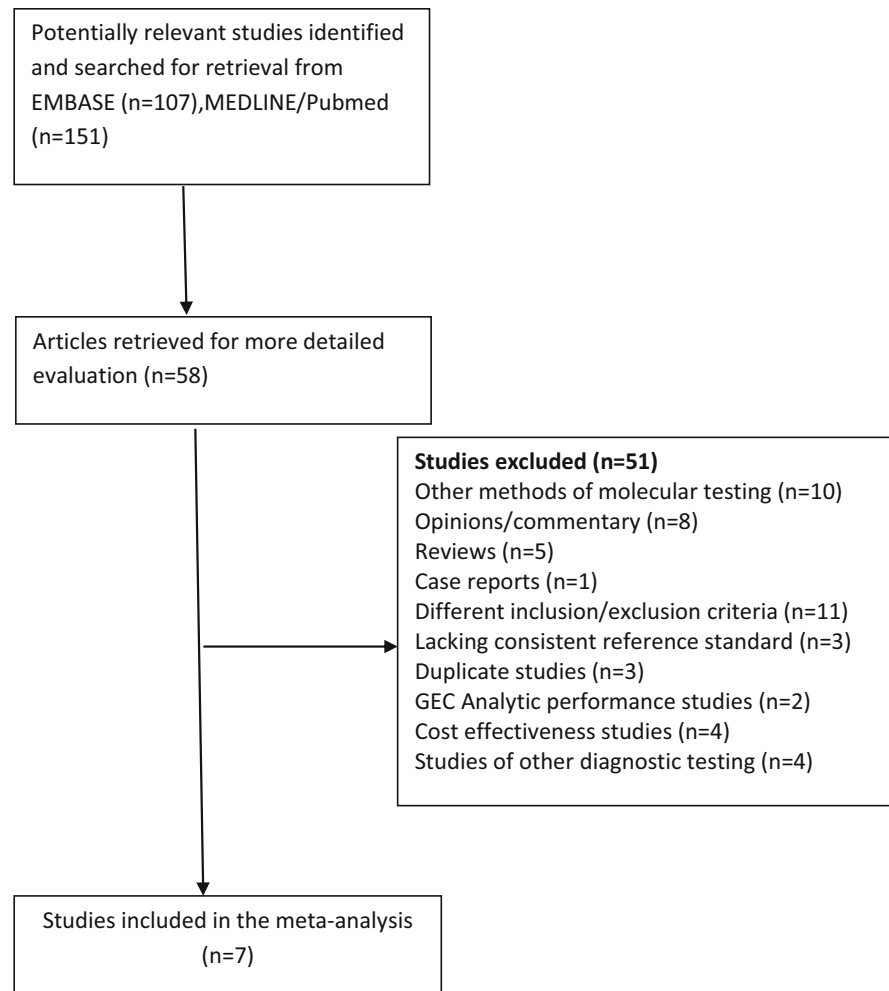
Flow diagram describing study selection

Potentially relevant studies identified and searched for retrieval from EMBASE (n=107),MEDLINE/Pubmed (n=151)

Articles retrieved for more detailed evaluation (n=58)

Studies excluded (n=51)
Other methods of molecular testing (n=10)
Opinions/commentary (n=8)
Reviews (n=5)
Case reports (n=1)
Different inclusion/exclusion criteria (n=11)
Lacking consistent reference standard (n=3)
Duplicate studies (n=3)
GEC Analytic performance studies (n=2)
Cost effectiveness studies (n=4)
Studies of other diagnostic testing (n=4)

Studies included in the meta-analysis (n=7)

Table 1 QUADAS-2 score reporting for all the studies included in the meta-analysis

Study ID	Risk of bias			Applicability			
	Patient selection	GEC test	Final histopathology	Flow risk	Patient selection	GEC test	Reference test Histopathology of thyroid sample
Alexander et al. [15]	Low	Low	Low	Unclear	Low	Low	High
McIver et al. [21]	Low	Low	Unclear	Low	Low	Low	Low
Harrell et al. [18]	Low	Low	High	Low	Unclear	Low	Low
Lastra et al. [19]	Low	Low	Unclear	Low	Low	Low	High
Marti et al. [20]	Low	Low	Low	Low	High	Low	Low
Han et al. [17]	Low	High	Low	Unclear	High	Unclear	High
Alexander et al. [16]	Low	Low	Unclear	Low	Low	Low	Low

The SROC curve has been shown in Fig. 3. The area under the curve (AUC) was 0.87 (SE 0.07). The forest plots showing the diagnostic odds ratio, positive likelihood ratio, negative likelihood ratio, pooled sensitivity as well as specificity have been shown in Fig. 4a–e.

The overall false-negative rate (1-sensitivity) was 0.04 (0.02–0.08), and the overall false-positive rate (1-specificity) was 0.69 (0.65–0.74).

The prediction ellipses for the HROC model have been shown in Fig. 5.

Table 2 Summary of the studies included in the analysis

Study ID	Site of study	GEC method	Research method
Alexander et al.	Local, regional, and tertiary	RNA	Prospective validation
McIver et al.	Tertiary	RNA	Prospective
Harrell et al.	Local center	RNA	Not known? Retrospective
Lastra et al.	Tertiary	RNA	Retrospective
Marti et al.	Tertiary	RNA	Retrospective
Han et al.	Tertiary	RNA, DNA, BRAF	Retrospective
EKA et al.	Tertiary	RNA	Retrospective

Table 3 Pooled sensitivities—confidence intervals and heterogeneity results (REM)

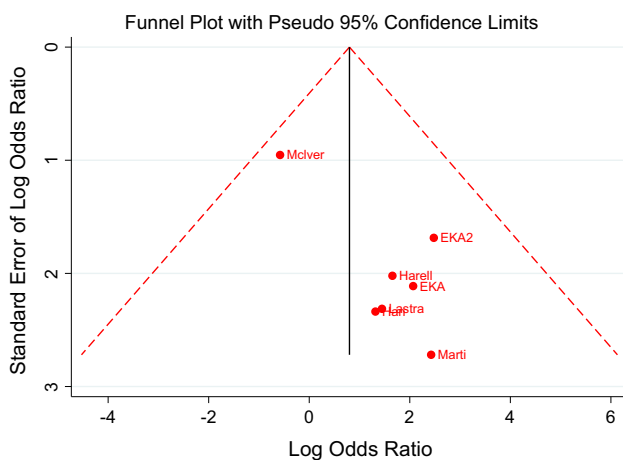| Test parameter (pooled values)	Value	Confidence interval	Heterogeneity Chi-square	p value	Inconsistency (I^2, %)
Sensitivity	0.957	0.922–0.979	10.99	0.09	45.40
Specificity	0.305	0.260–0.353	76.29	<0.01	92.10
Positive likelihood ratio	1.198	0.996–1.440	56.57	<0.01	89.40
Negative likelihood ratio	0.2	0.111–0.357	4.85	0.56	0.00
Diagnostic odds ratio	7.857	4.100–15.057	6	0.42	0.00

Fig. 2 Funnel plot outlining publication bias

The bivariate logistic regression showed a coefficient of 3.72 ($\pm$0.60) for sensitivity and 1.65 ($\pm$0.39) for specificity. The overall gradient-based log likelihood ratio was -28.14. This has been shown in Table 4 in detail. The hierarchical ordinal regression for ROC curves (HROC) showed $\gamma = 2.77\,(\pm 1.01)$, $\theta = 2.84\,(\pm 0.40)$, and $\beta = 0.50\,(\pm 0.61)$.

The covariance between estimates of E(logit Se) and E(logit Sp) was -0.09 in the HROC model.

Discussion

Fine-needle aspiration biopsy (FNAB) may not provide a definitive diagnosis (benign or malignant) for thyroid nodules [22]. The ambiguity is due to the fact that the sample of cells obtained from the procedure may not be
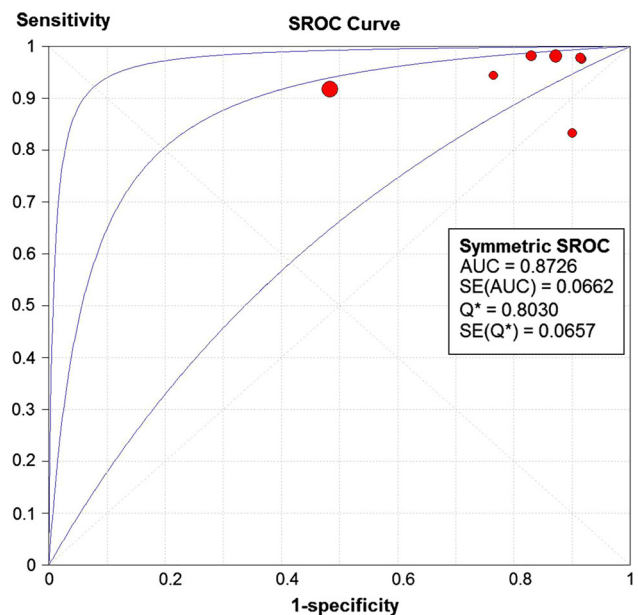

Fig. 3 Summary receiver operating characteristic (SROC) curve showing the overall result and area under the curve (AUC)

representative of the entire nodule. This is a particular problem with large nodules, leading some to recommend surgery for cytological benign nodules of size greater than 4 cm. In addition, the sample obtained by fine-needle aspiration does not provide information regarding capsular invasion, especially for follicular neoplasms. The Bethesda System for Reporting Thyroid Cytology (TBSRTC) has significantly improved the accuracy of malignancy detection from 25.9 to 53.7 % while lowering the rates of unnecessary thyroid surgeries [3]. Bethesda classification

Fig. 4 The forest plots showing the diagnostic odds ratio, positive likelihood ratio, negative likelihood ratio, pooled sensitivity as well as specificity
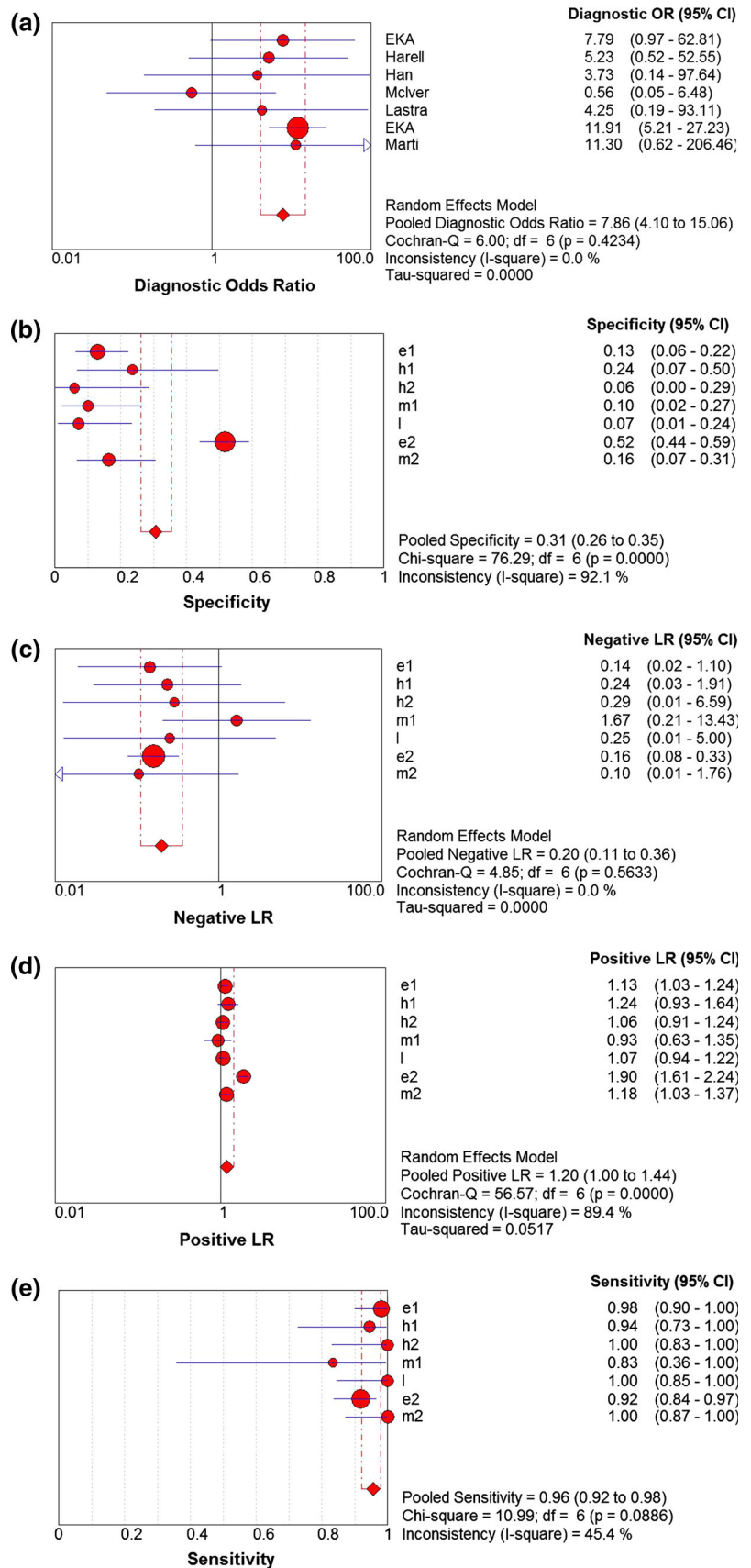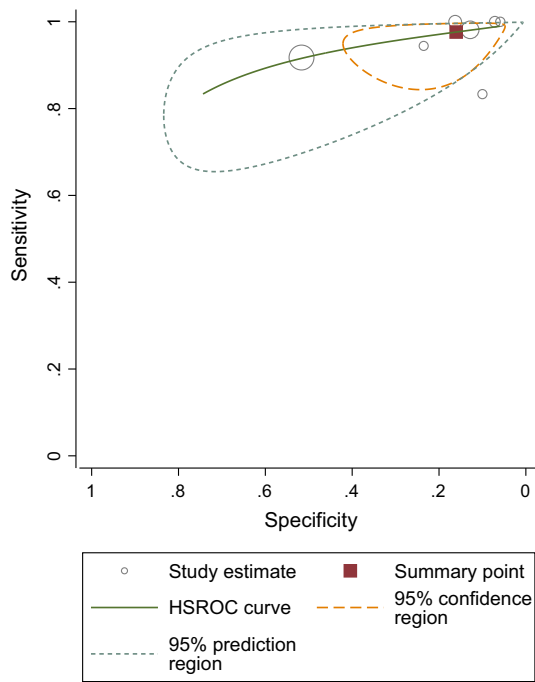
Fig. 5 The prediction ellipses of the HSROC model

Table 4 Coefficients in the bivariate model (HROC)

Bivariate parameter	Coefficient	Standard error	95 % CI
$E(logitSe)$	3.72	0.60	2.55–4.89
$E(logitSp)$	−1.65	0.39	2.42 to −0.88
Var(logitSe)	0.46	0.53	0.05–4.46
Var(logitSp)	0.77	0.51	0.21–2.80

for indeterminate FNAB has three subcategories: atypia of undetermined significance (AUS) or follicular lesion of undetermined significance (FLUS) with malignancy rate of 5–10 %, follicular neoplasm or suspicion for follicular neoplasm with malignancy rate of 20–30 %, and suspicious for malignancy with malignancy risk of 50–75 % [21]. Subsequent studies have found more variability in the diagnosis of malignancy mainly in patients with AUS or FLUS [22, 23]. In a study performed at a tertiary center of more than 1150 FNAs, the indeterminate thyroid nodules were reclassified into (1) indeterminate for neoplasia, (2) follicular neoplasm, and (3) suspicious for malignancy [23]. The corresponding malignancy rates on surgical excision were 6, 14, and 53 %, respectively [23].

Different methods have been studied to distinguish malignant nodules from benign ones in this heterogeneous group of cytological specimens. Studies showing the utility of core needle FNA for the diagnosis of indeterminate nodules have been performed [24]. In a multivariate logistic regression model, nodule calcification [odds ratio (OR) 6.37], size (OR 1.75), nuclear atypia on cytology (OR 4.91), and tobacco use (OR 4.59) were found to be significant predictors of malignancy in patients diagnosed with follicular neoplasm who underwent surgery [25]. A molecular technique of combining the expression of three different genes, cyclin D2 (CCND2), protein convertase 2 (PCSK2), and prostate differentiation factor (PLAB), has been shown to distinguish follicular carcinoma from follicular adenoma with a sensitivity of 100 % and a specificity of 94.7 % [26].

Radiologists have developed scoring systems (such as the Mayo, Park, and Kwak scores with area under the receiver operating characteristic (ROC) curve values of 0.827 for Park scores and 0.822 for Kwak scores) to identify the risk of malignancy in AUS/FLUS lesions [27]. The combination of real-time elastography and ultrasound has been shown to different cytology results with different malignancy rates [28, 29]. In the past, molecular testing for BRAF and RAS gene mutations and RET/PTC and PAX8/PPAR gamma gene rearrangements has shown the ability to separate high risk from low risk nodules in persons with AUS/FLUS [30]. However, even in an analysis of 1056 FNA samples with indeterminate cytology for multiple mutations (BRAF V600E, NRAS codon 61, HRAS codon 61, and KRAS codons 12/13 point mutations and RET/PTC1, RET/PTC3, and PAX8/PPAR gamma rearrangements), false-negative (mutation-negatives) rates were 6, 14, and 28 % for AUS/FLUS, follicular neoplasm, and suspicious for malignancy, respectively [31].

Our meta-analysis reveals a high pooled sensitivity and a low specificity (as reported in the results section) for the AFIRMA gene expression classifier test (AGEC) for indeterminate thyroid nodules, which makes it an excellent tool to rule out malignancy. The prevalence of malignancy was 37.1 % in the pooled cohort with a positive predictive value of 44.8 (95 % CI 40.4–49.4), suggesting that it is a very useful diagnostic test to rule out malignancy. Based on our data, one thyroid surgery could be avoided in patients with indeterminate FNAB for every two AGEC tests performed, assuming that >90 % of patients with benign AGEC are followed conservatively. The cost savings from including a genetic test like AGEC can be significant, realizing that the cost of hemithyroidectomy and total thyroidectomy performed at inpatient facilities often exceeds $20,000 in addition to the direct and indirect cost of complications [29]. Young et al. [30] reported that 11 % of patients undergoing thyroid surgery required a visit to the emergency room at least once within 30 days and nearly one quarter required a hospital admission.

There are several other factors to consider. According to the Surveillance, Epidemiology, and End Results (SEER) Program database (www.seer.cancer.gov), thyroid cancer accounts for only 3.8 % of all cancers and 0.3 % of all deaths due to cancer. The overall 5-year survival is greater

than 97.9 % (as of 2007) [32]. Interestingly, the overall 5-year survival has only marginally changed from 95.8 % in 1995 to 97.9 % in 2007 [32]. According to the American Cancer Society, the 5-year survival for stage I and stage II disease is nearly 100 % for follicular and papillary thyroid cancer (http://www.cancer.org/cancer/thyroidcancer/detailedguide/thyroid-cancer-survival-rates). Given the small number of persons with indeterminate cytology, the mortality as well as morbidity of this subgroup is likely to be very low. This was specifically reported in a study by Trimboli et al. [33] in endocrine. The diagnosis of suspicious and/or malignant nodules with the AGEC might quite simply represent a lead time bias with little change in overall survival. A study has suggested that molecular testing for indeterminate nodules only alters management in a small subset of persons [17]. Nevertheless, AGEC might have substantial benefits in lowering costs of hospitalization and surgery in light of the current practice of medicine [34].

The gene expression classifier may not perform well in Hurthle cell neoplasms, and a suspicious GEC result may not increase the probability of malignancy (25–35 % pretest rate) in this cytopathology subgroup [35]. Even more advanced multiplatform testings (with DNA, mRNA, and micro-RNA) are being used that may have a higher sensitivity and specificity of over 85 %, respectively [36].

Our study has limitations. Since there is heterogeneity in the different studies, the distribution of subgroups of indeterminate cytology (AUS, Hurthle cell neoplasm, follicular neoplasm, and suspicious for malignancy) might be different in the individual studies thereby resulting in different sensitivities and specificities. Since many of those with benign GEC do not undergo thyroidectomy, the true- and false-negative rate of GEC might be underestimated, especially since the progression and mortality of localized thyroid cancer are very low. It might take a few more years before the actual false-negative rates are assessed.

Conclusions

Gene expression classifier is a useful molecular test to rule out malignancy in thyroid nodules that have an indeterminate cytology, though the long-term benefits in reducing mortality and morbidity are unclear. The test might prevent some unnecessary thyroid surgeries, but for many persons it might represent an additional layer of testing prior to diagnostic thyroidectomy.

Overall GEC can be considered a risk assessment tool for a patient with follicular neoplasm, FLUS, and AUS on thyroid nodule FNA. Combining the results of the test with some other risk assessment modalities will probably be proven to increase the positive predictive value, which is

one of the limitations of the GEC (50–70 % with positive GEC results who undergo surgery will be proven to have benign thyroid nodule on surgery). The combined use of GEC and a specific genetic testing, which has high positive predictive value, will be a more cost-effective approach that may be used for these patients in the future.

Compliance with ethical standards

Conflict of interest   None.

References

1. Jarzab B, Wiench M, Fujarewicz K, Simek K, Jarzab M, Oczko-Wojciechowska M, et al. Gene expression profile of papillary thyroid cancer: sources of variability and diagnostic implications. Cancer Res. 2005;65(4):1587–97. doi:10.1158/0008-5472.can-04-3078.
2. Bernet V, Hupart KH, Parangi S, Woeber KA. AACE/ACE disease state commentary: molecular diagnostic testing of thyroid nodules with indeterminate cytopathology. Endocr Pract. 2014;20(4):360–3. doi:10.4158/ep14066.ps.
3. Hirsch D, Robenshtok E, Bachar G, Braslavsky D, Benbassat C. The implementation of the Bethesda system for reporting thyroid cytopathology improves malignancy detection despite lower rate of thyroidectomy in indeterminate nodules. World J Surg. 2015;. doi:10.1007/s00268-015-3032-6.
4. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. doi:10.7326/0003-4819-155-8-201110180-00009.
5. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ (Clin Res Ed). 1997;315(7109):629–34.
6. George BJ, Aban IB. An application of meta-analysis based on DerSimonian and Laird method. J Nucl Cardiol. 2015;. doi:10.1007/s12350-015-0249-6.
7. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. J Clin Epidemiol. 1991;44(2):127–39.
8. Ades AE, Lu G, Higgins JP. The interpretation of random-effects meta-analysis in decision models. Med Decis Mak. 2005;25(6):646–54. doi:10.1177/0272989x05282643.
9. Welton NJ, White IR, Lu G, Higgins JP, Hilden J, Ades AE. Correction: interpretation of random effects meta-analysis in decision models. Med Decis Mak. 2007;27(2):212–4. doi:10.1177/0272989x07300428.
10. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. BMJ (Clin Res Ed). 2001;323(7305):157–62.
11. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58. doi:10.1002/sim.1186.
12. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med. 1993;12(14):1293–316.

13. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol. 2006;59(12):1331–2. doi:10.1016/j.jclinepi.2006.06.011 (**author reply 2–3**).

14. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982–90. doi:10.1016/j.jclinepi.2005.02.022.

15. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. N Engl J Med. 2012;367(8):705–15. doi:10.1056/NEJMoa1203208.

16. Alexander EK, Schorr M, Klopper J, Kim C, Sipos J, Nabhan F, et al. Multicenter clinical experience with the Afirma gene expression classifier. J Clin Endocrinol Metab. 2014;99(1):119–25. doi:10.1210/jc.2013-2482.

17. Aragon Han P, Olson MT, Fazeli R, Prescott JD, Pai SI, Schneider EB, et al. The impact of molecular testing on the surgical management of patients with thyroid nodules. Ann Surg Oncol. 2014;21(6):1862–9. doi:10.1245/s10434-014-3508-x.

18. Harrell RM, Bimston DN. Surgical utility of Afirma: effects of high cancer prevalence and oncocytic cell types in patients with indeterminate thyroid cytology. Endocr Pract. 2014;20(4):364–9. doi:10.4158/ep13330.or.

19. Lastra RR, Pramick MR, Crammer CJ, LiVolsi VA, Baloch ZW. Implications of a suspicious afirma test result in thyroid fine-needle aspiration cytology: an institutional experience. Cancer Cytopathol. 2014;122(10):737–44. doi:10.1002/cncy.21455.

20. Marti JL, Avadhani V, Donatelli LA, Niyogi S, Wang B, Wong RJ, et al. Wide inter-institutional variation in performance of a molecular classifier for indeterminate thyroid nodules. Ann Surg Oncol. 2015;. doi:10.1245/s10434-015-4486-3.

21. McIver B, Castro MR, Morris JC, Bernet V, Smallridge R, Henry M, et al. An independent study of a gene expression classifier (Afirma) in the evaluation of cytologically indeterminate thyroid nodules. J Clin Endocrinol Metab. 2014;99(11):4069–77. doi:10.1210/jc.2013-3584.

22. Bhatia P, Deniwar A, Friedlander P, Aslam R, Kandil E. Diagnostic potential of ancillary molecular testing in differentiation of benign and malignant thyroid nodules. Anticancer Res. 2015;35(3):1237–41.

23. Nayar R, Ivanovic M. The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. Cancer. 2009;117(3):195–202. doi:10.1002/cncy.20029.

24. Choi WJ, Baek JH. Role of core needle biopsy for patients with indeterminate, fine-needle aspiration cytology. Endocrine. 2014;45(1):1–2. doi:10.1007/s12020-013-0071-3.

25. Macias CA, Arumugam D, Arlow RL, Eng OS, Lu SE, Javidian P, et al. A risk model to determine surgical treatment in patients with thyroid nodules with indeterminate cytology. Ann Surg Oncol. 2015;22(5):1527–32. doi:10.1245/s10434-014-4190-8.

26. Weber F, Shen L, Aldred MA, Morrison CD, Frilling A, Saji M, et al. Genetic classification of benign and malignant thyroid follicular neoplasia based on a three-gene combination. J Clin Endocrinol Metab. 2005;90(5):2512–21. doi:10.1210/jc.2004-2028.

27. Kamaya A, Lewis GH, Liu Y, Akatsu H, Kong C, Desser TS. Atypia of undetermined significance and follicular lesions of undetermined significance: sonographic assessment for prediction of the final diagnosis. J Ultrasound Med. 2015;34(5):767–74. doi:10.7863/ultra.34.5.767.

28. Garino F, Deandrea M, Motta M, Mormile A, Ragazzoni F, Palestini N, et al. Diagnostic performance of elastography in cytologically indeterminate thyroid nodules. Endocrine. 2015;49(1):175–83. doi:10.1007/s12020-014-0438-0.

29. Trimboli P, Treglia G, Sadeghi R, Romanelli F, Giovanella L. Reliability of real-time elastography to diagnose thyroid nodules previously read at FNAC as indeterminate: a meta-analysis. Endocrine. 2014;. doi:10.1007/s12020-014-0510-9.

30. Ohori NP, Nikiforova MN, Schoedel KE, LeBeau SO, Hodak SP, Seethala RR, et al. Contribution of molecular testing to thyroid fine-needle aspiration cytology of "follicular lesion of undetermined significance/atypia of undetermined significance". Cancer Cytopathol. 2010;118(1):17–23. doi:10.1002/cncy.20063.

31. Nikiforov YE, Ohori NP, Hodak SP, Carty SE, LeBeau SO, Ferris RL, et al. Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1056 FNA samples. J Clin Endocrinol Metab. 2011;96(11):3390–7. doi:10.1210/jc.2011-1469.

32. Howlader NA, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. SEER cancer statistics factsheets: thyroid cancer. National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/statfacts/html/thyro.html. SEER Cancer Statistics Review, 1975–2012, National Cancer Institute Bethesda, MD. 2012.

33. Trimboli P, Bongiovanni M, Rossi F, Guidobaldi L, Crescenzi A, Ceriani L, et al. Differentiated thyroid cancer patients with a previous indeterminate (Thy 3) cytology have a better prognosis than those with suspicious or malignant FNAC reports. Endocrine. 2015;49(1):191–5. doi:10.1007/s12020-014-0453-1.

34. Li H, Robinson KA, Anton B, Saldanha IJ, Ladenson PW. Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules. J Clin Endocrinol Metab. 2011;96(11):E1719–26. doi:10.1210/jc.2011-0459.

35. Brauner E, Holmes BJ, Krane J, Nishino M, Zurakowski D, Hennessey JV, et al. Performance of the afirma gene expression classifier in Hurthle cell thyroid nodules differs from other indeterminate thyroid nodules. Thyroid. 2015;. doi:10.1089/thy.2015.0049.

36. Labourier E, Shifrin A, Busseniers AE, Lupo MA, Manganelli ML, Andruss B, 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. 2015;. doi:10.1210/jc.2015-1158.