Title:
High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by down-regulation of antigen presentation pathway.

Authors:

The Japanese Ovarian Cancer Study Group, listed in random order
Mitsuaki Suzuki (Jichi Medical University), Yoshitaka Onishi, Kazunobu Sueyoshi, and Sumika Matsukida (Kagoshima City Hospital), Yoshiki Kudo (Hiroshima University), Hironori Tashiro (Kumamoto University), Tomoya Mizunoe (Kure Medical Center and Chugoku Cancer Center), Junzo Kigawa, Kanae Nosaka, and Hisao Ito (Tottori University), Sohei Yamamoto and Hideyuki Shimazaki (National Defense Medical College), Ken Takizawa (Cancer Institute Hospital), Kiyoko Kato and Satoru Takeda (Juntendo University), Yutaka Ueda, Yukari Miyoshi, Toshihiro Kimura, and Tadashi Kimura (Osaka University), Sosuke Adachi, Koji Nishino, Takehiro Serikawa, Tetsuro Yahata, Junko Sakurada, Go Hasegawa, and Nobutaka Kitamura (Niigata University)

1. Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan.
2. Laboratory for Medical Informatics, Center for Genomic Medicine, RIKEN Yokohama Institute, Yokohama, Japan.
3. Department of Obstetrics and Gynecology, Jichi Medical University, Shimotsuke, Japan.
4. Department of Obstetrics and Gynecology, Kagoshima City Hospital, Kagoshima, Japan.
5. Department of Obstetrics and Gynecology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan.
6. Department of Obstetrics and Gynecology, School of Medicine, Nagasaki University, Nagasaki, Japan.
7. Department of Gynecology and Obstetrics, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan.
8. Department of Gynecology, National Hospital Organization, Kure Medical Center and Chugoku Cancer Center, Kure, Japan.
9. Department of Obstetrics and Gynecology, Jikei University School of Medicine, Tokyo, Japan.
10. Department of Gynecology, National Hospital Organization, Shikoku Cancer Center, Matsuyama, Japan.
11. Department of Gynecology and Obstetrics, Kyoto University Graduate School of Medicine, Kyoto, Japan.
12. Department of Obstetrics and Gynecology, Fujita Health University School of Medicine, Toyoake, Japan.
13. Department of Obstetrics and Gynecology, Sapporo Medical University, Sapporo, Japan.
14. Department of Obstetrics and Gynecology, Tottori University School of Medicine, Yonago, Japan.
16. Department of Obstetrics and Gynecology, Yokohama City University Hospital, Yokohama, Japan.
17. Department of Gynecologic Oncology, Hyogo Cancer Center, Akashi, Japan.
18. Department of Obstetrics and Gynecology, Kurume University School of Medicine, Kurume, Japan.
19. Department of Gynecology, Cancer Institute Hospital, Tokyo, Japan.
20. Department of Obstetrics and Gynecology, Saitama Medical Center, Saitama Medical University, Kawagoe, Japan.
21. Department of Obstetrics and Gynecology, Faculty of Medicine, Juntendo University, Tokyo, Japan.
22. Departments of Obstetrics and Gynecology, Osaka University Graduate School of Medicine, Suita, Japan.
23. Department of Obstetrics and Gynecology, Tokai University School of Medicine, Isehara, Japan.
24. Department of Medical Informatics and Statistics, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan.
25. Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan.
26. Department of Pathology, Kawasaki Medical School, Kurashiki, Japan.
28. Division of Human Genetics, National Institute of Genetics, Mishima, Japan.

Running title
Novel risk classification in ovarian cancer

Keywords
Ovarian cancer, gene expression profiling, risk classification, antigen presentation pathway, immune activation
Financial support:
None

Corresponding author:
Kenichi Tanaka, M.D., Ph.D.
Address: 1-757 Asahimachi-dori, Chuo-ku, Niigata City, 951-8510, Japan,
Telephone: +81-25-227-2317, Fax: +81-25-227-0789, E-mail: tanaken@med.niigata-u.ac.jp

Conflicts of interests
The authors declare no competing financial interests.

Word count
Translational Relevance: 93 words
Abstract: 249 words
Text: 4189 words
Figures/Tables: 3/3 and supplementary information
References: 43
Translational relevance

Using large-scale microarray expression datasets (n = 1,054) by applying an elastic net method, a novel risk classification system for predicting overall survival of patients with advanced-stage high-grade serous ovarian cancer based on a 126-gene expression signature was developed and successfully validated. This study has profound significance in clarifying the down-regulation of HLA class I antigen presentation machinery that characterizes high-risk ovarian cancer. These results from comprehensive gene expression analysis using large-scale microarray data suggest that our risk classification system might have the potential to optimize treatment of high-grade serous ovarian cancer patients.
Abstract

Purpose: High-grade serous ovarian cancers are heterogeneous not only in term of clinical outcome but also at the molecular level. Our aim was to establish a novel risk classification system based on a gene expression signature for predicting overall survival, leading to suggesting novel therapeutic strategies for high-risk patients.

Experimental Design: In this large-scale cross-platform study of six microarray datasets consisting of 1,054 ovarian cancer patients, we developed a gene expression signature for predicting overall survival by applying elastic net and 10-fold cross-validation to a Japanese dataset A (n = 260) and evaluated the signature in five other datasets. Subsequently, we investigated differences in the biological characteristics between high- and low-risk ovarian cancer groups.

Results: An elastic net analysis identified a 126-gene expression signature for predicting overall survival in patients with ovarian cancer using the Japanese dataset A (multivariate analysis, $P = 4 \times 10^{-20}$). We validated its predictive ability with five other datasets using multivariate analysis (Tothill’s dataset, $P = 1 \times 10^{-5}$; Bonome’s dataset, $P = 0.0033$; Dressman’s dataset, $P = 0.0016$; TCGA dataset, $P = 0.0027$; Japanese dataset B, $P = 0.021$). Through gene ontology and pathway analyses, we identified a significant reduction in expression of immune-response related genes, especially on the antigen presentation pathway, in high-risk ovarian cancer patients.

Conclusions: This risk classification based on the 126-gene expression signature is an accurate predictor of clinical outcome in patients with advanced-stage high-grade serous ovarian cancer and has the potential to develop new therapeutic strategies for high-grade serous ovarian cancer patients.
Introduction

High-grade serous ovarian cancer comprises approximately 40% of epithelial ovarian cancer cases and is the most aggressive histologic type (1-4). This type of cancer usually presents as advanced-stage disease at the time of diagnosis, because there are no symptoms present at the early stage and no reliable screening test for early detection (1-4). Patients with advanced-stage high-grade serous ovarian cancer generally undergo primary debulking surgery followed by platinum-taxane chemotherapy. However, 30-40% of patients recur within 12 months after the standard treatment, and the overall five-year survival rate remains at approximately 30% (5,6). Clinicopathological characteristics, such as the International Federation of Gynecology and Obstetrics (FIGO) stage, histological grade, and debulking status after primary surgery, are clinically considered important clinical prognostic indicators of ovarian cancer but are insufficient for predicting survival time.

The development of microarray technology has provided new insights into cancer diagnosis and treatment. Large-scale microarray studies in breast cancer have succeeded in clarifying five molecular subtypes based on gene expression profiles, and in developing genomic biomarker for predicting recurrence in early breast cancer (MammaPrint™; Agendia, Amsterdam, the Netherlands) (7,8). Thus, breast cancer treatment strategies are being stratified according to molecular characteristics. In contrast, there are no gene expression signatures with high accuracy and reproducibility for clinical diagnosis and management in patients with ovarian cancer because there is a paucity of ovarian cancer samples available for microarray analysis compared to breast cancer. Although TP53 somatic mutation is present in almost all high-grade serous ovarian cancer and plays an important role in the pathogenesis (9,10), high-grade serous ovarian cancer exhibits much biological and molecular heterogeneity that should be considered when developing a novel therapeutic strategy for ovarian cancer (10,11).

In this study, we aimed to establish a novel system for predicting the prognosis of patients with advanced-stage high-grade serous ovarian cancer using large-scale microarray datasets (n=1,054) (10-13), leading to an optimal treatment based on molecular characteristics (14).
Methods

Clinical samples

Three hundred Japanese patients who were diagnosed with advanced-stage high-grade serous ovarian cancer between July 1997 and June 2010 were included in this study. All patients provided written informed consent for the collection of samples and subsequent analysis. Fresh frozen samples were obtained from primary tumor tissues during debulking surgery prior to chemotherapy. All patients with advanced-stage high-grade serous ovarian cancer were treated with platinum-taxane standard chemotherapy after surgery. In principle, patients were seen every 1 to 3 months for the first 2 years. Thereafter, follow-up visits had an interval of 3 to 6 months in the third to fifth year, and 6 to 12 months in the sixth to tenth year. At every follow-up visit, general physical and gynecologic examinations were performed. CA125 serum levels were routinely determined.

Staging of the disease was assessed according to the criteria of the International Federation of Gynecology and Obstetrics (FIGO) (15). Optimal debulking surgery was defined as less than 1 cm of gross residual disease, and suboptimal debulking surgery was defined as more than 1 cm of residual disease. Progression-free survival time was calculated as the interval from primary surgery to disease progression or recurrence. Based on the Response Evaluation Criteria In Solid Tumors (RECIST) version 1.1 (16), disease progression was defined as at least a 20% increase in the sum of the diameters of target lesions, as unequivocal progression of existing non-target lesions, or as the appearance of one or more new lesions. Overall survival time was calculated as the interval from primary surgery to the death due to ovarian cancer.

The histological characteristics of surgically resected specimens, which were diagnosed as serous histologic type by two pathologists at each institute, were at first assessed on formalin-fixed and paraffin-embedded hematoxylin and eosin sections by eight gynecological pathologists who belong to department of pathology in four institutes (Niigata University, National Defense Medical Collage, Tottori University, Kagoshima City Hospital). Next, these sections were reviewed under
central pathological review by two independent gynecological pathologists (H.T. and T.M.) with no knowledge of patients' clinical data. Histological subtype was diagnosed based on WHO classification of ovarian tumors (17). The degree of histological differentiation is determined according to Silverberg classification (18).

Microarray experiments

Frozen tissues containing more than 70% tumor cells upon histological evaluation were used for RNA extraction. Total RNA was extracted from tissue samples as previously described (19). Five hundred nanograms of total RNA were converted into labeled cRNA with nucleotides coupled to a cyanine 3-CTP (Cy3) (PerkinElmer, Boston, MA, USA) using the Quick Amp Labeling Kit, one-color (Agilent Technologies, Tokyo, Japan). Cy3-labeled cRNA (1.65 µg) was hybridized for 17 hours at 65°C to an Agilent Whole Human Genome Oligo Microarray (G4112F), which carries 60-mer probes to more than 40,000 human transcripts. The hybridized microarray was washed and then scanned in Cy3 channel with the Agilent DNA Microarray Scanner [model G2565CA, n = 260 (Japanese dataset A); G2565AA, n = 40 (Japanese dataset B)]. Signal intensity per spot was generated from the scanned image using Feature Extraction Software (version 10.1, Japanese dataset A; version 9.1, Japanese dataset B) (Agilent Technologies) with default settings. Spots that did not pass quality control procedures were flagged as “Not Detected”.

Next, we obtained Affymetrix HG-U133Plus2.0 microarray (Affymetrix, Santa Clara, CA, USA) data from 10 ovarian cancer samples that had been already been analyzed using the Agilent Whole Human Genome Oligo Microarray. Ten ovarian cancer samples were randomly selected from 260 samples in the Japanese dataset A. Microarray experiments were performed according to the Affymetrix recommended protocols. Briefly, biotinylated cRNA were synthesized by GeneChip 3'IVT Express Kit (Affymetrix) from 250 ng total RNA according to the manufacturer's instructions. Biotinylated cRNA yield were checked with the NanoDrop ND-1000 Spectrophotometer. Following fragmentation, 10 ug of cRNA were hybridized for 16 hr at 45°C on GeneChip Human Genome U133
Plus 2.0 Array. GeneChips were washed and stained in the Affymetrix Fluidics Station 450, and scanned using GeneChip Scanner 3000 7G.

The MIAME-compliant microarray data were deposited into the Gene Expression Omnibus data repository (accession number GSE32062 and GSE32063).

**Microarray Data analysis**

We prepared two our microarray datasets [Japanese dataset A (n = 260) and B (n = 40)] and four publicly available large sample-sized (n>100) microarray datasets [TCGA dataset (http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) (11), Tothill’s dataset (GSE9891) (12), Bonome’s dataset (GSE26712) (13), and Dressman’s dataset (http://data.genome.duke.edu/platinum.php) (14)] to discover predictive biomarkers. Clinical information of publicly available microarray datasets was obtained from their articles and websites. From Tothill’s original dataset (n = 285), we selected 131 samples that (i) were diagnosed as advanced-stage serous adenocarcinoma, (ii) were treated by platinum/taxane-based chemotherapy, and (iii) have clinical data about onset age, stage, grade, surgery, and survival time. Publicly available clinical information in TCGA was downloaded from TCGA Data portal (http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) at June 25, 2011. By the same methods above, 319 samples were selected from 562 microarray data as TCGA dataset (Affymetrix HT-HG-U133A). In this study, patients who were treated by adjuvant chemotherapy including molecular-targeted agents were excluded.

On the Agilent platform, data normalization was performed in GeneSpring GX 11.5 (Agilent Technologies) as follows: (i) Threshold raw signals were set to 1.0. (ii) 75th percentile normalization. Affymetrix microarray data was normalized and summarized using robust multiple average (RMA) in GeneSpring GX 11.5 (Agilent Technologies). To compare the microarray datasets measured with four different platforms (Agilent Whole Human Genome Oligo Microarray, Affymetrix HG-U133A, HG-U133Plus2.0, and HT-HG-U133A), we selected genes common to all platforms based on the Entrez Gene ID and used the Median Rank Score method (20) for
cross-platform normalization (Supplementary Fig. S1). Of 22,277 probes that were common among three Affymetrix platforms, 20,331 probes were selected to be in one to one relation between probe and gene. Using translation function based on Entrez Gene ID in GeneSpring GX 11.5, the 19,704 transcripts that matched to the 20,331 probes of the Affymetrix platform were extracted from all transcripts on the Agilent platform. Considering differences in microarray platforms, coefficient of correlation (r) in each gene between two microarray platforms were measured. Using 10 ovarian cancer microarray data obtained from both Agilent Whole Human Genome Oligo Microarray and Affymetrix HG-U133Plus2.0 (GSE32062), 9,141 genes with high correlation (r > 0.8) were extracted. After Median Rank Score analysis (20), we evaluated median value of each gene in both platforms and selected 3,553 genes with high correlation (r > 0.8) whose absolute value of subtracting median values between two platforms was less than 1. Similar analyses were performed using 16 breast cancer microarray data (GSE17700) (21) from both Affymetrix HG-U133Plus2.0 and HG-U133A. From the resulting 1,746 genes, we removed 60 that were not flagged as “Detected” in more than 90% of the Japanese dataset A samples (n = 260), considering them to have either missing or uncertain expression signals. In addition, the data were normalized per gene in each dataset by transforming the expression of each gene to obtain a mean of 0 and standard deviation of 1 (Z-transformation) for the cross-platform study.

We analyzed Japanese dataset A as a “training set”, Tothill’s dataset as a “test set”, and the other 4 datasets as “validation sets”. We applied elastic net analysis (22) using the R (23) package glmnet to identify survival-related genes for prediction of prognosis in patients with advanced-stage high-grade serous ovarian cancer. Using 10-fold cross-validation, we obtained regression coefficients with optimal penalty parameter for the penalized Cox model, and calculated a prognostic index for each patient as defined by

$$\text{Prognostic index} = \sum_{i=1}^{126} \hat{\beta}_i \times X_i$$

where $\hat{\beta}_i$ is the estimated regression coefficient of each gene in the Japanese dataset A under elastic net ($\alpha = 0.05$) and $X_i$ is the Z-transformed expression value of each gene. The estimated regression
coefficient of each survival-related gene given by elastic net ($\alpha = 0.05$) in the Japanese dataset A was also applied to calculate a prognostic index for each patient in five other datasets using the equation above. We classified all patients into the two groups (high- and low-risk groups) by the optimal cutoff value of the prognostic index in the Japanese dataset A. Patients were assigned to the “high-risk” group if their prognostic index was greater than or equal to cutoff value of prognostic index, whereas “low-risk” group was composed of cases with the prognostic indices that were less than cutoff value. Because risk classification divided by cutoff value 0.1517 indicated a minimum $P$ value ($P = 1 \times 10^{-30}$) when logrank test was used to compare differences in overall survival between high- and low-risk groups in the Japanese dataset A, this value was determined as optimal cutoff value.

Both hierarchical clustering and non-negative factorization (NMF) algorithm (24) were used to assess the similarity of gene expression profiles among 126 survival-related genes. In the two major datasets (Japanese dataset A and TCGA dataset), we constructed a heat map with hierarchical clustering for both samples and genes using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). Correlation (centered) and complete linkage were selected on similarity metrics and clustering method, respectively. A heat map was visualized with Java TreeView (http://jtreeview.sourceforge.net/). NMF-consensus matrices averaging 50 connectivity matrices were computed at $K=2$-$7$ (as the number of subclasses modeled) for 126 genes. With genes appearing along both the horizontal and vertical axes of the consensus matrices, consistency in the gene-pair clusters was visualized. We determined the optimal number of gene clusters by the cophenetic correlation coefficient that provides a scalar summary of global clustering robustness across the consensus matrix in the Japanese dataset A and TCGA dataset.

We performed a volcano plot analysis to extract differentially expressed transcripts between high- and low-risk ovarian cancer groups using the two major datasets [Japanese dataset A ($n = 260$) and TCGA dataset ($n = 319$)]. When the volcano plot analysis was performed in GeneSpring GX 11.5 (Agilent Technologies), we used gene expression data prior to $Z$-transformation normalization (Fig. 3).
To investigate the biological functions of gene expression signatures, we used GO Ontology Browser, embedded in GeneSpring GX11.5 (Agilent Technologies). The GO Ontology Browser was used to analyze which categories of gene ontology were statistically overrepresented among the gene list obtained. Statistical significance was determined by Fisher’s exact test, followed by multiple testing corrections by the Benjamini and Yekutieli false discovery rate method (25). To avoid bias of gene extraction by volcano plot analysis in GO analysis, Gene Set Enrichment Analysis (GSEA) (26) was performed using genes prior to gene selection by volcano plot analysis. Analysis settings in GSEA (software version 2.07) were as follows: (i) Gene sets database: c5.all.v2.5.symbols.gmt [Gene ontology], (ii) Number of permutations: 1000, (iii) Collapse dataset to gene symbol: true, (iv) Permutation type: phenotype, (v) Chip platform(s): Agilent_HumanGenome.chip or HT_HG_U133A.chip and (vi) other settings: default.

Furthermore, we used the Core Analysis tool in the Ingenuity Pathway Analysis (IPA) system to analyze networks and pathways for a set of genes. Q value was calculated using Fisher’s exact test with the Benjamini and Hochberg correction (27). Q value < 0.25 was considered as significant in GO, GSEA, and IPA.

SNP array experiments

We isolated genomic DNA from tumor tissues using a phenol-chloroform extraction method and from normal lymphocytes using the QIAamp DNA Blood Maxi Kit (Qiagen, Tokyo, Japan). SNP array experiments with the Genome-Wide Human SNP Array 6.0 (Affymetrix) were carried out at Niigata University (details in the Supplementary Methods). SNP array data were analyzed by using the Partek Genomic Suite 6.5 (Partek Inc., St. Louis, MO, USA) to investigate copy number variations in 30 genes involved in the antigen presentation pathway.

Immunohistochemical analysis

A monoclonal anti-human CD8 antibody (M7103; 1:100; DAKO, Tokyo, Japan) was used for
immunohistochemical staining of CD8 T lymphocyte in tumor tissues. Details of the immunohistochemistry method and sample selections are described in the Supplementary Methods.

Statistical analysis

Standard statistical tests including Pearson’s correlation analysis, unpaired t-tests, one-way ANOVA, Fisher’s exact tests, logrank tests, and Cox proportional hazard model analysis were used to analyze the clinical data, as appropriate. Analyses of clinical data were performed using JMP version 8 (SAS Institute, Cary, NC, USA) and GraphPad PRISM version 4.0 (GraphPad Software, San Diego, CA, USA).

Results

The distribution of the clinical variables for each microarray dataset is shown in Table 1. To compare the microarray datasets measured with four different platforms, 1,686 genes were selected (Supplementary Fig. S1). An elastic net analysis (22) using 10-fold cross-validation on Japanese dataset A (training set, n = 260) identified a 126-gene signature for predicting overall survival in patients with advanced-stage high-grade serous ovarian cancer (Supplementary Table S1). After calculating the prognostic index for each sample from the 126-gene expression signature as reported previously (19), we divided the training set into high- and low-risk groups based on the optimal cutoff value (0.1517) of the prognostic index and verified the high predictive power of this risk classification [Logrank P = 1 × 10^{-30}; Multivariate Cox P = 4 × 10^{-20}; hazard ratio (HR), 6.203; 95% confidence interval (CI), 4.239-9.123] (Fig. 1A and Table 2).

The predictive power of the 126-gene signature was tested with Tothill’s dataset (n = 131) (12). Both Kaplan-Meier survival analysis and multivariate analysis showed that this risk classification was significantly associated with overall survival time in Tothill’s dataset (Logrank P = 3 × 10^{-6}; Multivariate Cox P = 1 × 10^{-5}; HR, 3.728; 95% CI, 2.110-6.532) (Fig. 1B and Table 2). Next, we assessed the predictive power of the 126-gene expression signature on the three datasets in which
microarray data was obtained from Affymetrix platforms and confirmed that this risk classification indicated similar results in the three datasets (Fig. 1C-E and Table 2). Furthermore, to exclude the influence of differences in microarray platforms, we prepared Japanese dataset B from the same platform as the training set. Despite the small sample size, our risk classification demonstrated a significant association with overall survival time in the Japanese dataset B (Fig. 1F and Table 2). This risk classification showed high predictive accuracy even though overall survival was capped at 60 months (Supplementary Fig. S2) (10). Moreover, the high-risk groups had shorter progression-free survival times compared to the low-risk groups in four microarray datasets with available progression-free survival times (Supplementary Fig. S3). In addition, we applied 193 overall survival related gene signature that have been recently reported by TCGA research networks (10) to our large dataset (Japanese dataset A), and 191 of 193 genes were available in our dataset. High-risk ovarian cancer patients based on 191-gene signature model had significantly poor prognosis compared to low-risk patients (Supplementary Methods and Fig. S4).

We next investigated the molecular characteristics of the 126-gene signature with both hierarchical clustering and NMF (24) analyses using the two major datasets (Japanese dataset A and TCGA dataset). Hierarchical clustering of the expression of 126 genes in the Japanese dataset A (n = 260) resulted in two gene clusters (cluster 1 and 2) (Fig. 2A). Cluster 1 (outlined in yellow in Fig. 2A) comprised 23 genes, 22 of which were common genes in a small cluster of NMF class assignment for K = 2 (Supplementary Fig. S5). Similarly, this gene cluster was seen both in hierarchical clustering and NMF analyses of TCGA dataset, and the 20 genes in TCGA’s cluster were consistent with these 23 genes in cluster 1 (Fig. 2B and Supplementary Fig. S5). GO analysis of these 23 genes revealed 132 significantly overrepresented GO categories; the top 20 GO categories are shown in the Supplementary Table S2. Specifically, immunity-related categories were enriched in these 23 genes and were down-regulated in the high-risk groups compared with the low-risk groups. On the other hand, only one category [cytoplasm (GO0005737)] belonging to ‘cellular component’ was significantly over-represented in 103 genes of cluster 2.
To clarify the biological significance of our risk classification based on the 126-gene signature in advanced-stage high-grade serous ovarian cancer, we examined differences in molecular biological characteristics between high- and low-risk groups. For subsequent analyses, all transcripts on each platform were used to avoid the previous gene number limitations, which were for the cross-platform study. We extracted the genes differentially expressed between the two groups by performing a volcano plot analysis using the two major datasets (Japanese dataset A and TCGA dataset) (Supplementary Table S3). Figure 3A shows that 1,109 and 1,381 transcripts were differentially expressed between the high- and low-risk groups in the Japanese dataset A and TCGA dataset, respectively. GO analysis of these transcripts indicated that 9 out of 20 top-ranked GO categories were common between both datasets and involved in the immune system (Table 3).

To exclude the influence of cutoff in fold change or \( P \) value in volcano plot analysis, we re-evaluated differences in the biological characteristics between high- and low-risk groups using GSEA (26). Four immunity-related GO categories (immune system response, immune response, defense response, and inflammatory response) were also included in the list of 20 top-ranked GO categories when GSEA was performed using 21,785 (Japanese dataset A) and 14,023 (TCGA dataset) transcripts prior to gene extraction (Supplementary Table S4).

Furthermore, possible functional relations among differentially expressed genes between high- and low-groups were investigated using pathway analysis. Of 20 top-ranked pathways notably enriched in 1,109 transcripts of the Japanese dataset A and 1,381 transcripts of the TCGA dataset, 14 pathways were common between both datasets (Supplementary Table S5). In particular, the antigen presentation pathway (Fig. 3B) was the most significantly over-represented pathway in the two datasets (Japanese dataset A, \( Q = 2.0 \times 10^{-29} \); TCGA dataset, \( Q = 5.0 \times 10^{-14} \)). In the antigen presentation pathway, 30 transcripts in this pathway were significantly down-regulated in the high-risk group (Fig. 3B and Supplementary Table S6). We further examined whether molecular defects of HLA class I antigen presentation machinery components were caused by structural alterations using SNP array data. Our SNP array data showed that genes of HLA class I antigen
presentation machinery were not deleted (Supplementary Table S6). In the TCGA dataset, only 10% of cases (24/242) had deletions in HLA class I genes. On the other hand, 89% of high-risk cases (55/62) without deletion in HLA class I genes showed significantly lower expressions in HLA class I genes compared with those in low-risk groups (Supplementary Fig. S6).

On the basis of this result, we assessed the status of tumor-infiltrating lymphocyte reflecting an immune response against ovarian cancer. By immunohistochemically staining for CD8-positive T lymphocytes in a subset of the Japanese dataset (n = 30), we revealed that the number of CD8 T lymphocytes infiltrating into tumor tissues was significantly decreased in the high-risk group (n = 10) compared with the low-risk group (n = 20) and was clearly correlated with the expression level of HLA class I genes (Supplementary Fig. S7).

**Discussion**

In this study, we established a novel risk classification system based on the 126-gene expression signature for predicting overall survival time in patients with advanced-stage high-grade serous ovarian cancer. The significant association between our risk classification and overall survival time was indicated among the six microarray datasets.

In expression microarray analyses, there is a well-known “curse of dimensionality” problem that the number of genes is much larger than the number of samples. The curse of dimensionality leads to a concern about the reliability of the selected genes or an overfitting phenomenon, and using a large-scale microarray data is a simple and effective method to overcome this problem. From this theory, we planned a large-scale cross-platform study using four publicly available microarray datasets consisting of more than one hundred samples. To perform a cross-platform study, we took considerable care to reduce the influence of differences in microarray platforms using the Median Rank Score method and matching probes based on highly correlated expression (r>0.8) between the two platforms. As a result, this risk classification system showed a high predictive ability in the four microarray datasets from different platforms. Our risk classification
was suitably fitting to Japanese dataset B despite the small sample size. This might reflect the same platform in the microarray experiments and similarity in clinical backgrounds between Japanese dataset A and B. Intriguingly, 193 survival-related gene signature that has been reported by TCGA research networks (10) was also significantly associated with overall survival in our large dataset (Supplementary Fig.S4), and 5 genes (B4GALT5, CXCL9, ID4, MTRF1, and SLC7A11) were common between their and our gene signatures despite different analytical processes (Supplementary Table 2). These genes might contribute strongly to developing risk classification system for high-grade serous ovarian cancer patients.

We ascertained that alterations to the immune system in cancer cells are one of the most important factors affecting survival of patients with advanced-stage, high-grade serous ovarian cancer and that high-risk ovarian cancer was well characterized by the down-regulation of the antigen presentation pathway at the molecular level.

The two large-scale subclassification studies (10,11) based on gene expression analyses have indicated that there are at least four subclasses (immunoreactive, differentiated, proliferative, and mesenchymal) in high-grade serous ovarian cancer. Although there was no statistically significant difference in clinical outcome among four subclasses, this finding that immunity-related genes are identified in large-scale gene expression profiles despite different analytical approaches suggests that alteration of immune activity might play an important role in the molecular characterization of ovarian cancer.

Previous findings (28-30) that the presence of tumor-infiltrating T lymphocytes is associated with long survival in ovarian cancer patients are consistent with the results from our comprehensive gene expression analysis using large-scale microarray data, suggests that the presence of immune responses to cancer cells would influence the biological phenotype of ovarian cancer. Although several antigen-specific active immunotherapy studies have been conducted, the clinical benefit of this approach has not yet been shown in large randomized-controlled trials (31,32). Previous reports (28,33,34) and our data indicated that defects in HLA class I antigen presentation
machinery would decrease recruitment of tumor-infiltrating lymphocytes, leading to poor prognosis in cancer patients because of a reduction in antitumor immune activity. Therefore, up-regulation of HLA class I antigen presentation pathway might be one of efficient therapeutic approach for patients with high-risk serous ovarian cancer, especially when antigen-specific active immunotherapy is selected as a therapeutic strategy (35).

Defects of HLA class I gene expression occurs at the genetic, epigenetic, transcriptional, and posttranscriptional levels, and are classified into the two main groups: reversible and irreversible defects (36). Our data show that the frequency of deletion leading to irreversible defect of HLA class I gene expression was low in high-grade serous ovarian cancer. Only a few HLA class I gene mutations have been described thus far (36). Down-regulation of antigen presentation machinery components such as TAP1/2 or B2M also result in reversible defects in HLA class I molecules (37). Interestingly, the expression levels of genes in the antigen presentation pathway were positively correlated in the Japanese dataset A (Supplementary Fig. S8). IFN-γ or other cytokines stimulation induces the expression of genes in the antigen presentation pathway (38). Moreover, histone deacetylase (HDAC) inhibitors increase the expression of genes in the antigen presentation pathway such as TAP1/2 (39,40) and B2M (41) in several cancer cells leading to up-regulating the antigen presentation pathway. In addition, HDAC inhibitors that are promising anticancer drugs show a synergistic effect with taxane or platinum drugs, which were used as standard adjuvant chemotherapy for ovarian cancer, both in vitro and in vivo (42,43). These genetic and epigenetic activations of the antigen presentation pathway might induce immune recognition of ovarian cancer cells and enhance antitumor immune responses.

In summary, our data suggest that this predictive biomarker based on the 126-gene signature could identify patients who should not expect long-term survival by standard treatment and that activation of the antigen presentation pathway in tumor cells is an important key in new therapeutic strategies for ovarian cancer.
Acknowledgments

We greatly appreciate Professor Y. Nakamura for his help. We thank the tissue donors and supporting medical staff for making this study possible, C. Seki and A. Yukawa for their technical assistance, and T. Mizuochi for discussion. We are extremely grateful to Professor S. G. Silverberg for his enthusiastic help with the pathological review. We deeply appreciate K. Boroevich for helpful comments on the manuscript.

References

7. Dowsett M, Dunbier AK. Emerging biomarkers and new understanding of traditional markers in


23. R Development Core Team. R: A language and environment for statistical computing. 2011


Figure legends

Figure 1.
Kaplan-Meier survival analysis in six microarray datasets (A, Japanese dataset A; B, Tothill’s dataset; C, Bonome’s dataset; D, Dressman’s dataset; E, TCGA dataset; F, Japanese dataset B). Based on 126-gene signature, ovarian cancer patients were divided into two risk groups (red, high-risk; blue, low-risk). P values correspond to the logrank test comparing the survival curves.

Figure 2.
Molecular characteristics of 126 survival-related genes. (A) Heat map with a hierarchical classification of 260 patient samples with the expression profile of 126 survival-related genes in the Japanese dataset A. (B) Heat map with a hierarchical classification of 319 patient samples with the expression profile of 126 survival-related genes in TCGA dataset.

Figure 3.
Molecular characteristics of differentially expressed genes between high- and low-risk groups in the two major datasets (Japanese dataset A and TCGA dataset). (A) Using a volcano plot analysis, 1,109 and 1,381 transcripts were extracted as differentially expressed genes between high- and low-risk groups in the Japanese dataset A and TCGA dataset, respectively. (B) Ingenuity pathway analysis demonstrated that the antigen presentation pathway was significantly overrepresented in 1,109 Japanese dataset A and 1,381 TCGA transcripts. Green indicates that gene expression was down-regulated in high-risk group compared with in low-risk group.
Figure 2
Figure 3

A) Japanese dataset A
41,093 transcripts

21,785 transcripts

21,785 transcripts

Quality check

Volcano plot analysis

log2(Fold change)

log2(Fold change)

-3 -2 -1 0 1 2

-3 -2 -1 0 1 2

-3 -2 -1 0 1 2

-3 -2 -1 0 1 2

2

4

6

8

10

12

2

4

6

8

10

12

1,109 transcripts

909 genes

1,381 transcripts

1,077 genes

Fold change > 1.2

Welch’s t-Test

Benjamini-Hochberg FDR

Q < 0.001

137

86

443

434

252

296

B) TCGA dataset
22,277 transcripts

14,023 transcripts

14,023 transcripts

Quality check

Volcano plot analysis

log2(Fold change)

log2(Fold change)

-3 -2 -1 0 1 2

-3 -2 -1 0 1 2

-3 -2 -1 0 1 2

-3 -2 -1 0 1 2

2

4

6

8

10

12

2

4

6

8

10

12

1,109 transcripts

909 genes

1,381 transcripts

1,077 genes

Fold change > 1.2

Welch’s t-Test

Benjamini-Hochberg FDR

Q < 0.001

137

86

443

434

252

296
Table 1. Clinicopathological characteristics in six microarray datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Japanese dataset A</th>
<th>Tohill's dataset*</th>
<th>Bonome's dataset*</th>
<th>Dressman's dataset*</th>
<th>TCGA dataset‡</th>
<th>Japanese dataset B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>260</td>
<td>131</td>
<td>185</td>
<td>119</td>
<td>319</td>
<td>40</td>
</tr>
<tr>
<td>Age</td>
<td>58.2±10.8</td>
<td>58.4±9.8</td>
<td>62±12</td>
<td>N/A†</td>
<td>59.5±11.3</td>
<td>56.2±9.6</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serous</td>
<td>260</td>
<td>131</td>
<td>166</td>
<td>119</td>
<td>319</td>
<td>40</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage III</td>
<td>204</td>
<td>123</td>
<td>144</td>
<td>99</td>
<td>267</td>
<td>31</td>
</tr>
<tr>
<td>Stage IV</td>
<td>56</td>
<td>8</td>
<td>41</td>
<td>20</td>
<td>52</td>
<td>9</td>
</tr>
<tr>
<td>Grade</td>
<td>Silverberg</td>
<td>Silverberg</td>
<td>N/A†</td>
<td>N/A†</td>
<td>N/A†</td>
<td>Silverberg</td>
</tr>
<tr>
<td>grade 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>grade 2</td>
<td>131</td>
<td>51</td>
<td>40</td>
<td>57</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>grade 3</td>
<td>129</td>
<td>80</td>
<td>144</td>
<td>59</td>
<td>289</td>
<td>17</td>
</tr>
<tr>
<td>grade 4</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Surgery status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>103</td>
<td>78</td>
<td>92</td>
<td>63</td>
<td>236</td>
<td>19</td>
</tr>
<tr>
<td>Suboptimal</td>
<td>157</td>
<td>45</td>
<td>93</td>
<td>56</td>
<td>83</td>
<td>21</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platinum</td>
<td>260</td>
<td>131</td>
<td>185</td>
<td>119</td>
<td>319</td>
<td>40</td>
</tr>
<tr>
<td>Taxane</td>
<td>260</td>
<td>131</td>
<td>N/A</td>
<td>82</td>
<td>319</td>
<td>40</td>
</tr>
<tr>
<td>Follow-up period (months)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>42</td>
<td>29</td>
<td>38</td>
<td>34</td>
<td>31</td>
<td>39</td>
</tr>
<tr>
<td>Range</td>
<td>1-128</td>
<td>6-79</td>
<td>1-164</td>
<td>1-185</td>
<td>1-154</td>
<td>7-111</td>
</tr>
<tr>
<td>Number of deaths</td>
<td>121</td>
<td>60</td>
<td>129</td>
<td>69</td>
<td>187</td>
<td>22</td>
</tr>
<tr>
<td>Median survival (months)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microarray platform</td>
<td>Agilent Whole Human Genome Oligo Microarray</td>
<td>Affymetrix HG-U133 Plus2.0</td>
<td>Affymetrix HG-U133A</td>
<td>Affymetrix HG-U133A</td>
<td>Affymetrix HT-HG-U133A</td>
<td>Agilent Whole Human Genome Oligo Microarray</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Clinical information in these three dataset were obtained from their articles and website.
†No available information was described as N/A.
‡Publicly available clinical information in TCGA was downloaded from TCGA Data portal (http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) at June 25, 2011.
Table 2.

Univariate and multivariate Cox’s proportional hazard model analysis of prognostic factors for overall survival

<table>
<thead>
<tr>
<th></th>
<th>Univariate analysis</th>
<th></th>
<th></th>
<th>Multivariate analysis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazard ratio (95%CI*)</td>
<td>P value</td>
<td></td>
<td>Hazard ratio (95%CI)</td>
<td>P value</td>
</tr>
<tr>
<td>A) Japanese dataset A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.011 (0.993-1.029)</td>
<td>0.21</td>
<td></td>
<td>1.006 (0.988-1.024)</td>
<td>0.52</td>
</tr>
<tr>
<td>Stage IV (vs Stage III)</td>
<td>1.465 (0.968-2.165)</td>
<td>0.07</td>
<td></td>
<td>1.340 (0.884-1.984)</td>
<td>0.16</td>
</tr>
<tr>
<td>Optimal Surgery (vs suboptimal)</td>
<td>0.499 (0.334-0.730)</td>
<td>0.0003</td>
<td></td>
<td>0.643 (0.428-0.949)</td>
<td>0.026</td>
</tr>
<tr>
<td>B) Tothill’s dataset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.005 (0.977-1.035)</td>
<td>0.73</td>
<td></td>
<td>1.006 (0.978-1.035)</td>
<td>0.67</td>
</tr>
<tr>
<td>Stage IV (vs Stage III)</td>
<td>2.264 (0.785-5.176)</td>
<td>0.12</td>
<td></td>
<td>2.951 (0.983-7.235)</td>
<td>0.053</td>
</tr>
<tr>
<td>Optimal Surgery (vs suboptimal)</td>
<td>0.887 (0.525-1.517)</td>
<td>0.66</td>
<td></td>
<td>0.910 (0.528-1.592)</td>
<td>0.74</td>
</tr>
<tr>
<td>High (vs Low)</td>
<td>3.443 (1.967-5.962)</td>
<td>3×10^{-5}</td>
<td></td>
<td>3.728 (2.110-6.532)</td>
<td>1×10^{-5}</td>
</tr>
<tr>
<td>C) Bonome’s dataset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal Surgery (vs suboptimal)</td>
<td>0.592 (0.414-0.840)</td>
<td>0.0032</td>
<td></td>
<td>0.628 (0.438-0.894)</td>
<td>0.0097</td>
</tr>
<tr>
<td>High (vs Low)</td>
<td>2.034 (1.341-3.012)</td>
<td>0.0011</td>
<td></td>
<td>1.897 (1.247-2.818)</td>
<td>0.0033</td>
</tr>
<tr>
<td>D) Dressman’s dataset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal Surgery (vs suboptimal)</td>
<td>0.685 (0.425-1.102)</td>
<td>0.12</td>
<td></td>
<td>0.602 (0.370-0.977)</td>
<td>0.04</td>
</tr>
<tr>
<td>High (vs Low)</td>
<td>2.397 (1.335-4.140)</td>
<td>0.0042</td>
<td></td>
<td>2.687 (1.480-4.705)</td>
<td>0.0016</td>
</tr>
<tr>
<td>E) TCGA dataset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.014 (1.001-1.028)</td>
<td>0.038</td>
<td></td>
<td>1.011 (0.997-1.025)</td>
<td>0.13</td>
</tr>
<tr>
<td>Stage IV (vs Stage III)</td>
<td>1.018 (0.683-1.469)</td>
<td>0.93</td>
<td></td>
<td>1.034 (0.692-1.499)</td>
<td>0.86</td>
</tr>
<tr>
<td>Optimal Surgery (vs suboptimal)</td>
<td>0.909 (0.668-1.254)</td>
<td>0.56</td>
<td></td>
<td>0.903 (0.654-1.261)</td>
<td>0.54</td>
</tr>
<tr>
<td>High (vs Low)</td>
<td>1.712 (1.234-2.345)</td>
<td>0.0015</td>
<td></td>
<td>1.680 (1.202-2.321)</td>
<td>0.0027</td>
</tr>
<tr>
<td>F) Japanese dataset B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.059 (1.007-1.118)</td>
<td>0.024</td>
<td></td>
<td>1.058 (0.999-1.131)</td>
<td>0.057</td>
</tr>
<tr>
<td>Stage IV (vs Stage III)</td>
<td>1.696 (0.542-4.501)</td>
<td>0.34</td>
<td></td>
<td>2.998 (0.808-10.60)</td>
<td>0.098</td>
</tr>
<tr>
<td>Optimal Surgery (vs suboptimal)</td>
<td>0.582 (0.231-1.364)</td>
<td>0.22</td>
<td></td>
<td>0.510 (0.175-1.413)</td>
<td>0.20</td>
</tr>
<tr>
<td>High (vs Low)</td>
<td>4.100 (1.376-11.25)</td>
<td>0.013</td>
<td></td>
<td>4.468 (1.265-15.49)</td>
<td>0.021</td>
</tr>
</tbody>
</table>

*CI denotes confidence interval.
Table 3.
Enriched GO categories among the transcripts significantly differentially expressed between high- and low-risk groups.

A) Top-ranked 20 of 534 significantly overrepresented categories ($Q$ value < 0.25) among 1,109 transcripts in Japanese dataset A.

<table>
<thead>
<tr>
<th>GO category*</th>
<th>Genes within GO category</th>
<th>-log10$Q^\dagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td>immune system process (GO:0002376)</td>
<td>206</td>
<td>29.0</td>
</tr>
<tr>
<td>immune response (GO:0006955)</td>
<td>172</td>
<td>24.2</td>
</tr>
<tr>
<td>defense response (GO:0006952, 0022117, 0042829)</td>
<td>110</td>
<td>15.5</td>
</tr>
<tr>
<td>response to stimulus (GO:0050896, 0051869)</td>
<td>251</td>
<td>35.4</td>
</tr>
<tr>
<td>inflammatory response (GO:0006954)</td>
<td>68</td>
<td>9.6</td>
</tr>
<tr>
<td>positive regulation of immune system process (GO:0002684)</td>
<td>40</td>
<td>5.6</td>
</tr>
<tr>
<td>regulation of immune system process (GO:0002682)</td>
<td>52</td>
<td>7.3</td>
</tr>
<tr>
<td>antigen processing and presentation (GO:0019882, 0030333)</td>
<td>31</td>
<td>4.4</td>
</tr>
<tr>
<td>signal transducer activity (GO:0004871, 0005062, 0009369, 0009370)</td>
<td>169</td>
<td>23.8</td>
</tr>
<tr>
<td>molecular transducer activity (GO:0060089)</td>
<td>169</td>
<td>23.8</td>
</tr>
<tr>
<td>signal transduction (GO:0007165)</td>
<td>224</td>
<td>31.5</td>
</tr>
<tr>
<td>leukocyte activation (GO:0045321)</td>
<td>47</td>
<td>6.6</td>
</tr>
<tr>
<td>cell activation (GO:0001775)</td>
<td>48</td>
<td>6.8</td>
</tr>
<tr>
<td>regulation of immune response (GO:0050776)</td>
<td>32</td>
<td>4.5</td>
</tr>
<tr>
<td>response to wounding (GO:0009611, 0002245)</td>
<td>70</td>
<td>9.9</td>
</tr>
<tr>
<td>signal transmission (GO:0023060)</td>
<td>224</td>
<td>31.5</td>
</tr>
<tr>
<td>signaling process (GO:0023046)</td>
<td>224</td>
<td>31.5</td>
</tr>
<tr>
<td>regulation of cell activation (GO:0050865)</td>
<td>30</td>
<td>4.2</td>
</tr>
<tr>
<td>positive regulation of cell activation (GO:0050867)</td>
<td>24</td>
<td>3.4</td>
</tr>
<tr>
<td>regulation of lymphocyte activation (GO:0051249)</td>
<td>29</td>
<td>4.1</td>
</tr>
</tbody>
</table>

* Bold font denotes common categories included in top 20 lists both Japanese dataset A and TCGA dataset.
† $Q$ value was determined by Fisher’s exact test with Benjamini-Yekutieli correction.
B) Top-ranked 20 of 83 significantly overrepresented GO categories ($Q$ value < 0.25) among 1,381 transcripts in TCGA dataset.

<table>
<thead>
<tr>
<th>GO category*</th>
<th>Genes within GO category</th>
<th>-log10$Q$†</th>
</tr>
</thead>
<tbody>
<tr>
<td>immune system process (GO:0002376)</td>
<td>129</td>
<td>20.5</td>
</tr>
<tr>
<td>immune response (GO:0006955)</td>
<td>115</td>
<td>18.3</td>
</tr>
<tr>
<td>defense response (GO:0006952, 0002217, 0042829)</td>
<td>78</td>
<td>12.4</td>
</tr>
<tr>
<td>antigen processing and presentation (GO:0019882, 0030333)</td>
<td>25</td>
<td>4.0</td>
</tr>
<tr>
<td>inflammatory response (GO:0006954)</td>
<td>50</td>
<td>7.9</td>
</tr>
<tr>
<td>antigen processing and presentation of peptide antigen (GO:0048002)</td>
<td>13</td>
<td>2.1</td>
</tr>
<tr>
<td>antigen processing and presentation of exogenous peptide antigen (GO:0002478)</td>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>MHC class I peptide loading complex (GO:0042824)</td>
<td>9</td>
<td>1.4</td>
</tr>
<tr>
<td>MHC protein complex (GO:0042611)</td>
<td>18</td>
<td>2.9</td>
</tr>
<tr>
<td>TAP complex (GO:00042825)</td>
<td>8</td>
<td>1.3</td>
</tr>
<tr>
<td>MHC protein binding (GO:0042287)</td>
<td>12</td>
<td>1.9</td>
</tr>
<tr>
<td>antigen processing and presentation of exogenous antigen (GO:0019884)</td>
<td>7</td>
<td>1.1</td>
</tr>
<tr>
<td>response to stimulus (GO:0050896, 0051869)</td>
<td>183</td>
<td>29.1</td>
</tr>
<tr>
<td>antigen processing and presentation of peptide antigen via MHC class I (GO:0002474)</td>
<td>7</td>
<td>1.1</td>
</tr>
<tr>
<td>antigen processing and presentation of peptide or polysaccharide antigen via MHC class II (GO:0002504)</td>
<td>12</td>
<td>1.9</td>
</tr>
<tr>
<td>regulation of immune response (GO:0050776)</td>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>MHC class I protein binding (GO:0042288)</td>
<td>10</td>
<td>1.6</td>
</tr>
<tr>
<td>response to wounding (GO:0009611, 0002245)</td>
<td>50</td>
<td>7.9</td>
</tr>
<tr>
<td>positive regulation of immune response (GO:0050778)</td>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>positive regulation of immune system process (GO:0002684)</td>
<td>6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* Bold font denotes common categories included in top 20 lists both Japanese dataset A and TCGA dataset.
† $Q$ value was determined by Fisher’s exact test with Benjamini-Yekutieli correction.
High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by down-regulation of antigen presentation pathway

Kosuke Yoshihara, Tatsuhiko Tsunoda, Daichi Shigemizu, et al.

Clin Cancer Res  Published OnlineFirst January 12, 2012.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-11-2725

Supplementary Material  Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2012/01/12/1078-0432.CCR-11-2725.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.