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 Sterz1, Rene Schmidt3, Andreas Faldum4, Yvonne Kahlert1, Anne Engesser1, Shahab Asgharzadeh6, Robert Seeger4, Miki Ohira5, Akira Nakagawara4, Paola Scaruffi7, Gian Paolo Tonini8, Isabelle Janoueix-Lerosey9, Olivier Delattre9, Gudrun Schleiermacher10, Jo Vandesompele10, Frank Speleman10, Rosa Noguera11, Marta Piqueras11, Jean Bénard12, Alexander Valent12, Smadar Avigad13, Isaac Yaniv13, Richard G. Grundy14, Monika Ortmann15, Chunxuan Shao16, Manfred Schwab16, Roland Eils2, Thorsten Simon1, Jessica Theissen1, Frank Berthold1, Frank Westermann16, Benedikt Brors2 and Matthias Fischer1

1 Children’s Hospital, Department of Pediatric Oncology and Hematology and Center for Molecular Medicine Cologne (CMMC), University of Cologne, Germany
2 Department of Theoretical Bioinformatics (B080), German Cancer Research Center, Heidelberg, Germany
3 Institute of Biostatistics and Clinical Research, University of Muenster, Germany
4 Children’s Center for Cancer and Blood Diseases, Children’s Hospital Los Angeles, USA
5 Laboratory of Cancer Genomics, Chiba Cancer Center Research Institute, Chuoh-ku, Chiba, Japan
6 Division of Biochemistry and Innovative Cancer Therapeutics, Chiba Cancer Center Research Institute, Chuoh-ku, Chiba, Japan
7 Center of Physiopathology of Human Reproduction, Department of Obstetrics and Gynecology, IRCCS San Martino Hospital, National Cancer Research Institute (IST), Genoa, Italy
8 Laboratory of Neuroblastoma, Onco/Hematology Laboratory Department SDB University of Padua, Pediatric Research Institute, Padua35127, Italy
9 Institut Curie, INSERM Unit 830, Paris, France
10 Center for Medical Genetics, Ghent University Hospital, Belgium
11 Department of Pathology, University of Valencia, Spain
12 Department of Tumor Genetics, Institut Gustave Roussy, Villejuif, France
13 Schneider Children’s Medical Center of Israel, Pediatric Hematology Oncology, Petah Tikva; Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel
14 Children’s Cancer Leukaemia Group, University of Leicester, UK
15 Department of Pathology, University of Cologne, Germany
16 Department of Neuroblastoma Genomics (B087), German Cancer Research Center, Heidelberg, Germany

Corresponding Author:
Prof. Dr. med. Matthias Fischer
Children’s Hospital, Department of Pediatric Oncology and Hematology and Center for Molecular Medicine Cologne (CMMC), University of Cologne, Germany
Tel.: +49 221 478 6816
e-mail: matthias.fischer@uk-koeln.de

Running Title: Gene expression based risk stratification of neuroblastoma

Key words: cancer, genomic classifiers, supervised classification, outcome, therapy selection
Abstract

Background:

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

Material and Methods:

Gene expression profiles were generated from 709 neuroblastoma specimens using customized 4x44K 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 (LR: 5-year EFS 0.84±0.02 vs 0.29±0.10; 5-year OS 0.99±0.01 vs 0.76±0.11; both p<0.001) and intermediate-risk patients (IR: 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 LR/IR patients the classifier outperformed risk assessment of the current German trial NB2004 (EFS: HR 5.07, 95%-CI 3.20-8.02, OS: HR 25.54, 95%-CI 8.40-77.66; both p<0.001). Based on these findings, we propose to integrate the classifier into a revised risk stratification system for LR/IR 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.

Conclusion:

Combination of gene expression-based classification and established prognostic markers improves risk estimation of LR/IR neuroblastoma patients. We propose to implement our revised treatment stratification system in a prospective clinical trial.
Statement of translational relevance:

The clinical courses of neuroblastoma are heterogeneous and may range from spontaneous regression to fatal progression. Accordingly, accurate risk estimation of each individual patient at diagnosis is essential for appropriate treatment stratification. Gene expression-based classification has been demonstrated to precisely predict neuroblastoma outcome, however, no such classifier is used in clinical practice to date. Here, we report on the development of a revised risk estimation system for low- and intermediate-risk neuroblastoma patients, which integrates both a highly accurate gene expression-based classifier and established prognostic markers. According to this system, we identified novel subgroups of patients with poor and favorable outcome, who may benefit from intensified and de-escalated therapy regimens, respectively. Implementing the revised risk assessment and treatment stratification system in clinical practice may avoid patient over- and under-treatment in a substantial number of neuroblastoma patients, and will contribute to approaching the goal of biomarker-based individualized medicine in pediatric oncology.
Introduction

Neuroblastoma is the most frequent extracranial solid tumor in children accounting for 8-10% 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 \textit{MYCN} proto-oncogene (MNA)\(^5\),\(^6\), deletion or imbalance of chromosome 1p (del1p) and chromosome 11q (del11q)\(^7\),\(^8\), DNA ploidy\(^9\) and a histopathologic classification system proposed by Shimada\(^10\). Yet, despite elaborate risk estimation systems, stratification is still imperfect resulting in over- or undertreatment of neuroblastoma patients.

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 if genomic classification approaches will eventually lead to improved treatment concepts and outcome of neuroblastoma patients.

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 frontline therapy in newly diagnosed non-high risk neuroblastoma patients.
Material and Methods

Patients

This study comprised 709 newly diagnosed neuroblastoma patients from nine centers in nine countries for whom pre-treatment 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 to 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): stage 1: n=159 (MYCN amplified (MNA), n=5); stage 2: n=116 (MNA, n=4); stage 3: n=92 (MNA, n=15); stage 4: n=259 (MNA, n=94); stage 4S: 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). Chromosomal alterations were defined according to the guidelines of the European Neuroblastoma Quality Assessment Group. Detailed information on patients’ clinical co-variates and gene-expression based classification is given in supplementary table 1.

Gene Expression Analyses and Supervised Classification

Generation of gene-expression profiles: Single-color gene-expression profiles were generated using customized 4x44K oligonucleotide microarrays produced by Agilent Technologies (Palo Alto, CA, USA). Labeling and hybridization was performed as described. 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 pre-processing: raw gene expression data were normalized using the quantile algorithm from limma\textsuperscript{20}. In order to maintain the comparable scale of the training and validation data set, the validation set was pre-processed using the training data set as a reference. All calculations were performed in R v2.14.1\textsuperscript{21}. Subsequently, gene-expression based classifiers were generated.

Classifier training and evaluation: The classifiers were trained using recursive feature elimination method for feature selection\textsuperscript{22} and a linear SVM as classification algorithm. The nested cross validation (5xCV for outer loop, 3xCV for inner loop) was performed with ten repetitions. The average and the variance of classifier performance were evaluated using the following performance measures: accuracy\textsuperscript{23}, sensitivity (SEN), specificity (SPEC), Matthew’s 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\textsuperscript{22}. The initial set of features consisted of all probes (43,291, excluding the control probes). The features were then ranked based on 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\_th44, 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.

\textit{Statistical Analyses}
Kaplan-Meier estimates for EFS and OS were calculated from the time of diagnosis and compared by log-rank test. Recurrence, progression, and death from disease were considered as events. Cox’s regression models for low- and intermediate-risk patients were applied using a stepwise variable selection procedure recommended by Collett to analyze the prognostic value of potentially prognostic factors. The factors age (reference <18 vs ≥18 months), tumor stage (reference 1 vs 2/3 or 4 or 4S), status of chromosome 1p (reference normal vs deletion/imbalance), INPC/Shimada classification (reference favourable vs unfavorable), and the five top-performing genomic classifiers (reference favourable vs unfavourable) 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 (UF, n=22) vs event-free survival >1000 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 which had identical cross-validated classification values: an accuracy of 0.96, a sensitivity of 0.87, a specificity of 0.97 and a Matthew’s correlation coefficient (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 to n=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 (UF, n=138; F, n=187). As highlighted in supplementary table 2, all classifiers demonstrated comparably high classification accuracies (0.94 and 0.95, respectively) and showed a balanced ratio of sensitivity and specificity (supplementary table 2).

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 sub-cohorts 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 NB200425, 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 manuscript’s supplementaries (supplementary table 3).

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 1a¹). 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). By 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 to 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 sub-cohorts of non-high risk patients: (i) stage 1-3, MYCN non-amplified patients

¹ 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.
and (ii) MYCN non-amplified patients with metastasized disease (stage 4 or 4S) <18 months of age. In the first cohort, the classifier correctly identified all patients who succumbed to disease within the sub-cohort of 68 stage 1-3, MYCN non-amplified 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; Figure 2a). Moreover, the classifier was also able to discriminate patients with an unfavorable course of the disease in the cohort of 234 stage 1-3, MYCN non-amplified 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 0.86±0.13; p<0.001; Figure 2b).

Similarly, gene-expression based classification by the SVM_th10 predictor separated patients with divergent outcome in the cohorts of stage 4, MYCN non-amplified 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 p=0.217, respectively; Figure 2c) and stage 4S, MYCN non-amplified 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 p=0.38, respectively; Figure 2d).

In contrast, the classifier was not able to discriminate patients with divergent outcome in the two main sub-cohorts that define high risk disease27: (i) patients with MYCN amplified disease (n=114) and (ii) stage 4, MYCN non-amplified patients >18 months of age2 (n=102). In the sub-cohort of MYCN amplified cases it was observed that almost all patients (113/114) were predicted as unfavorable and had a poor outcome (EFS 0.31±0.05 and OS 0.37±0.05). Only one patient who carried a MYCN amplification was predicted as favorable, and this patient has survived event-free to date (both p-values not significant; Supplementary Figure 1a). Similarly,
the SVM_th10 classifier did not significantly discriminate patients with divergent outcome in the subgroup of stage 4, MYCN non-amplified patients >18 months of age (favorable (n=16): EFS, 0.38±0.12 and OS 0.68±12 versus unfavorable (n=86): EFS 0.34±0.05 and OS 0.54±0.06, both p-values not significant; Supplementary Figure 1b). We therefore concluded that the SVM_th10 classifier has only a limited potential for current high risk patients.

**Multivariate Cox's regression analyses**

As highlighted in supplementary table 2, the top five classifiers performed comparably in predicting outcome of neuroblastoma patients. 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, i.e. MYCN non-amplified patients with stage 1-3 of any age and patients with stage 4S or stage 4 disease <18 months of age, by applying a multivariable Cox regression selection method as proposed by Collett\textsuperscript{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 (SVM_th10 HR 5.11; 95%-CI, 3.04-8.59; p<0.001, Table 1a). 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 hazard ratio (HR SVM_th10 29.24, 95%-CI 9.77-87.54, p<0.001, Table 1b). Yet, a multivariate comparison of these markers with respect to OS could not be calculated due to 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 1c).

To furthermore visualize the contrasting transcriptomic characteristics of neuroblastoma patients 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 figure 3, a considerable correlation of distinct expression patterns with both clinical co-variables and patients’ outcome can be observed (Figure 3). To our minds, this graphical visualization further underscores that transcriptome information accurately reflect the individual tumor behaviour of neuroblastoma patients.
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 neuroblastoma patients than conventional risk estimation systems\textsuperscript{11-17}. In line with these reports, we here also observed that our novel SVM\textsubscript{th10} classifier significantly separated newly diagnosed, pre-treatment neuroblastoma patients 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 non-high risk neuroblastoma patients is at least comparable to that observed in other studies, including a 144-gene PAM classifier presented by our group a few years ago\textsuperscript{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 2x11K arrays vs 4x44K arrays in the present work) and using different experimental protocols and bioinformatics algorithms (two-color analyses vs single-color analyses; 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\textsuperscript{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, since 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-intense 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 neuroblastoma patients than microarray-based models.

In recent years, accumulating data indicated that treatment with reduced cytotoxic dose intensity is safely possible in neuroblastoma patients with intermediate risk of death from disease\textsuperscript{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\textsuperscript{30}. This data clearly documents that the underlying tumor biology of a substantial fraction of neuroblastoma tumors is little to non-aggressive 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\textsuperscript{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 4) overlap with existing genomic signatures or were reported to have a prognostic impact for neuroblastoma patients, such as WSB1\textsuperscript{31}, CHD5\textsuperscript{32} or CNR1\textsuperscript{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 biological 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, NXPH1 or DST)\textsuperscript{34, 35} and MYC signaling (supplementary table 4).
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 neuroblastoma patients 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 non-high risk neuroblastoma patients in the upcoming next German neuroblastoma trial as indicated in Figure 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 non-amplified disease ≥18 months of age, in 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 (Figure 4). Second, our data supports 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 Figure 4, we propose that treatment of these patients shall follow either an observational approach (for stage 1 and 2 patients) or an intermediate risk therapy of reduced intensity (IRG-reduced for stage 3 patients). A similar reduction of cytotoxic dose intensity will also be assessed for stage 4, MYCN non-amplified 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 non-reduced intermediate risk therapy. Likewise, stage 4S, MYCN non-amplified patients with unfavorable genomic classification will also receive the non-reduced intermediate risk therapy (Figure 4). Finally, no change in the first-line therapy is intended for the small cohort of neuroblastoma patients with localized, MYCN non-amplified disease <18 months of age, who are classified as unfavorable by the SVM_th10 predictor in order 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 (Figure 4). In our
opinion, this approach is supported by both the low number of events and the good overall outcome of these patients (5y-OS 0.86±0.13) as highlighted in Figure 2b.

With the proposed revision of risk stratification and treatment for non-high risk neuroblastoma patients 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 non-aggressive disease. To visualize the potential benefit of our proposed approach, Figure 5 highlights the outcome of the 413 non-high risk patients of this study stratified both according to the present German NB2004 trial protocol (Figure 5a) and according to the proposed revised approach (Figure 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=329) 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 et al.28, Rubie et al.29 and Hero et al.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 >1000 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 non-amplified disease >18 months of age (who will be treated according to the high risk protocol) or with stage 4S, MYCN non-amplified disease (who will be treated according to the standard intermediate-risk treatment protocol). The observation that due to intensified second line treatment the overall survival of these two patient subgroups was substantially better than event-free survival may argue against the hypothesis that treatment escalation will be ineffective in those sub-cohorts. We therefore hypothesize that intensified treatment of these patients will improve event-free survival at least to the level of the overall survival (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 turn-around 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 work-flow (e.g. a repetition of the analysis because of poor experimental quality), a maximum of 7-10 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 turn-around time that is currently required for the detection of genetic alterations, i.e. the determination of the genomic status of MYCN and chromosome 1p. Second, tumor heterogeneity is present in a small fraction of neuroblastoma tumor 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 conflictive results, it is planned to repeat the complete work flow. If the conflictive 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 ~5-10% 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 neuroblastoma patients 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 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, in order to improve the general outcome of low and intermediate risk patients by biomarker-based treatment stratification.
Legends:

**Figure 1**: Kaplan-Meier estimates for EFS and OS according to classification by the SVM_th10 predictor. Figure 1a highlights the EFS (left plot) and OS (right plot) for the complete validation cohort of neuroblastoma patients (n=634). Figure 1b-d highlight Kaplan-Meier survival estimates for the sub-cohorts of (b) low risk (n=313), (c) intermediate risk (n=69) and (d) high risk (n=234) patients as defined by the German neuroblastoma trial NB2004. F, favorable; UF, unfavorable.

**Figure 2**: Kaplan-Meier estimates EFS and OS according to classification by the SVM_th10 predictor for clinically relevant subgroups of neuroblastoma patients with MYCN non-amplified disease. Figure 2a presents EFS (left plot) and OS (right plot) for the cohort of 68 patients with localized (stage 1-3) disease ≥18 months of age, Figure 2b for 234 patients with localized disease <18 months of age, Figure 2c for 49 patients with disseminated stage 4 disease <18 months of age, and Figure 2d for stage 4S disease (n=62). F, favorable; UF, unfavorable.

**Figure 3**: Hierarchical cluster analysis and clinical covariates of all 709 patients of the present study using gene-expression data of the 194 features contributing to the SVM_th10 classifier. Lines represent genes, columns represent patients. Gene expression levels are visualized as z-scores ranging from yellow (-10.0) to blue (+10.0). Indicated on top of the clustering are both the grouping into training and test samples (black = training set unfavorable, white = training set favorable, grey = test set samples), the results of the SVM classification (black = unfavorable; white = favorable) and the clinical covariates for OS (white = alive, black = succumbed to disease), EFS (white = alive without event, black = relapse/progression), MYCN (white = non-amplified, black = amplified), age (white <1.5 years, black ≥1.5 years), tumor stage (white = stage 1-2 and 4S, grey = stage 3, black = stage 4) and risk grouping as defined by the criteria of the German NB2004 neuroblastoma trial (white = low-risk, grey = intermediate-risk, black = high-risk).

**Figure 4**: Schematic overview of the proposed revised risk stratification system for therapy selection of non-high risk neuroblastoma patients 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.

**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 manuscript (proposed reduced treatment intensity (n=44) vs. no change of treatment intensity (n=340) vs. proposed intensification of treatment intensity (n=29)). Due to 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, et al. Thereby, we found that for EFS, the difference of 0.095 (95%-CI: -0.027; 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.

**Table 1**: Univariate and multivariate Cox’s regression models for non-high risk neuroblastoma patients based on EFS and OS considering clinical prognostic markers and the top 5 genomic classifiers. Table 1a summarizes the Cox regression model based on EFS, Table 1b 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 months vs. ≥18 months” for age and “no aberration vs. imb/del of 1p”). Imb/del of 1p was defined according to the criteria of the European Neuroblastoma Quality Assessment Group. Table 1c 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.
Grant Information:

The study was supported by the following organisations: 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 FB, BB, MF and AO, and 01GS0896 to FW) and through the MYC-NET (CancerSys 0316076A to FW); the European Union (FP7) ASSET 259348 to FW, the Fördergesellschaft Kinderkrebs-Neuroblastom-Forschung e.V.; the Competence Network Pediatric Oncology and Hematology (KPOH). Furthermore the study was financed by grants from FIS (contract PI10/15), RTICC (Red Temática de Investigación Cooperativa en Cáncer, contracts RD06/0020/0102; RD12/0036/0020 to R.N.) and the Instituto Carlos III Madrid & ERDF (European Regional Development Fund), Spain (to R.N.), as well as the Italian Neuroblastoma Association (to G.P.T.).

Acknowledgments:

We thank the United States of America Children’s Oncology Group (COG) for providing tumor material and clinical co-variates of a substantial number of neuroblastoma patients of this study.
References:


31 Chen, QR, S Bilke, JS Wei, BT Greer, SM Steinberg, F Westermann, et al. Increased WSB1 copy number correlates with its over-expression which associates with increased survival in neuroblastoma. Genes Chromosomes Cancer. 2008; 45(9): 856-62.


Table 1A: Multivariate Cox regression model for non-high risk patients considering established prognostic markers and the top five SVM classifiers based on EFS

<table>
<thead>
<tr>
<th>Marker</th>
<th>available cases</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM_th44</td>
<td></td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>SVM_th26</td>
<td></td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>SVM_th24</td>
<td></td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>SVM_th22</td>
<td></td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>SVM_th10</td>
<td>363</td>
<td>5.11</td>
<td>[3.04; 8.59]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (&lt;18 months vs ≥18 months)</td>
<td></td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>1 (ref)</td>
<td>122</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 or 3</td>
<td>154</td>
<td>3.62</td>
<td>[1.75; 7.50]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>1.34</td>
<td>[0.46; 3.93]</td>
<td></td>
</tr>
<tr>
<td>4S</td>
<td>53</td>
<td>3.87</td>
<td>[1.69; 8.87]</td>
<td></td>
</tr>
<tr>
<td>1p (no aberration vs Imb/del)</td>
<td>363</td>
<td>0.4</td>
<td>[0.18; 0.90]</td>
<td>0.014</td>
</tr>
<tr>
<td>histology (F vs UF)</td>
<td></td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 1B: Univariate Cox regression models for non-high risk patients considering established prognostic markers and the top five SVM classifiers based on OS

<table>
<thead>
<tr>
<th>Marker</th>
<th>available cases</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM_th44</td>
<td>413</td>
<td>18.16</td>
<td>[6.97; 47.35]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th26</td>
<td>413</td>
<td>19.2</td>
<td>[6.42; 57.46]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th24</td>
<td>413</td>
<td>18.03</td>
<td>[6.02; 53.97]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th22</td>
<td>413</td>
<td>16.81</td>
<td>[5.62; 50.32]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SVM_th10</td>
<td>413</td>
<td>29.24</td>
<td>[9.77; 87.54]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (&lt;18 months vs ≥18 months)</td>
<td>413</td>
<td>8.55</td>
<td>[3.49; 20.95]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td>not calculable</td>
</tr>
<tr>
<td>1 (ref)</td>
<td>126</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 or 3</td>
<td>176</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4S</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1p (no aberration vs Imb/del)</td>
<td>363</td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>histology (F vs UF)</td>
<td>306</td>
<td>6.31</td>
<td>[1.76; 22.61]</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Table 1C: Multivariate Cox regression model for non-high risk patients based on OS considering the clinical NB2004 risk estimation system and the SVM_th10 classifier

<table>
<thead>
<tr>
<th>Marker</th>
<th>available cases</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB2004 (LR vs IR)</td>
<td>382</td>
<td>n.s.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM_th10</td>
<td>382</td>
<td>25.54</td>
<td>[8.40; 77.66]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Figure 1a

Event Free Survival

- Favorable N = 368
- Unfavorable N = 245

Time from diagnosis (years)

5-y EFS: 0.82 ± 0.02; 0.34 ± 0.03

Overall Survival

- Favorable N = 379
- Unfavorable N = 255

Time from diagnosis (years)

5-y OS: 0.98 ± 0.01; 0.51 ± 0.03

p < 0.001
Figure 1b

Event Free Survival

- Favorable N = 289
- Unfavorable N = 20

5-y EFS: 0.84 ± 0.02; 0.29 ± 0.10

Overall Survival

- Favorable N = 293
- Unfavorable N = 20

5-y OS: 0.99 ± 0.01; 0.76 ± 0.11

p < 0.001
Figure 1c

Event Free Survival

- Favorable N = 40
- Unfavorable N = 26

p < 0.001

Time from diagnosis (years)

5-y EFS: 0.88 ± 0.06; 0.41 ± 0.10

Overall Survival

- Favorable N = 41
- Unfavorable N = 28

p < 0.001

Time from diagnosis (years)

5-y OS: 1.00; 0.70 ± 0.09
Figure 1d

Event Free Survival

- Favorable N = 30
- Unfavorable N = 198

Time from diagnosis (years)

5-y EFS: 0.63 ± 0.09; 0.33 ± 0.03

Overall Survival

- Favorable N = 30
- Unfavorable N = 204

Time from diagnosis (years)

5-y OS: 0.83 ± 0.07; 0.46 ± 0.04
Figure 2a

Event Free Survival

- Favorable N = 44
- Unfavorable N = 19

Overall Survival

- Favorable N = 46
- Unfavorable N = 22

Time from diagnosis (years)

5-y EFS: 0.90 ± 0.05; 0.14 ± 0.08

5-y OS: 1.0; 0.51 ± 0.11
Figure 2b

Event Free Survival

Time from diagnosis (years)

5-y EFS: 0.84 ± 0.02; 0.56 ± 0.17

Overall Survival

Time from diagnosis (years)

5-y OS: 1.0; 0.86 ± 0.13

Favorable N = 218

Unfavorable N = 9

p = 0.018

Favorable N = 225

Unfavorable N = 9

p < 0.001
Figure 2c

**Event Free Survival**
- Favorable: N = 34
- Unfavorable: N = 14

5-y EFS: 0.88 ± 0.06; 0.64 ± 0.13

**Overall Survival**
- Favorable: N = 34
- Unfavorable: N = 15

5-y OS: 1.0; 0.87 ± 0.09

p = 0.043

p = 0.217
Figure 2d

Event Free Survival

Time from diagnosis (years)

5-y EFS: 0.80 ± 0.06; 0.29 ± 0.17

Overall Survival

Time from diagnosis (years)

5-y OS: 0.96 ± 0.03; 0.86 ± 0.13

Favorable N = 53

Unfavorable N = 7

p < 0.001

p = 0.380
1. Assessment of Stage, Age and MYCN Status

- Stage 1-3 MYCN single-copy
- Stage 4S MYCN single-copy
- Stage 4 MYCN single-copy
- any stage with amplified MYCN

2. Genomic Classification

- F
- UF
- F
- UF
- F
- UF
- F

3. Risk Stratification & Therapy Selection

- Stage 1/2: Observation
- Stage 3: IRG-reduced, HRG, Observation

If relapse or progression, then...
Figure 5a
Figure 5b

Event-Free Survival

Time from diagnosis (years)

Overall Survival

Time from diagnosis (years)

Proposed reduced therapy (n=43)

No change proposed (n=329)

Proposed intensified therapy (n=29)

Proposed reduced therapy (n=44)

No change proposed (n=340)

p < 0.001

p < 0.001
Revised risk estimation and treatment stratification of low- and intermediate-risk neuroblastoma patients by integrating clinical and molecular prognostic markers

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

*Clin Cancer Res* Published OnlineFirst September 17, 2014.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-14-0817</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2014/09/18/1078-0432.CCR-14-0817.DC1">http://clincancerres.aacrjournals.org/content/suppl/2014/09/18/1078-0432.CCR-14-0817.DC1</a></td>
</tr>
<tr>
<td>Author Manuscript</td>
<td>Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.</td>
</tr>
</tbody>
</table>