LETTER

nature

doi:10.1038/nature10825

Clonal selection drives genetic divergence of metastatic medulloblastoma

Xiaochong Wu [1], Paul A. Northcott [1], Adrian Dubuc [1], Adam J. Dupuy [2], David J. H. Shih [1], Hendrik Witt [3], Sidney Croul [4], Eric Bouffet [5], Daniel W. Fults [6], Charles G. Eberhart [7], Livia Garzia [1], Timothy Van Meter [8], David Zagzag [9], Nada Jabado [10], Jeremy Schwartzentruber [11], Jacek Majewski [10], Todd E. Scheetz [2], Stefan M. Pfister [3], Andrey Korshunov [12], Xiao-Nan Li [13], Stephen W. Scherer [14], Yoon-Jae Cho [15], Keiko Akagi [16], Tobey J. MacDonald [17], Jan Koster [18], Martin G. McCabe [19], Aaron L. Sarver [20], V. Peter Collins [21], William A. Weiss [22], David A. Largaespada [20], Lara S. Collier [23] & Michael D. Taylor [24]

Medulloblastoma, the most common malignant paediatric brain tumour, arises in the cerebellum and disseminates through the cerebrospinal fluid in the leptomeningeal space to coat the brain and spinal cord[1]. Dissemination, a marker of poor prognosis, is found in up to 40% of children at diagnosis and in most children at the time of recurrence. Affected children therefore are treated with radiation to the entire developing brain and spinal cord, followed by high-dose chemotherapy, with the ensuing deleterious effects on the developing nervous system[2]. The mechanisms of dissemination through the cerebrospinal fluid are poorly studied, and medulloblastoma metastases have been assumed to be biologically similar to the primary tumour[3,4]. Here we show that in both mouse and human medulloblastoma, the metastases from an individual are extremely similar to each other but are divergent from the matched primary tumour. Clonal genetic events in the metastases can be demonstrated in a restricted subclone of the primary tumour, suggesting that only rare cells within the primary tumour have the ability to metastasize. Failure to account for the bicompartmental nature of metastatic medulloblastoma could be a major barrier to the development of effective targeted therapies.

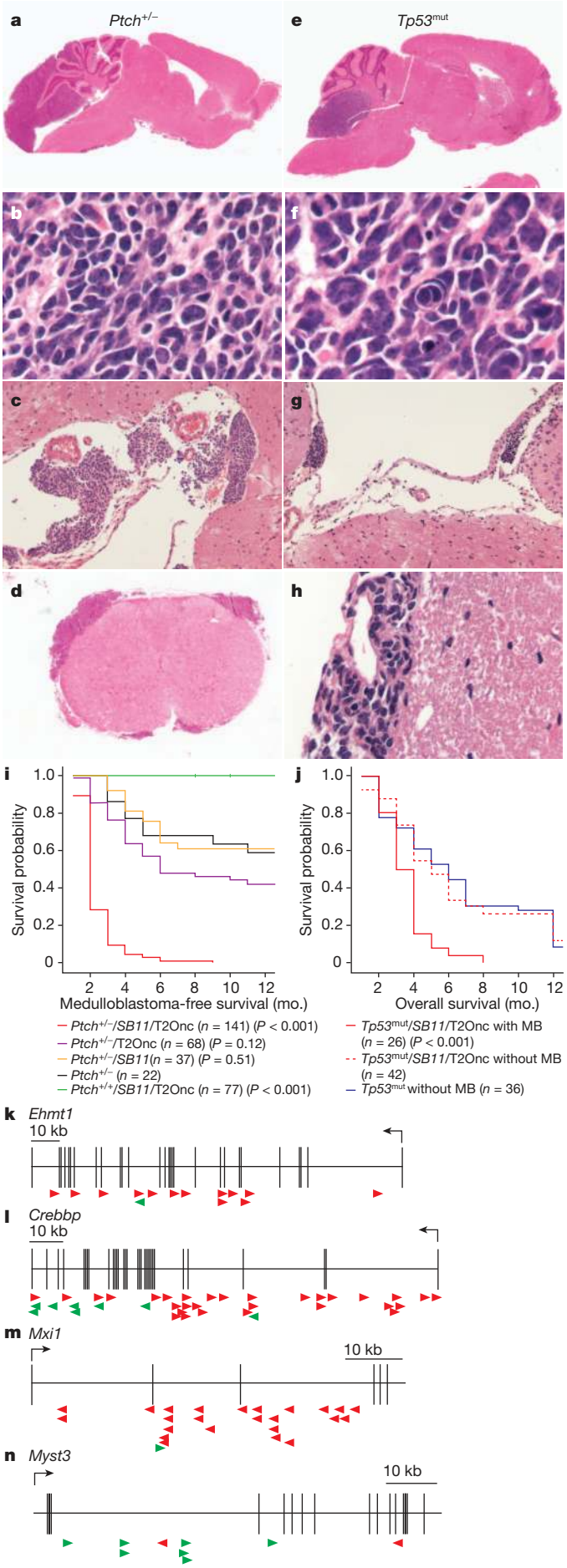
Thirty percent of patched-1-heterozygous ($Ptch^{1/2}$) mice develop non-disseminated medulloblastoma by 8 months of age[5]. Recently, the Sleeping Beauty (SB) transposon system was shown to be an effective tool for functional genomics studies of solid tumour initiation and progression[6,7]. We expressed the SB11 transposase in cerebellar progenitor cells in transgenic mice under the *Math1* (also known as *Atoh1*) enhancer/promoter, but we did not observe any tumours when these mice were bred with mice transgenic for a concatemer of the T2/Onc transposon[8] (Fig. 1a–j and Supplementary Figs 1 and 2). However, on a $Ptch^{1/2}$ background, these *Math1*-SB11/T2Onc mice showed increased penetrance of medulloblastoma (97%; 271 of 279 mice) compared with controls (39%; 54 of 139 mice), as well as decreased latency (2.5 months compared with 8 months) (Fig. 1 and Supplementary Fig. 2). Although $Ptch^{1/2}$ medulloblastomas are usually localized, the addition of SB transposition results in metastatic dissemination through the cerebrospinal fluid pathways, identical to the pattern that is seen in human children (Fisher's exact test, $P = 1.83 \times 10^{-27}$, odds ratio = 5.2; Supplementary Table 1) (Fig. 1c, d and Supplementary Fig. 2). As neither transposon nor transposase alone had an effect on tumour incidence, latency or dissemination, we conclude that SB-induced insertional mutagenesis drives medulloblastoma progression on the $Ptch^{1/2}$ background (Fig. 1i and Supplementary Fig. 2).

Humans with germline mutations in the tumour-suppressor gene *TP53* have Li–Fraumeni syndrome and have an increased risk of developing medulloblastoma. Although no medulloblastomas were found in mice with mutant *Tp53* (also known as *Trp53*) (denoted $Tp53^{mut}$ mice, which includes $Tp53^{1/2}$ and $Tp53^{2/2}$), 40% of $Tp53^{mut}$/*Math1*-SB11/T2Onc mice developed disseminated medulloblastoma[9] (Fig. 1e–h, j and Supplementary Fig. 2). Human medulloblastomas with *TP53* mutations frequently have large cell/anaplastic histology. $Tp53^{mut}$/*Math1*-SB11/T2Onc medulloblastomas have large cells, nuclear atypia and nuclear moulding that is typical of large cell/anaplastic histology (Fig. 1f). We conclude that SB transposition can drive the initiation and progression of metastatic medulloblastoma on a $Tp53^{mut}$ background.

We used linker-mediated PCR and 454 sequencing to identify the site of T2/Onc insertions in $Ptch^{1/2}$/*Math1*-SB11/T2Onc and $Tp53^{mut}$/*Math1*-SB11/T2Onc primary medulloblastomas and their matched metastases. Genes that contained insertions statistically more frequently than the background rate were identified as gene-centric common insertion sites (gCISs)[10]. We identified 359 gCISs in 139 primary tumours on the $Ptch^{1/2}$ background and 26 gCISs in 36 primary medulloblastomas on the $Tp53^{mut}$ background (Supplementary Tables 2–7 and Supplementary Figs 3–5). A large number of gCISs were candidate medulloblastoma oncogenes or tumour-suppressor genes[11] (Supplementary Table 8). Insertions in candidate tumour-suppressor genes, including *Ehmt1*, *Crebbp* and *Mxi1*, are predicted to cause a loss of function (Fig. 1k–m), whereas insertions in putative medulloblastoma oncogenes are largely gain of function, as exemplified by *Myst3* (Fig. 1n).

Many gCISs mapped to regions of amplification, focal hemizygous deletion and homozygous deletion (Supplementary Table 8) that we

[1] Arthur and Sonia Labatt Brain Tumour Research Center, and Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [2] Molecular & Cellular Biology Program, The University of Iowa, Iowa City, Iowa 52242, USA. [3] German Cancer Research Center (DKFZ), and Department of Pediatric Oncology, Hematology and Immunology, University Hospital Heidelberg, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [4] University Health Network Pathology, Arthur and Sonia Labatt Brain Tumour Research Centre, Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A1, Canada. [5] Brain Tumour Program, Haematology and Oncology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [6] Department of Neurosurgery, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA. [7] Johns Hopkins University, Baltimore, Maryland 21210, USA. [8] Virginia Commonwealth University, Richmond, Virginia 23284, USA. [9] Department of Pathology and Neurosurgery, Division of Neuropathology, New York University School of Medicine, New York, New York 10016, USA. [10] Departments of Human Genetics and Experimental Medicine, McGill University, Montreal H3Z 2Z3, Quebec, Canada. [11] McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 0G1, Canada. [12] German Cancer Research Institute (DKFZ) and Department of Neuropathology, University of Heidelberg, Im Neuenheimer Feld 220/221, 69120 Heidelberg, Germany. [13] Brain Tumor Program, Texas Children's Cancer Center, and Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA. [14] Program in Genetics and Genomic Biology and The Centre for Applied Genomics, The Hospital for Sick Children, and McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A1, Canada. [15] Departments of Neurology and Neurosurgery, Stanford University School of Medicine, Stanford, California 94305, USA. [16] Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210, USA. [17] Pediatric Neuro-Oncology Program, Emory University School of Medicine, Atlanta, Georgia 30307, USA. [18] Department of Human Genetics, Academic Medical Center, University of Amsterdam, 1100 DE Amsterdam, The Netherlands. [19] School of Cancer and Enabling Sciences, University of Manchester, Manchester M20 4BX, UK. [20] Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. [21] Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK. [22] Departments of Neurology, Pediatrics and Neurological Surgery, University of California, San Francisco, California 94143, USA. [23] School of Pharmacology, University of Wisconsin, Madison, Wisconsin 53715, USA. [24] Division of Neurosurgery, Arthur and Sonia Labatt Brain Tumour Research Center, and Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G 1X8, Canada.



recently reported in the genome of a large cohort of human medulloblastomas[11]. There is a high level of overlap between gCISs and known cancer genes (in the Catalogue of Somatic Mutations in Cancer (COSMIC) database) (Supplementary Tables 9 and 10), suggesting that many gCISs are bona fide driver genes in medulloblastoma (Fisher's exact test, $P = 0.0012$)[12]. Similarly, many of the mouse gCISs and the genes amplified in human medulloblastomas are overexpressed in human *SHH*-driven medulloblastomas (Supplementary Fig. 6). Conversely, mouse gCISs hemizygously deleted in human medulloblastomas were frequently expressed at a lower level in human medulloblastomas (Supplementary Fig. 6). The expression of six out of seven gCISs that had been studied by immunohistochemistry on a human medulloblastoma tissue microarray was associated with significantly worse overall and progression-free survival in human medulloblastoma[13] (Supplementary Table 11 and Supplementary Figs 7 and 8). We conclude that our SB-driven leptomeningeal-disseminated medulloblastoma model resembles the human disease anatomically, pathologically and genetically and thus is an accurate model of the human disease that can be used to identify candidate driver events and understand the pathogenesis of human medulloblastoma.

We compared the gCISs identified from *Ptch*$^{+/-}$/*Math1-SB11*/T2Onc and *Tp53*^{mut}/*Math1-SB11*/T2Onc primary medulloblastomas and matched metastases (Supplementary Table 2). Strikingly, the overlap between primary tumour gCISs (pri-gCISs) from *Ptch*$^{+/-}$/*Math1-SB11*/T2Onc tumours and those from metastases (met-gCISs) from the same animals was only 9.3% of all gCISs (Fig. 2a). Similarly, the overlap between pri-gCISs from *Tp53*^{mut}/*Math1-SB11*/T2Onc mice and the matching met-gCISs was only 8.9% (Fig. 2b). The leptomeningeal metastases and the matched primary tumour have identical, highly clonal insertion sites on both genetic backgrounds (Fig. 2c). The probability of two (or three) unrelated tumours having SB insertions in exactly the same TA dinucleotide is extremely low. We conclude that the leptomeningeal metastases and the matched primary tumour arise from a common transformed progenitor cell and have subsequently undergone genetic divergence.

Sequencing also identified insertions that are highly clonal in the metastases but are not observed in the matched primary tumour (data not shown). End-point PCR for these insertions in the matched primary and metastatic tumours shows that the insertion is highly clonal in the metastasis (or metastases) and is present in a very small subclone of the primary tumour (Fig. 2d and Supplementary Fig. 9). These data are consistent with a model in which metastatic disease arises from a minor restricted subclone of the primary tumour. Dissemination could occur repeatedly from the same subclone of the primary tumour, which seeds the rest of the central nervous system, or it could occur once, followed by reseeding of the rest of the leptomeningeal space by the initial metastasis. Insertions that are restricted to a minor subclone of the

Figure 1 | Transposon mutagenesis models of disseminated human medulloblastoma. a–d, The histology of transposon-driven medulloblastoma on the *Ptch*$^{+/-}$ background resembles human medulloblastoma, with leptomeningeal metastases on the surface of the brain (**c**) and spinal cord (**d**). Images show haematoxylin and eosin staining (**a**, entire brain; **b**, upper spinal cord). **e–h**, The histology of transposon-driven medulloblastoma on the *Tp53*^{mut} background shows histological features of large cell/anaplastic medulloblastoma, including nuclear pleomorphism and nuclear wrapping (**f**). Dissemination to the leptomeningeal spaces of the brain (**g**) and spinal cord (**h**) also occurs on this background. **i**, *Ptch*$^{+/-}$ mice with SB transposition develop more frequent medulloblastomas with a shorter latency than *Ptch*$^{+/-}$ mice without transposition. *P* values are from *t*-tests of survival comparing individual genotypes to *Ptch*$^{+/-}$ mice; *n*, number of mice per genotype. mo., months. **j**, Medulloblastoma (MB) was not observed in *Tp53*^{mut} mice without transposition but was observed in 42% of *Tp53*^{mut} mice with transposition. *P* values are from *t*-tests comparing survival between *Tp53*^{mut} mice and *Tp53*^{mut}/*SB11*/T2Onc mice with MB; *n*, number of mice. **k–n**, Insertion maps of notable gCISs. Insertions in the direction of transcription are denoted by green arrows, and those against the direction of transcription are denoted by red arrows. Transcription start sites are denoted by black arrows.

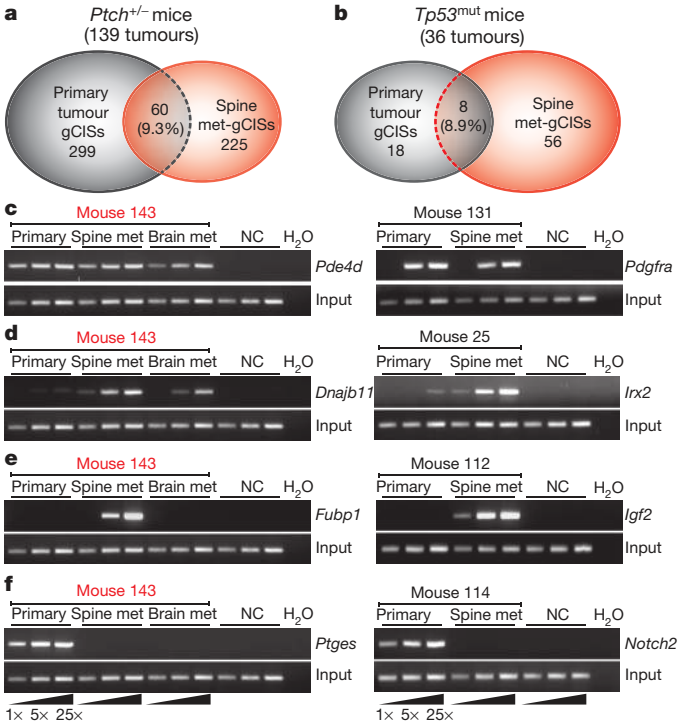


Figure 2 | Transposon-driven metastatic medulloblastoma genetically differs from the primary tumour. **a**, **b**, Venn diagrams depicting the degree of overlap and discordance in the gCISs in primary tumours and metastases, on the *Ptch*$^{+/-}$ and *Tp53*^{mut} backgrounds. **c**–**f**, Insertion-site end-point PCR was used to demonstrate the relative clonality of insertions between samples. Data for medulloblastoma in five mice are shown (mouse 143, left; and four mice, right). Three levels of input DNA were used for each sample (1×, 5× and 25×, with the increase depicted by a wedge). Shown are clonal events found in both the primary tumour and matching metastases (met) (**c**), insertions that are highly clonal in the metastases but very subclonal in the matching primary tumour (**d**), insertions that are highly clonal in the metastases but undetectable in the matching primary tumour (**e**), and insertions that are highly clonal in the primary tumour but undetectable in the matching metastases (**f**). NC, negative control; genomic DNA from a *Math1-SB11*/T2Onc double-transgenic mouse cerebellum.

primary tumour but that are clonal in the metastases could correspond to the previously described 'metastasis virulence' genes[14], which offer a genetic advantage during dissemination but not to the primary tumour. Another explanation for our data could be that the primary tumour was reseeded by a metastatic clone that had acquired additional genetic events in the periphery. This hypothesis is mitigated by the presence of highly clonal insertions in the metastasis that are completely absent from the primary tumour in the same animal[15] (Fig. 2e). As reseeding should be accompanied by contamination of the primary tumour with events found in the metastases, the absence of these events in the matched primary tumour makes reseeding much less likely (Fig. 2e). We propose that events found in only one metastasis represent progression events that are acquired post metastasis and that could lead to localized progression of metastatic disease, as is sometimes observed in human children.

We observed highly clonal insertions in the primary tumour, including in known medulloblastoma oncogenes such as *Notch2* and *Tert*, that were not found in the matching metastases (Fig. 2f). This pattern could be explained through remobilization of the SB transposon in the metastatic tumour; however, no signs of the DNA footprint remaining after SB remobilization at these loci were observed[16] (Supplementary Fig. 10). We suggest that these events, which may constitute driver events in the primary tumour, have arisen in the primary tumour after the metastases have disseminated (post-dispersion events). Although these known oncogenes are attractive targets for therapy, their utility as targets may be limited if the targets are not also found in the leptomeningeal compartment of the disease. Our data from two separate mouse lines support a model in which medulloblastoma disseminates early from a restricted subclone of the primary tumour and in which the primary tumour and the matched metastases then undergo differential clonal selection and evolution. Failure to account for the differences between the primary and leptomeningeal compartments could lead to the failure of targeted therapies. Failure to study the leptomeningeal disease (Fig. 2d, e) could result in systematically overlooking crucial targets for therapy in this compartment.

Examining the met-gCISs using Gene Set Enrichment Analysis (GSEA) demonstrated differences between the primary and metastatic disease, including enrichment for genes involved in the cytoskeleton in metastases (Supplementary Table 12). Targets that are present in both compartments and are maintenance genes, as exemplified by *Pdgfra*, will be optimal targets for treating both the primary tumour and the metastases (Fig. 2c and Supplementary Tables 7 and 9).

Pten, *Akt2*, *Igf2* and *Pik3r1* are all met-gCISs, implicating the phosphatidylinositol-3-OH kinase (PI(3)K) pathway in medulloblastoma progression. We injected the cerebellum of Nestin-TVA mice[17] with either an *Shh*-overexpressing retroviral vector (denoted *Shh* virus) or an *Shh*- and *Akt*-overexpressing retroviral vector (denoted *Shh* + *Akt* virus). Cerebellar injection of *Shh* virus alone resulted in medulloblastomas in 6 of 41 animals, compared with 20 of 42 animals injected with *Shh* + *Akt* virus ($P = 0.0018$). Although metastases were not observed with *Shh* virus alone (0 of 41), medulloblastoma metastases were observed in 9 of 42 animals injected with *Shh* + *Akt* virus ($P = 0.0024$) (Supplementary Fig. 11). *In vivo* modelling validates PI(3)K signalling as a putative contributor to leptomeningeal dissemination of medulloblastoma.

Previous publications and clinical approaches to human medulloblastoma have largely assumed that the primary tumour and its matched metastases are highly similar[3,4]. To test this assertion, we formally reviewed all cases of medulloblastoma from the past decade at The Hospital for Sick Children, in Ontario, Canada, and we identified 19 patients who had bulk residual primary tumour after surgery and metastases visible by magnetic resonance imaging, both of which could be followed for response to treatment (Supplementary Fig. 12 and Supplementary Table 13). Although it is possible that the metastases received less radiotherapy than the primary tumour in a subset of patients, in 58% of all cases (11 of 19) we observed a disparate response to therapy between the primary tumour and the matched metastases (binomial test, $P < 2.2 \times 10^{-16}$). Identification of definitive differences in the clinical response to standard therapy between the primary and the metastatic compartment awaits the completion of large, well-controlled, prospective clinical trials.

We examined seven matched primary and metastatic medulloblastomas for copy number aberrations (Fig. 3, Supplementary Figs 13 and 14, and Supplementary Tables 14 and 15). In each case, the primary tumour and the matched metastases shared complicated genetic events that provide strong support for their descent from a common transformed progenitor cell. Similar to our mouse data, in each case we observed clonal genetic events in the metastatic tumour(s) that were not present in the matched primary tumour (Fig. 3 and Supplementary Fig. 14). We also observed genetic events in the primary tumour that were absent from the matched metastases, consistent with a post-dispersion event (Fig. 3 and Supplementary Fig. 14). One patient with multiple leptomeningeal metastases had a deletion of chromosome 1p in only one of three examined metastases (Fig. 3a). This pattern of genetic events being present in only a subset of metastases could be a mechanism for the emergence of therapy-resistant metastatic clones.

We performed interphase fluorescence *in situ* hybridization (FISH) for the known medulloblastoma oncogenes *MYCN* and *MYC* on a collection of 17 paraffin-embedded primary and metastatic pairs of human medulloblastomas[18–20]. *MYCN* was amplified in three primary medulloblastomas but not in the matching metastases (Fig. 3b and

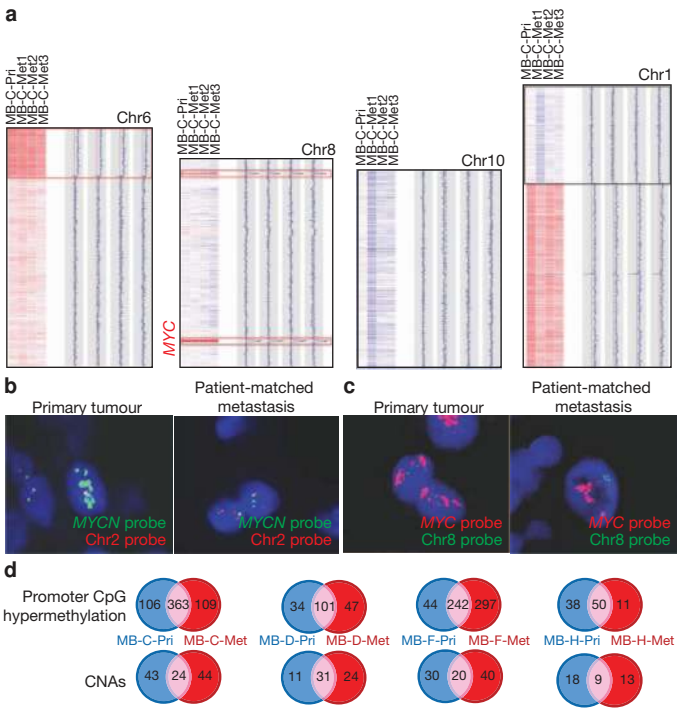


Figure 3 | Human medulloblastoma metastases are biologically distinct from their matched primary tumour. **a**, Copy number data from a primary medulloblastoma (MB-C-Pri) and three patient-matched metastases (MB-C-Met1, MB-C-Met2 and MB-C-Met3), with chromosomal regions in red representing genetic gain (amplification) and in blue denoting genetic loss (deletion). Examples of shared clonal events (red boxes) and events limited to one but not all metastases (black box) are shown. Chr, chromosome. **b**, Interphase FISH shows amplification of *MYCN* in a primary tumour but not the matched metastasis. Nuclei appear blue owing to 4′,6-diamidino-2-phenylindole (DAPI) staining. **c**, Interphase FISH for *MYC* demonstrates amplification in both the primary tumour and its matched metastases. **d**, Venn diagrams depicting the degree of overlap and discordance in promoter CpG methylation events and CNAs in primary medulloblastomas and their matched metastases, with MB-C, MB-D and MB-F and MB-H denoting different patients.

Supplementary Fig. 15). Conversely, *MYC* was amplified in two primary tumours and their matching metastases (Fig. 3c). These data are consistent with *MYCN* amplification being a post-dispersion event, similar to examples in SB-driven mouse medulloblastoma, and strongly indicate that anti-MYCN therapeutics may lack efficacy in the metastatic compartment of human medulloblastoma. The possibility that *MYCN* amplicons in the metastases have been 'lost' over time cannot be excluded.

We subsequently analysed promoter CpG methylation in these matched pairs and found much discordance between the primary tumour and matched metastases (Fig. 3d, Supplementary Figs 13 and 16, and Supplementary Tables 16 and 17). Finally, we performed whole-exome sequencing on a limited set of matched primary and metastatic medulloblastomas and found many single nucleotide variants (SNVs) that were restricted to a single compartment (Supplementary Fig. 13 and Supplementary Table 18). The discordance of CNAs, promoter CpG methylation events and SNVs between the primary tumour and its matched metastases supports a bicompartmental model for metastatic medulloblastoma. The mutational load in the human tumours (the combination of CNAs, CpG methylation and SNVs) compares favourably with the mutational load in our transposon-driven mouse models (in which the median number of gCISs is 25 per tumour; Supplementary Table 19). Validation of the individual CNAs that were restricted to the metastases showed that these CNAs can be detected in a very minor subclone of the primary tumour, in keeping with the relationship identified in the mouse model (Supplementary Fig. 17 and Supplementary Tables 20 and 21). Pathway analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) to compare mouse gCISs with the genes that were affected in the human metastases identified only one statistically significant shared signalling pathway: insulin signalling ($P = 0.027$) (Supplementary Table 22). The known role of insulin receptor signalling in primary medulloblastoma[21], together with the data presented here on the role of AKT in metastatic medulloblastoma, suggests that insulin signalling should be prioritized as a therapeutic target to be tested in clinical trials.

We performed unsupervised hierarchical clustering on the CpG methylation data, and we found that normal cerebellar controls cluster away from the medulloblastomas, whereas metastases cluster with their matching primary tumour (Fig. 4a). However, metastases cluster

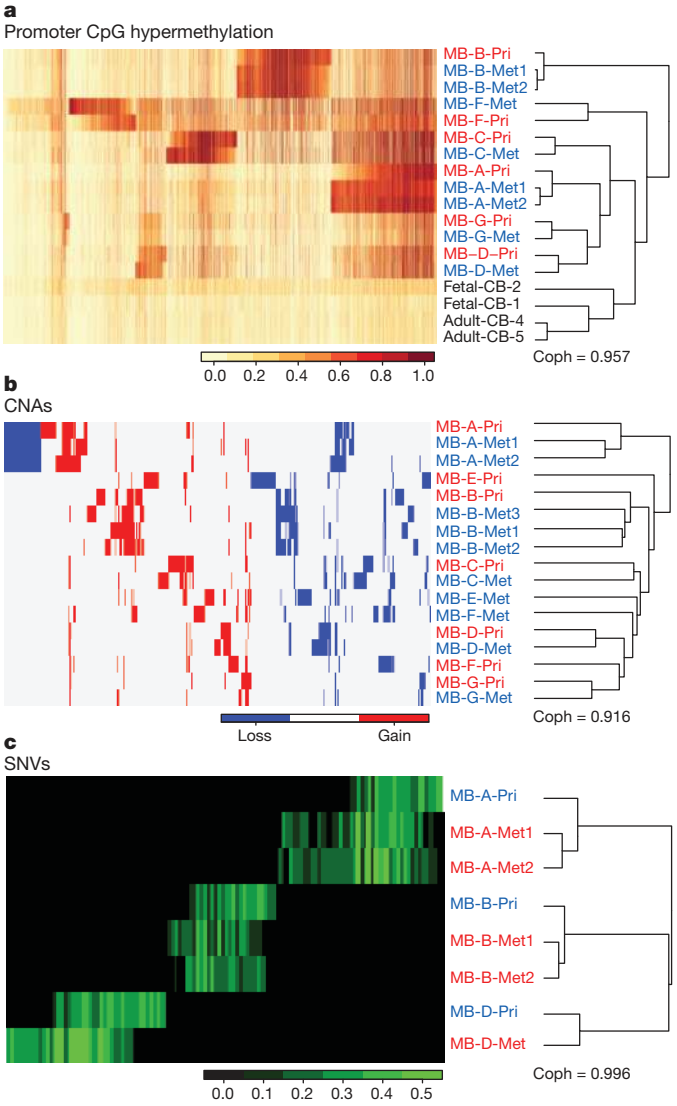


Figure 4 | Human medulloblastoma metastases are genetically distinct from their matched primary tumour. **a**, Profiling the methylation status of 27,578 CpG dinucleotide sites in the human genome in a collection of human matched primary and metastatic medulloblastomas; the top 2,000 genes are shown. Unsupervised hierarchical clustering by CpG methylation pattern demonstrates that patient-matched metastases are more similar to each other than to the matched primary tumour. **b**, Unsupervised clustering of regions of copy number gain and loss demonstrates that patient-matched metastases are more similar to each other than to the matched primary tumour. **c**, Unsupervised hierarchical clustering of SNV data from whole-exome sequencing demonstrates that patient-matched metastases are more similar to each other than to the matched primary tumour. SNVs that are found only in the primary compartment or only in both examined tumours in the metastatic compartment are evident. Coph, cophenetic correlation coefficient.

closer to each other than they do to the matched primary tumour (z-test, $P = 0.0014$) (Supplementary Fig. 18). Unsupervised hierarchical clustering of CNA and exome SNV data uncovered the same relationships (Fig. 4b, c). Evident within the exome data are many events that are shared only by patient-matched metastases (that is, metastases from a single patient), as well as events that are restricted to the primary tumour, both of which are similar to the genetic patterns observed in mice. These three data sets support a model in which patient-matched human medulloblastoma metastases are epigenetically and genetically very similar to each other but have substantially diverged from the primary tumour, resulting in two different disease compartments: the primary and metastatic compartments.

Our data from two mouse models, with support from initial data from human medulloblastoma, suggest that leptomeningeal metastases of medulloblastoma from a single human or mouse are genetically similar to each other but are highly divergent from the matched primary tumour, consistent with a bicompartmental model of disease. Our results are consistent with a model in which metastases arise from a restricted subclone of the primary tumour through a process of clonal selection in both humans and mice. That metastases might arise from a pre-existing minor subclone of the primary tumour through clonal selection was suggested more than three decades ago, but it remains a controversial hypothesis that might not be true of all cancers[22–25]. Failure to account for the divergent molecular pathology of the metastatic compartment may result in selection of therapeutic targets present in the primary tumour, which is more amenable to surgical control, but not the metastases, which are the more frequent cause of death.

METHODS SUMMARY

Generation of *Math1-SB11* construct. *SB11* cDNA was excised from the vector pCMV-*SB11* and ligated into the vector J2Q-*Math1* (refs 8, 26).

Linker-mediated PCR and 454 deep sequencing. Bar-coded, linker-mediated PCR was performed as previously described[6]. Sample preparation for the 454 sequencing and the subsequent procedures was performed as previously described[27].

Determination of gCISs. A chi-squared analysis was performed to determine whether the number of observed integration events within each transcription unit in the SB-driven medulloblastomas was significantly greater than expected given the following: the number of TA dinucleotide sites within the gene relative to the number of TA sites in the genome, the number of integration sites within each tumour, and the total number of tumours in each cohort. This gCIS analysis produced a P value for each of the ~19,000 mouse RefSeq genes, and Bonferroni correction was therefore used to adjust for multiple hypothesis testing.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 October 2010; accepted 3 January 2012.
Published online 15 February 2012.

1. Gajjar, A. *et al.* Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncol.* **7,** 813–820 (2006).
2. Mabbott, D. J. *et al.* Serial evaluation of academic and behavioral outcome after treatment with cranial radiation in childhood. *J. Clin. Oncol.* **23,** 2256–2263 (2005).
3. 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).
4. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genet.* **33,** 49–54 (2003).
5. Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277,** 1109–1113 (1997).
6. Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J. & Largaespada, D. A. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436,** 272–276 (2005).
7. Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic *Sleeping Beauty* transposon system. *Nature* **436,** 221–226 (2005).
8. Helms, A. W., Abney, A. L., Ben-Arie, N., Zoghbi, H. Y. & Johnson, J. E. Autoregulation and multiple enhancers control *Math1* expression in the developing nervous system. *Development* **127,** 1185–1196 (2000).
9. Wetmore, C., Eberhart, D. E. & Curran, T. Loss of *p53* but not *ARF* accelerates medulloblastoma in mice heterozygous for *patched*. *Cancer Res.* **61,** 513–516 (2001).
10. Brett, B. T. *et al.* Novel molecular and computational methods improve the accuracy of insertion site analysis in *Sleeping Beauty*-induced tumors. *PLoS ONE* **6,** e24668 (2011).
11. Northcott, P. A. *et al.* Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nature Genet.* **41,** 465–472 (2009).
12. Forbes, S. A. *et al.* in *Current Protocols in Human Genetics* Ch. 10.11 doi:10.1002/0471142905.hg1011s57 (2008).
13. Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29,** 1408–1414 (2011).
14. Nguyen, D. X. & Massague, J. Genetic determinants of cancer metastasis. *Nature Rev. Genet.* **8,** 341–352 (2007).
15. Norton, L. & Massague, J. Is cancer a disease of self-seeding? *Nature Med.* **12,** 875–878 (2006).
16. Luo, G., Ivics, Z., Izsvák, Z. & Bradley, A. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **95,** 10769–10773 (1998).
17. Rao, G., Pedone, C. A., Coffin, C. M., Holland, E. C. & Fults, D. W. c-Myc enhances sonic hedgehog-induced medulloblastoma formation from nestin-expressing neural progenitors in mice. *Neoplasia* **5,** 198–204 (2003).
18. Swartling, F. J. *et al.* Pleiotropic role for *MYCN* in medulloblastoma. *Genes Dev.* **24,** 1059–1072 (2010).
19. Pfister, S. *et al.* Outcome prediction in pediatric medulloblastoma based on DNA copy-number aberrations of chromosomes 6q and 17q and the *MYC* and *MYCN* loci. *J. Clin. Oncol.* **27,** 1627–1636 (2009).
20. Korshunov, A. *et al.* Accumulation of genomic aberrations during clinical progression of medulloblastoma. *Acta Neuropathol.* **116,** 383–390 (2008).
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. Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197,** 893–895 (1977).
23. Scheel, C., Onder, T., Karnoub, A. & Weinberg, R. A. Adaptation versus selection: the origins of metastatic behavior. *Cancer Res.* **67,** 11476–11480 (2007).
24. Talmadge, J. E. Clonal selection of metastasis within the life history of a tumor. *Cancer Res.* **67,** 11471–11475 (2007).
25. Talmadge, J. E. & Fidler, I. J. The biology of cancer metastasis: historical perspective. *Cancer Res.* **70,** 5649–5669 (2010).
26. Geurts, A. M. *et al.* Gene transfer into genomes of human cells by the *Sleeping Beauty* transposon system. *Mol. Ther.* **8,** 108–117 (2003).
27. Starr, T. K. *et al.* A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323,** 1747–1750 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M.D.T. holds a Canadian Institutes of Health Research Clinician-Scientist Phase II Award, was a Sontag Foundation Distinguished Scholar, and is supported by grants from the National Institutes of Health (R01CA148699), the Pediatric Brain Tumor Foundation, the Canadian Cancer Society, and Brainchild. X.W. was supported by a fellowship from the American Brain Tumor Association in tribute to Tracy Greenwood. L.G. was supported by a fellowship from the Davis M. Ferguson Fund from the American Brain Tumor Association. A.D. was supported by a Vanier Doctoral Fellowship from the Canadian Institutes of Health Research. L.S.C. was supported by a grant (K01CA122183) and a Kimmel Scholar award from the Kimmel Foundation. C.E. was supported by a grant from the National Institutes of Health (NS055089). We thank S. Archer for technical writing assistance.

Author Contributions M.D.T., X.W., P.A.N., L.S.C., A.D. and D.L. conceived the research and planned the experiments. X.W., P.A.N., A.D., L.G. and D.J.H.S. carried out the vast majority of the experiments under M.D.T.'s guidance. C.E., T.V.M., D.Z., Y.-J.C., T.M., X.-N.L., V.P.C., M.G.M. and W.A.W. provided human tumour materials. All authors contributed experimental expertise and participated in the writing of the manuscript. A.J.D., D.J.H.S., T.E.S., S.W.S., K.A., J.K., A.L.S., D.L. and L.S.C. provided biostatistical and bioinformatic expertise. E.B. provided the clinical data and analysis. D.W.F. carried out the *Akt* experiments. N.J., J.S. and J.M. carried out the exome sequencing. H.W., S.M.P. and A.K. carried out the immunostaining of human medulloblastoma tissue microarrays and FISH for *MYCN* and *MYC*. S.C. carried out the pathological analysis of mouse brain tumours.

Author Information CpG methylation data have been deposited in the Gene Expression Omnibus under accession number GSE34356. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.D.T. (mdtaylor@sickkids.ca).

METHODS

Linker-mediated PCR and 454 deep sequencing. Genomic DNA was isolated and purified from mouse tissues with a DNeasy Blood & Tissue Kit (QIAGEN). The subsequent bar-coded, linker-mediated PCR was performed as previously described[6]. Sample preparation for the 454 sequencing and the subsequent procedures was performed as previously described[27].

PCR for SB-tagged fragments. The primers for amplifying SB-transposon insertion sites were designed based on the chromosomal location of each independent insertion site and its orientation to transcription. The primers at the inverted repeats/direct repeats (left) (IRDRL) and inverted repeats/direct repeats (right) (IRDRR) of the transposon were 5′-CTGGGAATGTGATGAAAGAAATAAAA-3′ and 5′-TTGTGTCATGCACAAAGTAGATGT-3′, respectively. The input represents genomic DNA with SB transposition, which was illustrated by SB excision PCR that detected the transposon post transposition[6]. Three points of input (1×, 5× and 25×) were used. The following primers were used: *Pde4d*-143L, 5′-CACATAAAAACTGGACACCTAG-3′; *Pdgfra*-131R, 5′-CTATCATGACCACACGGAAGAGAGTGAAC-3′; *Dnajb11*-143L, 5′-CATGAGCTATGGCACAGATAC-3′; *Fubp1*-143R, 5′-CACTAGTGCCCATGGATTAGG-3′; *Ptges*-143R, 5′-CAGAACTGATAGAGGCCAAAG-3′; *Irx2*-25L, 5′-CAACACTTTCAGACACACATATATC-3′; *Igf2*-112R, 5′-GTGACCAGTGTGTATTCGTGGAATTTTTTGGG-3′; and *Notch2*-114R, 5′-CAGTGTCCAGGCAGTCATTTCAAAGAGTG-3′. Details about the primer design for specific insertion sites and the PCR protocol are available on request.

Review of clinical cases. We systematically reviewed all cases of medulloblastoma seen at The Hospital for Sick Children (Toronto, Ontario) over the past ten years. Cases that have both metastases and post-operative residual bulky disease at the primary site were identified on the basis of post-operative imaging obtained within 72 h of surgery. All radiology results were reviewed by a senior neuro-oncologist (E.B.). Objective responses of both the primary tumour and the metastatic disease were measured using the standard International Society of Paediatric Oncology (SIOP) criteria for clinical trials of paediatric brain tumours[28].

End-point PCR on human samples. For PCRs to confirm the deletion of the *CDKN2A* locus on chromosome 9, a genome-walking approach (GenomeWalker Universal Kit, Clontech, Catalogue number 638904) was taken to locate the specific deletion region based on single nucleotide polymorphism (SNP) coordinates. The following primers flanking the deletion region were used: forward, 5′-GCAATTAACCAAGACCACCCAATGGCAAG-3′; and reverse, 5′-GTAGCTATTGGGGAGGTTGAGAAGGAG-3′. Three points of input shown as *ACTB* (1×, 5× and 25×) were used. The PCR products were inserted into the pCR2.1 TA cloning vector (Invitrogen), sequenced and searched against the human genome in the blast database to confirm the deletion. For *REXO1L1* deletion on chromosome 8, specific primers flanking the deletion region were designed based on SNP microarray results. The PCR products were TA-cloned and sequenced as described above. The following primers were used: forward, 5′-GGCTGACTCCCTTCTGATATAG-3′; and reverse, 5′-CAATCACTTACAGTTACTAGGCAC-3′. Details about the primer design and PCR protocols are available on request.

Chromosomal mapping of gCISs. Chromosomal maps of gCIS-associated genes were obtained from the UCSC Mouse Genome Browser (assembly in July 2007). Each insertion site of a specific CIS was mapped to the gene with the same orientation as the direction of transcription (arrow in green) or the inverse orientation to the direction of transcription (arrow in red).

Human medulloblastoma tumour specimens. All tumour specimens were obtained in accordance with the Research Ethics Board at The Hospital for Sick Children. Surgically resected, fresh frozen samples were obtained from the Cooperative Human Tissue Network and the Brain Tumor Tissue Bank.

SB remobilization. Potential SB insertion sites at *Fubp1*, *Mnat1* or *Igf2* in primary tumours from mouse numbers 143, 14 or 11 or sites at *Ptges*, *Aof1* and *Notch2* in the matched spine metastases were tested for remobilization. The primers were designed to amplify each insertion site to produce approximately 300 base pairs (bp) with the insertion site in the middle. PCR products were either sequenced directly or after being TA-cloned. The resultant sequences were examined for 'scars' from potential remobilization. As positive controls for the scars, primers were used to amplify the T2/Onc transposon in each sample[8]. The products were sequenced and examined for the scars as described above. The following primers were used: *Aof1* forward (Fw), 5′-TACTCCAGACAGTCAGTCAGTG-3′; *Aof1* reverse (Rv), 5′-TAGTTCTGCCTCATGCCACAAG-3′; *Ptges* Fw, 5′-ACAGAGAAGGCTTCAGAGCTC-3′; *Ptges* Rv, 5′-GGTGCTCTCTGCTGTCCAATC-3′; *Notch2* Fw, 5′-CAAGCTTTCAAGTATAAACCACGC-3′; *Notch2* Rv, 5′-GAATGCATCATCCAGTGTCCAG-3′; *Fubp1* Fw, 5′-AGGAACGGGCTGGTGTTAAAATG-3′; *Fubp1* Rv, 5′-TCTAATACCATTTCCTTGGCTTGC-3′; *Mnat1* Fw, 5′-CTAACACATCAGAGTTGGACAAG-3′; *Mnat1* Rv, 5′-CATGAAGACCTGAGAGTGCAG-3′; *Igf2* Fw, 5′-GTGATTGGTGAATGTACTCTTTCC-3′; and *Igf2* Rv, 5′-GTGGAACACTAGATTCTGTAGTC-3′. Details about the primer design and PCR protocols are available on request.

Hierarchical clustering. Agglomerative hierarchical clustering analyses were performed in the R statistical programming environment (version 2.13). The average linkage method was used in all cases. Because different data types were used in the various analyses, the metric used for clustering differed between the analyses. The Manhattan distance metric was used for the copy number data because the data were encoded as {−1, 0, 1}. The magnitudes of the CNAs were not considered, owing to a multitude of confounding factors, including tumour heterogeneity and ploidy. The Kendall rank correlation was used for the SNV frequency data because the data distributions were not normal. The Pearson correlation was used for the methylation data, which were normally distributed.

Identification of CpG hypermethylation events. Human genomic DNA was isolated from matching primary and metastatic medulloblastomas obtained from Johns Hopkins University, the Virginia Commonwealth University and New York University. An EZ DNA Methylation Kit (Zymogen Research) was used to bisulphite convert 500ng each sample. The recovered DNA was profiled on HumanMethylation27 BeadChips (Illumina) at The Centre for Applied Genomics (TCAG). Subsequently 27,578 CpG dinucleotides spanning 14,495 genes were analysed. The probe signal intensity was corrected by using BeadStudio 3.2.0 software (Illumina). The background normalization and differential methylation analyses were performed against fetal cerebella using the custom error model (Illumina). Cancer-specific DNA hypermethylation events were defined as those with a 30% increase in methylation in at least one medulloblastoma sample relative to an average methylation level (less than 50%) in normal fetal and adult cerebellum samples. Unsupervised clustering using Euclidian hierarchical clustering metrics was then performed on 2,503 data points that were filtered for cancer-specific hypermethylation events. The CpG methylation data are available from the Gene Expression Omnibus under accession number GSE34356.

Bisulphite sequencing of CpG promoter methylation. Representative examples of primary-tumour- and metastasis-specific methylation events were identified from normalized Illumina Hg27 data. Bisulphite PCR (BSP) primers were designed using the EpiDesigner tool (SEQUENOM) (http://www.epidesigner.com/) to encompass a genomic region flanking the Illumina Hg27 gene-specific probe. DNA (500ng) from the primary tumour and the corresponding metastases was bisulphite converted using an EZ DNA Methylation Kit. Following PCR optimization, 10ng bisulphite-converted DNA was used to amplify the genomic regions of interest. Amplicons were subcloned into the pCR2.1-TOPO vector (Invitrogen), and plasmid DNA from 10–12 colonies was extracted using a PureLink Quick Plasmid Miniprep Kit (Invitrogen). Sequencing was performed at TCAG using the M13 reverse primer, 5′-CAGGAAACAGCTATGAC-3′. The following primers were also used: *MLH1* Fw, 5′-TTGTTGGAATGTTATTTATTATTTAGGA; *MLH1* Rv, 5′-CATAATATCCACCAAAAAACCAAAA-3′; *MRPS21* Fw, 5′-TTTTTGGTTTTTGTTGATTGTTTTT-3′; *MRPS21* Rv, 5′-CAAATCTCAAAAAATCTATCCTTTCC-3′; *RBP1* Fw, 5′-GTAGGGGAGGTATAGGTAGGTTGTG-3′; *RBP1* Rv, 5′-CTTAATCAAACCCCCTAAACAAAAA-3′; *WNK2* Fw, 5′-GTGTTTTTGGTTTATAGAGATGGA-3′; and *WNK2* Rv, 5′-ACTCCTCCTAATCCRACTCTAC-3′. Details about the primer design and PCR protocols are available on request.

Alignment and variant calling for whole-exome sequencing. Standard manufacturers' protocols were used to perform target capture with a TruSeq Exome Enrichment Kit (Illumina) and sequencing of 100-bp paired-end reads on a HiSeq sequencing system (Illumina). Approximately 10 gigabases of sequence was generated for each subject such that >90% of the coding bases of the exome defined by the Consensus CDS (CCDS) project were covered by at least ten reads. Adaptor sequences and quality trimmed reads were removed by using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and then a custom script was used to ensure that only read pairs with both mates present were subsequently used. Reads were aligned to Hg19 with BWA1, and duplicate reads were marked using Picard (http://picard.sourceforge.net/) and excluded from downstream analyses. SNVs and short insertions and deletions (indels) were called using SAMtools (http://samtools.sourceforge.net/) Pileup and varFilter2 with the base alignment quality (BAQ) adjustment disabled and were quality filtered to require at least 20% of reads supporting the variant call. Variants were annotated using both ANNOVAR3 and custom scripts to identify whether they affected protein coding sequence and whether they had previously been seen in dbSNP131, the 1,000 Genomes pilot release (November 2010) or in approximately 160 exomes that had previously been sequenced at our centre.

SNV analysis of whole-exome sequencing data. For clustering analysis, an SNV frequency matrix was constructed by calculating frequencies from the read counts of the reference and the alternative nucleotide. The matrix was not standardized (that is, converted to *z* scores) before clustering, because the absolute SNV frequencies were of interest.

For Venn analysis, the samples were grouped into primary–metastasis sets, and the filtered SNVs were used to identify SNVs that are enriched in one sample compared with all other samples of the same set, as determined by the hypergeometric test (P value threshold = 0.05). For sets consisting of three or more samples (A, B and C), an SNV was considered to be enriched in samples A and B if the SNV was enriched in A compared with C alone and also enriched in B compared with C alone. SNVs that were not enriched in any sample or subset of samples were considered to be common SNVs. Many of these common SNVs probably represented germline SNVs specific to the patient.

Analysis of CpG promoter methylation data. The similarities between the patient-matched metastatic and primary tumour samples and among patient-matched metastatic tumour samples were determined by using Pearson correlation analysis. As Pearson's r values are not normally distributed, they were standardized by Fisher's z transformation. Subsequently, the correlations between the metastatic samples and the matched primary tumour samples were compared with the correlations among the patient-matched metastatic samples, using the paired heteroscedastic Student's t-test.

Clustering analysis was performed as described above. The methylation matrix was not standardized before clustering, as doing so would entail discarding crucial information on the differences in the overall methylation profiles among samples or the average methylation among CpG promoters.

The stability of the CpG hypermethylation profile clusters was assessed using three methods. First, the clustering analysis was run for different numbers of CpG hypermethylation sites that varied most widely among samples. The partitions generated by each clustering run were compared with the reference partitions generated by the original clustering based on the 1,000 most variable hypermethylated CpG islands using the Jaccard similarity index. The same analysis was applied to a set of 100 background hypermethylation data matrices in which the sites are permuted independently in each sample. Second, the clustering analysis was performed for random subsamples of 1,000 sites, for 1,000 repeat runs. In each run, the resultant cluster was compared with the original cluster using the Jaccard index. Analysis on the original data matrix was compared with a set of 100 background matrices, permuted as described above. Third, the cluster stability was further assessed by bootstrap resampling of the samples using the pvclust R package (version 1.2).

28. Gnekow, A. K. Recommendations of the brain tumor subcommittee for the reporting of trials. *Med. Pediatr. Oncol.* **24,** 104–108 (1995).