

ORIGINAL PAPER

Altered expression of cell cycle genes distinguishes aggressive neuroblastomaAlexei L Krasnoselsky^{1,3}, Craig C Whiteford^{1,3}, Jun S Wei¹, Sven Bilke¹, Frank Westermann², Qing-Rong Chen¹ and Javed Khan^{*,1}¹Oncogenomics Section, Pediatric Oncology Branch, Advanced Technology Center, National Cancer Institute, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA; ²Department of Tumour Genetics – B030, German Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg D-69120, Germany

In this study, gene expression profiling was performed on 103 neuroblastoma (NB) tumors, stages 1–4 with and without *MYCN* amplification, using cDNA microarrays containing 42 578 elements. Using principal component analysis (PCA) to analyse the relationships among these samples, we confirm that the global patterns of gene expression reflect the phenotype of the tumors. To explore the biological processes that may contribute to increasing aggressive phenotype of the tumors, we utilized a statistical approach based on PCA. We identified a specific subset of the cell cycle and/or chromosome segregation genes that distinguish stage 4 NB tumors from all lower stage tumors, including stage 3. Furthermore, the control of the kinetochore assembly emerges from the Gene Ontology analysis as one of the key biological processes associated with an aggressive NB phenotype. Finally, we establish that these genes are further upregulated in the most aggressive *MYCN*-amplified tumors.

Oncogene advance online publication, 13 December 2004; doi:10.1038/sj.onc.1208341

Keywords: cDNA microarray; gene expression; principal component analysis; neuroblastoma; *MYCN*; gene ontology

Introduction

Neuroblastoma (NB) is a childhood tumor of the sympathetic nervous system accounting for 8–10% of all childhood malignancies (Westermann and Schwab, 2002; Brodeur, 2003; Schwab *et al.*, 2003; van Noesel and Versteeg, 2004). The International Neuroblastoma Staging System (INSS) places patients under 1 year of age or with lower stage diseases (stages 1 (ST1) and 2

(ST2)) in the better outcome group as opposed to older patients, or those with advanced stage diseases (stages 3 (ST3) and 4 (ST4)). To date, the most reliable molecular prognostic indicator in NB is the amplification of the *MYCN* gene, found in 20–30% of NBs (Westermann and Schwab, 2002; Schwab *et al.*, 2003; Tonini and Romani, 2003). The correlation between the expression level of *MYCN* and the proliferative activity of NB cells has been shown both *in vitro* using flow cytometry (Cohn *et al.*, 1990; Dominici *et al.*, 1992) and *in vivo* (Schweigerer *et al.*, 1990; Negrini *et al.*, 1991; Schmidt *et al.*, 1994). The elevated expression of *MYCN* has also been shown to induce the expression of a significant number of ribosomal proteins (RP) in NB cell lines as well as tumors (Boon *et al.*, 2001), suggesting that the genes of the protein synthesis machinery are a major target of the *MYCN* protein.

A recent study involving expression profiling of primarily advanced stage NB tumors has provided a broad view of the NB transcriptome and identified a few potential players associated with *MYCN* (Alaminos *et al.*, 2003). In our study, we have expanded expression analysis of NB to tumors of ST1–4, with and without *MYCN* amplification. We developed a principal component analysis (PCA)-based approach to data analysis, which allows one to associate gene expression profiles with gene ontology (GO) annotations to identify the potential biological processes occurring within the transcriptome of NB. Our analysis demonstrates that in the absence of *MYCN* amplification, ST4 tumors are characterized by increased expression of the genes involved in the cell cycle, which distinguishes them from all lower stage tumors, including ST3, and that these genes are further upregulated in the *MYCN*-amplified (NB_{*MYCN*-AMP}) tumors.

Results*Identification of stage-specific expression signatures in NB tumors*

Gene expression profiling was performed on 103 NB samples consisting of 23 ST1, 18 ST2, 20 ST3 and 42 ST4 tumors using a 42 578-clone cDNA microarray. Of

*Correspondence: J Khan, Oncogenomics Section, Pediatric Oncology Branch, Advanced Technology Center, National Cancer Institute, Room 134E, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA; E-mail: khanjav@mail.nih.gov

³These two authors contributed equally to this work

Received 13 August 2004; revised 1 November 2004; accepted 2 November 2004

these tumors, 21 were NB_{MYCN-AMP} (17 ST4, four ST3 and one of ST2). PCA was used on the gene expression data to reduce the dimensionality and visually classify the tumors according to stage and *MYCN* amplification. In this data set, PC1 captured 28% of the total variance and represents the patterns of maximum overall difference among the tumors. However, no separation of tumor types is observed in the PC1 dimension, as shown on Figure 1a, where the tumors are plotted in the space of the first two principal components (PC1 and PC2). The separation of ST1 tumors from ST4 and NB_{MYCN-AMP} tumors is achieved only in the PC2 dimension, which captures 11% of the variance. In addition, there is a significant degree of separation between ST4 tumors according to their *MYCN* amplification status. Therefore, based on the PC2 coordinate, we can distinguish ST1 from ST4 or *MYCN* not-amplified (NB_{MYCN-NA}) tumors from NB_{MYCN-AMP}.

Next, we identified the differentially expressed genes in NB_{MYCN-NA} tumors using a one-way ANOVA for the tumors of ST1–4 and identified 6103 clones with *P*-values less than 0.05. Figure 1b shows the PCA plot using these 6103 clones, and as expected, the stages are now more clearly separated along PC1 since we have subselected differentially expressed genes. We find clear separation of ST1 from ST4 and some separation from ST3 tumors. ST2 samples were markedly heterogeneous but the majority clustered with ST1 tumors, and ST3 tended to cluster between ST1 and ST4, although there was considerable admixture in these tumors along PC1. Figure 1c shows the PC1 values plotted using the median gene expression levels in each of the subgroups for these 6103 genes and gives a graphical depiction of the increasing variance as the tumors progress from ST1 to ST4.

In order to identify which of these 6103 clones distinguish one stage from another, we next performed a pairwise permutation *t*-test using these clones (see Materials and methods) on all combinations of stages of NB_{MYCN-NA}: ST1 vs ST2, ST1 vs ST3, ST1 vs ST4, ST2 vs ST3, ST2 vs ST4 and ST3 vs ST4 for NB_{MYCN-NA} tumors (Table 1). In this analysis, a gene was considered differentially expressed when the adjusted *P*-value obtained in the pairwise permutation *t*-test was less than 0.05, irrespective of the fold-change in gene expression (see Materials and methods). This analysis found differentially expressed genes representing 1434 distinct UniGene clusters when comparing ST1 vs ST4 NB_{MYCN-NA} tumors, 309 when comparing ST1 vs ST3, and 633 when comparing ST1 to ST3 and ST4. Using our stringent selection criteria, we found no differentially expressed genes when comparing ST1 vs ST2 or ST2 vs ST3, or ST3 vs ST4, which is not surprising given the closeness of these stages to one another in the PC1 plots (Figure 1c). These results establish a stage-specific expression signature in the NB tumors.

GO analysis of the stage-specific expression signatures

To explore the biological processes that are reflected in the differentially expressed genes in the stage-specific expression signature, we analysed the signatures using a

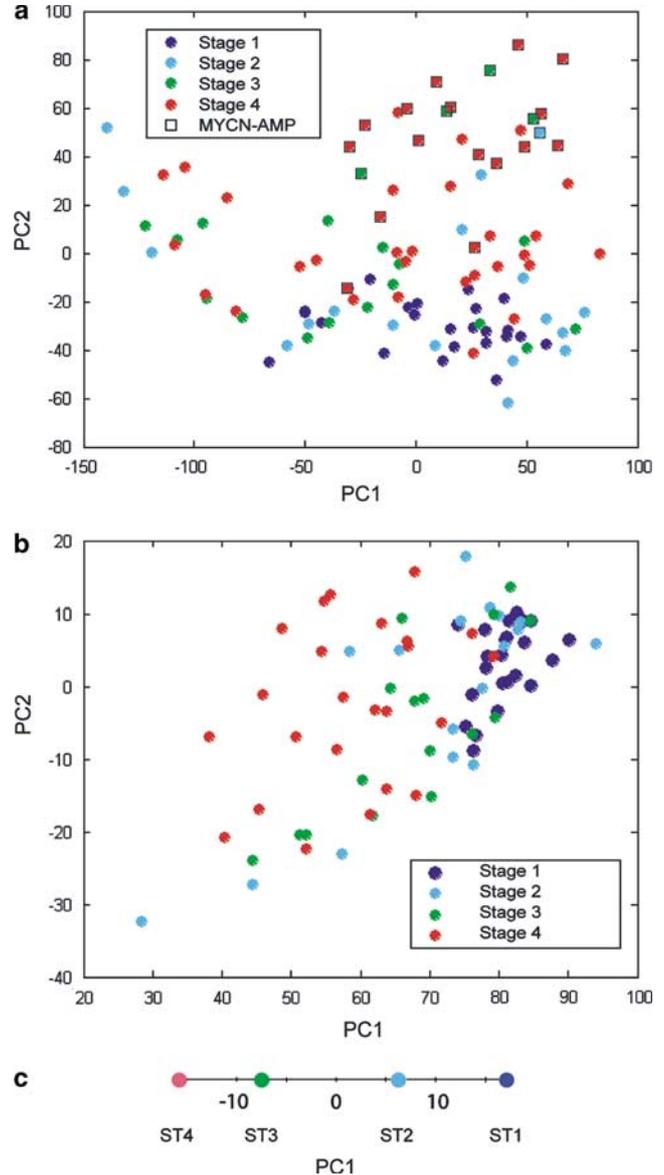


Figure 1 (a) The PCA separation of 103 NB tumors (see Materials and methods). The following labeling is used in all panels: blue – ST1; cyan – ST2; green – ST3; red – ST4; NB_{MYCN-AMP} tumors are enclosed in black squares. (b) PCA separation of NB_{MYCN-NA} tumors using only the 6103 ANOVA selected differentially expressed genes (using *P*-value cutoff, *P* < 0.05) (stage-specific signature). (c) PCA separation of median expression levels of the ANOVA genes for each of the four NB_{MYCN-NA} groups. PC1 is shown demonstrating increasing variance of the gene expression going from ST1, ST2 and ST3 to ST4

novel GO analysis based on PCA. Our motivation for this is as follows. We know that in the PCA plane, different expression patterns correspond to different directions and correlated genes lie in the same direction. We therefore postulated that genes that lie in the same direction have similar molecular functions contained within different biological processes. Therefore, we performed the PCA plots, using samples as variables. Thus, in these plots the genes whose expression profiles are correlated map to similar coordinates because they

Table 1 Results of the pairwise permutation *t*-test on all stages of NB_{MYCN-NA} tumors as well as NB_{MYCN-AMP}

Comparisons	Number of unique UniGenes		
	Upregulated	Downregulated	Total
ST1 vs ST4 (all MYCN-NA ^a)	800	634	1434
ST1 vs ST3 (all MYCN-NA ^a)	63	246	309
ST1 vs (ST3 and ST4) (all MYCN-NA ^b)	242	391	633
ST4 NB _{MYCN-NA} vs NB _{MYCN-AMP} ^c	1040	189	1229
ST1 vs NB _{MYCN-AMP} ^c	4547	2896	7443
ST4 NB _{MYCN-NA} vs ST4 NB _{MYCN-AMP} ^c	612	79	691

All using *P*-value cutoff of *P* < 0.05. ^aDerived from ANOVA-reduced gene set. ^bGenes differentially expressed in both ST3 and ST4 with respect to ST1. ^cDerived from the full 33K gene set

covary. Figure 2a shows the PCA plot of all 33 680 genes, represented by cyan dots, plotted in the space of PC1 and PC2. The overlaid blue dots represent the clones corresponding to the stage-specific signature common to both ST3 and ST4 compared to ST1 NB_{MYCN-NA} tumors (see Table 1) and the overlaid red dots are the clones corresponding to the unique ST4-specific signature in NB_{MYCN-NA} tumors. Thus, both red and blue clones constitute genes differentially expressed in ST4 compared to ST1 NB_{MYCN-NA} tumors.

We then divided the PC plot into 360 continuous bins (see Materials and methods and Figure 2a), and calculated the probability of the enrichment of the differentially expressed genes for a specific GO in each bin. In other words, an enrichment of a specific GO within a bin suggests the involvement of a common biological process for these correlated genes. We found that an overwhelming majority of the biological processes identified in this analysis are involved in cell proliferation:cell cycle, as represented by the probability heat map (Figure 2b and Table 2). This heat map show that the GO terms with the highest statistical significance, all of which involved in cell cycle, are clustered in the area corresponding to the bins located between 220° and 270° (Figure 2 and Table 2). This was the same region where the majority of upregulated genes in ST4 (but not ST3) were mapped to, marked by the red dots in Figure 2a. The genes in these bins with increased expression in ST4 NB_{MYCN-NA} tumors higher than 1.5-fold (compared to ST1) were selected and are shown in Table 3. Genes in this cluster participate in both S phase (GO:0000084) and M phase (GO:0000279) of the cell cycle, as well as the cell cycle checkpoint (GO:0000075). Interestingly, the majority of these cell cycle genes were composed of the genes upregulated in ST4 but not ST3 NB_{MYCN-NA} tumors (only 4.4% of these 159 clones with more than 1.5-fold upregulation are also upregulated in ST3).

MYCN amplification-specific expression signatures and GO analysis

Although several genes have been identified that are upregulated by the MYC proteins *in vitro* (Boon *et al.*, 2001; Fernandez *et al.*, 2003), the MYCN target genes critical for NB tumorigenesis and tumor progression have not yet been defined. In our analysis, *MYCN* is

found differentially expressed in NB_{MYCN-AMP} tumors, but not in NB_{MYCN-NA} tumors of ST1–4, which is evident from Figure 3a.

In order to identify the *MYCN* amplification-specific expression signature, a permutation *t*-test was performed only on ST4 NB_{MYCN-AMP} tumors vs ST4 NB_{MYCN-NA} tumors. This analysis found 691 differentially expressed genes, which are predominantly upregulated (89%). Furthermore, to determine the dependence of this signature on the stage of NB_{MYCN-AMP} tumors, the permutation *t*-test was also performed on all NB_{MYCN-AMP} (16 of ST4, four of ST3 and one of ST2) tumors vs ST4 NB_{MYCN-NA} tumors. In this analysis, we identified an additional 538 differentially expressed genes (Table 1). The PCA plot using these *MYCN* signature genes shows a nearly complete separation of the amplified samples from ST4 NB_{MYCN-NA} tumors (Figure 3b). This PCA plot demonstrates that *MYCN* amplification is associated with a unique *MYCN* amplification-specific expression signature, which appears to be independent of the stage of NB.

As with the MYCN-NA stage-specific profiles above, we next explored the biological processes that are reflected in the MYCN-AMP differentially expressed genes. The majority of the biological processes identified in GO analysis was carried out on the *MYCN* amplification-specific expression signatures involved in protein biosynthesis, as seen in the probability heat map (Figure 3c and Table 2). These results corroborated the finding by Boon *et al.* (2001) that genes encoding RP are upregulated in NB_{MYCN-AMP} tumors. The expression of RP is increased by 2-fold ± 0.3 (18 genes) compared to ST1 NB_{MYCN-NA} tumors. In addition to Boon's findings, we discovered an increase in the expression of mitochondrial RPs as well (*P* < 5 × 10⁻⁴), with upregulation of 1.6-fold ± 0.2 (10 genes).

Expression levels of the cell cycle genes in NB_{MYCN-AMP} tumors

The GO analyses identified that the predominate biological processes present in the stage and *MYCN* amplification-specific expression signatures were that of cell proliferation and protein biosynthesis, respectively. Since the cell cycle does not appear as the predominant GO process in NB_{MYCN-AMP} when compared to ST4

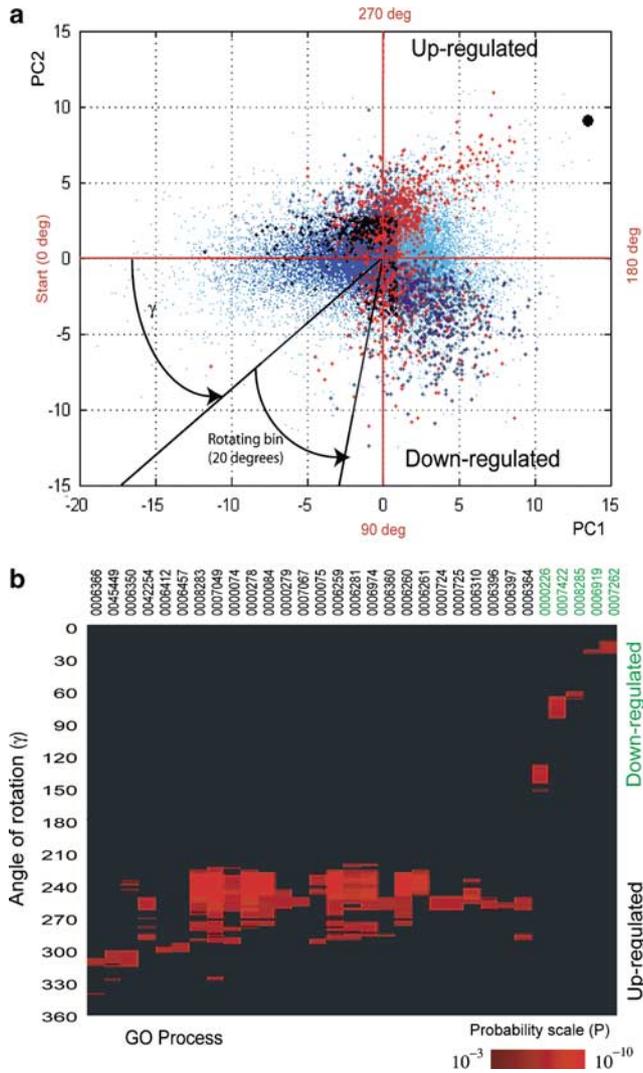


Figure 2 GO analysis of the genes of the ST4-specific expression signature in NB_{MYCN-NA} tumors compared to ST1. (a) All 33K clones (cyan dots) are plotted in the PC1 and PC2 space, where the tumors were used as variables in this PCA. Overlaid in blue are the clones corresponding to the stage-specific signature common to both ST3 and ST4 compared to ST1 NB_{MYCN-NA} tumors (see Table 1). Overlaid in red are the clones corresponding to the unique ST4-specific signature in NB_{MYCN-NA} tumors. Thus, both red and blue clones constitute genes differentially expressed in ST4 compared to ST1 NB_{MYCN-NA} tumors. This figure also illustrates the principle of the PCA-based GO analysis, where a segment (bin) is defined by a 20° angle that is used to select the genes plotted in the PCA space (only differentially expressed genes are considered, red and blue dots). The overlapping segments (bins, total of 360) are created by shifting the 20° angle by 1° starting at the 0° angle of rotation (negative PC1 axis) to the 360° angle of rotation, which makes a full circle. The angle γ defines the 'angle of rotation' formed between the PC1 axis and the 20° segment (bin). The genes in each segment are analysed by the GO analysis as described in Materials and methods. (b) The GO analysis was performed in the space of PC1 and 2 (see a) and the results are presented as a probability heat map, where the red intensity represents the probability of enrichment by random chance for each GO. The vertical axis numbers on the heat map refer to γ (angle of rotation in a). The result for the GO term 'process' is shown. The seven-digit numbers on top of the heat map denote the corresponding GO annotation, which are described in Table 2 in the same order

NB_{MYCN-NA} tumors, we next asked if the cell cycle genes identified in Table 3 are further upregulated in MYCN-AMP tumors. We found that these genes were further upregulated in MYCN-AMP tumors compared to ST1: the median is 2.7-fold (as compared to 1.6-fold in ST4NA vs ST1 tumors). It has been demonstrated previously that an association between the expression level of *MYCN* and the growth potential of NB cells exists (Schweigerer *et al.*, 1990; Negroni *et al.*, 1991; Schmidt *et al.*, 1994). Using this information, we asked if the *MYCN* expression level is correlated with the upregulation of the cell cycle genes. We found no difference in the mean level of *MYCN* expression between ST1 and ST4 NB_{MYCN-NA} tumors (Figure 3a) and only low levels of correlation (median correlation of 0.35 ± 0.12) were observed between *MYCN* expression and the cell cycle genes (data not shown), suggesting that *MYCN* expression alone cannot solely be responsible for the upregulation of the cell cycle genes in ST4 NB tumors.

Discussion

In this study, we performed gene expression profiling of 103 NB tumors, ST1–4 with and without *MYCN* amplification on cDNA microarrays containing 42,578 elements representing 25,933 unique UniGene clusters. To our knowledge, this represents the largest number of NB tumor samples profiled on the largest set of cDNA clones (Khan *et al.*, 2001; Berwanger *et al.*, 2002; Alaminos *et al.*, 2003; Mora *et al.*, 2003; McArdle *et al.*, 2004; Takita *et al.*, 2004). Similar to our study for the MYCN-NA tumors, Berwanger *et al.* found differentially expressed genes only when comparing 1 vs 4, but found no differences between 2 and 3. However, they found only 24 differentially expressed genes comparing ST1 and ST4, while in the same analysis we found 1434 unique genes. This was, in part, due to their use of a significantly smaller cDNA microarrays containing 4608 elements since the number of samples in both studies was comparable. Different from the previous study (Berwanger *et al.*, 2002), we identified 309 unique genes for ST3 NB_{MYCN-NA} tumors compared to ST1. Additionally, we identified 691 *MYCN* amplification-specific genes when we compared ST4 NB_{MYCN-AMP} tumors to ST4 NB_{MYCN-NA} tumors. Furthermore, to determine the dependence of this signature on stage of NB_{MYCN-AMP} tumors, we also compared all NB_{MYCN-AMP} (16 of ST4, four of ST3 and one of ST2) tumors vs ST4 NB_{MYCN-NA} tumors and found an additional 538 differentially expressed genes (Table 1). This result suggests that NB_{MYCN-AMP} tumors carry a stage-independent expression signature associated with *MYCN* amplification.

These results establish unique stage- and *MYCN* amplification-specific expression signatures in NB tumors; however, it exemplifies the propensity of these high-dimensional data sets to yield large numbers of differentially expressed genes. Therefore, although

Table 2 PCA-based GO analysis. Annotations that are significantly altered in ST4 MYCN-NA and MYCN-AMP tumors

GO term	Description
<i>Upregulated</i>	
Stage 4 vs 1	
0006366	Transcription from Pol II promoter
0045449	Regulation of transcription
0006350	Transcription
0042254	Ribosome biogenesis and assembly
0006412	Protein biosynthesis
0006457	Protein folding
0008283	Cell proliferation
0007049	Cell proliferation:cell cycle
0000074	Cell cycle:regulation of cell cycle
0000278	Cell cycle:mitotic cell cycle
0000084	Mitotic cell cycle:S phase of mitotic cell cycle
0000279	Cell proliferation:cell cycle:M phase
0007067	Cell cycle:M phase:nuclear division:mitosis
0000075	Regulation of cell cycle:cell cycle checkpoint
0006259	DNA metabolism
0006281	DNA metabolism:DNA repair
0006974	Response to DNA damage stimulus
0006360	Transcription, DNA-dependent:transcription from Pol I promoter
0006260	DNA metabolism:DNA replication
0006261	DNA replication:DNA-dependent DNA replication
0000724	DNA metabolism:double-strand break repair via homologous recombination
0000725	DNA metabolism: DNA repair:recombinational repair
0006310	DNA metabolism:DNA recombination
00063	RNA metabolism:RNA processing
0006397	RNA processing:mRNA processing
0006364	RNA metabolism:RNA processing:rRNA processing
<i>Downregulated</i>	
Stage 4 vs 1	
0000226	Microtubule cytoskeleton organization and biogenesis
0007422	Neurogenesis:peripheral nervous system development
0008285	Negative regulation of cell proliferation
0006919	Apoptosis:apoptotic program:caspase activation
0007262	JAK-STAT cascade:STAT protein nuclear translocation
<i>Upregulated</i>	
Stage 4 MYCN-AMP vs MYCN-NA	
0042147	Endosome organization and biogenesis:endosome transport
0007049	Cell cycle
0042254	Ribosome biogenesis and assembly
0006412	Protein biosynthesis
0006445	Regulation of translation
0006139	Nucleobase, nucleoside, nucleotide and nucleic acid metabolism
0006396	RNA metabolism:RNA processing
0006364	RNA processing:rRNA processing
0006350	Transcription
0009260	Ribonucleotide biosynthesis
0006941	Striated muscle contraction
0007275	Development
0007517	Organogenesis:muscle development
0006810	Cell growth and/or maintenance:transport
0006811	Transport:ion transport
0006626	Protein-mitochondrial targeting

several gene lists are generated and are likely to be a useful resource for the purpose of data mining, we have sought to discover the underlying processes represented by these differentially expressed genes using the GO database.

Our GO analysis of the stage-specific expression signatures discovered an enrichment of biological processes involved in proliferation and specifically in the cell cycle. We show that the cluster of genes corresponding to the cell cycle GO annotations are only

present in the ST4 NB_{MYCN-NA} tumors and distinguish them from all lower stage tumors, including ST3. Among the top five upregulated genes in this cluster, two are mapped to the commonly gained region of 17q23–25: survivin (*BIRC5*, 2.3-fold increased expression) and thymidine kinase (*TK1*, 2.5-fold). Survivin is an antiapoptotic protein, the expression of which has been shown to correlate with unfavorable histology and poor prognosis, to and promote survival of NB cell lines (Islam *et al.*, 2000). Its expression is known to be

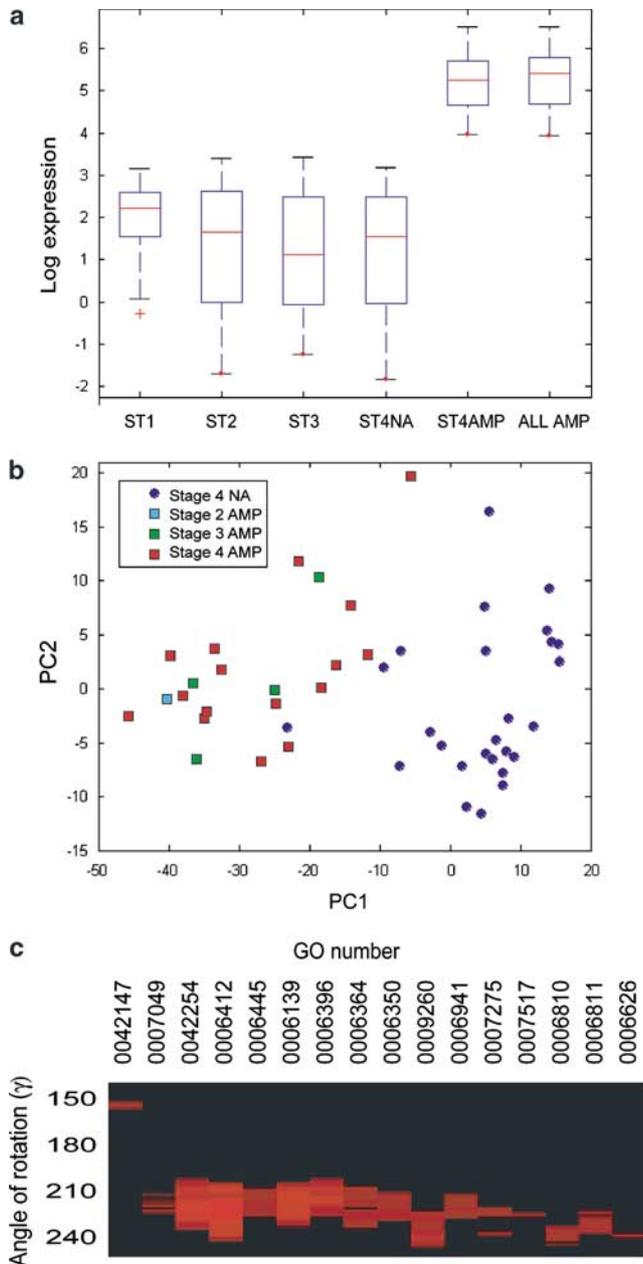


Figure 3 (a) *MYCN* expression. ST1 (23 tumors), ST2 (17 tumors), ST3 (16 tumors) and $NB_{MYCN-NA}$ ST4 (26 tumors) contain only nonamplified samples. ST4AMP represents 17 amplified tumors of ST4; AMP is composed of all the available amplified tumors (one of ST2, four of ST3 and 17 of ST4). The red bar denotes the median; the box encompasses the 50th percentile of the data. The error bars show the lower and upper quartiles. The cross represents an outlier. Log expression units are in reference to the microarray control expression. (b) PCA separation of $NB_{MYCN-AMP}$ tumors using only the *MYCN* amplification-specific expression signature. (c) The GO analysis of the genes of the *MYCN* amplification-specific expression signature. The GO analysis was performed in the space of PC1 and PC2 (in Figure 2a) and the results are presented as a probability heat map, where the red intensity represents the probability of enrichment by random chance for each cyto band. The vertical axis numbers on the heat map refer to γ (angle of rotation in Figure 2a). The result for the GO term "process" is shown. The seven-digit numbers on top of the heat map denote the corresponding GO annotation, which are described in Table 2 in the same order

regulated in a cell-cycle-dependent manner and shown to associate with centromere and mitotic spindle (Li *et al.*, 1998). In addition, this cluster includes all three groups of kinases involved in the regulation of the mitotic spindle checkpoint: NIMA-related kinases (*Nek2*), Aurora-A (*STK15*) and Polo kinases (*PLK1*, *PLK4*). The Aurora-A gene encodes a centrosome-associated kinase, which has been reported amplified and overexpressed in NB (Zhou *et al.*, 1998), as well as in several cancers (Ewart-Toland *et al.*, 2003). Its overexpression leads to centrosome amplification, abnormal regulation of the spindle checkpoint, chromosomal instability and transformation in mammalian cells. The majority of the genes in the cell cycle cluster are upregulated to the same or higher degree in the $NB_{MYCN-AMP}$ tumors. Among them, survivin is upregulated fourfold in $NB_{MYCN-AMP}$ tumors compared to ST1. In addition, we found that a partner of survivin (Aurora-B) in a complex regulating the kinetochore is upregulated in $NB_{MYCN-AMP}$ tumors, where Aurora-B requires survivin for its proper localization. Recently, Sugihara *et al.* (2004) showed that *MYCN* alone does not induce centrosome hyperamplification. However, in combination or followed by induced DNA damage, *MYCN* overexpression contributes to the centrosome hyperamplification, suggesting that higher levels of *MYCN* expression leads to dysregulation of centrosome maintenance. Our findings extend this observation further in pointing to a set of genes involved in centromere regulation.

Our analysis of *MYCN* amplification-specific expression signature corroborated the finding by Boon *et al.* (2001) that genes encoding RP are upregulated in $NB_{MYCN-AMP}$ tumors. In addition to Boon's findings, we discovered an increase in the expression of mitochondrial RP. It is possible that increased expression of RP is mediated by Dyskerin (*DKC1*), which is a direct target of *MYCN* (Boon *et al.*, 2001) and is a nuclear protein responsible for some early steps in ribosomal RNA processing (Ruggero and Pandolfi, 2003). *DKC1* is upregulated (1.9-fold in $NB_{MYCN-AMP}$ compared to ST4 $NB_{MYCN-NA}$ tumors and 2.5-fold compared to ST1) in an *MYCN*-dependent manner in both $NB_{MYCN-NA}$ and $NB_{MYCN-AMP}$ tumors (Pearson's correlation -0.77).

In conclusion, we have established stage- and *MYCN*-amplification-specific gene expression signatures and identified an association between the patterns of expression and the biological processes associated with an aggressive NB phenotype. Using our PCA-based statistical approach, we identified a specific subset of cell cycle and/or chromosome segregation genes that distinguish ST4 NB tumors from all lower stage tumors, including ST3. Furthermore, the control of the kinetochore assembly emerges from the GO analysis as one of the key biological processes associated with an aggressive NB phenotype. Additionally, we find that these genes are further upregulated or increased in expression in the most aggressive $NB_{MYCN-AMP}$ compared with the ST4 $NB_{MYCN-NA}$ tumors. Therefore, these cell cycle genes constitute important genes for biological validation and potential therapeutic targets in aggressive NB.

Table 3 Genes identified with the GO analysis that are associated with the cell cycle (GO:0007049) and upregulated in ST4 NBMYCN-NA tumors

Fold upregulation ^a		Unigene	Symbol	Cytoband	Description
ST4	NBMYCN-AMP				
NBMYCN-NA					
3.1	3.3	Hs.347524	MGC24665	16p13.2	Hypothetical protein MGC24665
2.5	1.9	Hs.164457	TK1	17q23.2–q25.3	Thymidine kinase 1, soluble
2.4	2.3	Hs.26516	ASF1B	19p13.13	ASF1 antisilencing function 1 homolog B (<i>S. cerevisiae</i>)
2.8	4.6	Hs.1578	BIRC5	17q25	Baculoviral IAP repeat-containing 5 (survivin)
2.3	3.2	Hs.250822	STK6/STK15	20q13.2–q13.3	Serine/threonine kinase 6 (Aurora-A)
2.2	3.8	Hs.104859	DKFZp762E1312	2q37.2	Hypothetical protein DKFZp762E1312
2.2	1.5	Hs.70937	HIST1H3H	6p22–p21.3	Histone 1, H3h
2.1	3	Hs.434886	CDCA5	11q12.1	Cell division cycle-associated 5
2.1	2.8	Hs.373547	CDCA2	8p21.2	Cell division cycle-associated 2
2.1	2.9	Hs.153479	ESPL1	12q13	Extraspindle poles like 1 (<i>S. cerevisiae</i>)
2	3	Hs.528654	FLJ11029	17q23.2	Hypothetical protein FLJ11029
2	1.9	Hs.153704	NEK2	1q32.2–q41	NIMA (never in mitosis gene a)-related kinase 2
2	4.4	Hs.38178	KLIP1	4q35.1	KSHV latent nuclear antigen interacting protein 1
1.9	4.8	Hs.70704	C20orf129	20q11.22–q12	Chromosome 20 open-reading frame 129
1.9	2.7	Hs.241517	POLQ	3q13.33	Polymerase (DNA directed), theta
1.8	2.8	Hs.52184	FLJ20618	22q12.2	Hypothetical protein FLJ20618
1.8	2.1	Hs.25292	JUNB	19p13.2	jun B proto-oncogene
1.8	3.6	Hs.96055	E2F1	20q11.2	E2F transcription factor 1
1.8	2.3	Hs.224137	HSPC109	9q34.13	Hypothetical protein HSPC109
1.8	2.5	Hs.283532	BM039	16q23.2	Uncharacterized bone marrow protein BM039
1.8	4.5	Hs.83758	CKS2	9q22	CDC28 protein kinase regulatory subunit 2
1.7	2.7	Hs.77171	MCM5	22q13.1	MCM5 minichromosome maintenance deficient 5, cell division cycle 46
1.7	2	Hs.335951	NY-REN-41	11p14.2	NY-REN-41 antigen
1.7	2.7	Hs.115474	RFC3	13q12.3–q13	Replication factor C (activator 1) 3, 38 kDa
1.7	2.1	Hs.31442	RECQL4	8q24.3	RecQ protein-like 4
1.7	2.4	Hs.48480	ZNF367	9q22	Zinc-finger protein 367
1.7	4	Hs.294088	GAJ	4q31.3	GAJ protein
1.7	3.5	Hs.99185	POLE2	14q21–q22	Polymerase (DNA directed), epsilon 2 (p59 subunit)
1.7	2.8	Hs.423348	MEN1	11q13	Multiple endocrine neoplasia I
1.7	2.9	Hs.298646	PRO2000	8q24.13	PRO2000 protein
1.6	3.1	Hs.511769	LOC113174	11p15.1	Hypothetical protein BC012010
1.6	3.1	Hs.435733	CDCA7	2q31	Cell division cycle-associated 7
1.6	4.2	Hs.122908	CDT1	16q24.3	DNA replication factor
1.6	3.9	Hs.443409	ODC1	2p25	Ornithine decarboxylase 1
1.6	4.7	Hs.462306	UBE2S	19q13.43	Ubiquitin-conjugating enzyme E2S
1.6	1.4	Hs.376792	CDH4	20q13.3	Cadherin 4, type 1, R-cadherin (retinal)
1.6	0.8	Hs.142442	HP1-BP74	1p36.12	HP1-BP74
1.6	1.6	Hs.290758	DDB1	11q12–q13	Damage-specific DNA binding protein 1, 127 kDa
1.6	2.4	Hs.110757	D21S2056E	21q22.3	DNA segment on chromosome 21 (unique) 2056 expressed sequence
1.6	0.9	Hs.56874	HSPB7	1p36.23–p34.3	Heat-shock 27 kDa protein family, member 7 (cardiovascular)
1.5	3	Hs.19114	HMGB3	Xq28	High-mobility group box 3
1.5	3.4	Hs.83765	DHFR	5q11.2–q13.2	Dihydrofolate reductase
1.5	2.1	Hs.437420	ABPI	7q34–q36	Amiloride binding protein 1 (amine oxidase (copper-containing))
1.5	1.6	Hs.48855	CDCA8	1p34.2	Cell division cycle-associated 8
1.5	2.9	Hs.172052	PLK4	4q27–q28	Polo-like kinase 4 (<i>Drosophila</i>)
1.5	1.4	Hs.194686	SLC25A14	Xq24	Solute carrier family 25 (mitochondrial carrier, brain), member 14
1.5	1.1	Hs.247551	MTX1	1q21	Metaxin 1
1.5	2	Hs.59461	DKFZP434C245	3p21.31	DKFZP434C245 protein
1.5	2.4	Hs.383913	BLM	15q26.1	Bloom syndrome
1.5	0.9	Hs.24395	CXCL14	5q31	Chemokine (C-X-C motif) ligand 14
1.5	3	Hs.34045	CDCA4	14q32.33	Cell division cycle-associated 4
1.5	2.1	Hs.99807	FLJ40629	2q14.1	Hypothetical protein FLJ40629
1.5	3.1	Hs.114311	CDC45L	22q11.21	CDC45 cell division cycle 45-like (<i>S. cerevisiae</i>)
1.5	1.4	Hs.283551	EIF2B5	3q27.3	Eukaryotic translation initiation factor 2B, subunit 5 epsilon, 82 kDa
1.5	1.6	Hs.155728	ATP5H	17q25	ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit d
1.5	1.8	Hs.47504	EXO1	1q42–q43	Exonuclease 1
1.5	1.9	Hs.87507	BRIP1	17q22–q24	BRCA1 interacting protein C-terminal helicase 1
1.5	2.7	Hs.179565	MCM3	6p12	MCM3 minichromosome maintenance deficient 3 (<i>S. cerevisiae</i>)
1.5	1.8	Hs.32748	SAE1	19q13.33	SUMO-1 activating enzyme subunit 1
1.5	1.9	Hs.329989	PLK1	16p12.3	Polo-like kinase 1 (<i>Drosophila</i>)
1.5	3.3	Hs.256301	MGC13170	19q13.41	Multidrug resistance-related protein

^aCompared to ST1 NBMYCN-NA

Materials and methods

NB tumors and microarrays

The data set consisted of 23 samples of ST1, 18 samples of ST2, 20 samples of ST3 and 47 samples of ST4 (103 total). NB_{MYCN-AMP} samples consisted of 17 samples of ST4, four samples of ST3 and one sample of ST2 (Supplemental Table 1). Pretreatment primary NB tumor samples were collected from The Children's Hospital at Westmead Tumour Bank (CHW, Australia), German Cancer Research Center (GCRC, Germany) and Cooperative Human Tissue Network (CHTN, Ohio, USA). The clinical and histological diagnosis was performed at the hospital where the tumors were obtained. Tumor samples were surgical specimens and snap-frozen in liquid nitrogen according to local procedures.

For microarray analysis, total RNA was extracted according to Wei and Khan (2002). The reference RNA used in these studies consisted of equal portions of total RNA obtained from the seven human cancer cell lines (Wei *et al.*, 2004). Total RNA was amplified as described by Wang *et al.* (2000); hybridization and washing of the microarrays were performed as described by Hegde *et al.* (2000). The microarrays used in these studies consisted of 42 578 cDNA clones, representing 25 933 UniGene clusters. The ratio data were normalized using quadrant normalization with minor modifications and quality filtered (Chen *et al.*, 2002), such that the average quality for a gene expression vector did not fall below 0.95. The total number of clones after filtering resulted in 33 680 clones that were used for the analyses described in this paper.

Permutation t-test and other statistical tests

To control a false-positive rate (FDR), a permutation *t*-test was performed with 1000 permutation cycles per test. The adjusted probability was calculated as the integrated probability for each tail. A paired *t*-test for each gene was performed as follows. First, two groups were formed according to the chosen partitioning of the full data set. Each group was filtered for the entries that did not meet the quality criterion, and then a homogeneity of variance test was performed for the two sample populations. For the populations with equal variance but unequal sample size, the Welch–Strathmore adjustment was performed to estimate the degrees of freedom (see, e.g. Zar, 1999). A Kolmogorov–Smirnov test was used to evaluate the normality of each sample population. When the normality distribution assumption was violated (<2% of all cases), a two-sample Kolmogorov–Smirnov test was performed instead of the *t*-test. A gene was considered differentially expressed when adjusted, $P < 0.05$. This false-discovery rate of 5% refers to the set of genes derived in the permutation *t*-test. Calculations of ANOVA, Pearson's correlation and *F*-test for the homogeneity of variance were carried out as described elsewhere (Zar, 1999).

Principal component analysis

The gene expression matrix X of the dimensions $n \times p$ was assembled with n rows containing individual gene expression vectors and p columns representing tumor expression profiles. The mean expression vector was subtracted by row centering the matrix. PCs were derived through the singular value decomposition (SVD) of the variance–covariance matrix of the filtered data set (33 680 clones \times N tumors, where $N \leq 103$). All computations were carried out in MATLAB (MathWorks Inc., Natick, MA, USA). In order to avoid ambiguity in the definitions of PCs, we refer to the eigenvectors derived in SVD as PCs, and newly derived variables as projections on PCs. If

SVD of a matrix X is presented as $X = USV^T$, where X is the centered matrix of gene expression, then the columns of V^T represent PCs in our definition. The vectors defined by V^T are referred to as eigentumors and the dimensions are defined by U as eigengenes (similar to the eigenarray nomenclature defined in Alter *et al.*, 2000).

For Figure 1c, we first calculated the median value of gene expression for the 6103 ANOVA genes (see above) for each of the four stages, and then performed PCA.

Probabilistic GO analysis

For all the clones that passed our quality filters (33 680 clones), projections of each clone on the plane formed by PC1 and PC2 were considered. Each clone, therefore, is represented by a coordinate pair $[pc_{1i}, pc_{2i}]$ on the plane. The genes in the PCA plane were binned such that each bin is defined by a 20° angle segment that is used to select the genes plotted in the PCA space. The use of a 20° angle was determined by as follows. Since our clones on the microarrays were redundant (33 680 clones representing 25 933 UniGene clusters), we first calculated the angle between multiple clones that represent a single UniGene cluster. Then, we calculated the average standard deviation for those angles and doubled it ($2 \times$ s.d.) and found that $2 \times$ s.d. of the angles was approximately a 20° angle. Thus, this angle will contain the majority of clones that represents a single UniGene cluster in the plane of PC1 and PC2. The overlapping segments (bins, total of 360) are created by shifting the 20° angle by 1° starting at 0° angle of rotation to 360° angle of rotation, which makes a full circle, see Figure 2a.

We analysed groups of genes preselected by the methods of analysis carried out in this study for an over-representation or enrichment of specific GO annotations. GO provides annotation for 47% of the clones in our experiment. From the directed acyclic graph structure of the GO, each node of annotation is coupled to over- or underlying nodes via an 'isa' or 'part-of' relation. A clone mapped to any given annotation is therefore also associated with the parent nodes. The GO provides a link from LocusLink identifiers to GO terms. We therefore mapped each clone to a UniGene cluster and used the locus link identifier to associate the clones with GO terms.

For the probabilistic analysis of GO annotations (Alter *et al.*, 2000; Ashburner *et al.*, 2000), each bin derived in PCA-based partitioning of the genes was analysed for over-representation of GO terms. For each GO term, the number of associated genes within each bin was compared with the number expected by random chance. By integration of the hypergeometric probability distribution, we estimated a *P*-value. We applied the Bonferroni correction to the threshold for significance by counting the number of GO annotations that were occupied by more than 50 genes in the set of differentially expressed unique genes. This resulted in an unadjusted *P*-value of 10^{-4} as a cutoff probability corresponding to 5% false-discovery rate in the final list of significant GO terms.

Abbreviations

NB, neuroblastoma; ST1, stage 1; ST2, stage 2; ST3, stage 3; NB_{MYCN-AMP}, *MYCN* amplified; NB_{MYCN-NA}, *MYCN* not amplified; PCA, principal component analysis; GO, gene ontology.

Acknowledgements

We thank Dr John Maris, and Children's Oncology Group (COG) and Dr Steven Qualman at the Cooperative Human Tissue Network for several of the tumors.

References

- Alaminos M, Mora J, Cheung NK, Smith A, Qin J, Chen L and Gerald WL. (2003). *Cancer Res.*, **63**, 4538–4546.
- Alter O, Brown PO and Botstein D. (2000). *Proc. Natl. Acad. Sci. USA*, **97**, 10101–10106.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G. (2000). *Nat. Genet.*, **25**, 25–29.
- 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.
- Boon K, Caron HN, van Asperen R, Valentijn L, Hermus MC, van Sluis P, Roobeek I, Weis I, Voute PA, Schwab M and Versteeg R. (2001). *EMBO J.*, **20**, 1383–1393.
- Brodeur GM. (2003). *Nat. Rev. Cancer*, **3**, 203–216.
- Chen Y, Kamat V, Dougherty ER, Bittner ML, Meltzer PS and Trent JM. (2002). *Bioinformatics*, **18**, 1207–1215.
- Cohn SL, Rademaker AW, Salwen HR, Franklin WA, Gonzales-Crussi F, Rosen ST and Bauer KD. (1990). *Am. J. Pathol.*, **136**, 1043–1052.
- Dominici C, Negroni A, Romeo A, McDowell H, Padula A, Pucci S, Cappelli C, Castello MA and Raschella G. (1992). *Anticancer Res.*, **12**, 59–63.
- Ewart-Toland A, Briassouli P, de Koning JP, Mao JH, Yuan J, Chan F, MacCarthy-Morrogh L, Ponder BA, Nagase H, Burn J, Ball S, Almeida M, Linardopoulos S and Balmain A. (2003). *Nat. Genet.*, **34**, 403–412.
- Fernandez PC, Frank SR, Wang L, Schroeder M, Liu S, Greene J, Cocito A and Amati B. (2003). *Genes Dev.*, **17**, 1115–1129.
- Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N and Quackenbush J. (2000). *Biotechniques*, **29**, 548–550, 552–554, 556 passim.
- Islam A, Kageyama H, Takada N, Kawamoto T, Takayasu H, Isogai E, Ohira M, Hashizume K, Kobayashi H, Kaneko Y and Nakagawara A. (2000). *Oncogene*, **19**, 617–623.
- Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C and Meltzer PS. (2001). *Nat. Med.*, **7**, 673–679.
- Li F, Ambrosini G, Chu EY, Plescia J, Tognin S, Marchisio PC and Altieri DC. (1998). *Nature*, **396**, 580–584.
- McArdle L, McDermott M, Purcell R, Grehan D, O'Meara A, Breatnach F, Catchpoole D, Culhane AC, Jeffery I, Gallagher WM and Stallings RL. (2004). *Carcinogenesis*, **9**, 1599–1609.
- Mora J, Gerald WL and Cheung NK. (2003). *Cancer Lett.*, **197**, 119–124.
- Negroni A, Scarpa S, Romeo A, Ferrari S, Modesti A and Raschella G. (1991). *Cell Growth Differ.*, **2**, 511–518.
- Ruggero D and Pandolfi PP. (2003). *Nat. Rev. Cancer*, **3**, 179–192.
- Schmidt ML, Salwen HR, Manohar CF, Ikegaki N and Cohn SL. (1994). *Cell Growth Differ.*, **5**, 171–178.
- Schwab M, Westermann F, Hero B and Berthold F. (2003). *Lancet Oncol.*, **4**, 472–480.
- Schweigerer L, Breit S, Wenzel A, Tsunamoto K, Ludwig R and Schwab M. (1990). *Cancer Res.*, **50**, 4411–4416.
- Sugihara E, Kanai M, Matsui A, Onodera M, Schwab M and Miwa M. (2004). *Oncogene*, **23**, 1005–1009.
- Takita J, Ishii M, Tsutsumi S, Tanaka Y, Kato K, Toyoda Y, Hanada R, Yamamoto K, Hayashi Y and Aburatani H. (2004). *Genes Chromosomes Cancer*, **40**, 120–132.
- Tonini GP and Romani M. (2003). *Cancer Lett.*, **197**, 69–73.
- van Noesel MM and Versteeg R. (2004). *Gene*, **325**, 1–15.
- Wang E, Miller LD, Ohnmacht GA, Liu ET and Marincola FM. (2000). *Nat. Biotechnol.*, **18**, 457–459.
- 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.
- Wei JS and Khan J. (2002). *DNA Microarrays: A Molecular Cloning Manual* Bowtell D and SJ (eds). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, pp. 110–119.
- Westermann F and Schwab M. (2002). *Cancer Lett.*, **184**, 127–147.
- Zar JH. (1999). *Biostatistical Analysis*. Prentice-Hall: Upper Saddle River, NJ.
- Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, Brinkley BR and Sen S. (1998). *Nat. Genet.*, **20**, 189–193.

Supplementary information accompanies the paper on Oncogene website (<http://www.nature.com/onc>).