Prediction of Clinical Outcome Using Gene Expression Profiling and Artificial Neural Networks for Patients with Neuroblastoma

Jun S. Wei,1 Braden T. Greer,1 Frank Westermann,2 Seth M. Steinberg,3 Chang-Gue Son,4,5 Qing-Rong Chen,1 Craig C. Whiteford,1 Sven Bilke,1 Alexei L. Krasnoselsky,1 Nicola Cenacchi,1 Daniel Catchpoole,5 Frank Berthold,6 Manfred Schwab,2 and Javed Khan1

1Advanced Technology Center, Oncogenomics Section, Pediatric Oncology Branch, National Cancer Institute, NIH, Gaithersburg, Maryland; 2Department of Tumour Genetics–B030, German Cancer Research Center, Heidelberg, Germany; 3Biostatistics and Data Management Section, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland; 4Department of Internal Medicine, College of Oriental Medicine, Daejeon University, Daejeon, Korea; 5Tumour Bank, The Children’s Hospital at Westmead, Westmead, New South Wales, Australia; and 6Department of Pediatrics, Klinik für Kinderheilkunde der Universität zu Köln, Köln, Germany

ABSTRACT

Currently, patients with neuroblastoma are classified into risk groups (e.g., according to the Children’s Oncology Group risk-stratification) to guide physicians in the choice of the most appropriate therapy. Despite this careful stratification, the survival rate for patients with high-risk neuroblastoma remains <30%, and it is not possible to predict which of these high-risk patients will survive or succumb to the disease. Therefore, we have performed gene expression profiling using cDNA microarrays containing 42,578 clones and used artificial neural networks to develop an accurate predictor of survival for each individual patient with neuroblastoma. Using principal component analysis we found that neuroblastoma tumors exhibited inherent prognostic specific gene expression profiles. Subsequent artificial neural network-based prognosis prediction using expression levels of all 37,920 good-quality clones achieved 88% accuracy. Moreover, using an artificial neural network-based gene minimization strategy in a separate analysis we identified 19 genes, including 2 prognostic markers reported previously, MYCN and CD44, which correctly predicted outcome for 98% of these patients. In addition, these 19 predictor genes were able to additionally partition Children’s Oncology Group-stratified high-risk patients into two subgroups according to their survival status (P = 0.0005). Our findings provide evidence of a gene expression signature that can predict prognosis independent of currently known risk factors and could assist physicians in the individual management of patients with high-risk neuroblastoma.

INTRODUCTION

Neuroblastoma is the most common solid extracranial tumor of childhood and is derived from the sympathetic nervous system. Patients in North America are currently stratified by the Children’s Oncology Group into high, intermediate, and low risk based on age, tumor staging, Shimada histology, MYCN amplification, and DNA ploidy (1). Patients <1 year of age or with lower stage diseases (International Neuroblastoma Staging System stages 1 and 2) usually have better outcome than older patients or those with advanced stage diseases (International Neuroblastoma Staging System stages 3 and 4). Certain consistent cytogenetic changes, including gain of 2p24 and 17q and loss of heterozygosity at 1p36 have been associated with a more aggressive phenotype (2, 3). The MYCN gene, located on 2p24, is amplified in ~22% of all neuroblastoma patients (4) and is an independent predictor for poor prognosis, especially for patients >1 year of age. Although other genes, such as TRKA, TRKB, hTERT, BCL-2, caspases, and FYN (4, 5) have been associated with neuroblastoma prognosis, they all lack the predictive power of MYCN and are not used currently in clinical practice. High-risk patients compose ~50% of all neuroblastoma cases; however, despite significant improvement in the therapy of neuroblastoma using neoadjuvant chemotherapy, surgery, and radiation, the death rate for these patients remains at 70% (6). Although the Children’s Oncology Group risk stratification has been carefully developed to take into account the above risk factors, it is primarily used to guide therapy and does not predict which individual patients will be cured from the disease.

DNA microarray technology has been proven to be an efficacious tool to molecularly classify cancers, to predict prognosis, and to identify genes that are potential therapeutic molecular targets (7–12). We have demonstrated previously that the combination of gene expression profiling and artificial neural networks is a powerful method that can accurately diagnose certain pediatric cancers including neuroblastoma (7). In this current study, we used gene expression profiles from cDNA microarrays to predict the outcome and identify an optimal gene set in patients with neuroblastoma using artificial neural networks.

MATERIALS AND METHODS

Tumor Samples. Fifty-six pretreatment primary neuroblastoma tumor samples from 49 neuroblastoma patients with outcome information were obtained retrospectively from three sources presenting between 1992 and 2000 (Table 1). All of the patients were treated according to local or national guidelines that followed similar protocols, which included “wait-and-see” after surgery or combinations of vincristine, doxorubicin, carboplatin, cisplatin, cyclophosphamide, melphalan, and etoposide, depending on the risk factors. All of the samples were anonymized, and our protocol was deemed exempt from the NIH Multiple Project Assurance. Pretreatment tumor samples were snap-frozen in liquid nitrogen after removal. Tumors were diagnosed as neuroblastoma by local centers experienced in the management of these cancers. Patients were divided into two outcome groups: the “good-outcome” group had event-free survival (i.e., neither relapse nor neuroblastoma progression) for at least 3 years (n = 30), and “poor-outcome” died due to neuroblastoma disease (n = 19). The median age for the good-outcome group was 0.9 years (range from 0.1 to 4.6 years) and for the poor-outcome group was 2.8 years (range from 0.8 to 10.5 years; Table 1).

RNA Extraction. Total RNA was extracted according to the published protocols (13). We used an Agilent BioAnalyzer 2100 (Agilent, Palo Alto, CA) to assess the integrity of total RNA from tumors. Total RNA from seven human cancer cell lines (CHP212, RD, HeLa, A204, K562, RDES, and CA46) was pooled in equal proportions to constitute a reference RNA, which was used in all of the cDNA microarray experiments.

RNA Amplification and Labeling of cDNA. mRNA was amplified one round using a modified Eberwine RNA amplification procedure (14). Next, an indirect fluorescent-labeling method was used to label cDNA as described by Hegde et al. (15). In brief, aminomethyl-dUTP (Sigma-Aldrich, St. Louis, MO) was first incorporated into cDNA in a reverse transcription reaction in which amplified antisense RNA was converted into cDNA by Superscript II reverse transcriptase enzyme (Invitrogen, Grand Island, NY) according to the manu-

Received 2/25/04; revised 7/27/04; accepted 8/3/04.

Grant support: Deutsche Krebshilfe (F. Westermann) and Deutsche Forschungsgemeinschaft (WE-2517/1–2, F. Westermann).

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

Note: J. S. Wei, B. T. Greer, and F. Westermann contributed equally to this work.

Requests for reprints: Javed Khan, Pediatric Oncology Branch, Oncogenomics Section, National Cancer Institute, 8717 Government Circle, Gaithersburg, MD 20877. Phone: 301-435-2937; Fax: 301-480-0314 or 301-402-3134; E-mail: khanjav@mail.nih.gov.

©2004 American Association for Cancer Research.
facturer’s instructions. Second, unincorporated aminomethyl-DUTP was removed with Qiagen PCR purification kits (Qiagen, Valencia, CA) according to the manufacturer’s instructions. Third, monoreactive-Cye5 or Cye3 dyes (Amer shamPharmacia, Piscataway, NJ) were conjugated with the aminomethyl-DUTP on the cDNA. Fluorescent-labeled cDNA was purified with Qiagen PCR purification kits.

Fabrication of cDNA Microarrays, Hybridization, Image Acquisition, and Image Analysis. Sequence-verified cDNA libraries were purchased from Research Genetics (Huntsville, AL), and a total of 42,578 cDNA clones, representing 25,933 unique genes (UniGene clusters; 13,606 known genes and 12,327 unknown expressed sequence tags), were printed on microarrays using a BioRobotics MicroGrid II spotter (Harvard Bioscience, Holliston, MA). Fabrication, hybridization, and washing of microarrays were performed as described by Hegde et al. (15). Images were acquired with an Agilent DNA microarray scanner (Agilent, Palo Alto, CA) and analyzed using the Microarray Suite program as described (16), coded in IPLab (Scanalytics, Fairfax, VA).

Data Normalization and Filtering. Gene expression ratios between tumor RNA and reference RNA on each microarray were normalized using a pin-based normalization method modified from Chen et al. (16). To include only high-quality data in the analysis, the quality of each individual cDNA spot was calculated according to Chen et al. (17). Next, spots with an average quality across all of the samples <0.95 were excluded from all of the analyses. There were 37,920 (90.3%) clones that passed this quality filter.
Architecture of Artificial Neural Networks. First, we used principal component analysis and reduced the dimensionality of the data to the top 10 principal components as inputs for artificial neural networks. This procedure reduced the number of variables from 37,920 to 10 to avoid over-fitting the data, which occurs when the number of variables exceeds the number of samples. We used feed-forward resilient back-propagation multilayer perceptron artificial neural networks (coded in Matlab, The Mathworks, Natick, MA) with three layers: an input layer of the top 10 principal components of the data (Fig. 1, A and B) or the gene expression ratios of each cDNA spot (for the minimized gene set, see Fig. 1B); a hidden layer with 3 nodes; and an output layer generating a committee vote that discriminates two classes (i.e., good- and poor-outcome groups). Average artificial neural network committee votes were used to classify samples, and 0.5 was used as the decision boundary for artificial neural network prediction throughout the study. The ideal vote was 0 for the good-outcome group (alive) and 1 for the poor-outcome group (dead). We trained the artificial neural networks using an 8-fold cross-validation scheme in all of the analyses similar to those described previously (7).

Prediction Using a Leave-One-Out Strategy. To test the generalizibility of the artificial neural network approach, we first performed a leave-one-out prediction strategy (Fig. 1A), where we left out each sample (of the 49 unique samples) one time during the training of artificial neural networks and tested it as an independent sample to predict the outcomes with all of the quality-filtered clones (n = 37,920) without additional clone selection.

Identification of Prognostic Signature Using Training and Test Sets. To identify the prognostic genes, we performed a separate artificial neural network analysis using a gene minimization procedure as described by Khan et al. (7). In brief, the 7 replicate samples were placed in the training set, and the remaining samples were then randomly partitioned into training (n = 35) and testing (n = 21) sets. None of the replicate samples were included in the test set to ensure that the selected genes did not bias the prediction outputs of the trained artificial neural networks. The minimal number of clones for outcome prediction was identified using only the training set. Quality-filtered clones were first ranked by determining the sensitivity of prediction of the 35 training samples with respect to a change in the gene expression level of each clone.

Fig. 1. Workflow diagrams. A, workflow for a complete leave-one-out artificial neural network (ANN) analysis using all 37,920 clones. Gene expression profiling was performed on tumors from 49 neuroblastoma (NB) patients (Alive, n = 30; Deceased, n = 19) using cDNA microarrays containing 42,578 clones. After a quality filter, 37,920 clones were used as a data matrix of high quality cDNA measurements for further data analysis. Principal component analysis was used to reduce the dimensionality of the data and reduce noise. The top 10 principal components were used for input to the ANN. One sample was left out as an independent test sample, and the ANNs were trained using the remaining 48 NB samples. ANN training scheme (gray box). 1. All remaining neuroblastoma samples were randomly partitioned into eight groups. 2. One of the eight groups (containing 6 samples) was selected as a validation set, whereas the remaining 7 groups (42 samples) were used to train the network. 3 and 4. The training weights were iteratively adjusted for 100 cycles (epochs). 5. The ANN output (0–1, where 0 = ideal good-outcome and 1 = ideal poor-outcome) was calculated for each sample in the validation set. 6. A different validation set was selected from the same partitioning in 1, and the remaining seven groups were used for training. Steps 2–6 were repeated until each of the eight groups from 1 had been used as a validation set exactly one time. 7. The samples were randomly partitioned into eight new groups, and steps 2–6 were repeated. Sample partitioning was performed 100 times in total. Thus, steps 1–6 were repeated 100 times. Eight hundred ANN models were, thus, trained and were used to predict the left out test sample. This scheme was repeated for each left out test sample. B, identifying prognostic gene expression signature and outcome prediction. Fifty-six neuroblastoma samples (7 replicates were added to the training group to examine the reproducibility of the results) were partitioned into a training (n = 35) and an independent test (n = 21) set. Principal component analysis was again performed, and ANNs were retrained using the 35 training samples based on the ANN training scheme detailed in the gray box in A. Gene minimization. Each of the input clones was ranked according to its importance to the prediction of ANNs (7). Increasing numbers of the top-ranked clones were used to train ANNs, and the resulting classification error was monitored. The minimal number of clones that yielded the minimal classification error (Fig. 3A) was identified, and the top-ranked clones for each gene were used to retrain the ANNs and predict the 21 test samples without performing a principal component analysis.
Then, using increasing numbers of the top artificial neural network-ranked clones, we identified the minimum number of clones that generated minimum prediction errors (Fig. 1B). Where multiple clones represented one gene, we selected the top-ranked clone to obtain a minimal predictor gene set. We recalibrated the artificial neural networks using the expression ratios of these genes with only the training samples (without performing principal component analysis). Finally, we predicted the survival status of the test samples using the trained artificial neural networks (Fig. 1B).

**Statistical Analysis for Survival.** Survival length was calculated for the 49 unique neuroblastoma patients from date of diagnosis until date of death or last follow-up as appropriate. The probability of survival and significance was calculated using the Kaplan-Meier and Mantel-Haenszel methods, respectively (18, 19). The Cox proportional hazards model (20) was used to determine the hazard ratios and confidence intervals (21) for survival between the dichotomized groups of patients and was used to assess which factors were jointly significant in the association with survival for the 24 high-risk patients (20). The Cox model parameters ($b$) were converted to hazard ratios by computing $\exp(b)$, where $\exp(a) = 2.7183^a$. The 95% confidence interval for the hazard ratio was computed as $[\exp(b_L), \exp(b_U)]$ where $b_L = b - 1.96$ [estimated SE ($b$)] and $b_U = b + 1.96$ [estimated SE ($b$)] (21). In this study, the hazard ratio indicates the risk associated with neuroblastoma-caused death while being in a greater-risk category compared with that of being in the lower-risk category. Using the procedure described by Simon and Altman (22), a likelihood ratio test was used to assess for importance of the microarray prediction after adjusting for standard prognostic factors such as MYCN amplification, age, or stage.

**RESULTS**

**Prediction of Outcome Using the Global Expression Profiles of All of the Clones.** Visualization of all 56 of the neuroblastoma samples using principal component analysis of all of the quality-filtered 37,920 clones revealed neuroblastoma samples generally grouped according to their clinical outcomes (Fig. 2A), clearly indicating a pre-existent prognostic signature. To demonstrate the generalizability of the artificial neural network approach, we next tested the ability of artificial neural networks to predict prognosis of the 49 unique individuals (excluding 7 replicated samples) with all 37,920 clones using a conservative unbiased leave-one-out prediction strat-
Identifying Prognostic Gene Expression Signature. To identify the optimal set of genes that results in the minimum classification errors, we performed a gene minimization procedure in a separate artificial neural network analysis using training and test sets as described previously (7). We first randomly partitioned all 56 of the samples into training (n = 35) and testing sets (n = 21) and used only the training set for the gene selection algorithm. We observed that the top 24 artificial neural network-ranked genes resulted in the minimal classification error (Fig. 3A). These 24 clones represented 19 unique genes, and we took the top-ranked clone for each gene and used this as our minimal gene set. When we visualized the overall variance of these genes using principal component analysis on all 56 of the samples we found a clearer separation of the poor- from the good-outcome samples when compared with the principal component analysis for all 37,920 clones (Fig. 3B).

We next recalibrated the artificial neural networks with the 35 training samples using the expression ratios for the 19 genes and correctly predicted the outcomes for 5 of 5 poor-outcome and 15 of 16 good-outcome patients in the independent test set, corresponding to a sensitivity of 100% and a specificity of 94% for predicting poor outcome (Fig. 3C; Table 2). The positive predictive values were 83% and 100% for the poor- and good-outcome groups, respectively, for the test samples and 95% and 100% for all of the patients (Table 2). The Kaplan-Meier curves demonstrated that patients with good and poor signatures based on the expression ratios of the 19 genes had significantly different survival probabilities (P < 0.0001, see Fig. 2C). The Cox proportional hazard ratio for the risk of death associated with the poor signature was 16.1 (95% confidence interval, 4.6 to 56.9, P < 0.0001), which was higher than those of all of the other risk factors we examined (stage, MYCN amplification, and age) except Shimada histology and was comparable with the Children’s Oncology Group risk stratification (Table 3; Fig. 2D).

Outcome Prediction for High-Risk Patients. We next investigated whether the gene expression signatures could predict the survival status of those patients in our study that are currently stratified as high risk (see Table 1). From our 49 patients, 24 were high risk (Table 1). The Kaplan-Meier curves demonstrated that artificial neural networks were able to additionally partition these high-risk patients according to their clinical outcomes using all 37,920 of the quality-filtered clones (P = 0.0067), as well as the top 19 artificial neural network-ranked genes (P = 0.0005; Fig. 5, A and B). As shown in Fig. 5B, the top 19 artificial neural network-ranked genes were able to correctly predict all 5 with good signature as surviving and 18 of 19 with poor signature as dying, suggesting a potential benefit for predicting outcome in these high-risk patients. The hazard ratio was again infinite, because all of the patients that we predicted to have a good outcome survived (Table 3).

To determine whether the gene expression signatures provide additional predictive power over the conventional risk factors, we first created a Cox model using age, stage, and MYCN amplification excluding the artificial neural network prediction results. The model showed that MYCN amplification (P = 0.0064) was the only significant factor (i.e., P < 0.05, see Fig. 5C). Therefore, we built another multivariate model using MYCN amplification and the prediction results based on all 37,920 clones (Fig. 5D). (We used the artificial neural network results based on the 37,920 clones, because there were no deaths in the good signature group using the 19 genes, and in these circumstances it is not possible to create models where the hazard ratios are infinite.) Applying the likelihood ratio test, we found that prediction by all of the clones added predictive ability to the model.
Additionally, the Kaplan-Meier curves (Fig. 5, E and F) illustrate that artificial neural network prediction can additionally separate the MYCN nonamplified patients according to their survival status based on either all of the clones (P = 0.047) or in particular the 19 genes (P = 0.0076, see Fig. 5F).

DISCUSSION

We have developed an artificial neural network-based method for predicting the outcome of patients with neuroblastoma using the expression profiles of only 19 genes that provides a significant improvement in prediction over the current known risk factors. Moreover, we found that the most important advantage of our approach was the ability to additionally partition Children’s Oncology Group stratified high-risk patients, in particular those without MYCN amplification, into two subgroups according to their survival status. The ability to predict the outcome of individual patients with high-risk neuroblastoma at initial diagnosis using gene expression signatures has major clinical implications, because ~70% of the patients in this group (~50% of all neuroblastoma patients) succumb to the disease (1). Firstly, patients that are identified to have a poor signature, i.e., predicted to die if given conventional therapy, may directly benefit from the newer therapeutic strategy trials that are currently under investigation by the cooperative study groups such as Children’s Oncology Group. Secondly, because treatment-related death rates have been reported to be as high as 23% (26), it may be possible to design future dose intensity reduction trials to minimize therapy-related morbidity and mortality for the high-risk patients who have a good signature. An example of such a patient in the latter category is NB14 (stage 4, MYCN-amplified) who, despite his high-risk status, experienced event-free survival for ~3 years as was predicted by our artificial neural networks. Although the survival rate for patients with Children’s Oncology Group-stratified low-risk disease is 95%, our approach may identify the few patients predicted to have a poor outcome by the artificial neural networks who may benefit from more aggressive therapy. For instance, although case NB18 was classified as low-risk (based on stage 2 and MYCN not amplified), our artificial neural networks predicted this sample as poor-outcome, and this patient died within 1.5 years after diagnosis. These results indicate the...
potential utility of using our approach for individualized management of patients with cancer. However, they need to be interpreted with some caution in view of the limited number of subjects in our study and some heterogeneity of their treatments, and confirmation is required in larger, prospective trials before these predictor genes are used in the clinic.

Because there was some overlap in the expression levels of the top 19 artificial neural network-ranked genes between the prognostic groups, the prospect of identifying a single gene that can accurately predict outcome is unlikely. Thus, a combinatorial approach using several genes and artificial machine learning algorithms was necessary for accurate outcome prediction. Among these 19 genes, 2 (MYCN and CD44) have been reported to correlate with neuroblastoma prognosis (23–25), thus validating our ability to identify prognostic-specific genes. MYCN amplification is an established marker for high stage and poor outcome (23) and plays a critical role in the aggressive phenotype of neuroblastoma tumors (27, 28). Our analysis confirmed MYCN as an important prognostic marker (ranked 16 of 19); however, the median expression level of this gene was similar in the two groups, in agreement with previous reports that MYCN expression levels are not consistently correlated with survival in patients with nonamplified tumors (29–31).

MYCN amplification is currently the only molecular marker used for risk stratification; however, it cannot be used as the sole risk predictor, because only 22% of neuroblastoma patients have this molecular trait.

Of the 19 predictor genes, 8 of the 12 known genes have been reported previously to be expressed in neural tissue. Of these, 5 were up-regulated in the poor-outcome group (DLK1, PRSS3, ARC, SLIT3, and MYCN), and 3 were down-regulated (CNR1, ROBO2, and BTBD3). DLK1 (ranked number 1) is the human homologue of the Drosophila Delta gene and is expressed by neuroblasts during neural development. The higher expression levels of these genes, in agreement with previous reports that MYCN expression levels are not consistently correlated with survival in patients with nonamplified tumors (29–31), MYCN amplification is currently the only molecular marker used for risk stratification; however, it cannot be used as the sole risk predictor, because only 22% of neuroblastoma patients have this molecular trait.
At this point, the predictive role of 
TRKA,
TRKB,
and
FYN
was not on our microarrays, and
FYN
is not a statistically significant conventional risk factor in the high-risk patients in our data set (n = 24). (H.R., hazard ratio; C.I., confidence interval). 

Of additional interest, the
ARHI
gene, which maps to 1p31, is a maternally imprinted tumor suppressor gene implicated in ovarian and breast cancer (39), possibly through methylation silencing (40), and is among the down-regulated genes for the poor-outcome group. An additional study of its role in tumorigenesis as a potential tumor suppressor gene in neuroblastoma is warranted particularly because of its proximity to the 1p36 region, which is frequently deleted in poor-outcome neuroblastoma patients.

We noted the absence of three prognostic related genes reported previously
TRKA,
TRKB,
and
FYN
(5, 41, 42), among our 19 genes.

In this study we have identified a small subset of 19 predictor genes from a pool of 25,933 unique genes with the majority of these 19 genes showing a >2-fold average differential expression between good- and poor-outcome tumors. This small number of genes can be developed into cost-effective clinical assays for outcome prediction. In addition, the products of 3 genes (DLK1, SLIT3, and PRSS3) are secreted proteins, raising the possibility of using these as serum markers for prognosis.

In this data set, our artificial neural network-based method provided a significant improvement in prediction over the current risk factors in patients with neuroblastoma. Moreover, the most important advantage of our approach was the ability to additionally partition Children’s Oncology Group-stratified high-risk patients, in particular those without MYCN amplification, into two subgroups according to their survival status. These findings merit confirmation on larger, prospective trials. We believe that our approach would allow physicians to tailor therapy for each individual patient according to their molecular profile, with the prospect of improving clinical outcome and survival rates in patients with neuroblastoma.

REFERENCES


Prediction of Clinical Outcome Using Gene Expression Profiling and Artificial Neural Networks for Patients with Neuroblastoma

Jun S. Wei, Braden T. Greer, Frank Westermann, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/64/19/6883

Cited articles
This article cites 37 articles, 16 of which you can access for free at:
http://cancerres.aacrjournals.org/content/64/19/6883.full.html#ref-list-1

Citing articles
This article has been cited by 30 HighWire-hosted articles. Access the articles at:
/content/64/19/6883.full.html#related-urls

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

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

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