Gene expression profiling of advanced-stage serous ovarian cancers distinguishes novel subclasses and implicates *ZEB2* in tumor progression and prognosis

Kosuke Yoshihara,[1,10] Atsushi Tajima,[2,10] Dai Komata,[1] Tadashi Yamamoto,[3] Shoji Kodama,[4] Hiroyuki Fujiwara,[5] Mitsuaki Suzuki,[5] Yoshitaka Onishi,[6] Masayuki Hatae,[6] Kazunobu Sueyoshi,[7] Hisaya Fujiwara,[8] Yoshiki Kudo,[8] Ituro Inoue[2,9] and Kenichi Tanaka[1,9]

[1]Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata; [2]Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan; [3]Department of Structural Pathology, Institute of Nephrology, Niigata University Graduate School of Medical and Dental Sciences; [4]Department of Gynecology, Niigata Cancer Center Hospital, Niigata; [5]Department of Obstetrics and Gynecology, Jichi Medical University, Shimotuke; [6]Department of Obstetrics and Gynecology, Kagoshima City Hospital; [7]Department of Pathology, Kagoshima City Hospital, Kagoshima; [8]Department of Obstetrics and Gynecology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan

(Received January 30, 2009/Revised April 16, 2009/Accepted April 24, 2009/Online publication May 26, 2009)

To elucidate the mechanisms of rapid progression of serous ovarian cancer, gene expression profiles from 43 ovarian cancer tissues comprising eight early stage and 35 advanced stage tissues were carried out using oligonucleotide microarrays of 18 716 genes. By non-negative matrix factorization analysis using 178 genes, which were extracted as stage-specific genes, 35 advanced stage cases were classified into two subclasses with superior ($n = 17$) and poor ($n = 18$) outcome evaluated by progression-free survival (log rank test, $P = 0.03$). Of the 178 stage-specific genes, 112 genes were identified as showing different expression between the two subclasses. Of the 48 genes selected for biological function by gene ontology analysis or Ingenuity Pathway Analysis, five genes (*ZEB2*, *CDH1*, *LTBP2*, *COL16A1*, and *ACTA2*) were extracted as candidates for prognostic factors associated with progression-free survival. The relationship between high *ZEB2* or low *CDH1* expression and shorter progression-free survival was validated by real-time RT-PCR experiments of 37 independent advanced stage cancer samples. *ZEB2* expression was negatively correlated with *CDH1* expression in advanced stage samples, whereas *ZEB2* knockdown in ovarian adenocarcinoma SKOV3 cells resulted in an increase in *CDH1* expression. Multivariate analysis showed that high *ZEB2* expression was independently associated with poor prognosis. Furthermore, the prognostic effect of E-cadherin encoded by *CDH1* was verified using immunohistochemical analysis of an independent advanced stage cancer samples set ($n = 74$). These findings suggest that the expression of epithelial–mesenchymal transition-related genes such as *ZEB2* and *CDH1* may play important roles in the invasion process of advanced stage serous ovarian cancer. (*Cancer Sci* 2009; 100: 1421–1428)

The serous type, comprising approximately 50% of ovarian cancers, is the most aggressive histology and has a tendency to be detected as advanced stage at the time of diagnosis.[1,2] Patients with advanced stage serous ovarian cancer are managed with surgical cytoreduction followed by platinum and taxane-based chemotherapy. Serous ovarian cancer is moderately chemosensitive and initially responds to postoperative chemotherapy, but the survival of patients with advanced stage remains poor. Because the majority of early stage ovarian cancers are asymptomatic and there is as yet no reliable screening test, it is difficult to diagnose early stage serous ovarian cancer. Therefore, the molecular mechanisms of progression in serous ovarian cancer should provide valuable clues for early detection and improved prognosis.

The development of microarray technology permits analysis of the expression levels of thousands of genes in cancer cells, and several studies have shown that microarrays can be used to identify gene expression profiles associated with surgery outcome, response to chemotherapy, grade, and survival in ovarian cancers.[3–17] However, there are limited reports of microarray analysis on tumor progression.[18–20] Serous ovarian cancer more rapidly progresses to advanced stage than other histological types.[21] In the present study, we used genome-wide expression microarray to distinguish between stage I (ovary confined) and stage III/IV serous ovarian cancers to focus on the molecular mechanisms of tumor progression and metastasis. Our microarray analysis identified 178 stage-specific genes, and also divided advanced stage (stage III/IV) ovarian cancers into two novel prognostic subclasses, by the NMF method. There were significant differences between the two subclasses in progression-free survival time. Furthermore, we extracted *CDH1* and its transcriptional repressor *ZEB2* from the 112 genes that were differentially expressed between the two novel subclasses, and found that the expression levels of these epithelial–mesenchymal transition-related genes[22,23] are associated with tumor progression and prognosis in advanced stage serous ovarian cancer patients.

Materials and Methods

Tissue samples. Eighty-nine patients (17 stage I; 72 stage III/IV) who were diagnosed with serous histological type ovarian cancer between July 1997 and October 2007 were recruited in this study. Fresh-frozen samples were obtained from primary tumor tissues at initial cytoreductive surgery. No patients received chemotherapy before surgery. All patients with advanced stage serous ovarian cancer ($n = 72$) were treated with platinum and taxane-based chemotherapy after surgery. The ethics committees of the participating institutions approved the study protocol, and each participant gave written, informed consent. Of the 89 samples, 43 were analyzed with microarray. The remaining 46 samples were used for subsequent validation analysis. There were no significant differences between the two samples sets regarding age of onset, stage, performance of optimal cytoreduction, histological grade, and follow-up period between the microarray set and validation set (Supplementary Table 1). Staging of the disease was assessed in accordance with the criteria of the International Federation of Gynecology and Obstetrics.[24] Optimal cytoreduction was defined as ≤1 cm of gross residual disease. The histological characteristics of surgically resected specimens

[9]To whom correspondence should be addressed.
E-mail: tanaken@med.niigata-u.ac.jp or ituro@is.icc.u-tokai.ac.jp
[10]These authors contributed equally to this work.

doi: 10.1111/j.1349-7006.2009.01204.x
© 2009 Japanese Cancer Association

were assessed on formalin-fixed and paraffin-embedded hematoxylin–eosin sections, and frozen tissues containing more than 80% tumor cells were used for RNA extraction. Normal peritoneum tissues were obtained from 10 patients having other procedures (such as hysterectomy for myoma uteri) at Niigata University. Tumors of 43 samples used for microarray analysis were screened for the presence of *TP53* somatic mutations using previously reported methods.[25] Four patients with family history of ovarian cancer in the microarray set were examined for germline mutations of *BRCA1* according to an in-house protocol,[26] and two patients showed mutations of *BRCA1*.

Microarray experiments. Total RNA, extracted from tissue samples using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) was examined with a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) using an RNA 6000 Nano LabChip (Agilent Technologies). Five hundred nanograms of total RNA was converted into labeled cRNA with nucleotides coupled to Cy3 (PerkinElmer, Boston, MA, USA) using the Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies). Cy3-labeled cRNA (1.5 µg) was hybridized for 17 h at 65°C to an Agilent Human 1A (v2) Oligo Microarray, which carries 60-mer probes to 18 716 human transcripts. The hybridized microarray was washed and then scanned in Cy3 channel with the Agilent DNA Microarray Scanner (model G2565AA). Signal intensity per spot was generated from the scanned image with Feature Extraction Software version 8.5 (Agilent Technologies) with the default settings. Spots that did not pass quality control procedures were flagged as 'absent'.

Microarray data analysis. Data normalization was carried out using GeneSpring GX 7.3 (Agilent Technologies) as follows: (i) values below 0.01 were set to 0.01, following background subtraction; and (ii) median percentile normalization was carried out using a per-chip 50th percentile of all measurements. Furthermore, genes with expression levels marked as 'absent' in more than 22 of 43 microarrays were excluded to analyze ovarian cancer-specific transcripts. When the gene expression patterns of two groups were compared, genes showing twofold or more mean expression differences between the groups were first determined by Welch's *t*-test in GeneSpring GX. For multiple testing corrections in this statistical analysis, the Benjamini–Hochberg procedure[27] of controlling the false discovery rate at the level of 0.05 was used.

To assess heterogeneity of the gene expression profile among serous ovarian cancer patients, we applied a NMF algorithm and hierarchical clustering using stage-specific gene expression profiles. NMF analysis was carried out according to Brunet *et al*.[28] as previously reported.[29]

To investigate the biological functions of the gene expression profiles, we used GO Ontology Browser, embedded in GeneSpring GX, and IPA (http://www.ingenuity.com). More detailed information about this analysis using the GO Ontology Browser and IPA is given in Supplementary methods.

Quantitative RT-PCR analysis. Total RNA (1 µg) from ovarian cancer was used as a template in first-strand cDNA synthesis with the SuperScript III First-Strand Synthesis System (Invitrogen). The cDNA was diluted one in ten for subsequent real-time PCR, which was carried out using TaqMan Gene Expression Assays (Applied Biosystems) with TaqMan Universal PCR Master Mix (Applied Biosystems) on a 7900HT Sequence Detection System (Applied Biosystems) according to the manufacturers' instructions. Detailed information on the 23 transcripts examined is summarized in Supplementary Table 2. The relative quantification method[30] was used to measure the amounts of the respective genes in serous ovarian cancer samples, normalized to *ACTB* and *TBP*.

Analysis of clinical and pathological parameters. All analyses except Cox's proportional hazard analysis were done using GraphPad PRISM version 4.0 (GraphPad Software, San Diego, CA, USA). Survival curves were investigated using the Kaplan–Meier method and log rank test (GraphPad PRISM). When clinicopathological parameters among ovarian cancer patients were compared, unpaired *t*-test, Fisher's exact test or χ^2-test was used depending on the purpose (GraphPad PRISM). Pearson's correlation coefficient was calculated for correlation between *ZEB2* expression and *CDH1* expression. Differences in gene expression levels between two subclasses were tested by Mann–Whitney test. Using a $\log_2$ transformation of expression data, Cox's proportional hazard model analysis was carried out using JMP version 6 (SAS Institute, Cary, NC, USA).

Results

Identification and characterization of molecular subclasses from advanced stage serous ovarian cancer cases. Using Agilent Human 1A(v2) Oligo microarray, we generated gene expression data for 43 serous ovarian cancers comprising eight stage I and 35 stage III/IV tumors, as well as 10 normal peritoneum tissues as a reference. First, 4275 ovarian cancer-specific genes that were differentially expressed between ovarian cancer and peritoneum tissues were isolated. Of these 4275 transcripts, 178 stage-specific genes showing significantly more than twofold upregulation or downregulation in stage III/IV samples compared to stage I samples; 107 transcripts were upregulated and 71 transcripts downregulated in stage III/IV serous ovarian cancers (Supplementary Fig. 1).

To clarify the heterogeneity of the samples at the transcriptome level, 43 serous ovarian cancer samples were analyzed by the NMF method[28,29,31] using the 178 transcriptomes that were differentially expressed between stage I samples and stage III/IV samples. Figure 1(A) shows reordered consensus matrices averaging 50 connective matrices generated for subclasses $K = 2$, 3, 4, and 5. The most distinct pattern of block partitioning was observed at the $K = 2$ model. Thus, the NMF method predicts the existence of robust subclasses of serous ovarian cancer samples for $K = 2$. This prediction was quantitatively supported by higher values of coph for NMF-clustered matrices. The NMF class assignment for $K = 2$ was the most robust with the highest coph value (coph = 0.999). Interestingly, one subclass in the $K = 2$ model was composed of eight stage I samples and 17 stage III/IV samples, whereas the other was composed of 18 stage III/IV samples. To verify the accuracy and robustness of the classification, a hierarchical clustering approach was also applied to log-transformed normalized data for stage-specific target genes. As depicted in Figure 1(B), 43 serous ovarian cancer samples were separated into two main branches showing similarity with the NMF-based subclassification. Thus, it was confirmed that the 35 advanced stage serous ovarian cancer samples were categorized into two distinct subclasses at the transcriptome level. A group composed of 17 stage III/IV samples with gene expression profiles similar to stage I samples was termed 'subclass 1', and the second group comprising 18 stage III/IV samples was termed 'subclass 2'. Two patients were identified as harboring *BRCA1* mutations: one patient belonged to stage I and the other to subclass 1 in the array analysis, but there was no particular gene expression pattern due to the mutations based on the expression levels of the 178 stage-specific genes.

We then investigated the possibility that the two subclasses of advanced stage serous ovarian cancers split by the NMF approach might represent clinically, pathologically, or genetically distinct characteristics. The distribution of several known prognostic factors is listed in Table 1. The two subclasses were similar in age of onset, stage, CA125 level before treatment, presence of tumor cells in ascites, histological grade, presence of lymph node metastasis, and frequency of *TP53* mutations, except that subclass 1 had a higher rate of optimal cytoreduction than subclass 2 (Fisher's exact test, $P = 0.09$). When the outcome of two

doi: 10.1111/j.1349-7006.2009.01204.x
© 2009 Japanese Cancer Association

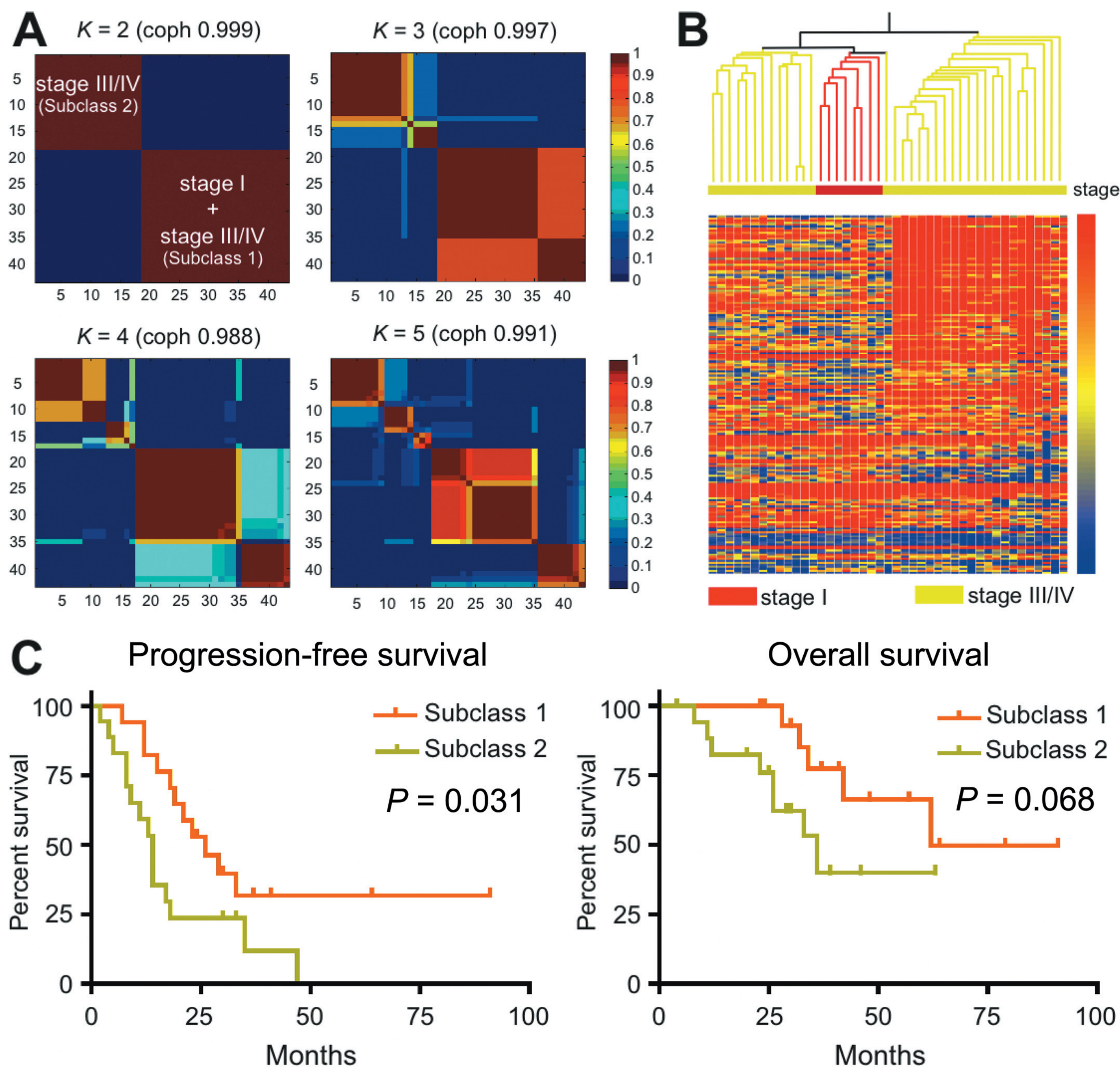


Fig. 1. Subclassification of 43 serous ovarian cancer samples and their prognosis. (A) By a non-negative matrix factorization (NMF) approach, NMF-consensus matrices averaging 50 connectivity matrices were computed at $K = 2–5$ (as the number of subclasses modeled) for the 43 serous ovarian cancer samples with 178 stage-specific genes. The NMF computation and model selection were carried out according to Brunet *et al.*[28] By accounting for the cophenetic correlation coefficients (coph) for NMF-clustered matrices, the NMF class assignment at $K = 2$ was the most robust. One subclass in the $K = 2$ model contained samples both from stage I ($n = 8$) and III/IV ($n = 17$), whereas the other contained only stage III/IV samples ($n = 18$). (B) A hierarchical clustering method was also used to classify serous ovarian cancer samples using 178 stage-specific genes. The 43 serous samples were largely separated into two clusters. The stage assignments for samples are: stage I, red; and stage III/IV, yellow. (C) Kaplan–Meier survival curves between two NMF-based subclasses of the 35 stage III/IV patients. Subclass 1, composed of 17 advanced stage patients with gene expression profiles similar to that of stage I, showed statistically prolonged progression-free survival (log rank test, $P = 0.031$), but no significant correlation with overall survival (log rank test, $P = 0.068$).

subclasses was compared for progression-free survival and overall survival, the Kaplan–Meier curves showed significantly better outcome in cases belonging to subclass 1 in progression-free survival (Fig. 1C, log rank test, $P = 0.031$) and fair outcome in overall survival (Fig. 1C, log rank test, $P = 0.068$).

Association of subclass-specific gene expression profile with prognosis of ovarian cancer patients. To characterize the gene expression differences associated with distinct prognoses between the two subclasses of advanced stage serous ovarian cancers, we identified 112 subclass-specific transcripts that were differentially expressed between the two subclasses; 25 transcripts were up-regulated in subclass 1 and 87 transcripts were upregulated in subclass 2 (Supplementary Table 3). We then examined the biological functions of the 112 subclass-specific genes using two analytic tools, GO analysis and IPA, to clarify the biological mechanism of tumor progression. The Gene Ontology Biological

© 2009 Japanese Cancer Association

Table 1. Clinical characteristics of two subclasses of advanced stage serous ovarian cancer samples

Characteristic	Subclass 1 (n = 17)	Subclass 2 (n = 18)	*P*-value
Age (years)	58.5 ± 8.6	61.1 ± 12.6	0.49†
Stage			
Stage III	16	15	1‡
Stage IV	1	3	
CA125 (IU)	1987 ± 2021	1178 ± 1057	0.14†
Cancer cell in abdominal fluid			
Positive	15	15	1‡
Negative	2	3	
Optimal cytoreduction			
Optimal (<1 cm)	12	7	0.09‡
Not optimal	5	11	
Lymph node metastasis			
Positive	6	4	1‡
Negative	9	6	
Unknown	2	8	
Grade			
Grade 1	6	4	0.18§
Grade 2	9	7	
Grade 3	2	7	
TP53 status			
Wild type	11	9	0.50‡
Mutated	6	9	

Differences in clinical characteristics between subclass 1 and subclass 2 were tested using the †unpaired *t*-test, ‡Fisher's exact test or §χ^2-test.

Table 2. Univariable Cox's proportional hazards model analysis of expression levels of five genes for progression-free survival and overall survival in patients with advanced stage serous ovarian cancers

Gene symbol	Hazard ratio (95% CI)	*P*-value
Microarray set (n = 35)		
Progression-free survival		
ZEB2	1.35 (1.06–1.77)	0.015*
CDH1	0.75 (0.62–0.94)	0.017*
LTBP2	1.63 (1.04–2.57)	0.032*
COL16A1	1.33 (1.02–1.74)	0.034*
ACTA2	1.21 (1.01–1.46)	0.036*
Overall survival		
ZEB2	1.56 (1.06–2.47)	0.023*
CDH1	0.81 (0.67–1.03)	0.081
LTBP2	2.53 (1.43–4.58)	0.0017*
COL16A1	1.66 (1.12–2.59)	0.012*
ACTA2	1.44 (1.10–1.95)	0.0087*
Validation set (n = 37)		
Progression-free survival		
ZEB2	1.74 (1.08–2.92)	0.023*
CDH1	0.20 (0.09–0.45)	0.00006*
LTBP2	1.16 (0.75–1.75)	0.49
COL16A1	1.18 (0.92–1.51)	0.20
ACTA2	1.22 (0.90–1.66)	0.19
Overall survival		
ZEB2	1.89 (1.06–3.64)	0.029*
CDH1	0.59 (0.26–1.30)	0.19
LTBP2	1.1 (0.70–1.66)	0.69
COL16A1	1.23 (0.93–1.66)	0.15
ACTA2	1.43 (0.99–2.13)	0.052

*P < 0.05.

Process categories over-represented among 112 subclass-specific genes are shown in Figure 2(A). After multiple testing corrections using the Benjamini–Hochberg FDR method, seven categories were significantly over-represented, and included 37 non-overlapping genes. Subclass-specific genes were involved in biological processes of transport (GO6817, GO15698, GO6820, and GO6811), development (GO48513 and GO1501), and cell adhesion (GO7155), and included a high proportion of extracellular matrix-related genes. In addition, when ID of Agilent probes of 112 subclass-specific transcripts were imported into the IPA software, a new pathway comprising 26 genes that were enriched in extracellular matrix genes was identified (Fig. 2B). Fifteen genes belonged to both the seven GO categories and the new network, and 48 non-redundant genes were biologically characterized.

To investigate whether the expression profile of the 48 genes extracted by GO analysis or IPA was implicated in the aggressive phenotype of ovarian cancer, we analyzed the association between the respective expression levels of the 48 genes and progression-free survival time using univariate Cox's proportional hazard model. The expression levels of *ZEB2*, *CDH1*, *LTBP2*, *COL16A1*, and *ACTA2* were significantly correlated with progression-free survival (Table 2). When overall survival also was evaluated by Cox's proportional hazard model, the expression of the above genes except *CDH1* was significantly correlated with overall survival.

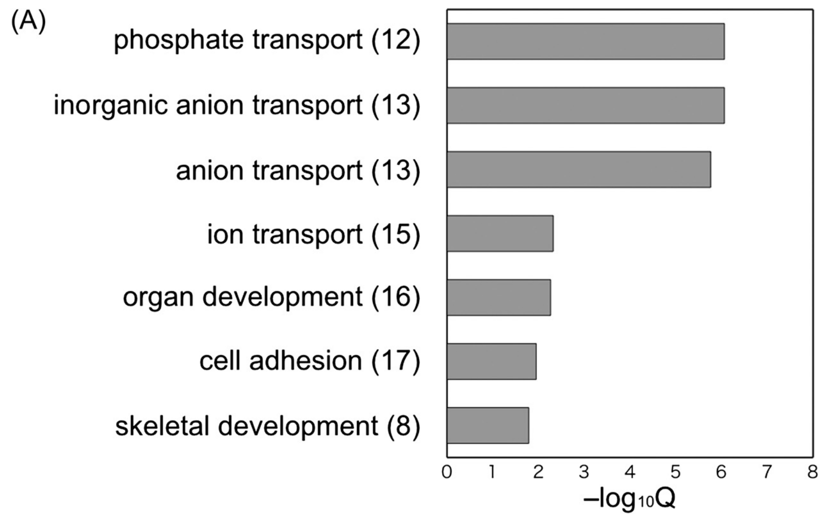
Validation by quantitative real-time RT-PCR. To validate the microarray expression data, we measured expression levels of 23 randomly selected transcripts from the 112 subclass-specific transcripts by real-time RT-PCR analysis. In agreement with microarray results, there was a significant difference between the expression levels of the 23 transcripts measured by real-time RT-PCR of subclass 1 and subclass 2 (Supplementary Table 4).

To validate the previous findings that the expression levels of *ZEB2*, *CDH1*, *LTBP2*, *COL16A1*, and *ACTA2* are associated with progression-free survival, quantitative real-time RT-PCR was carried out on 46 samples comprising nine stage I samples and 37 stage III/IV samples recruited as an independent validation set. Cox's proportional hazard analysis showed that the expression levels of *ZEB2* and *CDH1* were again correlated with progression-free survival (P = 0.023 and 0.00006, respectively) (Table 2). Moreover, *ZEB2* expression was significantly associated with overall survival (P = 0.029). At the protein level, an association of the expression of E-cadherin (encoded by *CDH1*) with prognosis of advanced stage serous ovarian cancer patients was further verified by immunohistochemical analysis of independent samples (n = 74) (Supplementary Fig. 3) as previously reported.[32–35]

Interaction between *ZEB2* and *CDH1*. *ZEB2* directly interacted with *CDH1* in the IPA network, as shown in Figure 2(B). We also found a significantly negative correlation between *ZEB2* expression and *CDH1* expression (Pearson's correlation coefficient: –0.432, P = 0.0002) in advanced stage serous ovarian cancers using real-time RT-PCR data (n = 72). *ZEB2* acts on the promoter of *CDH1*, a well-known epithelial marker, and reduces its expression.[23,36] To confirm the interaction between *ZEB2* and *CDH1* in ovarian cancer cells, a siRNA approach was used. For this purpose, we selected the SKOV3 cell line expressing endogenously higher *ZEB2* and lower *CDH1* mRNA than other ovarian cancer cell lines (Supplementary Fig. 3A,B). In SKOV3 cells, siRNA-mediated transient silencing of *ZEB2* expression resulted in upregulation of *CDH1* expression and downregulation of *FN1* and *VIM* expression (Supplementary Fig. 3C–F).

For multivariate analysis, we selected *ZEB2* from the two genes as likely to be the more important prognostic factor owing to its functional significance as an upstream repressor of *CDH1*.[23] The prognostic capability of *ZEB2* was further compared with other prognosis-related variables such as clinicopathological factors including age, performance of optimal cytoreduction, and histological grade using multivariate Cox's proportional

doi: 10.1111/j.1349-7006.2009.01204.x
© 2009 Japanese Cancer Association



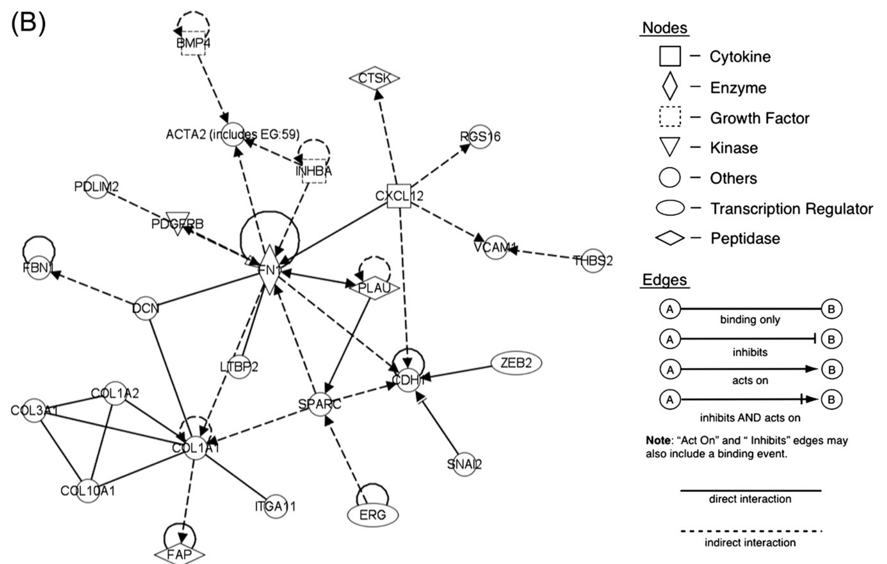


Fig. 2. Biological characterization of 112 subclass-specific genes using Gene Ontology analysis and Ingenuity Pathway Analysis (IPA). (A) Significant enrichments of gene ontology (GO) categories in GO-based profiling of 112 subclass-specific genes. Gray bars represent *q*-values (expressed as the negative logarithm [base 10]) after multiple testing correction of the Benjamini–Hochberg false discovery rate method for the significant ($q < 0.05$) GO categories over-represented in the 112 subclass-specific genes, using 4275 ovarian cancer-specific genes as a background set of genes for the determination of *q*-values. The actual number of the subclass-specific genes involved in each category is given in parentheses. (B) Twenty-six of 112 (24.1%) genes appeared in a new network based on the Ingenuity Pathway Knowledge base. Nodes represent genes, with their shapes showing IPA-defined functional classes of genes, and edges indicating biological relationships between nodes.

hazards analysis (Table 3). To increase the reliability of the multivariate analyses, all of the advanced stage serous ovarian cancer samples ($n = 72$) were analyzed by the real-time PCR technique. In Cox's proportional hazards model, *ZEB2* expression and the rate of optimal cytoreduction surgery were independent factors for progression-free survival time ($P = 0.014$ and 0.0011 respectively). The hazard ratio for relapse of *ZEB2* expression was 1.37 (95% confidence interval 1.07–1.78). Furthermore, when overall survival was evaluated by multivariate analysis, only the *ZEB2* expression level was independently associated with overall survival time ($P = 0.027$, hazard ratio = 1.53, 95% confidence interval 1.05–2.22).

***ZEB2* and *CDH1* expression and survival.** To clarify the details of the *ZEB2–CDH1* relationship, we analyzed the prognostic implications with regard to combinations of *ZEB2* and *CDH1* expression. For this purpose, we divided all of the samples into four groups, as shown in Table 4, when using a median expression level of each gene as a threshold for sample division. Multivariate Cox's proportional hazard model was used to compare survival among these four groups. The group showing high expression of *CDH1* and low expression of *ZEB2* served as a reference. Of the four groups, the only group with low expression of *CDH1* and high expression of *ZEB2* showed significantly poor prognosis in both progression-free survival and overall survival ($P = 0.0035$ and 0.013 respectively). When *ZEB2* expression was analyzed in combination with *CDH1* expression, the prognostic power of these genes became more significant.

Discussion

In this study, we evaluated the global gene expression profile to clarify the molecular etiology of the rapid progression specific to serous histological type cancers. We first attempted to subclassify our 43 serous type ovarian cancer tissues comprising eight stage I samples and 35 stage III/IV samples by a stepwise extraction of genes reflecting expression differences between samples (Supplementary Fig. 1). Although various classification methods have been proposed to characterize various cancer types at the molecular level using gene expression data, most of the methods tend to be unstable, producing different clusters with slightly different input or different choice of initial conditions.[37] Brunet *et al.*[28] showed that NMF is able to recover biologically significant phenotypes and appears superior to other methods especially when prior knowledge is lacking or undetermined. By applying the NMF algorithm, 35 patients with advanced stage serous ovarian cancer were grouped into two subclasses with 112 subclass-specific genes representing unique characteristics of tumor progression. Interestingly, one subclass, subclass 1 ($n = 17$), with a gene expression profile similar to that of stage I, showed a favorable outcome compared to the other subclass,

© 2009 Japanese Cancer Association

Table 3. Multivariable Cox's proportional hazards model analysis of prognostic factors for progression-free survival and overall survival in patients with advanced stage serous ovarian cancers ($n = 72$)

Variable	Hazard ratio (95% CI)	*P*-value
Progression-free survival		
ZEB2 expression	1.37 (1.07–1.78)	0.014*
Age	0.98 (0.96–1.00)	0.095
Optimal surgery (vs not optimal)	0.60 (0.44–0.82)	0.0011*
Grade 2 (vs Grade 1)	0.85 (0.58–1.24)	0.42
Grade 3 (vs Grade 1)	1.41 (0.98–2.06)	0.060
Overall survival		
ZEB2 expression	1.53 (1.05–2.22)	0.027*
Age	1.01 (0.96–1.04)	0.71
Optimal surgery (vs not optimal)	0.67 (0.41–1.05)	0.079
Grade 2 (vs Grade 1)	0.83 (0.47–1.50)	0.53
Grade 3 (vs Grade 1)	1.51 (0.93–2.62)	0.10

*Statistically significant ($P < 0.05$).

Table 4. Comparison of progression-free survival and overall survival in four groups with different expression profiles of *CDH1* and *ZEB2*

Serous ovarian cancer ($n = 72$)	Hazard ratio	95% CI	*P*-value
Progression-free survival			
CDH1 high/*ZEB2* low ($n = 23$)	1.00		
CDH1 high/*ZEB2* high ($n = 13$)	0.91	(0.53–1.43)	0.69
CDH1 low/*ZEB2* low ($n = 13$)	1.29	(0.83–1.94)	0.25
CDH1 low/*ZEB2* high ($n = 23$)	1.65	(1.18–2.35)	0.0035*
Overall survival			
CDH1 high/*ZEB2* low ($n = 23$)	1.00		
CDH1 high/*ZEB2* high ($n = 13$)	0.96	(0.37–1.96)	0.91
CDH1 low/*ZEB2* low ($n = 13$)	1.12	(0.63–1.95)	0.70
CDH1 low/*ZEB2* high ($n = 23$)	1.77	(1.12–2.92)	0.013*

subclass 2 ($n = 18$). This result was compatible with findings by Berchuck *et al.*[7] demonstrating similarities in gene expression between early stage serous ovarian cancers and a subset of advanced stage serous ovarian cancers that had favorable prognosis. Regarding the sample size in the current microarray analysis, one can realize that this may be first-stage evidence on ovarian expression profile associated with tumor progression. However, we successfully provided valuable insights that clarify the molecular mechanism of tumor progression using NMF algorithm.

Kurman *et al.* divide epithelial ovarian cancers into two groups designated type I and type II based on clinical, pathological, and molecular genetic studies.[21] Type I tumors are low grade and slow growing (including endometrioid, mucinous, and low-grade serous). Type II tumors (including high grade serous and undifferentiated) are rapidly growing, more aggressive, and are frequently associated with *TP53* mutation. In our experiments, the frequency of *TP53* mutation was higher in cases belonging to subclass 2 (9/18, 50%) compared to those belonging to subclass 1 + stage I (8/25, 32%). Although the frequency difference was not statistically significant, our novel subclassification based on gene expression profile might have a potential relationship with that of the two-type classification model of ovarian cancer proposed by Kurman *et al.*[21] Further study will be necessary to elucidate other biological and pathological implications except tumor progression in our subclassification.

After screening genes associated with tumor progression and subsequent validation of the association, we identified the expression of *ZEB2* and *CDH1* as prognostic factors for serous ovarian cancers. Although other genome-wide expression analyses[7–10] have identified gene expression profiles with prognosis values in patients with ovarian cancer, *ZEB2* and *CDH1* are not listed in their profiles. Previous studies using the expression microarrays investigate directly the association between gene expression level and survival time in patients with ovarian cancer, whereas we first extracted gene expression profiles reflecting tumor progression by a stepwise approach (Supplementary Fig. 1), and selected survival-associated genes with biological function from these genes. Furthermore, differences in microarray platforms, normalization methods, degrees of contamination by non-cancer cells in a given tumor specimen, and the patient populations under study[38] were observed between previous reports and ours. These points might contribute to the development of inconsistencies in lists of survival-associated genes from the microarray studies.

Our data also suggest that reduced *CDH1* expression is a key to subclassify advanced stage serous ovarian cancers. Recently Tothill *et al.* reported that six molecular subtypes of ovarian cancers, including serous and endometrioid histological types, were identified by a *k*-means clustering method according to genome-wide expression data from 285 ovarian cancer samples.[39] Of the six molecular subtypes, one subtype (C5 in the paper), comprising mainly high grade serous ovarian cancer samples, is characterized by reduced E-cadherin. Despite the difference in experimental design of the two studies, our data are compatible with their finding that a molecular subtype of ovarian cancers can be tagged by E-cadherin expression. E-cadherin is a hallmark of epithelial–mesenchymal transition, and a reduction of E-cadherin is thought to result in dysfunction of the cell–cell junction system, triggering cancer invasion in various human malignancies. In our experiment, E-cadherin expression was significantly associated with prognosis in patients with advanced stage serous ovarian cancer at both the mRNA and protein levels. Therefore, it is important to clarify the regulatory mechanisms of *CDH1* expression[40] in serous ovarian cancer in terms of tumor progression and prognosis, as well as subclassification.

Recent study shows that the interaction of Snail, ZEB, and bHLH factors regulates *CDH1* repression and epithelial–mesenchymal transition.[23] Besides *ZEB2*, other transcriptional repressors may reduce *CDH1* expression and lead to epithelial–mesenchymal transition.[41] Indeed, *SNAI2* was included in the 112 subclass-specific genes, and was found to directly interact with *CDH1* in the newly obtained IPA network (Fig. 2B). Previous reports show that other transcriptional repressors such as Snail 1 and Twist are related to prognosis in ovarian cancer, using immunohistochemical analysis.[35,42] Hosono *et al.*[42] have reported that expression of Twist is a significant prognostic factor in non-serous type but not in serous type tumors. Our results demonstrate that expression of *ZEB2* is negatively correlated with *CDH1* expression, and that the expression signature of increased *ZEB2* and reduced *CDH1* in ovarian tumor tissues is related to poor prognosis in serous ovarian cancer patients (Table 4). Furthermore, siRNA-mediated suppression of *ZEB2* in the serous type of ovarian cancer SKOV3 cells leads to an increase in *CDH1* expression (Supplementary Fig. 3), suggesting that *ZEB2* regulates *CDH1* expression in serous histological type tumors. To validate that *ZEB2* expression at the protein level is a significant prognostic factor, we would like to analyze *ZEB2* expression in a larger number of patients stratified according to individual histological types using immunohistochemical staining.

Park *et al.* have recently reported that microRNA-200 directly targets the mRNA of *ZEB2* as well as that of *ZEB1*, and indirectly controls the expression level of *CDH1* in cancer cell lines.[43] Further investigation is required to elucidate the more detailed mechanisms by which the *ZEB2*–*CDH1* axis in epithelial–mesenchymal transition is regulated in the process of ovarian cancer progression. Clarification of the mechanisms for the regulation of *ZEB2*–*CDH1* expression may provide plausible targets for the development of therapeutic strategies in the clinical management of serous ovarian cancers.

doi: 10.1111/j.1349-7006.2009.01204.x
© 2009 Japanese Cancer Association

Acknowledgments

This work was supported in part by a Grant-in-Aid for the Third-term Cancer Control Strategy Program from the Ministry of Health, Labor and Welfare, Japan. We are grateful to Hiroshi Kamiguchi and Tadayuki Satoh (Teaching and Research Support Center, Tokai University School of Medicine) for their technical support in the microarray experiment, and also thank Yoshiko Sakamoto, Eriko Tokubo, Hiromi Kamura, and Kozue Otaka for their technical assistance.

Abbreviations

ACTA2	actin, alpha 2, smooth muscle, aorta
ACTB	actin, beta
bHLH	basic helix-loop-helix
BRCA1	breast cancer 1, early onset
CA125	carbohydrate antigen 125
CDH1	cadherin 1
COL16A1	collagen, type XVI, alpha 1
coph	cophenetic correlation coefficient
Cy3	cyamine 3-CTP
FN1	fibronectin 1
GO	gene ontology
IPA	Ingenuity Pathway Analysis
LTBP2	latent transforming growth factor beta binding protein 2
NMF	non-negative matrix factorization
SNAI1	snail homolog 1
TBP	TATA box binding protein
TP53	Tumor Protein p53
VIM	vimentin
ZEB2	zinc finger E-box binding homeobox 2

References

1 Disaia PJ, Creasman WT. Epithelial ovarian cancer. In: Disaia PJ, Creasman WT, eds. *Clinical Gynecologic Oncology*, 6th edn. St Louis: Mosby, 2002; 289–350.
2 Cannistra SA. Cancer of the ovary. *N Engl J Med* 2004; **351**: 2519–29.
3 Agarwal R, Kaye SB. Expression profiling and individualization of treatment for ovarian cancer. *Curr Opin Pharmacol* 2006; **6**: 345–9.
4 Olivier RI, van Beurden M, van't Veer LJ. The role of gene expression profiling in the clinical management of ovarian cancer. *Eur J Cancer* 2006; **42**: 2930–8.
5 Fehrmann RS, Li XY, van der Zee AG *et al*. Profiling studies in ovarian cancer: a review. *Oncologist* 2007; **12**: 960–6.
6 Spentzos D, Levine DA, Ramoni MF *et al*. Gene expression signature with independent prognostic significance in epithelial ovarian cancer. *J Clin Oncol* 2004; **22**: 4700–10.
7 Berchuck A, Iversen ES, Lancaster JM *et al*. Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. *Clin Cancer Res* 2005; **11**: 3686–96.
8 Hartmann LC, Lu KH, Linette GP *et al*. Gene expression profiles predict early relapse in ovarian cancer after platinum-paclitaxel chemotherapy. *Clin Cancer Res* 2005; **11**: 2149–55.
9 Bonome T, Levine DA, Shih J *et al*. A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res* 2008; **68**: 5478–86.
10 Le Page C, Ouellet V, Quinn MC, Tonin PN, Provencher DM, Mes-Masson AM. BTF4/BTNA3.2 and GCS as candidate mRNA prognostic markers in epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 913–20.
11 Dressman HK, Berchuck A, Chan G *et al*. An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J Clin Oncol* 2007; **25**: 517–25.
12 Newton TR, Parsons PG, Lincoln DJ *et al*. Expression profiling correlates with treatment response in women with advanced serous epithelial ovarian cancer. *Int J Cancer* 2006; **119**: 875–83.
13 Partheen K, Levan K, Osterberg L, Horvath G. Expression analysis of stage III serous ovarian adenocarcinoma distinguishes a sub-group of survivors. *Eur J Cancer* 2006; **42**: 2846–54.
14 Okamoto A, Nikaido T, Ochiai K *et al*. Indoleamine 2,3-dioxygenase serves as a marker of poor prognosis in gene expression profiles of serous ovarian cancer cells. *Clin Cancer Res* 2005; **11**: 6030–9.
15 Donninger H, Bonome T, Radonovich M *et al*. Whole genome expression profiling of advance stage papillary serous ovarian cancer reveals activated pathways. *Oncogene* 2004; **23**: 8065–77.
16 Meinhold-Heerlein I, Bauerschlag D, Hilpert F *et al*. Molecular and prognostic distinction between serous ovarian carcinomas of varying grade and malignant potential. *Oncogene* 2005; **24**: 1053–65.
17 Bonome T, Lee JY, Park DC *et al*. Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. *Cancer Res* 2005; **65**: 10602–12.
18 Shridhar V, Lee J, Pandita A *et al*. Genetic analysis of early-versus late-stage ovarian tumors. *Cancer Res* 2001; **61**: 5895–904.
19 De Cecco L, Marchionni L, Gariboldi M *et al*. Gene expression profiling of advanced ovarian cancer: characterization of a molecular signature involving fibroblast growth factor 2. *Oncogene* 2004; **23**: 8171–83.
20 Lancaster JM, Dressman HK, Clarke JP *et al*. Identification of genes associated with ovarian cancer metastasis using microarray expression analysis. *Int J Gynecol Cancer* 2006; **16**: 1733–45.
21 Kurman RJ, Visvanathan K, Roden R, Wu TC, Shin IeM. Early detection and treatment of ovarian cancer: shifting from early stage to minimal volume disease based on a new model of carcinogenesis. *Am J Obstet Gynecol* 2008; **198**: 351–6.
22 Thiery JP. Epithelial mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2002; **2**: 442–54.
23 Peinado H, Olmeda D, Cano A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* 2007; **7**: 415–28.
24 FIGO Cancer Committee. Staging Announcement: FIGO Cancer Committee. *Gynecol Oncol* 1986; **25**: 383–5.
25 Amikura T, Sekine M, Hirai Y *et al*. Mutational analysis of TP53 and p21 in familial and sporadic ovarian cancer in Japan. *Gynecol Oncol* 2006; **100**: 365–71.
26 Sekine M, Nagata H, Tsuji S *et al*. Mutational analysis of BRCA1 and BRCA2 and clinicopathologic analysis of ovarian cancer in 82 ovarian cancer families: two common founder mutations of BRCA1 in Japanese population. *Clin Cancer Res* 2001; **7**: 3144–50.
27 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995; **57**: 289–300.
28 Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci USA* 2004; **101**: 4164–9.
29 Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, Inoue I. Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates ART3 in genetic susceptibility. *PLoS Genet* 2008; **4**: e26.
30 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* 2001; **25**: 402–8.
31 Inamura K, Fujiwara T, Hoshida Y *et al*. Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. *Oncogene* 2005; **24**: 7105–13.
32 Daraï E, Scoazec JY, Walker-Combrouze F *et al*. Expression of cadherins in benign, borderline, and malignant ovarian epithelial tumors: a clinicopathologic study of 60 cases. *Hum Pathol* 1997; **28**: 922–8.
33 Faleiro-Rodrigues C, Macedo-Pinto I, Pereira D, Lopes CS. Prognostic value of E-cadherin immunoexpression in patients with primary ovarian carcinomas. *Ann Oncol* 2004; **15**: 1535–42.
34 Voutilainen KA, Anttila MA, Sillanpää SM *et al*. Prognostic significance of E-cadherin–catenin complex in epithelial ovarian cancer. *J Clin Pathol* 2006; **59**: 460–7.
35 Blechschmidt K, Sassen S, Schmalfeldt B, Schuster T, Höfler H, Becker KF. The E-cadherin repressor Snail is associated with lower overall survival of ovarian cancer patients. *Br J Cancer* 2007; **98**: 489–95.
36 Imamichi Y, König A, Gress T, Menke A. Collagen type I-induced Smad-interacting protein 1 expression downregulates E-cadherin in pancreatic cancer. *Oncogene* 2007; **26**: 2381–5.
37 Gao Y, Church G. Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics* 2005; **21**: 3970–5.
38 Konstantinopoulos PA, Spentzos D, Cannistra SA. Gene-expression profiling in epithelial ovarian cancer. *Nat Clin Pract Oncol* 2008; **5**: 577–87.
39 Tothill RW, Tinker AV, George J *et al*. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* 2008; **14**: 5198–208.
40 Liu YN, Lee WW, Wang CY, Chao TH, Chen Y, Chen JH. Regulatory mechanisms controlling human E-cadherin gene expression. *Oncogene* 2005; **24**: 8277–90.
41 Imai T, Horiuchi A, Wang C *et al*. Hypoxia attenuates the expression of E-cadherin via up-regulation of SNAIL in ovarian carcinoma cells. *Am J Pathol* 2003; **163**: 1437–47.

© 2009 Japanese Cancer Association

42 Hosono S, Kajiyama H, Terauchi M *et al*. Expression of Twist increases the risk for recurrence and for poor survival in epithelial ovarian carcinoma patients. *Br J Cancer* 2007; **96**: 314–20.

43 Park SM, Gaur AB, Lengyel E, Peter ME. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 2008; **22**: 894–907.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Analytical process to extract 'subclass-specific genes'.

Fig. S2. Association between E-cadherin expression and prognosis of advanced stage serous ovarian cancers validated by immunohistochemical analyses.

Fig. S3. Interaction between *ZEB2* and *CDH1*.

Table S1. Comparison of clinicopathological characteristics between microarray set and validation set

Table S2. List of 23 transcripts analyzed by quantitative real-time RT-PCR in this study

Table S3. One hundred and twelve transcripts representing statistically significant expression differences between two subclasses of advanced stage serous ovarian cancers

Table S4. Expression levels of 23 genes by quantitative real-time RT-PCR were significantly different between subclass 1 (S1) and subclass 2 (S2)

Supplementary Methods Methods about GO analysis, Pathway analysis, siRNA experiments, and immunohistochemical analysis

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

doi: 10.1111/j.1349-7006.2009.01204.x
© 2009 Japanese Cancer Association