A multi-gene approach to differentiate papillary thyroid carcinoma from benign lesions: gene selection using support vector machines with bootstrapping

Krzysztof Fujarewicz¹, Michał Jarząb³,⁴, Markus Eszlinger⁵, Knut Krohn⁵,⁶, Ralf Paschke⁵, Małgorzata Oczko-Wojciechowska², Małgorzata Wiench², Aleksandra Kukulśka², Barbara Jarząb² and Andrzej Świerniak¹,²

¹Systems Engineering Group, Institute of Automatic Control, Silesian University of Technology, 44-100 Gliwice, Poland
²Department of Nuclear Medicine and Endocrine Oncology, Institute of Oncology, Maria Skłodowska-Curie Memorial Cancer Center, Gliwice Branch, Wybrzeże Armii Krajowej 15, 44-100 Gliwice, Poland
³Departments of Tumor Biology and ⁴Clinical Oncology, Institute of Oncology, Maria Skłodowska-Curie Memorial Cancer Center, Gliwice Branch, 44-100 Gliwice, Poland
⁵III. Medical Department and ⁶Interdisciplinary Center of Clinical Research Leipzig, University of Leipzig, 04103 Leipzig, Germany

Requests for offprints should be addressed to B Jarząb; Email: bjarzab@io.gliwice.pl

Abstract

Selection of novel molecular markers is an important goal of cancer genomics studies. The aim of our analysis was to apply the multivariate bioinformatical tools to rank the genes – potential markers of papillary thyroid cancer (PTC) according to their diagnostic usefulness. We also assessed the accuracy of benign/malignant classification, based on gene expression profiling, for PTC. We analyzed a 180-array dataset (90 HG-U95A and 90 HG-U133A oligonucleotide arrays), which included a collection of 57 PTCs, 61 benign thyroid tumors, and 62 apparently normal tissues. Gene selection was carried out by the support vector machines method with bootstrapping, which allowed us 1) ranking the genes that were most important for classification quality and appeared most frequently in the classifiers (bootstrap-based feature ranking, BBFR); 2) ranking the samples, and thus detecting cases that were most difficult to classify (bootstrap-based outlier detection). The accuracy of PTC diagnosis was 98.5% for a 20-gene classifier, its 95% confidence interval (CI) was 95.9–100%, with the lower limit of CI exceeding 95% already for five genes. Only 5 of 180 samples (2.8%) were misclassified in more than 10% of bootstrap iterations. We specified 43 genes which are most suitable as molecular markers of PTC, among them some well-known PTC markers (MET, fibronectin 1, dipeptidylpeptidase 4, or adenosine A1 receptor) and potential new ones (UDP-galactose-4-epimerase, cadherin 16, gap junction protein 3, sushi, nidogen, and EGF-like domains 1, inhibitor of DNA binding 3, RUNX1, leiomodin 1, F-box protein 9, and tripartite motif-containing 58). The highest ranking gene, metallophosphoesterase domain-containing protein 2, achieved 96.7% of the maximum BBFR score.

Endocrine-Related Cancer (2007) ¹⁴ 809–826

Introduction

Discrimination between benign thyroid nodules and cancer is an important aspect of determining the optimal extent of thyroid surgery. Currently, this is achieved by routine morphologic assessment of cytopathology samples. However, this method does not allow proper classification of all thyroid tumors (Baloch & Livolsi 2002, Franc et al. 2003). At several institutions, genomic studies have been undertaken which besides focusing on basic biological issues (Huang et al. 2001, Giordano et al. 2005), also explore potential diagnostic applications (Aldred et al. 2004, Chevillard et al. 2004, Finley et al. 2004a,b). Our recent microarray-based analysis brought a 20-gene classifier to differentiate between papillary thyroid cancer (PTC) and normal thyroid tissue (Jarząb et al. 2005), further verified using three independent datasets (Eszlinger et al. 2006). Very large and easily
distinguishable differences between the molecular profiles of PTC and normal thyroid have clearly demonstrated the applicability of gene expression findings to diagnostic purposes. However, even more desirable for the clinician would be genomic profiling-based capability to discriminate between malignant tumors and various benign lesions. Therefore, we decided to use a balanced mixture of samples from malignant and benign tumors and normal thyroid tissue to mimic the clinical situation, where the material from any of these may be obtained and shall be properly classified. This large 180-array dataset is derived respectively from de novo studies \( (n = 40) \), previously published own microarray data \( (n = 124; \text{Eszlinger et al. 2001, 2004, Jarzab et al. 2005}) \), and accessible datasets published by other authors \( (n = 16; \text{Huang et al. 2001}) \).

We set the following goals for the study:

1. To assess accuracy of benign/malignant classification of thyroid specimens in relation to gene set size, in the context of PTC and
2. To optimize the list of diagnostically relevant genes in PTC.

To answer both questions, we used the support vector machines (SVMs) method with bootstrapping. This approach relies on iterative construction of SVM classifiers based on randomly selected sets of specimens (bootstrap samples) and testing the classifiers on remaining samples. We applied bootstrap to obtain both gene (feature) ranking and outlier detection. The ranking of the genes that are most important for classification quality was based on the frequency of their occurrence in the classifiers of different size (bootstrap-based feature ranking, BBFR). The ranking of the misclassified samples allowed to detect outliers (bootstrap-based outlier detection, BBOD) and to obtain a reliable estimate of classification accuracy with appropriate confidence intervals (CI) for gene sets of different size.

**Material and methods**

**Microarray data used in the study**

Microarray datasets from three sources were included in the analysis:

1. Dataset obtained in Gliwice, Poland; in total, 90 specimens analyzed with GeneChip HG-U133A microarrays. The specimens were collected from 71 patients with PTC (9 males and 40 females; mean age 36 years, range 6–71 years) and 22 with other thyroid diseases, 6 with follicular adenoma, 13 with nodular or colloid goiter and 3 with chronic thyroiditis (9 males and 13 females; mean age 45 years, range 11–71 years). The thyroid tissue specimens included 49 PTC tumors and 41 normal/benign thyroid tissue samples. The latter samples were from patients with PTC \( (n = 17) \) or other benign thyroid lesions \( (n = 24) \), among them six follicular adenomas, four nodular goiters, nine colloid goiters, and five cases of thyroiditis, two of them taken from the contralateral lobe from patients with PTC. Fifty microarrays were included in our previously published study and publicly available at www.genomika.pl/thyroidcancer (Jarzab et al. 2005); 40 microarrays were from de novo studies. All new samples were processed according to description given in Jarzab et al. (2005).

2. Dataset obtained in Leipzig, Germany; 74 specimens analyzed with GeneChip HG-U95Av2 microarrays. The specimens included 15 autonomously functioning thyroid nodules, 22 cold thyroid nodules, and 37 samples of their respective surrounding thyroid tissues. The analysis of these datasets was published previously (Eszlinger et al. 2001, 2004) and the datasets are available at http://www.uni-leipzig.de/innere/_forschung/schwerpunkte/etiology.html.

3. Dataset obtained in Columbus, OH, USA; 16 specimens analyzed with GeneChip HG-U95A microarrays. The specimens were derived from eight patients and included both PTC tumors and their surrounding thyroid tissues. The dataset (Huang et al. 2001) is publicly available at http://thinker.med.ohio-state.edu.

In total, the three analyzed datasets comprised 57 PTCs, 61 benign thyroid lesions, and 62 apparently normal thyroid tissues analyzed on 180 GeneChips of two different generations. Half of them were U133A and the rest U95A platforms.

**Data pre-processing and generation of datasets**

Each dataset was pre-processed by the MAS5 algorithm. To compare the expression data generated using the U95A GeneChips (12 625 probe sets) with those from the U133A GeneChips (22 283 probe sets), we used the ‘Human Genome U95 to Human Genome U133 Best Match Comparison Spreadsheet’ (www.affymetrix.com/support/technical/comparison_spreadsheets.affx) which yielded an intersection of 9530 probe sets. The obtained data were log2 transformed.
Neighborhood analysis and recursive elimination in gene selection

For selection of gene sets with diagnostic potential, we applied here the recursive feature elimination (RFE) algorithm (Guyon et al. 2002) which is computationally less demanding than recursive feature replacement used in our previous studies (Jarzab et al. 2005, Eszlinger et al. 2006). The introductory gene selection was performed using neighborhood analysis (200 genes; Golub et al. 1999, Slonim et al. 2000), further selection of the 100 best genes set was carried out by RFE.

SVMs and classification

The linear SVM (Boser et al. 1992, Vapnik 1995) was used for developing the classification rule. As mentioned earlier, the classifier was independently trained for different numbers of selected genes (from 1 to 100).

Bootstrap for estimation of classifier accuracy and its CI

In order to determine the accuracy of the developed classifier, we performed classical bootstrap procedure in 500 resampling iterations (selection with equal probability and return of samples; Efron 1979). Iterations of all stages of the classifier construction (i.e. gene preselection, gene selection, and classifier learning) were performed in each bootstrap, as suggested previously (Simon et al. 2003). The accuracy of the classifier was calculated using the 0.632 bootstrap estimator (Efron 1983). The distribution of the misclassification rate obtained during all bootstrap runs was used to estimate the 95% CI. The accuracy of the classifier and the CI were calculated for different numbers of selected genes (up to 100).

Bootstrap based feature ranking (BBFR) and outlier detection (BBOD)

The primary purpose of the bootstrap used in this study was to estimate the accuracy of the molecular classifier for different sizes of gene subsets with appropriate CIs. However, the computational effort for the bootstrap technique may also be exploited to derive some additional information. We apply two methods that use the information collected during bootstrapping: BBFR and BBOD. They are similar to the methods of statistical learning based on resampling, such as bagging and boosting. In both techniques, an ensemble of many base classifiers is created. Each base classifier is trained on different bootstrap subsamples. The final decision is based on decisions of all base classifiers. The simplest approach is bagging (bootstrap aggregating) originally proposed by Breiman (1996).

In bagging, the subsamples are randomly drawn as in classical bootstrapping where each observation is picked with the same probability \( 1/m \), where \( m \) is the number of all observations. The final decision is the decision of most base classifiers. In boosting, different observations may be picked with different probability and the final decision is weighted sum of decisions of base classifiers. The well-known boosting algorithm is AdaBoost (Freund & Schapire 1996).

In our approach, we do not create an ensemble (committee) of many base classifiers but we use the information collected during bootstrap-based validation step of the SVM classifier.

Let the data contain \( m \) instances (observations). One instance is a vector of \( N_{\text{max}} \) features (gene expression values) with a corresponding class label specified by an expert. Let \( L_b \) be the number of bootstrap iterations. In each run, we select (with equal probability and return of samples) \( m \) instances from the dataset (bootstrap sample). Then, the bootstrap sample is used for feature selection and classifier learning. Finally, the classifier is tested on the test set containing all instances not belonging to the bootstrap sample.

To find the optimal size for the feature set, we select \( N \) feature sets \( \Omega_1, \Omega_2, \ldots, \Omega_N \) of sizes 1,2,\ldots,\( N \) respectively. In general, selected sets may not overlap, but in most commonly used feature selection methods, based on feature ranking or backward/forward searching, feature subsets satisfy the relation

\[
\Omega_1 \subset \Omega_2 \subset \ldots \subset \Omega_N
\]  

BBFR

Let \( r_j(i) \) be a number of subsets \( \Omega_i, i=1,2,\ldots,N \) where the gene \( j \) belongs to. For gene selection methods satisfying equation (1), we have

\[
\begin{align*}
n_j(i) &= N - \text{(position in the feature ranking)} + 1. \quad (2)
\end{align*}
\]

The BBFR score \( R_j \) of the feature \( j \) is defined as a sum of \( r_j(i) \) over all bootstrap runs as follows:

\[
R_j = \sum_{i=1}^{L_B} r_j(i) \quad (3)
\]

The maximum possible value of the BBFR score is \( L_B N \).

BBOD

Let \( q_k \) be the number of bootstrap iterations where the observation \( k \) is chosen as a test instance (not a member
of the bootstrap sample). Let \( q_k^{\text{true}} \) be the number of bootstrap iterations where the instance \( j \) is correctly classified at the test stage.

The \textit{BBOD score} for \( k \)-th observation is

\[
Q_k = \frac{q_k^{\text{true}}}{q_k}
\] (4)

The value of \( Q_k \) belongs to the interval \((0,1)\) and the low value indicates outliers.

\textbf{Comparison of different class prediction methods}

We used BRB ArrayTools (developed by Dr Richard Simon and Amy Peng Lam) to compare different class prediction algorithms (Compound Covariate Predictor, Linear Diagonal Discriminant Analysis, Nearest Centroid, 1-Nearest Neighbor, 3-Nearest Neighbors and SVMs). To compute misclassification rate, 0.632 bootstrap cross-validation method was used. All genes with univariate misclassification rate below 0.2 were used for this analysis.

\textbf{Results}

\textbf{Accuracy of malignant/benign classification and redundancy of PTC gene classifiers}

The huge difference in gene expression between PTC and benign/normal thyroid tissues implies that many multi-gene classifiers with similar classification ability may be created. For preliminary assessment of accuracy of the differentiation between PTC and benign lesions or normal thyroid, we randomly divided the 180-array dataset into two subgroups, according to sample number: A (odd numbers) and B (even numbers). Each subgroup contained data from similar number of benign and malignant tumor specimens analyzed with U133A or U95A GeneChips. We used set A to obtain a 20-gene classifier; this classifier was tested on set B and the procedure was repeated, using set B as a training set and testing the classifier on set A. Using the classifier obtained from set A, we were able to correctly predict 86 out of 90 samples (95.6%) within set B, while using the classifier obtained from set B, we accurately diagnosed 88 out of 90 samples in set A (97.8%). Both classifiers differed partly from our previous 20-gene classifier (37) obtained on a smaller dataset.

To avoid a bias in gene selection and accuracy estimation, related to the arbitrary selection of the training set, we carried out the procedure of accuracy estimation by bootstrapping, i.e. randomly selecting large numbers of slightly different training sets and validating them on the remaining samples. This procedure allows using sufficiently large training sets while simultaneously obtaining a reliable estimation of classification accuracy. By applying this method, we estimated the accuracy of discrimination between benign and malignant samples to be 98.6%, with a rather narrow CI (see Fig. 1). For small gene sets, the accuracy was a bit lower (93.7% for one-gene set, 96.9% for two-gene set, 97.9 for three-gene set, and from 98.3 to 98.6 for larger sets, up to \( n = 100 \)). For the 20-gene classifier, the accuracy was 98.5% and the estimated 95% CI was 95.9–100% for the classifiers built from more than five genes.

We compared the results of classification by the best 500 genes (Fig. 2) with the classification by consecutive 500-gene sets (i.e. first 500, 500–1000, 1000–1500, etc). We noted that only the first 500 genes allow accurately classifying samples by single genes or small gene sets. Genes ranked 500–1000 achieved 90% accuracy only for classifiers larger than 50 genes, while genes beyond the first 1000 hardly achieve this limit of accuracy. When we excluded all genes analyzed in Fig. 2 (8×500 = 4000), the accuracy obtained for small sets was only \( \sim 60\% \), close to random. However, the accuracy rose with gene set size, and for classifier sets larger than 700 genes it achieved 90% (data not shown). These results support the conclusion that the PTC transcriptome differs from the normal one in thousands of genes; they also provide evidence that optimizing a diagnostic gene set is a necessary step of analysis in order to make this set useful for molecular PTC classification.

\textbf{Ranking of PTC genes for their classification ability}

To obtain the ranking of genes based on their usefulness in the diagnostic context, we performed subsequent repetitive gene selection process by bootstrapping of the whole dataset. We ranked all genes according to the frequency of
appearance within the selected gene sets (BBFR). Genes important for the majority of diagnostic datasets were highly ranked, while less importance was given to complementing transcripts, which exhibited higher variability (Fig. 3). During the selection process, 365 transcripts occurred at least once within the obtained classifiers and some of them were present in nearly all classifiers. The maximum theoretical score to be obtained by a gene was $5 \times 10^4$ and the gene with the best rank, encoding metallophosphoesterase domain-containing protein 2 (MPPED2), had a score of $4.84 \times 10^4$, i.e. 96.7% of the maximum one. The first 20 genes were given scores $>3.74 \times 10^4$ (>77% of the maximum score), only slightly lower than the top gene, and the first 100 transcripts were characterized by scores $>0.64 \times 10^4$, which is >13.2% of the maximum score obtained. In total, 43 transcripts representing 41 genes scored higher than half of the value for the top gene ($>2.42 \times 10^4$, Fig. 3). Among them, there were both genes known for their changed expression in PTC or described in previous microarray studies, some used already as single markers, as well as new genes, not considered previously for their diagnostic potential (Table 1).

We analyzed fold-change differences between PTC and benign thyroid samples for the 43 selected transcripts to evaluate the potential influence of interplatform differences on the obtained gene selection. Twenty of them showed more than fourfold increase (log ratio >2) and four transcripts were increased more than twice, whereas the remaining 19 transcripts were decreased. Generally, the consistency between fold-changes observed in subsets from U95 and U133 arrays was good, although for some genes (e.g. the well-known thyroid cancer markers fibronectin 1 (FN1) and MET or novel genes cadherin 16 (CDH16) or gap junction protein β-3 (GJB3)) there were interplatform differences between the log ratios. However, 40 out of 43 selected genes exhibited more than twofold change in both the U133 and the U95 subsets. For all 43 genes, the PTC–benign difference was larger than the difference between fold-changes obtained with different GeneChip generation subsets. This confirms that the selection performed was robust to inter-array differences.

**Misclassified thyroid samples**

The algorithm with bootstrapping allows ranking the samples according to the frequency of their misclassification (Table 2). BBOF showed very frequent misclassifications for two samples. One of them was not properly classified by any gene set selected, and this was sample no. 154 from the U133 dataset no. 1, a small (10 mm in diameter) familial PTC found within a larger follicular adenoma. It was observed in an 18-year-old woman. A year later her mother, 43 years old, was diagnosed with 0.7 cm PTC (follicular
Table 1: Ranking of papillary thyroid cancer (PTC) genes as assessed by bootstrap-based feature ranking (BBFR) approach. For each transcript selected, rank and score obtained by the BBFR method are given, together with basic univariate statistics (log2 mean and log2 ratio).

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Affy_ID (U133)</th>
<th>Rank</th>
<th>Score</th>
<th>PTC mean log2</th>
<th>Benign mean log2</th>
<th>Log ratio</th>
<th>Log ratio U133</th>
<th>Log ratio U95</th>
<th>References of microarray or other high throughput studies*</th>
<th>Other data relevant for functional role in thyroid cancer</th>
<th>Gene function†</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPPED2</td>
<td>Metallophosphosterase domain-containing protein 2</td>
<td>205413_at</td>
<td>1</td>
<td>48449</td>
<td>4.66</td>
<td>7.82</td>
<td>−3.16</td>
<td>−3.46</td>
<td>−3.26</td>
<td>Aldred et al. (2003, 2004), Mazzanti et al. (2004) and Griffith et al. (2006)</td>
<td>Fetal brain protein of unknown function</td>
<td></td>
</tr>
<tr>
<td>HHA2</td>
<td>Hemoglobin, α1-hemoglobin, a-2</td>
<td>209458_x_at</td>
<td>2</td>
<td>45521</td>
<td>9.79</td>
<td>12.04</td>
<td>−2.25</td>
<td>−2.28</td>
<td>−1.87</td>
<td>Griffith et al. (2006)</td>
<td>Oxygen transport</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>Met proto-oncogene (hepatocyte growth factor receptor)</td>
<td>213807_x_at</td>
<td>3</td>
<td>43363</td>
<td>8.15</td>
<td>5.22</td>
<td>2.93</td>
<td>1.73</td>
<td>2.60</td>
<td>Barden et al. (2003), Wasenius et al. (2003), Finley et al. (2004a,b), Prasad et al. (2004), Zou et al. (2004) and Giordano et al. (2005)</td>
<td>Membrane tyrosine kinase receptor enhances cell motility, invasiveness, and chemokine production (Ruco et al. (2001))</td>
<td>Membrane tyrosine kinase receptor enhances cell motility, invasiveness, and chemokine production (Ruco et al. (2001))</td>
</tr>
<tr>
<td>FN1</td>
<td>Fibronectin 1</td>
<td>210495_x_at</td>
<td>4</td>
<td>44617</td>
<td>12.24</td>
<td>8.72</td>
<td>3.52</td>
<td>2.75</td>
<td>3.91</td>
<td>Chen et al. (2001), Barden et al. (2003), Wasenius et al. (2003), Finley et al. (2004a,b), Prasad et al. (2004), Giordano et al. (2005), Hamada et al. (2005) and Griffith et al. (2006)</td>
<td>Extracellular matrix glycoprotein participates in cell adhesion, regulates proliferation and survival of thyroid cells via integrin receptors (Iliari et al. (2003))</td>
<td></td>
</tr>
<tr>
<td>GALE</td>
<td>UDP-galactose-4-epimerase</td>
<td>202528_at</td>
<td>5</td>
<td>43974</td>
<td>7.12</td>
<td>3.70</td>
<td>3.42</td>
<td>2.41</td>
<td>3.50</td>
<td>Aldred et al. (2003) Chevillard et al. (2004), Finley et al. (2004a,b) and Griffith et al. (2004)</td>
<td>Converts glucose to galactose and N-acetylglucosamine to its UDP-derivatives</td>
<td></td>
</tr>
<tr>
<td>QPCT</td>
<td>Glutaminyl-peptide cyclotransferase (glutaminyl cyclase)</td>
<td>205174_s_at</td>
<td>6</td>
<td>43317</td>
<td>7.60</td>
<td>4.99</td>
<td>2.61</td>
<td>3.14</td>
<td>2.56</td>
<td>Barden et al. (2003), Chevillard et al. (2004), Finley et al. (2004a,b) and Griffith et al. (2004)</td>
<td>Converts glutaminyl peptides to cyclic pyroglutamyl ones</td>
<td></td>
</tr>
<tr>
<td>NELL2</td>
<td>NEL-like 2 (chicken)</td>
<td>203413_at</td>
<td>7</td>
<td>42953</td>
<td>9.68</td>
<td>7.56</td>
<td>2.12</td>
<td>2.02</td>
<td>2.53</td>
<td>Barden et al. (2003) and Finley et al. (2004a,b)</td>
<td>Brain protein with six EGF-like repeats</td>
<td></td>
</tr>
<tr>
<td>PGCP</td>
<td>Plasma glutamate carboxypeptidase</td>
<td>203501_at</td>
<td>8</td>
<td>42153</td>
<td>7.70</td>
<td>9.23</td>
<td>−1.53</td>
<td>−1.05</td>
<td>−1.33</td>
<td>Aldred et al. (2003, 2004), Barden et al. (2003), Finley et al. (2004a,b), Weber et al. (2005) and Barquis et al. (2006)</td>
<td>Breakdown of secreted peptides, hromocigous to prostate membrane-specific antigen (Gimbras et al. (1999))</td>
<td></td>
</tr>
<tr>
<td>ADORA1</td>
<td>Adenosine A1 receptor</td>
<td>205481_at</td>
<td>10</td>
<td>41699</td>
<td>7.16</td>
<td>4.85</td>
<td>2.30</td>
<td>2.00</td>
<td>2.81</td>
<td>Aldred et al. (2003, 2004) and Prasad et al. (2004)</td>
<td>Membrane receptor, stimulates motility and modulates proliferation</td>
<td></td>
</tr>
<tr>
<td>RYR1</td>
<td>Ryanodine receptor 1 (skeletai)</td>
<td>205485_at</td>
<td>12</td>
<td>40473</td>
<td>6.96</td>
<td>4.70</td>
<td>2.27</td>
<td>2.53</td>
<td>1.92</td>
<td>Barden et al. (2003) and Finley et al. (2004a,b)</td>
<td>Calcium release channel of the sarcoplasmic reticulum</td>
<td></td>
</tr>
<tr>
<td>CDH16</td>
<td>Cadherin 16, KSP-cadherin</td>
<td>206517_at</td>
<td>13</td>
<td>39770</td>
<td>3.47</td>
<td>8.07</td>
<td>−4.60</td>
<td>−4.68</td>
<td>−1.43</td>
<td>Present mainly in excitable cells Thought to be kidney specific (Thomson et al. (1995))</td>
<td>Calcium-dependent, membrane-associated glycoprotein, participates in cell adhesion</td>
<td></td>
</tr>
<tr>
<td>GJB3</td>
<td>Gap junction protein i-3, 31 kDa (connexin 31)</td>
<td>205490_x_at</td>
<td>14</td>
<td>39526</td>
<td>6.49</td>
<td>4.04</td>
<td>2.44</td>
<td>2.71</td>
<td>0.62</td>
<td>Barden et al. (2003) Finley et al. (2004a,b)</td>
<td>Does not normally appear in thyroid, in adult mouse becomes restricted to epidermis, testes and placenta (Tomlins et al. (2000), Plum et al. (2002) and Green et al. (2005))</td>
<td>Extracellular matrix protein, able to promote cell movements (Spessotto et al. (2003))</td>
</tr>
<tr>
<td>EMID1</td>
<td>EMI domain containing 1</td>
<td>213779_at</td>
<td>15</td>
<td>39505</td>
<td>6.44</td>
<td>8.12</td>
<td>−1.68</td>
<td>−1.09</td>
<td>−0.76</td>
<td>Barden et al. (2003), Cerutti et al. (2004) and Finley et al. (2004a,b)</td>
<td>Extracellular matrix protein, able to promote cell movements (Spessotto et al. (2003))</td>
<td></td>
</tr>
<tr>
<td>NRIP1</td>
<td>Nuclear receptor-interacting protein 1</td>
<td>202599_s_at</td>
<td>16</td>
<td>39358</td>
<td>8.31</td>
<td>6.33</td>
<td>1.98</td>
<td>1.36</td>
<td>2.06</td>
<td>Barden et al. (2003) and Finley et al. (2004a,b)</td>
<td>Interacts with nuclear receptors</td>
<td></td>
</tr>
</tbody>
</table>
**Table 1 (continued)**

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Affy_ID (U133)</th>
<th>Rank</th>
<th>Score</th>
<th>PTC mean log2</th>
<th>Benign mean log2</th>
<th>Log ratio</th>
<th>Log ratio U133</th>
<th>Log ratio U95</th>
<th>References of microarray or other high throughput studies</th>
<th>Referred to in single studies of thyroid cancer</th>
<th>Other data relevant for functional role in thyroid cancer</th>
<th>Gene function</th>
</tr>
</thead>
<tbody>
<tr>
<td>MET</td>
<td>Met proto-oncogene (hepatocyte growth factor receptor)</td>
<td>211599_x_at</td>
<td>17</td>
<td>39.348</td>
<td>8.44</td>
<td>5.68</td>
<td>2.76</td>
<td>1.53</td>
<td>2.54</td>
<td>Barden et al. (2003), Wasenius et al. (2003), Finley et al. (2004a,b), Prasad et al. (2004), Zou et al. (2004) and Giordano et al. (2005)</td>
<td>See the information given above for another probe set of the same gene</td>
<td>Participation in protein ubiquination</td>
<td>Anti-apoptotic protein</td>
</tr>
<tr>
<td>DTX4</td>
<td>Delta 4 homolog (Drosophila)</td>
<td>212611_at</td>
<td>18</td>
<td>39.298</td>
<td>10.24</td>
<td>8.24</td>
<td>2.00</td>
<td>2.07</td>
<td>1.45</td>
<td>Prasad et al. (2004)</td>
<td>Participation in protein ubiquination</td>
<td>Prenylated membrane bound protein with GTP-ase function</td>
<td></td>
</tr>
<tr>
<td>RAB27A</td>
<td>RAB27A, member RAS oncogene family</td>
<td>210991_x_at</td>
<td>19</td>
<td>39.913</td>
<td>8.62</td>
<td>6.52</td>
<td>3.00</td>
<td>1.60</td>
<td>1.24</td>
<td>Barden et al. (2003), Finley et al. (2000a,b), Weber et al. (2005), Musholt et al. (2006) and Sarquis et al. (2006)</td>
<td>Participation in protein ubiquination</td>
<td>Prenylated membrane bound protein with GTP-ase function</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>CDNA clone IMAGE:4152983</td>
<td>214803_at</td>
<td>20</td>
<td>37.397</td>
<td>7.30</td>
<td>5.39</td>
<td>1.90</td>
<td>1.94</td>
<td>1.13</td>
<td></td>
<td>Not identified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL2</td>
<td>B-cell CLL/lymphoma 2</td>
<td>203684_s_at</td>
<td>21</td>
<td>36.483</td>
<td>2.92</td>
<td>5.88</td>
<td>-2.95</td>
<td>-2.74</td>
<td>-1.39</td>
<td>Hoo et al. (2002), Baris et al. (2004), Bhaskar et al. (2004), Wreehmann et al. (2004), Giordano et al. (2005) and Jacques et al. (2005) and Giordano et al. (2005)</td>
<td></td>
<td>Anti-apoptotic protein</td>
<td></td>
</tr>
<tr>
<td>TACST2</td>
<td>Tumor-associated calcium signal transducer 2</td>
<td>202296_s_at</td>
<td>22</td>
<td>36.170</td>
<td>10.42</td>
<td>6.29</td>
<td>4.13</td>
<td>4.02</td>
<td>4.02</td>
<td></td>
<td>May serve as cell surface receptor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIO1</td>
<td>Deiodinase, iodothyronine, type 1</td>
<td>206457_s_at</td>
<td>23</td>
<td>35.971</td>
<td>6.19</td>
<td>9.94</td>
<td>-3.75</td>
<td>-3.79</td>
<td>-4.33</td>
<td>Ezzlinger et al. (2001, 2004), Huang et al. (2001), Barden et al. (2003), Finley et al. (2000a,b), Prasad et al. (2004), Wreehmann et al. (2004), Giordano et al. (2005) and Griffith et al. (2006)</td>
<td></td>
<td>5' Deiodination of thyroxine</td>
<td></td>
</tr>
<tr>
<td>ITPRI</td>
<td>Inositol 1,4,5-triphosphate receptor, type 1</td>
<td>203710_at</td>
<td>24</td>
<td>34.804</td>
<td>6.75</td>
<td>8.84</td>
<td>-2.09</td>
<td>-2.06</td>
<td>-1.83</td>
<td>Barden et al. (2003), Finley et al. (2000a,b), Prasad et al. (2004), Wreehmann et al. (2004) and Hamada et al. (2005)</td>
<td>Signal transducer coupled with calcium channels, participates in apoptosis (Sedlak &amp; Snyder (2006))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBB</td>
<td>Hemoglobin b</td>
<td>209116_x_at</td>
<td>25</td>
<td>34.591</td>
<td>9.62</td>
<td>12.13</td>
<td>-2.51</td>
<td>-2.48</td>
<td>-1.24</td>
<td>Aldred et al. (2003, 2004) and Orda et al. (2005)</td>
<td></td>
<td>See above the HBA gene</td>
<td></td>
</tr>
<tr>
<td>SNED1</td>
<td>Sushi, nidogen, and EGF-like domains 1</td>
<td>213493_at</td>
<td>26</td>
<td>33.625</td>
<td>2.87</td>
<td>5.88</td>
<td>-3.01</td>
<td>-2.14</td>
<td>-2.06</td>
<td></td>
<td>Participation in cell-matrix adhesion, contains sushi, nidogen and calcium-binding domains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHR</td>
<td>Aryl hydrocarbon receptor</td>
<td>202820_at</td>
<td>27</td>
<td>33.003</td>
<td>7.52</td>
<td>6.02</td>
<td>1.50</td>
<td>1.20</td>
<td>1.59</td>
<td>Barden et al. (2003), Wasenius et al. (2003) and Finley et al. (2004a,b)</td>
<td></td>
<td>AHR-dependent enzyme responsible for aromatic ring degrading</td>
<td></td>
</tr>
<tr>
<td>HGD</td>
<td>Homogentisic acid 1,2-dioxygenase (homogenisic oxdase)</td>
<td>205221_at</td>
<td>28</td>
<td>32.816</td>
<td>4.57</td>
<td>7.83</td>
<td>-3.26</td>
<td>-3.17</td>
<td>-3.92</td>
<td>Huang et al. (2001), Aldred et al. (2003), Barden et al. (2003), Aldred et al. (2004), Finley et al. (2004a,b), Prasad et al. (2004) and Giordano et al. (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RXRG</td>
<td>Retinoid X receptor, y</td>
<td>205954_at</td>
<td>29</td>
<td>32.444</td>
<td>7.35</td>
<td>4.69</td>
<td>2.66</td>
<td>2.80</td>
<td>2.62</td>
<td>Haugen et al. (2004)</td>
<td></td>
<td>Heterodimer partner of several nuclear receptors</td>
<td></td>
</tr>
<tr>
<td>CA4</td>
<td>Carbonic anhydrase IV</td>
<td>206209_s_at</td>
<td>30</td>
<td>31.332</td>
<td>6.33</td>
<td>8.51</td>
<td>-2.18</td>
<td>-2.62</td>
<td>-1.41</td>
<td>Barden et al. (2003), Finley et al. (2004a,b), Weber et al. (2005) and Sarquis et al. (2006)</td>
<td></td>
<td>An ancient isozyme</td>
<td></td>
</tr>
<tr>
<td>SDC4</td>
<td>Syndecan 4 (amphipathic, nedd26)</td>
<td>203071_at</td>
<td>31</td>
<td>28.036</td>
<td>10.76</td>
<td>8.31</td>
<td>2.45</td>
<td>1.86</td>
<td>2.41</td>
<td>Barden et al. (2003), Chevillard et al. (2004), Finley et al. (2004a,b), Prasad et al. (2004) and Griffith et al. (2006)</td>
<td></td>
<td>Transmembrane heparan sulfate proteoglycan involved in the organization of the actin cytoskeleton and in cell-matrix interactions, binds fibronectin, behaves as CXCL12 receptor (Lin et al. (2005))</td>
<td></td>
</tr>
<tr>
<td>Gene symbol</td>
<td>Gene name</td>
<td>Affy_ID (U133)</td>
<td>Rank</td>
<td>Score</td>
<td>PTC mean log2</td>
<td>Benign mean log2</td>
<td>Log ratio</td>
<td>Log ratio U133</td>
<td>Log ratio U95</td>
<td>References of microarray or other high throughput studies</td>
<td>Referred to in single studies of thyroid cancer</td>
<td>Other data relevant for functional role in thyroid cancer</td>
<td>Gene function</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>----------------</td>
<td>------</td>
<td>-------</td>
<td>---------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>----------------</td>
<td>-------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ENTPD1</td>
<td>Ectonucleoside triphosphate diphosphohydrolase 1</td>
<td>209473_at</td>
<td>32</td>
<td>27</td>
<td>859</td>
<td>8.71</td>
<td>6.75</td>
<td>1.97</td>
<td>1.49</td>
<td>Weber et al. (2005) and Sarquis et al. (2006)</td>
<td></td>
<td>Membrane bound enzyme converts adenine nucleotides to adenosine, interacts with caveolin 1 and 2 (Kittel et al. (2004))</td>
<td></td>
</tr>
<tr>
<td>TPO</td>
<td>Thyroid peroxidase</td>
<td>210342_s_at</td>
<td>33</td>
<td>27</td>
<td>658</td>
<td>7.29</td>
<td>12.24</td>
<td>-4.95</td>
<td>-4.93</td>
<td>Barden et al. (2003), Cerutti et al. (2004), Finley et al. (2004a,b) and Griffith et al. (2006)</td>
<td>Barden et al. (2003), Chevillard et al. (2004), Finley et al. (2004a,b), Prasad et al. (2004) and Griffith et al. (2006)</td>
<td>Thyroid-specific enzyme crucial for organization of iodine and synthesis of thyroid hormones</td>
<td></td>
</tr>
<tr>
<td>KRT19</td>
<td>Keratin 19</td>
<td>201650_at</td>
<td>34</td>
<td>27</td>
<td>398</td>
<td>8.92</td>
<td>5.71</td>
<td>3.22</td>
<td>3.55</td>
<td>Arturi et al. (1997), Lazar et al. (1999) and Schellhout et al. (1989)</td>
<td></td>
<td>The smallest known keratin expressed in some types of cancer</td>
<td></td>
</tr>
<tr>
<td>ID3</td>
<td>Inhibitor of DNA binding 3, dominant negative helix-loop-helix protein</td>
<td>207826_s_at</td>
<td>35</td>
<td>26</td>
<td>271</td>
<td>9.17</td>
<td>11.25</td>
<td>-2.08</td>
<td>-1.26</td>
<td></td>
<td></td>
<td>Downstream target of pituitary tumor transforming gene (PTTG)</td>
<td></td>
</tr>
<tr>
<td>RUNX1</td>
<td>Runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)</td>
<td>209360_s_at</td>
<td>36</td>
<td>26</td>
<td>202</td>
<td>7.37</td>
<td>12.24</td>
<td>4.95</td>
<td>4.93</td>
<td></td>
<td></td>
<td>Transcription factor may promote E-cadherin expression (Liu et al. (2005))</td>
<td></td>
</tr>
<tr>
<td>LMOD1</td>
<td>Leiomodin 1 (smooth muscle)</td>
<td>203766_s_at</td>
<td>37</td>
<td>26</td>
<td>044</td>
<td>5.60</td>
<td>7.80</td>
<td>-2.20</td>
<td>-2.77</td>
<td></td>
<td></td>
<td>Present both in thyroid cells and eye muscle (Kromminga et al. (1998))</td>
<td></td>
</tr>
<tr>
<td>RAB27A</td>
<td>RAB27A, member RAS oncogene family</td>
<td>209514_s_at</td>
<td>38</td>
<td>25</td>
<td>684</td>
<td>8.57</td>
<td>6.29</td>
<td>2.28</td>
<td>1.43</td>
<td>Barden et al. (2003), Finley et al. (2004a,b), Weber et al. (2005) and Sarquis et al. (2006)</td>
<td></td>
<td>See above information on the alternative probeset identifying the same gene</td>
<td></td>
</tr>
<tr>
<td>FBXO9</td>
<td>F-box protein 9</td>
<td>212987_at</td>
<td>39</td>
<td>25</td>
<td>331</td>
<td>8.47</td>
<td>9.29</td>
<td>-0.83</td>
<td>-0.50</td>
<td></td>
<td></td>
<td>Members of this gene family in complexes may act as protein–ubiquitin ligases</td>
<td></td>
</tr>
<tr>
<td>TRIM58</td>
<td>Tripartite motif-containing 58</td>
<td>215047_at</td>
<td>40</td>
<td>25</td>
<td>304</td>
<td>3.91</td>
<td>6.99</td>
<td>-3.08</td>
<td>-2.27</td>
<td></td>
<td></td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>ICAM1</td>
<td>Intercellular adhesion molecule 1 (CD54), human rhinovirus receptor</td>
<td>202638_s_at</td>
<td>42</td>
<td>24</td>
<td>534</td>
<td>8.18</td>
<td>5.61</td>
<td>2.57</td>
<td>1.70</td>
<td>Kawai et al. (1998)</td>
<td></td>
<td>Epithelial adhesion molecule plays a key role in lymphocyte infiltration into the thyroid</td>
<td></td>
</tr>
</tbody>
</table>

The original papers (Eszlinger et al. 2001, 2004, Huang et al. 2001, Jarzab et al. 2005) containing datasets included in the present study were not cited here. RXRG was listed in our previous microarray-based analysis (Jarzab et al. 2005), together with FN1, MET, KRT19, DPP4, HBB, QPCT, GJB3, and DTX4, also occurring in this table.

aOMIM-based information if not otherwise specified.

bDenotes immunohistochemistry studies.
variant). The other one, properly classified only in 8% of runs, was a benign follicular adenoma (diagnosed as atypical) from the same dataset (sample no. 97) which was derived from a 15-year-old boy of another family with familial PTC. In this family, there were two PTC cases (mother of the patient, diagnosed with pT2BNxM0 PTC and her aunt who died of a dissemination of PTC) and one follicular thyroid cancer case (pT2bNxM0, 11 years old, sister of the patient). These were the only two cases with a positive family history of thyroid cancer among 49 Polish patients included in the study. Two further samples were properly classified in 65–68% of runs (one from dataset no. 1 and one from dataset no. 3), again one benign adenoma and one PTC, respectively. For the fifth sample, the accuracy was much higher and it was properly classified in 88% of the runs. Thus, only 5 out of 180 samples (2.8%) were misclassified in more than 10% of the runs, while a total of 14 samples (7.8%) were misclassified in more than 1% of the runs. Seventy samples were classified with an excellent accuracy between 99 and 100%, and for further 64 cases no misclassification occurred during the bootstrapping process.

Comparison of classification accuracy by different class prediction methods

To evaluate our method, we compared the accuracy of prediction by different class prediction methods implemented in BRB-Array software. We based the class prediction on all genes that showed the univariate misclassification rate lower than 20%. We found out that the classification accuracy ranged from 89% (compound covariate predictor method) to 99%.

Table 2

Ranking of thyroid samples by tumor–normal misclassification frequency, assessed by bootstrap-based outlier detection (BBOD) approach. The BBOD rank and score $Q_k$, as defined in Material and methods, is given.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Status</th>
<th>Array</th>
<th>Set</th>
<th>Rank</th>
<th>Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>154</td>
<td>PTC</td>
<td>U133</td>
<td>B</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>97</td>
<td>Benign</td>
<td>U133</td>
<td>A</td>
<td>2</td>
<td>7.23</td>
</tr>
<tr>
<td>148</td>
<td>PTC</td>
<td>U133</td>
<td>B</td>
<td>3</td>
<td>65.34</td>
</tr>
<tr>
<td>95</td>
<td>Benign</td>
<td>U133</td>
<td>A</td>
<td>4</td>
<td>68.25</td>
</tr>
<tr>
<td>88</td>
<td>PTC</td>
<td>U95v1</td>
<td>B</td>
<td>5</td>
<td>88.28</td>
</tr>
<tr>
<td>166</td>
<td>PTC</td>
<td>U133</td>
<td>B</td>
<td>6</td>
<td>90.02</td>
</tr>
<tr>
<td>84</td>
<td>PTC</td>
<td>U95v1</td>
<td>B</td>
<td>7</td>
<td>93.11</td>
</tr>
<tr>
<td>161</td>
<td>PTC</td>
<td>U133</td>
<td>A</td>
<td>8</td>
<td>95.96</td>
</tr>
<tr>
<td>94</td>
<td>Benign</td>
<td>U133</td>
<td>B</td>
<td>9</td>
<td>97.26</td>
</tr>
<tr>
<td>116</td>
<td>Normal</td>
<td>U133</td>
<td>B</td>
<td>10</td>
<td>97.30</td>
</tr>
<tr>
<td>120</td>
<td>PTC</td>
<td>U133</td>
<td>A</td>
<td>11</td>
<td>97.98</td>
</tr>
<tr>
<td>77</td>
<td>Normal</td>
<td>U95v1</td>
<td>B</td>
<td>12</td>
<td>98.30</td>
</tr>
<tr>
<td>100</td>
<td>Benign</td>
<td>U133</td>
<td>B</td>
<td>13</td>
<td>98.70</td>
</tr>
<tr>
<td>139</td>
<td>PTC</td>
<td>U133</td>
<td>A</td>
<td>14</td>
<td>98.91</td>
</tr>
<tr>
<td>90</td>
<td>PTC</td>
<td>U95v1</td>
<td>B</td>
<td>15</td>
<td>99.09</td>
</tr>
<tr>
<td>42</td>
<td>CTN</td>
<td>U95v2</td>
<td>B</td>
<td>16</td>
<td>99.22</td>
</tr>
<tr>
<td>3</td>
<td>AFTN</td>
<td>U95v2</td>
<td>A</td>
<td>17</td>
<td>99.28</td>
</tr>
<tr>
<td>37</td>
<td>CTN</td>
<td>U95v2</td>
<td>A</td>
<td>18</td>
<td>99.36</td>
</tr>
<tr>
<td>147</td>
<td>PTC</td>
<td>U133</td>
<td>A</td>
<td>19</td>
<td>99.38</td>
</tr>
<tr>
<td>40</td>
<td>CTN</td>
<td>U95v2</td>
<td>B</td>
<td>20</td>
<td>99.41</td>
</tr>
</tbody>
</table>

70 samples were classified with an excellent accuracy between 99 and 100%, and for further 64 cases no misclassification occurred during the bootstrapping process.

Table 3

Comparison of results obtained by different class prediction methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound covariate predictor</td>
<td>89</td>
<td>85</td>
<td>88</td>
<td>77</td>
<td>93</td>
</tr>
<tr>
<td>Nearest centroid</td>
<td>90</td>
<td>86</td>
<td>89</td>
<td>79</td>
<td>93</td>
</tr>
<tr>
<td>Linear diagonal discriminant analysis</td>
<td>92</td>
<td>87</td>
<td>92</td>
<td>83</td>
<td>94</td>
</tr>
<tr>
<td>One-nearest neighbor</td>
<td>98</td>
<td>94</td>
<td>99</td>
<td>98</td>
<td>97</td>
</tr>
<tr>
<td>Three-nearest neighbors</td>
<td>98</td>
<td>93</td>
<td>100</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>Support vector machines</td>
<td>99</td>
<td>95</td>
<td>99</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

PPV, positive predictive value; NPV, negative predictive value.

www.endocrinology-journals.org
(SVM), and confirmed the best performance of SVM-based methods to analyze these data (Table 3).

Discussion

Transcripts important for discriminating PTC from benign and normal thyroid samples

In the study, we performed an advanced optimization of putative PTC markers using a large group of benign thyroid lesions and normal thyroid tissues and proposed a list of 43 transcripts, selected by their most frequent appearance in the classifiers. An additional proof of their efficacy was obtained by hierarchical clustering (all samples clustered correctly, data shown in the web appendix to this article, www.genomika.pl/thyroidcancer). Forty-one of them (95.3%) could be attributed to 39 known genes, 32 well-defined ones, and 7 of unknown or not well-defined function. There were 12 genes which had never before been related to the thyroid gland nor mentioned in genomic studies of thyroid cancer, while 29 genes (74%) were identified in previous thyroid microarray studies. However, only ten of them were discussed in the original papers for their putative role in thyroid carcinoma. Within the list of the well-known genes which received high scores by BBFR, one should mention gene encoding FN1, met proto-oncogene (MET; both scored 4.4 × 104), dipeptidylpeptidase 4 (DPP4), adenosine A1 receptor (ADORA1), keratin 19, and B-cell CLL (BCL2) genes (Huang et al. 2001, Wasenius et al. 2003, Baris et al. 2004, Cerutti et al. 2004, Finley et al. 2004a, Wreesmann et al. 2004, Giordano et al. 2005), all up-regulated with the exception of BCL2. Their inclusion in our classifier positively validates the applied criteria. All these genes except ADORA1 were previously found by single gene studies (see Table 1) and later confirmed by microarray approaches. Moreover, in the recent meta-analysis of thyroid cancer gene expression profile, MET and FN1 were included into top 12 candidates for consistent gene expression markers (Griffith et al. 2006). Similarly, thyroid-specific (down-regulated) genes, deiodinase, iodothyronine, type I and thyroid peroxidase, were widely recognized previously for their diagnostic significance both in microarray-based (Eszlenger et al. 2001, Huang et al. 2001, Baris et al. 2004, Cerutti et al. 2004, Finley et al. 2004a, Wreesmann et al. 2004) and single gene studies (Arturi et al. 1997, Lazar et al. 1999, De Micco et al. 1999, Czarnocka et al. 2001, Le Fourn et al. 2004, Ambroziak et al. 2005, Arnaldi et al. 2005). Nevertheless, neither our approach nor the meta-analysis mentioned earlier indicated other thyroid-specific genes, confirming the lesser diagnostic potency of sodium iodide symporter, thyroglobulin, thyrotrophin receptor, or thyroid-specific transcription factors, shown to be down-regulated in previous single gene studies (Arturi et al. 1997, Lazar et al. 1999, Shimura et al. 2001, Scouten et al. 2004, Ambroziak et al. 2005, Wagner et al. 2005).

The top gene identified by our effort, MPPED2, which is lost in PTC, was not previously considered for its role in PTC, although it was previously listed by Aldred et al. (2004, in the context of FTC) and by Mazzanti et al. (2004). It is an ancient gene highly conserved from Caenorhabditis elegans to mammals and expressed in fetal brain. Its function is unknown.

Already the first microarray-based analysis of a PTC gene expression profile (Huang et al. 2001) indicated the dominant position of genes controlling cell–matrix adhesion and cell–cell communication. Besides, FN1 mentioned earlier, and intercellular adhesion molecule 1 (ICAM-1; Kawai et al. 1998), it seems important to mention syndecan 4 (SDC4), a transmembrane heparan sulfate proteoglycan known to bind FN1 and functioning also as CXCL12 receptor in signal transduction (Huang et al. 2001, Chevillard et al. 2004, Finley et al. 2004a). Loss of CDH16 (kidney-specific cadherin; Thomson et al. 1998) was indicated for the first time in our study, a gene closely related to cadherin E (CDH1E), which is well known to be lost in a subgroup of PTCs with negative prognostic significance (Rocha et al. 2003), while cadherin P (CDH3) is up-regulated in PTC (Jarzab et al. 2005). Other genes involved in cell adhesion and present in our list comprise ectonucleoside triphosphate diphosphohydrolase 1 (ENTPD1) (up-regulated) and less known genes such as NEL-like 2 (up-regulated) and sushi, nidogen, and EGF-like domains 1 (down-regulated), both exhibiting EGF-like repeats (Watanabe et al. 1996). The GJB3 gene (connexin 31) encodes the protein subunit of gap junctions, essential for cell–cell communication.

DPP4 (CD26), ICAM1, and ENTPD1 (CD39) may be considered as immune-related genes, although their expression is not confined to immune or endothelial cells. ICAM1 was shown to be present in thyroid cancer cells (Kawai et al. 1998). ENTPD1 (ecto-ATPase), in turn, has not been described before for the thyroid gland; its expression was shown in some other organs like salivary glands or exocrine pancreas (Kittel et al. 2004). It converts adenine nucleotides to adenosine, thus participating in the control of signal transduction. DPP4, another membrane-bound enzyme which hydrolyzes peptides engaged in paracrine and autocrine
regulation, is up-regulated in PTCs both on RNA and protein level (Huang et al. 2001, Kholova et al. 2003). The contribution of various enzymes to our list is striking: others, not described previously in the context of thyroid gland, comprise UDP-galactose epimerase (GALE) and glutaminyl-peptide cyclotransferase (QPCT), both with virtually unknown expression patterns. The latter was also indicated by the meta-analysis of Griffith et al. Among gene encoding enzymes lost in PTC are plasma glutamate carboxypeptidase, plasma glutamate carboxypeptidase (Gingras et al. 1999), not mentioned in any thyroid-related study before; carbonic anhydrase 4 (CA4), and even the well-known homogentisate oxidase (encoding HGD), not previously related to the thyroid in any context, although listed in many microarray-based reports (Table 1).

Underexpression of hemoglobin transcripts (HBA1/A2 and HBB scored at positions 2 and 25 respectively) was already discussed in our papers as a very characteristic feature of PTC gene expression profile (Jarzab et al. 2005). We believe that the down-regulation of hemoglobin gene could be associated with tumor hypoxia; HBA has also been considered a tumor suppressor since transduction of this gene in an anaplastic thyroid cancer cell line induces an anti-proliferative effect (Onda et al. 2005).

Many of the genes listed in Table 1 participate in signal transduction; among them are MET, ADOR-A1, RAB27A as well as tumor-associated calcium signal transducer 2, inositol 1,4,5-triphosphate receptor, type 1 (ITPR1), ryanodine receptor 1, all up-regulated in PTC except for ITPRI. Some enzymes mentioned above (DPP4, ENTPD1, and QPCT) contribute to synthesis or breakdown of signaling molecules. On the other hand, the list also includes many genes participating in transcription regulation, among them high-mobility group AT-hook 2, aryl hydrocarbon receptor, retinoid X receptor, γ, ID3, nuclear receptor-interacting protein 1, and RUNX1. Both of these functional classes are typical for cancer genes. We noted only one gene clearly related to apoptosis (and lost in PTC), the well-known BCL2. Interestingly enough, some immunohistochemical studies report its up-regulation in PTC (Aksoy et al. 2005).

Although the selected genes were obtained by analysis of PTC, many of them may be found also in other types of thyroid tumors (M Oczko-Wojciechowska, J Starzyński, M Jarząb, Z Wygoda, A Czarniecka, G Gala, M Kalemba, E Gubala & B Jarząb, unpublished data). This is convincingly illustrated by the overlapping results of our analysis and one of the studies which dealt with follicular thyroid tumors only (Barden et al. 2003).

**Accuracy of discriminating PTC from benign/normal thyroid tissue**

Our study is the first to define the classification accuracy for thyroid cancer by 95% CIs and one of the few dealing with the problem of diagnostic accuracy of microarray-derived classifiers (Kerr & Churchill 2001). Although the estimation of CIs by Monte Carlo analysis has not gained a general acceptance still, it is necessary to stress the very good accuracy of PTC diagnosis in our study with the lower range of the CI at 95%, obtained using a sufficiently large study group, mimicking the real clinical setting. From a clinical point of view, for a PTC classifier, an even higher accuracy is required, as the risk of diagnosing PTC in a thyroid nodule is only about 5% (Hegedus 2004).

Our results stress the importance of multi-gene approaches for the molecular diagnosis of cancer. We observed that lower limits of accuracy CIs were decreased in case of classification by gene sets with less than ten genes. The initial conclusion from these data is that any combination of more than five to ten genes increases the reliability of distinguishing between malignant and benign tissue samples. This result is similar to that obtained by Hua et al. (2005), who demonstrated on simulated and real breast cancer data that for different classifiers the number of features lower than five was usually much less effective than larger classifiers. Recent paper reports a six-gene molecular classifier, efficient for molecular diagnosis of thyroid cancer (Kebebew et al. 2006).

**Bootstrap-based multi-gene classification of PTC microarray data**

Selection of genes is an important goal of microarray studies contributing to broader understanding of the cancer transcriptome as well as yielding novel molecular cancer markers. Such studies have been successfully performed in PTC and large numbers of discriminating physiologically relevant genes were proposed (Huang et al. 2001, Wasenius et al. 2003, Aldred et al. 2004, Chevillard et al. 2004, Finley et al. 2004a,b, Wresemann et al. 2004, Baris et al. 2005, Detours et al. 2005, Giordano et al. 2005). However, in the majority of these studies, the selection of important genes was based on either fold-change or significance criteria obtained using classical statistical tests. These approaches either favor genes with large amplitudes, sometimes coming from a
minor proportion of samples, or genes with low within-group variance, thus rather stably expressed in all analyzed tumor samples. Bearing in mind, complexity of molecular changes in tumors, the widespread skepticism about a single ‘cancer marker’ as well as possible differences in histological subtypes or other features of PTC, we decided to use SVM, a routine machine-learning approach to construct classifiers based on multiple features of the analyzed objects. This method allows integrating the information carried by many genes in the gene sets. Thus, effective molecular multi-gene classifiers may be built that rely on inter-gene interactions rather than on combining single ‘best markers’. SVMs have been confirmed as an effective method of multi-gene set selection and this is supported by our comparison to other class prediction methods. Our procedure helps us to optimize the list of markers which are to be implemented to real-time quantitative PCR-supported fine needle biopsy (Lubitz & Fahey 2006).

From the diagnostic point of view, the major drawback of the SVM-based methods are the fluctuations of gene content between classifiers of different size or based on slightly different training sets. To overcome this problem, we extended the original algorithm with bootstrap iterations, as recommended (Braga-Neto & Dougherty 2004). A bootstrap iteration depends on creating a temporary learning set (bootstrap sample) by performing selection from the original set with return of samples. Then, the classification rule is derived based on a bootstrap sample and applied to the rest of the original set. Multiple selections of slightly different training sets represent the variability, which may be observed between different thyroid cancer collections, laboratories, etc. Indeed, our current data generated using the bootstrap technique show much better agreement with the results of other thyroid cancer studies (Oczko-Wojciechowska et al. submitted) than data created by leave-one-out cross-validation of the whole dataset (Jarzab et al. 2005).

Originally, in a bootstrap iteration one counts only the number of misclassifications. Since in all bootstrap iterations every step of data processing (gene selection and classifier training) has to be repeated (Simon et al. 2003), some additional knowledge can be gained. The procedure used by us enables ranking of genes which are most often present in the classifiers obtained from the different subsets of the training set (BBFR). Furthermore, it also estimates the accuracy with appropriate CIs. Moreover, it allows ranking the samples according to the frequency of misclassifications (BBOD). The use of BBFR resulted in delineation of genes, which were either novel or not recognized before for their contribution to the PTC gene expression profile, even if they were included in the large gene lists given in previous genomic studies. BBOD allowed us to reveal ‘difficult’ samples in the analyzed group. The two thyroid samples with the poorest accuracy of diagnosis were derived from patients with familial thyroid tumors, which suggest that their gene expression profiles may differ from sporadic ones. For the remaining samples, in 175 out of 180 cases (>97%) the percentage of correct diagnoses was >90%.

Recently, Zhang et al. (2006) have published a SVM-based recursive method of gene selection. This method, called R-SVM, differs from the standard RFE algorithm, used here, in modified criteria applied in elimination steps. Moreover, the final gene subset is created on the basis of any resample method used at the validation stage, which is similar to our approach presented here. Nevertheless, our bootstrap-based method allows detecting outlier samples and provides the estimation of CIs for the classification accuracy, which is much more informative than the accuracy estimator alone.

PTC and normal/benign difference versus inter-platform difference

To assure a sufficient number of tissue samples, it was necessary to combine data obtained using different generations of GeneChips, which cannot be compared by a direct approach (Eszlinger et al. 2006). The use of multi-gene classifiers allows, however, overcoming this difficulty. We showed earlier that the classifier selected using the U133 platform (Jarzab et al. 2005) performs well on U95-obtained data and has high classification accuracy (Eszlinger et al. 2006). In the present paper, we demonstrate that it is possible, after correctly matching genes from two different generation microarrays, to derive an efficient multi-gene classifier. When we included both benign and malignant samples from both platforms, the vast majority of these samples were properly classified. Using Affymetrix GeneChips, Barden et al. (2003) and Finley et al. (2004a,b) had previously reported 20 of 43 genes now confirmed by us as diagnostically relevant for PTC. This is a level of agreement rarely noted for inter-group comparisons of microarray results.

Our analysis has been performed on microarray data pre-processed by the standard MAS5 algorithm. Although many authors demonstrate the superiority of other pre-processing methods (e.g. RMA or GC-RMA; Irizarry et al. 2003), for inter-platform comparisons, the MAS5 method still seems to be a reasonable approach. In the MAS5 algorithm, each
array is processed independently and the bootstrap procedure does not have to involve this step. Use of RMA pre-processing, which has to operate on the whole dataset, would pose the question of whether this step should also be bootstrapped. Presently, this is not feasible due to huge computational demand of pre-processing for large sample sets.

Redundancy of multi-gene cancer classifiers

This is inherently linked to the huge differences in gene expression profiles of several tumors, originating from the same tissue. This was indicated for the first time by Ein-Dor et al. (2005) in breast cancer. These authors re-analyzed the data of van’t Veer et al. (2002) and showed that multiple similar classifiers may be obtained; they have comparable classification potency as van’t Veer’s original 70-gene classifier but a different gene content. Ein-Dor et al. stressed also that even slight differences in the training set composition influenced the selected genes. Our analysis demonstrates that similar redundancy is present in PTC. This fact is frequently overlooked by authors interpreting the results of gene expression profile studies that involved only a few genes or which were obtained in small groups of patients. In this paper, we propose a method of ranking genes according to their importance in multi-gene classifiers and with appropriate CIs indicating the robustness of the result.

To conclude, the primary goal of this study was to validate a novel SVM-based approach to differentiation of PTC from benign thyroid lesions. This goal was achieved with a very satisfactory degree of accuracy, over 95%. Simultaneously, we were able to rank the genes most essential for the molecular diagnosis of PTC. Although the presented list of genes can be enlarged, we believe the first 40 genes are especially suitable for further prospective studies in fine needle biopsy material and may serve to construct multi-gene classifiers with potential application in clinical setting. The comparison with other published microarray studies yields sufficient validation for the vast majority of them.

Acknowledgements

We gratefully acknowledge Aleksander Sochanik, PhD, for the thorough language revision of the manuscript. This work was partially supported by Polish Ministry of Education and Science under grant 3T11A 019 29 (K F) and 2P05A 022 30 (B J). This work was partially supported by the Deutsche Krebshilfe grant 106542 (R P and K K) and the Interdisciplinary Center for Clinical Research at the Faculty of Medicine of the University of Leipzig (projects B20, Z03). This work was partially supported within GENRISK-T project, contract number 036495 (A S, B J). Authors declare no potential conflict of interest.

References


Aldred MA, Huang Y, Liyanarachchi S, Pellegrata 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.

Ambroziak M, Pachucki J, Stachlewlska-Nasfeter E, Nauman J & Nauman A 2005 Disturbed expression of type 1 and type 2 iodothyronine deiodinase as well as tif1/nkx2-1 and pax-8 transcription factor genes in papillary thyroid cancer. Thyroid 15 1137–1146.


Arnaldi LA, Borra RC, Maciel RM & Cerutti JM 2005 Gene expression profiles reveal that DCN, DIO1, and DIO2 are underexpressed in benign and malignant thyroid tumors. Thyroid 15 210–221.


Czarnocka B, Pastuszko D, Janota-Bzowski M, Weitman AP, Watson PF, Kemp EH, McIntosh RS, Asghar MS, Jarzab B, Gubala E et al. 2001 Is there loss or qualitative changes in the expression of thyroid peroxidase protein in thyroid epithelial cancer? *British Journal of Cancer* **85** 875–880.


Chastang C 2003 Interobserver and intraobserver reproducibility in the histopathology of follicular thyroid carcinoma. Human Pathology 34 1092–1100.


Ippolito A, Vella V, La Rosa GL, Pellegriti G, Vigneri R & Belfiore A 2001 Immunostaining for Met/HGF receptor may be useful to identify malignancies in thyroid lesions classified suspicious at fine-needle aspiration biopsy. Thyroid 11 783–787.


