ABSTRACT

Clinicopathological variables used at present for prognostication and treatment selection for papillary thyroid carcinomas (PTCs) do not uniformly predict tumor behavior, necessitating identification of novel prognostic markers. Complicating the assessment is the long natural history of PTC and our rudimentary knowledge of its genetic composition. In this study we took advantage of differences in clinical behavior of two distinct variants of PTC, the aggressive tall-cell variant (TCV) and indolent conventional PTC (cPTC), to identify molecular prognosticators of outcome using complementary genome wide analyses. Comparative genome hybridization (CGH) and cDNA microarray (17,840 genes) analyses were used to detect changes in DNA copy number and gene expression in pathological cPTC and TCV. The findings from CGH and cDNA microarray analyses were correlated and validated by real-time PCR and immunohistochemical analyses on a series of 100 cases of cPTC and TCV. Genes identified by this approach were evaluated as prognostic markers in cPTC by immunohistochemistry on tissue arrays. CGH identified significant differences in the presence and extent of DNA copy number aberrations in TCV compared with cPTC. Recurrent gains of 1p34–36, 1q21, 6p21–22, 9q34, 11q13, 17q25, 19, and 22 and losses of 2q11–31, 4, 5p14–q21, 6q11–22, 8q11–22, 8q11–32, 9q11–32, and 13q21–31 were unique to TCV. Hierarchical clustering of gene expression profiles revealed significant overlap between TCV and cPTC, but further analysis identified 82 dysregulated genes differentially expressed among the PTC variants. Of these, MUC1 was of particular interest because amplification of 1q by CGH correlated with MUC1 amplification by real-time PCR analysis and protein overexpression by immunohistochemistry in TCV (P = 0.005). Multivariate analysis revealed a significant association between MUC1 overexpression and treatment outcome, independent of histopathological categorization (P = 0.03). Analysis of a validation series containing a matched group of aggressive and indolent PTCs confirmed the association between MUC1 overexpression and survival (relative risk, 2.3; 95% confidence interval, 1.1–5.5; P = 0.03). Our data suggest that MUC1 dysregulation is associated with aggressive behavior of PTC and may serve as a prognostic marker and potential therapeutic target in this disease.

INTRODUCTION

Cancer results from a random accumulation of genetic and epigenetic aberrations that influence gene expression. Factors imparting growth advantages are propagated in a Darwinian-like selection process (1). At least 10–20% of gene expression changes in cancer cells directly result from DNA copy number abnormalities that are significant contributors to the oncogenesis of virtually all human malignancies (2–5). Moreover, the prognostic and therapeutic implications associated with these aberrations are becoming increasingly apparent, including the activation of several important oncogenes, such as ERBB2 (at chromosome 17q21.1) and epidermal growth factor receptor (EGFR; 7p12) from amplification events (6, 7). Recent work has shown that the expression of relatively few genes is altered in amplified regions (2, 4, 5). Accordingly, the simultaneous application of global genomic screening with high-throughput gene expression profiling has the potential to identify genes dysregulated as a result of primary genetic events among the large pool of secondary downstream changes in gene expression (8).

In the case of papillary thyroid carcinomas (PTCs), previous studies have independently demonstrated the presence of DNA copy number changes and gene expression changes (9–11). However, the genes driving selection for individual genomic aberrations have not been identified. In particular, the genetic basis of aggressive PTC, which may reveal outcome predictors and treatment targets, is unknown. We performed a genome-wide appraisal of clinically aggressive PTC, using comparative genomic hybridization (CGH) and cDNA microarray analyses to screen for DNA copy number aberrations and gene expression changes, respectively. To circumvent issues relating to the long natural history of PTC, we performed a comparative analysis of indolent conventional PTC (cPTC) and tall-cell variant PTC (TCV), an aggressive morphological variant of PTC (12, 13). Putative primary genetic events were identified by correlation of recurrent DNA copy number changes with gene expression changes and validated by real-time PCR and immunohistochemical analyses on tissue microarrays, respectively. The prognostic utility of candidate genes was validated in an independent patient group with cPTC. Our data identified several gene targets that may drive selection of DNA copy number changes in PTC. Of these, amplification and overexpression of MUC1 is significantly associated with outcome of PTC and may serve as a prognostic marker and therapeutic target in this disease.

MATERIALS AND METHODS

Case Collection and Pathological Classification. The study population included randomly selected cases of well-differentiated cPTC and TCV identified through a search of the Departments of Surgery and Pathology databases. Only previously untreated cases with available frozen primary tumor tissue obtained according to guidelines established by our Institutional Review Board were included in the study. A detailed histopathological review of all cases was performed by at least two pathologists with extensive experience in cancer pathology. Cases were grouped by the diagnostic criteria described by Rosai et al. (14) in the Armed Forces Institute of Pathology fascicle. Briefly, cPTC included cases of PTC of classical or follicular variants, characterized by the presence of “ground glass” nuclei (Orphan Annie Eyed Nuclei) without a significant degree of mitotic activity, necrosis, or nuclear pleomorphism (well differentiated). Cases of TCV were defined as PTC with typical nuclear differentiation. The high proportion of tall cells was used for the purpose of this study to allow more
representative genetic analyses. Tall cells were defined as cells with a height-to-width ratio of at least 2:1 in the presence of an acidophilic, granular cytoplasm. The clinicopathological characteristics of the study population were representative of those seen in the general population with cPTC and TCV (Table 1).

Although cases of cPTC (median, 2.5 cm; range, 0.6–7.5 cm) and TCV (median, 2.5 cm; range, 0.9–7.5 cm) had a similar median tumor size, several negative prognostic features, including extrathyroidal extension, distant metastasis, and patient age >45 years, were more common in TCV. Accordingly, a higher proportion of TCVs were stratified into higher risk groups and advanced TNM stages. The approaches to the treatment of cPTC and TCV were not significantly different and were derived from established institutional protocols based on risk group stratification and the extent of disease at presentation. Overall, TCVs were treated with total thyroidectomy more often because of the presence of poor prognostic features at presentation. An independent validation series of cPTCs was assembled from our surgical thyroid cancer database, including patients treated at our institution between 1960 and 1997 (15). This validation group included patients with cPTC who failed treatment matched for patient, tumor, and treatment characteristics in a 1:2 ratio with patients cured of disease. Only histologically confirmed cases with paraffin-embedded primary tumor tissue, undergoing all treatment at our institution were included. Cure of disease was defined as the lack of recurrence or persistent disease >15 years after primary treatment. Treatment failure was defined as the development of local (nodal recurrences excluded; Ref. 16) or distant recurrence after initial treatment. On the basis of these inclusion criteria, tissue samples from 12 cases of cPTC with treatment failure and 24 matched “cured” cases were eligible for analysis. The small number of cases in the treatment failure group predominantly reflects exclusions for treatment outside our institution, lack of available specimens, or the reclassification of the tumors from cPTC based on histopathological review (i.e., exclusion of all non-cPTC histological variants of PTC).

**CGH Analysis.** CGH was performed as described previously (17). Seven to 10 separate metaphases were captured and processed with the Quantitative Image Processing System (Quips Pathysysm System; Applied Imaging, Santa Clara, CA). Red, green, and blue fluorescence intensities were analyzed for all metaphase spreads, normalized to a standard length, and statistically combined to show the red:green signal ratio and 95% confidence intervals for the entire chromosome. Copy number changes were detected based on the variance of the red:green ratio profile from the standard of 1. Ratio values of 1.2 and 2.0 were defined as thresholds for gains and high-level amplifications, respectively, and losses were defined as a ratio value ≤0.8.

**Preparation of cDNA Microarrays.** A set of 17,840 sequence-verified human Gene Expression Microarray clones (Incyte Genomics, Palo Alto, CA) representing both known genes and expressed sequence tags were spotted on poly-L-lysine-coated microscope slides by a custom robot (18). Before hybridization, slides were prepared as described previously (18).8

**Labeling of cDNA and Hybridization to Arrays.** cDNA labeling and hybridization were performed as described previously (19). In brief, RNA was obtained through a Trizol extraction followed by isolation using RNeasy columns (Qiagen, Valencia, CA). One hundred µg of total RNA from each tumor sample were used to synthesize cDNA with 400 units of Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA), cyanine-5 (Cy5)-conjugated dUTP (Amersham Biosciences, Piscataway, NJ), and oligo(dT)12-18 primer (Invitrogen). To maximize the distance among the samples during cluster analysis, equal amounts of total RNA from matched normal thyroid tissues were used to make a pool of reference RNA.

Approximately 100 µg of total RNA from such a pool were labeled with cyanine 3 (Cy3)-conjugated dUTP as reference cDNA. Reactions were carried out at 42°C for 2 h in a buffer containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 0.5 mM deoxynucleotide triphosphates, 10 mM DTT, and 40 units of RNase OUT Recombinant RNase Inhibitor (Invitrogen). Hybridization to cDNA arrays was carried out overnight at 50°C in a buffer containing 30% formamide, 3× SSC, 0.75% SDS, and 100 ng of human Cot-1 DNA (Invitrogen). After hybridization, slides were washed in 0.2× SSC–0.1% SDS at room temperature for 15 min and in 0.2× SSC at room temperature for 15 min. Slides were dried and scanned with a GenePix 4000A microarray scanner (Axon Instruments, Foster City, CA).

**Acquisition and Analysis of Microarray Data.** Red (Cy5) and green (Cy3) fluorescent signal intensities for each spot on the array were calculated with the GenePix Pro 3.0 software (Axon Instruments). This software computes both the mean and median intensities per spot for each channel as well as the background mean and median. The background-corrected intensity for each channel was calculated to be the mean of the foreground intensity minus the median of the background, which we denote as “immunoglobulin” and “Ir” for the green and red channels, respectively. As a diagnostic of the hybridization, log-log scatter plots of the geometric mean versus the ratio of the intensity in each channel were examined, and those with anomalous scatter plots were removed from further analysis. The data were then normalized by use of an intensity-dependent nonlinear normalization scheme as described in previous publications (18–22). The final step before statistical analysis removed irrelevant noise-contributing genes by establishing the log fold change cutoff for each gene in each sample to an absolute value of log(3) (i.e., log changes >3-fold or <1/3-fold) setting. Values less than log(3) were set to zero. Only genes that had non-zero fold changes in at least four samples (25%) were used in subsequent analyses. These analyses included both a comparison of global gene expression profiles by clustering methodology and a comparison of expression of individual genes among the samples. Tumor samples were clustered by use of the dot product (angle) metric to measure the distance between the samples, which were then clustered by the hierarchical Ward linkage method (23, 24). To assess the significance of the cluster structure found, nonparametric bootstrap resampling was used (25–27). One hundred bootstrap resamples of the data were generated and clustered, giving 100 trees. A consensus tree was built using the standard CONSENSUS program from the PHYLIP package (28) and an online phylogeny program.10

A majority-rule consensus tree was built from the 100 bootstrap trees. For each node, the number of times (of 100) the subtree appeared in the bootstrap samples is shown. The closer this value is to 100, the more robust is the subtree/subcluster. We then applied a nonparametric two-sample test to find individual genes whose expression differentiated TCV from cPTC. We started with the same gene list used for clustering, using only genes that were dysregulated in at least four samples and in which the absolute value of the log fold change was greater than or equal to log(3). We computed the P for the Wilcoxon Rank-Sum test (also called the Mann–Whitney U test). We also examined the P computed with the standard t test on the log of the ratios. Significance was accepted at a two-tailed P < 0.05.

**Quantitative Real-Time PCR MUC1 Copy Number Analysis.** Quantitative real-time PCR MUC1 copy number analysis was performed with the iCycler System (Bio-Rad Laboratories, Hercules, CA) using SYBR Green A.8 http://sequence.acom.yu.edu/biosint/function genomic.html.

---

Table 1 Comparison of clinical characteristics of conventional and tall-cell variant papillary thyroid carcinomas at a median follow-up of 55 months

<table>
<thead>
<tr>
<th>Clinical factor</th>
<th>cPTC (n = 50)</th>
<th>TCV (n = 50)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex, n (%)</td>
<td>31 (62%)</td>
<td>30 (60%)</td>
<td>NS</td>
</tr>
<tr>
<td>Age &gt; 45 years, n (%)</td>
<td>19 (38%)</td>
<td>39 (78%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Extrathyroidal extension, n (%)</td>
<td>15 (30%)</td>
<td>48 (96%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Tumor size &gt; 4.0 cm, n (%)</td>
<td>6 (12%)</td>
<td>10 (20%)</td>
<td>NS</td>
</tr>
<tr>
<td>Distant metastasis, n (%)</td>
<td>0 (0%)</td>
<td>7 (14%)</td>
<td>0.012</td>
</tr>
<tr>
<td>Treatment (less than total thyroidectomy), n (%)</td>
<td>15 (30%)</td>
<td>7 (14%)</td>
<td>NS</td>
</tr>
<tr>
<td>5-year disease-free survival, n (%)</td>
<td>46 (92%)</td>
<td>39 (78%)</td>
<td>NS</td>
</tr>
<tr>
<td>5-year cause-specific survival, n (%)</td>
<td>49 (98%)</td>
<td>47 (94%)</td>
<td>NS</td>
</tr>
<tr>
<td>Overall survival, n (%)</td>
<td>49 (98%)</td>
<td>46 (92%)</td>
<td>NS</td>
</tr>
</tbody>
</table>

* cPTC, conventional papillary thyroid carcinoma; TCV, tall-cell carcinoma; NS, not significant.
MUC1 predicts outcome in PTC

The cellular localization (membranous, cytoplasmic, and intensity of tumor cells staining). The staining characteristics of the normal thyroid were used as negative controls. The expression of MUC1 antigen under study was used as positive control. Arrayed recombinant monoclonal antibodies were performed with a Ventana automated stainer according to the manufacturer’s recommendations (Ventana Medical Systems, Tucson, AZ). Diaminobenzidine was used as the chromogen and hematoxylin as the nuclear counterstain.

Five-μm sections were cut from the tissue array blocks and placed on charged polylysine-coated slides. Representative sections were stained with H&E and analyzed to confirm the presence of the desired antigen. Corresponding sections were subjected to immunohistochemical analysis. Sections from tissue arrays were deparaffinized, rehydrated in graded alcohols, and processed according to the avidin-biotin immunoperoxidase method. Immunostaining was performed with a Ventana automated stainer according to the manufacturer’s recommendations (Ventana Medical Systems, Tucson, AZ). Diaminobenzidine was used as the chromogen and hematoxylin as the nuclear counterstain.

Five-μm sections were cut from the tissue array blocks and placed on charged polylysine-coated slides. Representative sections were stained with H&E and analyzed to confirm the presence of the desired antigen. Corresponding sections were subjected to immunohistochemical analysis. Sections from tissue arrays were deparaffinized, rehydrated in graded alcohols, and processed according to the avidin-biotin immunoperoxidase method. Immunostaining was performed with a Ventana automated stainer according to the manufacturer’s recommendations (Ventana Medical Systems, Tucson, AZ). Diaminobenzidine was used as the chromogen and hematoxylin as the nuclear counterstain.

Comparison of Global Gene Expression Profiles of TCV and cPTC. We compared the global gene expression profiles of TCV and cPTC. To assess the global genomic composition of cPTC and TCV, CGH analysis was performed on 25 randomly selected cases of TCV and 45 cases of cPTC. This analysis revealed significant quantitative and qualitative differences in DNA copy number changes between TCV and cPTC. At least one CGH-detectable chromosomal aberration was present in 76% of TCVs compared with 27% of cPTCs (P < 0.0001), indirectly reflecting differences in the presence of genomic instability in these tumors (17). However, the degree of chromosomal instability was similar in cases with an abnormality, with medians of 3 and 2.5 contiguous chromosomal abnormalities in TCV and cPTC, respectively.

Comparison of Global Gene Expression Profiles of TCV and cPTC. We compared the global gene expression profiles of TCV (n = 9) and cPTC (n = 7), using cDNA microarrays containing 17,840 genes or expressed sequence tags. Cases were randomly selected from those analyzed by CGH based on availability of fresh-frozen tissue and the quality of the extracted total RNA. One cPTC case and six TCV cases demonstrated CGH-detectable chromosomal abnormalities, representative of the overall analyzed population. On the basis of criteria detailed in the methods section, we defined 448 genes (2%) dysregulated in these tumors (supplementary data). In concordance with the higher prevalence of genetic instability in TCV, we found a higher frequency of gene dysregulation in TCV relative to cPTC. However, hierarchical clustering based on the expression profile of the 448 genes did not segregate cPTC and TCV (Fig. 2A). Further analysis (see “Materials and Methods”) identified 82 genes that maximally differentiated TCV and cPTC (Fig. 2B; P < 0.05). The vast majority of these genes were dysregulated in TCV and had established oncogenic potential as determined by functional annotation analysis (supplementary data). Genes participating in known oncogenic pathways included BCL2 (apoptosis), EID-1 (retinoblastoma pathway), TFDPI1 (transforming growth factor-β pathway), PRDX5 (mitogen-activated protein kinase pathway), MAPK6 (mitogen-activated protein kinase pathway), FOSB, and JUN (epidermal growth factor signaling pathway). In addition, several genes were associated with oncogenic functions, such as cell adhesion, cell-extracellular matrix interaction, and cell migration, including TM4SF2, MUC1, FAT tumor suppressor homolog, SPARCL2, GPR30, CD36, SDPR, FN14, CD55, and TNC. Finally, the analysis also identified several genes functioning as transcription factors, including NUCB2, TF3, OLG2, EGR2, and ID4.

RESULTS

Comparison of DNA Copy Number Changes in cPTC and TCV. To compare the global genomic composition of cPTC and TCV, CGH analysis was performed on 25 randomly selected cases of TCV and 45 cases of cPTC. This analysis revealed significant quantitative and qualitative differences in DNA copy number changes between TCV and cPTC. At least one CGH-detectable chromosomal aberration was present in 76% of TCVs compared with 27% of cPTCs (P < 0.0001), indirectly reflecting differences in the presence of genomic instability in these tumors (17). However, the degree of chromosomal instability was similar in cases with an abnormality, with medians of 3 and 2.5 contiguous chromosomal abnormalities in TCV and cPTC, respectively.

Analysis of the overall pattern of abnormalities revealed several differences between TCV and cPTC, including gains of 1p34–36, 9q11–22, 3q43, 17q25, 19, and 22 and losses of 2q21–21; 4q11–12, 6q11–22, 8q11–22, 9q11–12, and 13q21–31 that were unique to TCV and gains of 2q21–24, 4q11–26, 5q14–21, 6q11–22, 8q21–23, and 13q21–31 and losses of 1p33–36, 9q34, 16q22–24, 17q22–25, 19p, and 22 that were exclusively detected in cPTC (Fig. 1). Gain and high-level amplification of 1q21 was the most frequent abnormality differentiating TCV from cPTC, occurring in 40 and 0% of cases, respectively (P < 0.0001).

Comparison of Global Gene Expression Profiles of TCV and cPTC. We compared the global gene expression profiles of TCV (n = 9) and cPTC (n = 7), using cDNA microarrays containing 17,840 genes or expressed sequence tags. Cases were randomly selected from those analyzed by CGH based on availability of fresh-frozen tissue and the quality of the extracted total RNA. One cPTC case and six TCV cases demonstrated CGH-detectable chromosomal abnormalities, representative of the overall analyzed population. On the basis of criteria detailed in the methods section, we defined 448 genes (2%) dysregulated in these tumors (supplementary data). In concordance with the higher prevalence of genetic instability in TCV, we found a higher frequency of gene dysregulation in TCV relative to cPTC. However, hierarchical clustering based on the expression profile of the 448 genes did not segregate cPTC and TCV (Fig. 2A). Further analysis (see “Materials and Methods”) identified 82 genes that maximally differentiated TCV and cPTC (Fig. 2B; P < 0.05). The vast majority of these genes were dysregulated in TCV and had established oncogenic potential as determined by functional annotation analysis (supplementary data). Genes participating in known oncogenic pathways included BCL2 (apoptosis), EID-1 (retinoblastoma pathway), TFDPI1 (transforming growth factor-β pathway), PRDX5 (mitogen-activated protein kinase pathway), MAPK6 (mitogen-activated protein kinase pathway), FOSB, and JUN (epidermal growth factor signaling pathway). In addition, several genes were associated with oncogenic functions, such as cell adhesion, cell-extracellular matrix interaction, and cell migration, including TM4SF2, MUC1, FAT tumor suppressor homolog, SPARCL2, GPR30, CD36, SDPR, FN14, CD55, and TNC. Finally, the analysis also identified several genes functioning as transcription factors, including NUCB2, TF3, OLG2, EGR2, and ID4.
Comparison of CGH and cDNA Microarray Findings. We correlated genomic aberrations identified by CGH with the 82 gene targets maximally segregating TCV from cPTC, based on chromosomal localization on the Human Genome Draft Sequence. Overall, 41 (50%) dysregulated genes were located within loci having CGH-detectable aberrations (Fig. 2B). We focused our analysis on the 1q21 region because its amplification was the most significant separator of TCV and cPTC based on CGH analysis (Fig. 1).

**MUC1** was the only gene dysregulated at the 1q21 locus and was one of the most significant genes segregating TCV from cPTC. **MUC1** was uniquely up-regulated in TCV, occurring in seven of nine (78%) cases with a range of 3.1–7.5-fold ($P_{/H11005} 0.005$). Cases of TCV with 1q gain identified by CGH had the highest levels of up-regulation (mean, 7.35-fold increase).

**Confirmation of MUC1 Amplification by Quantitative Real-Time PCR.** We confirmed the presence of MUC1 amplification by analysis of randomly selected cases of TCV ($n = 9$) and cPTC ($n = 14$), using real-time quantitative PCR. In TCV, copy number ratios ranged from 0.86 to 5.76 with a median of 2.27 (4–5 copies of the MUC1 gene) compared with 0.49–1.24 with a median of 0.94 for PTCs ($P < 0.001$; Fig. 3). Increased MUC1 copy number, defined based on previously validated cutoff ratios as a copy number ratio $\geq 1.8$ (34), was detected in 6 of 9 TCVs and none of the 14 cPTC ($P = 0.001$), correlating with the results of CGH analysis for these tumors.

**Validation of MUC1 Up-Regulation by Immunohistochemical Analysis.** We validated the cDNA findings by assessing the frequency and distribution of MUC1 up-regulation in cPTC ($n = 50$) and TCV ($n = 50$) by use of immunohistochemistry. On the basis of inherent tissue loss associated with tissue microarray analysis, 74 tumors with matched normal tissue were eligible for final analysis of MUC1 protein levels and localization by immunohistochemistry. **MUC1** expression was identified in 100% of both normal and tumor tissues. However, the degree of expression and cellular localization varied significantly between tumor and normal tissue and among tumor samples. In normal thyroid tissue, MUC1 was predominantly localized in the apical cell membrane. In tumor tissues, MUC1 expression was localized in both the apical cell membrane and cytoplasm at equal intensity (Fig. 4). All correlative analyses were based on cytoplasmic staining because this was uncommon in normal thyroid tissues and has previously been associated with aggressive behavior in breast and pancreatic cancer (35, 36). Our immunohisto-

---

**Fig. 1.** Ideogram showing DNA copy number changes identified by comparative genome hybridization in conventional papillary thyroid carcinomas (A) and tall-cell variant papillary thyroid carcinomas (B). Thin vertical lines on either side of the chromosomes indicate losses (left) and gains (right) of the chromosomal region. The high-level amplified chromosomal regions are represented by thick lines.
chemical analysis corroborated the cDNA microarray findings, with *MUC1* up-regulation present in 97.5% of TCVs compared with only 35% of conventional cPTCs ($P < 0.0001$).

**Prognostic Significance of MUC1 Up-Regulation in PTC.** To assess the potential clinical implications of *MUC1* up-regulation, we correlated *MUC1* status with relapse-free survival. The median follow-up time in this group was 53.7 months (range, 1–186.5 months), with a median relapse-free interval of 46.9 months (range, 1–156 months). The 5-year overall survival was 98.3%; the 5-year relapse-free survival was 84%. *MUC1* overexpression significantly correlated with decreased relapse-free survival in univariate analysis ($P = 0.02$). Multivariate analysis showed that this association was independent of histology and established clinical predictors ($P = 0.03$; Table 2).

**Validation of MUC1 as a Prognosticator of Outcome in cPTC.** To confirm the clinical significance of *MUC1* overexpression in cPTC, we analyzed protein expression by immunohistochemistry in an independent group of histologically confirmed cPTCs (see “Materials and Methods”). *MUC1* up-regulation was present in 3 of the 29 (10.3%) patients in this validation series (7 cases from the match group were lost during processing of the tissue array). *MUC1* overexpression was identified exclusively in patients with treatment failure and was associated with significantly poorer relapse-free survival ($P < 0.0001$; Fig. 5). This difference was even more pronounced when we analyzed patients dying as a consequence of disease, with the 3 cPTC cases having *MUC1* overexpression dying 1, 3, and 4 years after primary treatment. Given the small sample size, we pooled the cases of cPTC from the validation series ($n = 29$) and the...
Fig. 2 Continued.

MUC1 PREDICTS OUTCOME IN PTC
screening series \((n = 34; 16\) lost to processing) to allow assessment of the individual contribution of \(MUC1\) up-regulation to the outcome in cPTC by multivariate analysis. The median follow-up time for this group \((n = 63)\) was 77.9 months with a median relapse-free interval of 72.3 months. \(MUC1\) overexpression \((P = 0.03)\), the size of the primary tumor \((P = 0.03)\), and extrathyroidal extension \((P = 0.04)\) correlated with decreased relapse-free survival on univariate analysis, but only \(MUC1\) overexpression remained a significant predictor after multivariate analysis (relative risk, 2.3; 95\% confidence interval, 1.1–5.5; \(P = 0.03\); Table 3).

**DISCUSSION**

PTCs account for 80\% of thyroid malignancies and are characterized by slow growth and an excellent overall prognosis (37). However, 10–15\% of cases can exhibit aggressive behavior, hallmarked by local invasion, distant metastasis, treatment resistance, and mortality (38). As a result of the wide divergence in clinical behavior, the optimal management of PTC is dependent on the assessment of the malignant potential of individual tumors at presentation. Although several clinicopathological variables have been identified for this purpose, none consistently identifies patients at risk for poor outcome (15). Molecular factors underlying aggressive PTC behavior may represent more accurate outcome predictors and potential therapeutic targets (39–41). However, the identification of these markers has been limited by the long natural history of PTC combined with a rudimentary knowledge of its genetic composition. In this study we took advantage of differences in clinical behavior of two distinct variants of PTC, TCV and cPTC, to identify molecular prognosticators of outcome, using complementary genome-wide analyses. CGH

**Table 2**

<table>
<thead>
<tr>
<th>Prognostic factor</th>
<th>Variable</th>
<th>Patients ((n))*</th>
<th>5-year relapse-free survival (%)</th>
<th>(P), univariate analysis (log-rank test)</th>
<th>Multivariate analysis relative risk (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>(&lt;45) years</td>
<td>29</td>
<td>89.5</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\geq 45) years</td>
<td>45</td>
<td>80.2</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>28</td>
<td>84.3</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>46</td>
<td>84.5</td>
<td>0.06</td>
<td>1</td>
</tr>
<tr>
<td>Histological type</td>
<td>Classic</td>
<td>34</td>
<td>90.4</td>
<td>0.06</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TCV</td>
<td>40</td>
<td>75.9</td>
<td>0.2</td>
<td>1.08 (0.4–3.5); (P = 0.8)</td>
</tr>
<tr>
<td>Size of primary tumor</td>
<td>(\leq 4) cm</td>
<td>51</td>
<td>86.8</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;4 cm</td>
<td>11</td>
<td>79.5</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Extrathyroid extension of primary tumor</td>
<td>Absent</td>
<td>36</td>
<td>90.9</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>37</td>
<td>80.7</td>
<td>0.02</td>
<td>3333.3 (1.1–\infty); (P = 0.03)</td>
</tr>
<tr>
<td>(MUC1) expression</td>
<td>Baseline</td>
<td>23</td>
<td>100</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Overexpressed</td>
<td>51</td>
<td>74.2</td>
<td>0.02</td>
<td>3333.3 (1.1–\infty); (P = 0.03)</td>
</tr>
</tbody>
</table>

*Number differences reflect missing data.
analysis demonstrated that TCV is characterized by a significantly higher prevalence of genomic instability relative to cPTC. We have previously shown that genomic instability increases sequentially from the indolent well-differentiated to the aggressive poorly differentiated (PDTC) and anaplastic thyroid carcinomas (ATC), suggesting that the development of DNA copy number abnormalities is associated with progression and clinically aggressive behavior in thyroid cancer (42).

Extending this concept, the present data suggest that TCV has an intermediate position in the progression from cPTC to PDTC and ATC. TCV also differed significantly from cPTC in individual chromosomal abnormalities. Abnormalities uniquely detected in TCV, including chromosomal gains of 1p34, 9q34, 11q13, 17q25, 19, and 22 and losses of 6q11–22 and 13q21–31, are also commonly present in PDTC and ATC (42). Accordingly, these changes may underlie the observed differences in clinical behavior between TCV and PTC.

Overall, our gene expression analysis of PTC confirms previous findings by Huang et al. (9) showing a highly consistent gene expression profile despite the clinicopathological heterogeneity of PTC. We found significant overlap in the genes identified in our array analysis and those reported by Huang et al. These genes include CITED1, Syndecan-4, galectin-3, fibronectin 1, MTIG, TFF3, DPT, CRABP1, and DIO1. Dysregulation of these genes may be induced early in thyroid tumorigenesis; however, their exact functional role remains unclear.

In contrast to the CGH data, the cDNA microarray analysis suggested that TCV and PTC do not differ significantly in global gene expression profiles. The apparent differences between the CGH and cDNA microarray analyses may be explained by recent observations that changes in DNA copy number induce changes in the expression of only a limited number of genes, which is highlighted by the rarity of overexpressed genes in regions of amplification (2, 4, 5). Accordingly, chromosomally derived changes in DNA copy number have minimal impact on the overall expression profile. Several other factors may also explain the observed differences between the CGH and cDNA microarray findings, including the limited power to differentiate expression patterns derived from the number of cases we analyzed by cDNA microarray analysis and/or the lack of sensitivity of statistical approaches used to evaluate the array data.

A focused analysis of gene expression changes revealed that 82 of the 448 genes differentiated TCV and cPTC. In concordance with the CGH findings, cDNA microarray analysis demonstrated that a higher proportion of these genes were dysregulated in TCV. The majority of these genes have not been associated previously with thyroid cancer pathogenesis. Analysis of their functional annotation revealed several linked with cancer pathogenesis, including LCN2 (acute myelogenous leukemia; Ref. 43), HIP116 (colon cancer; Ref. 44), TCF3 (leukemia; Ref. 45), GAL53 (breast cancer; Ref. 46), CYR61 (breast cancer; Ref. 47), LAMC2 (head and neck squamous cell carcinoma and pancreatic adenocarcinoma; Refs. 48, 49), and TNC (cervical carcinoma, glioma, and non-small cell lung cancer; Refs. 50–52). In addition, we found several genes with oncogenic activities, including effects on cell adhesion and differentiation (see the "Results"). Therefore, it is conceivable that these genes may contribute to observed differences in clinical behavior between cPTC and TCV.

Several genes differentiating TCV from PTC identified by cDNA microarray analysis reside within loci of genomic aberrations identified by CGH (Fig. 2B). Many of these genes merit further consideration because they may be linked to oncogenesis and aggressive clinical behavior in patients with PTC (53). For example, LCN2, a gene involved in apoptosis, may drive selection for 9q34 amplification, an aberration also present in PDTC and ATC (42). Several genes with putative tumor-suppressive function residing on a region of recurrent loss on chromosome 4 were also identified, including FAT tumor suppressor homolog 2, SPARCL1, and CPE. Of the multiple putative prognostic targets identified in this study, we chose to further characterize the role of MUC1 because overrepresentation of 1q21 and MUC1 overexpression were the most common abnormalities differentiating TCV from cPTC. In addition, 1q21 amplification was previously associated with aggressive behavior in PTC, PDTC, and ATC, but no gene targets have been identified to date (11, 30, 42, 54). MUC1 constitutes a logical candidate for the 1q amplification for several reasons: (a) it was recently shown that MUC1 plays an

![Fig. 5. Kaplan–Meier analysis of the impact of MUC1 up-regulation on disease-free survival of patients with conventional papillary thyroid carcinoma. Twenty-nine patients (12 cases that failed treatment and 17 cured cases) were stratified according to MUC1 expression status. All cases with MUC1 up-regulation (n = 3) died of disease within 1 year after primary treatment. Of cases without MUC1 expression status. All cases with MUC1 (12 cases that failed treatment and 17 cured cases) were stratified according to including chromosomal gains of 1p34 and chromosomal abnormalities. Abnormalities uniquely detected in TCV, ATC. TCV also differed significantly from cPTC in individual chromosomal abnormalities. Abnormalities uniquely detected in TCV, including chromosomal gains of 1p34–36, 1q21, 6p21–22, 9q34, 11q13, 17q25, 19, and 22 and losses of 6q11–22 and 13q21–31, are also commonly present in PDTC and ATC (42). Accordingly, these changes may underlie the observed differences in clinical behavior between TCV and PTC.

Table 3  Factors influencing relapse-free survival in 63 patients with classic papillary thyroid carcinoma

<table>
<thead>
<tr>
<th>Prognostic factor</th>
<th>Variable</th>
<th>Patients (n)</th>
<th>5-Year relapse-free survival (%)</th>
<th>P, univariate analysis (log-rank test)</th>
<th>Multivariate analysis relative risk (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>&lt;45 years</td>
<td>30</td>
<td>85</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>≥45 years</td>
<td>33</td>
<td>80.3</td>
<td>0.4</td>
<td>0.5 (0.2–1.1); P = 0.1</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>25</td>
<td>73.7</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>38</td>
<td>88.4</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td>Size of primary tumor</td>
<td>≤4 cm</td>
<td>45</td>
<td>91.6</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>&gt;4 cm</td>
<td>10</td>
<td>70</td>
<td>1.0</td>
<td>0.5 (0.3–2.5); P = 0.2</td>
</tr>
<tr>
<td>Extrathyroid extension of primary tumor</td>
<td>Absent</td>
<td>36</td>
<td>86.6</td>
<td>0.04</td>
<td>2.1 (0.9–5.8); P = 0.06</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>27</td>
<td>6.9</td>
<td>0.04</td>
<td>1</td>
</tr>
<tr>
<td>MUC1 expression</td>
<td>Baseline</td>
<td>48</td>
<td>89.1</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Overexpressed</td>
<td>15</td>
<td>57.1</td>
<td>1.0</td>
<td>2.3 (1.1–5.5); P = 0.03</td>
</tr>
</tbody>
</table>

*Number differences reflect missing data.
oncogenic role and is up-regulated through amplification in aggressive B-cell lymphomas and breast cancer (55, 56); (b) up-regulation of MUC1 appears to be specifically associated with adenocarcinomas from various sources correlated with increased metastatic potential and poor prognosis of these tumors (33, 57); (c) MUC1 interferes with cell-cell and cell-matrix interactions, thereby promoting cellular disociation and oncogenic progression (58); and (d) MUC1 up-regulation may inhibit human T-cell proliferation, thereby contributing to cancer-propagated immunosuppression (59). Although MUC1 overexpression has been previously described in PTC, an association with aggressive tumor behavior has not been established (9, 60). All cases with MUC1 amplification also demonstrated MUC1 up-regulation in our study group, suggesting that MUC1 may drive selection for 1q amplification. MUC1 up-regulation was also identified in cases with normal MUC1 copy number. In this context, it is of interest that genomic rearrangement and hypomethylation of MUC1 have been associated with MUC1 up-regulation in B-cell lymphoma and breast cancer, respectively (56, 61). The most important evidence suggesting an association between MUC1 deregulation and aggressive behavior in thyroid cancer comes from correlation between overexpression and outcome in PTC, even after controlling for confounding variables. In our PTC tissue array, all patients with MUC1 overexpression died as a consequence of cancer, failing treatment within 1 year of primary intervention and dying of disease within 4 years of presentation. MUC1 status did not identify all cases with treatment failure, likely reflecting multifactorial influences on tumor behavior.

Our findings may have several biological and clinical implications. A diagnosis of TCV, which accounts for 10–15% of PTC, may have significant therapeutic consequences. However, several studies, as well as our own institutional experience, demonstrate that TCV is often misdiagnosed on routine histopathology (13). The detection of MUC1 up-regulation in virtually all (97.5%) cases of TCV indicates that it may be used as an adjunct to aid in diagnostic classification.

More importantly, the finding that MUC1 is an independent prognosticator in PTC suggests that this marker may also be used to refine prognostic stratification. In this context, preoperative evaluation of MUC1 status in fine-needle aspiration biopsy samples is feasible, and its value should be investigated further (62). In addition to these diagnostic and prognostic possibilities, MUC1 up-regulation may represent a target for therapeutic intervention in aggressive PTC. At present, MUC1 is being studied as a target for immunotherapy of human adenocarcinomas from various sources, and several MUC1-based vaccines have been tested in Phase I clinical trials (63, 64).

In summary, we identified several candidate genes that may underlie aggressive thyroid cancer. We showed that MUC1 may drive selection of 1q amplification in PTC and independently associated gene overexpression with aggressive behavior. These findings suggest that MUC1 aberrations should be considered further to delineate its role in clinical management. In addition, further research into the biological mechanisms underlying MUC1 activity in thyroid cancer is warranted.

REFERENCES


Genome-Wide Profiling of Papillary Thyroid Cancer Identifies MUC1 as an Independent Prognostic Marker

Volkert B. Wreesmann, Elizabeth M. Sieczka, Nicholas D. Socci, et al.