associated with migration of mesenchymal stem cells to conditioned medium from tumor cells or bone marrow cells. *Stem Cells* 2007;25:520–8.
37. Beckermann BM, Kallifatidis G, Groth A, Frommhold D, Apel A, Mattern J, Salnikov AV, Moldenhauer G, Wagner W, Diehlmann A, Saffrich R, Schubert M, et al. VEGF expression by mesenchymal stem cells contributes to angiogenesis in pancreatic carcinoma. *Br J Cancer* 2008;99:622–31.
38. Kallifatidis G, Beckermann BM, Groth A, Schubert M, Apel A, Khamidjanov A, Ryschich E, Wenger T, Wagner W, Diehlmann A, Saffrich R, Krause U, et al. Improved lentiviral transduction of human mesenchymal stem cells for therapeutic intervention in pancreatic cancer. *Cancer Gene Ther* 2008;15:231–40.
39. Dwyer RM, Potter-Beirne SM, Harrington KA, Lowery AJ, Hennessy E, Murphy JM, Barry FP, O'Brien T, Kerin MJ. Monocyte chemotactic protein-1 secreted by primary breast tumors stimulates migration of mesenchymal stem cells. *Clin Cancer Res* 2007;13:5020–7.
40. Klopp AH, Spaeth EL, Dembinski JL, Woodward WA, Munshi A, Meyn RE, Cox JD, Andreeff M, Marini FC. Tumor irradiation increases the recruitment of circulating mesenchymal stem cells into the tumor microenvironment. *Cancer Res* 2007;67: 11687–95.
41. Fidler IJ, Wilmanns C, Staroselsky A, Radinsky R, Dong Z, Fan D. Modulation of tumor cell response to chemotherapy by the organ environment. *Cancer Metastasis Rev* 1994;13:209–22.
42. Weaver VM, Gilbert P. Watch thy neighbor: cancer is a communal affair. *J Cell Sci* 2004;117:1287–90.
43. Haviv I, Polyak K, Qiu W, Hu M, Campbell I. Origin of carcinoma associated fibroblasts. *Cell Cycle* 2009;8:589–95.
44. Spaeth EL, Dembinski JL, Sasser AK, Watson K, Klopp A, Hall B, Andreeff M, Marini F Mesenchymal stem cell transition to tumor-associated fibroblasts contributes to fibrovascular network expansion and tumor progression. *PLoS One* 2009;4:e4992.
45. Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer* 2006;6:392–401.
46. Zeisberg EM, Potenta S, Xie L, Zeisberg M, Kalluri R. Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. *Cancer Res* 2007;67:10123–8.
47. Direkze NC, Hodivala-Dilke K, Jeffery R, Hunt T, Poulsom R, Oukrif D, Alison MR, Wright NA. Bone marrow contribution to tumor-associated myofibroblasts and fibroblasts. *Cancer Res* 2004;64:8492–5.

Journal of Pathology
J Pathol 2010; **221**: 275–284
Published online 7 May 2010 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/path.2717

ORIGINAL PAPER

Search for transmembrane protein in gastric cancer by the *Escherichia coli* ampicillin secretion trap: expression of DSC2 in gastric cancer with intestinal phenotype

Katsuhiro Anami,[1] Naohide Oue,[1] Tsuyoshi Noguchi,[2] Naoya Sakamoto,[1] Kazuhiro Sentani,[1] Tetsutaro Hayashi,[1] Takao Hinoi,[3] Masazumi Okajima,[3] Jonathan M Graff[4] and Wataru Yasui[1]*

[1] *Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan*
[2] *Department of Gastrointestinal Surgery, Oita University Faculty of Medicine, Oita, Japan*
[3] *Department of Endoscopic Surgery and Surgical Science, Graduate School of Biomedical Science, Hiroshima University, Hiroshima, Japan*
[4] *Department of Developmental Biology, University of Texas Southwestern Medical Center, Dallas, Texas, USA*

**Correspondence to: Wataru Yasui, Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan e-mail: wyasui@hiroshima-u.ac.jp*

Abstract

Gastric cancer (GC) is one of the most common malignancies worldwide. Genes expressed only in cancer tissue, and especially on the cell membrane, will be useful molecular markers for diagnosis and may also be good therapeutic targets. To identify genes that encode transmembrane proteins present in GC, we generated *Escherichia coli* ampicillin secretion trap (CAST) libraries from two GC cell lines and normal stomach. By sequencing 4320 colonies from CAST libraries, we identified 30 candidate genes that encode transmembrane proteins present in GC. Quantitative reverse transcription–polymerase chain reaction analysis of these candidates revealed that *ZDHHC14*, *BST2*, *DRAM2*, and *DSC2* were expressed much more highly in GC than in 14 kinds of normal tissues. Among these, *DSC2* encodes desmocollin 2, which is one of three known desmocollins. Immunohistochemical analysis demonstrated that 22 (28%) of 80 GC cases were positive for desmocollin 2, and desmocollin 2 expression was observed frequently in GC with the intestinal mucin phenotype. Furthermore, desmocollin 2 expression was correlated with CDX2 expression. These results suggest that expression of desmocollin 2, induced by CDX2, may be a key regulator for GC with the intestinal mucin phenotype. Our results provide a list of genes that have high potential as a diagnostic and therapeutic target for GC.

Keywords: CAST; gastric cancer; desmocollin 2; DSC2

Received 4 February 2010; Revised 9 March 2010; Accepted 29 March 2010

No conflicts of interest were declared.

Introduction

Gastric cancer (GC) is one of the most common human cancers. Cancer develops as a result of multiple genetic and epigenetic alterations [1]. Better knowledge of changes in gene expression that occur during gastric carcinogenesis may lead to improvements in diagnosis, treatment, and prevention. Identification of novel biomarkers for cancer diagnosis and novel targets for treatment is a major goal in this field [2]. Genes encoding transmembrane/secretory proteins expressed specifically in cancers may be ideal biomarkers for cancer diagnosis [3]. If the function of the gene product is involved in the neoplastic process, this gene may constitute a therapeutic target. We previously performed serial analysis of gene expression (SAGE) on four primary GC tissues [4] and identified several GC-specific genes [5]. Of these genes, *regenerating islet-derived family, member 4* (*REG4*, which encodes Reg IV) and *olfactomedin 4* (*OLFM4*, also known as GW112 or hGC-1) encode secreted proteins and serve as highly sensitive serum markers for GC [6,7]. However, through SAGE, we could not find GC-specific genes encoding transmembrane proteins.

In the present study, to identify genes that encode transmembrane proteins present in GC, we generated *Escherichia coli* ampicillin secretion trap (CAST) libraries from two GC cell lines, MKN-1 and MKN-28. CAST is a signal sequence trap method, developed by Ferguson *et al* [8]. Signal peptides target secreted and transmembrane proteins to their appropriate subcellular location, and typically consist of four to 15 hydrophobic amino acids flanked by a basic NH_2 terminus and a polar COOH terminus [9]. A consensus sequence for the signal peptide has not been identified and thus, standard molecular techniques are not well suited for identifying such proteins. CAST is a survival-based signal sequence trap method that exploits the ability of mammalian signal sequences to confer ampicillin

resistance to a mutant β-lactamase lacking the endogenous signal sequence [10]. We report here the identification of several genes that encode transmembrane proteins expressed in GC. Among these, we focused on the *DSC2* gene because this gene is frequently overexpressed in GC and *DSC2* expression is narrowly restricted in normal tissues. *DSC2* encodes desmocollin 2, which is one of the three known desmocollins. Desmocollins are membrane-spanning glycoproteins that form desmosomes along with desmogleins and function as Ca^{2+}-dependent cell adhesion molecules [11]. We examined the expression and distribution of desmocollin 2 in human GC by immunohistochemistry, and the relationship between desmocollin 2 staining and clinicopathological characteristics. Furthermore, because desmocollin 2 is expressed in intestinal metaplasia of the stomach, the association between desmocollin 2 expression and the mucin phenotype in GC was investigated.

Materials and methods

CAST library construction

CAST library construction was performed as described previously [8].

Tissue samples

In total, 457 primary tumour samples were collected from patients diagnosed with GC. Patients were treated at the Hiroshima University Hospital or an affiliated hospital. Because written informed consent was not obtained, for strict privacy protection, identifying information for all samples was removed before analysis. This procedure was in accordance with the Ethical Guidelines for Human Genome/Gene Research of the Japanese Government.

For quantitative reverse transcription-polymerase chain (RT-PCR) reaction, 50 GC samples and corresponding non-neoplastic mucosa samples were used. Samples were frozen immediately in liquid nitrogen and stored at −80 °C until use. Non-cancerous samples of heart, lung, stomach, small intestine, colon, liver, pancreas, kidney, bone marrow, peripheral leukocytes, spleen, skeletal muscle, brain, and spinal cord were purchased from Clontech (Palo Alto, CA, USA).

For immunohistochemical analysis, we used archival formalin-fixed, paraffin-embedded tissues from 407 patients who had undergone surgical excision for GC. Of 407 GC samples, one or two representative tumour blocks were examined from each patient by immunohistochemistry in 80 GC samples. Immunohistochemical analysis of the remaining 327 GC samples was carried out in a tissue microarray (TMA) [12]. Tumour staging was according to the TNM classification system [13]. Histological classification of GC was carried out according to the Lauren classification system [14].

Quantitative RT-PCR and western blot

Quantitative RT-PCR was performed with an ABI PRISM 7700 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) as described previously [15]. Western blot was performed as described previously [16]. Detailed information of quantitative RT-PCR and western blot is described in the Supporting information, Supplementary material.

Evaluation of the specificity of gene expression

To evaluate the specificity of expression of each gene, a specificity index was calculated as follows: first, we identified the normal tissue in which the target gene expression was highest among the 14 normal tissues analysed by quantitative RT-PCR (the mRNA expression level in this tissue was denoted as A). We then identified GC among the nine GC samples in which the target gene expression was highest by quantitative RT-PCR (the mRNA expression level in this tissue was denoted as B). The ratio B to A was defined as the specificity index. When the specificity index of the target gene was greater than or equal to 10, the gene was considered to show high specificity for GC. When the specificity index of the target gene was less than 10 and greater than or equal to 2, the gene was considered to show low specificity for GC. When the specificity index of the target gene was less than 2, the gene was considered to show no specificity for GC.

Immunohistochemistry

Immunohistochemical analysis was performed with a Dako Envision + Mouse Peroxidase Detection System (Dako Cytomation, Carpinteria, CA, USA). Antigen retrieval was done by heating in citrate buffer (pH 6.0) in a microwave oven for 30 min. After peroxidase activity was blocked with 3% H_2O_2–methanol for 10 min, sections were incubated with normal goat serum (Dako Cytomation) for 20 min to block nonspecific antibody binding sites. Sections were incubated with the following antibody dilutions: mouse monoclonal anti-desmocollin 2 (LifeSpan BioSciences, Inc, Seattle, WA, USA), 1 : 50, and mouse monoclonal anti-CDX2, 1 : 50 (BioGenex, San Ramon, CA, USA). Sections were incubated with primary antibody for 1 h at room temperature, followed by incubations with Envision+ anti-mouse peroxidase for 1 h. For colour reaction, except for the normal heart, sections were incubated with the DAB Substrate-Chromogen Solution (Dako Cytomation) for 10 min. For colour reaction of the normal heart, sections were incubated with the AEC Substrate-Chromogen Solution (Dako Cytomation) for 10 min. Sections were counterstained with 0.1% haematoxylin. A result was considered positive if at least 10% of the cells were stained. When fewer than 10% of cancer cells were stained, the immunostaining was considered negative.

J Pathol 2010; 221: 275–284
www.thejournalofpathology.com

Phenotypic analysis of GC

GCs were classified into four phenotypes: gastric (G) type, intestinal (I) type, gastric and intestinal mixed (GI) type, and unclassified (N) type. For phenotypic expression analysis of GC, we performed immunohistochemical analysis (as described above) with four antibodies: anti-MUC5AC (Novocastra, Newcastle, UK) as a marker of foveolar epithelial cells in the stomach; anti-MUC6 (Novocastra) as a marker of pyloric gland cells in the stomach; anti-MUC2 (Novocastra) as a marker of goblet cells in the small intestine and colorectum; and anti-CD10 (Novocastra) as a marker of microvilli of absorptive cells in the small intestine and colorectum. The criteria [17] for the classification of G-type and I-type GCs were as follows: GCs in which more than 10% of cells in the section expressed at least one gastric epithelial cell marker (MUC5AC or MUC6) or intestinal epithelial cell marker (MUC2 or CD10) were classified as G-type or I-type cancers, respectively. Sections that showed both gastric and intestinal phenotypes were classified as GI type, and those that lacked both the gastric and the intestinal phenotypes were classified as N type.

Cell lines

Eight cell lines derived from human GC were used. Detailed information of the GC cell lines is described in the Supporting information, Supplementary material. To assess *DSC2* as a CDX2-regulated gene, the HT-29 cell line was inoculated with the pCDX2-ER vector as described previously [18]. The pCDX2-ER vector encodes a chimeric protein in which full-length CDX2 sequences are fused upstream of a mutated oestrogen receptor (ER) ligand-binding domain. The mutated ER ligand-binding domain no longer binds oestrogen, but retains the ability to bind tamoxifen. In HT-29 cells expressing the CDX2–ER fusion protein (HT-29/CDX2-ER), CDX2 function was activated by the addition of 4-hydroxytamoxifen (4-OHT) (Sigma Chemical, St Louis, MO, USA) to the growth medium at a final concentration of 500 nmol/l.

RNA interference (RNAi) and cell growth and *in vitro* invasion assays

To knockdown the endogenous *DSC2*, RNAi was performed. Short interfering RNA (siRNA) oligonucleotides for *DSC2* and a negative control were purchased from Invitrogen (Carlsbad, CA, USA). 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) [19] and modified Boyden chamber assays were performed to examine cell growth and invasiveness, respectively (see Supporting information, Supplementary material).

Statistical method

Correlations between clinicopathological parameters and desmocollin 2 protein expression were analysed by Fisher's exact test. A p value of less than 0.05 was considered statistically significant.

Results

Generation of CAST libraries

To identify genes that encode transmembrane proteins present in GC, we generated CAST libraries from two GC cell lines (MKN-1 and MKN-28) and normal stomach. We sequenced 1440 ampicillin-resistant colonies from each CAST library. We compared these sequences with those deposited in the public databases using BLAST (accessed at http://blast.ncbi.nlm.nih.gov/Blast.cgi) and evaluated the subcellular localization of the gene products using data from GeneCards (accessed at http://www.genecards.org/index.shtml). Of 1440 colonies from the MKN-1 cell line, 1035 colonies were human named genes, which included 359 genes. In order for the *E. coli* to survive the antibiotic challenge, the signal sequence and translation initiator ATG codon must be cloned in-frame with the leaderless β-lactamase reporter. Among the 359 genes, 152 genes were cloned in-frame upstream of the leaderless β-lactamase gene. Of the 152 genes, ten genes encoded secreted proteins; 58 genes encoded transmembrane proteins; and 76 genes encoded proteins that were not secreted proteins or transmembrane proteins. The remaining eight genes were not well studied. Of 1440 colonies from the MKN-28 cell line, 1173 colonies expressed human named genes, which included 313 genes. Among the 313 genes, 154 genes were cloned in-frame upstream of the leaderless β-lactamase gene. Of these 154 genes, 13 genes encoded secreted proteins; 59 genes encoded transmembrane proteins; and 74 genes encoded proteins that were not secreted proteins or transmembrane proteins. The remaining eight genes were not well studied. Of 1440 colonies from normal stomach, 1248 colonies were human named genes, which included 158 genes. Among the 158 genes, 77 genes were cloned in-frame upstream of the leaderless β-lactamase gene. Of these 77 genes, 11 genes encoded secreted proteins; 29 genes encoded transmembrane proteins; and 30 genes encoded proteins that were not secreted proteins or transmembrane proteins. The remaining seven genes were not well studied. Because the purpose of this study was to identify genes that encode transmembrane proteins present in GC, we focused on transmembrane proteins.

Identification of genes expressed more highly in GC than in normal tissues

To identify genes expressed specifically in GC, we compared the gene list from each GC cell line CAST library with the normal stomach CAST library. We selected only genes that were detected at least twice

J Pathol 2010; 221: 275–284
www.thejournalofpathology.com

Table 1. Summary of quantitative RT-PCR analysis of candidate genes specifically expressed in gastric cancer

Gene name	Normal organ with highest expression: Organ name	mRNA expression level (A)	GC with highest expression mRNA expression level (B)	Specificity index (B/A)	Number of GC cases with mRNA expression level 2-fold greater than normal organ with highest expression	Number of GC cases with mRNA expression level 10-fold greater than normal organ with highest expression
High specificity						
PCDHB9	Skeletal muscle	2.9*	576*	198.6	7	4
C4orf34	Pancreas	6.6	418.8	63.5	2	1
ADAM17	Small intestine	3.6	187.4	52.1	3	1
TMEM50B	Pancreas	3	97	32.3	5	2
ENPP4	Leukocyte	4.9	96.3	19.7	1	1
SLC38A2	Skeletal muscle	16.2	162.6	10	2	0
Low specificity						
CD151	Small intestine	25.1	191.3	7.6	1	0
RFT1	Pancreas	11.6	78.8	6.8	2	0
CLDN7	Colon	28.9	150.2	5.2	2	0
DSC2	Heart	12.7	40.8	3.2	1	0
BST2	Liver	16.2	50.2	3.1	2	0
ZDHHC14	Brain	8.9	23.6	2.7	1	0
APP	Small intestine	21.1	56.5	2.7	1	0
DRAM2	Heart	2.3	5.8	2.5	1	0
No specificity						
ATP13A3	Heart	6.5	10.6	1.6	0	0
SLC12A2	Colon	1.2	1.8	1.5	0	0
AGPAT5	Pancreas	10	13.7	1.4	0	0
SLC7A1	Pancreas	19.6	28.2	1.4	0	0
ATP8B1	Pancreas	3.7	3.7	1	0	0
LNPEP	Skeletal muscle	4.7	4.3	0.9	0	0
PTP4A1	Skeletal muscle	29.9	24.1	0.8	0	0
XPR1	Heart	20.4	16.4	0.8	0	0
CD63	Pancreas	6.6	4.6	0.7	0	0
TFRC	Heart	15.3	10.9	0.7	0	0
SLC35F5	Skeletal muscle	9.1	5.1	0.6	0	0
SLC38A9	Pancreas	11.3	6.4	0.6	0	0
TM9SF4	Pancreas	13.1	7.6	0.6	0	0
C3orf1	Heart	32.4	12.9	0.4	0	0
SLC44A2	Skeletal muscle	112.2	42.2	0.4	0	0
TM7SF3	Kidney	9.3	2.1	0.2	0	0

*The units are arbitrary. Target mRNA expression levels were standardized to 1.0 μg total RNA from normal stomach as 1.0.

in each GC cell line CAST library. In addition, genes were selected that were not found in the normal stomach CAST library. We obtained 19 candidates from MKN-1 and 17 candidates from MKN-28. In total, 30 individual candidate genes were identified (Table 1). To confirm that these candidates were GC-specific, quantitative RT-PCR was performed to measure the expression of these candidates in nine GC tissue samples and in 14 normal tissues. Representative results are shown in Figure 1. Expression of the 30 candidate genes was not necessarily specific for GC. However, several genes showed much higher expression in GC than in normal tissues. We then focused on cancer specificity by calculating the specificity index for each gene. First, we identified the normal tissue in which the target gene expression was highest (mRNA expression levels are shown as A, Table 1). We then identified the GC among nine tissues in which the target gene expression was highest (mRNA expression levels are shown as B, Table 1). The specificity index (B/A ratio) for each gene is shown in Table 1. Of the 30 candidates, six genes—*PCDHB9, C4orf34, ADAM17, TMEM50B, ENPP4*, and *SLC38A2*—were found to show high specificity for GC, and eight genes—*CD151, RFT1, CLDN7, DSC2, BST2, ZDHHC14, APP*, and *DRAM2*—were found to show low specificity for GC.

mRNA expression of high- and low-specificity genes for GC

Expression of the six high-specificity and eight low-specificity genes for GC was analysed by quantitative RT-PCR in an additional 41 GC samples and corresponding non-neoplastic mucosa samples. We calculated the ratio of target gene mRNA expression levels between GC tissue (T) and corresponding non-neoplastic mucosa (N). T/N ratios greater than two-fold were considered to represent overexpression. Genes showing overexpression in more than 30% of the samples included *ZDHHC14* (19/41, 46%), *BST2* (16/41, 39%), *DRAM2* (15/41, 37%), and *DSC2* (13/41, 32%).

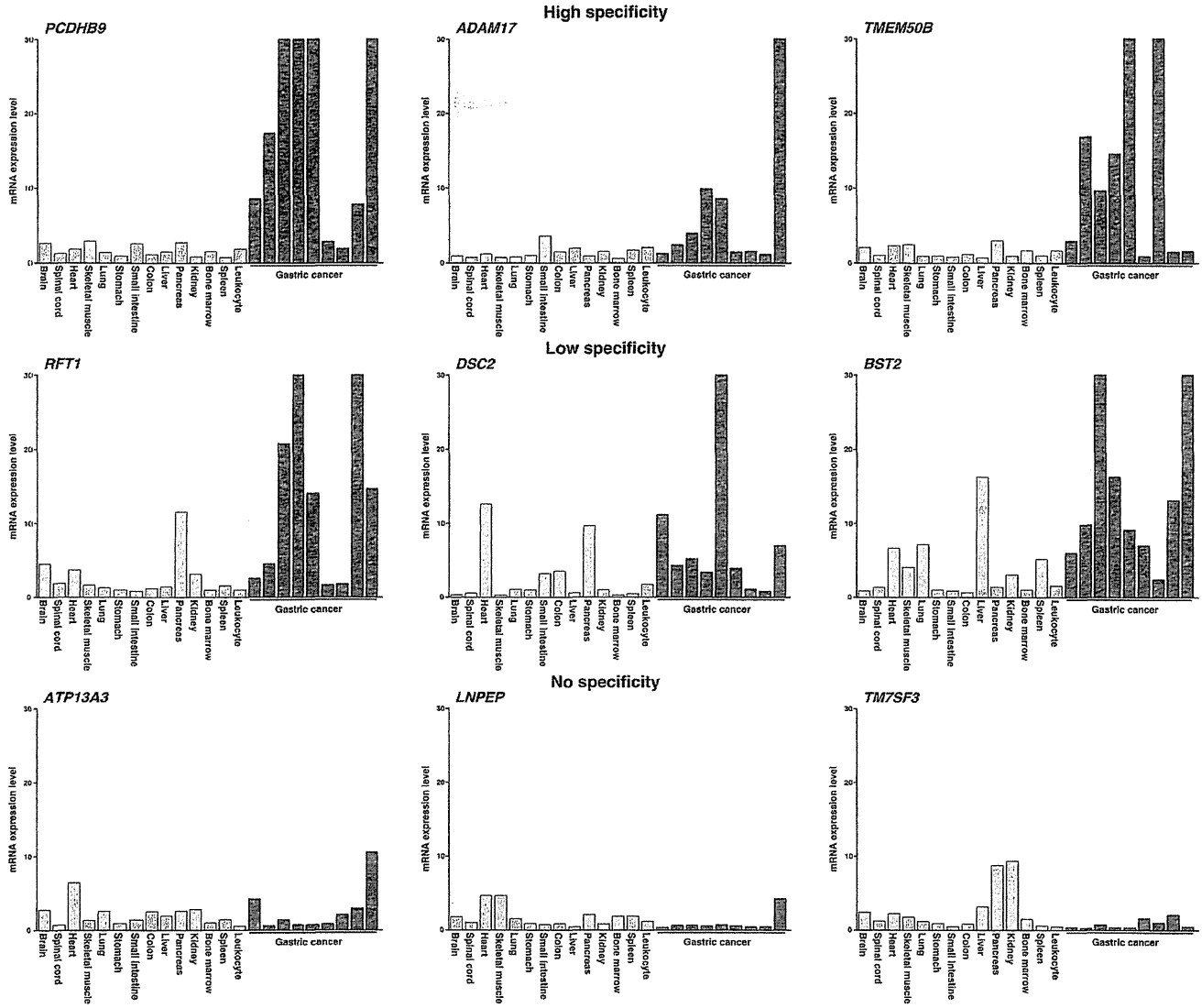

Figure 1. Quantitative RT-PCR analysis of candidate genes that encode cell surface proteins in 14 normal tissues and nine GC samples. Definitions of high specificity, low specificity, and no specificity are described in the Materials and methods section. mRNA expression levels of *PCDHB9*, *ADAM17*, and *TMEM50B* were much higher in GC samples than in normal tissues. In contrast, mRNA expression levels of *ATP13A3*, *LNPEP*, and *TM7SF3* were not significantly different between GC and normal tissues.

Other genes were overexpressed in less than 30% of the samples examined (*RFT1*, 10/41, 24%; *PCDHB9*, 8/41, 20%; *TMEM50B*, 8/41, 20%; *C4orf34*, 6/41, 15%; and *SLC38A2*, 5/41, 12%).

Immunohistochemical staining for DSC2 in GC and non-cancerous tissues

To confirm overexpression of genes whose expression by quantitative RT-PCR was much higher in GC than in normal tissues, we performed immunohistochemical analysis. Among four genes that were frequently overexpressed in GC, we focused on *DSC2* because an antibody against DSC2 protein is commercially available. *DSC2* encodes desmocollin 2, and desmocollins are membrane-spanning glycoproteins that function as Ca^{2+}-dependent cell adhesion molecules [11]. Although desmocollin 2 is expressed in normal colonic epithelium, its expression is reduced in colorectal cancer [20]. There are no reports of desmocollin 2 expression in GC. Immunohistochemical analysis was first performed in non-cancerous tissues with the obvious mRNA expression to serve as positive controls. Immunostaining of desmocollin 2 in the normal heart showed staining of cardiomyocytes (Figure 2A). Immunostaining in the normal pancreas revealed that duct epithelial cells expressed desmocollin 2 on cell membranes (Figure 2B). As reported previously, colonic epithelial cells expressed desmocollin 2 on cell membranes (Figure 2C). These results are consistent with our quantitative RT-PCR results.

Next, immunohistochemistry was performed on 80 GC samples. In non-neoplastic gastric mucosa, weak or no staining of desmocollin 2 was observed in the foveolar epithelium, whereas desmocollin 2 was expressed in the intestinal metaplasia (Figure 2D). Expression of desmocollin 2 was not detected in stromal cells, such as inflammatory cells and fibroblasts. In contrast, GC tissue showed stronger, more extensive

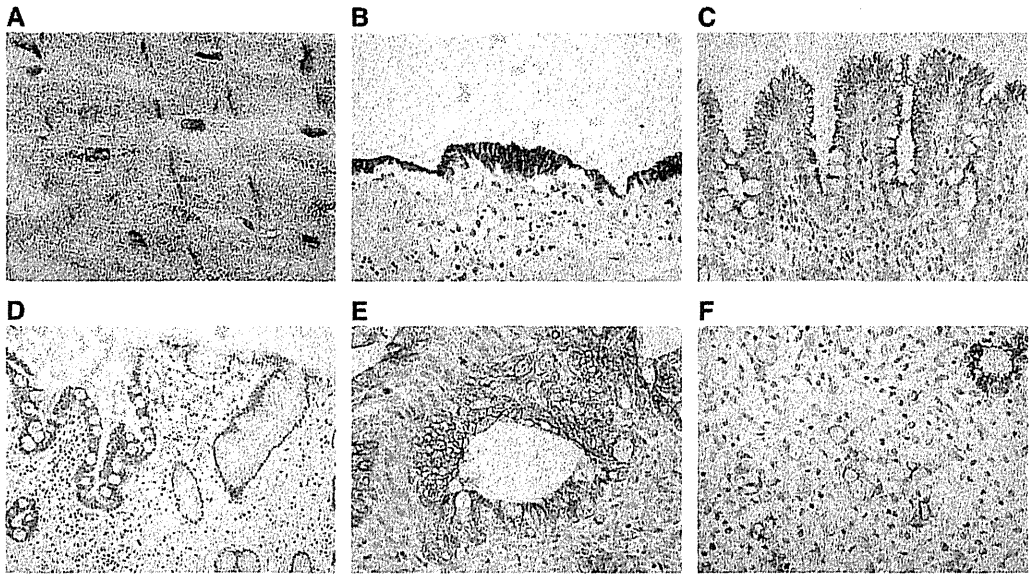

Figure 2. Immunohistochemical analysis of desmocollin 2 in non-neoplastic human tissues and GC tissues. (A) Immunostaining of desmocollin 2 in the normal heart showed staining of cardiomyocytes. Note that the signal is red (using AEC Substrate-Chromogen Solution) to differentiate the staining from the brown pigment in cardiomyocytes. Original magnification: 1000×. (B) Immunostaining of desmocollin 2 in non-neoplastic pancreas revealed that duct epithelial cells expressed desmocollin 2 on cell membranes. Original magnification: 400×. (C) Immunostaining of desmocollin 2 in non-neoplastic colon revealed that colonic epithelial cells expressed desmocollin 2 on cell membranes. Original magnification: 400×. (D) Immunostaining of desmocollin 2 in non-neoplastic gastric mucosa. Desmocollin 2 was present in intestinal metaplasia. Original magnification: 200×. (E) Immunostaining of desmocollin 2 in intestinal-type GC. Desmocollin 2 staining was observed in GC cells on cell membranes. Original magnification: 400×. (F) Immunostaining of desmocollin 2 in diffuse-type GC. Desmocollin 2 staining was observed in GC cells on cell membranes. Original magnification: 400×.

staining than corresponding non-neoplastic mucosa. Desmocollin 2 staining was frequently observed in intestinal-type GC (Figure 2E). Most diffuse-type GCs did not express desmocollin 2; however, some signet ring cell carcinoma cells did stain for desmocollin 2 (Figure 2F). Some GCs showed heterogeneity of desmocollin 2 immunostaining and the percentage of desmocollin 2-positive tumour cells ranged from 0% to 80%. The relationship of desmocollin 2 staining to clinicopathological characteristics was investigated (Table 2). When at least 10% of tumour cells were stained, the immunostaining was considered positive for desmocollin 2. In total, 22 (28%) of 80 GC cases were positive for desmocollin 2. Desmocollin 2 staining was observed more frequently in stage I/II cases (16/41, 39%) than in stage III/IV cases (6/39, 15%, $p = 0.0243$, Fisher's exact test). Moreover, desmocollin 2 staining was detected more frequently in intestinal-type GC (20/46, 46%) than in diffuse-type GC (2/34, 6%, $p < 0.0001$, Fisher's exact test). In the group of 59 advanced GC patients, no statistically significant prognostic impact was found (data not shown).

Desmocollin 2 is expressed in GC with the intestinal phenotype

Despite the usefulness of the Lauren classification, it was previously reported that GC can be subdivided into four phenotypes according to mucin expression (G type, I type, GI type, and N type) [21]. We further investigated the association between desmocollin 2 expression and the mucin phenotype, because desmocollin 2 was detected in intestinal metaplasia of the stomach and colon. Immunohistochemical analysis of gastric (MUC5AC and MUC6) and intestinal (MUC2 and CD10) markers was carried out in a TMA of 327 GC cases. In general, desmocollin 2 was frequently expressed in MUC2-positive GC cases (Figure 3A). Desmocollin 2 staining was detected more frequently in MUC2-positive GC (33/61, 54%) than in MUC2-negative GC (74/266, 28%, $p < 0.0001$, Fisher's exact test). However, desmocollin 2 was also expressed in tumour cells that did not express MUC2. There was no clear relationship between expression of desmocollin 2 and MUC5AC, MUC6, or CD10. On the basis of the expression of these four markers, we classified the 327 GC cases phenotypically as 133 (40%) G type, 55 (17%) I type, 64 (20%) GI type, and 75 (23%) N type. Expression of desmocollin 2 was observed more frequently in I-type GC than in other (G, GI, and N) GC types ($p = 0.0396$, Fisher's exact test) (Figure 3B).

CDX2 induces desmocollin 2 expression

We found that expression of desmocollin 2 is associated with the intestinal mucin phenotype in GC. It is known that intestinal metaplasia of the stomach and GC with the intestinal mucin phenotype are associated with ectopic CDX2 expression [21,22]. It has been reported that CDX2 interacts with the *DSC2* promoter and activates *DSC2* transcription [23]. To confirm the relationship between CDX2 function and *DSC2* gene expression, we studied desmocollin 2 expression in

Table 2. Relationship between desmocollin 2 expression and clinicopathological characteristics in gastric cancer

	Desmocollin 2 expression		
	Positive	Negative	*p* value*
Age, years			
≤65	6 (29%)	15	1.0000
>65	16 (27%)	43	
Sex			
Male	14 (26%)	39	0.7951
Female	8 (30%)	19	
T classification†			
T1	8 (38%)	13	0.2574
T2/3/4	14 (24%)	45	
N classification‡			
N0	10 (31%)	22	0.6130
N1/2/3	12 (25%)	36	
Stage			
I/II	16 (39%)	25	0.0243
III/IV	6 (15%)	33	
Histological classification			
Intestinal	20 (43%)	26	<0.0001
Diffuse	2 (6%)	32	

*Fisher's exact test. †T1, tumour invades lamina propria or submucosa; T2, tumour invades muscularis propria or subserosa; T3, tumour penetrates serosa; T4, tumour invades adjacent structures. ‡N0, no regional lymph node metastasis; N1, metastasis in one to six regional lymph nodes; N2, metastasis in seven to 15 regional lymph nodes; N3, metastasis in more than 15 regional lymph nodes.

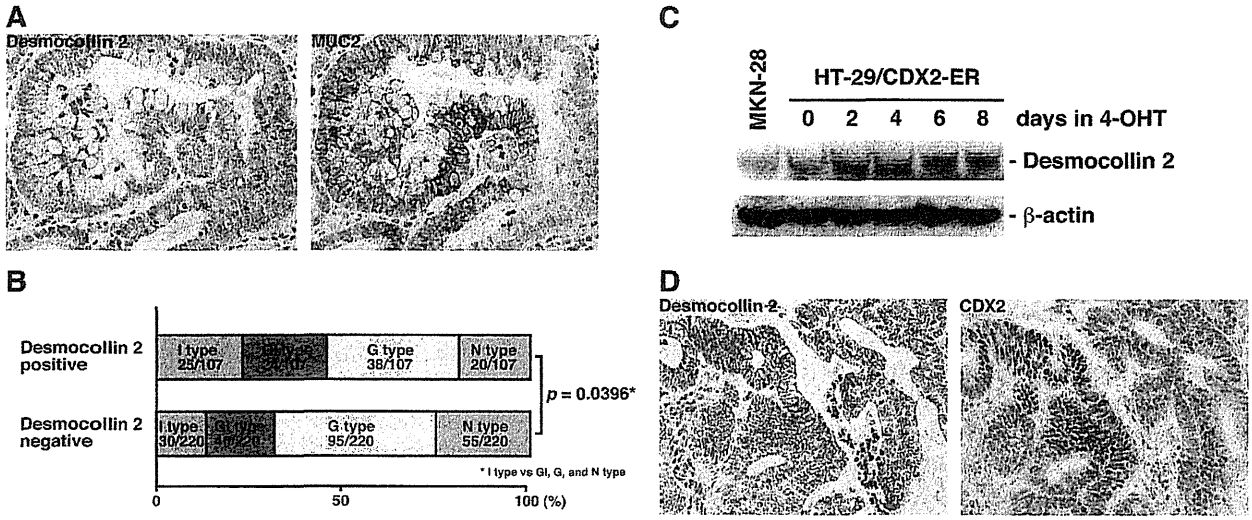

Figure 3. Expression of the mucin phenotype of GC. (A) Expression of desmocollin 2 and MUC2 in GC. In this case, desmocollin 2 was detected in GC cell membranes. MUC2 was also expressed, but several desmocollin 2-positive GC cells did not express MUC2. Original magnification: 400×. (B) Summary of desmocollin 2 expression and expression of the mucin phenotype. Desmocollin 2-positive GC cases were found more frequently in I-type GC than in the other (GI, G, and N) GC types. *Fisher's exact test. (C) Western blot analysis of desmocollin 2. Time course of desmocollin 2 induction in response to activation of a CDX2–ER fusion protein by 4-OHT. (D) Expression of desmocollin 2 and CDX2 in GC. Original magnification: 400×.

an HT-29-derived line with tightly regulated CDX2 activity [18,24]. We used a polyclonal HT-29 cell line that had been transduced with a vector encoding a chimeric CDX2–ER fusion protein. Treatment of the HT-29/CDX2-ER cell line with 4-OHT resulted in strong induction of desmocollin 2 protein expression within 2 days (Figure 3C). These results indicate that *DSC2* is a direct or primary target gene regulated by CDX2.

We then investigated whether CDX2 induces desmocollin 2 expression in GC tissue. Immunohistochemical analysis of CDX2 was carried out in a TMA (Figure 3D). Of 327 GC cases, 74 (23%) GC cases were positive for CDX2. Desmocollin 2 staining was found more frequently in CDX2-positive GC (44/74, 59%) than in CDX2-negative GC (63/253, 25%, $p < 0.0001$, Fisher's exact test). However, desmocollin 2 was also stained in tumour cells that were negative for CDX2.

Effect of desmocollin 2 inhibition on cell growth and invasive activity

Desmocollin 2-positive GC cases were observed more frequently in stage I/II than in stage III/IV, suggesting that desmocollin 2 may be associated with tumour progression; however, the biological significance of

desmocollin 2 in human cancer has not been studied. To investigate the possible anti-proliferative effect of desmocollin 2 knockdown, we performed an MTT assay 8 days after siRNA transfection in the MKN-1, MKN-28, and MKN-45 cell lines; however, cell viability was not significantly different between desmocollin 2 siRNA-transfected GC cells and negative control siRNA-transfected GC cells (data not shown). Next, to determine the possible role of desmocollin 2 in the invasiveness of GC cells, a transwell invasion assay was performed in the MKN-1, MKN-28, and MKN-45 cell lines. Invasion ability was not significantly different between desmocollin 2 knockdown GC cells and negative control siRNA-transfected GC cells (data not shown).

Discussion

In the present study, we generated CAST libraries from two GC cell lines and identified several genes that encode transmembrane proteins present in GC. Quantitative RT-PCR revealed that *PCDHB9, C4orf34, ADAM17, TMEM50B, ENPP4, SLC38A2, CD151, RFT1, CLDN7, DSC2, BST2, ZDHHC14, APP*, and *DRAM2* were expressed much more highly in GC than in 14 types of normal tissues. Quantitative RT-PCR in an additional 41 GC samples revealed that *ZDHHC14, BST2, DRAM2*, and *DSC2* were overexpressed in more than 30% of the samples. Because these genes were identified by CAST analysis of GC cell lines and quantitative RT-PCR analysis of bulk GC tissues, immunohistochemistry was required to determine which cells expressed these genes. Among these four genes, only an antibody against desmocollin 2 was commercially available. Therefore, we performed immunohistochemical analysis of desmocollin 2 in GC and found that 22 (28%) of 80 GC cases were positive for desmocollin 2. Desmocollin 2 was frequently expressed in MUC2-positive GC cases, and desmocollin 2 expression was observed frequently in I-type GC. Ectopic CDX2 expression plays an important role in the development of I-type GC [21,22] and it has been reported that CDX2 activates *DSC2* transcription [23]. In the present study, we showed that desmocollin 2 expression was correlated with CDX2 expression in GC tissue. Taken together, expression of desmocollin 2, induced by CDX2, may be a key factor mediating the biological behaviour of I-type GC.

In quantitative RT-PCR, *DSC2* was overexpressed in 32% of the samples examined. The observation is based on a comparison of GC with corresponding non-neoplastic mucosa. I-type GC contains I-type GC cells, whereas corresponding non-neoplastic mucosa is composed of a mixed population of foveolar, oxyntic, and endocrine cells. Because *DSC2*, similar to *CDX2*, is a marker of the intestinal mucin phenotype in both normal and GC cells, it is not surprising that *DSC2* is enriched in GC when compared with corresponding non-neoplastic mucosa containing foveolar epithelium. Therefore, it is correct to say that *DSC2* is ectopically expressed in GC, rather than overexpressed in GC.

In immunohistochemical analysis, desmocollin 2 is variably expressed in the normal upper gastrointestinal tract and thus, desmocollin 2 expression in I-type GC is not specific. Knockdown of desmocollin 2 by RNAi did not affect cell growth and invasiveness in GC cell lines. These results indicate that desmocollin 2 expression is not likely to be involved in carcinogenesis. Because desmocollin 2 is a membrane-spanning glycoprotein that functions as a Ca^{2+}-dependent cell adhesion molecule, there is a possibility that desmocollin 2 plays an important role in glandular formation of GC cells. In the present study, desmocollin 2 staining was observed more frequently in stage I/II cases than in stage III/IV cases. Therefore, loss of desmocollin 2 expression may be associated with tumour progression.

Several lines of evidence suggest that distinct cell surface or cell adhesion molecules are expressed in G-type GC and I-type GC. The claudin protein family, which comprises 24 members, is a component of tight junctions. All claudins are 20–27 kDa proteins with four transmembrane domains [25]. Expression of claudin-18 is frequently down-regulated in I-type GC [26]. In contrast, ectopic expression of claudin-3 and claudin-4 is found in I-type GC [27]. Liver-intestine cadherin (LI-cadherin), which is a member of the cadherin family of cell adhesion molecules [28], is one of the CDX2-regulated genes [18]. Ectopic expression of LI-cadherin is observed in I-type GC [29]. Taken together with the present results, it seems that the biological characteristics of I-type GC are different from those of G-type GC and thus, classification of GC based on the mucin phenotype is important for treatment of GC. In fact, distinctive responses to chemotherapy between G-type and I-type GC have been reported [30].

In summary, our present study yielded a list of genes that encode transmembrane proteins present in GC. We found that desmocollin 2 is expressed in GC and that expression of desmocollin 2 is associated with the intestinal mucin phenotype. We identified several genes by quantitative RT-PCR that have not been implicated previously in GC. Our current data also provide information with respect to the expression of these genes throughout the body. Additional examination including functional and immunohistochemical analysis will certify whether the genes identified in the present study may constitute not only a diagnostic but also a therapeutic target.

Acknowledgment

We thank Mr Shinichi Norimura for his excellent technical assistance and advice. This work was carried out with the kind cooperation of the Research Center for

Molecular Medicine, Faculty of Medicine, Hiroshima University. We thank the Analysis Center of Life Science, Hiroshima University for the use of their facilities. This work was supported in part by Grants-in-Aid for Cancer Research from the Ministry of Education, Culture, Science, Sports, and Technology of Japan; in part by a Grant-in-Aid for the Third Comprehensive 10-Year Strategy for Cancer Control and for Cancer Research from the Ministry of Health, Labour, and Welfare of Japan; and in part by a grant (07-23911) from the Princess Takamatsu Cancer Research Fund.

References

Note: References 31–33 are cited in the Supporting information to this article.

1. Yasui W, Oue N, Kitadai Y, Nakayama H. Recent advances in molecular pathobiology of gastric carcinoma. In *The Diversity of Gastric Carcinoma Pathogenesis: Diagnosis and Therapy*, Kaminishi M, Takubo K, Mafune K (eds). Springer: Tokyo, 2005; 51–71.
2. Yasui W, Oue N, Ito R, Kuraoka K, Nakayama H. Search for new biomarkers of gastric cancer through serial analysis of gene expression and its clinical implications. *Cancer Sci* 2004; **95**: 385–392.
3. Buckhaults P, Rago C, St Croix B, Romans KE, Saha S, Zhang L, *et al*. Secreted and cell surface genes expressed in benign and malignant colorectal tumors. *Cancer Res* 2001; **61**: 6996–7001.
4. Oue N, Hamai Y, Mitani Y, Matsumura S, Oshimo Y, Aung PP, *et al*. Gene expression profile of gastric carcinoma: identification of genes and tags potentially involved in invasion, metastasis, and carcinogenesis by serial analysis of gene expression. *Cancer Res* 2004; **64**: 2397–2405.
5. Aung PP, Oue N, Mitani Y, Nakayama H, Yoshida K, Noguchi T, *et al*. Systematic search for gastric cancer-specific genes based on SAGE data: melanoma inhibitory activity and matrix metalloproteinase-10 are novel prognostic factors in patients with gastric cancer. *Oncogene* 2006; **25**: 2546–2557.
6. Mitani Y, Oue N, Matsumura S, Yoshida K, Noguchi T, Ito M, *et al*. Reg IV is a serum biomarker for gastric cancer patients and predicts response to 5-fluorouracil-based chemotherapy. *Oncogene* 2007; **26**: 4383–4393.
7. Oue N, Sentani K, Noguchi T, Ohara S, Sakamoto N, Hayashi T, *et al*. Serum olfactomedin 4 (GW112, hGC-1) in combination with Reg IV is a highly sensitive biomarker for gastric cancer patients. *Int J Cancer* 2009; **125**: 2383–2392.
8. Ferguson DA, Muenster MR, Zang Q, Spencer JA, Schageman JJ, Lian Y, *et al*. Selective identification of secreted and transmembrane breast cancer markers using *Escherichia coli* ampicillin secretion trap. *Cancer Res* 2005; **65**: 8209–8217.
9. von Heijne G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res* 1986; **14**: 4683–4690.
10. Kadonaga JT, Gautier AE, Straus DR, Charles AD, Edge MD, Knowles JR. The role of the beta-lactamase signal sequence in the secretion of proteins by *Escherichia coli*. *J Biol Chem* 1984; **259**: 2149–2154.
11. Yin T, Green KJ. Regulation of desmosome assembly and adhesion. *Semin Cell Dev Biol* 2004; **15**: 665–677.
12. Sentani K, Oue N, Noguchi T, Sakamoto N, Matsusaki K, Yasui W. Immunostaining of gastric cancer with neuroendocrine differentiation: Reg IV-positive neuroendocrine cells are associated with gastrin, serotonin, pancreatic polypeptide and somatostatin. *Pathol Int* (in press).
13. Sobin LH, Wittekind CH (eds). *TNM Classification of Malignant Tumors* (6th edn). Wiley-Liss: New York, 2002; 65–68.
14. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* 1965; **64**: 31–49.
15. Kondo T, Oue N, Yoshida K, Mitani Y, Naka K, Nakayama H, *et al*. Expression of POT1 is associated with tumor stage and telomere length in gastric carcinoma. *Cancer Res* 2004; **64**: 523–529.
16. Yasui W, Ayhan A, Kitadai Y, Nishimura K, Yokozaki H, Ito H, *et al*. Increased expression of p34cdc2 and its kinase activity in human gastric and colonic carcinomas. *Int J Cancer* 1993; **53**: 36–41.
17. Mizoshita T, Tsukamoto T, Nakanishi H, Inada K, Ogasawara N, Joh T, *et al*. Expression of Cdx2 and the phenotype of advanced gastric cancers: relationship with prognosis. *J Cancer Res Clin Oncol* 2003; **129**: 727–734.
18. Hinoi T, Lucas PC, Kuick R, Hanash S, Cho KR, Fearon ER. CDX2 regulates liver intestine-cadherin expression in normal and malignant colon epithelium and intestinal metaplasia. *Gastroenterology* 2002; **123**: 1565–1577.
19. Alley MC, Scudiero DA, Monks A, Hursey ML, Czerwinski MJ, Fine DL, *et al*. Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. *Cancer Res* 1988; **48**: 589–601.
20. Khan K, Hardy R, Haq A, Ogunbiyi O, Morton D, Chidgey M. Desmocollin switching in colorectal cancer. *Br J Cancer* 2006; **95**: 1367–1370.
21. Tatematsu M, Tsukamoto T, Inada K. Stem cells and gastric cancer: role of gastric and intestinal mixed intestinal metaplasia. *Cancer Sci* 2003; **94**: 135–141.
22. Silberg DG, Sullivan J, Kang E, Swain GP, Moffett J, Sund NJ, *et al*. Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. *Gastroenterology* 2002; **122**: 689–696.
23. Funakoshi S, Ezaki T, Kong J, Guo RJ, Lynch JP. Repression of the desmocollin 2 gene expression in human colon cancer cells is relieved by the homeodomain transcription factors Cdx1 and Cdx2. *Mol Cancer Res* 2008; **6**: 1478–1490.
24. Hinoi T, Gesina G, Akyol A, Kuick R, Hanash S, Giordano TJ, *et al*. CDX2-regulated expression of iron transport protein hephaestin in intestinal and colonic epithelium. *Gastroenterology* 2005; **128**: 946–961.
25. Tsukita S, Furuse M, Itoh M. Multifunctional strands in tight junctions. *Nature Rev Mol Cell Biol* 2001; **2**: 285–293.
26. Sanada Y, Oue N, Mitani Y, Yoshida K, Nakayama H, Yasui W. Down-regulation of the claudin-18 gene, identified through serial analysis of gene expression data analysis, in gastric cancer with an intestinal phenotype. *J Pathol* 2006; **208**: 633–642.
27. Matsuda Y, Semba S, Ueda J, Fuku T, Hasuo T, Chiba H, *et al*. Gastric and intestinal claudin expression at the invasive front of gastric carcinoma. *Cancer Sci* 2007; **98**: 1014–1019.
28. Berndorff D, Gessner R, Kreft B, Schnoy N, Lajous-Petter AM, Loch N, *et al*. Liver-intestine cadherin: molecular cloning and characterization of a novel Ca(2+)-dependent cell adhesion molecule expressed in liver and intestine. *J Cell Biol* 1994; **125**: 1353–1369.
29. Motoshita J, Nakayama H, Taniyama K, Matsusaki K, Yasui W. Molecular characteristics of differentiated-type gastric carcinoma with distinct mucin phenotype: LI-cadherin is associated with intestinal phenotype. *Pathol Int* 2006; **56**: 200–205.
30. Tajima Y, Shimoda T, Nakanishi Y, Yokoyama N, Tanaka T, Shimizu K, *et al*. Association of gastric and intestinal phenotypic marker expression of gastric carcinomas with tumor thymidylate

synthase expression and response to postoperative chemotherapy with 5-fluorouracil. *J Cancer Res Clin Oncol* 2003; **129**: 683–690.
31. Ochiai A, Yasui W, Tahara E. Growth-promoting effect of gastrin on human gastric carcinoma cell line TMK-1. *Jpn J Cancer Res* 1985; **76**: 1064–1071.
32. Yanagihara K, Seyama T, Tsumuraya M, Kamada N, Yokoro K. Establishment and characterization of human signet ring cell gastric carcinoma cell lines with amplification of the c-myc oncogene. *Cancer Res* 1991; **51**: 381–386.
33. Sakamoto N, Oue N, Noguchi T, Sentani K, Anami K, Sanada Y, *et al*. Serial analysis of gene expression of esophageal squamous cell carcinoma: ADAMTS16 is up-regulated in esophageal squamous cell carcinoma. *Cancer Sci* (in press).

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article.

Supplementary data. Materials and methods.

Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis

Tetsuya Ueda, Stefano Volinia, Hiroshi Okumura, Masayoshi Shimizu, Cristian Taccioli, Simona Rossi, Hansjuerg Alder, Chang-gong Liu, Naohide Oue, Wataru Yasui, Kazuhiro Yoshida, Hiroki Sasaki, Sachiyo Nomura, Yasuyuki Seto, Michio Kaminishi, George A Calin, Carlo M Croce

Lancet Oncol 2010; 11: 136–46
Published Online December 21, 2009
DOI:10.1016/S1470-2045(09)70343-2

See Reflection and Reaction page 106

Department of Molecular Virology, Immunology, and Medical Genetics and Comprehensive Cancer Center, Ohio State University, Columbus, OH, USA (T Ueda MD, S Volinia PhD, H Okumura MD, C Taccioli PhD, H Alder PhD, Prof C-g Liu PhD, Prof C M Croce MD); Department of Surgical Oncology and Digestive Surgery, Graduate School of Medicine and Dental Science, Kagoshima University, Kagoshima, Japan (H Okumura); Department of Gastrointestinal Surgery, Graduate School of Medicine, University of Tokyo, Tokyo, Japan (T Ueda, S Nomura MD, Prof Y Seto MD, Prof M Kaminishi MD); Department of Surgery, Showa General Hospital, Tokyo, Japan (Prof M Kaminishi); Department of Morphology and Embryology, University of Ferrara, Ferrara, Italy (S Volinia, S Rossi PhD); Department of Experimental Therapeutics, Division of Cancer Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA (M Shimizu BS, S Rossi, G A Calin MD, Prof C-G Liu); Department of Molecular Pathology, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan (N Oue MD, Prof W Yasui MD); Department of Surgical Oncology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan (Prof K Yoshida MD); Department of Surgical Oncology, School of Medicine, Gifu University, Gifu, Japan (Prof K Yoshida); and Genetics Division, National Cancer Center Research Institute, Tokyo, Japan (H Sasaki PhD)

Summary

Background Analyses of microRNA expression profiles have shown that many microRNAs are expressed aberrantly and correlate with tumorigenesis, progression, and prognosis of various haematological and solid tumours. We aimed to assess the relation between microRNA expression and progression and prognosis of gastric cancer.

Methods 353 gastric samples from two independent subsets of patients from Japan were analysed by microRNA microarray. MicroRNA expression patterns were compared between non-tumour mucosa and cancer samples, graded by diffuse and intestinal histological types and by progression-related factors (eg, depth of invasion, metastasis, and stage). Disease outcome was calculated by multivariable regression analysis to establish whether microRNAs are independent prognostic factors.

Findings In 160 paired samples of non-tumour mucosa and cancer, 22 microRNAs were upregulated and 13 were downregulated in gastric cancer; 292 (83%) samples were distinguished correctly by this signature. The two histological subtypes of gastric cancer showed different microRNA signatures: eight microRNAs were upregulated in diffuse-type and four in intestinal-type cancer. In the progression-related signature, miR-125b, miR-199a, and miR-100 were the most important microRNAs involved. Low expression of let-7g (hazard ratio 2·6 [95% CI 1·3–4·9]) and miR-433 (2·1 [1·1–3·9]) and high expression of miR-214 (2·4 [1·2–4·5]) were associated with unfavourable outcome in overall survival independent of clinical covariates, including depth of invasion, lymph-node metastasis, and stage.

Interpretation MicroRNAs are expressed differentially in gastric cancers, and histological subtypes are characterised by specific microRNA signatures. Unique microRNAs are associated with progression and prognosis of gastric cancer.

Funding National Cancer Institute.

Introduction

Gastric cancer is the fourth most common human malignant disease and the second most frequent cause of cancer-related death worldwide.[1] Improvement of diagnosis and treatment has resulted in good long-term survival for patients with early gastric cancer, whereas the outlook for individuals with advanced disease remains poor.[2] Advanced gastric cancer frequently recurs as nodal and haematogenous metastases and peritoneal dissemination. Although several types of non-surgical treatment have been assessed, surgical resection is still the primary curative treatment for localised gastric cancer.

Data from several studies show that various genetic alterations cause tumorigenesis and progression of gastric cancer.[3,4] Inactivation of runt-related transcription factor 3 (*RUNX3*) by methylation has also been reported.[5] Several groups have undertaken high-throughput analyses of gastric cancer expression profiles by DNA microarrays[4] and microdissection.[6] However, markers for tumorigenesis and progression of gastric cancer have not yet been discovered and specific therapeutic targets have not been identified.

A new class of small non-coding RNAs—microRNAs—has been discovered.[7] Mature microRNAs are composed of 19–25 nucleotides and are cleaved from 60–110-nucleotide hairpin microRNA precursors in the cytoplasm by the RNase III enzyme Dicer.[8] Single-stranded microRNAs bind mRNAs of potentially hundreds of genes at the 3′ untranslated region with imperfect complementarity, resulting in degradation of target mRNAs and inhibition of translation.[8] Several target-prediction programs have been developed, but very few targets have been proved experimentally.[9] MicroRNAs play a part in crucial cellular processes, including development, differentiation, stress response, apoptosis, and proliferation.[8,10] 475 human microRNAs have been reported to date (miRBase version 9.2; University of Manchester, Manchester, UK);[11] this number could reach 800–1000 through experimental confirmation of predicted microRNA genes.[12]

Microarray platforms have been developed for analysis of microRNA expression, and data show that several microRNAs are expressed aberrantly in various haematological and solid malignant diseases.[13–16] MicroRNAs act as novel oncogenes or tumour-suppressor genes.[17,18] We and others have noted that alterations in microRNA expression correlate highly with progression and prognosis of human tumours.[19–24] Thus, focusing on microRNAs in gastric cancer could yield new insights

into the biological behaviour of this disease. For oncogenic microRNAs, antagomirs are a type of antisense oligonucleotide that inhibit microRNA function in vivo effectively;[25–27] for tumour-suppressive microRNAs, reconstitution with microRNA precursor sequences has an antitumour effect. Therefore, microRNAs are possible therapeutic targets for cancer.[22,28]

To ascertain whether microRNA expression signatures can differ between gastric cancer and non-tumour mucosa, we undertook genome-wide microRNA expression profiling in two sets of gastric tissues. With expression-profile results for these samples and associated clinical variables, we investigated the association between microRNAs and histological types, tumour progression, and prognosis of gastric cancer.

Methods

Tissue samples

For microRNA expression profiling, we obtained gastric tissue samples (cancer lesions and adjacent non-tumour mucosae) from patients who underwent gastrectomy between 2002 and 2005 at the University of Tokyo (group 1) and between 1998 and 2005 at Hiroshima University (group 2). We gathered all samples in the same manner, and they were snap-frozen immediately in liquid nitrogen and stored at −80°C until RNA and protein extraction could be done. Since microdissection is difficult to do in diffuse-type gastric cancer, for technical uniformity we used bulk tissue for all cases.

We obtained study approval from the ethics committee at the University of Tokyo and every patient from the University of Tokyo gave written informed consent for samples to be used. Because we did not obtain written informed consent for samples from Hiroshima University, for strict privacy protection, identification information was removed before analysis; this procedure is in accordance with ethical guidelines for human genome or gene research enacted by the Japanese Government and was approved by the ethics review committee of the Hiroshima University School of Medicine.

Correspondence to:
Prof Carlo M Croce, Department of Molecular Virology, Immunology, and Medical Genetics and Comprehensive Cancer Center, Ohio State University, Biomedical Research Tower, Room 1080, 460 W 12th Ave, Columbus, OH 43210, USA
carlo.croce@osumc.edu

For **miRBase** see http://www.mirbase.org

***Panel:* Patient cohorts and of analyses undertaken**

STEP 1: MicroRNA expression patterns in gastric cancer (non-tumour mucosa *vs* cancer)

Samples

61 pairs in group 1 and 99 in group 2 were analysed independently

Statistical methods

1 Class comparison by BRB-ArrayTools; paired *t* test ($p<0·01$)
2 Class prediction by BRB-ArrayTools; paired class prediction by the leave-one-out cross-validation method

Samples

169 non-tumour mucosae (64 samples from group 1 and 105 from group 2) and 184 cancers (81 samples from group 1 and 103 from group 2) (unpaired condition)

Statistical methods

Average linkage clustering with centred Pearson correlation with 35 microRNAs

STEP 2: MicroRNA expression patterns and histological types (diffuse-type *vs* intestinal-type gastric cancer)

Samples

103 diffuse-type and 81 intestinal-type gastric cancer samples

Statistical methods

1 Class comparison by BRB-ArrayTools; two-sample *t* test ($p<0·001$)
2 Average linkage clustering with centred Pearson correlation with the 19 most significant microRNAs ($p\leq 2\times10^{-6}$)

(Continues in next column)

(Continued from previous column)

STEP 3: MicroRNA expression and tumour progression correlation

Samples

- T3 and T4 *vs* T1 (101 *vs* 15 samples)
- Lymph-node metastasis (N) positive *vs* negative (126 *vs* 54 samples)
- Stage IV *vs* I (51 *vs* 37 samples)
- Peritoneal dissemination (P, CY) positive *vs* negative (33 *vs* 76 samples)
- Haematogenous metastasis (H, M) positive *vs* negative (12 *vs* 169 samples)

Statistical methods

1 Class comparison by BRB-ArrayTools; two-sample *t* test ($p<0·01$, for haematogenous metastasis, $p<0·05$)
2 Venn diagram of T, N, and stage
3 Significance analysis of microarrays (SAM) with rank-regression option for T and stage

STEP 4: MicroRNA expression and prognosis correlation

Samples

101 cases have information for disease outcome and underwent curative surgery. All 182 cases had surgery (curative or non-curative)

Overall survival

- Statistical methods
  1 Univariate Cox proportional hazards regression in BRB-ArrayTools
  2 Kaplan-Meier survival curves
  3 Multivariable Cox proportional hazards regression analysis

Disease-free survival

- Statistical methods
  1 Univariate Cox proportional hazards regression in BRB-ArrayTools
  2 Kaplan-Meier survival curves
  3 Multivariable Cox proportional hazards regression analysis

Procedures

We did RNA labelling and hybridisation on microRNA microarray chips and undertook postprocessing, as described previously.[13,15,19–21] Briefly, 5 μg of total RNA from every sample was reverse transcribed with biotin end-labelled random-octamer oligonucleotide primers. Hybridisation of biotin-labelled complementary DNA was done on the Ohio State University custom microRNA microarray chip (OSU_CCC version 3.0; ArrayExpress [European Bioinformatics Institute, Cambridge, UK], array design A-MEXP-620), which contains nearly 1100 microRNA probes, for 326 human and 249 mouse microRNA genes, spotted in duplicates. We washed and processed the hybridised chips to detect biotin-containing transcripts with streptavidin Alexa Fluor 647 conjugate (Invitrogen, Carlsbad, CA, USA) and scanned them on a microarray scanner (4000B; Axon Instruments, Sunnyvale, CA, USA).

We analysed microarray images with GenePix Pro 6.0 (Axon Instruments). Average values of the replicate spots for every microRNA sample were background subtracted, normalised, and subjected to further analysis. Only probes for human mature microRNAs were used for analysis. We implemented quantile normalisation with the Bioconductor 1.8 package affy 1.1.2.

MicroRNAs were retained when they were present in at least 20% of samples and when they had changes of more than 1·5-fold from the gene median in at least 20% of samples. Absent calls (background-level signals on the microarray) were removed at a threshold of 4·5 (log_2 scale) before statistical analysis. After the filtration, we included 237 microRNAs in further statistical analyses.

MicroRNA nomenclature is according to miRBase version 9.2.[11] The microarray dataset is deposited in ArrayExpress (experiment number E-TABM-341) according to MIAME (minimum information about a microarray experiment) guidelines.

For **ArrayExpress** see http://www.ebi.ac.uk/arrayexpress/

For the **Bioconductor 1.8 package affy 1.1.2** see http://www.bioconductor.org

See Online for webappendix

Statistical analysis

The panel summarises the analyses. We identified differentially expressed microRNAs with BRB-ArrayTools version 3.5.0 (Biometric Research Branch, National Cancer Institute, Bethesda, MD, USA),[29] and significance analysis of microarrays (SAM) version 3.0. The webappendix contains further descriptions of the methods used.

After filtration of microRNAs, we used the paired *t* test (level of significance, $p<0·01$) to independently analyse pairs of non-tumour mucosa and cancer samples from groups 1 and 2. We undertook class prediction with the leave-one-out cross-validation method, taking into account that samples were paired (eg, pairs of non-tumour mucosae and cancer lesions from the same patient).

We used hierarchical cluster analysis to generate a tree cluster showing the separation of every class. For hierarchical clustering, we used average linkage metrics and centred Pearson correlation of microRNAs identified between non-tumour mucosa and gastric cancer and between diffuse-type and intestinal-type gastric cancer (Cluster 3.0). For tree visualisation, we used Java Treeview version 1.1.1.

We identified microRNAs whose expression was related significantly to overall survival and disease-free survival of patients (endpoint of cancer-specific death and recurrence, respectively). We undertook univariate Cox proportional hazards regression in BRB-ArrayTools, and we judged microRNAs significant if $p<0·05$.

We used SPSS version 17.0.1 for Kaplan-Meier survival analysis and Cox proportional hazards regression. To generate survival curves, we converted continuous microRNA expression levels measured on microRNA array chips to a dichotomous variable, using the respective mean levels of expression as a threshold.[21] This procedure enabled division of samples into classes with high and low expression of microRNA. We compared survival curves by log-rank test and judged $p<0·05$ significant.

	Group 1 (n=79)	Group 2 (n=103)	p*	Total (n=182)
Age (years; mean [SD])	65·2 (9·8)	67·1 (11·6)	0·24	66·3 (10·9)
Sex			0·87	
Men	52/79 (66%)	66/102† (65%)		118/181 (65%)
Women	27/79 (34%)	36/102† (35%)		63/181 (35%)
Histological type‡			0·022	
Diffuse	53/81§ (65%)	50/103 (49%)		103/184 (56%)
Intestinal	28/81§ (35%)	53/103 (51%)		81/184 (44%)
Depth of invasion (T)‖			0·50	
T1	4/81§ (5%)	11/102† (11%)		15/183 (8%)
T2	29/81§ (36%)	38/102† (37%)		67/183 (37%)
T3	41/81§ (50%)	45/102† (44%)		86/183 (47%)
T4	7/81§ (9%)	8/102† (8%)		15/183 (8%)
Lymph-node metastasis (N)			0·028	
Negative (N0)	17/79 (22%)	37/101¶ (37%)		54/180 (30%)
Positive (N1–N3)	62/79 (78%)	64/101¶ (63%)		126/180 (70%)
Haematogenous metastasis (H, M)			0·69	
Negative	75/79 (95%)	94/102† (92%)		169/181 (93%)
Positive	4/79 (5%)	8/102† (8%)		12/181 (7%)
Peritoneal dissemination (P, CY)			<0·0001	
Negative	64/79 (81%)	12/30** (40%)		76/109 (70%)
Positive	15/79 (19%)	18/30** (60%)		33/109 (30%)
Stage††			0·13	
I	11/79 (14%)	26/102† (25%)		37/181 (21%)
II	14/79 (18%)	23/102† (23%)		37/181 (21%)
III	29/79 (37%)	27/102† (27%)		56/181 (30%)
IV	25/79 (31%)	26/102† (25%)		51/181 (28%)

Data are n (%) unless stated otherwise. *Differences between groups calculated by *t* test for age and χ^2 test for all others. †No information available for one patient. ‡Lauren's classification used for histological typing. Intestinal-type gastric cancer is almost the same as differentiated-type gastric cancer, and diffuse-type gastric cancer is almost the same as undifferentiated-type gastric cancer. §One patient had cancer in three regions. ‖Graded according to the International Union Against Cancer's TNM classification, 5th edn. ¶No information available for two patients. **No information on intraoperative cytology available for 73 patients. ††Graded according to the Japanese Classification of Gastric Cancer, 2nd English edn. Clinical stage is decided by the factors T, N, H, M, P, and CY. Stages IA and IB are regarded as stage I, and stages IIIA and IIIB as stage III.

Table 1: Characteristics of patients and tissues

We examined the joint effect of covariates with Cox proportional hazards regression to ascertain whether microRNAs are independent prognostic factors. We censored data for three patients who died of other diseases; data for one patient were censored before the first event (death) in overall survival and were included in the Kaplan-Meier analysis, but were removed for Cox regression analysis in overall survival.

We regarded age as a continuous covariate. T was dichotomised on the basis of absence (T1, T2) versus presence (T3, T4) of serosal invasion of tumour. Stage was dichotomised on the basis of a more than 65% 5-year survival (stages I and II) versus a less than 50% 5-year survival (stages III and IV). For all microRNAs, patients were categorised into groups with high and low expression, with respective mean levels of microRNA expression as a threshold.

We undertook univariate Cox regression to examine the effect of every clinical covariate on patient's survival. We did multivariable analysis by stepwise addition and removal of covariates found to be associated with survival in univariate models ($p<0\cdot10$). Conditions of the stepwise selection method were Score statistic ($p<0\cdot05$ for addition) and Wald statistic ($p<0\cdot05$ for removal). All stepwise addition models gave the same final models as did stepwise removal, and final models included only those

	p†	FDR (%)‡	Fold change	Chromosomal location	Gastric signature§	Proved targets‖	Cancer involvement¶
MicroRNAs upregulated in cancer							
miR-181d	$<1\times10^{-7}$	<0·01	2·3	19p13.12	Progression	*CDX2, GATA6, NLK*	Pancreas
miR-181a-1, miR-181a-2	$<1\times10^{-7}$	<0·01	2·2	1q31.3, 9q33.3	Progression	*HOXA11, BCL2, CD69, TRAα, PTPN11 (SHP2), PTPN22, DUSP5, DUSP6, KAT2B (PCAF), CDKN1B, CDX2, GATA6, NLK*	Breast, pancreas, liver, thyroid, uterus, brain
miR-181c	$<1\times10^{-7}$	<0·01	2·1	19p13.12	Progression	*CDX2, GATA6, NLK*	Lung, pancreas, liver, thyroid, uterus, brain
miR-181b-1, miR-181b-2	$<1\times10^{-7}$	<0·01	2·0	1q31.3, 9q33.3	Progression	*TCL1A, VSNL1, GRIA2, KAT2B (PCAF), AICDA (AID), CDX2, GATA6, NLK*	Breast, colon, pancreas, prostate, stomach, thyroid, uterus, brain, CLL
miR-21	$<1\times10^{-7}$	<0·01	2·0	17q23.2	Histotype, progression	*PTEN, TPM1, PDCD4, SERPINB5, BMPR2, BTG2, CDK6, IL6R, SOCS5, NFIB, SPRY2, RECK, TIMP3, TP63 (TP73L), DAXX, HNRNPK, TOPORS, TP53BP2, JMY, TGFBR2, TGFBR3, APAF1, PPIF, SPRY1, MTAP, SOX5, TGFB1, NCAPG, RTN4, DERL1, PLOD3, BASP1, MARCKS, IL12A, JAG1, LRRFIP1*	Breast, colon, lung, pancreas, prostate, stomach, liver, thyroid, uterus, ovary, brain, CLL, lymphoma
miR-25	$<1\times10^{-7}$	<0·01	1·7	7q22.1	Progression	*BCL2L11, KAT2B (PCAF), CDKN1C*	Pancreas, prostate, stomach, liver, thyroid, uterus, oesophagus, brain, AML
miR-92-1, miR-92-2	$<1\times10^{-7}$	<0·01	1·7	13q31.3, Xq26.2	..	*MYLIP, HIPK3, BCL2L11, VHL, ITGA5, TP63 (TP73L)*	Colon, pancreas, prostate, stomach, thyroid, CLL, AML
miR-93	$<1\times10^{-7}$	<0·01	1·6	7q22.1	Progression	*E2F1, CDKN1A, VEGFA, KAT2B (PCAF), STAT3, TP53INP1, TUSC2*	Colon, pancreas, prostate, stomach, ovary, AML
miR-17-5p	2×10^{-7}	<0·01	1·7	13q31.3	..	*E2F1, NCOA3 (AIB1), RUNX1 (AML1), RBL2, CDKN1A, PTEN, BCL2L11, TIMP1, VEGFA, HIF1A, CCND1, MAPK9, MAP3K8, PKD1, PKD2, PPARA, RBL1, STAT3, TSG101, KAT2B (PCAF), CRK, GAB1, MYCN, IRF1, NR4A3, RNF111, TP53INP1, APBB2, BRCA1, APP, RASSF2, TNFSF12, MAPK14, FN1, FNDC3A, BCL2, MEF2D, MAP3K12*	Breast, colon, lung, pancreas, prostate, stomach, bladder
miR-106a	3×10^{-7}	<0·01	1·7	Xq26.2	Progression	*RB1, RUNX1 (AML1), ARID4B (RBP1L1), MYLIP, HIPK3, CDKN1A, VEGFA, APP, IL10*	Colon, lung, pancreas, prostate, stomach, liver, AML
miR-20b	4×10^{-7}	<0·01	1·9	Xq26.2	Progression	*ARID4B (RBP1L1), MYLIP, HIPK3, CDKN1A, VEGFA*	..
miR-135a-1, miR-135a-2	7×10^{-7}	<0·01	2·1	3p21.1, 12q23.1	Progression	*APC, SMAD5, JAK2*	Colon, prostate, thyroid, uterus, AML, lymphoma
miR-425-5p	1×10^{-6}	<0·01	2·2	3p21.31	..	..	..
miR-106b	1×10^{-6}	<0·01	1·6	7q22.1	..	*E2F1, CDKN1A, VEGFA, KAT2B (PCAF), ITCH, APP, STAT3, MAPK14*	Colon, stomach, AML
miR-20a	3×10^{-6}	<0·01	1·8	13q31.3	..	*E2F1, E2F2, E2F3, TGFBR2, RUNX1 (AML1), CDKN1A, ZBTB7A (LRF), VEGFA, HIF1A, CCND1, STAT3, MYF5, APP, MAPK14, BCL2, MEF2D, MAP3K12*	Colon, pancreas, prostate, uterus, ovary, AML
miR-19b-1, miR-19b-2	5×10^{-6}	<0·01	1·7	13q31.3, Xq26.2	Histotype	*THBS1 (TSP1), MYLIP, HIPK3, SOCS1*	Prostate
miR-224	2×10^{-5}	0·02	2·2	Xq28	..	API5	Pancreas, liver, thyroid, ovary, AML
miR-18a	5×10^{-5}	0·04	1·7	13q31.3	..	*CTGF, CDKN1A, NR3C1 (GR), THBS1 (TSP1), ESR1, RUNX1 (AML1)*	Pancreas, liver, AML
miR-135b	5×10^{-5}	0·04	1·6	1q32.1	..	APC	Uterus
miR-19a	0·0008	0·5	1·5	13q31.3	Histotype, progression, prognostic	*PTEN, THBS1 (TSP1), SOCS1*	Uterus, CLL
miR-345	0·001	0·5	1·5	14q32.2	Progression	..	Prostate, thyroid
miR-191	0·002	1·0	1·3	3p21.31	..	..	Breast, colon, lung, pancreas, prostate, stomach

(Continues on next page)

	p†	FDR (%)‡	Fold change	Chromosomal location	Gastric signature§	Proved targets‖	Cancer involvement¶
(Continued from previous page)							
MicroRNAs downregulated in cancer							
miR-148a	$<1\times10^{-7}$	<0·01	0·2	7p15.2	..	*NR1I2 (PXR), DNMT3B, TGIF2*	Lung, pancreas, prostate
miR-148b	$<1\times10^{-7}$	<0·01	0·3	12q13.13	Histotype	DNMT3B	Colon, lung, pancreas, prostate
miR-375	$<1\times10^{-7}$	<0·01	0·3	2q35	..	*JAK2, MTPN, C1QBP, USP1, ADIPOR2, PDK1, AIFM1, RASD1, EEF1E1, GPHN, ELAVL4, CADM1, PLAG1*	Pancreas
miR-29b-1, miR-29b-2	1×10^{-6}	<0·01	0·7	7q32.3, 1q32.2	Histotype	*TCL1A, DNAJB11, SFPQ, MCL1, DNMT3A, DNMT3B, INSIG1, CAV2, BACE1, COL1A1, COL1A2, COL3A1, FBN1, ELN, YY1, PIK3R1 (p85-ALPHA), CDC42, COL4A2, COL5A3, HDAC4, TGFB3, ACVR2A, DUSP2, CTNNBIP1*	Breast, colon, lung, pancreas, prostate, thyroid, uterus, AML
miR-29c	1×10^{-5}	0·01	0·7	1q32.2	Histotype	*DNMT3A, DNMT3B, INSIG1, CAV2, COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, COL15A1, SFRS13A, LAMC1, SPARC, TDG, YY1, PIK3R1 (p85-ALPHA), CDC42*	Breast, pancreas, liver, thyroid, oesophagus, nasopharyngeal
miR-152	1×10^{-5}	0·01	0·7	17q21.32	Histotype, progression	..	Pancreas
miR-218-2	2×10^{-5}	<0·01	0·6	5q34	Histotype, progression	*LAMB3, MAFG*	Lung, pancreas, prostate, stomach, liver, uterus
miR-451	6×10^{-5}	<0·01	0·4	17q11.2	..	*GATA2, ABCB1 (MDR1), MIF*	..
miR-30d	7×10^{-5}	<0·01	0·7	8q24.22	Histotype	..	Lung, pancreas, thyroid, uterus
miR-30a-5p	7×10^{-5}	0·06	0·7	6q13	..	*NOTCH1, BDNF, BECN1*	*Lung, pancreas, prostate, thyroid*
miR-30b	8×10^{-5}	0·06	0·7	8q24.22	Progression	..	Pancreas, prostate, uterus, lymphoma
miR-30c-1, miR-30c-2	0·0003	0·2	0·7	1p34.2, 6q13	Histotype, progression	*CTGF, RUNX1 (AML1), UBE2I*	Breast, colon, pancreas, prostate
miR-422b	0·0008	0·5	0·7	5q32	Progression	..	..

FDR=false discovery rate. AML=acute myeloid leukaemia. CLL=chronic lymphocytic leukaemia. *These microRNAs were used in the clustering of webfigure 1. †Paired class comparison. ‡1% FDR predicts that this list is 99% accurate. §Similarities in gastric cancer signature and other (histotype, progression, and prognostic) signatures. ‖Information obtained from Tarbase (http://diana.cslab.ece.ntua.gr/tarbase), miRecords (http://mirecords.umn.edu/miRecords), and previous reports.[15,21,23,24,29,30,31] ¶Information obtained from previous reports.[14-16,18-23,31]

Table 2: **Frequent differentially expressed microRNAs (gastric cancer signature)***

covariates that were associated significantly with survival (Wald statistic, $p<0·05$). We tested proportional-hazard assumption by the log-minus-log plot, and no covariate violated assumption. All p values reported are two-sided.

Role of the funding source

The sponsor had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

81 gastric cancer samples (from 79 patients; one patient had cancer in three regions) were obtained at the University of Tokyo (group 1) and 103 samples were gathered at Hiroshima University (group 2) for microRNA expression profiling. Corresponding non-tumour mucosae were available for analysis for 61 cancers in group 1 and 99 in group 2. We also obtained three additional samples of non-tumour mucosa in group 1 and six in group 2, making 353 samples in total—184 cancers and 169 non-tumour mucosae.

Clinical features of patients and tumours are described in table 1 and the webappendix. Disease outcome was known for 101 patients who underwent curative surgery; 42 recurred and died of cancer within the follow-up period. The final follow-up date was Feb 25, 2007 (median follow-up 785 days [range 159–3070]). Most patients (disease stages IB–IV) were given anticancer drugs either orally or intravenously postoperatively as adjuvant chemotherapy. After disease recurrence, these individuals were given other anticancer drugs.

On microarray analysis, 35 microRNAs were expressed differentially in the paired non-tumour mucosa and cancer samples in groups 1 and 2 (table 2): 22 of these were upregulated and 13 were downregulated in cancer (designated as the gastric cancer signature). By paired class prediction, 97% of samples in group 1 and 94% in group 2 were classified correctly.

On the basis of the 35 microRNAs expressed differentially, cluster analysis with Pearson correlation of the 169 non-tumour mucosa and 184 cancer samples generated a tree showing good separation between non-tumour mucosa and cancer (page 4 of the webappendix). Despite the unpaired condition, 83% (292/353) of samples were classified correctly to non-tumour mucosa or cancer branches.

By quantitative reverse transcription-PCR (qRT-PCR), we analysed 24 pairs of samples investigated initially by microarray for miR-21 (upregulated) and miR-375 (downregulated). We compared the cancer:non-tumour mucosa expression ratio in qRT-PCR with that in the microRNA microarray. The microarray data were confirmed by qRT-PCR (page 5 of the webappendix).

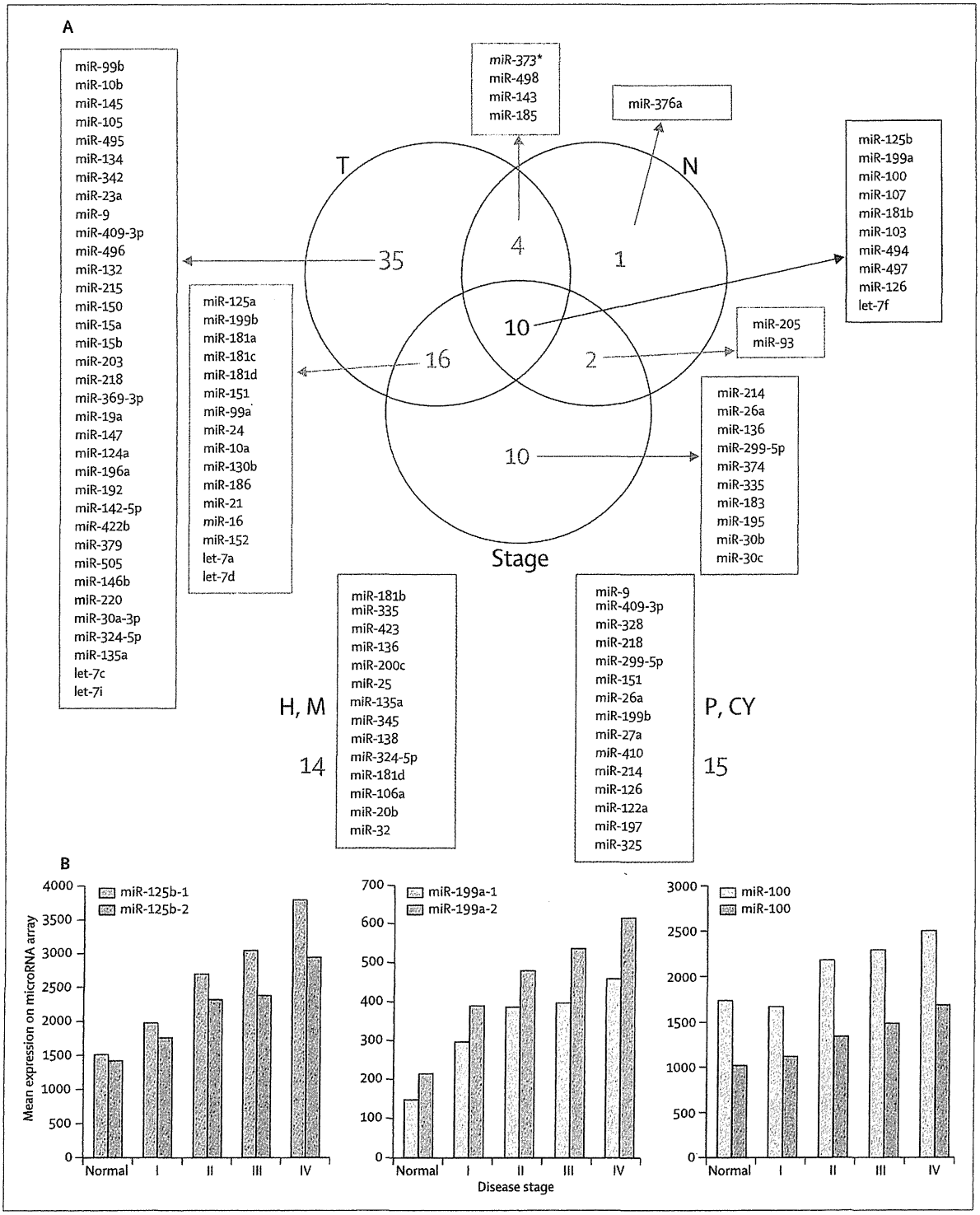

Figure 1: **MicroRNAs associated with progression of gastric cancer**

(A) Venn diagram of microRNAs related to T (depth of invasion), N (lymph-node metastasis), and stage. Listed microRNAs comprise the progression signature. Numerals indicate the number of microRNAs. Molecules corresponding to every part of the Venn diagram are shown. MicroRNAs in H and M (haematogenous metastasis) and P and CY (peritoneal dissemination) that are similar to those for T, N, or stage are shown in red. (B) Mean expression levels of miR-125b, miR-199a, and miR-100 on microRNA array according to progression in disease stage. Mean expression levels are shown as linear-scale data on microRNA array analysed with GenePix Pro 6.0; the calculation is based on the intensity (brightness) of each pixel on the microarray image. Mean expression levels of non-tumour mucosa (Normal) of group 1 are also shown. miR-125b-1 and miR-125b-2 are located on different chromosomes but the sequence of mature microRNA is the same; miR-199a-1 and miR-199a-2 are also the same. For miR-100, two probes were included on the microRNA array.

	Univariate analysis		Multivariable analysis†	
	Hazard ratio (95% CI)	p	Hazard ratio (95% CI)	p
Age	1·0 (0·9–1·0)	0·47	..	..
Sex				
Men	1·0 (reference)	0·33	..	..
Women	1·3 (0·7–2·5)		..	
Histological type				
Intestinal	1·0 (reference)	0·63	..	..
Diffuse	1·1 (0·6–2·1)		..	
T				
T1–T2	1·0 (reference)	0·001	..	..
T3–T4	3·0 (1·5–6·0)		..	
N				
Negative	1·0 (reference)	<0·0001	..	..
Positive	6·0 (2·3–15·5)		..	
Stage				
I–II	1·0 (reference)	<0·0001	1·0 (reference)	<0·0001
III–IV	5·2 (2·5–10·6)		4·3 (2·0–9·2)	
let-7g expression				
High	1·0 (reference)	0·003	1·0 (reference)	0·002
Low	2·6 (1·3–4·9)		2·9 (1·4–6·0)	
miR-214 expression				
Low	1·0 (reference)	0·007	1·0 (reference)	0·004
High	2·4 (1·2–4·5)		2·7 (1·3–5·6)	
miR-433 expression				
High	1·0 (reference)	0·015	1·0 (reference)	<0·0001
Low	2·1 (1·1–3·9)		3·4 (1·7–6·6)	
let-7e expression				
High	1·0 (reference)	0·009	..	..
Low	2·2 (1·2–4·2)		..	
let-7i expression				
High	1·0 (reference)	0·039	..	..
Low	1·9 (1·0–3·5)		..	

*One patient was censored before first event (patient's death) and these data were removed. †For the final model of multivariable analysis, stage, let-7g, miR-214, and miR-433 were included.

Table 3: **Univariate and multivariable Cox regression analysis of overall survival***

The similarity of the microRNA signature in groups 1 and 2 enabled us to merge all samples (184 cancers) into one group for further analyses. 103 diffuse-type and 81 intestinal-type specimens were used to establish whether microRNAs are differentially expressed between histological subtypes. By class comparison, 78 microRNAs were selected (false-discovery rate ≤0·42%), designated as the histotype signature.

We used the 19 most significant microRNAs (page 9 of the webappendix) in the histotype signature and undertook cluster analysis on the 184 cancer samples. These molecules were selected because they were identified also by SAM in the same order according to the absolute value of the SAM score (data not shown). Even though the histological characteristics of gastric cancer are complex (including seven histological types and mixtures of types), 74% (137/184) of tumours were distinguished successfully by the expression pattern of these 19 microRNAs (page 6 of the webappendix). Cluster analysis indicated that miR-105, miR-100, miR-125b, miR-199a, miR-99a, miR-143, miR-145, and miR-133a are upregulated in diffuse-type gastric cancer, and miR-373*, miR-498, miR-202*, and miR-494 are upregulated in intestinal-type lesions. These microRNAs are those expressed most differentially, characterising diffuse-type and intestinal-type tumours.

Next, we investigated the correlation between microRNA expression and gastric cancer progression. To identify microRNAs related to progression for every clinical feature, class comparisons were undertaken. 65 microRNAs were selected for T, 17 for N, 14 for H and M, 15 for P and CY, and 38 for stage (figure 1 A). False-discovery rate was 3·3% or less for T, 10·5% for N, 18·8% for P and CY, and 6·9% for stage. Because patients who have distant metastasis undergo surgery rarely, the sample number for positive H and M is just 12. This low number caused a reduction in power to detect microRNAs expressed differentially and a high false-discovery rate. However, six of 14 microRNAs were selected in T, N, or stage (shown in red in figure 1 A), and miR-25, miR-106a, miR-20b, miR-181b, miR-181d, and miR-135a—which were upregulated in gastric cancer relative to non-tumour mucosa—were also chosen. To identify the most important microRNAs associated with progression, we chose T and N as representative progression features and compared them with stage. Ten microRNAs—miR-125b, miR-199a, miR-100, miR-107, miR-181b, miR-103, miR-494, miR-497, miR-126, and let-7f—correlated with these variables (figure 1 A).

By SAM with rank-regression option, we selected 28 microRNAs whose expression was associated with progression from T1 to T4 and 47 microRNAs associated with progression from stage I to IV (data not shown). The q values in SAM of these microRNAs were 0% for T and 1·1% for stage. By comparison of these microRNAs with the ten identified in the previous step, we recorded miR-125b, miR-199a, and miR-100 as the most important microRNAs related to progression of gastric cancer. These three microRNAs showed increasing expression levels according to stage progression (figure 1 B).

We investigated the correlation between microRNA expression profiles and prognosis to establish the microRNAs that might signify unfavourable prognosis (independent of clinical factors). We used samples from 101 patients who underwent curative surgery and their associated prognostic information. Univariate Cox proportional hazards regression indicated that ten microRNAs (let-7c, let-7e, let-7g, let-7i, miR-19a, miR-214, miR-410, miR-433, miR-452, and miR-495) were related to overall survival of patients with gastric cancer. Kaplan-Meier survival curves were generated for every microRNA, and five (let-7e [p=0·007], let-7g [p=0·002], let-7i [p=0·038], miR-214 [p=0·005], and miR-433 [p=0·015]) were associated significantly with survival.

Table 3 shows univariate Cox proportional hazards regression analysis of overall survival relative to clinical factors. T, N, and stage were associated significantly with overall survival, as were five microRNAs. To elucidate whether these microRNAs are independent prognostic factors, multivariable analysis was done. The dichotomised expression values of these five microRNAs were not associated with clinical factors (Fisher's exact test). Because T and N were associated highly with stage by Fisher's exact test, and the same microRNAs were chosen in the final model of multivariable analysis including stage and in the final model including T and N, we showed only the stage model (table 3). In the final multivariable model, let-7g, miR-214, and miR-433 were associated with overall survival independent of clinical covariates (table 3). Patients with low expression of let-7g (hazard ratio 2·6 [95% CI 1·3–4·9]), low expression of miR-433 (2·1 [1·1–3·9]), or high expression of miR-214 (2·4 [1·2–4·5]) had poorer survival than did patients with high expression of let-7g, high expression of miR-433, or low expression of miR-214 (figure 2).

We validated the results for let-7g and miR-214 by qRT-PCR. 12 samples selected from the low-expression group showed low expression of let-7g and miR-214 by qRT-PCR, and 12 samples selected from the high-expression group showed high expression (page 7 of the webappendix). We analysed three additional specimens by qRT-PCR that were not used in microRNA array analysis because of low RNA yield. One sample with an unfavourable outcome showed high expression of miR-214 (higher than the mean of 12 samples from the high-expression group), and two with a favourable outcome showed low expression of miR-214 (lower than the mean of 12 samples from the low-expression group), consistent with our results.

We undertook the same analyses for disease-free survival in 101 patients. By univariate Cox proportional hazards regression, 12 microRNAs (let-7b, let-7c, let-7d, let-7g, miR-19a, miR-196a, miR-220, miR-373, miR-410, miR-433, miR-452, and miR-495) were related to disease-free survival of patients with gastric cancer. By log-rank analysis, six microRNAs (let-7b [p=0·001], let-7g [p=0·001], miR-19a [p=0·031], miR-410 [p=0·015], miR-433 [p=0·011], and miR-495 [p=0·035]) were related to survival. On univariate analysis, T, N, stage, and these six microRNAs were associated significantly with disease-free survival (table 4). The dichotomised expression values of six microRNAs were not associated with clinical factors (Fisher's exact test). Because T and N were associated highly with stage by Fisher's exact test, and the same microRNAs were chosen in the final model of multivariable analysis including stage and in the final model including T and N, we showed only the stage model (table 4). In the final multivariable Cox regression model, let-7b, let-7g, miR-19a, and miR-495 were associated with disease-free survival independent of clinical covariates (table 4). In

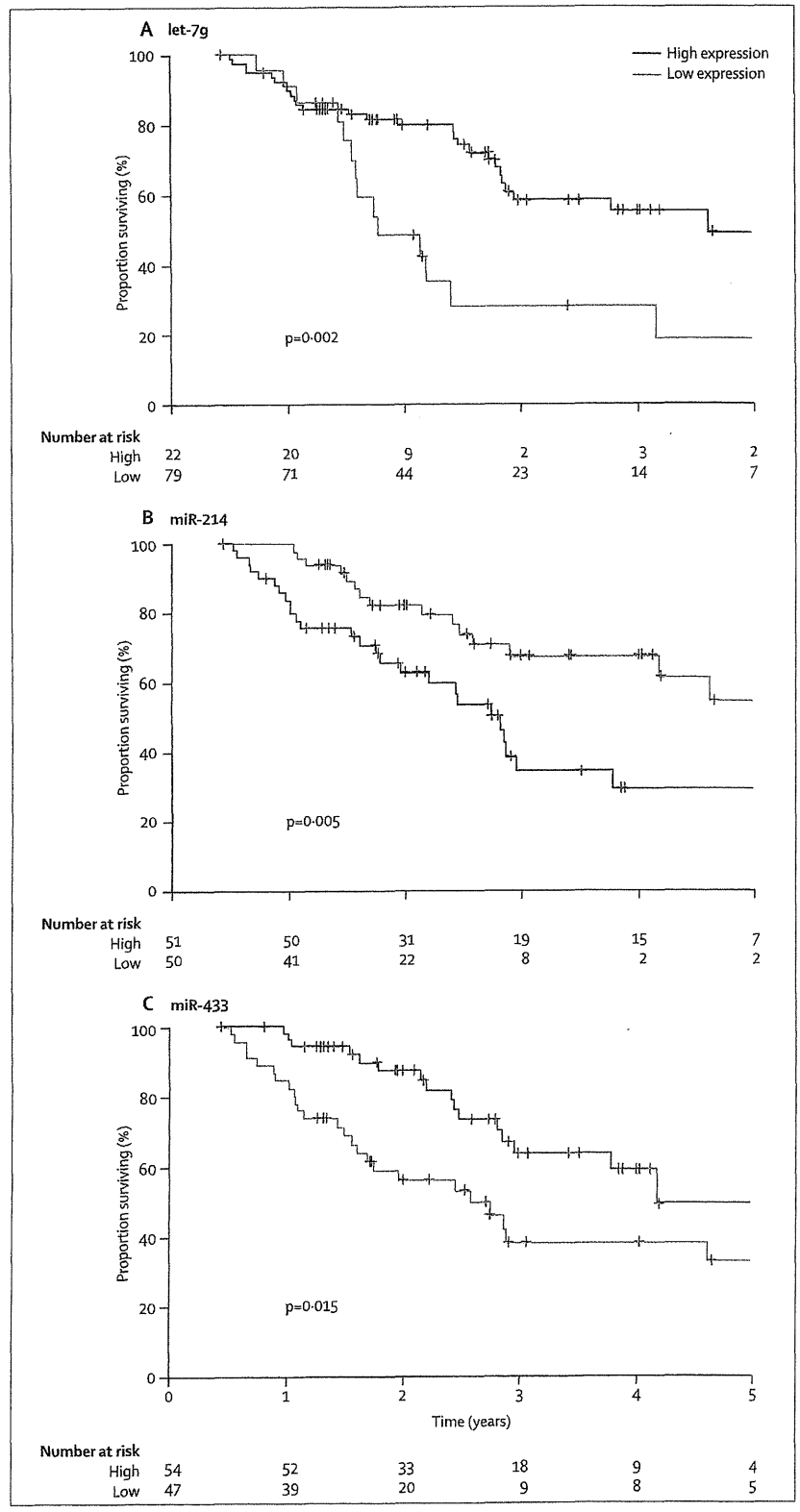

Figure 2: **Kaplan-Meier curves of independent prognostic factors for overall survival**
Curves are depicted with data for 101 patients. MicroRNA expression levels measured on the microarray were converted into discrete variables by division of samples into two classes (high and low expression), with the respective mean levels of microRNA expression as a threshold. Censored cases are shown on the curves. p values are log rank.

	Univariate analysis		Multivariable analysis†	
	Hazard ratio (95% CI)	p	Hazard ratio (95% CI)	p
Age	1·0 (0·9–1·0)	0·63	..	..
Sex				
Men	1·0 (reference)	0·31	..	..
Women	1·3 (0·7–2·5)		..	
Histological type				
Intestinal	1·0 (reference)	0·67	..	..
Diffuse	1·1 (0·6–2·1)		..	
T				
T1–T2	1·0 (reference)	0·001	..	..
T3–T4	3·1 (1·5–6·1)		..	
N				
Negative	1·0 (reference)	<0·0001	..	..
Positive	5·5 (2·1–14·2)		..	
Stage				
I–II	1·0 (reference)	<0·0001	1·0 (reference)	<0·0001
III–IV	4·5 (2·2–9·2)		5·2 (2·4–11·2)	
let-7b expression				
High	1·0 (reference)	0·003	1·0 (reference)	0·001
Low	2·7 (1·4–5·4)		3·2 (1·6–6·6)	
let-7g expression				
High	1·0 (reference)	0·002	1·0 (reference)	0·042
Low	2·7 (1·4–5·2)		2·0 (1·0–3·9)	
miR-19a expression				
High	1·0 (reference)	0·032	1·0 (reference)	<0·0001
Low	2·0 (1·0–3·6)		3·3 (1·7–6·5)	
miR-495 expression				
Low	1·0 (reference)	0·035	1·0 (reference)	0·007
High	1·9 (1·0–3·6)		2·4 (1·2–4·7)	
miR-410 expression				
Low	1·0 (reference)	0·016	..	..
High	2·2 (1·1–4·3)		..	
miR-433 expression				
High	1·0 (reference)	0·011	..	..
Low	2·1 (1·1–4·0)		..	

*No patients were censored before first event (disease recurrence). †For the final model of multivariable analysis, stage, let-7b, let-7g, miR-19a, and miR-495 were included.

Table 4: **Univariate and multivariable Cox regression analysis of disease-free survival***

both overall survival and disease-free survival, let-7g was selected as an independent prognostic factor (tables 3 and 4).

101 patients were divided into two groups by histological type (intestinal and diffuse) and multivariable Cox proportional hazards regression analysis was undertaken in the same way. The selected microRNAs remained as independent prognostic factors (table 5).

Discussion

Aberrant microRNA expression patterns have been described in various haematological and solid cancers,[14–16,20–22] and alterations in microRNA expression correlate highly with progression and prognosis of human malignant diseases.[19–24] However, profiles of microRNAs differ and need to be investigated in every type of tumour. In this study, we recorded substantial associations between differential expression of specific microRNAs and progression and prognosis of gastric cancer.

Antiapoptotic miR-21 is upregulated in various solid cancers and is related to tumour growth.[15,30] In previous work, miR-21 was overexpressed in gastric cancer and in *Helicobacter pylori*-infected gastric mucosa.[30] *H pylori* is an important pathogen for gastric cancer, and data are already starting to suggest the molecular mechanism of evolution of normal mucosa to chronic gastritis, atrophic gastritis, and intestinal metaplasia. Our sample set contained no detailed information about *H pylori* infection status because pathologists recorded histological types, depth of invasion, and status of lymph-node metastasis to decide clinical stage of cases. Non-tumour mucosae were obtained during surgery from resected stomach that seemed to be normal macroscopically. Therefore, in this study we could not investigate the correlation between microRNA expression and *H pylori* or chronic gastritis; however, we will investigate this important area in further studies.

We identified 35 differentially expressed microRNAs without use of microdissection. This procedure is difficult to adapt to some diffuse-type gastric cancers because cancer cells are localised singly. In a previous report, we analysed by microarray 20 pairs of intestinal-type gastric cancer and non-tumour mucosa samples from a white population and noted 14 upregulated and five downregulated microRNAs in cancers.[31] All the upregulated microRNAs and three of those down-regulated (60%) were similar to the molecules selected in this study, meaning that our method of using bulk samples of diffuse-type gastric cancer for microarray analysis can produce correct results, although they must be validated by in-situ hybridisation. This result also means that despite patients' different ethnic backgrounds in this and our previous study, the microRNA signature is linked to general mechanisms of gastric cancer tumorigenesis.

For some of the microRNAs we identified in gastric cancer samples, several targets have already been proven experimentally. We showed previously that molecules expressed differentially in the microRNA cluster miR-106b-25 are related to gastric cancer tumorigenesis,[31] suggesting that microRNAs have important roles in gastric cancer. Although gastric cancer is histologically complex and sometimes shows transition from differentiated to undifferentiated subtypes in the same tumour (ie, mixed type), we divided samples into diffuse and intestinal types and identified microRNAs expressed differentially, characterising these histological classes. A collaborator of ours reported that the Hedgehog signal is more active in diffuse-type than intestinal-type gastric cancer,[33] and glioma-associated oncogene homologue 1 (*GLI1*), a downstream target of the Hedgehog signal, is