Oncogene (2005) 24, 7902–7912

www.nature.com/onc

Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling

Alexander Schramm[1], Johannes H Schulte[1], Ludger Klein-Hitpass[2], Werner Havers[1], Hauke Sieverts[3,4], Bernd Berwanger[5], Holger Christiansen[5], Patrick Warnat[6], Benedikt Brors[6], Jürgen Eils[6], Roland Eils[6] and Angelika Eggert*[,1]

[1]Department of Pediatric Oncology and Hematology, University Hospital of Essen, Hufelandstr 55, Essen 45122, Germany; [2]Institute for Cell Biology, University Hospital of Essen, Essen, Germany; [3]Department of Pediatric Oncology, University Children's Hospital, Heidelberg, Germany; [4]Department of Hematology, University Children's Hospital, Heidelberg, Germany; [5]University Children's Hospital, Marburg, Germany; [6]Division of Intelligent Bioinformatics Systems, German Cancer Research Center (DKFZ), Heidelberg, Germany

Neuroblastoma is a common childhood tumor comprising cases with rapid disease progression as well as spontaneous regression. Although numerous prognostic factors have been identified, risk evaluation in individual patients remains difficult. To define a reliable prognostic predictor and gene signatures characteristic of biological subgroups, we performed mRNA expression profiling of 68 neuroblastomas of all stages. Expression data were analysed using support vector machines (SVM-rbf), prediction analysis of microarrays (PAM), k-nearest neighbors (k-NN) algorithms and multiple decision trees. SVM-rbf performed best of all methods, and predicted recurrence of neuroblastoma with an accuracy of 85% (sensitivity 77%, specificity 94%). PAM identified a classifier of 39 genes reliably predicting outcome with an accuracy of 80%. In comparison, conventional risk stratification based on stage, age and MYCN-status only reached a predictive accuracy of 64%. Kaplan–Meier analysis using the PAM classifier indicated a 5-year survival of 20 *versus* 78% for patients with unfavorably *versus* favorably predicted neuroblastomas, respectively ($P = 0.0001$). Significance analysis of microarrays (SAM) identified additional genes differentially expressed among subgroups. *MYCN*-amplification and high expression of *NTRK1/TrkA* demonstrated a strong association with specific gene expression patterns. Our data suggest that microarray-derived data in addition to traditional clinical factors will be useful for risk assessment and defining biological properties of neuroblastoma.

Oncogene (2005) **24,** 7902–7912. doi:10.1038/sj.onc.1208936; published online 15 August 2005

Keywords: neuroblastoma; expression profiling; outcome prediction; microarray

*Correspondence: A Eggert; E-mail: angelika.eggert@uni-essen.de
Received 11 February 2005; revised 31 May 2005; accepted 7 June 2005; published online 15 August 2005

Introduction

Neuroblastoma is a common childhood tumor derived from primitive sympathetic neuroblasts and characterized by its clinical and biological heterogeneity. Contemporary treatment protocols achieve long-term event-free survival (EFS) rates of approximately 30% for advanced stage neuroblastoma (Matthay *et al*., 1999). Numerous studies have demonstrated that molecular and cytogenetic features of clinically aggressive neuroblastomas differ from those observed in tumors associated with a favorable outcome (Schwab *et al*., 2003). Staging and risk assessment currently rely on histopathology and determination of a few clinical and biological risk factors including patient's age, tumor stage, amplification of the *MYCN*-oncogene, allelic deletions of regions in chromosomes 1p, 3p, 11q and 14q as well as differential expression of the neurotrophin receptors TrkA/TrkB and other genes (reviewed in Schwab *et al*., 2003). Additionally, amplification of chromosome 17q is frequent in neuroblastoma and is associated with a poor prognosis (Caron, 1995; Bown *et al*., 1999). *MYCN*-amplification correlates with high tumor stage and poor prognosis (Brodeur *et al*., 1984), and is used worldwide as a treatment stratification parameter in clinical neuroblastoma trials. Accordingly, patients with *MYCN*-amplified tumors receive a more intense treatment.

Accumulating evidence suggests that at least three biological subtypes of neuroblastoma can be distinguished (Maris and Matthay, 1999; Brodeur, 2003). The first group, characterized by hyperdiploid karyotype, lack of structural chromosomal changes, high TrkA expression and low tumor stage, is concomitant with a favorable prognosis. A second group of intermediate prognosis includes patients with near-diploid or tetraploid tumors, low TrkA expression, and structural chromosomal anomalies, but lacking *MYCN*-amplification. The last group exhibits highly malignant clinical behavior, and is characterized by *MYCN*-amplification, 1p deletion and elevated TrkB expression. In particular, the intermediate neuroblastoma type is still a biologically

heterogeneous group, which has to be more clearly subdivided by specific gene signatures.

Consistent with its genetic diversity, the hallmark of neuroblastoma is its clinical heterogeneity. Surgery combined with little or no adjuvant therapy is sufficient for curing children with localized disease (Yamamoto *et al*., 1998) or infants with a unique pattern of disseminated disease, often demonstrating complete spontaneous regression (stage 4s). In contrast, for the majority of older patients with metastatic disease, the outcome is usually fatal due to relapse within 3 years of diagnosis, despite good responses to first-line therapy.

Although risk stratification according to broad biological subgroups has led to marked therapeutic improvements, outcome prediction still remains unsatisfactory. To avoid overtreatment of spontaneously regressing tumors and undertreatment in aggressive neuroblastomas, precise prediction of tumor behavior and risk assessment of each individual patient at diagnosis is a major goal of current neuroblastoma research.

With the recent development of cDNA microarrays, it is now possible to take a genome-wide approach for improving neuroblastoma classification and/or risk prediction. In the current study, oligonucleotide-based microarray analysis was assessed as a tool for better risk stratification of neuroblastoma. This could serve as a prerequisite to a more individualized, patient-tailored therapy based on expression patterns characteristic of biological subgroups and/or associated with therapy response. The main goals of this study were prediction of EFS or early relapse and identification of subgroups defined by clinical or biological features.

Results

Prediction of early recurrence in neuroblastoma

We first assessed the correlation of known major risk factors with overall survival in our patient group. In agreement with previous studies, Kaplan–Meier analysis revealed a significant association of *MYCN*-amplification ($P = 0.008$) and tumor stage ($P = 0.02$) with reduced overall survival, as well as a strong trend for 1p LOH to correlate with overall survival ($P = 0.07$) (data not shown). However, for prediction of early relapse neither of these risk factors alone could be reliably used as determined by Fisher's exact test in our data set (Table 1). Combined conventional risk stratification based on stage, age and MYCN-status according to the German neuroblastoma trial NB-97 only reached a predictive accuracy of 64% for relapse within 2 years of diagnosis (='early relapse') in the patient cohort of this study. Patients with early relapse due to aggressive disease represent the highest challenge in neuroblastoma treatment. If identified at the time of diagnosis, this patient group might benefit most from more aggressive or additional treatment approaches. We, therefore, aimed at defining a reliable predictor for EFS or early relapse by expression profiling. Expression profiles from 68 primary neuroblastoma specimens were obtained using Affymetrix U95Av2 chips. For appropriate adjustment for known predictors, patients with no evidence of disease (NED) within 2 years of diagnosis, but demonstrating *MYCN*-amplification or 1p deletion, were excluded from training sets. These patients have a high risk for later relapse and interfere with prediction accuracy when grouped as favorable outcome. A total of 44 patients (Table 1), comprising 13 patients with a clinically defined event (relapse group) and 31 patients with NED and a sufficient follow-up time of >2 years (NED group), were eligible for predictor building. Exclusion of patients receiving chemotherapy prior to mRNA expression analysis did not improve prediction sensitivity. Using balanced sample numbers ($n = 10$ randomly chosen from both NED and relapse group), SVM-rbf correctly predicted early relapse in 85% of cases in the training set, while PAM resulted in 80% and multiple decision trees in 77% prediction accuracy. Class assignments using k-NN varied markedly between independent rounds of cross validation, resulting in a high standard deviation of prediction accuracy (data not shown). In all, two patients with recurrent disease were consistently misclassified as NED by all methods, resulting in a sensitivity of 77% and a specificity of 94% in terms of relapse prediction and reflecting the homogeneous gene expression patterns in the NED group.

All patients who died of disease were correctly predicted as relapse patients. Closer examination of the clinical follow-up revealed that the two patients wrongly classified as 'NED' had demonstrated localized recurrence of disease, but were in complete remission (CR) or very good partial remission (VGPR), respectively, after a second round of treatment. Patients with NED but remission time <2 years ($n = 15$) were not included in the training set, but were classified as NED for 13 out of 15 stage 1/2 patients and as 'early relapse' for four out of six patients with stage 3/4 when used as an independent test set. In multivariate analyses (logistic regression analysis), the SVM predictor proved to be independent of *MYCN*-amplification and 1p LOH. Interestingly, Kaplan–Meier analysis using the SVM predictor indicated a 5-year survival of 22 *versus* 80% of unfavorably *versus* favorably predicted neuroblastomas, respectively ($P = 0.007$, Figure 1a).

Supervised analysis using PAM with leave-one-out cross validation identified a 41-probe set classifier as the optimal set of genes resulting in the minimum classification errors. These 41 probe sets represented 39 genes (37 known genes and two expressed sequence tags) and reliably distinguished tumors with early relapse from those associated with EFS >2 years with an accuracy of 80% (Figures 1b and 3e). Two of the genes, *NME1* and *PFN2*, were represented by two independent oligonucleotide probes and, hence, acted as internal validation for these genes. We also validated the top-ranked 24 of the 37 known genes by quantitative reverse transcription PCR (data not shown). Only five of the 37 genes have previously been implicated in neuroblastoma biology (*LDHA, LDHB, NME1, NME2* and *HSPCA*). Kaplan–Meier analysis using the PAM classifier indicated a

Table 1 (a) Clinical characteristics of neuroblastoma patients and (b) correlation of clinical parameters with recurrent disease as determined by Fisher's exact test

(a) Characteristics of 68 neuroblastoma patients included in this study

Age (year)	*Number*	*(% of total)*	*No. of profiles used for SAM analyses*
<1	34	50.0	30
>1	34	50.0	28
INSS stasge			
1	20	29.4	20
2	16	23.5	16
3	7	10.3	NA
4	15	22.1	9
4s	10	14.7	9
MYCN status			
Normal	59	86.8	45
Amplified (>10 copies)	9	13.2	9
Status of chromosome 1p			
No LOH 1p	40	58.8	37
LOH 1p	14	20.6	9
Undetermined	14	20.6	NA
Recurrence of disease within 2 years following first-line therapy			*No. of profiles used for predictor building*
No evidence of disease >2 years	31	45.6	10
Recurrent disease	13	17.6	10

(b)

	Early relapse	*No evidence of disease >2 years*	P-*value*
MYCN (normal – amplified)	9–3	26–5	0.66
1p (normal – 1p LOH)	5–6	17–11	0.48
Age (<1 year – >1 year)	3–9	14–17	0.30
Stage 1, 2 – Stage 4	5–5	15–5	0.23
Stage 4 – Stage 4s	5–0	5–7	0.044

INSS staging was performed by local pathologists and confirmed independently. *MYCN*-amplification and 1p-LOH was determined by Southern blotting and fluorescence *in situ* hybridization (FISH), respectively. Patients receiving prior chemotherapy were excluded from SAM analysis. To predict EFS or relapse, patients with NED and follow-up <2 years were excluded. NA: not applied. 'Early relapse' = recurrent disease within 2 years of diagnosis

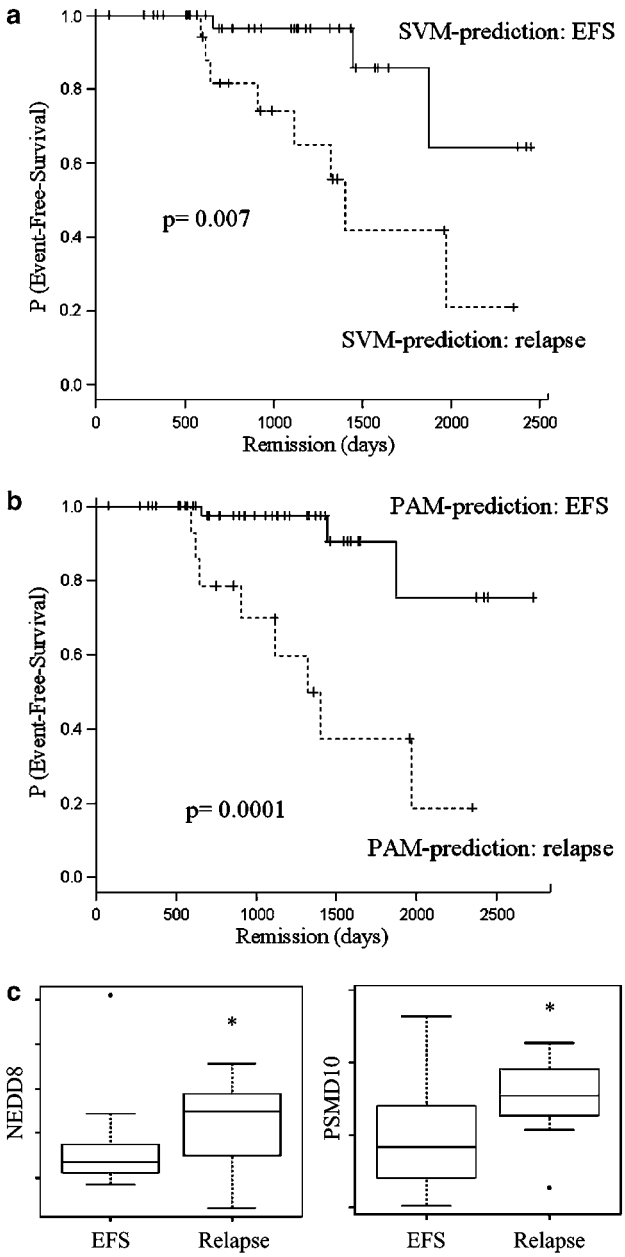

Figure 1 (**a** and **b**) Kaplan–Meier analyses for patients stratified by microarray-based predictions. Analysis was performed on 68 patients included in this study as outlined in Materials and methods. The application of an SVM-based predictor (**a**) and a PAM-based predictor (**b**) trained for predicting event-free survivors (EFS) *versus* relapse patients is shown. (**c**) Box plots of quantitative real-time PCR results. As a representative example, mRNA expression of PSMD10 and NEDD8 (top-ranked genes of the PAM-based predictor) was validated in 32 neuroblastomas, which were not used for PAM model building and had a follow-up time of >2 years. Expression of both genes was significantly different in patients with EFS *versus* patients with relapse ($P=0.01$ for PSMD10 and $P=0.05$ for NEDD8)

5-year survival of 20 *versus* 78% of unfavorably *versus* favorably predicted neuroblastomas, respectively ($P=0.0001$, Figure 1b). Of the patients, 23% with event were correctly predicted only by the PAM classifier, but not by conventional risk stratification. We additionally performed univariate cox regression analysis for each of the 41 probe sets in the PAM signature, and the expression values for all probe sets except one were significantly associated with EFS (log-rank, $P<0.05$). In contrast, when drawing 40 genes randomly from the 12 000 genes on the chip, none was associated with survival (data not shown). We also validated the 15 top-ranked genes of the PAM classifier by real-time PCR in tumors which were not used for predictor building. All

but one gene (*PFN2*) was differentially expressed between EFS and relapse patients as predicted by the PAM classifier. According to univariate analysis, differential mRNA expression in EFS *versus* relapse patients was statistically significant for two genes (*PSMD10* and *NEDD8*, $P<0.05$, Figure 1c).

Interestingly, patients with stage 4 could not be reliably distinguished from those with stage 4s using either SVM-rbf, PAM, k-NN or multiple decision trees. A training set learned by k-NN analysis resulted in a predictor failing to assign stage 4 patients correctly in an independent data set (data not shown).

Correlation of biological and clinical parameters with expression patterns

We used significance analysis of microarrays (SAM; Tusher *et al*., 2001) for molecular characterization of neuroblastoma subgroups of known biological and clinical relevance. Patient numbers for each group are depicted in Table 1. Scatter plots illustrate that biological parameters of prognostic relevance including *MYCN*-amplification (Figure 2a), expression level of *TrkA/NTRK1* (Figure 2b) and 1p LOH (data not shown) were strongly associated with broad specific gene signatures. A strong association with specific gene signatures was also found for local (stage 1 + 2) *versus* metastatic (stage 4) disease (Figure 2c). In contrast, pairwise comparison of other clinically defined groups of prognostic relevance including stages 4 *versus* 4s (Figure 2d) and age <1 year *versus* >1 year (data not shown) resulted in only a few significantly discriminating genes when the false discovery rate (FDR) was adjusted to 0.1 (Table 2). The 30 top-ranked genes distinguishing between the subgroups mentioned above are displayed as heatmaps (Figure 3a–d). Interestingly, TrkA was associated with the highest number of regulated genes compared to all other biological or clinical factors analysed. More than 1000 genes were coregulated with *TrkA/NTRK1* at an FDR <0.1, when the 20 tumors with highest *TrkA/NTRK1* expression were compared to those 20 with lowest *TrkA/NTRK1* expression. Classification of these genes on the basis of gene ontology (GO) revealed neurogenesis, cell adhesion and protein phosphorylation as major biological processes regulated by *TrkA/NTRK1* expression (data not shown).

Discussion

Although microarray technology is now available to many researchers, methods for evaluation and inter-

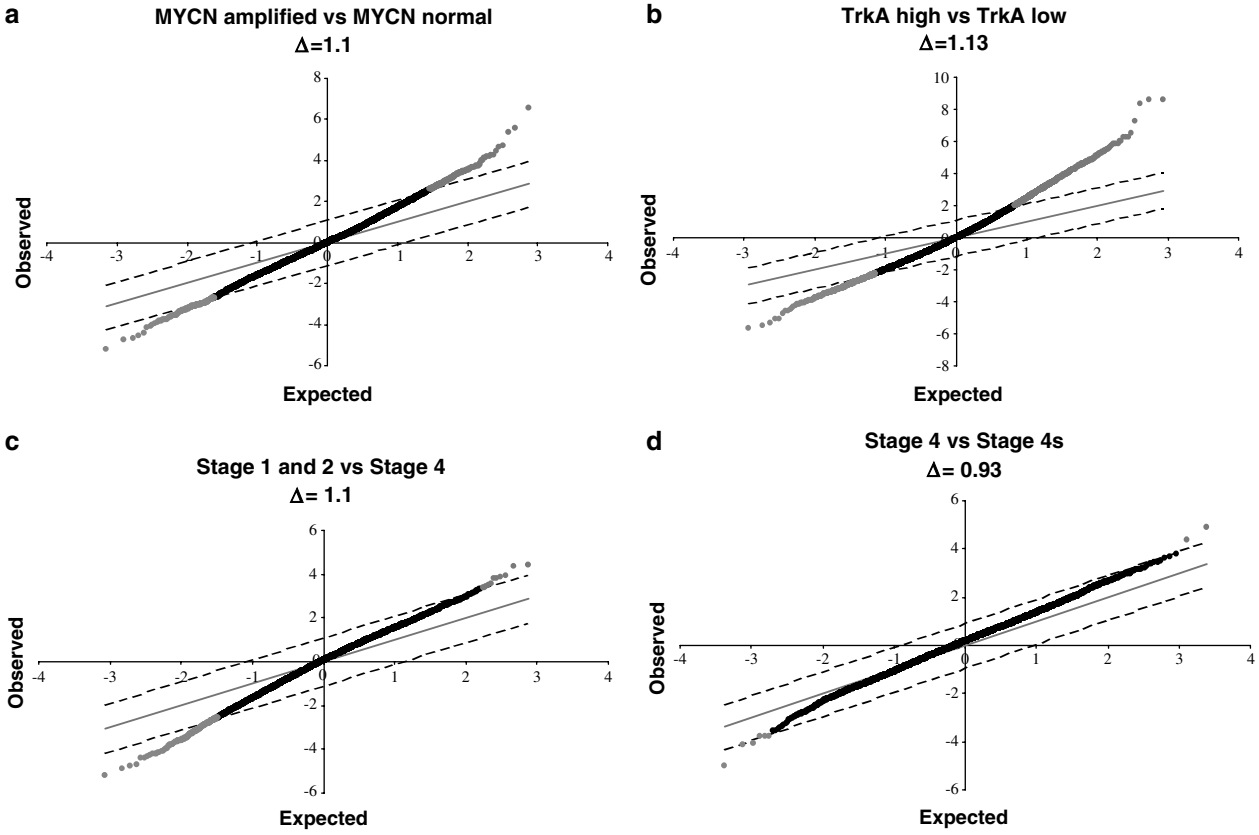

Figure 2 Scatter plots reveal differential expression of genes in pairwise comparisons of neuroblastoma subgroups: (**a**) tumors with *MYCN*-amplification *versus* tumors with normal *MYCN*, (**b**) 20 tumors with the highest TrkA/NTRK1-expression *versus* 20 tumors with the lowest TrkA-expression, (**c**) patients with stage 1 or 2 (localized) *versus* stage 4 (metastatic) neuroblastoma (**d**) patients with stage 4 *versus* stage 4s. Observed: observed relative difference; expected: expected relative difference; Δ values correspond to an FDR of <5% for (**a–c**) and <20% for (**d**) in the respective analyses

Table 2 Absolute gene numbers differentially expressed in pairwise comparison of clinically or biologically defined subgroups

	Biological subgroups				*Clinical subgroups*		
FDR (90th percentile)	*1p LOH/normal*	*MYCN amplified/normal*	*TrkA high/low*	*Stage 1 or 2/Stage 4*	*Stage 4/Stage 4s*	*Age < 1 year/> 1 year*	*Recurrent disease/NED*
<20%	517	1349	3108	956	8[a]	13	0
<10%	155	660	2359	364	0	0	0
<5%	65	360[a]	1661[a]	208[a]	0	0	0

FDR: false discovery rate. [a]Depicts analyses visualized in Figure 2

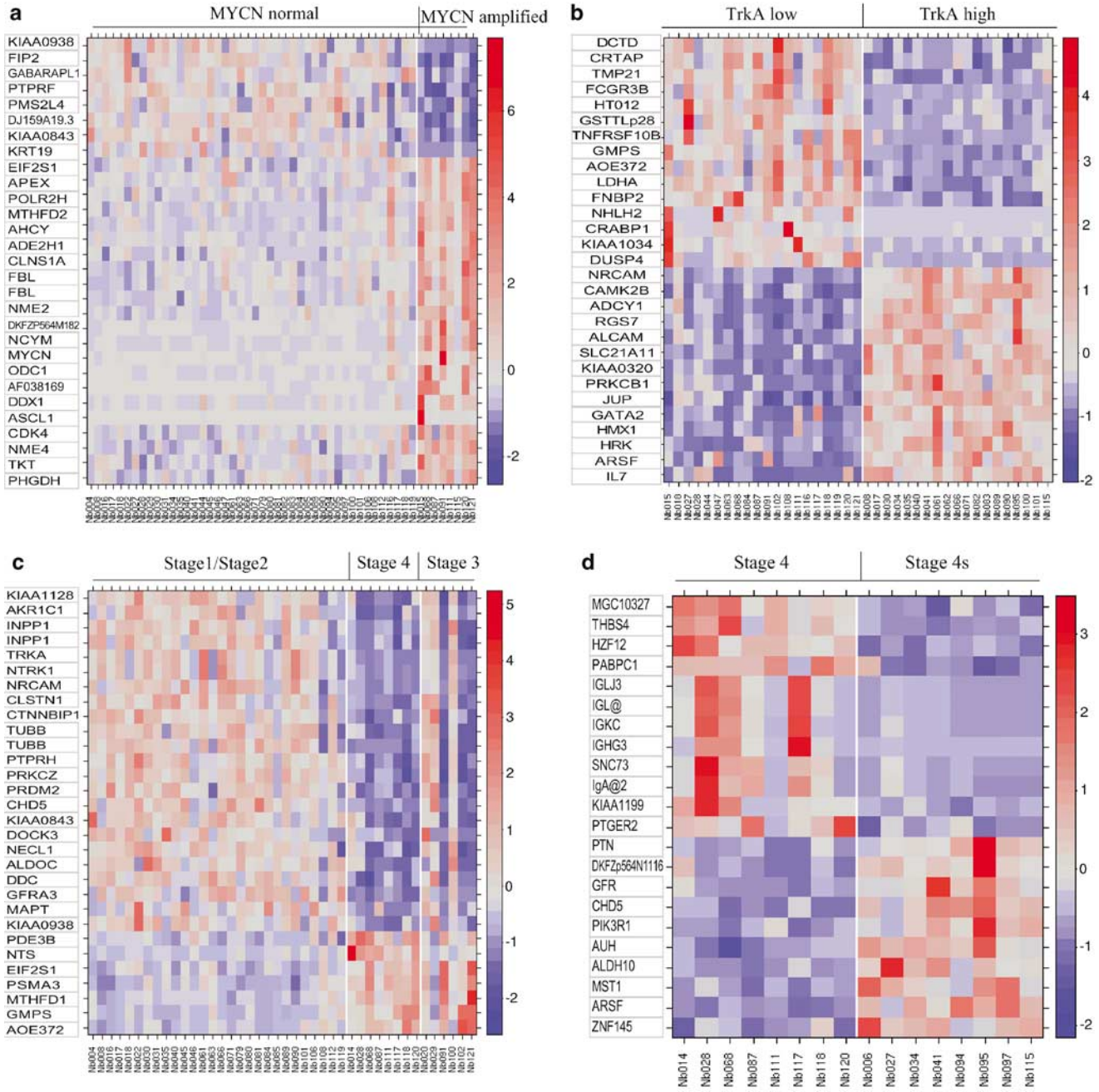

Figure 3 Visualization (heat map) of the 30 top-ranked genes identified by SAM analysis (**a–d**) and the 39 top-ranked genes identified by PAM analysis (**e**) to discriminate tumors with: (**a**) *MYCN*-amplification *versus* normal *MYCN* (genes known to be coamplified with *MYCN* are shown in red); false discovery rate (FDR = 2.1%). (**b**) High TrkA/NTRK1 expression *versus* low TrkA/NTRK1 expression; (FDR = 1.6%). (**c**) Stage 1/2 *versus* stage 4 (FDR = 2.2%). Note that stage 3 tumors have either characteristic signatures of stage 4 or stage 1/2. (**d**) Stage 4 *versus* stage 4s (FDR = 30%); (**e**) EFS *versus* early relapse (FDR was incalculable)

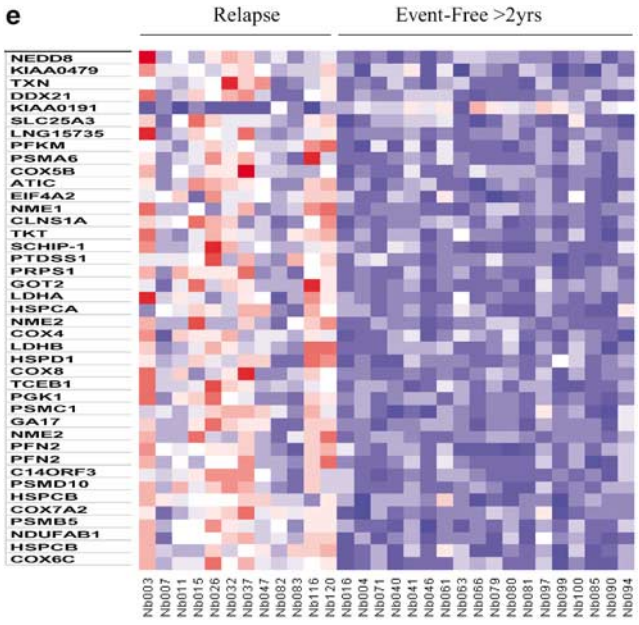

Figure 3 Continued

pretation of microarray data are still evolving. To date, most microarray studies presented from clinical settings concentrated on classification and/or pattern recognition for discrimination among different tumor types or subgroups. Prediction of clinical outcome has only recently come into the focus of microarray studies (Pomeroy *et al*., 2002; van de Vijver *et al*., 2002; Ntzani and Ioannidis, 2003; Nutt *et al*., 2003). Most recently, a data set has been published, in which 19 genes were sufficient to predict outcome in high-risk neuroblastoma (Wei *et al*., 2004). Since these data were obtained in a limited number of tumor samples (28 training samples and 21 test samples), there is clearly a need for testing this predictor prospectively in a larger study cohort. Other published microarray studies for neuroblastoma have dealt with group discrimination by the biological features, *MYCN*-amplification (Alaminos *et al*., 2003), telomerase activity (Hiyama *et al*., 2003) or expression of fyn kinase (Berwanger *et al*., 2002). Additionally, two studies identified gene expression patterns characteristic of favorable and unfavorable neuroblastomas (Hiyama *et al*., 2004; Ohira *et al*., 2003).

Here, we generated expression profiles of 68 tumors using HG-U95Av2 chips (Affymetrix) to characterize the neuroblastoma transcriptome comprising all clinical stages. We focused on molecular definition of clinical and biological subgroups as well as prediction of EFS or recurrent disease within 2 years of first-line therapy ('early relapse'). As neither known clinical risk factors nor SAM analyses were suitable to reliably discriminate between NED and 'early relapse' in our data set, the need for applying prediction tools based on machine learning was evident. We demonstrate the value of SVM-rbf and PAM for outcome prediction based on microarray data, in particular for the small sample size available for this study. Owing to the relatively small number of patients with clinically defined events ($n = 13$), for both methods the predictors were built based on crossvalidation rather than division of the data into training and test groups. As previously described, 10 independent rounds of crossvalidation are rigorous enough to obtain a robust assessment of a predictor even in small sample sets, and are more accurate than leave-one-out crossvalidation (Quinlan, 1993; Witten and Frank, 2002). We evaluated k-NN, SVM-rbf, PAM and multiple decision trees for predictor building in neuroblastoma. k-NN-based class assignment had the disadvantage of being most variable during independent rounds of crossvalidation (data not shown). SVM-rbf, PAM and multiple decision trees were more robust in our model, with SVM-rbf yielding the highest accuracy of class assignment (85%, PAM: 80%, multiple decision trees: 77%). However, class prediction by SVM-rbf does not allow the identification of a specific group of genes characterizing each group, which is important for the confirmation of our data by future independent studies. We therefore applied PAM analysis to generate a reliable classifier consisting of a specific gene signature. The PAM classifier identified here consisting of 39 genes provided a marked improvement in prediction over the conventional risk stratification currently used in the German neuroblastoma trial, as 23% of the patients with event were correctly predicted only by this PAM classifier, but not by conventional risk stratification. In particular, the additional partition of stratified high-risk patients by the PAM classifier into two subgroups according to their survival status has to be validated prospectively in a larger study cohort, before the statement of a major clinical implication of the 39 predictor genes identified here is justified.

Of the 37 known predictor genes (Figure 3e), seven genes have been previously reported to be expressed in neural tissue and/or during neuronal development (*NEDD8, HSPD1, PSMC1, PRPS1, PFK, TKT* and *SCHIP-1*) and five genes have been implicated in tumorigenesis (*NME1, NME2, HSPCA, TCEB1, PGK1* and *SCHIP-1*). A remarkable amassment of predictor genes in the fields, 'proteasome' (*PSMC1, PSMB5, NEDD8, PSMA6, PSMD10, GA17*) and 'heat shock response' (*HSPCA, HSPD1, AHA-1, HSPCB*), suggests that these areas may be promising targets for novel therapeutic approaches. Of additional interest is the chromosomal localization of five predictor genes, which map to regions frequently deleted or amplified in neuroblastoma: *NDUFAB1* (16p12), *NME1* and *NME2* (17q21), *TKT* (3p14) and *HSPCA* (14q32). Most recently, a 19-gene predictor for neuroblastoma was identified using artificial neural networks (Wei *et al*., 2004). Owing to the overall low numbers of neuroblastoma patients being treated each year, both expression-profiling studies were conducted using relatively limited sample sizes (49 primary neuroblastomas in the study by Wei *et al*. (2004) and 69 primary neuroblastomas in this study). Meta-analysis or validation of the predictors identified in both studies in an independent sample set would be most desirable. However, as two different microarray platforms and different data interpretation

methods have been used for the two studies, and nine of the 19 predictor genes identified by Wei *et al*. (2004) were not present on the Affymetrix array used here, a direct comparison of the results for validation is not possible. Expression of *DLK1*, the top-ranked gene of the 19-gene classifier, was not significantly different in the EFS *versus* early relapse group of our study cohort (as measured by Affymetrix arrays as well as quantitative real-time PCR, data not shown). Together with the fact that there is no overlap of the predictor genes identified in both studies, this indicates a strong influence of the choice of microarray platform, data interpretation method and study cohort on the results of expression profiling studies aiming to predict outcome in primary tumors. The obvious differences in the methods and the results of the studies performed here and by Wei *et al*. (2004) strongly emphasize the importance of a prospective analysis of both predictors in an independent larger study cohort. In a very recent paper, Ohira *et al*. (2005) reported on the prognosis of intermediate risk neuroblastomas predicted by expression profiling. Although this study was also performed using a different technological platform, we identified two genes, *AHCY* and *TKT*, as associated with unfavorable outcome in our analysis as well.

For all methods tested in this study, two patients with clinically defined events were consistently misclassified. These patients were wrongly assigned 'EFS', however, after a second round of treatment they had either VGPR or CR according to INSS criteria (Brodeur *et al*., 1993). This implies that patients with clinically defined events but good response to second-line therapy might be indistinguishable from patients with NED by our current predictor. However, both tumors were also not identified by any known clinical factor of adverse prognosis. The failure to identify the elevated relapse risk in these tumors by clinical and molecular approaches suggests that these tumors are biologically more similar to tumors from patients with NED. In contrast, all patients who died of disease were correctly classified as 'relapse' by all methods applied. Testing the predictors with independent data sets from tumors not included in the training sets grouped two of three relapse patients correctly, while nine out of twelve patients with NED and a follow-up >2 years were predicted as NED. Owing to the small number of events these prediction assignments need to be verified after extension of the follow-up time of all patients >3 years. At that time, the data of this study should also be reanalysed with regard to survival itself as an end point. This will enable us to further test the validity and develop our model predictor for clinical use in the future.

Surprisingly, reliable discrimination between patients with stages 4 and 4s tumors (all prior to chemotherapy) in our data set was not feasible by supervised learning methods. This may reflect biological heterogeneity in either group. SAM analysis identified a few differentially expressed genes between stages 4 and 4s only when the FDR was set to values $\geqslant 0.2$. Many of these genes are upregulated in stage 4s tumors and code for integral plasma membrane proteins involved in cell–cell interaction (NRCAM, ALCAM) or for membrane-bound receptors linked to cellular signalling pathways (PTPRH, AGTR1; Figure 4). The angiotensin II receptor (AGTR1) was previously described to interfere with cell proliferation by inactivating ERK2 signaling in a neuroblastoma cell line (Elbaz *et al*., 2000). Nevertheless, the overall similarity of stages 4 and 4s tumor expression profiles was striking. Expression analyses of larger study cohorts are needed to validate our data. We conclude that mechanisms causing progression (stage 4) or regression (stage 4s) of metastatic neuroblastoma are either not detectable by microarray analysis or not represented within the $\sim$12 000 genes present on the HG-U95Av2 array. Genetic components such as different innate immune responses or changes only detectable on the protein level, such as dysregulation in apoptotic or differentiation pathways, may account for the divergent clinical phenotypes of stages 4 and 4s tumors.

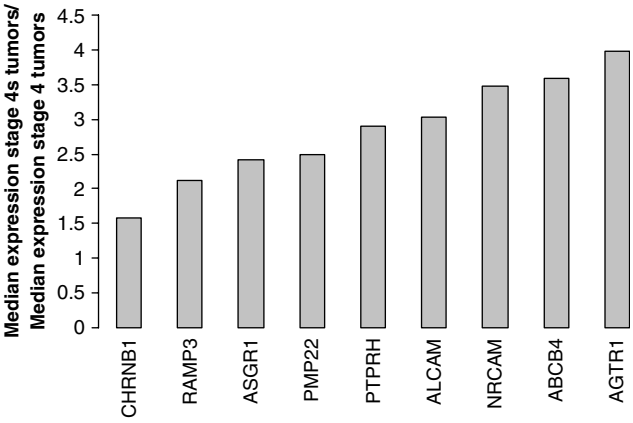

Figure 4 Genes discriminating between stage 4s and stage 4 neuroblastomas with a *q*-value $<10\%$ were used as input for OntoExpress. The most prominent group consisted of genes coding for integral plasma membrane proteins (GO5887). Values are given as the ratio of mean signal intensity in stage 4s *versus* mean signal intensity in stage 4 tumors

Using SAM analysis, we demonstrated a tight association between broad sets of genes and *MYCN*-amplification, 1p LOH or TrkA/NTRK1 expression (Table 2, Figures 2 and 3). *MYCN*-amplification was concomitant with the overexpression of other genes mapping to chromosome 2p24-2p25, including *ODC*, *NCYM* and *DDX1* (Figure 3a), confirming previous data (Godfried *et al*., 2002; Scott *et al*., 2003). This underscores the accuracy of patient assignment as well as the robustness of the data analysis. Interestingly, genes involved in global DNA methylation (*AHCY* and *MTHFR2*) were found to be significantly upregulated in *MYCN*-amplified tumors. Since aberrant methylation of a variety of genes including caspase-8 (Teitz *et al*., 2000) was previously shown in unfavorable neuroblastomas, the study of these genes may provide interesting insights into the role of DNA methylation patterns in neuroblastoma.

As expected, many genes located on chromosome 1 were differentially expressed in tumors with normal 1p

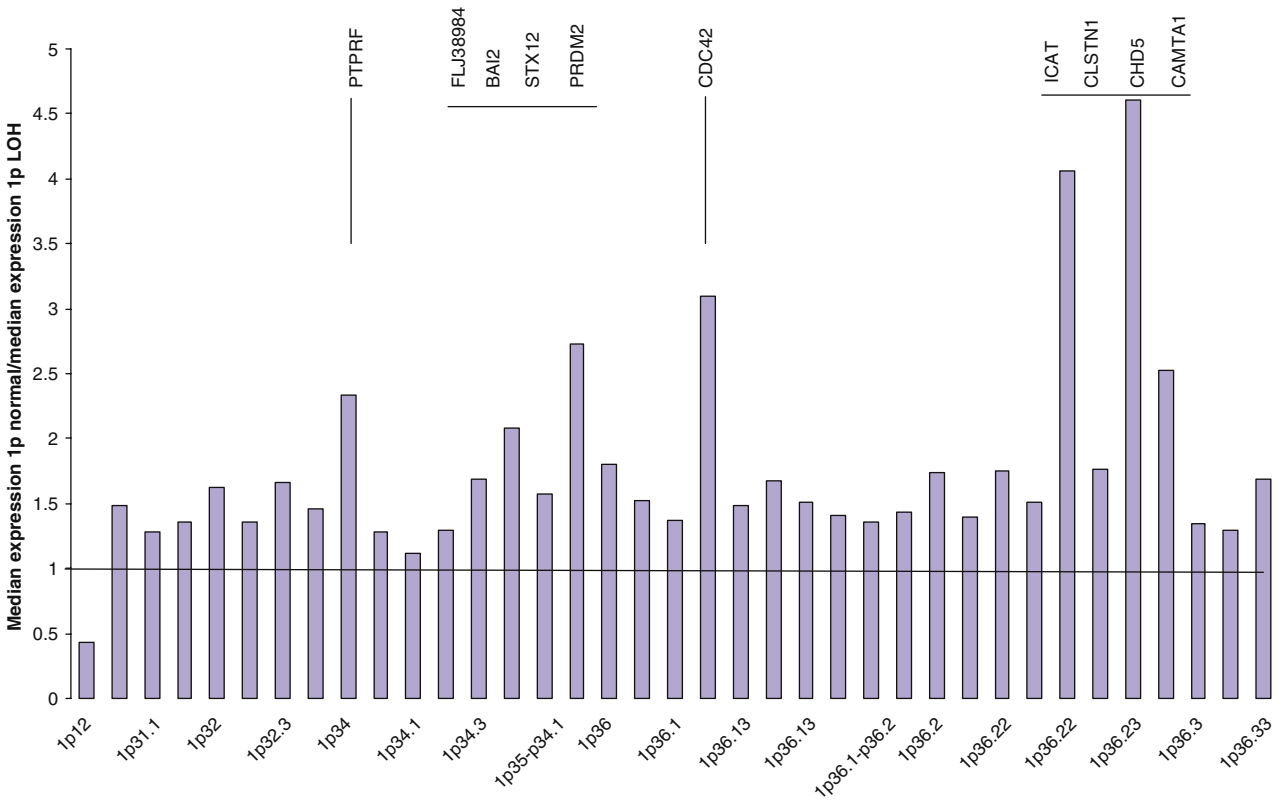

Figure 5 Expression scanning of genes on chromosome 1p in tumors with normal 1p *versus* 1p LOH. Differentially expressed genes identified by SAM analysis were sorted according to their chromosomal position. Values are given as the ratio of mean signal intensity in normal 1p *versus* 1p LOH tumors

compared to 1p LOH. Expression scanning of chromosome 1p (Figure 5) revealed highest expression differences for the recently identified *CHD5* (Thompson *et al*., 2003). This gene was mapped to 1p36.3, a core region of 1p deletion in neuroblastoma. Median gene expression was more than twofold higher in normal 1p *versus* 1p LOH tumors for *PTPRF*/*LAR* (1p34), *PRDM2* (1p36), *CDC42* (1p36.13), *ICAT* (1p36.22), and *CAMTA1* (1p36.3). *CAMTA1* was recently suggested as a putative tumor-suppressor gene in neuroblastoma (Katoh, 2003). In addition to the established major core region of 1p deletion at 1p36.3, at least a second core region has been postulated. Our data also show 1p36.3 as a cluster of elevated gene expression, as well as a second cluster between the *PTPRF* and *CDC42* loci on 1p34 and 1p36.1, respectively (Figure 5). Although tumor-suppressor genes cannot be mapped by expression profiling alone, these findings support the idea of at least two core regions commonly deleted in tumors with 1p LOH. Recently, this approach, which can be termed as 'positional expression mapping', led to the identification of rare 12q amplifications in neuroblastoma (Su *et al*., 2004).

One of the most striking features of our study was the impact of TrkA/NTRK1 expression on the neuroblastoma transcriptome. The tremendous number of diversely expressed genes in TrkA-high *versus* TrkA-low tumors suggests that these groups represent different molecular diseases, as has recently been proposed for a subset of B-cell lymphomas based on expression profiling data (Rosenwald and Staudt, 2002). Striking differences are observed in particular for the cell adhesion molecules, *NCAM* (coding for CD56), *ALCAM* (coding for CD166), *TLN2* and *CNTNAP2*, which are highly upregulated in TrkA-high tumors, while *RELN*, *THBS1* and *FN1* are upregulated in TrkA-low tumors. The contribution of these genes to the biological phenotypes of neuroblastoma cells should be analysed in more detail.

Taken together, this analysis provided new insights into neuroblastoma biology, and defined interesting novel target genes, which need to be functionally characterized in further studies. We presented evidence that SVM-rbf and PAM are suitable tools for the reliable prediction of clinical outcome in neuroblastoma, adding improved accuracy to conventional risk stratification (85 and 80% accuracy, respectively, *versus* 64% accuracy). Both might also prove useful for outcome prediction based on expression profiling in other diseases, in particular if the sample size available for analysis is rather limited. The practical usefulness and predictive power of the 39 top-ranked genes in our PAM classifier in comparison to established clinical risk factors should be evaluated prospectively in a larger, independent study as well as in the same patient cohort once a follow-up time of >5 years is reached. Precise

risk assessment using gene expression data in addition to clinical factors may be most valuable to clinicians faced with treatment decisions.

Materials and methods

Sample acquisition and patient cohort

The 70 primary neuroblastoma specimens from 68 patients analysed in this study were from tumor banks in Heidelberg, Marburg, Cologne or Essen. All patients were diagnosed between 1993 and 2002 and treated according to the German neuroblastoma trials NB90, NB95 or NB97. Median and mean follow-up time was 1108 and 1149 days, respectively, with an interquartile range of 810 days. Written informed consent was obtained from patients or their parents. The selection of tumors for study was based on the availability of a sufficient amount of high-quality RNA. As many patients had been identified by the general neuroblastoma screening (Schilling *et al*., 2002), a high number of cases with favorable biology and prognosis had to be included in this study. The investigators were blinded to all clinical data until mRNA-analysis completion. Patient characteristics are shown in Table 1.

RNA preparation

Representative areas of histologically confirmed, snap-frozen neuroblastomas were cut on dry ice. No preselection for tumor cell fraction or microenvironment was performed. Tumor pieces were mixed with Lysing Matrix D microbeads (Qbiogene, Carlsbad, CA, USA) and 700 μl RTL buffer (Qiagen, Hilden, Germany), and were homogenized using FastPrep FP220 (Qbiogene). RNA was isolated using the RNeasy Mini Kit (Qiagen) according to the manufacturers' instructions. RNA quality was controlled by gel electrophoresis and spectrophotometric measurement of $OD_{260/280}$. Samples of high quality were further processed and hybridized to an HG-U95Av2 array (Affymetrix, Santa Clara, CA, USA).

Array hybridizations

Fragmentation of cRNA, hybridization to HG-U95Av2 microarrays, washing, staining and scanning of the arrays in a GeneArray scanner (Agilent, Palo Alto, CA, USA) were performed as previously described (Schulte *et al*., 2003). Signal intensities and decision calls for further analysis were determined using the GeneChip Microarray Suite (MAS) 5.0 software (Affymetrix). Scaling across all probe sets of a given array to an average intensity of 1000 units compensated for variations in the amount and quality of the cRNA samples and other experimental variables. Reproducibility was ensured by analysis of independent replicates of three tumor samples, which were then excluded from any further data analysis.

Data processing and analysis

Expression data and gene annotations were stored in a relational database, iCHIP, developed at DKFZ Heidelberg, which complies with MIAME (minimal information about a microarray experiment) guidelines. Annotation was obtained from public databases, for example, Unigene, (http://www.ncbi.nlm.nih.gov/entrez/-query.fcgi?db = unigene) and Netaffx (Affymetrix, http://www.affymetrix.com/-analysis/index.affx). Data Mining Tool (DMT) 3.0 (Affymetrix) and scripts called within R environment (http://www.r-project.org/) were used for gene filtering and normalization. Gene ontology annotations were obtained using 'Onto-Express' (Draghici *et al*., 2003).

Significance analysis of microarrays

Normalized and log-transformed (base 2) expression data served as input for SAM. We used our own implementation of the published SAM method (Tusher *et al*., 2001) and the Microsoft Excel Add-In obtained from http://www-stat.stanford.edu/~tibs/SAM (version 1.21). Parameters were set as follows to analyse two-class, unpaired data: 1000 permutations, k-nearest neighbour imputer with 10 neighbours and newly initialized random number seeds for each analysis. FDR was varied between 0.05 and 0.2 for evaluation of changes in the number of significant genes.

Class prediction

We used four different methodological approaches for class prediction: support vector machines (SVM), prediction analysis of microarrays (PAM), neighborhood analysis and multiple-tree models. These supervised learning methods render it possible to automatically build classifiers that distinguish among specimens on the basis of predefined class label information. We used the libsvm implementation by Chang and Lin (2001) for classification by SVM, applying the radial basis function kernel (SVM-rbf). The hyperparameters C and gamma were tuned by cross-validating parameter combinations in a grid search over a two-dimensional parameter space with ranges from 2^5 to 2^{15} and from 2^{-25} to 2^{-15}, respectively.

In order to assess the classifiers produced by SVM and multiple decision trees we performed 10 iterations of a 10-fold crossvalidation to obtain a reliable estimate of the class prediction accuracy. No variable preselection was performed prior to classifier construction. The model-building process included a threefold crossvalidation for hyperparameter tuning (Figure 6), and was repeated in each crossvalidation training set, as recommended by Simon *et al*. (2003). This procedure allows for maximal reliability if no independent data set or only a few samples are available for model validation. For the use of multiple-tree models, we proceeded as described previously (Schoch *et al*., 2002). To build a classifier, a number of decision trees ($n = 30$) was generated using the C5.0 algorithm (Quinlan, 1993). Once a variable was selected for construction of a decision tree, it was excluded from construction of successive trees, thus utilizing each variable only once. The predictions from this set of decision trees were aggregated by a vote-by-majority rule. In order to avoid overfitting, the number of trees used in the resulting classifier was optimized by crossvalidation. Similarly, we applied PAM using the default settings of the program in R ('pamr'), including training, crossvalidation and testing of the model predictor (Tibshirani *et al*., 2002). Additionally, we used an 'off-the-shelf' program using k-nearest neighbors algorithms (GeneCluster 2.0, Whitehead Institute, Cambridge, MA, USA; (Golub *et al*., 1999)). Normalized data were log-transformed (base 10) before predictor building and application. Permutation between $n = 5$ and $n = 25$ was performed for optimal feature number selection (with $k = 3$).

Statistical analysis

We used Fisher's exact test to analyse correlation of clinical and biological factors with recurrence and non-recurrence after first-line therapy. Kaplan–Meier analysis was performed to assess association of *MYCN*-amplification, 1p LOH and

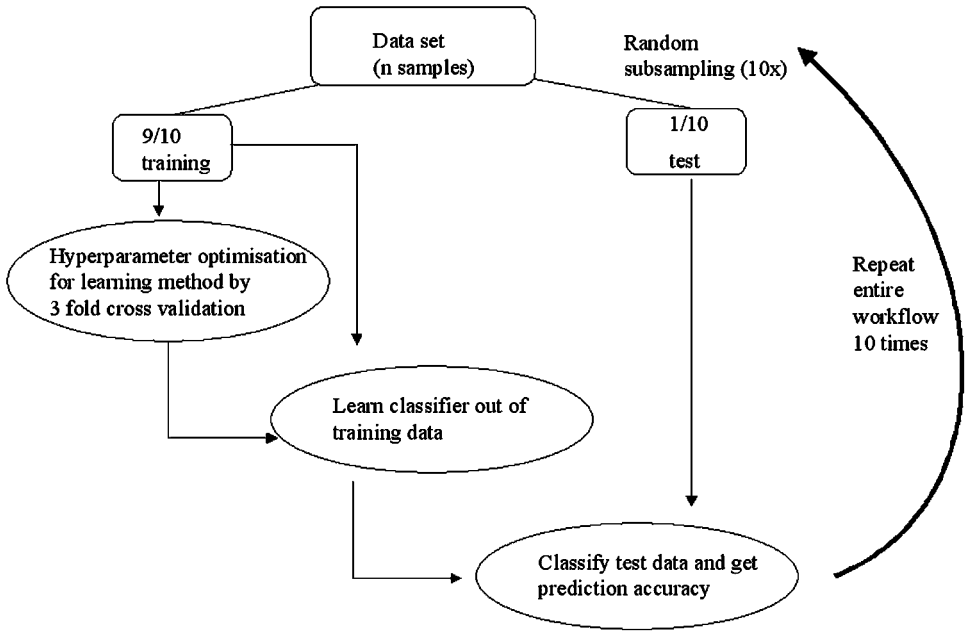

Figure 6 Schematic workflow of predictor building, crossvalidation and optimization by support vector machines

tumor stage with overall survival. We investigated independence between our predictor and known risk factors using logistic regression analysis. All data were analysed using the implementation of the respective tests in '*R*' (source code available at www.r-project.org). A *P*-value <0.05 was considered as significant.

Real-time PCR

For detection of gene expression, 'Assays on demand' (Applied Biosystems') were performed for PSMD10 (Assay ID Hs00829508_s1) and NEDD8 (Hs00362398_m1) using RNA from the neuroblastoma specimen described in this study. Expression values were normalized by geometric averaging of four housekeeping genes (SDH, GAPDH, UBC and HPRT) as proposed by Vandesompele *et al*. (2002). Correlation of gene expression and patient data was calculated using programs within the 'survival'-package of 'R' (www.r-project.org).

Abbreviations

AWD, alive with disease; CR, complete remission; EFS, event-free survival; FDR, false discovery rate; k-NN, k-nearest neighbors; NED, no evidence of disease; PAM, prediction analysis of microarrays; SAM, significance analysis of microarrays; SVM-rbf, support vector machines with a radial basis function kernel; VGPR, very good partial remission.

Acknowledgements

We thank Frank Berthold and Thorsten Simon from the German Neuroblastoma Study Trial Office at the University Children's Hospital of Cologne for providing the clinical patient data as well as a part of the primary tumor material from the neuroblastoma tumor bank of the German competence network 'Pediatric Oncology and Hematology'. This work was funded by a Grant from the Kind-Philipp Stiftung (to AE) and the German National Genome Network (BMBF/NGFN2) to AE and AS.

References

Alaminos M, Mora J, Cheung NK, Smith A, Qin J, Chen L and Gerald WL. (2003). *Cancer Res.*, **63,** 4538–4546.

Berwanger B, Hartmann O, Bergmann E, Bernard S, Nielsen D, Krause M, Kartal A, Flynn D, Wiedemeyer R, Schwab M, Schafer H, Christiansen H and Eilers M. (2002). *Cancer Cell*, **2,** 377–386.

Bown N, Cotterill S, Lastowska M, O'Neill S, Pearson AD, Plantaz D, Meddeb M, Danglot G, Brinkschmidt C, Christiansen H, Laureys G, Speleman F, Nicholson J, Bernheim A, Betts DR, Vandesompele J and Van Roy N. (1999). *N. Engl. J. Med.*, **340,** 1954–1961.

Brodeur GM. (2003). *Nat. Rev. Cancer*, **3,** 203–216.

Brodeur GM, Pritchard J, Berthold F, Carlsen NL, Castel V, Castelberry RP, de Bernardi B, Evans AE, Favrot M, Hedborg H, Kaneko M, Kemshead J, Lampert F, Lee REJ, Look ATh, Pearson AD, Philip T, Roald B, Sawada T, Seeger RC, Tsuchida Y and Voute PA. (1993). *J. Clin. Oncol.*, **11,** 1466–1477.

Brodeur GM, Seeger RC, Schwab M, Varmus HE and Bishop JM. (1984). *Science*, **224,** 1121–1124.

Caron H. (1995). *Med. Pediatr. Oncol.*, **24,** 215–221.

Chang CC and Lin CJ. (2001). *Neural. Comput.*, **13,** 2119–2147.

Draghici S, Khatri P, Bhavsar P, Shah A, Krawetz SA and Tainsky MA. (2003). *Nucleic Acids Res.*, **31,** 3775–3781.

Elbaz N, Bedecs K, Masson M, Sutren M, Strosberg AD and Nahmias C. (2000). *Mol. Endocrinol.*, **14,** 795–804.

Godfried MB, Veenstra M, v Sluis P, Boon K, v Asperen R, Hermus MC, v Schaik BD, Voute TP, Schwab M, Versteeg R and Caron HN. (2002). *Oncogene*, **21,** 2097–2101.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. (1999). *Science*, **286,** 531–537.

Hiyama E, Hiyama K, Nishiyama M, Reynolds CP, Shay JW and Yokoyama T. (2003). *J. Pediatr. Surg.*, **38,** 1730–1734.

Hiyama E, Hiyama K, Yamaoka H, Sueda T, Reynolds CP and Yokoyama T. (2004). *Pediatr. Surg. Int*, **20,** 33–38.

Katoh M. (2003). *Int. J. Oncol.*, **23,** 1219–1224.

Maris JM and Matthay KK. (1999). *J. Clin. Oncol.*, **17,** 2264–2279.

Matthay KK, Villablanca JG, Seeger RC, Stram DO, Harris RE, Ramsay NK, Swift P, Shimada H, Black CT, Brodeur GM, Gerbing RB and Reynolds CP. (1999). *N. Engl. J. Med.*, **341,** 1165–1173.

Ntzani EE and Ioannidis JP. (2003). *Lancet*, **362,** 1439–1444.

Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, Pohl U, Hartmann C, McLaughlin ME, Batchelor TT, Black PM, von Deimling A, Pomeroy SL, Golub TR and Louis DN. (2003). *Cancer Res.*, **63,** 1602–1607.

Ohira M, Morohashi A, Inuzuka H, Shishikura T, Kawamoto T, Kageyama H, Nakamura Y, Isogai E, Takayasu H, Sakiyama S, Suzuki Y, Sugano S, Goto T, Sato S and Nakagawara A. (2003). *Oncogene*, **22,** 5525–5536.

Ohira M, Oba S, Nakamura Y, Isogai E, Kaneko S, Nakagawa A, Hirata T, Kubo H, Goto T, Yamada S, Yoshida Y, Fuchioka M, Ishii S and Nakagawara A. (2005). *Cancer Cell*, **7,** 337–350.

Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, Curran T, Wetmore C, Biegel JA, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis DN, Mesirov JP, Lander ES and Golub TR. (2002). *Nature*, **415,** 436–442.

Quinlan JR. (1993). *Programs for Machine Learning*. C4.5 Morgan Kaufmann: San Francicso, CA.

Rosenwald A and Staudt LM. (2002). *Semin. Oncol.*, **29,** 258–263.

Schilling FH, Spix C, Berthold F, Erttmann R, Fehse N, Hero B, Klein G, Sander J, Schwarz K, Treuner J, Zorn U and Michaelis J. (2002). *N. Engl. J. Med.*, **346,** 1047–1053.

Schoch C, Kohlmann A, Schnittger S, Brors B, Dugas M, Mergenthaler S, Kern W, Hiddemann W, Eils R and Haferlach T. (2002). *Proc. Natl. Acad. Sci. USA*, **99,** 10008–10013.

Schulte JH, Schramm A, Pressel T, Klein-Hitpass L, Kremens B, Eils J, Havers W and Eggert A. (2003). *Klin. Padiatr.*, **215,** 298–302.

Schwab M, Westermann F, Hero B and Berthold F. (2003). *Lancet Oncol.*, **4,** 472–480.

Scott D, Elsden J, Pearson A and Lunec J. (2003). *Cancer Lett.*, **197,** 81–86.

Simon R, Radmacher MD, Dobbin K and McShane LM. (2003). *J. Natl. Cancer Inst.*, **95,** 14–18.

Su WT, Alaminos M, Mora J, Cheung NK, La Quaglia MP and Gerald WL. (2004). *Cancer Genet. Cytogenet.*, **154,** 131–137.

Teitz T, Wei T, Valentine MB, Vanin EF, Grenet J, Valentine VA, Behm FG, Look AT, Lahti JM and Kidd VJ. (2000). *Nat. Med.*, **6,** 529–535.

Thompson PM, Gotoh T, Kok M, White PS and Brodeur GM. (2003). *Oncogene*, **22,** 1002–1011.

Tibshirani R, Hastie T, Narasimhan B and Chu G. (2002). *Proc. Natl. Acad. Sci. USA*, **99,** 6567–6572.

Tusher VG, Tibshirani R and Chu G. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 5116–5121.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH and Bernards R. (2002). *N. Engl. J. Med.*, **347,** 1999–2009.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A and Speleman F. (2002). *Genome Biol.*, **3,** RESEARCH0034.1–RESEARCH0034.11.

Wei JS, Greer BT, Westermann F, Steinberg SM, Son CG, Chen QR, Whiteford CC, Bilke S, Krasnoselsky AL, Cenacchi N, Catchpoole D, Berthold F, Schwab M and Khan J. (2004). *Cancer Res.*, **64,** 6883–6891.

Witten A and Frank E. (2002). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implemenations*. Morgan Kaufmann: San Francicso, CA.

Yamamoto K, Hanada R, Kikuchi A, Ichikawa M, Aihara T, Oguma E, Moritani T, Shimanuki Y, Tanimura M and Hayashi Y. (1998). *J. Clin. Oncol.*, **16,** 1265–1269.