

# Global Gene Expression Profiling of Pleural Mesotheliomas: Overexpression of Aurora Kinases and *P16/CDKN2A* Deletion as Prognostic Factors and Critical Evaluation of Microarray-Based Prognostic Prediction

Fernando López-Ríos,<sup>1</sup> Shannon Chuai,<sup>2</sup> Raja Flores,<sup>3</sup> Shigeki Shimizu,<sup>1</sup> Takatoshi Ohno,<sup>5</sup> Kazuhiko Wakahara,<sup>5</sup> Peter B. Illei,<sup>1</sup> Sanaa Hussain,<sup>1</sup> Lee Krug,<sup>4</sup> Maureen F. Zakowski,<sup>1</sup> Valerie Rusch,<sup>3</sup> Adam B. Olshen,<sup>2</sup> and Marc Ladanyi<sup>1</sup>

Departments of <sup>1</sup>Pathology, <sup>2</sup>Epidemiology and Biostatistics, <sup>3</sup>Surgery, and <sup>4</sup>Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York; and <sup>5</sup>Department of Orthopaedic Surgery, Gifu University School of Medicine, Gifu, Japan

## Abstract

Most gene expression profiling studies of mesothelioma have been based on relatively small sample numbers, limiting their statistical power. We did Affymetrix U133A microarray analysis on 99 pleural mesotheliomas, in which multivariate analysis showed advanced-stage, sarcomatous histology and *P16/CDKN2A* homozygous deletion to be significant independent adverse prognostic factors. Comparison of the expression profiles of epithelioid versus sarcomatous mesotheliomas identified many genes significantly overexpressed among the former, including previously unrecognized ones, such as *uroplakins* and *kallikrein 11*, both confirmed by immunohistochemistry. Examination of the gene expression correlates of survival showed that more aggressive mesotheliomas expressed higher levels of Aurora kinases A and B and functionally related genes involved in mitosis and cell cycle control. Independent confirmation of the negative effect of Aurora kinase B was obtained by immunohistochemistry in a separate patient cohort. A role for Aurora kinases in the aggressive behavior of mesotheliomas is of potential clinical interest because of the recent development of small-molecule inhibitors. We then used our data to develop microarray-based predictors of 1 year survival; these achieved a maximal accuracy of 68% in cross-validation. However, this was inferior to prognostic prediction based on standard clinicopathologic variables and *P16/CDKN2A* status (accuracy, 73%), and adding the microarray model to the latter did not improve overall accuracy. Finally, we evaluated three recently published microarray-based outcome prediction models, but their accuracies ranged from 63% to 67%, consistently lower than reported. Gene expression profiling of mesotheliomas is an important discovery tool, but its power in clinical prognostication has been overestimated. (Cancer Res 2006; 66(6): 2970-9)

## Introduction

Many studies of human cancers that have used microarray data to develop diagnostic or prognostic predictors have done so with relatively few samples (1). Furthermore, early microarray-based predictors were often not subjected to full cross-validation or independent validation, and few have been compared systematically to conventional prognostic markers (1, 2). Thus, the composition of many microarray-based predictors of outcome is unstable, suggesting the persistence of varying degrees of "noise" in these signatures (3). This problem has been most extensively discussed in the setting of microarray-based prognostic predictors in breast cancer (4). Unfortunately, there is often limited enthusiasm for simply retesting correlations established by others due to the high costs of microarray studies and the low novelty of confirmatory work. However, given the strong interest in translating these data into clinical applications, there is a pressing need to independently establish the reproducibility of microarray-based predictors. In the present study, we have sought to address some of these issues in the setting of pleural mesothelioma.

A consideration of the clinical problem of mesothelioma highlights the need for improved prognostic prediction before definitive surgery and for new therapeutic targets. Current treatment of pleural mesothelioma has yet to significantly improve prognosis (5). Despite multimodal therapy, median survivals remain in the range of only 6 to 8 months (6). Response rates to either single-agent or combination chemotherapy have not exceeded 30% in most series, and the best available agents (e.g., pemetrexed) extend median survivals by only 3 months (6, 7). Given the modest survival advantage and the toxicity of chemotherapy, there is a need to identify patients with particularly poor prognosis for whom it is of little benefit and where new therapeutic modalities could be tested. At the other end of the clinical spectrum, there seem to be a small number of unexpectedly long-term survivors. The two most frequently identified favorable prognostic factors are stage I/II and epithelioid histology (8). Unfortunately, accurate staging and histologic subclassification are only possible after definitive surgery (8). Accurate preoperative subtyping of mesothelioma is becoming increasingly important because some do not consider patients with nonepithelioid mesotheliomas to be surgical candidates (8, 9). However, between 25% and 45% of patients with nonepithelioid components in their extrapleural pneumonectomy specimen are initially classified as pure epithelioid based on the pleural biopsy (5, 9).

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

F. López-Ríos is currently at the Departamento de Anatomía Patológica, Hospital Universitario "12 de Octubre" Avenida de Córdoba s/n, 28041 Madrid, Spain.

**Requests for reprints:** Marc Ladanyi, Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021. Phone: 212-639-6369; Fax: 212-717-3515; E-mail: ladanyim@mskcc.org.

©2006 American Association for Cancer Research.  
doi:10.1158/0008-5472.CAN-05-3907

Microarray-based global gene expression analysis can be a powerful approach to address these issues, but most gene expression profiling studies of mesothelioma have been done in cell lines or relatively small series of samples ( $n < 50$ ; refs. 10–14). Prognostic prediction based on gene expression profiles has been attempted by two groups, but the “prognostic gene lists” have shown little or no overlap (13, 15, 16). We report the analysis of a large expression profiling data set of pleural mesotheliomas, most with known *PI6/CDKN2A* deletion status (17), that we have mined for candidate diagnostic, prognostic, and therapeutic markers. We also describe the development of a new prognostic classifier and the independent evaluation of previously published prognostic classifiers.

## Materials and Methods

**Sample procurement and clinicopathologic characteristics.** Mesothelioma tumor samples were procured at surgery between 1990 and 2001 (except for one patient operated in 1989) at the Memorial Sloan-Kettering Cancer (MSKCC) under a protocol approved by the Institutional Review Board. The International Mesothelioma Interest Group staging system was used to determine T and N status and tumor stage. The clinical and pathologic features of the 99 patients whose tumors were in the final expression profiling data set were as follows. The median age was 63 years (range, 33–78 years). There were 24 women and 75 men. The histologic subtype was epithelioid in 69 cases, sarcomatoid in 10 cases, and biphasic in 20 cases. The T status was T<sub>1</sub>/T<sub>2</sub> in 44 patients and T<sub>3</sub>/T<sub>4</sub> in 52 patients. Stage I/II disease was present in 29 patients, and stage III/IV was present in 70 patients.

**Microarray procedures.** Total RNA was extracted from 116 snap-frozen tumor samples (RNeasy kit, Qiagen, Valencia, CA). Based on standard RNA quality checks, 102 of 116 samples (88%) showed no RNA degradation and were therefore used for microarray analysis. Of these 102 samples, 3 were excluded: one duplicate sample, one exclusively peritoneal mesothelioma, and one sample showing very high surfactant gene expression in the initial microarray analysis (frozen section of the remaining banked sample showed predominantly nonneoplastic lung tissue).

All microarray hybridization and scanning steps were done in the MSKCC Genomics Core Laboratory. Briefly, total RNA was converted to double-stranded cDNA with oligo d(T) primers and reverse transcriptase before *in vitro* transcription with biotinylated UTP and CTP. The resulting biotinylated cRNA was then fragmented and hybridized for 16 hours at 45°C to the Affymetrix oligonucleotide Human HG-U133A Genechip (Santa Clara, CA), containing 22,215 probe sets representing ~18,500 transcripts and 14,500 genes.

**Clinical statistical analysis.** Overall survival was calculated from date of surgery to date of last follow-up. Survival probabilities were estimated by the method of Kaplan and Meier. Variables, such as sex, histology, and stage, were related to overall survival using the log-rank test. Variables significant in univariate analysis ( $P < 0.10$ ) were entered into a multivariate Cox model to identify independent prognostic factors.

**Microarray data processing and analysis.** We used the robust multiplex average method to estimate expression for every probe set (18, 19). This algorithm was used, rather than the Affymetrix Microarray Suite 5.0 algorithm, because it has been shown to give more precise estimates, particularly for low-expressing genes.

For unsupervised clustering, we applied a two-dimensional hierarchical clustering algorithm on all 22,215 probe sets, with the Pearson correlation coefficient as the measure of similarity and average linkage as the method to join clusters. For supervised analyses, differentially expressed genes were identified by two-sample *t* tests for binary variables and the univariate Cox model for overall survival. *P*s were adjusted for multiple comparisons using the false discovery rate (FDR) method of Benjamini and Hochberg (20). The threshold for significance was set to control the expected FDR at 5%. We used the Ingenuity Pathway Analysis platform (Mountain View, CA) to examine functional associations between differentially expressed genes.

Outcome prediction rules for the binary variable of “short-term” (<1 year) or “long-term” (≥1 year) survival after surgery were developed

using the *k*-nearest neighbor (KNN) rule and support vector machines (SVM). In the KNN method, a sample is classified based on a majority vote of the classes of the *k* neighbors that are closest to it in terms of Euclidean distance. In the SVM method, a sample is classified based on its position relative to an optimal linear decision boundary constructed on a transformed feature space of the microarray data. For both methods, subsets of probe sets were chosen based on the largest absolute *t* statistics. The number of probe sets was between 5 and 100, incrementing by 5; between 200 and 500, incrementing by 100; and all 22,215 probe sets. For the KNN rule, we varied *k*, whereas for the SVM, we varied the kernel functions (linear, radial, polynomial, and sigmoidal; ref. 21).

Classification rates were estimated based on repeated 10-fold cross-validation. Nine groups were used for developing the model, which included choosing the best genes (a different set each time), which was then applied to the 10th group. This was repeated 10 times so that each sample was predicted using a model that did not include its own data for development. The whole cross-validation was repeated 100 times, and the results were averaged. The SE reported is the SD of the classification accuracy for the 100 repetitions.

To compare a model for survival with the clinical variables to a model consisting of the clinical variables and the microarray predictor, a logistic regression model was built both with and without the microarray predictor. Classification was based on the more probable group from the model. The same repeated cross-validation approach as above was used.

**Validation studies.** We did immunohistochemistry on a tissue microarray containing triplicate 0.6-mm cores of tumors from 65 additional patients (histology: 47 epithelioid, 4 sarcomatoid, and 14 biphasic). Immunohistochemistry studies were done using the avidin-biotin-peroxidase method. The following antibodies were tested: monoclonal uroplakin 3b [clone AU-1; Research Diagnostics, Inc., Concord, MA; 1:1 dilution (prediluted)], polyclonal kallikrein 11 (gift of Eleftherios P. Diamandis, Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Canada; 1:5,000 dilution), monoclonal Aurora kinase A (BD Transduction Laboratories, San Diego, CA; 1:200 dilution), and monoclonal Aurora kinase B (BD Transduction Laboratories; 1:200 dilution). Two authors (F.L.R. and S.H.) independently scored the immunohistochemistry results blinded to the other data. Because dichotomous scoring has been shown to enhance reproducibility (22), we defined such criteria for each antigen. For Uroplakin 3b, any membranous or luminal immunostaining of tumor cells was scored as positive (23, 24). Kallikrein 11 was considered positive when ≥10% of the cells showed an intense cytoplasmic (usually supranuclear) signal (25). Diffuse weak cytoplasmic staining was considered negative. Given that Aurora kinase A and Aurora kinase B are normally undetectable by immunohistochemistry in normal nonmitotic cells (26), any expression was considered positive, regardless of the number of positive cells (27). As described by others, Aurora kinase A shows cytoplasmic expression with occasional membranous enhancement (26), whereas Aurora kinase B exhibits a nuclear pattern, often highlighting mitotic figures (28).

For Western blotting, proteins (50 μg) were electrophoresed on a SDS/10% polyacrylamide gel and transferred to a polyvinylidene fluoride membrane. The Aurora kinase A protein was detected using a rabbit antihuman Aurora kinase A antibody (1:500 dilution; gift of Masashi Kimura, Department of Molecular Pathobiochemistry, Gifu University School of Medicine, Gifu, Japan).

For quantitative reverse transcriptase-PCR (RT-PCR), cDNA was made using the SuperScript III First-Strand Synthesis System (Invitrogen, Carlsbad, CA) on 1 μg of total RNA. We then used predesigned gene-specific primer and probe sets (Taqman Gene Expression Assays, Applied Biosystems, Inc., Foster City, CA) for Aurora kinase A (STK6) and Aurora kinase B, and the assays were run according to the manufacturer’s protocol on a Bio-Rad iCycler (Bio-Rad, Hercules, CA).

## Results and Discussion

**Univariate and multivariate analyses: prognostic significance of P16/CDKN2A status.** The median follow-up after surgery among the six subjects still alive was 25 months (range, 1–64 months). There were 44 patients who survived <1 year (median, 5.6

**Table 1.** Univariate and multivariate analysis for overall survival after surgery

Factors	Univariate hazard ratio (95% confidence interval)	Univariate <i>P</i>	Multivariate hazard ratio (95% confidence interval)	Multivariate <i>P</i>
Sex: female vs male	1.56 (0.95, 2.59)	0.08	1.43 (0.77, 2.66)	0.26
T status: T <sub>3</sub> /T <sub>4</sub> vs T <sub>1</sub> /T <sub>2</sub>	1.72 (1.12, 2.61)	0.01	1.03 (0.60, 1.78)	0.92
Stage: III/IV vs I/II	2.38 (1.46, 3.88)	<0.001	2.23 (1.18, 4.22)	0.01
Histology: B/S vs E	2.44 (1.53, 3.87)	<0.001	1.84 (1.08, 3.16)	0.03
P16 deletion status: D vs ND	2.43 (1.41-4.17)	0.001	0.49 (0.27, 0.88)	0.02

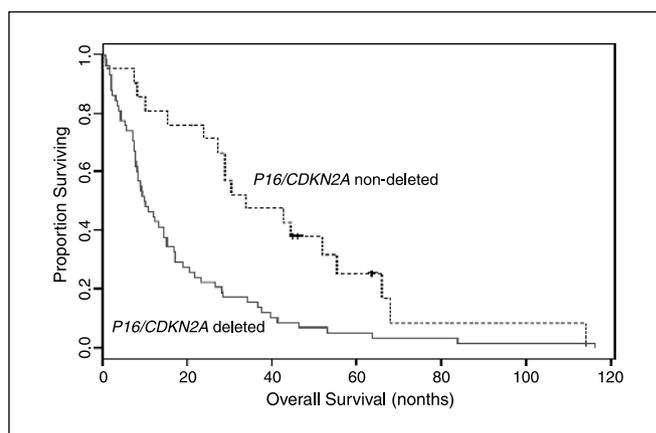
Abbreviations: E, epithelioid; B, biphasic; S, sarcomatoid; D, deleted; ND, nondeleted.

months; range, 0-11.9 months), and 53 who survived >1 year (median, 28.7 months). The total is <99 because two patients who were alive with <1 year follow-up were excluded from the survival analyses (samples MS-43 and MS-97). Univariate survival analysis showed a significant survival advantage for patients with T<sub>1</sub>/T<sub>2</sub> tumors, stage I/II, and exclusively epithelioid histology (Table 1). For *P16/CDKN2A* status, data from our previous study (17) were available in 80 cases, of which 59 showed homozygous deletion and 21 did not. Univariate analysis of the prognostic effect of *P16/CDKN2A* status showed a significant survival advantage for nondeleted cases (*P* = 0.001; median survival = 34 versus 10 months for deleted cases; Fig. 1), a novel observation in pleural mesothelioma. Multivariate analysis was then done in the subgroup of patients (*n* = 80) with data available on multiple factors. In the multivariate analysis, *P16/CDKN2A* deletion status, stage, and the presence of a sarcomatous component remained significant (Table 1). Interestingly, in spite of the known association of *P16/CDKN2A* deletion with the presence of sarcomatous component (17), these two factors were both prognostic in the multivariate model. Among other factors, neither N status (N<sub>0</sub> versus other), platelet count (<400 versus >400 × 10<sup>9</sup>/L), nor asbestos exposure history (present versus absent) were found to be significantly associated with survival in the univariate analysis (data not shown). The results obtained when survival from diagnosis was used instead of survival from operation were qualitatively identical. This is to our knowledge the first demonstration of *P16/CDKN2A* deletion as a negative prognostic factor in pleural mesothelioma. Consistent with our finding, loss of p16 immunoreactivity has recently been identified as an independent predictor of poorer survival in peritoneal mesothelioma (29).

**Overview of unsupervised hierarchical clustering and supervised analyses.** The unsupervised hierarchical clustering done on all 99 samples is shown in Supplementary Fig. S1. Mesothelioma has three main histologic types, with the epithelioid type being the most frequent (>50%) followed by the biphasic and sarcomatoid types. Although biphasic and sarcomatous cases were generally adjacent in the observed clusters (Supplementary Fig. S1), overall the unsupervised clustering did not display a robust separation of mesotheliomas according to histologic subtype. *P16/CDKN2A* deletion status, asbestos exposure history, stage, T status, N status, platelet count, or survival (<1 versus ≥1 year) showed little or no relation to the unsupervised clusters (data not shown). Table 2 lists the number of differentially expressed probe sets for the comparisons of interest. For clarity of data presentation, we provide separate tables for the top probe sets (see Table 2), in addition to the complete lists of significant probe sets for each

comparison (see Table 2). Aside from the comparisons described in detail below, we also examined gene expression correlates of asbestos exposure history, N<sub>0</sub> status versus other, stages I/II versus III/IV, and platelet count (<400 versus >400 × 10<sup>9</sup>/L) but found no significant associations (data not shown).

**Comparison of gene expression profiles of epithelioid versus sarcomatoid mesotheliomas.** Considering the difficulties in determining histologic subtype in pleural biopsies, an important distinction given that nonepithelioid cases might be excluded from aggressive treatment (8, 9), we examined genes whose expression differed between the histologic subclasses. We detected 1,039 probe sets differentially expressed between 69 epithelioid and 10 sarcomatoid mesotheliomas (Table 3; Supplementary Table S1; biphasic tumors were excluded to avoid the confounding effect of the varying proportions of epithelial components among cases thus classified). Significant gene expression differences were almost exclusively due to genes more highly expressed in epithelioid mesotheliomas, including genes typical of epithelial differentiation elsewhere (e.g., claudin 15 and cadherin 3). However, an unexpected observation in epithelioid mesotheliomas was the prominent expression of uroplakins 1B and 3B (*UPK1B* and *UPK3B*) and kallikrein 11 (*KLK11*). Because the expression of these proteins has not been previously studied in mesothelioma, these genes were selected for confirmation by immunohistochemistry. We found positive immunohistochemistry staining of mesothelioma cells (as defined for each antibody in Materials and Methods) for uroplakin 3B in 42% of cases (26 of 62; Fig. 2A)



**Figure 1.** Kaplan-Meier plots of overall survival after surgery according to *p16/CDKN2A* homozygous deletion status. *P* = 0.001, comparison between the curves.

**Table 2.** Number of differentially expressed genes for specific comparisons

Comparison	No. probe sets significantly differentially expressed using FDR method	No. samples in comparison	Top 25 probe sets	All probe sets
Histology: epithelioid vs sarcomatoid	1,039	79	Table 3*	Table S1
<i>P16/CDKN2A</i> deletion status: deleted vs nondeleted	32	80	—	Table S2
Overall survival after surgery	1,481	97	Table S3A	Table S3B
Overall survival after surgery (epithelioid histology only)	208	69	Table S4A	Table S4B
Survival after surgery: <1 vs >1 y	104	97	Table S5A	Table S5B
Survival after surgery: 25 shortest survivors vs 25 longest survivors	294	50	Table S6A	Table S6B

\*Top 20 probe sets only.

and kallikrein 11 in 79% of cases (48 of 61; Fig. 2B). The finding of uroplakin expression in epithelioid mesotheliomas is notable because these transmembrane proteins have been considered specific for the urothelial lineage (23, 30). Kallikreins are a family of secreted serine proteases that includes prostate-specific antigen. The expression of some kallikrein family members, including *KLK11*, may also be of prognostic value (reviewed in ref. 31).

**Comparison of gene expression profiles of *P16/CDKN2A* deleted versus *P16/CDKN2A* nondeleted mesotheliomas.** The comparison of *P16/CDKN2A* homozygously deleted versus nondeleted samples yielded 32 Affymetrix probe sets corresponding to 25 genes that were significantly differentially expressed (Supplemental Table S2). The top genes specifically expressed in *P16/CDKN2A*-deleted cases were *RHOBTB3*, *SHOX2*, and *DLC1*. Conversely, the top genes expressed in nondeleted cases were B-factor (*properdin*) and methylthioadenosine phosphorylase (*MTAP*), but surprisingly, did not include *P16/CDKN2A* itself. Of the 25 genes in Supplementary Table S2, only *MTAP* is located on chromosome 9 where it is ~100 kb telomeric to *P16/CDKN2A* at 9p21, its proximity explaining why it is often codeleted with the latter (17). That only 25 genes were significantly differentially expressed in this analysis suggests that *P16/CDKN2A* deletion does not have a strong effect on gene expression, consistent with its posttranscriptional mode of action (i.e., by controlling kinase signaling). The fact that *P16/CDKN2A* itself does not appear as a significantly differentially expressed gene in this comparison (in spite of three U133A probe sets) suggests loss of *P16/CDKN2A* expression through mechanisms other than homozygous deletion in some of the 21 nondeleted cases. Indeed, we have found evidence of *P16/CDKN2A* promoter methylation in some mesotheliomas without *P16/CDKN2A* homozygous deletion.<sup>6</sup> In contrast, *MTAP*, while frequently codeleted with *P16/CDKN2A*, is not known to be subject to promoter methylation, and this may explain the stronger association of *MTAP* gene expression with *P16/CDKN2A* deletion status. Indeed, the latter association is perhaps all the more remarkable given that up to 12% of cases with *P16/CDKN2A* homozygous deletion do not show loss of both copies of *MTAP* (17).

**Identification of gene expression correlates of outcome.** We used several approaches to define gene expression profiles related

to outcome in the subset of 97 patients considered suitable for survival analysis. As stated above, two of the original 99 patients were excluded because they were alive but had <1 year follow-up. In the first approach, we considered overall survival after surgery using a univariate Cox model. This identified 1,481 probe sets that were significantly correlated with overall survival (Supplementary Tables S3A and B). Next, given the known favorable prognostic effect of epithelioid histology, we restricted the same analysis to the 69 patients with epithelioid tumors to exclude histology as a confounding factor. This produced a reduced list of 208 significant probe sets (Supplementary Tables S4A and B), suggesting a major contribution of correlates of histology to the set of prognostic genes identified by global gene expression profiling.

In a second analysis, the 97 patients were divided into two groups: "short-term" ( $n = 44$ ) and "long-term" ( $n = 53$ ) survivors based on their actual survival from operation, either <1 or  $\geq 1$  year. In this comparison, 104 probe sets were significant (Supplementary Tables S5A and B). Reasons for the lower number of significant genes in this analysis include reduced statistical power due to the arbitrary nature of a 1-year survival cutoff (compared with the above analysis using a univariate Cox model) and the related issue of the large number of patients surviving only slightly <1 or >1 year, whose separation by this cutoff is less likely to reflect biological differences (median survival = 14.5 months). To explore the contribution of the latter, we compared the profiles of the 25 shortest survivors to the 25 longest ones (i.e.,  $n = 50$ ), essentially corresponding to the top and bottom quartiles in survival. This comparison yielded 294 significant probe sets (Supplementary Tables S6A and B), in spite of the smaller data set ( $n = 50$  versus  $n = 97$ ). Thus, this modified analysis enhanced the detection of gene expression differences between the most aggressive and the least aggressive mesotheliomas.

**Development of a novel mesothelioma prognostic classifier.** Next, we used microarray data from the above 97 samples for prediction of 1-year survival from operation. Gene selection in this classification model was the same as described in the preceding section. We fit KNN and SVM models for varying numbers of genes, neighbors (for the former), and kernels (for the latter) using a repeated cross-validation methodology discussed in Materials and Methods. The best KNN model was 68.4% accurate (SE = 2.7%) and consisted of 200 genes with  $k = 13$ . This corresponded to 56.7% accuracy for the long-term survivors and 78.0% for the short-term survivors. The best SVM model was 66.2% accurate (SE = 3.0%) and consisted of 500 genes with a radial kernel. It was again harder to

<sup>6</sup> P. Illei and M. Ladanyi, unpublished data.

**Table 3.** Top 20 probe sets differentially expressed according to histology

Probe set ID	Higher group	Fold change	P	Gene name	Gene symbol
206658_at	Epithelioid	20.1	1.99e-20	Uroplakin 3B	<i>UPK3B</i>
219411_at	Epithelioid	2.1	6.20e-18	Engulfment and cell motility 3	<i>ELMO3</i>
215066_at	Epithelioid	2.0	3.21e-13	Protein tyrosine phosphatase, receptor type, F	<i>PTPRF</i>
209293_x_at	Epithelioid	3.9	8.78e-13	Inhibitor of DNA binding 4	<i>ID4</i>
209631_s_at	Epithelioid	2.8	3.52e-12	G protein-coupled receptor 37	<i>GPR37</i>
219640_at	Epithelioid	4.5	7.63e-12	Claudin 15	<i>CLDN15</i>
206281_at	Epithelioid	1.9	8.06e-12	Adenylate cyclase activating polypeptide 1	<i>ADCYAP1</i>
207790_at	Epithelioid	2.6	9.14e-12	Leucine-rich repeat containing 1	<i>LRRC1</i>
215425_at	Epithelioid	1.7	1.51e-11	BTG family, member 3	<i>BTG3</i>
200635_s_at	Epithelioid	2.9	3.55e-11	Protein tyrosine phosphatase, receptor type, F	<i>PTPRF</i>
210065_s_at	Epithelioid	7.9	4.61e-11	Uroplakin 1B	<i>UPK1B</i>
219127_at	Epithelioid	2.7	5.23e-11	Hypothetical protein MGC11242	<i>MGC11242</i>
205470_s_at	Epithelioid	5.1	1.32e-10	Kallikrein 11	<i>KLK11</i>
213050_at	Epithelioid	3.4	1.41e-10	Cordon-bleu homologue (mouse)	<i>COBL</i>
211712_s_at	Epithelioid	2.2	1.58e-10	Annexin A9	<i>ANXA9</i>
208470_s_at	Epithelioid	6.6	1.75e-10	Haptoglobin-related protein or haptoglobin	<i>HPR/HP</i>
205710_at	Epithelioid	3.5	1.88e-10	Low-density lipoprotein-related protein 2	<i>LRP2</i>
215015_at	Epithelioid	3.0	2.99e-10	Hypothetical protein LOC92558	<i>LOC92558</i>
210064_s_at	Epithelioid	7.2	3.01e-10	Uroplakin 1B	<i>UPK1B</i>
210761_s_at	Epithelioid	1.6	3.75e-10	Growth factor receptor-bound protein 7	<i>GRB7</i>

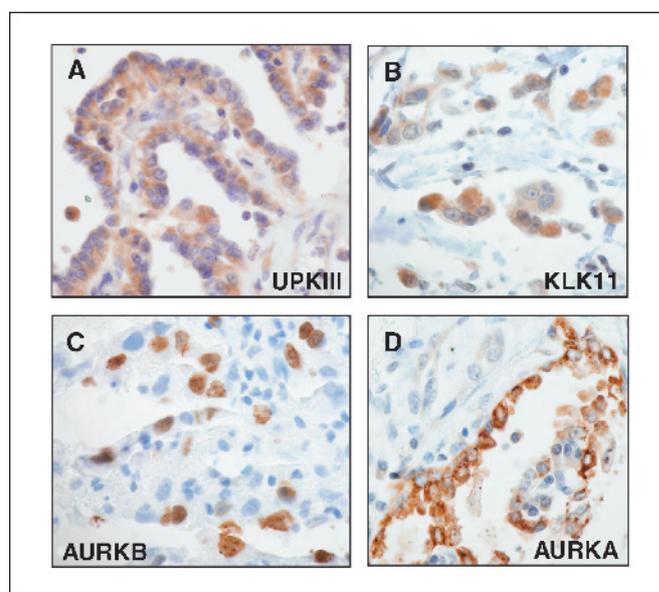
classify the long-term survivors, with 54.5% of the long-term survivors correctly predicted and 76.3% of the short-term survivors. To have a prediction model that was smaller but nearly as accurate, we chose a KNN model with 35 probe sets and  $k = 11$  that was 65.1% accurate (SE = 2.9%). This model was 56.5% accurate for the long-term survivors and 72.2% accurate for the short-term survivors. We also tested the partial least-squares method in this predictor development, but the results were very similar to those obtained with the t-statistic (data not shown).

Because a KNN model with 35 probe sets gave good results, and because cross-validation gives slightly different genes in each iteration, we designated the top 35 probe sets outside of cross-validation, corresponding to 29 genes, as the "MSKCC classifier" (Table 4). Because the gene selection in the prediction model was the same as described in the previous section, these are also the top 35 probe sets in Supplementary Table S5B. All genes except *NR4A2* were also present in the overall survival list (Supplementary Table S3B), and all genes except *NR4A2* and *STK39* also seem differentially expressed in the supervised comparison of short-term and long-term survivors (Supplementary Table S6B). Overall, the 29 genes include 19 unfavorable and 10 favorable ones.

**Evaluation of a microarray-based prognostic classification against clinicopathologic prognostic variables.** Next, to critically assess the practical clinical value of microarrays, we examined whether our microarray-based model would improve predictive accuracy when compared with a model built on standard clinical variables and p16/CDKN2A status. In a repeated cross-validated manner, we fit a logistic regression model using stage, histology, and p16/CDKN2A status and compared it with the same model combined with the above microarray model with 35 genes and  $k = 11$ . The clinical p16/CDKN2A model had an average classification accuracy of 73% (SE = 2.4%), whereas the clinical p16/CDKN2A plus microarray model had an average classification accuracy of 71.5% (SE = 1.7%). The former model was 84.4% accurate for the long-term survivors and 65% accurate for the short-term survivors, whereas the latter model was 64.3% accurate

for the long-term survivors and 75.7% accurate for the short-term survivors. Thus, the microarray variable tended to shift the classification in the direction of the short-term survivors, which was consistent with its performance in the pure microarray model.

**Contribution of gene expression correlates of histology to the MSKCC prognostic classifier.** Because histology has a strong effect on both prognosis and gene expression in mesothelioma, we next examined whether some genes in the MSKCC classifier reflected this association, either partly or predominantly.



**Figure 2.** Demonstration of proteins of interest by immunohistochemistry. A, uroplakin III expression in an epithelioid pleural mesothelioma ( $\times 40$ ). B, kallikrein 11 expression in an epithelioid pleural mesothelioma ( $\times 40$ ). C, Aurora kinase B expression in pleural mesothelioma ( $\times 40$ ). D, Aurora kinase A expression in pleural mesothelioma ( $\times 40$ ). This protein was also studied by Western blotting (Supplementary Fig. S3).

**Table 4.** MSKCC prognostic classifier

Probe set ID	Survival association	Gene	Gene symbol	Presence in supervised comparison lists*	Survival association histology independent vs dependent †
201236_s_at	Favorable	BTG family, member 2	<i>BTG2</i>	A, D	?
201425_at	Favorable	Aldehyde dehydrogenase 2 family (mitochondrial)	<i>ALDH2</i>	A, B, C, D	Both
201998_at	Favorable	Sialyltransferase 1	<i>SIAT1</i>	A, B, D	Independent
202094_at	Unfavorable	Baculoviral IAP repeat-containing 5 (survivin)	<i>BIRC5</i>	A, C, D	Dependent
202786_at	Unfavorable	Serine threonine kinase 39 (SPAK)	<i>STK39</i>	A	?
203623_at	Unfavorable	Likely orthologue of mouse plexin 3	<i>PLXN3</i>	A, B, D	Independent
203854_at	Favorable	I factor (complement)	<i>IF</i>	A, B, C, D	Both
204033_at	Unfavorable	Thyroid hormone receptor interactor 13	<i>TRIP13</i>	A, D	?
204092_s_at	Unfavorable	Aurora kinase A (serine/threonine kinase 6)	<i>AURKA</i>	A, C, D	Dependent
208079_s_at	Unfavorable	Aurora kinase A (serine/threonine kinase 6)	<i>AURKA</i>	A, C, D	Dependent
204621_s_at	Favorable	Nuclear receptor subfamily 4, group A, member 2	<i>NR4A2</i>	None	?
204825_at	Unfavorable	Maternal embryonic leucine zipper kinase	<i>MELK</i>	A, C, D	Dependent
205167_s_at	Unfavorable	Cell division cycle 25C	<i>CDC25C</i>	A, D	?
206067_s_at	Favorable	Wilms tumor 1	<i>WT1</i>	A, C, D	Dependent
216953_s_at	Favorable	Wilms tumor 1	<i>WT1</i>	A, C, D	Dependent
206391_at	Favorable	Retinoic acid receptor responder	<i>RARRES1</i>	A, C, D	Dependent
206392_s_at	Favorable	Retinoic acid receptor responder	<i>RARRES1</i>	A, C, D	Dependent
221872_at	Favorable	Retinoic acid receptor responder	<i>RARRES1</i>	A, D	?
206697_s_at	Favorable	Haptoglobin or haptoglobin-related protein ‡	<i>HP/HRP</i>	A, C, D	Dependent
208470_s_at	Favorable	Haptoglobin or haptoglobin-related protein ‡	<i>HP/HRP</i>	A, B, C, D	Both
207015_s_at	Favorable	Aldehyde dehydrogenase 1 family, member A2	<i>ALDH1A2</i>	A, C, D	Dependent
208002_s_at	Unfavorable	Brain acyl-CoA hydrolase	<i>BACH</i>	A, C, D	Dependent
208511_at	Unfavorable	Pituitary tumor-transforming 3 (securin)	<i>PTTG3</i>	A, D	?
208613_s_at	Unfavorable	Filamin B, beta (actin binding protein 278)	<i>FLNB</i>	A, B, C, D	Both
208711_s_at	Unfavorable	Cyclin D1	<i>CCND1</i>	A, B, D	Independent
208712_at	Unfavorable	Cyclin D1	<i>CCND1</i>	A, B, D	Independent
208936_x_at	Favorable	Lectin, galactoside-binding, soluble, 8 (galectin 8)	<i>LGALS8</i>	A, C, D	Dependent
209464_at	Unfavorable	Aurora kinase B	<i>AURKB</i>	A, B, C, D	Both
212949_at	Unfavorable	Barren homologue ( <i>Drosophila</i> )	<i>BRRN1</i>	A, C, D	Dependent
213288_at	Unfavorable	O-acyltransferase domain containing 2	<i>OACT2</i>	A, D	?
217755_at	Unfavorable	Hematologic and neurologic expressed 1	<i>HNI</i>	A, B, D	Independent
218355_at	Unfavorable	Kinesin family member 4A	<i>KIF4A</i>	A, B, C, D	Both
218586_at	Unfavorable	Chromosome 20 open reading frame 20	<i>C20orf20</i>	A, B, D	Independent
219490_s_at	Unfavorable	DNA cross-link repair 1B (PSO2 homologue)	<i>DCLRE1B</i>	A, C, D	Dependent
221591_s_at	Unfavorable	Hypothetical protein FLJ10156	<i>FLJ10156</i>	A, B, D	Independent

NOTE: Represents top 35 probe sets differentially expressed between short-term (<1 year) and long-term (>1 year) survivors (Table S5B), in numerical order by probe set number but with grouping of probe sets for the same gene.

\*A: overall survival after surgery (Table S3B); B: overall survival after surgery – epithelioid histology only (Table S4B); C: histology: epithelioid versus sarcomatoid (Table S1); D: survival after surgery: 25 shortest survivors vs 25 longest survivors (Table S6B).

† Interpretation of data in adjacent column as follows: B not C = independent, C not B = dependent, B and C = both (see text for explanation), not B not C = ? (inconclusive).

‡ Because the sequences of the genes for haptoglobin and haptoglobin-related protein are ~96% identical, it is unclear which of these two transcripts is measured by probe sets 206697\_s\_at and 208470\_s\_at. However, given their closely matching results in the present analysis, these probe sets are likely to be measuring expression of the same gene, either HP or HPR, at least in this data set.

We cross-referenced the 35 probe sets in the MSKCC classifier genes with the list of genes differentially expressed between epithelioid and sarcomatoid tumors (Supplementary Table S1). This identified 20 probe sets in common, corresponding to 16 genes, and as expected, among these 16 overlapping genes, all favorable genes were associated with epithelioid histology, and

all unfavorable genes were associated with sarcomatoid histology (Table 4). There were also 13 of 35 probe sets in common with the list of genes correlated with overall survival among epithelioid mesothelioma cases only. Interestingly, four unfavorable genes in the classifier (*AURKB*, *FLNB*, *SPC25*, and *KIF4A*) are both significantly overexpressed in sarcomatoid mesotheliomas and

associated with unfavorable outcome among epithelioid mesothelioma cases only. Thus, these genes are not only more highly expressed in the subset of mesotheliomas with unfavorable histology (sarcomatoid) but also are differentially expressed among epithelioid mesotheliomas, being higher in more aggressive cases. Overall, these analyses suggest that 11 of 29 genes in the MSKCC classifier reflect correlates of histology ("Dependent" in last column of Table 4), whereas the remaining 18 genes have a prognostic effect that seems partly or completely independent of histology. Notably, of the 10 favorable genes in the classifier, half (*WT1*, *RARRES1*, *HP/HRP*, *ALDH1A2*, and *LGALS8*) reflect an association with epithelioid histology (and are not present on Supplementary Table S4B) and may therefore have no direct biological effect on aggressive behavior. These observations highlight the need for large data sets to critically evaluate how gene expression correlates of standard clinicopathologic variables contribute to the composition of microarray-based prognostic classifiers.

**Independent evaluation of published mesothelioma prognostic classifiers.** Next, we compared the composition of the MSKCC classifier to three previously published prognostic classifiers or gene lists. In the first one, Pass et al. used a neural network approach trained on Affymetrix U95A microarray data from 21 mesotheliomas to develop a 27-gene classifier (designated here as the "Karmanos classifier") that separated patients into short-term and long-term survivors (survival from operation <12 or  $\geq 12$  months, respectively; ref. 13). However, there is almost no overlap between the 27 genes in the Karmanos classifier and the 29 genes in the MSKCC classifier, with only *BTG2* being present in both (Fig. 3). We therefore wished to evaluate the performance of the Karmanos classifier on our larger Affymetrix microarray data set. Because the Karmanos prognostic classifier was based on data from Affymetrix U95A arrays, we first did cross-platform gene mapping using the comparison spreadsheets provided by Affymetrix (<http://www.affymetrix.com>). Probe set redundancies were resolved using the "best match" option. The 27 probe sets on the U133A chip that matched the 27 probe sets from the U95A chip for the 27 genes in the Karmanos classifier are provided in Supplementary Table S7. Because the neural network software used by Pass et al. was not available to us, we used the same KNN and SVM algorithms presented above to evaluate the

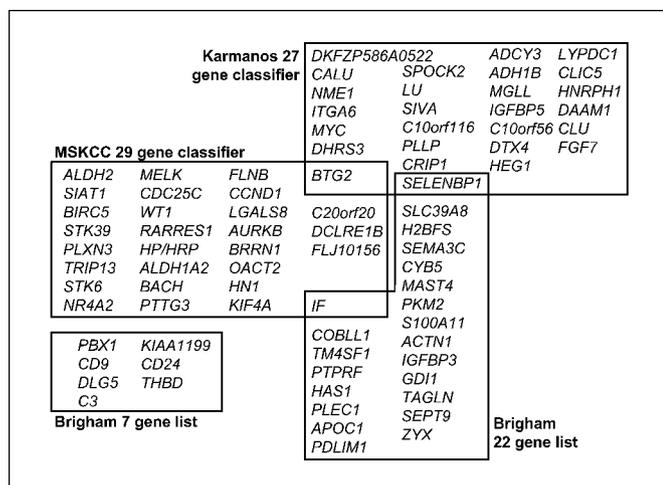
Karmanos classifier. Its accuracy in our data set ( $n = 97$ ) was 65.1% for  $k = 9$  (SE = 1.9%) and 63.9% for the sigmoid kernel (SE = 2.5%), with both models more accurate for predicting short-term versus long-term survivors. This compared with the reported accuracy of 76% (95% confidence interval, 51-92%) in their independent set of 17 patients (13).

Next, we examined two different sets of prognostic genes, a 22-gene list (15) and a nonoverlapping 7-gene list (16) recently published by Gordon et al., here designated the "Brigham lists." Because no Affymetrix probe set IDs were provided in these publications, Genbank accession numbers were used to identify the Affymetrix U133A probe set IDs corresponding to these genes. The Brigham 22-gene list shared one favorable gene with the MSKCC prognostic classifier, *complement factor I*, and another favorable gene, *SELENBP1*, with the Karmanos classifier (Fig. 3). Gordon et al. used four genes from their 22-gene list to derive a prognostic predictor of 1-year survival based on expression ratios obtained by quantitative RT-PCR (15) that functioned as follows: if the geometric mean of the gene ratios *KIAA0977* (also called *COBLL1*)/*GDI1*, *L6* (also called *TM4SF1*)/*CTHBP* (also called *PKM2*), and *L6/GDI1* was  $>1$ , this predicted long-term survival, whereas  $<1$  indicated short-term survival. This gene ratio predictor was reported to be 75% accurate in an independent set of 20 short-term (<6.8 months) or long-term (>24.8 months) survivors (16). We therefore simulated this gene expression ratio-based predictor using our data for the corresponding probe sets (203641\_s\_at for *KIAA0977*, 209386\_at for *TM4SF1*, 201251\_at for *CTHBP*, and 201864\_at for *GDI1*). This showed an accuracy of only 67% for prediction of 1-year survival.

In a subsequent study, the same group obtained a second list of genes correlated with prognosis, designated here as the Brigham 7-gene list. It does not overlap with any of the other classifiers (Fig. 2). This set of genes was used to derive a second prognostic predictor of 1-year survival based on expression ratios. Four ratios were based on four genes as follows: *CD9/KIAA1199*, *CD9/THBD*, *DLG5/KIAA1199*, and *DLG5/THBD*. Again, a geometric mean of the ratios of  $>1$  predicted long-term survival, whereas  $<1$  indicated short-term survival. Using our data from probe sets 201005\_at (*CD9*), 201681\_s\_at (*DLG5*), 203887\_s\_at (*THBD*), and 212942\_s\_at (*KIAA1199*), we sought to reproduce this gene expression ratio-based predictor to test its "cross-platform" performance. This showed an accuracy of only 63% for prediction of 1-year survival, compared with a published accuracy of 69% in an independent set of 13 short-term ( $\leq 5$  months) and 13 long-term ( $\geq 18$  months) survivors (16).

Overall, the results of the MSKCC classifier and our independent validation of the previously published classifier and gene ratio tests suggest an upper limit to the prognostic information contained within gene expression profiles in mesothelioma.

Finally, because both of the other groups had used patient sets enriched for very short term and very long term survivors to generate their prognostic gene lists, we examined whether genes in the Karmanos classifier and the Brigham 22-gene and 7-gene lists were present among the broader list of genes that we found to be differentially expressed between the most aggressive and the least aggressive mesotheliomas (Supplementary Table S6B). The Brigham 22-gene list had seven overlapping genes (*SELENBP1*, *COBLL1*, *PTPRF*, *IF*, *SLC39A8*, *S100A11*, and *GDI1*), and their 7-gene list had two (*CD24* and *complement component 3*). Of the 27 genes in the Karmanos classifier, only four were similarly correlated in our data (i.e., *BTG2*, *ADH1B*, *SELENBP1*, and *DKFZP586A0522*). Given that



**Figure 3.** MSKCC prognostic classifier. Graphical comparison with three other prognostic gene lists. See text for details.

the ordering of false positives (biological noise) is likely to be random, whereas the ordering of true positives is not, this comparison of overlapping genes and the analysis depicted in Fig. 3 identify genes of particular interest for future studies. The corollary of this observation is that the very limited overlaps observed suggest the presence of considerable “noise” in published prognostic gene lists.

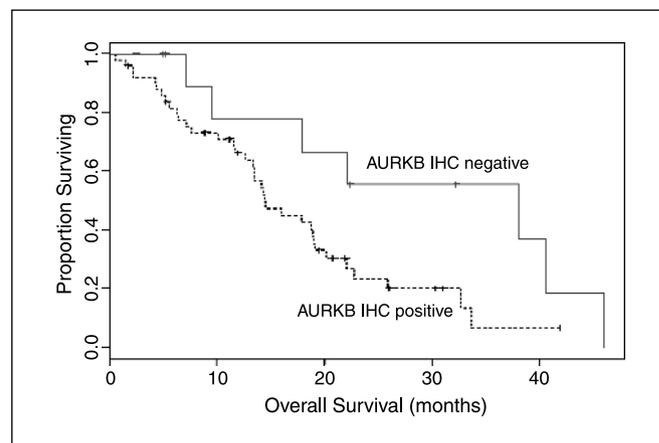
**Analysis of the core predictive genes in the MSKCC prognostic classifier reveals overrepresentation of genes involved in cell cycle control and mitosis.** When we analyzed the MSKCC classifier with the Ingenuity Pathway Analysis software, we found that 11 of the 29 genes (38%) were present in a network of genes linked directly or indirectly to each other by protein-protein or regulatory interactions: *cyclin D1*, *BTG2*, *WT1*, *CDC25C*, *ALDH1A2*, *HNI*, *KIF4A*, *HP*, *AURKA*, *AURKB*, and *BIRC5* (survivin) (Supplementary Fig. S2). This overrepresentation of 11 of 29 genes in one network was highly significant with an estimated  $P$  of  $10^{-22}$  by the Ingenuity software. This Ingenuity network is associated with cell cycle functions, cell death, and cancer and remarkably also includes *P16/CDKN2A* as one of its central nodes (Supplementary Fig. S2). This analysis highlights the remarkable overrepresentation of genes involved in cell cycle control and mitosis within the MSKCC prognostic classifier, including *CCND1*, *BTG2*, *CDC25C*, *BIRC5* (survivin), *PTTG3*/securin, *AURKA* and *AURKB*, and *KIF4A*, all except *BTG2* being associated with aggressive course. Survivin has previously been shown to be highly expressed in almost all mesothelioma cell lines and mesothelioma tissues tested (32, 33). The significance of *CCND1*, encoding cyclin D1, to mesothelioma biology is also reflected by the presence in our prognostic classifier of *BTG2*, a transcriptional repressor of *CCND1* (34), as a predictor of less aggressive behavior. Decreased expression of *BTG2* relative to corresponding normal tissue has been observed in other tumors (35, 36). The *CDC25C* phosphatase is involved in both entry into S phase and  $G_2$ -M progression (37). Overexpression of *CDC25* phosphatases in human cancers often correlates with aggressive features and poor prognosis, and a number of small molecule inhibitors of *CDC25* have recently been developed (37).

We selected Aurora kinases A and B for further analysis. We first sought to confirm and localize the expression of Aurora kinases A and B in mesothelioma, neither of which had been previously studied in this cancer. We found staining of mesothelioma cells for Aurora kinase B in 81% (50 of 62; Fig. 2C) and Aurora kinase A in 48% (29 of 60; Fig. 2D) of the cases on a tissue microarray of 65 independent tumors. To validate the Affymetrix data for Aurora kinases A and B (probe sets 208079\_s\_at and 209464\_at, respectively), we did quantitative RT-PCR on a subset of 24 samples of the original 99 samples used for the microarray analysis. For both genes, the Affymetrix and quantitative RT-PCR expression levels were significantly correlated (Spearman correlation coefficients  $r^2 = 0.58$ ,  $P = 0.004$  and  $r^2 = 0.46$ ,  $P = 0.024$ , respectively; data not shown). For Aurora kinase A, we also did Western blotting in a subset of 14 samples, confirming that protein levels, as measured by densitometry of the Western blot, paralleled the transcript levels determined by the microarray data (Spearman correlation coefficient  $r^2 = 0.74$ ,  $P = 0.002$ ; Supplementary Fig. S3). Finally, we examined whether immunohistochemistry expression of Aurora kinase A and Aurora kinase B in tumor cells correlated with survival in the independent group of 65 additional patients represented on the tissue microarray. This more recent cohort of patients had a shorter median follow-up with more patients alive

but was similar in other aspects to the set of 99 patients whose tumors were studied by expression profiling. In spite of the different detection method, different scoring (continuous for Affymetrix data, categorical for immunohistochemistry data), smaller study group, and the shorter available follow-up, Aurora kinase B expression by immunohistochemistry was nonetheless associated with a worse overall survival in this independent analysis ( $P = 0.05$ ; Fig. 4). Immunohistochemistry detection of Aurora kinase A did not show correlation with survival (not shown).

Aurora kinases are serine/threonine kinases with multiple roles in mitotic progression, including  $G_2$ -M transition, mitotic spindle organization, chromosome segregation, and cytokinesis (38, 39). Aurora kinase B and other chromosomal passenger proteins (Aurora C, INCENP, borealin, and survivin) are involved in the coordination of chromosome segregation and cytokinesis during mitosis (38, 40). Overexpression of several of these genes is known to contribute to oncogenesis. Overexpression of survivin (and of two of the Aurora kinases, Aurora kinases A and B; see below) is protumorigenic at least in part because it is associated with resistance to mitotic catastrophe, a type of apoptosis caused by aberrant mitosis (41–43). Indeed, an association of overexpression with aggressive cancer phenotypes and adverse clinical outcomes has been observed across multiple tumor types for Aurora kinase A (22, 44–46), Aurora kinase B (47), and survivin (48). Consistent with these data, we found Aurora kinase A and Aurora kinase B (and their binding partners survivin and *TPX2*, respectively) to be more highly expressed in sarcomatoid mesotheliomas and in the worse prognosis subgroups. Abrogation of p53 checkpoint function seems to be required for tumorigenesis mediated by Aurora kinase overexpression (40) and, notably, most mesotheliomas have at least partial functional inactivation of p53 through p16/CDKN2A deletion, which also results in p14<sup>ARF</sup> loss (17).

Remarkably, two other components of the MSKCC prognostic classifier, *KIF4A* and *PTTG3* (securin), also encode key mitotic proteins. *KIF4A* has been recently described as a chromosome-associated motor with essential roles in anaphase spindle dynamics and cytokinesis (49). Overexpression of *PTTG3*/securin causes aneuploidy (50, 51) and is associated with worse outcome in other cancers (52).



**Figure 4.** Kaplan-Meier plots of overall survival after surgery according to positivity for Aurora kinase B by immunohistochemistry.  $P = 0.05$ , comparison between the curves.

As determinants of aggressive behavior, genes overexpressed in poor prognosis tumors should be reviewed for potential therapeutic targets. Our expression profiling data identifying Aurora kinases as predictors of aggressive behavior in mesothelioma thus also highlight them as candidate therapeutic targets. Alterations in the spindle checkpoint and overexpression of Aurora kinase A have been related to resistance to chemotherapeutic agents, including taxanes (53, 54). Interestingly, taxanes are known to be ineffective in mesothelioma (55), and it is tempting to speculate that this clinical aspect of mesothelioma is related to its frequent overexpression of Aurora kinases. These considerations provide a rationale for testing a combination of a taxane and an Aurora kinase inhibitor in this cancer, as proposed for other similar settings (27, 56). Several small-molecule inhibitors of Aurora kinases A and B have recently been developed, as reviewed elsewhere (27, 57). A final point highlights the potential value of Aurora kinases as therapeutic targets in mesothelioma; that is, the observation that proliferating cells lacking functional p53 (would include most mesotheliomas because they lack p14<sup>ARF</sup> due to p16/CDKN2A deletion) may be especially sensitive to these inhibitors (27).

**Conclusions.** The results of prognostic prediction based on our microarray data set and the independent evaluation of similar published classifiers suggests an upper limit to the power of microarray data in this setting (~65%), below the level of clinical

usefulness. In our larger gene expression data set, we found the accuracy of published classifiers to be consistently lower than the estimates in the original publications. Moreover, microarray-based prognostic prediction was inferior to prognostic prediction based on standard clinicopathologic variables and *P16/CDKN2A* status. The very limited overlap between prognostic gene lists obtained by different groups emphasizes that such lists are “noisy” and that, absent independent validation, the significance of individual genes on these lists should be viewed with great caution. Nonetheless, supervised analyses of microarray data provide leads for new diagnostic markers and potential therapeutic targets, such as aurora kinases in the present study. Collectively, our predictive genes suggest cell cycle deregulation and aneuploidy as key determinants of aggressive behavior in mesothelioma. Gene expression profiling in mesothelioma is likely to remain a discovery tool rather than becoming a clinical testing platform.

## Acknowledgments

Received 11/1/2005; revised 1/6/2006; accepted 1/19/2006.

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.

We thank Agnes Viale, Irina Linkov, and the staffs of the Genomic Core Laboratory and of the Core Research IHC Laboratory for their help; Eleftherios P. Diamandis for providing the KLK11 polyclonal antibody; Masashi Kimura for providing the Aurora kinase A polyclonal antibody; and Eva Hernando (MSKCC) for helpful discussions.

## References

- Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005;23:7332-41.
- 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.
- Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;365:488-92.
- Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol* 2005;23:7350-60.
- Weder W, Kestenholtz P, Taverna C, et al. Neoadjuvant chemotherapy followed by extrapleural pneumonectomy in malignant pleural mesothelioma. *J Clin Oncol* 2004;22:3451-7.
- Rusch VW. Pemetrexed and cisplatin for malignant pleural mesothelioma: a new standard of care? *J Clin Oncol* 2003;21:2629-30.
- Berghmans T, Paesmans M, Lalami Y, et al. Activity of chemotherapy and immunotherapy on malignant mesothelioma: a systematic review of the literature with meta-analysis. *Lung Cancer* 2002;38:111-21.
- Steele JP. Prognostic factors in mesothelioma. *Semin Oncol* 2002;29:36-40.
- Bueno R, Reblando J, Glickman J, Jaklitsch MT, Lukanich JM, Sugarbaker DJ. Pleural biopsy: a reliable method for determining the diagnosis but not subtype in mesothelioma. *Ann Thorac Surg* 2004;78:1774-6.
- Gordon GJ, Rockwell GN, Jensen RV, et al. Identification of novel candidate oncogenes and tumor suppressors in malignant pleural mesothelioma using large-scale transcriptional profiling. *Am J Pathol* 2005;166:1827-40.
- Kettunen E, Nicholson AG, Nagy B, et al. L1CAM, INP10, P-cadherin, tPA and ITGB4 over-expression in malignant pleural mesotheliomas revealed by combined use of cDNA and tissue microarray. *Carcinogenesis* 2005;26:17-25.
- Hoang CD, D'Cunha J, Kratzke MG, et al. Gene expression profiling identifies matriptase overexpression in malignant mesothelioma. *Chest* 2004;125:1843-52.
- Pass HI, Liu Z, Wali A, et al. Gene expression profiles predict survival and progression of pleural mesothelioma. *Clin Cancer Res* 2004;10:849-59.
- Singhal S, Wiewrodt R, Malden LD, et al. Gene expression profiling of malignant mesothelioma. *Clin Cancer Res* 2003;9:3080-97.
- Gordon GJ, Jensen RV, Hsiao LL, et al. Using gene expression ratios to predict outcome among patients with mesothelioma. *J Natl Cancer Inst* 2003;95:598-605.
- Gordon GJ, Rockwell GN, Godfrey PA, et al. Validation of genomics-based prognostic tests in malignant pleural mesothelioma. *Clin Cancer Res* 2005;11:4406-14.
- Illei PB, Rusch VW, Zakowski MF, Ladanyi M. Homozygous deletion of CDKN2A and co-deletion of the methylthioadenosine phosphorylase gene in the majority of pleural mesotheliomas. *Clin Cancer Res* 2003;9:2108-13.
- Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;2:249-64.
- Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003;31:e15.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Soc Stat Soc B* 1995;57:289-300.
- Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000;16:906-14.
- 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.
- Kaufmann O, Volmerer J, Dietel M. Uroplakin III is a highly specific and moderately sensitive immunohistochemical marker for primary and metastatic urothelial carcinomas. *Am J Clin Pathol* 2000;113:683-7.
- Logani S, Oliva E, Amin MB, Folpe AL, Cohen C, Young RH. Immunoprofile of ovarian tumors with putative transitional cell urothelial differentiation using novel urothelial markers: histogenetic and diagnostic implications. *Am J Surg Pathol* 2003;27:1434-41.
- Diamandis EP, Okui A, Mitsui S, et al. Human kallikrein 11: a new biomarker of prostate and ovarian carcinoma. *Cancer Res* 2002;62:295-300.
- Tanaka T, Kimura M, Matsunaga K, Fukada D, Mori H, Okano Y. Centrosomal kinase AIK1 is overexpressed in invasive ductal carcinoma of the breast. *Cancer Res* 1999;59:2041-4.
- Keen N, Taylor S. Aurora-kinase inhibitors as anticancer agents. *Nat Rev Cancer* 2004;4:927-36.
- Chieffi P, Troncone G, Caleo A, et al. Aurora B expression in normal testis and seminomas. *J Endocrinol* 2004;181:263-70.
- Borcuzak AC, Taub RN, Hesdorffer M, et al. P16 loss and mitotic activity predict poor survival in patients with peritoneal malignant mesothelioma. *Clin Cancer Res* 2005;11:3303-8.
- Moll R, Wu XR, Lin JH, Sun TT. Uroplakins, specific membrane proteins of urothelial umbrella cells, as histological markers of metastatic transitional cell carcinomas. *Am J Pathol* 1995;147:1383-97.
- Borgono CA, Diamandis EP. The emerging roles of human tissue kallikreins in cancer. *Nat Rev Cancer* 2004;4:876-90.
- Xia C, Xu Z, Yuan X, et al. Induction of apoptosis in mesothelioma cells by antisurvivin oligonucleotides. *Mol Cancer Ther* 2002;1:687-94.
- Falleni M, Pellegrini C, Marchetti A, et al. Quantitative evaluation of the apoptosis regulating genes Survivin, Bcl-2 and Bax in inflammatory and malignant pleural lesions. *Lung Cancer* 2005;48:211-6.
- Tirone F. The gene PC3 (TIS21/BTG2), prototype member of the PC3/BTG/TOB family: regulator in control of cell growth, differentiation, and DNA repair? *J Cell Physiol* 2001;187:155-65.
- Struckmann K, Schraml P, Simon R, et al. Impaired expression of the cell cycle regulator BTG2 is common in clear cell renal cell carcinoma. *Cancer Res* 2004;64:1632-8.

36. Ficazzola MA, Fraiman M, Gitlin J, et al. Antiproliferative B cell translocation gene 2 protein is down-regulated post-transcriptionally as an early event in prostate carcinogenesis. *Carcinogenesis* 2001;22:1271-9.
37. Kristjansdottir K, Rudolph J. Cdc25 phosphatases and cancer. *Chem Biol* 2004;11:1043-51.
38. Carmena M, Earnshaw WC. The cellular geography of aurora kinases. *Nat Rev Mol Cell Biol* 2003;4:842-54.
39. Giet R, Petretti C, Prigent C. Aurora kinases, aneuploidy and cancer, a coincidence or a real link? *Trends Cell Biol* 2005;15:241-50.
40. Marumoto T, Zhang D, Saya H. Aurora-A: a guardian of poles. *Nat Rev Cancer* 2005;5:42-50.
41. Tatsuka M, Katayama H, Ota T, et al. Multinuclearity and increased ploidy caused by overexpression of the aurora- and Ipl1-like midbody-associated protein mitotic kinase in human cancer cells. *Cancer Res* 1998;58:4811-6.
42. Zhang D, Hirota T, Marumoto T, et al. Cre-loxP-controlled periodic Aurora-A overexpression induces mitotic abnormalities and hyperplasia in mammary glands of mouse models. *Oncogene* 2004;23:8720-30.
43. Castedo M, Perfettini JL, Roumier T, Andreau K, Medema R, Kroemer G. Cell death by mitotic catastrophe: a molecular definition. *Oncogene* 2004;23:2825-37.
44. Jeng YM, Peng SY, Lin CY, Hsu HC. Overexpression and amplification of Aurora-A in hepatocellular carcinoma. *Clin Cancer Res* 2004;10:2065-71.
45. Sen S, Zhou H, Zhang RD, et al. Amplification/overexpression of a mitotic kinase gene in human bladder cancer. *J Natl Cancer Inst* 2002;94:1320-9.
46. Tong T, Zhong Y, Kong J, et al. Overexpression of Aurora-A contributes to malignant development of human esophageal squamous cell carcinoma. *Clin Cancer Res* 2004;10:7304-10.
47. Katayama H, Ota T, Jisaki F, et al. Mitotic kinase expression and colorectal cancer progression. *J Natl Cancer Inst* 1999;91:1160-2.
48. Altieri DC. Validating survivin as a cancer therapeutic target. *Nat Rev Cancer* 2003;3:46-54.
49. Mazumdar M, Sundareshan S, Misteli T. Human chromokinesin KIF4A functions in chromosome condensation and segregation. *J Cell Biol* 2004;166:613-20.
50. Zou H, McGarry TJ, Bernal T, Kirschner MW. Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* 1999;285:418-22.
51. Christopoulou L, Moore JD, Tyler-Smith C. Overexpression of wild-type Securin leads to aneuploidy in human cells. *Cancer Lett* 2003;202:213-8.
52. Ogbagabriel S, Fernando M, Waldman FM, Bose S, Heaney AP. Securin is overexpressed in breast cancer. *Mod Pathol* 2005;18:985-90.
53. Anand S, Penrhyn-Lowe S, Venkitaraman AR. AUORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. *Cancer Cell* 2003;3:51-62.
54. Sudo T, Nitta M, Saya H, Ueno NT. Dependence of paclitaxel sensitivity on a functional spindle assembly checkpoint. *Cancer Res* 2004;64:2502-8.
55. van Meerbeeck J, Debruyne C, van Zandwijk N, et al. Paclitaxel for malignant pleural mesothelioma: a phase II study of the EORTC Lung Cancer Cooperative Group. *Br J Cancer* 1996;74:961-3.
56. Hata T, Furukawa T, Sunamura M, et al. RNA interference targeting aurora kinase suppresses tumor growth and enhances the taxane chemosensitivity in human pancreatic cancer cells. *Cancer Res* 2005;65:2899-905.
57. Andrews PD. Aurora kinases: shining lights on the therapeutic horizon? *Oncogene* 2005;24:5005-15.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Global Gene Expression Profiling of Pleural Mesotheliomas: Overexpression of Aurora Kinases and *P16/CDKN2A* Deletion as Prognostic Factors and Critical Evaluation of Microarray-Based Prognostic Prediction

Fernando López-Ríos, Shannon Chuai, Raja Flores, et al.

*Cancer Res* 2006;66:2970-2979.

<b>Updated version</b>	Access the most recent version of this article at: <a href="http://cancerres.aacrjournals.org/content/66/6/2970">http://cancerres.aacrjournals.org/content/66/6/2970</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://cancerres.aacrjournals.org/content/suppl/2009/09/09/66.6.2970.DC1">http://cancerres.aacrjournals.org/content/suppl/2009/09/09/66.6.2970.DC1</a>

<b>Cited articles</b>	This article cites 57 articles, 30 of which you can access for free at: <a href="http://cancerres.aacrjournals.org/content/66/6/2970.full#ref-list-1">http://cancerres.aacrjournals.org/content/66/6/2970.full#ref-list-1</a>
<b>Citing articles</b>	This article has been cited by 20 HighWire-hosted articles. Access the articles at: <a href="http://cancerres.aacrjournals.org/content/66/6/2970.full#related-urls">http://cancerres.aacrjournals.org/content/66/6/2970.full#related-urls</a>

<b>E-mail alerts</b>	<a href="#">Sign up to receive free email-alerts</a> related to this article or journal.
<b>Reprints and Subscriptions</b>	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .
<b>Permissions</b>	To request permission to re-use all or part of this article, contact the AACR Publications Department at <a href="mailto:permissions@aacr.org">permissions@aacr.org</a> .