Featured Article

Gene-Expression Profile Changes Correlated with Tumor Progression and Lymph Node Metastasis in Esophageal Cancer

Eiji Tamoto,1 Mitsuhiro Tada,3
Katsuhiko Murakawa,1 Minoru Takada,1
Gaku Shindo,1 Ken-ichi Teramoto,1
Akihiro Matsunaga,1 Kazuteru Komuro,1
Motoshi Kanai,1 Akiko Kawakami,1
Yosie Fujiwara,4 Nozomi Kobayashi,4
Katsutoshi Shirata,4 Norihiro Nishimura,4
Shun-ichi Okushiba,1 Satoshi Kondo,2
Jun-ichi Hamada,3 Takashi Yoshiki,2
Tetsuya Moriuchi,3 and Hiroyuki Katoh1

1Department of Surgical Oncology, Division of Cancer Medicine, Graduate School of Medicine, 2Department of Pathology/Pathophysiology, Division of Pathophysiological Science, and 3Division of Cancer-Related Genes, Institute for Genetic Medicine, Hokkaido University, and 4GeneticLab Co., Ltd., Sapporo, Japan

ABSTRACT

Purpose: The purpose of this research was to identify molecular clues to tumor progression and lymph node metastasis in esophageal cancer and to test their value as predictive markers.

Experimental Design: We explored the gene expression profiles in cDNA array data of a 36-tissue training set of esophageal squamous cell carcinoma (ESCC) by using generalized linear model-based regression analysis and a feature subset selection algorithm. By applying the identified optimal feature sets (predictive gene sets), we trained and developed ensemble classifiers consisting of multiple probabilistic neural networks combined with AdaBoosting to predict tumor stages and lymph node metastasis. We validated the classifier abilities with 18 independent cases of ESCC.

Results: We identified 71 genes of 1289 cancer-related genes of which the expression correlated with tumor stages. Of the 71 genes, 47 significantly differed between the Tumor-Node-Metastasis pT1/2 and pT3/4 stages. Cell cycle regulators and transcriptional factors possibly promoting the growth of tumor cells were highly expressed in the early stages of ESCC, whereas adhesion molecules and extracellular matrix-related molecules possibly promoting invasiveness increased in the later stages. For lymph node metastasis, we identified 44 genes with predictive values, which included cell adhesion molecules and cell membrane receptors showing higher expression in node-positive cases and cell cycle regulators and intracellular signaling molecules showing higher expression in node-negative cases. The ensemble classifiers trained with the selected features predicted tumor stage and lymph node metastasis in the 18 validation cases with respective accuracies of 94.4% and 88.9%. This demonstrated the reproducibility and predictive value of the identified features.

Conclusion: We suggest that these characteristic genes will provide useful information for understanding the malignant nature of ESCC as well as information useful for personalizing the treatments.

INTRODUCTION

Esophageal cancer shows the poorest prognosis among the malignant tumors of the digestive tract. Although advances in diagnostic methods and health education have contributed to discovery of the disease in its early stages, many cases are still not detected until the advanced stage. Despite the use of modern surgical techniques in conjunction with multitreatment modalities such as radio- and chemotherapy, the overall 5-year survival rate remains 40–60% (1–4). In an effort to better understand this disease and predict its clinical outcome, numerous molecular markers for tumor progression and prognosis, e.g., STMY3 (5), interleukin 6 (6), C1orf10 (7), and EC45 (8), and for lymph node metastasis, e.g., cavelolin-1 (9), EphA2 (10), FAK (11), cystain B (12), and MMP-12 (13), have been identified. This limited information, however, is not enough to clarify the carcinogenesis, tumor progression, and invasiveness of esophageal cancer; just as in the allegory about the five blind men and the elephant, the complete picture of the pathophysiology of the disease remains elusive.

As distinct from these “single gene-oriented” approaches, expression profiling that collectively analyzes the expression of many genes using a cDNA array has drawn attention as a promising approach for uncovering molecular mechanisms independently of previous knowledge (14–16). In the current study, we analyzed gene expression profiles in a total of 36 esophageal cancers and their relation to pathological features based on the Tumor-Node-Metastasis staging. By using a generalized linear model-based regression analysis that allowed us to extract information on features relevant to discrete graded categories, we identified genes of which the expression was altered in association with tumor progression. Also, we identified genes of which the expressions were associated with lymph node metastasis by a feature-subset selection algorithm: this algorithm enabled us to optimize the statistical explanatory...
model for distinction, which we demonstrated in the classification of 18 independent cases for validation.

MATERIALS AND METHODS

Patients and Samples. We collected the primary cancer tissues obtained from a total of 54 patients with histologically verified esophageal squamous cell carcinomas (ESCCs) who underwent surgery from January 2001 to September 2003 in the Hokkaido University Hospital and 15 affiliated hospitals in Hokkaido prefecture, Japan. Only patients who agreed with the aim and contents of this study and who provided their written informed consent were included. One to five bulk tumor tissue samples of 5 mm-size were immediately cut from the esophagus resected by a standard surgical procedure, snap frozen in liquid nitrogen, and stored at -80 °C until use. Part of each sample piece was cut and stained with H&E for verification of the presence of squamous cell carcinoma cells, and the sample pieces, of which the evaluated areas were composed of at least 50% tumor cells, were used for RNA extraction. All of the procedures in this portion of the study were approved by the Ethics Committee of Hokkaido University and the independent internal ethics committees of the affiliated hospitals.

Clinicopathological Parameters. Histological subclassification and staging of the tumors was done by reviewing the specimens taken for pathological diagnosis, according to the Tumor-Node-Metastasis classification (17). The tumor status of each case was categorized based on the Tumor-Node-Metastasis classification (Unio Internationale Contra Cancrum, 6th edition) for the pT, pN, and pM stages. Pertinent major clinicopathological parameters are shown in Table 1.

Analysis of Gene Expression. Each frozen tissue was crushed in liquid nitrogen to powder by using a CRYO-PRESS compressor (Microtec Nition, Chiba, Japan). Total RNA was extracted using TRIZol reagent (Invitrogen, Tokyo, Japan). Before mRNA extraction, total RNA was treated with 1 unit/μl DNase I and 10 units/μl RNase inhibitor (TOYOBO, Osaka, Japan) at 4 °C for 15 min. Polyadenylated mRNA was extracted from total RNA using oligodeoxythymidylic acid magnetic beads (MagExtracter-mRNA-kit; TOYOBO). The mRNA (1.0 μg) was then reverse-transcribed using ReverTraAce reverse transcriptase (TOYOBO). The quality of the sample cDNA was checked by amplifying β-actin, G3PDH, and α-tubulin with PCR amplification. The cDNA was ethanol-precipitated and substituted to polyadenylated tailing with terminal deoxynucleotidyl transferase (TOYOBO). The Polyadenylated cDNA was then amplified by PCR (25 cycles) for biotin-16-dUTP labeling with KOD dash DNA polymerase (TOYOBO). The labeled cDNA was ethanol-precipitated, denatured at 68 °C, and used as a hybridization probe in 10 ml of PerfectHyb solution (TOYOBO). Hybridization was done overnight at 68 °C. Chemiluminescence was done using a streptavidin-biotinylated alkaline phosphatase system (Imaging High- Chemilum-Gene Navigator; TOYOBO) using CDP-Star as the luminogen, and the signals were photographed by a Fluor-S MultiImager (Nippon Bio-Rad Laboratories, Tokyo, Japan). Image analysis and quantification were performed with Imagine 4.2 software (BioDiscovery Inc., Los Angeles, CA), averaging the signals from the adjusted regions of two symmetrically arranged spots for each gene, relative to background values.

Statistical Analyses and Array Data Analysis. In this study we investigated changes in the gene-expression profile during tumor progression. Analysis of gene expression profiles

<table>
<thead>
<tr>
<th>Lymph node metastasis</th>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD)</td>
<td>66.2 ± 6.1</td>
<td>65.9 ± 7.3</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Depth of invasion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>pT2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>pT3</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>pT4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>pTNM M-status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>M1a</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M1b</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pTNM stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>II A</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>II B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IV A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IV B</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinicopathological features in the 36 cases for training and the 18 cases for validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymph node metastasis</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Depth of invasion</td>
</tr>
<tr>
<td>pT1</td>
</tr>
<tr>
<td>pT2</td>
</tr>
<tr>
<td>pT3</td>
</tr>
<tr>
<td>pT4</td>
</tr>
<tr>
<td>pTNM M-status</td>
</tr>
<tr>
<td>M0</td>
</tr>
<tr>
<td>M1a</td>
</tr>
<tr>
<td>M1b</td>
</tr>
<tr>
<td>pTNM stage</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II A</td>
</tr>
<tr>
<td>II B</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>IV A</td>
</tr>
<tr>
<td>IV B</td>
</tr>
</tbody>
</table>
and related statistical analyses were performed using programs originally developed in the MATLAB language, ver. 6.5 (MathWorks, Tokyo, Japan). To determine confounding relations among the histopathological parameters, a contingency table analysis ($\chi^2$ test) was performed between pN status and pT status.

The array data in the present study consisted of 1289 dimensional variables (genes) that showed an exponential distribution. The data distribution of each sample exhibited a cumulative distribution curve of different shape, reflecting a biased size-shift (nonlinear offset effect) because of a different condition in each hybridization. Normalizations using expression values of housekeeping genes were not considered appropriate, because they often lost linearity due to overhybridization. Thus, we normalized the data by dividing the expression value of each sample by the mean value of the genes of which the values were within 0.1–1.2-fold of the median expression. We set this interval because of the stability of data values observed.

**Fig. 1** The cumulative distributions of cDNA array data of the 54 esophageal squamous cell carcinomas. Each ——— in the cascade-like picture represents a cumulative distribution curve of the expression values of 1289 genes in a case. A, before normalization, note that distributions are uneven across the cases, forming an irregularly shaped "cascade"; B, after normalization, the shape of the cascade was smoothed, especially in the lower 80 percentiles of genes.

**Fig. 2** Changes in the expression of genes correlated with tumor progression (pT stages). A, 24 genes (top) were decreased, and 47 genes (bottom) were increased along with the advance of pT stage by the generalized linear model-based regression analysis. Brighter color indicates higher expression. B, heat-map view of the 88 genes of which the expressions significantly differed between the pT1/2 (n = 16) and pT3/4 (n = 20) groups in the 36-case training set. The normalized relative expression for each transcript (rows) in each sample (columns) is shown in color (brighter color indicates higher expression). For comparison, the expressions of the genes in the pT1/2 (n = 7) and pT3/4 (n = 11) groups in the 18-case validation set are shown on the right of the panel. Note that the similar gene-expression patterns in the validation cases demonstrate the reproducibility of the expression profile in the independent set of samples. C, the process of feature-subset gene selection with predictive value for pT stages (pT1/2 versus pT3/4). Sequential addition of the most significant genes at each step, with a total of 3916 models tested, resulted in a minimal error rate of 0.0%.
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Symbol</th>
<th>Accession</th>
<th>β Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive correlation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell adhesion protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibronectin</td>
<td>FN1</td>
<td>M10905</td>
<td>0.26</td>
</tr>
<tr>
<td>Integrin α5</td>
<td>ITGA5</td>
<td>X06256</td>
<td>0.26</td>
</tr>
<tr>
<td>Integrin β1</td>
<td>ITGB1</td>
<td>X07979</td>
<td>1.19</td>
</tr>
<tr>
<td>Integrin β3</td>
<td>ITGB3</td>
<td>M20311</td>
<td>2.35</td>
</tr>
<tr>
<td>Integrin β6</td>
<td>ITGB6</td>
<td>M35198</td>
<td>1.65</td>
</tr>
<tr>
<td>Laminin β1</td>
<td>LAMB1</td>
<td>M61951</td>
<td>0.85</td>
</tr>
<tr>
<td>Laminin γ</td>
<td>LAMC1</td>
<td>M55210</td>
<td>0.95</td>
</tr>
<tr>
<td>Cell cycle regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell division cycle27</td>
<td>CDC27</td>
<td>U00001</td>
<td>1.94</td>
</tr>
<tr>
<td>Nucleolar protein 3</td>
<td>NOL3</td>
<td>AF043244</td>
<td>0.43</td>
</tr>
<tr>
<td>Apoptosis regulator</td>
<td>BID</td>
<td>AF042083</td>
<td>1.16</td>
</tr>
<tr>
<td>Secreted frizzled-related protein 2</td>
<td>SFRP2</td>
<td>AF017986</td>
<td>0.26</td>
</tr>
<tr>
<td>Galectin 1</td>
<td>LGALS1</td>
<td>J04456</td>
<td>0.06</td>
</tr>
<tr>
<td>Cell membrane protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 antigen</td>
<td>CD4</td>
<td>M12807</td>
<td>1.16</td>
</tr>
<tr>
<td>CD58 antigen</td>
<td>CD58</td>
<td>Y00636</td>
<td>1.08</td>
</tr>
<tr>
<td>Lactadherin</td>
<td>MPGE8</td>
<td>U58516</td>
<td>0.45</td>
</tr>
<tr>
<td>Member RAS oncogene family</td>
<td>RAB5A</td>
<td>M28215</td>
<td>0.72</td>
</tr>
<tr>
<td>Cell membrane receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-1D-adrenergic receptor</td>
<td>ADRA1D</td>
<td>U03864</td>
<td>0.06</td>
</tr>
<tr>
<td>CD47 antigen</td>
<td>CD47</td>
<td>Z25521</td>
<td>0.99</td>
</tr>
<tr>
<td>Dopamine receptor D2</td>
<td>DRD2</td>
<td>M30625</td>
<td>0.23</td>
</tr>
<tr>
<td>Dopamine receptor D3</td>
<td>DRD3</td>
<td>U32499</td>
<td>0.40</td>
</tr>
<tr>
<td>IFN-α receptor</td>
<td>IFNAR1</td>
<td>J03171</td>
<td>2.20</td>
</tr>
<tr>
<td>Cholinergic receptor, muscarinic 1</td>
<td>CHRM1</td>
<td>X15263</td>
<td>0.86</td>
</tr>
<tr>
<td>Glutamate receptor N-methyl d-aspartate 2C</td>
<td>GRIN2C</td>
<td>U77782</td>
<td>0.09</td>
</tr>
<tr>
<td>PY2 purinocceptor 1</td>
<td>P2RY1</td>
<td>U42030</td>
<td>0.47</td>
</tr>
<tr>
<td>Extracellular matrix related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsin K</td>
<td>CTSK</td>
<td>X82153</td>
<td>0.13</td>
</tr>
<tr>
<td>Matrix metalloproteinase 1</td>
<td>MMP1</td>
<td>X54925</td>
<td>0.22</td>
</tr>
<tr>
<td>Matrix metalloproteinase 7</td>
<td>MMP7</td>
<td>Z11887</td>
<td>0.47</td>
</tr>
<tr>
<td>Matrix metalloproteinase 10</td>
<td>MMP10</td>
<td>X07820</td>
<td>0.51</td>
</tr>
<tr>
<td>Matrix metalloproteinase 14</td>
<td>MMP14</td>
<td>Z48481</td>
<td>0.26</td>
</tr>
<tr>
<td>Mucin 2</td>
<td>MUC2</td>
<td>M74027</td>
<td>0.37</td>
</tr>
<tr>
<td>Protein C</td>
<td>PROC</td>
<td>K02059</td>
<td>1.03</td>
</tr>
<tr>
<td>Tissue inhibitor of metalloproteinase 1</td>
<td>TIMP1</td>
<td>X03124</td>
<td>0.03</td>
</tr>
<tr>
<td>Tissue inhibitor of metalloproteinase 3</td>
<td>TIMP3</td>
<td>U14394</td>
<td>0.77</td>
</tr>
<tr>
<td>Growth/differentiation factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth differentiation factor</td>
<td>GDF1</td>
<td>M62302</td>
<td>1.96</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 6</td>
<td>IGFBP6</td>
<td>M62402</td>
<td>0.32</td>
</tr>
<tr>
<td>Inhibin βA</td>
<td>INHBA</td>
<td>M13436</td>
<td>0.14</td>
</tr>
<tr>
<td>Chemokine ligand 4</td>
<td>CCL4</td>
<td>J04130</td>
<td>1.07</td>
</tr>
<tr>
<td>Chemokine (C-C motif) ligand 1</td>
<td>CCL17</td>
<td>D43767</td>
<td>1.12</td>
</tr>
<tr>
<td>Vascular endothelial growth factor C</td>
<td>VEGFC</td>
<td>U43142</td>
<td>0.98</td>
</tr>
<tr>
<td>Intracellular signaling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitric oxide synthase 2A</td>
<td>NOS2A</td>
<td>AB022318</td>
<td>1.10</td>
</tr>
<tr>
<td>Ribosomal protein S6 kinase, 90kD, 1</td>
<td>RPS6KA1</td>
<td>BC014966</td>
<td>0.63</td>
</tr>
<tr>
<td>Transforming growth factor, β-induced, 68kDa</td>
<td>TGFB1</td>
<td>M77349</td>
<td>0.15</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molecular chaperone</td>
<td>SERPINH1</td>
<td>X61598</td>
<td>0.16</td>
</tr>
<tr>
<td>Caveolin 2</td>
<td>CAV2</td>
<td>AA878149</td>
<td>0.47</td>
</tr>
<tr>
<td>Activating transcription factor 3</td>
<td>ATF3</td>
<td>L19871</td>
<td>0.48</td>
</tr>
<tr>
<td>DNA repair protein XRCC2</td>
<td>XRCC2</td>
<td>AF035587</td>
<td>0.35</td>
</tr>
<tr>
<td>Cathepsin L</td>
<td>CTSL</td>
<td>X12451</td>
<td>0.66</td>
</tr>
<tr>
<td>Negative correlation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell cycle regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDC-like kinase 2</td>
<td>CLK2</td>
<td>L29216</td>
<td>−0.51</td>
</tr>
<tr>
<td>p53-binding protein MDM2</td>
<td>MDM2</td>
<td>Z12020</td>
<td>−1.48</td>
</tr>
<tr>
<td>PCTAIRE protein kinase 2</td>
<td>PCTK2</td>
<td>X66360</td>
<td>−1.24</td>
</tr>
<tr>
<td>Cell division cycle 2-like</td>
<td>CDC2L1</td>
<td>U04815</td>
<td>−0.76</td>
</tr>
<tr>
<td>Retinoblastoma-like 2 (p130)</td>
<td>RBL2</td>
<td>X76061</td>
<td>−0.99</td>
</tr>
<tr>
<td>Cell membrane receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD20 antigen</td>
<td>MS4A1</td>
<td>X12530</td>
<td>−0.53</td>
</tr>
<tr>
<td>G protein-coupled receptor 65</td>
<td>GPR65</td>
<td>U95218</td>
<td>−1.20</td>
</tr>
<tr>
<td>EphB6</td>
<td>EPHB6</td>
<td>D83492</td>
<td>−2.58</td>
</tr>
</tbody>
</table>
in a number of empirical cumulative distribution curves of the array data.

We divided the 54 cases into a training set of 36 cases and a validation set of 18. The pertinent clinicopathological data of the constituents of each set are shown in Table I. They indicate an unbiased division of cases. We used the data of the training set to explore the characteristics of tumor progression (advancing pT stages) and metastatic potential (pN stages). We then used the validation set to evaluate the predictive ability of the selected features.

To detect genes that increased or decreased in correlation with the graded depth of invasion, we adopted a regression analysis on a generalized linear model (18). This model allows a reduction of the relation between an explanatory variable X and a response variable Y into a linear relation if the distribution of Y belongs to the exponential distribution function, by transforming Y with an appropriate link (kernel) function:

\[ Y = \alpha + \beta x, \]

where \( \alpha \) is the intercept and \( \beta \) is the regression coefficient. Probit function (inverse Gaussian distribution function) was used for the link function. In the estimation of \( \beta \), the maximum likelihood method is used to minimize the residuals (the weighted least-square fit). We applied this method to each of 1289 genes and extracted those that showed \( P \) of regression coefficient \( \beta < 0.05 \).

On the other hand, to determine the expression profiles that are characteristic of the presence or absence of lymph node metastasis and those that are characteristic of the pT1/2 versus pT3/4 stages, we selected genes on the basis of statistical difference between mean expressions of the classes (two-sided \( t \) test, \( P \leq 0.05 \)) in the 36-case training set. To determine whether the selected genes were classified accurately as samples in the various classes, we performed a classification using an expectation maximization algorithm that separates a mixture of different data distributions in iterative steps of maximum likelihood estimation (19). We also tested the separability of the distributions with a k-means clustering (20). In addition, we tested the differences of gene expressions between the classes by a nonparametric bootstrap resampling test that permitted us to check the stability of statistical differences independent of the data distribution (21). This procedure randomly and multiply resamples the data with replacement (allowing overlap) and compares the means of the resampled data with the means of the original samples (pivotal test, two-sided, \( \alpha \) level 0.05). This tests the apparent by-chance difference attributable to the presence of outlier values or biased distributions of values, to which the parametric \( t \) test is insensitive.

We then performed a feature-subset selection that allows identification of a set of genes optimal for pattern classification from the selected genes to extract the essential part of the gene set. As the algorithm for the feature-subset selection, we used a sequential forward selection, which extensively explored a combination of genes minimizing the leave-one-out error rate of a k-nearest neighbor classifier. This algorithm, proposed by Whitney (22), adds the most significant feature (gene) that gives a minimal error rate for each step, from an empty set to the full set of the given data, and finally selects a series of gene sets yielding the grand minimal error rate. As a classifier we adopted the k-nearest neighbor method, which classifies each sample

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Symbola</th>
<th>Accessionb</th>
<th>( \beta ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth/differentiation factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemokine receptor 7</td>
<td>CCR7</td>
<td>L31581</td>
<td>−0.92</td>
</tr>
<tr>
<td>Interleukin 18</td>
<td>IL18</td>
<td>D40950</td>
<td>−1.84</td>
</tr>
<tr>
<td>Colony stimulating factor 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>CSF1</td>
<td>M27087</td>
<td>−2.16</td>
</tr>
<tr>
<td>Intracellular signaling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dishevelled 3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>DVL3</td>
<td>AF006013</td>
<td>−0.36</td>
</tr>
<tr>
<td>G protein-coupled receptor kinase 5&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>GPRK5</td>
<td>L15388</td>
<td>−1.39</td>
</tr>
<tr>
<td>G protein-coupled receptor kinase 6&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>GPRK6</td>
<td>L16862</td>
<td>−1.00</td>
</tr>
<tr>
<td>RalA binding protein 1&lt;sup&gt;d&lt;/sup&gt;</td>
<td>RALBP1</td>
<td>L42542</td>
<td>−1.92</td>
</tr>
<tr>
<td>Transcriptional factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear receptor coactivator 2&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>NCOA2</td>
<td>X97674</td>
<td>−2.67</td>
</tr>
<tr>
<td>Nuclear receptor coactivator 1</td>
<td>NCOA1</td>
<td>U40396</td>
<td>−1.61</td>
</tr>
<tr>
<td>Signal transducer and activator of transcription 6&lt;sup&gt;d&lt;/sup&gt;</td>
<td>STAT6</td>
<td>U16031</td>
<td>−0.46</td>
</tr>
<tr>
<td>Paired box gene 8&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>PAX8</td>
<td>L19606</td>
<td>−1.11</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptor interacting kinase 1&lt;sup&gt;d&lt;/sup&gt;</td>
<td>RIPK1</td>
<td>U50062</td>
<td>−0.80</td>
</tr>
<tr>
<td>CD8 antigen</td>
<td>CD8A</td>
<td>M12828</td>
<td>−0.97</td>
</tr>
<tr>
<td>Topoisomerase (DNA) 1&lt;sup&gt;′&lt;/sup&gt;</td>
<td>TOP1</td>
<td>J03250</td>
<td>−0.58</td>
</tr>
<tr>
<td>Cytochrome P450, subfamily IID, polypeptide 6</td>
<td>CYP2D6</td>
<td>M20403</td>
<td>−0.63</td>
</tr>
<tr>
<td>Keratin 18&lt;sup&gt;c&lt;/sup&gt;</td>
<td>KRT18</td>
<td>M26326</td>
<td>−0.23</td>
</tr>
</tbody>
</table>

<sup>a</sup> Symbol in LocusLink database.
<sup>b</sup> GenBank ID.
<sup>c</sup> Genes increased their expression in association with the depth of invasion (pT stage).
<sup>d</sup> Genes of which the expressions significantly differed between pT1/2 and pT3/4 stages.
<sup>e</sup> Features selected for two-class discrimination for pT1/2 versus pT3/4 stages.
<sup>f</sup> Genes decreased their expression in association with the depth of invasion (pT stage).
<sup>g</sup> \( \beta \) value: regression coefficient.
according to the class memberships of its k nearest points based on Euclidean distance in the N-dimensional space (20, 23). The leave-one-out error cross-validation procedure designates one sample as a test set and the remaining N - 1 samples as a training set. The k-nearest neighbor rule developed on the training set was repeatedly applied for each left-one sample, and a cumulative training error was calculated.

The feature subset selection yields a number of feature (gene) sets that give the same grand minimal error rate. The gene sets were then analyzed for their selection tree structure with Dijkstra and depth-first search algorithms to extract, by pruning leaves and branches, the connected parts (vertices) that form the tree trunk (24, 25). This procedure based on the graph theory permits extraction of the essential part of the feature subsets, by discarding the possible local minimal solutions (prediction models) that are considered the major cause of deterioration of the generalization ability of a classifier. The obtained sets of genes were then used as the final diagnostic sets to form an ensemble classifier consisting of multiple probabilistic neural networks (PNN; Ref. 26) in which each component classifier was weighted with the AdaBoost algorithm (27). PNN is a space-dividing classifier like k-nearest neighbor, but uses the radial basis function to simulate Bayesian posterior probabilities for each data point instead of evaluating simple Euclidean distances as in k-nearest neighbor. With the use of AdaBoosting, each classifier (PNN) was weighted with leave-one-out in-sample errors in the training procedure to optimize the training and voting. Finally, to determine the performance of the ensemble classifier, we tested it with the 18-case validation set independent of the 36-case training set and evaluated the overall classification error.

RESULTS

In this study we attempted to extract gene expression profiles characteristic of the invasiveness and progression status of ESCCs, with a particular focus on lymph node metastasis (pN stage) and depth of tumor invasion (pT stage). A contingency table analysis ($\chi^2$ test) showed no significant relationship between the pN stages and the pT stages (pN0: pT1, 7; pT2, 3; pT3, 14; pT4, 1; pN1: pT1, 3; pT2, 10; pT3, 14; pT4, 2 cases; $\chi^2 = 5.436, P = 0.1425$). The cumulative distribution curves of the array data of the 54 cases before and after normalization are shown in Fig. 1. After normalization, the different shapes in the array data of the 54 cases before and after normalization are similar to those in the 36-case training set. Among these 88 genes, 47 were in common with the above 71 genes selected by the regression analysis. We confirmed the stability of statistical differences for these 88 genes with a nonparametric bootstrap test (10,000 times resampling). We tested the separability of the data subspaces (distributions) formed by the expression values of the 88 genes by using an expectation maximization algorithm. This algorithm misclassified 1 of the 36 cases (error rate, 2.8%). Another classifier, k-means clustering, misclassified 11 of 36 cases (error rate, 30.6%). To identify the essential features to optimally classify the training set data, we applied a sequential forward-selection algorithm that sequentially selected a better combination of genes based on leave-one-out error rates of a k-nearest neighbor classifier. The algorithm and the subsequent tree analysis using the Dijkstra and depth-first search algorithms finally selected 33 combinations of features (genes), consisting of 26–62 genes that classified the two classes with 100% accuracy (Fig. 2C).

As the genes characteristic of lymph node metastasis, we extracted a total of 87 genes of which the expression showed a significant difference (two-sided t test, $P < 0.05$) between pN0 cases ($n = 16$) and pN1 cases ($n = 20$; Fig. 3A). Again, the independent validation set (pN0, 9 cases; pN1, 9 cases) showed expression patterns similar to those of the training set, demonstrating the reproducibility of the expression profiles across the independent sets of samples. We confirmed the statistical differences by the bootstrap resampling test. A test of the separability of the data subspaces (87 dimensions) using the expectation maximization algorithm classified 31 of the 36 in-sample cases correctly (error rate, 13.9%). Also, the k-means clustering misclassified 11 of 36 cases (error rate, 30.6%). In the same way as described above, to prune features that made the margin between pN0 and pN1 classes and obtain an optimal set of features that characterize the lymph node status, we performed a feature-subset selection. The algorithm and the subsequent tree analysis finally selected 19 sets consisting of 11–44 genes that classified the two classes with 97.2% accuracy (Fig. 3B; Table 3).

The 71 genes of which the expressions were altered in association with T stages (Table 2) and the 87 genes that were differentially expressed between node-positive and node-negative cases shared 14 common genes (NOL3, CAV2, CD47, CYP2D6, DRD2, DRD3, EPHB6, ITGB6, MDM2, MMP14, MUC2, GRIN2C, P2RY1, and TGFBI). We estimated the probability of the coincidental selection of these 14 genes with a Monte-Carlo simulation (100,000 times), which gave a $P$ of $1.0 \times 10^{-5}$. The 57 genes that excluded these 14 genes from the 71 genes may specifically reflect the molecular characteristics of depth of tumor invasion (pT stage). Only 4 (CYP2D6, EPHB6, ITGB6, and MDM2) of the 71 genes were also included among the 44 genes that characterized lymph node metastasis ($P = 0.07695$, Monte-Carlo simulation). Whereas these may represent the common molecular events between graded depth of invasion and lymph node metastasis, most genes were different between the two categories, suggesting that depth of invasion and lymph node metastasis had distinct underlying molecular pathophysologies.

To test the validity of the selected features, we tested the feature subsets composed of 33 combinations of genes (26–62)
for pT1/2 versus pT3/4 classification, as well as 19 combinations (11–44 genes) for pN1 and pN0, by using a PNN ensemble classifier. For each subset of features, a PNN classifier was formed, and its leave-one-out classification error was evaluated with the 36-case training set. The AdaBoost algorithm increased the weight of the misclassified cases and accordingly modified the successive training. Finally, the ensemble classification was done by summing up the weighted votes of all of the PNN classifiers for the independent validation-set cases and then by classifying each test case to a class given a larger vote-sum. As shown in Table 4, the overall prediction error rates for pT stages and pN stages were respectively 5.6% (17 of 18 cases correctly classified) and 11.1% (16 of 18 cases). Furthermore, we tested whether or not the 71 genes selected by the regression analysis could discriminate between the pT1/2 and pT3/4 classes in the 18-case validation set. The feature subset selection algorithm selected a total of 20 subsets consisting of 10–29 genes, giving an in-sample error rate of 2.8%. The final performance of the formed ensemble classifier was a 16.7% overall error rate (16 of 18 cases correctly classified, 2 of 7 pT1/2 and 1 of 11 pT3/4 misclassified). The finally selected 29 genes contained 18 in common with the above 62 finally selected features.

**DISCUSSION**

Multiple genes are involved in the complex multistep process of carcinogenesis, tumor progression, and acquisition of invasiveness in esophageal cancer. The cDNA array technology has enabled us to analyze the expression profiles of thousands of genes simultaneously and to investigate the correlation between clinicopathological phenotypes and gene expression status. This technique provides a powerful means to stratify (or personalize) tumors into classes with distinct molecular pathophysiology that has not been completely proven. In esophageal cancer, some authors have reported that the gene expression profile of ESCC is different from those of pericancerous, normal, or dysplastic epithelia, implying the potential usefulness of the technique for the differential diagnosis of early ESCC from precancerous benign lesions (28, 29). To date, however, there has been no report that investigates expression profiles associated with specific malignant characters of ESCC. In the present study, we for the first time identified genes of which the expression was altered in correlation with the graded depth of invasion (pT stages) by applying a generalized linear model-based regression analysis to cDNA array data. The changes in gene expression were most conspicuous between pT1 and pT3 (invasion beyond the musculus propius), suggesting that a major alteration in molecular pathophysiology occurs at a rather late stage of tumor progression. The validity of the identified genes was confirmed by the reproducibility of the expression patterns in the independent set of samples, and by the overall performance of the classifier starting from both feature sets selected by the regression analysis and binary-class comparison.

As shown in Table 2, the identified genes that increased their expression in association with advance of pT stages had a wide spectrum of functions, including cell adhesion molecules, extracellular matrix-related proteins, cell death regulators, and growth/differentiation factors. Among them, for instance, integrin α5 is known to activate signal transduction, enhance cell proliferation, suppress apoptosis, and thereby contribute to tumor progression in lung squamous cell carcinoma (30). **ITGB6** also promotes the progression of oral squamous cell carcinoma by activating Fyn on binding to fibronectin (31), which showed higher expression in the progressed tumors in our study. Among cell death regulators, **NOL3** inhibits apoptosis (32). **LGALS1** is involved in neoplastic progression and proliferative activities (33). The matrix metalloproteinase **MMP1** is associated with depth of invasion in esophageal cancer (34), whereas **MMP7** and **MMP10** have been shown to be prognostic factors in esophageal cancer (35, 36). **TIMP1**, which is an inhibitor of matrix metalloproteinases, correlated with the tumor progression in the present results, as one previous report indicated a positive correlation with tumor progression (37). **VEGFC**, which is a growth/differentiation factor, is known to promote esophageal tumor progression via lymphangiogenesis and angiogenesis (38). Taken together, these results indicate that tumor progression of ESCC is associated with factors that are relevant to the invasive characters of tumors.

The genes of which the expressions were decreased in
### Table 3
The selected optimal set of genes for classification into each class (lymph node positive and negative)

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Symbol</th>
<th>Accession</th>
<th>pN1:pN0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell adhesion proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD33 antigen (gp67)</td>
<td>CD33</td>
<td>M23197</td>
<td>1.21</td>
</tr>
<tr>
<td>Gap junction protein, β1, 32kDa</td>
<td>GB1</td>
<td>X04325</td>
<td>1.22</td>
</tr>
<tr>
<td>Integrin αIb</td>
<td>ITGA2B</td>
<td>J02764</td>
<td>1.17</td>
</tr>
<tr>
<td>Integrin αX</td>
<td>ITGAX</td>
<td>Y00093</td>
<td>1.17</td>
</tr>
<tr>
<td>Integrin β6</td>
<td>ITGB6</td>
<td>M53198</td>
<td>1.15</td>
</tr>
<tr>
<td>Lymphocyte-function antigen-2</td>
<td>CD2</td>
<td>M14362</td>
<td>1.23</td>
</tr>
<tr>
<td>Avian erythroblastic leukemia viral oncogene homolog 4</td>
<td>ERBB4</td>
<td>L07868</td>
<td>1.17</td>
</tr>
<tr>
<td>Colony-stimulating factor-2 receptor</td>
<td>CSF2RB</td>
<td>M59941</td>
<td>1.24</td>
</tr>
<tr>
<td>Cell death regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitotic arrest deficient-like 1</td>
<td>MAD1L1</td>
<td>L06895</td>
<td>1.20</td>
</tr>
<tr>
<td>BCL2/adenovirus E1B 19kDa interacting protein 3</td>
<td>BNIP3</td>
<td>U15172</td>
<td>1.21</td>
</tr>
<tr>
<td>Intracellular signaling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD38 antigen (p45)</td>
<td>CD38</td>
<td>M34461</td>
<td>1.38</td>
</tr>
<tr>
<td>Rho GDP dissociation inhibitor γ</td>
<td>ARHGDIG</td>
<td>AF080237</td>
<td>1.36</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytochrome p450 XVIIA1</td>
<td>CYP17A1</td>
<td>M14564</td>
<td>1.17</td>
</tr>
<tr>
<td>Retroviral sequences NP2</td>
<td>RVNP2</td>
<td>M15971</td>
<td>1.19</td>
</tr>
<tr>
<td>NADPH:quinone reductase</td>
<td>CRYZ</td>
<td>L31521</td>
<td>1.20</td>
</tr>
<tr>
<td>Matrix metalloproteinase 13</td>
<td>MMP13</td>
<td>X75308</td>
<td>1.45</td>
</tr>
<tr>
<td>Cell cycle regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDC16</td>
<td>CDC16</td>
<td>AF164598</td>
<td>0.856</td>
</tr>
<tr>
<td>Cyclin-dependent kinase 8</td>
<td>CDK8</td>
<td>X85753</td>
<td>0.781</td>
</tr>
<tr>
<td>Cyclin F</td>
<td>CCNF</td>
<td>U17105</td>
<td>0.831</td>
</tr>
<tr>
<td>Protein phosphatase 1D</td>
<td>PPM1D</td>
<td>U78305</td>
<td>0.801</td>
</tr>
<tr>
<td>Intracellular signaling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein D</td>
<td>APOD</td>
<td>J02611</td>
<td>0.714</td>
</tr>
<tr>
<td>Apolipoprotein H</td>
<td>APOH</td>
<td>X53595</td>
<td>0.815</td>
</tr>
<tr>
<td>Rapamycin target protein</td>
<td>FRAP1</td>
<td>L34075</td>
<td>0.845</td>
</tr>
<tr>
<td>GRB2-associated binder-1</td>
<td>GAB1</td>
<td>U43885</td>
<td>0.715</td>
</tr>
<tr>
<td>NCK adaptor protein 1</td>
<td>NCK1</td>
<td>X17576</td>
<td>0.794</td>
</tr>
<tr>
<td>Mitogen-activated protein kinase 14</td>
<td>MAPK14</td>
<td>U19775</td>
<td>0.855</td>
</tr>
<tr>
<td>Cell membrane receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Androgen receptor</td>
<td>AR</td>
<td>M20132</td>
<td>0.810</td>
</tr>
<tr>
<td>EphB 6</td>
<td>EPHB6</td>
<td>D83402</td>
<td>0.875</td>
</tr>
<tr>
<td>Insulin receptor</td>
<td>INSR</td>
<td>M10051</td>
<td>0.864</td>
</tr>
<tr>
<td>Toll-like receptor 3</td>
<td>TLR3</td>
<td>U88879</td>
<td>0.777</td>
</tr>
<tr>
<td>Toll-like receptor 6</td>
<td>TLR6</td>
<td>AB020807</td>
<td>0.763</td>
</tr>
<tr>
<td>Metabolic enzyme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytochrome P450, family 24</td>
<td>CYP24A1</td>
<td>L13286</td>
<td>0.772</td>
</tr>
<tr>
<td>Cytochrome P450, subfamily IID, polypeptide 6</td>
<td>CYP2D6</td>
<td>M20403</td>
<td>0.769</td>
</tr>
<tr>
<td>Phosphodiesterase 4A</td>
<td>PDE4A</td>
<td>L20965</td>
<td>0.842</td>
</tr>
<tr>
<td>Phosphodiesterase 4C</td>
<td>PDE4C</td>
<td>Z46632</td>
<td>0.703</td>
</tr>
<tr>
<td>Cell death regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caspase-activated DNase</td>
<td>DFFB</td>
<td>AF039210</td>
<td>0.841</td>
</tr>
<tr>
<td>Programmed cell death 1</td>
<td>PDCD1</td>
<td>U64863</td>
<td>0.779</td>
</tr>
<tr>
<td>DNA damage response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRAP-related protein 1</td>
<td>ATR</td>
<td>U49844</td>
<td>0.883</td>
</tr>
<tr>
<td>Mouse double min 2</td>
<td>MDM2</td>
<td>Z12020</td>
<td>0.688</td>
</tr>
<tr>
<td>Growth/differentiation factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone morphogenetic protein 5</td>
<td>BMP5</td>
<td>M38693</td>
<td>0.831</td>
</tr>
<tr>
<td>Chemokine ligand 1</td>
<td>CCL1</td>
<td>M57502</td>
<td>0.861</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat shock 70kDa protein 4</td>
<td>HSPA4</td>
<td>AB023420</td>
<td>0.793</td>
</tr>
<tr>
<td>Retinoid X receptor, β</td>
<td>RXRB</td>
<td>M84820</td>
<td>0.873</td>
</tr>
<tr>
<td>IFN-α</td>
<td>IFNA1</td>
<td>V00537</td>
<td>0.766</td>
</tr>
</tbody>
</table>

* Symbol in LocusLink database.
* GenBank accession number.
* pN1:pN0: ratio of mean expression values (node positive cases to negative cases).
* Expression status in node positive cases compared with node negative cases.
association with advancing pT stage included several cell cycle regulators and transcription factors. Among the cell cycle regulators, CDC2L1 is known to induce Fas-mediated apoptosis in malignant melanoma cells (39), and RBL2, which is a member of the retinoblastoma gene family, is known to be a tumor suppressor gene that is positively correlated with the prognosis of the retinoblastoma gene family, is known to be a tumor suppressor gene that is positively correlated with the prognosis of ESCC (40). NCOA1 and NCOA2 are ligand-dependent transcription factors involved in a wide array of biological processes, including cell differentiation, reproduction, and homeostasis (41). PAX8 promotes thyroid follicular carcinoma progression by forming fusion proteins with PPARγ (42). These results suggest that the expression of cell cycle regulators acting as a tumor suppressor is maintained in the early pT stages of ESCC. Also, the transcription factors are implicated in the promotion of tumor cell proliferation through activation of multiple signal transductions. Additional investigation of these genes may provide a clue for interpretation of the mechanism by which malignancy is acquired in concert with tumor progression.

In regard to lymph node metastasis, which is another important factor for determination of clinical stage, we also confirmed the validity of the identified genes according to the reproducibility and performance of a classifier beyond the sample sets. It was here shown that cell adhesion proteins (CD33, GJB1, ITGAX, and ITGB6) and cell membrane receptors (CD2, ERBB4, and CSF2RB) showed higher expression in node-positive cases. In contrast, cell cycle regulators (CDC16, CDC8, CCNF, and PPM1D) and intracellular signaling molecules (APOD, APOH, FRAP1, GAB1, NCK1, and MAPK14) were expressed in lower levels in node-positive cases. These results might indicate that the invasive character of tumor cells is determined by cellular interaction with the extracellular environment rather than the proliferative potentials. It is interesting that the spectrum of expressed genes in relation to lymph node metastasis is distinct from that for tumor progression.

In the present study, we used feature-subset selection to extract genes relevant to specific characteristics of esophageal cancers. This approach has been successfully used to extract essential features to classify cDNA array samples into categories (28, 43). If one merely selects features according to statistical differences, the selected data still contain noise that is not well separated as distinct spatial distributions, as was seen in the present study using expectation maximization-algorithm and k-means clustering (respective separation rates: 86.1% and 69.4% for lymph node status). This may often be the case in cDNA array data, which are rather susceptible to randomness in expressions and their measurements. Feature-subset selection that searches for optimal features in separation of the feature subspaces can extract genes of which the expression is consistently different in the given classes. In this regard, we used the k-nearest neighbor classifier rather than the more sophisticated classifiers that transform feature subspaces or unequally place weights on the feature values. We consider that such transformation or weighting can result in a local minimal solution that overestimates trivial features that do not have true biological significance. For the final validation of the selected features, we used a similar strategy, the use of a PNN that divides the feature space with simulated Bayesian posterior probabilities. The ensemble classification using the AdaBoost algorithm combined the multiple feature sets (models) and performed well. The good classification performance obtained for the independent set of samples, by the use of a classifier system different from that used in feature selection, indeed demonstrates the validity of the selected features. We propose that the presently shown method provides a useful means to extract biologically significant differences from noisy cDNA array data.

Although the selected genes that are characteristic of tumor progression and invasiveness in ESCC displayed some interesting trends, these features do not uncover the whole appearance of the malignant nature underlying complex gene network, because these were selected from only a part (3–4%; 1289 genes) of the whole human genes. Hence, additional investigation should be warranted to better understand the malignant nature of esophageal SCC by making the selected features as a clue.

**ACKNOWLEDGMENTS**

We thank the following hospitals in Hokkaido prefecture, Japan for providing esophageal cancer samples: Asahikawa City Hospital, Hakodate Central General Hospital, Hakodate Medical Association Hospital, Hokkaido Gastroenterology Hospital, Keiyukai Sapporo Hospital, Kitami Red Cross Hospital, Kiyota Hospital, Kushiro City Hospital, Kushiro Red Cross Hospital, National Hakodate Hospital, Obihiro-Kosei General Hospital, Oji General Hospital, Shinminton Muroran General Hospital, Teine Kijinkai Hospital, and Tonan Hospital.

**REFERENCES**


Clinical Cancer Research

Gene-Expression Profile Changes Correlated with Tumor Progression and Lymph Node Metastasis in Esophageal Cancer

Eiji Tamoto, Mitsuhiro Tada, Katsuhiro Murakawa, et al.


Updated version Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/10/11/3629

Cited articles This article cites 37 articles, 11 of which you can access for free at: http://clincancerres.aacrjournals.org/content/10/11/3629.full#ref-list-1

Citing articles This article has been cited by 7 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/10/11/3629.full#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.