vernier. For the same reason, activity corresponding to four isolated (that is, unbound) contextual elements (Fig. 2a) might not be decreased by neighbouring contextual elements and, thus, might produce high neural activity inhibiting the vernier target. For the extended standard grating, shine through is possible because little neural activity occurs in the vernier's close neighbourhood. Analogously, extended contextual gratings might not interfere with the vernier because activity corresponding to their inner elements is weak. Therefore, context–context inhibition prevents context–target suppression of the vernier target.

Contextual elements might directly influence the firing pattern of cortical neurons that code the vernier. For example, the activity of a neuron that is responding vigorously to a vernier target should diminish when isolated contextual elements are presented together with the standard grating. By contrast, increasing the number of contextual elements should yield a rebound of activity. Because vernier display times are short, we expect strongest modulation effects in the transient response of the neurons—that is, in the activity immediately after the presentation of the target. □

1. Knierim, J. J. & van Essen, D. C. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophys.* **67,** 961–980 (1992).
2. Li, W., Thier, P. & Wehrhahn, C. Contextual influence on orientation discrimination of humans and responses of neurons in V1 of alert monkeys. *J. Neurophys.* **83,** 941–954 (2000).
3. Nothdurft, H.-C., Gallant, J. L. & van Essen, D. C. Response modulation by texture surround in primate area V1: Correlates of pop-out under anesthesia. *Vis. Neurosci.* **16,** 15–34 (1999).
4. Herzog, M. H. & Koch, C. Seeing properties of an invisible element: feature inheritance and shine-through. *Proc. Natl Acad. Sci. USA* **98,** 4271–4275 (2001).
5. Herzog, M. H., Fahle, M. & Koch, C. Spatial aspects of object formation revealed by a new illusion, shine-through. *Vis. Res.* **41,** 2325–2335 (2001).
6. Mareschal, I., Sceniak, M. P. & Shapley, M. Contextual influences on orientation discrimination: binding local and global cues. *Vis. Res.* **41,** 1915–1930 (2001).
7. Wehrhahn, C., Li, W. & Westheimer, G. Patterns that impair discrimination of orientation in human vision. *Perception* **25,** 1053–1064 (1996).
8. Stemmler, M., Usher, M. & Niebur, E. Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics. *Science* **269,** 1877–1880 (1995).
9. Li, Z. Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc. Natl Acad. Sci. USA* **96,** 10530–10535 (1999).
10. Wolfson, S. S. & Landy, M. S. Long range interactions between oriented texture elements. *Vis. Res.* **39,** 933–945 (1999).
11. Nothdurft, H. C. Texture segmentation and pop-out from orientation contrast. *Vis. Res.* **31,** 1073–1078 (1991).
12. Cannon, M. W. & Fullenkamp, S. C. Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. *Vis. Res.* **31,** 1985–1998 (1991).
13. Lamme, V. A. The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* **15,** 1605–1615 (1995).
14. Zipser, K., Lamme, V. A. F. & Schiller, P. H. Contextual modulation in primary visual cortex. *J. Neurosci.* **16,** 7376–7389 (1996).
15. Caputo, G. The role of the background: Texture segregation and figure-ground segmentation. *Vis. Res.* **36,** 2815–2826 (1996).
16. Sengpiel, F., Sen A. & Blakemore, C. Characteristics of surround inhibition in cat area 17. *Exp. Brain Res.* **116,** 216–228 (1997).
17. Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J. & Davis, J. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* **378,** 492–496 (1995).
18. Macknik, S. L. & Livingstone, M. S. Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neurosci.* **1,** 144–149 (1998).
19. Polat, U. & Sagi, D. Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vis. Res.* **33,** 993–999 (1993).
20. Kapadia, M. K., Ito, M., Gilbert, C. D. & Westheimer, G. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron* **15,** 843–856 (1995).
21. Levitt, J. B. & Lund, J. S. Contrast dependence of contextual effects in primate visual cortex. *Nature* **387,** 73–76 (1997).
22. Bonneh, Y. & Sagi, D. Effects of spatial configuration on contrast detection. *Vis. Res.* **38,** 3541–53 (1998).
23. Yu, C. & Levi, D. Surround modulation in human vision unmasked by masking experiments. *Nature Neurosci.* **3,** 724–728 (2000).
24. Das, A. & Gilbert, C. D. Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature* **399,** 655–661 (1999).
25. Chubb, C., Sperling, G. & Solomon, J. A. Texture interactions determine perceived contrast. *Proc. Natl Acad. Sci. USA* **86,** 9631–9635 (1989).
26. Weisstein, N. & Harris, C. S. Visual detection of lines segments: an object-superiority effect. *Science* **186,** 752–755 (1974).
27. Banks, W. P. & White, H. Lateral interference and perceptual grouping in visual detection. *Percept. Psychophys.* **36,** 285–295 (1984).
28. Kurylo, D. D. Time course of perceptual grouping. *Percept. Psychophys.* **59,** 142–147 (1997).
29. Ernst, U. A., Pawelzik, K. R., Sahar-Pikielny, C. & Tsodyks, M. V. Intracortical origin of visual maps. *Nature Neurosci.* **4,** 431–436 (2001).

....

Prediction of central nervous system embryonal tumour outcome based on gene expression

Scott L. Pomeroy*, Pablo Tamayo†, Michelle Gaasenbeek†, Lisa M. Sturla*, Michael Angelo†, Margaret E. McLaughlin‡, John Y. H. Kim*§, Liliana C. Goumnerova‖, Peter M. Black‖, Ching Lau¶, Jeffrey C. Allen#, David Zagzag☆, James M. Olson**, Tom Curran††, Cynthia Wetmore††, Jaclyn A. Biegel‡‡, Tomaso Poggio§§, Shayan Mukherjee§§, Ryan Rifkin§§, Andrea Califano‖‖, Gustavo Stolovitzky‖‖, David N. Louis¶¶, Jill P. Mesirov†, Eric S. Lander†## & Todd R. Golub†§☆☆

* *Division of Neuroscience, Department of Neurology;* ‡ *Department of Pathology;* ‖ *Department of Neurosurgery;* ☆☆ *Department of Medicine, Children's Hospital;* § *Department of Pediatric Oncology, Dana-Farber Cancer Institute; and* ¶¶ *Department of Pathology and Neurosurgical Service, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA*
† *Whitehead Institute/MIT Center for Genome Research, and* §§ *McGovern Institute, Center for Biological and Computational Learning, AI Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
¶ *Division of Pediatric Oncology, Baylor College of Medicine, Houston, Texas 77030, USA*
\# *Beth Israel Medical Center, New York 10128, USA*
☆ *Department of Pathology, New York University School of Medicine, New York 10016, USA*
** *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA*
†† *Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA*
‡‡ *Division of Human Genetics, The Children's Hospital of Philadelphia, Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA*
\## *Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
‖‖ *IBM Watson Research Center, Yorktown Heights, New York 10598, USA*

....

Embryonal tumours of the central nervous system (CNS) represent a heterogeneous group of tumours about which little is known biologically, and whose diagnosis, on the basis of morphologic appearance alone, is controversial. Medulloblastomas, for example, are the most common malignant brain tumour of childhood, but their pathogenesis is unknown, their relationship to other embryonal CNS tumours is debated[1,2], and patients' response to therapy is difficult to predict[3]. We approached these problems by developing a classification system based on DNA microarray gene expression data derived from 99 patient samples. Here we demonstrate that medulloblastomas are molecularly distinct from other brain tumours including primitive neuro-ectodermal tumours (PNETs), atypical teratoid/rhabdoid tumours
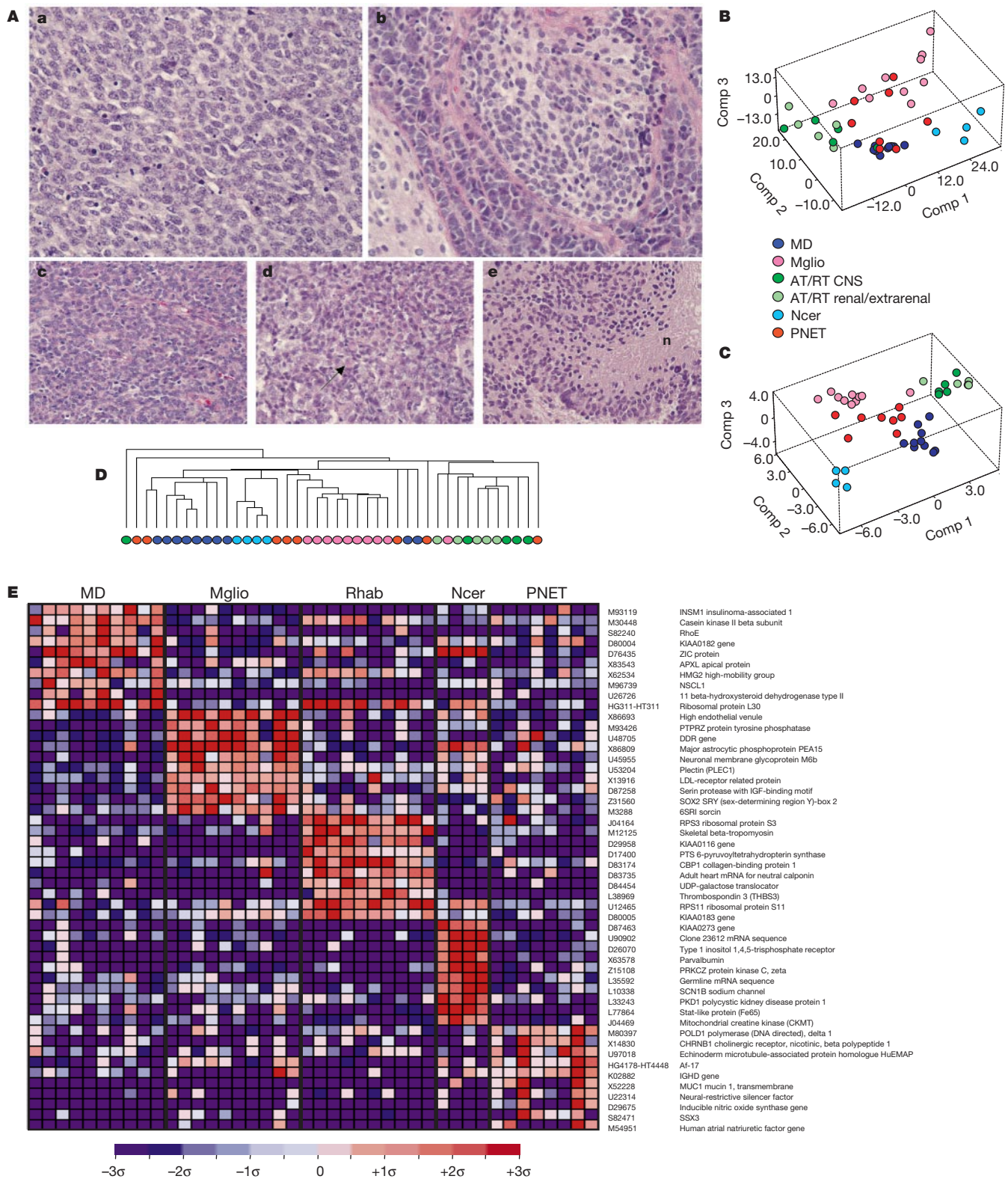
Figure 1 Classification of embryonal brain tumours by gene expression. **A**, Representative photomicrographs of embryonal and non-embryonal tumours. **a**, Classic medulloblastoma; **b**, desmoplastic medulloblastoma; **c**, supratentorial primitive neuroectodermal tumour (PNET); **d**, atypical teratoid/rhabdoid tumour (AT/RT; arrow indicates rhabdoid cell morphology); and **e**, glioblastoma with pseudopalisading necrosis (n). Magnification at 400×. **B**, Principal component analysis (PCA) of tumour samples using all genes exhibiting variation across the data set. The axes represent the three linear combinations of genes that account for most of the variance in the original data set (see Supplementary Information I and III). MD, medulloblastoma; Mglio, malignant glioma; Ncer, normal cerebella. **C**, PCA using 50 genes selected by signal-to-noise metric to be most highly associated with each tumour type (the top 10 for each tumour are listed in **E**). **D**, Clustering of tumour samples by hierarchical clustering using all genes exhibiting variation across the data set. **E**, Signal-to-noise rankings of genes comparing each tumour type to all other types combined (see Supplementary Information I). For each gene, red indicates a high level of expression relative to the mean; blue indicates a low level of expression relative to the mean. Rhab, rhabdoid. The standard deviation (σ) from the mean is indicated.
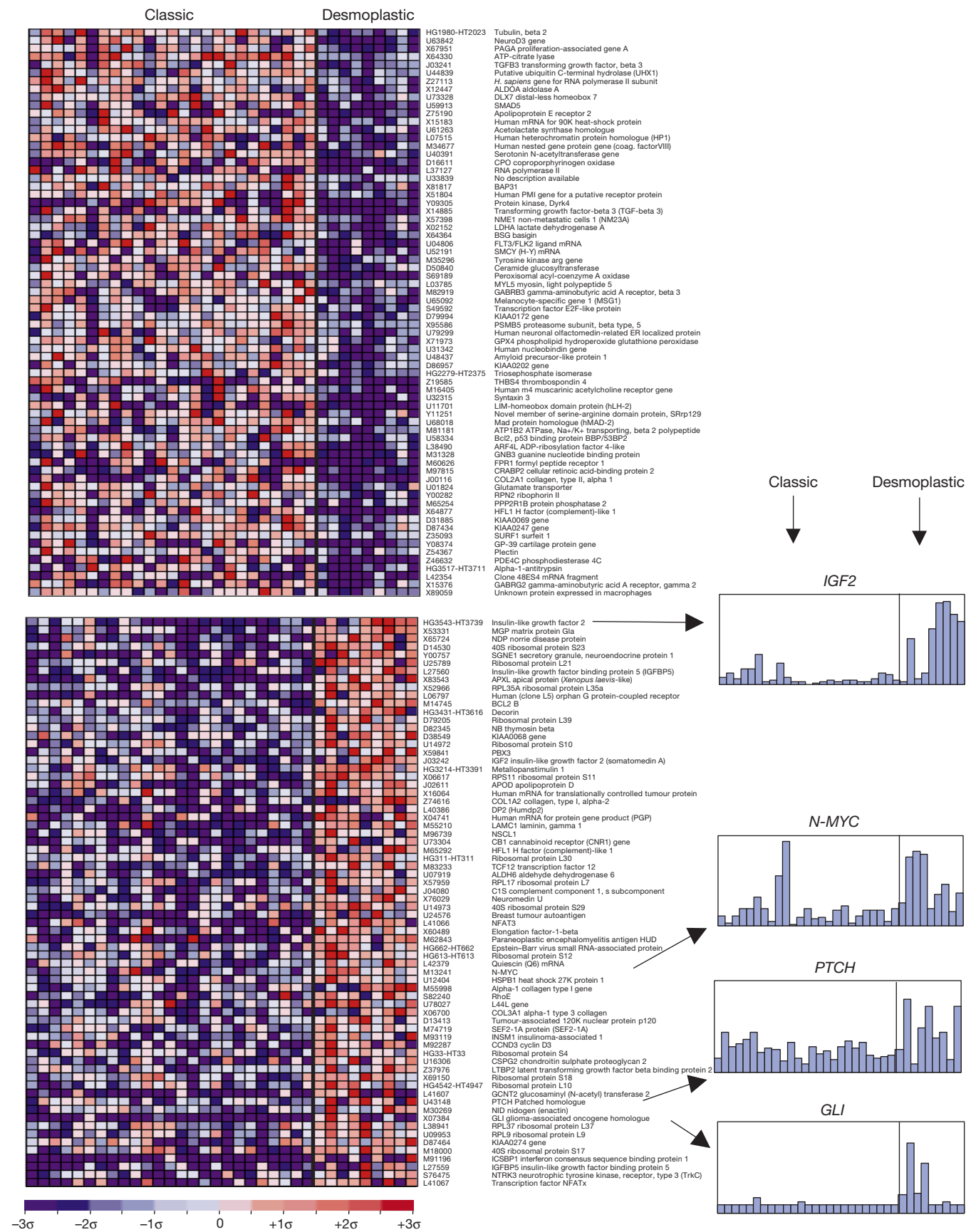
437

Figure 2 Differential expression of genes in classic versus desmoplastic medullo-blastomas. Genes were ranked by the signal-to-noise metric according to their correlation with the classic versus desmoplastic distinction. Genes shown are those more highly correlated with the distinction than 99% of permutations of the class labels ($P < 0.01$; see Supplementary Information III). GenBank accession numbers and gene descriptions are shown. Genes regulated by SHH are shown at the right. Normalized level of expression of selected SHH-regulated genes is shown at right. Each bar represents a different tumour.

438

NATURE | VOL 415 | 24 JANUARY 2002 | www.nature.com

(AT/RTs) and malignant gliomas. **Previously unrecognized evidence supporting the derivation of medulloblastomas from cerebellar granule cells through activation of the Sonic Hedgehog (SHH) pathway was also revealed. We show further that the clinical outcome of children with medulloblastomas is highly predictable on the basis of the gene expression profiles of their tumours at diagnosis.**

We first addressed the problem of distinguishing different embryonal CNS tumours from each other. This is important as the classification of these tumours based on histopathological appearance is debated (Fig. 1A). There are two hypotheses regarding the classification of medulloblastomas: the first is that they are part of a larger class of PNETs arising from a common cell type in the subventricular germinal matrix[1], the second is that they arise from cerebellar granule cell progenitors[2]. To begin to generate a molecular taxonomy of CNS embryonal tumours, we analysed the gene expression profiles of 42 patient samples (data set A: 10 medulloblastomas, 5 CNS AT/RTs, 5 renal and extrarenal rhabdoid tumours, and 8 supratentorial PNETs, as well as 10 non-embryonal brain tumours (malignant glioma) and 4 normal human cerebella). RNA extracted from frozen specimens was analysed with oligonucleotide microarrays containing probes for 6,817 genes. The gene expression data are available as Supplementary Information II (see also http://www.genome.wi.mit.edu/MPR/CNS).

To determine whether the different types of tumours could be molecularly distinguished, we used a method of data reduction called principal component analysis in which the high dimensionality of the data was reduced to three viewable dimensions representing linear combinations of variables (genes) that account for most of the variance in the original data set (Fig. 1B)[4]. Normal brain was easily separable from the brain tumours, and the different tumour types were similarly separable. Separation of tumour types was also seen using hierarchical clustering (Fig. 1D)[5]. A more appropriate strategy for distinguishing known tumour types, however, is to use supervised learning methods to identify the genes most highly correlated with the tumour type distinctions (Fig. 1C, E). Analysis of 1,000 random permutations of the data failed to yield a separation of tumour classes to the extent observed in Fig. 1C, indicating that the observed gene expression patterns could not be explained by chance (Supplementary Information III). The robustness of these markers for classification was further investigated using a weighted-voting algorithm and evaluated by cross validation testing[6]. Correct classification of the tumours could be achieved with accuracy (35 out of 42 correct classifications, $P < 10^{-10}$ compared with random classification; see Supplementary Information III).

As expected, malignant gliomas were clearly separable from medulloblastomas, reflecting the derivation of gliomas from cells of non-neuronal origin. Consistent with this, the gliomas expressed

genes typical of the astrocytic and oligodendrocytic lineage (*PEA15*, *SOX2*, *PMP2*, *Olig-2*, TrkB kinase-negative splice variant, *S100*, *GFAP*), genes related to metabolism (fructose 2,6-bisphosphatase, glutamate dehydrogenase), and genes involved in cell differentiation (*ID2*, *GDF1*, *TYK2*; Fig. 1E and Supplementary Information III). The medulloblastomas form a cluster that is also separate from the PNETs (Fig. 1C), supporting the hypothesis that these two classes of embryonal tumours are molecularly distinct. Among the genes most highly correlated with the medulloblastoma class were *ZIC* and *NSCL1*, encoding transcription factors that are specific for cerebellar granule cells (Fig. 1E)[7,8]. This result suggests that medulloblastomas, but not PNETs, arise from cerebellar granule cells, or alternatively, have activated the transcriptional programme of cerebellar granule cells.

We next analysed the AT/RT tumours, which have only recently been distinguished from medulloblastoma[9]. Accurate identification of AT/RT is particularly important because patients with these tumours have an extremely poor prognosis. AT/RT tumours arise in the CNS or in other organs such as the kidney, where they are referred to as rhabdoid tumours. Most tumours harbour *hSNF5/INI1* mutations, but it is unknown whether AT/RTs arising in different anatomical locations are molecularly distinct[9–11]. As shown in Fig. 1C, the AT/RTs and rhabdoid tumours were easily distinguishable from the other tumour types in the study. Of note, the CNS AT/RTs and abdominal rhabdoid tumours were molecularly similar despite having arisen in different anatomical locations. This finding supports the idea that they arise from a similar cell of origin. Alternatively, a common mechanism of transformation may yield similar transcriptional programmes in cells of distinct origin. Markers of the distinction between AT/RT and rhabdoid tumours include genes specifically expressed during myogenesis, including skeletal β-tropomyosin, neutral calponin, *NFAT3*, and myosin regulatory light chain (Fig. 1E and Supplementary Information III). This finding is consistent with the hypothesis that the tumours have a mesenchymal origin.

We next focused on molecular heterogeneity within a single tumour type, medulloblastoma. The principal histological subclass of medulloblastoma is desmoplastic medulloblastoma, although its diagnosis is highly subjective (Fig. 1A). Desmoplastic medulloblastoma is of interest because it is seen with high frequency in patients with Gorlin's syndrome, a rare autosomal dominant disorder resulting from mutation of the SHH receptor *PTCH*[12,13]. It is unclear whether dysregulation of the SHH pathway, known to be mitogenic for cerebellar granule cells, is also involved in the pathogenesis of sporadic desmoplastic medulloblastoma[14–18].

To determine whether desmoplastic and classic medulloblastoma are distinguishable by gene expression, we analysed 34 medulloblastoma samples (data set B) whose histology was scored using World Health Organization (WHO) criteria[19]. As shown in Fig. 2, a sharp and statistically significant gene expression signature of desmoplastic histology was evident, and this signature was sufficient for correct classification of 33 out of 34 tumours ($P = 8.6 \times 10^{-7}$ compared with random classification; see Supplementary Information III). Notably, among the genes most highly correlated with desmoplastic medulloblastoma was *PTCH* (itself a transcriptional target of SHH), as well as two other SHH downstream targets: *GLI*[20] and *N-MYC* (A. Kenney and D. Rowitch, personal communication). Furthermore, insulin-like growth factor II expression was correlated with desmoplastic histology, and its expression is essential for SHH-mediated tumorigenesis in mice[21]. Taken together, the transcriptional profiling indicates that sporadic desmoplastic medulloblastomas, like tumours associated with Gorlin's syndrome, are characterized by activation of the SHH signalling pathway, further supporting the proposal that SHH dysregulation may be important in the pathogenesis of medulloblastoma.

A clinical challenge concerning medulloblastoma is the highly variable response of patients to therapy. Whereas some patients are
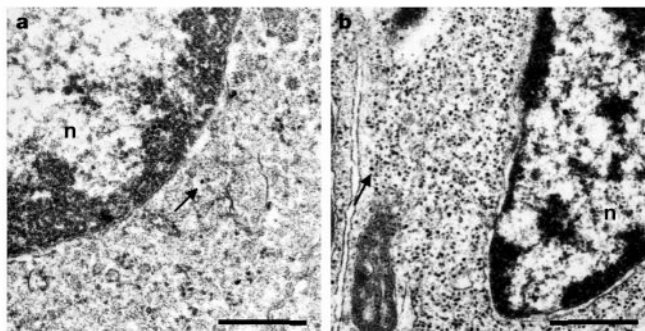

Figure 3 Representative electron micrographs showing medulloblastomas with low ribosome (**a**) and high ribosome (**b**) content. Each panel shows a portion of a single cell with a portion of the nucleus (n) (arrows designate ribosomes). Scale bars, 0.5 μm.
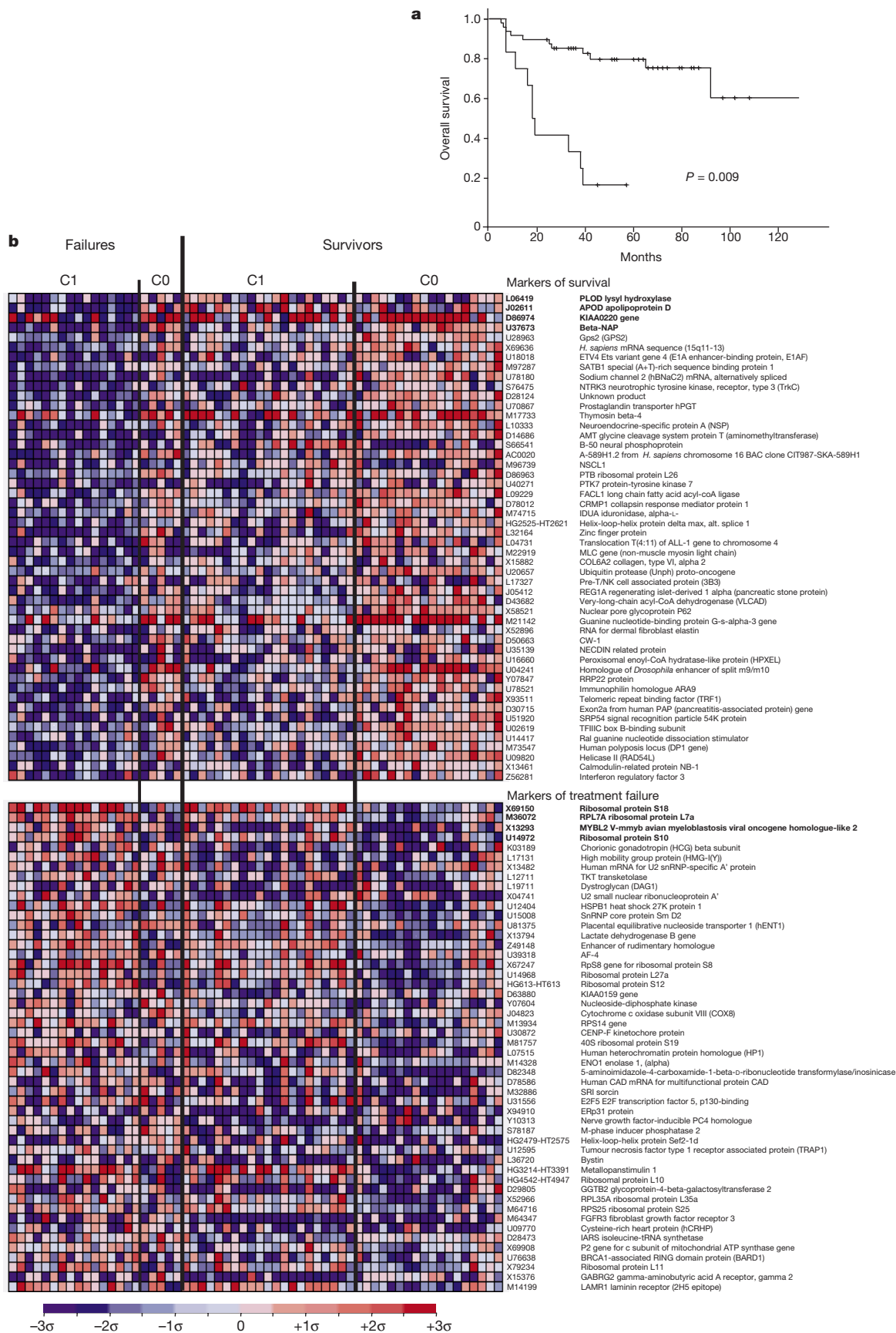
Figure 4 Predicting medulloblastoma outcome by gene expression profiling. **a**, Kaplan–Meier overall survival curves for patients predicted to survive and patients predicted to be treatment failures using an 8-gene *k*-NN model. Vertical hash marks indicate time of censorship. **b**, Fifty genes most highly associated with favourable outcome (top panel) or with treatment failure (bottom panel) according to the signal-to-noise metric. Samples are further sorted according to their membership in the two unsupervised SOM-derived clusters (C0, C1). Class C1 tumours are notable for their high ribosomal content. The 8 genes most frequently used by the *k*-NN outcome predictor are indicated in bold. The colour scheme is the same as Fig. 1E.

cured by chemotherapy and radiation, others have progressive disease[3]. Currently, the only prognostic factor used in clinical practice is tumour staging; a reflection of postoperative tumour size and the presence of metastases. Unfortunately, staging-based prognostication is imperfect in that many patients with low-stage disease still succumb to their disease. There are currently no molecular markers of outcome used in clinical practice for any brain tumour. High levels of expression of the neurotrophin-3 receptor (TRKC), however, have been reported to correlate with a favourable medulloblastoma outcome[22,23], suggesting a molecular basis for the variability of medulloblastoma outcome. Molecular correlates of medulloblastoma metastasis have also been recently reported[24].

To explore the heterogeneity in response to treatment of medulloblastomas, we expanded our analysis to include 60 similarly treated patients from whom biopsies were obtained before receiving treatment, and for whom clinical follow-up was available (data set C). We first investigated whether clustering methods would identify biologically distinct subsets of the tumours. The tumours were clustered into two groups using self-organizing maps (SOM); an unsupervised algorithm that groups samples into a predetermined number of clusters on the basis of their gene expression patterns[6,25]. The genes most highly correlated with the SOM clusters were primarily ribosomal protein-encoding genes (Supplementary Information III), suggesting differences in ribosome biogenesis. Blinded electron microscopic examination of 9 samples by 3 observers confirmed that tumours falling into the cluster characterized by high expression of ribosomal protein genes contained higher numbers of ribosomes ($P = 0.03$, Fisher's exact test; Fig. 3). We next investigated whether the SOM-derived clusters were correlated with patient survival. No statistically significant difference in the proportion of survivors compared with treatment failures in each cluster was observed (Fisher's exact test, $P = 0.1$; see Supplementary Information III). Because unsupervised methods are generally not the most appropriate analytical approach to predicting known distinctions such as outcome, we developed a supervised learning outcome predictor based on gene expression in which the classifier 'learns' the distinction between patients who are alive after treatment ('survivors') compared with those who succumbed to their disease ('failures', minimum follow-up of 24 months for surviving patients, overall median 41.5 months).

We used a k-nearest neighbours (k-NN) algorithm[26] that computes the distance of a test sample to each of the training set samples, each of which has an associated class (in this case, survivor or failure), and then predicts the class of the test sample to be that of the majority of the k-closest samples. The k-NN classifier was evaluated by cross-validation, whereby one sample is randomly withheld, a model is trained on the remaining samples, and the model is then used to predict the class of the withheld sample. The process is repeated until all of the samples are tested.

Outcome predictions based on gene expression were statistically significant for k-NN models ranging from 2 to 21 genes, with optimal predictions made by an 8-gene model that made only 13/60 classification errors (Fisher's exact test, $P = 0.0002$). Shown most clearly by a Kaplan–Meier survival analysis in Fig. 4a, patients that were predicted to be survivors had a 5 year overall survival of 80% compared with 17% for patients predicted to have a poor outcome ($P = 0.000003$, log rank test). A more conservative method of assessing statistical significance is to attempt to optimize classifiers of random permutations of the survivor/failure class labels. We performed 1,000 such permutations and found only 9 for which prediction accuracy matched or exceeded our observed result (Supplementary Information III), indicating that the result is unlikely to be achieved by chance ($P = 0.009$). We subsequently tested several other classification algorithms including weighted voting[6,27], support vector machines[28,29] and IBM SPLASH[30], all of which performed with similarly high accuracy (Supplementary Information I and III).

We explored further the clinical value of the predictor by considering existing prognostic factors for medulloblastoma outcome. Patients with localized disease (M0) had a more favourable outcome compared with patients with involvement of the cerebrospinal fluid or with distant metastases (M+) ($P = 0.03$ comparing M0 with M+ by Kaplan–Meier analysis), although not all M0 patients survived. When our outcome predictor was applied only to the 42 M0 patients, the prediction of outcome remained significant ($P = 0.002$), indicating that the expression-based predictor substantially improved staging-based prognostication. Similarly, prediction based on TRKC expression was imperfect in this series in that not all patients in the unfavourable (TRKC-low) category died. When our gene expression-based predictor was applied to the 33 TRKC-low patients, the surviving patients could be significantly separated from those who succumbed to their disease ($P = 0.01$, Supplementary Information III). Of note, not all patients in this study received identical therapy. However, restricting our analysis to the 35 patients that received surgery, vincristine, cisplatin and cyclophosphamide, the predictor continued to yield a significant Kaplan–Meier survival distinction ($P = 0.0012$). Taken together, these results demonstrate that the outcome predictor based on gene expression exceeds other approaches to prognosis determination.

A number of genes not previously associated with clinical outcome were identified (Fig. 4b). Those correlated with favourable outcome included many genes characteristic of cerebellar differentiation (vesicle coat protein β-NAP, NSCL1, TRKC, sodium channels), and genes encoding extracellular matrix proteins (PLOD lysyl hydroxylase, collagen type V αI, elastin). As expected, TRKC expression was correlated with a favourable outcome, consistent with previous reports of this association[22,23]. In contrast, genes related to cerebellar differentiation were underexpressed in poor prognosis tumours, which were dominated by the expression of genes related to cell proliferation and metabolism (MYBL2, enolase 1, LDH, HMG1(Y), cytochrome C oxidase) and multidrug resistance (sorcin). Genes correlated with poor outcome included a number of the ribosomal protein-encoding genes identified by the SOM clustering experiments (Fig. 4b). This indicates that whereas this ribosomal signature is correlated with poor outcome, optimal outcome prediction requires not only these genes, but also genes correlated with a favourable outcome that were not identified by the unsupervised clustering analysis.

The routine clinical implementation of genomics-based outcome predictors must await confirmation in independent data sets, and the models may need to be modified as treatment regimens evolve. For patients predicted to have a favourable outcome, efforts to minimize toxicity of therapy might be indicated, whereas for those predicted not to respond to standard therapy, earlier treatment with experimental regimens might be considered. This work illustrates how genomic technologies have the potential to advance treatment planning beyond the empiric, towards a more molecularly defined, individualized approach to medicine. □

Methods

Patient samples

Patients included 60 children with medulloblastomas, 10 young adults with malignant gliomas (WHO grades III and IV), 5 children with AT/RTs, 5 with renal/extrarenal rhabdoid tumours, and 8 children with supratentorial PNETs (see Supplementary Information I). Medulloblastoma patients were treated with craniospinal irradiation to 2,400–3,600 centiGray (cGy) with a tumour dose of 5,300–7,200 cGy. All patients with medulloblastoma were treated with chemotherapy consisting of cisplatin and vincristine, plus combinations of carboplatin, etoposide, cyclophosphamide or lumustine (1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea, CCNU) (details in Supplementary Information II). Samples were snap frozen in liquid nitrogen and stored at −80 °C. Studies were done with approval of the Committee for Clinical Investigation of Boston Children's Hospital. The data were organized into three sets: data set A (42 samples containing 10 medulloblastomas, 10 malignant gliomas, 10 AT/RTs, 8 PNETs and 4 normal cerebella); data set B (34 samples containing 9 desmoplastic medulloblastomas and 25 classic medulloblastomas); and data set C (60 samples containing 39 medulloblastoma survivors and 21 treatment failures). The clinical attributes of each of the patients in the study are available

in Supplementary Information II. Tissues were homogenized in guanidinium isothio-cyanate and RNA was isolated by centrifugation over a CsCl gradient. RNA integrity was assessed either by northern blotting or by gel electrophoresis. Ten–twelve micrograms total RNA was used to generate biotinylated antisense RNAs, which were hybridized overnight on HuGeneFL arrays containing 5,920 known genes and 897 expressed sequence tags, as described previously[6]. Arrays were scanned on Affymetrix scanners and the expression value for each gene was calculated using GENECHIP software (Affymetrix, Santa Clara, California). Minor differences in microarray intensity were corrected using a linear scaling method as detailed in Supplementary Information I. Scans were rejected if the scaling factor exceeded 3, fewer than 1,000 genes received 'present' calls, or microarray artefacts were visible.

Preprocessing and clustering

The gene expression data were subjected to a variation filter that excluded genes showing minimal variation across the samples being analysed, as detailed in Supplementary Information I.

The data were first normalized by standardizing each column (sample) to mean 0 and variance 1. SOMs were performed using our GeneCluster clustering package (http://www.genome.wi.mit.edu/MPR/software). Hierarchical clustering was performed using Cluster and TreeView software[5]. Principal component analysis (PCA) was performed by computing and then plotting the three principal components using the S-Plus statistical software package (www.insightful.com/products/desktop.asp) using default settings.

Supervised learning

Genes correlated with particular class distinctions (for example, classic versus desmo-plastic medulloblastoma) were identified by sorting all of the genes on the array according the signal-to-noise statistic $(\mu_0 - \mu_1)/(\sigma_0 + \sigma_1)$, where μ and σ represent the median and standard deviation of expression, respectively, for each class. Similar results were obtained using a standard t-statistic as the metric $((\mu_0 - \mu_1)/\sqrt{(\sigma_0^2/N_0 + \sigma_1^2/N_1)})$, where N represents the number of samples in each class (see Supplementary Information). Permutation of the column (sample) labels was performed to compare these correlations to what would be expected by chance in 99% of the permutations. For classification, we developed a modification of the k-NN algorithm[26] that predicts the class of a new data point by calculating the euclidean distance (d) of the new sample to the k nearest samples (for these experiments we set $k = 5$) in the training set using normalized gene expression data, and selecting the class to be that of most of the k samples. The weight given to each neighbour was $1/d$. The k-NN models were evaluated by 60-fold 'leave-one-out' cross-validation, whereby a training set of 59 samples was used to predict the class of a randomly withheld sample, and the cumulative error rate was recorded. We tested models with variable numbers of genes (1–200, selected according to their correlation with the survivor versus treatment failure distinction in the training set) in this manner. An 8-gene k-NN outcome prediction model yielded the lowest error rate, and was therefore used to generate Kaplan–Meier survival plots using S-Plus. Predictors using metastatic staging or $TRKC$ expression were constructed by finding the decision boundary half way between the classes, (μclass0 + μclass1)/2 using either the staging values 0 versus 1, 2, 3, 4 or the continuous $TRKC$ microarray gene expression levels, and then predicting the unknown sample according to its location with respect to that boundary.

1. Rorke, L. B. The cerebellar medulloblastoma and its relationship to primitive neuroectodermal tumors. *J. Neuropathol. Exp. Neurol.* **42,** 1–15 (1983).
2. Kadin, M. E., Rubinstein, L. J. & Nelson, J. S. Neonatal cerebellar medulloblastoma originating from the fetal external granular layer. *J. Neuropath. Exp. Neurol.* **29,** 583–600 (1970).
3. Packer, R. J. *et al.* Treatment of children with medulloblastomas with reduced-dose craniospinal radiation therapy and adjuvant chemotherapy: a children's cancer group study. *J. Clin. Oncol.* **17,** 2127–2136 (1999).
4. Mardia, K. V., Kent, J. T. & Bibby, J. M. *Multivariate Analysis* (Academic, London, 1979).
5. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).
6. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286,** 531–537 (1999).
7. Aruga, J. *et al.* A novel zinc finger protein, Zic, is involved in neurogenesis, especially in the cell lineage of cerebellar granule cells. *J. Neurochem.* **63,** 1880–1890 (1994).
8. Yokota, N. *et al.* Predominant expression of human Zic in cerebellar granule cell lineage and medulloblastoma. *Cancer Res.* **56,** 377–383 (1996).
9. Rorke, L. B., Packer, R. J. & Biegel, J. A. Central nervous system atypical teratoid/rhabdoid tumors of infancy and childhood: definition of an entity. *J. Neurosurg.* **85,** 56–65 (1996).
10. Biegel, J. A. *et al.* Germ-line and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. *Cancer Res.* **59,** 74–79 (1999).
11. Versteege, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394,** 203–206 (1998).
12. Hahn, H. *et al.* Mutations of the human homolog of *Drosophila* patched in the nevoid basal cell carcinoma syndrome. *Cell* **85,** 841–851 (1996).
13. Johnson, R. L. *et al.* Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. *Science* **272,** 1668–1671 (1996).
14. Pietsch, T. *et al.* Medulloblastomas of the desmoplastic variant carry mutations of the human homologue of *Drosophila patched. Cancer Res.* **57,** 2085–2088 (1997).
15. Raffel, C. *et al.* Sporadic medulloblastomas contain *PTCH* mutations. *Cancer Res.* **57,** 842–845 (1997).
16. Xie, J. *et al.* Mutations of the *PATCHED* gene in several types of sporadic extracutaneous tumors. *Cancer Res.* **57,** 2369–2372 (1997).
17. Wechsler-Reya, R. J. & Scott, M. P. Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron* **22,** 103–114 (1999).
18. Wetmore, C., Eberhart, D. E. & Curran, T. The normal patched allele is expressed in medulloblastomas from mice with heterozygous germ-line mutation of patched. *Cancer Res.* **60,** 2239–2246 (2000).
19. Giangaspero, F. *et al.* in *World Health Organization Histological Classification of Tumours of the Nervous System* (eds Kleihues, P. & Cavenee, W. K.) 129–137 (International Agency for Research on Cancer, Lyon, 2000).
20. Murone, M., Rosenthal, A. & deSauvage, F. J. Sonic hedgehog signaling by the patched-smoothened receptor complex. *Curr. Biol.* **28,** 76–84 (1999).
21. Hahn, H. *et al.* Patched target IGF2 is indispensable for the formation of medulloblastoma and rhabdomyosarcoma. *J. Biol. Chem.* **275,** 28341–28344 (2000).
22. Segal, R. A., Goumnerova, L. C., Kwon, Y. K., Stiles, C. D. & Pomeroy, S. L. Expression of the neurotrophin receptor *TRKC* is linked to a favorable outcome in medulloblastoma. *Proc. Natl Acad. Sci. USA* **91,** 12867–12871 (1994).
23. Kim, J. Y. H. *et al.* Activation of neurotrophin-3 receptor TRKC induces apoptosis in medulloblastomas. *Cancer Res.* **59,** 711–719 (1999).
24. MacDonald, T. J. *et al.* Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nature Genet.* **29,** 143–152 (2001).
25. Tamayo, P. *et al.* Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl Acad. Sci. USA* **96,** 2907–2912 (1999).
26. Dasarathy, V. B. (ed.) *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques* (IEEE Computer Society Press, Los Alamitos, California, 1991).
27. Slonim, D. K. *et al.* in *Proc. 4th Annu. Int. Conf. Computational Mol. Biol.* 263–272 (ACM Press, New York, 2000).
28. Mukherjee, S. *et al.* Support vector machine classification of microarray data. CBCL paper 182/AI memo 1676 (Massachusetts Institute of Technology, Cambridge, Massachusetts, 1999); also at http://www.ai.mit.edu/projects/cbcl/publications/ps/cancer.ps.
29. Brown, M. P. S. *et al.* Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl Acad. Sci. USA* **97,** 262–267 (2000).
30. Califano, *et al.* in *Proc. 8th Int. Conf. Intel. Syst. Mol. Biol.* (eds Bourne, P. *et al.*) 75–85 (AAAI Press, Menlo Park, CA, 2000).

....

A calcium sensor in the sodium channel modulates cardiac excitability

Hanno L. Tan*†, **Sabina Kupershmidt**‡, **Rong Zhang**‡§,
Svetlana Stepanovic*, **Dan M. Roden**‡§, **Arthur A. M. Wilde**†,
Mark E. Anderson‡§ & **Jeffrey R. Balser***‡

*Departments of * Anesthesiology, § Medicine, and ‡ Pharmacology,
Vanderbilt University School of Medicine, Nashville 37232, Tennessee, USA
† The Experimental and Molecular Cardiology Group, Academic Medical Center,
University of Amsterdam, 1105 AZ, Amsterdam, the Netherlands*

....

Sodium channels are principal molecular determinants responsible for myocardial conduction and maintenance of the cardiac rhythm. Calcium ions (Ca^{2+}) have a fundamental role in the coupling of cardiac myocyte excitation and contraction, yet mechanisms whereby intracellular Ca^{2+} may directly modulate Na channel function have yet to be identified. Here we show that calmodulin (CaM), a ubiquitous Ca^{2+}-sensing protein, binds to