A PCR-based expression signature of malignancy in follicular thyroid tumors

Theodoros Foukakis1, Arief Gusnanto2,4, Amy YM Au5, Anders Höög3, Weng-Onn Lui1, Catharina Larsson1, Göran Wallin1 and Jan Zedenius1

Departments of 1Molecular Medicine and Surgery, 2Medical Epidemiology and Biostatistics and 3Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden 4Medical Research Council – Biostatistics Unit, Institute of Public Health Robinson Way, Cambridge, United Kingdom 5Cancer Genetics Unit, Kolling Institute of Medical Research, University of Sydney, New South Wales, Australia

(Requests for offprints should be addressed to T Foukakis; Email: theodoros.foukakis@ki.se)

Abstract

The diagnosis of follicular thyroid carcinoma (FTC) in the absence of metastasis can only be established postoperatively. Moreover, high-risk FTCs are often not identifiable at the time of diagnosis. In this study, we aimed to identify transcriptional markers of malignancy and high-risk disease in follicular thyroid tumors. The expression levels of 26 potential markers of malignancy were determined in a panel of 75 follicular thyroid tumors by a TaqMan quantitative RT-PCR approach. Logistic regression analysis (LRA) was used for gene selection and generation of diagnostic and prognostic algorithms. An algorithm based on the expression levels of five genes (TERT, TFF3, PPARγ, CITED1, and EGR2) could effectively predict high-risk disease with a specificity of 98.5%. The metastatic potential could be predicted in all four cases with apparently benign or minimally invasive (MI) disease at the time of diagnosis, but poor long-term outcome. In addition, a second model was produced by implementing two genes (TERT and TFF3), which was able to distinguish adenomas from de facto carcinomas. When this model was tested in an independent series of atypical adenomas (AFTA) and MI-FTCs, 16 out of 17 AFTAs were classified as ‘benign’, while MI-FTCs with vascular invasion (sometimes referred to as ‘moderately invasive’) and/or large tumor size tended to classify in the ‘malignant’ group. The reported models can be the foundation for the development of reliable preoperative diagnostic and prognostic tests that can guide the therapeutic approach of follicular thyroid neoplasms with indeterminate cytology.

Endocrine-Related Cancer (2007) 14 381–391

Introduction

The distinction between follicular thyroid cancer (FTC) and its benign counterpart follicular thyroid adenoma (FTA) has for years been a challenge to surgeons, cytologists, and histopathologists. The main reason is that the fine needle aspiration (FNA) biopsy, which is the initial diagnostic approach for thyroid nodules, fails to discriminate between FTC and FTA. In the absence of metastasis, the definite diagnosis of FTC relies on the demonstration of capsular and/or vascular invasion, which is demonstrable only in the postoperative surgical specimen.

Moreover, postoperative difficulties in establishing the correct diagnosis are not uncommon. For minimally invasive FTCs (MI-FTCs), identification of the rare invasion sites requires thorough and meticulous examination of the surgical specimen by an experienced histopathologist. Similarly, in atypical FTAs (AFTA), cell proliferation is pronounced and irregular architecture is evident, and the ‘benign’ diagnosis is merely based on the absence of obvious invasion. For tumors in which malignancy cannot be determined at histopathology, the term ‘follicular tumors of uncertain malignancy’ was recently proposed (Williams 2000). Furthermore, the aggressiveness spectrum of FTCs is wide. While patients with metastatic and/or widely invasive FTC (WI-FTC) have a substantial long-term morbidity and mortality, patients with MI-FTC rarely have an aggressive course of the disease (Shaha et al. 1997). On the other hand, in some cases where the histopathological diagnosis is benign, metastases are present or develop during follow-up, motivating a
refinement of the diagnosis to FTC based on clinical evidence.

The shortfalls in the preoperative diagnosis of follicular thyroid tumors are reflected in the treatment of patients. The typical therapeutic approach of a patient with a thyroid nodule and suspicious follicular cytology is a lobectomy. In more than 80% of the cases, the histopathological diagnosis will be that of adenoma, after which no further treatment is required. When the surgical specimen is diagnosed as FTC, the patient is re-operated with removal of the contralateral thyroid lobe. Alternatives to a completion thyroidectomy are $^{131}$I ablation of the remaining lobe (rarely used) and/or levothyroxine suppressive therapy (Singer et al. 1996). Nevertheless, both alternatives are restricted to low-grade MI-FTCs and are not indicated for high-risk cases, wherein a second operation is unavoidable. Hence, as there are no reliable markers to predict the natural course of AFTA and MI-FTC, clinical decisions can be hard even after lobectomy.

The distinction of FTC from FTA at the molecular level has been an attractive approach to improve the diagnostics of follicular thyroid tumors on FNA. Several immunohistochemical and molecular biomarkers have been proposed for the distinction of FTA from FTC, including thyroid peroxidase (TPO), galectin-3, telomerase, PAX8-peroxidase proliferator-activated receptor γ (PAX8-PPARγ), and HMBE-1 (de Micco et al. 1994, Sack et al. 1997, Saji et al. 1999, Bartolazzi et al. 2001). However, none of these findings have yet led to the development of markers that are sensitive and specific enough to be used in routine laboratories. Recently, several groups have reported that global gene expression patterns determined by microarray analysis or serial analysis of gene expression (SAGE) are distinct in FTA and FTC (Barden et al. 2003, Cerutti et al. 2004, Finley et al. 2004, Puskas et al. 2005, Weber et al. 2005, Fryknäs et al. 2006). Although the specificity in those cases was high, the high cost, low reproducibility, and large amounts of starting RNA required presently exclude their clinical application in preoperative FNA diagnostics. On the other hand, PCR-based methods offer a simpler and more practical approach to quantify gene expression also in limited amounts of material. Gene expression patterns obtained by quantitative RT-PCR (qRT-PCR) have been linked to diagnosis or prognosis in various tumor types (Brabender et al. 2004, Lossos et al. 2004, Paik et al. 2004) including thyroid (Rosen et al. 2005, Weber et al. 2005). We here report a TaqMan qRT-PCR-based signature of malignancy in follicular thyroid tumors.

### Materials and methods

#### Patient data and tissue samples

This study includes 75 primary thyroid tumor samples obtained from 75 patients who underwent thyroid surgery at Karolinska University Hospital, Stockholm, Sweden, during the period 1986–2004. In 47 of the 68 cases where a preoperative FNA was performed and available to us, the cytological diagnosis was that of a follicular tumor suspicious for malignancy (69%). Histopathologically verified normal thyroid tissue was obtained from the contralateral lobe of eight patients operated for FTC and used as a reference in the qRT-PCR analyses (normal thyroid pool). All tissue samples were collected in direct connection to surgery, macroscopically dissected by a histopathologist, snap frozen in liquid nitrogen, and stored at $-70^\circ$C until required. For each frozen sample used in this study, a section was cut and subjected to histopathological evaluation to confirm the high tumor cell representativeness (>70%) in the tumor samples, and the lack of neoplastic cells in the normal thyroid tissues. Samples with excessive lymphocytic infiltration were excluded. All samples were obtained with informed consent from the patients and the study of the tissue material was approved by the local ethics committee.

Histopathological evaluation of the primary surgical specimens was done according to the criteria of World Health Organization (Hedinger et al. 1998), whereby 38 tumors were diagnosed as FTC, 18 as FTA, and 19 as AFTA. The clinical diagnosis was subsequently changed from AFTA to FTC in two cases, since the patients developed local recurrence and distant metastasis during follow-up. Vascular or capsular invasion was not detected in any of the FTAs or AFTAs. However, the AFTAs frequently exhibited irregular architecture and cellular pattern, as well as signs of considerable proliferation, pleomorphism, and frequent mitoses. Among the FTCs, 5 were classified as WI-FTC and 35 as MI-FTC. In four cases of MI-FTC (including the two reclassified AFTAs), the diagnosis of cancer was based on the presence of distant metastasis rather than on histopathological proof of invasion. In total, four patients presented with metastatic disease at diagnosis, and in three cases the tumors were not totally resected.

All patients were operated at the Department of Surgery and the malignant cases were referred to the Department of Oncology for follow-up and possible adjuvant treatment with radiiodine ($^{131}$I). In 29 of the 40 FTCs (72.5%), a total thyroidectomy was performed either per se or as a completion thyroidectomy. Sixteen FTC patients (40%) received $^{131}$I
postoperatively. Four patients with benign tumors (11%) were operated with total thyroidectomy due to cytological suspicion of papillary cancer or concomitant goiter or Graves’ disease.

The patients were followed at the Department of Surgery or the Department of Oncology at Karolinska University Hospital. The FTA and some of the AFTA patients were followed in a primary health care setting, but were admitted back to the Karolinska University Hospital in every case of thyroid disease-related event. For benign cases without any events during follow up, patient files were backtracked to confirm the absence of disease recurrence. All FTC patients were followed 3, 6, and 12 months postoperatively, thereafter at a 6-month basis for 5 years and then yearly, with a clinical examination and measurement of serum thyroglobulin (Tg). For patients receiving levothyroxine suppressive therapy after lobectomy, non-stimulated serum Tg was determined, while for patients operated with a total thyroidectomy, Tg was initially measured following thyroid-stimulating hormone stimulation. If this was negative (not detectable), non-stimulated Tg test was subsequently used for these patients as well. Neck ultrasonography and diagnostic 131I whole body scan were only used upon clinical indication and/or abnormal Tg values (> 1 ng/ml). The average follow-up time (till December 2004, till the patient’s death, or till the last date patient resided in Stockholm County) was 8 years (range 0–18 years) for FTCs, 10.5 years (range 4–17 years) for FTAs, and 9.5 years (range 4–15 years) for AFTAs. Local recurrence or metastasis was not detected in any FTA or AFTA patients as the two AFTA patients with metastasis at follow-up were reclassified to FTC.

Local recurrence or metastasis was not detected in any FTC patients. The average follow-up time (till December 2004, till the patient’s death, or till the last date patient resided in Stockholm County) was 8 years (range 0–18 years) for FTCs, 10.5 years (range 4–17 years) for FTAs, and 9.5 years (range 4–15 years) for AFTAs. Twenty-six genes were selected for analysis in the extended series of follicular thyroid tumorigenesis, previously reported as markers in FTC; or recurrently found differentially expressed in expression profiling of thyroid tumors (microarray or SAGE). In addition, β-actin and 18S were evaluated as endogenous controls. 18S showed a higher consistency across the samples and was chosen as the endogenous control for the experiments. After the initial qRT-PCR screening, ten of the genes were selected for analysis in the extended series of follicular thyroid tumors (genes marked in bold in Table 1). The selection was based on the ability of each gene to discriminate the benign from the malignant tumors, the low redundancy and the efficiency of the PCR (data not shown).

**RNA extraction and cDNA synthesis**

Total RNA was isolated from all samples using the TRIzol Reagent (GIBCO BRL). After column purification with the RNaseasy kit (Qiagen) and DNase treatment (Qiagen), cDNA was synthesized from 1 μg of each RNA sample using MultiScribe reverse transcriptase primed with random hexamers (High capacity cDNA archive kit; Applied Biosystems), according to the protocol recommended by the manufacturer.

**Quantitative real-time RT-PCR (qRT-PCR)**

The gene expression levels of the 26 test genes and 2 control genes were quantified using TaqMan technology on an ABI PRISM 7700 sequence detection system (Applied Biosystems). Gene-specific primers and probes were available as TaqMan Gene Expression Assays (Applied Biosystems; Table 1). The qRT-PCR was performed with 25 μl reaction mixture containing 1 × TaqMan Universal Master Mix, 1 × Target Assay Mix (Applied Biosystems), and 10 μl of first-strand cDNA from each sample as a template, using standard thermocycling conditions, as described previously (Foukakis et al. 2006). For the qRT-PCR of the highly abundant 18S, only 2 μl template were used in order to avoid early and inconsistent amplification. A standard curve was used for relative quantification and the expression values of each gene were normalized to the respective 18S values for the same sample. For each gene, a value of 1 was assigned to the normal thyroid pool. In samples where no expression was detected for a particular gene, an arbitrary value was given (i.e. the lowest detected value among the samples divided by 8 ($2^{3}$)).

**Initial gene selection**

Twenty-six genes were selected (Table 1) and their mRNA expression was measured in a randomly chosen subset of tumors (21 FTC, 11 FTA, and 10 AFTA) and a pool of normal thyroid tissues. The selection of the genes was based on the following criteria: involvement in thyroid development and/or function; implication in follicular thyroid tumorigenesis, previously reported as markers in FTC; or recurrently found differentially expressed in expression profiling of thyroid tumors (microarray or SAGE). In addition, β-actin and 18S were evaluated as endogenous controls. 18S showed a higher consistency across the samples and was chosen as the endogenous control for the experiments. After the initial qRT-PCR screening, ten of the genes were selected for analysis in the extended series of follicular thyroid tumors (genes marked in bold in Table 1). The selection was based on the ability of each gene to discriminate the benign from the malignant tumors, the low redundancy and the efficiency of the PCR (data not shown).

**Statistical analysis**

The objective of the statistical modeling was the classification and prediction of the binary response variable, i.e. high or low risk in the first model and the presence or not of malignancy in the second model. For this purpose, LRA was implemented (Hosmer & Lemeshow 2000). Using LRA, a classification rule equation is generated using the expression values of selected genes as variables. The equation has the
general form

\[ X = \chi_0 + \chi_aA + \chi_bB + \ldots + \chi_nN, \]

where \( X \) is the logistic regression parameter, \( A, B, \ldots, N \) are the expression values of the included genes (variables), \( \chi_0 \) is the intercept (constant), and \( \chi_a, \chi_b, \ldots, \chi_n \) are the regression coefficients (constants). In accordance with the theoretical basis of logistic regression, the classification rule can be written as

\[ \mu = \frac{1}{1 + e^{-X}}, \]

where \( e \) is the natural number and \( X \) the logistic regression parameter calculated as shown above. The value of \( \mu \) (here referred to as score) is bounded between 0 and 1 and the unknown samples are assigned to one class if \( \mu > 0.5 \) (i.e. when \( X > 0 \)), and to the second if \( \mu < 0.5 \) (i.e. when \( X < 0 \)).

To generate the classification models, the gene expressions were log-transformed and a two-sided \( t \)-test was performed for each gene. Then, a variable selection was performed by a forward stepwise approach, where the variables (genes) were entered one at a time into the model, starting with the gene with the highest \( t \)-statistic (absolute value). The process continued until the entry of additional genes in the model did not reduce the number of misclassifications, as assessed by a leave-one-out cross-validation. This is an ‘out-of-sample’ prediction method, where one observation is taken out from the dataset, and a model is built based on the remaining observations. Based on the model parameter estimates, a prediction of the taken out observation is made from its gene expressions. The use of this method is of utmost importance, since the application of the model is for the prediction of new cases; thus its predictive power is more objectively measured by samples not used in generating the model. The statistical analysis was performed in R (R Foundation for Statistical Computing 2004), using the routine glm (R-Development-Core-Team 2004).

---

Table 1: The genes surveyed in the study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Symbol(^a)</th>
<th>Cytogenetic location</th>
<th>Assay id(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes suggested by expression profiling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cbp/p300-interacting transactivator 1</td>
<td><strong>CITED1</strong></td>
<td>Xq13.1</td>
<td>Hs00366310_m1</td>
</tr>
<tr>
<td>NESH-binding protein (TARSH)</td>
<td><strong>ABI3BP</strong></td>
<td>3q12</td>
<td>Hs00248821_m1</td>
</tr>
<tr>
<td>Early growth response 2</td>
<td><strong>EGR2</strong></td>
<td>10q21.1</td>
<td>Hs00166165_m1</td>
</tr>
<tr>
<td>GPI-anchored metastasis-associated protein homolog</td>
<td><strong>C4.4A</strong></td>
<td>19q13.32</td>
<td>Hs00205233_m1</td>
</tr>
<tr>
<td>Gelsolin (amyloidosis, Finnish type)</td>
<td><strong>GSN</strong></td>
<td>9q33</td>
<td>Hs00609276_m1</td>
</tr>
<tr>
<td>Caveolin 1</td>
<td><strong>CAV1</strong></td>
<td>7q31.1</td>
<td>Hs00184697_m1</td>
</tr>
<tr>
<td>Arginase, type II</td>
<td><strong>ARG2</strong></td>
<td>14q24</td>
<td>Hs00265750_m1</td>
</tr>
<tr>
<td>DNA damage-inducible transcript 3</td>
<td><strong>DDIT3</strong></td>
<td>12q13</td>
<td>Hs00358796_g1</td>
</tr>
<tr>
<td>v-jun sarcoma virus 17 oncogene homolog (avian)</td>
<td><strong>JUN</strong></td>
<td>1p32-p31</td>
<td>Hs00277190_s1</td>
</tr>
<tr>
<td>Integral membrane protein 1</td>
<td><strong>ITM1</strong></td>
<td>11q23.3</td>
<td>Hs00537619_m1</td>
</tr>
<tr>
<td>Genes previously implicated as FTC markers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telomerase reverse transcriptase</td>
<td><strong>TERT</strong></td>
<td>5p15.33</td>
<td>Hs00162669_m1</td>
</tr>
<tr>
<td>Trehalose factor 3 (intestinal)</td>
<td><strong>TFF3</strong></td>
<td>21q22.3</td>
<td>Hs00173625_m1</td>
</tr>
<tr>
<td>Lectin, galactoside binding, soluble, 3 (galectin 3)</td>
<td><strong>LGALS3</strong></td>
<td>14q21-22</td>
<td>Hs00173587_m1</td>
</tr>
<tr>
<td>Genes related to thyroid development/function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solute carrier family 5 (sodium iodide symporter)</td>
<td><strong>NIS</strong></td>
<td>19p13-12</td>
<td>Hs00166567_m1</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td><strong>TG</strong></td>
<td>8q24</td>
<td>Hs00794359_m1</td>
</tr>
<tr>
<td>Thyroid peroxidase</td>
<td><strong>TPO</strong></td>
<td>2p25</td>
<td>Hs00174927_m1</td>
</tr>
<tr>
<td>Deiodinase, iodothyronine, type I</td>
<td><strong>DIO1</strong></td>
<td>1p32-p33</td>
<td>Hs00174944_m1</td>
</tr>
<tr>
<td>Thyroid stimulating hormone receptor</td>
<td><strong>TSHR</strong></td>
<td>14q31</td>
<td>Hs00174910_m1</td>
</tr>
<tr>
<td>Thyroid transcription factor 1</td>
<td><strong>TTF1</strong></td>
<td>14q13</td>
<td>Hs00183037_m1</td>
</tr>
<tr>
<td>Paired box gene 8</td>
<td><strong>PAX8</strong></td>
<td>2q13</td>
<td>Hs00247586_m1</td>
</tr>
<tr>
<td>Deiodinase, iodothyronine, type II</td>
<td><strong>DIO2</strong></td>
<td>14q24</td>
<td>Hs00255341_m1</td>
</tr>
<tr>
<td>Genes implicated in thyroid carcinogenesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peroxisome proliferator-activated receptor, ( \gamma )</td>
<td><strong>PPARG</strong></td>
<td>3p25</td>
<td>Hs00234592_m1</td>
</tr>
<tr>
<td>Ras association (RatGDS/AF-6) domain family 1</td>
<td><strong>RASSF1</strong></td>
<td>3p21.3</td>
<td>Hs00945257_m1</td>
</tr>
<tr>
<td>Phosphatase and tensin homolog</td>
<td><strong>PTEN</strong></td>
<td>10q23.3</td>
<td>Hs00829813_s1</td>
</tr>
<tr>
<td>Ras association (RatGDS/AF-6) domain family 5</td>
<td><strong>RASSF5</strong></td>
<td>1q31.3</td>
<td>Hs00417514_m1</td>
</tr>
<tr>
<td>Cyclin D1</td>
<td><strong>CCND1</strong></td>
<td>11q13</td>
<td>Hs00277039_m1</td>
</tr>
<tr>
<td>Reference genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic 18S rRNA</td>
<td><strong>18S</strong></td>
<td></td>
<td>Hs99999901_s1</td>
</tr>
<tr>
<td>( \beta )-actin</td>
<td><strong>ACTB</strong></td>
<td>7p15-12</td>
<td>Hs99999903_m1</td>
</tr>
</tbody>
</table>

\(^a\)In bold are the genes selected after the initial qRT-PCR screening.

\(^b\)TaqMan Gene Expression Assays (Applied Biosystems).
**Results**

The patients were prospectively divided into four groups (groups I–IV; Table 2). The referral of cases to different groups was based on the final diagnosis determined both from the histopathological classification of the primary tumors and the clinical follow-up. Groups I \( (n=18) \) and II \( (n=17) \) consisted of patients with FTA and AFTA respectively who had a benign histopathology at diagnosis and a long follow-up without recurrent disease. Group III \( (n=27) \) comprised all MI-FTC patients who did not have metastasis at diagnosis and remained disease free during follow-up (i.e. low-risk FTCs). Group IV \( (n=13) \) was composed of all patients with WI-FTC and all MI-FTCs with metastatic disease, at diagnosis or during follow-up (i.e. high-risk FTCs).

After an initial screening of 26 potential malignancy markers of thyroid tumors, we selected ten genes and investigated their mRNA expression in the 75 follicular thyroid tumors by qRT-PCR. The expression levels of all the ten genes showed variation among the samples and also to a varying extent among the different groups (I–IV). The results for the five genes used later in the algorithms are illustrated as boxplots in Fig. 1, and for the five remaining genes in Fig. 2.

**A classification model of five genes is specific for high-risk FTCs (model 1)**

Since the majority of follicular thyroid neoplasms have an indolent course, we first investigated whether the high-risk tumors can be genetically predicted. Therefore, we compared the expression of the ten genes between group IV (high-risk) and groups I–III (low-risk), designated as model 1. An initial comparison by unpaired \( t \)-test unmasked the genes that could best distinguish between the two and a classification model was created by LRA using the expression values of five genes (Fig. 3A).

Application of this classification model to the 75 samples led to a ‘correct’ assignment in 72 out of 75 cases \( (P=6.7 \times 10^{-8}, \text{ with } \chi^2 \text{ on deviance}=43.2, \text{ df}=5) \). More specifically, all but one of the low-risk tumors (groups I–III) were classified correctly (61/62; 98.5%), with only one FTA being misclassified as high risk, while 11 out of 13 (84.5%) of the tumors of group IV were recognized as high risk. After the cross-validation, the classification was changed in two MI-FTCs (both angioinvasive) of group III, which were now predicted as high-risk tumors (Fig. 3B). All six patients of group IV who were initially treated...
with lobectomy were correctly identified as having a high-risk disease. Notably, in four of these cases, the patients were operated with completion thyroidectomy only after the development of clinical metastasis. The \( m \) scores for the two metastatic FTCs that were initially diagnosed as adenomas were 0.99 and 0.78, clearly above the 0.5 threshold (Fig. 3B).

### Two genes effectively distinguish benign from malignant tumors (model 2)

Model 1 was shown to be highly specific for high-risk FTCs (group IV). However, a considerable proportion of FTC patients in group III would also benefit from a total thyroidectomy. To determine whether the mRNA expression of a different set of genes could point out these cases, we excluded groups II and III from the training set and created a prediction model based on the differences between groups I and IV (model 2). The tumors of groups II and III were then merely used as validation samples. In this respect, the model addresses the more general question of the presence or not of malignancy, rather than risk assessment. A logistic regression approach was undertaken as before and a classification model comprising two genes was created (Fig. 4A). Application of this classification rule to groups I and IV led to the same classifications obtained by the first model, i.e. the same single FTA was misclassified as ‘malignant’ and the same two FTCs were misclassified as ‘benign’ (\( P = 1.04 \times 10^{-6} \), \( \chi^2 \) on deviance = 28.93646, df = 2).

In model 2, the cross-validation for groups I and IV was similarly performed with the leave-one-out method, and furthermore tumors of groups II and III were classified according to the generated algorithm as external test samples (Fig. 4B). The classifications in

---

**Figure 1** Boxplots illustrating the mRNA expression in tumor groups I–IV for the five genes used in the prediction models including (A) TERT, (B) TFF3, (C) PPAR\( \gamma \), (D) CITED1, and (E) EGR2. The expression values for TERT are presented in the logarithmic scale. A value of 1 in the linear scale represents the expression level of the normal thyroid pool.
each group were as follows: in group I (FTA), 17 out of 18 of the tumors had a ‘benign’ signature. Similarly, in group II (AFTA), 16 out of 17 of the tumors were classified as ‘benign’ (the only misclassified AFTA had a positive cytology and was operated with total thyroidectomy). In group IV, the ‘malignancy’ was correctly recognized in 11 out of the 13 cases (the same cases were misclassified as in model 1). Here, the two FTCs that were initially diagnosed and treated as benign tumors were correctly predicted to be malignant, both with a $\mu$ score of 1. A more complex pattern was seen in group III, where 9 out of 27 tumors were classified as ‘malignant’ and 18 out of 27 as ‘benign’. Notably, among the nine MI-FTC tumors assigned to the ‘malignant’ group, four exhibited both vascular and capsular invasion (out of five in total; 80%), one only vascular invasion (out of three), two were larger than 5 cm in diameter (out of three; 67%), and only two had only capsular invasion and were smaller than 5 cm (out of 16; 12%).

**Figure 2** Boxplot diagrams of the mRNA expression of the five genes that did not enter the regression models (A) TPO, (B) NIS, (C) DIO1, (D) TG, and (E) ABI3BP. A value of 1 represents the expression level of the normal thyroid pool.

### Discussion

FNA biopsy has been the initial diagnostic approach for thyroid nodules for the last three decades (Gharib & Goellner 1993, Hegedus 2004). FNA is simple, inexpensive, devoid of serious complications and diagnostically useful in about 80% of all cases. It can reliably determine malignancy in papillary, medullary, and undifferentiated thyroid cancer. However, when the picture is that of a follicular neoplasm, the cytology is termed ‘suspicious’ or ‘indeterminate’, due to the inability of the method to differentiate between FTC and FTA. Furthermore, cytology is incapable of identifying high-risk tumors and thereby selecting the patients that would benefit from a total thyroidectomy.

A multitude of FTC markers have been proposed, but in subsequent studies none of them has proved sensitive and specific enough for use in routine diagnostics. This is probably due to the many borderline cases that are present in clinical practice, but were usually not considered in the initial studies.
Moreover, the proportion of FTAs among the follicular tumors is much higher than that of FTCs, which requires a very high specificity of a diagnostic test to avoid overtreatment of benign lesions.

In the current study, 26 potential molecular markers were evaluated by qRT-PCR in a panel of 75 follicular tumors that covered the entire clinical spectrum, from innocent FTAs, borderline AFTAs and MI-FTCs, to pernicious WI-FTCs and metastatic tumors. An extended follow-up of the patients was also considered to support, or in few cases to refine, the initial histopathological diagnosis. To increase the information provided by the gene expressions, we prospectively divided the tumors into four groups based on the histopathological and clinical data for each case (Table 2). As single genes could not independently stand as surrogates for malignancy or risk stratification, we searched for combinations of genes that could serve this purpose. To this end, we developed classification/prediction models using LRA. The algorithm for LRA is relatively simple, its theoretical basis is well established, and does not assume normality in the data.

We first investigated whether the mRNA expression of a combination of the selected genes could predict high-risk tumors. Five genes were selected by the LRA for this purpose (model 1). TERT (Fig. 1A) encodes for the reverse transcriptase domain of telomerase, which is necessary for telomerase activity. It was very early recognized as a potential marker of malignancy in follicular tumors by non-quantitative PCR methods (Saji et al. 1999). Its use as a surrogate marker was however limited by the high percentage of false-positive results, as it is detectable in some FTAs and also in the presence of inflammation (Umbricht et al. 2004). Trefoil factor 3 (TFF3; Fig. 1B) was first identified as an FTC marker by SAGE and was subsequently confirmed by qRT-PCR (Takano et al. 2004). Microarray studies also demonstrated its reduced expression in FTC (Barden et al. 2003, Finley et al. 2004). The peroxisome proliferator-activated receptor β (PPARβ) is a key player in follicular thyroid tumorigenesis. A PAX8-PPARβ fusion that inhibits wild-type PPARβ action is found in 19–63% of FTCs and seldom in FTAs (Kroll et al. 2000, Nikiforova et al. 2002, Dwight et al. 2003).

Figure 3 The classification model for high-risk disease (model 1). (A) The expression of the ten genes was compared between groups I–III versus group IV and the five genes with the highest t-score (absolute value) are shown in black staples: the reverse transcriptase domain of telomerase (TERT), the trefoil factor 3 (intestinal; TFF3), the peroxisome proliferator-activated receptor β (PPARβ), the Cbp/p300-interacting transactivator 1 (CITED1), and the early growth response 2 gene (EGR2).

Using coefficients defined by the regression analysis, the following classification rule equation was generated (model 1):

\[ X_\alpha = -4.1423 + 0.7876 \times \ln \text{TERT} - 0.7271 \times \ln \text{PPAR} - 0.7238 \times \ln \text{TFF3} + 0.1601 \times \ln \text{CITED1} - 0.1191 \times \ln \text{EGR2}, \]

where \( \ln \) is the natural logarithm of the expression value for the respective gene. A plus sign of the coefficient indicates that increased expression is associated with high-risk disease and a minus sign indicates that increased expression is associated with low-risk disease. The \( \mu \) score is \( \mu = 1/(1 + e^{-X_\alpha}) \), where \( e \) is the natural number and \( X_\alpha \) the logistic regression parameter for the model. The value of \( \mu \) is bounded between 0 and 1, and the tumor is classified as high risk if \( \mu > 0.5 \) (i.e. when \( X_\alpha > 0 \)) and low risk if \( \mu < 0.5 \) (i.e. when \( X_\alpha < 0 \)). (B) Scatterplot of the \( \mu \) scores in the tumors of groups I–IV obtained by the leave-one-out cross-validation of model 1.
We and others have previously shown that the expression of PPARγ is high in the presence of the fusion (due to the expression of the fusion product), FTCs and low in without a PPARγ fusion (Aldred et al. 2003, Lui et al. 2005). As shown in Fig. 1C, reduction of PPARγ expression is pronounced in advanced carcinomas (group IV). The Cbp/p300-interacting transactivator 1 (CITED1) has previously been implicated in papillary thyroid cancer, while only a subset of FTCs was shown to express it in high levels (Huang et al., 2001, Aldred et al. 2004). Higher levels of CITED1 in FTC compared with FTA were recently reported (Fryknäs et al. 2006). In this study, a highly increased expression of CITED1 mRNA was strongly correlated with advanced FTCs (group IV; Fig. 1D). Finally, the early growth response 2 (EGR2; Fig. 1E) was reported down-regulated in FTCs by Barden et al. (2003). It is also shown to possess tumor inhibitory properties and is involved in PTEN signaling (Unoki & Nakamura 2001), which in turn is dysregulated in highly malignant thyroid tumors (Gimm et al. 2000, Frisk et al. 2002). The classification model generated by the combination of the five genes (model 1) proved highly significant (robust) and highly specific (98.5%) for the prediction of aggressiveness in FTC. It can become the basis for the development of both a preoperative test for selecting high-risk patients for extensive surgery but also a postoperative test for prognostic purpose. It should be noted that all six patients in group IV whose initial treatment was suboptimal, including the two whose tumors were initially diagnosed as adenomas, were identified as high risk by the model (Table 2).

The second model (model 2) addressed the question of malignancy and was developed by comparing the unambiguously benign and malignant cases (groups I and IV). Thereby a tumor was classified as ‘benign’ if its gene expressions resembled those of the clearly benign FTAs and as ‘malignant’ when it was closer to the de facto aggressive FTCs. The use of only two genes (TERT, TFF3) renders this model less robust than the first one, but at the same time more sensitive in detecting malignancy potential in less obvious cases. It is noteworthy that all four tumors without histopathological signs of invasion in which malignancy was determined by the presence or later development of metastasis were correctly diagnosed as malignant by the model.

Interestingly, MI-FTCs were not uniformly classified. This reflects a common situation in clinical practice, where clinicians often have difficulties in treatment planning for these cases of uncertain malignancy potential, by applying evidence-based criteria. Notably, MI-FTCs as well as AFTAs were under-represented or not included in most of the published gene expression studies of follicular tumors (Barden et al. 2003, Cerutti et al. 2004, Finley et al. 2004, Weber et al. 2005). Furthermore, when used in microarray analyses, the
majority of MI-FTCs cluster together with adenomas (Barden et al. 2003, Lubitz et al. 2005).

The second model exhibited a strong tendency of classifying angioinvasive and large tumors as ‘malignant’. MI-FTCs with vascular invasion may indeed have a worse prognosis and are considered as a separate group named ‘moderately invasive FTC’ by some authors (D’Avanzo et al. 2004). The specificity of this model remained high (94%) as only two cases among FTAs and AFTAs were mispredicted as ‘malignant’. The transition of the test to the FNA level could therefore detect a large proportion of patients indicated for extensive surgery, without the risk of over-treating benign tumors. As long as the problem of inadequate or contaminated sampling is surpassed, the qRT-PCR method is relatively simple and inexpensive. Moreover, the development of a two-gene assay for routine use is more convenient and can be performed in a single multiplex qRT-PCR, by implementing two different fluorescent probes.

In summary, we report two algorithms for the prediction of high-risk disease and malignancy in follicular thyroid tumors, based on the mRNA expression of five and two genes respectively. The very promising specificity and robustness of the predictions indicate that the suggested models can be the foundations for the future development of diagnostic and prognostic tests in follicular thyroid tumors.

Acknowledgements

We thank Ms Lisa Ånfalk for her excellent assistance in collection of the tumor samples. This study was supported by grants from the Swedish Cancer Foundation, the Swedish Research Council, the Åke Wiberg Foundation, the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, the Nilsson-Ehle foundation, the Cancer Society in Stockholm, the Swedish Society for Medical Research, and the Stockholm County Council.

Funding

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References

Aldred MA, Morrison C, Gimm O, Huang-Yu, Liyanarachchi S, Pellegata NS, Gimm O, Jhiang S, Davuluri RV, de la Chapelle A, & Eng C. 2004 Papillary and follicular thyroid carcinomas show distinctly different microarray expression profiles and can be distinguished by a minimum of five genes. *Journal of Clinical Oncology* 22 3531–3539.


