CANCER

Childhood cancer mutagenesis caused by transposase-derived PGBD5

Makiko Yamada[1,2]†, Ross R. Keller[1,2]†‡, Rodrigo Lopez Gutierrez[3], Daniel Cameron[1,2], Hiromichi Suzuki[4], Reeti Sanghrajka[5], Jake Vaynshteyn[6], Jeffrey Gerwin[6], Francesco Maura[7], William Hooper[8], Minita Shah[8], Nicolas Robine[8], Phillip Demarest[1,2], N. Sumru Bayin[5,9], Luz Jubierre Zapater[1,2], Casie Reed[1,2], Steven Hébert[10], Ignas Masilionis[11], Ronan Chaligne[11], Nicholas D. Socci[12], Michael D. Taylor[13,14,15], Claudia L. Kleinman[3,10], Alexandra L. Joyner[5,16], G. Praveen Raju[6]§, Alex Kentsis[1,2,17]*

Genomic rearrangements are a hallmark of most childhood tumors, including medulloblastoma, one of the most common brain tumors in children, but their causes remain largely unknown. Here, we show that PiggyBac transposable element derived 5 (Pgbd5) promotes tumor development in multiple developmentally accurate mouse models of Sonic Hedgehog (SHH) medulloblastoma. Most Pgbd5-deficient mice do not develop tumors, while maintaining normal cerebellar development. Ectopic activation of SHH signaling is sufficient to enforce cerebellar granule cell progenitor–like cell states, which exhibit Pgbd5-dependent expression of distinct DNA repair and neurodevelopmental factors. Mouse medulloblastomas expressing Pgbd5 have increased numbers of somatic structural DNA rearrangements, some of which carry PGBD5-specific sequences at their breakpoints. Similar sequence breakpoints recurrently affect somatic DNA rearrangements of known tumor suppressors and oncogenes in medulloblastomas in 329 children. This identifies PGBD5 as a medulloblastoma mutator and provides a genetic mechanism for the generation of oncogenic DNA rearrangements in childhood cancer.

INTRODUCTION

Cancer development is caused by the acquisition of somatic mutations affecting tumor suppressors and oncogenes that encode factors regulating cell proliferation, differentiation, and other hallmarks of cancer (*1*). Compared to aging-associated cancers, childhood tumors are characterized by significantly lower total numbers of genetic mutations, although they exhibit chromosomal deletions, amplifications, translocations, and other complex oncogenic genomic rearrangements (*2*). The origins of these oncogenic mutations in childhood and young adult tumors remain obscure. A quintessential example is medulloblastoma (MB), one of the most frequent childhood brain tumors. For example, MBs with activation of Sonic Hedgehog (SHH) signaling are caused by the aberrant proliferation of cerebellar granule cell precursors (GCPs) due to activating mutations in *SMO* or somatic deletions of the tumor suppressors *PTCH1*, *SUFU*, and *TP53*. MBs also exhibit frequent amplifications of oncogenes such as *GLI2*, *MYC*, and *MYCN*, and other complex genomic rearrangements. MBs can be caused by Li-Fraumeni syndrome with germline deficiency of *TP53*, which dysregulates chromosome replication and repair, thereby causing complex oncogenic DNA rearrangements known as chromothripsis. However, the causes of somatic DNA rearrangements in sporadic MBs are not known, despite being a defining feature of this and most other childhood and young adult tumors.

Recently, we found that PGBD5 (PiggyBac transposable element derived 5), the most evolutionarily conserved domesticated DNA transposase–derived gene in vertebrates, can mediate sequence-specific DNA rearrangements dependent on its putative nuclease activity and end-joining DNA repair in human cells (*3*–*5*). Although PGBD5 can support genomic DNA integration in cells, its cellular activity predominantly involves double-strand DNA breaks, deletions, and other DNA rearrangements. While the specific enzymatic mechanisms of PGBD5-induced genome rearrangements need to be defined, PGBD5 can mediate rearrangements of both heterologous transposon substrates as well as sequence elements in the human and mouse genomes (*3*, *4*), as validated by Helou *et al.* (*6*, *7*) and recently confirmed by Bigot *et al.* (*8*).

RESULTS

In surveying the expression of PGBD5 across human cancers, we found that most human MBs express high levels of *PGBD5* (*4*). Therefore, we hypothesized that dysregulation of PGBD5 might

[1]Tow Center for Developmental Oncology, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Molecular Pharmacology Program, Sloan Kettering Institute, New York, NY, USA. [3]Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [4]Division of Brain Tumor Translational Research, National Cancer Center Research Institute, Tokyo, Japan. [5]Developmental Biology Program, Sloan Kettering Institute, New York, NY, USA. [6]Departments of Neurology and Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [7]Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, USA. [8]Computational Biology, New York Genome Center, New York, NY, USA. [9]Wellcome Trust/Cancer Research UK Gurdon Institute, Cambridge University, Cambridge, UK. [10]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada. [11]Single-Cell Analytics Innovation Lab, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [12]Bioinformatics Core, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [13]Department of Pediatrics—Hematology/Oncology and Neurosurgery, Baylor College of Medicine, Houston, TX, USA. [14]Hematology-Oncology Section, Texas Children's Cancer Center, Houston, TX, USA. [15]The Arthur and Sonia Labatt Brain Tumour Research Centre and the Developmental and Stem Cell Biology Program, The Hospital for Sick Children, Toronto, ON, Canada. [16]Biochemistry, Cell and Molecular Biology Program and Neuroscience Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA. [17]Departments of Pediatrics, Pharmacology, and Physiology & Biophysics, Weill Medical College of Cornell University, New York, NY, USA.
*Corresponding author. Email: kentsisresearchgroup@gmail.com
†These authors contributed equally to this work.
‡Present address: Private Health Management, New York, NY, USA.
§Present address: Rady Children's Hospital and Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA.

contribute to the somatic induction of oncogenic DNA rearrangements. Among the four major types of MBs, *PGBD5* gene expression is highest in tumors with constitutive SHH signaling (*4*). Here, we used mouse models of sporadic SHH MB, induced by constitutive activation of SHH signaling in developing mouse cerebellar GCPs, to investigate the possibility of developmental mutator activity of PGBD5 (Fig. 1A).

First, we engineered mice with *loxP* sites flanking exon 4 of mouse *Pgbd5*, generating $Pgbd5^{fl/fl}$ and knockout $Pgbd5^{-/-}$ mice in which translation of the protein is out of frame after exon 4 (fig. S1A). In situ hybridization (ISH) with a *Pgbd5* exon 4–specific probe set confirmed complete loss of exon 4 transcripts in $Pgbd5^{-/-}$ mice (figs. S1B and S3B). We then crossed the $Pgbd5^{-/-}$ alleles into $Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$ mice, which develop tumors with the highest known penetrance among mouse models of sporadic SHH MB (Fig. 1, B and C) (*9*). Mosaic loss of *Ptch1* leads to preneoplastic hyperplasia of developing cerebellar GCPs due to the constitutive activation of SHH signaling, similar to sporadic human SHH MBs with somatic mutations of *PTCH1*, which affect more than 40% of patients with SHH MBs (*10*). Analysis of three independent cohorts of $Pgbd5^{+/+}$;$Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$ and $Pgbd5^{-/-}$;$Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$ mice showed that *Pgbd5* significantly promoted MB development (Fig. 1D and fig. S1C). As observed with other mouse cancer models (*11*), tumorigenesis was more penetrant in mice with mixed SW genetic background as compared to the congenic C57BL/6J strain. Among *Pgbd5*-knockout tumor model mice, as many as 70% of animals did not develop tumors after 1 year of life (mean 61%, log-rank $P = 1.4 \times 10^{-8}$), whereas most $Pgbd5^{+/+}$ mice (79%) rapidly succumbed to MBs with a median latency of 5 months.

To exclude the possibility of tumor cell–extrinsic effects of germline $Pgbd5^{-/-}$ deletion, we used *Pgbd5*-floxed mice in which *Pgbd5* loss was primarily confined to cerebellar progenitor cells due to conditional Cre expression and *loxP* recombination. Both $Pgbd5^{fl/-}$;$Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$ mice and their $Pgbd5^{+/-}$;$Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$ littermates developed MBs with similar penetrance and latency (Fig. 1E). However, genomic polymerase chain reaction (PCR) analysis using primers specific for the *Pgbd5*-floxed allele demonstrated that seven of nine analyzed tumors (78%) retained a substantial amount of intact *Pgbd5*, indicating a selective advantage for *Pgbd5*-expressing tumor progenitor cells (Fisher's exact test $P = 2.3 \times 10^{-3}$; Fig. 1F and fig. S1D).

We performed a similar analysis using a developmentally accurate mouse model of sporadic *SMO*-mutant MB, corresponding to a mutation of *SMO* that aberrantly activates SHH signaling in human patients (*10*, *12*). This model leverages a system for mosaic mutagenesis with spatial and temporal control of recombination (MASTR) (*12*, *13*). In this system, green fluorescent protein (GFP)–Cre is induced by tamoxifen at postnatal day 0 (P0) in cerebellar GCPs and further induces oncogenic *SmoM2* expression and concurrent deletion of *Pgbd5* (fig. S2A). Both $Pgbd5^{fl/fl}$ and their $Pgbd5^{fl/+}$ littermates developed tumors with similar latencies (Fig. 1G). However, genomic PCR analysis again showed that most analyzed $Pgbd5^{fl/fl}$ tumors (four of five) retained intact *Pgbd5* alleles, indicating that *Pgbd5* enhances *SmoM2*-mutant SHH MB development (Fisher's exact test $P = 2.1 \times 10^{-2}$; Fig. 1H and fig. S2B). To exclude the possibility that apparent *Pgbd5* expression was due to the infiltration of stromal or immune cells, we confirmed the requirement for *Pgbd5* in MB tumor cell development using three of the same *SmoM2*-mutant tumors for ISH with a *Pgbd5* exon 4–specific probe and *Cre*-specific probe as a positive control for tumor cells (fig. S3, C to F). This revealed specific *Pgbd5* transcript expression in tumor cells for two of three analyzed *MASTR-SmoM2;*$Pgbd5^{fl/fl}$ tumors. In all, these results indicate that *Pgbd5* promotes tumor development in diverse developmentally accurate mouse models of SHH MBs.

SHH MBs are thought to originate in developing cerebellar GCPs, which in turn are dependent on SHH signaling (*13*–*20*). To exclude the possibility that PGBD5-induced tumorigenesis is due to its control of normal cerebellar development, we analyzed the cerebellar cytoarchitecture of $Pgbd5^{-/-}$ mice. We observed grossly intact medial cerebellar vermis and lateral hemispheres, including normal cytoarchitecture and morphology, which are essential hallmarks of cerebellar development (Fig. 2A). To examine the effects of *Pgbd5* deficiency on SHH signaling directly, we isolated GCPs from the cerebella of 5-day-old mice, when SHH signaling is required for cerebellar development, and measured SHH pathway activity using quantitative reverse transcription PCR of the canonical SHH signaling biomarker *Gli1* (*17*). We observed no significant differences in *Gli1* expression between $Pgbd5^{-/-}$ and wild-type developing cerebellar GCPs ($P = 0.87$; Fig. 2B). Thus, the requirement of *Pgbd5* for SHH MB development cannot be explained by the effects of SHH signaling on growth and survival of mutant GCPs, which initiate tumor progression.

Oncogenic SHH signaling leads to hyperplasia of cerebellar GCPs. These preneoplastic cells express Atoh1 and persist in the external granule layer between 3 and 8 weeks of age (*20*, *21*). To examine the effects of Pgbd5 on preneoplastic GCP hyperplasia, we used an *Atoh1-GFP* reporter transgene and fluorescence-activated cell sorting (FACS) to isolate preneoplastic GCPs developing in the cerebellar external granule layer in 3- to 8-week-old $Pgbd5^{-/-}$;$Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$;*Atoh1-GFP* mice (*22*). We found that 21 of 28 (75%) of $Pgbd5^{+/+}$;$Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$;*Atoh1-GFP* mice harbored preneoplastic cells in their cerebella, which was similar to *Pgbd5*-knockout mice (Fig. 2, C and D). The preneoplastic populations also showed no significant differences in the numbers of *Atoh1*-expressing cells ($P = 0.16$; fig. S4). Thus, *Pgbd5* is dispensable for normal cerebellar development, physiologic SHH signaling, and the growth of preneoplastic SHH MB progenitor cells.

Vertebrate *PGBD5* is derived from *piggyBac* DNA transposases, which induce double-strand DNA breaks and rearrangements at specific sequences via their ribonuclease H (RNase H)–like domain, which is highly conserved among vertebrate *PGBD5* genes (*3*, *4*, *23*, *24*). Does PGBD5 promote MB development by inducing sequence-specific somatic mutations? To test this hypothesis, we performed whole-genome PCR-free DNA sequencing of MBs from both *Ptch1*- and *SmoM2*-mutant tumors, as compared to their matched normal tissues. We analyzed resultant sequencing data using recently developed methods optimized for the accurate detection of somatic cancer genome rearrangements (*25*).

Pgbd5-expressing MBs from $Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$ mice exhibited nearly twice as many somatic DNA rearrangements, including deletions, insertions, translocations, and amplifications, as compared to those from *Pgbd5*-deficient mice (mean 69 and 44, respectively; *t* test $P = 0.036$; Fig. 3A). This increase could not be attributed to the age of tumor development, as *Pgbd5*-expressing MBs were significantly younger (mean 140 and 201 days, respectively; *t* test $P = 0.024$; Fig. 3B). There was no correlation between the number of somatic genome rearrangements and tumor age ($r = -0.18$, $P = 0.55$; fig. S5A), consistent with a distinct somatic mutational process responsible for tumor development. $Pgbd5^{+/+}$ and $Pgbd5^{-/-}$ tumors

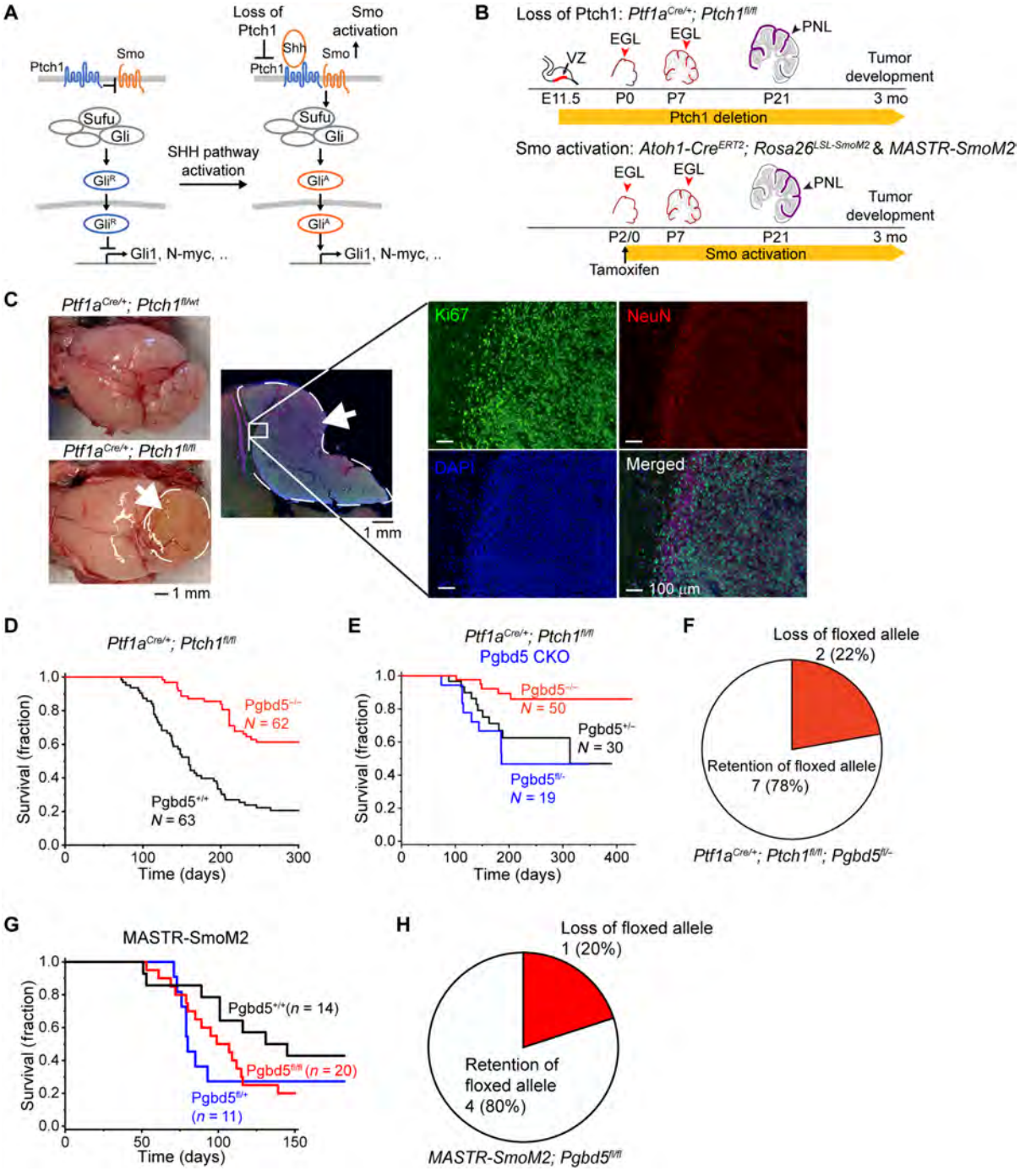

Fig. 1. *Pgbd5* promotes tumorigenesis in diverse developmentally accurate mouse models of SHH MB. (**A**) Schematic of aberrant mechanisms of SHH signaling in cerebellar GCPs in MB development (left). In *Ptch1*-mutant *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$ MB, by deletion of *Ptch1* encoding a receptor for SHH, SMO signaling is disinhibited and highly activated, leading to the generation of activated GLI (GLI^{A}). In *Smo*-mutant *MASTR-SmoM2* or *Atoh1-CreER*^{T2};*Rosa26*$^{LSL\text{-}SmoM2}$ MB, oncogenic constitutively activated form of SMO results in GLI activation and aberrant SHH signaling. (**B**) Schematic of cerebellar tumor development in *Ptch1*- (top) and *SmoM2*-mutant (bottom) MB. Red arrowheads mark conditionally gene-targeted cell populations. PNLs (purple) lead to MB development. E, embryonic day; P, postnatal day. (**C**) Representative photographs of dissected brains of *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$ (bottom left) mice that develop MBs marked by white arrows and dashed circles, as compared to *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/wt}$ mice (top left) that do not develop tumors. Immunofluorescence microscopy (right) shows high Ki67 (green) and low NeuN (red) expression in MB tumors, with nuclei marked with DAPI (blue). The edge of the tumor (white inset) is magnified with NeuN-positive cells on tumor edge corresponding to normal cerebellum. (**D**) Survival of *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$;*Pgbd5*$^{+/+}$ (black) and *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$;*Pgbd5*$^{-/-}$ (red) mice. (**E**) Survival of *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$ mice with conditional knockout (CKO) of *Pgbd5*$^{fl/-}$ (blue) or control *Pgbd5*$^{+/-}$ (black) or *Pgbd5*$^{-/-}$ (red) littermates. (**F**) Genomic PCR analysis of conditional *Pgbd5* excision in *Pgbd5*$^{fl/-}$ CKO *Ptch1*-mutant tumors demonstrates that seven of nine (79%) analyzed tumors retain intact *Pgbd5*, as detailed in fig. S1D. (**G**) Survival of *MASTR-SmoM2* mice with CKO of *Pgbd5*$^{fl/fl}$ (red), as compared to control *Pgbd5*$^{+/+}$ (black) or *Pgbd5*$^{fl/+}$ (blue) littermates. (**H**) Genomic PCR analysis of conditional *Pgbd5* excision in *MASTR-SmoM2; Pgbd5*$^{fl/fl}$ tumors demonstrates that four of five (80%) analyzed tumors retain *Pgbd5* floxed alleles, as detailed in fig. S2B.

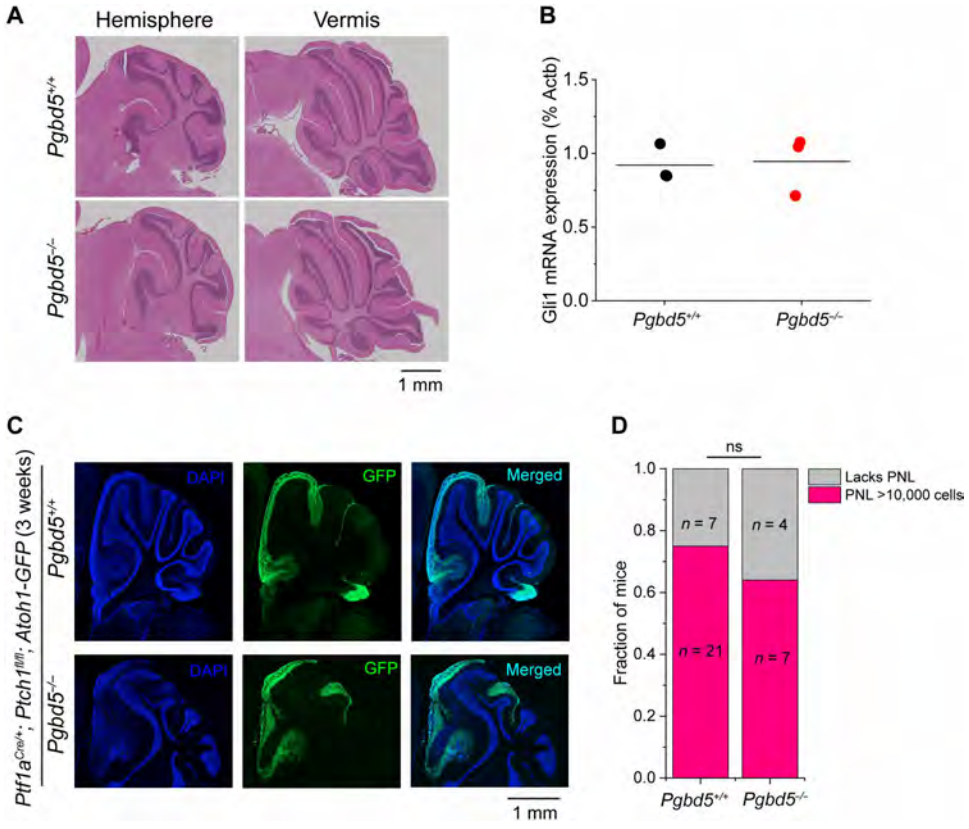

Fig. 2. *Pgbd5* is dispensable for normal SHH signaling and cerebellar development. (**A**) Representative immunohistochemistry micrographs of sagittal sections of cerebellum of *Pgbd5*$^{+/+}$ (top) and *Pgbd5*$^{-/-}$ (bottom) mice at 6 weeks of age show normal cytoarchitecture and morphology of cerebellar hemispheres (left) and vermis (right). (**B**) Expression of *Gli1* mRNA in purified cerebellar GCPs from 5-day-old *Pgbd5*$^{+/+}$ (black) and *Pgbd5*$^{-/-}$ (red) mice. Bars represent means of three biologic replicates ($P = 0.87$). (**C**) Representative fluorescence images of cerebellar hemispheres of 3-week-old *Pgbd5*$^{+/+}$ (top) versus *Pgbd5*$^{-/-}$ (bottom) *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$;*Atoh1-GFP* mice showing PNLs (green) marked by Atoh1-GFP expression, with nuclei marked with DAPI (blue). (**D**) Fraction of mice harboring PNLs (red) in *Pgbd5*$^{wt/wt}$;*Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$;*Atoh1-GFP* and *Pgbd5*$^{-/-}$;*Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$;*Atoh1-GFP* mice between 3 and 8 weeks of age. Both groups harbor similar fractions of PNLs that are defined as at least 10,000 Atoh1-GFP–positive cells (Fisher's exact test $P = 0.69$); ns, not significant.

exhibited similar types of somatic structural rearrangements (fig. S5B). Similar to human MBs, both *Ptch1*- and *SmoM2*-mutant mouse MBs exhibited relatively low numbers of single-nucleotide variants (SNVs), consistent with their early embryonal age of onset, regardless of *Pgbd5* expression (mean 1.6 and 0.28 mutations/megabase, respectively; fig. S6, A and B). There were also no significant differences in single- or double-nucleotide mutational signatures between *Pgbd5*$^{-/-}$ and *Pgbd5*$^{+/+}$ tumors, with the predominance of SBS5 and SBS18 signatures, currently attributed to chronological age and radical oxygen stress damage, respectively (fig. S6C) (*26*).

We identified putative MB tumor suppressor and oncogenes arising from *Pgbd5*-induced genomic rearrangements by analyzing their recurrence in independent mouse tumors, as compared to genes recurrently mutated in human MBs (Fig. 3C, fig. S7, and data S1, S2, and S5). This identified several genes, including *Fbxw7*, *Tbr1*, *Gfi1*, *Pik3ca*, and others known to be recurrently mutated in human SHH and non-SHH MBs (data S8) (*10*). The many of the same genes were also recurrently mutated in *SmoM2*-mutant mouse MB tumors, but not in the rare *Pgbd5*-deficient tumors (fig. S8A). Mutated genes were distributed across multiple distinct chromosomal regions, consistent with the specific genome-wide activity of their Pgbd5-dependent somatic mutagenesis (fig. S9 and data S20). This suggests that mouse SHH tumors model salient mutational features of human MBs, including specific developmental PGBD5-induced mutations.

To elucidate the specific mutational processes responsible for PGBD5-induced somatic genomic rearrangements in SHH MBs, we extracted 50–base pair (bp) sequences flanking all somatic DNA rearrangement breakpoints and analyzed their composition using supervised and de novo sequence motif analysis algorithms (fig. S10, A and B, and data S6). This analysis showed that most of the somatic DNA rearrangements contained repetitive sequences at their breakpoints, without apparent differences in rearrangement lengths (fig. S11), consistent with involvement of nonallelic homologous recombination or microhomology-mediated end-joining, which showed modest but not significant differences between *Pgbd5*-expressing and *Pgbd5*-deficient SHH MBs.

In contrast, specific nonrepetitive sequence breakpoints were significantly enriched at breakpoints of somatic DNA rearrangements in *Pgbd5*-expressing SHH mouse MBs as compared to *Pgbd5*-deficient tumors (mean 18 versus 6 per tumor, respectively; *t* test $P = 2.7 \times 10^{-3}$; fig. S11, A and B). Of the 2480 breakpoint sequences derived from somatic DNA rearrangements among nine *Ptch1*-mutant *Pgbd5*$^{+/+}$ MB tumors, 149 exhibited distinct sequence motifs (Fig. 3D). Some

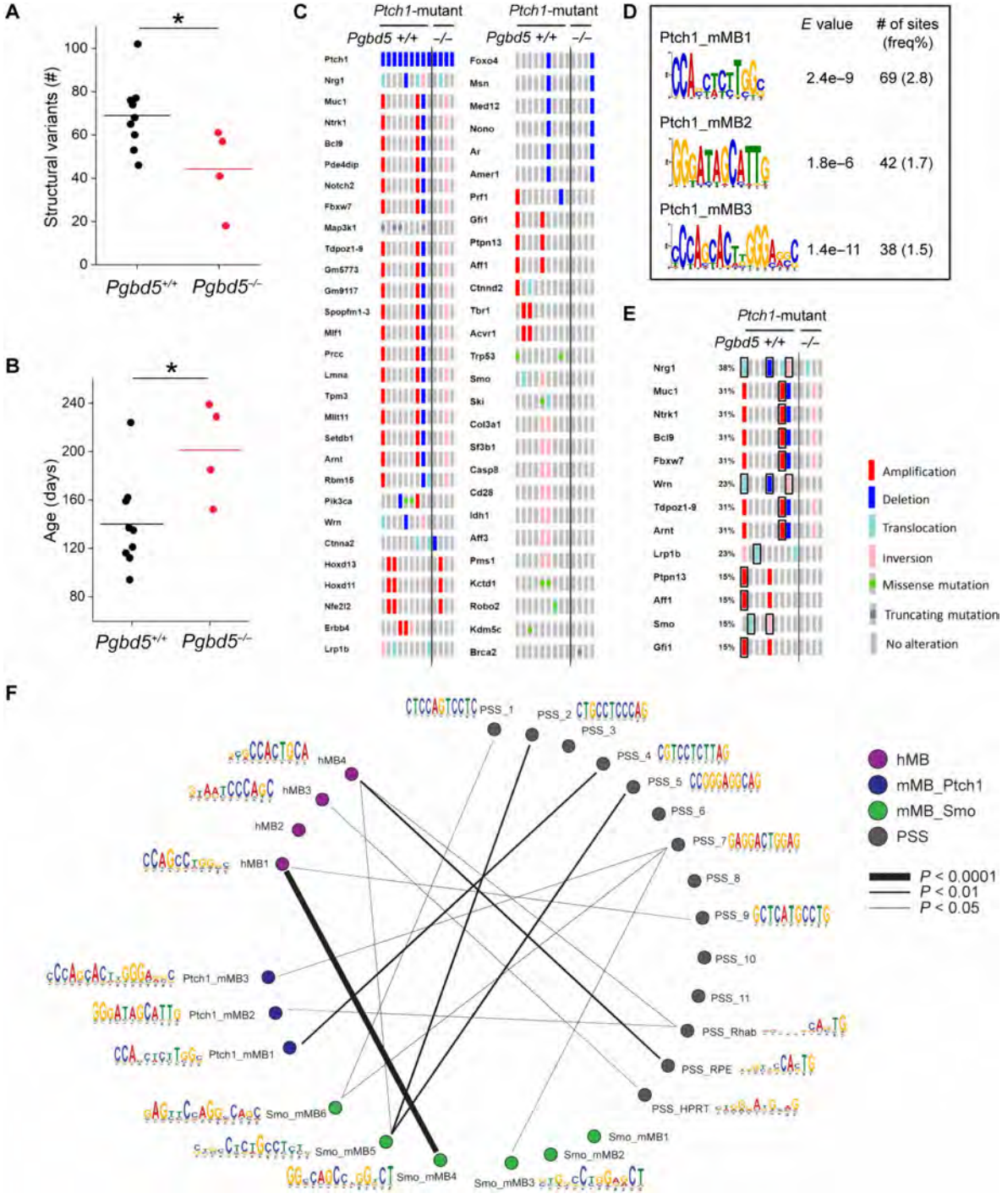

Fig. 3. *Pgbd5* promotes somatic mutagenesis of recurrently mutated tumor suppressor and oncogenes in mouse SHH MBs. (**A**) Numbers of SVs in *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$ tumors. *Pgbd5*$^{+/+}$ tumors (black, $n = 9$) harbor more SVs than *Pgbd5*$^{-/-}$ (red, $n = 4$). Lines indicate mean (69 and 44, respectively), and significance is measured using *t* test (*$P = 0.036$). (**B**) Age of tumors (days) in *Ptch1*-mutant tumors. *Pgbd5*$^{+/+}$ tumors (black, $n = 9$) are younger than *Pgbd5*$^{-/-}$ tumors (red, $n = 4$; mean 140 and 201 days, *$P = 0.024$). (**C**) Oncoprint showing genes recurrently affected by SVs and SNVs in independent *Ptch1*-mutant tumors. Genes are curated based on likelihood that SVs or SNVs affect gene function (see Materials and Methods for details). The left nine and right four columns indicate tumors from *Pgbd5*$^{+/+}$ and *Pgbd5*$^{-/-}$ mice, respectively. Red, blue, light blue, pink, and gray symbols indicate amplifications, deletions, translocations, inversions, and no alteration, respectively. Green and dark gray squares in gray symbols indicate missense and truncating mutations, respectively. (**D**) Three *Pgbd5*$^{+/+}$-specific motifs are identified at SV breakpoints in *Ptch1*-mutant tumors, using discriminative MEME with *Pgbd5*$^{-/-}$ tumors as controls. *E* values indicate MEME discriminative algorithm significance (see Materials and Methods for details). The frequency shown was calculated by dividing the number of sites by total numbers of 50-bp breakpoint sequences extracted from SVs. (**E**) These motifs in (D) and the previously identified PSS_RPE and PSS_Rhab motifs were identified at SV breakpoints affecting known tumor suppressor and oncogenes in six of nine (66%) of *Pgbd5*$^{+/+}$ tumors. The affected tumor suppressors and oncogenes are boxed black. (**F**) Circos plot showing similarities among all motifs (see Materials and Methods for details).

of these breakpoint sequences exhibited significant similarity to the PGBD5-specific signal sequence (PSS) motifs previously observed to be rearranged by PGBD5 in genomic transposition and forward genetic assays ($P = 7.6 \times 10^{-3}$, 4.5×10^{-2}, and 2.4×10^{-2}, respectively; Fig. 3F and table S1) (*4*, *24*). We confirmed the specificity of this PSS breakpoint detection using shuffled sequences that showed no significant associations despite having identical sequence composition (fig. S13 and table S1). We observed similar results in *SmoM2*-mutant MB tumors (252 of 5036 PSS-like breakpoints; $P = 5.2 \times 10^{-3}$ to 3.4×10^{-2}; Fig. 3E, figs. S12 and S13, and table S1). The relatively modest association of MB genomic rearrangement breakpoints with previously observed PSS motifs suggests that other developmental mutational processes remain to be found.

The occurrence of somatic genomic rearrangements with specific PSS motif breakpoints frequently involved known genetic drivers of MB, including *Smo*, *Nrg1*, *Wrn*, *Ntrk1*, *Fbxw7*, and *Gfi1* (Fig. 3E and table S2). In some cases, we observed recurrently mutated regions, such as, for example, mutations in chromosome 8 affecting both *Nrg1* and *Wrn* genes with distinct breakpoints in three independent mouse MB tumors (table S2). In total, six of nine (67%) mouse MB tumors exhibit recurrent Pgbd5-associated genetic somatic mutations affecting known MB tumor suppressor and oncogenes (table S2). These findings indicate that Pgbd5-induced sequence-specific somatic mutagenesis contributes to mouse SHH MB development.

To examine the contribution of PGBD5-induced sequence-specific somatic mutagenesis in human MBs, we used de novo local sequence assembly–based methods to analyze 329 tumor genomes isolated from patients with all four major MB tumor subtypes, including SHH MBs (Fig. 4, A and B) (*27*). We found that nearly one in five somatic MB DNA rearrangement breakpoints contained PSS motifs, previously observed to be rearranged by PGBD5 in genomic transposition and forward genetic assays (Fig. 4C and data S9 to S12). This enrichment was significant when compared to somatic breakpoints in non–PGBD5-expressing but highly somatically rearranged human breast carcinomas (χ^2 test $P = 7.7 \times 10^{-117}$; Fig. 4C) (*4*, *28*). Using unsupervised de novo sequence motif analysis, we also identified four sequence motifs, termed hMB1 to hMB4, which were also specifically and significantly enriched in breakpoint sequences of somatic human DNA MB rearrangements, but not in those from human breast carcinomas that are also highly somatically rearranged (χ^2 test $P = 1.5 \times 10^{-155}$; Fig. 4D and figs. S14 and S15, A and B).

In total, nearly one in three somatic human MB DNA rearrangements exhibited specific sequence breakpoints (Fig. 4D). Multiple human tumor DNA rearrangement breakpoints were similar to those detected in Pgbd5-induced mouse SHH MBs ($P = 5.3 \times 10^{-6}$ and 1.0×10^{-2} for hMB1 versus Smo_mMB4 and hMB4 versus Smo_mMB5, respectively; Fig. 3F, fig. S13, and table S2). In particular, hMB4 motif showed significant similarity to PSS_Rhab and PSS_RPE sequences identified as PGBD5 substrates in cellular assays ($P = 3.3 \times 10^{-2}$ and 6.0×10^{-4}, respectively; Fig. 3F, figs. S13 and S15C, and table S1). Consistent with the oncogenic activity of PGBD5-induced somatic mutagenesis, many recurrently mutated key SHH and non-SHH MB tumor suppressors and oncogenes, including *GLI2*, *PPM1D*, and *MYC* (*29*, *30*), involved breakpoints with specific PSS-like sequences (Fig. 4E and data S5). Therefore, human MBs are defined by somatic DNA deletions, amplifications, and other chromosomal rearrangements, marked in part by specific breakpoints with similarity to PSSs, which recurrently affect MB tumor suppressors and oncogenes.

We reasoned that PGBD5 may induce somatic mutations in preneoplastic lesions (PNLs) in transformed neuronal progenitor cells and/or their progeny, as they develop into fulminant MB tumors. To investigate this process, we analyzed γH2AX as a surrogate marker of DNA damage signaling in $Pgbd5^{+/+}$ and $Pgbd5^{-/-}$ cerebellar GCP cells transformed by oncogenic SHH signaling in *Atoh1-CreER*^{T2}*; R26SmoM2* mice. We observed prominent preneoplastic expansions in both $Pgbd5^{+/+}$and $Pgbd5^{-/-}$ mice at P22 and P23 of development (fig. S16A). We used the external cell layer structure to identify the PNLs and 5-ethynyl-2′-deoxyuridine (EdU) pulse-chase labeling to identify proliferating cells (fig. S16, A and B). $Pgbd5^{+/+}$ preneoplastic cells did not exhibit higher level of γH2AX nuclear foci in both dividing EdU-positive and nondividing EdU-negative cells compared with $Pgbd5^{-/-}$ preneoplastic cells (fig. S16, C and D). This suggests that Pgbd5-dependent mutagenesis occurs upon later tumorigenesis of PNLs and/or its early activity in transformed GCPs does not lead to global DNA damage signaling.

To define potential Pgbd5-dependent mechanisms in cerebellar GCP transformation induced by oncogenic SHH signaling, we performed single-nucleus RNA-sequencing (snRNA-seq) gene expression analysis of $Pgbd5^{+/+}$ and $Pgbd5^{-/-}$ MBs isolated from $Ptf1a^{Cre/+}$;$Ptch1^{fl/fl}$ mice (Fig. 5A). Upon identifying MB tumor cells by detecting their DNA copy number alterations induced by somatic mutations, we mapped the observed MB tumor cell gene expression onto the developmental ontogeny of normal mouse cerebellum (*31*). We observed that gene expression states of both $Pgbd5^{+/+}$ and $Pgbd5^{-/-}$ MB tumor cells resembled developing cerebellar GCPs and granule cells, with similar abundance of inferred cell states between $Pgbd5^{+/+}$ and $Pgbd5^{-/-}$ MB tumors (Fig. 5, B and C; figs. S17 and S18; and data S21). Thus, Pgbd5-dependent neuronal progenitor cell transformation involves GCP-like cells and their subsequent tumorigenic evolution (Fig. 5E).

Can Pgbd5 contribute to the transformation of cerebellar GCP transformation induced by oncogenic SHH signaling? To explore this question, we compared single-nucleus gene expression profiles of $Pgbd5^{+/+}$ and $Pgbd5^{-/-}$ for each cell population within the tumors. We observed significantly differentially expressed *Pgbd5*-associated genes known to encode regulators of DNA damage repair, neurogenesis, and cell migration (Fig. 5D and data S22). This included *Hjurp*, which can regulate double-strand DNA repair (*32*–*34*), neurogenesis factors *Dll3* and *Arx* (*35*, *36*), as well as neuronal cell migration regulator *Nhsl1*a (Fig. 5D and data S22) (*37*). Thus, in addition to its oncogenic mutator functions via sequence-specific somatic mutagenesis, Pgbd5 may also regulate neuronal progenitor cell development and DNA damage repair.

DISCUSSION

Although somatic mutational mechanisms have been extensively documented in aging-associated cancers, how oncogenic DNA rearrangements occur during dysregulated development in childhood and young-onset tumors is poorly understood. Here, we demonstrate that a domesticated DNA transposase–derived gene causes oncogenic sequence-specific DNA rearrangements in medulloblastomas, a common childhood brain tumor (*38*). We provide evidence that oncogenic developmental signaling not only can induce preneoplastic cell expansion but also involves endogenous mutators to generate

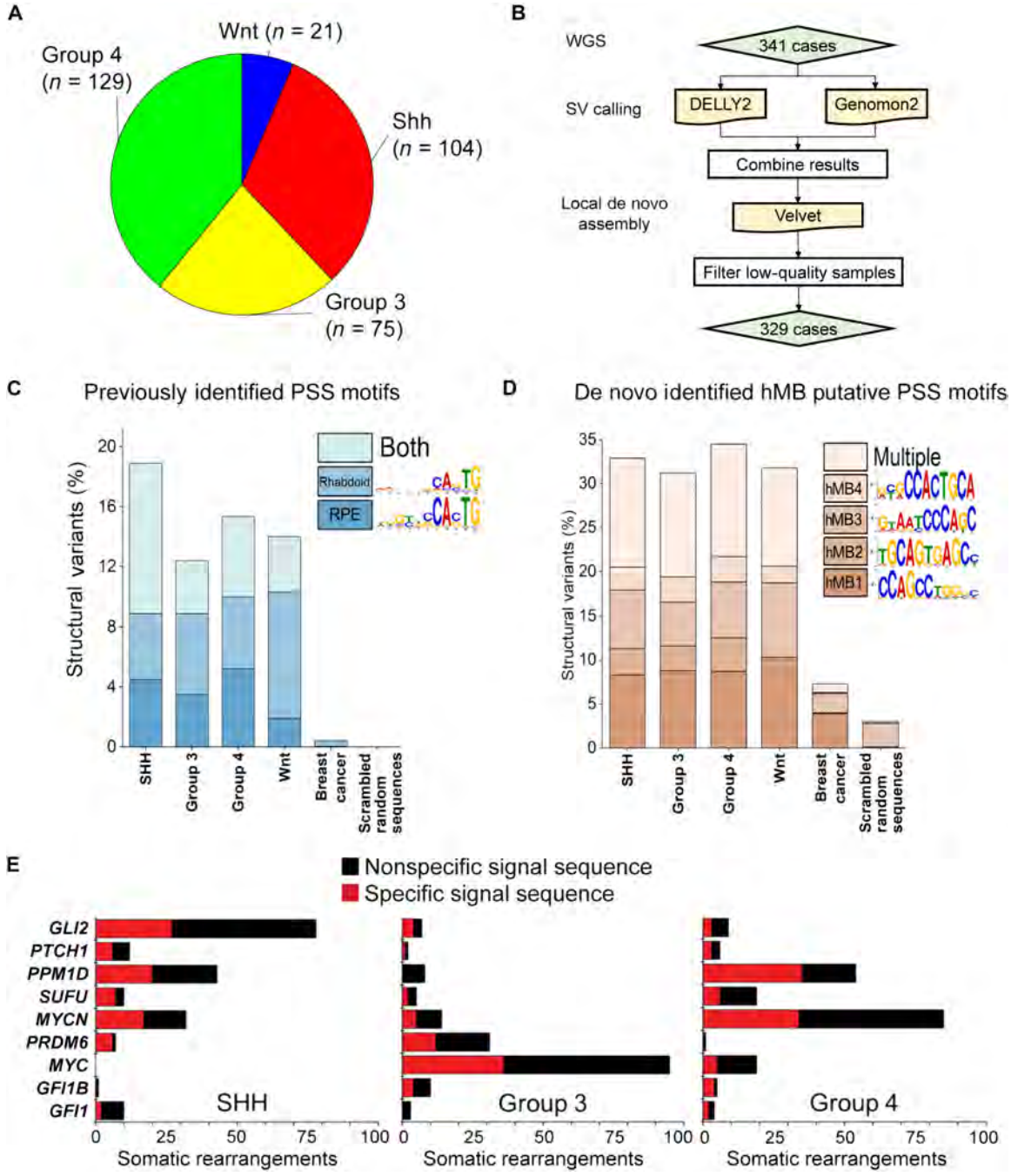

Fig. 4. PGBD5-associated sequence breakpoints recurrently affect somatic DNA rearrangements of known tumor suppressor and oncogenes in human MBs. (**A**) Representation of human patient cohort showing the four major subgroups of MB that were included in the analysis ($n = 329$). (**B**) Pipeline to identify somatic SVs in human MB. (**C**) Previously identified PSS sequences (*4*, *24*) are enriched at SV breakpoints in human MB as compared to somatic SVs in human breast carcinomas (χ^2 test $P = 7.7 \times 10^{-117}$). Percentages represent the frequency of SVs (each SV has two breakpoints and 4 × 50-mers) where the motif was identified using a FIMO *q* value threshold of 0.3 based on a ROC curve analysis (fig. S14, B and C). Multiple indicates that more than one motif was identified at one SV, in either distinct or the same 50-mers. Scrambled sequences showed no enrichment and represent the background of the FIMO algorithm. (**D**) A set of four de novo motifs identified at SV breakpoints in human MB is enriched relative to breast carcinoma and scrambled sequences. hMB1 to hMB4 were identified as being specific using MEME and eliminating repetitive motifs. In addition, discriminative MEME, where control sequences were a set of 50,000 randomly selected 50-mers from the hg19 reference genome, was used to determine whether the motif was enriched at breakpoints relative to the genome (fig. S14). Percentages represent motif frequency among SVs as in (C) and are compared to SV breakpoints in human breast carcinoma (χ^2 test $P = 1.5 \times 10^{-155}$) and scrambled sequences, which represent the background of the FIMO algorithm. (**E**) Recurrently mutated MB tumor suppressors and oncogenes in diverse tumor subtypes involve somatic DNA rearrangements with specific (red) sequence breakpoints, including PSS motifs. Numbers refer to the SVs detected in human patient cohort described in (A).

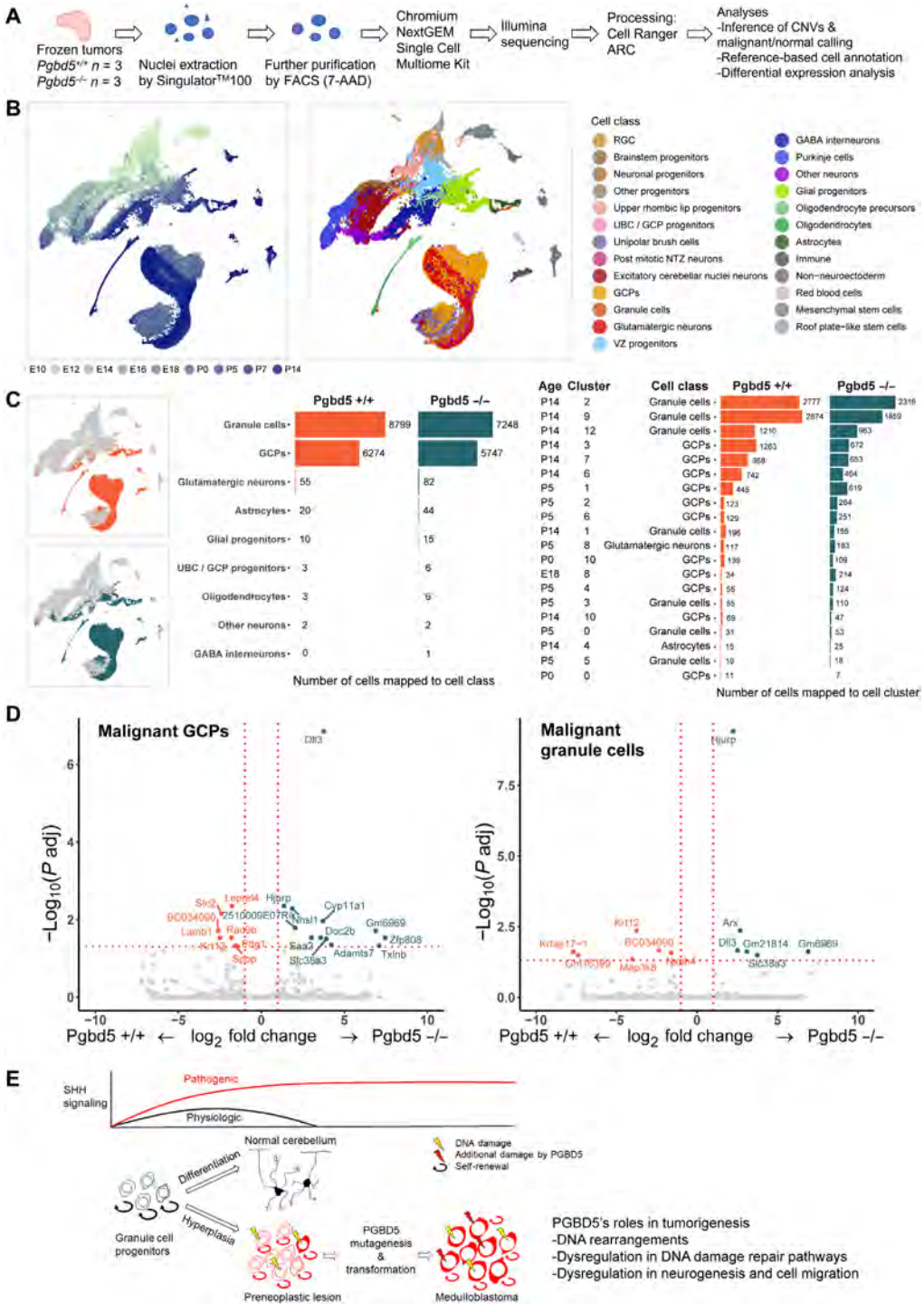

Fig. 5. Pgbd5-dependent mechanisms of cerebellar GCP transformation. (**A**) Schematic of experimental procedure for snRNA-seq of three *Pgbd5*$^{+/+}$ and three *Pgbd5*$^{-/-}$ fresh-frozen *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$ SHH MBs; same tumors as were analyzed by whole-genome DNA sequencing were used. (**B**) Uniform Manifold Approximation and Projection (UMAP) plots of the cerebellar reference ($N = 62{,}040$) used to annotate the cells from *Pgbd5*$^{+/+}$ and *Pgbd5*$^{-/-}$ *Ptf1a*$^{Cre/+}$;*Ptch1*$^{fl/fl}$ SHH MBs. Cells are colored by age (left) and by refined cell class ontology (right). (**C**) Cell class and cluster consensus predictions for *Pgbd5*$^{+/+}$ (orange) and *Pgbd5*$^{-/-}$ (blue) MB tumor cells. Left: UMAP plots of the cerebellar reference highlighting the mapped clusters of malignant cells of each genotype. Middle: Bar plot depicting the cell class consensus predictions for the tumor cells for each genotype. Right: Bar plot depicting the cluster-specific consensus predictions for tumor cells for each genotype. Only clusters with more than 10 cells mapped to by either genotype are shown. (**D**) Volcano plot showing differential gene expression between *Pgbd5*$^{+/+}$ and *Pgbd5*$^{-/-}$ malignant GCPs and granule cells. Genes significantly up-regulated in *Pgbd5*$^{+/+}$ MBs are highlighted in orange, while significantly up-regulated genes in *Pgbd5*$^{-/-}$ tumor cells are highlighted in blue ($\log_2$FC > 0.25, adjusted $P < 0.05$). (**E**) Model of PGBD5-dependent tumorigenesis, illustrating how pathogenic SHH signaling is associated with hyperplasia of cerebellar granule cell progenitor cells, leading to PNLs that undergo PGBD5-dependent somatic mutagenesis and malignant transformation.

genomic rearrangements that are predicted to affect tumor suppressors and oncogenes. We also derived the first principles of how a domesticated DNA transposase can promote somatic genome mutagenesis, providing a foundation for the identification of other sequence-specific mutational processes in other cancers. This raises the possibility that incorporation of sequence-specific DNA rearrangements into cancer mutational profiling could improve diagnosis and treatment (*39*, *40*).

PGBD5-mediated somatic mutagenesis offers a genetic mechanism for sequence-specific DNA rearrangements in developmental cancers, including neuroblastomas, Ewing sarcomas, desmoplastic small round cell tumors, rhabdomyosarcomas, small cell carcinomas, and many other young-onset cancers that express PGBD5. Aberrant PGBD5 activity also offers a plausible mechanism for the generation of complex DNA rearrangements typically observed in these tumors, including chromoplexy, and other complex DNA rearrangements (*41*).

Many studies have implicated replication stress as a cause of somatic mutations in cancer, including that induced by high-level SHH signaling in cerebellar GCPs (*42*). However, this concept does not explain how sequence-specific mutations, including those affecting tumor suppressors and oncogenes, initially occur. The results presented here offer a mechanism by which dysregulated SHH developmental signaling can activate PGBD5 mutagenic activity and/or impair its efficient repair. It is also possible that dysregulation of PGBD5 and other sequence-specific nucleases can induce not only somatic mutagenesis by virtue of their cellular nuclease activities but also epigenetic dysregulation via its interactions with chromatin and cellular factors, such as those identified in single-nucleus gene expression profiling of specific tumor cells.

While we favor the idea that PGBD5 acts directly on DNA (*3*, *4*, *6*–*8*), direct biochemical and structural studies will be needed to define the exact enzymatic mechanisms of PGBD5 cellular activities and their developmental regulatory factors, including the possibility that PGBD5 promotes somatic DNA rearrangements through recruitment of other nucleases and chromatin remodeling factors. This is an important direction of future work because two recent publications reported an inability to detect DNA transposition of PGBD5 in cellular assays (*43*, *44*). While PGBD5 can support genomic DNA integration in cells, its cellular activity predominantly involves double-strand DNA breaks, deletions, and other DNA rearrangements, with relatively few precise transposon-specific excisions and insertions (*8*). Thus, human PGBD5 is divergent from "cut-and-paste" transposases such as *Trichoplusia ni* piggyBac, but rather should be considered a transposase-derived protein with domesticated genomic integration and DNA rearrangement activities in human cells.

Although we focused on the domesticated DNA transposase–derived PGBD5 in this study, we propose that the principles and implications of somatic developmental mutators revealed herein extend beyond cancer development (*38*). First, PGBD5 has been recently implicated in normal brain development, where PGBD5 is required for the induction of double-strand DNA breaks in developing neurons and proper neuronal migration and development (*45*, *46*). Second, transposable element–derived genes are among the most ubiquitously present genes in living organisms (*47*), with many transposases domesticated in diverse genetic species and somatic tissues (*48*). One can imagine how domesticated transposases can provide molecular mechanisms for somatic genetic diversification during normal tissue development and, when dysregulated, cause somatic mutations that are increasingly being found as causes of diverse sporadic human diseases.

MATERIALS AND METHODS

Mice

Ptf1a-Cre/+, *Ptch1*$^{fl/fl}$, and *Atoh1-GFP* mice were obtained from M. Hoshino (*49*), B. Wainwright (*50*), and J. Johnson (*22*), respectively. All three lines were initially maintained on mixed backgrounds (SW and C57BL/6J) and subsequently backcrossed with C57BL/6J mice to generate C57BL/6J-background mice. *Atoh1-CreER*^{T2} (*Math1-CreER*^{T2}) (007684) and *R26SmoM2* (005130) mice were obtained from the Jackson Laboratory (*51*, *52*). *Atoh1-CreER*^{T2} mice were maintained on SW background. *Pgbd5*-floxed mice were generated by targeting exon 4 of mouse *Pgbd5* (InGenious Targeting Laboratory; fig. S1A). Targeting vector consisting of *LacZ* and *NeoR* cassettes was electroporated in C57BL/6 embryonic stem cells, and targeted clones were microinjected into Balb/c blastocysts. Resulting chimeras were crossed with C57BL/6 FLP mice to remove the Neo cassette to generate *Pgbd5*-floxed mice, as confirmed using genotyping with SC2 (GAGAGCACCGTTGGTGCATATCAG) and SC4 (AGAGTATGAGCGGGAGAGGAGCAG) (fig. S1A). *Pgbd5*-floxed mice were then crossed with *B6.FVB-Tg(EIIa-cre)C5379Lmgd/J (EIIa-Cre)* mice to generate global *Pgbd5*-deficient mice, as confirmed by genotyping with SC2 and SC5 (TTTCTCAGCTGTCCCCAGCATAGC) primers. *Pgbd5*-deficient mice were backcrossed with C57BL/6J mice for six generations. The MASTR model was adapted from Wojcinski *et al.* (*53*) and Lao *et al.* (*13*) (fig. S2A). *Atoh1-FlpoER/+*, *R26*^{MASTR} *(MA)*, and *R26SmoM2 (SmoM2)* mice were maintained on a mixed background and subsequently outbred into an SW background. Genotyping of *Atoh1-FlpoER/+* and *R26*^{MASTR} were done by PCR using following pairs of forward and reverse primers: GCTCTACTTCATCGCATTCCTTGC (forward) and ATTATTTTTGACACCAGACCAAC (reverse) and GATATCTCACGTACTGACGG (forward) and TGACCAGAGTCATCCTTAGC (reverse), respectively. Offspring were obtained by crossing either *Atoh1-FlpoER/+*; *SmoM2/SmoM2; Pgbd5*$^{fl/+}$ with *MA/+*; Pgbd5$^{fl/fl}$, *Atoh1-FlpoER/+*; *SmoM2/SmoM2*; *Pgbd5*$^{fl/+}$ with *MA/MA* or *Atoh1-FlpoER/+*; *SmoM2/SmoM2*; *Pgbd5*$^{fl/+}$ with *MA/MA*; *Pgbd5*$^{fl/+}$. All experiments were conducted in compliance with protocols approved by the Memorial Sloan Kettering Cancer Center Institutional Animal Care and Use Committee.

Tumor induction

On P2, pups from *Atoh1-CreER*^{T2} × *R26SmoM2* and crossings from the MASTR model were injected with tamoxifen (2 and 200 mg/kg; Sigma-Aldrich, T5648), respectively, in corn oil subcutaneously. Some pups from the MASTR model were injected on P0. For the γH2AX analysis using p22/23 *Atoh1-CreERT; R26SmoM2*, pups were injected with tamoxifen (200 mg/kg) on P0. Study end points included reduced activity, ataxia, and/or domed skulls.

Histology, immunohistochemistry, and immunofluorescence of mouse cerebellum and tumors

Under deep anesthesia, animals were perfused with intracardiac 0.9% saline followed by 4% paraformaldehyde (PFA)/0.1 M phosphate buffer (PB). Brains were dissected, further fixed in 4% PFA/0.1 M PB overnight, and embedded in paraffin blocks. Sagittal sections of cerebella were stained with hematoxylin and eosin. Immunohistochemistry was

performed using anti-Ki67 (ab15580, Abcam) and anti-NeuN (A60, EMD Millipore), respectively. For γH2AX staining of the PNLs, cryosections were used. On P22/23, cerebella were dissected and fixed in 4% PFA/0.1 M PB overnight. After cryopreservation in sucrose, cerebella were embedded in OCT (Thermo Fisher Scientific), and blocks were sectioned sagittally in 10-μm sections using a rotary microtome cryostat (Leica). Sections were stored at −20°C. Antigen retrieval was done in sodium citrate buffer [10 mM sodium citrate and 0.05% Tween20 (pH6.0)] for an hour at 99°C. Anti–phospho-histone H2A.X (Ser^{139}) antibody (JBW301, MilliporeSigma) was used in combination with an Alexa Fluor 488–conjugated anti-mouse immunoglobulin G antibody (Invitrogen).

EdU staining

On P22/23, at 2 hours before sacrifice, EdU (50 mg/kg, Invitrogen) diluted in 0.9% saline was injected intraperitoneally. Cryosections were prepared as detailed above. For double staining with γH2AX, antigen retrieval was performed before EdU staining. Click-iT Plus EdU Alexa647 Imaging kit (Invitrogen) was used to stain EdU^+ nuclei.

Counting of γH2AX foci

The PNLs between lobule VI and VII of the vermis from *Atoh1-CreER^{T2}; R26SmoM2; Pgbd5$^{+/+}$* and *Pgbd5$^{-/-}$* mice were imaged using LSM800 (Zeiss) with a 63× oil lens. A total of 90 nuclei per animal (3 images × 30 nuclei) were counted for γH2AX foci and for the presence of EdU.

Copy number analysis

Upon euthanasia and tumor dissection, genomic DNA was extracted from tumors and matched normal tissues (Transnetyx). Genomic DNA real-time PCR was performed using probes for *Pgbd5* wild-type (Pgbd5-1 WT), *Pgbd5*-floxed (Pgbd5-1 FL), and *Pbgd5*-null (Pgbd5-1 EX) alleles, with probes for *jun* as reference (Transnetyx).

Isolation of mouse cerebellar GCPs

The protocol was adapted from Nakashima *et al.* (*54*) and Lee *et al.* (*55*). Briefly, P5 cerebella were trypsinized at 37°C, followed by deoxyribonuclease (DNase) I treatment. GCPs were then isolated by a Percoll gradient consisting of 60 and 35% Percoll solutions. After centrifugation at 2000*g*, cells at the interface between the 35 and 60% Percoll were collected. The cells were washed twice in phosphate-buffered saline (PBS) before analysis.

Isolation and sorting of PNLs

After euthanasia, cerebella at 3 to 8 weeks were isolated and dissociated by trypsin followed by DNase I treatment (*55*). The cells were resuspended in 0.1% bovine serum albumin/PBS with DNase I. The GFP-positive population representing PNLs were collected by FACS using the FACSMelody system (BD Biosciences). After doublet/triplet and dead cells were excluded, GFP-positive fractions were collected for analysis. Data were analyzed using FlowJo version 10 (BD Biosciences).

Gene expression analysis

Total RNA was extracted from P5 GCPs using RNeasy Plus Micro Kit using on-column DNA digestion protocol (Qiagen). Reverse transcription was performed using qScript (Quanta Biosciences) followed by real-time PCR using KAPA SYBR FAST ROX Low (Roche) on the ViiA 7 Real-Time PCR instrument (Applied Biosystems). The following primers were used for *Gli1* (forward, GAGGTTGGGATGAAGAAGCA; reverse, CTTGTGGTGGAGTCATTGGA), *Pgbd5* (forward, GCGGCCGGAAAGAACTATATC; reverse, CACAGCAGTAGATCCCTTGC), and *Actb* (forward, GAGAAGATCTGGCACCACACC; reverse, GGTCTCAAACATGATCTGGGTC).

RNA ISH/FISH

BaseScope hybridization probes specific for *Pgbd5* exon 4 were generated as per the manufacturer's instructions (ACD, catalog no. 1181898-C1). Upon cardiac perfusion and fixation in 4% PFA/0.1 M PB overnight, dissected brains were washed twice with 30% sucrose/PBS and incubated overnight at 4°C. After cryopreservation in sucrose, brains were embedded in OCT (Thermo Fisher Scientific), and blocks were sectioned sagittally in 10-μm sections using a rotary microtome cryostat (Leica). Sections were stored at −80°C. Cryosections were baked for 1 hour at 60°C and fixed in 4% PFA for 15 min followed by washing in PBS. After dehydration, epitope retrieval treatment with ER2 for 5 min at 95°C and subsequent Protease III treatment for 15 min at 40°C were performed. The probe set was hybridized for 2 hours at 42°C. Signal amplification steps were performed according to the manufacturer's protocol. Fast Red (Leica Bond Polymer Refine Red Detection kit DS9390) was used as chromogen. Hematoxylin was used as a counterstain. Mouse *Ppib* (ACD, catalog no. 701078) and *Bacillus subtilis dapB* (ACD, catalog no. 701018) probe sets were used as positive and negative controls, respectively. An adjacent section of the *Pgbd5* ISH section was used to identify *Cre*-expressing cells, i.e., tumor cells. A *Cre*-specific probe set (ACD, catalog no. 312288-C2) was used for fluorescence ISH (FISH). Cryosections from frozen samples were baked for 1 hour at 60°C. Sections were fixed in 4% PFA for 15 min and washed in PBS followed by dehydration. Epitope retrieval treatment with ER2 for 10 min at 95°C, followed by 10 min of incubation with ACD 2.5 LS Hydrogen Peroxide, was performed. The probe set was hybridized for 2 hours at 42°C. Mouse *Ppib* (ACD, catalog no. 313918) and bacterial *dapB* (ACD, catalog no. 312038) probes were used as positive and negative controls, respectively. The hybridized probes were detected using the RNAscope 2.5 LS Reagent Kit–Brown (ACD, catalog no. 322100) according to the manufacturer's instructions with the following modifications. 3,3′-diaminobenzidine application was omitted and replaced with Alexa Fluor 488 Tyramide signal amplification reagent for 20 min at room temperature (Life Technologies, B40953). After staining, slides were washed in PBS and incubated in 4′,6-diamidino-2-phenylindole (DAPI; 5 μg/ml; Sigma-Aldrich) in PBS for 5 min, rinsed in PBS, and mounted in Mowiol 4-88 (Calbiochem). Slides were stored at −20°C before imaging.

Whole-genome sequencing of mouse MBs

Tissues, including tumor and matched skin or spleen, were harvested from symptomatic mice and flash-frozen using a dry ice and ethanol bath. DNA and RNA were extracted using the Qiagen AllPrep kit, according to the manufacturer's instructions. Paired end libraries (2 × 150 bp) were prepared using Illumina TruSeq DNA PCR-Free kit and sequenced using Illumina HiSeq X at a depth of 80× for tumors and 40× for matched normal tissues. Reads were aligned to the mm10 reference genome with BWA-MEM (version 0.7.15) and processed by eliminating duplicate reads with NovoSort MarkDuplicates (version 3.08.02). SNVs were detected using Mutect2

(version 4.0.5.1), Strelka2 (version 2.9.3), and Lancet (exonic) (version 1.0.7). High-confidence SNVs included those that were detected by more than one caller. SVs were detected using Manta, SvABA (version 0.2.1), and Lumpy (version 0.2.13). For SVs, all those that passed quality filters were included. Sequencing data are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA).

Single-nucleotide mutational signature analysis of mouse MBs

Mutational signature analysis was performed using three main steps: de novo extraction, assignment, and fitting (*56*). For the first step, we ran hierarchical Dirichlet process (https://github.com/nicolaroberts/hdp, c78989b). Then, all extracted signatures were assigned to the COSMIC reference (https://cancer.sanger.ac.uk/signatures/sbs/) (*39*) to define which known mutational processes are active. Last, we applied our recently developed fitting algorithm (https://github.com/UM-Myeloma-Genomics/mmsig, df771ba) to estimate both the presence and the contribution of each mutational signature in each sample (*57*).

Identification of genes affected by SVs in mouse MBs

Genes analyzed included those from the COSMIC cancer gene census (*58*). We considered genes to be putatively affected by SVs if (i) the gene was affected by an SV breakpoint; (ii) the gene overlapped with a duplication, inversion, or deletion; or (iii) the gene resided within 25 kb of a translocation breakpoint. For the Oncoprint depicted in Fig. 3C, we curated putative driver alterations arising from *Pgbd5*-induced SVs by analyzing for recurrence, as well as comparing to genes known to be recurrently mutated in human SHH MBs. Specifically, the putative tumor suppressor and oncogenes were identified as (i) genes affected by SVs in ≥4 *Ptch1*-mutant tumors, (ii) genes affected by point mutations in ≥2 tumors, (iii) genes affected by SVs in ≥2 tumors that were also affected by SVs in ≥10% of human SHH MBs, or (iv) genes affected by SVs that were demonstrated to be recurrently mutated by Northcott *et al.* (*10*). These SVs were further analyzed for their chromosomal loci. By using the UCSC cytoband annotation file, the cytobands corresponding to the region in which each breakpoint of SVs resides were annotated (fig. S9 and data S20).

Whole-genome sequencing of human MBs

All patient material was collected after receiving written informed consent, which includes consent to publish the data, as approved by the Medulloblastoma Advanced Genomics International Consortium and the International Cancer Genome Consortium (*27*). Somatic SVs were called using Genomon-SV (v0.4.1) and DELLY2 (v0.7.5) as described by Skowron *et al.* (*27*). For each SV detected by either algorithm, we used Velvet to assemble reads around the detected breakpoints, and the resultant contigs were then remapped locally using human genome reference sequences with and without detected SVs using blat. This approach ensures detection of heterozygous SVs, which would be mapped both to reference sequence without incorporated variant contigs and to reference sequence that incorporates them. Subsequently, we selected variants for which assembled contigs could be mapped to the reference sequence containing somatic SVs, and variants from matched normal tissue that could not be mapped to the reference sequence that incorporates variant contig sequences. Sequencing data are available from European Genome-phenome Archive (EGA) using the following accession numbers: EGAD00001003125, EGAD00001004347, and EGAD00001003127.

Identification of breakpoint sequence motifs in mouse MBs

Sequences ±50 bp flanking both breakpoints from identified SVs were extracted with bedtools (version 2.29.2) using the mm10 reference genome. To identify motifs that are putatively associated with Pgbd5 activity in mouse tumors, we used de novo motif analysis of the breakpoint sequences from *Pgbd5*-expressing tumors and the breakpoint sequences from *Pgbd5*-deficient tumors as background. We used discriminative MEME (v5.4.1) (*59*), which determines relative sequence enrichment in an experimental set of sequences relative to a control set. *Pgbd5*-deficient breakpoint sequences were used as controls. Discriminative MEME was used with default parameters but limited to between 11 and 16 bp. Repetitive motifs were eliminated, and the remainder were chosen as putative *Pgbd5*-associated motifs. For SmoM2-mutant tumors, motifs were identified using classic MEME with default parameters and limited to between 11 and 16 bp.

Identification of breakpoint sequence motifs in human MBs

Sequences ±50 bp flanking both breakpoints from identified SVs were extracted with bedtools (version 2.29.2) using the hg19 reference genome. To identify de novo motifs, classic MEME (version 5.1.1) (*60*) with default parameters but limited to 11 bp was used (http://meme-suite.org/). Putative PSS motifs were selected by eliminating repetitive motifs and by determining whether candidate motifs were found when discriminative MEME was run relative to 50,000 randomly selected 50-mers from the hg19 reference genome without repeat masking. To quantify the abundance of motifs at SV breakpoints, we used FIMO (version 5.1.1) (*60*) with default parameters (fig. S14). *Q*-value cutoffs for quantitation were determined by identifying a threshold where previously identified PSS motifs (Rhabdoid and RPE) could be specifically detected, but negative control motifs (RAG1/2 and scrambled) were not. This cutoff was characterized by a receiver operating characteristic (ROC) curve and chosen for sensitivity of 100% and specificity of 75% for the previously identified PSS motifs. For the quantifications, each SV is counted only once (i.e., each SV has 4 × 50-mers and 2× breakpoints). As a comparator, we included human breast carcinoma SVs (*4*, *28*). To mimic the background distribution of the FIMO algorithm, we used randomly scrambled sequences.

Sequence motif comparisons

The similarity of motifs was evaluated using TomTom (https://meme-suite.org/meme/tools/tomtom) (*61*). The first set of query motifs that included those from three *Ptch1*-mutant tumors and six *SmoM2*-mutant tumors were compared with previously identified target motifs [PSS motifs: including 11 motifs (PSS_1 to 11), PSS_Rhab, PSS_RPE, and PSS_HPRT]. To test for similarities, 10,000 shuffled target motifs were generated and their *P* values were examined. If the *P* value of the target motif ranked within the lowest 5% of all 10,001 *P* values, motif was deemed significantly similar. In the same way, the second set of query motifs (hMB1 to hMB4) were compared with nine mouse motifs and their shuffled sequences. Last, the query motif sets (hMB1 to hMB4) were compared with the previously identified 14 PSS motifs and their shuffled sequences. A Circos plot (Fig. 3F) was generated to show similarities among all

motifs. Line thickness is proportional to degree of similarity, as indicated. The mouse MB motifs include three *Ptch1*-mutant motifs in blue (Ptch1_mMB1-3) and six *SmoM2*-mutant motifs in green (Smo_mMB1-6). The human MB motifs are colored purple (hMB1 to hMB4). The previously identified PSS motifs include 14 motifs in gray (PSS_1-11, PSS_Rhab, PSS_RPE, and PSS_HPRT) (*4, 24*).

Single-nuclei multiome ATAC + gene expression of mouse MBs

Nuclei extraction

Frozen pieces (~50 mg) of tumors from *Ptf1a-Cre*/+; *Ptch1*$^{fl/fl}$; *Pgbd5*$^{+/+}$ ($n = 3$) and *Pgbd5*$^{-/-}$ ($n = 3$) were used for nuclei isolation. Nuclei extraction was performed according to the protocol from Masilionis *et al.* (*62*). Singlulator100 (S2 Genomics) was used for nuclei extraction. Extracted nuclei were stained with 7-aminoactinomycin D (Invitrogen) and FACS-sorted for further nuclei purification. Ten thousand nuclei were targeted for library construction using ChromiumNextGEM Multiome ATAC + Gene Expression kit (10x Genomics).

Single-cell multiome data processing and QC

Cell Ranger ARC v2.0.0 (10x Genomics) was used ("count" option with default parameters) to filter and align raw reads, identify transposase cut sites, detect accessible chromatin peaks, call cells, and generate raw count matrices for scMultiome samples. Alignment was performed using the mm10 reference genome build coupled with the Ensembl 98 gene annotation. Reads that mapped to the intronic regions were excluded for the RNA modality.

Quality control (QC) and data processing steps were performed using Signac v1.3.0 (*63*) and Seurat v4.3.0 (*64*). QC metrics for RNA and Assay for Transposase-Accessible Chromatin (ATAC) modalities were calculated independently but were jointly used to filter cells. A combination of thresholds was established for each sample based on hard cutoffs or on the distribution of each metric within the sample (data S21). In the RNA modality, cells were filtered on the number of genes, unique molecular identifiers (UMIs), and mitochondrial content. In the ATAC modality, cells were filtered on the number of peaks detected, transcription start site enrichment, and nucleosome signal.

Normalization and dimensionality reduction in single-cell multiome data

For the RNA modality, libraries were scaled to 10,000 UMIs per cell and log-normalized. UMI counts and mitochondrial content were regressed out from normalized gene counts, and the residuals were *z*-scored gene-wise. Dimensionality reduction was performed using principal components analysis (PCA) on the top 2000 most variable features. For the ATAC modality, peaks were called using MACS2 (v2.2.7.1) (*65*) using the CallPeaks function from Signac library with default parameters. ATAC reads were quantified in each peak per cell, and a resulting count matrix was generated. Dimensionality reduction was performed using latent semantic indexing (LSI) (*66*).

A weighted nearest neighbor graph was constructed between all cells using the first 30 principal components from the RNA data and the top six dimensions of the LSI reduction from the ATAC data with the following default parameters: 20 multimodal neighbors, 200 approximate neighbors, and L2 normalization enabled. This weighted nearest neighbor graph was used as input for projection into two dimensions using Uniform Manifold Approximation and Projection (UMAP) (*67*) and for clustering using a shared nearest-neighbor (SNN) algorithm (*64*) based on the Louvain algorithm on a *k*-nearest neighbor graph with $k = 20$ and resolution 0.2.

Joint sample integration and visualization in single-nuclei RNA data

RNA libraries from all samples were merged and processed as described above for normalization and dimensionality reduction with a minor change (no variables were regressed out). To visualize the samples in a shared UMAP space, Harmony (v0.1.1) (*68*) was run to integrate the samples using the first 30 principal components as input and regressing out the differences between samples. The resulting batch-corrected embedding and the top 30 dimensions were used as input for projection into two dimensions (UMAP) and for clustering (SNN algorithm) with $k = 20$ and resolution 0.5.

Cell type annotation in single-nuclei RNA data

Annotation of cell types was performed using four different reference-based annotation tools: three machine learning–based prediction methods [SciBet (*69*), SingleCellNet (*70*), and SingleR (*71*)] and a statistical model [Spearman correlation as previously described (*72*)]. A consensus label annotation was assigned when at least two methods agreed.

Cell type annotation was performed using a murine cerebellar developmental atlas (*31*) [nine developmental samples ranging from time points E10 (embryonic day 10) to P14; $N = 62{,}040$ cells] as the reference. To obtain more granular clusters in the reference, we reprocessed and clustered each individual reference sample as described above, resulting in 209 new clusters. To label these new clusters, we used a combination of age, new cluster number, and cell class information from the original publication (*31*). Cell class was assigned on the basis of the proportion of the original cell type in the new cluster. If the proportion of one original cell type was greater than 75% in the new cluster, that label was used. In cases where this was not achieved, a label consisting of two prominent cell type labels was used. The new cluster labels were used to perform annotation of cell types in the mouse MB samples.

Malignant cell identification in single-nuclei RNA data

Copy number variants (CNVs) were inferred on a sample basis using inferCNV (v1.10.1) (*73*) with the following parameters: cutoff = 0.1, window_length = 101, analysis_mode = "samples," cluster_by_groups = FALSE, sd_amplifier = 1.5, HMM = FALSE, and denoise = TRUE. The mitochondrial chromosome (defined as having gene symbols starting with "mt"-), ribosomal genes (defined as having gene symbols matching "Rps," "Rpl," "Mrps," "Mrpl"), and MHC genes (defined as having gene symbols starting with "H2"-) were excluded from the CNV inference. A normal reference was assembled selecting clusters composed of a single cell type and originating from multiple samples (immune and oligodendrocyte precursor cell clusters) in the integrated snRNA-seq space. Last, hierarchical clustering of cells based on their CNV profiles was used to identify subtrees of malignant cells with prominent copy number signal, distinguishing them from normal cells lacking CNV signal.

Differential expression in single-nuclei RNA data

To identify differentially expressed genes (DEGs) between different genotypes within malignant cell populations, we used two different strategies. The first approach used a Wilcoxon rank sum test applied through the FindMarkers function from the Seurat package (logFC.threshold = 0, min.pct = 0.05). For the second approach, pseudobulk RNA files were generated and processed for the targeted cell populations of each sample. Adaptor sequences and the first four nucleotides of each read were removed from the pseudobulk read

sets using Trimmomatic (v0.39) (*74*). Reads were scanned from the 5′ end and truncated when the average quality of a 4-nucleotide sliding window fell below a threshold (phred33 < 30). Short reads after trimming (< 30 bp) were discarded. High-quality reads were aligned to the mouse reference genome build mm10 using STAR (v2.7.9a) (*75*) with default parameters. Multimapping reads (MAPQ < 1) were discarded from downstream analysis. Gene expression levels were estimated by quantifying reads mapping to at most two locations (MAPQ ≥ 3) to exonic regions (the maximal genomic locus of each gene and its known isoforms) using featureCounts (v2.0.3) (*76*) and the mm10 ensGene annotation set from Ensembl. Gene expression levels were estimated by quantifying primary alignments mapping to at most two locations (MAPQ ≥ 3) to exonic regions (the maximal genomic locus of each gene and its known isoforms) using featureCounts (v2.0.3) (*76*) and the mm10 ensGene annotation set from Ensembl. Normalization (mean of ratios) of the data and differential gene expression analysis were performed using DESeq2 (v1.30.1) (*77*) with the Wald test. Genes that pass significance in both strategies ($\log_2$FC > 0.25 and adjusted *P* value > 0.05) are DEGs with strongest confidence.

Supplementary Materials

This PDF file includes:
Materials and Methods
Figs. S1 to S8
Tables S1 and S2
Legends for data S1 to S22
References

Other Supplementary Material for this manuscript includes the following:
Data S1 to S22

REFERENCES AND NOTES

1. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
2. N. D. Anderson, R. de Borja, M. D. Young, F. Fuligni, A. Rosic, N. D. Roberts, S. Hajjar, M. Layeghifard, A. Novokmet, P. E. Kowalski, M. Anaka, S. Davidson, M. Zarrei, B. Id Said, L. C. Schreiner, R. Marchand, J. Sitter, N. Gokgoz, L. Brunga, G. T. Graham, A. Fullam, N. Pillay, J. A. Toretsky, A. Yoshida, T. Shibata, M. Metzler, G. R. Somers, S. W. Scherer, A. M. Flanagan, P. J. Campbell, J. D. Schiffman, M. Shago, L. B. Alexandrov, J. S. Wunder, I. L. Andrulis, D. Malkin, S. Behjati, A. Shlien, Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science* **361**, eaam8419 (2018).
3. A. G. Henssen, E. Henaff, E. Jiang, A. R. Eisenberg, J. R. Carson, C. M. Villasante, M. Ray, E. Still, M. Burns, J. Gandara, C. Feschotte, C. E. Mason, A. Kentsis, Genomic DNA transposition induced by human PGBD5. *eLife* **4**, e10565 (2015).
4. A. G. Henssen, R. Koche, J. Zhuang, E. Jiang, C. Reed, A. Eisenberg, E. Still, I. C. MacArthur, E. Rodríguez-Fos, S. Gonzalez, M. Puiggròs, A. N. Blackford, C. E. Mason, E. de Stanchina, M. Gönen, A. K. Emde, M. Shah, K. Arora, C. Reeves, N. D. Socci, E. Perlman, C. R. Antonescu, C. W. M. Roberts, H. Steen, E. Mullen, S. P. Jackson, D. Torrents, Z. Weng, S. A. Armstrong, A. Kentsis, PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nat. Genet.* **49**, 1005–1014 (2017).
5. A. G. Henssen, C. Reed, E. Jiang, H. D. Garcia, J. von Stebut, I. C. MacArthur, P. Hundsdoerfer, J. H. Kim, E. de Stanchina, Y. Kuwahara, H. Hosoi, N. J. Ganem, F. dela Cruz, A. L. Kung, J. H. Schulte, J. H. Petrini, A. Kentsis, Therapeutic targeting of PGBD5-induced DNA repair dependency in pediatric solid tumors. *Sci. Transl. Med.* **9**, eaam9078 (2017).
6. L. Helou, L. Beauclair, H. Dardente, P. Arensburger, N. Buisine, Y. Jaszczyszyn, F. Guillou, T. Lecomte, A. Kentsis, Y. Bigot, The C-terminal domain of piggyBac transposase is not required for DNA transposition. *J. Mol. Biol.* **433**, 166805 (2021).
7. L. Helou, L. Beauclair, H. Dardente, B. Piégu, L. Tsakou-Ngouafo, T. Lecomte, A. Kentsis, P. Pontarotti, Y. Bigot, The piggyBac-derived protein 5 (PGBD5) transposes both the closely and the distantly related piggyBac-like elements Tcr-pble and Ifp2. *J. Mol. Biol.* **433**, 166839 (2021).
8. Y. Bigot, M. Yamada, H. Mueller, V. Morell, S. Alves, T. Lecomte, A. Kentsis, Analysis of DNA transposition by DNA transposases in human cells. bioRxiv 2023.04.26.538406 [Preprint] (2023). https://doi.org/10.1101/2023.04.26.538406.
9. G. A. Suero-Abreu, G. Praveen Raju, O. Aristizábal, E. Volkova, A. Wojcinski, E. J. Houston, D. Pham, K. U. Szulc, D. Colon, A. L. Joyner, D. H. Turnbull, In vivo Mn-enhanced MRI for early tumor detection and growth rate analysis in a mouse medulloblastoma model. *Neoplasia* **16**, 993–1006 (2014).
10. P. A. Northcott, I. Buchhalter, A. S. Morrissy, V. Hovestadt, J. Weischenfeldt, T. Ehrenberger, S. Gröbner, M. Segura-Wang, T. Zichner, V. A. Rudneva, H. J. Warnatz, N. Sidiropoulos, A. H. Phillips, S. Schumacher, K. Kleinheinz, S. M. Waszak, S. Erkek, D. T. W. Jones, B. C. Worst, M. Kool, M. Zapatka, N. Jäger, L. Chavez, B. Hutter, M. Bieg, N. Paramasivam, M. Heinold, Z. Gu, N. Ishaque, C. Jäger-Schmidt, C. D. Imbusch, A. Jugold, D. Hübschmann, T. Risch, V. Amstislavskiy, F. G. R. Gonzalez, U. D. Weber, S. Wolf, G. W. Robinson, X. Zhou, G. Wu, D. Finkelstein, Y. Liu, F. M. G. Cavalli, B. Luu, V. Ramaswamy, X. Wu, J. Koster, M. Ryzhova, Y. J. Cho, S. L. Pomeroy, C. Herold-Mende, M. Schuhmann, M. Ebinger, L. M. Liau, J. Mora, R. E. McLendon, N. Jabado, T. Kumabe, E. Chuah, Y. Ma, R. A. Moore, A. J. Mungall, K. L. Mungall, N. Thiessen, K. Tse, T. Wong, S. J. M. Jones, O. Witt, T. Milde, A. von Deimling, D. Capper, A. Korshunov, M. L. Yaspo, R. Kriwacki, A. Gajjar, J. Zhang, R. Beroukhim, E. Fraenkel, J. O. Korbel, B. Brors, M. Schlesner, R. Eils, M. A. Marra, S. M. Pfister, M. D. Taylor, P. Lichter, The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
11. K. M. Reilly, The effects of genetic background of mouse models of cancer: Friend or foe? *Cold Spring Harb. Protoc.* **2016**, pdb top076273 (2016).
12. J. Mao, K. L. Ligon, E. Y. Rakhlin, S. P. Thayer, R. T. Bronson, D. Rowitch, A. P. McMahon, A novel somatic mouse model to survey tumorigenic potential applied to the Hedgehog pathway. *Cancer Res.* **66**, 10171–10178 (2006).
13. Z. Lao, G. P. Raju, C. B. Bai, A. L. Joyner, MASTR: A technique for mosaic mutant analysis with spatial and temporal control of recombination using conditional floxed alleles in mice. *Cell Rep.* **2**, 386–396 (2012).
14. I. L. Tan, A. Wojcinski, H. Rallapalli, Z. Lao, R. M. Sanghrajka, D. Stephen, E. Volkova, A. Korshunov, M. Remke, M. D. Taylor, D. H. Turnbull, A. L. Joyner, Lateral cerebellum is preferentially sensitive to high sonic hedgehog signaling and medulloblastoma formation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3392–3397 (2018).
15. P. Gibson, Y. Tong, G. Robinson, M. C. Thompson, D. S. Currle, C. Eden, T. A. Kranenburg, T. Hogg, H. Poppleton, J. Martin, D. Finkelstein, S. Pounds, A. Weiss, Z. Patay, M. Scoggins, R. Ogg, Y. Pei, Z. J. Yang, S. Brun, Y. Lee, F. Zindy, J. C. Lindsey, M. M. Taketo, F. A. Boop, R. A. Sanford, A. Gajjar, S. C. Clifford, M. F. Roussel, P. J. McKinnon, D. H. Gutmann, D. W. Ellison, R. Wechsler-Reya, R. J. Gilbertson, Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**, 1095–1099 (2010).
16. S. Jessa, A. Blanchet-Cohen, B. Krug, M. Vladoiu, M. Coutelier, D. Faury, B. Poreau, N. de Jay, S. Hébert, J. Monlong, W. T. Farmer, L. K. Donovan, Y. Hu, M. K. McConechy, F. M. G. Cavalli, L. G. Mikael, B. Ellezam, M. Richer, A. Allaire, A. G. Weil, J. Atkinson, J. P. Farmer, R. W. R. Dudley, V. Larouche, L. Crevier, S. Albrecht, M. G. Filbin, H. Sartelet, P. E. Lutz, C. Nagy, G. Turecki, S. Costantino, P. B. Dirks, K. K. Murai, G. Bourque, J. Ragoussis, L. Garzia, M. D. Taylor, N. Jabado, C. L. Kleinman, Stalled developmental programs at the root of pediatric brain tumors. *Nat. Genet.* **51**, 1702–1713 (2019).
17. R. J. Wechsler-Reya, M. P. Scott, Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron* **22**, 103–114 (1999).
18. L. V. Goodrich, L. Milenkovic, K. M. Higgins, M. P. Scott, Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109–1113 (1997).
19. J. D. Corrales, S. Blaess, E. M. Mahoney, A. L. Joyner, The level of sonic hedgehog signaling regulates the complexity of cerebellar foliation. *Development* **133**, 1811–1821 (2006).
20. A. Flora, T. J. Klisch, G. Schuster, H. Y. Zoghbi, Deletion of *Atoh1* disrupts Sonic Hedgehog signaling in the developing cerebellum and prevents medulloblastoma. *Science* **326**, 1424–1427 (2009).
21. T. G. Oliver, T. A. Read, J. D. Kessler, A. Mehmeti, J. F. Wells, T. T. T. Huynh, S. M. Lin, R. J. Wechsler-Reya, Loss of patched and disruption of granule cell development in a pre-neoplastic stage of medulloblastoma. *Development* **132**, 2425–2439 (2005).
22. E. A. Lumpkin, T. Collisson, P. Parab, A. Omer-Abdalla, H. Haeberle, P. Chen, A. Doetzlhofer, P. White, A. Groves, N. Segil, J. E. Johnson, Math1-driven GFP expression in the developing nervous system of transgenic mice. *Gene Expr. Patterns* **3**, 389–395 (2003).
23. T. Pavelitz, L. T. Gray, S. L. Padilla, A. D. Bailey, A. M. Weiner, PGBD5: A neural-specific intron-containing piggyBac transposase domesticated over 500 million years ago and conserved from cephalochordates to humans. *Mob. DNA* **4**, 23 (2013).
24. A. G. Henssen, E. Jiang, J. Zhuang, L. Pinello, N. D. Socci, R. Koche, M. Gonen, C. M. Villasante, S. A. Armstrong, D. E. Bauer, Z. Weng, A. Kentsis, Forward genetic screen of human transposase genomic rearrangements. *BMC Genomics* **17**, 548 (2016).
25. K. Arora, M. Shah, M. Johnson, R. Sanghvi, J. Shelton, K. Nagulapalli, D. M. Oschwald, M. C. Zody, S. Germer, V. Jobanputra, J. Carter, N. Robine, Deep whole-genome sequencing of 3 cancer cell lines on 2 sequencing platforms. *Sci. Rep.* **9**, 19123 (2019).
26. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. J. R. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A.-L. Børresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjörd, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jäger, D. T. W. Jones, D. Jones,

S. Knappskog, M. Kool, S. R. Lakhani, C. López-Otín, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. J. Tutt, R. Valdés-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain, J. Zucman-Rossi, P. A. Futreal, U. M. Dermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, M. R. Stratton, Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

27. P. Skowron, H. Farooq, F. M. G. Cavalli, A. S. Morrissy, M. Ly, L. D. Hendrikse, E. Y. Wang, H. Djambazian, H. Zhu, K. L. Mungall, Q. M. Trinh, T. Zheng, S. Dai, A. S. G. Stucklin, M. C. Vladoiu, V. Fong, B. L. Holgado, C. Nor, X. Wu, D. Abd-Rabbo, P. Bérubé, Y. C. Wang, B. Luu, R. A. Suarez, A. Rastan, A. H. Gillmor, J. J. Y. Lee, X. Y. Zhang, C. Daniels, P. Dirks, D. Malkin, E. Bouffet, U. Tabori, J. Loukides, F. P. Doz, F. Bourdeaut, O. O. Delattre, J. Masliah-Planchon, O. Ayrault, S. K. Kim, D. Meyronet, W. A. Grajkowska, C. G. Carlotti, C. de Torres, J. Mora, C. G. Eberhart, E. G. van Meir, T. Kumabe, P. J. French, J. M. Kros, N. Jabado, B. Lach, I. F. Pollack, R. L. Hamilton, A. A. N. Rao, C. Giannini, J. M. Olson, L. Bognár, A. Klekner, K. Zitterbart, J. J. Phillips, R. C. Thompson, M. K. Cooper, J. B. Rubin, L. M. Liau, M. Garami, P. Hauser, K. K. W. Li, H. K. Ng, W. S. Poon, G. Yancey Gillespie, J. A. Chan, S. Jung, R. E. McLendon, E. M. Thompson, D. Zagzag, R. Vibhakar, Y. S. Ra, M. L. Garre, U. Schüller, T. Shofuda, C. C. Faria, E. López-Aguilar, G. Zadeh, C. C. Hui, V. Ramaswamy, S. D. Bailey, S. J. Jones, A. J. Mungall, R. A. Moore, J. A. Calarco, L. D. Stein, G. D. Bader, J. Reimand, J. Ragoussis, W. A. Weiss, M. A. Marra, H. Suzuki, M. D. Taylor, The transcriptional landscape of Shh medulloblastoma. *Nat. Commun.* **12**, 1749 (2021).
28. J. Zhuang, Z. Weng, Local sequence assembly reveals a high-resolution profile of somatic structural variations in 97 cancer genomes. *Nucleic Acids Res.* **43**, 8146–8156 (2015).
29. T. J. Pugh, S. D. Weeraratne, T. C. Archer, D. A. Pomeranz Krummel, D. Auclair, J. Bochicchio, M. O. Carneiro, S. L. Carter, K. Cibulskis, R. L. Erlich, H. Greulich, M. S. Lawrence, N. J. Lennon, A. McKenna, J. Meldrim, A. H. Ramos, M. G. Ross, C. Russ, E. Shefler, A. Sivachenko, B. Sogoloff, P. Stojanov, P. Tamayo, J. P. Mesirov, V. Amani, N. Teider, S. Sengupta, J. P. Francois, P. A. Northcott, M. D. Taylor, F. Yu, G. R. Crabtree, A. G. Kautzman, S. B. Gabriel, G. Getz, N. Jäger, D. T. W. Jones, P. Lichter, S. M. Pfister, T. M. Roberts, M. Meyerson, S. L. Pomeroy, Y. J. Cho, Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110 (2012).
30. M. Kool, D. T. Jones, N. Jäger, P. A. Northcott, T. J. Pugh, V. Hovestadt, R. M. Piro, L. A. Esparza, S. L. Markant, M. Remke, T. Milde, F. Bourdeaut, M. Ryzhova, D. Sturm, E. Pfaff, S. Stark, S. Hutter, H. Seker-Cin, P. Johann, S. Bender, C. Schmidt, T. Rausch, D. Shih, J. Reimand, L. Sieber, A. Wittmann, L. Linke, H. Witt, U. D. Weber, M. Zapatka, R. König, R. Beroukhim, G. Bergthold, P. van Sluis, R. Volckmann, J. Koster, R. Versteeg, S. Schmidt, S. Wolf, C. Lawerenz, C. C. Bartholomae, C. von Kalle, A. Unterberg, C. Herold-Mende, S. Hofer, A. E. Kulozik, A. von Deimling, W. Scheurlen, J. Felsberg, G. Reifenberger, M. Hasselblatt, J. R. Crawford, G. A. Grant, N. Jabado, A. Perry, C. Cowdrey, S. Croul, G. Zadeh, J. O. Korbel, F. Doz, O. Delattre, G. D. Bader, M. McCabe, V. P. Collins, M. W. Kieran, Y. J. Cho, S. L. Pomeroy, O. Witt, B. Brors, M. D. Taylor, U. Schüller, A. Korshunov, R. Eils, R. J. Wechsler-Reya, P. Lichter, S. M. Pfister; ICGC PedBrain Tumor Project, Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**, 393–405 (2014).
31. M. C. Vladoiu, I. el-Hamamy, L. K. Donovan, H. Farooq, B. L. Holgado, Y. Sundaravadanam, V. Ramaswamy, L. D. Hendrikse, S. Kumar, S. C. Mack, J. J. Y. Lee, V. Fong, K. Juraschka, D. Przelicki, A. Michealraj, P. Skowron, B. Luu, H. Suzuki, A. S. Morrissy, F. M. G. Cavalli, L. Garzia, C. Daniels, X. Wu, M. A. Qazi, S. K. Singh, J. A. Chan, M. A. Marra, D. Malkin, P. Dirks, L. Heisler, T. Pugh, K. Ng, F. Notta, E. M. Thompson, C. L. Kleinman, A. L. Joyner, N. Jabado, L. Stein, M. D. Taylor, Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* **572**, 67–73 (2019).
32. T. Kato, N. Sato, S. Hayama, T. Yamabuki, T. Ito, M. Miyamoto, S. Kondo, Y. Nakamura, Y. Daigo, Activation of Holliday junction recognizing protein involved in the chromosomal stability and immortality of cancer cells. *Cancer Res.* **67**, 8544–8553 (2007).
33. D. R. Foltz, L. E. T. Jansen, A. O. Bailey, J. R. Yates III, E. A. Bassett, S. Wood, B. E. Black, D. W. Cleveland, Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell* **137**, 472–484 (2009).
34. E. M. Dunleavy, D. Roche, H. Tagami, N. Lacoste, D. Ray-Gallet, Y. Nakamura, Y. Daigo, Y. Nakatani, G. Almouzni-Pettinotti, HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485–497 (2009).
35. S. Lutolf, F. Radtke, M. Aguet, U. Suter, V. Taylor, Notch1 is required for neuronal and glial differentiation in the cerebellum. *Development* **129**, 373–385 (2002).
36. K. Kitamura, M. Yanazawa, N. Sugiyama, H. Miura, A. Iizuka-Kogo, M. Kusaka, K. Omichi, R. Suzuki, Y. Kato-Fukui, K. Kamiirisa, M. Matsuo, S. I. Kamijo, M. Kasahara, H. Yoshioka, T. Ogata, T. Fukuda, I. Kondo, M. Kato, W. B. Dobyns, M. Yokoyama, K. I. Morohashi, Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nat. Genet.* **32**, 359–369 (2002).
37. A. L. Law, S. Jalal, T. Pallett, F. Mosis, A. Guni, S. Brayford, L. Yolland, S. Marcotti, J. A. Levitt, S. P. Poland, M. Rowe-Sampson, A. Jandke, R. Köchl, G. Pula, S. M. Ameer-Beg, B. M. Stramer, M. Krause, Nance-Horan Syndrome-like 1 protein negatively regulates Scar/WAVE-Arp2/3 activity and inhibits lamellipodia stability and cell migration. *Nat. Commun.* **12**, 5687 (2021).
38. A. Kentsis, S. A. Frank, Developmental mutators and early onset cancer. *Front. Pediatr.* **8**, 189 (2020).
39. L. B. Alexandrov, J. Kim, N. J. Haradhvala, M. N. Huang, A. W. T. Ng, Y. Wu, A. Boot, K. R. Covington, D. A. Gordenin, E. N. Bergstrom, S. M. A. Islam, N. Lopez-Bigas, L. J. Klimczak, J. R. McPherson, S. Morganella, R. Sabarinathan, D. A. Wheeler, V. Mustonen; PCAWG Mutational Signatures Working Group, G. Getz, S. G. Rozen, M. R. Stratton; PCAWG Consortium, The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
40. A. Degasperi, X. Zou, T. D. Amarante, A. Martinez-Martinez, G. C. C. Koh, J. M. L. Dias, L. Heskin, L. Chmelova, G. Rinaldi, V. Y. W. Wang, A. S. Nanda, A. Bernstein, S. E. Momen, J. Young, D. Perez-Gil, Y. Memari, C. Badja, S. Shooter, J. Czarnecki, M. A. Brown, H. R. Davies; Genomics England Research Consortium, S. Nik-Zainal, Substitution mutational signatures in whole-genome–Sequenced cancers in the UK population. *Science* **376**, science.abl.9283 (2022).
41. K. Hadi, X. Yao, J. M. Behr, A. Deshpande, C. Xanthopoulakis, H. Tian, S. Kudman, J. Rosiene, M. Darmofal, J. DeRose, R. Mortensen, E. M. Adney, A. Shaiber, Z. Gajic, M. Sigouros, K. Eng, J. A. Wala, K. O. Wrzeszczyński, K. Arora, M. Shah, A. K. Emde, V. Felice, M. O. Frank, R. B. Darnell, M. Ghandi, F. Huang, S. Dewhurst, J. Maciejowski, T. de Lange, J. Setton, N. Riaz, J. S. Reis-Filho, S. Powell, D. A. Knowles, E. Reznik, B. Mishra, R. Beroukhim, M. C. Zody, N. Robine, K. M. Oman, C. A. Sanchez, M. K. Kuhner, L. P. Smith, P. C. Galipeau, T. G. Paulson, B. J. Reid, X. Li, D. Wilkes, A. Sboner, J. M. Mosquera, O. Elemento, M. Imielinski, Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. *Cell* **183**, 197–210.e32 (2020).
42. L. Tamayo-Orrego, D. Gallo, F. Racicot, A. Bemmo, S. Mohan, B. Ho, S. Salameh, T. Hoang, A. P. Jackson, G. W. Brown, F. Charron, Sonic hedgehog accelerates DNA replication to cause replication stress promoting cancer initiation in medulloblastoma. *Nat. Cancer* **1**, 840–854 (2021).
43. T. M. Beckermann, W. Luo, C. M. Wilson, R. A. Veach, M. H. Wilson, Cognate restriction of transposition by piggyBac-like proteins. *Nucleic Acids Res.* **49**, 8135–8144 (2021).
44. O. Kolacsek, G. Wachtl, Á. Fóthi, A. Schamberger, S. Sándor, E. Pergel, N. Varga, T. Raskó, Z. Izsvák, Á. Apáti, T. I. Orbán, Functional indications for transposase domestications—Characterization of the human piggyBac transposase derived (PGBD) activities. *Gene* **834**, 146609 (2022).
45. A. Simi, F. Ansaloni, D. Damiani, A. Codino, D. Mangoni, P. Lau, D. Vozzi, L. Pandolfini, R. Sanges, S. Gustincich, The Pgbd5 DNA transposase is required for mouse cerebral cortex development through DNA double-strand breaks formation. bioRxiv 2023.05.09.539730 [Preprint] (2023). https://doi.org/10.1101/2023.05.09.539730.
46. L. J. Zapater, E. Rodriguez-Fos, M. Planas-Felix, S. Lewis, Daniel, Cameron, P. Demarest, A. Nabila, J. Zhao, P. Bergin, C. Reed, Makiko, Yamada, N. D. Socci, M. Hayes, R. Rabadan, D. Torrents, Michael C., Kruer, M. Toth, A. Kentsis, A transposase-derived gene required for human brain development. bioRxiv 2023.04.28.538770 [Preprint] (2023). https://doi.org/10.1101/2023.04.28.538770**.**
47. R. K. Aziz, M. Breitbart, R. A. Edwards, Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* **38**, 4207–4217 (2010).
48. R. L. Cosby, J. Judd, R. Zhang, A. Zhong, N. Garry, E. J. Pritham, C. Feschotte, Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* **371**, eabc6405 (2021).
49. M. Hoshino, S. Nakamura, K. Mori, T. Kawauchi, M. Terao, Y. V. Nishimura, A. Fukuda, T. Fuse, N. Matsuo, M. Sone, M. Watanabe, H. Bito, T. Terashima, C. V. E. Wright, Y. Kawaguchi, K. Nakao, Y. I. Nabeshima, Ptf1a, a bHLH transcriptional gene, defines GABAergic neuronal fates in cerebellum. *Neuron* **47**, 201–213 (2005).
50. T. Ellis, I. Smyth, E. Riley, S. Graham, K. Elliot, M. Narang, G. F. Kay, C. Wicking, B. Wainwright, Patched 1 conditional null allele in mice. *Genesis* **36**, 158–161 (2003).
51. R. Machold, G. Fishell, Math1 is expressed in temporally discrete pools of cerebellar rhombic-lip neural progenitors. *Neuron* **48**, 17–24 (2005).
52. J. Jeong, J. Mao, T. Tenzen, A. H. Kottmann, A. P. McMahon, Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev.* **18**, 937–951 (2004).
53. A. Wojcinski, A. K. Lawton, N. S. Bayin, Z. Lao, D. N. Stephen, A. L. Joyner, Cerebellar granule cell replenishment postinjury by adaptive reprogramming of Nestin$^+$ progenitors. *Nat. Neurosci.* **20**, 1361–1370 (2017).
54. K. Nakashima, H. Umeshima, M. Kengaku, Cerebellar granule cells are predominantly generated by terminal symmetric divisions of granule cell precursors. *Dev. Dyn.* **244**, 748–758 (2015).

55. H. Y. Lee, L. A. Greene, C. A. Mason, M. C. Manzini, Isolation and culture of post-natal mouse cerebellar granule neuron progenitor cells and neurons. *J. Vis. Exp.*, 990 (2009).
56. F. Maura, A. Degasperi, F. Nadeu, D. Leongamornlert, H. Davies, L. Moore, R. Royo, B. Ziccheddu, X. S. Puente, H. Avet-Loiseau, P. J. Campbell, S. Nik-Zainal, E. Campo, N. Munshi, N. Bolli, A practical guide for mutational signature analysis in hematological malignancies. *Nat. Commun.* **10**, 2969 (2019).
57. E. H. Rustad, V. Yellapantula, D. Leongamornlert, N. Bolli, G. Ledergor, F. Nadeu, N. Angelopoulos, K. J. Dawson, T. J. Mitchell, R. J. Osborne, B. Ziccheddu, C. Carniti, V. Montefusco, P. Corradini, K. C. Anderson, P. Moreau, E. Papaemmanuil, L. B. Alexandrov, X. S. Puente, E. Campo, R. Siebert, H. Avet-Loiseau, O. Landgren, N. Munshi, P. J. Campbell, F. Maura, Timing the initiation of multiple myeloma. *Nat. Commun.* **11**, 1917 (2020).
58. Z. Sondka, S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, S. A. Forbes, The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
59. T. L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
60. T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
61. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. S. Noble, Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).
62. I. Masilionis, O. Chaudhary, B. M. Urben, R. Chaligne, Nuclei extraction for 10x Genomics Single Cell Multiome ATAC + Gene Expression from frozen tissue using Singulator™ 100 or 200 (S2 Genomics) V2.0. protocols.io (2023).
63. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
64. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
65. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
66. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
67. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv:1802.03426 [stat.ML] (2018).
68. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
69. C. Li, B. Liu, B. Kang, Z. Liu, Y. Liu, C. Chen, X. Ren, Z. Zhang, SciBet as a portable and fast single cell type identifier. *Nat. Commun.* **11**, 1818 (2020).
70. Y. Tan, P. Cahan, SingleCellNet: A computational tool to classify single cell RNA-seq data across platforms and across species. *Cell Syst.* **9**, 207–213.e2 (2019).
71. D. Aran, A. P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R. P. Naikawadi, P. J. Wolters, A. R. Abate, A. J. Butte, M. Bhattacharya, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
72. S. Jessa, A. Mohammadnia, A. S. Harutyunyan, M. Hulswit, S. Varadharajan, H. Lakkis, N. Kabir, Z. Bashardanesh, S. Hébert, D. Faury, M. C. Vladoiu, S. Worme, M. Coutelier, B. Krug, A. Faria Andrade, M. Pathania, A. Bajic, A. G. Weil, B. Ellezam, J. Atkinson, R. W. R. Dudley, J. P. Farmer, S. Perreault, B. A. Garcia, V. Larouche, M. Blanchette, L. Garzia, A. Bhaduri, K. L. Ligon, P. Bandopadhayay, M. D. Taylor, S. C. Mack, N. Jabado, C. L. Kleinman, K27M in canonical and noncanonical H3 variants occurs in distinct oligodendroglial cell lineages in brain midline gliomas. *Nat. Genet.* **54**, 1865–1880 (2022).
73. T. Tickle, I. Tirosh, C. Georgescu, M. Brown, B. Haas, inferCNV of the Trinity CTAT Project. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA (2019).
74. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
75. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
76. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
77. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
78. P. A. Northcott, D. J. H. Shih, J. Peacock, L. Garzia, A. Sorana Morrissy, T. Zichner, A. M. Stütz, A. Korshunov, J. Reimand, S. E. Schumacher, R. Beroukhim, D. W. Ellison, C. R. Marshall, A. C. Lionel, S. Mack, A. Dubuc, Y. Yao, V. Ramaswamy, B. Luu, A. Rolider, F. M. G. Cavalli, X. Wang, M. Remke, X. Wu, R. Y. B. Chiu, A. Chu, E. Chuah, R. D. Corbett, G. R. Hoad, S. D. Jackman, Y. Li, A. Lo, K. L. Mungall, K. Ming Nip, J. Q. Qian, A. G. J. Raymond, N. Thiessen, R. J. Varhol, I. Birol, R. A. Moore, A. J. Mungall, R. Holt, D. Kawauchi, M. F. Roussel, M. Kool, D. T. W. Jones, H. Witt, A. Fernandez-L, A. M. Kenney, R. J. Wechsler-Reya, P. Dirks, T. Aviv, W. A. Grajkowska, M. Perek-Polnik, C. C. Haberler, O. Delattre, S. S. Reynaud, F. F. Doz, S. S. Pernet-Fattet, B. K. Cho, S. K. Kim, K. C. Wang, W. Scheurlen, C. G. Eberhart, M. Fèvre-Montange, A. Jouvet, I. F. Pollack, X. Fan, K. M. Muraszko, G. Yancey Gillespie, C. di Rocco, L. Massimi, E. M. C. Michiels, N. K. Kloosterhof, P. J. French, J. M. Kros, J. M. Olson, R. G. Ellenbogen, K. Zitterbart, L. Kren, R. C. Thompson, M. K. Cooper, B. Lach, R. E. McLendon, D. D. Bigner, A. Fontebasso, S. Albrecht, N. Jabado, J. C. Lindsey, S. Bailey, N. Gupta, W. A. Weiss, L. Bognár, A. Klekner, T. E. van Meter, T. Kumabe, T. Tominaga, S. K. Elbabaa, J. R. Leonard, J. B. Rubin, L. M. Liau, E. G. van Meir, M. Fouladi, H. Nakamura, G. Cinalli, M. Garami, P. Hauser, A. G. Saad, A. Iolascon, S. Jung, C. G. Carlotti, R. Vibhakar, Y. Shin Ra, S. Robinson, M. Zollo, C. C. Faria, J. A. Chan, M. L. Levy, P. H. B. Sorensen, M. Meyerson, S. L. Pomeroy, Y. J. Cho, G. D. Bader, U. Tabori, C. E. Hawkins, E. Bouffet, S. W. Scherer, J. T. Rutka, D. Malkin, S. C. Clifford, S. J. M. Jones, J. O. Korbel, S. M. Pfister, M. A. Marra, M. D. Taylor, Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).
79. R. J. Gilbertson, S. C. Clifford, W. MacMeekin, W. Meekin, C. Wright, R. H. Perry, P. Kelly, A. D. Pearson, J. Lunec, Expression of the ErbB-neuregulin signaling network during human cerebellar development: Implications for the biology of medulloblastoma. *Cancer Res.* **58**, 3932–3941 (1998).
80. J. Aldaregia, P. Errarte, A. Olazagoitia-Garmendia, M. Gimeno, J. J. Uriz, T. R. Gershon, I. Garcia, A. Matheu, Erbb4 is required for cerebellar developmentand malignant phenotype of medulloblastoma. *Cancers* **12**, 997 (2020).
81. C. Gu, N. Yokota, Y. Gao, J. Yamamoto, T. Tokuyama, H. Namba, Gene expression of growth signaling pathways is up-regulated in CD133-positive medulloblastoma cells. *Oncol. Lett.* **2**, 357–361 (2011).
82. H. Fukushima, R. Suzuki, Y. Yamaki, S. Hosaka, M. Inaba, A. Muroi, T. Tsurubuchi, W. Morii, E. Noguchi, H. Takada, Cancer-predisposition genetic analysis in children with brain tumors treated at a single institution in Japan. *Oncology* **100**, 163–172 (2022).
83. W. P. Kloosterman, J. Koster, J. J. Molenaar, Prevalence and clinical implications of chromothripsis in cancer genomes. *Curr. Opin. Oncol.* **26**, 64–72 (2014).
84. A. Thomaz, M. Jaeger, A. L. Brunetto, A. T. Brunetto, L. Gregianin, C. B. de Farias, V. Ramaswamy, C. Nör, M. D. Taylor, R. Roesler, Neurotrophin signaling in medulloblastoma. *Cancers* **12**, 2542 (2020).
85. P. A. Northcott, D. T. W. Jones, M. Kool, G. W. Robinson, R. J. Gilbertson, Y. J. Cho, S. L. Pomeroy, A. Korshunov, P. Lichter, M. D. Taylor, S. M. Pfister, Medulloblastomics: The end of the beginning. *Nat. Rev. Cancer* **12**, 818–834 (2012).
86. P. A. Northcott, C. Lee, T. Zichner, A. M. Stütz, S. Erkek, D. Kawauchi, D. J. H. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D. T. W. Jones, M. Kool, M. Remke, F. M. G. Cavalli, S. Zuyderduyn, G. D. Bader, S. VandenBerg, L. A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jäger, I. Buchhalter, C. D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H. J. Warnatz, T. Risch, M. L. Yaspo, U. D. Weber, C. C. Bartholomae, C. von Kalle, E. Turányi, P. Hauser, E. Sanden, A. Darabi, P. Siesjö, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M. U. Schuhmann, M. Ebinger, H. L. Grimes, G. W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A. E. Kulozik, A. von Deimling, O. Witt, R. Eils, R. J. Gilbertson, A. Korshunov, M. D. Taylor, P. Lichter, J. O. Korbel, R. J. Wechsler-Reya, S. M. Pfister, Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

Acknowledgments: We thank Y. Bigot, A. Henssen, G. Casalena, H. Mueller, S. Takao, S. Cheng, M. Kharas, H. Zhu, A. Gutierrez, and M. Mansour for helpful suggestions; M. Gönen for statistical advice; W. Kang, E. Rosiek, K. Manova-Todorova, V. Morell, the MSK Molecular Cytology, Integrated Genomics and Bioinformatics Core facilities, and the Center for Comparative Medicine and Pathology for technical assistance; and N. Kentsis for editing help. Data analyses were enabled by compute and storage resources provided by the Digital Research Alliance of Canada. A.K. is a Scholar of the Leukemia & Lymphoma Society. **Funding:** This work was supported by NIH grants R01 CA214812 (A.K.), R01 CA192176 (A.L.J.), and P30 CA008748 (A.K., A.L.J., and N.D.S.); St. Baldrick's Foundation (A.K.); Burroughs Wellcome Fund (A.K.); Rita Allen Foundation (A.K.); Pershing Square Sohn Cancer Research Alliance and the G. Harold and Leila Y. Mathers Foundation (A.K.); Starr Cancer Consortium (A.K.); Cookies for Kids' Cancer (A.K. and G.P.R.); MSK Brain Tumor Center (A.K. and G.P.R.); MSK Functional Genomics Initiative (A.L.J.); MSK Cycle for Survival (A.L.J.); and Canadian Institutes of Health Research grant PJT-190271 (C.L.K.). **Author contributions:** Conceptualization: A.K., G.P.R., A.L.J., M.D.T., R.R.K., and M.Y. Methodology: R.R.K., M.Y., D.C., H.S., R.S., J.V., J.G., F.M., W.H., M.S., N.R., P.D., N.S.B., L.J.Z., C.R., I.M., R.C., N.D.S., S.H., R.L.G., and C.L.K. Investigation: R.R.K., M.Y., D.C., H.S., R.S., J.V., J.G., F.M., W.H., M.S., N.R., P.D., N.S.B., L.J.Z., C.R., N.D.S., R.L.G., and C.L.K. Visualization: M.Y., R.R.K., F.M.,

H.S., J.V., N.S.B., and R.L.G. Funding acquisition: A.K., G.P.R., A.L.J., and M.D.T. Project administration: A.K. Supervision: A.K., G.P.R., A.L.J., and M.D.T. Writing—original draft: A.K., M.Y., and R.R.K. Writing—review and editing: all authors. **Competing interests:** A.K. is a consultant for Novartis, Rgenta, Blueprint, and Syndax. The authors declare that they have no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Sequencing data are available from the NCBI SRA (mouse) and the EGA (human) using respective accession numbers listed in Materials and Methods. All processed data, including specific structural variants and their breakpoint sequences, are openly available via Zenodo (https://zenodo.org/records/10666885). *Atoh1-FlpoER/+* and *R26*^{MASTR} can be requested from A.L.J.; all other strains are available from the Jackson Laboratory.

Submitted 11 December 2023
Accepted 16 February 2024
Published 22 March 2024
10.1126/sciadv.adn4649