*Int Surg* 2011;**96:**207–216

# Gene Expression of Mesenchyme Forkhead 1 (FOXC2) Significantly Correlates With the Degree of Lymph Node Metastasis in Colorectal Cancer

Toshiaki Watanabe[1], Takashi Kobunai[1,2], Yoko Yamamoto[1], Keiji Matsuda[1], Soichiro Ishihara[1], Keijiro Nozawa[1], Hisae Iinuma[1], Takamitsu Kanazawa[3], Toshiaki Tanaka[3], Tsuyoshi Konishi[3,4], Hiroki Ikeuchi[5], Kiyoshi Eshima[2], Tetsuichiro Muto[4], Hirokazu Nagawa[3]

[1]*Department of Surgery, Teikyo University School of Medicine, Itabashi-ku, Tokyo, Japan*

[2]*Tokushima Research Center, Taiho Pharmaceutical Co., Ltd., Kawauchi-cho, Tokushima, Japan*

[3]*Department of Systematic Clinical Oncology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan*

[4]*The Cancer Institute Hospital of JFCR, Kotoh-ku, and Department of Surgical Oncology, University of Tokyo, Bunkyo-ku, Tokyo, Japan*

[5]*Second Department of Surgery, Hyogo College of Medicine, Nishinomiya, Hyogo, Japan*

**In stage III colorectal cancer, patients with N1 stage tumors show poorer outcome than patients with N2 stage tumors. Our objective was to identify genes that are predictive for the presence of lymph node metastasis, and to characterize the aggressiveness of lymph node metastases. Gene expression profiles of colorectal cancer were determined by microarray in training (n = 116) and test (n = 25) sets of patients. We identified 40 discriminating probes in patients with and without lymph node metastases. Using these probes, we could predict the presence of lymph node metastasis with an accuracy of 87.1% (training set) and 76.0% (test set). Among discriminating probes, FOXC2**

Reprint requests: Toshiaki Watanabe, MD, PhD, Department of Surgery, Teikyo University School of Medicine, 2-11-1 Kaga, Itabashi-ku, Tokyo, 173-8605, Japan.
Tel.: +81 3 3964 1231; Fax: +81 3 5375 6097; E-mail: toshwatanabe@yahoo.co.jp

**expression was significantly correlated with the degree of lymph node metastasis. FOXC2 was expressed significantly and disparately in patients with N1 and N2 stage tumors as analyzed by real-time reverse transcriptase–polymerase chain reaction. FOXC2 appears to be involved in determining the aggressiveness of lymph node metastasis in colorectal cancer.**

*Key words:* FOXC2 – GUCY2C – Lymph node metastasis – Colorectal cancer – Microarray – RT-PCR

Lymph node metastasis is a poor prognostic factor for colorectal cancer patients.[1,2] Once lymph node metastasis is confirmed by pathologic examination of surgically resected specimens, adjuvant chemotherapy is usually given. However, the capability to predict the presence of lymph node metastases before surgery is a powerful tool to reveal the optimal extent of surgical lymph node dissection as well as to determine which additional treatments, such as neoadjuvant chemotherapy or chemoradiotherapy, will be required. Furthermore, the postoperative outcome can differ significantly depending on the number of positive lymph node metastases present. Lymph node metastasis is classified into two stages—the N1 stage, consisting of 1 to 3 positive nodes, and the N2 stage, which has 4 or more positive nodes.[3] It is well known that patients with tumors of the N2 stage have a poorer survival outcome than patients at the N1 stage.[1,2]

Therefore, the presence of lymph node metastasis and the degree of lymph node metastasis are important clinical prognostic factors. However, at present there has been sparse data regarding the factors that may determine the aggressiveness of lymph node metastasis. Identifying these factors will help establish a new system for identifying lymph node metastasis.

Previous studies have shown that DNA microarray analysis is useful to predict lymph node metastases in various cancers including breast, head and neck, gastric, and pancreatic cancers.[4–11] Several studies have explored the possibility of predicting lymph node metastases in colorectal cancer patients.[12–15] However, these studies examined only a small number of patients, or otherwise the validation of the predictive model was not performed in a separate test sample set of patients (between 12 and 21 patients). We have recently reported that gene expression profiles can predict the presence of lymph node metastases in colorectal cancer patients.[16] In the present study, we examined a larger sample size of patients and constructed a predictive model of lymph node metastases. Furthermore, we compared gene expression between patients with N1 stage tumors and those with N2 stage tumors in search for markers that may characterize the aggressiveness of lymph node metastasis.

Recently, Waldman *et al*[17] reported that guanylyl cyclase 2C (GUCY2C) is a specific molecular marker for lymph node metastasis in patients with colorectal cancer. GUCY2C is up-regulated and has been characterized as a marker for occult metastases in patients with colorectal cancer who also have histologically negative lymph nodes.[17] Therefore, in the present study, we examined whether or not GUCY2C plays a role in determining the aggressiveness of lymph node metastasis by comparing the expression of GUCY2C in patients with N1 and N2 stage tumors using real-time polymerase chain reaction (PCR).

We herein attempted to identify a set of discriminating genes for characterization of lymph node metastases, and to predict lymph node metastases. Furthermore, we identified a novel gene set that may characterize the degree of lymph node metastasis. To our knowledge this is the first study to examine the differences in gene expression in between N1 and N2 colorectal cancer patients using DNA microarray and real-time PCR.

## Materials and Methods

### *Patient samples*

Informed consent for the collection of specimens was obtained from all participating patients with colorectal cancer, and the study protocol was approved by the local Ethics Committee. One hundred forty-one patients with colorectal cancer who had undergone surgical resections were included in the study. Specimens were removed from colorectal cancer tumors and from the normal colonic mucosa in surgically resected specimens. Samples were snap-frozen in liquid nitrogen and stored at $-80^{\circ}$C until use. Duplicate tumor speci-

mens were formalin-fixed and paraffin embedded for histologic examination. Microscopic examination of normal mucosa verified that no neoplastic cells were present in any of the samples. RNA was extracted from tumor samples after verification that parallel specimens contained more than 70% tumor cells, as described previously.[6]

*RNA isolation and microarray procedures*

Total RNA was isolated from each of the frozen samples using RNeasy Mini Kit (Qiagen, Chatswort, California) for gene expression analysis. Gene expression profiles were analyzed using Affymetrix HG-U133 Plus 2.0 Gene Chip (Affymetrix, Santa Clara, California) according to the manufacturer's recommendations as previously described.[6]

*Selection of discriminating genes and prediction of lymph node metastasis*

A comparative analysis between gene expression in patients with and without lymph node metastases was carried with GeneSpring Software, version 7.2 (Silicongenetics, Redwood, California). Gene expression data, when classified as either flag-P or flag-M in more than 50% of all samples, was uploaded into the software program and then was subsequently normalized in 2 ways: per chip normalization and per gene normalization. For per chip normalization, all expression data on a chip were normalized to the 50th percentile of all values on that chip. Per gene normalization involved individual gene data normalization to the median expression level of that gene across all samples. Samples from 141 patients were divided into a training set (116 samples) and a testing set (25 samples) (Fig. 1). We used only training samples in the original statistical analysis to distinguish a gene set and to build a predictive signature. The testing set was used for independent validation. We first collected 116 training samples to build a predictive model. After establishing this model, we prospectively collected 25 samples for validation of the model. We classified samples into training and test samples in an exact chronologic order. To identify discriminating gene markers in training samples, expression profiles were compared between patients with and without lymph node metastases using unpaired *t* tests (with Welch's correction for unequal variances, using the Benjamini and Hochberg false discovery rate (FDR) controlling procedure) and fold-change analysis. We then carried out supervised class predictions using

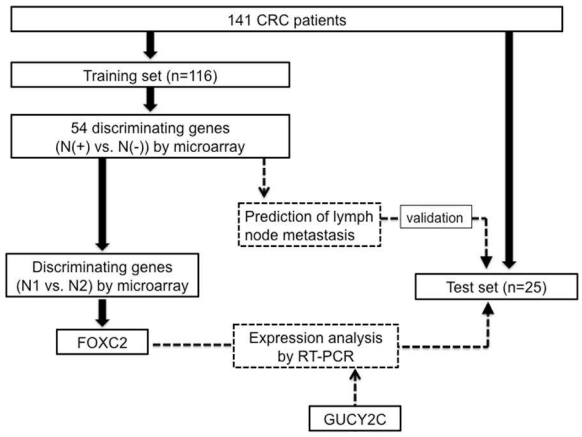


**Fig. 1** Study design. N (+): lymph node metastasis (+); N (−): lymph node metastasis (−).

the k-nearest-neighbor method and a leave-one-out cross-validation with discriminating genes in 116 training samples.[18] We performed a validation test by applying the same method to 25 test samples. Two-dimensional hierarchical clustering was applied to log-transformed data with average-linkage clustering and standard correlation as the similarity metric for the discriminating genes that were differentially expressed between patients with and without lymph node metastases. Any variations in multigene expression between patients with and without lymph node metastases were also analyzed by principal component analysis (PCA).

*Selection of discriminating genes between patients with N1 and N2 tumor stage*

We compared the gene expression and identified discriminating gene markers between patients with N1 and N2 tumor stage in the training set samples with the same method as noted previously to predict lymph node metastasis (Fig. 1).

*Real-time reverse transcriptase–PCR*

The GUCY2C gene expression levels and discriminating genes of patients with N1 and N2 tumor stage were determined by *TaqMan* real-time reverse transcriptase (RT)-PCR (Applied Biosystems, Foster City, California) in a test set of patients, as described previously.[19] Briefly, first-strand complementary DNA (cDNA) was synthesized according to the manufacturer's instructions from total RNA using the High Capacity cDNA Archive kit (Applied Biosystems) in 50-μL reactions, yielding cDNAs for *TaqMan* real-time PCR analysis. Two microliters of

cDNA samples (2 ng/μL) were added to 1.6 μL of RNase-free water, 4 μL of 2× *TaqMan* Universal PCR Master Mix (Applied Biosystems), and 0.4 μL of 20× Primer Probe mix. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as an endogenous control gene (Applied Biosystems). The assay identifications of these genes were Hs00270951_s1, Hs00192035_m1, and Hs99999905_m1 for FOX2C, GUCY2C, and GAPDH, respectively. PCR amplification was performed with an Applied Biosystems Prism 7900HT Sequence Detection System and a 384-well plate format under the following thermal cycler conditions: 2 minutes at 50°C and 10 minutes at 94.5°C for 40 cycles (30 seconds at 97°C and 1 minute at 59.7°C). The relative gene expression was calculated by comparing the ΔCT values as previously described.[20]

### *Statistical analysis*

To analyze the clinicopathologic factors, Fisher's exact test was used for qualitative variables with discrete categories and the Mann-Whitney *U* test for continuous variables. Differences with *P* values of less than 0.05 were considered to be statistically significant.

## Results

### *Gene expression profiling*

#### Class comparison between patients with and without lymph node metastases

Gene expression profiling was performed using a DNA array for all samples consisting of training and test samples. Of 141 patients, 62 were positive and 79 were negative for lymph node metastases. There was no significant difference in the mean number of pathologically examined lymph nodes in patients with or without lymph node metastases (Table 1). Among 116 training samples, 53 were positive and 63 were negative for lymph node metastases. Conversely, 9 samples were positive and 16 negative for lymph node metastases among the test samples. There was no significant difference in the clinicopathologic factors between patients of training and test samples (Table 1). To identify molecular signatures of lymph node metastasis, gene expression profiles from patients with or without lymph node metastases were compared. Using class-comparison analysis, we identified a list of 54 genes that were differentially expressed at statistically significant levels ($P < 0.05$) between patients with lymph node metastases and patients without metastases (Table 2). Thirty-six genes exhibited higher expression levels, and 18 genes were found to decrease in patients with lymph node metastases in comparison to those without metastases.

*Table 1 Characteristics of the patients*

| | | Training patients (n = 116) | Test patients (n = 25) |
|---|---|---|---|
| Age[a] | Mean (STD) | 63.9 (12.5) | 62.9 (14.0) |
| | Range | 31–92 | 22–83 |
| Gender[b] | Men | 77 | 16 |
| | Women | 39 | 9 |
| Tumor location[b] | Colon | 76 | 16 |
| | Rectum | 40 | 9 |
| T classification[b] | T1 | 15 | 5 |
| | T2 | 21 | 3 |
| | T3 | 60 | 12 |
| | T4 | 20 | 5 |
| Lymph node metastasis[b] | N0 | 63 | 16 |
| | N1 | 33 | 6 |
| | N2 | 20 | 3 |
| Number of examined lymph nodes[a] | Mean (STD) | 23.9 (15.6) | 21.9 (14.5) |
| | Range | 2–89 | 4–68 |
| Histologic type[b] | Well | 75 | 16 |
| | Moderate | 24 | 7 |
| | Poor | 11 | 1 |
| | Others | 6 | 1 |

T classification, TNM classification.

[a]Not statistically significant as measured by Mann-Whitney *U* test.

[b]Not statistically significant according to the Fisher's exact test.

#### Class prediction of patients with and without lymph node metastases

We next examined whether expression profiling by microarray analysis was useful to predict whether lymph node metastases are present in patients with cancer. Using 116 training samples, we established a prediction model using the k-nearest-neighbor method and a leave-one-out cross-validation with the 54 genes that were differentially expressed.[18] To determine the number of probes that provided the best separation of patients with and without lymph node metastases, we ranked the 54 genes on the basis of the significance of their FDR *P* values and in decrements of one starting at the bottom of the rank-ordered list (54, 53, etc.). Figure 2 shows the various prediction rates when the number of discriminating probes was changed. The best accuracy rate, 87.1%, was obtained when we used the 40 top-ranked genes (Table 2). The sensitivity, specificity, positive predictive value, and negative predictive value were 92.5%, 82.5%, 81.7%, and 92.3%, respectively. The

predictive performance of the model was evaluated by a permutation-based procedure. The false-finding rate was only 0.026 (*i.e.*, only 26 among 1000 permutations exhibited a higher prediction accuracy than in the actual data).

### *Two-dimensional hierarchic clustering analysis and principal component analysis*

In a hierarchic cluster analysis of the 40 top-ranked genes that were used for prediction of lymph node metastasis, patients with or without lymph node metastases were clustered into two distinct groups (Fig. 3). We also used the top 40 ranked genes to generate a 3-dimensional (from 40-dimensional) plot of the data (Fig. 3). The 3 axes are the first 3 principal components fitted to the patient's molecular profile data. PCA-based multidimensional scaling visualization separated patients with lymph node metastasis and those without into a linearly separable gene expression data space (Fig. 3).

### *Selection of genes discriminating between patients with N1 and N2 tumor stage*

In a training set of 53 patients with lymph node metastases, 33 were stage N1 and 20 stage N2. Gene expression profiles of 54 discriminating genes were compared between patients with N1 and N2 tumor stage, and we identified 3 genes (FOXC2, NM003182; LOC100131039, AF086027; and BF446723) that were differentially and significantly expressed ($P < 0.05$) when comparing patients with N1 and N2 stage. Expression of 3 genes showed correlation with the degree of lymph node metastasis as shown in Fig. 4. Among 3 probes, the mesenchyme forkhead 1 (FOXC2) gene is also annotated at the protein level, and appears to be involved in the metastatic spread of breast cancers.[21] Therefore, FOXC2 was further examined by real-time PCR analysis.

### *FOXC2 and GUCY2C messenger RNA expression analysis by RT-PCR*

In accordance with the results by microarray, RT-PCR analysis revealed that gene expression levels of FOXC2 significantly differed between patients with lymph node metastases and those without metastases. The lymph node-positive group displayed a more than twofold level of FOX2C messenger RNA expression than patients in the lymph node-negative group ($P < 0.05$) (Fig. 5). GUCY2C expression, however, was not differentially expressed between the two groups (Fig. 4).

## Discussion

In the present study we demonstrated that a microarray analysis of gene expression profiles in primary colorectal cancer samples can enable prediction of lymph node metastases with an accuracy rate of 87.1%. We further validated the predictive ability of this model using a new set of test samples, which yielded an accuracy rate of 76.0%. There have so far been few studies predicting lymph node metastasis in colorectal cancer.[12–15] However, these previous studies had several limitations, such as a small number of patients, a low sensitivity of the model, or a lack of validation of these models in an independent set of patients. We herein examined a large number of patients and validated the accuracy of the predictive model in a test set of patients and the accuracy of the model compared with previous studies.

In the clinical setting, the prediction of lymph node lesions may be useful for the treatment of both early and advanced colorectal cancers. In advanced cancers, the prediction of lymph node involvement is essential to determine the magnitude of lymph node dissection that is required. This is of particular importance in the treatment of rectal cancer because extensive surgery can cause impairment of urinary function or sexual dysfunction.[22–24] In early cancers the prediction of lymph node metastasis is most intriguing for the endoscopic treatment of colorectal cancer with submucosal invasion. Because ~10% of early colonic cancers with coincident submucosal invasion exhibit lymph node metastases, additional surgery with lymph node dissection is recommended for endoscopically resected submucosal cancers.[25,26] However, additional surgery for these lesions would require ~90% of patients to undergo unnecessary operations. At present, various histopathologic risk factors have been studied to predict lymph node metastasis. However, no specific markers have yet been established that help avoid unnecessary surgery, which are of great use in clinical practice. The present study suggests that a predictive model that relies on gene signatures may be a novel way to avoid unnecessary surgery for patients with early cancer. However, a limitation of the present study is because only a small number of test set patients were included in the study. Therefore, we believe that a prospective trial with

*Table 2 List of 54 discriminating probes between patients with and without lymph node metastases*

| Probe ID | Fold change[a] (N/Y) | FDR P | Fold change[b] (N1/N2) | Gene symbol | GenBank | Gene title | 40 Predictor probes[c] |
|---|---|---|---|---|---|---|---|
| 1553613_s_at | 0.492 | 0.021 | 0.714 | FOXC1 | AI911330 | Forkhead box C1 | Yes |
| 1555989_at | 0.623 | 0.035 | 1.086 | DAAM1 | AK093733 | Disheveled-associated activator of morphogenesis 1 | Yes |
| 1556034_s_at | 0.632 | 0.021 | 1.101 | MTMR11 | NM_000156 | Myotubularin-related protein 11 | Yes |
| 1556043_a_at | 0.664 | 0.050 | 0.686 | LOC100131039 | AF086027 | Hypothetical protein LOC100131039 | |
| 1556253_s_at | 1.825 | 0.021 | 0.792 | | AI341686 | *Homo sapiens* cDNA FLJ37989 fis, clone CTONG2011676. | Yes |
| 1557363_a_at | 1.968 | 0.030 | 1.633 | PHIP | AV699786 | Pleckstrin homology domain interacting protein | Yes |
| 1559697_a_at | 1.568 | 0.050 | 1.420 | | AW117498 | *Homo sapiens* full length insert cDNA clone YW24B11 | |
| 1562598_at | 0.597 | 0.018 | 1.135 | | BE045998 | *Homo sapiens* LOC374338 (LOC374338), mRNA | Yes |
| 202723_s_at | 0.597 | 0.049 | 0.847 | FOXO1A | AA633619 | Forkhead box O1 | |
| 203841_x_at | 0.530 | 0.030 | 1.063 | MAPRE3 | AA528140 | Microtubule-associated protein, RP/EB family, member 3 | Yes |
| 204537_s_at | 0.654 | 0.030 | 0.912 | GABRE | AU149712 | Gamma-aminobutyric acid (GABA) A receptor, epsilon | Yes |
| 204636_at | 0.387 | 0.022 | 1.577 | COL17A1 | BF056280 | Collagen, type XVII, alpha 1 | Yes |
| 204836_at | 1.730 | 0.021 | 1.258 | GLDC | AW469777 | Glycine dehydrogenase (decarboxylating) | Yes |
| 204885_s_at | 0.427 | 0.022 | 1.832 | MSLN | BF433657 | Mesothelin | Yes |
| 205354_at | 1.524 | 0.050 | 0.992 | GAMT | BE857425 | Guanidinoacetate *N*-methyltransferase | |
| 205727_at | 0.665 | 0.030 | 0.832 | TEP1 | AK000470 | Telomerase-associated protein 1 | Yes |
| 205978_at | 1.774 | 0.039 | 1.003 | KL | BE048919 | Klotho | Yes |
| 206552_s_at | 0.346 | 0.018 | 0.662 | TAC1 | NM_004795 | Tachykinin, precursor 1 | Yes |
| 207596_at | 1.667 | 0.030 | 1.623 | LOC100130703 | BE222220 | Similar to hCG2042168 | Yes |
| 209667_at | 0.666 | 0.036 | 1.160 | CES2 | BE552428 | Carboxylesterase 2 (intestine, liver) | Yes |
| 215102_at | 0.606 | 0.035 | 1.412 | DPY19L1P1 | AB051513 | dpy-19-like 1 pseudogene 1 (*C. elegans*) | Yes |
| 215554_at | 1.663 | 0.049 | 0.572 | GPLD1 | BF033242 | Glycosylphosphatidylinositol-specific phospholipase D1 | |
| 217081_at | 1.613 | 0.036 | 0.806 | OR2H2 | AL031983 | Olfactory receptor, family 2, subfamily H, member 2 | Yes |
| 220803_at | 0.508 | 0.008 | 0.945 | STAMBPL1 | AI732824 | STAM-binding protein-like 1 | Yes |
| 221040_at | 0.620 | 0.022 | 0.934 | CAPN10 | AW014593 | Calpain 10 | Yes |
| 221530_s_at | 0.610 | 0.039 | 1.080 | BHLHB3 | AA890373 | Basic helix-loop-helix domain containing, class B, 3 | |
| 221887_s_at | 0.641 | 0.042 | 1.554 | DFNB31 | AK026768 | Deafness, autosomal recessive 31 | |
| 222071_s_at | 1.699 | 0.039 | 1.088 | SLCO4C1 | BF224430 | Solute carrier organic anion transporter family, member 4C1 | Yes |
| 222196_at | 0.547 | 0.039 | 0.869 | LOC286434 | BG222594 | Hypothetical protein LOC286434 | |
| 223185_s_at | 0.455 | 0.018 | 1.118 | BHLHB3 | NM_004961 | Basic helix-loop-helix domain containing, class B, 3 | Yes |
| 226548_at | 1.747 | 0.026 | 0.812 | SBK1 | NM_007110 | SH3-binding domain kinase 1 | Yes |
| 228057_at | 1.555 | 0.041 | 1.486 | DDIT4L | AU147091 | DNA-damage-inducible transcript 4-like | |
| 229546_at | 0.622 | 0.018 | 0.743 | LOC653602 | AI969112 | Hypothetical LOC653602 | Yes |
| 229927_at | 0.600 | 0.039 | 1.145 | LEMD1 | AK026853 | LEM domain containing 1 | Yes |
| 231029_at | 0.523 | 0.022 | 1.028 | F5 | NM_018515 | Coagulation factor V (proaccelerin, labile factor) | Yes |
| 231578_at | 1.584 | 0.035 | 0.722 | GBP1 | AI935915 | Guanylate-binding protein 1, interferon-inducible, 67 kDa | Yes |
| 231899_at | 1.824 | 0.039 | 0.742 | ZC3H12C | AI740541 | Zinc finger CCCH-type containing 12C | Yes |

*Table 2 Continued*

| Probe ID | Fold change[a] (N/Y) | FDR P | Fold change[b] (N1/N2) | Gene symbol | GenBank | Gene title | 40 Predictor probes[c] |
|---|---|---|---|---|---|---|---|
| 232315_at | 0.571 | 0.039 | 1.104 | LOC400713 | NM_005823 | Zinc finger-like | Yes |
| 232818_at | 1.595 | 0.030 | 1.065 | | NM_023089 | *Homo sapiens* cDNA FLJ12073 fis, clone HEMBB1002387. | Yes |
| 234179_at | 0.650 | 0.030 | 0.677 | | BF446723 | *Homo sapiens* cDNA: FLJ23200 fis, clone KAIA38871. | Yes |
| 234437_at | 0.639 | 0.018 | 1.214 | | NM_000494 | *Homo sapiens* clone 25220 mRNA sequence | Yes |
| 235568_at | 0.645 | 0.039 | 0.819 | C19orf59 | NM_001453 | Chromosome 19 open reading frame 59 | Yes |
| 235746_s_at | 0.627 | 0.039 | 1.089 | PLA2R1 | AW452656 | Phospholipase A2 receptor 1, 180 kDa | |
| 237137_at | 0.615 | 0.039 | 1.336 | SCARNA2 | AK097000 | Small Cajal body-specific RNA 2 | |
| 238803_at | 0.639 | 0.042 | 0.832 | HECTD2 | NM_000170 | HECT domain containing 2 | |
| 239058_at | 0.647 | 0.022 | 0.675 | FOXC2 | NM_003182 | Forkhead box C2 (MFH-1, mesenchyme forkhead 1) | Yes |
| 239492_at | 2.015 | 0.018 | 0.575 | SEC14L4 | AI125859 | SEC14-like 4 (*S. cerevisiae*) | Yes |
| 241252_at | 0.612 | 0.035 | 1.006 | ESCO2 | AW449728 | Establishment of cohesion 1 homolog 2 (*S. cerevisiae*) | Yes |
| 241534_at | 1.552 | 0.039 | 1.061 | | AI378035 | | Yes |
| 242540_at | 0.657 | 0.050 | 0.730 | C11orf47 | AK094329 | Chromosome 11 open reading frame 47 | |
| 242708_at | 0.638 | 0.030 | 0.901 | PEX1 | AI819798 | Peroxisome biogenesis factor 1 | Yes |
| 242822_at | 0.655 | 0.018 | 1.099 | MGC39584 | AF131786 | Hypothetical gene supported by BC029568 | Yes |
| 242996_at | 0.615 | 0.050 | 0.788 | MTRF1 | R12499 | Mitochondrial translational release factor 1 | |
| 244509_at | 1.648 | 0.018 | 1.126 | GPR155 | NM_017597 | G protein-coupled receptor 155 | Yes |

[a]Fold change: lymph metastasis negative/lymph metastasis positive.

[b]Fold change: lymph metastasis N1/lymph metastasis N2.

[c]40 Predictor probes: probes denoted ''yes'' are included in 40 probes that gave the best predictive accuracy.

a larger number of independent patients will confirm the present results in future studies.

By comparing gene expression profiling between patients with and without lymph node metastases, we identified 54 discriminating probes. These 54 probes were used as predictors for the present model. Furthermore, to identify genes that contribute to the aggressiveness of lymph node metastasis, we identified 43 probes whose expression significantly differed between N1 and N2 patients. Among these 43 probes, 3 probes were also included in the 54 discriminating probes between patients with and without lymph node metastases. One of these 3 probes included FOXC2. FOXC2 belongs to the forkhead family of transcription factors, which is characterized by a distinctive DNA-binding forkhead domain. A recent study reported that FOXC2 is central to promoting invasion and metastasis of human breast cancers. Mani *et al*[21] demonstrated that FOXC2 is associated with metastatic capabilities of cancer cells, that FOXC2 expression is required for murine mammary carcinoma cells to metastasize to the lung, and that overexpression of FOXC2 enhances the metastatic ability of mouse mammary carcinoma cells. However, to our knowledge no study has outlined a relationship between FOX2C

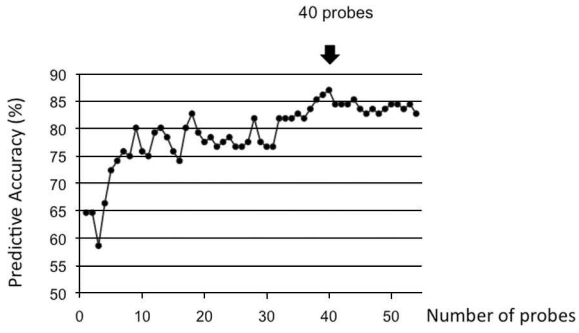


**Fig. 2** Predictive accuracy of different probe sets. Predictive models were established based on different numbers of probes (54, 53, 52, and so on) to determine the number of genes that provided the best separation between patients with and without lymph node metastasis. The best accuracy rate was obtained when the 40 top-ranked probes were used.

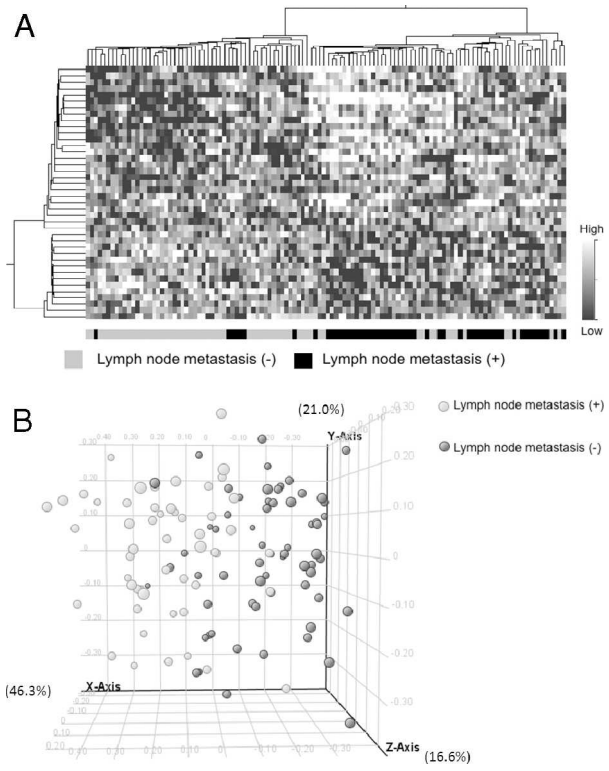


**Fig. 3** Two-way hierarchical clustering and principal component analysis. Supervised clustering of colorectal cancer and 54 genes. Two-way hierarchical clustering was used to order samples (columns) and array targets (rows). Red indicates an overexpression, green indicates an underexpression. At the bottom, yellow indicates patients with lymph node metastases and red indicates those without. (A) Discriminating genes were used to generate a 3-dimensional (from 40-dimensional) plot of the data. The 3 axes are the first 3 principal components fitted to the patients' molecular profile data. The cumulative proportion of the variance captured by each principal component axis is: (*a*) principal component axis 46.3%; (*b*) principal component axis 21.0%; and (*c*) principal component axis 16.6%. PCA-based multidimensional scaling visualization separated patients with lymph node metastasis (red) and those without (yellow) into linearly separable gene expression data space (B).

and human colorectal cancers, particularly for the latter's metastatic capability. We were able to demonstrate that expression level of FOX2C correlates with the degree of lymph node metastasis in patients with N2 stage tumor showing the highest expression and patients without lymph node metastases expressing the lowest level of FOX2C. Furthermore, we were able to confirm that node-positive patients exhibit a higher level of FOXC2 expression than node-negative patients, as determined by RT-PCR analysis. These results suggest

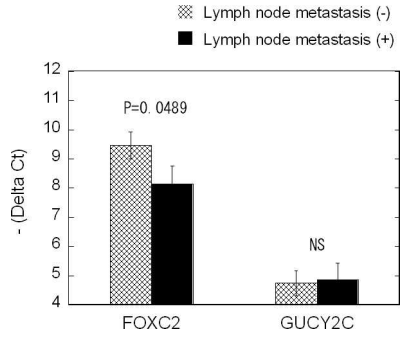


**Fig. 4** Gene expression profiles of 3 genes, NM003182 (FOXC2), AF086027 (LOC100131039), and BF446723, that were differentially and significantly expressed ($P < 0.05$) between patients with tumors of the N1 and N2 stages. The expression of these 3 genes correlated with lymph node metastases.

that FOXC2 may be important for the development of lymph node metastases in colorectal cancer. Future studies will confirm the role of FOXC2 in colorectal cancer carcinogenesis.

Guanylyl cyclase 2C, an intestinal tumor suppressor, is the receptor for the paracrine hormones guanylin and uroguanylin and is a specific molecular marker for metastatic colorectal cancer.[27–31] Recently Waldman *et al*[17] demonstrated that expression of GUCY2C was associated with time to recurrence as well as with disease-free survival in patients with colorectal cancer who also had lymph nodes free of tumor cells as determined by histopathology (pN0). Furthermore, Waldman *et al*[17] suggested that GUCY2C is a marker for occult metastases in lymph nodes in colorectal cancer. These results show the importance of GUCY2C in the development of lymph node metastasis. Therefore, in the present study we examined the levels of

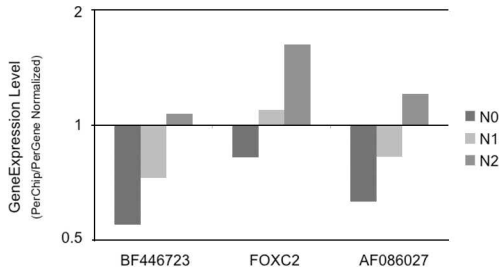


**Fig. 5** FOXC2 and GUCY2C mRNA quantification by RT-PCR analysis. The gene expression levels of FOXC2 differed significantly between patients with and without lymph node metastases. GUCY2C was not significantly different between these two groups.

GUCY2C expression in patients with or without lymph node metastases. Furthermore, we compared the GUCY2C expression between patients with N1 and N2 tumor stage. However, no significant relationship was observed between lymph node metastatic status and GUCY2C expression.

In conclusion, we demonstrated that gene expression profiling is useful for predicting the presence of lymph node metastasis in patients with colorectal cancer. Taken together, data suggest that FOXC2 is contributing to the development of lymph node metastasis in colorectal cancer. Future studies are expected to further confirm the role of FOXC2 in the development of colorectal cancers.

## Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and Intractable Diseases, the Health and Labour Sciences Research Grants from Ministry of Health, Labor and Welfare of Japan. The authors are indebted to Eriko Hashimoto and Riyo Kakimoto for technical and secretarial support.

## References

1. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin* 2007;**57**(1):43–66
2. Greene FL. *AJCC Cancer Staging Manual*. 6th ed. New York, NY: Springer, 2002
3. Compton CC, Greene FL. The staging of colorectal cancer: 2004 and beyond. *CA Cancer J Clin* 2004;**54**(6):295–308
4. Lu X, Lu X, Wang ZC, Iglehart JD, Zhang X, Richardson AL. Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res Treat* 2008;**108**(2):191–201
5. Kim TJ, Choi JJ, Kim WY, Choi CH, Lee JW, Bae DS *et al*. Gene expression profiling for the prediction of lymph node metastasis in patients with cervical cancer. *Cancer Sci* 2008;**99**(1):31–38
6. Watanabe T, Komuro Y, Kiyomatsu T, Kanazawa T, Kazama Y, Tanaka J *et al*. Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene-expression profiles. *Cancer Res* 2006;**66**(7): 3370–3374
7. Watanabe T, Kobunai T, Toda E, Kanazawa T, Kazama Y, Tanaka J *et al*. Gene expression signature and the prediction of ulcerative colitis-associated colorectal cancer by DNA microarray. *Clin Cancer Res* 2007;**13**(2 Pt 1):415–420
8. Watanabe T, Kobunai T, Toda E, Yamamoto Y, Kanazawa T, Kazama Y *et al*. Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers. *Cancer Res* 2006;**66**(20):9804–9808
9. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al*. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**(25):1999–2009
10. Hansel DE, Rahman A, House M, Ashfaq R, Berg K, Yeo CJ *et al*. Met proto-oncogene and insulin-like growth factor binding protein 3 overexpression correlates with metastatic ability in well-differentiated pancreatic endocrine neoplasms. *Clin Cancer Res* 2004;**10**(18 Pt1):6152–6158
11. Hasegawa S, Furukawa Y, Li M, Satoh S, Kato T, Watanabe T *et al*. Genome-wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA microarray representing 23,040 genes. *Cancer Res* 2002;**62**(23):7012–7017
12. Croner RS, Peters A, Brueckl WM, Matzel KE, Klein-Hitpass L, Brabletz T *et al*. Microarray versus conventional prediction of lymph node metastasis in colorectal carcinoma. *Cancer* 2005;**104**(2):395–404
13. Bandres E, Catalan V, Sola I, Honorato B, Cubedo E, Cordeu L *et al*. Dysregulation of apoptosis is a major mechanism in the lymph node involvement in colorectal carcinoma. *Oncol Rep* 2004;**12**(2):287–292
14. Kwon HC, Kim SH, Roh MS, Kim JS, Lee HS, Choi HJ *et al*. Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. *Dis Colon Rectum* 2004;**47**(2): 141–152
15. Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C *et al*. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* 2004;**23**(7):1377–1391
16. Watanabe T, Kobunai T, Tanaka T, Ishihara S, Matsuda K, Nagawa H. Gene expression signature and the prediction of lymph node metastasis in colorectal cancer by DNA microarray. *Dis Colon Rectum* 2009;**52**(12):1941–1948
17. Waldman SA, Hyslop T, Schulz S, Barkun A, Nielsen K, Haaf J *et al*. Association of GUCY2C expression in lymph nodes with time to recurrence and disease-free survival in pN0 colorectal cancer. *JAMA* 2009;**301**(7):745–752
18. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS *et al*. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A* 2000;**97**(1):262–267
19. Camós M, Esteve J, Jares P, Colomer D, Rozman M, Villamor N *et al*. Gene expression profiling of acute myeloid leukemia with translocation t(8;16)(p11;p13) and MYST3-CREBBP rearrangement reveals a distinctive signature with a specific pattern of HOX gene expression. *Cancer Res* 2006;**66**(14):6947–6954
20. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* 2001;**25**(4):402–408

21. Mani SA, Yang J, Brooks M, Schwaniger G, Zhou A, Miura N *et al*. Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proc Natl Acad Sci U S A* 2007;**104**(24):10069–10074
22. Moriya Y, Hojo K, Sawada T, Koyama Y. Significance of lateral node dissection for advanced rectal carcinoma at or below the peritoneal reflection. *Dis Colon Rectum* 1989;**32**(4):307–315
23. Hojo K, Sawada T, Moriya Y. An analysis of survival and voiding, sexual function after wide iliopelvic lymphadenectomy in patients with carcinoma of the rectum, compared with conventional lymphadenectomy. *Dis Colon Rectum* 1989;**32**(2):128–133
24. Nagawa H, Muto T, Sunouchi K, Higuchi Y, Tsurita G, Watanabe T *et al*. Randomized, controlled trial of lateral node dissection vs. nerve-preserving resection in patients with rectal cancer after preoperative radiotherapy. *Dis Colon Rectum* 2001;**44**(9):1274–1280
25. Muto T, Sawada T, Sugihara K. Treatment of carcinoma in adenomas. *World J Surg* 1991;**15**(1):35–40
26. Wilcox GM, Anderson PB, Colacchio TA. Early invasive carcinoma in colonic polyps. A review of the literature with emphasis on the assessment of the risk of metastasis. *Cancer* 1986;**57**(1):160–171
27. Li P, Schulz S, Bombonati A, Palazzo JP, Hyslop TM, Xu Y *et al*. Guanylyl cyclase C suppresses intestinal tumorigenesis by restricting proliferation and maintaining genomic integrity. *Gastroenterology* 2007;**133**(2):599–607
28. Pitari GM, Li P, Lin JE, Zuzga D, Gibbons AV, Snook AE *et al*. The paracrine hormone hypothesis of colorectal cancer. *Clin Pharmacol Ther* 2007;**82**(4):441–447
29. Cagir B, Gelmann A, Park J, Fava T, Tankelevitch A, Bittner EW *et al*. Guanylyl cyclase C messenger RNA is a biomarker for recurrent stage II colorectal cancer. *Ann Intern Med* 1999;**131**(11):805–812
30. Carrithers SL, Barber MT, Biswas S, Parkinson SJ, Park PK, Goldstein SD *et al*. Guanylyl cyclase C is a selective marker for metastatic colorectal tumors in human extraintestinal tissues. *Proc Natl Acad Sci U S A* 1996;**93**(25):14827–14832
31. Frick GS, Pitari GM, Weinberg DS, Hyslop T, Schulz S, Waldman SA. Guanylyl cyclase C: a molecular marker for staging and postoperative surveillance of patients with colorectal cancer. *Expert Rev Mol Diagn* 2005;**5**(5):701–713