ARTICLE

doi:10.1038/nature13379

Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma

Paul A. Northcott[1]*, Catherine Lee[2,3]*, Thomas Zichner[4]*, Adrian M. Stütz[4], Serap Erkek[1,4], Daisuke Kawauchi[1], David J. H. Shih[5], Volker Hovestadt[6], Marc Zapatka[6], Dominik Sturm[1], David T. W. Jones[1], Marcel Kool[1], Marc Remke[5], Florence M. G. Cavalli[5], Scott Zuyderduyn[7], Gary D. Bader[7], Scott VandenBerg[8], Lourdes Adriana Esparza[3], Marina Ryzhova[9], Wei Wang[6], Andrea Wittmann[1], Sebastian Stark[1], Laura Sieber[1], Huriye Seker-Cin[1], Linda Linke[1], Fabian Kratochwil[1], Natalie Jäger[10], Ivo Buchhalter[10], Charles D. Imbusch[11], Gideon Zipprich[11], Benjamin Raeder[4], Sabine Schmidt[12], Nicolle Diessl[12], Stephan Wolf[12], Stefan Wiemann[12], Benedikt Brors[10], Chris Lawerenz[11], Jürgen Eils[11], Hans-Jörg Warnatz[13], Thomas Risch[13], Marie-Laure Yaspo[13], Ursula D. Weber[6], Cynthia C. Bartholomae[14], Christof von Kalle[14,15], Eszter Turányi[16], Peter Hauser[17], Emma Sanden[18,19], Anna Darabi[18,19], Peter Siesjö[18,19], Jaroslav Sterba[20], Karel Zitterbart[20], David Sumerauer[21], Peter van Sluis[22], Rogier Versteeg[22], Richard Volckmann[22], Jan Koster[22], Martin U. Schuhmann[23], Martin Ebinger[23], H. Leighton Grimes[24], Giles W. Robinson[25,26], Amar Gajjar[26], Martin Mynarek[27], Katja von Hoff[27], Stefan Rutkowski[27], Torsten Pietsch[28], Wolfram Scheurlen[29], Jörg Felsberg[30], Guido Reifenberger[30], Andreas E. Kulozik[31], Andreas von Deimling[32], Olaf Witt[31], Roland Eils[10,15], Richard J. Gilbertson[25,26], Andrey Korshunov[32], Michael D. Taylor[5,33], Peter Lichter[6,15], Jan O. Korbel[4,34], Robert J. Wechsler-Reya[3] & Stefan M. Pfister[1,31]

Medulloblastoma is a highly malignant paediatric brain tumour currently treated with a combination of surgery, radiation and chemotherapy, posing a considerable burden of toxicity to the developing child. Genomics has illuminated the extensive intertumoral heterogeneity of medulloblastoma, identifying four distinct molecular subgroups. Group 3 and group 4 subgroup medulloblastomas account for most paediatric cases; yet, oncogenic drivers for these subtypes remain largely unidentified. Here we describe a series of prevalent, highly disparate genomic structural variants, restricted to groups 3 and 4, resulting in specific and mutually exclusive activation of the growth factor independent 1 family proto-oncogenes, *GFI1* and *GFI1B*. Somatic structural variants juxtapose *GFI1* or *GFI1B* coding sequences proximal to active enhancer elements, including super-enhancers, instigating oncogenic activity. Our results, supported by evidence from mouse models, identify *GFI1* and *GFI1B* as prominent medulloblastoma oncogenes and implicate 'enhancer hijacking' as an efficient mechanism driving oncogene activation in a childhood cancer.

Recent genome sequencing studies of medulloblastoma, a leading cause of cancer-related mortality in children[1], have yielded considerable insight into the genes, pathways and overall mutational landscape contributing to its pathogenesis[2–4]. Despite these advances, medulloblastoma remains a vastly heterogeneous disease characterized by very few recurrently mutated genes[5]. Medulloblastoma comprises at least four distinct molecular subgroups—wingless (WNT), sonic hedgehog (SHH), group 3 and group 4—each of which exhibits unique clinical and biological attributes, consistent with the concept of medulloblastoma existing not as a single entity, but more aptly a collection of different diseases[6,7].

Of the consensus subgroups, group 3 and 4 medulloblastomas have the poorest outcomes and remain least understood in terms of underlying genetics and biology[5]. Somatic *MYC* and *MYCN* amplifications rank among the most prevalent driver events known in these subgroups,

[1]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [2]Biomedical Sciences Graduate Program, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0685, USA. [3]Tumor Initiation and Maintenance Program, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. [4]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Meyerhofstrasse 1, Heidelberg 69117, Germany. [5]The Arthur and Sonia Labatt Brain Tumor Research Centre, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [6]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [7]The Donnelly Centre, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. [8]Department of Pathology, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. [9]Department of Neuropathology, NN Burdenko Neurosurgical Institute, 4th Tverskaya-Yamskaya 16, Moscow 125047, Russia. [10]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [11]Data Management Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [12]Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [13]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, Berlin 14195, Germany. [14]Division of Translational Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Im Neuenheimer Feld 460, Heidelberg 69120, Germany. [15]Heidelberg Center for Personalised Oncology (DKFZ-HIPO), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [16]1st Department of Pathology and Experimental Cancer Research, Semmelweis University SE, II.sz. Gyermekklinika, Budapest 1094, Hungary. [17]2nd Department of Pediatrics, Semmelweis University, SE, II.sz. Gyermekklinika, Budapest 1094, Hungary. [18]Glioma Immunotherapy Group, Division of Neurosurgery, Lund University, Paradisgatan 2, Lund 221 00, Sweden. [19]Department of Clinical Sciences, Lund University, Paradisgatan 2, Lund 221 00, Sweden. [20]Department of Pediatric Oncology, Masaryk University and University Hospital, Brno, Cernopolni 9 Brno 613 00, Czech Republic. [21]Department of Pediatric Hematology and Oncology, 2nd Faculty of Medicine, Charles University and University Hospital Motol, V Úvalu 84, Prague 150 06, Czech Republic. [22]Department of Oncogenomics, AMC, University of Amsterdam, Meibergdreef 9, Amsterdam 1105, AZ Netherlands. [23]Department of Neurosurgery, Tübingen University Hospital, Hoppe-Seyler Strasse 3, Tübingen 72076, Germany. [24]Division of Immunobiology, Program in Cancer Pathology of the Divisions of Experimental Hematology and Pathology, Program in Hematologic Malignancies of the Cancer and Blood Disease Insitute, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 452229, USA. [25]Department of Developmental Neurobiology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [26]Department of Oncology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [27]Department of Paediatric Haematology and Oncology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany. [28]Department of Neuropathology, University of Bonn, Sigmund-Freud-Str. 25, Bonn 53105, Germany. [29]Cnopf'sche Kinderklinik, Nürnberg Children's Hospital, St-Johannis-Mühlgasse 19, Nürnberg 90419, Germany. [30]Department of Neuropathology, Heinrich-Heine-University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. [31]Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Im Neuenheimer Feld 430, Heidelberg 69120, Germany. [32]Department of Neuropathology, University of Heidelberg, Im Neuenheimer Feld 220, Heidelberg 69120, Germany. [33]Division of Neurosurgery, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [34]EMBL, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Saffron Walden CB10 1SD, UK.
*These authors contributed equally to this work.

altered in just 17% (*MYC*) and 6% (*MYCN*) of group 3 and group 4 medulloblastomas, respectively[8]. Recurrent, somatically mutated genes are equally scarce, and for most cases, no obvious somatic 'drivers' have yet been revealed[5].

By analysing medulloblastoma genome sequencing data from different initiatives[2,4], we identified a series of spatially clustered somatic genomic structural variants (SVs) involving diverse SV classes that are linked to activation of *GFI1B* or its paralogue *GFI1* in group 3 and group 4 medulloblastomas. Further genomic and epigenomic analyses revealed a varied yet consistent interplay between SVs and the underlying epigenome that can explain *GFI1* and *GFI1B* activation in most cases. Functional analyses performed in mice confirmed the oncogenicity of GFI1 and GFI1B in the context of medulloblastoma. Collectively, these studies establish *GFI1* and *GFI1B* as novel, highly prevalent medulloblastoma oncogenes specifically activated in group 3 and group 4.

Diverse SVs activate GFI1B in medulloblastoma

Whole-genome sequencing (WGS; standard 100-base-pair (bp), paired-end and large-insert paired-end sequencing, see Methods) of 137 primary group 3 and 4 medulloblastoma samples (46 published[2,4] and 91 newly generated; Supplementary Table 1) facilitated a systematic, high-resolution screen for somatic SVs targeting novel medulloblastoma drivers. Rather than limiting our search to minimal common regions of recurrent amplification or deletion, a well-established approach for identifying somatically altered cancer genes[9,10], we considered all chromosomal rearrangements (that is, breakpoint clusters) detectable by WGS, including deletions, insertions, tandem duplications, amplifications, inversions and complex variants involving different SV classes (see Methods). Loci harbouring known medulloblastoma-related genes, including *MYCN* (2p24.3), *MYC* (8q24.21) and *SNCAIP* (5q23.2)[8], were readily recovered using this strategy (Fig. 1a). A novel prominent region of

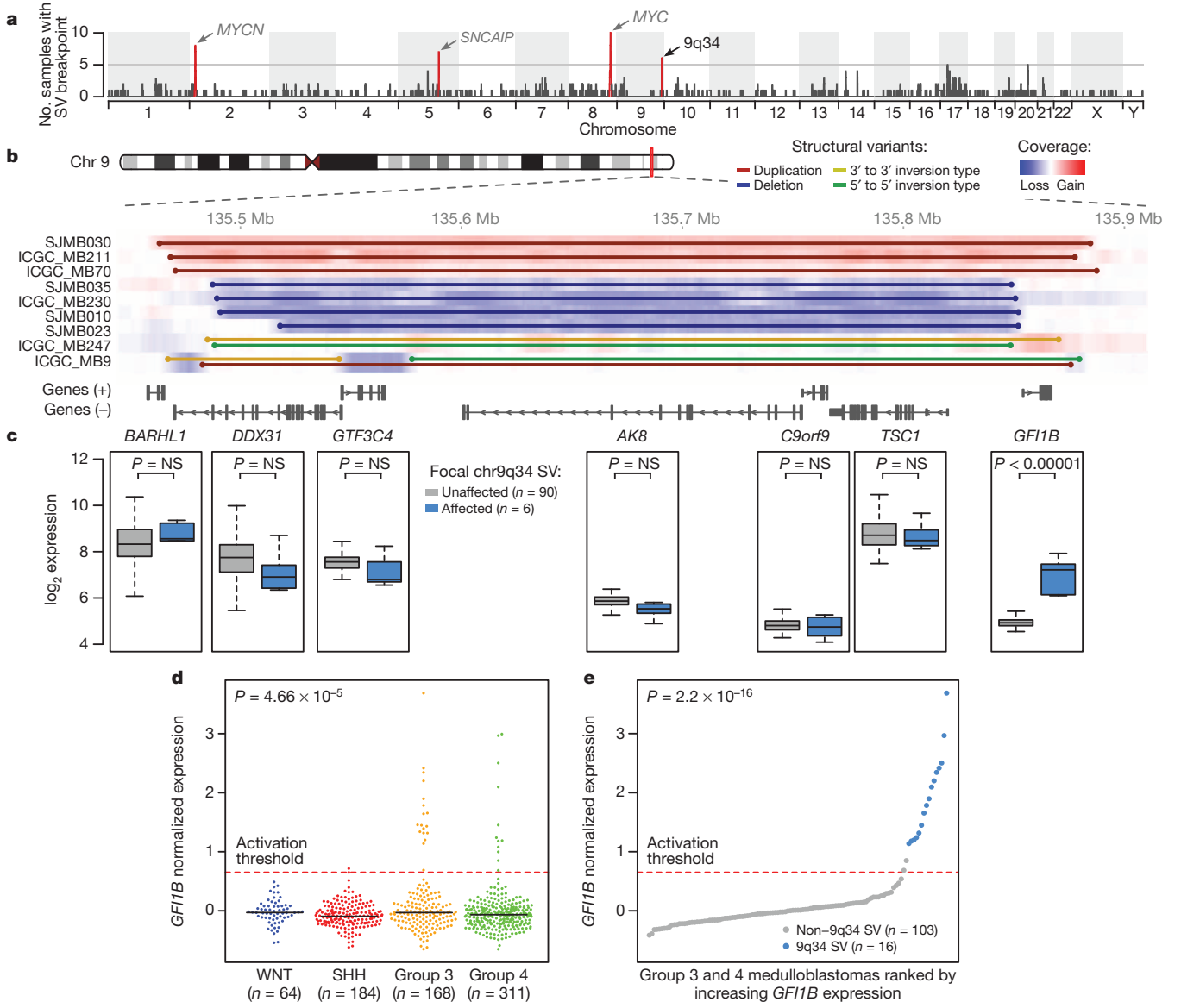


Figure 1 | Recurrent SVs activate the *GFI1B* proto-oncogene in medulloblastoma. **a**, Genome-wide SVs identified by WGS in a discovery cohort of group 3 and group 4 medulloblastomas ($n = 137$). **b**, Summary of SVs affecting a common locus of aberration on 9q34. **c**, Expression box-plots ($n = 96$) for the 7 genes contained within the 9q34 region of interest. Middle bar, median; lower and upper box limits, 25th and 75th percentiles, respectively; whiskers, 1.5 times the interquartile range from the 25th and 75th percentiles. **d**, *GFI1B* expression across medulloblastoma subgroups ($n = 727$). Dashed line delineates the threshold for detectable expression (see Methods). **e**, *GFI1B* expression for group 3 and group 4 medulloblastomas ($n = 119$) coloured according to 9q34 SV state. Dashed line indicates the threshold for detectable expression (see Methods).

interest mapped to chromosome 9q34.13 (Fig. 1a). Further assessment of our entire discovery series identified 9 of 137 (6.6%) cases with evidence of focal SV spanning this region of interest on chromosome 9 (135.46–135.89 megabases (Mb), ~425 kilobases (kb); Fig. 1b).

Instead of showing a predilection for a particular SV type, we observed a range of different SV classes at 9q34, including focal deletion ($n = 4$), tandem duplication ($n = 3$), and complex variants exhibiting inversions and focal deletions ($n = 2$). Examination of microarray-based copy-number data from our recent medulloblastoma genomics study[8] revealed additional evidence for subgroup-specific incidence of recurrent SVs affecting this region (Extended Data Fig. 1).

The region of interest on 9q34 harbours seven known genes (Fig. 1b), including the *TSC1* tumour suppressor gene previously implicated in medulloblastoma[11]. Integration of SV status with sample-matched gene expression data, however, uncovered highly specific transcriptional up-regulation of *GFI1B* in samples harbouring 9q34 SV compared to non-affected counterparts ($P < 0.00001$, Fig. 1c). In contrast, neither *TSC1* nor any of the other remaining candidate genes exhibited a significant difference in expression in this context (Fig. 1c). Analysis of *GFI1B* expression in a large series of medulloblastomas ($n = 727$)[4,8] further substantiated this candidate, confirming restriction of *GFI1B* activation to groups 3 and 4, affecting 10.7% and 3.5% of cases from these subgroups, respectively (Fig. 1d).

To further characterize the relationship between somatic SVs at 9q34 and *GFI1B* transcriptional activation, we sequenced a validation set of 11 group 3 and group 4 medulloblastomas exhibiting *GFI1B* expression, confirming the existence of somatic SVs in 10 of 11 cases (Supplementary Table 2). In just one case (MAGIC_MB179), we failed to detect an underlying SV, suggesting that *GFI1B* overexpression might, in rare instances, be driven by an alternative non-SV-associated mechanism. Collectively, among 119 group 3 and group 4 medulloblastomas for which both WGS and matched expression array data were available, 16 of 18 (89%) *GFI1B*-activated cases displayed a detectable underlying SV (Fig. 1e).

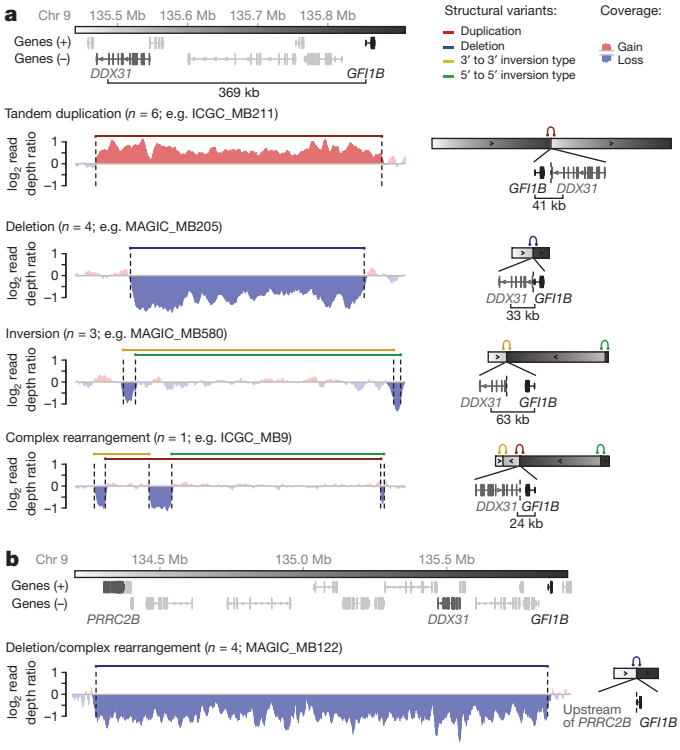


Figure 2 | Summary of recurrent SVs identified in *GFI1B*-activated medulloblastomas. **a**, **b**, Representative WGS coverage plots and associated schematics summarizing the different mechanisms of SV observed in *GFI1B*-activated medulloblastomas.

Importantly, every case showing SV at the 9q34 locus was associated with accompanying activation of *GFI1B* expression.

We next investigated each of the somatic SVs occurring at this locus in further detail to determine the mechanisms of *GFI1B* activation. Interestingly, irrespective of the underlying SV class, in 14 of 18 unique cases these events repositioned *GFI1B* proximal to the terminal sequence of the *DDX31* gene, a region that is normally positioned ~370 kb upstream of the *GFI1B* transcriptional start site (TSS) (Fig. 2a and Supplementary Table 2). Most affected samples annotated in our series juxtapose *GFI1B* within ~40 kb of *DDX31*, with the most distal introns of *DDX31* positioned either upstream or downstream of *GFI1B*, depending on the individual SV. Additionally, a smaller subset of examined cases (4 of 18) was found to exhibit broader deletions (~1.6 Mb) and complex rearrangements spanning a consistent region that starts upstream of the *PRRC2B* gene (chromosome 9, 134.27–134.28 Mb) and extends into the first intron (upstream of the first coding exon) of *GFI1B* (Fig. 2b).

The pattern of observed SVs does not support fusions of the *DDX31* coding sequence or its promoter with *GFI1B* as a common means of gene activation (see Fig. 2a). *DDX31–GFI1B* fusion transcripts were detected in two cases (2 of 4 *GFI1B*-activated medulloblastomas with available RNA-seq data: ICGC_MB9 and ICGC_MB247) but these were predicted to constitute non-functional (antisense or out-of-frame) alternative transcripts, not resulting in *GFI1B* activation (Extended Data Fig. 2).

Active enhancers drive GFI1B expression

The unexpected yet consistent observation of SVs resulting in juxtaposition of *GFI1B* to DNA elements normally located several hundred kilobases upstream suggested that rearrangements of *cis*-acting regulatory elements (such as enhancers) within these regions might be responsible for *GFI1B* activation. *DDX31* is highly expressed in group 3 and group 4 medulloblastomas and shows a correlated expression pattern with its two closest neighbours, *BARHL1* (downstream) and *GTF3C4* (upstream), suggesting that this locus exists in a transcriptionally permissive chromatin state in these medulloblastoma subgroups (Extended Data Fig. 3). To examine this locus and the surrounding region for evidence of enhancer activity, we used chromatin immunoprecipitation coupled with sequencing (ChIP-seq) for H3K27ac and H3K9ac, both known to mark active enhancers[12], in six primary group 3 medulloblastomas (Supplementary Table 1), including three *GFI1B*-activated cases (MAGIC_MB399, MAGIC_MB360 and ICGC_MB9; marked with an asterisk; Fig. 3a). Peak identification of these histone modification data predicted the presence of multiple enhancer clusters in this region, with peak H3K27ac and H3K9ac signals prominently overlapping or found immediately adjacent to the SV breakpoints observed in *GFI1B*-activated cases (Fig. 3a). Such clustering of highly active enhancers, and the overall H3K27 acetylation signal measured for these regions, is consistent with the recently described super-enhancers—regulatory elements associated with the expression of cell identity genes and master transcriptional regulators[13]. Super-enhancer identification (see Methods) performed on our H3K27ac ChIP-seq data suggested the presence of two such elements within the 9q34 region of interest (designated *PRRC2B* super-enhancer and *BARHL1/DDX31* super-enhancer, Fig. 3a and Extended Data Fig. 3).

GFI1B-activated medulloblastomas with SV breakpoints overlapping the inferred *BARHL1/DDX31* super-enhancer (MAGIC_MB360 and ICGC_MB9) showed markedly higher levels of H3K27ac and H3K9ac within this region (compared to non-*GFI1B*-activated samples), indicative of a potential feedback mechanism that increases the local enhancer signal (Fig. 3a). Moreover, H3K27ac and H3K9ac both mark the *GFI1B* locus in these two cases, suggesting 'spreading' of the activating enhancer marks from within the predicted super-enhancer to *GFI1B* consequent to genomic rearrangement (Fig. 3a). Allelic analysis of RNA-seq and enhancer ChIP-seq data for ICGC_MB9 demonstrated that both *GFI1B* expression and the active enhancer signals spanning *GFI1B* originate from the same allele (Fig. 3b), presumably the allele residing on the rearranged haplotype. Whole-genome bisulphite sequencing (WGBS)

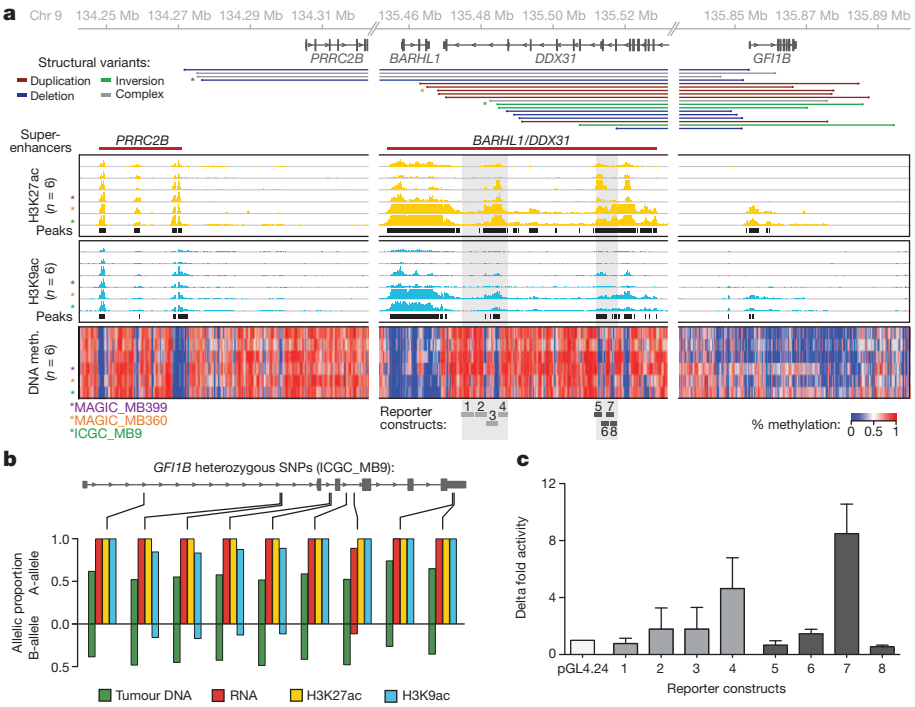


Figure 3 | Recurrent SVs juxtapose *GFI1B* proximal to active enhancers on 9q34. **a**, SV breakpoints ($n = 18$), enhancer-histone marks (H3K27ac and H3K9ac; $n = 6$) and whole-genome DNA methylation data ($n = 6$) overlapping the 9q34 locus in a subset of analysed medulloblastomas. **b**, Allelic analysis of RNA-seq and enhancer ChIP-seq reads overlapping *GFI1B*. **c**, Luciferase reporter activity for regions encompassed within the predicted enhancers indicated in panel **a** compared to empty vector. Error bars represent standard deviation from 3–4 independent experiments.

analysis of the same cases revealed profound DNA hypomethylation overlapping the putative enhancers identified by ChIP-seq, further supporting the accessibility of this chromatin to the transcriptional machinery in group 3 medulloblastomas (Fig. 3a).

To assess directly the capacity of identified enhancer elements to potentiate gene expression, a series of genomic fragments (~2 kb each) tiling two of the constituent enhancers (Fig. 3a, shaded regions) that contribute to the *BARHL1/DDX31* super-enhancer were independently tested for enhancer activity. Assays performed in the D425 group 3 medulloblastoma cell line confirmed robust reporter activity for constructs derived from either region, whereas constructs mapping outside of these peak regions failed to yield any detectable signal (Fig. 3c).

Mutually exclusive GFI1 and GFI1B activation

GFI1B is a paralogue of growth factor independence 1 (*GFI1*), with both genes functioning as SNAG-domain-containing zinc-finger transcriptional repressors essential for a variety of developmental processes, most notably haematopoiesis[14–16]. Importantly, extensive mouse genetics and insertional mutagenesis screens have established *Gfi1* and *Gfi1b* as potent proto-oncogenes in subtypes of leukaemias and lymphomas[17,18]. However, no recurrent somatic SVs affecting *GFI1* or *GFI1B* have been reported in these or any other cancers. Transcriptional analysis showed clear activation of *GFI1* in a subset of medulloblastomas (29 of 724, 4.0%), with expression tightly restricted to group 3 medulloblastomas ($P < 2 \times 10^{-16}$; Fig. 4a). Comparison of *GFI1* and *GFI1B* expression

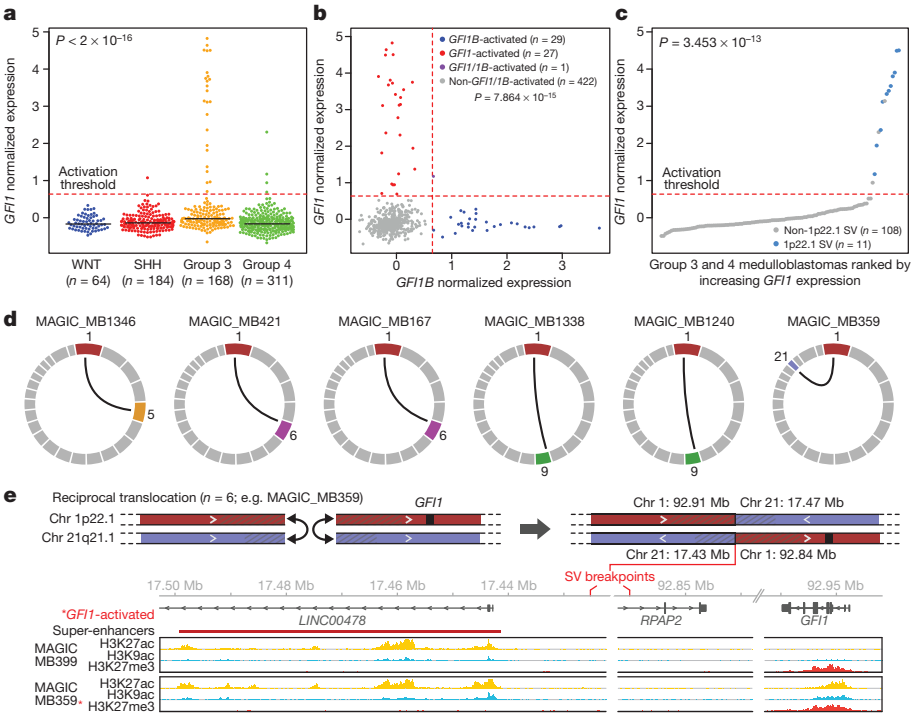


Figure 4 | Mutually exclusive activation of *GFI1* and *GFI1B* in medulloblastoma. **a**, **b**, *GFI1* expression is largely restricted to group 3 (**a**) and is mutually exclusive from *GFI1B* expression (**b**). **c**, *GFI1* expression for group 3 and group 4 medulloblastomas ($n = 119$) coloured according to underlying SV state. Dashed line indicates the threshold for detectable gene expression (see Methods). **d**, Summary of *GFI1* translocations ($n = 6$) observed in group 3 medulloblastoma. **e**, Schematic of the reciprocal t(1:21) translocation observed in *GFI1*-activated MAGIC_MB359. Histone marks overlapping the breakpoints proximal to *GFI1* and the partner chromosome 21 translocation region are shown for a non-*GFI1*-activated case (MAGIC_MB399) and the translocation case (MAGIC_MB359).

among group 3 and group 4 medulloblastomas showed a mutually exclusive pattern of activation ($P = 7.864 \times 10^{-15}$; Fig. 4b), further supportive of their oncogenic roles in medulloblastoma. Collectively, *GFI1* and *GFI1B* expression was observed in 25% and 5% of group 3 and group 4 medulloblastomas, respectively (Extended Data Fig. 4). These findings were validated in an independent series of medulloblastomas ($n = 156$) by immunohistochemical analysis, confirming mutually exclusive GFI1 and GFI1B expression that contributed to 41% and 10% of group 3 and group 4 cases analysed, respectively (Extended Data Fig. 4).

GFI1/GFI1B-expressing medulloblastomas did not form their own discrete subtype within the group 3 subgroup as evaluated by clustering of either gene expression or DNA methylation data (Extended Data Fig. 5). Activation status of GFI1 and GFI1B was associated with patient age in group 3, occurring exclusively in non-infant cases in the gene expression cohort ($P < 0.0001$, chi-squared test; Extended Data Fig. 5). However, no association with patient outcome or other clinical/demographic variables was observed in either our combined gene expression or formalin-fixed paraffin-embedded cohorts (Extended Data Fig. 5).

To investigate whether *GFI1* activation in medulloblastoma is attributable to SV mechanisms similar to those targeting *GFI1B*, we examined the *GFI1* locus in our discovery WGS series of group 3 and group 4 medulloblastomas ($n = 137$) and sequenced an additional validation set consisting of 11 non-overlapping *GFI1*-activated cases. This strategy revealed a diversity of SVs affecting the *GFI1* locus or surrounding genomic regions, including interchromosomal translocations ($n = 6$), tandem duplications ($n = 4$) and a complex rearrangement ($n = 1$), respectively, in medulloblastomas exhibiting *GFI1* expression (Fig. 4d, e and Extended Data Fig. 6). We confirmed somatic SVs in 11 of 14 *GFI1*-activated cases analysed (Fig. 4c), demonstrating that, similar to *GFI1B*, *GFI1* activation is typically associated with an underlying SV.

RNA-seq analysis did not disclose evidence for possible *GFI1* fusion genes (data not shown), suggesting that the detected rearrangements contribute to *GFI1* activation by alternative mechanisms. Observed translocation partners showed no apparent preference for intragenic or intergenic breakpoints (Supplementary Table 2). Overlaying histone ChIP-seq data with translocation breakpoint regions revealed activating enhancer-histone modification states close to the observed breakpoints (Fig. 4e and Extended Data Fig. 7), suggesting translocations of the normally repressed *GFI1* locus into actively transcribed regions as the likely mechanism of gene activation. Importantly, most *GFI1* translocation partners were confirmed to harbour clusters of highly active enhancers consistent with super-enhancers that were situated proximal to sequenced breakpoints, analogous to what we observed for *GFI1B*-activated cases (Fig. 4e and Extended Data Fig. 7).

Despite identifying multiple distinct t(1:6) and t(1:9) translocations (Fig. 4d), the only recurrent SV detected in *GFI1*-activated medulloblastomas was a focal (~6 kb) tandem duplication located ~45 kb downstream of *GFI1*, identified in three *GFI1*-activated samples but not in any other sequenced sample (Extended Data Figs 6 and 7). Enhancer mark ChIP-seq analysis confirmed that this focal region was profoundly marked by the active H3K27ac mark in the context of tandem duplication but not in non-activated group 3 medulloblastomas (Extended Data Fig. 7), suggesting that this region downstream of *GFI1* can, when duplicated, promote its activation.

GFI1 and GFI1B promote medulloblastoma formation *in vivo*

Mouse models of medulloblastoma have given important insights into disease biology[19,20]. Recently, two group 3 models have been described[21,22]. Each of these involves overexpression of *Myc* with *Trp53* loss-of-function—a combination not typically observed in human medulloblastomas, as *MYC* amplification/overexpression (group 3) and *TP53* mutations (WNT and SHH subgroups) occur in different subgroups[5,23].

Group 3 medulloblastoma expression data confirmed significant up-regulation of *MYC* in *GFI1*-activated cases versus non-*GFI1/GFI1B*-activated, subgroup-matched counterparts (Extended Data Fig. 8). Pathway analysis identified MYC target gene sets as being highly enriched in *GFI1/GFI1B*-activated group 3 medulloblastomas (Extended Data Fig. 8). Additionally, co-occurrence of *MYC* amplification and *GFI1* activation was noted in a subset of group 3 medulloblastomas (Extended Data Fig. 8), further suggesting that GFI1 and MYC may cooperate to promote group 3 medulloblastoma. Indeed, *Gfi1* and *Myc* are known to function as synergistic oncogenes and enhance T-cell lymphomagenesis in transgenic mouse models[24,25].

To evaluate whether *GFI1* and *GFI1B* can function as novel medulloblastoma oncogenes, we used an orthotopic transplantation model[21] whereby retroviruses encoding GFI1 or GFI1B were transduced either alone or in combination with viruses encoding MYC into neural stem cells, followed by their transplantation into the cerebella of immunocompromised mice (Fig. 5a). Neither GFI1 nor GFI1B alone was sufficient to promote

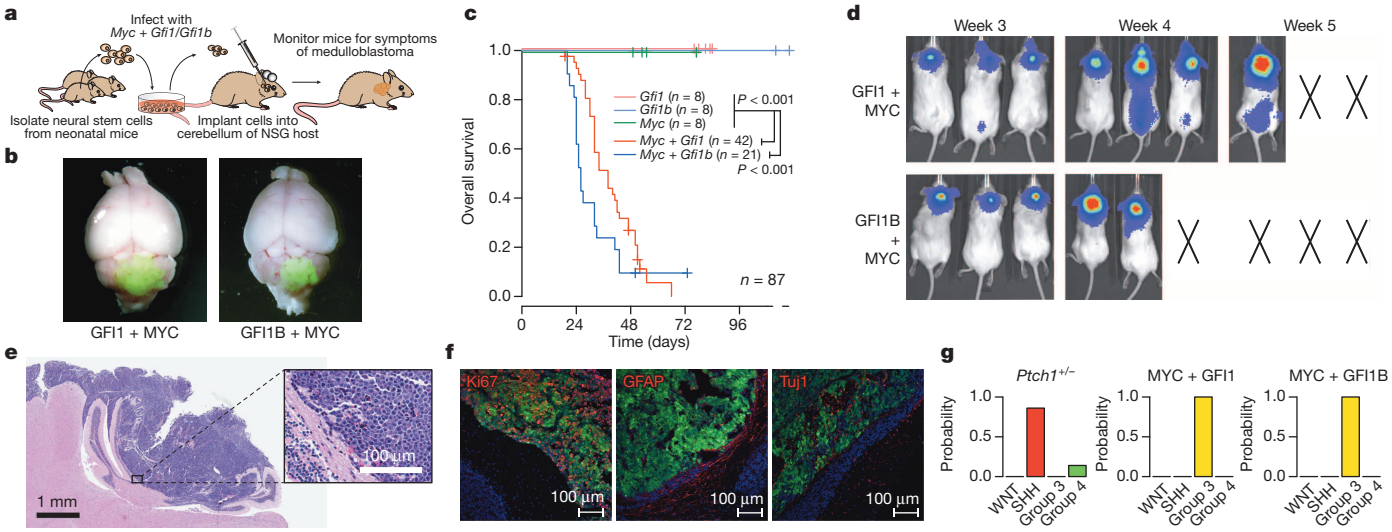


Figure 5 | GFI1 and GFI1B cooperate with MYC to promote medulloblastoma formation in mice. **a**, Strategy for evaluating *Gfi1* and *Gfi1b* as putative medulloblastoma oncogenes. **b**, Whole-mount images of GFP-expressing MYC + GFI1 (left) and MYC + GFI1B (right) tumours. **c**, Survival curves for animals receiving 1×10^5 cells infected with viruses carrying the indicated transgenes. **d**, Bioluminescent imaging of recipient animals at the indicated time points. X symbols denote animals that were euthanized before reaching the indicated time point. **e**, Haematoxylin and eosin staining of cerebellar sections derived from MYC + GFI1 tumour-bearing mice. **f**, Immunofluorescence imaging of MYC + GFI1 tumours stained with the indicated antibodies. **g**, Subgroup probabilities for *Ptch1*$^{+/-}$, MYC + GFI1, and MYC + GFI1B models based on cross-species molecular classification.

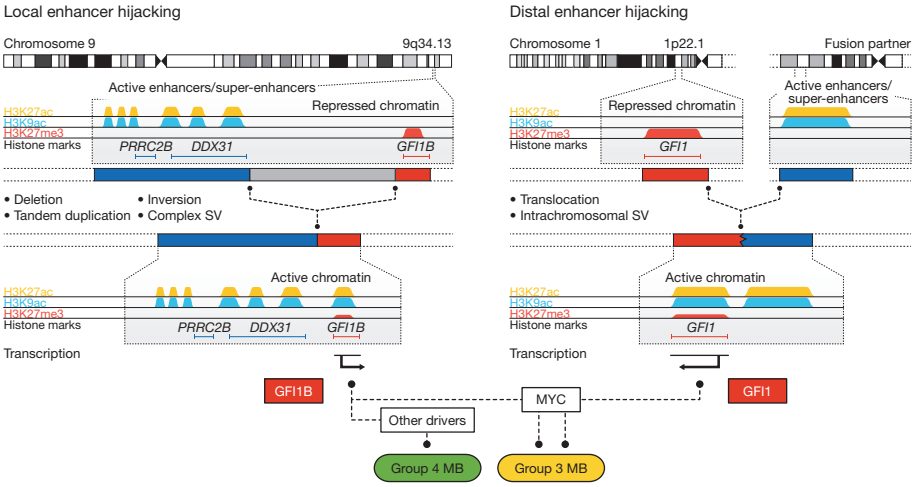


Figure 6 | Summary of inferred mechanisms underlying *GFI1* and *GFI1B* activation in medulloblastoma. Predominant mechanisms of SV and corresponding genomic redistribution of active enhancers, including super-enhancers, observed in *GFI1/GFI1B*-activated medulloblastomas. Activation of *GFI1* and *GFI1B* occurs in a mutually exclusive manner in either group 3 or group 4 and both oncogenes can cooperate with MYC to promote medulloblastoma pathogenesis.

tumorigenesis in this system (Fig. 5c and Extended Data Fig. 9). When combined with MYC, however (which is also insufficient to generate medulloblastoma on its own in this system[21]), both GFI1 (that is, MYC + GFI1) and GFI1B (that is, MYC + GFI1B) rapidly produced highly aggressive cerebellar tumours in nearly all recipient mice within 4–5 weeks ($n = 37$ of 42 and 19 of 21 with median survival time of 38 days and 26 days for MYC + GFI1 and MYC + GFI1B, respectively; Fig. 5b–f).

Cerebellar sections derived from either MYC + GFI1 or MYC + GFI1B recipient mice showed large masses of infiltrating tumour cells with marked cellular pleomorphism, morphologically consistent with large cell, anaplastic (LCA) medulloblastoma (Fig. 5e and Extended Data Fig. 9). LCA histology is significantly more prevalent in group 3 medulloblastoma (~20–25% of cases) than in other medulloblastoma subgroups[6,7]. Metastatic dissemination was also noted in 30–50% of MYC + GFI1 and MYC + GFI1B tumour-bearing mice (Fig. 5d and data not shown), paralleling the high frequency of metastasis seen in group 3 medulloblastoma patients[6,7]. Moreover, immunofluorescence microscopy confirmed that MYC + GFI1 and MYC + GFI1B tumours are highly proliferative and express neuronal but not glial lineage markers (Fig. 5f and Extended Data Fig. 9), consistent with a medulloblastoma-like immunophenotype. Transcriptional profiling and subsequent multidimensional scaling analysis demonstrated a notable similarity between the GFI1- and GFI1B-driven models and confirmed an expression signature consistent with human group 3 medulloblastoma counterparts, suggesting that these models recapitulate molecular characteristics of the human disease (Fig. 5g and Extended Data Fig. 9).

Discussion

Medulloblastoma sequencing studies have highlighted the intertumoral molecular heterogeneity underpinning this malignancy, revealing very few recurrently mutated driver genes, especially in group 3 and group 4 (ref. 5). Here, we have identified somatic genomic rearrangements in association with mutually exclusive *GFI1* and *GFI1B* activation in approximately one-third of group 3 medulloblastomas—now qualifying these oncogenes as the most prevalent drivers in this subgroup (Fig. 6). Moreover, 5–10% of group 4 medulloblastomas harbour analogous SV associated with *GFI1/GFI1B* activation, reinforcing the notion that these subgroups share some biological similarities[6,7].

The verification of diverse SVs in nearly all *GFI1/GFI1B*-activated medulloblastomas analysed in this study has implications for future cancer genome studies. Conventional approaches for identifying genes recurrently targeted by SV in cancer usually focus on minimal common regions of aberration and require that putative gene targets are (at least partially) included within these altered regions[9]. In contrast to the high-level amplifications known to target *MYC*, *MYCN* and other recognized medulloblastoma oncogenes[3,8], *GFI1* and *GFI1B* are not amplified in medulloblastoma. Observations extracted from the current study revealed that (1) a considerable proportion of SVs leading to *GFI1* and *GFI1B* activation do not actually include the target gene itself; and (2) multiple distinct classes of SV including duplication, deletion, inversion and other complex rearrangements can converge on activation of a single target, often without associated gene-level copy-number change. Our findings suggest that similar mechanisms leading to gene deregulation (that is, activation of oncogenic drivers) might have thus far been overlooked in other cancers.

SV-dependent redistribution of *GFI1* and *GFI1B* from regions of transcriptionally silent chromatin to regions populated with active enhancers, such as super-enhancers (Fig. 6), underscores the diversity of interplay between the cancer genome and epigenome[26–28]. Activation of *GFI1* and *GFI1B* seemingly does not rely on specific epigenetic deregulation but rather implicates a form of 'enhancer hijacking' whereby oncogene activation hinges on the appropriation of a physiologically active epigenetic state from proximal or distant loci, including those mapping to other chromosomes. This concept of merging oncogenes with active regulatory elements has long been observed in lymphoid malignancies, where translocations are known to relocate *MYC*, *BCL2* and other oncogenes adjacent to highly active promoter or enhancer loci, most commonly those belonging to the immunoglobulin genes (that is, IgH/IgL loci) or T-cell receptors (that is, TCR-α/β loci)[29]. To the best of our knowledge, this is the first report to substantiate such a phenomenon in brain tumours.

We have discovered a series of highly variable genomic rearrangements leading to oncogene activation in a significant proportion of cases from poorly understood medulloblastoma subgroups, implicating *GFI1* and *GFI1B* as novel oncogenic drivers worthy of pursuit as candidates for molecularly targeted therapy. The patterns of rearrangement associated with *GFI1* and *GFI1B* activation described here have broad-reaching implications for cancer genomics, and warrant the implementation of similar efforts to revisit existing sequencing data using analytical approaches that extend beyond the coding genome. On the basis of our observations, it is tempting to speculate that similar 'enhancer hijacking' may be equally prevalent in other solid cancers.

METHODS SUMMARY

All patient material included in this study was collected after receiving informed consent from the patients and their families. Medulloblastoma samples were collected at first resection, before adjuvant chemotherapy or radiotherapy. Full details on the sequencing cohorts included in this report are summarized in Supplementary Table 1. Medulloblastoma subgroups were assigned using gene expression array data, a custom nanoString CodeSet, DNA methylation profiling, or a combination of the above, as previously described[7,30,31]. WGS and long-range paired-end mapping were performed as described[2,3]. WGBS and DNA methylation analysis was conducted as described[32]. Chromatin extraction, immunoprecipitation and library preparation for ChIP-seq studies was performed using proprietary methods at Active Motif (Carlsbad). H3K27ac and H3K9ac peaks were called using

BayesPeak[33]. Super-enhancers were inferred using the ROSE algorithm with default parameters as described[28]. Affymetrix expression array profiling of human and mouse tumour RNAs was performed at core facilities within the Amsterdam Medical Centre (Amsterdam, Netherlands), German Cancer Research Center (Heidelberg, Germany), and The Hospital for Sick Children (Toronto, Canada). Mouse studies were conducted at the Sanford-Burnham Medical Research Institute and Sanford Consortium for Regenerative Medicine Animal Facilities in accordance with national regulations using procedures approved by the Institutional Animal Care and Use Committees at Sanford-Burnham and the University of California San Diego.

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 12 January; accepted 15 April 2014.
Published online 22 June 2014.

1. Ostrom, Q. T. *et al.* CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. *Neuro-oncol.* **15** (Suppl 2), 1–56 (2013).
2. Jones, D. T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488,** 100–105 (2012).
3. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148,** 59–71 (2012).
4. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488,** 43–48 (2012).
5. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nature Rev. Cancer* **12,** 818–834 (2012).
6. 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).
7. Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29,** 1408–1414 (2011).
8. Northcott, P. A. *et al.* Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488,** 49–56 (2012).
9. Santarius, T., Shipley, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nature Rev. Cancer* **10,** 59–64 (2010).
10. Kim, T. M. *et al.* Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Res.* **23,** 217–227 (2013).
11. Bhatia, B. *et al.* Tuberous sclerosis complex suppression in cerebellar development and medulloblastoma: separate regulation of mammalian target of rapamycin activity and p27 Kip1 localization. *Cancer Res.* **69,** 7224–7234 (2009).
12. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
13. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
14. Hock, H. *et al.* Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature* **431,** 1002–1007 (2004).
15. Person, R. E. *et al.* Mutations in proto-oncogene GFI1 cause human neutropenia and target ELA2. *Nature Genet.* **34,** 308–312 (2003).
16. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev.* **16,** 301–306 (2002).
17. Gilks, C. B., Bear, S. E., Grimes, H. L. & Tsichlis, P. N. Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma lines to IL-2-independent growth following activation of a gene (Gfi-1) encoding a novel zinc finger protein. *Mol. Cell. Biol.* **13,** 1759–1768 (1993).
18. Scheijen, B., Jonkers, J., Acton, D. & Berns, A. Characterization of pal-1, a common proviral insertion site in murine leukemia virus-induced lymphomas of c-myc and Pim-1 transgenic mice. *J. Virol.* **71,** 9–16 (1997).
19. Gibson, P. *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468,** 1095–1099 (2010).
20. 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).
21. Pei, Y. *et al.* An animal model of MYC-driven medulloblastoma. *Cancer Cell* **21,** 155–167 (2012).
22. Kawauchi, D. *et al.* A mouse model of the most aggressive subgroup of human medulloblastoma. *Cancer Cell* **21,** 168–180 (2012).
23. Zhukova, N. *et al.* Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma. *J. Clin. Oncol.* **31,** 2927–2935 (2013).
24. Zornig, M., Schmidt, T., Karsunky, H., Grzeschiczek, A. & Moroy, T. Zinc finger protein GFI-1 cooperates with myc and pim-1 in T-cell lymphomagenesis by reducing the requirements for IL-2. *Oncogene* **12,** 1789–1801 (1996).
25. Schmidt, T. *et al.* Zinc finger protein GFI-1 has low oncogenic potential but cooperates strongly with pim and myc genes in T-cell lymphomagenesis. *Oncogene* **17,** 2661–2667 (1998).
26. Plass, C. *et al.* Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nature Rev. Genet.* **14,** 765–780 (2013).
27. Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153,** 38–55 (2013).
28. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
29. Nambiar, M., Kari, V. & Raghavan, S. C. Chromosomal translocations in cancer. *Biochim. Biophys. Acta* **1786,** 139–152 (2008).
30. Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125,** 913–916 (2013).
31. Northcott, P. A. *et al.* Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathol.* **123,** 615–626 (2012).
32. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* http://dx.doi.org/10.1038/nature13268 (18 May 2014).
33. Cairns, J. *et al.* BayesPeak–an R package for analysing ChIP-seq data. *Bioinformatics* **27,** 713–714 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements For technical support and expertise we thank: the DKFZ Genomics and Proteomics Core Facility; B. Haase, D. Pavlinic and B. Baying (EMBL Genomics Core Facility); M. Knopf (NCT Heidelberg); the Sanford-Burnham Animal Facility and Cell Imaging, Tissue & Histopathology Shared Resource; and the UCSD Flow Cytometry Core Facility. We also thank Active Motif for the preparation of histone ChIP libraries. This work was principally supported by the PedBrain Tumor Project contributing to the International Cancer Genome Consortium, funded by the German Cancer Aid (109252) and by the German Federal Ministry of Education and Research (BMBF, grants 01KU1201A, MedSys 0315416C and NGFNplus 01GS0883). Additional support came from the German Cancer Research Center–Heidelberg Center for Personalized Oncology (DKFZ-HIPO), the EMBL International PhD Programme (T.Z.), Dutch Cancer Foundations KWF (2010-4713) and KIKA (M.Ko.), the US National Institutes of Health, National Center for Research Resources (P41 GM103504; G.D.B.), the CancerSys grant MYC-NET (German Federal Ministry of Education and Research, BMBF, 0316076A), the European Commission (Health-F2-2010-260791), and the Helmholtz Alliance PCCC (grant number HA-305). PAN is a Roman Herzog Postdoctoral Fellow funded by the Hertie Foundation and the DKFZ. R.J.W.-R. is the recipient of a Research Leadership Award from the California Institute for Regenerative Medicine (CIRM LA1-01747) and obtained additional support from the National Cancer Institute (5P30CA030199 and R01 CA159859), and the CureSearch for Children's Cancer Foundation.

Author Contributions P.A.N., C.L., T.Z., A.M.S., D.K., L.A.E., W.W., A.W., S.St., L.S., H.S.-C., L.L., F.K., J.F., B.R., S.Sc., N.D., S.Wo., T.R., C.C.B., P.v.S. and A.K. performed and/or coordinated experimental or technical work. P.A.N., T.Z., S.E., D.J.H.S., V.H., M.Z., S.Z., G.D.B., N.J., I.B., C.D.I., G.Z., J.E., R.Vo., J.K. and J.O.K. performed and/or coordinated data analysis. M.Re., F.M.G.C., S.V., M.Ry., E.T., P.H., E.S., A.D., P.S., J.S., K.Z., D.Su., M.U.S., M.E., H.L.G., G.W.R., A.G., M.M., K.v.H., S.R., T.P., W.S., R.J.G., A.K. and M.D.T. contributed data, provided reagents, or patient materials. P.A.N., C.L., T.Z., S.E., D.J.H.S., V.H., D.St., D.T.W.J., M.K., S.Z., H.-J.W., R.J.G., M.D.T., P.Li., J.O.K., R.J.W.-R. and S.M.P. prepared the initial manuscript and display items. P.A.N., G.D.B., S.Wi., B.B., C.L., M-L.Y., U.D.W., C.v.K., R.V., G.R., A.E.K., A.v.D., O.W., R.E., P.Li., J.O.K., R.J.W.-R. and S.M.P. provided project leadership. P.A.N., J.O.K., R.J.W.-R. and S.M.P. co-conceived and led the study. P.Li., J.O.K., R.J.W.-R. and S.M.P are co-senior authors of this study.

Author Information Short-read sequencing data have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) hosted by the EBI, under accession number EGAS00001000215. 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 the paper. Correspondence and requests for materials should be addressed to S.M.P. (s.pfister@dkfz-heidelberg.de), R.J.W.-R. (rwreya@sanfordburnham.org), J.O.K. (korbel@embl.de) or P.Li. (m.macleod@dkfz-heidelberg.de).

METHODS

General statistical methods. All statistical tests were performed in the R Statistical Environment (R version 3.0.0) unless otherwise specified. The Kolmogorov–Smirnov test was used to compare candidate gene expression in chr9q34 SV cases to non-SV cases. Differential expression of *GFI1* and *GFI1B* across medulloblastoma subgroups was calculated using ANOVA. Enrichment of underlying locus-specific SVs in *GFI1/GFI1B*-expressing cases was calculated using Fisher's exact test. Mutual exclusivity of *GFI1* and *GFI1B* expression in group 3 and group 4 medulloblastomas was determined using Fisher's exact test. Survival analyses were performed in GraphPad Prism 5 using the log-rank (Mantel-Cox) test to compare survival differences between groups.

Sample collection and preparation. An Institutional Review Board ethical vote (Ethics Committee of the Medical Faculty of Heidelberg) was obtained according to ICGC guidelines (http://www.icgc.org), along with informed consent for all participants. No patient underwent chemotherapy or radiotherapy before surgical removal of the primary tumour. Tumour tissues were subjected to neuropathological review for confirmation of histology and for tumour cell content >80%. Analytes were isolated as previously described[2]. Cells were cultured at 37 °C with 5% CO_2. D425_Med MB cells (D425; a gift from Professor Darrell D. Bigner) were cultured in DMEM with 10% FCS (Life Technologies) and regularly authenticated and tested for mycoplasma (Multiplexion, Heidelberg, Germany). Validation samples for WGS were obtained in accordance with the Research Ethics Board at The Hospital for Sick Children (Toronto, Canada).

High-throughput sequencing data generation. *Short-insert paired-end sequencing.* Samples were processed and libraries sequenced as previously described[2].

Medulloblastoma and germline WGS data[4] generated by the Paediatric Cancer Genome Project (http://explore.pediatriccancergenomeproject.org/) was accessed from The European Genome-phenome Archive (Study ID EGAS00001000347). The original alignments of this WGS data were performed against either reference genome hg18 or hg19. For comparability with our data, the alignment files in hg18 have been converted to FASTQ files using Picard tools (http://picard.sourceforge.net) providing the 'SamToFastq' option. For the alignment of the FASTQ files, the same reference genome as used in the creation of the original hg19 BAM files has been used along with BWA for alignment and Picard for merging and duplicate read filtering.

Long-range paired-end sequencing data generation. Long-range (or 'Mate-pair') DNA library preparation was carried out as previously described[2] or using the newer Nextera Mate Pair Sample Preparation Kit (Illumina). In brief, 4 μg of high-molecular-mass genomic DNA were fragmented by the Tagmentation reaction in 400 μl, followed by the strand displacement and AMPure XP (Agencourt) clean-up reaction. Samples were size selected to 4–6 kb with a gel step following the Gel-Plus path of the protocol. 300–550 ng of size-selected DNA were circularized in 400 μl for 16 h at 30 °C. The library was then constructed after an exonuclease digestion step to get rid of remaining linear DNA, fragmentation to 300–700 bp with a Covaris S2 instrument (LGC Genomics), binding to streptavidin beads and Illumina Truseq adaptor ligation. Final library was obtained after PCR for 1 min at 98 °C, followed by 9 cycles of 30 s at 98 °C, 30 s at 60 °C, 1 min at 72 °C and a final 5 min at 72 °C step. Deep sequencing was carried out with the Illumina HiSeq2000 (2 × 101 bp) instrument to reach an average physical coverage of 20–30×.

ChIP sequencing. Chromatin extraction, immunoprecipitation and library preparation for ChIP-seq were performed at Active Motif (Carlsbad, CA) according to proprietary methods. Briefly, 15 μg of chromatin were used as input for ChIP with ChIP-grade antibodies recognizing H3K27ac (AM#39133, Active Motif), H3K9ac (AM#39918, Active Motif), or H3K27me3 (#07-449, Millipore). Libraries were sequenced on the Illumina HiSeq 2000 platform using 2 × 101 cycles according to the manufacturer's instructions.

Whole-genome bisulphite sequencing. Whole-genome bisulphite library preparation was carried out as recently described[34], with modifications to a previously published protocol[35]. In brief, 5 μg of genomic DNA were sheared using a Covaris device (Covaris Inc.). After adaptor ligation, DNA fragments with insert lengths of 200–250 bp were isolated using an E-Gel electrophoresis system (Life Technologies) and bisulphite converted using the EZ DNA Methylation kit (Zymo Research). PCR amplification of the fragments was performed in six parallel reactions per sample using the FastStart High Fidelity PCR kit (Roche). Library aliquots were then pooled per sample and sequenced on an Illumina HiSeq 2000 machine.

RNA sequencing. RNA quality control was performed using the 2100 Bioanalyzer platform (Agilent). RNA sequencing libraries were prepared using the TruSeq stranded protocol with Ribo-Zero Gold (Illumina) and sequenced on the Illumina HiSeq 2000 platform with 2 × 51 cycles according to the manufacturer's instructions.

High-throughput sequencing data analysis. *Whole-genome sequencing.* Short-insert WGS data was analysed as previously described[2]. Long-range paired-end sequencing reads were aligned to the hg19 assembly of the human reference genome using the Illumina-provided alignment software (ELAND, version 2).

Structural variant discovery and filtering. Deletions, tandem duplications, inversion, translocations, as well as complex rearrangements resulting in the corresponding paired-end signatures were inferred using DELLY v0.0.11 (ref. 36). We considered all those predictions as somatic that were not present in a set of 1,000 Genomes Project (1000GP; http://1000genomes.org)[37] samples corresponding to germline samples taken from normal healthy individuals. Specifically, we used DELLY to infer variants in 1,106 healthy samples belonging to phase 1 of the 1000GP. Furthermore, we inferred variants in the germline samples belonging to the studied tumours. For a given tumour sample, we considered all those variants as somatic that were present neither in any of the 1000GP samples nor in any of the additional germline samples. Two SVs were considered as identical if their start and end coordinates differed by less than 5.0 kb (approximate insert size of a long-range paired-end library) and if their reciprocal overlap was larger than 50%. Variants that were present in the control samples were either true germline variants or represented artefacts caused by misalignment of reads (for example, due to inaccuracies within the human reference genome). To consider a variant prediction as high-confidence we further required at least four supporting read pairs with a minimum median mapping quality of 20 for each event to exclude false-positive predictions caused by randomly mapping low-quality reads.

Region identification. We divided the human reference genome into overlapping 1-Mb windows (100-kb offset). For each window, we counted the number of samples with at least one SV breakpoint in the given region (based on short-insert as well as long-insert paired-end sequencing data). Only focal high-confidence SV predictions were used in this analysis (20 kb to 10 Mb in size). Windows affected in at least five samples were investigated manually.

Copy-number analysis. We determined the number of sequencing reads per non-overlapping genomic window of size 250 bp (high-coverage paired-end data) or 1,000 bp (low-coverage long-range paired-end data) for tumour samples with chr9q34 or chr1p22 SV and their corresponding controls. Tumour values were normalized by the ratio of read counts between tumour and controls within a 500-kb region. Subsequently, for each window, the $\log_2$ ratio between normalized tumour and control counts was determined. These values were averaged along a sliding window of 5 kb (short-insert paired-end data) or 10 kb (long-range paired-end data). For tumour samples without a matching control sample, the control of ICGC_MB230 was used.

ChIP sequencing. Histone ChIP-seq data for H3K27ac, H3K9ac and H3K27me3 was processed by the Illumina analysis pipeline (version 1.8.3) and aligned to the Human Reference Genome (assembly hg19, GRCh37) using BWA version 0.5.9-r16 (ref. 38). Putative PCR duplicates were filtered using Picard MarkDuplicates (http://picard.sourceforge.net). For downstream analyses, we generated whole-genome coverage tracks with reads normalized to all properly paired reads (RPM, paired-end reads/fragments per million). We used igvtools version 2.2.2 (http://www.broadinstitute.org/igv/igvtools) and the non-default parameter–pairs and a window size of 25. For peak-calling of histone marks, ChIP-seq data for each histone modification (H3K27ac or H3K9ac) was used to generate individual BED files for analysed samples using BEDTools[39]. Individual BED files were then combined for each histone modification and peaks were called using the Bioconductor BayesPeak package in R[33]. Super-enhancers were identified using the ROSE algorithm with default parameters (stitching distance of 12,500 bp and promoter exclusion region of ±2,000 bp around TSS)[28]. Briefly, peaks called via BayesPeak were used as constituent enhancers to run the algorithm and super-enhancers were called by ranking of H3K27ac signal at stitched constituent enhancers.

Whole-genome bisulphite sequencing. WGBS sequencing data was analysed using methylCtools (Hovestadt *et al.*, manuscript in preparation). In brief, methylCtools builds on BWA and adds functionality for aligning bisulphite treated DNA to a reference genome in a similar manner as described previously[40]. Sequencing reads were adaptor-trimmed using SeqPrep (https://github.com/jstjohn/SeqPrep) and translated to a fully C-to-T converted state. Alignments were performed against a single index of both *in silico* bisulphite-converted strands of the human reference genome (hg19, NCBI build 37.1) using BWA version 0.6.1-r104 (ref. 38) and the non-default parameters -q 20 -s. Previously translated bases were translated back to their original state, and reads mapping antisense to the respective reference strand were removed. Putative PCR duplicates were filtered using Picard MarkDuplicates (http://picard.sourceforge.net). Non-conversion rates were estimated on the basis of lambda phage genome spike-ins. Single base pair methylation ratios (beta-values) were determined by quantifying evidence for methylated (unconverted) and unmethylated (converted) cytosines at all CpG positions. Only properly paired or singleton reads with mapping quality of ≥1 and bases with Phred-scaled quality score of ≥20 were considered. To account for population variability, we filtered CpGs for which more than 25% of reads at a given position (on either strand) were not supportive of this CpG being in fact a CpG in the sample being analysed. Subsequently, information from both strands was combined and CpGs with coverage less than five reads were set as NA.

RNA sequencing. Demultiplexed FASTQ files were generated using the Bcl2FastQ conversion software (Illumina, version 1.8.4). The resulting sequencing reads were aligned to the human genome reference build hg19 (version human_g1k_v37 – 1,000 Genomes Project Phase 1) using BWA version 0.5.9-r16 (ref. 38) with default parameters. Only the chromosomes 1–22, X, Y and M were used for the mapping. Read coverage plots were prepared using the UCSC Genome Browser showing the number of aligned reads for each genomic position per million mapped reads (RPM) with mapping quality MAPQ >1. The sequencing reads were also used as input for the TopHat2-Fusion algorithm[41] for detection of gene fusion breakpoints.

Allelic analysis. Germline SNPs were determined using Samtools and BCFtools. For each SNP, the number of reads in the tumour DNA-, RNA-, and ChIP-seq data supporting the alternative or the reference allele were counted using Samtools mpileup. Only bases with phred score >20 were considered. Only heterozygous SNPs covered by at least 4 sequencing reads in each data set were included in the final summary.

PCR and Sanger sequencing validation of structural variants. PCR experiments were performed as follows: 10 ng of genomic DNA were used with the SequalPrep Long PCR Kit (Invitrogen) in 20 µl volumes using the following PCR conditions in a MJ Mini thermocycler (BioRad): 94 °C for 3 min, followed by 10 cycles of 94 °C for 10 s, 62 °C for 30 s and 68 °C for 6 min and 25 cycles of 94 °C for 10 s, 60 °C for 30 s and 68 °C for 7 min, followed by a final cycle of 72 °C for 10 min. PCR products were analysed on a 1% agarose gel stained with Sybr Safe Dye (Invitrogen). Gel-extracted bands using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) were capillary sequenced at GATC Biotech AG to analyse SV breakpoints.

Expression array processing and data analysis. *General array processing*. For gene expression array profiling of human medulloblastomas and normal cerebellar controls, high-quality RNAs were processed and hybridized to either (i) the Affymetrix Gene 1.1 ST array at The Centre for Applied Genomics (TCAG, Toronto, Canada) or (ii) the Affymetrix U133 Plus2.0 expression array at the Microarray Department of the University of Amsterdam (Amsterdam, the Netherlands). Sample library preparation, hybridization, and quality control were performed according to protocols recommended by the manufacturer. The CEL files were quantile normalized using Expression Console (v1.1.2; Affymetrix, USA) and signal estimates determined using the RMA algorithm.

Mouse medulloblastomas, non-neoplastic cerebellar stem cells (NSCs), and normal mouse cerebella were analysed using the Affymetrix Mouse Genome 430 2.0 expression array according to the manufacturer's instructions at the DKFZ Genomics and Proteomics Core Facility (Heidelberg, Germany). The CEL files were quantile normalized using Expression Console (v1.1.2; Affymetrix, USA) and signal estimates determined using the RMA algorithm.

Merging of expression array platforms. Gene expression array data generated using the Affymetrix Gene 1.1 ST array and U133 Plus2.0 array platforms was merged in order to generate a combined series that would facilitate more streamlined downstream analyses. For each platform, a contrast value per gene was calculated by subtracting the mean expression of that gene across all samples hybridized on that platform from each individual sample (see formula below), and the resulting contrast values of the two platforms were then combined.

$$\text{Contrast}_{\text{geneA in SampleX}} = \text{Gene}_{\text{A}}\ \text{expression in Sample}_{\text{X}} - \text{mean (Gene}_{\text{A}}\ \text{expression)}$$

This method minimized possible batch effects existing between the two array platforms and allowed for downstream analyses containing the combined series.

Identification of GFI1- and GFI1B-activated medulloblastomas. After combining the gene expression data for the two expression array platforms, for both *GFI1* and *GFI1B*, expression values were modelled by fitting two normal distributions to the data using the R package 'mclust'[42]. With a P value cut-off of $P < 0.0001$, threshold expressions for *GFI1* and *GFI1B* were identified as contrast scores of 0.64 and 0.65, respectively. Samples having expression greater than or equal to the thresholds were called as *GFI1*- or *GFI1B*-activated.

Pathway analysis. Medulloblastoma expression array profiles (Affymetrix Gene 1.1 ST) were used to fit a linear model for each gene using group 3 status, *GFI1* expression, and *GFI1B* expression as covariates. The R package 'limma' was used to perform these fits. The average rank of the statistical significance of the *GFI1* and *GFI1B* coefficients was used to perform a Mann–Whitney *U*-test for a given collection of genes (the null hypothesis being that the genes in a gene set are not ranked any higher than those which are not). In cases where multiple probes matched a single gene, the higher-ranking probe was used. The gene sets contained in the c2-c6 collections from the Molecular Signatures Database (MSigDB) were tested[43]. The *P* values obtained for each gene set in a collection underwent a Benjamini–Hochberg correction to correct for multiple testing.

Cross-species comparisons of human and mouse medulloblastomas. Human medulloblastoma samples were analysed on the Affymetrix U133Plus2 platform and normalized by the MAS5 algorithm. Mouse tumours were analysed on the Affymetrix Mouse Genome 430 2.0 platform and similarly normalized by MAS5 using the 'affy' (v1.38) package within the R Statistical Environment (v 3.0.2). Human and mouse expression profiles were matched by homologues using official gene symbols and filtered for genes that exhibit conserved expression across 32 matched human and mouse tissues[44] as determined by Pearson correlation tests with multiple hypotheses correction using the Benjamini–Hochberg false discovery rate method (FDR <0.1). Mouse adult cerebellum, fetal cerebellum and *Ptch1*$^{+/-}$ medulloblastoma samples were matched against the most similar human adult cerebellum, fetal cerebellum and SHH medulloblastoma samples, respectively, by Pearson correlation of expression profiles. Subsequently, these matched sample pairs were designated as replicate samples for cross-platform calibration by the Linear Cross-Platform Integration of Microarray Data (LTR) algorithm[45] as implemented in the 'LTR' package (v 1.0.0).

Following gene filtering and expression calibration, the human and mouse expression profiles were combined and analysed by multidimensional scaling. The first two dimensions were disregarded, as expression differences between human and mouse dominated them. The third dimension was identified as the medulloblastoma subgroup spectrum, as the coordinate values discriminate samples from different human medulloblastoma subgroups. Using this molecular subgroup spectrum, mouse samples were classified using a Bayesian classifier initialized with a uniform prior. The posterior probabilities were calculated as the normalized product of the prior and the likelihood of Gaussian distribution parameters with mean and variance estimates from each of the human medulloblastoma subgroups.

Luciferase enhancer assays. Candidate enhancer regions were amplified by PCR using the primer sets listed below and cloned into the pGL4.24[luc2P/minP] Vector (Promega) containing a multiple cloning region for insertion of a response element of interest upstream of a minimal promoter and the luciferase reporter gene, *luc2P*.

Primer sequences. The following primer sequences were used: BARHL1/DDX31 SE: region 1, forward primer GAAGGTACCATCCCCACTTCCTGGTAAGG, reverse primer GAAGGTACCTTCTTTGGGGAAATCATTGG; BARHL1/DDX31 SE: region 2, forward primer GAAGGTACCCTGAGAGTTTGGGCTTCAGG, reverse primer GAAGGTACCGCCTGCCAATTTTTATGTGG; BARHL1/DDX31 SE: region 3, forward primer GAAGGTACCTGTCTCCAAGTGTGGTTTCG, reverse primer GAAGGTACCTGAGCAGGGGATTTAACAGG; BARHL1/DDX31 SE: region 4, forward primer GAAGGTACCAGGGGTATCGTGGTCTTGG, reverse primer GAAGGTACCGGAAAGCACACGTGAAAAGG; BARHL1/DDX31 SE: region 5, forward primer AAGGTACCAGTGTGTCAACCACCCACAA, reverse primer AAGAGCTCGGATGGAGTGCAGTCACCTT; BARHL1/DDX31 SE: region 6, forward primer AAGGTACCGAAATTCCCCAGGAGGAGAG, reverse primer AAGAGCTCCCAATGCACCCTACGTTTCT; BARHL1/DDX31 SE: region 7, forward primer AAGGTACCCACCCAGCTCTTCTCCAGTC, reverse primer AAGAGCTCCTCCTCCAGCACAACACTGA; BARHL1/DDX31 SE: region 8, forward primer AAGGTACCCTGTAGCCTCGACCTTCTGG, reverse primer AAGAGCTCCCTTCAGAGCACTTGTAGGAGAA.

For evaluation of enhancer activity, D425 group 3 medulloblastoma cells were plated on 6-well plates. At 50% confluence, cells were transfected in triplicate with 2.25 µg of the pGL4.24 reporters carrying the *DDX31* DNA fragments plus 0.25 µg of phRL-TK encoding Renilla luciferase. Two days post-transfection, the cells were harvested, followed by measurement of luciferase activities using the Dual-Glo Luciferase Assay System (Promega). As a control, the pGL4.24 empty vector was included for calibration of activity obtained with the experimental constructs. The luminescence of the Firefly Luciferase was normalized to the Renilla Luciferase signal obtained from the phRL-TK vector and data was presented as the mean delta-fold activity (Firefly Luciferase/Renilla Luciferase) of experimental transfectants compared to the pGL4.24 empty vector transfectants.

Immunohistochemical and FISH analysis of human medulloblastoma samples. Immunohistochemistry (IHC) and FISH were performed on formalin-fixed paraffin-embedded MB sections as previously described[7]. Monoclonal GFI1 (clone 3G8, Sigma) and polyclonal GFI1B (HPA007012, Sigma) antibodies were used at working dilutions of 1:100 with an incubation time of 1 h @ 32°C using the Ventana protocol cc1.

Mouse models. *Animals*. C57BL/6 mice (males and females) were used as a source of cerebellar stem cells and immunocompromised (NOD-*scid* IL2Rgamma^{null}, NSG) female mice were used as transplantation hosts. Mice were bred and maintained at the Sanford-Burnham and Sanford Consortium Animal Facilities. Experiments were performed in accordance with national regulations using procedures approved by the Institutional Animal Care and Use Committees at Sanford-Burnham and the University of California San Diego. No a priori calculations related to sample size were performed. No specific randomization or blinding was performed.

Isolation of cerebellar stem cells. Cerebellar stem cells were isolated as previously described[46]. Briefly, neonatal (p4-p6) cerebella from wild-type C57BL/6 mice were dissected and enzymatically dissociated into single cell suspension. Cells were

subjected to Percoll fractionation (GE Healthcare Life Sciences 17-0891-02) and stained (anti-mouse CD133 PE, eBioscience 12-4301-82) and sorted for the Prominin1+ (Prom1$^+$) population (approximately 3–4% of cells).

Retroviral constructs. Retroviruses employed in this study included MSCV-c-Myc^{T58A}-IRES-GFP[21], MSCV-c-Myc T58A-IRES-Luc, MSCV-Gfi1-IRES-GFP, MSCV-Gfi1-IRES-Luc, MSCV-Gfi1b-IRES-GFP and MSCV-Gfi1b-IRES-Luc. To create the *Gfi1* and *Gfi1b* viral constructs, cDNAs were PCR-amplified and cloned into MSCV-IRES-GFP and MSCV-IRES-Luc. *Gfi1* and *Gfi1b* were PCR-amplified from pCMV6-Gfi1 (MC208542, OriGene) and pCMV6-Gfi1b (MC201880, OriGene), respectively, and EcoRI and XhoI restriction sites were added to the cDNA ends.

Gfi1 PCR primers: forward primer GAATTCACCATGCCGCGCTCATTCC TGGTC, reverse primer CTCGAGTCATTTGAGTCCATGCTGACTCTC. *Gfi1b* PCR primers: forward primer GAATTCACCATGCCACGGTCCTTTCTAGTG, reverse primer CTCGAGTCACTTGAGATTGTGTTGACTCTC.

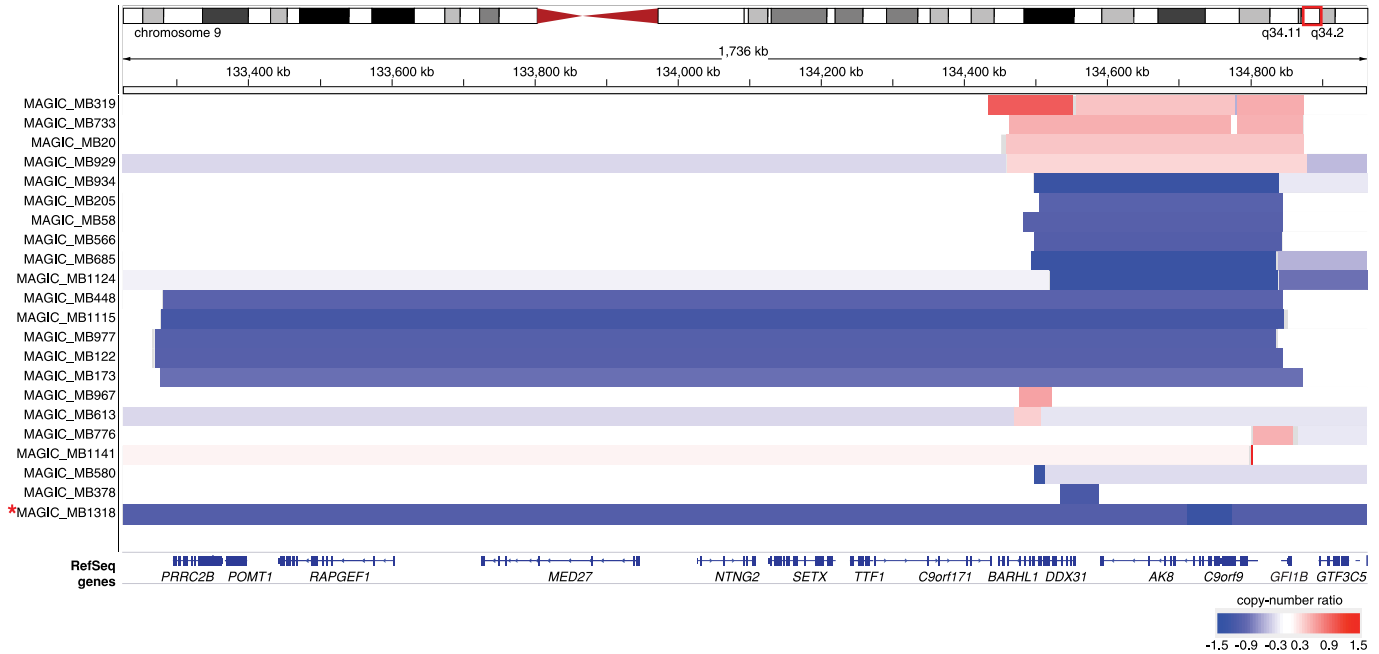
The PCR-amplified products were blunt-end-ligated into pJET1.2 (CloneJET PCR Cloning Kit, Thermo Scientific K1231) and then cut with EcoRI and XhoI. The sticky-ended fragments were then ligated into the EcoRI/XhoI-digested MSCV-IRES-GFP and MSCV-IRES-Luc vectors.

Orthotopic transplantation and tumour formation. Before transplantation, cerebellar stem cells (Prom1$^+$ cells) were infected with retroviruses encoding *Myc^{T58A}* and *Gfi1*- or *Gfi1b* for 20 h. Next, 1×10^5 transduced cells were re-suspended in Neurocult NSC Basal medium (Stem Cell Technologies, cat #05700) with Neurocult NSC Proliferation Supplement (Stem Cell Technologies, cat #05701) and injected into the cerebella of NSG mice (6–8 weeks old) using a stereotactic frame equipped with mouse adaptor (David Kopf Instruments). Animals were monitored weekly and euthanized when they showed signs of medulloblastoma. At time of euthanasia, brains were removed for tumour dissection and dissociation or for embedding and sectioning.

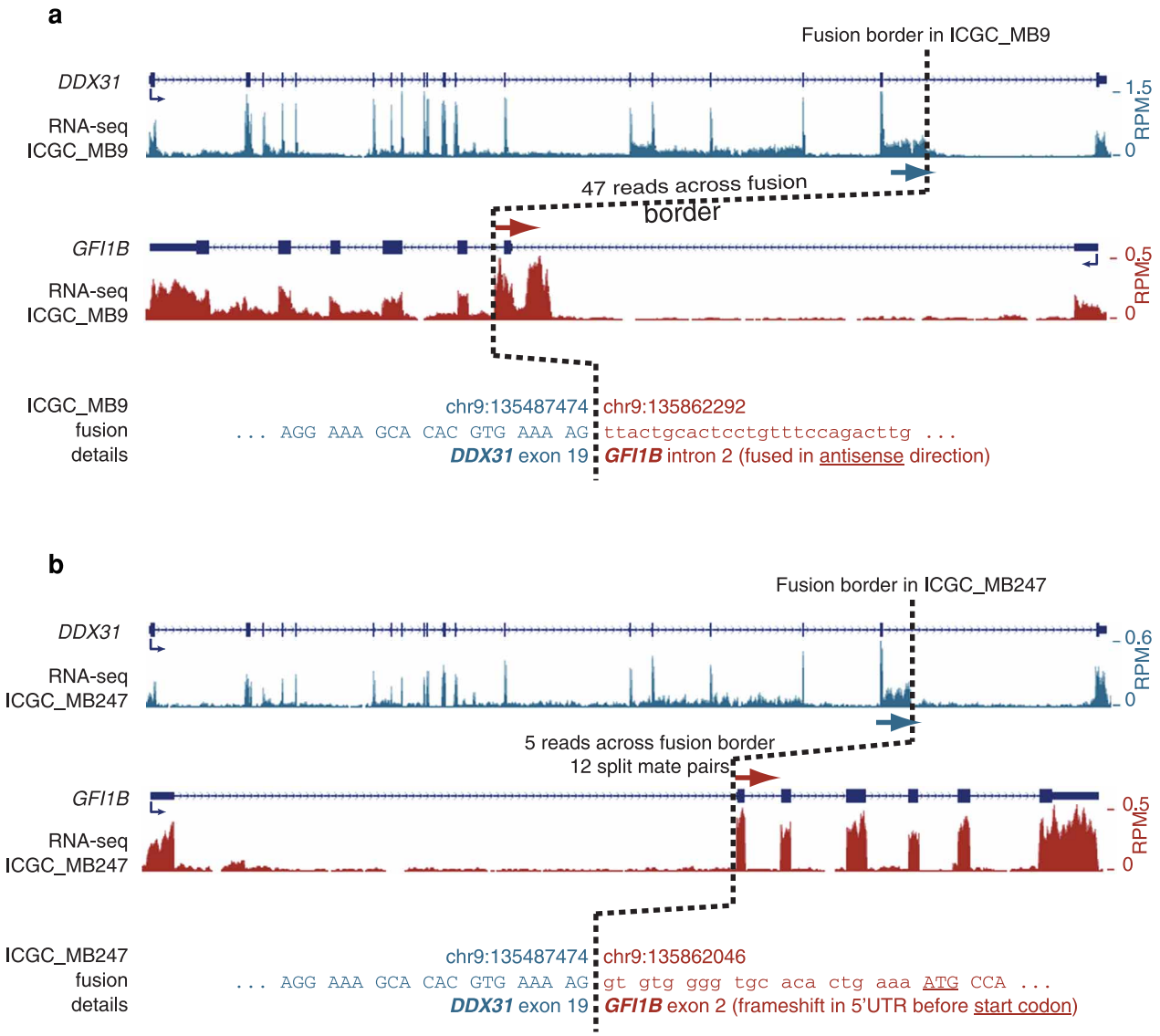
Tissue sectioning and staining. Mouse brains were fixed with 4% paraformaldehyde and embedded in either paraffin or OCT. Samples for histological analysis were paraffin-embedded, sectioned and stained with H&E by the Sanford-Burnham Histopathology Core Facility. Samples frozen in OCT were sectioned using a Leica CM3050S cryostat. Cryosections were stained overnight with primary antibodies against proliferation (anti-Ki67, Abcam ab15580) and lineage markers (anti-GFAP, Novus Biologicals NB300-141; anti-β3-Tubulin, Cell Signaling 5568) and stained for 1 h with fluorescent secondary antibodies (Alexa Fluor 568 Donkey Anti-Rabbit IgG, Invitrogen A10042). Sections were then counter-stained with DAPI (Cell Signaling 4083), mounted using Fluoromount G (Southern Biotech #0100-01), and imaged on a confocal (Zeiss LSM700) fluorescent microscope.

In vivo bioluminescent imaging. Mice were anaesthetized with 2.5% isoflurane and given intraperitoneal injections of 150 ng g^{-1} D-Luciferin (Caliper Life Sciences, cat 12279). Five minutes after injection, animals were imaged using the Xenogen Spectrum (IVIS-200) imaging system (Sanford-Burnham and Sanford Consortium Animal Facilities).

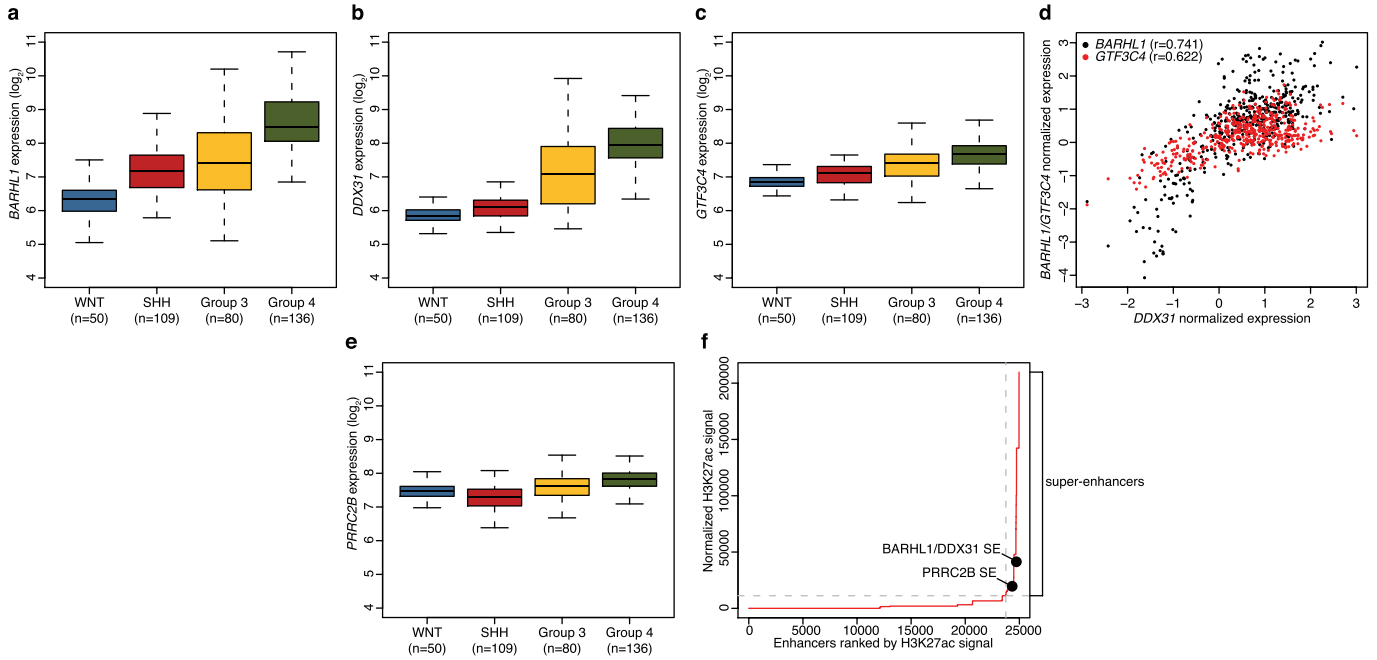
34. Richter, J. *et al.* Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nature Genet.* **44,** 1316–1320 (2012).
35. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).
36. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28,** 333–339 (2012).
37. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
39. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
40. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).
41. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12,** R72 (2011).
42. Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. & Ruzzo, W. L. Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17,** 977–987 (2001).
43. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
44. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101,** 606–6067 (2004).
45. Boutros, P. C. LTR: Linear cross-platform integration of microarray data. *Cancer Inform.* **9,** 197–208 (2010).
46. Lee, A. *et al.* Isolation of neural stem cells from the postnatal cerebellum. *Nature Neurosci.* **8,** 723–729 (2005).



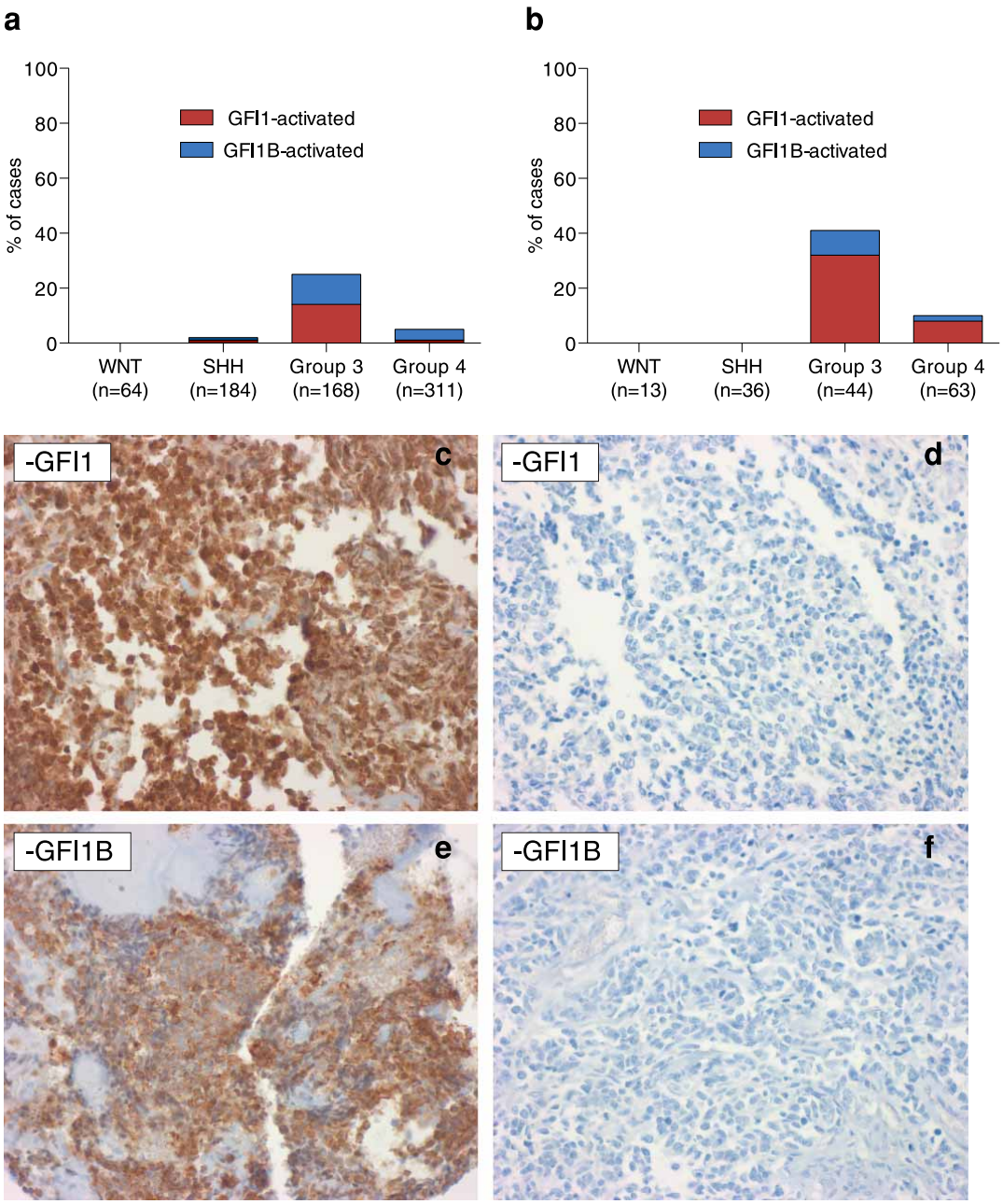
Extended Data Figure 1 | Recurrent somatic copy-number aberrations target a common region on 9q34. Affymetrix SNP6 copy-number output for 22 primary medulloblastomas from the published[8] MAGIC series exhibiting focal somatic copy-number aberrations within the 9q34 region of interest defined by WGS in the current study. Of the affected samples, medulloblastoma subgroup information was available for 15 of 22 cases: SHH ($n = 1$*), group 3 ($n = 11$) and group 4 ($n = 3$). Close examination of the single non-group 3/group 4 medulloblastoma affected by a focal copy-number event in the region (MAGIC_MB1318, SHH) revealed that this sample exhibits a homozygous deletion (in the context of broad chr9q deletion) specifically overlapping *TSC1* and is therefore unlikely to be related to the events which target *GFI1B* for transcriptional activation. Indicated coordinates are based on the hg18 reference genome (NCBI Build 36.1) that was used in the original MAGIC study.



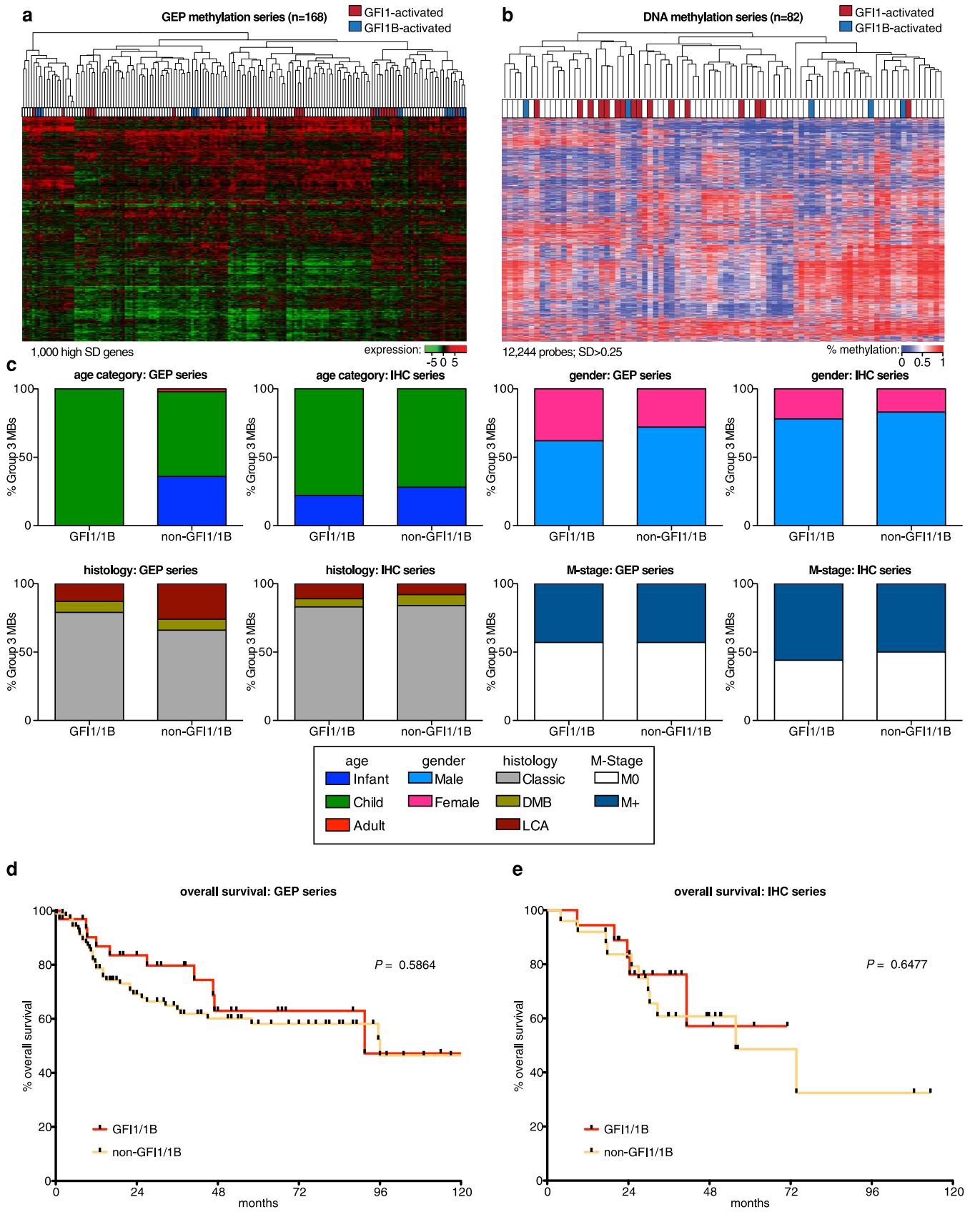
Extended Data Figure 2 | Non-functional *DDX31–GFI1B* fusion transcripts detected by RNA-seq. **a**, A complex SV on 9q34 in ICGC_MB9 resulted in expression of *DDX31* (exon 19) fused to *GFI1B* (intron 2, antisense orientation). Note the intronic reads in *GFI1B* after the fusion breakpoint. **b**, 9q34 inversions in ICGC_MB247 resulted in expression of *DDX31* (exon 19) fused to *GFI1B* (exon 2, sense orientation). This fusion transcript included a frameshift, inferred to generate a C-terminal-truncated DDX31 protein and no GFI1B protein from this fused allele.



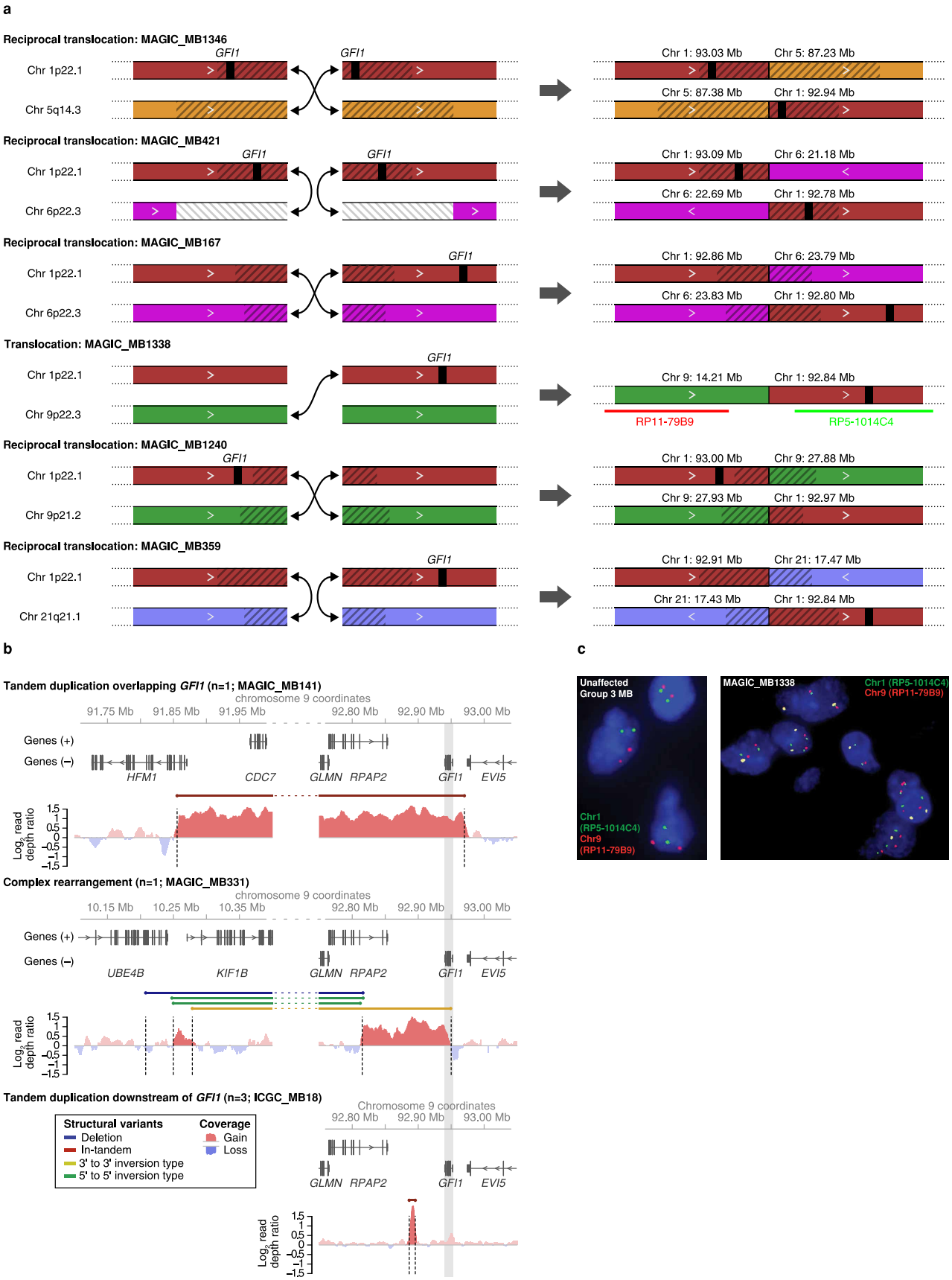
Extended Data Figure 3 | Expression and correlation of 9q34 genes in medulloblastoma subgroups. **a**–**c**, Box-plots summarizing expression of *BARHL1* (**a**), *DDX31* (**b**) and *GTF3C4* (**c**) according to medulloblastoma subgroup. Data set includes 375 medulloblastomas profiled on the Affymetrix U133plus2 array. **d**, Pearson correlation analysis showing correlated expression of *DDX31* with *BARHL1* and *GTF3C4* in group 3 and group 4 medulloblastomas. *DDX31* expression is positively correlated with both *BARHL1* ($r = 0.741$) and *GTF3C4* ($r = 0.622$). **e**, *PRRC2B* expression in medulloblastoma subgroups. Samples are from the same series summarized in **a**–**c**. **f**, Distribution of H3K27ac ChIP-seq signal at predicted enhancers in group 3 medulloblastomas (data for MAGIC_MB360 are shown). Enhancer regions are plotted in increasing order based on their input-normalized H3K27ac signal. Super-enhancers are defined as the population of enhancers above the inflection point of the curve (horizontal dashed grey line). Positions of the predicted *BARHL1/DDX31* and *PRRC2B* super-enhancers described in the text are highlighted.



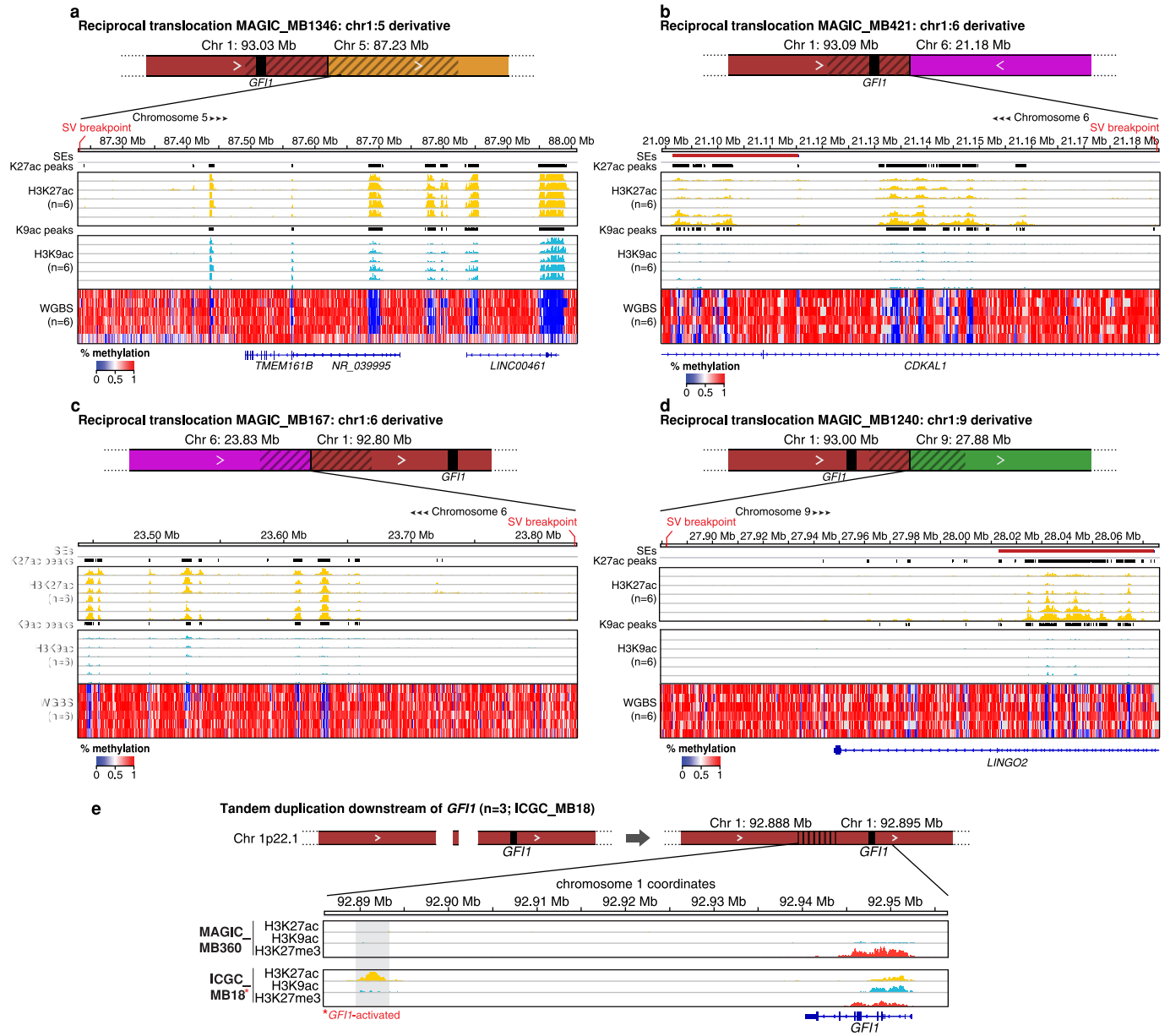
Extended Data Figure 4 | Frequency and distribution of GFI1/GFI1B activation in medulloblastoma subgroups. **a**, Stacked bar graph indicates the proportion of *GFI1/GFI1B*-expressing cases in each of the four medulloblastoma subgroups, as determined by Affymetrix gene expression profiling of two independent cohorts ($n = 727$). **b**, Stacked bar graph indicates the proportion of GFI1/GFI1B-positive cases in each of the four medulloblastoma subgroups, as determined by immunohistochemistry performed with anti-GFI1 and anti-GFI1B antibodies on formalin-fixed paraffin-embedded sections derived from a medulloblastoma clinical trial cohort (HIT2000, NCT00303810; $n = 156$). **c**–**f**, Representative positive and negative immunohistochemistry results for group 3 medulloblastomas stained with anti-GFI1 (**c**, **d**) and anti-GFI1B (**e**, **f**) antibodies, respectively.



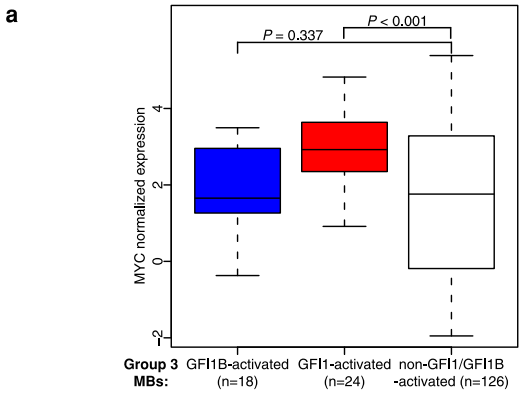
Extended Data Figure 5 | Demographic and clinical characteristics of GFI1/GFI1B-activated group 3 medulloblastoma. **a**, **b**, Unsupervised hierarchical clustering of group 3 medulloblastoma samples profiled by Affymetrix gene expression array (**a**) or Illumina 450K DNA methylation array (**b**). **c**, Patient characteristics, including age, gender, histological subtype (histology) and metastatic status (M-stage) for group 3 medulloblastomas stratified according to GFI1 and GFI1B expression status. Both gene expression and immunohistochemistry cohorts are summarized. **d**, **e**, Overall survival of group 3 medulloblastomas stratified by GFI1 and GFI1B expression status for both our gene expression (**d**) and immunohistochemistry series (**e**).



Extended Data Figure 6 | Summary of *GFI1* SVs detected by WGS in group 3 medulloblastoma. **a**, Schematics depicting the six different *GFI1* translocations detected by large-insert paired-end sequencing of our *GFI1*-activated validation series. **b**, WGS coverage plots showing SVs affecting the *GFI1* locus in *GFI1*-activated medulloblastomas sequenced in our series. **c**, Fluorescence *in situ* hybridization (FISH) analysis of MAGIC_MB1338 validating the unbalanced t(1:9) translocation (shown in **a**) predicted by WGS.



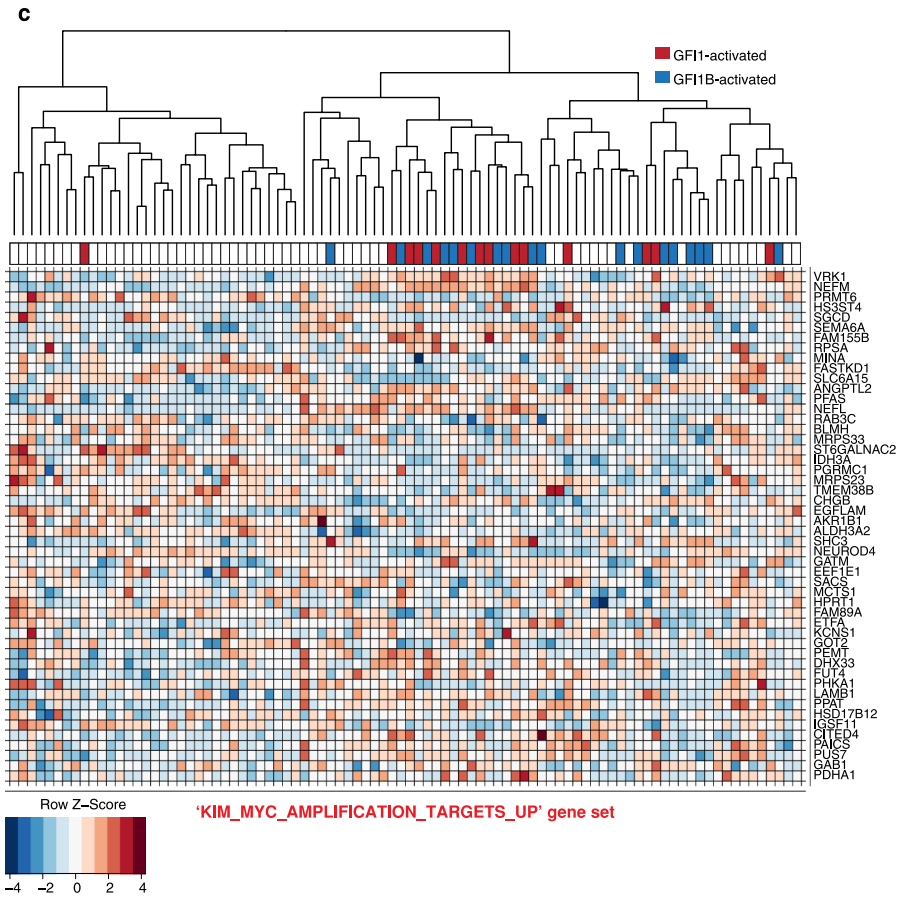
Extended Data Figure 7 | Chromatin states proximal to SVs observed in *GFI1*-activated group 3 medulloblastomas. a–d, ChIP-seq (H3K27ac and H3K9ac) and WGBS data respectively highlighting the active chromatin and methylation states present in the regions proximal to SV breakpoints identified in *GFI1* translocation cases. **e**, Schematic summarizing the series of focal tandem duplications observed approximately ~45 kb downstream of *GFI1* in group 3 medulloblastomas ($n = 3$; ICGC_MB18 is shown as a representative case). Activating and repressive histone marks overlapping the region of interest are shown for a non-*GFI1*-activated group 3 medulloblastoma (MAGIC_MB360) and the tandem duplication case (ICGC_MB18).

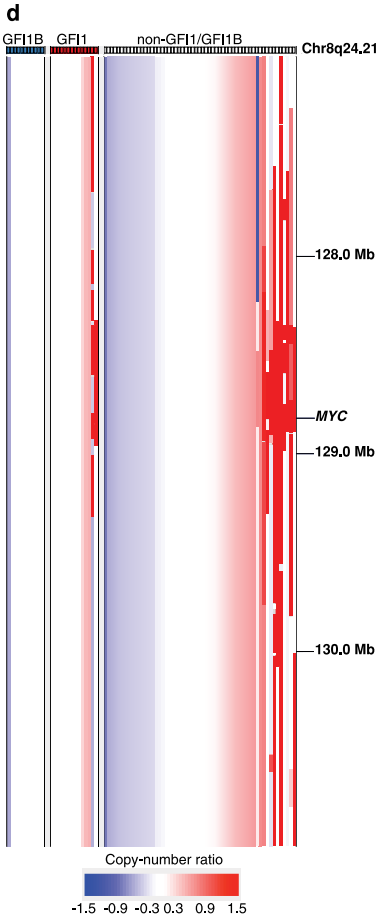


b

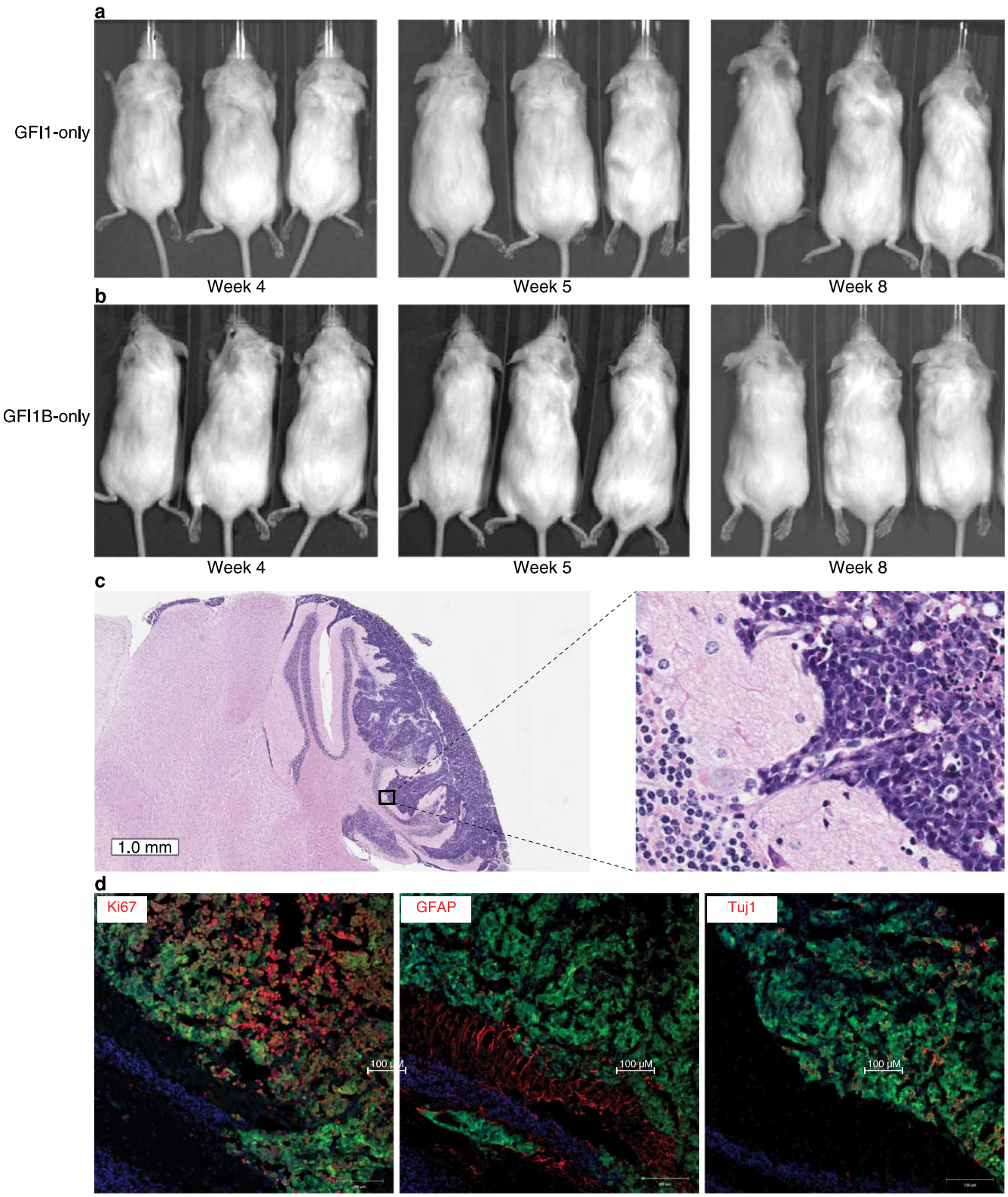
c2 (CGP, Biocarta, KEGG, Reactome)

gene set	# of observed genes	FDR (0.25)
CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_2	430	<<0.0001
KIM_MYC_AMPLIFICATION_TARGETS_UP	187	0.00159
VERHAAK_GLIOBLASTOMA_CLASSICAL	196	0.0123





Extended Data Figure 8 | Association between *GFI1/GFI1B* activation and *MYC* in group 3 medulloblastoma. **a**, *MYC* expression in group 3 medulloblastomas ($n = 168$) according to *GFI1* and *GFI1B* activation status. **b**, Gene sets with significant enrichment in *GFI1/GFI1B*-associated genes from the MSigDB c2 gene set collection. The collection highlighted in red is the only result found that shows a significant enrichment in both *GFI1* and *GFI1B* associated genes and a clear connection to a known pathway. **c**, Heat-map of the expression values for the 50 genes in the KIM_MYC_AMPLIFICATION_TARGETS_UP gene set with the most significant association with *GFI1* or *GFI1B* expression (the complete gene set contains 187 profiled genes). Genes are ordered top to bottom from most to least significant. A set of 90 group 3 medulloblastomas included in the analysis is displayed. Sample-wise hierarchical clustering was performed only to enhance the visual organization of the heat map. **d**, Affymetrix SNP6 copy-number output for 82 primary group 3 medulloblastomas from the published MAGIC series, highlighting the incidence of *MYC* amplification in the context of *GFI1/GFI1B*-activation. *MYC* amplification was found at a comparable frequency in both *GFI1*-activated ($n = 2$ of 14, 14.3%) and non-*GFI1/GFI1B*-activated ($n = 10$ of 57, 17.5%) group 3 medulloblastomas. Indicated coordinates are based on the hg18 (NCBI Build 36.1) reference genome that was used in the original MAGIC study.



Extended Data Figure 9 | Phenotypic characteristics of novel *GFI1/GFI1B* orthotopic mouse models. **a**, **b**, Bioluminescent imaging of animals injected with either *GFI1*- (**a**) or *GFI1B*-expressing (**b**) neural stem cells at the indicated time points. No tumour signal was detectable in these animals. **c**, Haematoxylin and eosin staining of cerebellar sections derived from MYC + GFI1B tumour-bearing mice. **d**, Immunofluorescence imaging of cerebellar sections from MYC + GFI1B tumours stained with the indicated antibodies.