The whole-genome landscape of medulloblastoma subtypes

Paul A. Northcott1–4, Ivo Buchhalter5–7,8, A. Sorana Morrissy5,9, Volker Hovestadt14, Joachim Weischenfeldt15, Tobias Ehrenberger16, Susanne Gröbner16, Maia Segura-Wang17, Thomas Zichner14, Vasilisa A. Rudneva1,2, Hans-Jörg Warnatz12, Nikos Sidiropoulos9, Aaron H. Phillips11, Steven Schumacher14, Kortine Kleinheinz1, Sebastian M. Waszak11, Serap Erkek1,18, David T. W. Jones1,10, Barbara C. Worst1,10, Marcel Kool1,13, Marc Zapata1, Natalie Jäger1, Lukas Chavez1–10, Barbara Hutter4, Matthias Bieg1,3,19, Nagarajan Paramasivam1,16, Michael Heindel5,20, Zuguang Gu1,3,15, Naveed Ishaque15,19, Christina Jäger–Schmidt14, Charles D. Imbusch4, Alke Jugold1, Daniel Hübschmann1,5,17, Thomas Risch12, Vyacheslav Amstislavskiy12, Francisco German Rodriguez Gonzalez2, Ursula D. Weber2, Stephan Wolf18, Giles W. Robinson19, Xin Zhou20, Gang Wu20, David Finkelstein20, Yanliu Liu20, Florence M. G. Cavallia8, Betty Luu7, Vijay Ramaswamy6, Xiaochong Wu4, Jan Koster21, Mariza Ryzhova2, Yoon-Jae Cho23, Scott L. Pomeroy24, Christel Herold-Mende25, Martin Schumann26, Martin Ebinge27, Linda M. Liu28, Maume Mora29, Roger E. McLendon30, Nada Jabado31, Yoshihiro Kume32, Eric Chuah33, Yussanne Ma33, Richard A. Moore33, Andrew J. Mungall33, Karen L. Mungall33, Tina Wong33, Steven J. M. Jones33, Olaf Witt17, Till Milder12, Andreas Von Deimling34, David Capper34, Andrey Korshunov34, Marie-Laure Yapo34, Richard Kriwacki13, Amar Gajjar19, Jinghui Zhang20, Rameen Beroukhim14, Ernest Fraenkel9, Jan O. Korbel31, Benedikt Brors3,4,10, Matthias Schlesner3, Roland Eils3,5,10, Marco A. Marra3,6, Stefan M. Pfister1,10,17, Michael D. Taylor6,5,35, and Peter Lichter7,10.

Current therapies for medulloblastoma, a highly malignant childhood brain tumour, impose debilitating effects on the developing child, and highlight the need for molecularly targeted treatments with reduced toxicity. Previous studies have been unable to identify the full spectrum of driver genes and molecular processes that operate in medulloblastoma subgroups. Here we analyse the somatic landscape across 491 sequenced medulloblastoma samples and the molecular heterogeneity among 1,256 epigenetically analysed cases, and identify subgroup-specific driver alterations that include previously undiscovered actionable targets. Driver mutations were confidently assigned to most patients belonging to Group 3 and Group 4 medulloblastoma subgroups, greatly enhancing previous knowledge. New molecular subtypes were differentially enriched for specific driver events, including hotspot in-frame insertions that target PRDM6. Thus, the application of integrative genomics to an extensive cohort of clinical samples derived from a single childhood cancer entity revealed a series of cancer genes and biologically relevant subtypes that represent attractive therapeutic targets for the treatment of patients with medulloblastoma.

Next-generation sequencing (NGS) studies have tremendously advanced our understanding of the genes, pathways and molecular processes that underly most commonly diagnosed human cancers. These efforts have identified core sets of ‘driver’ genes that are frequently mutated across a wide spectrum of different cancer entities1–2. Although the genetic underpinnings of some cancers were largely resolved during the first ‘wave’ of NGS studies, especially for comparatively simple malignancies driven by deregulation of a single pathway3,4, others remain enigmatic and require further interrogation with sufficient power to overcome confounding molecular heterogeneity and diversity.

Medulloblastoma (MB) (World Health Organization grade IV) is a highly malignant childhood brain tumour that has been the subject of several NGS studies conducted by the International Cancer Genome Consortium (ICGC)5–8, the Paediatric Cancer Genome Project (PCGP)9, and the Medulloblastoma Advanced Genomics Consortium (MAGIC)10,11. Consensus molecular subgroups of MB, namely WNT, SHH, Group 3 and Group 4, exhibit distinctive transcriptional and

1Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 2Department of Developmental Neurobiology, St Jude Children’s Research Hospital, Memphis, Tennessee, USA. 3Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 4Division of Applied Bioinformatics, German Cancer Research Center, Children’s Research Hospital, Memphis, Tennessee, USA. 5Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 6Developmental Tumor Biology Laboratory, Hospital Sant Joan de Déu, Barcelona, Spain. 7Division of Cancer and Tumor Biology, King’s College London, London, United Kingdom. 8Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. 9Division of Neurosurgery, Hospital for Sick Children, Toronto, Ontario, Canada. 10Department of Oncology, St Jude Children’s Research Hospital, Memphis, Tennessee, USA. 11Department of Pathology, Brigham and Women’s Hospital, Boston, Massachusetts, USA. 12Department of Neurology, Boston Children’s Hospital and Harvard Medical School, Boston, Massachusetts, USA. 13Department of Neurosurgery, University of California, Los Angeles, California, USA. 14Division of Neurosurgery, University of British Columbia, Vancouver, British Columbia, Canada. 15Division of Pathology, Duke University, Durham, North Carolina, USA. 16Department of Pediatrics, McGill University, Montreal, Quebec, Canada. 17Department of Oncology, Kitasato University School of Medicine, Sagamihara, Japan. 18Division of Neurosurgery, Harvard Medical School, Boston, Massachusetts, USA. 19Department of Pediatrics, St Jude Children’s Research Hospital, Memphis, Tennessee, USA. 20Department of Neurosurgery, University Hospital Tübingen, Tübingen, Germany. 21Department of Hematology and Oncology, Children’s University Hospital Tübingen, Tübingen, Germany. 22Department of Neurosurgery, David Geffen School of Medicine at UCLA, Los Angeles, California, USA. 23Developmental Tumor Biology Laboratory, Hospital Sant Joan de Déu, Barcelona, Spain. 24Department of Pathology, Duke University, Durham, North Carolina, USA. 25Department of Pediatrics, McGill University, Montreal, Quebec, Canada. 26Department of Neurosurgery, Kitasato University School of Medicine, Sagamihara, Japan. 27Department of Medicine, Stanford University, Stanford, California, USA. 28Department of Pediatrics, St Jude Children’s Research Hospital, Memphis, Tennessee, USA. 29Division of Neurology, Heidelberg University Hospital, Heidelberg, Germany. 30Department of Pathology, Heidelberg University Hospital, Heidelberg, Germany. 31Department of Pediatrics, McGill University, Montreal, Quebec, Canada. 32Department of Neurosurgery, Kitasato University School of Medicine, Sagamihara, Japan. 33Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada. 34Department of Oncology, St Jude Children’s Research Hospital, Memphis, Tennessee, USA. 35Division of Neurosurgery, Hospital for Sick Children, Toronto, Ontario, Canada.

*These authors contributed equally to this work.
§These authors jointly supervised this work.
epigenetic signatures that define clinically relevant patient subsets. WNT and SHH subgroup MBs are primarily driven by mutations leading to constitutive activation of the Wingless and Sonic hedgehog signalling pathways, respectively. By contrast, the genetics and biology underlying the Group 3 and Group 4 MB subgroups remain less clear. Targeted therapies for MB are scarce yet desperately needed, warranting intensive investigation into the full range of genetic lesions and molecular heterogeneity that contribute to MB subgroups, especially as it relates to poorly characterized Group 3 and Group 4 disease. Here we report the genomic landscape across a series of 491 previously untreated MBs. Our comprehensive and integrative approach to this multilayered dataset provides considerable biological insight into each of the core subgroups, including the identification of new subgroup-specific driver genes, epigenetic subtypes, and candidate targets for therapy. This dataset provides a rich resource for the cancer genomics community and will serve as the foundation of ongoing and future candidate-driven functional studies focused on resolving MB aetiology.

Patient cohorts and genomic datasets
Patient-matched tumour and non-tumour (blood) Illumina DNA sequences were collected from a total of 579 untreated patients diagnosed with MB that were sequenced at one of four participating institutions (see Methods). After eliminating samples with poor quality sequencing data (based on quality control measures), samples sequenced more than once (that is, duplicate cases analysed at different sequencing sites or sequencing of patient-matched primary and relapse pairs), samples lacking molecular subgroup annotation, and cases with clear molecular evidence for misdiagnosis, we amassed a final cohort of 491 diagnostic MBs with matched normal samples for further analysis, including standardized sequence alignment and filtering, as well as harmonized single nucleotide variant (SNV), indel, and structural variant calling (Fig. 1). Germline and somatic alterations were annotated primarily from whole genomes (n = 390; n = 190 published5,6,5 and 200 unpublished) while the remaining alterations were derived from published whole exomes (n = 101)5,14. Patient ages ranged from 1 month to 50 years (median age = 8 years; Supplementary Table 1). Verification of MB diagnosis and subgroup status was established using a molecular classification approach based on DNA methylation arrays13 (see Methods). Illumina 450 k methylation array data were generated for 1,256 MBs, including 396 out of 491 (80.7%) of the NGS cohort. Transcriptome data were acquired through RNA sequencing (RNA-seq; n = 164) and Affymetrix expression arrays (n = 392). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) data were generated for several chromatin marks (such as H3K27ac and CTCF) on a subset of the cohort (Supplementary Table 1). Mutations, structural variants and supporting epigenetic and transcriptional data can be freely explored online through multiple data portals (see ‘Data availability’ in Methods).

Mutational signatures operative in MB
Mutational signatures have been extensively catalogued across a broad spectrum of cancerous tissues, and for many of these, underlying exogenous and endogenous processes have been described. A seminal pan-cancer analysis that included 100 MBs (not split by subgroup) revealed three predominant signatures active in MB: signatures 1, 5 and 8. Here we analysed a total of 440,459 somatic mutations across 385 MB genomes divided into molecular subgroups (5 out of 390 whole-genome sequencing (WGS) cases were excluded owing to quality control issues) and detected 24 signatures with a mutation contribution of at least 5% in one or more samples (Extended Data Fig. 1a). Signature 1, associated with patient age at diagnosis, was the most common signature in all subgroups (Extended Data Fig. 1a, b). Signature 3, which was not previously detected in MB16 and has been linked with underlying BRCA1 and BRCA2 mutations in breast, aggressive prostate, and pancreatic cancers16,17, was unexpectedly observed in most of the patients with Group 3 and Group 4 MB, and

Figure 1 | Summary of MB genomic datasets. Graphical summary of genomic, epigenomic, and transcriptomic MB datasets analysed in the study.
extensively detailed in a parallel study investigating more than 1,000 MB germ lines (Supplementary Table 2) (S.M.W. et al., manuscript submitted).

To discriminate potential drivers from passengers, we annotated our dataset using MutSigCV, GenomeMuSiC and IntOGen significance algorithms (Fig. 2d; Supplementary Table 2). Overlapping the output derived from these analyses identified a core set of high-confidence somatic drivers detected by all three algorithms (Fig. 2d; Supplementary Table 2). Overlapping genes; n = 491 samples).

WNT subgroup MB

All 36 WNT MBs sequenced in this study were confidently explained by mutations in at least one or more driver genes (Extended Data Fig. 6a). Somatic CTNNB1 mutations, the hallmark feature of WNT-driven MB, were found in 86% of patients. Three CTNNB1 wild-type MBs harboured pathogenic APC germline variants, explaining the WNT pathway activation seen in these patients and underscoring the need to perform genetic testing for APC carrier status (that is, Turcot syndrome) when WNT MB is suspected despite failure to detect mutant CTNNB1. Monosomy 6, a signature chromosomal alteration characteristic of patients with WNT MB, was confirmed in 83% of cases (Extended Data Figs 6a, 7a–c), demonstrating that neither CTNNB1 mutation nor chromosome 6 loss are universally present in all patients with WNT MB. The latter has direct clinical implications, driving the need for genetic testing for CTNNB1 mutations in patients with WNT MB (Extended Data Fig. 5b).

Consistent with our previous studies, genetic events targeting histone modifiers, especially those regulating lysine methylation and/or acetylation, were found across subgroups and contributed to a considerable proportion of cases (Extended Data Fig. 5c, d), further corroborating the hypothesis that deregulation of the epigenetic machinery is fundamental to MB development.

Figure 2 | Driver genes and pathways altered in MB. a, Oncoprint summarizing recurrently altered genes according to MB subgroup (n = 390; WGS series only). b, Top, Venn diagram summarizing the subgroup overlap of recurrently mutated genes (≥3 affected cases). Bottom, incidence plot of recurrently mutated genes (≥3 affected cases) detected in the series (n = 356 genes; n = 491 samples). c, Graphical summary of the most frequently mutated genes (≥10 affected cases) and their subgroup distribution. d, Venn diagram summarizing the significantly mutated gene lists output from multiple significance algorithms. e, Results from d restricted to Cancer Gene Census (CGC) genes.
trials (NCT02601937) and based on our findings here, could represent a rational targeted therapy for treating WNT MB.

**SHH subgroup MB**

Building on our previous work, we reliably assigned at least one driver gene to more than 95% of patients with SHH MB, and revealed several insights that extend beyond genetic events that target the canonical SHH signalling pathway (Extended Data Fig. 6b). *IDH1* mutations represent a hallmark genetic event in adult patients with glioma that exhibit a distinct hypermethylated phenotype (that is, glioma CpG island methylator phenotype, G-CIMP)\(^27,28\). We identified six *IDH1* R132C mutations (five SHH, one WNT), consistent with a recent case report\(^29\) (Extended Data Fig. 6e). *IDH1*-mutant SHH MBs were determined to be CIMP\(^+\) (Extended Data Fig. 6f, g), confirming that these mutations lead to epigenetic consequences reminiscent of those reported in other IDH1/2-mutant cancers.

Systematic Gene Ontology (GO) and pathway analysis demonstrated significant overrepresentation of somatic alterations that target histone acetyltransferase (HAT) complexes in SHH compared to other subgroups (\(q = 2.2 \times 10^{-5}\); Extended Data Fig. 6h). Genes that encode HATs, namely *CREBBP, KAT6B* and *EP300*, as well as HAT complex regulatory components *BRPF1* and *KANSL1*, all exhibit recurrent, mostly SHH subgroup-restricted mutations in our series (19% of patients with SHH MB). The mechanism(s) by which deregulation of HAT activity cooperates with constitutively active SHH signalling remains poorly defined, warranting further studies to determine whether this epigenetic pathway can be exploited therapeutically.

**Group 3 and Group 4 subgroup MB**

Re-analysis of NGS data derived from previous studies (\(n = 173\))\(^5,9,14\) indicated that less than one-third of Group 3 (32%) and Group 4 (30%) cases could be explained by at least one probable driver event (Fig. 3c), consistent with the individual estimates reported in those publications.

In the considerably larger, more integrated dataset analysed here, we confidently assigned potential driver events to 76% and 82% of Group 3 and Group 4, respectively, more than doubling the proportion of explained cases per subgroup (Fig. 3a–c). As expected, MYC amplifications were restricted to patients with Group 3 MB (17%), whereas MYCN amplifications were found at a comparable frequency in patients with either Group 3 (5%) or Group 4 (6%) MB (Fig. 3a, b). Similarly, structural variants leading to aberrant induction of *GFI1* or *GFI1B\(^3\) were mutually exclusive and distributed in both subgroups (Fig. 3a, b). Mutually exclusive analysis disclosed that the most prominent candidate driver events were largely non-overlapping and very few cooperating events were identified (Fig. 3d and data not shown).

Pathway analysis of recurrent genetic events revealed significant overrepresentation of genes involved in the Notch and TGFβ3 signalling pathways in Group 3, and in chromatin modification in Group 4 (Extended Data Fig. 5c). Aberrant Notch signalling has been repeatedly suggested in the MB literature\(^30,31\); however, this is, to our knowledge, the first report documenting Notch pathway mutations in samples from patients with MB. A role for deregulated TGFβ3 signalling in Group 3 has been suggested in our previous genomic/epigenomic studies\(^11,32\); however, functional studies that further substantiate these observations are still lacking.

**Epigenetic refinement of MB substructure**

The molecular composition and boundaries defining Group 3 and Group 4 MB subgroups are not as clearly demarcated as their WNT and SHH counterparts. Notable similarities between Group 3 and Group 4 have been discussed, including ambiguous cases that exhibit features characteristic of either subgroup\(^33–35\). To refine the inter- and intra-subgroup heterogeneity underlying MB subgroups, we sought to investigate molecular substructure in a series of 1,256 MBs profiled by Illumina 450 k methylation array (Extended Data Fig. 8a). Analysis of pairwise sample similarities using t-distributed stochastic neighbour...
embedding (t-SNE) uncovered notable heterogeneity across the cohort, especially in non-WNT subgroups. Restricting our analysis to Group 3 and Group 4 (n = 740) separated the parental subgroups into up to eight subtypes (Fig. 4a, b). Iterative down-sampling performed on the same dataset verified that molecular substructure seems to stabilize once 500 or more samples are included in the analysis (Extended Data Fig. 8b), exemplifying the power afforded by the high sample number included here.

Integration of methylation subtypes with sample-matched genomic and transcriptomic data revealed marked enrichment of probable driver events in specific subtypes (Fig. 4c; Extended Data Fig. 8c–e). For example, somatic events targeting the known MB drivers GFI1B (subtype I) and MYC (subtype II), as well as chromatin-modifying genes KDM6A and ZMYM3 (both subtype VIII), all demonstrated remarkable subtype specificity. Broad copy-number alterations were also differentially distributed among subtypes (Extended Data Fig. 8c). Analysis of case-matched gene expression array data (n = 248) confirmed discriminatory transcriptional features associated with these subtypes, including aberrant expression of the MYC and GFI1 family oncogenes (Extended Data Fig. 8d, e).

**Hotspot insertions target KBTBD4**

Recurrent, in-frame insertions targeting KBTBD4 in Group 3 and Group 4 were among the most compelling single-gene discoveries in this large dataset. Of 20 somatic KBTBD4 variants we identified, 18 (90%) were determined to be in-frame insertions clustered across just six amino acids within the KBTBD4 Kelch domain (Fig. 5a). Notably, the predominant insertion inferred in Group 3, of proline and arginine at Arg313 (R313>PRR) differed from that observed in Group 4, an insertion of proline at Pro311 (P311>PP). Overlaying KBTBD4 mutation status with methylation subtype assignments revealed two tightly clustered mutation groups within subtypes II (21%) and VII (14%) (Fig. 5b), ranking KBTBD4 as the most prevalent candidate driver identified in these subtypes. KBTBD4 encodes a BTB–BACK–Kelch domain protein belonging to a large family of cullin-RING ubiquitin ligase adaptors that facilitate the ubiquitination of target substrates\(^5\). Homology modelling of the KBTBD4 Kelch domain with known structures derived from other family members (n = 12) verified that the MB-specific insertions observed here are unlikely to disrupt the overall structure of the Kelch domain but instead converge on the known substrate-binding interface described for other family members (Fig. 5c).

**Enhancer hijacking activates PRDM6**

We previously identified\(^6\) GFI1 and GFI1B as new MB oncogenes recurrently activated by ‘enhancer hijacking’ in Group 3 and Group 4. Expanding on this previous work, we recently developed cis expression structural alteration mapping (CESAM)\(^7\), an approach for systematically inferring enhancer hijacking events genome-wide by integrating gene expression and structural variant data that we interpret in the context of topologically associated domains (TADs) and enhancer annotations. Application of CESAM to MB (n = 164) confidently identified GFI1B among the most highly significant candidate genes subject to enhancer hijacking, substantiating the robustness of our method (Extended Data Fig. 9a). The top-ranking gene uncovered by CESAM, however, was PRDM6 (chr5q23), encoding a poorly characterized SET-domain containing protein. Notably, PRDM6 maps approximately 600 kb downstream of SNAIP1, a gene known to be targeted by highly recurrent, stereotypical tandem duplications uniquely restricted to patients with Group 4 MB\(^8\) (Fig. 6a). In the context of Group 4 patients harbouring SNAIP1-associated structural variants analysed here, PRDM6 expression was markedly upregulated (at least 20-fold), considerably more than any other gene mapping within the proximal TADs including SNAIP1 (Fig. 6b; Extended Data Fig. 9b).
Using our recently published MB enhancer data\textsuperscript{32} and structural variant breakpoints to identify putative promoter–enhancer juxtaposition as a consequence of structural variants, we identified a significant enrichment of structural variants associated with rearrangements that link PRDM6 to Group 4 enhancer elements ($P < 0.0001$, permutation test). The SNCAIP locus overlaps a strong Group 4-specific super-enhancer (Fig. 6c). Notably, the structural variants observed in PRDM6-activated Group 4 converge on the SNCAIP super-enhancer, consistent with enhancer hijacking (Fig. 6c). Integrative analysis of CTCF chromatin data revealed notable clustering of structural variant breakpoints proximal to CTCF-binding sites associated with the TAD boundary separating the SNCAIP and PRDM6 loci (Fig. 6c). Collectively, these data suggest that structural variants targeting the SNCAIP locus disrupt the local chromatin environment to promote de novo interactions between the SNCAIP super-enhancer and gene promoters in the neighbouring TAD, thus leading to aberrant gene induction, most notably PRDM6 (Fig. 6d).

**Discussion**

Our highly integrative genomic analysis of the paediatric brain tumour MB has enabled the discovery of new cancer genes and actionable pathways, effectively assigning candidate drivers to most tumours in a rapid fashion. Most notably, KBTBD4 gene promoters in the neighbouring TAD, thus leading to aberrant gene expression. PRDM6 insertions were highly specific for discrete patient subtypes and were devoid of other obvious oncogenic driver events, suggesting that these mutations are functional. Likewise, PRDM6—a presumed histone methyltransferase\textsuperscript{38} not previously implicated in MB—was identified as the probable target of SNCAIP-associated enhancer hijacking in Group 4, now representing the most prevalent driver alteration in this subgroup. Studies further detailed the normal, physiological cellular functions of KBTBD4 and PRDM6 and how somatic alterations targeting these genes specifically contribute to MB pathogenesis are essential and will be required to determine their potential ‘actionability’ in affected patients.

The relatively recent recognition of consensus MB subgroups has rapidly changed the way MB is studied in the research setting and how it is diagnosed and treated in the clinic\textsuperscript{39}. Still, considerable molecular and clinical heterogeneity has been demonstrated\textsuperscript{11,40}, suggesting that currently defined MB subgroups are likely to be an oversimplification of true molecular substructure. Methylation analysis of over 1,250 MBs discovered new tumour subtypes enriched for specific genetic and transcriptional signatures, especially those underlying Group 3 and Group 4. Definitive de-convolution of these subtypes will enable a better understanding of the developmental origins of MB, creating a path towards the efficient modelling of each individual subtype in the correct cellular context using subtype-relevant genetic perturbations. Moreover, by redefining molecular substructure as we have described here, new opportunities for improved risk-stratification tailored to treat individual patient subtypes according to their genotype are likely to emerge. 

**Figure 6 | Enhancer hijacking activates PRDM6 in Group 4 MB.** a, Summary of structural variants (SVs) targeting the SNCAIP locus in Group 4. b, Group 4 MB expression box plots of genes mapping proximal to SNCAIP-associated structural variants. NS, not significant. c, Summary of annotated chromatin interactions (Hi-C), TADs (brown bars), CTCF chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and CTCF binding orientation (red and blue arrowheads), as well as SNCAIP-associated structural variants, CTCF ChIP–seq peaks, Group 4-specific super-enhancers (SEs), H3K27ac ChIP–seq peaks, and RNA-seq data derived from a subset of Group 4 MBs, stratified according to underlying SNCAIP structural variant status. d, Proposed model depicting inferred molecular basis of SNCAIP/PRDM6-associated enhancer hijacking.
In conclusion, this study embodies an unparalleled resource of high-resolution genetic, epigenetic and transcriptional data for the childhood brain tumour MB. Our data underscore the heterogeneous, complex nature of disease subgroups and the utility of continued efforts to divulge the full spectrum of molecular mechanisms underlying MB aetiology. We anticipate that the findings reported here, combined with the future exploration and mining of this large genomics resource, will undoubtably advance treatments and the outlook for children and families affected by this devastating malignancy.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 September 2016; accepted 10 May 2017.**

1. Kandoth, C. et al. Mutational landscape and significance across 12 major cancer types. Nature 502, 333–339 (2013).
2. Tamborero, D. et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci. Rep. 3, 2650 (2013).
3. Jones, D. T. et al. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytomas. Nat. Genet. 45, 927–932 (2013).
4. Lee, R. S. et al. A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. J. Clin. Invest. 122, 2983–2988 (2012).
5. Jones, D. T. et al. Dissecting the genomic complexity underlying medulloblastoma. Nature 488, 100–105 (2012).
6. Kool, M. et al. Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. Cancer Cell 25, 393–405 (2014).
7. Northcott, P. A. et al. Enhancer hijacking activates GF1 family oncogenes in medulloblastoma. Nature 511, 428–434 (2014).
8. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 148, 59–71 (2012).
9. Robinson, G. et al. Novel mutations target distinct subgroups of medulloblastoma. Nature 488, 43–48 (2012).
10. Morrissey, A. S. et al. Divergent clonal selection dominates medulloblastoma at recurrence. Nature 529, 351–357 (2016).
11. Northcott, P. A. et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. Nature 488, 49–56 (2012).
12. Northcott, P. A. et al. Medulloblastomics: the end of the beginning. Nat. Rev. Cancer 12, 818–834 (2012).
13. Taylor, M. D. et al. Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol. 123, 465–472 (2012).
14. Pugh, T. J. et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature 488, 106–110 (2012).
15. Hovestadt, V. et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density array comparative genomic hybridization (aCGH) analysis. Nat. Genet. 45, 913–916 (2013).
16. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. Nature 500, 415–421 (2013).
17. Deckter, B. et al. Biallelic BRCA2 mutations shape the somatic mutational landscape of aggressive prostate tumors. Am. J. Hum. Genet. 98, 818–829 (2016).
18. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 497, 330–337 (2012).
19. Kandoth, C. et al. Integrated genomic characterization of endometrial carcinoma. Nature 497, 67–73 (2013).
20. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–218 (2013).
21. Dees, N. D. et al. MuSiC: identifying mutational significance in cancer genomes. Genome Res. 22, 1589–1598 (2012).
22. Gonzalez-Perez, A. et al. IntogenMutations identifies cancer drivers across tumor types. Nat. Methods 10, 1081–1082 (2013).
23. Northcott, P. A. et al. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. Nat. Genet. 41, 465–472 (2009).
24. Parsons, D. W. et al. The genetic landscape of the childhood cancer medulloblastoma. Science 331, 439–443 (2011).
25. Fackenthal, J. D. et al. Somatic mutations of the APC gene in patients with Tucor syndrome and analysis of APC mutations in brain tumors. Genes Chromosom. Cancer 9, 168–172 (1994).
26. Wilson, B. G. et al. Epigenetic antagonism between polycomb and SWI/SNF complexes disrupts genomic transformation. Cancer Cell 18, 316–328 (2010).
27. Nosuchmehr, H. et al. Identification of a Cpg island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell 17, 510–522 (2010).
28. Cuccarelli, M. et al. Molecular profiling reveals bio logically discrete subsets and pathways of progression in diffuse Glioma. Clin. Cancer Res. 16, 550–563 (2010).
29. Snuderl, M. et al. Deep sequencing identifies IDH1 R132S mutation in adult medulloblastoma. J. Clin. Oncol. 33, e27–e31 (2015).
30. Fan, X. et al. Notch1 and Notch2 have opposite effects on embryonal brain tumor growth. Cancer Res. 64, 7787–7793 (2004).
31. Garzia, L. et al. MicroRNA-199a-5p impairs cancer stem cells through negative regulation of HES1 in medulloblastoma. PLoS One 4, e4998 (2009).
32. Lin, C. Y. et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. Nature 530, 57–62 (2016).
33. Cho, Y. J. et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. J. Clin. Oncol. 29, 1424–1430 (2011).
34. Kool, M. et al. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. PLoS One 3, e3896 (2008).
35. Northcott, P. A. et al. Medulloblastoma comprises four distinct molecular variants. J. Clin. Oncol. 29, 1408–1411 (2011).
36. Canning, P. et al. Structural basis for Cui3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases. J. Biol. Chem. 288, 7803–7814 (2013).
37. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates IR54 and IG2 in enhancer hijacking. Nat. Genet. 49, 65–74 (2017).
38. Davis, C. A. et al. PRISM/PRDM6, a transcriptional repressor that promotes the proliferative gene program in smooth muscle cells. Mol. Cell. Biol. 26, 2626–2636 (2006).
39. Louis, D. N. et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 131, 803–820 (2016).
40. Shih, D. J. et al. Cytogenetic prognostication within medulloblastoma subgroups. J. Clin. Oncol. 32, 886–896 (2014).
METHODS
No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment unless stated otherwise.

Patient consent. ICGC samples: all patient material was collected after receiving informed consent according to ICGC guidelines and as approved by the institutional review board of contributing centres. Broad and MAGIC samples: informed consent was provided by the families of patients with medulloblastoma treated at Children’s Hospital Boston (Boston, Massachusetts, USA), The Hospital for Sick Children (Toronto, Canada), and institutions contributing to the Children’s Oncology Group/Cooperative Human Tissue Network, under approval and oversight by their respective internal review boards.

St Jude samples: human tumour and matched blood samples were obtained with informed consent through an institutional review board approved protocol at St Jude Children's Research Hospital (Memphis, Tennessee, USA).

Bam to FASTQ and alignment. NGS data were collected from four primary sources (ICGC PedBrain38, PCGP39, MAGIC, and the Broad Institute40). To ensure all samples were processed with the same analysis pipelines, sequences that were not available as FASTQ files were unaligned using the SamToFASTQ tool from Picard (http://broadinstitute.github.io/picard). To avoid biases in the insert size estimation of the realignment the Bam files were name sorted before the unalignment. The subsequent alignments were done according to the standards defined for ICGC PanCancer41. All reads were aligned against the phase II reference of the 1000 Genomes Project including decoy sequences d5 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/phase2_reference_assembly/his37d5.fa.gz) using BWA-MEM (v.0.7.8 using standard values except for invoking -T 0)42. The raw Bam files were sorted and duplicates were marked using biobambam (v.0.0.148). Sequencing coverage was calculated using custom scripts43. For annotations we chose the latest compatible GENCODE version 19 (http://www.genecodesgenes.org/releases/19.html).

Variant calling. Somatic variant calling (SNVs, indels, structural variants and CNVs) was done using the DKFZ/EMBL core pipelines in accordance with ICGC PanCancer44. The workflow is available on the Dockstore webpage: https://dockstore.org/containers/quay.io/pancan/pcawg-dkfz-workflow.

SNVs. SNVs were called using the DKFZ samtools-based calling pipeline as described38 using the ICGC PanCancer version. In short, variants were first called in the tumour sample and then queried in the control sample. The raw calls were then annotated using multiple publicly available tracks such as 1000 Genome variants, single nucleotide polymorphism database (dbSNP), repeats and other elements. The functional effect of the mutations was annotated using Annovar45 and the variants were assessed for their confidence and split into somatic and non-somatic calls. Owing to the poor coverage of the TERT promoter region, variants were called with relaxed stringency manually using custom scripts.

Indels. Raw calls for indels were obtained from Platypus (v.0.7.4)46. Annotation and confidence assessment was done similar to SNV processing.

SNV and indel integration. SNVs and indels were integrated using custom scripts. Variant frequencies were calculated for the whole cohort and for each subgroup individually. To increase the already high specificity of the workflows38, we manually checked all functional variants (non-synonymous, stop-gain, stop-loss and splice-site SNVs and in-frame, frame-shift and splice-site indels), in genes that had at least three hits in the cohort. For manual curation, we used a custom script to take screenshots for each variant and then scored the confidence randomly at least three times for each call.

Structural variants. Structural variants were inferred using DELLLY47 that ensured high precision and recall values based on precision-recall analysis. Recall was calculated as the number of regions that satisfy the cut-off in results obtained on both halves of the dataset; precision was calculated as a fraction of the recalled regions to the total number of regions satisfying the cut-off in the data. We then used the chosen parameters and executed the pipeline on the complete dataset.

Oncoplots. The data from the variant calling workflows were summarized using custom scripts and plotted into oncoplots using the R package complex heat maps50. Frequencies of events were adjusted to the number of samples that could be annotated for the respective event (that is, samples where we could not call CNVs were not counted and shaded light grey for CNV relevant genes). Subgroup enrichment for specific genes was determined using Fisher's exact test and a threshold of the Benjamini–Hochberg-adjusted P value (P < 0.05).

Significance workflows. The significance of somatic mutations (SNVs, indels) was determined using three published methods: MuSiCv2, IntOGen22 v.2.4.2 and MuSiCv3 v.0.4. The corresponding input data formats were parsed from our custom VCF files and loaded into the respective tools. The programs were run using default settings. Significant genes were determined using the recommended significance thresholds for each of the output files.
Copy number integration/significance. Significant copy number gains and losses (WGS samples; n = 352) were calculated using GISTIC\textsuperscript{14} v.2.0.22. We used a custom script to parse the region based output from ACESeq into a segmented data format suitable for GISTIC. Regions containing false positive recurrent events mainly around centromeres and telomeres were excluded from the analysis. The following 38 samples were excluded from the analysis owing to low data quality: ICGC\_MB126, ICGC\_MB143, ICGC\_MB147, ICGC\_MB149, ICGC\_MB246, ICGC\_MB256, ICGC\_MB304, ICGC\_MB305, ICGC\_MB306, ICGC\_MB62, ICGC\_MB800, ICGC\_MB889, ICGC\_MB92, ICGC\_MB94, MDT\_AP-0009, MDT\_AP-1200, MDT\_AP-1367, MDT\_AP-1369, MDT\_AP-1403, MDT\_AP-1405, MDT\_AP-2073, MDT\_AP-2110, MDT\_AP-2111, MDT\_AP-2116, MDT\_AP-2307, MDT\_AP-2514, MDT\_AP-2532, MDT\_AP-2673, MDT\_AP-2719, MDT\_AP-2745, MDT\_AP-2774, MDT\_AP-3017, MDT\_AP-3399, MDT\_AP-3402, SJMB015 and SJMB019. GISTIC was run separately for each subgroup using a length cut-off of 0.5 chromosome arms, a noise threshold of 0.3 copies, a cap of 1.5, a confidence level of 0.95 and gene GISTIC for the delineation analysis.

Mutation signatures. Mutation signatures are calculated based on trinucleotides centred at somatic SNVs. Therefore, the immediate 3’ and 5’ nucleotides of all somatic SNVs were extracted from the reference genome and the obtained trinucleotides were converted to pyrimidine context resulting in 96 possible mutation types. Directly adjacent SNVs (multiple nucleotide variants, MNVs) were excluded for mutation signature analysis. For each sample, its mutational profile was calculated by counting the number of occurrences of each of the possible 96 mutation types. By combining mutational profiles of all samples per entity, mutational catalogues for signature extraction were compiled. The mutational profile of a tumour, and therefore, the mutational catalogue of a tumour type, is supposed to reflect a pattern of exposures for one or more subgroups (that is, significant enrichment as unknown variables. The resulting expression exposures were used for further downstream analyses and visualization. Signature probability distributions are displayed for the 96 mutation types according to the representation described previously\textsuperscript{16}. Association of signature exposures and age at diagnosis were calculated by generalized linear models in all subgroups. Specificity of exposures for one or more subgroups (that is, significant enrichment of exposure compared to the other groups) was determined using ANOVA and post hoc Tukey’s test.

DNA methylation array processing. DNA methylation profiling of MB samples was performed using the Infinium HumanMethylation450 BeadChip array (450k array) according to the manufacturer’s instructions (Illumina). Data were primarily generated at the DKFZ Genomics and Proteomics Core Facility (Heidelberg, Germany) and The Hartwell Center at St Jude Children’s Research Hospital (Memphis, USA). MB subgroup status was inferred as previously described\textsuperscript{15} or inherited from published annotations\textsuperscript{8,14}. DNA methylation data of 1,256 samples presented in this study were generated from both fresh-frozen and formalin-fixed paraffin-embedded (FFPE) tissue samples. For most fresh-frozen samples, more than 500 ng of DNA was used as input material. DNA was used in a ratio of 1:1. DNA was used within a 24-hr window of primary culture. DNA metrics of all samples were carefully controlled. Samples were also checked for unexpected genotype matches by pairwise comparison of the 65 genotyping probes on the 450k array.

All DNA methylation analyses were performed in R v.3.3.0 (R Development Core Team, 2016). Raw signal intensities were obtained from IDAT-files using the minfi Bioconductor package\textsuperscript{51} v.1.18.0 using default settings. A correction for the type of material (FFPE/frozen) was performed by using the removeBatchEffect function of the limma package v.3.24.15. The log\textsubscript{2}-transformed intensities of the methylated and unmethylated signal were corrected individually. Beta-values were calculated from the retransformed intensities using an offset of 100 (as recommended by Illumina) and used for all downstream analyses.

The following criteria were applied to filter out probes prone to yield inaccurate methylation levels: removal of probes targeting the X and Y chromosomes (n = 11,551), removal of probes containing an SNP (dSNP132Common) within five base pairs of and including the targeted CpG site (n = 24,536), and probes not mapping uniquely to the human reference genome (hg19) allowing for one mismatch (n = 9,993). In total, 438,370 probes were kept for analysis.

For unsupervised t-SNE analysis of 1,256 MB samples, we selected the 22,349 most variably methylated probes across the dataset (s.d. > 0.25). Pairwise sample distances were calculated by using 1 minus the weighted Pearson correlation coefficient as the distance measure. Pairwise Pearson correlation was calculated using the wtd.corr function of the weights package v.0.85. We used the probe standard deviation subtracted by 0.25 as the weight, giving more variable probes greater influence. The resulting distance matrix was used to perform the t-SNE analysis (Rtsne package v.0.11). The following non-default parameters were used: theta = 0, is_distance = T, pca = E, max_iter = 2000. Resulting clusters were annotated as WNT, SHH, Group 3 and Group 4 based on classification using a previously described 48 CpG signature\textsuperscript{15}.

A similar approach was used for the unsupervised analysis restricted to Group 3 and 4 samples (n = 740, 12,454 most variable probes, s.d. > 0.25), and for the downsampling analysis. To ensure a similar rotation of samples, t-SNE analysis was performed by using the sample coordinates obtained after 150 iterations of the analysis of all MB samples as initialization points, and then performing an additional 1,850 iterations for the respective subset of samples. For the analysis of Group 3 and Group 4 samples, clusters were annotated using the DBSCAN algorithm as implemented in the dbscan package v.0.9-7. The following non-default parameters were used: minPts = 16, eps = 3.9. Subsequently, samples not assigned to any cluster were iteratively merged to their nearest cluster.

For the comparison of IDH1-mutated samples, we restricted the analysis to samples of the SHH subgroup that were also part of the sequencing cohort (n = 89). The 16,946 most variably methylated probes were used (s.d. > 0.25). One minus the Pearson correlation coefficient was used as the distance measure, and average linkage was used for hierarchical clustering.

CVN analysis from 450 k methylation array data was performed using the comuse Bioconductor package v.1.4.0. A set of 50 control samples displaying a balanced copy-number profile was used for normalization.

Gene expression array processing. Samples for which RNA of sufficient quality and quality was available were analysed on the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array. Sample library preparation, hybridization and quality control were performed according to manufacturer’s protocols. Expression data were normalized using the MAS5.0 algorithm of the GCOS program (Affymetric).

Gene expression differences between MB Group 3 and group 4 subtypes were analysed using available data for samples that were classified in the initial DBSCAN annotation (n = 219 samples). Excluding genes located on chromosomes X and Y, differentially expressed genes between subtypes were determined by the ANOVA procedure, and using Tukey’s post hoc test. Genes were considered differentially expressed if the false discovery rate (FDR)-adjusted P value across subtypes is < 0.01, and if for at least one comparison between subtypes the absolute difference of mean expression levels was larger than 2 and P < 0.01 (n = 869 genes).

RNA-seq and ChIP–seq data generation and analysis. RNA-seq and ChIP–seq data were generated as described previously\textsuperscript{32,38}. CESAM. CESAM integrates structural variant-derived breakpoints with RNA-seq data to identify expression changes associated with breakpoints in cis as described in previously\textsuperscript{32}, by performing linear regression of expression (molecular phenotype) on structural variant-derived breakpoint (somatic genotype) data. We used CESAM as described previously\textsuperscript{32}, with some modifications. In brief, locally recurring structural variant breakpoints were assigned to bins if they fell into the same pre-annotated TAD, using TAD data from the IMR90 cell line\textsuperscript{54} (mean TAD size = 830 kb). A somatic genotype matrix based on ‘TAD bins’ was constructed using BEDTools (v.2.24.0)\textsuperscript{33} by annotating for every sample the presence/absence of these breakpoints within TADs. We then performed GISTIC\textsuperscript{44} on the somatic SCNA-derived breakpoint matrix. We used an FDR of 5% using the Benjamini–Hochberg procedure, and required more than 2-fold expression upregulation relative to controls for reporting CESAM candidate genes. Fold change was computed as the median expression in the group of structural variant
ArticulereSeArcH

[37x567]related pathways and biological processes, considering the interaction-

{1, 2, 3}, and

{0.0005, 0.001, 0.005, 0.007, 0.008, 0.009, 0.01},

Collecting Steiner Forest (PCSF) problem to reduce the size of the network around

structural variants each with two breakpoints

For the final display items, we limited the algorithm to only include genes in the search

k × n × m

Predicting cancer driver genes using machine learning approaches requiring a

‘carriers’ compared to the median of ‘non-carrier’ control donors (median values

have in the interactome (‘degree’) should be about the same for

Considering the interaction-neighbourhood of the alterations (along with the reductionist approach of PCSF)

Considering the interaction-neighbourhood of the alterations (along with the reductionist approach of PCSF)

CoMet analysis. Combinations of Mutually Exclusive Alterations (CoMEt) is

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Data availability. Short-read sequencing and microarray data have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) hosted by the EBI, under accession number EGAS00001001953. Genetic, epigenetic and transcriptional data can be freely explored using the PeCan (http://pecan.stjude.org/proteinpaint/study/BT.MB...Pfister%20pan-MB), R2 (https://hgserver1.amc.nl/cgi-bin/r2/main.cgi?&dscope=MB500&option=about_dscope), and PedBio (http://pedbiportal.org/study.do?cancer_study_id=medullo_pa_01#summary) data portals. All other data are available from the corresponding authors upon reasonable request.

41. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korbel, J. O. Data analysis: Create a cloud commons. Nature 523, 149–151 (2015).
42. Li, H. et al. The Sequence Alignment/Map format and SAM tools. Bioinformatics 25, 2078–2079 (2009).
43. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).
44. Rimmer, A. et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat. Genet. 46, 912–918 (2014).
45. Alioto, T. S. et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. Nat. Commun. 6, 10001 (2015).
46. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28, i333–i339 (2012).
47. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311 (2001).
48. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
49. Olshef, A. B. et al. Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. Bioinformatics 27, 2038–2046 (2011).
50. Johann, P. D. et al. Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. Cancer Cell 29, 379–393 (2016).
51. Merzel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 12, R41 (2011).
52. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. Cell Reports 3, 246–259 (2013).
53. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363–1369 (2014).
54. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380 (2012).
55. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
56. Ønæ, H., Bull, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 32, 1479–1485 (2015).
57. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
58. Tuncbag, N. et al. Network-Based interpretation of diverse high-throughput datasets through the Omics Integrator software package. PLOS Comput. Biol. 12, e1004879 (2016).
59. Razick, S., Magklaras, G. & Donaldson, I. M. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics 9, 405 (2008).
60. Wishart, D. S. et al. HMDB 3.0—The Human Metabolome Database in 2013. Nucleic Acids Res. 41, D801–D807 (2013).
61. Thiele, I. et al. A community-driven global reconstruction of human metabolism. Nat. Biotechnol. 31, 419–425 (2013).
62. Cline, M. S. et al. Integration of biological networks and gene expression data using Cytoscape. Nat. Protocols 2, 2366–2382 (2007).
63. Leiserson, M. D., Wu, H. T., Vandin, F. & Raphael, B. J. CoMEt: a statistical approach to identify combinations of mutually exclusive alterations in cancer. Genome Biol. 16, 160 (2015).
64. Bassini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252–W258 (2014).
65. Dobin, A. et al. STAR: ultratast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
Extended Data Figure 1 | Mutational signatures in MB. a, b, Exposure plot (a) and heatmap (b) summarizing mutation signatures, contributing \( \geq 5\% \) of the overall mutation burden per sample are depicted. Asterisks in b indicate subgroup-enriched signatures. c, d, Box plots showing the subgroup specificity of signatures 18 (c) and 5 (d).

e, Correlation of signatures 1 and 5 with patient age. f, Summary of total somatic mutation counts observed in the series. g, Bar plot summarizing distribution of mutation signatures in MBs with outlier mutation counts. h, Rainfall plots depicting somatic mutation burden in typical (top) and hypermutated (bottom) SHH MBs.
Extended Data Figure 2 | Genome-wide summary of somatic SNVs.

a, Precision-recall curves for different binomial P value cut-offs. Minimal and maximal precision values are shown in colour, mean precision is shown as dotted line. P value cut-offs for 200 bp window sizes are indicated. b, Manhattan plot showing the $-\log_{10}$ test statistic of 200 bp genomic windows plotted against their respective chromosomal positions. Red line indicates the genome-wide significance threshold ($P = 10^{-25}$). High-confidence regions are shown in red; regions representing probable false positives are shown in blue. c, Summary of TERT promoter mutations observed in the series.
Extended Data Figure 3 | Prevalent candidate driver mutations observed in MB. a, b. Gene-level summaries of SNVs/indels inferred to predominantly result in loss-of-function (LOF) (a) or gain-of-function (GOF) (b) of known and putative MB driver genes.
Extended Data Figure 4 | Summary of fusion gene transcripts detected by RNA-seq. a–c, Schematic summaries of high-confidence fusion transcripts targeting known or putative MB driver genes, organized according to MB subgroup.
Extended Data Figure 5 | Candidate driver genes and pathways in MB subgroups. a, b, Box plots summarizing allelic expression fractions (a) and estimated clonality (b) inferred for prominent MB driver gene mutations. c, GO and pathway summary of recurrently mutated genes in MB. GO and pathway categories are grouped according to functional theme and the proportion of cases affected by individual pathway alterations are plotted per subgroup and across the series. d, Network summary of recurrently mutated genes involved in histone lysine methylation (GO accession 0034968).
Extended Data Figure 6 | Mutational landscape of WNT and SHH MB.

| Extended Data Figure 6 | Mutational landscape of WNT and SHH MB. |
|------------------------|----------------------------------------|
| a, b, Oncoprint summaries of recurrently mutated genes and cytogenetic alterations in WNT (a; n = 36 samples) and SHH (b; n = 131 samples). |
| c, Gene-level summary of WNT subgroup-enriched CSNK2B and EPHA7 mutations. |
| d, Summary of SWI/SNF superfamily-type complex (GO accession 0070603) mutations observed in patients with WNT MB. |
| e, Gene-level summary of somatic IDH1 R132C mutations detected in MB. |
| f, Quantification of methylcytosine beta-values detected in IDH1-mutant versus wild-type SHH MBs. |
| g, Unsupervised hierarchical clustering of SHH MB methylation data, confirming CIMP in IDH1-mutated SHH MBs. |
| h, Summary of histone acetyltransferase complex (GO accession 0000123) mutations observed in patients with SHH MB. |

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 7 | Somatic copy-number alterations in MB. 

a, Copy-number heat maps for individual MB subgroups derived from WGS series. 
b, Genome-wide copy-number summary plots for the MB dataset shown in a. 
c, GISTIC plots summarizing significant CNVs according to MB subgroup.
Extended Data Figure 8 | t-SNE analysis of Group 3 and Group 4 methylation data. a, t-SNE plot of DNA methylation array data for 1,256 analysed MBs. b, t-SNE analysis of iteratively down-sampled Group 3 and Group 4 methylation data. c, Genome-wide copy-number summary plots for Group 3/Group 4 methylation subtypes. d, t-SNE plots showing the relative, normalized expression intensities of GFI1, GFI1B, MYC, MYCN and PRDM6 in methylation subtypes (n = 219). e, Expression heat map showing transcriptional diversity among new MB subtypes (n = 248).
Extended Data Figure 9 | SNCAIP-associated enhancer hijacking in Group 4 MB. a, Quantile–quantile plot depicting the statistical inference of CESAM applied to systematically identify loci targeted by enhancer hijacking in Group 3 and Group 4 (n = 164) MB. b, Ascending PRDM6 expression in Group 4 MB annotated according to SNCAIP structural variant status.