Combining Gene Expression Profiles and Clinical Parameters for Risk Stratification in Medulloblastomas

Ana Fernandez-Teijeiro, Rebecca A. Betensky, Lisa M. Sturla, John Y.H. Kim, Pablo Tamayo, and Scott L. Pomeroy

ABSTRACT

Purpose

Stratification of risk in patients with medulloblastoma remains a challenge. As clinical parameters have been proven insufficient for accurately defining disease risk, molecular markers have become the focus of interest. Outcome predictions on the basis of microarray gene expression profiles have been the most accurate to date. We ask in a multivariate model whether clinical parameters enhance survival predictions of gene expression profiles.

Patients and Methods

In a cohort of 55 young patients (whose medulloblastoma samples have been analyzed previously for gene expression profile), associations between clinical and gene expression variables and survival were assessed using Cox proportional hazards models. Available clinical variables included age, stage (ie, the presence of disseminated disease at diagnosis), sex, histologic subtype, treatment, and status.

Results

Univariate analysis demonstrated expression profiles to be the only significant clinical prognostic factor (P = 0.03). In multivariate analysis, gene expression profiles predicted outcome independent of other criteria. Clinical criteria did not significantly contribute additional information for outcome predictions, although an exploratory analysis noted a trend for decreased survival of patients with metastases at diagnosis but favorable gene expression profile.

Conclusion

Gene expression profiling predicts medulloblastoma outcome independent of clinical variables. These results need to be validated in a larger prospective study.

INTRODUCTION

Medulloblastomas are the most common malignant brain tumors in children. Current management strategies combine surgery, radiotherapy, and chemotherapy, and long-term survival rates are 60% to 80%. Unfortunately, cognitive deficits and other sequelae of therapy are common among survivors [1-3]. A major challenge, therefore, is to differentiate high- from low-risk patients to tailor therapy to the degree of biologic aggressiveness.

Current clinical high-risk prognostic indicators include age less than 3 years, more than 1.5 cm³ of residual disease after surgery, and/or evidence of metastasis [4-6]. Recent reports indicate that these clinical variables are an inadequate method of defining disease risk [3,7]. Expression of the neurotrophin-3 receptor, trkC, correlates with a favorable outcome, whereas c-myc expression, erb2 expression, chromosome 17p loss, or anaplastic histology may identify high-risk patients [8-13].

The most accurate outcome predictions to date have been obtained through microarray gene expression profiling [14-16]. The aim of this article is to investigate whether microarray gene expression–based outcome predictions can be improved by combining clinical and molecular analysis for disease risk stratification. To do this, we combined gene expression–based outcome
predictions with clinical parameters in a multivariate model, using the Cox proportional hazards regression model to test whether gene expression predicts outcome independently or whether clinical parameters substantially add to the accuracy of survival predictions by microarray gene expression profiling.

**PATIENTS AND METHODS**

**Patients**

This retrospective study was conducted using the data set of 60 newly diagnosed patients from eight different institutions whose medulloblastoma tissue specimens had been previously investigated for gene expression profile analysis [14]. Thirty-five patients were part of a cohort described in previous publications as well [8,9]. To build a robust multivariate survival model in a group of patients with comparable clinical parameters, we focused our attention on the pediatric age range. Only five patients were older than 18 years in the cohort of 60 patients, significantly skewing the age range distribution, so we did not include them in the analysis. Available variables for the remaining 55 pediatric patients were sex, age, stage, histologic subtype, chemotherapy, follow-up, and status when the study was closed. Clinical details of the 55 patients are summarized in Table 1. The data set included 37 boys (67%) and 18 girls (33%). The median age was 6 years (range, 0.6 to 14.4 years). Ten patients (18%) were younger than 3 years. The project was approved by the Institutional Review Board of Children’s Hospital Boston (Boston, MA).

The histologic diagnosis of medulloblastoma was confirmed according to WHO criteria [17]: 42 samples (76%) fulfilled criteria for classic medulloblastoma, whereas 13 samples (24%) had the desmoplastic variant. The distribution of M stage, as defined by Chang [6], was as follows: M0 (39 patients), M1 (five patients), M2 (one patient), M3 (nine patients), and M4 (one patient). In view of the small numbers, analysis was limited to two stage groups: 39 patients (71%) were M0, whereas the remaining 16 patients (29%) were considered M+. All patients were treated with radiotherapy and chemotherapy. Craniospinal irradiation was 24 to 36 Gy with a tumor dose of 53 to 72 Gy. Chemotherapy consisted of cisplatin and vincristine, and combinations of carmustine, etoposide, cyclophosphamide, procarbazine, or lomustine. Only one patient received a two-drug regimen with cisplatin and vincristine. Two patients received high-dose chemotherapy at relapse, including methotrexate and thiopeta, followed by autologous bone marrow transplantation. The 55 patients included were observed for a median of 3.2 years (range, 5 months to 11 years) from date of diagnosis to last contact or death. With 17 deaths by the close of the study, median follow-up for the 38 surviving patients was 4.6 years (range, 2 to 11 years).

**Outcome Gene Profiling Model**

Gene profiling expression classification was based on an eight-gene model, established through k-nearest neighbors algorithm, previously described [14,18]. The eight-gene outcome prediction model of the 55-patient cohort was comparable to that of the 60-patient cohort, and the same eight genes most highly predictive of outcome were selected by both models. These genes and their relative expression levels are shown in Figure 1. Reanalysis of gene profiles belonging to the 55 patients younger than 18 years also showed two resulting groups with significantly different survival outcomes (P = .03; Fig 2A). Although the low-risk gene profiling group included 42 patients, the remaining 13 patients presented a high-risk gene expression pattern (Table 1).

**Statistical Analysis**

Wilcoxon rank sum tests were used to compare age at diagnosis among groups defined by categoric variables. Relationships among categoric markers were assessed using Fisher’s exact tests. Survival distributions were estimated using Kaplan-Meier curves [19]. Univariate and multivariate analyses were conducted using Cox proportional hazards regression models. For each of these tests, the significance level was taken to be 0.05, and all P values were two-sided. The software package StatView (version 5.0.1; SAS Institute, Cary, NC) was used for these analyses. Exploratory classification tree analyses [20] were conducted using the tssa software for S-plus (version 3.3; Statistical Sciences, Seattle, WA) available at the Statlib Web site at Carnegie Mellon University (http://lib.stat.cmu.edu/).

**RESULTS**

No significant associations between sex, stage, subtype, or gene expression profile were observed. In contingency table analysis, no significant associations among variables could be established. In particular, sex, stage, and subtype were not significantly associated with gene expression pattern.

**Univariate Survival Analysis**

Table 1 lists the P values for the comparisons of interest. Among the clinical variables analyzed, neither metastatic di-
ease stage at diagnosis (M0 vs M+; \( P = .1 \)), histologic subtype (classic vs desmoplastic; \( P = .5 \)), sex (\( P = .7 \); Figs 2B, 2C, and 2D), nor age at diagnosis (\( P = .7 \)) were found to be significantly associated with survival. As already noted, gene profiling expression analysis distinguished two groups with significantly different survival outcomes (\( P = .03 \); Fig 2A). According to this univariate model, the hazard for death for patients in the low-risk gene expression group is 66% lower than that for patients in the high-risk gene expression group (Table 2).

**Multivariate Regression Analysis**

A principal aim of this study was to establish whether clinical stratification provides prognostic information for patients with medulloblastoma in addition to that afforded
by gene expression profile analysis. We used backward elimination, with a threshold of 0.10, to select a multivariate model. None of the clinical covariates was a significant prognostic factor. Only gene expression group remained significant in the model. We considered possible age effects in the multivariate model with several possible cutpoints for age, but found none of them to be significantly predictive of survival.

**Survival Analysis After Risk Stratification**

An exploratory classification tree analysis suggested that the clinical variable of stage might be important among patients in the high-risk gene expression group [20]. However, in a proportional hazards model, stage did not significantly distinguish among patients within the high-risk gene group ($P = .08$), perhaps because of the small numbers. Nevertheless, as shown in Figure 3, there was a nonsignificant trend for good-risk patients by gene expression profile but with M+ disease to have more early deaths than those with M0 disease. A larger data set will be needed to determine whether M stage improves outcome predictions in good-risk patients by gene expression profiles.

**DISCUSSION**

The aim of this study was to investigate the hypothesis that the combined assessment of clinical and molecular markers identified by gene expression profiling will allow increased accuracy of disease risk stratification for patients with medulloblastoma over the use of either type of marker alone. We found that gene expression profiles are powerful predictors of survival independent of clinical parameters, although a trend for worse prognosis was noted for a subset of good-prognosis patients by gene expression pattern but with metastatic disease at initial diagnosis. Although our series did not show disseminated disease to be significantly associated with poor outcome, the size of the patient cohort may be insufficient to identify metastasis as a prognostic factor. A larger patient cohort will be required to definitively determine whether metastasis at diagnosis improves the accuracy of outcome markers defined by gene expression profiling.

Tumor size, an important element of the T stage of Chang, also was not found to be a prognostic factor, as has been previously reported [2,4,6,21,22] (data not shown). Although in our series boys are clearly predominant, no difference in survival related to sex is observed in univariate analysis. Previous studies of the effect of sex on clinical outcome in patients with medulloblastoma reached various conclusions. In at least two large studies, the survival advantage for girls was statistically significant [23,24], whereas other studies showed either borderline significance or no significance to the association of sex and outcome [13,25-30]. Finally, we did not find a survival advantage for patients with desmoplastic medulloblastomas.

We conclude that gene expression profiling provides prognostic information that cannot be obtained from histologic or clinical criteria. For future studies, it will be important to analyze the interaction between gene expression patterns and other outcome measures such as time to progression or treatment response, and to validate gene expression predictors prospectively in an independent test set [31]. If these results are confirmed in larger prospective studies, molecular profiling should become the new standard for risk classification in future clinical trials.

**REFERENCES**