Analysis of Gene Expression in Early-Stage Ovarian Cancer

Sergio Marchini,[1] Pietro Mariani,[1] Giovanna Chiorino,[2] Eleonora Marrazzo,[1] Riccardo Bonomi,[1] Robert Fruscio,[3] Luca Clivio,[1] Annalisa Garbi,[2] Valter Torri,[1] Michela Cinquini,[1] Tiziana Dell'Anna,[3] Giovanni Apolone,[1] Massimo Broggini,[1] and Maurizio D'Incalci[1]

Abstract

Purpose: Gene expression profile was analyzed in 68 stage I and 15 borderline ovarian cancers to determine if different clinical features of stage I ovarian cancer such as histotype, grade, and survival are related to differential gene expression.

Experimental Design: Tumors were obtained directly at surgery and immediately frozen in liquid nitrogen until analysis. Glass arrays containing 16,000 genes were used in a dual-color assay labeling protocol.

Results: Unsupervised analysis identified eight major patient partitions, one of which was statistically associated to overall survival, grading, and histotype and another with grading and histotype. Supervised analysis allowed detection of gene profiles clearly associated to histotype or to degree of differentiation. No difference was found between borderline and grade 1 tumors. As to recurrence, a subset of genes able to differentiate relapsers from nonrelapsers was identified. Among these, cyclin E and minichromosome maintenance protein 5 were found particularly relevant, as their expression was inversely correlated to progression-free survival ($P = 0.00033$ and 0.017, respectively).

Conclusions: Specific molecular signatures define different histotypes and prognosis of stage I ovarian cancer. Mucinous and clear cells histotypes can be distinguished from the others regardless of tumor grade. Cyclin E and minichromosome maintenance protein 5, whose expression was found previously to be related to a bad prognosis of advanced ovarian cancer, appear to be potential prognostic markers in stage I ovarian cancer too, independent of other pathologic and clinical variables.

In spite of much recent preclinical and clinical efforts to identify novel markers and investigate new therapies, epithelial ovarian cancer (EOC) remains the fourth most common cause of cancer-related death in women in the western world, with ~70% of patients dying within 5 years of diagnosis (1–4). This high death rate results from late diagnosis, when the disease has spread beyond the pelvis (stage III and IV). In contrast, when EOC is diagnosed at stage I, when it is still organ confined, the 5-year survival rate exceeds 90%. Unfortunately, only ~10% of all ovarian cancers are diagnosed at this early stage (5–7). Fewer than 20% to 25% of stage I patients relapse within 5 years from primary surgical-medical therapy. A similar percentage of stage I patients present with a more aggressive disease, which is difficult to predict based on currently known clinicopathologic features. These facts highlight the need for better tools for screening and staging ovarian cancer.

The exploitation of the knowledge of EOC genetic changes to benefit early disease detection or specific therapy has been limited (8, 9). Predictions based on changes in single genes or sets of genes have yet to be used in the clinic, and most of the genetic insights have been gained in small numbers of patients that do not allow firm conclusions (10, 11). The situation is complicated by the heterogeneity among ovarian cancer subtypes. Based on morphologic criteria, EOC is classified into four major subtypes: serous, mucinous, endometrioid, and clear cell. Each class is further divided into benign, malignant, and borderline with different degrees of differentiation (2). Recent studies have shown that molecular defects appear to differ among the most common types of EOC, supporting the notion that they represent histopathologically, genetically, and biologically distinct, albeit related, diseases (8, 12). The histologic grade of the tumor is probably the variable that is most significantly associated with the prognosis of stage I ovarian carcinoma, but its predictive value is far from being satisfactory (13).

Many groups have applied cDNA microarray technology to the study of advanced EOC or cell lines derived from EOC (14–18), but very little is known about stage I and borderline carcinomas, which, although relatively rare, are potentially more

Authors' Affiliations: [1]Department of Oncology, Istituto di Ricerche Farmacologiche "Mario Negri," Milan, Italy; [2]Cancer Genomic Laboratory, "Edo Tempia" Foundation, Biella, Italy; and [3]San Gerardo Hospital, University of Milan, Monza, Italy

Received 2/27/08; revised 6/4/08; accepted 7/28/08.

Grant support: Fondazione Cariplo, Compagnia di San Paolo, and FIRB-MIUR grant RBINO48RHA.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

S. Marchini, P. Mariani, and G. Chiorino contributed equally to the work.

Requests for reprints: Sergio Marchini, Laboratory of Molecular Pharmacology, Department of Oncology, via La Masa N°19, 20156 Milan, Italy. Phone: 39-2-39014236; Fax: 39-2-3546277; E-mail: marchini@marionegri.it.

doi:10.1158/1078-0432.CCR-08-0523

Translational Relevance

The gene expression profile of 16,000 genes was analyzed in 68 stage I EOC and 15 borderline tumors selected from a single tumor tissue bank collection of frozen biopsies. Gene profiling was associated to grading, histotype, and survival. Borderline tumors revealed a molecular signature overlapping grade 1 but not grade 2 or 3. We defined a subset of genes involved in cell cycle regulation and chromosome maintenance whose expression discriminated between relapsers from nonrelapsers. Of these, cyclin E and MCM5 were found up-regulated in relapsing patients before chemotherapy and their expression levels correlated with overall survival and progression-free survival. This study suggests the potential use of cyclin E and MCM5 as predictive biomarkers for relapse in stage I EOC.

useful sources of information concerning early events associated with tumor pathogenesis. This lack of information prompted us to test the hypothesis that cDNA microarray analysis can detect properties of stage I and borderline tumors that might not be discovered by the commonly used clinical or histopathologic analyses at diagnosis. Using a single center tumor bank of >1,200 ovarian carcinoma biopsies frozen immediately after surgical removal, 68 cases of stage I and 15 borderline tumors were selected to assess the expression of 16,000 genes. The aim of the present study was to determine if different clinical features of stage I ovarian cancer such as histotype, grade, and survival are related to differential gene expression.

Materials and Methods

Tissue processing and tumor selection

A total of 83 patients were recruited among the 1,200 samples stored into the frozen tissue bank between September 1992 and March 2003. Sixty-eight were diagnosed as having stage I ovarian cancer and 15 as borderline. The histology, grade, and stage of each tumor are listed in Table 1. Tumors were staged according to the International Federation of Gynecology and Obstetrics criteria; 63% of the stage I cases (43 of 68) were treated after surgery with platinum-based therapy (carboplatin, cisplatin, or CDBCA + paclitaxel), whereas the remaining 37% (25 of 68) did not receive adjuvant therapy. One of 15 borderline tumors (serous, substage c) received carboplatin therapy, whereas the remaining were untreated (Table 1). Ovarian tumor samples were collected in the operating theater and frozen within 15 min in liquid nitrogen after surgery and stored at -80°C. The tumor content of the specimens was assessed by H&E stain at San Gerardo anatomopathologic unit. Only specimens containing >70% of tumor were used for these experiments. The collection and use of tumor samples was approved by the local scientific ethical committee and written consent was obtained from the patients.

RNA isolation

Frozen primary tumor tissues (30 mg) were homogenized in a Ultra-turrax at 4°C and total RNA was purified using SVTotal RNA isolation system according to the manufacturer's instructions (Promega). RNA was measured by spectrophotometer and quality assessed by a Bioanalyzer (Agilent). Only samples with a RNA quality score (RIN) larger than 5 were further processed and aliquots were stored at -80°C until use.

Real-time reverse transcription-PCR and primer design

Real-time quantitative reverse transcription-PCR (qRT-PCR) was used to validate the differential expression of selected genes in RNA samples from primary EOC. Automatic liquid handling station (epMotion 5075 LH; Eppendorf) and real-time PCR (ABI-7900; Applied Biosystems) were used as described (19, 20). The comparative threshold cycle (C_t) method was used for the calculation of amplification fold as specified by the manufacturer. Differences among histologic types in qRT-PCR expression data were analyzed by Welch t test. Arrays and qRT-PCR data correlation were assessed by Spearman's ρ and test. Primers pair sequences were designed as described (20) and their sequences were reported in Supplementary Table S1. Actin B and cyclophilin A were used as internal control.

cDNA preparation and microarray hybridization

Gene expression was analyzed using Agilent slides onto which 16,000 cDNAs were spotted (Agilent). Human Atlas control RNA from a pool of human cell lines was used as a reference for the whole experimental protocol (BD Biosciences). The T7 amplification was done according to Message Amp aRNA kit (Ambion) and cRNA was reverse transcribed in Cy3/Cy5 cDNA (GEH) using the amino allyl cDNA labeling kit (Ambion). For each slide, Cy3-labeled samples were compared with their Cy5-labeled controls and labels were swapped to increase the robustness of our data.

Image acquisition and data processing

Arrays were scanned using a laser scanner Axon 4000-B (Axon Instruments) using 5 μm resolution and with photomultiplier variables that defined the quanta intensity density distribution of the array to give a Cy3/Cy5 ratio close to 1.0. Separate 16-bit images were acquired for Cy3 and Cy5. Two similar but not equivalent procedures were adopted for image analysis and data processing. Details of the two procedures are attached as Supplementary Data. Briefly, the first method consisted in analyzing images with Spot and processing data with R (Procedure 1), whereas the second one used GenePixPro (Axon Instruments) for image analysis and the Rosetta Resolver SE software (Rosetta Biosoftware) for data processing and analysis (Procedure 2). Raw and processed data are available in the Gene Expression Omnibus repository (GSE8841 and GSE8842).[4]

Data analysis

Available clinical and pathologic variables were tested for association to gene expression profiles in a supervised approach (*a priori* defining classes to compare) and in an unsupervised one (clustering based on expression data only; ref. 21).

Unsupervised analysis: cluster reproducibility and statistical correlation to clinical variables. Self-organizing maps analysis (as described in Supplementary Data) and hierarchical clustering (with average linkage and Euclidean distance) were applied to cluster samples. The reproducibility of clustering results was assessed using BRB-Array Tools version 3.3.0.[5] Measures of reproducibility (R) and discrepancy (D) were used to evaluate the robustness of the clusters as explained in the program manual (22).

The unsupervised analysis was focused on the detection of clusters, their correlation with known prognostic factors, and the prognostic role of each cluster in terms of overall survival (OS; primary endpoint) and progression-free survival (PFS) benefit. Correlations between each cluster and known prognostic variables (substage, grading, and histology) were described with relative and absolute frequencies and analyzed with the χ^2 test for association or trend, as appropriate. Cumulative OS and PFS curves were constructed as time-to-event plots by the Kaplan-Meier method. Comparison of the curves was done using log-rank statistics. Hazard ratio was calculated for each cluster by using

[4] http://www.ncbi.nlm.nih.gov/geo/
[5] http://linus.nci.nih.gov/BRB-ArrayTools.html

Table 1. Clinical variables and chemotherapy regimens of the 68 stage I and 15 borderlines patients enrolled in the study

Median age at diagnosis (y) = 52		Chemotherapeutic regimens			
Histopathologic variables	**No. (%) cases**	**Carboplatin**	**Cisplatin**	**Carboplatin + paclitaxel**	**Untreated**
Stage I (n = 68)					
A	17 (25)	7	1		9
B	4 (5.88)	3			1
C	47 (69.12)	30	1	1	15
Grade					
1	13 (19.12)	3	1		9
2	20 (29.41)	14			6
3	35 (51.47)	23	1	1	10
Histotype					
Serous	24 (35.29)	12	1		11
Mucinous	10 (14.71)	4			6
Endometrioid	17 (25)	11	1	1	4
Clear cell	16 (23.53)	13			3
Undifferentiated	1 (1.47)				1
Borderline (n = 15)					
a	8 (53.33)				8
b	1 (6.67)				1
c	6 (40)	1			5
Histotype					
Serous	7 (46.67)	1			6
Mucinous	7 (46.67)				7
Not determined	1 (6.66)				1

the Cox model. Results are reported as hazard ratio estimates with 95% confidence intervals and P value at Wald's test. A hazard ratio of 1 denotes the absence of a difference between the two arms (or the two categories of a compared covariate), whereas a hazard ratio of >1 or <1 denotes an increase or decrease in the risk in a given patient group with respect to the reference. The proportional hazards assumption implied by Cox's model was checked by analysis of Schoenfeld residuals. All the computations were made using SAS software.

Supervised analysis. Supervised analysis, which directly links gene expression and clinical variables, was accomplished using BRB-Array Tools, PAM (23), Significance Analysis of Microarrays (24), and the Rosetta Resolver SE software.

(a) Class comparison. With BRB-Array Tools, genes that were differentially expressed in two or more classes were identified by using a multivariate permutation test (25) to provide 95% confidence interval that the false discovery rate was ≤10%. We also applied the random variance model (26) and performed the global test of significance (with default algorithm settings) described in the program manual. One-way error-weighted ANOVA with Benjamini and Hochberg false discovery rate method for multiple test correction implemented on the Resolver System was applied to extract differentially expressed genes between two or more classes. This algorithm was applied to expression data obtained with the second analysis method. Significance Analysis of Microarrays[6] was also applied on patient Z scores (analysis method 1) and log ratios (analysis method 2) to extract differentially expressed genes. For the latter, the *k-nn* input algorithm available within R was adopted to complete missing values in the gene per patient matrix before submission to Significance Analysis of Microarrays. When there were two classes, the "two-class unpaired data" option was applied, and when there were more than two, the "multiclass" option was chosen.

(b) Correlation with survival outcomes. Using BRB-Array Tools, we searched for genes whose expression was significantly related to survival. Algorithm settings were the same as in the class comparison case. OS and PFS analysis was also done by using the "survival" option in the SAM package using the ratio experiment matrix as input and specifying each patient's survival time and status. Association between gene expression and survival was described using Kaplan-Meier curves and quantified using hazard ratio computed by the Cox model as summary measure.

(c) Class prediction. Models to predict the class of samples with unknown class, using gene expression profiles, were developed. These were based on different algorithms (27, 28) and incorporated genes differentially expressed at different significance levels as assessed by the random variance t test or F test. The level of significance (among 0.01, 0.005, 0.001, and 0.0005) is automatically selected in a fully cross-validated way for each algorithm as a tuning variable value. We estimated the prediction error of each model using two methods: cross-validation (leave-one-out and 10-fold coefficient of variance) and bootstrap (0.632 bootstrap) as described in refs. 29, 30. For each coefficient of variance (and bootstrap) training set, we repeated the entire model building process, including gene selection. We also evaluated whether the cross-validated error rate estimate for a model was significantly less than one would expect from random prediction as reported in BRB-Array Tools manual.

(d) Gene ontology and pathways analysis. Gene ontology (GO) groups of genes whose expression was differentially regulated among classes were identified using the "Gene set comparison" tool in BRB-Array Tools with default settings. We considered a GO category significantly differentially regulated if significance levels of both the Fisher and the Kolmogorov-Smirnov statistics were <0.005 (31).[7] Pathway analysis is similar to the GO analysis, except that genes are grouped by Kegg or BioCarta pathways. Functional annotation of gene lists was done using DAVID.[8]

Comparison of the gene lists and correlation. Gene lists obtained by applying supervised procedures on the expression data from analysis methods 1 and 2 were tested for correlation if their intersections contained at least two genes. The rankings of differential statistics in the

[6] http://www-stat.stanford.edu/~tibs/SAM/

[7] http://linus.nci.nih.gov/brb

[8] http://david.abcc.ncifcrf.gov/

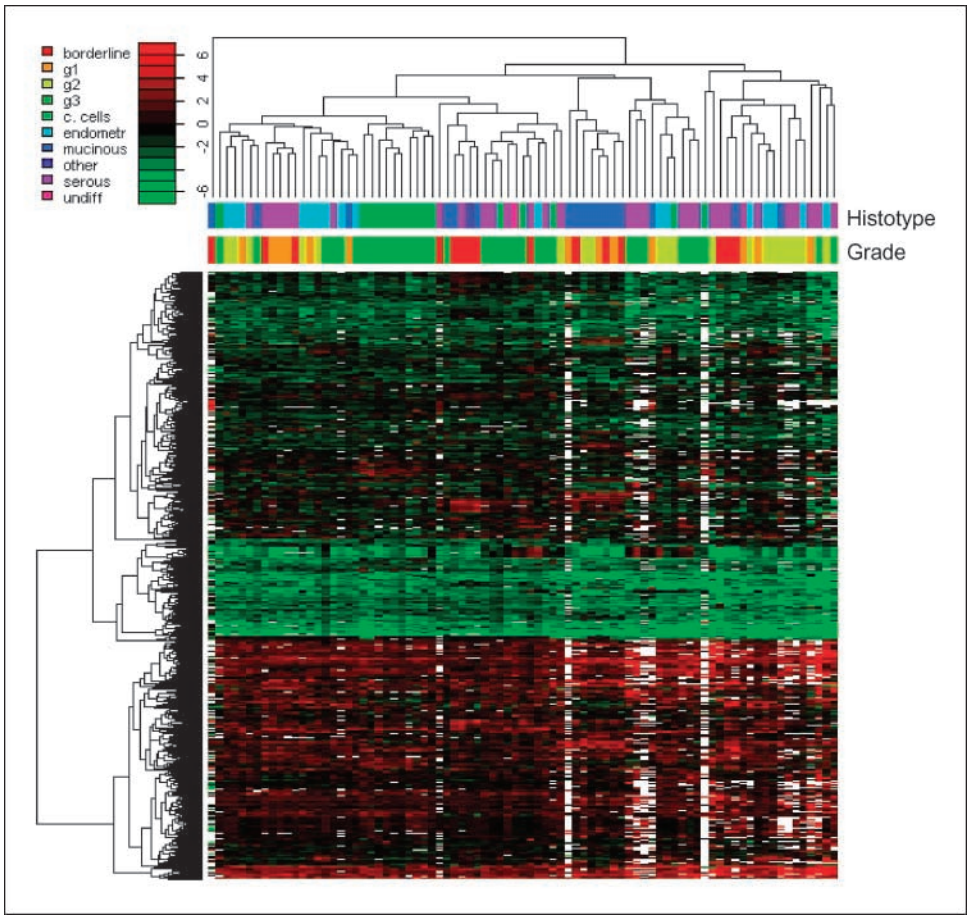

Fig. 1. Unsupervised two-dimensional hierarchical cluster analysis of 68 ovarian stage I cancer and 15 borderline patients based on variation of expression of 1783 genes obtained by feature filtering based on percentage of missing values (at most 20%) and log ratio SD (at least 1.2). Each row represents a single gene and each column represents a tumor. The two colored bars below the dendogram report the different histologic subtypes and grades, respectively, as indicated (*top left*). The color contrast of the vertical bar indicates the fold of gene expression changes in $\log_2$ scale (numbers beside the bar). *Red,* up-regulation; *green,* down-regulation; *black,* no changes; *white,* missing value.

original lists were taken as input for the algorithm. A nonparametric correlation test was done in R[9] by the cor test function in the stats package and using both Spearman's ρ and Kendall's τ to find correlations statistically different from 0 (32).

Results

Agilent human arrays with 16,000 cDNAs were employed to identify global gene expression changes in 68 stage I EOC biopsies and in 15 borderline tumors. Table 1 shows histopathologic features and chemotherapeutic regimens of the cohort of patients enrolled in this study. Cox proportional hazards model shows that our series appears to be representative of stage I ovarian cancer, as PFS and OS are similar to those reported in literature according to substage, grading, and histotype (Supplementary Table S2A and S2B; ref. 2). Patients enrolled in the study described here did not harbor any mutation in the p53 gene (20).

The use of two different preprocessing procedures, which gave Z scores and log ratios, respectively, did not markedly affect unsupervised or supervised analysis outputs (see Supplementary Data). Reported here are results that showed high consistency within the two procedures.

Molecular profiling of ovarian cancer subtypes. Hierarchical cluster analysis on a subset of 1,783 cDNAs obtained by feature filtering based on percentage of missing values and SD (Fig. 1) shows that the 83 samples are overall molecularly homogeneous. The selected genes can be roughly subdivided into three distinct subsets: slightly differentially expressed compared with the universal reference (light green and light red spots) at the top of the figure, strongly down-regulated (green spots) in the middle, and strongly up-regulated at the bottom. Notably, the clear cell and mucinous histotypes are readily distinguished from the serous, endometrioid, and undifferentiated ones, which appear to be intermingled among themselves regardless of tumor grade. In particular, 10 of 16 clear cells, grade 3, show an overlapping signature. Borderline samples are scattered around and cocluster with stage I tumors regardless of their grade.

To unravel the complexity of the molecular fingerprint of the 68 stage I tumors, self-organizing maps analysis was applied to a gene expression matrix of 1,116 rows (similar filtering criteria as above), which yielded 34 patient partitions showing statistically significant associations to clinical variables (P = 0.001-0.0001; see Supplementary Table S3).

We focused our attention on two partitions characterized by genes up-regulated in the clear cell grade 3 subgroup, named partition 1 (Fig. 2A), or in the mucinous subtypes, named partition 2 (Fig. 2B). To support the robustness of our observations, we validated partition 1 signature by qRT-PCR in an independent set of total RNA purified from the same cohort of patients. Table 2 shows a representative panel of observed differences in 6 genes between the clear cells and the

[9] http://www.R-project.org

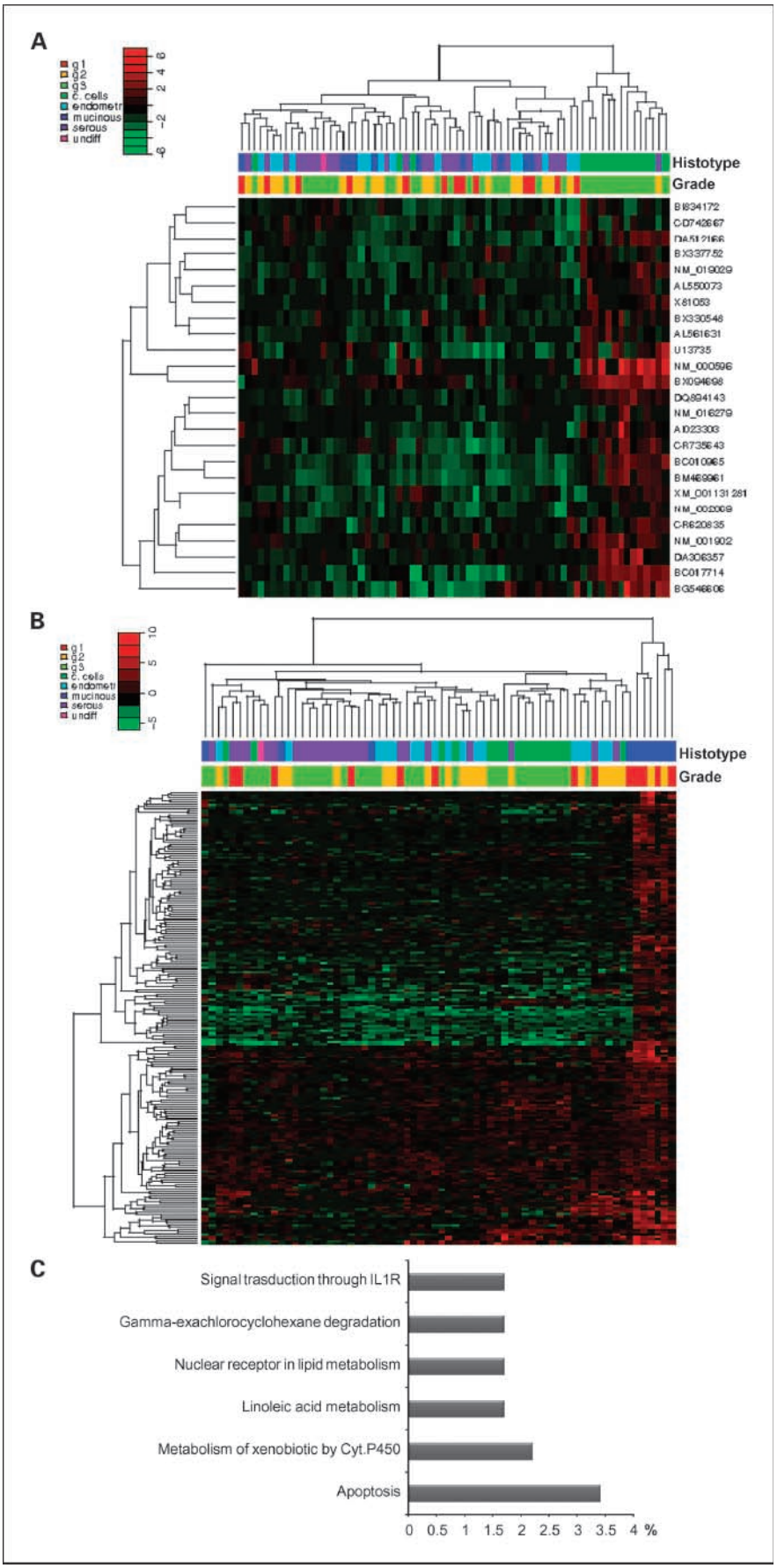

Fig. 2. Dendogram obtained by hierarchical clustering of the 68 stage I samples (columns) and of the genes (rows) defining partition 1 (*A*) and partition 2 (*B*) signatures. The two colored bars below the dendogram report the different histologic subtypes and grades, respectively, as indicated (*top left*). The color contrast of the vertical bar indicates the fold of gene expression changes in $\log_2$ scale (numbers beside the bar). *Red,* up-regulation; *green,* down-regulation; *black,* no changes. The GenBank accession no. of the 25 genes associated to partition 1 is reported. *C,* functional annotation of the 202 genes of partition 2 according to pathways obtained using *H. sapiens* as background. Reported are the categories with significant DAVID score ($P < 0.05$). Percentages refer to the number of genes with that function within the list with respect to the total number of genes of the list.

Table 2. Partition 1 data expression validation by qRT-PCR

Gene name (accession no.)	Arrays (Cl-nCl)	qRT-PCR (Cl-nCl)	ρ
ADAMTS9 (BX094698)	2.6882026 (P = 1.018e^{-7})	2.0976887 (P = 0.00009077)	0.7248384 ($P \leq$ 2.2e^{-16})
CDH9 (NM_016279)	3.7667154 (P = 0.05345)	3.537528 (P = 0.009844)	0.8809524 (P = 0.007242)
CTH (NM_001902)	1.7961264 (P = 0.00009343)	1.5353106 (P = 0.002965)	0.6839744 ($P \leq$ 2.2e^{-16})
GRB14 (XM_001131281)	1.6008336 (P = 0.00011)	1.557095 (P = 0.00335)	0.7481759 ($P \leq$ 2.2e^{-16})
IGFBP1 (NM_000596)	3.8279291 (P = 0.000003487)	1.5353106 (P = 1.587e^{-7})	0.868902 (P = 4.256e^{-14})
NDRG1 (AL550073)	0.88948457 (P = 0.005667)	1.4797459 (P = 0.0001684)	0.4770345 (P = 0.000312)

NOTE: For 6 analyzed genes, mean group differences between clear cells (Cl) and non-clear cells (nCl) samples using microarray technology or qRT-PCR are reported. From left to right, the columns refer to gene name with GenBank accession no.; arrays log ratio values; qRT-PCR results calculated with the -$\Delta\Delta C_t$ protocol and the correlation value (ρ) between the two assays using the Spearman's test. P is the P value referred to Welch t test or correlation test as appropriate ($P < 0.05$).

other histotypes analyzed using microarray or qRT-PCR. With a P value threshold of 0.05 and a ρ value ranging from 0.4 to 0.8, in 75% of the analyzed samples, the results obtained by qRT-PCR mirror those gained by microarray in both qualitative and quantitative terms. This suggests that most array probe sets are likely to accurately measure single transcript levels within a complex mixture of transcripts.

Partition 2 is characterized by a subset of 202 genes, the expression of which was found consistently modified in 6 mucinous tumors. To ascertain whether particular functional categories could be overrepresented, GO enrichment analysis was applied to the differentially regulated probe sets identified for mucinous histotypes. This provided a list of GO categories having more genes differentially expressed between mucinous and nonmucinous histotypes than expected by chance. DAVID functional annotation, according to GO molecular functions level 4, yielded an enrichment of 6 functions ($P < 0.05$; Fig. 2C). Attention was focused on genes involved in "metabolism of xenobiotics by cytochrome $P450$" and "signal transduction through the interleukin-1 receptor pathway." In 70% of the selected genes, differences in gene expression between mucinous and nonmucinous subtypes were confirmed by qRT-PCR in the same cohort of patients (Table 3). The mucinous histotype was characterized by an increase in the expression of different families of genes involved in the metabolisms of xenobiotics as the cytochrome $P450$ family (CYP2S1 and CYP4F12), glutathione metabolism isozymes (MGST2), nuclear orphan receptor (NR1I2), and NADH dehydrogenase quinone 1 (NQO1). These differences were also confirmed in an independent cohort of stage I patients ("test" set) selected from our tumor bank. Of these, only 9 of 21 tumors were diagnosed as mucinous (Supplementary Table S4). Data obtained in this test set (Table 3, *right column*) confirmed the trend in mean differences obtained previously in the original cohort of patients, although the P value did not reach significance. Nevertheless, the tumor necrosis factor gene, which belongs to the interleukin-1 receptor signal transduction pathway, was confirmed to be strongly down-regulated in the mucinous histotype compared with the nonmucinous one in both data sets.

Relationship between gene expression patterns and histopathology. Supervised analysis was done to define molecular features associated to histopathologic variables. Analyses were carried out on the complete set of cDNAs without any kind of feature filtering. This allowed detection of cDNAs with small variance among samples but homogeneous behavior within predefined sample categories. Figure 3 shows a multidimensional scaling (MDS) three-dimensional scatter plot for 67 stage I samples, all but one with undifferentiated histotype, built on 692 genes obtained by one-way ANOVA on histotypes. In particular, it can be noted that serous (red) and endometrioid samples (blue) have a much lower distance between each other than clear cells (cyan) or mucinous samples (green).

Further supervised analysis was done by class comparison between grade 1 and grade 3 tumors. Hierarchical clustering applied to 128 genes shared between results from two different class comparison algorithms is shown in Fig. 4A. Grade 1 (red) are separated from grade 3 (cyan) stage I tumors. Only 1 of 13

Table 3. Partition 2 gene expression validation by qRT-PCR in the original and in the "test" set

Gene name (accession no.)	Arrays (Muc-nMuc)	qRT-PCR original data set (Muc-nMuc)	qRT-PCR test set (Muc-nMuc)	ρ
CYP2S1 (NM_030622)	2.93354 (P = 0.006038)	3.102301879 (P = 0.006447)	1.480718333 (P = 0.2131)	0.7029887 (P = 2.68e^{-10})
CYP4F12 (AK091995)	2.2993218 (P = 0.04075)	3.687911810 (P = 0.01114)	2.046051111 (P = 0.08361)	0.5599081 (P = 0.0003233)
MGST2 (NM_002413)	0.7332621 (P = 0.01469)	1.002780741 (P = 0.01973)	0.161748333 (P = 0.7292)	0.4793297 (P = 4.41e^{-05})
NQO1 (BU167840)	1.692737 (P = 0.02557)	1.373538534 (P = 0.0989)	0.585560278 (P = 0.4972)	0.6731583 ($P \leq$ 2.2e^{-16})
NR1I2 (AK122990)	3.5072617 (P = 0.01273)	3.690360466 (P = 0.01349)	1.25111399 (P = 0.3434)	0.7692308 (P = 0.003084)
TNF (NM_000594)	-1.178148 (P = 0.03833)	-2.200605362 (P = 0.0008089)	-1.463160278 (P = 0.04814)	0.5295689 (P = 6.78e^{-06})

NOTE: For 6 analyzed genes, mean group differences between mucinous (Muc) and nonmucinous (nMuc) samples using microarray or qRT-PCR are reported. From left to right, the columns refer to gene name with GenBank accession no.; the arrays log ratio value; qRT-PCR results calculated with the -$\Delta\Delta C_t$ protocol in the original cohort and in the new "test" set; and the correlation value (ρ) between the two assays using the Spearman's test. P is the P value referred to Welch t test or correlation test as appropriate ($P < 0.05$).

grade 1 samples coclusters with grade 3 tumors. Functional annotation done on these 128 genes revealed that 6 of 10 biological processes with significant enrichment score are related to cell cycle regulation, supporting the notion that changes in expression profiles of genes involved in cellular proliferation account for different histologic grade within stage I (Fig. 4B).

To refine and improve the quality of the comparison among histologic grades, grade 2 tumors were included into the analysis and the borderlines were compared with the other grades. The MDS three-dimensional scatter plot built on 139 genes obtained from one-way ANOVA on grades 1 to 3 (Fig. 4C) shows that grade 2 tumors (blue) are spread between grade 1 (green) and grade 3 (red), which are distributed at the two edges of the plot. Interestingly, the 15 borderlines (cyan) overlap grade 1 samples and are separated from the more undifferentiated tumors. Statistical analysis confirmed the MDS scatter plot results.

Gene expression pattern in relapsed patients. Class comparison was applied to the 68 stage I patients to discriminate the recurrence status of the samples. One hundred ninety-one genes distinguish relapsed samples from those who did not relapse. The MDS scatter plot built on these genes is shown in Fig. 5A, whereas Fig. 5B reports the GO biological processes with significant DAVID score ($P < 0.05$) obtained using *Homo sapiens* as background. The complete set of genes subdivided for the seven GO biological processes is reported in Supplementary Table 5. Attention was focused on genes involved in cell cycle control and chromosome organization, particularly on CCNE1 (cyclin E) and minichromosome maintenance protein 5 (MCM5), a member of the minichromosome maintenance complex. The $\log_2$ ratio and $-\Delta\Delta C_t$ values for the group mean differences were 1.12 and 1.16, respectively, for cyclin E and 0.76 and 0.44 for MCM5. Both qRT-PCR and array data confirmed that CCNE1 and MCM5 were significantly ($P < 0.05$) up-regulated in relapsing patients compared with nonrelapsing ones (Table 4). Absence of relapsing patients in our 21 test sample set confounded confirmation of this observation in an independent cohort of patients.

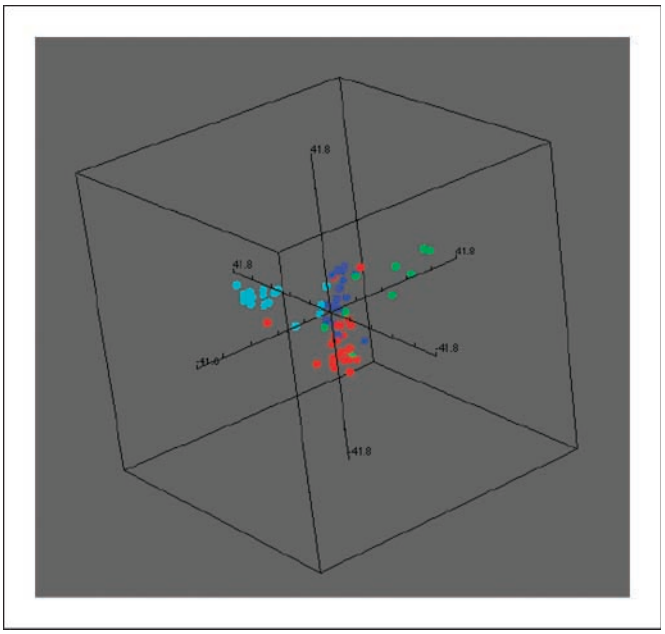

Fig. 3. MDS three-dimensional scatter plot of 67 samples on 692 genes distinguishing the four principal histotypes as obtained by one-way ANOVA. *Cyan,* clear cells; *blue,* endometrioid; *green,* mucinous; *red,* serous.

Survival analysis. Survival analysis was done in 68 patients. Recurrence occurred in 18 patients, 13 (19%) of which died for progression. We correlated the expression levels of CCNE1 and MCM5 with OS and PFS, plotted with the Kaplan-Meier method (Fig. 6). Even if the survival curves for CCNE1 and MCM5 did not show a significant difference of OS ($P = 0.203$ and 0.295, respectively), patients with expression levels of CCNE1 higher than 0.124 had a worst prognosis than those with levels lower than 0.124. The power test ($1 - \beta = 80\%$) suggested that a cohort of at least 2,675 patients should be analyzed to reach significance of 95%. However, when expression levels of CCNE1 and MCM5 were correlated with PFS, patients with expression levels higher than 0.34 and 0.58, respectively, had a shorter PFS than patients with lower levels, and such differences reached statistically significance ($P = 0.0003$ and 0.018 for CCNE1 and MCM5, respectively). Patients were stratified according to chemotherapy and results reported in Supplementary Fig. S1 confirm similar differences in both chemotherapy-treated and untreated group. To assess the effect of each variable on PFS and OS, a Cox's model was done using univariate analysis. Variables of interest were gene expression (CCNE1 and MCM5), grading, histotype, and chemotherapy. In addition, a further Cox's model was done according to the level of expression of CCNE1 and MCM5 gene, adjusted for chemotherapy. Grading was found associated with both OS and PFS ($P = 0.0098$ and 0.046 for PFS and OS, respectively), whereas gene expression of CCNE1 was found significantly associated with PFS only ($P = 0.0069$). Chemotherapy alone did not affect OS and PFS in the whole sample ($P = 0.09$ and 0.066 for OS and PFS, respectively) or in stratified patients according to CCNE1 or MCM5 levels ($P = 0.059$ and 0.08 for PFS and $P = 0.52$ and 0.56 for OS, respectively; Supplementary Table S6, sections a and b). A multivariate Cox's regression model for PFS was built using variables that were found significant on univariate analysis. In this case, both CCNE1 and grading were not statistically significant (Supplementary Table S6, section c).

Discussion

The data presented above represent to our knowledge the first gene profiling study focused on stage I ovarian cancer. Our results show that gene expression profiles are related to both tumor pathologic features and patients' survival. As to tumor morphology, it should be noted that the histologic origin of the different ovarian cancer subtypes is still a matter of debate. According to the most recent studies, it appears that clear cell, endometrioid, and mucinous EOCs arise probably from the metaplastic Mullerian epithelium rather than directly from ovarian surface celomatic sheets, which is more likely in the case of serous carcinomas (12). The strong correlation of stage IC and clear cell histotype with inferior prognosis in our series is consistent with previous reports indicating that the gene profiling approach adopted here was done on a series adequately representing stage I ovarian cancer (2). Morphologic studies suggest there are at least two distinct ovarian cancer subtypes, defined as weakly and highly malignant with

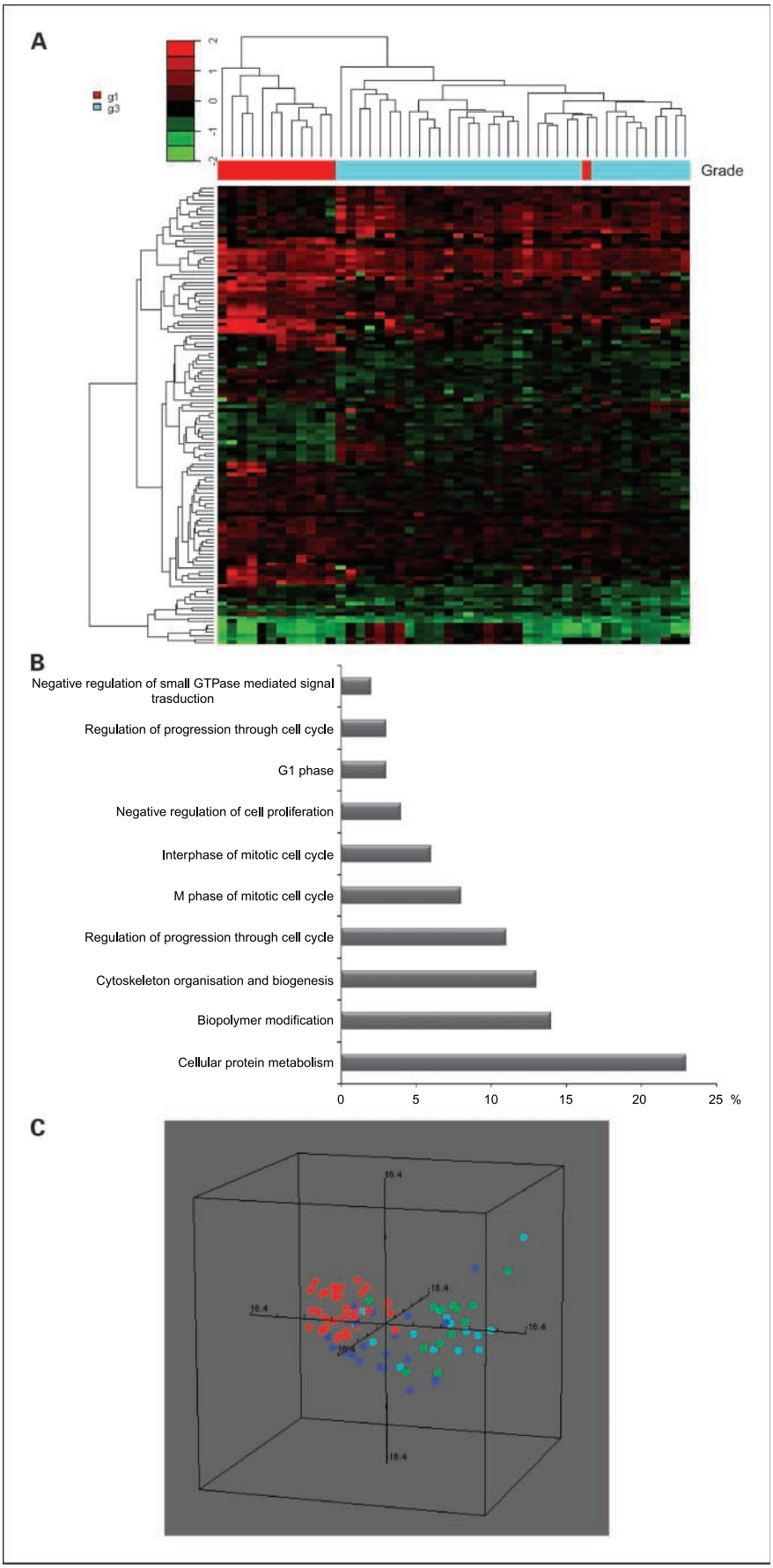

Fig. 4. *A,* hierarchical clustering applied on grade 1 and 3 samples using 128 genes shared between results from two class comparison algorithms (Significance Analysis of Microarrays and BRB ArrayTools). Each row represents a single gene and each column represents a tumor. The colored bar below the dendogram reports grades: grade 1 (*red square*) and grade 3 (*cyan square*). The color contrast of the vertical bar indicates the fold of gene expression changes in $\log_2$ scale (numbers beside the bar). *Red,* up-regulation; *green,* down-regulation; *black,* no changes. *B,* functional annotation of the 128 genes according to GO level 5 biological processes obtained using *H. sapiens* as background. Reported are the categories with significant DAVID score ($P < 0.05$). Percentages refer to the number of genes with that function within the list with respect to the total number of genes of the list. *C,* MDS three-dimensional scatter plots of 83 samples (15 borderline, 13 grade 1, 20 grade 2, and 35 grade 3) built on 139 genes obtained from one-way ANOVA on grades 1 to 3. *Cyan,* borderline; *green,* grade 1; *blue,* grade 2; *red,* grade 3.

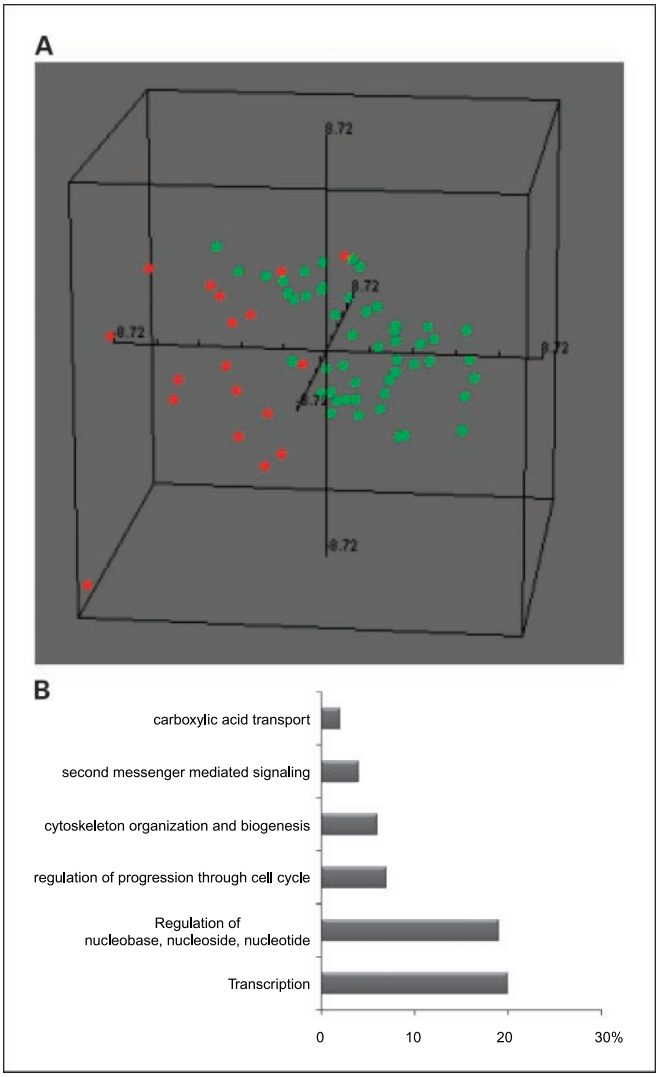

Fig. 5. *A,* MDS three-dimensional scatter plot of 68 stage I samples built on 191 genes distinguishing relapsed from nonrelapsed samples. *Green,* not relapsed; *red,* relapsed. *B,* functional annotation of the 191 genes according to GO level 5 biological processes obtained using *H. sapiens* as background. Reported are the categories with significant DAVID score ($P < 0.05$). Percentages refer to the number of genes with that function within the list with respect to the total number of genes of the list.

different pathogenesis. Consistent with this definition, the study presented here indicates that changes between the different histologic subtypes are associated with specific gene expression patterns regardless of the substage and grade of the disease.

Unsupervised and supervised analysis of these homogeneous samples identified a sizeable number of genes associated with grading and histotype. Regardless of the tumor grade, mucinous and clear cells are largely separable and distinct from the other histotypes. Partition 1, which strongly identifies patients with a poor prognosis, is also associated with the clear cell histotype. Partition 2 strongly associates with the mucinous type. The nature and function of the specific genes associated with these clusters needs thorough analysis to identify those most likely to predict whether a patient belongs to one group or another. The identification of several clear cell or mucinous specific markers constitutes the basis for future studies in which some of these biomarkers can be tested for their suitability to serve in diagnosis. For example, the overexpression of genes coding for insulin-like growth factor binding proteins and its intracellular adaptor protein GRB14 in clear cell tumors may explain the poor prognosis associated with this histotype compared with other types of EOC. The insulin-like growth factor binding protein family of proteins is often associated with increased tumor aggressiveness (33, 34) and has been tentatively linked to resistance against cisplatin treatment (35).

Previous studies done in advanced ovarian cancer have shown that the gene expression profile in mucinous histotype cancer differs from that in other histotypes (36). The present study confirms this finding and shows that clear-cut differences between histotypes other than the mucinous are based on the expression of oxidative stress response genes (CYP2S1, CYP4F12, MGST2, NQO1, and NR1I2). The fact that the tumor necrosis factor gene was down-regulated in the mucinous histotype may be clinically significant. The contribution of tumor necrosis factor to human ovarian cancer progression has long been recognized (37, 38). Models predict that constitutive tumor necrosis factor-α production by tumor cells may generate and sustain a tumor-promoting cytokine network in the ovarian cancer microenvironment (39). The present study suggests that this cytokine is less expressed in mucinous tumors, a finding possibly relevant for the selection of patients undergoing therapeutic approaches against this target.

As observed previously for late-stage EOC (40) also within stage I, grade 1 tumors profile was similar to that of borderline tumors, both being different from grade 2 or 3 tumors. This is a novel observation that might justify similar clinical management of borderline and grade 1 stage I ovarian cancer perhaps obviating postsurgical chemotherapy.

As far as the relationship between gene profile and prognosis, class comparison assessing differential expression between relapsers and nonrelapsers in conjunction with GO analysis showed a different expression of genes involved in cell cycle

Table 4. Data expression validation by qRT-PCR of two genes found differently expressed between stage I relapsing (R) or nonrelapsing (nR) tumor samples

Gene name (accession no.)	Arrays (R-nR)	qRT-PCR (R-nR)	ρ
CCNE1 (M74093)	1.1216554 ($P = 0.0152$)	1.16741623 ($P = 0.02158$)	0.789648 ($P < 2.2e^{-16}$)
MCM5 (NM_006739)	0.7649194 ($P = 0.002128$)	0.44442057 ($P = 0.03524$)	0.4496941 ($P = 0.0001996$)

NOTE: From left to right, the columns refer to gene name with GenBank accession no.; arrays log ratio values; qRT-PCR results calculated with the $-\Delta\Delta C_t$ protocol; and the correlation value (ρ) between the two assays using the Spearman's test. *P* is the *P* value referred to Welch *t* test or correlation test as appropriate ($P < 0.05$).

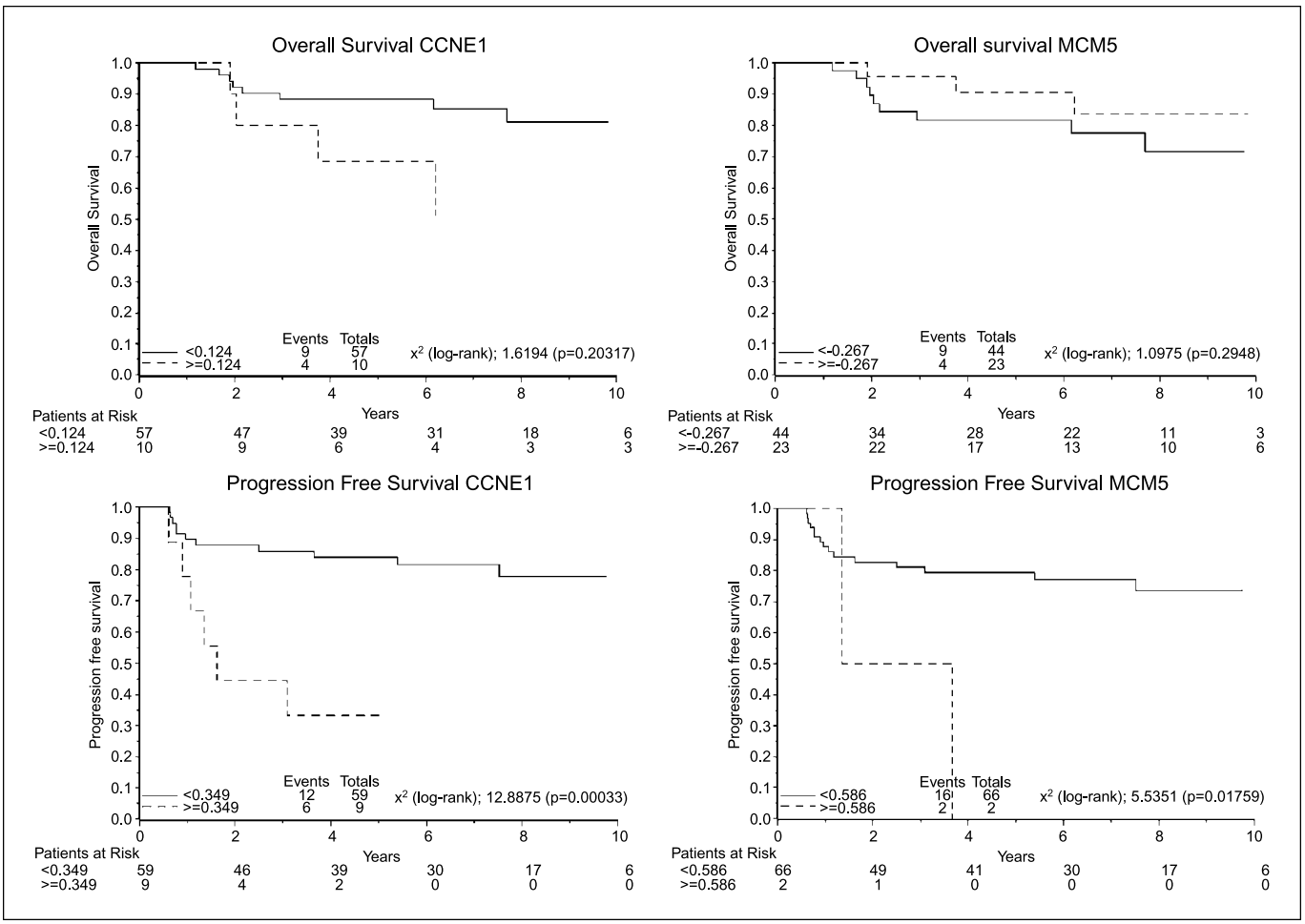

Fig. 6. Kaplan-Meier survival curves in relation to OS and PFS for CCNE1 and MCM5 expression in EOC stage I patients. Patients were divided according to the expression levels of the two genes calculated as reported in Materials and Methods.

regulation (those associated with cytoskeleton organization and biogenesis, regulation, and progression through the cell cycle and regulation of nucleotide metabolism). One of the most interesting finding of our analysis is the inverse relationship between the expression of cyclin E and MCM5 on one side and patients' disease-free survival on the other. These differences were found in both patients treated with chemotherapy and untreated patients, supporting the idea that, in our cohort of patients, the prognostic value of MCM5 and CCNE1 is not affected by treatment.

The importance of cyclin E as a predictor of tumor progression and poor prognosis of ovarian cancer patients has been reported previously (41–43). Cyclin E levels were assessed by immunohistochemistry in archived paraffin-embedded tissue blocks from a relative large series of patients, the majority of whom were at an advanced stage. The present study confirms this observation in stage I ovarian cancer and corroborates the idea that cyclin E could be an important prognostic marker in the management of ovarian cancer (41, 44). Less information is available in the literature on MCM5, a protein recently found to be a good proliferation marker related to the prognosis in different tumors (45–48). Very recently, a study done in ovarian cancer indicated a strong correlation between the expression of MCM2 and MCM5 and Ki-67 LI and p53 levels, presumably associated to mutations of this protein, as well as with patients' survival (49). Because all our cases expressed wild-type p53, it appears that the relevance of MCM5 is independent of p53 status and might be of prognostic value also in patients affected by stage I ovarian cancer. Studies on a larger cohort of patients are mandatory to confirm our observation.

In conclusion, our study shows the genetic diversity of ovarian cancer according to histotype and degree of differentiation as reflected by their profile signature. Evaluation of tumor gene profiling can provide immensely useful information germane to prognosis in early-stage ovarian cancer. Cyclin E and MCM5 expression levels appear to be promising as predictors of disease-free survival in stage I ovarian cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank the Nerina and Mario Mattioli Foundation, Italian Association for Cancer Research, and Luigi and Francesca Lucchetti, for the generous contribution and Prof. Andreas Gescher for insightful comments, suggestions, and critical review of the article.

References

1. Bankhead C. For ovarian cancer, an optimal treatment remains to be found. J Natl Cancer Inst 2004; 96:96–7.
2. Cannistra SA. Cancer of the ovary. N Engl J Med 2004;351:2519–29.
3. Gatta G, Lasota MB, Verdecchia A. Survival of European women with gynaecological tumours, during the period 1978-1989. Eur J Cancer 1998;34:2218–25.
4. Ozols RF, Bookman MA, Connolly DC, et al. Focus on epithelial ovarian cancer. Cancer Cell 2004;5: 19–24.
5. Powell DE, Puls L, van Nagell J, Jr. Current concepts in epithelial ovarian tumors: does benign to malignant transformation occur? Hum Pathol 1992;23:846–7.
6. Puls LE, Powell DE, DePriest PD, et al. Transition from benign to malignant epithelium in mucinous and serous ovarian cystadenocarcinoma. Gynecol Oncol 1992;47:53–7.
7. Scully RE. Pathology of ovarian cancer precursors. J Cell Biochem Suppl 1995;23:208–18.
8. Aunoble B, Sanches R, Didier E, Bignon YJ. Major oncogenes and tumor suppressor genes involved in epithelial ovarian cancer [review]. Int J Oncol 2000; 16:567–76.
9. Menon U, Jacobs IJ. Ovarian cancer screening in the general population: current status. Int J Gynecol Cancer 2001;11 Suppl 1:3–6.
10. Crijns AP, Duiker EW, de Jong S, Willemse PH, van der Zee AG, de Vries EG. Molecular prognostic markers in ovarian cancer: toward patient-tailored therapy. Int J Gynecol Cancer 2006;16 Suppl 1:152–65.
11. See HT, Kavanagh JJ, Hu W, Bast RC. Targeted therapy for epithelial ovarian cancer: current status and future prospects. Int J Gynecol Cancer 2003;13:701–34.
12. Feeley KM, Wells M. Precursor lesions of ovarian epithelial malignancy. Histopathology 2001;38:87–95.
13. Trope C, Kaern J. Adjuvant chemotherapy for early-stage ovarian cancer: review of the literature. J Clin Oncol 2007;25:2909–20.
14. Hough CD, Sherman-Baust CA, Pizer ES, et al. Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. Cancer Res 2000;60:6281–7.
15. Ono K, Tanaka T, Tsunoda T, et al. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res 2000;60:5007–11.
16. Tonin PN, Hudson TJ, Rodier F, et al. Microarray analysis of gene expression mirrors the biology of an ovarian cancer model. Oncogene 2001;20:6617–26.
17. Wang K, Gan L, Jeffery E, et al. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. Gene 1999;229:101–8.
18. Welsh JB, Zarrinkar PP, Sapinoso LM, et al. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci U S A 2001;98:1176–81.
19. Marabese M, Marchini S, Marrazzo E, et al. Expression levels of p53 and p73 isoforms in stage I and stage III ovarian cancer. Eur J Cancer 2008;44: 131–41.
20. Marchini S, Marabese M, Marrazzo E, et al. ΔNp63 expression is associated with poor survival in ovarian cancer. Ann Oncol 2008;19:501–7.
21. Raychaudhuri S, Sutphin PD, Chang JT, Altman RB. Basic microarray analysis: grouping and feature reduction. Trends Biotechnol 2001;19:189–93.
22. McShane LM, Freidlin B, Yu R, Li MC, Simon R. Methods of assessing reproducibility of clustering patterns observed in analyses of microarray data. Bioinformatics 2002;18:1462–9.
23. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002; 99:6567–72.
24. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98: 5116–21.
25. Korn EL, McShane LM, Simon R. Controlling the number of false discoveries: application to high-dimensional genomic data. Stat Plann Inference 2003; 124:2448–55.
26. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 2003;19: 2448–55.
27. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. J Comput Biol 2002;9:505–11.
28. Dudoit S, Fridlyand J, Speed T. Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Stat Assoc 2002; 97:77–87.
29. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 1983;78:316–31.
30. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 2003;95:14–8.
31. Simon RKE, McShane L, Radmacher M, Wright G, Zhao Y. Design and analysis of DNA microarray investigations. New York: Springer Verlag, 2003.
32. R Development Core Team. A language and environment for statistical computing. Vienna (Austria); 2006.
33. Baron-Hay S, Boyle F, Ferrier A, Scott C. Elevated serum insulin-like growth factor binding protein-2 as a prognostic marker in patients with ovarian cancer. Clin Cancer Res 2004;10:1796–806.
34. Holt LJ, Siddle K. Grb10 and Grb14: enigmatic regulators of insulin action—and more? Biochem J 2005; 388:393–406.
35. Chakrabarty S, Kondratick L. Insulin-like growth factor binding protein-2 stimulates proliferation and activates multiple cascades of the mitogen-activated protein kinase pathways in NIH-OVCAR3 human epithelial ovarian cancer cells. Cancer Biol Ther 2006;5: 189–97.
36. Heinzelmann-Schwarz VA, Gardiner-Garden M, Henshall SM, et al. A distinct molecular profile associated with mucinous epithelial ovarian cancer. Br J Cancer 2006;94:904–13.
37. Balkwill FR. Tumour necrosis factor and cancer. Prog Growth Factor Res 1992;4:121–37.
38. Kulbe H, Thompson R, Wilson JL, et al. The inflammatory cytokine tumor necrosis factor-α generates an autocrine tumor-promoting network in epithelial ovarian cancer cells. Cancer Res 2007;67:585–92.
39. Naylor MS, Stamp GW, Foulkes WD, Eccles D, Balkwill FR. Tumor necrosis factor and its receptors in human ovarian cancer. Potential role in disease progression. J Clin Invest 1993;91:2194–206.
40. 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.
41. Farley J, Smith LM, Darcy KM, et al. Cyclin E expression is a significant predictor of survival in advanced, suboptimally debulked ovarian epithelial cancers: a Gynecologic Oncology Group study. Cancer Res 2003;63:1235–41.
42. Rosen DG, Yang G, Deavers MT, et al. Cyclin E expression is correlated with tumor progression and predicts a poor prognosis in patients with ovarian carcinoma. Cancer 2006;106:1925–32.
43. Rosenberg E, Demopoulos RI, Zeleniuch-Jacquotte A, et al. Expression of cell cycle regulators p57(KIP2), cyclin D1, and cyclin E in epithelial ovarian tumors and survival. Hum Pathol 2001;32:808–13.
44. Sui L, Dong Y, Ohno M, et al. Implication of malignancy and prognosis of p27(kip1), cyclin E, and Cdk2 expression in epithelial ovarian tumors. Gynecol Oncol 2001;83:56–63.
45. Korkolopoulou P, Givalos N, Saetta A, et al. Minichromosome maintenance proteins 2 and 5 expression in muscle-invasive urothelial cancer: a multivariate survival study including proliferation markers and cell cycle regulators. Hum Pathol 2005; 36:899–907.
46. Neben K, Korshunov A, Benner A, et al. Microarray-based screening for molecular markers in medulloblastoma revealed STK15 as independent predictor for survival. Cancer Res 2004;64:3103–11.
47. Rosenwald A, Wright G, Wiestner A, et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell 2003;3:185–97.
48. Yu Z, Feng D, Liang C. Pairwise interactions of the six human MCM protein subunits. J Mol Biol 2004; 340:1197–206.
49. Gakiopoulou H, Korkolopoulou P, Levidou G, et al. Minichromosome maintenance proteins 2 and 5 in non-benign epithelial ovarian tumours: relationship with cell cycle regulators and prognostic implications. Br J Cancer 2007;97:1124–34.

Clinical Cancer Research

Analysis of Gene Expression in Early-Stage Ovarian Cancer

Sergio Marchini, Pietro Mariani, Giovanna Chiorino, et al.

Clin Cancer Res 2008;14:7850-7860.