Revised Risk Estimation and Treatment Stratification of Low- and Intermediate-Risk Neuroblastoma Patients by Integrating Clinical and Molecular Prognostic Markers

André Oberthuer1, Dilafruz Juraeva2, Barbara Hero1, Ruth Volland1, Carolina Sterz2, Rene Schmidt3, Andreas Faldum2, Yvonne Kahlert4, Anne Engesser5, Shahab Asgharzadeh4, Robert Seeger4, Miki Ohira5, Akira Nakagawara5, Paola Scaruffi6, Gian Paolo Tonini8, Isabelle Janoueix-Lerosey9, Olivier Delattre9, Gudrun Schleiermacher6, Jo Vandesompele10, Frank Speelman10, Rosa Noguera11, Marta Piqueras11, Jean Bénard12, Alexander Valent12, Smadar Avigad13,14, Isaac Yani13,14, Richard G. Grundy15, Monika Ortmann16, Chunxuan Shao17, Manfred Schwab17, Roland Ellis2,18, Thorsten Simon1, Jessica Theissen1, Frank Berthold1, Frank Westermann17, Benedikt Brors2, and Matthias Fischer1

Abstract

Purpose: To optimize neuroblastoma treatment stratification, we aimed at developing a novel risk estimation system by integrating gene expression–based classification and established prognostic markers.

Experimental Design: Gene expression profiles were generated from 709 neuroblastoma specimens using customized 4 × 44 K microarrays. Classification models were built using 75 tumors with contrasting courses of disease. Validation was performed in an independent test set (n = 634) by Kaplan–Meier estimates and Cox regression analyses.

Results: The best-performing classifier predicted patient outcome with an accuracy of 0.95 (sensitivity, 0.93; specificity, 0.97) in the validation cohort. The highest potential clinical value of this predictor was observed for current low-risk patients [5-year event-free survival (EFS), 0.84 ± 0.02 vs. 0.29 ± 0.10; 5-year overall survival (OS), 0.99 ± 0.01 vs. 0.76 ± 0.11; both P < 0.001] and intermediate-risk patients [5-year EFS, 0.88 ± 0.06 vs. 0.41 ± 0.10; 5-year OS, 1.0 vs. 0.70 ± 0.09; both P < 0.001]. In multivariate Cox regression models for low-risk/intermediate-risk patients, the classifier outperformed risk assessment of the current German trial NB2004 [EFS: hazard ratio (HR), 5.07; 95% confidence interval (CI), 3.20–8.02; OS: HR, 25.54; 95% CI, 8.40–77.66; both P < 0.001]. On the basis of these findings, we propose to integrate the classifier into a revised risk stratification system for low-risk/intermediate-risk patients. According to this system, we identified novel subgroups with poor outcome [5-year EFS, 0.19 ± 0.08; 5-year OS, 0.59 ± 0.1], for whom we propose intensified treatment, and with beneficial outcome [5-year EFS, 0.87 ± 0.05; 5-year OS, 1.0], who may benefit from treatment de-escalation.


© 2014 AACR.
Introduction

Neuroblastoma is the most frequent extracranial solid tumor in children accounting for 8% to 10% of all childhood cancers and for 15% of pediatric oncology deaths (1). A major hallmark of the disease is its broad variety of clinical behavior ranging from spontaneous regression or maturation of disease to fatal tumor progression despite intensive multimodal treatment. To appropriately tailor therapy, current trials use different combinations of clinical and genetic markers to discriminate patients with low-, intermediate-, and high-risk of death from disease. These markers include age at diagnosis (2), tumor stage (3, 4), genomic amplification of the MYCN protooncogene (MNA; refs. 5, 6), deletion or imbalance of chromosome 1p (del1p1) and chromosome 11q (del11q1) (7, 8), DNA ploidy (9), and a histopathologic classification system proposed by Shimada and colleagues (10). Yet, despite elaborate risk estimation systems, stratification is still imperfect resulting in over- or undertreatment of patients with neuroblastoma.

In recent years, several well-conducted studies demonstrated that molecular classification using mRNA expression information more accurately reflects the individual tumor behavior at the time of diagnosis than traditional clinical markers and might therefore allow for improved risk stratification and therapy selection (11–17). However, to date, none of the published gene expression–based classifiers has been incorporated into clinical classification systems or validated prospectively in a controlled trial. The reasons for this insufficient transfer of promising basic research findings into clinical applications are diverse and comprise high logistic and bureaucratic efforts to implement genomic classifiers into clinical practice, difficulties to setup randomized controlled trials for relatively small patient numbers, and the challenge to appropriately adjust cytotoxic dosing according to binary genomic classification results. At present, it is therefore still unclear whether genomic classification approaches will eventually lead to improved treatment concepts and outcome of patients with neuroblastoma.

To overcome this limitation, we here report on the generation of a novel gene expression–based classifier that accurately classified patients in a comprehensive neuroblastoma cohort of internationally collected specimens. We determined subgroups of patients for whom genomic classification using our predictor might offer the biggest clinical benefit. Finally, we describe in detail how our genomic classification model can be integrated into a revised risk stratification system as a therapy stratifying marker that allows for an improved determination of the intensity of first-line therapy in newly diagnosed patients with non–high-risk neuroblastoma.

Materials and Methods

Patients

This study comprised 709 newly diagnosed patients with neuroblastoma from nine centers in nine countries for whom pretreatment tumor material was available: Belgium (n = 5; 0.7%); France (n = 19; 2.7%); Germany (n = 517; 72.9%); Israel (n = 12; 1.7%); Italy (n = 24; 3.4%); Japan (n = 20; 2.8%); Spain (n = 14; 2.0%); United Kingdom (n = 5; 0.7%), and the United States (n = 93; 13.1%). All patients were registered in the respective clinical trials with informed consent. Patients’ age at diagnosis ranged from 0 to 305 months (median age, 14.2 months). Median follow-up for patients without fatal events was 6.7 years (range, 0–19 years). Five-year event-free survival (EFS) of the total cohort was 0.64 ± 0.02, and 5-year overall survival (OS) was 0.78 ± 0.02. Data on EFS were available for 688 and data on OS were available for all 709 patients. Stage was classified according to the International Neuroblastoma Staging System (INSS; 4): stage I: n = 159 (MNA, n = 5); stage II: n = 116 (MNA, n = 4); stage III: n = 92 (MNA, n = 15); stage IV: n = 259 (MNA, n = 94); stage IVs: n = 80 (MNA, n = 4); and patients with multilocalized primary tumors: n = 3 (MNA; n = 0). Response to treatment was defined according to the revised criteria of the International Neuroblastoma Response Criteria (INRC; 4). Chromosomal alterations were defined according to the guidelines of the European Neuroblastoma Quality Assessment Group (18). Detailed information on patients’ clinical covariates and gene expression–based classification is given in Supplementary Table S1.

Gene expression analyses and supervised classification

Generation of gene expression profiles. Single-color gene expression profiles were generated using customized 4 × 44 K oligonucleotide microarrays produced by Agilent Technologies. Labeling and hybridization was performed as described previously (19). After washing and scanning, resulting TIFF images were processed using Agilent’s Feature Extraction software Version 9.5.1. Both the raw and the processed expression profiling data and basic clinical information are available through ArrayExpress (http://www.ebi.ac.uk/arrayexpress; accession: E-MTAB-1781).

Data preprocessing. Raw gene expression data were normalized using the quantile algorithm from limma (20). To maintain the comparable scale of the training and validation dataset, the validation set was preprocessed using the training dataset as a...
reference. All calculations were performed in R v2.14.1 (21). Subsequently, gene expression–based classifiers were generated.

**Classifier training and evaluation.** The classifiers were trained using recursive feature elimination method for feature selection (22) and a linear support vector machine (SVM) as classification algorithm. The nested cross-validation (5xCV for outer loop, 3xCV for inner loop) was performed with 10 repetitions. The average and the variance of classifier performance were evaluated using the following performance measures: accuracy (23), sensitivity (SEN), specificity (SPEC), Matthew correlation coefficient (MCC), root mean squared error, and area under the curve (AUC) of a receiver-operating characteristics curve.

**Feature selection.** Features (i.e., microarray probes for transcripts whose expression values were considered for classification) were selected using SVM-based recursive feature elimination (SVM-RFE) method (22). The initial set of features consisted of all probes (43,291, excluding the control probes). The features were then ranked on the basis of how frequently they have been selected in 50 cross-validation runs. The upper threshold for selecting the features for building a classifier was 65% of all cross-validation runs. This threshold was gradually lowered from 64% to 10% in 2% increments to evaluate the classifiers’ performance of the larger feature space. The classifiers SVM_th4, SVM_th26, SVM_th24, SVM_th22, and SVM_th10 were trained on all available training data using variables that have been selected in 44%, 26%, 24%, 22%, and 10% of all cross-validation runs, respectively.

**Statistical analyses**

Kaplan–Meier estimates for EFS and OS were calculated from the time of diagnosis and compared by the log-rank test. Recurrence, progression, and death from disease were considered as events. Cox regression models for low- and intermediate-risk patients were applied using a stepwise variable selection procedure recommended by Collett (24) to analyze the prognostic value of potentially prognostic factors. The factors age (reference <18 vs. ≥18 months), tumor stage (reference I vs. II/III or IV or IVS), status of chromosome 1p (reference normal vs. deletion/imbalance), INPC/Shimada classification [reference favorable (F) vs. unfavourable (UF)], and the five top-performing genomic classifiers [reference favorable vs. unfavorable] were fitted into a stepwise-backward selection. The likelihood-ratio test P value for inclusion was less than 0.1 and for exclusion more than 0.05.

**Results**

**Generation and validation of single-color gene-expression classifiers for neuroblastoma patients**

We generated more than 200 different classification models from expression profiles of a training set of 75 patients with maximally divergent courses of the disease [death from disease (IF, n = 22) vs. EFS >1,000 days without cytotoxic treatment (F, n = 53)]. All models were evaluated by a complete 10 times repeated 5-fold cross-validation. From this internal validation, we selected the five best-performing classifiers that had identical cross-validated classification values: an accuracy of 0.96, a SEN of 0.87, a SPEC of 0.97, and a MCC of 0.86. Intriguingly, all of the top five classifying models were generated using a support vector machine learning algorithm but comprised a variable number of microarray probes (n = 10–194). Subsequently, external validation of these classifiers’ performance values was conducted using those 325 patients of the complete test set (n = 634) who fulfilled the criteria for classifier training (IF, n = 138; F, n = 187). As highlighted in Supplementary Table S2, all classifiers demonstrated comparably high classification accuracies (0.94 and 0.95, respectively) and showed a balanced ratio of SEN and SPEC (Supplementary Table S2).

**Performance of gene expression–based classification in the entire validation cohort and clinical risk groups defined according to the German neuroblastoma trial NB2004**

To further evaluate the performance of the five selected SVM classifiers, Kaplan–Meier analyses for EFS and OS were performed for both the complete cohort of test set patients (n = 634) and for subcohorts of patients considered to have a low (n = 313), intermediate (n = 69), or high risk (n = 234) of death from disease as determined by the criteria of the current German neuroblastoma trial NB2004 (25, 26). Of note, 18 patients of the total cohort could not be categorized according to these criteria because of either missing chromosomal 1p status (n = 12), heterogeneous 1p status (n = 3), or heterogeneous MYCN status (n = 3). We here report on the results of the SVM_th10 classifier, while survival estimates for the remaining four classifiers are summarized in the Supplementary Table S3.

In the total cohort, the SVM_th10 predictor classified 379 patients as favorable. These patients had a 5-year EFS of 0.82 ± 0.02 and a 5-year OS of 0.98 ± 0.01. In contrast, survival estimates were significantly worse in 255 patients classified as unfavorable (5-year EFS, 0.34 ± 0.03 and OS, 0.51 ± 0.03; both P <0.001; Fig. 1A; for 21 patients of the total cohort only data on OS but not on EFS were available, resulting in differing patient numbers in the Kaplan–Meier estimates for EFS and OS.). Analysis of clinical risk groups revealed that 293 of 313 patients classified as low risk by NB2004 were also classified as favorable by the SVM_th10 classifier and had an excellent outcome (5-year EFS, 0.84 ± 0.02 and 5-year OS, 0.99 ± 0.01). In contrast, the 20 patients classified as unfavorable within this subgroup had a significantly worse EFS and OS (0.29 ± 0.1 and 0.76 ± 0.11, respectively; both P <0.001; Fig. 1B). Likewise, 41 of 69 NB2004 intermediate-risk patients were classified as favorable and had an EFS of 0.88 ± 0.06 and an OS of 1.0 as compared with an EFS of 0.41 ± 0.10 and an OS of 0.70 ± 0.09, respectively, of those 28 patients who had an unfavorable prediction (P < 0.001; Fig. 1C). Finally, within the cohort of 234 high-risk patients, the gene expression–based votes also separated subgroups with significantly differing EFS and OS [favorable (n = 30): EFS, 0.63 ± 0.09 and OS, 0.83 ± 0.07 vs. unfavorable (n = 204): EFS, 0.33 ± 0.03 and OS, 0.46 ± 0.04; both P < 0.001; Fig. 1D].

**Performance of gene expression–based classification in further clinically relevant patient subgroups**

To identify patients who might benefit the most from gene expression–based risk estimation, we assessed the prognostic value of the SVM_th10 predictor in clinically relevant subgroups of patients defined by combinations of prognostic markers. First, we determined the classifiers’ power in main subcohorts of non–high-risk patients: (i) stage I–III, MYCN nonamplified patients, and (ii) MYCN nonamplified patients with metastasized disease (stage IV or IVS) <18 months of age. In the first cohort, the
Figure 1.
Kaplan-Meier estimates for EFS and OS according to classification by the SVM_th10 predictor. A, the EFS (left) and OS (right) for the complete validation cohort of patients with neuroblastoma ($n = 634$). Kaplan-Meier survival estimates for the subcohorts of low-risk ($n = 313$; B), intermediate-risk ($n = 69$; C), and high-risk ($n = 234$; D) patients as defined by the German neuroblastoma trial NB2004.
classifier correctly identified all patients who succumbed to disease within the subcohort of 68 stage I–III, MYCN nonamplified patients ≥18 months of age. Here, 46 patients classified as favorable by the SVM_th10 predictor had an excellent outcome (EFS, 0.90 ± 0.05 and OS, 1.0) as opposed to an EFS of 0.14 ± 0.08 and an OS of 0.51 ± 0.11 of those 22 patients who were classified as unfavorable (both P < 0.001; Fig 2A). Moreover, the classifier was also able to discriminate patients with an unfavorable course of the disease in the cohort of 234 stage I–III, MYCN nonamplified patients <18 months of age, thereby identifying those few patients within this subgroup whose tumors demonstrated a more aggressive behavior (favorable (n = 225) EFS, 0.84 ± 0.02 vs. unfavorable (n = 9), 0.56 ± 0.17; P = 0.018; OS, 1.0 vs. 0.66 ± 0.13; P < 0.001; Fig. 2B).

Similarly, gene expression–based classification by the SVM_th10 predictor separated patients with divergent outcome in the cohorts of stage IV, MYCN nonamplified patients <18 months of age (favorable, n = 34, EFS, 0.88 ± 0.06 and OS, 1.0 vs. unfavorable, n = 15, EFS, 0.64 ± 0.13 and OS, 0.87 ± 0.09; P = 0.043 and 0.217, respectively; Fig. 2C) and stage IVS, MYCN nonamplified patients (favorable (n = 55) EFS, 0.8 ± 0.06 and OS, 0.96 ± 0.03 vs. unfavorable (n = 7) EFS, 0.29 ± 0.17 and OS, 0.86 ± 0.13; P < 0.001 and 0.38, respectively; Fig. 2D).

In contrast, the classifier was not able to discriminate patients with divergent outcome in the two main subcohorts that define high-risk disease (27): (i) patients with MYCN-amplified disease (n = 114) and (ii) stage IV, MYCN nonamplified patients >18 months of age (n = 102). In the German NB2004 trial, stage IV MYCN nonamplified patients were considered high risk when older than 12 months of age. However, since an age cutoff of 18 months excelled as a prognostic marker in recent years, we also used an 18 months cutoff in the present study, thereby leaving out 18 stage IV patients between 12 and 18 months of age who had been stratified as high risk by the NB2004 risk estimation system; refs 25 and 26.) In the subcohort of MYCN-amplified cases, it was observed that almost all patients (113 of 114) were predicted as unfavorable and had a poor outcome (EFS, 0.31 ± 0.05 and OS, 0.37 ± 0.05). Only 1 patient who carried an MYCN amplification was predicted as favorable, and this patient has survived event-free to date (both P values not significant; Supplementary Fig S1A).

Similarly, the SVM_th10 classifier did not significantly discriminate patients with divergent outcome in the subgroup of stage IV, MYCN nonamplified patients >18 months of age (favorable (n = 16): EFS, 0.38 ± 0.12 and OS, 0.68 ± 0.12 versus unfavorable (n = 86): EFS, 0.34 ± 0.05 and OS, 0.54 ± 0.06, both P values not significant; Supplementary Fig S1B). We therefore concluded that the SVM_th10 classifier has only a limited potential for current high-risk patients.

Multivariate Cox regression analyses

As highlighted in Supplementary Table S2, the top five classifiers performed comparably in predicting outcome of patients with neuroblastoma. Moreover, all genomic classifiers worked particularly well in non–high-risk patient cohorts. Therefore, we compared the predictive power of all five genomic classification models for non–high-risk patients, that is, MYCN nonamplified patients with stage I–III of any age and patients with stage IVS or stage IV disease <18 months of age, by applying a multivariable Cox regression selection method as proposed by Collett (24). Following this approach, we analyzed the prognostic value of the following potentially explanatory prognostic factors with respect to EFS and OS: age at diagnosis, tumor stage, chromosome 1p status, the Shimada/INPC classification, and the five gene expression classifiers, SVM_th10, SVM_th22, SVM_th24, SVM_th26, and SVM_th44. In the model based on EFS the SVM_th10 classifier, tumor stage and chromosome 1p status were independent prognostic markers, with the SVM_th10 predictor presenting the highest hazard ratio [HR; SVM_th10 HR, 5.11; 95% confidence interval (CI), 3.04–8.59; P < 0.001; Table 1]. In the model based on OS, all parameters except for chromosome 1p status were significant prognostic markers in the first step of univariate marker assessment, again with the SVM_th10 classifier demonstrating the highest HR (HR SVM_th10, 29.24; 95% CI, 9.77–87.54; P < 0.001; Table 1). Yet, a multivariate comparison of these markers with respect to OS could not be calculated because of the low absolute number of deaths in this non–high-risk patient cohort. However, these findings were further supported by very similar results observed in another multivariate model in which only the potentially best-performing SVM_th10 classifier and the currently established markers age, tumor stage, and chromosomal 1p status were included (data not shown). Finally, we performed an additional multivariate comparison using only the variables (i) risk stratification according to the German neuroblastoma trial NB2004 (low risk vs. intermediate risk), and (ii) the genomic SVM_th10 classifier. Here, the SVM_th10 classifier was the only significant predictor for both EFS and OS (EFS: HR, 5.07; 95% CI, 3.20–8.02; and OS: HR, 25.54; 95% CI, 8.40–77.66; both P < 0.001; Table 1).

To further visualize the contrasting transcriptomic characteristics of patients with neuroblastoma with favorable and unfavorable outcome, we performed a hierarchical cluster analysis using expression data of the 194 classifying features of the SVM_th10 predictor. As shown in Fig. 3, a considerable correlation of distinct expression patterns with both clinical covariables and patients’ outcome can be observed (Fig. 3). To our minds, this graphical visualization further underscores that transcriptome information accurately reflects the individual tumor behavior of patients with neuroblastoma.

Discussion

In recent years, several studies have demonstrated that genomic classification models, in particular those based on gene expression information, more accurately predict outcome of patients with neuroblastoma than conventional risk estimation systems (11–17). In line with these reports, we here also observed that our novel SVM_th10 classifier significantly separated newly diagnosed patients with pretreatment neuroblastoma with divergent outcome both within risk groups defined by the criteria of the current German neuroblastoma trial NB2004 and in additional clinically relevant subgroups of the disease. Intriguingly, we noticed a particular high classification performance in the large cohort of low- and intermediate-risk patients, in which our genomic classifier reliably identified patients with adverse outcome. This finding underlines the capability of gene expression information to predict aggressive tumor behavior where conventional risk stratification fails.

The high accuracy of the present classifier to discriminate patients with non–high-risk neuroblastoma is at least comparable with that observed in other studies, including a 144-gene PAM
Figure 2.
Kaplan-Meier estimates for EFS and OS according to classification by the SVM_th10 predictor for clinically relevant subgroups of patients with neuroblastoma with MYCN nonamplified disease. A, EFS (left) and OS (right) for the cohort of 68 patients with localized (stage I–III) disease ≥18 months of age; B, for 234 patients with localized disease <18 months of age; C, for 49 patients with disseminated stage IV disease <18 months of age; and D, for stage IVS disease (n = 62).
From a clinical perspective, the single-color protocol that was used in the present trial may be beneficial because of reduced diagnostic costs per patient, because no dye-flipped replicates need to be performed. In view of the increasing utilization of next-generation sequencing approaches for the generation of expression profiles, the comparably lower prices of microarray analyses can be considered a major argument supporting the use of microarrays instead of more cost-intensive and more intricate sequencing-based approaches. However, future studies are required to clarify whether RNA sequencing-based approaches, which deliver expression profiles with an unprecedented level of detail, will allow for a higher classification accuracy of patients with neuroblastoma than microarray-based models.

In recent years, accumulating data indicated that treatment with reduced cytotoxic dose intensity is safely possible in patients with neuroblastoma with intermediate risk of death from disease (28, 29). In addition, it was also shown that a high percentage of infant neuroblastoma tumors undergo spontaneous regression with first signs of regression remarkably appearing beyond the first year of life in some patients (30). These data clearly document that the underlying tumor biology of a substantial fraction of neuroblastoma tumors is little to nonaggressive despite being considered as intermediate-risk by current markers. Together with the plausible presumption that the sum of expressed genes in a tumor reflects its biological behavior, it is therefore not surprising that genomic classification approaches more reliably distinguish low- and intermediate-risk tumors with contrasting behavior as shown both by the present study and by previous reports from several other groups (11, 13, 16, 17). In this context, it is reassuring to see that some of the 194 features of our novel SVM_th10 classifier (summarized in Supplementary Table S4) overlap with existing genomic signatures or were reported to have a prognostic impact for patients with neuroblastoma, such as WSB1 (31), CHD5 (32), or CNR1 (33). Although it has to be stressed that selection of a specific feature for the SVM_th10 does not necessarily indicate mechanistic relevance for neuroblastoma tumor behavior, an analysis of the reported functions of the classifying transcripts revealed that several features exert biologic functions that have been shown to reflect neuroblastoma tumor behavior and to contribute to most genomic signatures for neuroblastoma, such as neuronal differentiation (e.g., AGRN, NXP1, or DST; 34, 35) and MYC signaling (Supplementary Table S4).

Integration of gene expression–based classification into a revised risk estimation and therapy stratification system for non–high-risk neuroblastoma patients

To eventually prove in a prospective clinical trial that our molecular classifier will inure to the benefit of those subgroups of patients with neuroblastoma for which our classifier appears to offer a more accurate view of the underlying tumor behavior than current risk estimation approaches, we intend to implement the SVM_th10 classifier into a clinical protocol. Thus, we here propose to revise both risk stratification and treatment concepts of patients with non–high-risk neuroblastoma in the upcoming next German neuroblastoma trial as indicated in Fig. 4. This proposed revision is based on the following findings. First, the SVM_th10 classifier was able to identify patients with highly aggressive tumor biology within the cohort of patients with localized, MYCN nonamplified disease ≥18 months of age, in

Table 1. Univariate and multivariate Cox regression models for patients with non–high-risk neuroblastoma based on EFS and OS considering clinical prognostic markers and the top five genomic classifiers

<table>
<thead>
<tr>
<th>Marker</th>
<th>Available cases</th>
<th>HR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM_th26</td>
<td>n.s.</td>
<td>18.16 (6.97–47.55)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th24</td>
<td>n.s.</td>
<td>18.03 (6.02–53.97)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th22</td>
<td>n.s.</td>
<td>16.81 (5.62–50.32)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (&lt;18 vs. ≥18 mo)</td>
<td>513</td>
<td>29.24 (9.77–87.54)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stage</td>
<td>n.s.</td>
<td>8.55 (3.49–20.95)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>I (ref)</td>
<td>122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II or III</td>
<td>154</td>
<td>3.62 (1.75–7.50)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IV</td>
<td>34</td>
<td>1.34 (0.46–3.93)</td>
<td></td>
</tr>
<tr>
<td>IIV</td>
<td>53</td>
<td>3.87 (1.69–8.87)</td>
<td></td>
</tr>
<tr>
<td>1p (no aberration vs. amb/def)</td>
<td>363</td>
<td>0.4 (0.18–0.90)</td>
<td>0.014</td>
</tr>
<tr>
<td>Histology (F vs. UF)</td>
<td>n.s.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C: Multivariate Cox regression model for non–high-risk patients considering the clinical NB2004 risk estimation system and the SVM_th10 classifier

<table>
<thead>
<tr>
<th>Marker</th>
<th>Available cases</th>
<th>HR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM_th26</td>
<td>413</td>
<td>19.2 (6.42–57.46)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th24</td>
<td>413</td>
<td>18.03 (6.02–53.97)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th22</td>
<td>413</td>
<td>16.81 (5.62–50.32)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th10</td>
<td>413</td>
<td>29.24 (9.77–87.54)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (&lt;18 vs. ≥18 mo)</td>
<td>413</td>
<td>8.55 (3.49–20.95)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stage</td>
<td>Not calculable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I (ref)</td>
<td>126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II or III</td>
<td>176</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIV</td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1p (no aberration vs. amb/def)</td>
<td>363</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histology (F vs. UF)</td>
<td>306</td>
<td>6.31 (1.76–22.61)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

NOTE: A, the Cox regression model based on EFS, B, the model based on OS (univariate models only). The lower number of cases in the model for EFS results from a reduction of those patients for whom all variables were available (n = 363). For each factor, the reference level to which the marker is compared is indicated first and underscored (e.g., “F vs. UF” for histology; “<18 vs. ≥18 mo” for age and “no aberration vs. amb/def of 1p”). Amb/def of 1p was defined according to the criteria of the European Neuroblastoma Quality Assessment Group (18). C, highlights the multivariate Cox regression model based on OS using the variables risk stratification according to the German neuroblastoma trial NB2004 (low risk vs. intermediate risk) and the potentially best-performing genomic predictor SVM_th10.

classifier presented by our group a few years ago (11, 13, 15, 16).

With regard to the technical differences of these two classifiers, it is reassuring to note that the predictive performance of both classifiers is similar although they were generated on different versions of microarrays (previously 2 × 11K arrays vs. 4 × 44K arrays in the present work) and using different experimental protocols and bioinformatics algorithms (two-color analyses vs. single-color analyses; prediction analysis for microarrays (PAM) vs. SVM algorithm). In our minds, the observed analogue classification accuracy of both predictors not only supports the notion that classifiers built from single- or dual-color microarray experiments perform similarly well (19) but also proves the often-doubted robustness and prognostic reliability of array-based expression signatures.

From a clinical perspective, the single-color protocol that was used in the present trial may be beneficial because of reduced diagnostic costs per patient, because no dye-flipped replicates need to be performed. In view of the increasing utilization of next-generation sequencing approaches for the generation of expression profiles, the comparably lower prices of microarray analyses can be considered a major argument supporting the use of microarrays instead of more cost-intensive and more intricate sequencing-based approaches. However, future studies are required to clarify whether RNA sequencing-based approaches, which deliver expression profiles with an unprecedented level of detail, will allow for a higher classification accuracy of patients with neuroblastoma than microarray-based models.
whom a more intensive first-line treatment appears to be justified. Therefore, we propose to consider these patients as high risk and to treat them accordingly in the upcoming German NB trial protocol (Fig. 4). Second, our data support the hypothesis that therapy reduction might safely be possible in those patients of the same subgroup who receive a favorable gene expression–based prediction. Thus, as shown in Fig. 4, we propose that treatment of these patients shall follow either an observational approach (for stage I and II patients) or an intermediate risk therapy of reduced intensity (IRG-reduced for stage III patients). A similar reduction of cytotoxic dose intensity will also be assessed for stage IV, MYCN nonamplified patients <18 months of age, who receive a favorable genomic classification, while patients of this group with an unfavorable classification result will be treated with a nonreduced intermediate-risk therapy. Likewise, stage IVS, MYCN nonamplified patients with unfavorable genomic classification will also receive the nonreduced intermediate-risk therapy (Fig. 4). Finally, no change in the first-line therapy is intended for the small cohort of neuroblastoma patients with localized, MYCN nonamplified disease <18 months of age, who are classified as unfavorable by the SVM_th10 predictor to evaluate the classifier’s accuracy for these patients without a potential treatment bias. However, to prevent putting them at risk, they will continue to initially follow an observational approach but will be treated according to the intermediate-risk protocol in case of progression of disease (Fig. 4). In our opinion, this approach is supported by both the low number of events and the good overall outcome of these patients (5-year OS, 0.86 ± 0.13) as highlighted in Fig. 2B.

With the proposed revision of risk stratification and treatment for patients with non–high-risk neuroblastoma, we intend to both improve outcome of patients whose aggressive tumor behavior is not captured by current risk stratification concepts, and to safely reduce treatment in those patients who are currently considered as intermediate risk but whose molecular profile suggests nonaggressive disease. To visualize the potential benefit of our proposed approach, Fig. 5 highlights the outcome of the 413 non–high-risk patients of this study stratified both according to the present German NB2004 trial protocol (Fig. 5A) and according to the proposed revised approach (Fig. 5B). In total, 382 of the 413 non–high-risk patients could be stratified into either low or intermediate risk of death from disease according to the current NB2004 risk stratification system. The 5-year EFS and OS for patients of the low-risk group (n = 313) was 0.80 ± 0.02 and 0.98 ± 0.01, respectively, and patients of the intermediate-risk cohort (n = 69) had an EFS of 0.69 ± 0.06 and an OS of 0.87 ± 0.04. In contrast, the outcome for patients intended to receive intensified therapy according to the revised protocol (n = 29) was remarkably poor (EFS, 0.19 ± 0.08 and OS, 0.59 ± 0.10) while outcome of the reduced treatment cohort (n = 44) was excellent [EFS, 0.87 ± 0.05 and OS, 1.0; unchanged therapy (n = 340) EFS, 0.83 ± 0.02 and OS, 0.98 ± 0.01]. Although it has to be stressed that patients were not yet treated according to the proposed revised protocol, both the poor outcome in the intensified treatment cohort and the excellent outcome in the reduced treatment subgroup strongly demand a prospective validation of our revised protocol in a clinical trial.

Three clinical trials by Baker and colleagues (28), Rubie and colleagues (29), and Hero and colleagues (30) proved that therapy reduction in different subgroups of non–high-risk patients did not result in inferior patient outcome, and thus underline the feasibility of our concept to offer reduced cytotoxic dose intensity for non–high-risk patients with a favorable
molecular prediction. In contrast, it remains to be determined whether an intensification of cytotoxic treatment for patients whose tumors are molecularly unfavorable will result in reduced rates of relapse or death from disease. Considering the fact that our classifier was trained to discriminate patients whose tumors have the potential to regress spontaneously (as indicated by >1,000 days of EFS without chemotherapy) from those who died despite comprehensive treatment efforts, it is conceivable that an unfavorable prediction may indicate tumors that are incurable despite best currently available therapy. Yet, in our proposed revised treatment protocol, the therapeutic intensity will be increased for unfavorably classified patients with either localized, MYCN nonamplified disease >18 months of age (who will be treated according to the high risk protocol), or with stage IVS, MYCN nonamplified disease (who will be treated according to the standard intermediate-risk treatment protocol). The observation that due to intensified second-line treatment the OS of these two patient subgroups was substantially better than EFS, may argue against the hypothesis that treatment escalation will be ineffective in those subcohorts. We therefore hypothesize that intensified treatment of these patients will improve EFS at least to the level of the OS (50% at 5 years) observed in this study.

Practical issues of performing RNA-based biomarker analysis for neuroblastoma patients

The implementation of our gene expression classifier into a revised risk stratification system of a clinical trial requires consideration of some practical issues. First, the turnaround time that is required for the genomic classification result is an important aspect. In our experience, 3 working days are required to isolate RNA from fresh-frozen tumor material, to assess tumor histology and RNA quality, and to perform the microarray experiments including quality control and running the classification algorithm. Thus, considering potential delays in this workflow (e.g., a repetition of the analysis because of poor experimental quality), a maximum of 7 to 10 working days from diagnosis appears a realistic time period in which expression-based classification results can be obtained for each patient. Of note, this time frame matches the turnaround time that is currently required for the detection of genetic alterations, that is, the determination of the genomic status of MYCN and chromosome 1p. Second, tumor heterogeneity is present in a small fraction of neuroblastoma tumors (36), raising the possibility that the genomic profile might not adequately reflect the underlying tumor behavior. To prevent misclassifications due to tumor heterogeneity, it is intended to perform expression profiles from RNA of at least two separate parts of the tumor specimens of each patient. In case of conflicting results, it is planned to repeat the complete workflow. If the conflicting results persist, it is planned to stratify the patient according to the conventional risk classification approach. This practice may also be applicable to a small fraction of patients for whom no adequate tumor or RNA specimens can be obtained. The latter problem, however, appears to be infrequent in low- and intermediate-risk patients. From our experience, we expect a total dropout rate of approximately 5% to 10% of all patients due to insufficient specimen quality or heterogeneity following implementation of our molecular classifier in the upcoming neuroblastoma trial.

In conclusion, we comprehensively validated a newly built gene expression–based classifier for patients with neuroblastoma using a very large cohort of neuroblastoma tumor samples. Subsequent thorough analyses of this molecular classifier revealed that the highest clinical potential can be assumed for non–high-risk patients, and indicated that the classifier distinguishes two major neuroblastoma subgroups, one with a high potential to regress or differentiate either spontaneously or after .

Figure 4. Schematic overview of the proposed revised risk stratification system for therapy selection of patients with non–high-risk neuroblastoma in the upcoming NB2013-LR/IR trial. F, favorable genomic classification; UF, unfavorable genomic classification; IRG-reduced, intermediate risk with reduced treatment intensity; IRG-standard, intermediate-risk group with standard treatment intensity (as in NB2004); HRG, high-risk group.
limited treatment and another with a high propensity to progress or relapse after current therapeutic concepts. Finally, we propose to evaluate the prognostic power of our molecular classifier in a clinical setting using an innovative revised risk estimation approach, to improve the general outcome of low- and intermediate-risk patients by biomarker-based treatment stratification.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Figure 5.
Kaplan–Meier estimates for EFS and OS of the 413 non-high-risk test set patients of this study according to (A) the present NB2004 (low risk \( n = 313 \) vs. intermediate risk \( n = 69 \)), and (B) the proposed revised NB2013-LR/IR risk assessment outlined in this article (proposed reduced treatment intensity \( n = 44 \) vs. no change of treatment intensity \( n = 340 \) vs. proposed intensification of treatment intensity \( n = 29 \)). Because of a lack of proportional hazards between the curves in A, we applied a test of proportions by comparing the point estimates for EFS and OS at 5 years after diagnosis according to the approach of Klein and colleagues (37). Thereby, we found that for EFS, the difference of 0.095 (95% CI, −0.027 to 0.236) between the two cohorts was not statistically significant \( (P = 0.144) \), while we observed a difference of 0.103 (95% CI, 0.02–0.22) for the point estimates for OS at 5 years after diagnosis with a \( P \) value of 0.004. LR, low-risk; IR, intermediate-risk.
Oberthuer et al.


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Oberthuer, B. Hero, Y. Kahler, A. Engesser, S. Aghazadeh, A. Nakagawara, R. Noguer, J. Theisens, F. Westermann, M. Fischer


Other (laboratory assistant): Y. Kahler

Acknowledgments

The authors thank the United States of America Children's Oncology Group (COG) for providing tumor material and clinical covariates of a substantial number of neuroblastoma patients for this study.

References


Grant Support

The study was supported by the following organizations: the Center for Molecular Medicine Cologne (CMMC) University of Cologne, Germany; Bundesministerium für Bildung und Forschung (BMBF) through the National Genome Research Network plus (NGFNplus, Grant No. 01GS0895 to F. Berthold, B. Brots, M. Fischer, and A. Oberthuer, and 01GS0896 to F. Westermann) and through the MYC-NET [CancerSys 0316676A to F. Westermann]; the European Union (FP7) ASSET 259348 to F. Westermann, the Fordegesellschaft Kinderkrebs-Neuroblastom-Forschung e.V.; the Competence Network Pediatric Oncology and Hematology (KPOH). Furthermore, the study was financed by grants from FIS [ Fondo de Investigaciones Sanitarias] (contract PI10/115), RTICC [Red Temática de Investigación Cooperativa en Cáncer, contracts RD06/0020/0102, RD12/0036/0020 to R. Noguer], and the Instituto Carlos III Madrid and ERDF (European Regional Development Fund), Spain (to R. Noguer), as well as the Italian Neuroblastoma Association (to G.P. Tonini).

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.

Received April 7, 2014; revised July 28, 2014; accepted August 5, 2014; published OnlineFirst September 17, 2014.
Revised Risk Estimation and Treatment Stratification of Low- and Intermediate-Risk Neuroblastoma Patients by Integrating Clinical and Molecular Prognostic Markers

André Oberthuer, Dilafruz Juraeva, Barbara Hero, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-14-0817

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/09/18/1078-0432.CCR-14-0817.DC1

Cited articles
This article cites 34 articles, 14 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/21/8/1904.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/21/8/1904.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.