

Molecular signatures of thyroid follicular neoplasia

Rehannah Borup¹, Maria Rossing¹, Ricardo Henao⁷, Yohei Yamamoto^{1,8}, Annelise Krogdahl⁵, Christian Godballe⁶, Ole Winther^{7,9}, Katalin Kiss³, Lise Christensen⁴, Estrid Høgdall³, Finn Bennedbaek² and Finn Cilius Nielsen¹

Departments of ¹Clinical Biochemistry, Rigshospitalet, ²Endocrinology and Metabolism, ³Pathology, Herlev Hospital and ⁴Pathology, Bispebjerg Hospital, University of Copenhagen, 2100 Copenhagen, Denmark

Departments of ⁵Pathology and ⁶ENT Head and Neck Surgery, Odense University Hospital, 2000 Odense, Denmark

⁷Bioinformatics Centre, University of Copenhagen, 5000 Copenhagen, Denmark

⁸Department of Molecular Pathology and Tumor Pathology, Akita University Graduate School of Medicine, Akita, 10-8502 Japan

⁹Danish Technical University, Informatics, 2800 Lyngby, Denmark

(Correspondence should be addressed to F C Nielsen; Email: fcn@rh.dk)

Abstract

The molecular pathways leading to thyroid follicular neoplasia are incompletely understood, and the diagnosis of follicular tumors is a clinical challenge. To provide leads to the pathogenesis and diagnosis of the tumors, we examined the global transcriptome signatures of follicular thyroid carcinoma (FC) and normofollicular adenoma (FA) as well as fetal/microFA (fetal adenoma). Carcinomas were strongly enriched in transcripts encoding proteins involved in DNA replication and mitosis corresponding to increased number of proliferating cells and depleted number of transcripts encoding factors involved in growth arrest and apoptosis. In the latter group, the combined loss of transcripts encoding the nuclear orphan receptors NR4A1 and NR4A3, which were recently shown to play a causal role in hematopoietic neoplasia, was noteworthy. The analysis of differentially expressed transcripts provided a mechanism for cancer progression, which is why we exploited the results in order to generate a molecular classifier that could identify 95% of all carcinomas. Validation employing public domain and cross-platform data demonstrated that the signature was robust and could diagnose follicular nodules originating from different geographical locations and platforms with similar accuracy. We came to the conclusion that down-regulation of factors involved in growth arrest and apoptosis may represent a decisive step in the pathogenesis of FC. Moreover, the described molecular pathways provide an accurate and robust genetic signature for the diagnosis of FA and FC.

Endocrine-Related Cancer (2010) 17 691–708

Introduction

Thyroid nodules are a common clinical finding (Hegedus *et al.* 2003, Hegedus 2004). In Western Europe, ~6% of all women have palpable nodules, and the number of silent thyroid nodules is several folds higher. In addition to alleviating local compressive symptoms or thyroid hyperfunction, the major clinical challenge is to exclude the possibility of malignancy. Only about 5% of cold thyroid nodules become malignant, and it is therefore important that the diagnostic procedures exhibit a high sensitivity and

specificity (Utiger 2005, Ruggeri *et al.* 2008). Follicular thyroid carcinomas (FCs) comprise about 15% of all malignant nodules and they may be overlooked, since the diagnosis mainly relies on the exclusion of capsular and/or vascular invasion. Moreover, it is difficult to distinguish benign follicular adenoma (FA) from carcinoma.

The road to follicular neoplasia is not completely understood. In contrast to the well-defined *RET* and *BRAF* mutations found in medullary and papillary thyroid cancers, follicular tumors do not exhibit

consistent mutations (Fagin & Mitsiades 2008), although individuals exhibiting variations in *FOXE1* (*TTF2*) and *NKX2-1* (*TTF1*) have recently been reported to have increased risk of developing follicular carcinoma (FC; Gudmundsson *et al.* 2009). *RAS* is mutated in up to half of the tumors and a recurrent PAX8–PPAR γ translocation has been identified in 26–56% of the cancers, but also in a number of adenomas (Nikiforova *et al.* 2003, Delellis 2006, Fagin & Mitsiades 2008). The microFAs or fetal adenomas (FEAs) represent a subtype of follicular nodules exhibiting a high degree of aneuploidy, which renders these tumors more likely to become malignant (Castro *et al.* 2001). A number of studies have successfully exploited global expression profiling to identify molecular markers or signatures of thyroid neoplasia (Barden *et al.* 2003, Finley *et al.* 2004, Mazzanti *et al.* 2004, Lubitz *et al.* 2005, Weber *et al.* 2005, Fryknas *et al.* 2006, Griffith *et al.* 2006, Fujarewicz *et al.* 2007, Prasad *et al.* 2008, Hinsch *et al.* 2009). Among others, cyclin D2 (*CCND2*), protein convertase 2 (*PCSK2*), and prostate differentiation factor (*PLAB*) have been reported to differentiate between FC and FA (Weber *et al.* 2005). With few exceptions (Weber *et al.* 2005, Fryknas *et al.* 2006, Fujarewicz *et al.* 2007, Prasad *et al.* 2008), most studies have relied on unsupervised methods, such as hierarchical clustering of a list of differentially expressed genes, which are not appropriate methods to provide signatures that are robust across geographical locations or platforms that may affect the accuracy of the predictions made by a particular classifier (Simon 2006).

To explore if it was possible to identify molecular pathways implicated in follicular neoplasia and further improve diagnosis, we performed a global expression profiling of follicular nodules and applied supervised learning by support vector machines (SVM) to generate diagnostic signatures based on the major cancer-specific changes. We report that thyroid FCs are characterized by transcripts encoding factors involved in DNA replication and mitosis and loss of growth arrest, and proapoptotic factors such as NR4A1 and NR4A3, FOSB and JUN, which previously have been causally associated to stem cell proliferation and defective extrinsic apoptotic signaling (Mullican *et al.* 2007). Based on the analysis of differentially expressed transcripts, we generated a molecular classifier that could identify carcinomas with a high accuracy. Validation employing public domain and cross-platform data demonstrated that the signature was robust and worked equally well on follicular nodules originating from different geographical locations and platforms.

Materials and methods

Collection of tumor samples

Sixty-nine tumor samples were collected from patients who underwent thyroidectomy at the Copenhagen University Hospital, Rigshospitalet, and the Odense University Hospital from 1989 to 2008. The sampling at the Copenhagen University Hospital was part of an ongoing quality assurance programme, and all patients had been informed about and agreed to the sampling. Handling and usage of all the samples obtained from the Odense University Hospital were approved by the ethics committee of the County of Funen. After surgical excision, the tumor samples were snap frozen in liquid nitrogen and stored at -80°C . The tumors included 22 benign FAs (7 from Copenhagen and 15 from Odense), 18 FCs (3 from Copenhagen and 15 from Odense), 12 microFAs (all obtained from Odense), 4 anaplastic carcinomas (AC; from Copenhagen), 2 papillary carcinomas (PC; from Copenhagen), and 9 nodular goiters (NG; from Copenhagen). Twenty-three samples were obtained from the expression profile repository, Array Express (<http://www.ebi.ac.uk/arrayexpress/>); these amounted to 14 PC and 9 normal thyroid samples.

Microarray analysis

Total RNA was isolated using TRIzol reagent (Invitrogen), and purified over RNeasy columns (Qiagen). The quantity and integrity of the extracted RNA were determined using Nanodrop (Nanodrop Technologies, Wilmington, DE, USA) and the Bioanalyzer LabChips (Agilent Technologies, Santa Clara, CA, USA) respectively. Samples were labeled according to the manufacturer's guidelines. In short, 2 μg of total RNA were transcribed into cDNA using an oligo-dT primer containing a T7 RNA polymerase promoter. cDNA was used as a template in the *in vitro* transcription reaction driven by the T7 promoter, under which biotin-labeled oligonucleotides were incorporated into the synthesized cRNAs. The labeled cRNAs were hybridized to the HG-U133plus2 GeneChip array (Affymetrix, Santa Clara, CA, USA), which query close to 48 000 well-substantiated genes by ~ 56 000 probe sets. The arrays were washed and stained with phycoerythrin-conjugated streptavidin (SAPE) using the Affymetrix Fluidics Station 450, and scanned in the Affymetrix GeneArray 3000 7G scanner to generate fluorescent images as described in the Affymetrix GeneChip protocol.

Microarray data analysis

Cel files were imported into the statistical software package R v. 2.7.2 using BioConductor v. 2.8 (Gentleman *et al.* 2004) and gcRMA modeled using quantile normalization and ‘lowess’ summarization (Bolstad *et al.* 2003). The modeled log-intensity of 56 400 probe sets was used for high-level analysis for selecting differentially expressed genes and formulating the classifier. Model construction and optimization were written in R (v. 2.7.2). Various functions from the BioConductor packages, Biobase, affy, multtest, MASS, class, e1071, mda, grid and RocR were applied in the code (Gentleman *et al.* 2004). The microarray data were submitted to the gene expression repository at Array Express (<http://www.ebi.ac.uk/arrayexpress/>) with accession number E-MEXP-2442.

Differential expression analysis

Genes were defined as being differentially expressed in a class comparison analysis if they were selected in the univariate two-sample *t*-test or *F*-test with equal variance as described below. Statistical hypothesis testing was performed using the multtest package in Bioconductor v. 2.7.2. Equal variance two-sample *t*-statistic or multi-sample *F*-statistic for tests of equality of population means was performed on each gene. Control of type I error rate was performed by computing adjusted *P* values for simple multiple testing procedures from a vector of raw (unadjusted) *P* values. The procedures include the Bonferroni, Holm, Hochberg, and Sidak procedures for strong control of the family-wise type I error rate (FWER), and the Benjamini & Hochberg and Benjamini & Yekutieli procedures for (strong) control of the false discovery rate (FDR). The FWER methods provide a very conservative control of error rates, and hence the resulting number of rejections (discoveries of differentially expressed genes) is in practice close to zero. In comparison, the FDR methods give more power in the analysis, and since we wished to make as many discoveries as possible to enhance the chance of defining the molecular change in the samples and a small proportion of errors will not change the overall result, we chose to apply the Benjamini & Hochberg (1995) FDR analysis. A probe set is defined as being differentially expressed if the adjusted *P* value is below 0.05 applying Benjamini & Hochberg (1995) controlling procedure, and has a fold change larger than 1.5 and a difference of means larger than 100 (real unlogged values) between (mutual) classes of samples (FC versus FA, FEA versus FA and FC versus FEA).

The differentially expressed genes were grouped according to their functional categories in cell cycle, cytoskeleton and extracellular matrix (ECM), DNA binding and transcription, metabolism, RNA processing and translation, and secretion and signaling.

Formulation of classifiers

The diagnostic classifiers were developed in R v. 2.7. For all classification problems, the training of the classifiers inside the leave-one-out (LOO) loop consists of two steps: a univariate probe ranking and selection step, and a fitting step in which a SVM is fit on the sample division using the selected probes as covariates. All models were optimized by a grid search of *P* value cut-offs, and the cut-off resulting in a gene signature of optimal performance was used in the final model. In classifiers 1 and 3, the gene signature was selected with Student’s *t*-test with *P* values below 1e-4 and 1e-6, and in classifier 2, it was selected with an *F*-test with a *P* value below 5e-6. Model fitting was done by training an SVM with a Gaussian kernel (Vapnik 1998). The parameters of the classifier (cost and gamma) were selected by grid search using different combinations of values and cross-validation within LOO loops to ensure that the estimation of the classifier parameters was unbiased. The grid search optimization showed that a spectrum of values of cost and gamma provided similar performance, and the median values were used in the algorithm. For each cross-validation loop, the percentage of genes selected is reported, and we applied this measure to enhance the robustness of the model in classifier 1 by using only the probe sets that have 100% cross-validation support in the final classifier.

The trained SVM model was turned into a probabilistic classifier giving an estimate of the probability of the predicted class label, i.e. quantification of the prediction uncertainty, or the predictive probability of a sample being one or the other type using logit estimates (Platt 1999). The predictive probability is graphed as the function $p(\text{FA})$ by plotting the predictive probability on the *y*-axis and samples on the *x*-axis (classifiers 1 and 3). In the three-class problem (classifier 2), a ternary plot was produced combining the probabilities for a sample being either one of the three classes: FA, FC or FEA.

Estimation of misclassification rate

The misclassification rate for each classifier was evaluated using LOO cross-validation (LOOCV) during which we applied *t*-tests (classifiers 1 and 3)

or the *F*-test (classifier 2) for feature selection of probe sets to include in each model. The correct classification rate was calculated as the percentage of correctly classified samples of the total number of samples examined. Furthermore, the performance of each classifier during LOOCV is described by the following parameters: sensitivity, which is the probability for a class A sample to be correctly predicted as class A; specificity, which is the probability for a non-class A sample to be correctly predicted as non-A; positive predictive value (PPV) is the probability that a sample predicted as class A actually belongs to class A, and negative predictive value (NPV), the probability that a sample predicted as non-class A actually does not belong to class A (Simon et al. 2007).

Statistical significance of the error rate

A permutation test was performed in order to determine if the cross-validated misclassification rate was lower than expected by chance (Tusher et al. 2001, Simon et al. 2007). In 1000 random permutations of the class label, the entire cross-validation was repeated for classifying the random classes of samples. The proportion of the 1000 random permutations that gives a smaller or similar cross-validation misclassification rate as that obtained with the real data determines the permutation *P* value. The statistical significance of the error rate was determined for the SVM classifier in the two-class cases, and it was determined using the 3-nearest neighbors (3-NN) method for the three-class case due to computational limitations.

Comparison of different classification models

The performance of the SVM classifier was compared with that of other classifiers based on different algorithms, these being diagonal linear discriminant analysis (DLDA), compound covariate predictor (CCP), and 1-NN and 3-NN using BRB-Array tools (Simon et al. 2007). Sensitivity, specificity, PPV, and NPV were calculated for all algorithms.

External validation data sets

To validate whether classifier 1 could distinguish between FA and FC outside the training setting, we included two external data sets (Weber et al. 2005, Hinsch et al. 2009), which included expression profiles of 24 samples (12 FA and 12 FC) analyzed with Affymetrix HG-U133A arrays and 12 samples (4 FA and 8 FC) analyzed with ABI Human Genome Survey Microarray version 2 from Applied Biosystem

(Carlsbad, CA, USA) respectively. The raw data files (cel files) from the Weber et al. (2005) study were normalized and summarized, and expression values were calculated using invariant set normalization and the PM-MM model implemented in the dChip software (Li & Wong 2001) as described. We identified the ~22,200 probe sets that were shared between the HG-U133A and the HG-U133_plus2 arrays. The normalized data from the Hinsch et al. study were downloaded from the expression profile warehouse at the NCBI Gene expression Omnibus (Barrett et al. 2009) GEO accession number 'GEO15045', and the unique ids were coupled with gene name and gene symbol. The gene symbol was used to find the overlap of the genes between the ABI Human Genome Survey Microarray version 2 array and the other arrays.

Validation using external data sets

We tested the performance of the gene signature developed in the FA versus FC classifier 1 based on Affymetrix HG-U133_plus2 array on the external data sets. Of the 76 probe sets (66 unique genes) in the classifier 1 signature, 45 unique genes were represented on the older HG-U133A array. In order to classify Weber's data, we imported the 24 samples along with our 40 samples using only the ~22,200 probe sets that were shared between the HG-U133A and the HG-U133_plus2 arrays. The 24 samples were used (Weber et al. 2005) as an independent test set, and we applied the classifier using the reduced signature of 45 genes. The same procedure was used for the 12-sample validation data (Hinsch et al. 2009) employing the samples as an independent test set implementing the gene signature from classifier 1 using the 53 genes of the 66 genes in our classifier that were shared between the Affymetrix and the ABI platforms.

Test of gene signatures across platforms

In order to determine the overall performance of our model using the 76-probe set signature compared to other published signatures and classifiers, we tested the two signatures reported by Weber et al. (3 and 80 genes) and one signature reported by Hinsch et al. (21 genes) using our data as the independent test set.

In addition, we tested signatures from other recent publications (Griffith et al. 2006, Foukakis et al. 2007, Prasad et al. 2008). These signatures included a 5-gene signature (Foukakis et al. 2007), and a 12- and 32-gene signature of benign versus malignant thyroid tumors from a meta-study (Griffith et al. 2006) and a 75-gene signature to discriminate benign thyroid nodules from

malignant thyroid nodules (Prasad *et al.* 2008). In this way, a total of 8-gene signatures were tested to assess their cross-platform classification accuracy. All gene signatures were used in classifiers comparing the performance of DLDA, 1-NN, 3-NN and SVM respectively as implemented in the BRB-Array tools (Simon *et al.* 2007).

Immunohistochemistry

Resected tumors from thyroid glands were fixed by immersion with formalin. Paraffin sections were cut to a thickness of 4 μ m. The sections were placed in Target Retrieval Solution (DAKO, Glostrup, Denmark), and microwaved three times for 3 min to improve staining by antigen unmasking. After washing and quenching of endogenous peroxidase, the sections were blocked and incubated for 1 h at room temperature with antibodies against human Ki67 (Abcam, Cambridge, MA, USA), TOP2A (DAKO), NR4A1 (Lifespan Biosciences, Seattle, WA, USA), and NR4A3 (MBL, Woburn, MA, USA). The labeling was visualized with peroxidase-labeled polymer conjugated to goat anti-rabbit (or anti-mouse) immunoglobulins (DAKO), followed by incubation with diaminobenzidine and counterstaining with hematoxylin.

Results

Clinical representation of patients and tumors in the study

We generated global expression profiles of 69 thyroid samples comprising 2 normal thyroid (NT), 9 NG, 2 PC, and 4 AC samples as well as 52 follicular neoplasia, 22 FA, 12 FEA, and 18 FC samples, including 10 widely and 7 minimally invasive carcinoma samples as well as 1 trabecular cancer sample (Table 1). Furthermore, we collected 23 samples (14 PC and 9 NT) from external sites (E-GEOD-6004 and E-GEOD-7307) submitted to the expression profile repository at Array Express (<http://www.ebi.ac.uk/arrayexpress/>). The 69 samples were diagnosed by a pathologist to assign histopathological diagnosis (according to the WHO classification). Among the follicular neoplasia patients, women constituted the majority, 35 (67.3%), whereas only 17 (32.7%) of the patients were men. The median age of patients with FC was 65 years, which was higher than the median age of patients diagnosed with FEA (50 years) and FA (54 years) respectively. The median size of the nodules was 5.4 cm in diameter (range 2–10 cm)

Table 1 Clinical representation of the data included in the training set. A total of 52 samples of follicular neoplasia were analyzed including 22 follicular adenoma (FA), 12 fetal adenoma (FEA), and 18 follicular carcinoma (FC) samples (6 minimally invasive (FC-M), 11 widely invasive (FC-W), and 1 trabecular (FC-T)). The median age of the patients with FC was 65 years compared with 50 years for fetal carcinoma patients and 54 years for FA patients, with median size of nodules being 5.4 cm (2;10) for FC patients compared with 3.8 (2;8) cm for FEA nodule patients and 4.1 (2;11) for FA nodule patients

Diagnosis	Age (years)	Sex (M/F)	Nodule size (cm; max diameter)	Relapse (Y/N)	Years from diagnosis
FC-M	78	F	4	N	1
FC-W	74	M	9	N	2
FC-W	49	M	7	N	1
FC-W	61	F	9.5	N	1
FC-W	58	F	2.5	Y	13
FC-W	63	F	5	Y	5
FC-M	67	M	8	Y	10
FC-M	52	F	5	Y	20
FC-W	61	M	6	Y	5
FC-W	56	M	6	Y	3
FC-W	71	F	5	Y	1
FC-W	88	F	2	Y	3
FC-M	24	F	2.5	N	1
FC-W	71	F	2.5	Y	7
FC-M	76	M	4.5	N	4
FC-M	63	F	2.0	N	18
FC-W	90	M	6.0	Y ^a	3
FC-T	77	F	10	Y ^a	1 (deceased)
FEA (<i>n</i> =12) ^b	50 (30;69)	6M/6F	3.8 (2;8)	–	–
FA (<i>n</i> =22) ^b	54 (31;77)	4M/18F	4.1 (2;11)	–	–

^aMetastasis found at the time of operation.

^bData given as the median values, and the range of values is given in the parentheses.

in FC patients compared with 3.8 cm (range 2–8 cm) in the FEA nodule patients and 4.1 cm (range 2–11) in FA patients (Table 1).

In order to obtain an overview of the molecular differences and similarities of the thyroid nodules, we compared the gene expression of all samples with a principal component analysis (PCA) using all transcripts (Fig. 1). The three-dimensional plot captures 36% of the variance measured across the 92 samples. The analysis showed that AC is clearly distinguishable from the other thyroid nodules, except for a few of the FCs. NG is a specific entity and is easily distinguished from the remaining samples, although it is located next to normal thyroid and FA samples. The FEAs

are clustered together, but share similarity with FAs and FCs as expected. The FCs are the most heterogeneous group ranging from being in proximity to FAs and normal thyroid to the ACs. Taken together, the PCA based on the global expression profiles is in agreement with the consensus, stating that AC may be safely diagnosed, whereas the major clinical challenge is to distinguish FC from adenoma. Moreover, the PCA indicates that FEAs are likely to represent a separate biological entity – distinguishable from FA and FC, although they have many shared features.

Differentially expressed transcripts

To achieve knowledge about the molecular perturbations leading to follicular neoplasia, we compared the gene expression patterns of FC, adenoma and microfollicular nodules. Differentially expressed genes were identified by class comparison analysis as described in Materials and methods. The comparative analyses of FC versus FA, FEA versus FA, and FC versus FEA resulted in the identification of 117, 240, and 512 differentially expressed probe sets respectively (Supplementary Table 1, see section on supplementary data given at the end of this article). Forty-five probe sets were overlapping between FC versus FA and FC versus FEA; however, there was no overlap among the differentially expressed transcripts between FC and FA compared to FEA versus FA. Taken together, we inferred that the transcripts, which are changed in FC, are different from the transcripts altered in FEA, indicating that the differences between FA and FC are likely to be cancer related. Moreover, the high number of selective FEA transcripts emphasizes the unique biological properties of this histopathological entity.

To provide an overview of the molecular function of the differentially expressed transcripts of the FC group compared to FA group, we categorized the encoded proteins as either DNA- or RNA-binding factors, extracellular matrix and adhesion and cytoskeletal components, and proteins involved in metabolism or cell signaling, protein secretion, cell cycle regulation and apoptosis as well as in DNA repair. Moreover, a few proteins or transcripts of unknown function were grouped together. The relative distribution of the functional groups is shown in Fig. 2. Compared with the entire collection of probe sets on the chip (Jonson *et al.* 2007), where about 10% of all encoded proteins may be connected to cell cycle control and apoptosis, we noted that 21% of the differentially expressed mRNAs in FCs encoded proteins categorized under

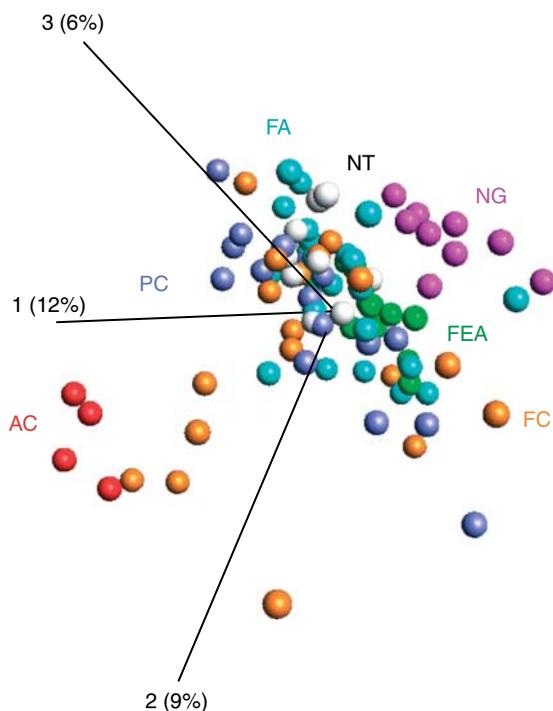


Figure 1 Results obtained from the PCA done on 92 different thyroid samples based on the entire transcriptome (all genes on the array). Each dot represents a sample that is colored according to the histological grouping. Follicular adenoma (FA), light blue; follicular carcinoma (FC), orange; fetal adenoma (FEA), green; anaplastic carcinoma (AC), red; papillary carcinoma (PC), dark blue; nodular goiter (NG), pink; normal thyroid (NT), white. The PCA plot of the first three principal components captures 27% of the variance across the samples. The percentage of the total variance described by each of the three principal components is given in parentheses at each axis. Both the AC group and the NG are well separated from the remaining samples. FA, FC, and FEA are difficult to distinguish, although FEAs are mainly localized in the lower part of the cluster. The FC nodules are very heterogeneous, showing similarity with PC, FA and FEA. The PCA and visualization were performed using the QluCore Gene Expression Explorer 1.1 (www.qlucore.com).

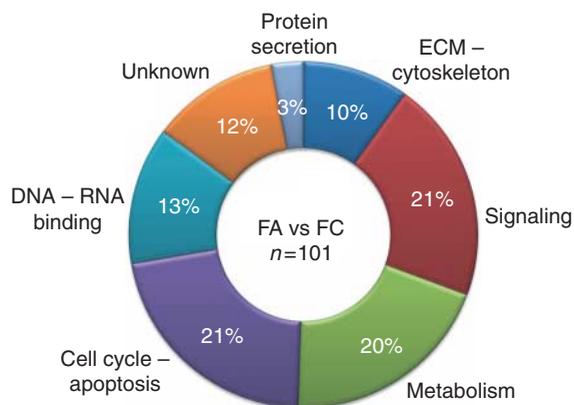


Figure 2 Functional categories of the genes that were differentially expressed between follicular adenoma and carcinoma. A total of 117 probe sets were selected as being differentially expressed with the Benjamini–Hochberg corrected *P* values below 0.05, absolute fold changes above 1.5, and an absolute change in the expression levels of more than 100. The probe sets were pruned to remove redundant probe set, which resulted in a list of 101 unique regulated genes. The putative role of every differentially expressed factor was derived from the comprehensive cDNA-supported gene and transcripts annotation Ace View (Genome Biology 2006) (www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html), and substantiated by a search in PubMed and Google. The genes were grouped according to the functional categories: DNA–RNA binding, cell cycle and apoptosis, metabolism, signaling, ECM and cytoskeleton, protein secretion, and unknown function. The percentage of genes in each category is shown.

this group. Moreover, transcripts encoding factors involved in protein secretion or DNA repair, which normally represent more than 10%, only constituted 3% of the mRNAs, and the number of mRNAs encoding DNA- and RNA-binding proteins was 1.5 times lower than their representation in the total transcriptome (19%). Transcripts encoding proteins involved in signaling and metabolism were represented similarly to the entire transcriptome.

Since cell cycle control and apoptosis are major cancer-related pathways, we examined these pathways in more detail. All redundant probe sets were excluded, and this left us with 101 unique differentially expressed factors. The function of every differentially expressed factor was derived from the comprehensive cDNA-supported gene and transcript annotation Ace View (Thierry-Mieg & Thierry-Mieg 2006) (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html>), and was substantiated by a search in PubMed and Google. This revealed a striking enrichment of transcripts encoding proteins involved in DNA replication and mitosis as well as in apoptosis among the up- and down-regulated mRNAs respectively (Table 2). In addition to the proteins directly involved

in cytokinesis, the increased level of RRM2, which previously has been implicated in carcinogenesis (Boukovinas *et al.* 2008, Souglakos *et al.* 2008), and TOP2A, which may represent a possible treatment target, was notable (Pritchard *et al.* 2008, O'Malley *et al.* 2009). Transcripts encoding the nuclear orphan receptors NR4A1 and NR4A3 were heavily down-regulated together with *JUN* and *FOSB* and the two transcripts encoding growth inhibitory factors ERG2 and SDPR. Moreover, the down-regulation of the transcripts encoding the mitochondrial potassium voltage-gated channel KCNAB1 and the solute carrier organic anion transporter family 2A1 factors, both of which are implicated in Fas-mediated apoptosis, was observed.

We corroborated the correlation of the mitotic transcripts to cell division by Ki67 staining of a series of histological sections of FA and FC, and confirmed the up- and down-regulation of TOP2A, NR4A1 and NR4A3 respectively (Fig. 3A). Nuclear Ki67 staining was mainly observed in the malignant epithelial follicular cells. A few scattered surrounding mesenchymal cells also stained positive in both adenoma and carcinoma. In agreement with the microarray data shown in Fig. 3B, FC exhibits a wide range of mitotic activities, but in general, adenomas had fewer Ki67-positive cells. A similar pattern was observed for the TOP2A, which was significantly elevated in the FC. Also in agreement with the microarray results, nuclear and cytoplasmic staining of NR4A1 and NR4A3 was reduced in carcinoma. Compared with the variable number of mitotic cells, the absence of NR4A1 and NR4A3 was observed in all carcinomas (Fig. 3A). This led us to generate a hierarchical cluster of the transcripts in order to examine if this was a general pattern during transition from adenoma to carcinoma. As shown in Fig. 3B, transcripts encoding apoptotic factors are consistently lost in carcinoma, whereas mitotic factors provide a gradient of increased expression among the malignant nodules. To determine whether the increased expression of mitotic genes was correlated to tumor size, we clustered the same transcripts again, but this time, the samples were ranked from small to large tumor size within FA and FC samples respectively. The cluster shows that there is no clear correlation between large tumors and the increased expression of mitotic gene (Supplementary Figure 2, see section on supplementary data given at the end of this article).

To investigate if the loss of apoptotic factors was a primary sign of malignancy, we compared the expression of the factors in normal thyroid gland with that of the factors in FA and FC as well as in AC

Table 2 List of gene products related to cell cycle control, specifically S-phase and mitosis, and apoptosis, which were found to be highly enriched within the genes that were regulated between follicular adenoma and follicular carcinoma (FC). The genes in the two categories are shown along with fold change (FC) and *P* value. Positive and negative fold changes represent genes that are up- and down-regulated in FC respectively

	FC	<i>P</i> value
<i>S-phase and mitosis</i>		
ANLN: anillin	26	0.013
ARPC5L: actin-related protein 2/3 complex, subunit 5-like	2	0.025
ASPM: asp (abnormal spindle) homolog	26	0.034
BUB1B: BUB1 budding uninhibited by benzimidazoles 1	9	0.029
CBX3: chromobox homolog 3	2	0.020
CCNB2: cyclin B2	14	0.017
CDCA5: cell division cycle associated 5	12	0.038
CENPF: centromere protein F	18	0.027
CEP55: centrosomal protein 55 kDa	16	0.038
CKS2: CDC28 protein kinase regulatory subunit 2	4	0.038
CTD: carboxy-terminal domain, RNA polymerase II	2	0.021
H2A histone family, member Y	2	0.017
KIF4A: kinesin family member 4A	18	0.037
MELK: maternal embryonic leucine zipper kinase	16	0.026
NEK2: NIMA (never in mitosis gene a)-related kinase 2	15	0.031
NUSAP1: nucleolar and spindle associated protein 1	10	0.020
PBK: PDZ-binding kinase	36	0.038
PRC1: protein regulator of cytokinesis 1	10	0.043
RCC2: regulator of chromosome condensation 2	2	0.020
RRM2: ribonucleotide reductase M2 polypeptide	20	0.005
SAC3D1: SAC3 domain containing 1	3	0.043
TMPO: thymopoietin	4	0.046
TOP2A: topoisomerase (DNA) II alpha 170 kDa	18	0.002
TPX2, microtubule-associated, homolog	23	0.031
UBE2C: ubiquitin-conjugating enzyme E2C	16	0.030
<i>Apoptosis and growth arrest</i>		
AGTR1: angiotensin II receptor	-7	0.046
CCDC85A: coiled-coil domain containing 85A	-2	0.018
CDH16: cadherin 16	-4	0.041
CITED2: Cbp/p300-interacting transactivator	-3	0.008
CTGF: connective tissue growth factor	-6	0.011
CYR61: cysteine-rich, angiogenic inducer, 61	-4	0.038
DLC1: deleted in liver cancer 1	-2	0.042
DNASE1L3: DNase I-like 3	-7	0.005
DUSP14: dual specificity phosphatase 14	-2	0.035
EGR2: early growth response 2	-8	0.005
FOSB: FBJ murine osteosarcoma viral oncogene	-8	0.010
JUN: jun oncogene	-3	0.034
KCNAB1: potassium voltage-gated channel	-4	0.045
MAN1C1: mannosidase, alpha	-3	0.026
MATN2: matrilin 2	-3	0.043
NR4A1: nuclear receptor subfamily 4	-7	0.009
NR4A3: nuclear receptor subfamily 4, group A	-24	0.032
PLA2R1: phospholipase A2 receptor 1	-4	0.007
PTPRN2: protein tyrosine phosphatase	-2	0.027
SDPR: serum deprivation response	-5	0.010
SLC26A4: solute carrier family 26	-3	0.038
SLCO2A1: solute carrier organic anion transporter family	-2	0.032

and PC. As shown in Fig. 4, the expression of the proteins was similar in adenoma and normal thyroid tissue, whereas the activity was lost in the carcinoma. Oppositely, mitotic factors are increased in the carcinoma group in agreement with the expansion of malignant cells.

Generation of a robust classifier for follicular neoplasia

Since the analysis of differentially expressed transcripts provided a possible mechanism for cancer progression, we aimed towards exploiting the results to generate an accurate molecular classifier that could

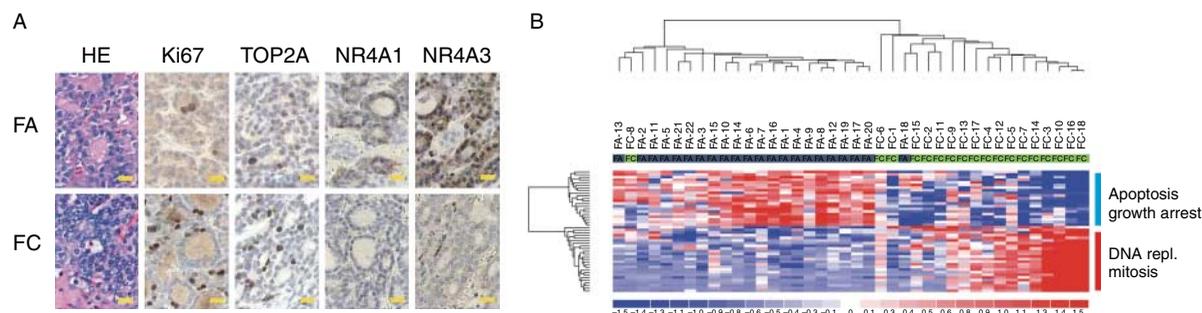


Figure 3 (A) Immunostaining of follicular adenoma (FA) and follicular carcinoma (FC). The upper panels indicate FA and lower panels indicate FC. Sections of formalin-fixed paraffin-embedded tissues were stained with hematoxylin and eosin (HE) for pathological study (the left panels). Peripheral lesions of the tumors with lower magnifications are shown in [Supplementary Figure 1](#), see section on [supplementary data](#) given at the end of this article. Immunostainings with antibodies against Ki67, TOP2A, NR4A1, and NR4A3 are shown respectively. The scale bar is 20 μ m, and the original magnification was $\times 400$. (B) Hierarchical cluster visualization of genes involved in cell cycle, mitosis, and apoptosis. The genes were clustered using rank correlation and average linkage clustering. Each row represents a gene, and each column represents a sample. Follicular adenoma samples are labeled in gray and follicular carcinoma samples are labeled in brown, where A denotes adenoma samples and C denotes carcinoma samples. The color bar indicates the degree of up- or down-regulation, with dark blue representing down-regulation of two s.d.s from the mean, and red represents up-regulation of two s.d.s or more from the mean. The heat map illustrates a loss of apoptotic and growth arrest factors during transformation to malignancy, i.e. up-regulated in FA samples and down-regulated in FC samples. Up-regulation of cell-cycle-associated transcripts in FC was gradually compared to the loss of apoptotic factors, which was observed in all samples.

differentiate between benign and malignant follicular thyroid lesions. Three classes of follicular nodules were included in the analysis. First, we focused on improving the ability to discriminate between FA and FC. This part is referred to as classifier 1. Secondly, we included the FEAs and built a classifier that could distinguish between all three types and subtypes of follicular lesions, since the differential expression analysis indicated that the FEA might be a unique histopathological entity. This is referred to as classifier 2. Furthermore, in contradiction to the differential expression analysis between FA and FEA, we observed that the FEA samples mainly classified as FA samples when introduced as independent test set in classifier 1 (FA versus FC), [Fig. 5b](#). Based on this observation,

which underscores the close relationship between FA and the subtype FEA, we decided to build a classifier that could distinguish adenomas (FAs and FEAs combined) from carcinomas (FCs).

As a result, three different classifiers were generated and evaluated: classifier 1 (FA versus FC), classifier 2 (FA versus FEA versus FC), and classifier 3 (FA and FEA merged versus FC).

All classifiers were based on the SVM algorithm developed in our R script. We compared the accuracy of the prediction by using the different classification methods implemented in BRB-Array tools. We selected 76 probe sets in order to discriminate between the two groups of 22 FA and 18 FC samples. Of the prediction algorithms that were tested, 3-NNs and

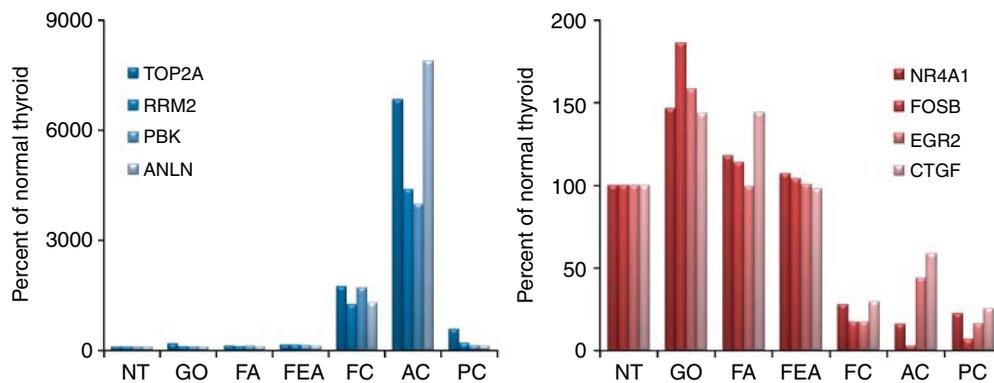
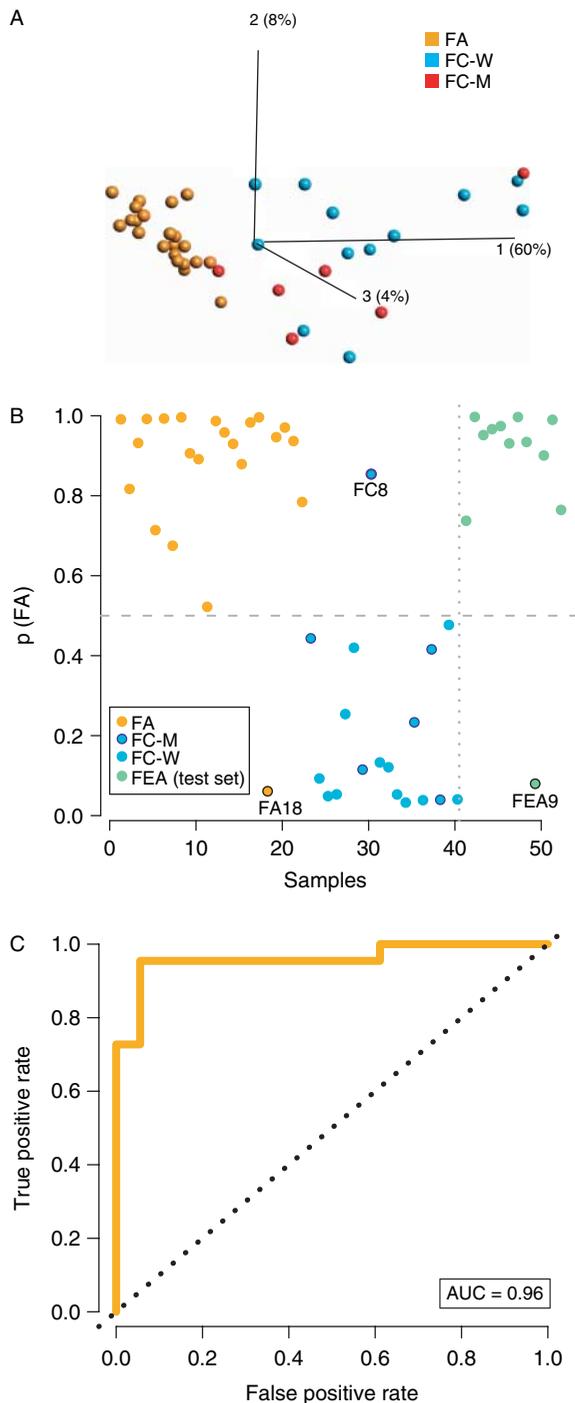


Figure 4 Bar graph of average expression level of deregulated apoptotic and mitotic factors in benign and malignant samples. Blue graphs show apoptotic factors that are up-regulated in follicular and anaplastic carcinomas, and red graphs show mitotic factors that are down-regulated in cancer. NT, normal thyroid; NG, nodular goiter; FA, follicular adenoma; FEA, fetal adenoma; FC, follicular carcinoma; AC, anaplastic carcinoma; PC, papillary carcinoma.

SVM classification had the best overall performance in this study. The cross-validation misclassification rate using the LOOCV procedure ranged from 5 to 10% with SVM outperforming all other classifiers. The performance of all six predictive algorithms is shown in Supplementary Table 2, see section on supplementary data given at the end of this article.



Essentially, all the above-described transcripts involved in cell division and apoptosis were included in the classifier. By applying cross-validation, we ensured that the data used for evaluating the predictive accuracy of the classifier were distinct from the data used to select the genes to be included in and build the classifier. The overall accuracy and performance of the classifiers during cross-validation are listed in Table 3. Classifier 1 achieved an accuracy of 95% during cross-validation. Two of the 40 samples were misclassified, which include one FA (FA18) sample and one FC (FC8) sample respectively. Classifier 2 achieved a correct classification in 85% of the samples (44 of 52 samples) during LOOCV. The eight misclassified samples were two FA, two FC, and four FEA samples. Consistent with classifier 1, the samples FA18 and FC8 were misclassified. In classifier 3, FA and FEA merged versus FC; we observed an accuracy of 90%. Fifty-two samples were used in this analysis, and five of these received a wrong label from the classifier. Again, FA18 was classified as an FC sample, as were FA5 and FEA9. The samples FC8 and FC17 were classified as FAs.

Figure 5 (A) Results obtained from the principal component analysis (PCA) of 76 probe sets defined in the FA classifier versus FC classifier. The PCA and visualization were performed using the Qlucore Gene Expression Explorer 1.1 (www.qlucore.com). Each dot represents a sample, FA, orange; FC-W, blue; FC-M, red. The three-dimensional PCA plot captures 72% of the variance described by the gene signature. The percentage of the total variance described by each of the three principal components is given in the parentheses at each axis. The FA and FC groups are well separated, except for two samples (FC8 and FA18), which are located in the borderline area between the FA and FC clusters. These samples are misclassified in the classification analysis. In the FC group, minimally invasive cancers are labeled in red, and widely invasive cancers are labeled in blue. The PCA and visualization were performed using the Qlucore Gene Expression Explorer 1.1 (www.qlucore.com). (B) Plot of the predictive probability output for classifier 1. The first 40 samples constitute the training set of 22 FA samples and 18 FC samples. Samples 41 to 52 are included as a test set, which consists of 12 FEA samples. If a sample has a predictive value above 0.5 ($p(\text{FA}) > 0.5$), it is classified as follicular adenoma, otherwise as a follicular carcinoma. Each dot represents a sample, and the color indicates the true class. FA, orange; FC, blue; FEA, green. The samples FC8 and FA18 from the training set were misclassified during cross-validation. Only one FEA sample, FEA9, was classified as a follicular carcinoma. Minimally invasive carcinomas are indicated with a blue circle. (C) Receiver operating characteristic (ROC) curve for the binary classifier built to distinguish between follicular adenoma and carcinoma nodules. The curve shows the true positive rate versus false positive rate, i.e. the tradeoff between sensitivity and specificity. The area under the curve (AUC), which captures the ability of the classifier to correctly group the patients with follicular adenoma and those with follicular carcinoma, is equal to 0.96. A perfect classifier will have an AUC of 1.0, whereas an AUC value of 0.5 indicates that the classification is random.

Table 3 Performance of the classifiers. Performance of the three classifiers during leave-one-out (LOO) training and cross-validation. Classifier 1 was trained to predict and discriminate between follicular adenoma (FA) and follicular carcinoma (FC). Classifier 2 was trained to predict and discriminate between FA, FC and fetal adenoma (FEA). Classifier 3 was trained to discriminate between the merged adenomas and the FC. The overall classification performance of 95, 85, and 90% of the three classifiers respectively is given. PPV is the positive predictive value, and NPV is the negative predictive value. Accuracy, sensitivity, specificity, PPV, and NPV are reported for each of the follicular subtypes taking part in each classification analysis during LOO training and cross-validation

Model	Accuracy per class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Statistical significance ^a	Accuracy/error rate (%)
Classifier 1 Class FA	95.5 (21/22)	95.5	94.4	95.5	94.4	$P < 0.01$	95% (2/40)
Classifier 1 Class FC	94.4 (17/18)	94.4	95.5	94.4	95.5		
Classifier 2 Class FA	90.9 (20/22)	90.9	86.7	83.3	92.9	$P < 0.03$	85% (8/52)
Classifier 2 Class FC	88.9 (16/18)	88.9	91.2	84.2	93.9		
Classifier 2 Class FEA	66.7 (8/12)	66.7	97.5	88.9	91.0		
Classifier 3 Class FA + FEA	91.2 (31/34)	91.2	88.9	93.9	84.2	$P < 0.03$	90% (5/52)
Classifier 3 Class FC	88.9 (16/18)	88.9	91.2	84.2	93.9		

^aStatistical significance of error rate (1000 permutations).

For FC, the sensitivity of classifier 1 was 0.94 and the specificity was 0.96, resulting in a NPV of 0.96 (Table 3). The high performance of the classifier is also shown by the relative receiver operating characteristic curve, where the AUC is 0.96 for classifier 1 (Supplementary Table 5C). Classifier 3, which was built to distinguish between the two merged adenoma samples and the carcinomas, had a sensitivity of 0.89 for FC and a specificity of 0.91 for FA. When translating these numbers into the positive prediction values for FC, which were 0.94 for FC in classifier 1 and 0.84 for FC in classifiers 2 and 3, a better performance of classifier 1 to predict the FCs correctly was observed. To assess whether or not the classifiers predicted more accurately than by chance (Simon *et al.* 2007), we computed misclassification rates of 1000 random permutations in order to calculate a P value of the global test that the classifier is picking up random noise in the data (Simon *et al.* 2007). The error rate estimate was statistically significant with P values for the three classifiers of 0.01, 0.03 and 0.03 respectively (Table 3).

Predictive probability and PCA of the classifiers

We used the trained SVM classifiers to derive the predictive probability of a sample being one or the other type using logit estimates (Platt 1999). A sample is classified as an FA sample if the predictive probability is above 0.5 and as an FC sample if the predictive probability is below 0.5 (Fig. 5B). It should be noted that a probability of 1.0 could be interpreted as the classifier being completely certain about its prediction, whereas a value of 0.5 reflects total uncertainty. The analysis showed that one sample (FC11) was close to the borderline ($p(\text{FA})=0.5$),

although it was labeled correctly during LOO. One sample, FA18, was very similar to the FC samples having a probability of 0.96 of being FC samples. This sample was misclassified in every analysis. Similarly, sample FC8 was classified as an FA sample, with a probability of belonging to the FA group of 0.83. When the FEA samples were included in the analysis as a test set, all but one FEA sample had a high probability of being FA samples (see samples labeled in green, Fig. 5B). Interestingly, PCA (which does not use sample labels) of the 76 probe sets that constitute the gene signature of classifier 1 was in agreement with the classifier, and showed full separation of the FA and FC classes, except for samples FA18 and FC8, which were located in the area between the FA and FC sample clusters (Fig. 5A). The PCA also demonstrated that the classifier provided a clear separation of both minimally invasive carcinoma (FC-M) and widely invasive carcinoma (FC-W) compared with adenoma. Moreover, there were no differences in the predictive probabilities between the two subtypes. We also derived the prediction uncertainty (predictive probability) for assigning class labels to the FA, FC or FEA samples in classifier 2. The results are summarized in a triangular diagram of the probabilities (Fig. 6B). Each vertex of the triangle represents a subclass. Samples plotted close to a vertex have a high probability of belonging to this particular class, whereas samples that are plotted in the center are fully uncertain (Fig. 6B). The eight samples that were misclassified were in concordance with the cross-validation error obtained during the training of the model: four FEA samples and two FA and FC samples respectively. The plot shows that two-thirds of the 12 samples were correctly

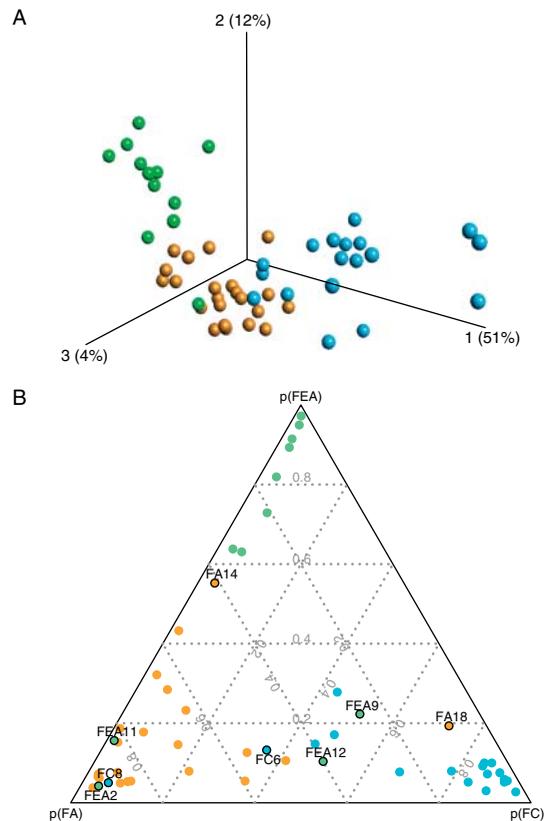


Figure 6 (A) Results obtained from the principal component analysis (PCA) of the 116 signatures defined in classifier 2 built to separate the three nodules: FA, FEA and FC. Each dot represents a sample: follicular adenoma (FA), light blue; follicular carcinoma (FC), orange; fetal adenoma (FEA), green. The three-dimensional PCA plot captures 67% of the variance described by the gene signature. The percentage of the total variance described by each of the three principal components is given in the parentheses at each axis. All groups can be easily distinguished from each other, except for a few samples that are the FA and FC groups, and are well separated, except for a few samples, which are located in the borderline area between the FA and FC clusters. The FEA samples overlap with the FA samples. The PCA and visualization were performed using the Qlucore Gene Expression Explorer 1.1 (www.qlucore.com). (B) Ternary plot of the predictive probability output obtained for classifier 2. Each dot is a sample, where the color code represents the true class: follicular adenoma (FA), light blue; follicular carcinoma (FC), orange; fetal adenoma (FEA), green. Each vertex of the triangle represents a subclass. Samples plotted close to a vertex have a high probability of belonging to this particular class, and samples plotted in the center are fully uncertain. The dotted lines indicate the predictive probability with respect to each of the three groups. Using the SVM model as a probabilistic classifier, eight samples were misclassified: four FEA, two FA and two FC samples. The misclassified samples are labeled.

classified as FEA. This was in concordance with the PCA of the gene signature of classifier 2. Here, we observed that the three classes were separable, although a few samples from each class were overlapping with other classes (Fig. 6A). Driven by

the observation that FEA indeed is a subtype of FAs, which is supported by classifier 1 results (Fig. 5B and Supplementary Figure 3A, see section on supplementary data given at the end of this article), we constructed a third classifier, classifier 3, where the two adenoma subclasses were merged.

We tested the probabilistic classifier's ability to distinguish between the merged adenoma group and the carcinomas. The results are shown in Table 3. Five samples were misclassified: three adenoma samples and two FC samples. In agreement with classifiers 1 and 2, the sample, FA18, was placed as an FC sample (Supplementary Figure 4B, see section on supplementary data given at the end of this article). According to the probabilistic classifier, ten of the twelve FEA samples had a high probability of belonging to the FA class (Supplementary Figure 4B). This was expected, and suggests that these two classes have a higher degree of similarity than FC and FEA. This is also supported by the differential expression analysis where we found 240 genes differentially expressed between FA and FEAs compared with 512 genes regulated between FEAs and FCs. Placing the FEA samples with the FA group is justifiable according to the results from classifier 2. Since classifier 1 accurately classified the FEA samples as FA and showed the optimal results as well as had the benefit of a balanced design between FA and FC, we used this signature for further analysis.

Test of classifier on external validation data

One downside of the microarray classifiers is that different studies analyzing the same outcome report different genes used in the classifier. Thus, we examined whether or not classifier 1 could accurately classify independent data. The validation data made publically available by Hinsch *et al.* were produced with oligonucleotide arrays obtained from Applied Biosystems, and moreover, they provided a means of validating the model cross-platform. The data set consisted of four FA and eight FC samples. We downloaded an expression matrix of preprocessed data and used gene symbols to match the genes in our model to the probe id on the ABI array. Of the 76 genes in our signature, 53 were represented on this array. Applying the SVM classifier on the data resulted in an accuracy of 83% (10/12), whereas the qLDA classifier had an accuracy of 92% (11/12) with a sensitivity of 1.0 for FC, validating the performance and robustness of the signature (Table 4, Supplementary Table 18, see section on supplementary data given at the end of this article).

Table 4 Classifier performance of gene signatures. Results were obtained for classifier 1 on the internal 40-sample data set and the external data reported by Weber *et al.* (24 samples) and Hinsch *et al.* (12 samples), seven additional signatures were tested on training and validation data. The support vector machine classifier was built to distinguish between follicular adenoma (FA) and follicular carcinoma (FC) samples. Accuracy and sensitivity for FC are given

Gene signature used	Data set	Accuracy	Sensitivity (FC)	Extra data
76-gene (76 probes) classifier 1	40 samples	95%	0.94	Table 2, Supplementary Table 2
76-gene (45 probes) classifier 1	24 samples	92%	0.83	Supplementary Table 10
76-gene (53 probes) classifier 1	12 samples	83% (92% qLDA)	0.88 (1.0)	Supplementary Table 18
Weber <i>et al.</i> 's 80 genes (96PS)	40 samples	72%	0.67	Supplementary Table 3
Weber <i>et al.</i> 's 80 genes (96PS)	24 samples	92%	0.91	Supplementary Table 11
Weber <i>et al.</i> 's 3 genes (5PS)	40 samples	43%	0.28	Supplementary Table 4
Weber <i>et al.</i> 's 3 genes (5PS)	24 samples	83%	0.92	Supplementary Table 12
Foukakis <i>et al.</i> 's 5 genes (6PS)	40 samples	85%	0.78	Supplementary Table 5
Foukakis <i>et al.</i> 's 5 genes (6PS)	24 samples	71%	0.67	Supplementary Table 13
Griffith <i>et al.</i> 's 12 genes (28PS)	40 samples	55%	0.44	Supplementary Table 6
Griffith <i>et al.</i> 's 12 genes (28PS)	24 samples	88%	0.83	Supplementary Table 14
Griffith <i>et al.</i> 's 32 genes (32PS)	40 samples	83%	0.78	Supplementary Table 7
Griffith <i>et al.</i> 's 32 genes (26PS)	24 samples	75%	0.75	Supplementary Table 15
Hinsch <i>et al.</i> 's 21 genes (32PS)	40 samples	70%	0.67	Supplementary Table 8
Hinsch <i>et al.</i> 's 21 genes (27PS)	24 samples	71%	0.58	Supplementary Table 16
Prasad <i>et al.</i> 's 75 genes (158PS)	40 samples	80%	0.67	Supplementary Table 9
Prasad <i>et al.</i> 's 75 genes (93PS)	24 samples	92%	0.83	Supplementary Table 17

Supplementary Tables, see section on supplementary data given at the end of this article.

Moreover, we obtained raw data files of 12 FA and 12 FC samples analyzed with Affymetrix HG-U133A arrays, which were preprocessed and re-analyzed as described by Weber *et al.* Initially, we reproduced the result reported by Weber *et al.*, i.e. obtaining an accuracy of 96% (23/24) in discriminating 12 FA samples from 12 FC samples using a linear discriminant analysis, based on 80 regulated genes (Table 4, Supplementary Table 11, see section on supplementary data given at the end of this article). These samples were analyzed on an older generation of Affymetrix arrays, the HG-U133A array, and ~22,500 transcripts are shared between the two generations of arrays. Of the 76 probe sets in our classifier, 45 were represented on the older array and used in the analysis. When applying our SVM classifier on the Weber's validation data, we obtained an accuracy of 92% (22/24) with a sensitivity of 0.83 and a specificity of 1.0 for FC (Supplementary Tables 3 and 10, see section on supplementary data given at the end of this article), although this data set was preprocessed and normalized differently from our data. Multiple studies have shown that different normalization strategies have a great impact on data analysis and end results (Hoffmann *et al.* 2002, Ploner *et al.* 2005, Shedden *et al.* 2005), and the high accuracy on the validation data emphasizes the robustness of the classifier and signature. Although the reduced gene signature from classifier 1 performed well on the older generation of Affymetrix arrays, we observed a

decrease in performance when we substituted this reduced gene list into our classifier and applied it on the HG-U133_plus_2 array, resulting in an accuracy of 88% for SVM and 92% using the LDA algorithm (data not shown). This suggests that we would get a even better classification of the validation data had they been hybridized on the next generation arrays, since the subset from the HG-U133A array did not perform as well on the new array as the full set of 76 probe sets.

Cross-platform and cross-laboratory use of signatures

In order to further test the overall performance of the FA classifier versus FC classifier, we applied a selection of recently published gene signatures and classifiers on both our 40 follicular neoplasia samples and the 24 validation samples to test if our classifier gives comparable or better results. Besides the 80-gene signature mentioned above, we tested additional 6-gene signatures (Griffith *et al.* 2006, Foukakis *et al.* 2007, Prasad *et al.* 2008, Hinsch *et al.* 2009). The results for all the signatures are given in Table 4, and they are given in detail for all the employed algorithms in Supplementary Tables 3–18, see section on supplementary data given at the end of this article. First, we applied the two signatures (3 and 80 genes) published by Weber *et al.* using a model framework that was the same as that used in the previous analyses. Weber's 80-gene signature did not perform as well on

our 40 samples (accuracy of 72% (31/40), Table 4, Supplementary Table 3) as our signature did on their data (accuracy of 92%, Table 4, Supplementary Table 10) although all genes from Weber's signature were represented on the array used in our analysis. When applying the optimized 3-gene signature of Weber *et al.* on their own data, we obtained an accuracy of 83% (20/24) compared with 43% (17/40) with our 40 samples (Table 4, Supplementary Tables 4 and 12), indicating that the 3-gene signature is too small to show any discriminating power on external data. On the contrary, we obtained better results when applying the 5-gene signature published by Foukakis *et al.* on our 40-sample data set, namely an accuracy of 85% (34/40). Four of the misplaced samples were FCs, which resulted in a sensitivity and specificity of 0.78 and 0.91 respectively for classifying FC samples (Table 4, Supplementary Table 5). When applied on the 24-sample validation data set, the 5-gene signature had an accuracy of 71% (17/24), Supplementary Table 13.

Also, we tested a 32-gene signature optimized to distinguish benign thyroid lesion from malignant thyroid lesion, which was derived in a large meta-study by taking the lessons from recent papers on thyroid microarray analysis into account (Griffith *et al.* 2006). Based on a ranking system giving higher weight to genes that were selected in three or more of the evaluated expression studies, a top 12-gene signature was devised (Griffith *et al.* 2006). For each signature, we performed classification by applying both SVM learning and other algorithms for comparison. The SVM model gave the worst outcome with the top 12-gene signature applied on our data of only 55% accuracy, i.e. sensitivity and specificity of 0.44 and 0.64 respectively for FC. Nearest centroid showed improved performance with 70% accuracy (Table 4 and Supplementary Table 6). Better results were obtained when applying the signature to the validation data set reported by Weber, showing 88% accuracy, reflecting a sensitivity and specificity of 0.83 and 0.92 respectively (Table 4 and Supplementary Table 14). Notably, this 12-gene signature performed as well as Weber's own 80-gene signature when applying the LDA algorithm resulting in an accuracy of 92% (22/24).

The poor results of the top 12 meta-genes on our data were improved somewhat when the full 32-gene signature was applied (Griffith *et al.* 2006), increasing the accuracy to 83% (33/40), see Table 4 and Supplementary Table 7. Lastly, we tested the performance of the SVM classification based on the 25 (21-annotated)-gene signature published by Hinsch *et al.* on our 40-sample data set and the 24-sample set,

which resulted in accuracies of 70 and 71% compared with 80 and 92% for the 75-gene signature published by Prasad *et al.* (Table 4, Supplementary Tables 8, 9, 16, and 17).

In general, we observed that even if a signature showed good performance on one data set, it performed poorly on the other data set (Table 4). Overall, classifier 1, built to classify FA and FC, showed the best cross-platform and cross-laboratory performance, both on the training set and on validation data sets (Weber *et al.* 2005, Hinsch *et al.* 2009) with PPVs of 0.94, 0.89 and 1.0 respectively for malignancy for FC (Table 4, Supplementary Table 3).

Discussion

We showed that FC is characterized by increased levels of mRNAs encoding proteins involved in DNA replication and mitosis corresponding to increased numbers of dividing cells, as well as to the loss of transcripts encoding proteins involved in growth arrest and apoptosis. Taken together, these aberrations may provide a minimal platform for malignant transformation (Evan & Vousden 2001).

Poorly differentiated and invasive carcinomas are known to exhibit a high proliferative grading, and it has been debated whether the mitotic index is useful to diagnose FC (Perez-Montiel & Suster 2008, Ghossein 2009). In agreement with the clinical experience, we noted that cell-cycle mRNAs followed a gradient ranging from a few fold to more than 50-fold up-regulation, which may limit the isolated use of this parameter for diagnostic purposes. A number of the up-regulated transcripts including anillin (Hall *et al.* 2005), ARP 2/3 complex (Otsubo *et al.* 2004), abnormal spindle homolog (Ayllon & O'connor 2007, Lin *et al.* 2008), centromere protein F (Campone *et al.* 2008), KIF4A (Taniwaki *et al.* 2007), maternal embryonic leucine zipper kinase (Gray *et al.* 2005), NIMA-related kinase 2 (Hayward & Fry 2006), PDZ-binding kinase, protein regulator of cytokinesis 1 (Boukarabila *et al.* 2009), regulator of chromosome condensation 2 (Stacey *et al.* 2008), encoded proteins that are directly involved in mitosis and several of them are over-expressed in other cancers. In particular, transcripts encoding ribonucleotide reductase small subunit, RRM2, and topoisomerase 2 α , TOP2A respectively, are intimately connected to cancer. RRM2 promotes invasion and metastasis of tumors, and its over-expression is associated with gemcitabine resistance (Boukovinas *et al.* 2008, Souglakos *et al.* 2008), whereas TOP2A is frequently amplified in breast cancer, where it is

an independent predictor of survival and a marker for anthracycline-based chemotherapy (Pritchard *et al.* 2008, O'Malley *et al.* 2009).

Apoptosis is important in both benign and malignant thyroid diseases (Mitsiades *et al.* 2000, 2003, 2006, Chen *et al.* 2004). Under normal conditions, epithelial cells are strictly organized, and detachment from the epithelial lining and basal membrane triggers apoptosis (Evan & Vousden 2001). In this way, proliferation and migration of preneoplastic cells are suppressed. In comparison with normal thyroid tissue and benign goiter and adenoma, we found that loss of apoptotic and growth arrest factors first occur during transformation to malignancy. Compared with papillary cancers that only exhibit a moderate decrease in apoptotic factors, down-regulation was marked in follicular cancers and ACs. In contrast to the up-regulation of cell-cycle-associated transcripts, loss of apoptotic mRNAs was prominent in all samples, implying that this event precedes proliferation. The coordinated down-regulation of *NR4A1* and *NR4A3* and *JUN*, *FOSB*, and *CITED2* is striking, since these factors have previously been shown to be part of a common proapoptotic and cancer-predisposing pathways (Mullican *et al.* 2007). Moreover, this finding is supported by two recent studies where *NR4A1* was found to be down-regulated in FC (Fryknas *et al.* 2006, Camacho *et al.* 2009). *NR4A1* and *NR4A3* also known as Nur77 and Nor-1 respectively are homologous orphan nuclear receptors that regulate the transcription of a common set of target genes (Li *et al.* 2006), and both have been described as homeostatic regulators of proliferation and apoptosis (Moll *et al.* 2006, Zhan *et al.* 2008). While *NR4A1*- and *NR4A3*-deficient mice respectively exhibit subtle phenotypes, it was recently shown that double knock-out quickly leads to acute myeloid leukemia. The mice exhibited abnormal expansion of hematopoietic stem cells and myeloid progenitors as well as decreased expression of the AP-1 transcription factors JunB and c-Jun and defective extrinsic apoptotic Fas-L and TRAIL signaling (Mullican *et al.* 2007). *NR4A1* and *NR4A3* translocate to mitochondria and stimulate the release of cytochrome *c* in a BCL2-dependent manner. The observed down-regulation of mitochondrial ion channels could promote these processes, since they also participate in apoptosis (Yu & Choi 2000, Szabo *et al.* 2004).

FEA has many similarities to FA, but due to its morphological resemblance to fetal thyroid it has been described as a separate follicular variant. As shown in the PCA of all transcripts, both FC and FA are heterogeneous, and this is also the case for FEA. FCs

may roughly be distinguished from FEA by the transcripts that are the same as those that differentiate FCs from FAS, and in this way, they may classify FEA as adenoma. On the other hand, a few hundred transcripts differ between FAs and FEAs, supporting that FEA represents a distinct variant of the adenoma. However, we found no evidence of any particular changes in fetal markers such as *PAX8*, *TTF-1*, and *HEX* or markers of differentiated thyroid cells such as *DUOX*, *NIS*, *TPO*, or *PDS* to support a fetal origin, so perhaps it is not justified to withhold the present nomenclature. Moreover, it should be noted that the difference among the transcriptomes of FA and FEA is merely a quantitative difference. We cannot detect mRNAs that are specific to either tumor, so taken together, we propose that FEAs are categorized as FAs until a unique difference in clinical outcome or biology is demonstrated.

Since the global expression data identified a number of transcripts encoding factors intimately correlated to transformation, we explored if it was possible to generate a robust diagnostic signature. We compared the performance of several algorithms (Table 3), and although the majority performed well, the support vector was efficient in all combinations. Classifier 1, which was designed to distinguish between FA and FC, exhibited a sensitivity of 0.94 and specificity of 0.96 for FC. Both widely and minimally invasive carcinomas were accurately predicted, indicating that this histopathological distinction is not related to the transcripts included in the classifier. Although with lower sensitivity and specificity, it was also possible to classify FEA supporting the unique nature of these tumors. One of the major challenges for microarray-based diagnostics is to eliminate cross-platform variation and exploit public domain data in the development of robust signatures. Hence, we compared the efficacy of previously generated signatures on our material, and used the publically available expression profiles as the validation set. In agreement with previous experiences from analysis of breast cancer samples (Sotiriou & Pusztai 2009), there was limited overlap between the genetic signatures that different research groups have employed for classification. Nevertheless, there was an encouraging agreement in their ability to predict the correct diagnosis (Prasad *et al.* 2008, Hinsch *et al.* 2009). Classifier 1 correctly determined the diagnosis of FA and FC in 92% of the tumors examined by two independent laboratories (Weber *et al.* 2005, Hinsch *et al.* 2009), and stands out as a very robust signature. In general, classifiers consisting of few transcripts were less accurate, probably reflecting that small numbers of

mRNAs in a classifier may be more sensitive to geographical and platform variations (Table 4). We trust that the high accuracy of our signature may be related to improved mathematical tools and the fact that we have had the possibility to select more optimal probe sets, since the whole genome U133 2.0 array contains about double as many probe sets compared with previous generations of arrays. Moreover, the signature used for classification is very similar to the set of differentially expressed mRNAs, which as described above, may reflect biological changes that are intimately connected to transformation.

In conclusion, we propose that down-regulation of factors involved in growth arrest and apoptosis may represent a decisive step in the pathogenesis of FC. The coordinated loss of NR4A1 and NR4A3 may play a central role in transformation since this pathway previously has been causally associated with malignancy. Finally, additional clinico-pathological studies with follow-up are needed to assess the predictive power of this apparently robust classifier in differentiating FA from FC in the clinical setting.

Supplementary data

This is linked to the online version of the paper at <http://dx.doi.org/10.1677/ERC-09-0288>.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

The Danish Medical Research Councils, the Danish Cancer Society, the Novo Nordisk Foundation, and the Toyota Foundation supported the research.

Acknowledgements

We would like to thank Susanne Smed and Elisabeth Schiefloe for technical assistance.

References

Ayllon V & O'connor R 2007 PBK/TOPK promotes tumour cell proliferation through p38 MAPK activity and regulation of the DNA damage response. *Oncogene* **26** 3451–3461.

Barden CB, Shister KW, Zhu B, Guiter G, Greenblatt DY, Zeiger MA & Fahey TJ III 2003 Classification of follicular thyroid tumors by molecular signature: results of gene profiling. *Clinical Cancer Research* **9** 1792–1800.

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA *et al.* 2009 NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Research* **37** D885–D890.

Benjamini Y & Hochberg Y 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* **57** 289–300.

Bolstad BM, Irizarry RA, Astrand M & Speed TP 2003 A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193.

Boukarabila H, Saurin AJ, Batsche E, Mossadegh N, van LM, Otte AP, Pradel J, Muchardt C, Sieweke M & Duprez E 2009 The PRC1 polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes and Development* **23** 1195–1206.

Boukovinas I, Papadaki C, Mendez P, Taron M, Mavroudis D, Koutsopoulos A, Sanchez-Ronco M, Sanchez JJ, Trypaki M, Staphopoulos E *et al.* 2008 Tumor BRCA1, RRM1 and RRM2 mRNA expression levels and clinical response to first-line gemcitabine plus docetaxel in non-small-cell lung cancer patients. *PLoS ONE* **3** e3695.

Camacho CP, Latini FR, Oler G, Hojaij FC, Maciel RM, Riggins GJ & Cerutti JM 2009 Down-regulation of NR4A1 in follicular thyroid carcinomas is restored following lithium treatment. *Clinical Endocrinology* **70** 475–483.

Campone M, Campion L, Roche H, Gouraud W, Charbonnel C, Magrangeas F, Minvielle S, Geneve J, Martin AL, Bataille R *et al.* 2008 Prediction of metastatic relapse in node-positive breast cancer: establishment of a clinicogenomic model after FEC100 adjuvant regimen. *Breast Cancer Research and Treatment* **109** 491–501.

Castro P, Sansonetty F, Soares P, Dias A & Sobrinho-Simoes M 2001 Fetal adenomas and minimally invasive follicular carcinomas of the thyroid frequently display a triploid or near triploid DNA pattern. *Virchows Archiv* **438** 336–342.

Chen S, Fazle Akbar SM, Zhen Z, Luo Y, Deng L, Huang H, Chen L & Li W 2004 Analysis of the expression of Fas, FasL and Bcl-2 in the pathogenesis of autoimmune thyroid disorders. *Cellular & Molecular Immunology* **1** 224–228.

Delellis RA 2006 Pathology and genetics of thyroid carcinoma. *Journal of Surgical Oncology* **94** 662–669.

Evan GI & Vousden KH 2001 Proliferation, cell cycle and apoptosis in cancer. *Nature* **411** 342–348.

Fagin JA & Mitsiades N 2008 Molecular pathology of thyroid cancer: diagnostic and clinical implications. *Best Practice & Research. Clinical Endocrinology & Metabolism* **22** 955–969.

Finley DJ, Zhu B, Barden CB & Fahey TJ III 2004 Discrimination of benign and malignant thyroid nodules by molecular profiling. *Annals of Surgery* **240** 425–436.

- Foukakis T, Gusnanto A, Au AY, Hoog A, Lui WO, Larsson C, Wallin G & Zedenius J 2007 A PCR-based expression signature of malignancy in follicular thyroid tumors. *Endocrine-Related Cancer* **14** 381–391.
- Fryknas M, Wickenberg-Bolin U, Goransson H, Gustafsson MG, Foukakis T, Lee JJ, Landegren U, Hoog A, Larsson C, Grimelius L *et al.* 2006 Molecular markers for discrimination of benign and malignant follicular thyroid tumors. *Tumour Biology* **27** 211–220.
- Fujarewicz K, Jarzab M, Eszlinger M, Krohn K, Paschke R, Oczko-Wojciechowska M, Wiench M, Kukulska A, Jarzab B & Swierniak A 2007 A multi-gene approach to differentiate papillary thyroid carcinoma from benign lesions: gene selection using support vector machines with bootstrapping. *Endocrine-Related Cancer* **14** 809–826.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al.* 2004 Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5** R80.
- Ghossein R 2009 Problems and controversies in the histopathology of thyroid carcinomas of follicular cell origin. *Archives of Pathology & Laboratory Medicine* **133** 683–691.
- Gray D, Jubb AM, Hogue D, Dowd P, Kljavin N, Yi S, Bai W, Frantz G, Zhang Z, Koepfen H *et al.* 2005 Maternal embryonic leucine zipper kinase/murine protein serine–threonine kinase 38 is a promising therapeutic target for multiple cancers. *Cancer Research* **65** 9751–9761.
- Griffith OL, Melck A, Jones SJ & Wiseman SM 2006 Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *Journal of Clinical Oncology* **24** 5043–5051.
- Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M *et al.* 2009 Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nature Genetics* **41** 460–464.
- Hall PA, Todd CB, Hyland PL, McDade SS, Grabsch H, Dattani M, Hillan KJ & Russell SE 2005 The septin-binding protein anillin is overexpressed in diverse human tumors. *Clinical Cancer Research* **11** 6780–6786.
- Hayward DG & Fry AM 2006 Nek2 kinase in chromosome instability and cancer. *Cancer Letters* **237** 155–166.
- Hegedus L 2004 Clinical practice. The thyroid nodule. *New England Journal of Medicine* **351** 1764–1771.
- Hegedus L, Bonnema SJ & Bennedbaek FN 2003 Management of simple nodular goiter: current status and future perspectives. *Endocrine Reviews* **24** 102–132.
- Hinsch N, Frank M, Doring C, Vorlander C & Hansmann ML 2009 QPRT: a potential marker for follicular thyroid carcinoma including minimal invasive variant; a gene expression, RNA and immunohistochemical study. *BMC Cancer* **9** 93.
- Hoffmann R, Seidl T & Dugas M 2002 Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. *Genome Biology* **3** RESEARCH0033.
- Jonson L, Vikesaa J, Krogh A, Nielsen LK, Hansen T, Borup R, Johnsen AH, Christiansen J & Nielsen FC 2007 Molecular composition of IMP1 ribonucleoprotein granules. *Molecular & Cellular Proteomics* **6** 798–811.
- Li C & Wong WH 2001 Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *PNAS* **98** 31–36.
- Li QX, Ke N, Sundaram R & Wong-Staal F 2006 NR4A1, 2, 3 – an orphan nuclear hormone receptor family involved in cell apoptosis and carcinogenesis. *Histology and Histopathology* **21** 533–540.
- Lin SY, Pan HW, Liu SH, Jeng YM, Hu FC, Peng SY, Lai PL & Hsu HC 2008 ASPM is a novel marker for vascular invasion, early recurrence, and poor prognosis of hepatocellular carcinoma. *Clinical Cancer Research* **14** 4814–4820.
- Lubitzz CC, Gallagher LA, Finley DJ, Zhu B & Fahey TJ III 2005 Molecular analysis of minimally invasive follicular carcinomas by gene profiling. *Surgery* **138** 1042–1048.
- Mazzanti C, Zeiger MA, Costouros NG, Umbricht C, Westra WH, Smith D, Somervell H, Bevilacqua G, Alexander HR & Libutti SK 2004 Using gene expression profiling to differentiate benign versus malignant thyroid tumors. *Cancer Research* **64** 2898–2903.
- Mitsiades N, Poulaki V, Tseleni-Balafouta S, Koutras DA & Stamenkovic I 2000 Thyroid carcinoma cells are resistant to FAS-mediated apoptosis but sensitive to tumor necrosis factor-related apoptosis-inducing ligand. *Cancer Research* **60** 4122–4129.
- Mitsiades CS, Poulaki V & Mitsiades N 2003 The role of apoptosis-inducing receptors of the tumor necrosis factor family in thyroid cancer. *Journal of Endocrinology* **178** 205–216.
- Mitsiades CS, Poulaki V, Fanourakis G, Sozopoulos E, McMillin D, Wen Z, Voutsinas G, Tseleni-Balafouta S & Mitsiades N 2006 Fas signaling in thyroid carcinomas is diverted from apoptosis to proliferation. *Clinical Cancer Research* **12** 3705–3712.
- Moll UM, Marchenko N & Zhang XK 2006 p53 and Nur77/p53 and Nur77/TR3 – transcription factors that directly target mitochondria for cell death induction. *Oncogene* **25** 4725–4743.
- Mullican SE, Zhang S, Konopleva M, Ruvolo V, Andreeff M, Milbrandt J & Conneely OM 2007 Abrogation of nuclear receptors Nr4a3 and Nr4a1 leads to development of acute myeloid leukemia. *Nature Medicine* **13** 730–735.
- Nikiforova MN, Lynch RA, Biddinger PW, Alexander EK, Dorn GW, Tallini G, Kroll TG & Nikiforov YE 2003 RAS point mutations and PAX8–PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma. *Journal of Clinical Endocrinology and Metabolism* **88** 2318–2326.

- O'Malley FP, Chia S, Tu D, Shepherd LE, Levine MN, Bramwell VH, Andrulis IL & Pritchard KI 2009 Topoisomerase II alpha and responsiveness of breast cancer to adjuvant chemotherapy. *Journal of the National Cancer Institute* **101** 644–650.
- Otsubo T, Iwaya K, Mukai Y, Mizokami Y, Serizawa H, Matsuoka T & Mukai K 2004 Involvement of Arp2/3 complex in the process of colorectal carcinogenesis. *Modern Pathology* **17** 461–467.
- Perez-Montiel MD & Suster S 2008 The spectrum of histologic changes in thyroid hyperplasia: a clinicopathologic study of 300 cases. *Human Pathology* **39** 1080–1087.
- Platt JC 1999 Probabilities for SV machines. *Advances in Large Margin Classifiers*, pp 61–74. MIT Press: Cambridge, MA, USA.
- Ploner A, Miller LD, Hall P, Bergh J & Pawitan Y 2005 Correlation test to assess low-level processing of high-density oligonucleotide microarray data. *BMC Bioinformatics* **6** 80.
- Prasad NB, Somervell H, Tufano RP, Dackiw AP, Marohn MR, Califano JA, Wang Y, Westra WH, Clark DP, Umbricht CB *et al.* 2008 Identification of genes differentially expressed in benign versus malignant thyroid tumors. *Clinical Cancer Research* **14** 3327–3337.
- Pritchard KI, Messersmith H, Elavathil L, Trudeau M, O'Malley F & Dhesy-Thind B 2008 HER-2 and topoisomerase II as predictors of response to chemotherapy. *Journal of Clinical Oncology* **26** 736–744.
- Ruggeri RM, Campenni A, Baldari S, Trimarchi F & Trovato M 2008 What is new on thyroid cancer biomarkers. *Biomarker Insights* **3** 237–252.
- Shedden K, Chen W, Kuick R, Ghosh D, Macdonald J, Cho KR, Giordano TJ, Gruber SB, Fearon ER, Taylor JM *et al.* 2005 Comparison of seven methods for producing Affymetrix expression scores based on false discovery rates in disease profiling data. *BMC Bioinformatics* **6** 26.
- Simon R 2006 A checklist for evaluating reports of expression profiling for treatment selection. *Clinical Advances in Hematology and Oncology* **4** 219–224.
- Simon R, Lam A, Li MC, Ngan M, Menezes S & Zhao Y 2007 Analysis of gene expression data using BRB-Array tools. *Cancer Informatics* **2** 11–17.
- Sotiropoulos C & Pusztai L 2009 Gene-expression signatures in breast cancer. *New England Journal of Medicine* **360** 790–800.
- Souglakos J, Boukovinas I, Taron M, Mendez P, Mavroudis D, Tripaki M, Hatzidaki D, Koutsopoulos A, Stathopoulos E, Georgoulas V *et al.* 2008 Ribonucleotide reductase subunits M1 and M2 mRNA expression levels and clinical outcome of lung adenocarcinoma patients treated with docetaxel/gemcitabine. *British Journal of Cancer* **98** 1710–1715.
- Stacey SN, Gudbjartsson DF, Sulem P, Bergthorsson JT, Kumar R, Thorleifsson G, Sigurdsson A, Jakobsdottir M, Sigurgeirsson B, Benediktsdottir KR *et al.* 2008 Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nature Genetics* **40** 1313–1318.
- Szabo I, Adams C & Gulbins E 2004 Ion channels and membrane rafts in apoptosis. *Pflügers Archiv* **448** 304–312.
- Taniwaki M, Takano A, Ishikawa N, Yasui W, Inai K, Nishimura H, Tsuchiya E, Kohno N, Nakamura Y & Daigo Y 2007 Activation of KIF4A as a prognostic biomarker and therapeutic target for lung cancer. *Clinical Cancer Research* **13** 6624–6631.
- Thierry-Mieg D & Thierry-Mieg J 2006 AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biology* **7** (Supplement 1) S12–S14.
- Tusher VG, Tibshirani R & Chu G 2001 Significance analysis of microarrays applied to the ionizing radiation response. *PNAS* **98** 5116–5121.
- Utiger RD 2005 The multiplicity of thyroid nodules and carcinomas. *New England Journal of Medicine* **352** 2376–2378.
- Vapnik V 1998 *Statistical Learning Theory*. New York, NY, USA: Wiley-Interscience.
- Weber F, Shen L, Aldred MA, Morrison CD, Frilling A, Saji M, Schuppert F, Broelsch CE, Ringel MD & Eng C 2005 Genetic classification of benign and malignant thyroid follicular neoplasia based on a three-gene combination. *Journal of Clinical Endocrinology and Metabolism* **90** 2512–2521.
- Yu SP & Choi DW 2000 Ions, cell volume, and apoptosis. *PNAS* **97** 9360–9362.
- Zhan Y, Du X, Chen H, Liu J, Zhao B, Huang D, Li G, Xu Q, Zhang M, Weimer BC *et al.* 2008 Cyclosporin B is an agonist for nuclear orphan receptor Nur77. *Nature Chemical Biology* **4** 548–556.