High Diagnostic Accuracy Based on *CLDN10*, *HMGA2*, and *LAMB3* Transcripts in Papillary Thyroid Carcinoma

Mateus Camargo Barros-Filho, Fabio Albuquerque Marchi, Clóvis Antônio Pinto, Silvia Regina Rogatto,* and Luiz Paulo Kowalski*

International Research Center/AC Camargo Cancer Center (M.C.B.F., F.A.M., C.A.P., S.R.R., S.R.R.), Sao Paulo 01509-010, SP, Brazil; and Faculty of Medicine (S.R.R.), Sao Paulo State University, Botucatu 18618-970, SP, Brazil

Context: Thyroid nodules are common in adult population and papillary thyroid carcinoma (PTC) is the most frequent malignant finding. The natural history of PTC remains poorly understood and current diagnostic methods limitations are responsible for a significant number of potentially avoidable surgeries.

Objective: This study aimed to identify molecular markers to improve the diagnosis of thyroid lesions.

Design: Gene expression profiling was performed using microarray in 61 PTC and 13 surrounding normal tissues (NT). A reliable gene list was established using cross-study validation (138 matched PTC/NT from external databases). Results were collectively interpreted by in silico analysis. A panel of 28 transcripts was evaluated by RT-qPCR, including benign thyroid lesions (BTL) and other follicular cell-derived thyroid carcinomas (OFDTC). A diagnostic algorithm was developed (training set: 23 NT, 8 BTL, and 86 PTC), validated (independent set: 10 NT, 140 BTL, 120 PTC, and 12 OFDTC) and associated with clinical features.

Results: *GABRB2* was ranked as the most frequently up-regulated gene in PTC (cross-study validation). Altered genes in PTC suggested a loss of T_4 responsiveness and dysregulation of retinoic acid metabolism, highlighting the putative activation of *EZH2* and histone deacetylases (predicted in silico). An algorithm combining *CLDN10*, *HMGA2*, and *LAMB3* transcripts was able to discriminate tumors from BTL samples (94% sensitivity and 96% specificity in validation set). High algorithm scores were associated with regional lymph node metastases.

Conclusions: A promising tool with high performance for PTC diagnosis based on three transcripts was designed with the potential to predict lymph node metastasis risk. **(*J Clin Endocrinol Metab* 100: E890–E899, 2015)**

Well-differentiated thyroid carcinomas are generally associated with an indolent disease course and are comprised principally by papillary thyroid carcinoma (PTC) subtype (80–85%) (1). Aside from radiation exposure and low consumption of iodine, the risk factors that contribute to the development of these tumors are not well understood (2). The molecular pathogenesis of PTC is related to genetic and epigenetic alterations, which lead to MAPK and phosphatidylinositol 3-kinase pathway activation (1). Predominant genetic alterations include mutations on *BRAF* and *RAS*, and *RET/PTC* rearrangements, which are mutually exclusive, and found in nearly

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA

Received November 10, 2014. Accepted April 8, 2015.
First Published Online April 13, 2015

* S.R.R. and L.P.K. contributed equally to the study.

Abbreviations: AUC, area under the curve; BTL, benign thyroid lesion; NT, normal tissues; CI, confidence interval; FC, fold change; FNAB, fine-needle aspiration biopsy; FTC, follicular thyroid carcinoma; GEO, Gene Expression Omnibus; IPA, Ingenuity Pathways Analysis; NIS, sodium-iodide symporter; NT, normal thyroid; OFDTC, other follicular cell-derived thyroid carcinomas; PTC, papillary thyroid carcinoma; RA, retinoic acid; RAR, retinoic acid receptor; RARA, retinoic acid receptor alpha; RXR, retinoid X receptor; TLDA, TaqMan Low Density Arrays; TR, thyroid hormone receptor.

70% of PTC (3). Overexpression of several growth factors and their receptors, dysregulation of cell cycle regulators, and adhesion molecules are well-known events in thyroid tumorigenesis and progression (2). Nonetheless, only few studies have comprehensively investigated the biological functions related to differentially expressed genes in PTC, which may influence carcinogenesis and have potential as treatment targets (4, 5).

In recent years, PTC has presented the largest increase in overall incidence among all cancers (6). It has been suggested that this is a consequence of improvement in small tumor detection, in particular because of the widespread use of ultrasound and fine-needle aspiration biopsy (FNAB) (7). In fact, PTC is a common autopsy finding in individuals who have died from other causes (8). Some cases have no propensity to develop into clinically apparent thyroid disease (9). Furthermore, approximately 17% (10–26%) of FNAB are reported as indeterminate, of which only one third is proven to be malignant, mostly represented by the follicular variant of PTC (10). Consequently, many patients are exposed to unnecessary surgical risk (11), requiring lifelong thyroid hormone replacement therapy and considerably increasing the healthcare burden (12). Therefore, it is paramount to identify molecular markers to aid in the diagnosis of thyroid lesions (13).

Well-differentiated thyroid carcinomas have a high occurrence of known genetic alterations, mainly *BRAF*V600E in PTC. Therefore, its detection via FNAB has been proposed; however, low sensitivity is frequently seen (14). Thyroid cancer diagnostic profiles have been investigated using mRNA, miRNA, and methylation assays in tumor samples obtained from both surgical and FNAB specimens (13, 15). Although a high accuracy has been demonstrated by these studies, most have failed to show optimal sensitivity and specificity in the validation analysis using independent sample sets. Molecular profiling analyses usually yield a large number of potential biomarkers; however, small sets would be more viable in clinical practice, with the use of RT-qPCR or immunohistochemistry (16).

Therefore, mRNA global expression analysis comparing PTC to normal thyroid tissue (NT) was performed in this study. Data were confirmed and ranked by a novel cross-study validation method. Molecular in silico analysis identified critical dysregulated canonical pathways in PTC, revealing potential therapeutic targets for advanced PTC. In addition, a combination of three targets and two reference transcripts was able to discriminate malignant tumors from normal/benign thyroid tissues, with both high sensitivity and specificity.

Materials and Methods

Sample population

PTC samples from 206 patients submitted to total thyroidectomy were studied (Supplemental Table 1). Samples from 148 benign thyroid lesions (BTL) (follicular adenoma, nodular hyperplasia, nodular goiter, lymphocytic thyroiditis), as well as 12 other follicular cell-derived thyroid carcinomas (OFDTC) (poorly differentiated, anaplastic, follicular, and Hürthle cell carcinoma) were also included. A total of 33 normal thyroid tissues (NT) were obtained from morphologically normal areas adjacent to the tumors, with 32 having been paired to their corresponding PTC. The histological evaluation of the samples was performed by an experienced pathologist (C.A.P.) using blinded interpretation. In case of discrepancies, a final consensus was reached by three specialized pathologists from A.C. Camargo Cancer Center. The Ethics Committee in Human Research for the Institution approved this study (Protocol No. 1410/10). The general study design is summarized in Figure 1.

Microarray expression analysis

RNA was isolated from fresh frozen thyroid lesions and NT using TRIZOL (Invitrogen) or the RNeasy mini kit (QIAGEN), according to the manufacturers' instructions. RNA integrity was verified using the Agilent 2100 Bioanalyzer RNA 6000 LabChip kit (Agilent Technologies, Inc.).

Global gene expression analysis was conducted using the SurePrint G3 8 × 60K microarray platform (Agilent Technologies, Inc.) for 57 PTC samples and nine NT samples (pool), which were used as a reference. Probes with a $\log_2$ mean ratio (Cy3/Cy5) >1 and <-1 with a 95% confidence interval (CI) were considered as differentially expressed, for up and down-regulated genes, respectively. Similarly, other four PTC and their corresponding NT samples were compared by Significance Analysis of Microarray (SAM) (MeV v.4.8.1). Transcripts presenting at least 2-fold change (FC) and less than 5% of false discovery ratio were defined as significant. Probes identified by both comparisons (57 PTC vs nine unpaired NT and four PTC vs four paired NT) were further investigated. Microarray data have been deposited into the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/gds/) with the accession number GSE50901. Details of labeling, hybridization, normalization, and filters quality are described in the Supplemental Methods.

Cross-study validation

Transcription profiling series from 138 matched PTC and NT were obtained from the publicly available databases GEO and The Cancer Genome Atlas (TCGA) (http://tcga-data.nci.nih.gov/tcga/) to confirm our gene list. Data from both microarray (82 paired samples) and RNA sequencing (56 paired samples) analyses were included (Supplemental Table 2). Statistical analysis was applied separately for each platform, employing the nonparametric sign test with Bonferroni correction. For ranking purposes, data from both platforms were merged. As the method relies on sign test, data adjustments from different studies were dismissed. To identify the most promising candidate transcripts, a ranking criterion was established, primarily based on the sign test *P* value. Detailed descriptions can be found in the Supplemental Methods and Supplemental Figure 1.

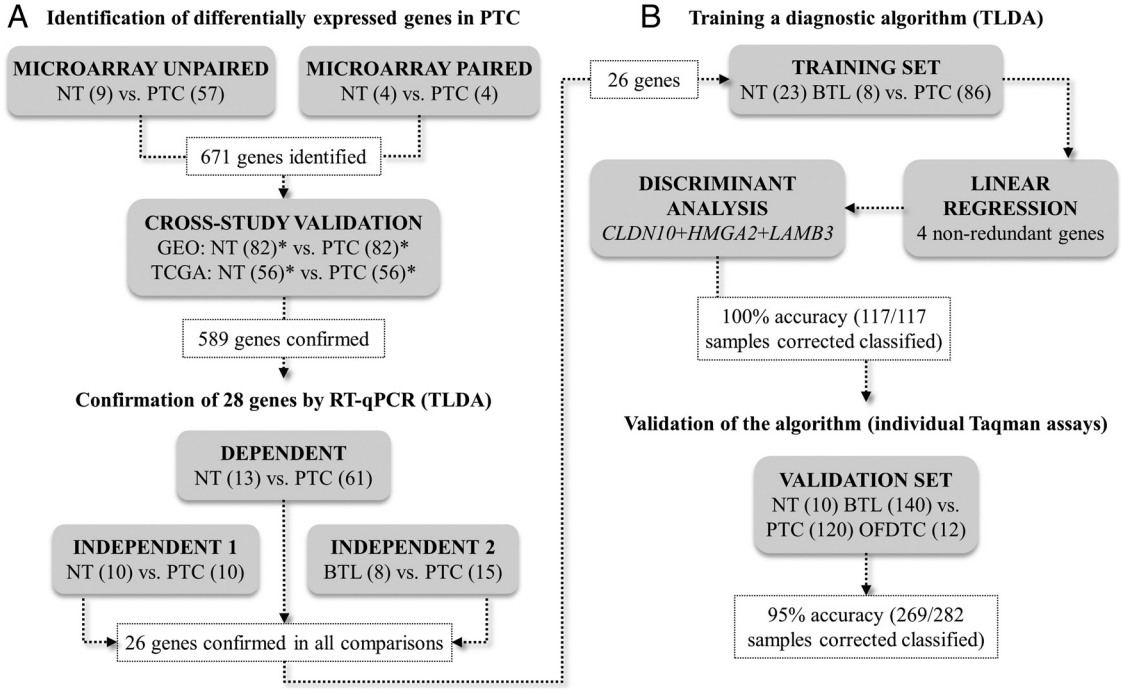


Figure 1. Flowchart illustrating the study design. A, By combining two independent microarray analyses, using paired and unpaired tumor (PTC) and nontumor (NT) samples, 671 genes were identified as differentially expressed in PTC. A list of 589 genes was confirmed with a cross-study validation method. Twenty-eight selected genes were further evaluated by RT-qPCR (TLDA) in both microarray-dependent and independent sample sets (including both paired and unpaired samples) and 26 genes were confirmed in all comparisons. B, A diagnostic algorithm was trained, filtering four nonredundant markers from the 26 genes tested (linear regression) and using *CLDN10*, *LAMB3*, and *HMGA2* relative expression (discriminant analysis), which could classify correctly all tumors from normal/benign samples. The models were validated in an independent set (validation set), classifying correctly 95% of the samples. PTC, papillary thyroid carcinoma; NT, surrounding normal thyroid tissue; BTL, benign thyroid lesions; OFDTC, other follicular cell-derived thyroid carcinomas; *, in silico data.

In silico molecular interaction analysis

Cross-study validated genes were submitted to biological function, canonical pathway, and upstream regulator analysis using the Ingenuity Pathways Analysis software (IPA; Ingenuity Systems). Pathway analysis was performed via KOBAS (v2.0; http://kobas.cbi.pku.edu.cn/home.do), which uses the BioCyc, Kyoto Encyclopedia of Genes and Genomes Pathway, Pathway Interaction Database, Reactome, and Panther databases. Protein–protein interactions were assessed with known and predicted physical interactions using I2D v2.0 (http://ophid.utoronto.ca/i2d). The resulting networks were visualized using NAViGaTOR (v2.3; http://ophid.utoronto.ca/navigator).

RT-qPCR

Twenty-eight altered genes in PTC were selected for RT-qPCR analysis. The selection criteria were based on: 1) cross-study validation ranks; 2) Additional cross-study validation, which compared BTL with PTC, to uncover markers with an improved diagnostic value (detailed in Supplemental Methods); 3) In silico molecular analysis; 4) Highest fold changes. RT-qPCR analysis was carried out using TaqMan Low Density Arrays (TLDA) to verify the expression of the selected target and reference transcripts (Supplemental Table 3) in 117 samples (86 PTC, 23 NT, and 8 BTL). Four targets (*CLDN10*, *GABRB2*, *HMGA2*, and *LAMB3*) and two reference transcripts (*EIF2B1* and *PUM1*, selected by geNorm) were further assessed in 282 additional samples (120 PTC, 10 NT, 140 BTL, and 12 OFDTC) using individual Taqman assays (detailed in Supplemental Methods).

Statistical analysis

Statistical analysis was performed using the SPSS (v21.0; SPSS) and GraphPad Prism software (v5.0; GraphPad Software Inc.), adopting a two-tailed $P < .05$ value as significant. RT-qPCR results were correlated with microarray data (Pearson correlation test) and compared between samples (NT, BTL, PTC, and OFDTC) using unpaired and paired *t* test and one-way ANOVA (Tukey post-hoc test). Clinical and histopathological parameters were associated with expression data (unpaired *t* test). The diagnostic algorithm was developed with linear discriminant analysis and the algorithm scores were associated with clinical parameters (linear-by-linear association test) (detailed in Supplemental Methods).

Results

Identification of genes consistently differentially expressed in PTC

Gene expression analysis of the initial set of 57 PTC samples against a pool of nine unpaired NT samples revealed 1382 significant probes (FC > 2.0; CI 95%). An additional analysis of four paired PTC and NT samples revealed 1413 probes (FC > 2.0; false discovery ratio < 5%). A list of 774 common probes, corresponding to 671 genes, was identified as differentially expressed in PTC in both comparisons. *GABRB2* was the main up-regulated

(FC = 117) and *MYOC* the main down-regulated (FC= −31) genes in PTC, respectively.

A cross-study validation was performed to confirm the initial gene list (671 genes) (Supplemental Figure 1). Microarray expression data from PTC and NT cases collected from the GEO database supported the differential expression of 631 genes, whereas the RNA sequencing data from TCGA confirmed 612 genes. Probes/isoforms representing 589 genes, confirmed by both GEO and TCGA analyses, were merged; statistical tests were then reapplied and ranked. The genes within the first decile (ranks 1–59), including the top-ranked gene, *GABRB2*, are shown in Table 1. Genes rated on the remaining deciles are presented in Supplemental Table 4. The location of these genes was evenly distributed through all chromosomes (Supplemental Figure 2).

In silico prediction of dysregulated biological functions and pathways in PTC

In silico prediction by the IPA software (589 genes) revealed an inhibited state of anoikis and apoptosis, as well as an activated state of the metastatic process, adhesion of connective tissue cells, metabolism of reactive oxygen species, and inflammatory related-functions (Supplemental Table 5). Mechanisms involved in the inflammatory process were identified as the main dysregulated canonical pathways when using IPA and KOBAS (Supplemental Table 6 and 7). Canonical pathways involving retinoic acid (RA) receptors also seemed to be overrepresented. Dysregulation of these pathways may result in the loss of RA and T_3 signaling (Supplemental Figure 3A).

Upstream transcriptional regulator analysis predicted an inhibition of SMAD7 and NANOG and activation of the HDAC family and EZH2, which transcription levels were shown to be overexpressed in PTC compared with NT samples. This prediction was supported by the expected expression direction of the downstream factors in our gene list (Supplemental Figure 3B). The EZH2 protein–protein physical interaction analysis indicated RA receptor alpha (RARA) and HDAC1 as interaction partners, which could display links with retinoic acid receptor (RAR) and retinoid X receptor (RXR)/thyroid hormone receptor (TR) activation. In addition, both EZH2 and HDAC are susceptible to pharmacological inhibition (Supplemental Figure 3C).

Validation of the selected transcripts by RT-qPCR

All 28 selected transcripts were confirmed by analyzing both microarray dependent (61 PTC and 13 NT samples, Pearson test $P < .05$; median r = 0.923) and independent group of samples (10 paired PTC/NT samples; paired *t* test $P < .05$). A third comparison, including eight BTL and 15 PTC samples, revealed 26/28 genes with differential expression (*t* test $P < .05$), with the exception of *FOXP2* and *TFF3*. In a multivariable analysis, *CDLN10*, *GABRB2*, *HMGA2*, and *LAMB3* showed low redundancy ($P < .05$) with a high diagnostic value (area under the curve (AUC) > 0.9; CI 95%) (Supplemental Table 8). All genes showed significant differential expression when NT (n = 23) was grouped with BTL (n = 8) and they were then compared with PTC (n = 86) (unpaired *t* test $P < .05$) (Figure 2).

The association between gene expression (RT-qPCR) and clinical parameters revealed that *CDLN10*, *HMGA2*, and *LAMB3* were related to more aggressive tumor features, including lymph node involvement, extrathyroidal extension, vascular invasion, and a larger primary tumor (Supplemental Table 9).

Molecular diagnostic algorithm

Expression levels of *CLDN10*, *GABRB2*, *HMGA2*, and *LAMB3* were assessed increasing the sample size (Figure 3A). An algorithm was trained (23 NT, 8 BTL, and 86 PTC) and the results revealed that the combination of *CLDN10*, *HMGA2*, and *LAMB3* genes expression were able to differentiate the samples according to malignant status. Applying the same algorithm in an independent set of cases (validation set: 10 NT, 140 BLT, 120 PTC, and 12 OFDTC), 95% of them were correctly classified (94% sensitivity and 96% specificity in separating tumors from BTL) (Figure 3, B and C). *GABRB2* alone was able to correctly discriminate 96% of training samples and 91% from the validation set (85% sensitivity and 97% specificity in separating tumors from BTL) (Figure 3C). Classification according to histological type is detailed in Supplemental Table 10. Consistent with the association observed between higher *CLDN10*, *HMGA2*, and *LAMB3* expression levels and poor prognostic characteristics, higher algorithm scores were strongly associated with lymph node involvement (*t* test $P < .001$). By categorizing the algorithm scores in benign range, low, intermediate, and high, the risk of lymph node disease at diagnosis were 13, 23, 36, and 50%, respectively (linear by linear association $P < .001$) (Figure 4A).

Molecular diagnostic algorithm in external data

The use of *GABRB2* as a single diagnostic marker and the three gene–based algorithm (*CLDN10*/*HMGA2*/*LAMB3*) were tested using TCGA database (500 PTC and 58 NT samples). *GABRB2* showed 97% of accuracy (98% of sensitivity and 91% of specificity in distinguish PTC from NT) and *CLDN10*/*HMGA2*/*LAMB3* presented 92% of accuracy (91% of sensitivity and 93% of specificity in distinguish PTC from NT) (Supplemental

Table 1. Differentially Expressed Genes in PTC Confirmed with External Data in the First Decile of the Ranking Procedure

		Current study (Agilent 8×60K)		Cross-study (Affymetrix microarray/Illumina HiSeq)				Cross-study Ranking	
Gene Symbol	Description	Sig/Total Probes	FC	Total Transcripts	Down	Up	Sign test *P*[a]	Rank	Decile
GABRB2	γ-aminobutyric acid (GABA) A receptor, β 2	1/1	116.6	10	0	8	5.74E-42	1	1
TMEM171	transmembrane protein 171	1/1	−5.4	4	4	0	5.74E-42	2	1
TFF3	trefoil factor 3 (intestinal)	3/3	−23.9	2	2	0	5.74E-42	3	1
LIPH	lipase, member H	1/1	19.7	4	0	3	9.18E-41	4	1
GALE	UDP-galactose-4-epimerase	1/1	4.8	12	0	11	7.98E-40	5	1
CSGALNACT1	chondroitin sulfate N-acetylgalactosaminyltransferase 1	2/2	−4.3	8	8	0	7.98E-40	6	1
MPZL2	myelin protein zero-like 2	1/1	5.6	6	0	6	7.98E-40	7	1
TFCP2L1	transcription factor CP2-like 1	1/1	−11.6	7	6	0	7.98E-40	8	1
P4HA2	prolyl 4-hydroxylase, α polypeptide II	2/2	3.5	8	0	6	7.98E-40	9	1
COL13A1	collagen, type XIII, α 1	1/1	5.6	25	0	6	7.98E-40	10	1
GGCT	γ-glutamyl cyclotransferase	2/2	5.4	6	0	5	7.98E-40	11	1
XPR1	xenotropic and polytropic retrovirus receptor	1/2	3.2	6	0	5	7.98E-40	12	1
TPPP	tubulin polymerization promoting protein	1/1	−7.2	6	5	0	7.98E-40	13	1
EPHB1	EPH receptor B1	1/1	−3.3	8	5	0	7.98E-40	14	1
EPS8	epidermal growth factor receptor pathway substrate 8	1/1	5.2	8	0	4	7.98E-40	15	1
CITED2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	2/2	−3.3	4	3	0	7.98E-40	16	1
GDF15	growth differentiation factor 15	1/1	15.5	4	0	3	7.98E-40	17	1
PLCH1	phospholipase C, η 1	2/2	−3.7	8	3	0	7.98E-40	18	1
METTL7B	methyltransferase like 7B	1/1	7.7	2	0	2	7.98E-40	19	1
MMRN1	multimerin 1	1/1	−15.4	4	4	0	1.58E-39	20	1
OCA2	oculocutaneous albinism II	1/1	−10.2	3	2	0	1.58E-39	21	1
SLC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4	1/1	−11.1	13	10	0	5.51E-38	22	1
NRCAM	neuronal cell adhesion molecule	2/2	5.0	13	0	10	5.51E-38	23	1
CDH6	cadherin 6, type 2, K-cadherin (fetal kidney)	1/1	3.7	9	0	8	5.51E-38	24	1
PPARGC1A	peroxisome proliferator-activated receptor γ, coactivator 1 α	1/1	−8.6	7	7	0	5.51E-38	25	1
PLA2R1	phospholipase A2 receptor 1, 180 kDa	1/1	−8.9	7	7	0	5.51E-38	26	1
STK32A	serine/threonine kinase 32A	1/1	6.2	8	0	7	5.51E-38	27	1
ETV4	ets variant 4	1/1	5.1	10	0	7	5.51E-38	28	1
MTHFD1 liter	methylenetetrahydrofolate dehydrogenase (NADP + dependent) 1-like	1/2	2.5	12	0	7	5.51E-38	29	1
MET	met proto-oncogene (hepatocyte growth factor receptor)	1/1	6.3	19	0	7	5.51E-38	30	1
ANGPTL1	angiopoietin-like 1	1/2	−4.2	6	6	0	5.51E-38	31	1
RYR2	ryanodine receptor 2 (cardiac)	2/2	−6.6	7	6	0	5.51E-38	32	1
MPPED2	metallophosphoesterase domain containing 2	3/3	−5.7	5	5	0	5.51E-38	33	1
WSCD2	WSC domain containing 2	1/2	−8.9	5	5	0	5.51E-38	34	1
KLHDC8A	kelch domain containing 8A	1/1	8.6	5	0	5	5.51E-38	35	1
PLSCR4	phospholipid scramblase 4	1/1	−2.5	6	5	0	5.51E-38	36	1
IPCEF1	interaction protein for cytohesin exchange factors 1	2/2	−9.3	7	5	0	5.51E-38	37	1
PAPSS2	3′-phosphoadenosine 5′-phosphosulfate synthase 2	1/1	−8.9	7	5	0	5.51E-38	38	1
SIPA1L2	signal-induced proliferation-associated 1 like 2	1/1	3.2	7	0	5	5.51E-38	39	1
DPP6	dipeptidyl-peptidase 6	1/2	−5.3	9	5	0	5.51E-38	40	1
DPT	dermatopontin	1/1	−9.9	4	4	0	5.51E-38	41	1
LRP4	low density lipoprotein receptor-related protein 4	1/1	18.6	5	0	4	5.51E-38	42	1
QPCT	glutaminyl-peptide cyclotransferase	2/2	16.0	3	0	3	5.51E-38	43	1
C4ORF48	chromosome 4 open reading frame 48	1/1	4.2	3	0	3	5.51E-38	44	1
CCND1	cyclin D1	1/1	2.5	4	0	3	5.51E-38	45	1
TBX22	T-box 22	1/1	−18.6	4	3	0	5.51E-38	46	1
LRP1B	low density lipoprotein-related protein 1B (deleted in tumors)	1/1	−22.1	5	3	0	5.51E-38	47	1
GPC3	glypican 3	1/2	−8.7	12	3	0	5.51E-38	48	1
PRR15	proline rich 15	1/1	5.1	2	0	2	5.51E-38	49	1
NPC2	Niemann-Pick disease, type C2	1/1	3.8	3	0	2	5.51E-38	50	1
HMGA2	high mobility group AT-hook 2	1/3	64.9	18	0	13	2.51E-36	51	1
ABCC3	ATP-binding cassette, sub-family C (CFTR/MRP), member 3	1/2	8.8	10	0	9	2.51E-36	52	1
ADH1A	alcohol dehydrogenase 1A (class I), α polypeptide	2/2	−10.1	15	9	0	2.51E-36	53	1
LIFR	leukemia inhibitory factor receptor α	1/1	−7.9	10	8	0	2.51E-36	54	1
MATN2	matrilin 2	1/1	−6.1	8	7	0	2.51E-36	55	1
C8ORF79	chromosome 8 open reading frame 79	1/1	−4.0	8	7	0	2.51E-36	56	1
MRO	maestro	1/1	−8.9	9	7	0	2.51E-36	57	1
FAM84A	family with sequence similarity 84, member A	1/2	3.4	9	0	7	2.51E-36	58	1
ARHGAP24	ρ GTPase activating protein 24	1/1	−5.2	10	7	0	2.51E-36	59	1

Abbreviation: FC, fold change.

[a] Sign test without correction.

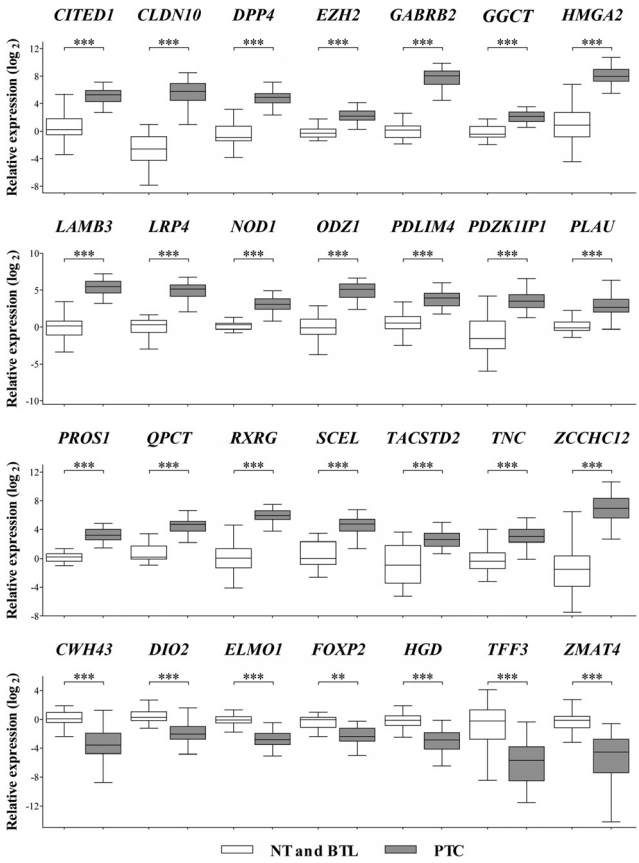


Figure 2. RT-qPCR analysis of the selected transcripts by Taqman Low Density Arrays (TLDA). Expression of 28 transcripts in all samples assessed by TLDA and statistical comparison between PTC and NT/BTL. **, $P < .01$; ***, $P < .001$ (unpaired t test).

Figure 4). In addition, a linear association with risk of lymph node involvement was similarly noticed in external data (benign range, 5%; low, 26%; intermediate, 51%; high, 66%; $P < .001$) (Figure 4B).

As *GABRB2* was included in the diagnostic classifier previously described by Alexander et al (32), composed by 167 genes, it was possible to test its performance using FNAB (available on GEO datasets). An accuracy of 92% (95% sensitivity and 89% specificity) and 74% (60% of sensitivity and 80% of specificity) was observed when evaluating conclusive and indeterminate results, respectively (Supplemental Figure 4).

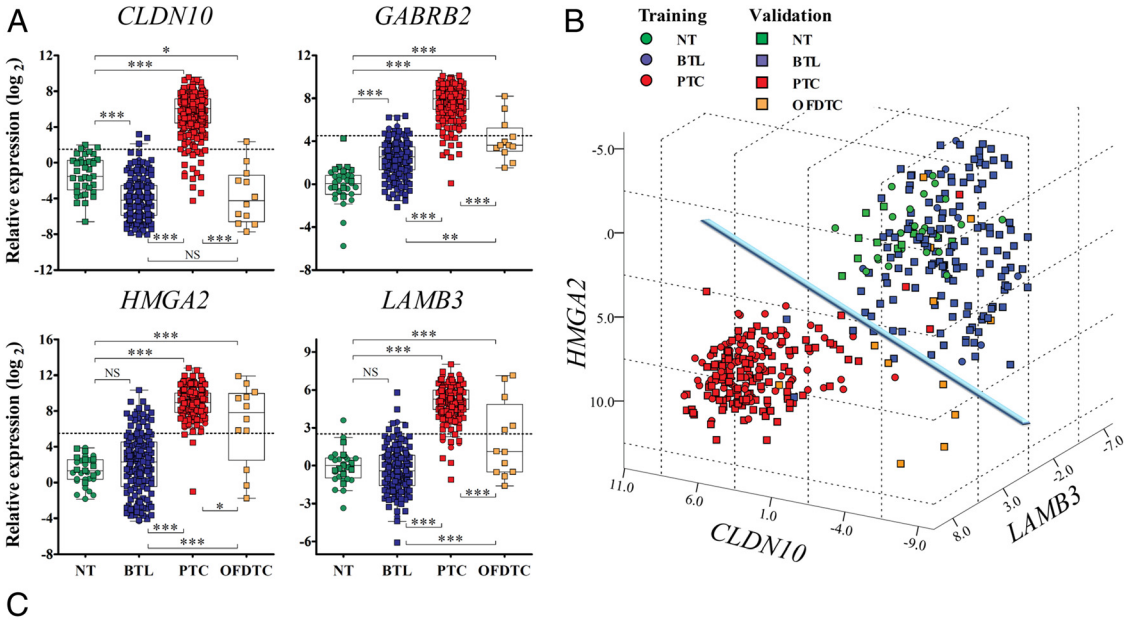
Discussion

In this study, trustworthy differences in gene expression profiles in PTC have been reported. In contrast with other studies, the consistency of our results was ensured by a novel and straightforward cross-study validation strategy. Transcriptome studies often lack reproducibility, which is usually due to a small sample size. Validation of the findings using independent external data is an excellent and reliable way to increase the statistical power of single-study analysis (17). A previous meta-analysis/meta-review study described 38 genes as consistently altered among several lists published for thyroid cancer (18). All of these genes, with the exception of four, were included in our final gene list (589 genes). In addition, three of these genes (*TIMP1*, *TGFA*, and *FABP4*) would have been included after adjustment of the microarray quality control filters (data not shown).

In silico molecule interaction analysis allowed a collected view, limited to differentially expressed genes in PTC validated by external records. Loss of anoikis, increased metastasis, and adhesion of connective tissue cells were predicted to be the altered biological functions by IPA. Anoikis is an extracellular matrix deprivation–induced apoptosis process that represents a critical barrier to the metastatic cascade (19). The role of cell-to-cell contacts in AKT activation and survival of thyroid cancer cells in low-adherent conditions has already been demonstrated (20). Similarly, inflammatory processes were overrepresented, as previously reported (4). Increased metabolism and synthesis of reactive oxygen species could be the result of the inflammatory microenvironment, which may in turn stimulate the MAPK pathway and enhance thyroid tumorigenesis (21).

RAR and RXR/TR signaling pathways were also observed. RA plays a key role in cell differentiation and fundamentally requires RAR and RXR nuclear receptor families for signal transduction (22). RA has been investigated for its potential to induce cell redifferentiation, which can restore type I iodothyronine-5′-deiodinase enzyme and sodium-iodide symporter (NIS) expression, promoting radioiodine sensitivity (23). In our PTC cases, *RBP7*, *ADH1A*, *CRABP1*, and *RARB* were found to have a lower expression, whereas the *STRA6*, *CYP1B1* (both absent in the IPA database), *RDH5*, *ALDH3B1*, and *RXRG* genes were observed to have higher expression levels. The down-regulation of the thyroid hormone transporter (MCT8) and deiodinase genes (*DIO1* and *DIO2*) could represent lower levels of both T_4 and T_3. The actual function of the thyroid hormone intracellular transport, deiodination, and bioactive action at nuclear receptors have been highlighted as mechanisms required for an adequate thyroid hormone action and metabolic function (24). Higher expression of *RXRG* could enhance the formation of heterodimers with the thyroid hormone receptors (TR), which block target gene transcription without T_3 (25). In mice, induction of follicular carcinomas by TRβ mutations suggests a protective effect of thyroid hormone over the gland itself (26).

The prediction of upstream regulators, used to clarify the observed gene expression results, underlined a putative activation of the *EZH2* and *HDAC* genes family. *EZH2* is described as specifically overexpressed in anaplastic thy-



C

TRAINING SET		*GABRB2*		*CLDN10/LAMB3/HMGA2*	
		Non-neoplastic	Malignant	Non-neoplastic	Malignant
Histology	Normal	23	0	23	0
	Benign	7	1	8	0
	Malignant	4	82	0	86

***GABRB2*:** Sens: 95% (89-99); Spec: 88% (47-98); PPV: 99% (93-100); NPV: 63% (31-89); AUC: 0.978 (0.951-1.000)
3-gene classifier: Sens: 100% (96-100); Spec: 100% (63-100); PPV: 100% (96-100); NPV: 100% (63-100); AUC:1.000 (1.000-1.000)

VALIDATION SET		*GABRB2*		*CLDN10/LAMB3/HMGA2*	
		Non-neoplastic	Malignant	Non-neoplastic	Malignant
Histology	Normal	10	0	10	0
	Benign	137	3	135	5
	Malignant	22	110	8	124

***GABRB2*:** Sens: 85% (78-90); Spec: 97% (92-99); PPV: 98% (93-99); NPV: 83% (76-89); AUC: 0.950 (0.924-0.976)
3-gene classifier: Sens: 94% (88-97); Spec: 96% (92-99); PPV: 96% (91-99); NPV: 94% (89-98); AUC: 0.970 (0.948-0.992)

Figure 3. Diagnostic models with RT-qPCR assessed transcripts. A, Expression of *CDLN10*, *HMGA2*, *LAMB3*, and *GABRB2* in all samples included in the study with adjusted threshold (dashed line) from Receiver Operating Characteristic curves. B, Three-dimensional graph plotting relative expression of *CLDN10*, *HMGA2*, and *LAMB3*. The classifier (cyan) was trained with Fisher linear discriminant analysis. C, Classification performance in discriminating tumors BTL over training and validation sets of *CLDN10*, *HMGA2*, and *LAMB3* and the single use of *GABRB2*. NS, not significant; *, $P < .05$; **, $P < .01$; ***, $P < .001$ (ANOVA with Tukey's post-hoc test); Sens, sensitivity; Spec, specificity; PPV, positive predictive value (PPV); NPV, negative predictive value; AUC, area under the curve.

roid carcinoma, and its inhibition promotes differentiation, anoikis, and inhibition of cell growth and invasion (27). HDAC inhibitors have also been demonstrated to induce differentiation, apoptosis and cell cycle arrest in many thyroid cancer cell lines (28). Even in low doses, HDAC inhibitors are able to induce NIS expression and restore radioiodine uptake in undifferentiated thyroid tumors cells (29). Protein–protein interaction analysis demonstrated a direct interaction of EZH2 and HDAC on RAR and RXR/TR activation pathways. *EZH2* knockdown has been previously shown to be able to restore RAR signaling in melanoma cells (30) that, similar to thyroid cancer, frequently harbor the *BRAF* mutation.

A set of 28 genes confirmed from cross-study validation, including *EZH2*, was additionally submitted to RT-qPCR. All transcripts were confirmed, with very high correlation with microarray and reproducible in an independent set of samples. Although several genes have been previously associated with PTC, overexpression of *NOD1* and *EZH2* and underexpression of *CWH43* and *FOXP2* have not been previously reported. *GABRB2* (γ-aminobutyric acid [GABA] receptor) presented the highest expression in PTC using the microarray, and was the most important transcript in meta-analysis, reaching the highest AUC when used individually by RT-qPCR. Despite this finding, only one study has reported it as overexpressed in thyroid carcinomas (31).

In this study, an effort was made to identify a limited number of transcripts to be used in the diagnosis of thyroid cancer. Given the morphological similarities of thyroid lesions, particularly follicular thyroid carcinoma (FTC) and follicular variant of PTC from follicular adenoma,

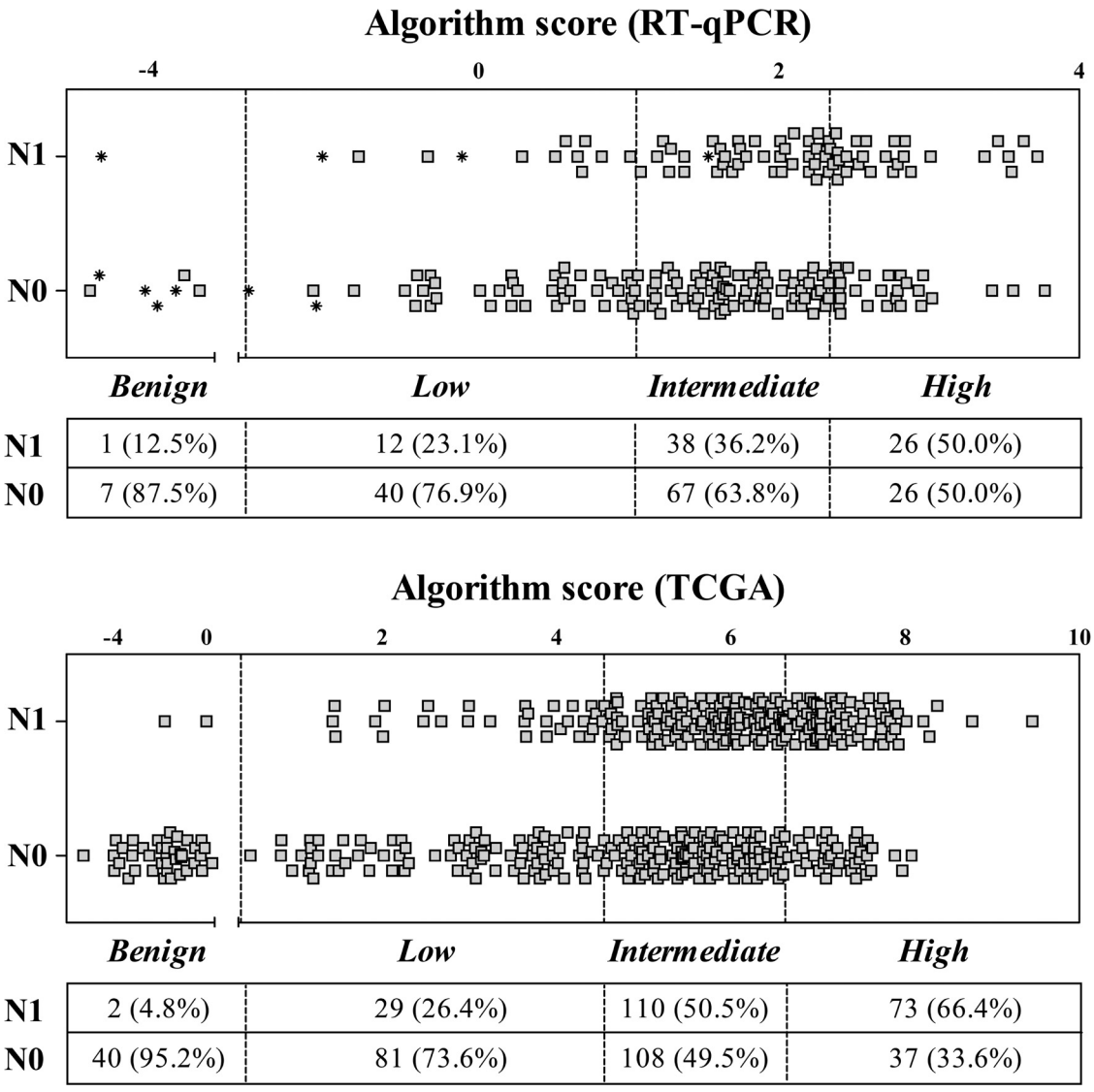


Figure 4. Lymph node disease risk according to algorithm score. The figure demonstrates a linear association of node involvement risk according to algorithm scores values in thyroid patients with cancer in both internal and external data (TCGA). *, Other follicular derived thyroid carcinomas.

diagnosis is frequently constrained to surgical specimens. A prospective study evaluated a diagnostic molecular test based on the measurement of 167 genes by microarray in 265 indeterminate nodules (32). The classifier presented high sensitivity (92%) but low specificity (52%), with additional reports corroborating these findings (33, 34). However, a recent independent study reported a confirmation of malignancy rate of 17% in indeterminate nodules classified as suspicious by the test (35). Furthermore, it was demonstrated that outside of North America, the test is not cost-effective (36). Albeit not yet validated for FNAB, our method is more viable via the inclusion of only a small number of transcripts evaluated by an accessible technique.

The algorithm, which combines *CLDN10*, *HMGA2* and *LAMB3*, and *GABRB2* as markers presented a high performance in discriminating thyroid carcinomas from benign samples, despite most FTC was considered indolent. FTC diagnosis is difficult and its reproducibility among experienced pathologists is low, especially for minimally invasive cases, where discrepancies can reach 57% (37). *GABRB2* demonstrated 60% of sensitivity and 80% of negative predictive value using publicly available database of indeterminate FNAB. Given that *CLDN10*, *HMGA2*, and *LAMB3* were not publicly available from a database of indeterminate FNAB, the algorithm could not be efficiently tested. Genes included in the algorithm had been previously associated with PTC (16, 38, 39). *CLDN10*, which encodes an integral membrane protein and component of tight junction strands, was clearly overexpressed in PTC when compared with NT. Nonetheless, it was found to be down-regulated in BTL and OFDTC, principally in FTC, as previously described (38). Not surprisingly, it was an important classifier member in the

discrimination of BTL and PTC, although the probability of identifying FTC was lower. However, overexpression of *HMGA2* and *LAMB3* and down-regulation of *CLDN10* can be indicative of FTC. *HMGA2*, revealed as overexpressed in PTC, alters the chromatin structure regulating the transcription of several genes. Despite high sensitivity and specificity, *HMGA2* as a unique marker by RT-qPCR was not sufficient to discriminate between benign and malignant lesions. However, its expression was still much higher in PTC, as previously described in the comparison between borderline/benign and malignant lesions (39). *LAMB3* belongs to a family of basement membrane proteins (laminins) and has been previously described as overexpressed in PTC (38). A strong association between *LAMB3* overexpression with lymph node metastases was found. This finding reflected on the higher risk of nodal metastases in high algorithm scores found in PTC patients using internal and external data. Remarkably, half of these patients (and 66% of the TCGA cohort) presented lymph node metastasis at diagnosis, in contrast with 23% of low scoring patients (26% of the TCGA cohort). In our cohort, the only case presenting lymph node metastases classified in the benign range was a Hürthle cell carcinoma, which is a distinct thyroid malignancy (40). Other cases classified as benign (false negative) showed neither aggressive clinical characteristics nor recurrence after a follow-up period ranging from 3.6 to 6.3 years. If proven to be useful when evaluated preoperatively, this test could be an important factor in decisions such as central neck compartment dissection.

Taken together, the results of the present study contribute to the understanding of the molecular pathobiology of PTC and highlight a possible role for the RA and T_4 metabolism dysregulation in the processes of tumorigenesis and tumor maintenance. EZH2 and HDAC inhibitors are potential therapeutic targets for use in advanced disease, although experimental studies are required to demonstrate this hypothesis. An algorithm combining five transcripts (*CLDN10*, *HMGA2*, and *LAMB3* as targets and *EIF2B1* and *PUM1* as references) presented high diagnostic accuracy and was associated with regional lymph node metastasis risk.

Acknowledgments

The authors thank the A.C. Camargo Cancer Center Biobank for providing and processing the samples.

Address all correspondence and requests for reprints to: Luiz Paulo Kowalski, MD, PhD, and Silvia Regina Rogatto, PhD, International Research Center–A.C. Camargo Cancer Center, Rua Taguá 440, Liberdade, 01508-010, São Paulo, Brazil. E-mail: lp_kowalski@uol.com.br and rogatto@fmb.unesp.br.

This work was supported by grants from the National Institute of Science and Technology in Oncogenomics (Fundação de Amparo à Pesquisa do Estado de São Paulo [FAPESP] Grant 2008/57887-9 and Conselho Nacional de Desenvolvimento Científico e Tecnológico Grant CNPq 573589/08-9) and FAPESP (2010/09526-7 and 2010/18370-0).

Disclosure Summary: The authors have nothing to disclose.

References

1. **Xing M.** Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer*. 2013;13:184–199.
2. **Kondo T, Ezzat S, Asa SL.** Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat Rev Cancer*. 2006;6:292–306.
3. **Kimura ET, Nikiforova MN, Zhu Z, Knauf JA, Nikiforov YE, Fagin JA.** High prevalence of BRAF mutations in thyroid cancer: Genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Res*. 2003; 63:1454–1457.
4. **Delys L, Detours V, Franc B, et al.** Gene expression and the biological phenotype of papillary thyroid carcinomas. *Oncogene*. 2007; 26:7894–7903.
5. **Smallridge RC, Chindris AM, Asmann YW, et al.** RNA sequencing identifies multiple fusion transcripts, differentially expressed genes, and reduced expression of immune function genes in BRAF (V600E) mutant vs BRAF wild-type papillary thyroid carcinoma. *J Clin Endocrinol Metab*. 2014;99:338–347.
6. **Siegel R, Ma J, Zou Z, Jemal A.** Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64:9–29.
7. **Davies L, Welch HG.** Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg*. 2014;140:317–322.
8. **Lang W, Borrusch H, Bauer L.** Occult carcinomas of the thyroid. Evaluation of 1,020 sequential autopsies. *Am J Clin Pathol*. 1988; 90:72–76.
9. **Ito Y, Uruno T, Nakano K, et al.** An observation trial without surgical treatment in patients with papillary microcarcinoma of the thyroid. *Thyroid*. 2003;13:381–387.
10. **Yang J, Schnadig V, Logrono R, Wasserman PG.** Fine-needle aspiration of thyroid nodules: A study of 4703 patients with histologic and clinical correlations. *Cancer*. 2007;111:306–315.
11. **Gonçalves Filho J, Kowalski LP.** Surgical complications after thyroid surgery performed in a cancer hospital. *Otolaryngol Head Neck Surg*. 2005;132:490–494.
12. **Hundahl SA, Cady B, Cunningham MP, et al.** Initial results from a prospective cohort study of 5583 cases of thyroid carcinoma treated in the United States during 1996. US and German Thyroid Cancer Study Group. An American College of Surgeons Commission on Cancer Patient Care Evaluation study. *Cancer*. 2000;89:202–217.
13. **Rossing M.** Classification of follicular cell-derived thyroid cancer by global RNA profiling. *J Mol Endocrinol*. 2013;50:39–51.
14. **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;96:3390–3397.
15. **Rodríguez-Rodero S, Fernández AF, Fernández-Morera JL, et al.** DNA methylation signatures identify biologically distinct thyroid cancer subtypes. *J Clin Endocrinol Metab*. 2013;98:2811–2821.
16. **Prasad NB, Kowalski J, Tsai HL, et al.** Three-gene molecular diagnostic model for thyroid cancer. *Thyroid*. 2012;22:275–284.
17. **Ramasamy A, Mondry A, Holmes CC, Altman DG.** Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med*. 2008;5:e184.
18. **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;24:5043–5051.
19. **Frisch SM, Screaton RA.** Anoikis mechanisms. *Curr Opin Cell Biol*. 2001;13:555–562.
20. **Jensen K, Patel A, Klubo-Gwiezdzinska J, Bauer A, Vasko V.** Inhibition of gap junction transfer sensitizes thyroid cancer cells to anoikis. *Endocr Relat Cancer*. 2011;18:613–626.
21. **Xing M.** Oxidative stress: A new risk factor for thyroid cancer. *Endocr Relat Cancer*. 2012;19:7–11.
22. **Niederreither K, Dollé P.** Retinoic acid in development: Towards an integrated view. *Nat Rev Genet*. 2008;9:541–553.
23. **Kim WG, Kim EY, Kim TY, et al.** Redifferentiation therapy with 13-cis retinoic acids in radioiodine-resistant thyroid cancer. *Endocr J*. 2009;56:105–112.
24. **Visser WE, van Mullem AA, Visser TJ, Peeters RP.** Different causes of reduced sensitivity to thyroid hormone: Diagnosis and clinical management. *Clin Endocrinol (Oxf)*. 2013;79:595–605.
25. **Li J, Wang J, Wang J, et al.** Both corepressor proteins SMRT and N-CoR exist in large protein complexes containing HDAC3. *EMBO J*. 2000;19:4342–4350.
26. **Köhrle J.** Guard your master: Thyroid hormone receptors protect their gland of origin from thyroid cancer. *Endocrinology*. 2004;145:4427–4429.
27. **Borbone E, Troncone G, Ferraro A, et al.** Enhancer of zeste homolog 2 overexpression has a role in the development of anaplastic thyroid carcinomas. *J Clin Endocrinol Metab*. 2011;96:1029–1038.
28. **Mitsiades CS, Poulaki V, McMullan C, et al.** Novel histone deacetylase inhibitors in the treatment of thyroid cancer. *Clin Cancer Res*. 2005;11:3958–3965.
29. **Kitazono M, Robey R, Zhan Z, et al.** Low concentrations of the histone deacetylase inhibitor, depsipeptide (FR901228), increase expression of the Na(+)/I(-) symporter and iodine accumulation in poorly differentiated thyroid carcinoma cells. *J Clin Endocrinol Metab*. 2001;86:3430–3435.
30. **Epping MT, Wang L, Edel MJ, Carlée L, Hernandez M, Bernards R.** The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. *Cell*. 2005;122:835–847.
31. **Wiseman SM, Haddad Z, Walker B, et al.** Whole-transcriptome profiling of thyroid nodules identifies expression-based signatures for accurate thyroid cancer diagnosis. *J Clin Endocrinol Metab*. 2013;98:4072–4079.
32. **Alexander EK, Kennedy GC, Baloch ZW, et al.** Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med*. 2012;367:705–715.
33. **Duick DS, Klopper JP, Diggans JC, et al.** The impact of benign gene expression classifier test results on the endocrinologist-patient decision to operate on patients with thyroid nodules with indeterminate fine-needle aspiration cytopathology. *Thyroid*. 2012;22:996–1001.
34. **Alexander EK, Schorr M, Klopper J, et al.** Multicenter clinical experience with the Afirma gene expression classifier. *J Clin Endocrinol Metab*. 2014;99:119–125.
35. **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;99:4069–4077.
36. **Lee L, How J, Tabah RJ, Mitmaker EJ.** Cost-effectiveness of molecular testing for thyroid nodules with atypia of undetermined significance cytology. *J Clin Endocrinol Metab*. 2014;99:2674–2682.
37. **Franc B, de la Salmonière P, Lange F, et al.** Interobserver and intraobserver reproducibility in the histopathology of follicular thyroid carcinoma. *Hum Pathol*. 2003;34:1092–1100.
38. **Aldred MA, Huang Y, Liyanarachchi S, et al.** Papillary and follicular thyroid carcinomas show distinctly different microarray expression profiles and can be distinguished by a minimum of five genes. *J Clin Oncol*. 2004;22:3531–3539.
39. **Arora N, Scognamiglio T, Lubitz CC, et al.** Identification of borderline thyroid tumors by gene expression array analysis. *Cancer*. 2009;115:5421–5431.
40. **Ganly I, Ricarte Filho J, Eng S, et al.** Genomic dissection of Hurthle cell carcinoma reveals a unique class of thyroid malignancy. *J Clin Endocrinol Metab*. 2013;98:962–972.