Results

Genome-Scanning for Differentially Methylated CpG Islands. MS-RDA was done using five primary neuroblastomas with a good prognosis and five neuroblastoma cell lines established from cases with a poor prognosis. Seven DNA fragments, derived from CGIs of *PCDHB16*, *PCDHA1*, *HLP*, *DKFZp451I127*, *FLJ37440*, *ZNF297*, and *CYP26C1*, were isolated as methylated in the latter samples. No DNA fragments were isolated as methylated in the former samples. Methylation statuses of (i) 17 CGIs of the *PCDHB* family (detailed structure in Supplementary Fig. 1), (ii) 13 CGIs of the *PCDHA* family, (iii) *HLP* and its pseudogene, and (iv) other four unique CGIs were examined by MSP. This revealed that the *PCDHB* family (5q31), the *PCDHA* family (5q31), *HLP* (3p21) and its pseudogene (1p36), *DKFZp451I127* (5q14), and *CYP26C1* (10q23) were specifically methylated in the latter samples (Fig. 1*A* and *B*).

Close Association between Methylation and Poor Prognosis in 140 Independent Primary Samples. To analyze the significance of the differential methylation of the above five CGI (groups) in primary neuroblastomas, 140 primary samples, all different from the initial five samples, were analyzed by quantitative MSP. When distributions of methylation indices were analyzed (Fig. 2), a clear bimodal distribution was observed for (i) the CGI group in the *PCDHB* family (17 CGIs), (ii) the CGIs of *HLP* and its pseudogene, and (iii) the *CYP26C1* CGI. The results thus indicated that the cases could be classified into two groups, one with high methylation and the other with low methylation. The dose-response relationships between high *PCDHB* methylation and poor prognosis were analyzed by the Cox proportional model using the methylation index as a continuous value, and the association was confirmed with a trend $P < 0.0001$. Normal adrenal medulla had a methylation index of 4%.

According to the bimodal distribution, the effect of high methylation was assessed by dichotomous groups. For the *PCDHB* family, cutoff values of 30%, 40%, 50%, 60%, 70%, and 80% were tested, and HRs of 16.8 [95% confidence interval (95% CI), 4.0-70.9], 22.1 (95% CI, 5.3-93.4; Fig. 3), 13.1 (95% CI, 4.5-37.9), 9.1 (95% CI, 3.8-23.4), 7.0 (95% CI, 3.1-15.8), and 7.8 (95% CI, 3.4-17.6), respectively, were obtained ($P < 0.001$ for all cutoff values). This showed that cases can be classified into two groups with distinct prognoses, and we adopted a cutoff value of 40%, which gave the highest HR, for convenience in the following analysis.

The dose-response relationships were also confirmed for other four CGI (groups), *PCDHA* ($P = 0.004$), *HLP* ($P < 0.0001$), *DKFZp451I127* ($P = 0.02$), and *CYP26C1* ($P < 0.0001$). Cutoff values were similarly tested, and those for *PCDHA*, *HLP*, *DKFZp451I127*, and *CYP26C1* were set at 80%, 10%, 20%, and 70%, respectively, with HRs of 5.7 (95%CI, 1.4-24.0; $P = 0.07$), 21.7 (95% CI, 5.1-91.4; $P < 0.0001$), 3.2 (95% CI, 1.0-10.5; $P = 0.045$), and 8.7 (95% CI, 4.1-18.1; $P < 0.0001$), respectively (Fig. 3).

Existence of the CpG Island Methylator Phenotype in Neuroblastomas. Methylation of the different CGI (groups) had shown close associations with each other (Table 1). When correlation was analyzed as a continuous value, Pearson correlation coefficients between *PCDHB* and *PCDHA*, *HLP*, *DKFZp451I127* and *CYP26C1* were 0.55, 0.70, 0.26 and 0.77, respectively. This showed that multiple CGIs were simultaneously methylated in

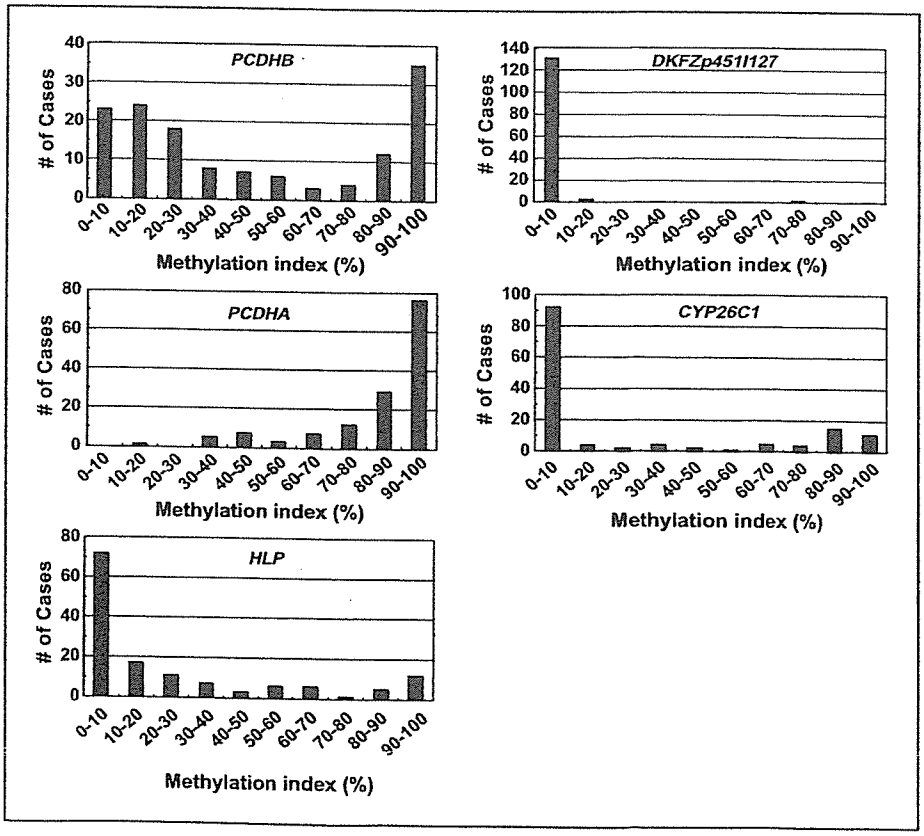

Figure 2. The distribution of methylation indices among the 140 cases analyzed: (*i*) 17 CGIs of the *PCDHB* family, (*ii*) 13 CGIs of the *PCDHA* family, (*iii*) CGIs of *HLP* and its pseudogene, (*iv*) *DKFZp451I127*, and (*v*) *CYP26C1*.

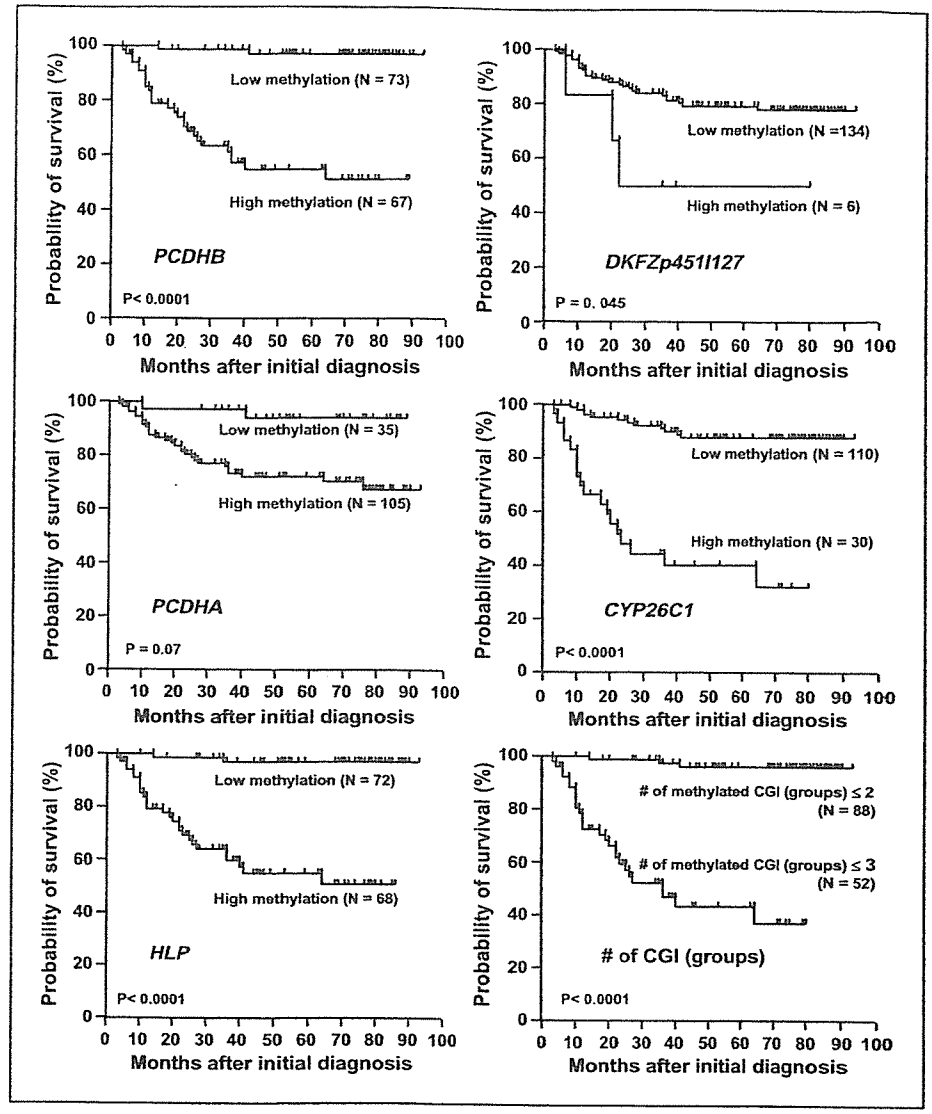

Figure 3. Predictive powers of methylation of the five CGI (groups) identified, and their multiple methylation: (*i*) 17 CGIs of the *PCDHB* family, (*ii*) 13 CGIs of the *PCDHA* family, (*iii*) CGIs of *HLP* and its pseudogene, (*iv*) *DKFZp451I127*, (*v*) *CYP26C1*, and (*vi*) methylation of three of these or more were analyzed by the Kaplan-Meier method using 140 primary samples. The *PCDHB* family, *HLP*, *DKFZp451I127*, *CYP26C1*, and methylation of multiple CGI (groups) had significant influence on survival.

neuroblastomas with a poor prognosis (Supplementary Fig. 2*A*). The simultaneous methylation of (i) 17 CGIs of the *PCDHB* family, (ii) 13 CGIs of the *PCDHA* family, (iii) CGIs of *HLP* and its pseudogene, (iv) *DKFZp451I127* CGI, and (v) *CYP26C1* CGI conformed with the concept of the CpG island methylator phenotype (CIMP; ref. 16).

Associations between CIMP and poor prognosis were examined by defining CIMP as cases with methylation of two CGI (groups) or more, those with three or more, those with four or five, and those with five. When CIMP was defined as cases with methylation of three CGI (groups) or more, the largest association with poor prognosis was observed, with a HR of 25.4 (95% CI, 7.6-84.5; Fig. 3). However, the HR (22.1) given by 17 CGIs of the *PCDHB* gene family approximated to this, and the *PCDHB* methylation level closely correlated with the number of methylated CGI (groups; Supplementary Fig. 2*B*). Therefore, for simplicity of analysis, we defined CIMP in neuroblastomas on the basis of high methylation of the *PCDHB* family, tentatively with a cutoff value of 40%.

Predictive Power of CIMP, Compared with Known Prognostic Factors. Univariate analyses showed that N-*myc* amplification, low *TrkA* expression, DNA diploidy, and an age no younger than 1 year gave HRs of 9.5 (95% CI, 4.4-20.5), 3.9 (95% CI, 1.7-9.3), 4.2 (95% CI, 1.65-10.8), and 12.3 (95% CI, 3.7-41.7). Cases were stratified by these known factors (Table 2). In those without N-*myc* amplification, CIMP also showed an influence with a HR of 12.4 (95% CI, 2.6-58.9), but almost all cases with N-*myc* amplification (37 of the 38 cases) showed CIMP. It was suggested that cases with N-*myc* amplification were contained in the cases with CIMP. CIMP was independent from *TrkA* overexpression, DNA ploidy, and age at diagnosis. Stage seemed to be a stronger prognostic factor. Notably, even when limited to cases in stages III and IV without N-*myc* amplification, which are classified into the intermediate risk group and clinically important, CIMP gave a HR of 4.8 (95% CI, 1.0-23.0; $P = 0.048$).

Multivariate analyses were finally done taking all the five known prognostic factors into account. Although CIMP gave a HR of 5.0 (95% CI, 0.47-52.7), it was not significant ($P = 0.18$), possibly due to limitation in the number of cases.

Table 1. Association between the *PCDHB* methylation and methylation of other CGIs

Variables	Methylation level of *PCDHB* family gene		*P**
	High (≥40%)	Low (<40%)	
No. cases (n = 140)	67	73	
Methylation of CGIs outside promoter regions (n = 140)			
PCDHA gene family (exon 1)†	65/67	41/73	<0.0001
HLP (exons 2-13)‡	52/67	16/73	<0.0001
CYP26C1 (exon 2)§	30/67	0/73	<0.0001
p41Arc (intron 8)	1/67	1/73	0.48
SIM2 (exon 2)	0/67	0/73	
Methylation of CGIs in promoter regions (n = 140)			
DKFZp451I127‖	6/67	0/73	0.011
RASSF1A	51/67	10/73	<0.0001
BLU	25/67	3/73	<0.0001
p16	0/67	0/73	
hMLH1	0/67	0/73	
PCDHB1	0/67	0/73	
TAF7	0/67	0/73	
p41Arc	0/67	0/73	
SIM2	0/67	0/73	

*Fisher's exact test.
†Boundaries for high methylation and low methylation of *PCDHA* gene family were set at 80% of the methylation index.
‡Boundaries for high methylation and low methylation of *HLP* were set at 10% of the methylation index.
§Boundaries for high methylation and low methylation of *CYP26C1* were set at 70% of the methylation index.
‖Boundaries for high methylation and low methylation of *DKFZ-p451I127* were set at 20% of the methylation index.

Effects of PCDHB Methylation on Gene Expression and Chromatin Structure. The CGIs of the *PCDHB* family were located in their gene bodies, whose methylation generally does not block gene transcription (17). The actual effects of methylation on expression were examined for 16 genes of the *PCDHB* family using 10 primary neuroblastomas with low methylation and five primary neuroblastomas with high methylation. The methylation was not associated with loss of expression (a representative result is shown in Fig. 4*A*). The effect of methylation of the *PCDHB16* CGI on the histone modification was further examined by chromatin immunoprecipitation assay. It was found that DNA methylation of the *PCDHB16* CGI did not induce histone H3 lysine 9 methylation or histone H3 deacetylation (data not shown).

Association between CIMP and Promoter Methylation. High methylation of *PCDHB* CGIs, a sensitive surrogate marker of CIMP in neuroblastomas, did not repress gene expression or induce histone modification. This indicated that CIMP is involved in the poor prognosis of neuroblastomas by causing methylation of promoter CGIs, although it is known that promoter CGIs are resistant to *de novo* methylation (18, 19).

Among the five CGI (groups) identified in this study, only that of *DKFZp451I127* was located in a promoter region. Although its methylation was infrequent, the methylation was observed only in neuroblastomas with CIMP (Table 1), and was associated with expression loss (Fig. 4*B*). To make the association clearer, methylation statuses were analyzed for eight additional CGIs in promoter regions. It was shown that methylation of promoter CGIs of *RASSF1A* (3p21) and *BLU* (3p21) was far more frequently observed in neuroblastomas with CIMP (Table 1, $P < 0.0001$). At the same time, there was a preference for CGIs affected by CIMP among CGIs in promoter regions, and also among those outside promoter regions (Table 2).

Discussion

Extensive methylation of multiple CGIs, conforming with the concept of CIMP, was here found specifically present in neuroblastomas with a poor prognosis and could be sensitively detected by focusing on the *PCDHB* family. *PCDHB* methylation did not suppress gene expression or induce histone modification. However, CIMP was associated with promoter methylation of *RASSF1A* and *BLU* genes and one of the mechanisms underlying the poor prognosis of neuroblastomas seemed to be silencing of these and possibly other tumor suppressor genes and genes important for differentiation.

CIMP was originally identified in colon cancers (16), but there has been some dispute over its presence (20). The clear correlation between CIMP and a poor prognosis found here for neuroblastomas was unequivocal and presumably reflects an intrinsic tendency for methylation of CGIs. This is because, first, neuroblastomas have a much shorter history than colon cancers, and the accumulated number of methylated CGIs in neuroblastomas is expected to parallel the speed of occurrence of

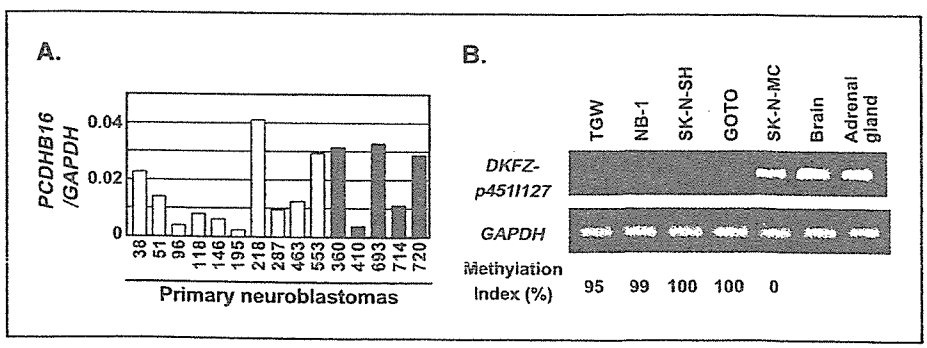

Figure 4. Effects of methylation of the *PCDHB* family and *DKFZp451I127* on gene expression. *A*, *PCDHB16* expression was analyzed by quantitative RT-PCR in 10 primary samples with low methylation (*open columns*) and five primary samples with high methylation (*closed columns*), and no difference was observed between the two groups. *B*, silencing of *DKFZp451I127* by methylation of its promoter CGI. The CGI was methylated in four cell lines, TGW, NB-1, SK-N-SH, and GOTO, whereas it was unmethylated in one cell line, SK-N-MC. *DKFZp451I127* was expressed in SK-N-MC, but not expressed at all in the four cell lines with the promoter methylation.

Table 2. HRs of death by *PCDHB* methylation status in subgroup of known prognostic factors

Stratified by		*PCDHB* methylation	No. cases	No. deaths	HR* (95% CI)	$P^{\dagger}$
Overall (n = 140)		High	67	1	22.1 (5.3-93.4)	< 0.0001
		Low	73	2	1	
N-*myc* amplification (n = 140)	No	High	30	8	12.4 (2.6-58.9)	0.002
		Low	72	2	1	
	Yes	High	37	20	NE	—
		Low	1	0		
TrkA overexpression (n = 130)	Yes	High	20	6	18.3 (2.2-152.6)	0.007
		Low	49	1	1	
	No	High	40	19	NE	
		Low	21	0		
DNA ploidy (n = 125)	Aneuploid	High	17	5	18.3 (2.1-156.7)	0.008
		Low	49	1	1	
	Diploid	High	38	17	NE	
		Low	21	0		
Clinical stages (n = 140)	Stages I, II, and IVS	High	8	0	NE	—
		Low	52	0		
	Stages III and IV	High	59	28	7.4 (1.8-31.3)	0.006
		Low	21	2	1	
Age at diagnosis (n = 140)	<1	High	11	3	NE	
	≥1	Low	59	0		
		High	56	25	4.5 (1.1-18.9)	0.043
		Low	14	2	1	

*HR of death for a case with high *PCDHB* methylation compared with a case with low methylation. NE shows not estimable due to no events in at least one category.
†Significance level for a high *PCDHB* methylation to low methylation using Cox proportional model.

methylation. Second, methylation of the *PCDHB* family did not affect gene expression, and there should have been no selection of cells with the *PCDHB* methylation, in contrast to the case of promoter methylation of tumor suppressor genes. Investigation into the mechanism of the intrinsic tendency for methylation of multiple CGIs is necessary. Furthermore, alleviation of the intrinsic tendency could block progression of neuroblastomas and have potential therapeutic value.

Among the six CGI (groups) outside promoter regions analyzed here, CIMP in neuroblastomas preferentially affected four CGI (groups); those of the *PCDHB* family, the *PCDHA* family, *HLP*, and *CYP26C1*. Unexpectedly, three CGIs that are known to be frequently methylated in human colon cancers with CIMP, MINT1, MINT2, and MINT17 (16) were not methylated in neuroblastoma cell lines (data not shown). Among the nine CGIs in promoter regions analyzed, CIMP in neuroblastomas affected only three, those of *RASSF1A*, *BLU*, and *DKFZp451I127*. The nine CGIs were selected based upon previous reports as tumor suppressor genes (*RASSF1A*, *BLU*, *p16*, and *hMLH1*; refs. 21–23), the chromosomal location flanking the *PCDHB* family (*PCDHB1* and *TAF7*), our previous report on the fidelity in inheriting methylation patterns (*p41Arc* and *SIM2*; ref. 19), and the findings here (*DKFZp451I127*). Because gene expression and possibly chromatin structures affect the frequency of *de novo* methylation (24, 25), the available data suggest that CGIs useful to sensitively detect CIMP might vary according to the tumor type.

The influence of CIMP on prognosis was here found to be comparable to that of the currently most reliable marker, N-*myc* amplification, and stronger than *TrkA* overexpression and DNA ploidy on univariate analysis. Subgroup analysis showed that the influence was independent of *TrkA* overexpression, DNA ploidy and age at diagnosis and CIMP had influence even in cases without N-*myc* amplification and in advanced stages. These points strongly indicated CIMP to be a promising new prognostic marker. However, the cutoff values adopted here are tentative, and the HRs obtained could have been overestimated. A validation study using independent samples is necessary for further evaluation. The fact that cases with CIMP contained almost all the cases with N-*myc* amplification suggested that a common molecular mechanism caused both alterations, or that CIMP may lead to N-*myc*

amplification. Whatever the case, the findings might provide clues to molecular mechanisms of neuroblastoma development.

In summary, the present study showed that CIMP is present specifically in neuroblastomas with poor prognosis and that can be sensitively detected by focusing on *PCDHB* methylation. CIMP seems to be a promising new prognostic marker, and its evaluation and investigations into the mechanisms underlying CIMP in neuroblastomas seem warranted.

Acknowledgments

Received 7/27/2004; revised 11/14/2004; accepted 11/24/2004.

Grant support: Grant-in-aid for the Third-term Cancer Control Strategy Program from the Ministry of Health, Labour, and Welfare, Japan and Research Resident Fellowship from the Foundation for Promotion of Cancer Research (M. Abe).

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

We thank Drs. E. Okochi-Takada and G. S. Goldberg for critical reading of the article and the institutions for participation in the collection of clinical materials.

References

1. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet 2002;3:415–28.
2. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA hypomethylation leads to elevated mutation rates. Nature 1998;395:89–93.
3. Kondo Y, Kanai Y, Sakamoto M, et al. Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis-A comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma. Hepatology 2000;32:970–9.
4. Ushijima T, Morimura K, Hosoya Y, et al. Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. Proc Natl Acad Sci U S A 1997;94:2284–9.
5. Kaneda A, Takai D, Kaminishi M, Okochi E, Ushijima T. Methylation-sensitive representational difference analysis and its application to cancer research. Ann N Y Acad Sci 2003;983:131–41.
6. Takai D, Yagi Y, Wakazono K, et al. Silencing of *HTR1B* and reduced expression of *EDN1* in human lung cancers, revealed by methylation-sensitive representational difference analysis. Oncogene 2001; 20:7505–13.
7. Kaneda A, Kaminishi M, Yanagihara K, Sugimura T, Ushijima T. Identification of silencing of nine genes in human gastric cancers. Cancer Res 2002;62:6645–50.
8. Miyamoto K, Asada K, Fukutomi T, et al. Methylation-associated silencing of heparan sulfate *D-glucosaminyl 3-O-sulfotransferase-2* (*3-OST-2*) in human breast, colon, lung and pancreatic cancers. Oncogene 2003;22:274–80.
9. Hagihara A, Miyamoto K, Furuta J, et al. Identification of 27 5′ CpG islands aberrantly methylated and 13 genes silenced in human pancreatic cancers. Oncogene 2004;23:8705–10.
10. Brodeur GM. Neuroblastoma: biological insights into a clinical enigma. Nat Rev. Cancer 2003;3:203–16.
11. Schwab M, Westermann F, Hero B, Berthold F. Neuroblastoma: biology and molecular and chromosomal pathology. Lancet Oncol 2003;4:472–80.
12. Nakagawara A, Arima-Nakagawara M, Scavarda NJ, et al. Association between high levels of expression of the *TRK* gene and favorable outcome in human neuroblastoma. N Engl J Med 1993;328:847–54.
13. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 2003;33:245–54.
14. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 2002;3:662–73.
15. Kaneda A, Kaminishi M, Sugimura T, Ushijima T. Decreased expression of the seven ARP2/3 complex genes in human gastric cancers. Cancer Lett 2004; 212:203–10.
16. Toyota M, Ahuja N, Ohe-Toyota M, et al. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 1999;96:8681–6.
17. Gonzalgo ML, Hayashida T, Bender CM, et al. The role of DNA methylation in expression of the *p19/p16* locus in human bladder cancer cell lines. Cancer Res 1998;58:1245–52.
18. Nguyen C, Liang G, Nguyen TT, et al. Susceptibility of nonpromoter CpG islands to *de novo* methylation in normal and neoplastic cells. J Natl Cancer Inst 2001; 93:1465–72.
19. Ushijima T, Watanabe N, Okochi E, et al. Fidelity of the methylation pattern and its variation in the genome. Genome Res 2003;13:868–74.
20. Yamashita K, Dai T, Dai Y, Yamamoto F, Perucho M. Genetics supersedes epigenetics in colon cancer phenotype. Cancer Cell 2003;4:121–31.
21. Agathanggelou A, Dallol A, Zochbauer-Muller S, et al. Epigenetic inactivation of the candidate 3p21.3 suppressor gene *BLU* in human cancers. Oncogene 2003; 22:1580–8.
22. Takita J, Hayashi Y, Nakajima T, et al. The *p16* (*CDKN2A*) gene is involved in the growth of neuroblastoma cells and its expression is associated with prognosis of neuroblastoma patients. Oncogene 1998; 17:3137–43.
23. Harada K, Toyooka S, Maitra A, et al. Aberrant promoter methylation and silencing of the *RASSF1A* gene in pediatric tumors and cell lines. Oncogene 2002; 21:4345–9.
24. De Smet C, Loriot A, Boon T. Promoter-dependent mechanism leading to selective hypomethylation within the 5′ region of gene *MAGE-A1* in tumor cells. Mol Cell Biol 2004;24:4781–90.
25. Richards EJ, Elgin SC. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell 2002;108:489–500.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cancer Letters 228 (2005) 5–11
CANCER Letters
www.elsevier.com/locate/canlet

Mini-review

A review of DNA microarray analysis of human neuroblastomas

Miki Ohira[a], Shigeyuki Oba[b], Yoko Nakamura[a], Takahiro Hirata[c], Shin Ishii[b], Akira Nakagawara[a,*]

[a]*Division of Biochemistry, Chiba Cancer Center Research Institute, 666-2 Nitona, Chuoh-ku, Chiba 260-8717, Japan*
[b]*Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma 630-0192, Japan*
[c]*Hisamitsu Pharmaceutical Co. Inc., Tokyo 100-622, Japan*

Received 20 December 2004; accepted 12 January 2005

Abstract

Neuroblastoma (NBL) is an enigmatic tumor with heterogeneous clinical behaviors including maturation, regression, and aggressive growth. Despite recent progress in therapeutic strategies against advanced NBL, long-term outcomes still remain very poor. The prediction of cancer prognosis is one of the most urgent demands to initiate the suitable treatment of NBL. Recent papers have demonstrated that cancers can be diagnosed on the basis of gene expression profiling. We have been proceeded NBL cDNA project to collect a large number of genes expressed in NBLs, to identify the genes differentially expressed between favorable and unfavorable NBLs, and to make an NBL-proper cDNA chip for large-scale analysis of NBL tumors. Computational analysis of gene expression data in NBLs identified many prognosis-related genes and provided a classifier to predict the patient prognosis with high efficiency. Conversion of these findings into better diagnosis and treatment is now underway. Thus, molecular profiling of NBL has become a feasible tool for clinical applications.

Keywords: Neuroblastoma; Expression profile; Microarray; Diagnosis; Prognosis prediction; Differential expression

1. Introduction

Neuroblastoma (NBL) is one of the most frequent solid cancers in young children and has variable clinical and biological characteristics [1]. Favorable type of tumors frequently regress spontaneously, while unfavorable type of tumors are often resistant to intensive chemotherapy and lead patients to fatal outcome. The poor prognosis of NBL patients depends on age at diagnosis (older than 12 months), advanced tumor stage (3 or 4), presence of *MYCN* amplification, low *TRKA* expression, unfavorable histology, diploidy, and chromosomal loss of 1p36 in tumors [2]. However, even these markers sometimes fail to classify the aggressiveness of tumors, especially for the intermediate type of NBL (stage 3 or 4 patients with single copy of *MYCN*). Moreover, some *MYCN*-amplified tumors can be distinguished by a better response to the combined treatment resulting in a better prognosis for

* Corresponding author. Tel.: +81 43 264 5431; fax: +81 43 265 4459.
E-mail address: akiranak@chiba-cc.jp (A. Nakagawara).

0304-3835/$ - see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.canlet.2005.01.054

Table 1
Genes whose expression is differential between neuroblastoma subsets and related to patient prognosis

Genes	Definition	Pattern[a]	Reference
TRKA	neurotrophic tyrosine kinase, receptor, type 1	F>UF	[3]
CD44	CD44 antigen	F>UF	[6]
PTN	Pleiotrophin	F>UF	[7]
CDC10	cell division cycle 10	F>UF	[10]
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene	F>UF	[9]
XCE	endothelin-converting enzyme-like 1	F>UF	[17]
NLRR3	neuronal leucine-rich repeat 3	F>UF	[18]
TTL	tubulin tyrosine ligase	F>UF	[19]
BMCC1	novel putative apoptosis-related gene with BCH domain	F>UF	
FOG2	Friend of GATA protein 2	F>UF	[16]
NEDL1	NEDD4-like ubiquitin ligase 1	F>UF	
NEDL2	NEDD4-like ubiquitin ligase 2	F>UF	
GABARAP	gamma-aminobutyric acid receptor-associated protein gene	F>UF	[20]
GABA(A) family	GABA(A) receptor subunit gene family	F>UF	[20]
TRKB	neurotrophic tyrosine kinase, receptor, type 2	F<UF	[4]
hTERT	telomerase reverse transcriptase	F<UF	[5]
NM23A	non-metastatic cells 1 (*NM23-H1*)	F<UF	[12]
NM23B	non-metastatic cells 2 (*NM23-H2*)	F<UF	[13]
BIRC5	baculoviral IAP repeat-containing 5 (survivin)	F<UF	[11]
PPM1D	protein phosphatase 1D	F<UF	[14]
NLRR1	neuronal leucine-rich repeat 1	F<UF	[18]
LMO3	LIM-only protein 3	F<UF	

[a] F;favorable NBL, UF;unfavorable NBL.

the patient. Therefore, additional potent markers for predicting the NBL prognosis should be discovered to construct a more effective as well as less toxic therapeutic strategy. Recent works have demonstrated that cancers can be diagnosed on the basis of gene expression profiling using cDNA microarrays with calculation by computational algorithms. This process needs (1) mass collection or identification of genes expressed in NBLs, (2) construction of a DNA chip and analysis of tumor samples, (3) computational analysis of gene expression data and identification of prognosis-related genes, and (4) conversion of these findings into better diagnosis and treatment. In this review, we discuss the recent attempts of large-scale molecular profiling of NBL and their future applications to the clinic.

2. Prognostic markers for neuroblastoma

In addition to conventional prognostic markers such as age, INSS stage, *MYCN* copy number, histology, and DNA ploidy, expression levels of several genes have recently been added as new indicators. They include inverse relationship of *TRKA* and *TRKB* expression [3,4], *telomerase* [5], *CD44* [6], *pleiotrophin* [7], *N-cadherin* [8], *H-RAS* [9], and *CDC10* [10] (Table 1). From the analyses of genomic aberrations occurred in NBLs, several candidate genes that exhibit overexpression in advanced NBL were identified such as *survivin* [11], *NM23-H1* and *NM23-H2* [12,13], and *PPM1D* [14]. These genes are located on chromosome 17q, which is known to be frequently increased chromosomal copies in advanced NBLs.

3. Identification of novel prognosis-related genes

Various genomic approaches have been used to identify differentially expressed genes among different tissue and tumor types, including differential hybridization screening, representational difference analysis (RDA), gene counting using cDNA libraries followed by semi-quantitative reverse transcriptase polymerase chain reaction (RT-PCR) screening, serial analysis of

gene expression (SAGE), suppression subtractive hybridization (SSH), and cDNA and oligonucleotide microarrays.

In order to collect a large number of genes expressed in various type of NBLs, we have constructed oligo-capping cDNA libraries from primary NBL tissues with different biological characteristics: the tumors with favorable (F; stage 1, single copy of *MYCN*, high *TrkA* expression) and unfavorable (UF; stage 3 or 4, amplification of *MYCN*, no expression of *TrkA*) characteristics and the stage 4 S tumor [15,16]. Ten thousands of clones in total were isolated from those libraries, which corresponded 5,340 independent genes, and approximately 40% of those were shown to contain novel sequences by database search [16]. To identify the genes expressed differentially between the F and UF subsets, all independent clones except housekeeping genes were subjected to semi-quantitative RT-PCR analysis using RNAs obtained from 16 F and 16 UF NBL tissues as templates. From this project, we have identified more than 500 genes differentially expressed between F and UF NBLs. These included many novel genes with unknown functions, including endothelin-converting enzyme-like 1 (*XCE/ECEL1*) [17], neuronal leucine-rich repeat family members (*NLRR1* and *NLRR3*) [18], tubulin tyrosine ligase (*TTL*) [19], novel putative apoptosis-related gene with BCH domain (*BMCC1*, Machida et al., manuscript in preparation), a member of LIM-only protein (*LMO3*) (Aoyama et al., under submission), and NEDD4-like ubiquitin E3 ligase (*NEDL1* and *NEDL2*) (Miyazaki et al., manuscript in preparation) (Table 1). All these genes were analyzed by quantitative real-time RT-PCR method and confirmed to be strongly related to patient prognosis of NBLs. These genes are now being investigated by functional analysis.

Roberts et al. [20] have applied SSH technique to identify potential NBL biomarkers that may improve outcome prediction, and identified differential expression of members of the GABAergic gene family in NBL. They found that low levels of gamma-aminobutyric acid (GABA) receptor-associated protein (*GABARAP*) gene expression predict decreased survival, and that GABA(A) delta receptor subunit gene expression was predictive of a poor outcome among stage 4S patients.

4. Expression profiling of neuroblastomas by microarray approach

Recently, the DNA microarray method has been applied to comprehensively demonstrate the expression profiles of primary NBLs and cell lines (Table 2). The first microarray-based gene expression profiling study of NBL was reported by Khan et al. [21], who demonstrated that the small, round blue-cell tumors (SRBCTs), including NBL, rhabdomyosarcoma, non-Hodgkin lymphoma, and the Ewing family of tumors, which is often present diagnostic dilemmas in clinical practice, could be distinguished on the basis of their patterns of gene expression using artificial neural networks (ANNs). Among these SRBCTs, they showed that 6 of 7 test samples with NBL were classified correctly by using 93 unique genes. Among the 93 classifier genes, 15 genes were highly and specifically expressed in NBLs. This finding is very important when diagnostic tool is going to apply the gene expression data, because it is the first point to be confirmed whether the tumor, which is to be examined, is NBL or not.

Subsequent microarray studies have facilitated class separation of differentiating NBL tumors from poorly differentiated tumors and of high-risk tumors from low-risk tumors. Yamanaka et al. [22] examined 14 NBLs by 23,040 cDNAs microarray, and identified 78 genes whose expression levels were significantly different between differentiating NBLs and poorly differentiated NBLs. The 78 genes included those associated with cell maturation and apoptosis; 15 genes that were up-regulated in stage 4 tumors included those encoding cell adhesion molecules and cytoskeleton proteins. Berwanger et al. [23] examined expression profiles from 94 primary NBL specimens using a 4,608 cDNA human unigene chip. They found 24 significant genes differentially expressed between stage 1 ($n=19$) and stage 4 ($n=21$) *MYCN* non-amplified tumors. Interestingly, a significant percentage of the 24 genes encoded those involved in signaling through the nonreceptor tyrosine kinase Fyn and the actin cytoskeleton. These genes were coordinately down-regulated in advanced stage NBL, both in *MYCN* amplified and nonamplified tumors (Table 2). They also showed that expression of *FYN* predicts long term survival of NBL patients, independently of *MYCN* amplification. Takita et al. [24] performed DNA microarray analysis on 20

Table 2
Differentially expressed genes identified by gene expression profiling using microarray

	Genes identified	Genes on microarray	Sample number	Category
NBL diagnosis				
Khan et al. [21]	15 genes	6,567cDNAs	16 NBLs	For SRBCTs diagnosis *DPYSL4, CDH2, AF1Q, CRMP1, KIF3C, GAP43, MAP1B, RCV1, SFRP1, GATA2, PFN2, FHL1* (highly and specifically expressed in NBLs)
Expression profiling				
Yamanaka et al. [22]	78 genes	23,040 cDNAs	14 NBLs	Differentiating NBLs vs. poorly differentiated NBLs *ITGE, SYP, OLIG2, MADH2, DFFB, CASP8, CASP9* (up-regulated in differentiating NBLs) *CLDN5, CCND1, NFKBIL2* (up-regulated in poorly differentiated NBLs)
Berwanger et al. [23]	36 genes	4,608 cDNAs	94 NBLs	Stage 1 vs. stage 4 *MYCN* non-amplified tumors *FYN, AFAP, CTNNA1, NRCAM, tropomodulin, MARCKS* (down-regulated in advanced stage NBL)
Takita et al. [24]	3 genes	1,700 genes	20 NBLs	Stage 1 vs. stage 4 *BIRC3, CDKN2D* (up-regulated in the early-stage group) *SMARCD3* (down-regulated in the early-stage group)
Hiyama et al. [25]	123 genes	6,272 cDNAs	20 NBLs	Unfavorable vs. favorable (regressing and maturing) 43 genes including *MYCN, hTERT, NME1, CCND1, CCNE1, E1, BIRC5, BIRC1* (up-regulated in unfavorable NBLs) 80 genes (up-regulated in favorable NBLs) *CD44, IGF2, TRKA, ANK1* (highly expressed in maturing NBLs) *CASP8, CASP9, TNFSF10, NGFA, GDF10* (highly expressed in regressing NBLs)
McArdle et al. [26]		14,500 genes	20 NBLs	Differentially expressed in the 11q-, MYCN non-amplified and hyperdiploid subtypes of NBL
Janoueix-Lerosey et al.[27]		320 genes (on 1p35-36)	43 NBLs	1p loss vs. 1p-normal *CDC42, VAMP3, CLSTN1, GNB1, STMN1, RPA2, RBAF600, FBXO6, MAD2L2* (decreased expression in NBLs with 1p deletion)
Prognosis prediction				
Wei et al. [28]	19 genes	42,578 cDNA	56 NBLs	To develop an accurate predictor of survival for patient with NBL *DLK1, PRSS3, ARC, SLIT3, MYCN, JPH1* (up-regulated in the poor-outcome group) *ARHI, CNR1, CD44, ROBO2, BTBD3, KLRC3* (down-regulated in the poor-outcome group)
Ohira et al. [29]	70 genes	5,340 cDNAs	136 NBLs	To develop an accurate predictor of survival for patient with NBL

primary tumors (stage 1 versus 4) and identified that the expression of *BIRC3* and *CDKN2D* genes were significantly higher in the early-stage group than in the advanced-stage group and that the expression of the *SMARCD3* gene was significantly reduced in the early-stage group. The *BIRC3*, *CDKN2D*, and *SMARCD3* genes have been reported to be associated with apoptosis, cell cycles, and the transcriptional activator, respectively. Hiyama et al. [25] analyzed 20 NBLs with 6272 cDNAs microarray, and revealed that 43 genes, including *MYCN*, *hTERT*, *NM23-H1*, cell cycle regulatory protein-coding genes (*CCND1*, *CCNE1*, *E1*), and apoptosis-escape genes (*BIRC5*, *BIRC1*) were highly expressed in unfavorable neuroblastomas, while another 80 genes, including neuronal differentiating genes and apoptotic inducing genes (*CD44*, *IGF2*, *TRKA*, *ANK1* in maturing NBLs, *CASP8*, *CASP9*, *TNFSF10*, *NGFA*, *GDF10* in regressing NBLs) were detected as highly expressed in favorable tumors.

There have been several studies by focusing the certain genomic aberrations and corresponding gene expression profiles. McArdle et al. [26] identified transcripts that are differentially expressed in the 11q-, *MYCN* nonamplified and hyperdiploid subtypes of NBL. Janoueix-Lerosey et al. [27] compared the expression profiles between the tumors with 1p loss and those with normal 1p status and identified the genes with decreased expression in NBLs with 1p deletion (Table 2).

5. Microarray-based system for predicting prognosis of neuroblastoma patients

We now have in our hands the increasing number of information for genes which can distinguish prognosis of the patient with NBL as described above. These data should subsequently be integrated and organized to construct a simple prediction system which is practical for the clinic. Such efforts are now ongoing.

Wei et al. [28] have performed gene expression profiling of 56 NBLs using cDNA microarrays containing 42,578 cDNA clones and used artificial neural networks (ANNs) to develop an accurate predictor of survival for each individual patient with NBL. ANN-based prognosis prediction has been accomplished by using expression levels of only 19 genes including *MYCN* and *CD44*. In addition, these 19 predictor genes were able to additionally classify high-risk patients into two subgroups according to their survival status. Among these predictor genes, *DLK1*, *PRSS3*, *ARC*, *SLIT3*, *MYCN* and *JPH1* were up-regulated in the poor-outcome group included, whereas *ARHI*, *CNR1*, *CD44*, *ROBO2*, *BTBD3* and *KLRC3* were down-regulated.

We also have constructed an in-house, microceramic pump-based ink-jet-printed cDNA microarray carrying 5,340 genes obtained from primary NBL cDNA libraries and applied it for the analysis of 136 tumors. A probabilistic output computational analysis using learning samples has selected 70 genes which constructed a classifier for patient outcome, and provided a correct prognosis of test samples with high efficiency [29]. Of clinical interest, Kaplan-Meier analysis indicated that the classifier can divide significantly 5-year survivals of NBL patients, even for the intermediate risk type. Furthermore, our microarray prediction exhibited the best balance between sensitivity and specificity among prognostic factors including *MYCN* amplification and *TrkA* expression.

These findings provide evidence of a gene expression signature that can predict prognosis independent of currently known risk factors and could assist physicians in the individual management of patients with high-risk NBL. Such gene expression-based diagnosis system should be highly accurate and reproducible, as well as simple, easy to analyze, and with low cost. Based on this, we subsequently made a mini-chip carrying top-ranked 200 genes for clinical use. We are now investigating its (1) reproducibility from the original 5,340 genes chip, (2) potential of predicting prognosis for the newly added test samples, (3) potential of predicting prognosis when whole experimental process (RNA isolation, sample labeling, hybridization, and calculation of the probability of patient survival by using the classifier) is conducted in an independent laboratory. After confirming these, the microarray system will come to prove its practical use to be feasible in the clinic to predict the prognosis of the patient with NBL.

6. Conclusion

Using the gene expression profiling, we are now able to distinguish a group of high-risk patients with high efficiency who will not respond to conventional therapy and therefore require alternative treatment strategies. Although further prospective studies will be necessary, we may also be able to reduce the toxicity of treatment regime for the patients who have been predicted to survive according to gene expression profile. Recently, in addition to the gene expression profiling, certain genome aberrations as well as epigenetic alterations have been reported to be strongly related to the patient prognosis with NBL. Abe et al. [30] indicated that poor NBLs suffer the increased methylation pressure in their tumor genome, and that the methylation of certain CpG island, such as *Protocadherin beta* family, can predict poor NBL prognosis with high sensitivity. Furthermore, screening of prognosis-related proteins secreted into serum of NBL patients has been started by many groups. Combination of such independent systems to predict

prognosis will further improve the accuracy of the diagnostic system for NBL.

Acknowledgements

This work was supported in part by a fund from Hisamitsu Pharmaceutical Co., Inc. and by Grant-in-Aids for Scientific Research on Priority Areas (C) 'Medical Genome Science' and 'Genome Information Science', and for Scientific Research (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and by Grant-in Aids for Cancer Research and for the 2nd Term Comprehensive 10-year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan.

References

[1] R.P. Bolande, The neurocristopathies: a unifying concept of disease arising in neural crest maldevelopment, Hum. Pathol. 5 (1974) 409–429.
[2] J.L. Weinstein, H.M. Katzenstein, S.L. Cohn, Advances in the diagnosis and treatment of neuroblastoma, Oncologist 8 (2003) 278–292.
[3] A. Nakagawara, M. Arima-Nakagawara, N.J. Scavarda, C.G. Azar, A.B. Cantor, G.M. Brodeur, Association between high levels of expression of the TRK gene and favorable outcome in human neuroblastoma, N. Engl. J. Med. 328 (1993) 847–854.
[4] A. Nakagawara, C.G. Azar, N.J. Scavarda, G.M. Brodeur, Expression and function of TRK-B and BDNF in human neuroblastomas, Mol. Cell Biol. 14 (1994) 759–767.
[5] E. Hiyama, K. Hiyama, T. Yokoyama, Y. Matsuura, M.A. Piatyszek, J.W. Shay, Correlating telomerase activity levels with human neuroblastoma outcomes, Nat. Med. 1 (1995) 249–255.
[6] M.C. Favrot, V. Combaret, C. Lasset, CD44–a new prognostic marker for neuroblastoma, N. Engl. J. Med. 1993; 329.
[7] A. Nakagawara, J. Milbrandt, T. Muramatsu, T.F. Deuel, H. Zhao, A. Cnaan, G.M. Brodeur, Differential expression of pleiotrophin and midkine in advanced neuroblastomas, Cancer Res. 55 (1995) 1792–1797.
[8] R. Shimono, S. Matsubara, H. Takamatsu, T. Fukushige, M. Ozawa, The expression of cadherins in human neuroblastoma cell lines and clinical tumors, Anticancer Res. 20 (2000) 917–923.
[9] T. Tanaka, T. Sugimoto, T. Sawada, Prognostic discrimination among neuroblastomas according to Ha-ras/trk A gene expression: a comparison of the profiles of neuroblastomas detected clinically and those detected through mass screening, Cancer 83 (1998) 1626–1633.
[10] T. Nagata, Y. Takahashi, S. Asai, Y. Ishii, H. Mugishima, T. Suzuki, et al., The high level of hCDC10 gene expression in neuroblastoma may be associated with favorable characteristics of the tumor, J. Surg. Res. 92 (2000) 267–275.
[11] A. Islam, H. Kageyama, N. Takada, T. Kawamoto, H. Takayasu, E. Isogai, et al., High expression of Survivin, mapped to 17q25, is significantly associated with poor prognostic factors and promotes cell survival in human neuroblastoma, Oncogene 19 (2000) 617–623.
[12] N. Hailat, D.R. Keim, R.F. Melhem, X.X. Zhu, C. Eckerskorn, G.M. Brodeur, et al., High levels of p19/nm23 protein in neuroblastoma are associated with advanced stage disease and with N-myc gene amplification, J. Clin. Invest. 88 (1991) 341–345.
[13] A. Leone, R.C. Seeger, C.M. Hong, Y.Y. Hu, M.J. Arboleda, G.M. Brodeur, et al., Evidence for nm23 RNA overexpression, DNA amplification and mutation in aggressive childhood neuroblastomas, Oncogene 8 (1993) 855–865.
[14] F. Saito-Ohara, I. Imoto, J. Inoue, H. Hosoi, A. Nakagawara, T. Sugimoto, J. Inazawa, PPM1D is a potential target for 17q gain in neuroblastoma, Cancer Res. 63 (2003) 1876–1883.
[15] M. Ohira, A. Morohashi, Y. Nakamura, E. Isogai, K. Furuya, S. Hamano, et al., Neuroblastoma oligo-capping cDNA project: toward the understanding of the genesis and biology of neuroblastoma, Cancer Lett. 197 (2003) 63–68.
[16] M. Ohira, A. Morohashi, H. Inuzuka, T. Shishikura, T. Kawamoto, H. Kageyama, et al., Expression profiling and characterization of 4200 genes cloned from primary neuroblastomas: identification of 305 genes differentially expressed between favorable and unfavorable subsets, Oncogene 22 (2003) 5525–5536.
[17] T. Kawamoto, M. Ohira, S. Hamano, T. Hori, A. Nakagawara, High expression of the novel endothelin-converting enzyme genes, Nbla03145/ECEL1alpha and beta, is associated with favorable prognosis in human neuroblastomas, Int. J. Oncol. 22 (2003) 815–822.
[18] S. Hamano, M. Ohira, E. Isogai, K. Nakada, A. Nakagawara, Identification of novel human neuronal leucine-rich repeat (hNLRR) family genes and inverse association of expression of Nbla10449/hNLRR-1 and Nbla10677/hNLRR-3 with the prognosis of primary neuroblastomas, Int. J. Oncol. 24 (2004) 1457–1466.
[19] C. Kato, K. Miyazaki, A. Nakagawa, M. Ohira, Y. Nakamura, T. Ozaki, et al., Low expression of human tubulin tyrosine ligase and suppressed tubulin tyrosination/detyrosination cycle are associated with impaired neuronal differentiation in neuroblastomas with poor prognosis, Int. J. Cancer 112 (2004) 365–375.
[20] S.S. Roberts, M. Mori, P. Pattee, J. Lapidus, R. Mathews, J.P. O'Malley, et al., GABAergic system gene expression predicts clinical outcome in patients with neuroblastoma, J. Clin. Oncol. 22 (2004) 4127–4134.

[21] J. Khan, J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, et al., Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nat. Med. 7 (2001) 673–679.
[22] Y. Yamanaka, Y. Hamazaki, Y. Sato, K. Ito, K. Watanabe, T. Heike, et al., Maturational sequence of neuroblastoma revealed by molecular analysis on cDNA microarrays, Int. J. Oncol. 21 (2002) 803–807.
[23] B. Berwanger, O. Hartmann, E. Bergmann, S. Bernard, D. Nielsen, M. Krause, et al., Loss of a FYN-regulated differentiation and growth arrest pathway in advanced stage neuroblastoma, Cancer Cell 2 (2002) 377–386.
[24] J. Takita, M. Ishii, S. Tsutsumi, Y. Tanaka, K. Kato, Y. Toyoda, et al., Gene expression profiling and identification of novel prognostic marker genes in neuroblastoma, Genes Chromosomes Cancer 40 (2004) 120–132.
[25] E. Hiyama, K. Hiyama, H. Yamaoka, T. Sueda, C.P. Reynolds, T. Yokoyama, Expression profiling of favorable and unfavorable neuroblastomas, Pediatr. Surg. Int. 20 (2004) 33–38.
[26] L. McArdle, M. McDermott, R. Purcell, D. Grehan, A. O'Meara, F. Breatnach, et al., Oligonucleotide microarray analysis of gene expression in neuroblastoma displaying loss of chromosome 11q, Carcinogenesis 25 (2004) 1599–1609.
[27] I. Janoueix-Lerosey, E. Novikov, M. Monteiro, N. Gruel, G. Schleiermacher, B. Loriod, et al., Gene expression profiling of 1p35-36 genes in neuroblastoma, Oncogene 23 (2004) 5912–5922.
[28] J.S. Wei, B.T. Greer, F. Westermann, S.M. Steinberg, C.G. Son, Q.R. Chen, et al., Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma, Cancer Res. 64 (2004) 6883–6891.
[29] M. Ohira, S. Oba, Y. Nakamura, E. Isogai, S. Kaneko, A. Nakagawa, et al., Expression profiling using a tumor-specific cDNA microarray prediols the prognosis of intermediate-risk neuroblastomas, Cancer Cell 7 (2005) 337-350.
[30] M. Abe, M. Ohira, A. Kaneda, Y. Yagi, S. Yamamoto, Y. Kitano, et al., CpG island methylator phenotype is a strong determinant of poor prognosis in neuroblastomas, Cancer Res., 65 (2005) 828–834.

Expression profiling using a tumor-specific cDNA microarray predicts the prognosis of intermediate risk neuroblastomas

Miki Ohira,[1,8] Shigeyuki Oba,[2,8] Yohko Nakamura,[1] Eriko Isogai,[1] Setsuko Kaneko,[3] Atsuko Nakagawa,[4] Takahiro Hirata,[5] Hiroyuki Kubo,[5] Takeshi Goto,[5] Saichi Yamada,[6] Yasuko Yoshida,[6] Misa Fuchioka,[7] Shin Ishii,[2] and Akira Nakagawara[1,*]

[1]Division of Biochemistry, Chiba Cancer Center Research Institute, Chiba 260-8717, Japan
[2]Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma 630-0192, Japan
[3]Department of Pediatric Surgery, University of Tsukuba School of Medicine, Tsukuba 305-8575, Japan
[4]Second Department of Pathology, Aichi Medical University, Nagakute 480-1195, Japan
[5]Hisamitsu Pharmaceutical Co. Inc., Tokyo 100-622, Japan
[6]Micro Ceramics Laboratory, R & D Center, NGK Insulators, LTD, Nagoya 467-8530, Japan
[7]Center for Molecular Biology and Cytogenetics SRL Inc., Tokyo 191-0002, Japan
[8]These authors contributed equally to this work.
*Correspondence: akiranak@chiba-cc.jp

Summary

To predict the prognosis of neuroblastoma patients and choose a better therapeutic protocol, we developed a cDNA microarray carrying 5340 genes obtained from primary neuroblastomas and examined 136 tumor samples. We made a probabilistic output statistical classifier that provided a high accuracy in prognosis prediction (89% at 5 years) and a highly reliable method to validate it. Kaplan-Meier analysis indicated that the patients in an intermediate group defined by existing markers are divided by microarray into two further groups with 5 year survivals for 36% and 89% of patients ($p < 10^{-4}$), i.e., with unfavorably and favorably predicted neuroblastomas, respectively. According to these results, we developed a gene subset chip for a clinical tool, for which our classifier exhibited 88% prediction accuracy.

Introduction

Neuroblastoma is one of the most common solid tumors in children and originates from the sympathoadrenal lineage of the neural crest (Bolande, 1974). Its clinical behaviors are heterogeneous. The tumor, when developed in infants, frequently regresses spontaneously by inducing differentiation and/or programmed cell death. When developed in children over 1 year of age, however, the tumor is often aggressive and acquires resistance to intensive chemotherapy. Although recent progress in therapeutic strategies against advanced neuroblastoma has improved patient survival, long-term outcomes still remain very poor. Furthermore, part of neuroblastomas categorized to the intermediate group (stage 3 or 4 tumors that possess a single copy of the *MYCN* gene) often recurs after complete response to initial therapy. Such differences in the final outcomes of the tumor are considered presumably attributable to differences in genetic and biological abnormalities, which are reflected in the gene and protein expression profiles of the tumor.

The prediction of cancer prognosis is one of the most urgent demands to initiate the treatment of neuroblastoma. As expected from the natural course of neuroblastoma, patient age at diagnosis (over or under 1 year of age) is an important prognostic factor (Evans et al., 1971). Disease stage is also a powerful indicator for neuroblastoma prognosis (Brodeur et al., 1993). Moreover, recent advances in basic research have discovered several molecular markers that are useful in clinical practice, including amplification of the *MYCN* oncogene (Schwab et al., 1983; Brodeur et al., 1984), DNA ploidy (Look et al., 1984; Look et al., 1991), deletion of chromosome 1p (Brodeur et al., 1988), and *TrkA* expression (Nakagawara et al., 1992; Nakagawara et al., 1993). Other indicators also include *telomerase* (Hiyama et al., 1995), *CD44* (Favrot et al., 1993), *pleiotrophin* (Nakagawara et al., 1995), *N-cadherin* (Shimono et al., 2000), *CDC10* (Nagata et al., 2000), and *Fyn* (Berwanger et al., 2002). However, the combinations thereof still frequently fail to predict patient outcome. In the post-genome sequence era, therefore, the advent of new diagnostic tools has been ex-

SIGNIFICANCE

Neuroblastoma is an enigmatic tumor with heterogeneous clinical behaviors including maturation, regression, and growth. Despite recent improvements in the cure rate of many pediatric tumors, the prognosis of advanced neuroblastoma is still poor. In addition, it is usually difficult to predict the prognosis of the intermediate risk group in advanced stages without *MYCN* amplification. Through our supervised machine learning and highly reliable statistical validation procedure with the 5 year prognosis of the patients, we established a simple, low-cost microarray system carrying top-ranked genes, which exhibited high accuracy (88%) to predict the neuroblastoma prognosis and is highly feasible as a clinical tool.

pected. Recently, the DNA microarray method, applied to comprehensively demonstrate expression profiles of primary neuroblastomas and cell lines, has already identified the following: (1) differences in gene expression between favorable and unfavorable subsets (Yamanaka et al., 2002; Berwanger et al., 2002); and (2) differences in gene expression that occur during retinoic acid-induced neuronal differentiation (Ueda, 2001). However, a study to predict neuroblastoma prognosis with a microarray using a large number of neuroblastoma samples has never been reported. We have recently isolated 5500 genes from the cDNA libraries, which were generated from primary neuroblastomas, part of which has previously been reported (Ohira et al., 2003a; Ohira et al., 2003b). In this study, to identify genes strongly associated with neuroblastoma prognosis and to apply them to make a really practical cDNA microarray for neuroblastoma diagnosis, we constructed an in-house, ink-jet-printed cDNA microarray carrying 5340 genes proper to neuroblastoma and applied it to analyze 136 samples. After selecting genes significantly related to patient prognosis, we made a mini-chip carrying 200 top-ranked genes to apply for the clinic.

There have been many attempts to predict cancer outcome using microarray. A reliable prediction for outcomes of cancer patients naturally demands its reproducibility, and it is quite important to use sound and highly reliable statistical methodologies; a complete crossvalidation analysis without introducing any information leakage and an independent test using new samples are necessary. As Ntzani and Ioannidis (2003) pointed out, however, such a careful methodology has often been ignored in most microarray studies. We here developed a supervised classification method without any information leakage as a statistic tool and demonstrated that the probabilistic output of the analysis defines the molecular signature of neuroblastoma to predict its prognosis. Although the construction of the statistical tool was based on one of the most reliable statistical tests, we also consulted a validation test for an independent experiment examining 50 samples (whose RNAs were prepared in an independent laboratory) by using the mini-chip. The high performance for the outcome prediction by the mini-chip system suggests the high feasibility of developing a clinical tool based on molecular signature.

Results

Neuroblastoma proper cDNA microarray

The whole scheme of our study is summarized in Figure 1. We first constructed a neuroblastoma proper cDNA microarray harboring the spots of 5340 genes on a slide glass by using a ceramics-based ink-jet printing system (the 5340 genes system). This in-house cDNA microarray appeared to have overcome the previous problems caused by pin-spotting, e.g., uneven quantity or shape of individual spots on an array. Ten micrograms each of the total RNA extracted from 136 frozen tissues of primary neuroblastomas were labeled with Cy3 dye. As a common reference, the mixture of the total RNA obtained from four neuroblastoma cell lines with a single copy of *MYCN* (NB69, NBLS, SK-N-AS, and SH-SY5Y) was labeled with Cy5 dye.

We first evaluated the quality of our cDNA microarray, the 5340 genes system. The log Cy3/Cy5 fluorescence ratio of each gene spot was normalized to eliminate intensity-dependent biases. Since the 5340 genes array contains 260 duplicated or multiplicated genes, the expression ratio of such a duplicated gene was represented by the average of multiple spots. Based on estimation performance for missing values (see the Supplemental Data available with this article online) and on reproduction variance of the duplicated genes, the standard deviation for the log ratio of a single gene was sufficiently small, ranging between about 0.2 and about 0.3 (Figure S1A). The scatter plots of the log Cy3/Cy5 fluorescence ratio between duplicated gene spots in the 136 experiments and those between repeated experiments also indicated high reproducibility of spotting and experiment (Figures S1B and S1C). These suggest that the production of and experiments by our cDNA microarray are highly reproducible.

Supervised classification

To develop a statistical tool that predicts the prognosis of a new patient with neuroblastoma, we introduced a supervised classification. In the development, we used 136 neuroblastomas, randomly selected tumor samples from the neuroblastoma tissue bank, consisting of 41 stage 1 tumors, 22 stage 2 tumors, 33 stage 3 tumors, 28 stage 4 tumors, and 12 stage 4s tumors. The follow-up duration ranged between 3 and 241 months (median, 56 months, mean, 57.3 months) after diagnosis. The left panel in Figure 2 compiles summary information of each sample, including survival time and important prognosis markers (see Experimental Procedures for details). Since variations in follow-up duration generated noises in the supervised classification, we used patient outcome (dead or alive) at 5 years after diagnosis as the target label to be predicted. Since the outcomes of 40 of 136 samples were unknown at 5 years after diagnosis, data for 96 remaining samples were used subsequently. When we were interested in short-term outcome prediction, the target label was set at 2 years after diagnosis, for which purpose 126 samples out of the 136 samples were used.

We constructed the weighted voting as a supervised classifier after important genes were selected according to pairwise *F* scores. To estimate the prediction accuracy for new data, we consulted leave two out (LTO) analysis, which obtains almost unbiased estimation of prediction accuracy for new data while avoiding overestimation due to information leakage (Figure S2A). Although it is known that the prediction accuracy of a supervised classifier depends on the number of genes to be used (Figure S3), the LTO procedure enables us to optimize it without introducing information leakage, by using a sample left out at the outer loop of the double-loop procedure (see Experimental Procedures). The crossvalidation accuracy for the 5 year prognosis prediction was as high as 88.5% (sensitivity of 86.7% and specificity of 89.4%) (Table 1, "Whole cases"). In the LTO analysis, we selected genes and constructed the corresponding classifier individually for the outcome prediction of each sample. The average number of the selected genes, n, was 30.7. If we applied the same procedure to the short-term (2 year) prediction, the accuracy, sensitivity, and specificity were 89.8%, 88.0%, and 90.2%, respectively (data not shown).

Construction of a probabilistic output

According to the LTO analysis, we can obtain weighted vote values and the corresponding survival rates. After approximat-

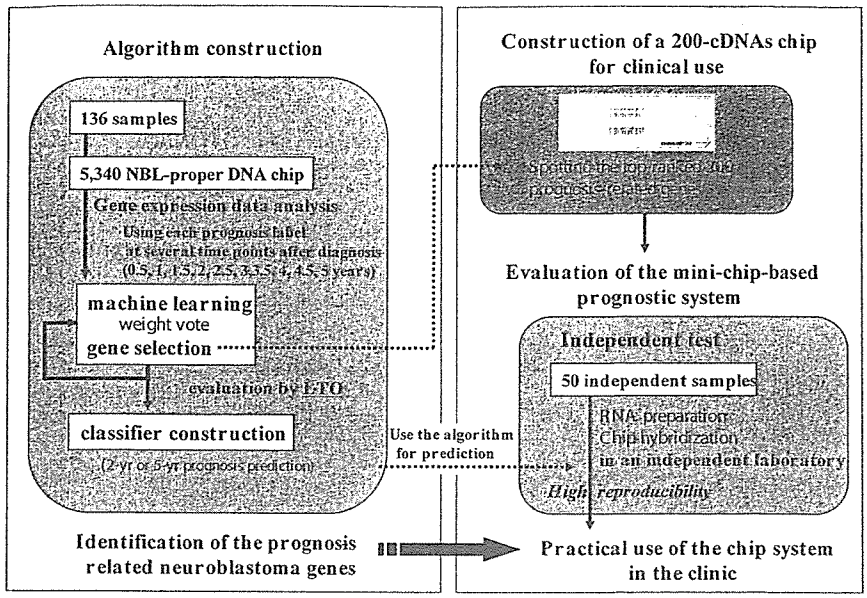

Figure 1. Schematic diagram of this study

ing a nonlinear transformation from weighted vote values to the survival rates, the transformation outputs the reliability of each sample's outcome prediction as a probabilistic output, posterior probability. We suppose each posterior probability, a real number between 0 and 1, corresponds to the expected 5 year survival rate. The right upper panel of Figure 2 shows the predictions for the 136 samples as posterior probabilities. Most of the samples alive at 5 years after diagnosis (blue mark) are found to have posterior values near 1, while most of the dead samples (red mark) have those near 0. It is known that it is difficult to predict the prognosis of neuroblastoma patients of the intermediate risk group (the type II subset: stage 3 or 4, without amplification of *MYCN*), denoted by green area. The posterior values are likely to take intermediate values near 0.5; however, their binarization after being separated by threshold 0.5 shows good accordance with the actual outcome. Frequencies of posterior values for alive and dead samples are shown in the right middle panel. The rate of alive samples among the whole samples, which denotes the actual survival rate, is plotted against each posterior value in the right bottom panel in Figure 2; this panel shows the good correspondence between the posterior value and the survival rate.

Probabilistic outputs are considered to be advantageously useful as compared with conventional binary outputs when used in making a clinical assessment and may be considered identical to them if establishing an appropriate threshold value. The real-valued posterior can be used for categorization into arbitrary number of groups. For example, dividing the posterior values into three by setting thresholds 0.3 and 0.7, we obtain three groups whose survival curves are significantly different from each other; this tertiary categorization provides another definition of intermediate risk group based only on expression patterns (Figure S4).

Comparing the survival curves

Figure 3A shows survival curves for favorable and unfavorable patients predicted by the classifier with a binary threshold (0.5). The 5 year survival rate for the former (n = 98) was as good as 94%, while that for the latter (n = 38) was as poor as 33% ($p < 10^{-5}$). When 70 sporadic neuroblastomas were evaluated after excluding the tumors found by mass screening, the 5 year survival rate for the former (n = 40) was 85%, while that for the latter (n = 30) was 20% ($p < 10^{-5}$) (Figure 3B). To further evaluate the efficiency of our system, we calculated the posterior value for the intermediate subset of neuroblastoma (type II), whose prognosis is usually difficult to predict. As shown in Figure 3C, the survival curves were significantly categorized into two groups. The 5 year survival rate of patients who were predicted as favorable was 89%, while that for unfavorable patients was 36% ($p = 0.000067$). Since the age at diagnosis (≥ 1 year) is currently used as a poor prognostic factor for the type II tumors, we examined the ability of the classifier for the older patients with type II tumors. Even for such patients whose prognosis is difficult to predict, the survival rate (45%) of all 18 patients was divided solely by gene expression into the group with favorable prognosis (n = 10; 73%) and that with poor outcome (n = 8; 13%) (Figure 3D). In addition, if the intermediate risk group was further separated into stage 3 tumor group and stage 4 tumor group, the posterior value was significantly related to the survival, especially for stage 3 tumors (Figure S5). These results suggest that the posterior value obtained by our statistical analysis highly efficaciously allows the classification of patient outcomes, even when the tumor is of the intermediate type.

We further compared our results to existing prognosis markers in Table 1 and found that the supervised microarray analysis showed the best sensitivity-specificity balance among the prognostic factors for predicting the outcome of neuroblastoma. When the classifier is combined with the age at diagnosis, the disease stage (stage 1, 2, or 4s versus stage 3 or 4) and the *MYCN* amplification, accuracy, sensitivity, and specificity increased up to 95.8%, 93.3%, and 97.0%, respectively. Although the currently used markers (age, stage, and *MYCN*)

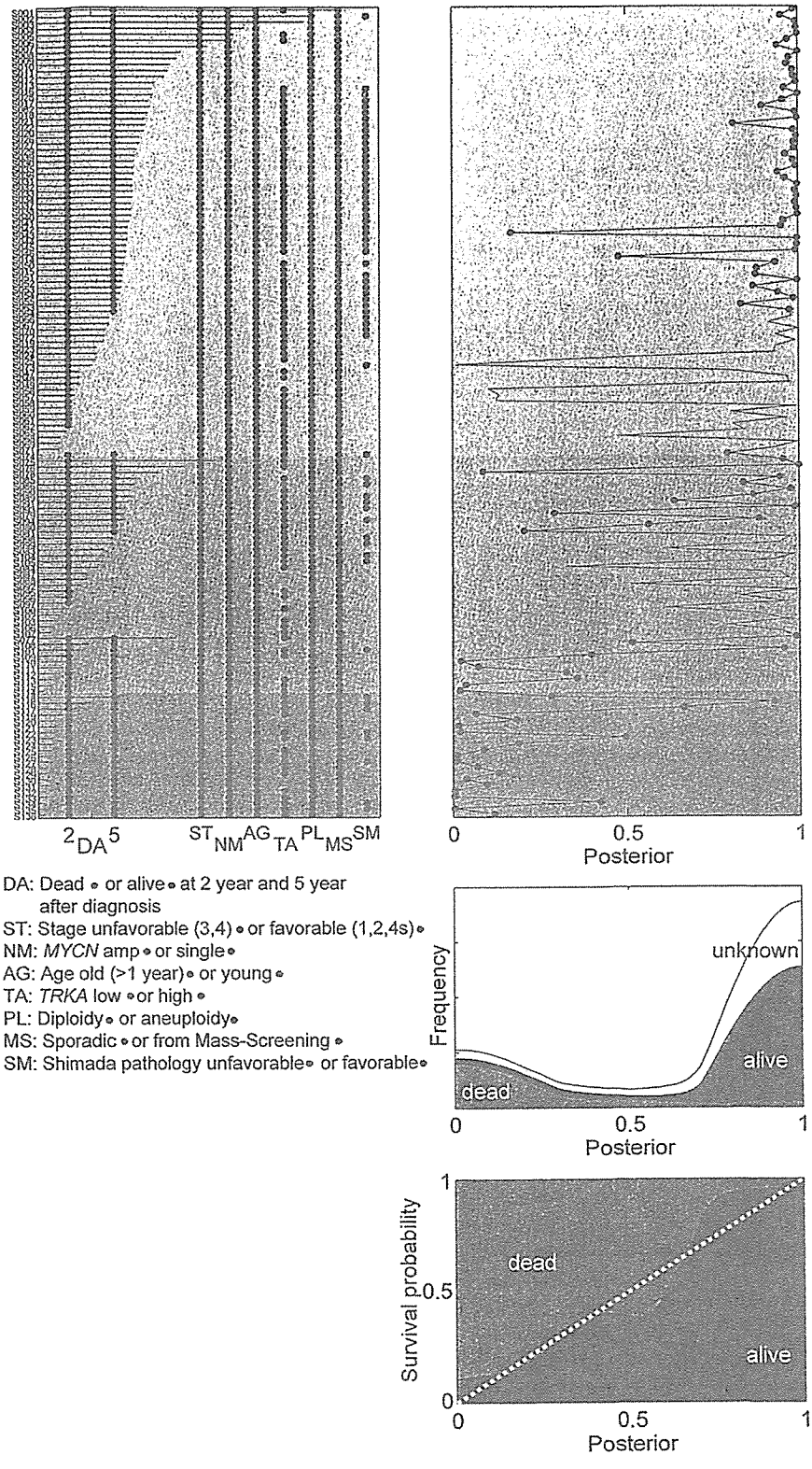

Figure 2. Posterior probability of survival at 5 years

Posterior probability of survival at 5 years for 136 training data samples, output by the leave two out (LTO) crossvalidation without any information leakage. Left panel: Neuroblastoma samples. A red or blue horizontal line denotes survival period after diagnosis for a dead or alive patient, respectively. Red and blue marks denote various clinical properties of patients; see text below the panel for detailed explanation. Background colors show groups determined by stage and *MYCN* amplification status: red, type III, with *MYCN* amplification; green, type II, with single copy of *MYCN* at unfavorable stage (3 or 4); and blue, type I, with single copy of *MYCN* and at favorable stage (1, 2, or 4s). Right upper panel: The LTO crossvalidated prediction (posterior) for each patient; a red or a blue mark denotes that the patient is dead or alive at 5 years, respectively. Right middle panel: Cumulative smooth histogram of posterior probabilities for patients of dead (red), alive (blue), and unknown (white) at 5 years after diagnosis. Right lower panel: The horizontal and vertical axes denote the posterior and the empirical probability of 5 year survival, i.e., the ratio of the smooth histogram values between dead and alive patients, shown in the middle panel, respectively. Because the border between dead and alive samples is close to the white broken line (x = y), the posterior can be regarded as a 5 year survival chance rate.

Table 1. Accuracy of each marker for prognosis prediction (5 years after diagnosis)

	Whole cases				Sporadic cases		Intermediate and old age[a]	
	n	accuracy	sensitivity	specificity	n	accuracy	n	accuracy
Microarray classifier	136	89%	87%	89%	56	82%	14	86%
Age (less than 1 year old)	136	81%	83%	80%	56	71%	14	64%
Stages (1, 2, and 4s)	136	83%	97%	77%	56	84%	14	64%
Shimada classification (unfavorable)	62	87%	78%	89%	25	72%	(n < 10)	—
Hyperdiploidy (aneuploidy)	62	72%	67%	73%	27	56%	(n < 10)	—
MYCN amplification	136	89%	67%	99%	56	80%	14	36%
Microarray + age + stages + MYCN*	136	96%	93%	97%	56	93%	14	86%

Sensitivity/specificity is the rate of unfavorably/favorably predicted samples, i.e., LTO posterior <0.5/>0.5, among actually unfavorable/favorable samples. Microarray classifier, supervised classification based on the microarray data. *By this classifier, all patients with the *MYCN* amplification are predicted as unfavorable, and all patients with a single copy of *MYCN* and at stage 1, 2, or 4s are predicted as favorable. In the remaining intermediate samples (with a single copy of *MYCN* and at stage 3 or 4), the patients with age <1 year are predicted as favorable, and the microarray predictions are applied for those with age >1 year.
[a]Age at diagnosis >1 year.

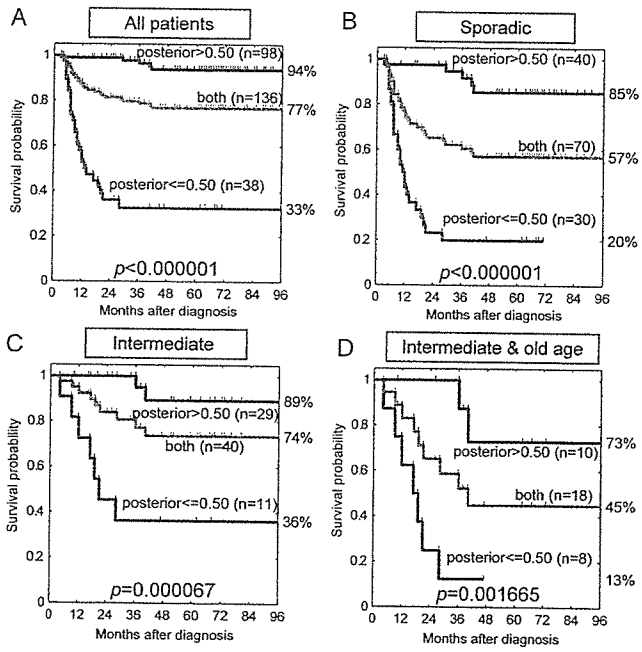

Figure 3. Disease-free survival of patients who were stratified based on the gene expression patterns

For each of the four figures, whole objective patients (green) are divided into favorable (blue) or unfavorable (red) based on the posterior values with threshold 0.5, which are calculated from gene expression patterns, and statistical features of their survival times are denoted by the Kaplan-Meier survival curves. The differences of the survival curves between the favorable (blue) and unfavorable (red) groups are evaluated by p values of the log rank test.

A and B: Survival analysis of whole and sporadic patients, respectively, divided by the supervised classifier based on microarray data.

C and D: Survival analysis of patients in the intermediate risk group with different definitions, divided by the supervised classifier. The intermediate risk group shown in (C) is defined as *MYCN* single and stage 3 or 4 (type II), and that in (D) is defined as *MYCN* single, stage 3 or 4, and older than 1 year of age.

also showed good potential to predict generally but less than the microarray, these exhibited only 64% accuracy of prediction for the type II tumors with ≥1 year of age (Table 1). Together with the results of survival analysis, the microarray classifier is revealed to be a powerful predictor to classify such group of neuroblastomas (86% accuracy; Table 1).

Practical application of 200 cDNAs microarray and independent test

For the practical use in the clinic, a cDNA microarray system that contains cDNA spots of a relatively small number and hence is easy to treat is expected. According to our gene selection based on the pairwise *F* score, the numbers of genes that were appropriate for the 5 year and 2 year prognosis prediction for all available samples were 10 and 70, respectively. In order for the system to reserve the applicability to short-term and long-term outcome prediction simultaneously, we selected 200 top-ranked genes according to the pairwise *F* scores in the 2 year prediction, because the 2 year prediction required larger variety of genes, and then made a smaller cDNA microarray system carrying the 200 genes. The newly designed microarray system (the mini-chip system) was evaluated by being hybridized with 5 μg total RNA obtained from 50 independent test samples. To preserve the independence of experimental procedure, these RNAs were prepared and hybridized in a different laboratory from the original experiments of 136 samples with the 5340 genes system (see Experimental Procedures). Although the weight values in the weighted voting classifier were determined by the 5340 genes system without any information leakage from the 50 independent samples, the result was as good as that obtained by the 5340 cDNA microarray analysis (90% [45/50] for 2 year, and 87.8% [43/50] for 5 year prognosis prediction; Figure 4B). This test validated not only the prediction robustness of our classifier constructed by the 5340 genes system, but also the construction procedure of the mini-chip system according to our gene selection based on pairwise *F* scores. When we reconstructed another supervised classifier by applying the LTO analysis to the 50 samples measured by the mini-chip system, the accuracy of the 5 year prediction was 91.8% (45/49) (Figure 4C). These results suggest

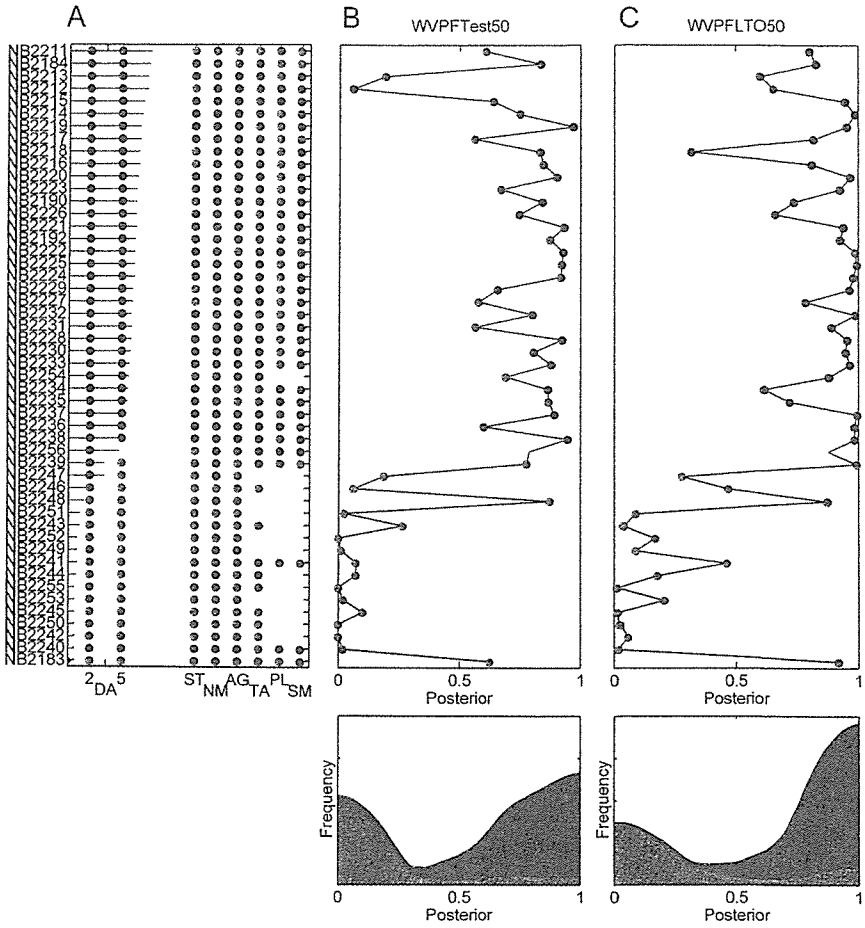

Figure 4. Posterior probability of survival at 5 years for test samples

Posterior probability of survival at 5 years for 50 independent test samples measured by newly developed 200 genes chip (the mini-chip system). Left panel: Neuroblastoma samples; see also Figure 2 legend. Center panel: Prediction results when the supervised classifier constructed from 96 training samples is applied to the 50 independent samples (independent test for the classifier's reproducibility). Right panel: LTO crossvalidation analysis using the new 50 samples (test for the procedure's reproducibility). Both tests do not introduce any information leakage. Lower panels: Smooth histograms of posterior probabilities for dead (red) and alive (blue) patients.

three things. (1) The supervised classifier obtained by the statistical analysis by the 5340 genes system is reproducible even if it is applied to the data measured by the reduced 200 genes system. Note that the 50 samples were completely new data for the classifier in this case. (2) Our procedure to construct a supervised classifier according to the LTO analysis is also reproducible, because the same procedure was successful in making another classifier with a high prediction accuracy when applied to the data taken by the mini-chip system. (3) A simple, low-cost microarray system, the mini-chip system, is highly feasible for predicting the prognosis of neuroblastoma.

Genes selected for prognosis prediction

To assess the relationship between the clinically defined subsets of neuroblastoma and the expression of 70 genes that were selected as top scored with 2 year prognosis according to the pairwise *F* score, we conducted an unsupervised clustering analysis (Figure 5). As expected, part of the type II (intermediate) tumors of patients with a poor prognosis showed an expression pattern that was similar to that of the type III (unfavorable) tumors, and many of them died. On the other hand, expression profiles of the remaining type II tumors seemed to be heterogeneous similarly to those of the type I (favorable) tumors with a good outcome. Most of the tumors with highly expressed *TrkA* and hyperdiploidy, as well as tumors detected by mass screening, were included in the latter group. Table 2 shows a list of 41 genes that corresponded to the 70 top-scored genes and their p and q values (Storey and Tibshirani, 2003) in the log rank test, since we found that several genes were duplicated in the selected 70 genes. Based on the above clustering, these genes were categorized into two groups (group F and group UF; the gene groups strongly correlated with favorable and unfavorable prognosis, respectively) (Figure 5 and Table 2).

The genes in group F tended to show high levels of expression in the type I tumors, while those in group UF were highly expressed in the type III tumors. The former contained genes that were related to neuronal differentiation (*tubulin α*, *peripherin*, *neuromodulin* [*GAP43*], and *HMP19*) and genes that were related to catecholamine metabolism (*dopa decarboxylase* [*DDC*], *dopamine β-hydroxylase* [*DBH*], and *tyrosine hydroxylase* [*TH*]). On the other hand, the latter involved many members of genes that are related to protein synthesis (ribosomal protein genes such as *RPL18A*, *RPLP0*, *RPL5*, *RPL4*, and *RPL7A* as well as translation initiation and elongation factor genes *EEF1G* and *EIF3S5*) and genes that are related to me-

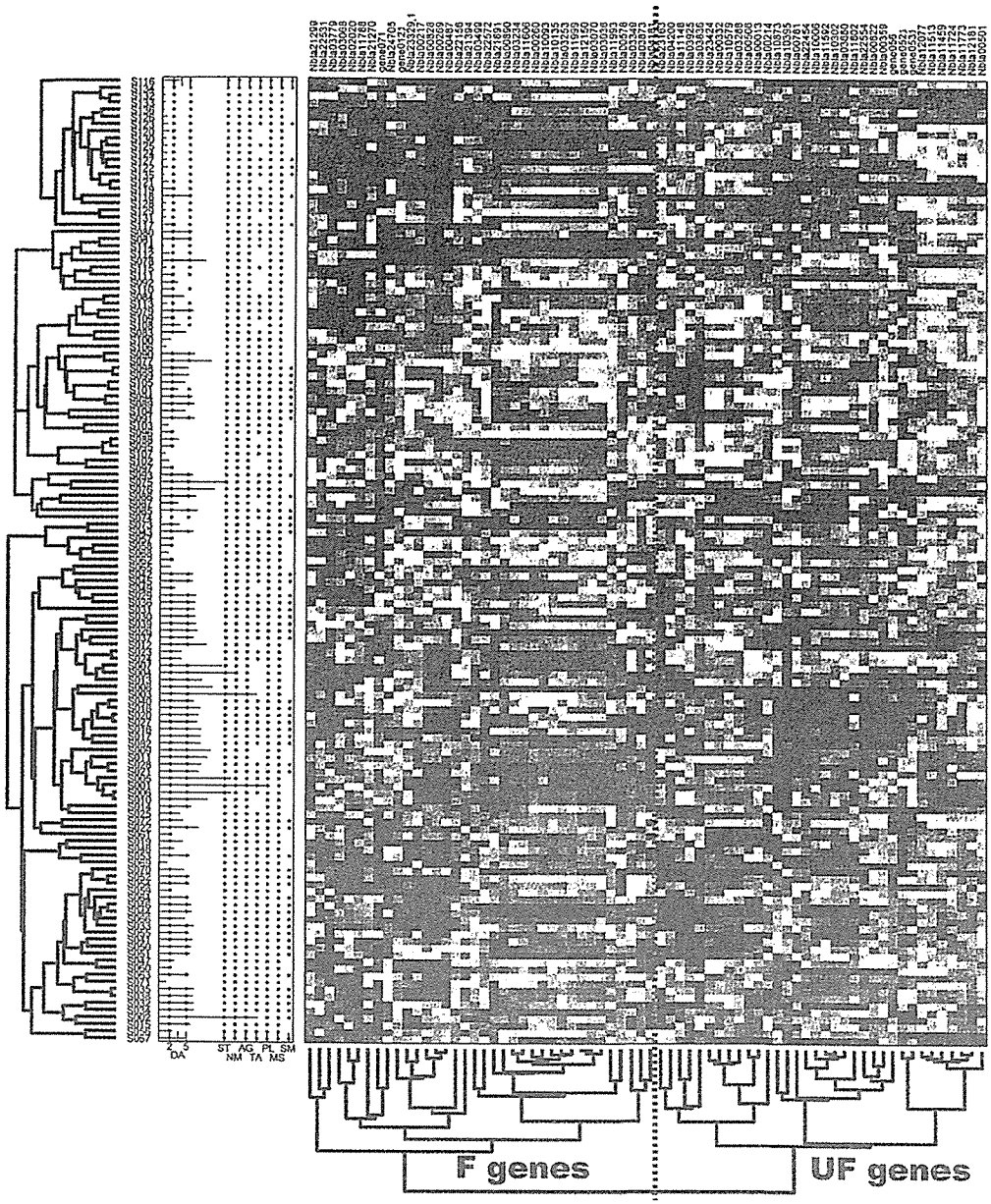

Figure 5. Expression profiles of 70 genes selected for predicting neuroblastoma prognosis at 2 years

Note that 10 genes for predicting prognosis at 5 years are also included in the 70 genes. The left and lower trees depict hierarchical clustering of the 136 neuroblastoma samples and the 70 genes selected in the present study, respectively. In the left tree, blue, green, and red colors denote "*MYCN* single and stage 1, 2 or 4s tumor" (type I, favorable), "*MYCN* single and stage 3, 4 tumor" (type II, intermediate), and "*MYCN* amplified tumor" (type III, unfavorable), respectively. The blue and red colors in the expression matrix show the high and low expression, respectively. A gene showing high expression level likely for unfavorable samples belongs to the group "UF" (red subtree in the lower tree), while one showing high expression likely for favorable samples belongs to the group "F" (blue subtree in the lower tree).

tabolism (*enolase 1* [*ENO1*] and *transketolase* [*TKT*]). The top 10 genes selected for the 5 year outcome prediction were *RPL18A*, *ENO1*, *EEF1G*, *TUBA3*, *GNB2L1*, *ARHGEF7*, *GCC2*, *DDX1* (duplicated), and *PRPH*. The *MYCN* gene was also a member of 70 genes (group UF) as expected; however, it was outside of the top 10 genes for the 5 year label. Instead, *DDX1*, which is frequently coamplified with *MYCN* on chromosome 2p24, was a member of the top 10 genes (UF group) for both of the 2 year and 5 year labels. Confirmation of the differential expression of the selected genes was further conducted by

Table 2. Top-ranked genes used for prediction of 2 year and 5 year prognosis of neuroblastoma

	Spot name	Accession number	Gene code	Chromosome map	Pattern	Log rank p	q value
F group							
	Nbla11606	NM_006009	*TUBA3*	12q13.12	F > UF	0	0.000674
	Nbla00890	NM_003899	*ARHGEF7*	13q34	F > UF	0.000001	0.000743
	Nbla00260	NM_006082	*K-ALPHA-1*	12q13.12	F > UF	0.000003	0.000926
	Nbla21891	U87309	*VPS41*	7p14.1	F > UF	0.000006	0.001096
	Nbla03873	NM_006054	*RTN3*	11q13.1	F > UF	0.00001	0.001282
	Nbla11788	NM_006262	*PRPH*	12q13.12	F > UF	0.000017	0.001522
	Nbla10093	NM_000183	*HADHB*	2p23.3	F > UF	0.000018	0.001541
	Nbla22572	NM_000790	*DDC*	7p12.2	F > UF	0.000035	0.00213
	Nbla21270	NM_001915	*CYB561*	17q23.3	F > UF	0.00016	0.00495
	gene071	NM_000360	*TH*	11p15.5	F > UF	0.000787	0.012173
	Nbla03499	NM_002074	*GNB1*	1p36.33	F > UF	0.000795	0.012237
	Nbla04181	AK55112	*AK55112*	5q13.2	F > UF	0.001425	0.017462
	Nbla00487	AB075512	*C6orf134*	6p21.33	F > UF	0.002751	0.025273
	Nbla00269	NM_000787	*DBH*	9q34.2	F > UF	0.00362	0.030407
	Nbla22531	NM_002045	*GAP43*	3q13.31	F > UF	0.004394	0.034175
	Nbla22156	NM_014944	*CLSTN1*	1p36.22	F > UF	0.005233	0.038274
	Nbla00578	NM_006818	*AF1Q*	1q21.3	F > UF	0.009397	0.05354
	Nbla00217	NM_032638	*GATA2*	3q21.3	F > UF	0.010245	0.056301
	Nbla21394	NM_000743	*CHRNA3*	15q25.1	F > UF	0.072464	0.162629
	Nbla11993	NM_015980	*HMP19*	5q35.2	F > UF	0.204274	0.282486
UF group							
	Nbla00214	NM_000980	*RPL18A*	19p13.11	F < UF	0.000002	0.001107
	Nbla00013	NM_006098	*GNB2L1*	5q35.3	F < UF	0.000006	0.001051
	Nbla11459	NM_004939	*DDX1*	2p24.3	F < UF	0.000024	0.001795
	Nbla11148	NM_001002	*RPLP0*	12q24.23	F < UF	0.000049	0.002549
	Nbla00332	NM_001404	*EEF1G*	11q12.3	F < UF	0.000055	0.002696
	Nbla10395	NM_002593	*PCOLCE*	7q22.1	F < UF	0.000164	0.005009
	Nbla03286	NM_020198	*GK001*	17q23.3	F < UF	0.000175	0.005204
	Nbla23163	NM_003754	*EIF3S5*	11p15.4	F < UF	0.000341	0.007105
	Nbla10579	NM_181453	*GCC2*	2q12.3	F < UF	0.000962	0.01407
	Nbla00359	NM_003550	*MAD1L1*	7p22.3	F < UF	0.00112	0.01525
	gene052-1	NM_005378	*MYCN*	2p24.3	F < UF	0.001253	0.016367
	Nbla03925	NM_002295	*LAMR1*	3p22.2	F < UF	0.001773	0.01931
	Nbla23424	NM_001404	*EEF1G*	11q12.3	F < UF	0.003579	0.030326
	Nbla22554	NM_000687	*AHCY*	20q11.22	F < UF	0.003946	0.032409
	gene056	NM_000546	*TP53*	17p13.1	F < UF	0.004087	0.032829
	Nbla10873	NM_005762	*TRIM28*	19q13.43	F < UF	0.004984	0.037476
	Nbla00501	NM_000969	*RPL5*	1p22.1	F < UF	0.005786	0.04012
	Nbla10302	NM_001428	*ENO1*	1p36.23	F < UF	0.007702	0.048179
	Nbla04200	NM_000968	*RPL4*	15q22.31	F < UF	0.04097	0.120453
	Nbla03836	NM_000972	*RPL7A*	9q34.2	F < UF	0.048031	0.132345
	Nbla00781	NM_001064	*TKT*	3p21.1	F < UF	0.048075	0.132342

Although 70 clones were selected as important genes for the supervised classifier, duplicated and multiplicated clones are omitted in this table. The 41 genes are classified into two groups, "F > UF" and "F < UF," when the expression in favorable samples is higher than that in unfavorable samples, and vice versa, respectively. In each group, genes are sorted by log rank p values. The log rank p value for each gene was calculated by comparing survival curves of two patient groups, in which the expression of the gene is higher and lower, respectively, than the median over the samples. A "q value" of a gene denotes the estimated false discovery rate among the genes whose p value is the same or smaller than that of the gene, and is a p-like value while incorporating multiplicity of the statistical test.

using representative 16 favorable and 16 unfavorable tumor samples that were independent of the 136 samples used in the present analysis, by semiquantitative RT-PCR (Figure S6; refer also to Ohira et al., 2003a). We also conducted immunohistochemical analysis for peripherin antibody using tissue sections prepared from primary neuroblastoma with favorable and unfavorable histology, since peripherin gene is a member of the top 10 genes for both 2 year and 5 year outcome prediction (Table 2). Peripherin protein was positively detected in the cytoplasm of neuroblastic cells as well as neuritis in all three favorable histology tumors (Figure S7, FH&NA). Two unfavorable histology tumors with poorly differentiated subtype, regardless of *MYCN* status, showed sporadic staining (less than 20% of the favorable histology tumor) for peripherin protein in neurites. Peripherin was completely negative in the unfavorable histology tumor of undifferentiated subtype (Figure S7, UF&NA). These results indicate the reliability of our gene selection. In the log rank test, p values of 18 of 20 genes in group F and of all 21 genes in group UF were less than 0.05 (Table 2), indicating that these 39 genes can be independent prognostic factors for primary neuroblastomas.

Discussion

Our study has disclosed the molecular signature of neuroblastoma that predicts patient outcomes by using RNA ob-