

# Genome-wide gene expression profile analysis of esophageal squamous cell carcinomas

TAKUMI YAMABUKI<sup>1,2</sup>, YATARO DAIGO<sup>1</sup>, TATSUYA KATO<sup>1</sup>, SATOSHI HAYAMA<sup>1</sup>,  
TATSUHIKO TSUNODA<sup>4</sup>, MASAKI MIYAMOTO<sup>2</sup>, TOMOO ITO<sup>3</sup>, MASAHIRO FUJITA<sup>5</sup>,  
MASAO HOSOKAWA<sup>5</sup>, SATOSHI KONDO<sup>2</sup> and YUSUKE NAKAMURA<sup>1</sup>

<sup>1</sup>Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639; Departments of <sup>2</sup>Surgical Oncology and <sup>3</sup>Surgical Pathology, Hokkaido University Graduate School of Medicine, Sapporo 060-8638; <sup>4</sup>Laboratory for Medical Informatics, SNP Research Center, Riken (Institute of Physical and Chemical Research), Kanagawa 230-0045; <sup>5</sup>Keiyukai Sapporo Hospital, Sapporo 003-0027, Japan

Received December 19, 2005; Accepted February 6, 2006

**Abstract.** To identify the molecules involved in esophageal carcinogenesis and those applicable as novel tumor markers and for the development of new molecular therapies, we performed gene expression profile analysis of 19 esophageal squamous cell carcinoma (ESCC) cells purified by laser microbeam microdissection (LMM). Using a cDNA microarray representing 32,256 genes, we identified 147 genes that were commonly up-regulated and 376 transcripts that were down-regulated in ESCC cells compared with non-cancerous esophageal epithelial cells. A comparison of clinicopathological data with the expression profiles of the 19 ESCCs identified 20 genes whose expression levels could most significantly separate cases with lymph node metastasis from those without. In addition, immunohistochemical analysis of candidate tumor markers on tissue microarrays demonstrated transactivation of a secretory protein, transforming growth factor  $\alpha$  (TGFA) in the great majority of 228 ESCC cases and an association of their expression with the poor prognosis of patients. Our data provide valuable information for establishing novel diagnostic markers for early diagnosis and choice of therapy, and identifying therapeutic target molecules for the development of novel anti-cancer drugs and immunotherapy in esophageal cancer treatment.

## Introduction

Esophageal squamous cell carcinoma (ESCC) is a tumor with a very poor prognosis, and most patients are at advanced stages

at the time of diagnosis (1). Despite using modern surgical techniques combined with various treatment modalities, such as radiotherapy and chemotherapy, the overall 5-year survival rate remains at 40-60% (2). Several tumor markers, such as squamous cell carcinoma antigen (SCC), carcinoembryonic antigen (CEA), and cytokeratin 19-fragment (CYFRA 21-1), are used in clinical diagnosis as well as in patient follow-up. In addition, serum levels of midkine (MDK), CD147, matrix metalloproteinase-2 (MMP-2), MMP-9, and MMP-26 in patients with ESCC were reported to be associated with poor prognosis (3-6). However, no tumor marker has proven to be useful for the detection of ESCC at a potentially curative stage (early stage), and no practical prediction marker is presently available for the selection of treatment modalities for individual patients. Therefore, new diagnostic and therapeutic strategies are urgently needed, i.e. novel tumor markers that can detect this disease at an early stage and be applied to individualized treatments based on the biological characteristics of cancer cells.

Analysis of gene expression profiles on cDNA microarray enables us to perform a comprehensive analysis of gene expression profiles in cancer cells for selecting candidates for the development of novel anti-cancer drugs and tumor markers (7-9), and some studies describing gene expression profiles of human ESCC tissues have been performed (2,10,11). However, since ESCCs contain various types of cells, such as mesenchymal and inflammatory cells, at different proportions (12), all of the previous expression data on human ESCC obtained by use of bulk tumor tissue do not accurately reflect gene expression changes during esophageal carcinogenesis.

In this study, we performed a genome-wide analysis of gene expression profiles of 19 ESCC cells purified by laser microbeam microdissection (LMM) using a cDNA microarray containing 32,256 genes. In the process, we identified a number of genes that were potentially good candidates for the development of novel diagnostic markers, therapeutic drugs, and/or immunotherapy as well as a small subset of biomarkers for predicting the presence of lymph node metastasis. In addition, we confirmed that overexpression of transforming growth factor  $\alpha$  (TGFA) has independent

---

*Correspondence to:* Dr Yusuke Nakamura, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan  
E-mail: yusuke@ims.u-tokyo.ac.jp

**Key words:** gene expression profile, esophageal squamous cell carcinoma, lymph node metastasis

prognostic value for esophageal cancer by tissue microarray analysis of 228 cases of archived ESCCs.

## Materials and methods

**Tissue samples and microdissection.** Nineteen ESCC tissue samples (5 female and 14 male patients; median age, 66.6 years; range, 51-76 years) were obtained with written informed consent along with adjacent normal esophageal tissue samples from patients undergoing surgery at Hokkaido University and its affiliated hospitals (Sapporo, Japan). All cancer tissues were histopathologically diagnosed as squamous cell carcinoma of the esophagus. Clinical information was obtained from medical records. Clinical stage was judged according to the UICC TNM classification. All specimens were embedded in TissueTek OCT medium (Sakura, Tokyo, Japan) immediately after surgical resection and stored at  $-80^{\circ}\text{C}$  until use. These frozen tissues were cut into  $8\text{ }\mu\text{m}$  sections using a cryostat (Sakura, Tokyo, Japan) and then stained with hematoxylin and eosin (H&E) for histological examination. ESCC cells and corresponding normal esophageal epithelial cells were selectively collected using the EZ cut system with a pulsed ultraviolet narrow beam-focus laser (SL Microtest GmbH, Jena, Germany) according to the manufacturer's protocols. RNA isolated from the microdissected normal esophageal epithelial cells of five individuals was mixed and used as a 'universal control' for all 19 cancer samples on microarray hybridization.

A total of 228 formalin-fixed primary ESCCs (15 female and 213 male patients; median age, 62.1 years; range, 42-81 years) and adjacent normal esophageal tissue samples used for immunostaining on tissue microarrays were obtained from patients undergoing surgery at Keiyukai Sapporo Hospital (Sapporo, Japan). This study and the use of all mentioned clinical materials were approved by individual institutional ethics committees.

**Cell lines.** The human esophageal carcinoma cell lines used in this study were as follows: nine SCC cell lines, TE1, TE2, TE3, TE4, TE5, TE6, TE8, TE9, and TE10; and one adenocarcinoma (ADC) cell line, TE7 (13). All cells were grown in monolayer in appropriate media supplemented with 10% fetal calf serum (FCS) and maintained at  $37^{\circ}\text{C}$  in a humidified atmosphere of 5%  $\text{CO}_2$ .

**cDNA microarray.** We fabricated a genome-wide cDNA microarray with 32,256 cDNAs selected from the UniGene database (build #186) of the National Center for Biotechnology Information (NCBI). This microarray system was essentially constructed as described previously (14). Briefly, the cDNAs were amplified by RT-PCR using poly(A)<sup>+</sup> RNAs isolated from various human organs as templates; the lengths of the amplicons ranged from 200 to 1100 bp, without any repetitive or poly(A) sequences.

**RNA extraction, T7-based RNA amplification, and hybridization.** Total RNA was extracted from each sample of laser-microdissected cells into 350  $\mu\text{l}$  of RLT lysis buffer (Qiagen, Hilden, Germany). The extracted RNAs were treated for 30 min at room temperature with 30 U of DNase I (Roche Diagnostics, Basel, Switzerland) in the presence of 1 U of

RNase inhibitor (Toyobo, Osaka, Japan) to remove any contaminating genomic DNA. After inactivation at  $70^{\circ}\text{C}$  for 10 min, the RNAs were purified using an RNeasy Mini kit (Qiagen) according to the manufacturer's recommendations. All of the DNase I-treated RNAs were subjected to T7-based RNA amplification; two rounds of amplification yielded 50-100  $\mu\text{g}$  of aRNA from each sample. The 2.5- $\mu\text{g}$  aliquots of aRNA from cancer cells or normal esophageal epithelial cells were then labeled by reverse transcription with Cy5-dCTP or Cy3-dCTP (GE Healthcare/Amersham Biosciences Corp., Piscataway, NJ), respectively, as described elsewhere (8,14). Hybridization, washing, and scanning were also carried out according to methods described previously (8,14).

**Data analysis.** Signal intensities of Cy3 and Cy5 from the 32,256 spots were quantified and analyzed by substituting backgrounds, using ArrayVision software (Imaging Research, Inc., St. Catharines, Ontario, Canada). Subsequently, the fluorescent intensities of Cy5 (tumor) and Cy3 (control) for each target spot were adjusted so that the mean Cy5/Cy3 ratio of 52 housekeeping genes on the array was equal to 1. Because data derived from low signal intensities are less reliable, we determined a cutoff value on each slide as described previously (8,14) and excluded genes for further analysis when both Cy3 and Cy5 dyes yielded signal intensities lower than the cutoff. For other genes, we calculated the Cy5: Cy3 ratio using the raw data of each sample.

**Identification of genes associated with a risk of lymph node metastasis.** Genes associated with clinicopathological features, such as lymph node metastasis-positive (node-positive) (p) or node-negative (n), were chosen according to the following two criteria: i) signal intensities are higher than the cutoff value in at least 50% of either group; and ii)  $|\text{Med}_p/\text{Med}_n| \geq 1.2$ , where Med indicates the median derived from log-transformed relative expression ratios in two groups. Genes were selected as candidates when they met the criteria with a permutation P-value  $<0.01$  in each clinicopathological status.

We applied a random permutation test to identify genes that were expressed differently in the two groups. The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) were calculated from the log-transformed relative expression ratios of each gene in node-positive (p) and node-negative (n) cases. A discrimination score (DS) for each gene was defined using the following equation:  $\text{DS} = (\mu_p - \mu_n) / (\sigma_p + \sigma_n)$ .

We carried out permutation tests to estimate the ability of individual genes to distinguish between two groups; samples were randomly permuted between the two classes 10,000 times. Since the DS data set of each gene showed a normal distribution, we calculated a P-value for the user-defined grouping (15). For this analysis, we applied the expression data of 19 cases whose T factors were either 2 or 3, consisting of 13 lymph node-positive and 6 negative cases.

**Calculation of prediction score.** We further calculated the prediction score of lymph node metastasis according to procedures described previously (15). Each gene ( $g_i$ ) votes for either node-positive (p) or node-negative (n) cases depending on whether the expression level ( $x_i$ ) in the sample is closer to the mean expression level of node-positive cases

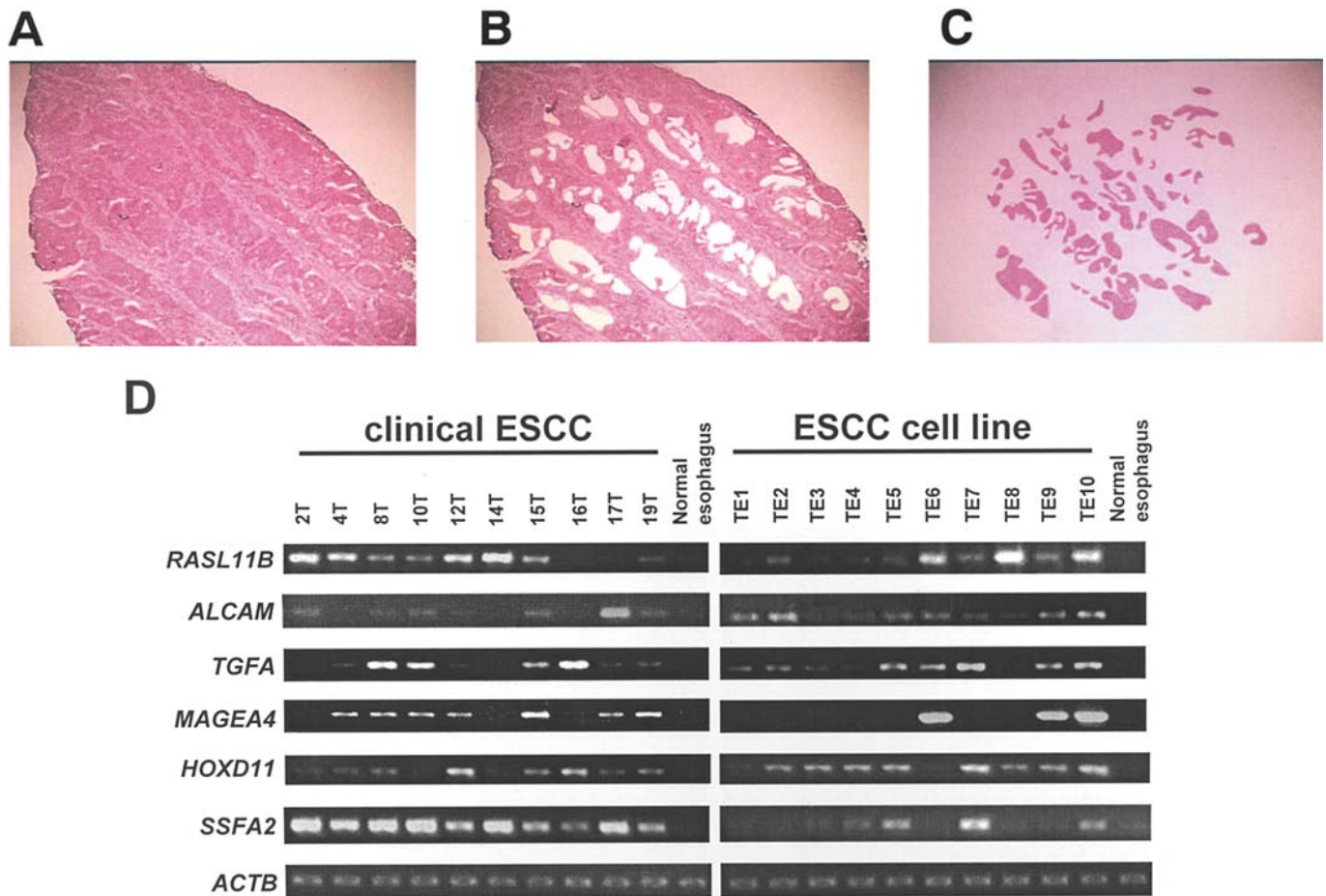


Figure 1. Images illustrating laser-microbeam microdissection (LMM) of a representative ESCC. (A) The samples before dissection, (B) The same sections after microdissection (H&E stain; original magnification, x100). (C) The microdissected cancer cells captured on the collecting cap were also shown. (D) Validation of the six representative gene expression profiles by semi-quantitative RT-PCR.

or node-negative cases in reference samples. The magnitude of the vote ( $v_i$ ) reflects the deviation of the expression level in the sample from the average of the two classes:  $V_i = l \times i - (\mu_n + \mu_p) / 2l$ .

We calculated the total for node-positive cases ( $V_p$ ) and node-negative cases ( $V_n$ ), and the PS value using the following equation:  $PS = [(V_n - V_p) / (V_n + V_p)] \times 100$ , reflecting the margin of victory in the direction of either node-positive cases or node-negative cases. PS values range from -100 to 100, and a higher absolute value of PS reflects a stronger prediction of lymph node metastasis.

**Cross-validation of scores and evaluation of classification.** The prediction scores of all samples were obtained by a leave-one-out approach, in which one sample is withheld, the permutation P-value and mean expression levels are calculated using remaining samples, and the class of the withheld sample is subsequently evaluated by calculating its prediction score. We repeated this procedure for each of the 19 samples. We calculated the classification score (CS) by using the prediction score of node-positive cases ( $PS_p$ ) and node-negative cases ( $PS_n$ ) in each gene set:  $CS = (\mu PS_n - \mu PS_p) / (\sigma PS_n + \sigma PS_p)$ . A larger value of CS indicates a better separation of the two groups when using the predictive scoring system.

**Semi-quantitative RT-PCR.** We selected highly up-regulated genes and examined their expression levels by means of semi-quantitative RT-PCR experiments as previously described (9). A total of 3  $\mu$ g aliquot of amplified RNA from each sample was reverse-transcribed to single-stranded cDNAs using random primer (Roche Diagnostics) and Superscript II (Invitrogen, Carlsbad, CA). Semi-quantitative RT-PCR experiments were carried out with the following sets of synthesized primers specific to the six representative genes up-regulated in ESCCs or with  $\beta$ -actin (ACTB)-specific primers as an internal control: RAS-like, family 11, member B (*RASL11B*), 5'-GAGGAAGAATTGCTTTTCTCTTACC-3' and 5'-TTTTAAAGTGCATCTGTGGAGG-3'; activated leukocyte cell adhesion molecule (*ALCAM*), 5'-CAGCCTCAATGGATACTGGC-3' and 5'-GCTAGAAAGCAAACCTCATGCTCTG-3'; transforming growth factor  $\alpha$  (*TGFA*), 5'-GGAAGATAGACAGCAGCCAAC-3' and 5'-TCTCTCTCTCTGTGTGAACCAC-3'; melanoma antigen, family A, 4 (*MAGEA4*), 5'-CTCCGAGTCCCTGAAGATGAT-3' and 5'-TTCTCTTACATTACGAAGCCA-3'; homeobox D11 (*HOXD11*), 5'-CAGAATCGCAGGATGAAAGATA-3' and 5'-GTGACTCATGCCTTGATATGACA-3'; sperm-specific antigen 2 (*SSFA2*), 5'-GCTCCTTCTCTCATGGATTACCT-3' and 5'-CAAGTGGGTAAAATGCTGTCTTC-3'; *ACTB*, 5'-GAGGTGATAGCATTGCTTTTCG-3' and 5'-CAAGTCAGT



Table I. Representative up-regulated genes with known function.

| Accession no.                  | Symbol         | Gene name  |
|--------------------------------|----------------|--|
| Cell adhesion and cytoskeleton |                |  |
| AL833566                       | <i>ALCAM</i>   | Activated leukocyte cell adhesion molecule   |
| NM_005964                      | <i>MYH10</i>   | Myosin, heavy polypeptide 10, non-muscle   |
| X02761                         | <i>FN1</i>     | Fibronectin 1  |
| L10678                         | <i>PFN2</i>    | Profilin 2   |
| Signal transduction            |                |  |
| NM_002193                      | <i>INHBB</i>   | Inhibin, beta B (activin AB $\beta$ polypeptide)                                       |
| BM677885                       | <i>RASL11B</i> | RAS-like, family 11, member B  |
| BM994359                       | <i>FGFR1</i>   | Fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome) |
| NM_002391                      | <i>MDK</i>     | Midkine (neurite growth-promoting factor 2)  |
| CR596974                       | <i>MLP</i>     | MARCKS-like protein  |
| AK093143                       | <i>SSFA2</i>   | Sperm-specific antigen 2   |
| BC021290                       | <i>IMP-2</i>   | IGF-II mRNA-binding protein 2  |
| Cell proliferation             |                |  |
| NM_021809                      | <i>TGIF2</i>   | TGFB-induced factor 2 (TALE family homeobox)   |
| NM_003236                      | <i>TGFA</i>    | Transforming growth factor, $\alpha$   |
| AA102634                       | <i>TRAF5</i>   | TNF receptor-associated factor 5   |
| BC005963                       | <i>MAGEA3</i>  | Melanoma antigen, family A, 3  |
| NM_002362                      | <i>MAGEA4</i>  | Melanoma antigen, family A, 4  |
| AA781195                       | <i>PRAME</i>   | Preferentially expressed antigen in melanoma   |
| Enzymatic activity             |                |  |
| U07620                         | <i>MAPK10</i>  | Mitogen-activated protein kinase 10  |
| AA255900                       | <i>STK38L</i>  | Serine/threonine kinase 38-like  |
| AB011109                       | <i>ARK5</i>    | AMP-activated protein kinase family member 5   |
| NM_182734                      | <i>PLCB1</i>   | Phospholipase C, $\beta$ 1 (phosphoinositide-specific)                                 |
| NM_002526                      | <i>NT5E</i>    | 5'-nucleotidase, ecto (CD73)   |
| Transcription factors          |                |  |
| AA043381                       | <i>HOXD10</i>  | Homeobox D10   |
| BC001873                       | <i>HEY1</i>    | Hairy/enhancer of split-related with YRPW motif 1                                      |
| Transporter activity           |                |  |
| BU737730                       | <i>RBP1</i>    | Retinol binding protein 1, cellular  |
| N46424                         | <i>RAI14</i>   | Retinoic acid-induced 14   |

GTACAGGTAAGC-3'. PCR reactions were optimized for the number of cycles to ensure product intensity within the linear phase of amplification.

**Immunohistochemistry and tissue microarray.** Tumor tissue microarrays were constructed with 228 formalin-fixed primary esophageal cancers, as described elsewhere (16-18). The tissue area for sampling was selected based on visual alignment with the corresponding H&E-stained section on a slide. Three tissue cores (diameter, 0.6 mm; depth, 3-4 mm) taken from a donor tumor block were placed into a recipient paraffin block with a tissue microarrayer (Beecher Instruments, Sun Prairie, WI). A core of normal tissue was punched from each case, and 5- $\mu$ m sections of the resulting microarray block were used for immunohistochemical analysis.

To investigate the presence of TGFA protein in clinical samples that had been embedded in paraffin blocks, we stained the sections in the following manner. Briefly, 10  $\mu$ g/ml of a mouse monoclonal anti-human TGFA antibody (EMD Biosciences, Inc., San Diego, CA) was added to each slide after the blocking of endogenous peroxidase and proteins, and sections were incubated with horseradish peroxidase-labeled anti-mouse IgG [Histofine Simple Stain MAX PO (G), Nichirei, Tokyo, Japan] as the secondary antibody. Substrate-chromogen was added, and the specimens were counter-stained with hematoxylin. Three independent investigators semi-quantitatively assessed the TGFA positivity as reported previously (19); cytoplasmic staining intensity was recorded as absent (scored as 0), weakly positive (scored as 1+), or strongly positive (scored as 2+), without prior knowledge of

Table II. List of 34 candidate genes associated with lymph node metastasis.

| Rank                 | Permutational<br>P-value | 2-group<br>gap | Accession no. | Symbol               | Gene name  |
|----------------------|--------------------------|----------------|---------------|----------------------|--|
| Up-regulated genes   |                          |                |               |                      |  |
| 1                    | 6.280E-17                | 3.3            | S79880        | <i>NFIL3</i>         | Nuclear factor, interleukin 3 regulated  |
| 2                    | 2.060E-13                | 2.6            | BC007754      | <i>SUV39H2</i>       | Suppressor of variegation 3-9 homolog 2<br>( <i>Drosophila</i> )   |
| 3                    | 2.700E-11                | 8.5            | AK092260      | <i>FLJ34941</i>      | cDNA FLJ34941 fis, clone NT2RP7007480  |
| 4                    | 1.960E-09                | 4.8            | U81787        | <i>WNT10B</i>        | Wingless-type MMTV integration site family,<br>member 10B  |
| 5                    | 2.060E-08                | 5.5            | NM_018372     | <i>RIF1</i>          | Receptor-interacting factor 1  |
| 6                    | 1.310E-06                | 6.8            | AI096953      |                      | EST  |
| 7                    | 1.950E-06                | 6.5            | AW290884      |                      | EST  |
| 8                    | 1.359E-04                | 5.8            | AA909008      | <i>CCNT2</i>         | Cyclin T2  |
| 9                    | 2.595E-04                | 4.2            | AF081466      | <i>KCNJ14</i>        | Potassium inwardly-rectifying channel, subfamily J,<br>member 14   |
| 10*                  | 7.537E-04                | 3.9            | NM_138967     | <i>SCAMP5</i>        | Secretory carrier membrane protein 5   |
| 11                   | 1.597E-03                | 8.5            | L08238        | <i>SUV39H1</i>       | Suppressor of variegation 3-9 homolog 1<br>( <i>Drosophila</i> )   |
| 12                   | 2.186E-03                | 2.4            | NM_005915     | <i>MCM6</i>          | MCM6 minichromosome maintenance deficient 6<br>(MIS5 homolog, <i>S. pombe</i> ) ( <i>S. cerevisiae</i> ) |
| 13                   | 3.941E-03                | 3.3            | BX090035      |                      | Transcribed locus  |
| 14                   | 5.772E-03                | 8.5            | NM_182964     | <i>NAV2</i>          | Neuron navigator 2   |
| 15                   | 6.538E-03                | 1.7            | NM_001011663  | <i>PCGF6</i>         | Polycomb group ring finger 6   |
| 16                   | 6.877E-03                | 4.1            | AK022379      | <i>B2M</i>           | Beta-2-microglobulin   |
| 17                   | 8.639E-03                | 3.2            | BX640908      | <i>EVII</i>          | Ecotropic viral integration site 1   |
| 18                   | 8.913E-03                | 2.4            | BC001076      | <i>C7orf26</i>       | Chromosome 7 open reading frame 26   |
| Down-regulated genes |                          |                |               |                      |  |
| 1                    | 2.94E-05                 | 2.7            | AK000383      | <i>DUSP22</i>        | Dual specificity phosphatase 22  |
| 2                    | 2.99E-05                 | 3.3            | BQ001493      | <i>EHD3</i>          | EH-domain containing 3   |
| 3                    | 5.26E-05                 | 5.2            | AL359590      | <i>PGS1</i>          | Phosphatidylglycerophosphate synthase  |
| 4                    | 0.00010                  | 3.2            | NM_000112     | <i>SLC26A2</i>       | Solute carrier family 26 (sulfate transporter),<br>member 2  |
| 5                    | 0.00056                  | 2.9            | BG939434      | <i>LOC128977</i>     | Hypothetical protein LOC128977   |
| 6                    | 0.00079                  | 4.2            | AK022171      | <i>DCTN2</i>         | Dynactin 2 (p50)   |
| 7                    | 0.00080                  | 1.3            | AL137575      | <i>DKFZp564C1964</i> | mRNA; cDNA DKFZp564C1964 (from clone<br>DKFZp564C1964)   |
| 8                    | 0.00080                  | 2.0            | AW868740      | <i>SYNPO2</i>        | Synaptopodin 2   |
| 9                    | 0.00099                  | 4.9            | CA308965      | <i>RABAC1</i>        | Rab acceptor 1 (prenylated)  |
| 10*                  | 0.00142                  | 3.2            | AA700303      | <i>GGTL3</i>         | Gamma-glutamyltransferase-like 3   |
| 11                   | 0.00496                  | 8.5            | AF024714      | <i>AIM2</i>          | Absent in melanoma 2   |
| 12                   | 0.00583                  | 2.3            | CR605567      | <i>CXorf40</i>       | Chromosome X open reading frame 40   |
| 13                   | 0.00647                  | 1.9            | T62626        | <i>LRP2</i>          | Low density lipoprotein-related protein 2  |
| 14                   | 0.00720                  | 1.2            | BX640898      | <i>THUMPD1</i>       | THUMP domain containing 1  |
| 15                   | 0.00729                  | 1.9            | BC069145      | <i>ATOH1</i>         | Atonal homolog 1 ( <i>Drosophila</i> )   |
| 16                   | 0.00745                  | 1.4            | AF090988      | <i>HRP8BP</i>        | WD repeat domain 57 (U5 snRNP-specific)  |

\*The 20 top-ranked genes whose expression levels could most significantly separate cases with lymph node metastasis from those without metastasis.

clinicopathological data. Cases were accepted as being strongly positive only if the reviewers independently defined them as such.

**Statistical analysis.** Statistical analyses were performed using the StatView statistical program (SAS, Cary, NC). Tumor-specific survival curves were calculated from the date of

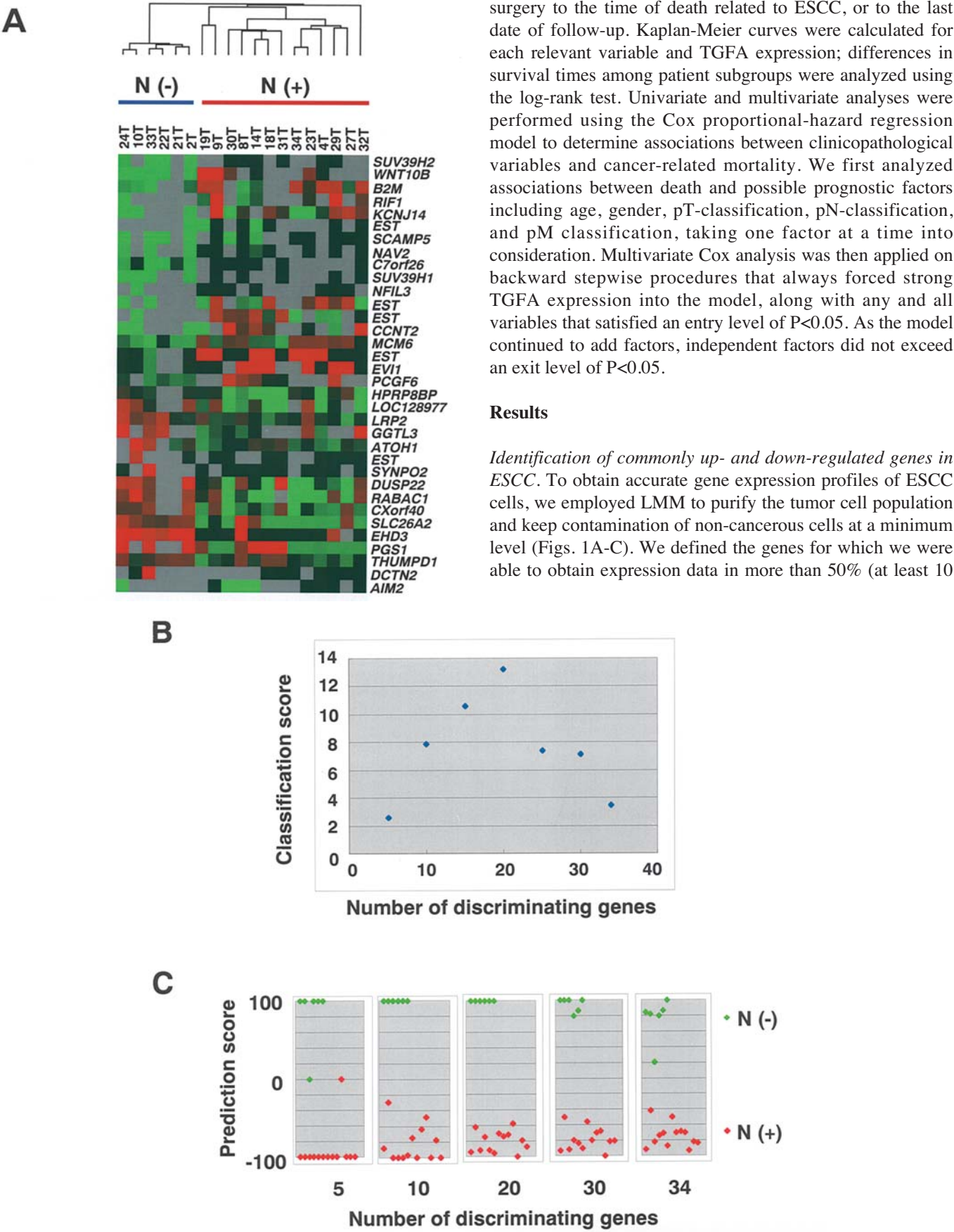


Figure 2. (A) Supervised two-dimensional hierarchical clustering analysis using 34 genes associated with lymph node metastasis that were selected by a random permutation test ( $P$ -value  $< 0.01$ ). (B) Optimization of the number of discriminating genes. Different prediction scores appear when the number of discriminating genes is changed. The number of discriminating gene sets (from 5 to 34) corresponds to the number of selected genes from the top of the rank-ordered list in Table II. A larger classification score (CS) indicates a better separation of the two groups. The best CS was obtained when the scores using only the 20 top-ranked genes were calculated ( $P$ -value  $< 0.0015$ ). (C) The schematic distinction of negative and positive lymph node metastasis cases was verified on the basis of the CS. Green diamonds represent node-negative cases, N(-); red diamonds represent node-positive cases, N(+).

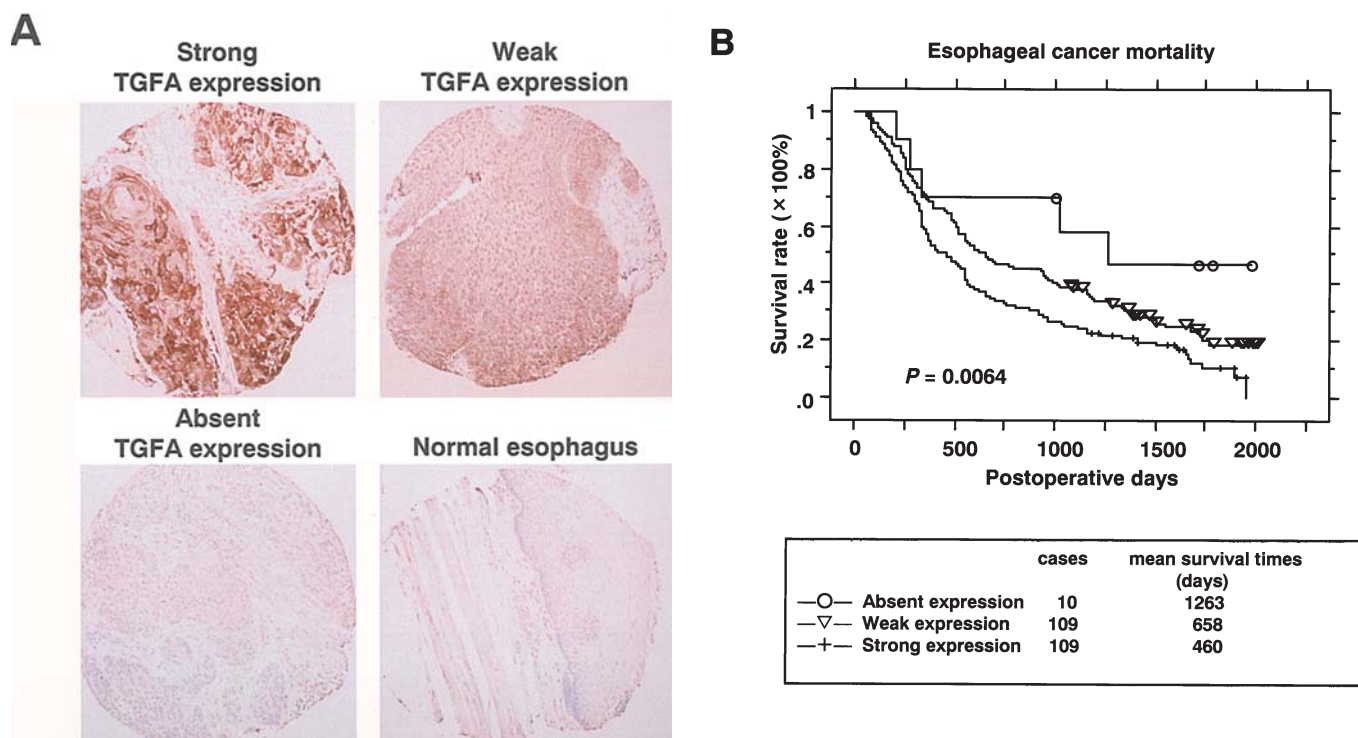


Figure 3. (A) TGFA expression on tissue microarrays. Examples of strong, weak, and absent TGFA expression in esophageal cancer and no expression in normal esophagus are shown (original magnification, x100). (B) Kaplan-Meier analysis survival curves of esophageal cancer patients according to TGFA expression.

of the 19 cases) of the cancers examined as commonly up- or down-regulated according to the following criteria: genes whose expression ratio was  $>5.0$  in ESCC cells were defined as up-regulated and genes whose expression ratio was  $<0.2$  were defined as down-regulated. A total of 147 genes were identified as commonly up-regulated in ESCC (representative known genes are listed in Table I), and 376 genes were commonly down-regulated (data not shown). The up-regulated genes included genes associated with signal transduction, cell proliferation, enzymatic activities, gene transcription, and transporter activity. Some known activated genes in ESCC, such as midkine (*MDK*); *TGFA*; activated leukocyte cell adhesion molecule (*ALCAM*); melanoma antigen, family A, 3 (*MAGEA3*); and melanoma antigen, family A, 4 (*MAGEA4*), appear on our list of up-regulated genes (4,20-22).

**Validation of selected genes by semi-quantitative RT-PCR.** To validate the expression data obtained by microarray analysis, we performed semi-quantitative RT-PCR experiments for a total of 38 representative genes, which were frequently overexpressed in ESCC cases (6 genes were shown in Fig. 1D). The results of RT-PCR experiments using ESCC materials were concordant with microarray data.

**Identification of genes associated with lymph node metastasis.** Status of lymph node metastasis at surgery is an important determinant for the prognosis of patients with ESCC. Therefore, new therapeutic strategies based on information of a biological nature from individual tumors are expected to improve survival for ESCC patients. To find genes that can be used to predict lymph node metastasis, we compared the expression profiles of 13 lymph node-positive cases with those of 6 node-negative cases, and identified 34 genes

that were associated with lymph node status by a random permutation test ( $P < 0.01$ ). Of these, 18 genes were relatively up-regulated, and 16 were down-regulated in node-positive tumors (Table II). Supervised hierarchical clustering analyses using these identified gene sets clearly classified the individual groups according to lymph node status (Fig. 2A). To further determine the minimum number of discriminating genes giving the best separation of the two groups, we rank-ordered the above 34 genes by the magnitude of their permutation P-values, and calculated a classification score (CS) by the leave-one-out test for cross-validation using the top 5, 10, 15, 20, 25, 30, and 34 genes on the rank-ordered list (Fig. 2B), and obtained the best separation when we used the top 20 genes (10 up-regulated and 10 down-regulated genes) for score calculation ( $P$ -value  $< 0.0015$ ; Fig. 2C).

**Evaluation of TGFA as a diagnostic marker for ESCC.** To validate the possibility of applying the overexpressed genes as diagnostic protein markers for ESCC, we carried out immunohistochemical staining on tissue microarray containing tissue sections from 228 ESCC cases that underwent curative surgical resection, which were on our list of overexpressed genes (Table I), with antibody for a secretory protein (*TGFA*). *TGFA* staining was mainly observed in the cytoplasm of tumor cells but was not detected in normal esophageal epithelial cells (Figs. 3A). Of the 228 cases examined, *TGFA* was strongly stained in 109 (47.8%), weakly stained in 109 (47.8%) and not stained in 10 cases (4.4%). The median survival time of ESCC patients was significantly shorter in accordance with the higher expression levels of *TGFA* ( $P = 0.0064$  by log-rank test; Fig. 3B). We also applied univariate analysis to evaluate associations between patient prognosis and several factors, including age, gender, pT stage (tumor depth; T1 versus



T2+T3+T4), pN stage (N0 versus N1), pM stage (metastasis to coeliac nodes or cervical nodes, but no remote organ metastasis; M0 versus M1), and TGFA status (strong positive versus weak or absent). Among those parameters, TGFA status, pT stage, pN stage, and pM stage were significantly associated with poor prognosis. Multivariate analysis using a Cox proportional-hazard model determined that TGFA ( $P=0.0121$ ) was an independent prognostic factor for surgically-treated ESCC patients.

## Discussion

ESCC is known to have the worst prognosis among malignant tumors generated in the digestive tract. It was reported that nearly half of ESCC patients who had curative resection have recurrent disease at a median follow-up of 37.3 months (23). However, the majority of esophageal cancers are diagnosed at an advanced stage and are unlikely to experience a cure, especially by surgical treatment alone. Hence, extensive research has been directed towards studying adjuvant chemotherapy and chemoradiation, concentrating particularly on defining the best regimens from the standpoint of efficacy, minimal toxicity, and predictive response to treatment. However, previous investigations of neoadjuvant and adjuvant therapies have led to controversial conclusions. In fact, most studies failed to determine an optimal neoadjuvant or adjuvant regimen in terms of patient survival benefit (24,25). Therefore, the development of novel diagnostic tools for the early detection of cancer and the better selection of adjuvant treatment modalities for appropriate patients, as well as novel molecular-targeted therapies involving small-molecule and antibody-based approaches and cancer-vaccine immunotherapies, are urgently required.

Analysis of expression profiles by means of cDNA microarray is now widely used in various cancer cells (7,8,26) and could be the first step to effectively screen diagnostic and therapeutic targets that are specifically expressed in cancer cells (9,19,27-30). Several studies reported gene expression profiles of human ESCC (2,10,11), but these studies used bulk ESCC tissues and appeared to have limitations in distinguishing accurate expressional changes in cancer cells, since ESCC tissues contain various types of non-cancerous cells, such as mesenchymal and inflammatory cells. To avoid the contamination of these non-cancerous cells, we used an LMM system to purify populations of cancerous cells and normal epithelial cells from surgical specimens (8). To our knowledge, this is the first study on gene expression profiles of human ESCC on cDNA microarray combined with an LMM system. We established an in-house genome-wide expression analysis system consisting of 32,256 genes and applied it to ESCCs. The 147 transcripts commonly up-regulated in ESCCs are known to have a variety of functions and include genes encoding cancer-testis or oncofetal antigens as well as those important for signal transduction, cell proliferation, enzymatic activities, gene transcription, and transporter activity. A portion of such genes should be useful as diagnostic/prognostic markers and as a therapeutic target for the development of novel treatments for esophageal cancer. Among them, the genes encoding putative tumor-specific transmembrane or secretory proteins are considered

to have significant advantages, as they are present on the cell surface or within the extracellular space, and/or in serum, making them easily accessible as molecular markers and therapeutic targets. Some tumor-specific markers available at present, such as CYFRA or Pro-GRP, are transmembrane/secretory proteins (31,32); rituximab (Rituxan), a humanized monoclonal antibody against CD20-positive lymphomas, provides proof that targeting specific cell-surface proteins can result in significant clinical benefits (33). In this study, we selected an up-regulated gene (*TGFA*) encoding secretory protein and examined the protein expression status by means of tissue microarray analysis. *TGFA*, one of the ligands for EGFR, is a type I transmembrane protein. Previous reports showed a correlation between high-level expression of *TGFA* and poor prognosis in esophageal cancer (34,35), while another study reported that low-level expression of *TGFA* was associated with a worse prognosis in patients with node-negative esophageal cancer (36). Our study suggested that a higher level of *TGFA* expression was an independent prognostic factor for surgically-treated ESCC patients. In this study, we evaluated the prognostic potential of the marker by using tissue microarray containing archived samples from 228 patients who underwent standardized curative esophagectomy at a single institute. Although further evaluation of *TGFA* for clinical use will be necessary, this marker may support clinicians in selecting the appropriate therapies for individual ESCC patients in advance.

We also tried to establish a predictive scoring system for lymph node metastasis with a limited number of genes by comparing expression profiles of node-positive cases with those of node-negative cases, as lymph node metastasis is a key step in tumor progression and a risk factor for poor prognosis. We identified 20 genes, a combination of which significantly distinguished node-positive from node-negative cases ( $P$ -value  $<0.0015$ ; Fig. 2 and Table II). Among these genes, 10 were relatively up-regulated, and the other 10 were down-regulated in node-positive cases compared with those in node-negative cases. The former genes included key molecules that were indicated to worsen patient prognosis, such as wingless-type MMTV integration site family and member 10B (*WNT10B*) and cyclin T2 (*CCNT2*) (37,38). In logarithmically growing cells, CDK9/*CCNT2* and pRb are located in a nuclear multiprotein complex likely involved in the transduction of cellular signals and regulation of cell cycle progression (38). The 20 genes could be useful in selecting patients for adjuvant chemotherapy. It was reported that clinically unsuspected metastasis to the cervicothoracic lymph nodes was present in 36% of patients with esophageal cancer, regardless of cell type (39). A diagnosis based on the minimum gene expression may also have great potential in providing information about the biological nature of cancer cells that cannot be obtained by conventional diagnostic tools; thus improving the survival benefit of adjuvant treatments. However, further validation using larger sets of patients is necessary.

Among the tumor antigens identified to date, cancer-testis antigens (CTAs) have been recognized as a group of highly attractive targets for cancer vaccination (40). Although other factors, such as the *in vivo* immunogenicity of the protein, are also important (41), and further examination is necessary, our candidate genes include known CTAs such as MAGEA3,



MAGEA4 and the preferentially expressed antigen in melanoma (PRAME). Further study using this expression profile should enable us to identify novel CTAs that could be a good target for immunotherapy of ESCC.

In summary, our cDNA microarray analysis combined with an LMM system revealed precise gene expression profiles of ESCC that may be associated with carcinogenesis and lymph node metastasis, yielding valuable insights into the molecular events underlying esophageal carcinogenesis. The application of the gene expression data of ESCC to the selection of candidate molecular targets using tissue microarray analysis could offer a powerful strategy for rapid identification and further evaluation of target molecules for a personalized therapy of this type of tumor.

### Acknowledgements

This work was supported in part by a 'Research for the Future' Program Grant from the Japan Society for the Promotion of Science (#00L01402) to Y.N.

### References

- Shimada H, Nabeya Y, Okazumi S, Matsubara H, Shiratori T, Gunji Y, Kobayashi S, Hayashi H and Ochiai T: Prediction of survival with squamous cell carcinoma antigen in patients with resectable esophageal squamous cell carcinoma. *Surgery* 133: 486-494, 2003.
- Tamoto E, Tada M, Murakawa K, Takada M, Shindo G, Teramoto K, Matsunaga A, Komuro K, Kanai M, Kawakami A, Fujiwara Y, Kobayashi N, Shirata K, Nishimura N, Okushiba S, Kondo S, Hamada J, Yoshiki T, Moriuchi T and Katoh H: Gene-expression profile changes correlated with tumor progression and lymph node metastasis in esophageal cancer. *Clin Cancer Res* 10: 3629-3638, 2004.
- Kawaguchi H, Ohno S, Miyazaki M, Hashimoto K, Egashira A, Saeki H, Watanabe M and Sugimachi K: CYFRA 21-1 determination in patients with esophageal squamous cell carcinoma: clinical utility for detection of recurrences. *Cancer* 89: 1413-1417, 2000.
- Shimada H, Nabeya Y, Tagawa M, Okazumi S, Matsubara H, Kadomatsu K, Muramatsu T, Ikematsu S, Sakuma S and Ochiai T: Preoperative serum midkine concentration is a prognostic marker for esophageal squamous cell carcinoma. *Cancer Sci* 94: 628-632, 2003.
- Ishibashi Y, Matsumoto T, Niwa M, *et al*: CD147 and matrix metallo-proteinase-2 protein expression as significant prognostic factors in esophageal squamous cell carcinoma. *Cancer* 101: 1994-2000, 2004.
- Yamamoto H, Vinitketkumnuen A, Adachi Y, Taniguchi H, Hirata T, Miyamoto N, Noshio K, Imsumran A, Fujita M, Hosokawa M, Hinoda Y and Imai K: Association of matrilysin-2 (MMP-26) expression with tumor progression and activation of MMP-9 in esophageal squamous cell carcinoma. *Carcinogenesis* 25: 2353-2360, 2004.
- Ochi K, Daigo Y, Katagiri T, Nagayama S, Tsunoda T, Myoui A, Naka N, Araki N, Kudawara I, Ieguchi M, Toyama Y, Toguchida J, Yoshikawa H and Nakamura Y: Prediction of response to neoadjuvant chemotherapy for osteosarcoma by gene-expression profiles. *Int J Oncol* 24: 647-655, 2004.
- Kakiuchi S, Daigo Y, Ishikawa N, Furukawa C, Tsunoda T, Yano S, Nakagawa K, Tsuruo T, Kohno N, Fukuoka M, Sone S and Nakamura Y: Prediction of sensitivity of advanced non-small cell lung cancers to gefitinib (Iressa, ZD1839). *Hum Mol Genet* 13: 3029-3043, 2004.
- Kato T, Daigo Y, Hayama S, *et al*: A novel human tRNA-dihydrouridine synthase involved in pulmonary carcinogenesis. *Cancer Res* 65: 5638-5646, 2005.
- Luo A, Kong J, Hu G, Liew CC, Xiong M, Wang X, Ji J, Wang T, Zhi H, Wu M and Liu Z: Discovery of Ca<sup>2+</sup>-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. *Oncogene* 23: 1291-1299, 2004.
- Kihara C, Tsunoda T, Tanaka T, Yamana H, Furukawa Y, Ono K, Kitahara O, Zembutsu H, Yanagawa R, Hirata K, Takagi T and Nakamura Y: Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res* 61: 6474-6479, 2001.
- Nishida K, Mine S, Utsunomiya T, Inoue H, Okamoto M, Udagawa H, Hanai T and Mori M: Global analysis of altered gene expression during the process of esophageal squamous cell carcinogenesis in the rat: a study combined with a laser microdissection and a cDNA microarray. *Cancer Res* 65: 401-409, 2005.
- Nishihira T, Hashimoto Y, Katayama M, Mori S and Kuroki T: Molecular and cellular features of esophageal cancer cells. *J Cancer Res Clin Oncol* 119: 441-449, 1993.
- Ono K, Tanaka T, Tsunoda T, Kitahara O, Kihara C, Okamoto A, Ochiai K, Takagi T and Nakamura Y: Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Res* 60: 5007-5011, 2000.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537, 1999.
- Chin SF, Daigo Y, Huang HE, Iyer NG, Callagy G, Kranjac T, Gonzalez M, Sangan T, Earl H and Caldas C: A simple and reliable pretreatment protocol facilitates fluorescent *in situ* hybridisation on tissue microarrays of paraffin wax embedded tumour samples. *Mol Pathol* 56: 275-279, 2003.
- Callagy G, Cattaneo E, Daigo Y, Happerfield L, Bobrow LG, Pharoah PD and Caldas C: Molecular classification of breast carcinomas using tissue microarrays. *Diagn Mol Pathol* 12: 27-34, 2003.
- Callagy G, Pharoah P, Chin SF, *et al*: Identification and validation of prognostic markers in breast cancer with the complementary use of array-CGH and tissue microarrays. *J Pathol* 205: 388-396, 2005.
- Ishikawa N, Daigo Y, Takano A, Taniwaki M, Kato T, Hayama S, Murakami H, Takeshima Y, Inai K, Nishimura H, Tsuchiya E, Kohno N and Nakamura Y: Increases of amphiregulin and transforming growth factor- $\alpha$  in serum as predictors of poor response to gefitinib among patients with advanced non-small cell lung cancers. *Cancer Res* 65: 9176-9184, 2005.
- Iihara K, Shiozaki H, Oku K, Tahara H, Doki Y, Oka H, Kadowaki T, Iwazawa T, Inoue M and Mori T: Growth-regulatory mechanism of two human esophageal-cancer cell lines in protein-free conditions. *Int J Cancer* 55: 364-370, 1993.
- Verma A, Shukla NK, Deo SV, Gupta SD and Ralhan R: MEMD/ALCAM: a potential marker for tumor invasion and nodal metastasis in esophageal squamous cell carcinoma. *Oncology* 68: 462-470, 2005.
- Zamboni A, Mandruzzato S, Parenti A, Macino B, Dalerba P, Ruol A, Merigliano S, Zaninotto G and Zanovello P: MAGE, BAGE, and GAGE gene expression in patients with esophageal squamous cell carcinoma and adenocarcinoma of the gastric cardia. *Cancer* 91: 1882-1888, 2001.
- Mariette C, Balon JM, Piessen G, Fabre S, Van Seuningen I and Triboulet JP: Pattern of recurrence following complete resection of esophageal carcinoma and factors predictive of recurrent disease. *Cancer* 97: 1616-1623, 2003.
- Law S, Fok M, Chow S, Chu KM and Wong J: Preoperative chemotherapy versus surgical therapy alone for squamous cell carcinoma of the esophagus: a prospective randomized trial. *J Thorac Cardiovasc Surg* 114: 210-217, 1997.
- Ando N, Iizuka T, Ide H, Ishida K, Shinoda M, Nishimaki T, Takiyama Y, Watanabe H, Isono K, Aoyama N, Makuuchi H, Tanaka O, Yamana H, Ikeuchi S, Kabuto T, Nagai K, Shimada Y, Kinjo Y and Fukuda H; Japan Clinical Oncology Group: Surgery plus chemotherapy compared with surgery alone for localized squamous cell carcinoma of the thoracic esophagus: a Japan Clinical Oncology Group Study JCOG9204. *J Clin Oncol* 21: 4592-4596, 2003.
- Kitahara O, Furukawa Y, Tanaka T, *et al*: Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture micro-dissection of tumor tissues and normal epithelia. *Cancer Res* 61: 3544-3549, 2001.
- Suzuki C, Daigo Y, Kikuchi T, Katagiri T and Nakamura Y: Identification of COX17 as a therapeutic target for non-small cell lung cancer. *Cancer Res* 63: 7038-7041, 2003.

28. Suzuki C, Daigo Y, Ishikawa N, Kato T, Hayama S, Ito T, Tsuchiya E and Nakamura Y: ANLN plays a critical role in human lung carcinogenesis through activation of RHOA and by involvement in PI3K/AKT pathway. *Cancer Res* 65: 11314-11325, 2005.
29. Ishikawa N, Daigo Y, Yasui W, Inai K, Nishimura H, Tsuchiya E, Kohno N and Nakamura Y: ADAM8 as a novel serological and histochemical marker for lung cancer. *Clin Cancer Res* 10: 8363-8370, 2004.
30. Furukawa C, Daigo Y, Ishikawa N, Kato T, Ito T, Tsuchiya E, Sone S and Nakamura Y: PKP3 oncogene as prognostic marker and therapeutic target for lung cancer. *Cancer Res* 65: 7102-7110, 2005.
31. Pujol JL, Grenier J, Daures JP, Daver A, Pujol H and Michel FB: Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer. *Cancer Res* 53: 61-66, 1993.
32. Miyake Y, Kodama T and Yamaguchi K: Pro-gastrin-releasing peptide (31-98) is a specific tumor marker in patients with small cell lung carcinoma. *Cancer Res* 54: 2136-2140, 1994.
33. Hennessy BT, Hanrahan EO and Daly PA: Non-Hodgkin's lymphoma: an update. *Lancet Oncol* 5: 341-353, 2004.
34. Yoshida K, Kuniyasu H, Yasui W, Kitadai Y, Toge T and Tahara E: Expression of growth factors and their receptors in human esophageal carcinomas: regulation of expression by epidermal growth factor and transforming growth factor alpha. *J Cancer Res Clin Oncol* 119: 401-407, 1993.
35. Iihara K, Shiozaki H, Tahara H, Kobayashi K, Inoue M, Tamura S, Miyata M, Oka H, Doki Y and Mori T: Prognostic significance of transforming growth factor-alpha in human esophageal carcinoma. Implication for the autocrine proliferation. *Cancer* 71: 2902-2909, 1993.
36. Aloia TA, Harpole DH Jr, Reed CE, Allegra C, Moore MB, Herndon JE II and D'Amico TA: Tumor marker expression is predictive of survival in patients with esophageal cancer. *Ann Thorac Surg* 72: 859-866, 2001.
37. Kasat K, Go V and Pogo BG: Effects of pyrethroid insecticides and estrogen on WNT10B proto-oncogene expression. *Environ Int* 28: 429-432, 2002.
38. Simone C, Bagella L, Bellan C and Giordano A: Physical interaction between pRb and cdk9/cyclinT2 complex. *Oncogene* 21: 4158-4165, 2002.
39. Altorki N, Kent M, Ferrara C and Port J: Three-field lymph node dissection for squamous cell and adenocarcinoma of the esophagus. *Ann Surg* 236: 177-183, 2002.
40. Li M, Yuan YH, Han Y, Liu YX, Yan L, Wang Y and Gu J: Expression profile of cancer-testis genes in 121 human colorectal cancer tissue and adjacent normal tissue. *Clin Cancer Res* 11: 1809-1814, 2005.
41. Wang Z, Zhang Y, Mandal A, Zhang J, Giles FJ, Herr JC and Lim SH: The spermatozoa protein, SLLP1, is a novel cancer-testis antigen in hematologic malignancies. *Clin Cancer Res* 10: 6544-6550, 2004.