Prognostic Significance of Gene Expression Profiles of Metastatic Neuroblastomas Lacking MYCN Gene Amplification

Shahab Asgharzadeh, Roger Pique-Regi, Richard Sposto, Hong Wang, Yujun Yang, Hiroyuki Shimada, Katherine Matthay, Jonathan Buckley, Antonio Ortega, Robert C. Seeger

Background: The aggressiveness of metastatic neuroblastomas that lack MYCN gene amplification varies with age—they are least aggressive when diagnosed in patients younger than 12 months and most aggressive when diagnosed in patients older than 24 months. However, age at diagnosis is not always associated with patient survival. We examined whether molecular classification of metastatic neuroblastomas without MYCN gene amplification at diagnosis using gene expression profiling could improve the prediction of risk of disease progression. Methods: We used Affymetrix microarrays to determine the gene expression profiles of 102 untreated primary neuroblastomas without MYCN gene amplification obtained from children whose ages at diagnosis ranged from 0.1 to 151 months. A supervised method using diagonal linear discriminant analysis was devised to build a multigene model for predicting risk of disease progression. The accuracy of the model was evaluated using nested cross-validations, permutation analyses, and gene expression data from 15 additional tumors obtained at disease progression. Results: An expression profile model using 55 genes defined a tumor signature that distinguished two groups of patients from among those older than 12 months at diagnosis and clinically classified as having high-risk disease, those with a progression-free survival (PFS) rate of 16% (95% confidence interval [CI] = 8% to 28%), and those with a PFS rate of 79% (95% CI = 57% to 91%) (P<.01). These tumor signatures also identified two groups of patients with PFS of 15% (95% CI = 7% to 27%) and 69% (95% CI = 40% to 86%) (P<.01) from among patients who were older than 18 months at diagnosis. The gene expression signature of untreated molecular high-risk tumors was also present in progressively growing tumors. Conclusion: Gene expression signatures of tumors obtained at diagnosis from patients with clinically indistinguishable high-risk, metastatic neuroblastomas identify subgroups with different outcomes. Accurate identification of these subgroups with gene expression profiles may facilitate development, implementation, and analysis of clinical trials aimed at improving outcome. [J Natl Cancer Inst 2006;98:1193–203]
with favorable outcomes in patients with tumors that lack MYCN gene amplification who were 12–24 months of age at diagnosis but not in patients older than 24 months. Recently, unbalanced chromosome 11q loss of heterozygosity was shown to be statistically significantly associated with a poor outcome in patients with intermediate-risk disease (13). Several epigenetic (14,15) and gene expression profiling (16–18) studies of neuroblastoma have also reported associations with patient outcomes. However, these latter studies (16–18) analyzed groups of patients who were heterogeneous with respect to clinical stage, MYCN gene amplification status, and therapy, and they included only a small number of high-risk patients whose tumors lacked MYCN gene amplification.

The goal of this study was to use tumor gene expression profiles (signatures) to develop an algorithm for predicting the outcomes of children who were classified clinically as having high-risk disease because they were diagnosed with metastatic neuroblastoma that lacked MYCN gene amplification. To date, no clinical or biologic markers have been associated with outcome for this group of neuroblastoma patients, especially for those who are older than 18 months at diagnosis. Accurate molecular classification of these tumors could be used to identify patients with true high-risk disease (e.g., those with a progression-free survival [PFS] rate of 15% or lower), who might benefit from new and possibly more effective therapeutic strategies. Conversely, patients with low-risk disease (e.g., those with PFS rate of >70%) may be successfully treated with strategies that also aim to reduce therapy-related toxicities. Our study is unique because we focused on a relatively homogeneous group of patients by controlling for MYCN gene amplification status and clinical stage. We hypothesized that the gene expression signatures of tumors from patients who were younger than 12 months at diagnosis (and who thus have generally excellent outcomes) could be used to help identify the least aggressive tumors in patients who were 12 months or older at diagnosis. We used the gene expression signatures of untreated primary neuroblastoma tumors to build a multigene model for defining risk groups that predict outcome as defined by the presence or absence of disease progression.

### Subjects and Methods

#### Study Population

All neuroblastoma patients included in this study (102 patients whose tumors were obtained at time of diagnosis and 15 patients whose tumors were obtained at the time of progression) had metastatic tumors that lacked MYCN gene amplification. There were three patients from whom we obtained tumor samples at diagnosis and at relapse. All patients, at the time of diagnosis, were enrolled in one of the following CCG studies: CCG-3881, CCG-323P, CCG-321-P2, CCG-321-P3, or CCG-3891 (9). CCG risk stratification for this group of patients, which is based solely on the patient’s age at diagnosis, dictated the type of therapy the patients received. Patients younger than 12 months at diagnosis, who were classified as the CCG intermediate-risk group and enrolled in CCG-3881, received multiagent chemotherapy. Patients who were 12 months or older at diagnosis (the CCG high-risk group) were enrolled in CCG pilot studies from 1984 through 1994 (CCG-323P, CCG-321-P2, CCG-321-P3) or in the group-wide study from 1991 through 1996 (CCG-3891). CCG-3891 used the same induction regimen as the pilot studies (CCG-321-P2 and CCG-321-P3) followed by consolidation therapy in which patients who randomly assigned to receive either myeloablative chemoradiotherapy and autologous bone marrow transplantation or chemotherapy only. After completion of consolidation therapy, CCG-3891 patients were further randomly assigned to receive 13-cis-retinoic acid or no further therapy (9).

The median age at diagnosis of the 102 patients whose tumors were obtained at time of diagnosis and used to generate our model was 34.7 months (range = 0.1–151 months), and the median length of follow-up among the 56 (55%) patients who did not experience disease progression was 7.8 years (range = 1.6–12.8 years). The 15 patients whose tumors were obtained at progression (i.e., the test set) were older than 24 months at diagnosis and classified as high-risk patients. Detailed clinical information for the 102 patients used in generating the model are provided in Table 1.

All patients in this study were also reclassified according to their risk of progression using recently revised COG risk.

### Table 1. Characteristics of patients with metastatic neuroblastoma lacking MYCN gene amplification*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>&lt;12 mo (n = 28)</th>
<th>12–24 mo (n = 20)</th>
<th>&gt;24 mo (n = 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age at diagnosis, mo (range)</td>
<td>7.0 (0.1–11.8)</td>
<td>16.7 (12.3–23.7)</td>
<td>55.7 (25–151)</td>
</tr>
<tr>
<td>No. of females (%)</td>
<td>15 (54)</td>
<td>11 (55)</td>
<td>21 (39)</td>
</tr>
<tr>
<td>CCG risk stratification</td>
<td>Intermediate</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>INPC status, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favorable</td>
<td>27 (96)</td>
<td>8 (40)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Unfavorable</td>
<td>1 (4)</td>
<td>12 (60)</td>
<td>52 (96)</td>
</tr>
<tr>
<td>CCG study enrollment, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>323P</td>
<td>1 (4)</td>
<td>1 (5)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>321-P2</td>
<td>0</td>
<td>3 (15)</td>
<td>5 (9)</td>
</tr>
<tr>
<td>321-P3</td>
<td>0</td>
<td>2 (10)</td>
<td>5 (9)</td>
</tr>
<tr>
<td>3881</td>
<td>27 (96)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3891</td>
<td>0</td>
<td>14 (70)</td>
<td>43 (80)</td>
</tr>
<tr>
<td>Consolidation therapy received, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy only</td>
<td>28 (100)</td>
<td>9 (45)</td>
<td>22 (41)</td>
</tr>
<tr>
<td>Autologous bone marrow transplant</td>
<td>0</td>
<td>6 (30)</td>
<td>17 (31)</td>
</tr>
<tr>
<td>Progression of disease, No. (%)</td>
<td>0</td>
<td>8 (40)</td>
<td>38 (70)</td>
</tr>
<tr>
<td>Rapid progression of disease, No. (%)</td>
<td>0</td>
<td>5 (25)</td>
<td>15 (28)</td>
</tr>
</tbody>
</table>

*COG = Children’s Oncology Group; INPC = International Neuroblastoma Pathology Classification; CCG = Children’s Cancer Group.

†Defined as progression of disease within 8 months of diagnosis (before completion of consolidation therapy).

Downloaded from https://academic.oup.com/jnci/article-abstract/98/17/1193/2521741 by guest on 26 January 2019
stratification criteria to facilitate comparison of the accuracy of clinically based outcome prediction to the proposed molecular risk classifier. The revised COG risk stratification scheme expands the classification of the intermediate-risk group to include patients who are 12–18 months old at diagnosis and whose tumors have favorable pathology and a hyperdiploid genome. Although we did not have information about tumor ploidy for the patients in our data set, we assigned patients who were 12–18 months old at diagnosis and whose tumors had favorable pathology to the COG intermediate-risk group. All of these patients are progression-free survivors, and thus, reassignment did not contribute to any classification errors for the COG classifier.

Tumor tissues obtained at diagnosis or at disease progression were snap frozen and shipped to the Neuroblastoma Biology Reference Laboratory at Childrens Hospital Los Angeles (Los Angeles, CA) for storage at −80 °C. All samples were evaluated histologically to confirm the diagnosis and to assess features used in the INPC system, including neuroblastic differentiation and the mitosis-karyorrhexis index, to classify tumor histology as favorable or unfavorable (Table 1). Written informed consent was obtained from patients’ parents or guardians in accordance with institutional review board policies and procedures for research dealing with tumor specimens and clinical information. The institutional review board at Children’s Hospital Los Angeles approved the study.

**Gene Expression Profiling**

Total RNA was isolated from frozen tumor sections (15–30 sections, 20 μm thick) using TRIzol reagent (Invitrogen, Carlsbad, CA), and its quality was assessed by gel electrophoresis. Twenty micrograms of total RNA from each sample was reverse transcribed into complementary DNA followed by in vitro transcription and labeling to produce biotinylated complementary RNA using an IVT labeling kit (Affymetrix, Santa Clara, CA) according to the manufacturer’s protocol. The labeled complementary RNAs were hybridized to U133A and U133B gene chip microarrays (Affymetrix) according to the manufacturer’s protocol, and the chips were scanned using a GeneChip Scanner 3000 (Affymetrix) to obtain fluorescence intensity values for oligonucleotide probe sets (11 probes per set) representing gene transcripts. Together, the U133A and U133B gene chips contain 45,000 probe sets, representing approximately 33,000 genes. The probe sets’ fluorescence intensity values were normalized using the invariant probe set method and modeled using the perfect match minus mismatch algorithm as implemented in the dChip program (19). The gene expression values obtained from these probe sets were log(base 10) transformed before the analyses. In this report, the term “gene” is used to denote one probe set.

**Statistical Analysis**

The log-transformed fluorescence intensity values were imported into GenePix (Epicenter Software Inc, Los Angeles, CA) and Matlab programs (version 7.0; The Mathworks Inc, Natick, MA) for further analysis. Kaplan–Meier survival analyses were performed using STATA software (version 8.0; StataCorp LP, College Station, TX). Hierarchical clustering of tumor samples and gene expression values was performed using the average linkage method and Euclidean distance measurements as implemented in the dChip program (19). We used two databases, Gene Ontology (GO; http://www.geneontology.org) and the Broad/MIT Molecular Signature Database (http://www.broad.mit.edu/gsea), to identify groupings of genes. The GO database groups genes on the basis of their associated biologic processes, cellular components, and molecular functions. The Broad/MIT Molecular Signature Database groups genes on the basis of 475 metabolic and signaling pathways assembled from 10 publicly available, manually curated databases, including Signal Transduction Knowledge Environment (http://stke.sciencemag.org), BioCarta (http://www.biocarta.com), and Gene arrays (GEArray) BioScience corporation (http://www.superarray.com). We analyzed each gene group and selected those that have more genes differentially expressed among the classes (molecular risk groups) than expected by chance. This analysis was carried out with the use of the Gene Set Expression Comparison tool in BRB Array Tools (version 3.4; developed by Dr Richard Simon and Amy Peng Lam, http://linus.nci.nih.gov/~brb/tool.htm). The statistical tests used in this tool include functional class score and Kolmogorov–Smirnov permutation tests. Gene sets from the GO groupings or the Broad/MIT Molecular Signature Database with P values less than .001 in either test were considered statistically significant. The microarray data set is accessible through Gene Expression Omnibus (GEO accession number GSE3446; http://www.ncbi.nlm.nih.gov/geo/).

**Unsupervised and supervised analyses.** Principal component analysis was used to identify patterns in gene expression levels by calculating major directions of gene expression variability on each sample’s data without regard to any clinical or biologic features. The scatter diagram of the first two principal components represented an unsupervised analysis method, whereby samples with similar principal components cluster together, as shown in Fig. 1, B. These groupings were used to identify associations between global gene expression of their clinical or biologic features in an unsupervised fashion.

We then performed a supervised analysis using the schema shown in Fig. 2, A. A modified t test that was similar to other methods described previously (19) was used to examine the statistical significance of differentially expressed genes on the basis of whether disease progression had or had not occurred (20). All probe sets were used in the analysis without any preselection or filtering outside the validation loop to avoid biasing the results. Prediction analysis was carried out using diagonal linear discriminant analysis (21). The number of genes used for the diagonal linear discriminant analysis model was determined by minimizing the cross-validation prediction error. A sequence of models was built by successively adding the next highest scoring gene as determined by the absolute value of the modified t test. Each model was evaluated using 100 repetitions of a 10-fold cross-validation. The training and test sets generated during each cross-validation had the same ratio of disease-free patients to those with progressive disease as that of the entire group of 102 patients. The number of genes that minimized the average cross-validation error rate was determined, and a final model comprising that number of highest scoring genes was computed using the entire data set. The ranking of genes in this final model was based on the absolute value of the modified t test. Molecular risk classification was based on the diagonal linear discriminant analysis class prediction posterior probability. The molecular low-risk group was defined as having a class prediction probability of less than .5, and the molecular high-risk group was defined as having a class prediction probability greater than .5. To ensure that the multigene model with the lowest classification error rate was not due to chance, we applied the procedure described...
random class permutation assignments of the original data set.

An external outer nest of leave-one-out cross-validation independent from the internal nest (i.e., the iterative 10-fold cross-validation used to choose the model) was performed to obtain cross-validated molecular risk assignments used to compute the misclassification error estimate. These risk assignments were also used for outcome analyses and for presenting prediction statistics. The precision of the cross-validated prediction error was assessed by performing the full cycle of the 100 × 10–fold inner loop cross-validation followed by a leave-one-out outer loop cross-validation on 200 sample sets each with 102 samples bootstrapped from the original data set (22). A detailed description of our model selection and evaluation strategy and a comparison of the model to other prediction algorithms, including support vector machines and nearest shrunken centroids (20), are included in the Supplementary Data (available at http://jncicancerspectrum.oxfordjournals.org/jnci/content/vol98/issue17).

Comparison of molecular model to CCG and COG risk classification. Risk classification using the molecular model was compared with clinically based CCG and COG risk classifications to determine if the molecular model improved prediction of PFS. We also evaluated hybrid classifiers that combined CCG or COG risk classification and information from the molecular model. The hybrid classifications were not used to change the assignment of patients who were classified as intermediate risk by either CCG or COG criteria, but patients who were considered to be at high risk by CCG or COG criteria were reclassified to a hybrid intermediate-risk group if the tumor exhibited a low-risk gene expression signature (i.e., clinical high risk but molecular low risk). The false-positive, false-negative, and total misclassification errors were computed for each of these classifiers, as was the absolute decrease in total errors compared with the clinical classifiers. P values were determined by comparing the observed error rates to the null permutation distribution of these quantities.

A permutation experiment was performed to compute P values appropriate for the log-rank test and for the test of the predictive ability of the proposed classifier. Specifically, 100 random permutations of the molecular data against the clinical annotation and outcome data were created, and the full cycle of the 100 × 10–fold inner loop cross-validation followed by a leave-one-out outer loop cross-validation was performed to generate 100 permutation-based molecular risk group assignment sets. P values presented for the unstratified and stratified log-rank analyses were in reference to the permutation distribution generated by performing these analyses for each of these assignment sets (log-rank P values based on a maximum likelihood gamma distribution estimate of the null permutation distribution is included in the Supplementary Data; available at http://jncicancerspectrum.oxfordjournals.org/jnci/content/vol98/issue17). The number of misclassification errors for the proposed molecular classifier alone or in conjunction with known clinical and biologic factors was computed for each of these assignment sets in a similar fashion, resulting in the permutation distribution from which P values were determined.

Outcome analysis. The primary endpoint for outcome analysis was disease progression, which was defined a priori according to the CCG-3891 protocol (9) as 1) the development of any new lesion, 2) a greater than 25% increase in the mass of any measurable tumor, or 3) a previously negative bone marrow sample that was positive for tumor cells. The classification model was based simply on whether or not the patient had experienced disease progression according to this definition. The effective period for risk of recurrence in our patient population was 4 years from diagnosis. Because few patients were censored before this effective period for risk of recurrence, ignoring that censoring had little practical effect on the results. Analyses of PFS defined the time to progression from the date of diagnosis and censored progression-free survivors at the time of last follow-up. One patient who died 5.5 years after diagnosis without experiencing disease progression was also censored. The Kaplan–Meier (product-limit) method was used to compute PFS probabilities and to produce survival curves. The reported probabilities were based on 5-year PFS. The log-rank statistic, with stratification when appropriate, was used to test the statistical significance of differences in PFS between prognostic groups (23), with P values determined as described above. In the log-rank analysis, age groups were defined as 12–18 months old at diagnosis and older than 18 months.

![Fig. 1. Progression-free survival (PFS) of patients with metastatic neuroblastoma lacking MYCN gene amplification by age at diagnosis and tumor clustering based on gene expression. A) PFS of 102 patients stratified by age at diagnosis. B) Unsupervised principal component analysis based on global gene expression profiles (45,000 probe sets) of tumors obtained at diagnosis from the same 102 patients. In the scatter diagram of the first two principal components, symbols indicate age groups at diagnosis and colors denote disease progression status of patients (blue = disease free; red = disease progression).](https://academic.oup.com/jnci/article-abstract/98/17/1193/2521741)
RESPECTS

To identify molecular biologic differences among metastatic neuroblastomas that lack MYCN gene amplification, we used an Affymetrix gene microarray platform to obtain the gene expression profiles of primary tumors from 102 patients diagnosed with metastatic disease. The patients were enrolled in CCG trials from 1984 through 1996, and the therapy they received in those trials was based on risk categories that were defined according to age at diagnosis, pathology classification, and MYCN gene amplification status (Table 1). Our estimates of PFS for these 102 patients according to age at diagnosis (Fig. 1, A) were similar to published estimates of PFS for other patients in these age groups (2). Among the patients who were 12 months or older at diagnosis to facilitate comparisons with the current COG risk stratification scheme.

RESULTS

To identify molecular biologic differences among metastatic neuroblastomas that lack MYCN gene amplification, we used an Affymetrix gene microarray platform to obtain the gene expression profiles of primary tumors from 102 patients diagnosed with metastatic disease. The patients were enrolled in CCG trials from 1984 through 1996, and the therapy they received in those trials was based on risk categories that were defined according to age at diagnosis, pathology classification, and MYCN gene amplification status (Table 1). Our estimates of PFS for these 102 patients according to age at diagnosis (Fig. 1, A) were similar to published estimates of PFS for other patients in these age groups (2). Among the patients who were 12 months or older at diagnosis to facilitate comparisons with the current COG risk stratification scheme.

Construction of a Gene Expression–Based Outcome Predictor

We first examined the global gene expression profiles of each tumor in the entire data set using the first two principal components and found that the 102 tumors clustered by patient age at diagnosis and by disease progression status (Fig. 1, B). Unsupervised principal component analysis based on global gene expression profiles revealed that tumors from patients who were 12 months or older at diagnosis who did not experience disease progression tended to cluster with those of patients who were younger than 12 months at diagnosis (and typically unlikely to experience disease progression), suggesting that the overall gene
expression pattern among these two groups of tumors was similar and independent of age at diagnosis.

We next devised a supervised analysis strategy to identify the minimum number of genes that could accurately predict whether a patient would experience disease progression (Fig. 2, A). The analysis used cross-validation techniques to identify the smallest number of genes with the greatest variability in expression levels between tumors of patients who had progression of disease versus those without progression that should be included in the model to minimize risk classification errors (17). We used diagonal linear discriminant analysis as a prediction algorithm because it has been shown to have good performance for microarray data (21). We found that the average cross-validation error rate using diagonal linear discriminant analysis was first minimized for multigene models that used probes for 55 genes (52 unique genes) (Fig. 2, B). We further performed the nested leave-one-out cross-validation of the entire model-building procedure shown in Fig. 2, A and used the derived gene expression model to assign each patient a probability of survival. The molecular low- and high-risk groups were designated based on patient's probability of survival using the top 55 genes in the model. The leave-one-out cross-validation misclassification rate estimate (±permutation standard error) obtained from our nested algorithm used to predict progression status was 15.7% (±3.95%) for all patients and 17.9% (±5.09%) for patients older than 12 months at diagnosis. The misclassification rate of 9.8% using the 55 genes without performing cross-validation is probably an underestimate and reflects the need for cross-validation (24).

To examine whether the 55-gene model that minimized the classification error rate could have occurred by chance, we created 1000 permuted sample sets in which each of the 102 samples within a set was randomly mislabeled with respect to progression status. Each sample set was then used to rerun the model and generate 1000 multigene models. The misclassification rates generated from the models in these permuted sample sets were statistically significantly higher than those generated from our correctly labeled sample set (P<.01).

Molecular Classification of Risk of Progression

Figure 3 shows the Kaplan–Meier estimates of PFS for the patients classified as molecular low and high risks, as defined in the nested cross-validation classification, along with the distribution of patients by age at diagnosis for each risk category. Fifty-two patients were classified as molecular high risk and 50 patients were categorized as molecular low risk (Fig. 3, A). Among the 74 patients who were older than 12 months at diagnosis and treated on high-risk CCG protocols, 25 (34%) were classified as molecular low risk and 49 (66%) were assigned to the molecular high-risk group and had a PFS rate of 16% (95% confidence interval [CI] = 57% to 91%) and 49 (66%) were assigned to the molecular high-risk group and had a PFS rate of 16% (95% CI = 5% to 28%) (P<.01, permutation-based stratified log-rank test) (Fig. 3, B). All patients who were classified as molecular low risk and were 12–24 months old at diagnosis (n = 11) were progression free at a median follow-up of 7.2 years (range = 2.2–12.9 years). For patients who had completed consolidation therapy, there was no difference between the two molecular risk groups with respect to the type of consolidation therapy. Of the 20 patients who developed progressive disease within 8 months of diagnosis and did not complete consolidation therapy, 19 (95%) were categorized as molecular high risk (Fig. 3, B). Examination of tumors from patients who were older than 18 months at diagnosis and for whom no clinical or biologic markers of outcome have previously been identified revealed that they too could be classified into
nullifying MYCN gene amplification. Children’s Cancer Group (CCG) risk classification as being at high risk for disease progression according to the Journal of the National Cancer Institute, Vol. 98, No. 17, September 6, 2006 ARTICLES 1199

molecular low-risk and high-risk groups, with PFS rates of 69% (95% CI = 40% to 86%) and 15% (95% CI = 7% to 27%), respectively ($P < .01$, permutation-based stratified log-rank; Fig. 3, C).

Molecular risk classification was a statistically significant predictor of PFS independent of age at diagnosis and tumor histology (according to INPC) ($P < .01$, permutation-based stratified log-rank test). For each of the analyses shown in Fig. 3, the observed log-rank statistic was 1.5–2 times greater than the maximum observed value from each of the 100 random permutations described in the “Subjects and Methods” section.

We next compared risk classification error rate (i.e., predictive accuracy) of the molecular risk classifier to that of the two clinical risk classifiers (CCG and COG) and to that of two hybrid risk classifiers that combined clinical and molecular data (CCG-M and COG-M) (Fig. 4). The permutation-based $P$ values for the reduction in total errors after using or including the molecular classifier are also shown. Use of the molecular risk classifier alone or in combination with either of the clinical risk classifiers was associated with a statistically significant decrease in the number of total risk classification errors compared with either clinical risk classifier alone (all permutation-based $P$ values for comparisons <.01). Moreover, the number of false-positive errors was substantially reduced, and the number of false-negative errors increased.

Molecular Risk Classification of Relapsed Tumors

We analyzed an additional set of tumors obtained from 15 patients at the time of disease progression as part of strategy to validate the molecular risk classifier. These 15 patients were all older than 24 months when they were initially diagnosed with metastatic tumors that lacked MYCN gene amplification. Application of the 55-gene model predicted that 14 of the 15 tumors were in the molecular high-risk group (error rate = 6.7%; 95% CI = 0.2% to 32%). Hierarchical clustering using data from the 55-gene model revealed that the tumors obtained at relapse had gene expression profiles that were similar to those of tumors obtained at initial diagnosis and classified as molecular high risk (Fig. 5). Three of the 15 tumors obtained at disease progression were from patients from whom we had obtained tumor specimens at diagnosis and who were included in the training set. All tumors obtained from these three patients at diagnosis and progression were classified as molecular high risk.

Gene Set Expression Comparison

Among the 55 genes included in our model, the gene encoding neurotrophic tyrosine kinase receptor 2 (NTRK2 or TrkB) was the most statistically significantly associated with the risk of progression ($P < .001$) (Fig. 5). NTRK2 and neurotrophic tyrosine kinase receptor 1 (NTRK1 or TrkA) are involved in neurotrophin-mediated pathways that play important roles in the survival and differentiation of neuroblastoma cells (25–28). The gene encoding brain-derived neurotrophic factor (BDNF), which is a ligand for NTRK2, was also among the 3% of genes that displayed the most variability in expression between tumors classified as molecular high versus low risk; however, this variability did not reach the statistical threshold to be included in our model. The 55-gene signature also included several genes worthy of further research based on either their chromosomal location, their known function (tumor suppressor), or their possible role in the microenvironment (e.g., calmodulin-binding transcription activator 1 [CAMTA1]; harakiri [HRK], a BCL2-interacting protein; and LOC391427, which encodes a protein similar to immunoglobulin kappa chain V-I region Walker precursor).

To gain insight into the biologic differences that may govern neuroblastoma tumor behavior, we also analyzed the two molecular risk groups of tumors for the differential expression of sets of genes that were categorized by the GO Consortium or according to the MIT/Broad Molecular Signature Database. Gene categories were considered to be statistically significantly associated with the molecular risk groups if the number of genes that were differentially expressed were greater than would be expected by chance for the category. Table 2 lists the GO categories and pathways that met this statistical significance threshold using two statistical tests. Genes involved in antigen binding (a GO category) were statistically significantly associated with the molecular high-risk classification in both tests ($P < .001$), and genes involved in two pathways (the differentiation pathway in PC12 cells and the Wnt signaling pathway) were statistically significantly associated with the molecular low-risk classification in both tests ($P < .001$).

DISCUSSION

We found that gene expression profiles of metastatic neuroblastomas that lack MYCN gene amplification identified two distinct groups of patients who were at low and high risk of disease progression. Importantly, we found that molecular categorization of neuroblastomas at diagnosis using a 55-gene model identified subgroups of patients with different outcomes, even...
though all of these patients would be classified according to current criteria as having high-risk disease. The estimated PFS rates for patients older than 12 months at diagnosis whose tumors were classified as molecular low risk and molecular high risk were 79% and 16%, respectively. Extensive cross-validation analyses demonstrated that our 55-gene model, which was derived from the gene expression profiles of the largest number of metastatic neuroblastomas without MYCN gene amplification tested to date, was not due to chance and had a misclassification error rate of only 17.9%. Several genes identified by our 55-gene model have been shown to play roles in determining neuroblastoma tumor behavior (i.e., NTRK1 and NTRK2). Thus, gene expression profiling reveals biologic markers that are differentially expressed in tumors from a group of patients with clinically indistinguishable disease at diagnosis. The 55-gene model derived in this study should improve prognostication for patients diagnosed with metastatic neuroblastoma and may help guide therapy.

We used four strategies to ensure the validity of our results. First, we used an iterative 10-fold cross-validation algorithm, which avoided the selection bias that is often introduced in microarray analyses (24). Second, we performed nested cross-validation, which provided a less biased estimate of the misclassification error rate and assigned posterior probabilities of survival to each patient. Third, we used random permutation analysis to show that the misclassification error rate for defining our multigene model could not be obtained by chance. Fourth, our findings gain credence from the confirmation that some of the genes included in the predictive model are known to be involved in neuroblastoma cell differentiation, survival, and proliferation in experimental models and are associated with distinct tumor behaviors in patients.

Accurate risk stratification is likely governed by the biology of neuroblastomas, whereas clinical variables, such as age at diagnosis, likely act as surrogate markers. The COG has recently expanded its criteria for classifying neuroblastoma patients as being at intermediate risk for disease progression to include all patients who are younger than 18 months at diagnosis and have a tumor that is hyperdiploid with favorable histology according to the INPC (29). This change was the result of a reanalysis of the prognostic importance of age at diagnosis, which indicated that any cutpoint for age at diagnosis between 15 and 19 months was statistically significantly associated with good prognosis among patients diagnosed with metastatic neuroblastoma without MYCN gene amplification (29). Our experiments demonstrate that
grouping tumors by their gene expression signature was generally associated with the patient’s age at diagnosis but that these biologic features distinguish clinical behavior more accurately than age or histologic features. Our analyses also reveal that patients who are older than 12 months at diagnosis can have tumors with gene expression profiles that suggest that they may have either a low or a high risk of disease progression. In general, expression profiles of tumors in the molecular low-risk group were similar to those of tumors diagnosed during the first year of life, whereas tumors with molecular high-risk features had expression profiles that were similar to those of tumors obtained at relapse. The predictive accuracy of this molecular risk classifier was independent of current standard prognostic factors. More importantly, both the molecular risk classifier and the hybrid molecular/cellular risk classifiers statistically significantly improved on CCG or COG risk stratification.

Our demonstration that tumors obtained at disease progression had molecular high-risk gene expression profiles provides evidence that our model is promising as a predictor of disease progression; however, the small number of samples in this unique positive test set limits the statistical power of this interpretation. The presence of the high-risk molecular signature in tumors at diagnosis and those obtained at progression does not contradict the assumption that there are biologic differences among them, but it does suggest that gene expression profiles can reflect important core genomic features that are present at diagnosis and that persist during relapse. Other examples of core tumor features that remain unaltered from diagnosis to relapse include the karyotypes of acute myeloid leukemia cells (30) and genomic alterations, such as the BCR–ABL fusion gene in leukemia or MYCN gene amplification in neuroblastoma (31,32). Core genomic alterations that are present both at diagnosis and at disease progression may confer survival advantages for tumors. Additional survival advantages could be provided by acquired alterations, including increased expression of the multidrug resistance–associated protein 1 (33) or mutations in the TP53 gene (34,35).

Several of the 55 genes included in our model are likely to have roles in determining neuroblastoma behavior. NTRK1 (TrkA) expression was greater in the molecular low-risk group than in the molecular high-risk group. This finding extends previous observations that NTRK1 is, in general, most highly expressed by neuroblastomas with favorable biologic, pathologic, and clinical features, to specifically include in this group the molecular low-risk subset of metastatic neuroblastomas that lack MYCN gene amplification (25,26). By contrast, NTRK2 (TrkB) and its ligand, BDNF, were most highly expressed in molecular high-risk tumors. Expression of NTRK2 and BDNF has been associated with increased survival of neuroblastoma cells and with unfavorable clinical and biologic features (27,28). The reported clinical association (27,28) was found in heterogeneous cohorts of patients with respect to extent of disease (clinical stage) and MYCN gene amplification status, whereas our data are the first to demonstrate that high expression of both TrkB and BDNF is associated with metastatic, MYCN gene nonamplified tumors with aggressive behavior.

Some genes that were most highly expressed by molecular low-risk tumors are potential suppressors of tumor growth. For example, molecular low-risk tumors had high expression of CAMTA1, which resides within the smallest region of overlap in the loss-of-heterozygosity region of chromosome 1p36 that is associated with aggressive forms of neuroblastomas (36–38). A recent study among a heterogeneous group of neuroblastoma patients showed that CAMTA1 gene expression was statistically significantly correlated with outcome but not with loss of...
heterozygosity at chromosome 1p (39). The gene for another potential tumor suppressor, HRK, an activator of apoptosis in neuronal cells that exhibits its effects through interactions with death repressor proteins Bcl-2 and Bcl-XL (40) and is silenced through epigenetic changes in colon and gastric cancers (41) displayed high expression in molecular low-risk tumors and was included in the 55-gene model.

The two pathways that were the most statistically significantly associated with the risk of progression in our study (PC12 cell differentiation and Wnt signaling) also reflect the balance between survival and differentiation that likely governs neuroblastoma behavior. Rat pheochromocytoma PC12 cells have characteristics of sympathoadrenal progenitors (42), similar to neuroblastomas. Our finding that genes involved in antigen binding were more highly expressed in high-risk neuroblastomas than in low-risk neuroblastomas is novel; one member of this group of genes (LOC391427, which encodes a protein similar to immunoglobulin kappa chain) was among those included in our 55-gene model. The antigen-binding gene expression signature was associated with expression of a larger set of genes representing an inflammatory gene expression signature observed among the molecular high-risk group lacking MYCN amplification (data not shown). However, we advise caution in interpreting results of these preliminary GO and pathway analyses because knowledge of the functions of and interactions among the genes that define their groupings is evolving and further investigation is required to elucidate their roles in tumor behavior.

Our prediction model using gene expression profiling of a relatively homogenous group of patients was able to define subgroups of patients with distinctly different risks of disease progression with statistically significant predictive accuracy. The internal validation strategy (outer leave-one-out cross-validation) provides an estimate of the misclassification error rate for future populations of patients with clinical and biologic characteristics similar to those of the patients in our study. However, the performance of this risk classifier needs to be externally validated in a larger, independent group of patients. Furthermore, external validation studies should also assess the benefits of therapy among patients who are categorized using these molecular predictors. In this context, it will be important to develop strategies for reducing therapy-related morbidity while maintaining efficacy in the molecular low-risk group and for improving outcome for the molecular high-risk patients. The design of such trials has been discussed elsewhere (43,44).

In conclusion, we have demonstrated that gene expression profiles of metastatic neuroblastomas that lack MYCN gene amplification provide new definitions of high and low risks of disease progression. Molecular risk classification at diagnosis may ultimately improve risk stratification for patients who are older than 12 months at diagnosis and is especially important for patients who are older than 18 months at diagnosis because their outcomes currently cannot be estimated by any other means. Current therapeutic strategies were successful for the patients included in our study who were classified clinically as having high-risk disease but whose tumors had gene expression profiles associated with PFS (molecular low-risk tumors). However, therapeutic modifications aimed at reducing toxicities should be considered for this group. By contrast, novel strategies will be required to improve the outcomes of patients with molecular high-risk tumors. For example, high expression of NTRK2 (TrkB) by molecular high-risk tumors supports the therapeutic investigation of inhibitors of Trk tyrosine kinases. The new opportunities offered by gene expression profiling should contribute to improving the care of children with neuroblastoma.

REFERENCES


Notes

This study was supported in part by grant CA60104 from the National Cancer Institute (R. C. Seeger); by training grants from the Children’s Cancer Research Fund, the National Institute of Health (T32-CA09659), Childrens Hospital Los Angeles, and CureSearch National Childhood Cancer Foundation—Hammond Research Fellowship (S. Asgharzadeh); by La Caixa fellowship program (R. Pique-Regi); and by training grants from the Children’s Cancer Research Fund, the National Institute of Health (T32-CA09659), Childrens Hospital Los Angeles, and AIDS Research (R. C. Seeger). The sponsors of this study did not participate in the collection, analysis, or interpretation of the data. The sponsors also did not write any part of the manuscript or have any influence on the decision to submit the manuscript for publication.

J. Buckley owns Epicenter Software, Inc, which manufactures the Genetrix program used in the analysis.

We thank Dr John Maris, Dr Emil Bogenmann, and Dr Leonid Metelitsa for their critical review of the manuscript and Robert Gerbing and Peter Wakamatsu for assisting with compilation of clinical data.

Funding to pay the Open Access publication charges for this article was provided by grant CA60104 from the National Cancer Institute.

Manuscript received January 10, 2006; revised April 18, 2006; accepted July 6, 2006.