IJC
International Journal of Cancer

A three-gene panel that distinguishes benign from malignant thyroid nodules

Bing Zheng[1,2,3]*, Jun Liu[4]*, Jianlei Gu[1,2]*, Yao Lu[1,2], Wei Zhang[3], Min Li[3] and Hui Lu[1,2,5]

[1] Shanghai Institute of Medical Genetics, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai, China
[2] Key Laboratory of Molecular Embryology, Ministry of Health and Shanghai Key Laboratory of Embryo and Reproduction Engineering, Shanghai, China
[3] Department of Laboratory Medicine, Renji Hospital, Shanghai Jiao Tong University, Shanghai, China
[4] Department of Otolaryngology, Renji Hospital, Shanghai Jiao Tong University, Shanghai, China
[5] Department of Bioengineering, University of Illinois at Chicago, Chicago, IL

Reliable preoperative diagnosis of malignant thyroid tumors remains challenging because of the inconclusive cytological examination of fine-needle aspiration biopsies. Although numerous studies have successfully demonstrated the use of high-throughput molecular diagnostics in cancer prediction, the application of microarrays in routine clinical use remains limited. Our aim was, therefore, to identify a small subset of genes to develop a practical and inexpensive diagnostic tool for clinical use. We developed a two-step feature selection method composed of a linear models for microarray data (LIMMA) linear model and an iterative Bayesian model averaging model to identify a suitable gene set signature. Using one public dataset for training, we discovered a three-gene signature dipeptidyl-peptidase 4 (DPP4), secretogranin V (SCG5) and carbonic anhydrase XII (CA12). We then evaluated the robustness of our gene set using three other independent public datasets. The gene signature accuracy was 85.7, 78.8 and 85.7%, respectively. For experimental validation, we collected 70 thyroid samples from surgery and our three-gene signature method achieved an accuracy of 94.3% by quantitative polymerase chain reaction (QPCR) experiment. Furthermore, immunohistochemistry in 29 samples showed proteins expressed by these three genes are also differentially expressed in thyroid samples. Our protocol discovered a robust three-gene signature that can distinguish benign from malignant thyroid tumors, which will have daily clinical application.

Key words: thyroid cancer, prediction model, machine learning, diagnostic panel, biomarkers

Abbreviations: ACT: anaplastic thyroid carcinoma; BMA: Bayesian model averaging; BSS/WSS: the ratio of between-group to within-group sum of squares; CA12: carbonic anhydrase XII; DPP4: dipeptidyl-peptidase 4; FA: follicular adenoma; FNA: fine-needle aspiration; FTC: follicular thyroid carcinoma; FVPC: follicular variants of papillary thyroid carcinoma; GEO: gene expression omnibus; HCA: Hurthle cell adenoma; HMGA2: high mobility group AT-hook 2; HT: Hashimoto thyroiditis; MRC2: mannose receptor, C type 2; PTC: papillary thyroid carcinoma; QPCR: quantitative real-time polymerase chain reaction; SAM: significance analysis of microarrays; SCG5: secretogranin V; SFN: stratifin; US: ultrasonography; TH: thyroid hyperplasia

Additional Supporting Information may be found in the online version of this article.

*B.Z., J.L. and J.G. contributed equally to this paper.

Grant sponsor: National Natural Science Foundation of China; **Grant number:** 31071167 and 31370751; **Grant sponsor:** Medical Engineering Cross Research Foundation of Shanghai Jiaotong University

DOI: 10.1002/ijc.29172

History: Received 30 Dec 2013; Accepted 3 July 2014; Online 30 Aug 2014

Correspondence to: Hui Lu, Shanghai Institute of Medical Genetics, Shanghai Children's Hospital, Shanghai Jiaotong University and University of Illinois at Chicago, E-mail: huilu.bioinfo@gmail.com

Thyroid cancer is the most common endocrine malignancy, accounting for 86.8% of all endocrine system tumors[1] with an annual incidence of 1.0–1.5% of newly diagnosed systemic cancers in the United States. As a result of the increased frequency of medical examinations, and the availability of advanced diagnostic instruments, clinicians are faced with an ever-increasing number of patients with thyroid nodules. Notably, only 5–7.7% of palpable or nonpalpable thyroid lesions harbor thyroid carcinomas,[2] with many benign thyroid nodules only diagnosed as such after surgical removal. It is, therefore, important to preoperatively diagnose thyroid nodules to avoid unnecessary surgical procedures.

Currently, the main clinical diagnostics for thyroid cancer include high-resolution ultrasonography (US) and fine-needle aspiration (FNA),[2] while cytological examination by ultrasound-guided FNA (US-FNA) is an especially accurate and cost-effective preoperative diagnostic method. However, 15–30% indeterminate or suspicious cytological findings inspected by US-FNA have been reported,[3] which poses dilemmas for clinicians. Even after repeated US-FNA, the uncertainty rate is still 20–58%.[4–6] Therefore, there is a strong drive to develop new diagnostic tools to address indeterminate results (of which 20–30% are in fact malignant[7]). In recent years, microarray-based gene-expression analysis of thyroid FNA biopsy or tissue samples have shown promise as a diagnostic tool.[8–10]

What's new?

Today a key challenge in thyroid cancer research lies in distinguishing benign thyroid nodules from malignant tumors in order to avoid unnecessary surgery. While many researchers have focused on molecular classification based on oligonucleotide microarray gene-expression patterns, the high cost and poor reproducibility are barriers for daily clinical application. Here, using a two-step feature selection method, the authors constructed a small gene-expression set that can distinguish benign from malignant thyroid tumors with high prediction accuracy. The three-gene signature was validated experimentally. The prediction model is simpler and more affordable than microarray-based gene-expression patterns and more suitable for daily clinical application.

One potential problem in microarray-based diagnostics is that the number of genes assessed is usually much larger than the number of samples analyzed. Furthermore, the majority of feature selection algorithms, such as significance analysis of microarrays (SAM)[11] and the *t*-test,[12] are univariate classification methods (where each gene is considered independently), which disregard the potential relationships between genes. Therefore, multivariate gene selection methods (that consider a number of genes simultaneously), which reduce the number of correlative genes or redundant removal method,[13] are more suited for establishing a cost-effective diagnostic method with clinical application.

In this article, we applied Bayesian model averaging (BMA)[14] as a multivariate feature selection and a binary classification method, to differentiate malignant from benign tumors, and discovered a three-gene signature. The BMA algorithm has several advantages: one can fully consider model uncertainty by averaging the posterior probability of each possible model; it has been shown to produce satisfactory prediction accuracy with only a few signature genes[14,15]; it is highly efficient computationally; and the posterior probability of candidate genes are available for use in biological interpretation. Numerous classification signatures for thyroid nodules exist, but these each consist of around 100 genes,[8,10,16] which is not affordable for clinical use in terms of cost or time. We identified a three-gene (DPP4, SCG5, CA12) predictor model and validated its ability to distinguish thyroid tumors using three other public microarray databases. Most importantly, we were also able to validate the three-gene signature in 70 thyroid nodule samples from surgery by QPCR and achieve high accuracy. Furthermore, the proteins produced from these three genes showed distinguished pattern between malignant and benign nodules using immunohistochemistry experiment.

Materials and Methods

Patients and tissue samples

Thyroid tissue specimens were excised intraoperatively from 70 patients undergoing primary thyroidectomies in Renji Hospital (Supporting Information Table S1). Samples were immediately preserved in RNALater reagent and stored at −80°C to prevent RNA degradation. Of the 70 patients, 51 were female (aged 20–78 years) and 19 were male (aged 26–65 years). Histological classification of the samples, stained with hematoxylin and eosin, was performed by two independent clinical pathologists. There were 39 benign [20 follicular adenomas (FAs), 11 cases of Hashimoto thyroiditis (HT) and eight thyroid hyperplasias (TH)] and 31 malignant [29 papillary thyroid carcinomas (PTC), 1 follicular thyroid carcinoma (FTC) and one anaplastic thyroid carcinoma (ACT)] thyroid samples. All patients' personal information were deidentified and unknown to the authors, and the sample collection has been approved by IRB of the Renji Hospital.

Gene-expression data sets

Four published datasets from the gene expression omnibus (GEO) series GSE29315,[17] GSE33630, GSE27155[18,19] and GSE3678[20] were used in this study. We selected the training set (GSE29315) based on the proportion of benign and malignant tumors and the diversity of tumor subtypes in each dataset. The training set contained nine hurthle cell adenomas (HCAs), 17 FAs, nine FTCs, 13 follicular variants of papillary thyroid carcinomas (FVPCs), six HTs, eight THs and nine PTCs. The remaining three datasets were used as independent validation sets.

Microarray preprocessing

Because of the different platforms used in each dataset, and variability between various laboratories and experiments, the expression values in the training and test sets were normalized to 0–1. In other words, the expression value of a gene is normalized with the formula below:

$N_{\text{adjusted}} = (N - \min(\text{dataset}))/(\max(\text{dataset}) - \min(\text{dataset}))$, where N_{adjusted} is the adjusted value used in the training and test sets; N is the primitive value in the microarray dataset; min (dataset) is the minimum expression level in one microarray; and max (dataset) is the maximum expression level in one microarray.

Feature selection methods

GSE29315, which included 71 samples, was selected as the training set for reasons previously described. The feature selection method to distinguish benign from malignant samples consisted of two steps: a LIMMA linear model, and an iterative BMA algorithm.[14] The R package LIMMA[21] was used to select significantly differentially expressed genes

(fold-change cutoff of ≥ 2 and a p-value < 0.01). To minimize the number of signature genes, we applied the iterative BMA R package,[22] which appropriately accounts for model uncertainty as well as the dependency between signature genes.

The main steps of this iterative method are: First, rank all differentially expressed genes, selected by LIMMA, by the ratio of between-group to within-group sum of squares (BSS/WSS),[23] so that genes with a greater significant difference between benign and malignant groups and smaller variation within groups are ranked higher. Then, the leaps and bounds algorithm[24] is applied to the top 30 ranked genes. Genes with posterior probabilities less than 1% are removed and replaced by genes from the BSS/WSS ratio order until all genes are processed.

RNA isolation

For each clinical sample (preserved in RNALater) 100–150 mg of tissue was retrieved, excess RNALater solution was removed, and the samples were homogenized in lysis solution according to the manufacturers' instructions. RNA was extracted using TRIzol Reagent (Invitrogen) and the quantity and integrity of RNA was evaluated by NanoDrop 2000c (Thermo Scientific) and confirmed by 1% agarose gel electrophoresis. DNase I (Fermentas) was used to remove genomic DNA, after which about 1 ug RNA was reverse transcribed into cDNA using the RevertAid first strand cDNA synthesis kit (Fermentas).

QPCR analysis

Following reverse transcription, real-time QPCR was performed using the SYBR Green method (Roche) on an ABI 7500 (Applied Biosystems). cDNA samples were amplified in duplicate with the following primers: DPP4 (forward 5′-CTC CTT CTC TGA ACG CTC-3′; reverse 5′-TCA TCT GTG CCT TTG TTC-3′); SCG5 (forward 5′-GAC TTC AGC ATT TGG GTC C-3′; reverse 5′-ATT TGG AGG GTC TGG GTA C-3′); CA12 (forward 5′- TGC TCC TGC TGG TGA TCT-3′; reverse 5′-TGG AGG ATG TCA CTG TGC-3′); GAPDH (forward 5'-TGA CAA CTT TGG TAT CGT GGA AGG-3′; reverse 5′-AGG CAG GGA TGA TGT TCT GGA G-3′).[25] All the primers were designed to span exon-exon junctions or intronic sequences, to reduce any interference from genomic DNA. The average Ct value of the target gene was normalized to that of a reference gene, glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The value of $2^{-\Delta Ct} \times 100$ was used to measure the relative gene expression [15], where $\Delta Ct = Ct_{\text{target gene}} - Ct_{GAPDH}$.

Classification performance and cross validation

The signature genes derived from the training set were evaluated in the validation sets (GSE3678, GSE33630, GSE27155) as well as in 70 samples by QPCR analysis to verify the performance of the BMA classification model, which predicts the probabilities of a test sample. For the prediction probabilities restricted to 0–1, we designated the benign thyroid nodule group as 1 and the malignant group as 0. For a binary classification problem, 0.5 is usually the line of demarcation––if the calculated probability is more than 0.5, the test sample is assigned to the benign group and vice versa. Finally, the true classifications were compared with the classifications predicted by the model to analyze the prediction accuracy of the classifier. In addition, the Brier score,[26] which assesses the mean squared difference between the predicted probability and the actual outcome for the event, was applied to estimate the prediction accuracy of probabilistic forecasts.

Leave-one-out cross validation was performed to indicate the robustness of the iterative BMA model in predicting benign and malignant thyroid nodules in unseen samples. Hereby, one test sample is separated from the training set and used as validation data, while the remaining samples serve as the training set. This is repeated until all the samples in the original training set have been used as validation data.

Immunohistochemistry

Antibodies of the three proteins, DPP4, SCG5 and CA12 were all commercially available. Immunohistochemical analysis was performed to evaluate the protein expression between malignant and benign groups. Twenty nine paraffin-embedded formalin-fixed tissue samples (14 malignant + 15 benign thyroid nodules) were immunostained by the labeled streptavidin-biotin complex method. After the antigen retrival, the specimens were incubated with anti-DPP4/CD26 antibody (Abcam/ab114033), anti-SCG5 antibody (Abcam/ab22699) and anti-CA12 antibody (Abcam/ab140385) at the dilution of 1:50, 1:400 and 1:150, respectively. Smears incubated without primary antibody were used as negative controls. As DPP4/CD26 were expressed in luminal borders or secreted as intraluminal pattern, the positive results were determined when the immunoreaction exhibited any secretory expression. For SCG5 and CA12, positive findings were identified as the immunostainings were discernible in cytoplasm or membrane.

Two pathologists (Drs. Qiang L. and Luying Z.) blindly and independently evaluated the tissue sections to obtain the pathological results. DPP4, SCG5 and CA12 immunostainings were classified as follows: (-) = no reaction, (1+) = faint or weak positivity, (2+) = moderately positivity, (3+) = strong positivity. For statistical analysis, (2+) to (3+) were considered as "high" and (-) to (1+) were defined as "low."

Results

Genes differentially expressed in benign vs. malignant thyroid nodules identified by LIMMA

The original microarray expression training set (GSE29315) containing seven subtypes of thyroid neoplasias (40 benign and 31 malignant thyroid neoplasias) was preprocessed by using the R package simpleaffy[27] and normalized by robust multiarray averaging. Using LIMMA, a total of 43 probes, corresponding to 37 genes were found to be significantly differentially expressed between benign and malignant thyroid

tumors. Thirty and seven genes were up or downregulated in malignant compared with benign nodules, respectively (Supporting Information Table S2). Many of these genes are well-known markers of thyroid cancer, including FN1, TPO, MET, TFF3, TIMP1 and SERPINA1.[28]

A further feature selection step identified by the iterative BMA method associated with benign and malignant thyroid tumors

Using a small set of genes to diagnose thyroid tumors reduces the costs associated with microarrays and time-consuming data analysis; and will undoubtedly have clinical application. BMA outperformed standard microarray analysis by identifying fewer relevant genes, while still supplying better average predictive performance. The uncertainty in models are accounted for by averaging over all possible sets of models, weighted by their posterior model probabilities. It is a multivariate gene selection method which considers the relationship between various genes. LIMMA and BMA were therefore employed as a two-step feature selection method (Fig. 1).

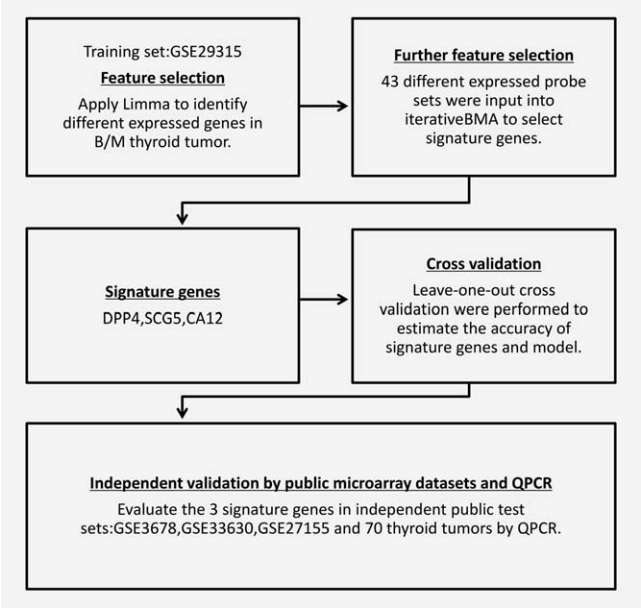

Figure 1. Study design. We combined two feature selection methods, LIMMA and iterative BMA. LIMMA was used to identity all genes differentially expressed between benign and malignant thyroid samples, after which iterative BMA was applied to reduce the number of signature genes and achieve relatively high prediction accuracy.

Iterative BMA identified three probe sets in three models (Supporting Information Fig. S1) to distinguish between benign and malignant thyroid nodules in our training set. The three probe sets 34823_at, 34265_at and 36464_at represent the genes DPP4, SCG5 and CA12, respectively. According to the BSS/WSS ratios, DPP4, SCG5 and CA12 were ranked as 1st, 19th and 37th, respectively (Table 1). From Supporting Information Figure S1 it is clear that each of the three models employed by iterative BMA contains one or two signature genes in the calculation. The posterior probability of each gene is the sum of the posterior probabilities of the models, which applied the signature gene in the BMA paradigm. For example, the posterior probability of DPP4 is the sum of the posterior probabilities of models 1, 2 and 3. Or, the posterior probabilities of DPP4 = 67.8% + 23.0% + 9.2% = 100%. The posterior probabilities of each signature gene are listed in Table 1. Next, a BMA classifier was built for the test samples by averaging over the three models, weighted by their posterior probabilities.

We adopted the leave-one-out classification method to assess the predictive performance of the three signature genes to verify the robustness of the BMA model. The 71 microarrays in the training set were randomly divided into 70 training samples and one testing sample until all samples served as a testing sample once. The BMA framework calculates the predicted probability of the testing sample according to the model computed by the training set. The cutoff value of the predicted probability to distinguish benign and malignant thyroid nodules is 0.5, so that samples with a predicted probability < 0.5 are classified as malignant and vice versa. The prediction results were compared with the true classifier of the test sample. Of the 71 samples, 31 were malignant and 40 were benign; and BMA correctly classified 36 of the 40 benign nodules and 23 of the 31 malignant nodules with prediction accuracy of 83.1% {85.2% sensitivity [95% confidence

Table 1. The three signature genes selected by iterative BMA, used to distinguish benign from malignant thyroid nodules

Gene symbol	Probe_id	Chr_location	Gene title	Gene function	Posterior probabilities	BSS/WSS rank
DPP4	34823_at	2q24.3	Dipeptidyl-peptidase 4	Aminopeptidase activity; protease binding	100%	1
SCG5	34265_at	15q13-q14	Secretogranin V	GTP binding; enzyme inhibitor activity	67.8%	19
CA12	36464_at	15q22	Carbonic anhydrase XII	Carbonate dehydratase activity; zinc ion binding	23.0%	37

Table 2. Prediction performance of the three signature genes detected by QPCR and in three public datasets

QPCR result[1]	Malignant reference standard ($N = 31$)	Benign reference standard ($N = 39$)
Malignant	29	2
Benign	2	37
BMA result[2,3]	**Malignant reference standard ($N = 60$)**	**Benign reference standard ($N = 45$)**
Malignant	55	10
Benign	5	35
BMA result[4,5]	**Malignant reference standard ($N = 78$)**	**Benign reference standard ($N = 21$)**
Malignant	62	5
Benign	16	16
BMA result[6,7]	**Malignant reference standard ($N = 7$)**	**Benign reference standard ($N = 7$)**
Malignant	5	0
Benign	2	7

[1]Sensitivity: 93.5%; specificity: 94.9%; accuracy: 94.3%.
[2]GSE33630 ($N = 105$)
[3]Sensitivity: 91.7%; specificity: 77.8%; accuracy: 85.7%; average Brier score: 0.1048.
[4]GSE27155 ($N = 99$)
[5]Sensitivity: 79.5%; specificity: 76.2%; accuracy: 78.8%; average Brier score: 0.182.
[6]GSE3678 ($N=14$)
[7]Sensitivity: 71.4%; specificity: 100%; accuracy: 85.7%; average Brier score: 0.135.

interval (CI), 77.0–93.5]; 81.8% specificity [95% CI, 72.8–90.8]} using only three genes.

The prediction performance of three signature genes in three independent public datasets

To reliably estimate the prediction performance using the three signature genes, we extracted expression data of the three signature genes from three independent public datasets from the GEO, for input into the BMA algorithm. Because one gene symbol can be represented by several probe sets, we chose the probe set with the highest *p*-value between benign and malignant tumor groups as calculated using LIMMA. All gene-expression values were normalized to minimize the effect of different microarray platforms and experimenters. For the test set (GSE33630), with 105 samples, the gene-expression classifier accurately identified 55 of the 60 malignant thyroid tumors with a sensitivity of 91.7% (95% CI, 86.4–97.0) and 35 of 45 benign nodules with a specificity of 77.8% (95% CI, 69.8–85.8). The average Brier score of the 105 test samples was 0.1048. For the GSE27155 dataset, the model correctly recognized 62 of 78 malignant samples with a sensitivity of 79.5% (95% CI, 71.5–87.5), and 16 of 21 benign samples with a specificity of 76.2% (95% CI, 67.8–84.6). The average Brier score of the 99 test samples was 0.182. For the test set GSE3678 with 14 microarrays, the classifier distinguished five of the seven malignant samples with a sensitivity of 71.4% (95% CI, 47.7–95.1) and all of the benign nodules, with a specificity of 100%. The average Brier score of the 14-sample test set was 0.135. As shown in Table 2, the prediction accuracies of the three datasets were 85.7, 78.8 and 85.7%, respectively, indicating good performance of the three signature genes as a gene-expression classifier.

Validation of the three signature genes in distinguishing benign from malignant thyroid tumors by quantitative polymerase chain reaction

To validate the performance of the three signature genes and the BMA model, we measured the expression of these genes by QPCR analysis for 70 samples. When the original samples were measured by QPCR, the relative expression level of DPP4 and SCG5 were significant upregulated in malignant group than in benign group and CA12 did not show much difference (Supporting Information Table S3 and Fig. 2). Because there was a significant discrepancy in the data scale between QPCR and microarray expression data, we had to refit the BMA classifier by leave-one-out cross validation. As shown in Figure 3*a*, the gene-expression classifier achieved high accuracy in all 70 samples. Most of the prediction probabilities of malignant tumors were extremely close to 0 and that of benign nodules were markedly near 1, which demonstrates great reliability. Notably, only two malignant thyroid tumors and two benign nodules were misclassified among the 70 samples. The performance of the QPCR-trained classifier was satisfactory with a sensitivity of 93.5% (95% CI, 87.8–99.2), a specificity of 94.9% (95% CI, 89.7–100), and a prediction accuracy of 94.3% (95% CI, 88.9–99.7; Table 2). To further confirm the discriminatory power of the three signature genes, we log-transformed the QPCR gene-expression data to plot the data in three-dimensional (3D) space. As shown in Figure 3*b*, the malignant samples could clearly be distinguished from the benign ones.

Validation by immunohistochemistry analysis

As the significance of the biological functions, we further validated the protein expression levels of DPP4, SCG5 and CA12 between benign and malignant tumors in 29 of our 70

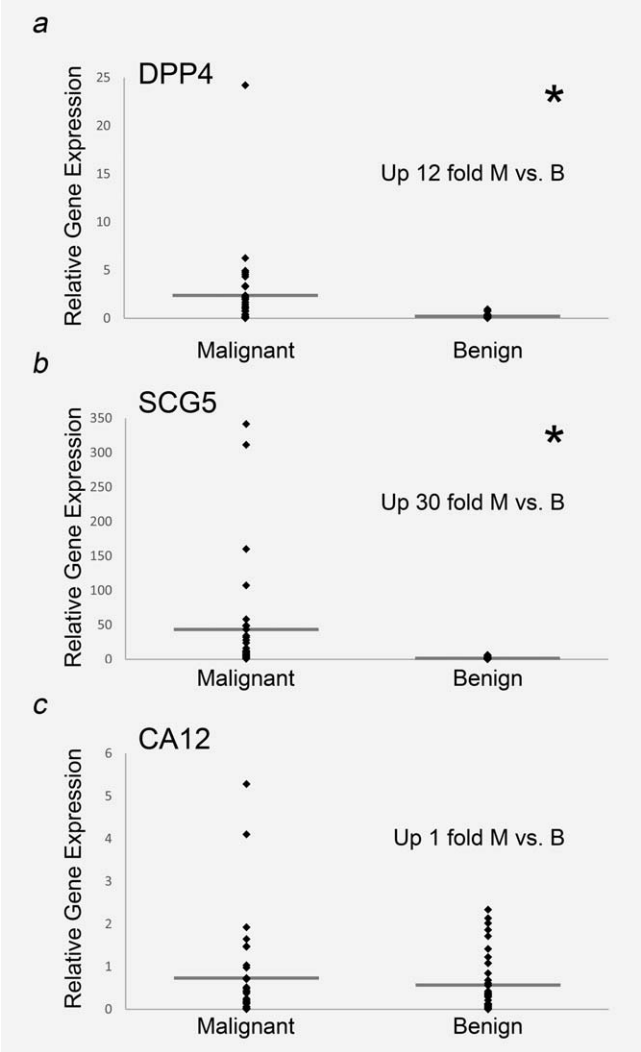

Figure 2. Scatter plot illustrating the relative levels of DPP4, SCG5 and CA12 gene expression determined by QPCR between malignant and benign thyroid nodules in 70 tumor samples. Horizontal lines indicate group means. B, benign thyroid tumor; M, malignant cancer. *, $p < 0.01$ by two-tailed t test between benign and malignant thyroid tumor types.

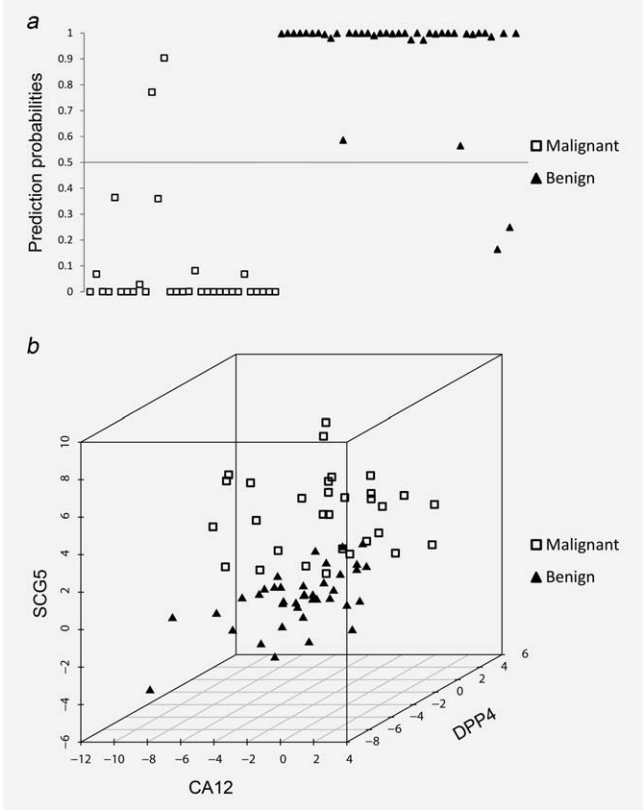

Figure 3. The performance of the three signature genes in predicting benign and malignant thyroid nodules by QPCR. (*a*) The black line in at $y = 0.5$ represents the cutoff of the prediction probability to classify benign and malignant nodules. (*b*) Scatter diagram of the gene expression profiles of the three signature genes in three-dimensional space. The QPCR gene-expression data were log-transformed. Squares and dots represent malignant and benign samples, respectively. These two groups can be visually divided into two clusters in 3D space, which suggests that the three signature genes perform well in predicting thyroid tumors.

original thyroid tissue samples by immunohistochemistry. All the three proteins were overexpressed in malignant tumors compared with benign thyroid nodules, which was in consonance with the result of QPCR (Fig.4*a* and Supporting Information Fig.S2). As we defined (2+) to (3+) as "high" expressions and (-) to (1+) as "low" expressions, immunostaining for DPP4, SCG5 and CA12 showed the individual diagnosis accuracy of malignant thyroid tumors with 86.2% (25/29), 82.8% (24/29) and 72.4% (21/29), respectively (Fig. 4*b*).

Discussion

The most pressing clinical challenge in thyroid tumor surgery is to avoid unnecessary thyroid lobectomies, especially total thyroidectomies, thereby reducing surgery-associated risks and avoiding the need for long-term thyroid hormone replacement therapy.[29] Clearly, there is a compelling need for improved diagnostic tools to distinguish benign from malignant tumors. Numerous studies have harnessed high-throughput technologies to develop diagnostics[8,17,25,30,31]––a strategy that could decrease the rate of indeterminate or suspicious outcomes diagnosed by FNA cytology,[8] as well as help elucidate the pathogenesis of thyroid carcinoma.

Machine learning and knowledge mining have been applied in bioinformatics field for many years.[13,17] Previously, diagnostic studies have applied machine-learning techniques to all significantly differentially expressed genes. For example, Alexander *et al.*[8] recently screened 142 differentially expressed genes to diagnose benign and malignant thyroid nodules by support vector machines (SVM); Prasad *et al.*[32] revealed 75 differentially expressed transcripts and predicted thyroid tumors using nearest-neighbor classification; Cerutti *et al.*[33] identified 73 differentially expressed genes to preoperatively diagnose thyroid carcinoma. However, these

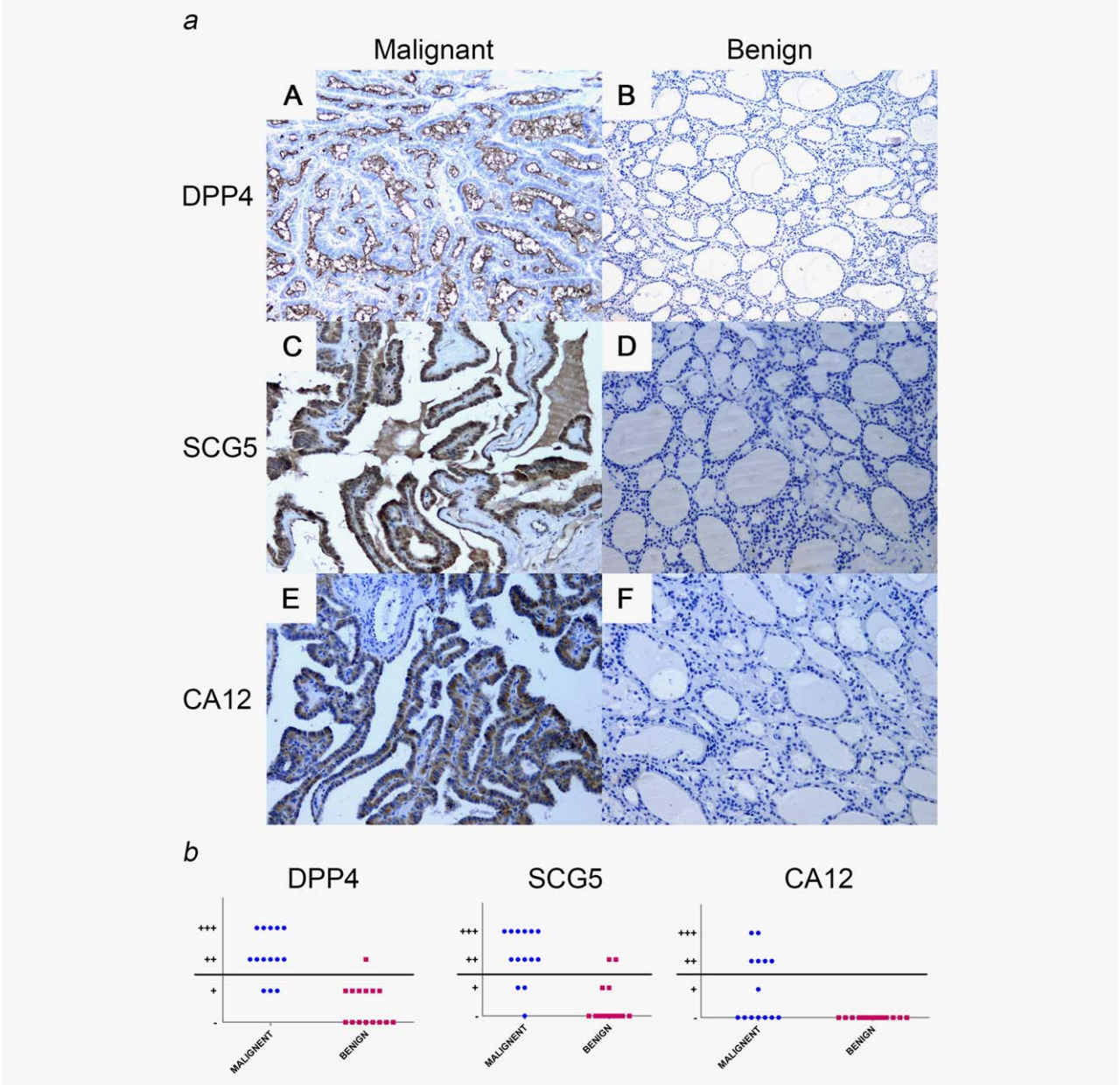

Figure 4. Immunohistochemical analysis of DPP4/CD26, SCG5 and CA12. (*a*) Immunohistochemical staining of some representative thyroid tissue samples from papillary thyroid carcinoma compared with follicular adenoma. Magnification 200×. (*b*) Scatter diagram of the Immunohistochemistry results from 14 papillary thyroid carcinoma and 15 follicular adenoma samples. The black line between (+) and (++) were the boundary between low and high immunostaining as we defined in the methods.

microarray-based gene-expression profiles may be too complex and time-consuming to apply in a clinical setting to diagnose thyroid cancer.

In this study, we developed a simple and affordable diagnostic by selecting a small group of signature genes to distinguish benign from malignant thyroid nodules and their prediction performances were further validated by immunohistochemistry analysis. A public dataset composed of the majority of common thyroid lesions (FA, FTC, PTC, FVPC, HT, TH and HCA) was used as the training set to find signature genes and build the classifier. To this end, we used a two-step feature selection procedure: in addition to the widely used LIMMA method (used to assess differential gene-expression), we added the iterative BMA algorithm because it is a multivariate gene selection method, considering the relationship between each gene to eliminate highly correlated ones. Iterative BMA has been successfully applied to breast cancer and leukemia data as a gene signature selection method, with typically only a few genes being selected.[14] After the two-step feature selection procedure, three signature genes, DPP4, SCG5 and CA12, with respective posterior probabilities of 100, 67.8 and 23%, respectively, were

identified. The univariate rankings (BSS/WSS rank) of the three genes were 1st, 19th and 37th indicating that genes used in combination, even with poor individual rankings, can still have good predictive power when combine together with other genes.

Of the three signature genes, DPP4 had the highest univariate ranking and posterior probability. DPP4 (CD26) encodes a membrane-bound serine protease, a member of the dipeptidyl prolyl peptidase (DPP) family.[34] Our findings are supported by literature reports that this gene significant increases in differentiated carcinomas vs. normal or benign thyroid nodules at average 46.19 times.[28,35] In addition, some studies showed that DPP4 was downregulated in ATC with unclear mechanism,[36,37] which was consistent with the expression level of our ACT specimen by QPCR. As DPP4 is involved in immune function, cell adhesion, and modulating cell proliferation and extracellular matrix disassembly, many studies have reported it as a marker of various cancers, inflammatory diseases, and diabetes mellitus.[34] Currently, there is considerable interest in investigating the use of DPP4 as a biomarker in cancer diagnosis and prognosis because it is increased in various tumors (such as prostate cancer, mesothelioma, hepatocellular cancer and thyroid cancer), and the DPP4 protein can be detected in plasma––which is particularly useful in the clinical setting.[38] Interestingly, in contrast to the elevated levels of DPP4 in tumor tissue, plasma concentrations of DPP4 were lower in cancer patients than in healthy subjects[34]; this apparent contradiction requires further investigation.

The second gene in our signature panel, SCG5, encodes the neuroendocrine protein secretogranin V (7B2 protein). Importantly, SCG5 has been reported to be significantly upregulated more frequently in papillary thyroid carcinoma[39,40] compared with other subtypes of thyroid cancers. This result is in agreement with our study because our malignant thyroid tissue samples tested by QPCR consisted primarily of PTC, and SCG5 expression levels were largely responsible for the classification of benign vs. malignant thyroid tumors. The 7B2 protein encoded by SCG5 is a small neuroendocrine protein, which can be detected in plasma; Ohashi *et al.*[41] demonstrated the use of 7B2 detection in plasma as a biomarker of medullary carcinoma of the thyroid and showed that 7B2 is released from parafollicular cells of the thyroid, which lead to high 7B2 plasma levels in patients with medullary carcinoma of the thyroid.

The third gene in our signature panel, CA12, encodes the protein that belongs to the carbonic anhydrases family, which is involved in various biological processes like one-carbon metabolism and small molecule metabolic processes. Despite the relatively low posterior probability of CA12 and modest performance between malignant and benign groups measured by QPCR in our results, the literature supports the oncogenic role of CA12 in thyroid cancer,[42] and the usefulness of CA12 as a tumor marker.[40,43,44]

Despite the small number of genes in our panel, the ability to predict different categories of thyroid nodules was similar to those published data from over 100 genes.[8,31] Another advantage of our approach was that the small number of genes could be found on various microarray platforms, which made it possible to take full advantage of the public microarray datasets to identify and verify the panel. Last and most intriguingly, all three of our signature genes could be detected in plasma, which has important potential as a noninvasive and affordable clinical diagnostic tool. Prasad *et al.*[25] has also identified a three-gene molecular diagnostic model, including high mobility group AT-Hook 2 (HMGA2) + mannose receptor, C type 2 (MRC2) + stratifin (SFN), to distinguish thyroid tumor with similar accuracy as our panel. However, HMGA2 does not exist in plasma and this would determinately set a limit for its application in noninvasive diagnosis. Further research may benefit from studying these genes' role in cancer as well as in other diseases.[38]

In summary, we developed a two-step feature selection procedure, which implements a multivariate BMA model, to identify a small signature gene panel to preoperatively diagnose benign vs. malignant thyroid nodules. Because this signature gene panel consists of only three genes, expression can easily be assessed by QPCR, we believe this can form the basis for a robust, practical, and inexpensive molecular diagnostic tool for clinical use in the near future. Furthermore, we will test the relationship between the concentrations of each of the three proteins in our panel against each thyroid nodule category to develop the use of plasma-based biomarker detection.

Acknowledgements

We are especially grateful to Dr. Chunmei Ying, Dr. Xiaojie Jin for their advice in QPCR experiments, to pathologists Dr. Qiang Liu and Dr. Luying Zheng for their help in immunohistochemistry analysis.

References

1. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2011 Incidence and Mortality Web-based Report.
2. Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Prac* 2010;16:1–43.
3. Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association Management Guidelines for patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association (ATA) guidelines taskforce on thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009;19:1167–214.
4. Choi YS. Clinical and ultrasonographic findings affecting nondiagnostic results upon the second fine needle aspiration for thyroid nodules. *Ann Surg Oncol* 2012;19:2304–9.
5. Park KT, Ahn SH, Mo JH, et al. Role of core needle biopsy and ultrasonographic finding in management of indeterminate thyroid nodules. *Head Neck* 2011;33:160–5.
6. Orija IB, Piñeyro M, Biscotti C, et al. Value of repeating a nondiagnostic thyroid fine-needle aspiration biopsy. *Endocr Pract* 2007;13:735–42.
7. Kim MI, Alexander EK. Diagnostic use of molecular markers in the evaluation of thyroid nodules. *Endocr Pract* 2012;18:796–802.
8. 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.
9. Chudova D, Wilde JI, Wang ET, et al. Molecular classification of thyroid nodules using high-dimensionality genomic data. *J Clin Endocrinol Metab* 2010;95:5296–304.
10. Barden CB, Shister KW, Zhu B, et al. Classification of follicular thyroid tumors by molecular signature results of gene profiling. *Clin Cancer Res* 2003;9:1792–800.
11. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001; 98:5116–21.
12. Nguyen DV, Rocke DM. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* 2002;18:39–50.
13. Gu JL, Lu Y, Liu C, Lu H. Multiclass classification of sarcomas using pathway based feature selection method. *J Theor Biol* 2014; published online. [Eput ahead of print]
14. Yeung KY, Bumgarner RE, Raftery AE. Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics* 2005;21:2394–402.
15. Oehler VG, Yeung KY, Choi YE, et al. The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data. *Blood* 2009;114:3292–8.
16. Chevillard S, Ugolin N, Vielh P, et al. Gene expression profiling of differentiated thyroid neoplasms diagnostic and clinical implications. *Clin Cancer Res* 2004;10:6586–97.
17. Finley DJ, Zhu B, Barden CB, Fahey TJ, III. Discrimination of benign and malignant thyroid nodules by molecular profiling. *Ann Surg* 2004; 240: 425–36.
18. Giordano TJ, Kuick R, Thomas DG, et al. Molecular classification of papillary thyroid carcinoma: distinct BRAF, RAS, and RET/PTC mutation-specific gene expression profiles discovered by DNA microarray analysis. *Oncogene* 2005;24: 6646–56.
19. Giordano TJ, Au AY, Kuick R, et al. Delineation, functional validation, and bioinformatic evaluation of gene expression in thyroid follicular carcinomas with the PAX8-PPARG translocation. *Clin Cancer Res* 2006;12:1983–93.
20. Data downloaded at NCBI GEO database. at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3678.
21. Gentleman R, Carey V, Huber W, et al. Bioinformatics and computational biology solutions using R and Bioconductored., vol. 746718470. New York: Springer, 2005.
22. Amalia A, Roger EB, Adrian ER, and Ka YY. Iterative Bayesian Model Averaging: a method for the application of survival analysis to high-dimensional microarray data. *BMC Bioinformatics* 2009;10:72.
23. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc* 2002;97:77–87.
24. Furnival GM, Wilson RW. Regressions by leaps and bounds. *Technometrics* 1974;16:499–511.
25. Prasad NB, Kowalski J, Tsai H-L, et al. Three-gene molecular diagnostic model for thyroid cancer. *Thyroid* 2012;22:275–84.
26. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 1950;78:1–3.
27. Miller CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics* 2005;21:3683–3685.
28. Griffith OL, Melck A, Jones SJ, Wiseman SM. Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *J Clin Oncol* 2006;24: 5043–51.
29. Gharib H, Goellner JR, Zinsmeister AR, et al. Fine-needle aspiration biopsy of the thyroid: the problem of suspicious cytologic findings. *Ann Inter Med* 1984;101:25–8.
30. Mazzanti C, Zeiger MA, Costourous N, et al. Using gene expression profiling to differentiate benign versus malignant thyroid tumors. *Cancer Res* 2004;64:2898–903.
31. Finley DJ, Zhu B, Barden CB, Fahey TJ, III. Discrimination of benign and malignant thyroid nodules by molecular profiling. *Ann Surg* 2004; 240:425–36.
32. Prasad NB, Somervell H, Tufano RP, et al. Identification of genes differentially expressed in benign versus malignant thyroid tumors. *Clin Cancer Res* 2008;14:3327–37.
33. Cerutti JM, Delcelo R, Amadei MJ, et al. A preoperative diagnostic test that distinguishes benign from malignant thyroid carcinoma based on gene expression. *J Clin Invest* 2004;113:1234–42.
34. Javidroozi M, Zucker S, Chen W-T. Plasma seprase and DPP4 levels as markers of disease and prognosis in cancer. *Dis Markers* 2012;32: 309–20.
35. Ducena K, Ābols A, Vilmanis J, et al. Validity of multiplex biomarker model of 6 genes for the differential diagnosis of thyroid nodules. *Thyroid Res* 2011;4:11.
36. Aratake Y, Nomura H, Kotani T, et al. Coexistent anaplastic and differentiated thyroid carcinoma an immunohistochemical study. *Am J Clin Pathol* 2006;125:399–406.
37. Aratake Y, Umeki K, Kiyoyama K, et al. Diagnostic utility of galectin-3 and CD26/DPPIV as preoperative diagnostic markers for thyroid nodules. *Diagnos Cytopathol* 2002;26:366–72.
38. Sedo A, Stremenova J, Busek P, Duke-Cohan JS. Dipeptidyl peptidase-IV and related molecules: markers of malignancy? *Expert Opin Med Diagnos* 2008;2:677–89.
39. Nikolova DN, Zembutsu H, Sechanov T, et al. Genome-wide gene expression profiles of thyroid carcinoma: identification of molecular targets for treatment of thyroid carcinoma. *Oncol Rep* 2008; 20:105–21.
40. Stein L, Rothschild J, Luce J, et al. Copy number and gene expression alterations in radiation-induced papillary thyroid carcinoma from chernobyl pediatric patients. *Thyroid* 2010;20:475–87.
41. Ohashi M NS, Fujio N, Iguchi H, Nawata H. Secretory protein 7B2. A novel tumor marker of medullary carcinoma of the thyroid. *Horm Metab Res* 1990;22:114–6.
42. Ilie MI, Lassalle S, Long-Mira E, et al. In papillary thyroid carcinoma, TIMP-1 expression correlates with BRAF V600E mutation status and together with hypoxia-related proteins predicts aggressive behavior. *Virchows Arch* 2013:1–8.
43. Fryknäs M, Wickenberg-Bolin U, Göransson H, et al. Molecular markers for discrimination of benign and malignant follicular thyroid tumors. *Tumor Biol* 2006;27:211–20.
44. Gomez Saez JM. Diagnostic usefulness of tumor markers in the thyroid cytological samples extracted by fine-needle aspiration biopsy. *Endocr Metab Immune Disord Drug Targets* 2010;10: 47–56.