Molecular Profiling of Matched Samples Identifies Biomarkers of Papillary Thyroid Carcinoma Lymph Node Metastasis

Janete M. Cerutti,1,2 Gisele Oler,2 Pedro Michaluart, Jr.,1 Rosana Delcelo,3 Robert M. Beaty,1 Jennifer Shoemaker,1 and Gregory J. Riggins1

1Department of Neurosurgery, Johns Hopkins University Medical School, Baltimore, Maryland; 2Genetic Bases of Thyroid Tumors Laboratory, Division of Genetics, Department of Pathology, Federal University of São Paulo, and 3Division of Head and Neck Surgery, Department of Surgery, University of São Paulo Medical School, São Paulo, SP, Brazil; and 4Division of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina

Abstract

Biomarkers of papillary thyroid carcinoma (PTC) metastasis can accurately identify metastatic cells and aggressive tumor behavior. To find new markers, serial analysis of gene expression (SAGE) was done on three samples from the same patient: normal thyroid tissue, primary PTC, and a PTC lymph node metastasis. This genomewide expression analysis identified 31 genes expressed in lymph node metastasis, but not in the primary tumor. Eleven genes were evaluated by quantitative real-time reverse transcription-PCR (qPCR) on independent sets of matched samples to find genes that were consistently different between the tumor and metastatic samples. LIMD2 and PTPRC (CD45) showed a statistically significant difference in expression between tumor and metastatic samples (P < 0.0045), and an additional gene (LTB) had borderline significance. PTPRC and LTB were tested by immunohistochemistry in an independent set of paired samples, with both markers showing a difference in protein expression. All 20 metastases from 6 patients showed expression in both markers, with little or no expression in primary tumor. Some of these markers could provide an improved means to detect metastatic PTC cells during initial staging of a newly diagnosed carcinoma and/or to rule out recurrence. The functional role of these genes may also provide insight into mechanisms of thyroid cancer metastasis.

Introduction

Papillary thyroid carcinoma (PTC) is the most common thyroid cancer, accounting for about 80% of all thyroid cancers. Although PTCs are usually curable with standard surgical and adjuvant radioiodine treatment, neck lymph node metastases (LNM) are found in 30% to 65% of cases at initial diagnosis (1–3). Unfortunately, about 15% of cases with LNMs also display a very aggressive behavior, characterized by local invasion, distant metastasis, treatment resistance, and increased mortality. Due to this clinical heterogeneity, the management of PTC is often controversial and depends on the detection of distant metastasis (2–4). An earlier and more accurate detection of the aggressive PTC phenotype would help improve individualized treatment for this cancer.

Several groups have done expression profile of metastatic primary PTC or in vitro model systems to identify clinical outcome markers (5, 6). Although these approaches could identify useful prognostic markers, they do not directly identify gene expression changes that occur in the metastatic cells. Identification of genes that are consistently expressed in metastatic PTC cells could yield useful biomarkers. These markers have potential prognostic use and may also help identify occult metastatic cells in lymph node biopsies. Eventually, they could also be investigated as serum markers. In addition to the practical uses as markers, genes associated with metastasis could help reveal the molecular mechanisms of the metastatic process.

To identify gene expression changes that occur subsequent to thyroid LNM, we did SAGE on matched normal thyroid (NT), primary PTC, and its LNM and in a normal lymph node. Serial analysis of gene expression (SAGE; ref. 7) was employed because of its ability to accurately produce comprehensive expression profiles from small samples and because there are archived databases of SAGE expression profiles of human tissues freely available for comparison.

To our knowledge, this is the first gene expression comparison on matched NT, primary PTC, and metastasis samples. We identified transcripts exclusively expressed in a LNM library and, therefore, potentially related to the metastatic process of PTC. The expression of the selected transcripts were investigated in a series of matched-normal, primary tumor and LNMs by real-time reverse transcription-PCR (qPCR). The transcripts that were found consistently overexpressed in LNMs were evaluated by immunohistochemistry in an independent set of paired samples for confirmation. The markers located in this study may have eventual utility for better prediction and detection of PTC metastasis.

Materials and Methods

Generation of SAGE libraries. Matched tissues of a NT, a PTC, and its LNM were chosen for SAGE (Table 1, case 1). This matched set was chosen in part because the sample quality was high, measured by the high percentage of tumor cells observed by H&E histopathology done on frozen sections from primary and metastatic tumors. The primary sample was from the tumor core, in an attempt to avoid the capsule and surrounding normal tissue. To eliminate the expression of normal lymph node cells without any metastasis, a SAGE library was generated from a normal lymph node (NL; Stratagene). The libraries were constructed using NlaIII as the anchoring enzyme as described in the original SAGE procedure (7), and the ditag containing plasmid inserts were sequenced through the SAGE portion of the Cancer Genome Anatomy Project (8). Tags

Requests for reprints: Janete Cerutti, Rua Pedro de Toledo 781, 12° andar., Federal University of São Paulo, 04039-032, São Paulo, SP, Brazil. Phone: 55-11-5081-5233; Fax: 55-11-5084-5231; E-mail: cerutti@endo@pesquisa.epm.br.

DOI: 10.1158/0008-5472.CAN-06-4771
were extracted from sequence text files and processed to remove duplicate ditags, linker sequences, and repetitive tags using SAGE 2002 software version 4.12.6,7

SAGE analysis. The metastasis-derived library was compared with primary tumor using SAGE 2000 software. Pairwise and Monte Carlo simulations were used to identify transcripts in which difference was statistically different at a \( P = 0.001 \). Transcripts that were overexpressed in the metastasis library when compared with primary tumor and, therefore, candidate genes as associated with metastasis process, were assessed in LNM and LN libraries.

Tissue samples for validation of candidate genes. To test our hypothesis, the metastasis candidate genes were first analyzed in a series of matched normal thyroid tissue, primary PTC, and LNM provided by the Hospital das Clínicas, Universidade de São Paulo. The neck dissection was done by a single surgeon (P.M.) during thyroidectomy procedure. All samples were frozen immediately after surgical biopsy and stored at \(-80^\circ C\). Tissue histology of H&E-stained, paraffin-embedded sections obtained from frozen samples was evaluated by an experienced pathologist (R.D.) from frozen samples was evaluated by an experienced pathologist (R.D.). About 1 \( \mu g \) of total RNA was treated with a DNase (Ambion) and was reverse transcribed to cDNA using Super-Script II Reverse Transcriptase kit with an oligo(dT)12–18 primer and 10 units of RNase inhibitor (Invitrogen Corp.). An aliquot of cDNA was used in 20 \( \mu L \) PCR reactions containing TaqMan Universal PCR Master mix, 10 \( \mu M \) each specific primer and FAM-labeled probes for the target genes or reference gene (QP-C) and VIC-labeled probe for the second reference gene (RBS; TaqMan Gene Assays on Demand; Applied Biosystems). Quantitative PCR reactions were done in triplicate, the threshold cycle (\( C_t \)) was obtained using the Applied Biosystem software and were averaged (SD \( \leq 1 \)). Relative expression levels were calculated according to the formula \( 2^{(C_t - C_{ref})} \times 2^{(C_t - C_{ref})} \), where Ce is the \( C_t \) cycle number observed in the

Table 1. Summary of clinical data

<table>
<thead>
<tr>
<th>Cases</th>
<th>Diagnosis</th>
<th>Sex</th>
<th>Age at diagnosis (y)</th>
<th>Nodule size (cm)</th>
<th>Extrathyroidal extension</th>
<th>Metastasis at presentation</th>
<th>Metastasis (follow-up)</th>
<th>Persistent disease</th>
<th>Iodine</th>
<th>BRAF mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>First validation set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>PTC (^b)</td>
<td>F</td>
<td>19</td>
<td>3.5</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>V600E + K601del</td>
</tr>
<tr>
<td>2</td>
<td>FVPTC (^c)</td>
<td>F</td>
<td>76</td>
<td>1.0</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>V600E</td>
</tr>
<tr>
<td>3</td>
<td>PTC (^c)</td>
<td>M</td>
<td>41</td>
<td>1.5</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>V600E</td>
</tr>
<tr>
<td>4</td>
<td>PTC (^c)</td>
<td>F</td>
<td>55</td>
<td>2.5</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>FVPTC (^c)</td>
<td>F</td>
<td>52</td>
<td>5.0</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>PTC (^c)</td>
<td>F</td>
<td>29</td>
<td>3.5</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>V600E</td>
</tr>
<tr>
<td>7</td>
<td>PTC (^c)</td>
<td>F</td>
<td>39</td>
<td>NA</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>NA</td>
<td>V600E + K601del</td>
</tr>
<tr>
<td>8</td>
<td>FVPTC (^c)</td>
<td>F</td>
<td>49</td>
<td>2.0</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>9</td>
<td>PTC (^c)</td>
<td>F</td>
<td>49</td>
<td>2.2</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
</tr>
<tr>
<td>10</td>
<td>PTC (^c)</td>
<td>F</td>
<td>46</td>
<td>1.7</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>V600E</td>
</tr>
<tr>
<td>11</td>
<td>PTC (^c)</td>
<td>F</td>
<td>42</td>
<td>1.5</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>V600E</td>
</tr>
<tr>
<td>12</td>
<td>FVPTC (^c)</td>
<td>F</td>
<td>15</td>
<td>4.5</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N</td>
</tr>
<tr>
<td>Second validation set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>PTC (^c)</td>
<td>M</td>
<td>34</td>
<td>5.5</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>V600E + K601del</td>
</tr>
<tr>
<td>14</td>
<td>PTC (^c)</td>
<td>M</td>
<td>62</td>
<td>2.8</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
</tr>
<tr>
<td>15</td>
<td>PTC (^c)</td>
<td>M</td>
<td>34</td>
<td>0.7</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
</tr>
<tr>
<td>16</td>
<td>FVPTC (^c)</td>
<td>M</td>
<td>23</td>
<td>1.3</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
</tr>
<tr>
<td>17</td>
<td>PTC (^c)</td>
<td>F</td>
<td>19</td>
<td>3.5</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
</tr>
<tr>
<td>18</td>
<td>PTC (^c)</td>
<td>F</td>
<td>33</td>
<td>1.2</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
</tr>
</tbody>
</table>

Abbreviations: FVPTC, follicular variant of PTC; NA, not available. Y, yes; N, no.

\(^a\)Case 1 was chosen for SAGE.

\(^b\)Matched normal, tumor, and lymph node metastasis.

\(^c\)Matched normal and lymph node metastasis.

\(^d\)Matched tumor and lymph node metastasis.

\(^e\)Lymph node metastases obtained during reoperation for recurrence.

2 paired NT and LNM (Table 1, first validation set). Additionally, a normal lymph node was included as a negative control.

For further confirmation of metastasis genes, a second validation set of paraffin-embedded sections was obtained from six matched samples of normal lymph nodes, normal thyroid, primary tumor, and LNM (Table 1, second validation set). The follow-up of these patients showed that three out of six patients had regional recurrence (Table 1, cases 15–17). There were at least two lymph nodes with metastasis studied from each case in the second validation set. Additionally, in the three patients who recurred at least two more lymph nodes were studied from the recurrence. A total of 20 LNMs were investigated. A third set of samples included 15 primary classic and follicular variant of PTC (metastatic and nonmetastatic tumors, not shown in Table 2). The paraffin-embedded sections were selected from the archives of the Department of Pathology, Federal University of São Paulo. The study was approved by the Ethic and Research Committees from both universities and was conducted in accordance with the Declaration of Helsinki Principles.

RNA isolation, cDNA synthesis, and qPCR. Total RNA was isolated by TRIzol (Invitrogen Corp.). About 1 \( \mu g \) of total RNA was treated with a DNase (Ambion) and was reverse transcribed to cDNA using Super-Script II Reverse Transcriptase kit with an oligo(dT)12–18 primer and 10 units of RNase inhibitor (Invitrogen Corp.). An aliquot of cDNA was used in 20 \( \mu L \) PCR reactions containing TaqMan Universal PCR Master mix, 10 \( \mu M \) of each specific primer and FAM-labeled probes for the target genes or reference gene (QP-C), and VIC-labeled probe for the second reference gene (RBS; TaqMan Gene Assays on Demand; Applied Biosystems). Quantitative PCR reactions were done in triplicate, the threshold cycle (\( C_t \)) was obtained using the Applied Biosystem software and were averaged (SD \( \leq 1 \)). Relative expression levels were calculated according to the formula \( 2^{(C_t - C_{ref})} \times 2^{(C_t - C_{ref})} \), where Ce is the \( C_t \) cycle number observed in the

---


The full set of tag counts for all four libraries is available for downloading or analysis at the SAGE Genie Web site at [http://cgap.nci.nih.gov/SAGE](http://cgap.nci.nih.gov/SAGE) (9).
Experimental sample for controls, Genes involved in thyroid function

Genes in the experimental sample for the reference gene, Cn is the average of the cycle number observed in the normal thyroid tissues for control genes, and Rn is the average of the cycle number observed in the normal thyroid tissues for the reference gene (10, 11). The results obtained from relative expression levels were log transformed and used for statistical analysis. Representative results are shown in Fig. 1.

Statistical analysis. The first objective of this analysis was to determine if the expression values, as measured by qPCR, for 11 genes were different between tumor and metastases (n = 10 pairs) or between normal and metastases (n = 11 pairs), using paired data (Table 1). The comparison of the expression levels was carried out using a paired Student’s t test. To correct for multiple tests, a Bonferroni correction was used. A comparison was designated as statistically significant if the t statistic was found to be significant using an α level that had been adjusted (using a Bonferroni adjustment) to keep the family-wise error rate at 0.05. A one-sided test was used because of the expectation that the gene expression in the metastatic samples would be higher than the gene expression in the tumor or normal samples. The second objective was to determine whether a class predictor (tumor versus metastasis) could be developed using the qPCR data from these 11 genes. To investigate the development of an expression-based predictor that could be used to predict tumor or metastatic class, we followed the framework outlined by Radmacher et al. (12) using the compound covariate predictor for gene expression data (12, 13). The performance of the predictor was tested using leave-one-out cross-validation for all steps of the prediction procedure (i.e., selection of differentially express genes as well as creation of the prediction rule; refs. 12, 14). We assessed the significance of the performance of the predictor using the permutation-based test outlined by Radmacher et al. (12), in which the class labels were randomly permuted and the proportion of data sets that had a cross-validated error rate and is small as the error rate observed in the data set was calculated (12). The only difference in the permutation test from that used by Radmacher et al. (12) was that, because the data were paired in the present study, we used a paired permutation. In addition, because of the small size of the data set, we were able to generate the complete permutation distribution.

Table 2. Genes induced in LNMs selected for verification and normal thyroid-specific genes

<table>
<thead>
<tr>
<th>Tag sequence</th>
<th>Normal lymph node*</th>
<th>Normal thyroid*</th>
<th>Primary tumor*</th>
<th>LNMs*</th>
<th>Transcript description</th>
<th>Aliases</th>
<th>GenBank accession no.</th>
<th>Location</th>
<th>Gene ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTCAACAGTA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>ABC23, ATP-binding</td>
<td>MRP3</td>
<td>NM_005786</td>
<td>17q22</td>
<td>Transporter activity</td>
</tr>
<tr>
<td>GCAGTGGGAA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>LTB, lymphotoxin β</td>
<td>TNFC</td>
<td>NM_002341</td>
<td>6p21.3</td>
<td>Tumor necrosis factor receptor-binding activity</td>
</tr>
<tr>
<td>GTAGCGCCTC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>CST7, cystatin F</td>
<td>CMAP</td>
<td>BC015057</td>
<td>20p11.21</td>
<td>Cysteine protease inhibitor activity</td>
</tr>
<tr>
<td>TTAACTGTTG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>SYT2, synaptotagmin XII</td>
<td>SRG1</td>
<td>BC037406</td>
<td>11q13.2</td>
<td>Transporter activity</td>
</tr>
<tr>
<td>CTITTTTCCC</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>23</td>
<td>CD48, B-cell membrane protein</td>
<td>BLAST1</td>
<td>BC016182</td>
<td>1q21</td>
<td>GPI anchor binding activity</td>
</tr>
<tr>
<td>TTAATCCCA</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>29</td>
<td>PTPRC, protein tyrosine phosphatase, receptor type C</td>
<td>SLAMF2</td>
<td>NM_002838</td>
<td>1q31</td>
<td>Protein tyrosine phosphatase activity</td>
</tr>
<tr>
<td>AAAGCAAAA</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>23</td>
<td>PTPN4, protein tyrosine phosphatase, nonreceptor type 4</td>
<td>ICA, CD45</td>
<td>NM_002830</td>
<td>2q14.2</td>
<td>Protein tyrosine phosphatase activity</td>
</tr>
<tr>
<td>TTTCATAGA</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>LIMD2</td>
<td>MGC10986</td>
<td>BC004400</td>
<td>17q23.3</td>
<td>Metal ion binding</td>
</tr>
<tr>
<td>GAGGCCATCC</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>26</td>
<td>LSM7, U6 small rRNA associated</td>
<td>No</td>
<td>BC018621</td>
<td>19p13.3</td>
<td>RNA binding</td>
</tr>
<tr>
<td>CAATCAAAT</td>
<td>0</td>
<td>3</td>
<td>25</td>
<td>94</td>
<td>MET proto-oncogene</td>
<td>HGF, RCCP2</td>
<td>NM_00245.2</td>
<td>7q31</td>
<td>Protein tyrosine-kinase activity</td>
</tr>
<tr>
<td>CAGGCCACAC</td>
<td>4</td>
<td>17</td>
<td>18</td>
<td>72</td>
<td>S100A11, S100 calcium binding protein A11</td>
<td>Calgizzarin</td>
<td>BC001410</td>
<td>1q21</td>
<td>Calcium ion binding</td>
</tr>
</tbody>
</table>

Genes previously associated with papillary metastatic process

<table>
<thead>
<tr>
<th>Tag sequence</th>
<th>Normal lymph node*</th>
<th>Normal thyroid*</th>
<th>Primary tumor*</th>
<th>LNMs*</th>
<th>Transcript description</th>
<th>Aliases</th>
<th>GenBank accession no.</th>
<th>Location</th>
<th>Gene ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATGAAAT</td>
<td>0</td>
<td>75</td>
<td>0</td>
<td>0</td>
<td>TPO, thyroid peroxidase</td>
<td>No</td>
<td>M17755</td>
<td>2p25</td>
<td>Thyroid hormone generation</td>
</tr>
<tr>
<td>CGGGAAAGCA</td>
<td>0</td>
<td>134</td>
<td>16</td>
<td>57</td>
<td>TG, thyroglobulin</td>
<td>No</td>
<td>NM_003235</td>
<td>8q24.2</td>
<td>Thyroid hormone generation</td>
</tr>
<tr>
<td>ATGCTAAGAG</td>
<td>0</td>
<td>30</td>
<td>2</td>
<td>0</td>
<td>DI02, deiodinase, iodothyronine, type II</td>
<td>No</td>
<td>NM_000793</td>
<td>14q24.2</td>
<td>Thyroid hormone generation</td>
</tr>
</tbody>
</table>


*SAGE tag counts shown in each column refer to the abundance of SAGE tags in the libraries after normalization to 200,000 total tags.

†Transcript description refers to the gene name to which tag was attributed, according to the HUGO/GDB nomenclature committee–approved symbols.

‡Gene classification was by molecular function (http://cgap.ncbi.nlm.nih.gov/Genes/AllAboutGO).
Relative levels of expression determined by qPCR in matched samples of normal thyroid (white columns), PTC (gray columns), and LNMs (black columns; samples 1–9), matched normal and metastases (samples 10 and 11), matched tumor and metastases (sample 12), and normal lymph node (sample 13). Transcript levels were normalized to the average of RS8 and OP-C control genes, which were uniformly expressed in all three thyroid SAGE libraries. Numbers correspond to cases described in Table 1. The qPCR data were log 2 transformed for statistical analysis. The paired t test (normal versus metastases) showed that PTPRC, \( P = 0.001 \); LIMD2, \( P = 0.0016 \); LTB, \( P = 0.00096 \); and CD48, \( P = 0.00042 \). ABC23 (\( P = 0.0094 \)) was very close to significance. The paired t test (tumor versus metastases) showed that PTPRC (\( P = 0.00059 \)) and LIMD2 (\( P = 0.00056 \)) expression was significantly different at the 0.05 level.

**Immunohistochemical analysis.** Sections (3 \( \mu \)m) were deparaffinized and rehydrated through a graded series of ethanol. Endogenous peroxidase was quenched using a 3% solution of hydrogen peroxide in methanol for 30 min. Steaming retrieval was done in buffer AR-10 (BioGenex) for 10 min and then allowed to cool for 30 min. The sections were incubated with primary antibodies for at least 16 h at 4°C, followed by incubation with the labeled polymer DAKO EnVision+ System, HRP (DAKO Laboratories). Hematoxylin was used as the nuclear counterstain. Anti-CD45 (T29/33; DAKO Laboratories) was used at a dilution of 1:800, and anti-LTβ (FL244; Santa Cruz Biotechnologies) was used at a dilution of 1:50. The control for antibody specificity included incubation with rat immunoglobulin G used at the same concentration as the first antibody. Positive and negative controls were used in each run. All slides were scored in a blinded fashion, with immunopositivity evaluated semiquantitatively as follows: negative (−) when <10% of the cells were immunoreactive, and positive (+) when more than 10% of cells were immunoreactive. Immunohistochemical staining was evaluated independently by two investigators.

**Results**

**Analysis of SAGE data.** A total of 481,775 SAGE tags were obtained from four libraries, representing 160,951 unique transcript tags. The number of SAGE tags per library ranged from 99,911 (NL) to 143,689 (PTC). Tag numbers were normalized to 200,000 tags per library.

To identify genes potentially involved in metastasis, a comparison between primary tumor and LNMs libraries was done. Monte Carlo simulations yielded 498 tags statistically significant at a \( P \) value of ≤0.001 or 319 tags at a \( P \) value of ≤0.0001. A total of 31 of the 498 transcripts were highly expressed in the metastasis library and not expressed or expressed at low levels in primary tumor, whereas 47 transcripts were underexpressed in the metastasis library. To refine our analysis, the transcript expression of the metastasis-associated genes was assessed in the normal lymph node and normal thyroid libraries. Those transcripts with the greatest fold induction in metastasis library were chosen for validation by qPCR because they could not only help better understand the metastatic process, but also had potential as prognostic markers. Additionally, we chose three genes (MET, LSM7, and S100A11) whose expression was previously reported in primary tumors with metastatic potential, although expression within the metastasis was not reported (15–17). Table 2 lists the 11 transcripts selected for validation and their tag counts in our four SAGE libraries. For comparison, we show in Table 2 the transcript levels for those genes known to be involved in normal thyroid physiology. According to gene ontology databases, the differentially expressed genes are mainly involved in transport and cell signaling.

Among the transcripts highly expressed in LNMs library and not expressed in primary tumor were six transcripts (CXXC4, SNC73, PRSS1, CXCL13, STAT5A, and SDC1) that were previously associated with invasion and metastases in other cancers. Although not selected for validation by qPCR or immunohistochemistry, SAGE analysis suggested that their expression correlated with metastasis in this patient (\( P < 0.001 \)). However, further analysis will be needed to determine a role in LNM of PTC.

**Relative levels of gene expression for selected candidate genes.** We used qPCR to test the expression of the transcripts shown in Table 1 in matched sets of normal thyroid, primary PTC, and LNMs. This was the first validation set. We compared the results obtained from SAGE with the qPCR results for the samples used to generate NT, PTC, and LNMs libraries (case 1, Table 1 and Fig. 1). When the initial samples were used, the difference predicted by SAGE was confirmed for all 11 genes.

PTPRC was found overexpressed in all metastases analyzed when compared with paired normal thyroid and primary tumor (Fig. 1, samples 1–12) and normal lymph node (Fig. 1, sample 13). LIMD2 and CD48 were consistently expressed in the LNMs and were not expressed in the normal lymph node (Fig. 1, sample 13) and not expressed or expressed at very low levels in most of the matched normal thyroid tissues and/or primary tumors.

LTB and ABC3 were remarkably higher in all metastases and expressed at much lower levels in a number of primary tumors.
Predicted Markers for Lymph Node Metastasis

Although CST7 expression was elevated in the majority of paired primary tumor, its expression was markedly higher in the LNMs. Therefore, some genes analyzed here may start expression earlier than at the onset of metastasis, but still prove useful as markers of metastasis.

LSM7 and SYT12 showed a similar level of expression in normal thyroid, primary tumor, and LNM in most samples tested. Contrary results were obtained for PTPN4. The validation data showed that PTPN4 was underexpressed in most LNMs rather than overexpressed.

For those genes previously reported to be associated with metastasis, but not implicated in the late stages of metastasis by our SAGE data, qPCR data also showed a lack of overexpression in the metastases. For example, S100A11 was observed in LNMs, and its expression was higher in most primary tumors than in the matched metastases. MET was expressed in high levels in about 60% of LNMs analyzed and in intermediate levels in primary tumors from these cases. In the remaining cases, however, the expression level of MET was higher in primary tumor and normal thyroid tissue than in corresponding metastases. Although we did not find MET specific for metastasis, our results are in agreement with the literature where MET was found overexpressed in about 60% to 70% of metastatic primary PTC (18, 19).

Statistical analysis. In the first comparison PTPRC, LIMD2, LTB, and CD48 were found to be significantly different between the normal and metastatic samples (P value <0.0045). PTPRC and LIMD2 were found to be statistically significant between tumor and metastatic classes (P value <0.0045).

The second analysis determined whether a subset of these 11 genes could be used to predict metastasis versus tumor class. The class predictor used genes whose expression levels were declared significantly different at the 0.05 family-wise error level using the t test (P value <0.0045). The sample t statistics were used as weights in the compound covariate predictor (12). To evaluate the predictor, we used leave-one-out cross-validation: for each run, one tumor-metastatic pair of samples was left out, and the predictor developed on the remaining nine pairs of samples. The two left-out samples were predicted. We used all the steps of the prediction procedure, including the selection of differentially expressed genes, as well as the creation of the prediction rule (14). Using leave-one-out cross-validation, 4 of the 20 samples were misclassified for a prediction accuracy of 80% with a 95% two-sided confidence interval of 0.56, 0.94. To assess the significance of these prediction results, we implemented a permutation test. The proportion of random permutations that classified four or fewer misclassifications was 0.013. Thus, the results of the prediction analysis are statistically significant. PTPRC and LIMD2 were always selected in each step of the cross-validation procedure (10 out of 10 times, i.e., each time a pair of samples was left out).

Immunohistochemical analysis. To ascertain if our candidate metastasis-associated markers had increased protein levels in the LNM cells compared with other tissues, immunohistochemistry analysis was done in a second validation set of paired normal lymph node, normal thyroid, and primary tumor and LNMs (Table 1). Because there is no commercially available antibody for LIMD2, immunohistochemistry analysis was done for PTPRC (CD45) and LTB.

As shown by qPCR, PTPRC, also named CD45 or LCA, was highly expressed in LNMs and was not expressed in matched normal thyroid and primary tumors. In the metastatic lymph nodes, a strong staining was observed in lymphocytes rather than in the metastatic cancer cells. Of note, lymphocytes were the predominant cells in the metastatic lymph node, compared with normal lymph node, which may explain the strong brown staining observed (Fig. 2) and the qPCR results (Fig. 1). To ascertain that PTPRC was found expressed at lower levels in the normal lymph nodes, we additionally investigated the expression of PTPRC in six normal axillary lymph nodes. Although PTPRC expression was found in the lymphocytes in the normal lymph nodes, the number of lymphocytes with a positive expression of PTPRC was inferior to the metastatic lymph node (Fig. 2). We further assessed the expression of PTPRC in 15 primary PTCs (follicular variant and classic). Immunohistochemistry analysis revealed that 13 out of 15 PTCs were negative for PTPRC. In one case, however, tumor-infiltrating lymphocytes were positive. In the remaining case of PTC with a trabecular-insular area, a focal staining was observed in the epithelial cells.

qPCR data showed that LTB was highly expressed in most LNMs, although it was expressed at very low levels in a primary tumors. Because it was very close to the significance, we tested LTB expression by immunohistochemistry. In the second set of validation, LTB immunoreactivity was positive in all tumors cells within the lymph node in all 20 metastases analyzed but was not detected in any adjacent cells (Fig. 2). Additionally, LTB was negative in normal lymph nodes and matched normal thyroid tissue and primary tumors (Fig. 2). When LT was investigated by immunohistochemistry in 15 primary PTC, as suggested by qPCR, 3 out of 15 primary tumors showed a weak staining of LTB in the epithelial tumor cells. LT was negative in the surrounding normal cells. Interestingly, all positive tumors were highly invasive and had metastasis to the lymph nodes.

Discussion

PTC is the most frequent thyroid carcinoma. Lymph nodes are the most frequent site of PTC metastasis and, if found, predict recurrence and poor survival (20). LNMs have also been identified in the absence of clinically detectable primary tumor. Like most lethal cancers, death from PTC is from metastatic spread rather than invasion of the primary tumor. Therefore, it is of particular importance to quickly identify those patients with aggressive disease, so that the patient can be treated before metastatic spread.

Over the last decade, several biological markers tested in primary tumors have been explored for their value in predicting lymph node involvement in PTC (6, 21–23). Among all candidate prognostic markers, BRAF mutation is claimed as one of the most effective markers in predicting clinical outcome in classic PTC. However, BRAF mutation as a high-risk marker is controversial. First, a number of groups found an association between BRAF V600E mutation in PTC and high-risk tumor features (24, 25), and others found no association (26–28). Second, BRAF mutation is found in about 30% of classic variant of PTC. In the BRAF mutation-negative group of patients, which includes the remaining classic PTC and other variants, one cannot exclude lymph node involvement or distant metastases. Lastly, we and others have identified BRAF de novo mutations in the LNM cells that were absent in matched primary tumor (29, 30).

Thus far, no clinical markers associated with lymph node or distant metastases of thyroid metastasis have reached clinical practice. A question arising is why the reported prognostic markers have failed to appear in the clinic. One possible explanation is that most studies to date have attempted to make prediction from sets of primary tumors and have not employed an analysis of matched primary/metastatic pairs which can be difficult to obtain clinically.
Comparing a matched pair may help control against expression changes that are unrelated to the metastatic process.

To understand PTC metastasis at a molecular level and identify potential prognostic markers and therapeutic target, we obtained SAGE expression profiles of a normal lymph node and from the same patient: a normal thyroid, primary tumor, and LNM. This is the first report of gene expression profiling of matched samples in thyroid.

The comparison of the mRNA from these four tissues revealed 31 transcripts potentially associated with the metastasis. Eleven transcripts were tested by qPCR in a series of matched normal, primary tumor, and LNMs. LIMD2 and PTPRC were differentially expressed between metastasis and primary tumor. LIMD2, PTPRC, CD48, and LTB were significant between normal thyroid and metastasis. The markers validated here were positive not only in the LNMs from classic PTC, but also in the metastases from the follicular variant of PTC, independent of BRAF mutation status.

The present study shows, for the first time, that the expression of LIMD2 is associated with the metastatic process of PTC. LIMD2 derives its name from having two tandem copies of a LIM domain. In the LIM domain, there are seven conserved cysteine residues and a histidine sequence found in proteins that play important roles in a variety of fundamental biological processes, including cytoskeleton organization, cell lineage specification, and organ development, and in oncogenesis (31). Thus far, there are no data on the functional role of LIMD2 in metastasis. Although our findings suggest that LIMD2 is associated with metastases process, in vitro and in vivo analyses are necessary to test this hypothesis.

Several observations made in this study may be assembled to reveal a new pathway associated with the LNM of PTC. First, our qPCR results showed an association between PTPRC (CD45) overexpression and LNM. Immunohistochemistry confirmed that all LNMs express PTPRC, although its expression was found mainly in the surrounding immune cells. One might suggest that these findings may reflect immune modulation, rather than a role in the metastatic process. It has been suggested that inflammation associated with cancer development may be distinct from the normal inflammatory process on the basis of activation of the immune escape (32). Although the role of the adaptive immune response in controlling the growth and recurrence of human tumors has been controversial, several studies have shown the sustaining role of inflammatory mediators at distinct phases of malignant progression. It has been suggested that immune cells can release...
inflammatory mediators for invasion, migration, and metastasis (33), perhaps making the "soil fertile" for metastatic growth.

PTPRC is a prototype of the receptor-like protein tyrosine phosphatase that has been previously shown to play a significant role in CXCL12-induced and CXCR4-mediated chemotaxis through efficient coupling of CXCR4 with its signaling complex (34). It has been shown that CXCR4 may also play a role in the outgrowth of the carcinoma metastases in lymph nodes by chemotraction of the CXCR4-positive cells to the lymph nodes that highly express its ligand CXCL12 (35). Second, we have identified in the metastasis library an elevated expression of CXCR4. These results are in agreement with others where, in spite of tumor heterogeneity, most primary tumor cells were CXCR4 negative, with only a few CXCR4-positive cells that were destined to metastasize (35, 36). However, PTPRC plays a role in the outgrowth of carcinoma metastases in lymph nodes by facilitating the growth of CXCR4-expressing tumor cells, but further investigation of this hypothesis is required. Third, an additional indication that this signaling pathway may play a crucial role in determining the metastatic phenotype is the fact that STAT5A, a member of the STAT family activated by CXCR4, was found highly expressed in our LNM library. Interestingly, it has been suggested that thyroid tumors with RET/PTC1 rearrangement, a hallmark of PTC, use the CXCR4/CXCL12 receptor-ligand pathway to proliferate, survive, and migrate (33, 37).

There were also other genes we implicated in PTC metastasis. Although LTB mRNA expression was initially of borderline statistical significance, we also looked at protein expression. Immunohistochemistry showed expression from the epithelial cells within the metastatic lesion in all metastases analyzed. Although LTB was found expressed in three invasive and metastatic primary tumors, it was not expressed in normal thyroid and nonmetastatic PTC. One likely explanation is that LTB may be necessary for the early phases such as invasion as those required for the late phase of this complex process. Interestingly, prevention of LTB-LTBR signaling is reported to inhibit tumor angiogenesis and neovascularization, which results in tumor growth arrest (38). These findings have prompted others to target the LTBR with agonist antibodies as a potential anticancer therapy (39). These results suggest that targeting the LTBR signaling could also be a novel approach to the treatment of metastatic PTC.

In addition to a better understanding of the metastatic process in thyroid cancers, these markers have potential clinical use for monitoring disease progression. Monitoring of serum thyroglobulin (TG) levels has been used as gold-standard method in the follow-up of patients after total thyroidectomy and radioiodine therapy. However, the value of this immunoassay is limited to patients with total thyroidectomy. Interference by antithyroglobulin antibodies in the blood is also present in ~25% of patients. Circulating thyroid-specific transcripts such as TG, TPO, TSHR, NIS, and PDS have been suggested as potential molecular markers of residual or recurrent thyroid cancer (40–42). Conversely, several reports suggested that these markers could not be used in the follow-up of patients (43, 44). Therefore, there is much need for better markers. The new markers described here are candidates for possible use in conjunction with thyroglobulin or alone in the follow-up of patients with thyroid tumor recurrence.

Although we focused on genes highly expressed in LNsMs, these gene expression profiles may have other purposes. For example, local invasion in the primary tumor likely precedes distant metastasis. Therefore, the primary tumor may already contain genes that predict poor prognosis. In this sense, a comparison done among the matched normal and primary PTC would allow us to identify genes highly expressed in primary tumors that could be an indication of host stromal response and poor prognosis. As expected, this analysis revealed similarities with microarray data done in primary thyroid tumors (16, 45, 46). For example, we identified and observed the previously reported transcripts highly expressed in primary PTC, MET and S100A4. The validation analysis, however, proved that they are not associated with the late stages of the metastatic process.

In conclusion, our study identifies and validates a list of novel candidate markers associated with LNsMs from PTC. Given the critical role of tumor cell-host cell interactions in the metastatic process, the genes identified here could be helpful to ascertain the molecular basis of these cellular interactions. The genes and molecular pathways consistently associated with tumor invasion and metastasis could also provide new targets for therapy. However, we should first determine if these biomarkers can improve patient care by earlier or better detection of metastasis in patients with a diagnosis of PTC.

Acknowledgments

Received 12/27/2006; revised 5/13/2007; accepted 6/14/2007.

Grant support: NIH grant CA113461, the Virginia and D.K. Ludwig Fund for Cancer Research, and the São Paulo State Research Foundation (FAPESP) from grants 04/15288-0 and 05/60330-8. J.M. Cerutti is investigator of the Brazilian Research Council (CNPq), G. Oler is a scholar from FAPESP, and G.J. Riggin is the recipient of the Irving J. Sherman M.D. Research Professorship. The project has been funded in part with federal funds from the National Cancer Institute, NIH, under contract N01-CO-12400.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The content of this publication does not necessarily reflect the views of policies of the Department of Health and Human Service, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.
Molecular Profiling of Matched Samples Identifies Biomarkers of Papillary Thyroid Carcinoma Lymph Node Metastasis

Janete M. Cerutti, Gisele Oler, Pedro Michaluart, Jr., et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/67/16/7885

Cited articles
This article cites 46 articles, 15 of which you can access for free at:
http://cancerres.aacrjournals.org/content/67/16/7885.full.html#ref-list-1

Citing articles
This article has been cited by 4 HighWire-hosted articles. Access the articles at:
/content/67/16/7885.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.