Predicting thyroid nodule malignancy at several prevalence values with a combined Bethesda-molecular test

HÉLÈNE LASOLLE, BENJAMIN RICHE, MYRIAM DECAUSSIN-PETRUCCI, EMMANUELLE DANTONY, VÉRONIQUE LAPRAS, CATHERINE CORNU, JOËL LACHUER, JEAN-LOUIS PEIX, JEAN-CHRISTOPHE LIFANTE, OANA-MARIA CAPRARU, SAMIA SELMI-RUBY, BERNARD ROUSSET, FRANÇOISE BORSON-CHAZOT[1], and PASCAL ROY[1]

BRON, LYON, VILLEURBANNE, PIERRE-BÉNITE, FRANCE, AND TIRGU MURES, ROMANIA

Investigation of thyroid nodules using fine-needle aspiration cytology (FNAC) gives indeterminate results in up to 30% of samples using the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC). We present a combined Bethesda-molecular predictor of nodule malignancy to improve the accuracy of the preoperative diagnosis of thyroid nodules. To detect a molecular signature of thyroid nodule malignancy, a molecular test was performed on FNACs from 128 thyroid nodules from prospectively included patients, collected in a tertiary center. The test relied on a transcriptomic array of 20 genes selected from a previous study. An optimal set of seven genes was identified using a logistic regression model. Comparison between the combined predictor (TBSRTC + molecular) and TBSRTC alone used the area under the ROC curve (AUC). Performance of the combined predictor was calculated according to various malignancy prevalence values and benefit-to-harm ratios (B/Hr) (favoring sensitivity or specificity). In our population (36% malignancy prevalence) and with a B/Hr of 1, the combined predictor achieved 95% specificity and 76% sensitivity. The AUC was 93.5%; higher than that of TBSRTC ($P = 0.004$). Among indeterminate nodules (30% malignancy prevalence), sensitivity and specificity were 52.2% and 96.2%, respectively, with a B/Hr of 1, or 95.7% and 64.2% with a B/Hr of 4 (favoring sensitivity), allowing avoidance of 64% of unnecessary surgeries at the cost of only one false-positive result. In conclusion, this predictor could improve the detection of thyroid nodule malignancy, taking into account malignancy prevalence and B/Hr, and reduce the number of unnecessary thyroidectomies. (Translational Research 2017;188:58–66)

[1] Both are last authors.

From the Fédération d'Endocrinologie, Hospices Civils de Lyon, Groupement Hospitalier Est, Bron, France; Université Lyon 1, Lyon, France; Service de Biostatistique, Hospices Civils de Lyon, Lyon, France; CNRS UMR5558, Laboratoire de Biométrie et Biologie Évolutive, Équipe Biostatistique-Santé, Villeurbanne, France; Service d'anatomie-pathologique, Hospices Civils de Lyon, Pierre-Bénite, France; Service de Radiologie, Centre Hospitalier Lyon-Sud, Hospices Civils de Lyon, Pierre-Bénite, France; INSERM, CIC1407, Bron, France; Service de Pharmacologie Clinique, Hospices Civils de Lyon, Bron, France; Centre de Recherche en Cancérologie de Lyon, INSERM, U 1052, Lyon, France; ProfileXpert, SFR santé Lyon Est, UMS 3453 CNRS – US7 INSERM, Lyon, France; Service de Chirurgie Digestive et Endocrinienne, Hôpital Lyon-Sud, Hospices Civils de Lyon, Pierre-Bénite, France; University of Medicine and Pharmacy, Tirgu Mures, Romania.

Submitted for publication March 6, 2017; revision submitted July 6, 2017; accepted for publication July 18, 2017.

Reprint requests: Hélène Lasolle, Département d'Endocrinologie - Fédération d'Endocrinologie, Hospices Civils de Lyon - Groupement hospitalier Est, 59 Boulevard Pinel 69677 Bron Cedex; e-mail: helene.lasolle@chu-lyon.fr.

1931-5244/$ - see front matter

http://dx.doi.org/10.1016/j.trsl.2017.07.005

Abbreviations: FNAC = fine-needle aspiration cytology; TBSRTC = The Bethesda System for Reporting Thyroid Cytopathology; DC = diagnostic category; GEC = gene expression classifier; NPV = negative predictive value; AUC = area under the ROC curve; B/Hr = benefit-to-harm ratio; RIN = RNA integrity number

AT A GLANCE COMMENTARY

Lasolle H, et al.

Background

Investigation of thyroid nodules using fine-needle aspiration cytology gives indeterminate results in up to 30% of samples using the Bethesda System for Reporting Thyroid Cytopathology. In such cases, surgery is often performed in order to obtain a definitive diagnosis. Improving the preoperative estimation of malignancy is necessary to reduce the number of avoidable surgical procedures and their potential deleterious effects.

Translational Significance

To improve the pre-operative diagnosis of thyroid nodules, we built a molecular test based on transcriptomic analysis searching for genes having their expression modified in thyroid cancers. A statistical model was fitted to be adapted in clinical practice, in various populations of patients.

INTRODUCTION

Thyroid nodules are quite common,[1] and radiology incidentalomas are now a common way of diagnosis.[2,3] The rate of malignancy of detected thyroid nodules has been estimated at 4%.[4]

Nodule evaluation and malignancy probability estimation rely mainly on ultrasonography and fine-needle aspiration cytology (FNAC).[4] Ultrasound, which provides information on nodule characteristics, is used to select suspicious nodules for FNAC, the cornerstone of preoperative nodule evaluation.[5-7]

The 2007 Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) is now commonly used for FNAC interpretation. This classification has 6 diagnostic categories (DCs) of increasing risk of malignancy.[8-10] It is very accurate in the malignant category (DC VI), in which surgery is indicated, and in the benign category (DC II), in which ultrasound follow-up is discussed, as Durante *et al.* proved the risk of progression was extremely low.[11] However, in up to 30% of FNACs, the sample is inadequate for cytology (DC1) or has an indeterminate risk of malignancy (DCs III to V).[9,12] In such cases, surgery is often performed to obtain a definitive diagnosis, although it may have deleterious effects.[13] Thus, improving the preoperative estimation of malignancy would dramatically reduce the number of avoidable surgical procedures.

Advances in the genetics of thyroid tumorigenesis have led to the development of molecular tests to detect structural genomic abnormalities (e.g., mutations) or abnormal gene expression.[14] Regarding genomic abnormalities, the discovery of the key role of the MAP-kinases pathway has led to the development of 7-gene mutation panels, which showed high specificity ("ruling in" strategy) but low sensitivity and limited therapeutic impact.[15-17] More recently, a next-generation sequencing panel of almost 60 genes has shown promising results.[18] Regarding gene expression, microarray technology for transcriptome analysis has focused on identifying molecular signatures to distinguish benign from malignant nodules. One gene expression classifier (GEC) developed to identify benign nodules in cases of indeterminate cytology ("ruling out" strategy) has shown 92% sensitivity and negative predictive values of 95% and 94% in DC III and DC IV, respectively.[19] However, put into practice, GEC was less useful, depending on the malignancy prevalence in the population of use.[14,17,20,21]

We describe herein the development of a combined Bethesda-molecular predictor that takes into account the prevalence of the disease and the benefit-to-harm ratio (B/Hr) in an attempt to improve the accuracy of preoperative diagnosis of thyroid nodules in clinical practice.

SUBJECTS AND METHODS

Patients. Patients with thyroid nodules 1 cm or larger requiring FNAC were prospectively included between December 2007 and March 2010 in a single University Hospital referral center. Patients with cystic or confluent nodules were excluded. Finally, 722 patients were included, of whom 225 (31%) underwent surgical excisions (Fig 1). The research was carried out according to ethical guidelines (Declaration of Helsinki), and the study received the approval of the local ethics committee.

Material and data. Several clinical and ultrasound data were first collected. For each participant, ultrasound-guided fine-needle aspirations were carried out by an experienced specialized radiologist with 2–5 passes per nodule using 23-gauge needles. Slides were then prepared by direct smear. One sample was expelled in lysis buffer (Qiagen, Courtaboeuf, France)

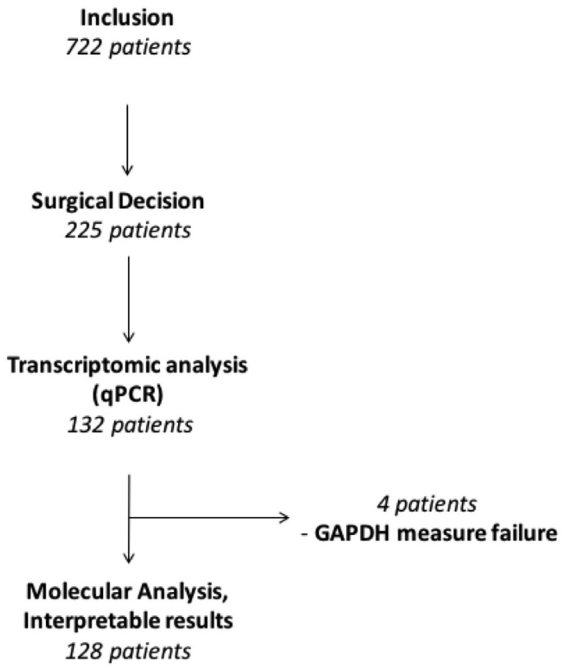

Fig 1. Flowchart of the study.

for RNA preservation and kept at −20°C until laboratory analyses. The assessed population included 128 patients with RNA samples quantitatively and qualitatively suitable for transcriptome analysis.

Cytological analysis. All FNAC results were interpreted by local cytologists using 3 slides per nodule on average and Papanicolaou staining, and classified as benign, malignant, indeterminate, nondiagnostic, or other. This presurgical classification (together with other nodule and patient characteristics) was used only for the surgeon's decision.

TBSRTC scores were established by a second blind examination of preoperative cytology results by an expert cytologist (M.D-P.).[8,10] Only this classification was used for development of the new test.

Histological analysis. Only aspirated nodules were analyzed; incidentally discovered tumors were excluded. The histopathological analysis was performed by a local pathologist, but specimens with uncertain malignant potential were independently reviewed by an expert in thyroid tumors (M.D-P.) Finally, all nodules were classified as malignant or nonmalignant based on pathology.[22]

Molecular analysis. RNA extraction was carried out using the RNeasy Micro kit (Qiagen, Courtaboeuf, France). Potential genomic DNA contamination was eliminated using RNase free DNase I (Qiagen). The RNA quantity was measured by a NanoDrop spectrophotometer (Wilmington, Del), and its quality was assessed by determination of the RNA integrity number with a bioanalyzer (Agilent, Santa Clara, Calif). Only samples with a RNA integrity number >5 and a minimum of 40 ng RNA were further analyzed.

A twenty-gene DNA array was carried out (plus GAPDH as housekeeping gene) (Supplementary Table 1). Nineteen of these genes were previously selected in a work on gene expression in carcinomas and adenomas[23]; the twentieth (GDF15) was added after literature review.[24]

Quantitative polymerase chain reaction was performed in duplicates on the profileXpert platform (Lyon, France) by an operator blind to the histological status. This quantitative polymerase chain reaction used TaqMan Array Microfluidic Cards on a 7900HT fast real-time polymerase chain reaction system (Applied Biosystems, Foster City, Calif) and complementary DNAs obtained after reverse transcription and amplification with a QuantiTect Whole Transcriptome kit (Qiagen) (see supplementary data for details).

Statistical analysis. The analyses used R software (http://www.R-project.org/).

Group comparisons. Comparisons of the characteristics of the "surgery with molecular test" group (n = 128) vs the "surgery without molecular test" group (n = 97) were carried out with the Fisher test for categorical variables and the Wilcoxon rank test for continuous variables.

Molecular predictor. Logistic regression models, with $\log_{10}$(RQ) as linear covariates, were built using a forward selection process based on Akaike information criterion to obtain the molecular predictor. Missing and discarded gene measurements were imputed with appropriate methods (Supplementary Table 1).[25] To avoid overfitting due to selection of too many variables, the optimal number of genes in the model was estimated with a bootstrap method (Supplementary data).

The combined Bethesda-molecular predictor. The BSRTC score of each patient was used to build a receiver-operating characteristic (ROC) curve. The area under the ROC curve (AUC) and its 95% confidence interval were estimated from a logistic regression model adjusted only on the BSRTC score (model A) and from the molecular predictor (model B). The combined predictor (Bethesda-molecular predictor) was then built after adjustment on the linear predictors stemming from models A and B. The estimated AUC of the combined model was compared with that of model A using the DeLong method.[26]

Using the prevalence of malignancy is necessary to minimize the number of misclassified patients. For each patient, the combined predictor gave a probability of malignancy, thus a nonbinary response. A cutoff (or threshold) value of this probability was then necessary to classify a nodule as malignant or nonmalignant.

Consequently, sensitivity and specificity had to be calculated at several thresholds. Traditionally, the optimal threshold is the one that maximizes sensitivity plus specificity. Here, we considered a threshold c that maximized the expression:

$$Sensitivity(c) \times Prevalence + Specificity(c) \times (1 - Prevalence)$$

This choice minimizes the expected number of misclassified nodules. Hence, the sensitivity and specificity of the combined predictor can be estimated at several malignancy prevalence values after setting an optimal threshold. To estimate the performance of the predictor in the whole data set, we used the prevalence value of the whole study data set (36%), and used 30%, the prevalence value of the indeterminate nodule group only, for estimation of the performance in that subset.

The clinician's expectations may differ according to the clinical context, and a "ruling out" (minimize false-negatives) or "ruling in" (minimize false-positives) strategy may be preferred. This variable was integrated using the concepts of "net harm" and "net benefit".[27] Finally, determination of the optimal predictor threshold c took into account the prevalence of malignancy in the study population and the benefit-to-harm ratio (B/Hr) by maximizing the following expression:

$$[Sensitivity(c) \times Prevalence] + [(Specificity(c) \times (1 - Prevalence)) \times Ratio].$$

This is a minimization of a "weighted" expected number of misdiagnosed nodules to favor either sensitivity or specificity. We tested 2 B/Hr, a B/Hr of 1 being no weighting and a B/Hr of 4 favoring sensitivity.

The 95% confidence intervals of sensitivity and specificity values were calculated using bootstrap sampling.[28]

RESULTS

Population characteristics. The clinical characteristics of the participants are given in Table I. Of the 225 patients who underwent thyroidectomy, 128 were assessed with the molecular test. There were no significant differences between the latter patients (n = 128) and those who did not have the molecular test (n = 97) regarding malignancy prevalence ($P = 0.579$), age ($P = 0.836$), or sex ratio ($P = 0.775$); however, the ultrasound-determined nodule size was lower in assessed than that in the nonassessed patients ($P = 0.029$).

According to TBSRTC, samples from 18 (14%) of the 128 patients were classified as DC II (benign) and 24 (19%) were DC VI (malignant). Samples from 10 patients (8%) were classified as DC I (nondiagnostic) and 76 (59%) were indeterminate (i.e., DC III, IV, and V).[10]

On histological examination, 46 nodules (36%) were malignant. Of these, 39 were papillary thyroid carcinomas. The other nodules were follicular thyroid carcinomas (3 cases), anaplastic thyroid cancer (1 case), poorly differentiated carcinoma (1 case), and medullary thyroid carcinoma (2 cases). The malignancy

Table I. Participant clinical characteristics

Characteristic	Enrollment	Thyroidectomies		
		All	With the combined Bethesda-molecular test	Without the combined Bethesda-molecular test
Patients, n	722	225	128	97
Females, n (%)	536 (74)	153 (68)	86 (67)	67 (69)
Males, n (%)	186 (26)	72 (32)	42 (33)	30 (31)
Age at inclusion; yr, mean (SD)	53.0 (13.5)	50.7 (14.0)	50.5 (14.4)	51.0 (13.7)
Largest nodule diameter; mm, mean (SD)	28.3 (12.6)	31.4 (14.3)	29.6 (13.4)	33.9 (15.0)
Cytology/Bethesda diagnostic category				
Benign or DC II, n (%)	342 (48)	37 (16)	18 (14)	19 (20)
Malignant or DC VI, n (%)	50 (7)	46 (20)	24 (19)	22 (23)
Indeterminate or DC III to V, n (%)	161 (22)	116 (52)	76 (59)	40 (41)
Inconclusive or DC I, n (%)	135 (19)	24 (11)	10 (8)	14 (14)
Other, n (%)	31 (4)	2 (1)	0	2 (2)
Missing data, n (%)	3 (0)	0	0	0
Surgical procedure				
Total thyroidectomy, n (%)	179 (25)	179 (80)	101 (79)	78 (80)
Partial thyroidectomy, n (%)	46 (6)	46 (20)	27 (21)	19 (20)
No surgery, n (%)	497 (69)	—	—	
Malignancy				
Malignant, n (%)	—	85 (38)	46 (36)	39 (40)
Nonmalignant, n (%)	—	140 (62)	82 (64)	58 (60)

rates within TBSRTC categories III to VI were 2/21 (9.5%), 9/37 (24.3%), 12/18 (66.7%), and 22/24 (92%), respectively (Table II).

The molecular predictor. The optimal number of genes for the logistic regression model was 7; a 7-gene molecular predictor was then considered to be the best predictive model.

The adjusted odds ratios of the 11 selected genes ranged between 0.31 and 0.50 for genes whose overexpression could be considered to be "protective" and between 1.71 and 2.60 for genes whose overexpression could be associated with increased risk of malignancy (Table III).

The AUC of the molecular predictor was estimated at 85.3% (95% CI: [78.2%–92.4%]).

The combined predictor. The estimated AUC of the combined predictor was significantly higher than that of TBSRTC alone: 93.5% (95% CI: [89.1%–97.9%]) vs 88.1% (95% CI: [81.9%–94.3%]) ($P = 0.004$, Delong test) (Fig 2).

The combined predictor performance in the whole data set. The performance of the combined predictor is shown in Table IV.

With 36% prevalence and a B/Hr of 1, sensitivity of the combined predictor was 76.1% and specificity 95.1%, with estimated positive and negative predictive values of 90% and 88%, respectively. This corresponds to 86 and 31 false-negative and false-positive results, respectively, of 1000 aspirated nodules in a cohort having a distribution of malignant nodules and TBSRTC DCs similar to that of the study population.

As expected, with 36% prevalence but a B/Hr of 4, the sensitivity increased up to 95.7%, but the specificity decreased to 74.4%. In a similar cohort, false-negatives would decrease to 15 per 1000 nodules, but the sum of false-negatives and false-positives would increase to 179 per 1000 nodules.

The combined predictor performance in indeterminate nodules. In indeterminate nodules, using the prevalence of malignancy of 30% and a B/Hr of 1, specificity was 96.2% and sensitivity 52.2%, which means 11 false-negative and 2 false-positive results among 76 indeterminate nodules. With a B/Hr of 4, sensitivity rose to 95.7%, at the cost of 64.2% specificity. There were thus 19 false-positives but 1 false-negative among the 76 indeterminate nodules. If we consider only the TBSRTC DC 4 subset (37 tumors), sensitivity was 44.4% and specificity was 100% for a B/Hr of 1, compared to 100% sensitivity and 53.6% specificity for a B/Hr of 4.

Table III. Adjusted odds ratios for the study molecular predictor

Gene reference	Adjusted OR (95% CI)	*P*-value*
FN1 Hs01549976m1	2.01 [1.12–3.58]	0.015
CITED2 Hs01897804s1	0.31 [0.17–0.58]	<0.001
CITED1 Hs00366310m1	2.60 [1.45–4.65]	<0.001
CHI3L1 Hs00609691m1	0.50 [0.29–0.87]	0.012
TFF3 Hs00173625m1	0.46 [0.26–0.82]	0.005
CDKN1A Hs00355782m1	1.71 [0.99–2.98]	0.046
CSGALNACT1 Hs00218054m1	1.79 [0.98–3.28]	0.047

*According to the likelihood ratio test.

DISCUSSION

We present here the construction of a combined Bethesda-molecular test of thyroid nodule malignancy that showed a higher overall performance than TBSRTC alone.

These promising results were obtained in a representative sample of patients who underwent ultrasound-guided FNAC for solid nodules ≥1 cm, and of whom 31% underwent surgery, which is similar to the percentages typically reported for tertiary centers.[9,12] As expected, the rate of indeterminate aspirations according to our first presurgical classification (22%) was equivalent to that of indeterminate nodules in recent publications (DC III–V).[9] Most operated nodules were classified as indeterminate or malignant;

Table II. Rates of malignancy according to The Bethesda System for Reporting Thyroid Cytopathology (BSRTC)

Bethesda diagnostic category	Malignant nodules in the study, n (%)	Malignant nodules in the literature, (%)*
DC I (nondiagnostic)	1/10 (10.0)	16.8
DC II (benign)	0/18 (0.0)	3.7
DC III (atypia/follicular lesion of undetermined significance)	2/21 (9.5)	15.9
DC IV (follicular neoplasm or suspicious for a follicular neoplasm)	9/37 (24.3)	26.1
DC V (suspicious for malignancy)	12/18 (66.7)	75.2
DC VI (malignant)	22/24 (91.7)	98.6
Total	46/128 (36.0)	33.8

*Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta cytologica 2012; 56: 333–339.

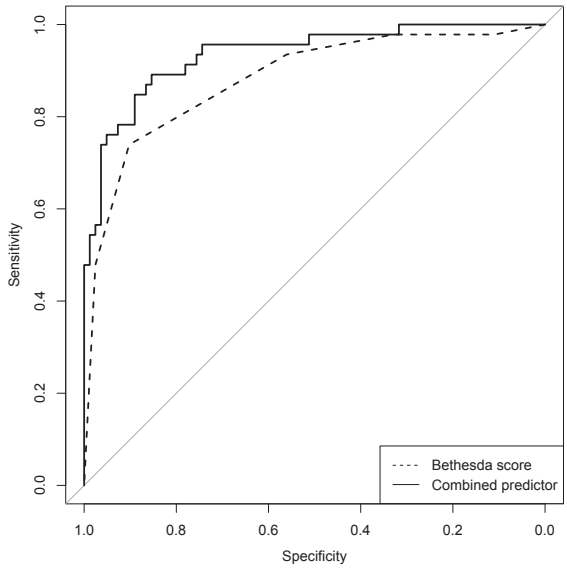

Fig 2. Bethesda System for Reporting Thyroid Cytopathology scores and combined Bethesda-molecular predictor ROC curve. ROC curve, receiver-operating characteristic curve.

the surgery rate for cytological malignant nodules was 92%, and 72% in indeterminate aspirations. Here, the malignancy prevalence among all operated patients was 38%, which is close to rates reported in literature.[9,12]

The molecular analysis involved 128 patients who had undergone surgery. In this subset, the malignancy rate was 36%, and the distribution of malignant nodules vs TBSRTC DC was as expected.[8,10] All DCs, and not only indeterminate categories, were used to build the new predictor, because it enabled the overall expected malignancy prevalence to be considered, and any potential classification errors made in the DC II and DC VI categories could be amended.

The 7 genes selected were based on a previous expression analysis of 200 genes.[23] Modified expression of 6 of these 7 genes has previously been described in thyroid cancer. They seem to be implicated in various cell functions, i.e., cell cycle regulation for *CHI3L1* and *CDKN1A*,[29] and cell migratory ability for FN1[30] and *TFF3*.[31] Less is known about the functions of the *CITED1* and 2 genes and the *CSGALNACT1* gene, although some studies have reported modifications of CITED1 and CSGALNACT1 expression in thyroid carcinomas.[32-34]

The incidence of thyroid nodules and cancers has markedly increased over the past decades, which is mainly due to an increased detection rate.[35,36] Indeed, thyroid nodules are quite common imaging incidentalomas: 16% with computed tomography scans and magnetic resonance imagings but >50% with ultrasound.[3,37] The current preoperative diagnosis of a nodule relies mainly on FNAC, and surgery is proposed for patients with suspicious findings. This strategy results in detection of a malignancy in only 30% of operated patients[9]; i.e., in France, yearly, 18,500 of 26,500 thyroidectomies would be "unnecessary". In a recent retrospective analysis, the rate of surgical complications was 9%: hypocalcemia, hematoma, vocal cord dysfunction, tracheostomy, and death (0.3%).[38] Improving preoperative assessment is therefore required to avoid unnecessary interventions and complications. Besides, extensive thyroid surgeries increase incidental discoveries of papillary microcarcinomas, which has increased the incidence of thyroid cancer in Western countries over the past 30 years,[39] although the death rate from thyroid cancer remained unchanged (0.47/100,000 persons in the USA), and the overall 10-year specific survival remained close to 95%.[1,39,40] This suggests an increased detection of subclinical disease rather than a true increase in thyroid cancer incidence.[3,35,36,41]

These facts highlight the need to improve the specificity of TBSRTC. Because FNAC is still unavoidable, the combined Bethesda-molecular test is welcome and should result in the avoidance of a number of unnecessary surgeries, as proved by the significant gain of AUC. Given the excellent performance of TBSRTC for the DC II and DC VI subsets, the gain in performance with the new predictor obviously concerned indeterminate nodules (DCs III-V).

The combined predictor may be indicated in cases of indeterminate DC or questionable DC II (for example with ultrasound suspicion of malignancy). In clinical practice, a sample of cytological material could be kept after fine-needle aspiration, and molecular examination performed only in the case of an indeterminate cytological result. If available, an estimation of the malignancy prevalence in the concerned center could

Table IV. Performance of the combined Bethesda-molecular predictor of nodule malignancy on the whole data set

Prevalence	B/Hr	Sensitivity (95% CI)	Specificity (95% CI)	FN*	FP*	PPV	NPV
36%	4	95.7% [87.0–100.0]	74.4% [66.4–93.9]	15	164	68%	97%
36%	1	76.1% [65.2–97.8]	95.1% [85.4–100.0]	86	31	90%	88%

Abbreviations: *B/Hr*, benefit/harm ratio; *FN*, false-negatives; *FP*, false-positives; *PPV*, positive predictive value; *NPV*, negative predictive value. An elevated B/Hr favors sensitivity.
*Per 1000 nodule aspirates.

be considered to fix the most appropriate threshold before classifying nodules; if not, literature malignancy prevalence may be considered. Then, the B/Hr should be chosen according to the physician's objective. Physicians are often anxious about missing cancer. Thus, the B/H of 4 seems to be generally the most convenient, as it gives a small number of false-negative results while reducing the number of false-positive results of the TBSRTC. Actually, in our population, with 30% malignancy prevalence among indeterminate nodules, the new predictor using a B/Hr of 4 would have resulted in only 1 cancer being missed, but the number of unnecessary surgeries would have decreased from 53 to 19 in DC III-V, thus avoiding 34 surgeries. A lower B/Hr can be used when a false-negative result is less harmful, for example, in the case of a small nodule. In a given nodule, malignancy risk is not linked to nodule size. However, in cases of thyroid cancer, the prognosis is poorer in large nodules, and this variable is integrated to tumor, lymph nodes and metastasis staging. Also, a false-negative result is obviously less damaging when a nodule is smaller than 2 cm, which corresponds to tumor, lymph nodes and metastasis stage 1, compared to a larger nodule.

As previously published tests did not include TBSRTC DC or malignancy prevalence, comparisons are limited. However, the few current molecular tests to "rule-in" (low sensitivity but high specificity) do not contribute much to the therapeutic strategy due to their low sensitivity. Indeed, in indeterminate mutation-negative nodules, the risk of malignancy was 6%–28% with the gene mutation panel of Nikiforov *et al.*,[16] and 10%–79% with that of Eszinlger *et al.*[42] The recent Multi-Gene ThyroSeq Next-Generation Sequencing Assay has shown excellent performances but has only been assessed in DC III and IV cases of a monocentric cohort and not yet in clinical practice.[18,43] The very sensitive GEC[19] was developed to "rule out" but displayed lower accuracy in clinical practice than first evaluated, mainly because of variable prevalence.[20,21,44] Thus, we adapted the test threshold to various cancer prevalence values to minimize the number of prediction errors (i.e., misclassifications) in the target population. This could also allow incorporation of the performance characteristics of other tools; e.g., ultrasound classifications such as thyroid imaging-reporting and database system or American Thyroid Associations, whose categories are associated with known malignancy probabilities and play an ever increasing part in nodule assessment.[5,45]

One limitation of this study is the small number of thyroid nodules evaluated; further prospective assessment of the new predictor should be performed in a larger study to confirm the present results. Moreover, major modifications of the international histological classification of thyroid tumors, the gold standard of diagnosis, are likely to be available soon and could alter the performances of all molecular tests currently estimated. It has been suggested that a new category of tumor, noninvasive follicular thyroid neoplasm with papillary-like nuclear features, should be created to classify the current encapsulated noninvasive follicular variant papillary thyroid carcinomas as nonmalignant tumors.[46] However, this is based on moderate quality evidence, and further prospective studies are needed to confirm good patient outcomes. Currently, it requires surgical excision with meticulous histological examination and a prolonged follow-up.[47]

In conclusion, the new combined Bethesda-molecular predictor of nodule malignancy allows for the prevalence of the disease and B/Hr. We believe this will enhance its performance in everyday clinical practice and avoid unnecessary thyroidectomies.

ACKNOWLEDGMENTS

Conflicts of Interest: All authors have read the journal's policy on disclosure of potential conflicts of interest and have none to declare.

This study was supported by the French Ministry of Health (PHRC 2006).

Author contributions are as follow: Study design was done by B.Ri., E.D., C.C., B.Ro., F.B-C., and P.R. Laboratory work and data were carried out by J.L., S.S-R., and B.Ro. Cytological and histological examinations were conducted by M.D-P. Clinical follow-up and data were carried out by H.L., V.L., C.C., J-L.P., J-C.L., and F.B-C. Data collection and management was carried out by H.L., C.C., J.L., O.C., S.S-R., B.Ro., and F.B-C. Data analysis was carried out by H.L., B.Ri., E.D., and P.R. Result interpretation was carried out by H.L., B.Ri., E.D., F.B-C., and P.R. Manuscript drafting was performed by H.L., B.Ri., and E.D. Manuscript revision was conducted by F.B-C. and P.R.

The manuscript has been revised for the English by an independent scientific language editing service (Anglo-scribe). The authors thank Dr Marie-Laure Denier for the ultrasonographic examinations, Pr David Wynford-Thomas for many helpful comments, and Jean Iwaz (Hospices Civils de Lyon) for the editing the manuscript. The final version manuscript has been reviewed by and approved by all named authors. They all have read the journal's authorship agreement.

REFERENCES

1. McLeod DSA, Sawka AM, Cooper DS. Controversies in primary treatment of low-risk papillary thyroid cancer. Lancet Lond Engl 2013;381:1046–57.

2. Valeix P, Zarebska M, Bensimon M, et al. Ultrasonic assessment of thyroid nodules, and iodine status of French adults participating in the SU.VI.MAX study. Ann Endocrinol 2001;62:499–506.
3. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. Eur J Clin Invest 2009;39: 699–706.
4. Wémeau J-L, Sadoul J-L, d'Herbomez M, et al. Recommendations of the French Society of Endocrinology for the management of thyroid nodules. Presse Medicale Paris Fr 2011;40: 793–826.
5. Kwak JY, Han KH, Yoon JH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. Radiology 2011;260:892–9.
6. Brito JP, Gionfriddo MR, Al Nofal A, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. J Clin Endocrinol Metab 2014;99:1253–63.
7. Russ G, Bigorgne C, Royer B, Rouxel A, Bienvenu-Perrard M. The thyroid imaging reporting and data system (TIRADS) for ultrasound of the thyroid. J Radiol 2011;92:701–13.
8. Ali SZ. Thyroid cytopathology: Bethesda and beyond. Acta Cytol 2011;55:4–12.
9. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. Acta Cytol 2012;56:333–9.
10. Cibas ES, Ali SZ, NCI Thyroid FNA State of the Science Conference. The Bethesda system for reporting thyroid cytopathology. Am J Clin Pathol 2009;132:658–65.
11. Durante C, Costante G, Lucisano G, et al. The natural history of benign thyroid nodules. JAMA 2015;313:926–35.
12. Jo VY, Stelow EB, Dustin SM, Hanley KZ. Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda System for Reporting Thyroid Cytopathology. Am J Clin Pathol 2010;134:450–6.
13. Biondi B, Wartofsky L. Treatment with thyroid hormone. Endocr Rev 2014;35:433–512.
14. Nishino M. Molecular cytopathology for thyroid nodules: a review of methodology and test performance. Cancer Cytopathol 2016;124:14–27.
15. Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. Nat Rev Cancer 2006;6:292–306.
16. Nikiforov YE, Ohori NP, Hodak SP, et al. Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1056 FNA samples. J Clin Endocrinol Metab 2011;96: 3390–7.
17. 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 2015;25:760–8.
18. 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 2014;120:3627–34.
19. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. N Engl J Med 2012;367:705–15.
20. Harrell RM, Bimston DN. Surgical utility of Afirma: effects of high cancer prevalence and oncocytic cell types in patients with indeterminate thyroid cytology. Endocr Pract 2014;20:364–9.
21. McIver B, Castro MR, Morris JC, et al. An independent study of a gene expression classifier (Afirma) in the evaluation of cytologically indeterminate thyroid nodules. J Clin Endocrinol Metab 2014;99:4069–77.
22. Kakudo K, Bai Y, Liu Z, Li Y, Ito Y, Ozaki T. Classification of thyroid follicular cell tumors: with special reference to borderline lesions. Endocr J 2012;59:1–12.
23. Durand S, Ferraro-Peyret C, Selmi-Ruby S, et al. Evaluation of gene expression profiles in thyroid nodule biopsy material to diagnose thyroid cancer. J Clin Endocrinol Metab 2008; 93:1195–202.
24. Weber F, Shen L, Aldred MA, et al. Genetic classification of benign and malignant thyroid follicular neoplasia based on a three-gene combination. J Clin Endocrinol Metab 2005;90: 2512–21.
25. Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. Bioinforma Oxf Engl 2001;17:520–5.
26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
27. DeNeef P, Kent DL. Using treatment-tradeoff preferences to select diagnostic strategies: linking the ROC curve to threshold analysis. Med Decis Making 1993;13:126–32.
28. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26: 565–74.
29. Lei S-T, Shen F, Chen J-W, et al. MiR-639 promoted cell proliferation and cell cycle in human thyroid cancer by suppressing CDKN1A expression. Biomed Pharmacother 2016;84:1834–40.
30. Sponziello M, Rosignolo F, Celano M, et al. Fibronectin-1 expression is increased in aggressive thyroid cancer and favors the migration and invasion of cancer cells. Mol Cell Endocrinol 2016;431:123–32.
31. Abols A, Ducena K, Andrejeva D, et al. Trefoil factor 3 is required for differentiation of thyroid follicular cells and acts as a context-dependent tumor suppressor. Neoplasma 2015;62:914–24.
32. Huang Y, Prasad M, Lemon WJ, et al. Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. Proc Natl Acad Sci U S A 2001;98:15044–9.
33. Sassa M, Hayashi Y, Watanabe R, et al. Aberrant promoter methylation in overexpression of CITED1 in papillary thyroid cancer. Thyroid 2011;21:511–7.
34. Schulten H-J, Al-Mansouri Z, Baghallab I, et al. Comparison of microarray expression profiles between follicular variant of papillary thyroid carcinomas and follicular adenomas of the thyroid. BMC Genomics 2015;16:S7.
35. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic"–screening and overdiagnosis. N Engl J Med 2014; 371:1765–7.
36. Ahn HS, Welch HG. South Korea's thyroid-cancer "epidemic"–turning the tide. N Engl J Med 2015;373:2389–90.
37. Yoon DY, Chang SK, Choi CS, et al. The prevalence and significance of incidental thyroid nodules identified on computed tomography. J Comput Assist Tomogr 2008;32:810–5.
38. Weiss A, Parina RP, Tang JA, Brumund KT, Chang DC, Bouvet M. Outcomes of thyroidectomy from a large California state database. Am J Surg 2015;210:1170–6. 1177.
39. Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973-2002. JAMA 2006;295:2164–7.
40. Ho AS, Davies L, Nixon IJ, et al. Increasing diagnosis of subclinical thyroid cancers leads to spurious improvements in survival rates. Cancer 2015;121:1793–9.
41. Brito JP, Davies L. Is there really an increased incidence of thyroid cancer? Curr Opin Endocrinol Diabetes Obes 2014;21: 405–8.
42. Eszlinger M, Böhme K, Ullmann M, et al. Evaluation of a two-year routine application of molecular testing of thyroid fine-

needle aspirations using a seven-gene panel in a primary referral setting in Germany. Thyroid 2017;27:402–11.
43. 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 Off J Am Thyroid Assoc 2015;25:1217–23.
44. Marti JL, Avadhani V, Donatelli LA, et al. Wide inter-institutional variation in performance of a molecular classifier for indeterminate thyroid nodules. Ann Surg Oncol 2015;22:3996–4001.
45. Yoon JH, Lee HS, Kim E-K, Moon HJ, Kwak JY. Malignancy risk stratification of thyroid nodules: comparison between the thyroid imaging reporting and data system and the 2014 American thyroid association management guidelines. Radiology 2016;278:917–24.
46. 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 2016;2:1023–9.
47. Haugen BR, Sawka AM, Alexander EK, et al. The ATA guidelines on management of thyroid nodules and differentiated thyroid cancer task force review and recommendation on the proposed renaming of eFVPTC without invasion to NIFTP. Thyroid 2017;27:481–3.

Appendix

Supplementary Data

Quantitative PCR. A gene was considered "nonexpressed" when the number of PCR cycles was ≥32. When the duplicate values differed markedly (standard deviation > 0.5), they were discarded for insufficient quality. This occurred in 1.6%–18.8% of patients, depending on the gene (Supplementary data, Table 1). A standardized determination of PCR cycles (vs the GAPDH housekeeping gene) was used to correct for RNA quantity. The statistical analysis used a log transformation of this standardized number as $\log_{10}(RQ)$, where $RQ = 2^{\text{standardized measure for the calibrator} - \text{standardized measure for the patient}}$.

Statistical performance of the logistic regression model. The optimal number of genes was defined as the median number of genes kept after 1000 bootstrap simulations.

The final molecular predictor was obtained after selection of the optimal number of genes using the forward selection process on the whole data set.

The first step of the selection process was the determination of the optimal number "k" of genes to retain for the logistic regression model. This number was obtained using 1000 bootstrap samples with the Akaike information criterion as a stopping rule for including a new gene in the predictor equation.

The second step was to bootstrap 1000 pseudo-data sets from the whole data set to identify the best-fit k-gene model. Because this optimal number was 7 genes in 41.9%, 5 genes in 2.4%, 6 genes in 15.8%, and 8 genes in 26.3% of the 1000 bootstrap replicates, we decided to also investigate models with k-2, k-1, and k+1 genes. Thus, we investigated a total of 4000 logistic regression models.

To obtain representative pseudo-data sets, we considered only bootstrap replicates in which the number of patients and the proportion of malignant tumors are the same as in the whole data set (i.e., 128% and 36%, respectively). Also, to evaluate the unbiased predictive performance of the model, each pseudo-data set was spilt into 2 parts: 90% of the data (patients) were used to build the model and 10% to estimate the sensitivity, the specificity, and the percentage of good status prediction.

Supplementary Table I. Number of patients per gene with imputed and real PCR data

Gene	Real data (number of patients, %)	Imputed data (number of patients, %)
FN1	122 (95.3)	6 (4.7)
TFF3	120 (63.7)	8 (6.3)
CITED2	126 (98.4)	2 (1.6)
PROS1	116 (90.6)	12 (9.4)
CITED1	118 (96.2)	10 (7.8)
MPPED2 (C11orf8)	119 (93.0)	9 (7.0)
DUSP6	104 (81.2)	24 (18.8)
TIMP1	123 (96.6)	5 (3.4)
CDKN1A	123 (96.6)	5 (3.4)
CTSB	126 (98.4)	2 (1.6)
MATN2	111 (96.7)	17 (13.3)
GPX3 (PAB)	126 (98.4)	2 (1.6)
GDF15	119 (93.0)	9 (7.0)
TNFRSF11B	118 (92.2)	10 (7.8)
PGCP	125 (97.7)	3 (2.3)
FGFR2	116 (90.6)	12 (9.4)
CSGALNACT1 (ChGn)	124 (96.9)	4 (3.1)
CHI3L1	123 (96.1)	5 (3.9)
TGFBI	126 (98.4)	2 (1.6)
MAN1C1	122 (95.3)	6 (4.7)