Subclonal architecture, evolutionary trajectories and patterns of inheritance of germline variants in pediatric glioblastoma

Mary Hoffman[1,2,3], Aaron H Gillmor[1,2,3], Daniel J Kunz[4,5,6], Michael Johnston[2,3], Ana Nikolic[1,2,3], Kiran Narta[1,2,3], Jennifer King[2,3], Katrina Ellestad[2,3], Ngoc Ha Dang[2,3], Florence MG Cavalli[7], Michelle M Kushida[7], Fiona J Coutinho[7], Betty Luu[7], Yussanne Ma[8], Andrew J Mungall[8], Richard Moore[8], Marco A Marra[8], Michael D Taylor[7], Trevor J Pugh[10], Peter B Dirks[7,9], Donna L Senger[2,3,11], Benjamin D Simons[4,5,6], Jennifer A Chan[2,3,12], A Sorana Morrissy[1,2,3], Marco Gallo[1,2,3,13]*.

[1]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB T2N 4N1, Canada.

[2]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB T2N 4N1, Canada.

[3]Charbonneau Cancer Institute, University of Calgary, Calgary, AB T2N 4N1, Canada.

[4]Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK.

[5]The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK.

[6]The Wellcome Trust/Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, UK.

[7]The Hospital for Sick Children Research Institute, Arthur and Sonia Labatt Brain Tumour Research Centre, Toronto, ON M5G 0A4, Canada.

[8]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC VSZ 4S6, Canada.

[9]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3H7, Canada.

[10]Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 2C4, Canada.

[11]Department of Oncology, University of Calgary, Calgary, AB T2N 4N1, Canada.

[12]Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada.

[13]Department of Physiology and Pharmacology, University of Calgary, Calgary, AB T2N 4N1, Canada.

*Correspondence: marco.gallo@ucalgary.ca.

Abstract

Pediatric glioblastoma (pGBM) is a lethal cancer with no effective therapies. Intratumoral genetic heterogeneity and mode of tumor evolution have not been systematically addressed for this cancer. Whole-genome sequencing of germline-tumor pairs showed that pGBM is characterized by intratumoral genetic heterogeneity and consequent subclonal architecture. We found that pGBM undergoes extreme evolutionary trajectories, with primary and recurrent tumors having different subclonal compositions. Analysis of variant allele frequencies supported a model of tumor growth involving slow-cycling cancer stem cells that give rise to fast-proliferating progenitor-like cells and to non-dividing cells. pGBM patients' germlines had subclonal structural variants, some of which underwent dynamic frequency fluctuations during tumor evolution. By sequencing germlines of mother-father-patient trios, we found that inheritance of deleterious germline variants from healthy parents cooperate with *de novo* germline and somatic events to the tumorigenic process. Our studies therefore challenge the current notion that pGBM is a relatively homogeneous molecular entity.

Pediatric glioblastoma (pGBM) is a lethal brain tumor. Unlike adult GBM, no effective therapeutic intervention or standard of care currently exists for this cancer. Despite their histological similarities, pGBM is clinically, biologically and molecularly distinct from adult GBM *(1)*. Unfortunately, the biological principles uncovered by studying the adult disease have not translated to the pediatric malignancy *(2)*. A major breakthrough in understanding the molecular principles underlying pGBM stemmed from the discovery that a third of patients harbor somatic mutations in *H3F3A*, a gene encoding the histone variant H3.3 *(3, 4)*. Mutations in this gene cause one of two amino acid substitutions in H3.3: (i) Mutation of lysine 27 to methionine (K27M); (ii) mutation of glycine 34 to arginine (G34R) or valine (G34V). Both types of mutations interfere with the wildtype function of H3.3 *(5, 6)*, but the H3.3K27M variant has been studied more extensively because it is more frequent and trimethylation of K27 has an established and important role in repressing gene transcription and compacting chromatin. Several experimental models have been used to show that *H3F3A* mutations can induce cellular hyperproliferation on their own, but they need to cooperate with other drivers, including *TP53* mutations, to produce overt malignancies *(6–8)*. For the majority of pGBM patients without H3.3 mutations, the early genetic events leading to overt malignancy are not fully understood.

It is currently believed that pGBM, similarly to other pediatric cancers, is characterized by relatively bland genomes *(9)*. Furthermore, *H3F3A* mutations were shown to be clonal in pGBM *(10)*, and to result in global loss of trimethylation of lysine 27 of histone 3 (H3K27me3) in virtually all cells in a tumor *(6)*. These findings have contributed to the view that pGBM might have a relatively low level of intratumoral complexity at the genomic and functional level. However, a

4

recent report identified differences in mutational signatures between matched primary and recurrent pGBM samples using exome sequencing *(10)*.

Evidence of subclonal architecture in diffuse intrinsic pontine glioma and pGBM was recently provided *(11)*. The authors reanalyzed previously published exome datasets for 36 pGBM samples and identified evidence for the coexistence of subclones in each tumor. To the best of our knowledge, no study to date has systematically assessed subclonal composition and evolution in matched primary and recurrent samples from the same patient using whole-genome information. Such studies are important because they may allow inference of biological principles responsible for tumor evolution and recurrence, and they might identify intergenic variants with important roles in the etiology of this cancer. Furthermore, treatment regimens for pGBM differ between centers and individual oncologists, raising the possibility that different therapeutic approaches might skew recurrences toward the acquisition of specific molecular features. The relative rarity of pGBM (adjusted incidence rate of 0.15/0.16 in 100,000 population) *(12)* and the lack of standardized practices for surgical resection contribute to the scarce availability of diagnostic-recurrent sample pairs for study.

New whole-genome technologies that allow sensitive measurements of cellular heterogeneity in a single tumor, and documentation of tumor evolution over time, have yet to be systematically employed for this malignancy. We hypothesized that whole-genome sequencing (WGS) of primary and recurrent tumors, coupled with targeted bioinformatic approaches to specifically test for subclonal architecture of each tumor, might bring new insight into the intratumoral organization of pGBM and its evolution. Understanding the principles exploited by the tumor to escape treatment may have important implications for rationally selecting or designing personalized treatment options for these rare and aggressive cancers.

Results

pGBM samples and whole-genome sequencing with linked-reads. We characterized longitudinally collected pGBM samples that include primary tumors and their recurrences, in addition to peripheral blood (used as germline control), and were biobanked at our institution (Calgary cohort). For 4 out of 5 patients we had at least two longitudinal samples and for one patient we had three recurrent samples (**Table S1**). For the purposes of this manuscript, diagnostic samples are referred to as "sample 1" and each consecutive resection from the same patient is referred to as "sample *n*." Treatment history for each patient is known (**Table S2**). We also obtained a set of matched germline and primary pGBM samples biobanked in a second institution (Toronto cohort) (**Table S3**) and used these for validation purposes. We extracted high molecular weight DNA (average length distribution > 40 kbp) from each sample and generated libraries for WGS with linked-read technology (10xGenomics). With this system, short next-generation sequencing reads originating from the same DNA molecule (ie an individual chromosome fragment) are tagged with an identical barcode, enabling the assembly of Mb-size haplotypes from individual chromosomes. Average sequencing depth for all samples was 35.6x, while average physical coverage was 178x because of the use of linked reads (**Table S4**). The longest phase block generated was 71 Mbp, with an average maximum length for all samples of 16.6 Mbp and a median of 12 Mbp (**Table S4**).

Diagnostic and recurrent pGBM samples share few single nucleotide variants. We used linked-read WGS data for pGBM samples and matched germlines to call somatic variants in diagnostic samples and their recurrences. Focusing specifically on non-synonymous single-nucleotide variants (SNVs), our data are in agreement with previous reports of low mutation burden *(3, 13, 14)* in pGBM (range: 26-124 SNVs per tumor; average: 65.6 SNVs per tumor in primary samples, 60 SNVs per tumor in the recurrences). Overall, we found that recurrent tumors shared few somatic mutations with their matched primary malignancies, with only between 1 and 5 non-synonymous SNVs shared between longitudinal samples from the same patient. This translated to an average of 7.8% non-synonymous SNVs that were shared between a primary tumor and its recurrence. A similar amount (11.8%) of somatic SNVs and indels were shared between diagnostic and recurrent samples in the pediatric brain tumor medulloblastoma *(15)*, in contrast with the much higher proportion shared between primary and recurrent adult glioma samples (54%) *(16)*. Surprisingly, we observed that the total number of somatic SNVs tended to be stable in recurrences post-treatment (**Fig. 1A-1E; Supplemental file 1**), whereas mutational load is known to increase 5-fold at recurrence in pediatric medulloblastoma *(15)* as well as in adult gliomas *(16)*. Patient 1 was one exception (**Fig. 1A**), but this individual harbored a germline deletion of the last exon of *TP53* (**Fig. S1A and S1B**), which might have contributed to accelerated tumor evolution by increasing mutational events. Sample 2 had 126 somatic non-synonymous SNVs, and only lesions in *ARMC9* and *DSE* were shared with the diagnostic sample. Interestingly, *DSE* encodes a dermatan sulfarase epimerase and is expressed in the human brain at higher levels before birth than after birth (**Fig. S1C and S1D**). Patient 2 had a somatic mutation in *H3F3A* that was retained in the second sample (**Fig. 1B**), together with a mutation in *MUC22* and two deleterious SNVs in *TP53*. This patient was treated with temozolomide and local radiation

(**Supplementary Table S2**), which is the standard of care for adult GBM patients. However, the overall number of somatic, non-synonymous SNVs was lower in the second sample (69 SNVs) than in the diagnostic sample (109 SNVs), whereas this treatment is known to increase mutational burden in adults. This difference in response to chemotherapy and radiation between pGBM and adult GBM was also noted in an independent cohort of patients in a recent publication *(10)*. Samples 1 and 2 from patient 3 only shared a mutation in *PCNX* (**Fig. 1C**). Patient 4 had 84 somatic non-synonymous SNVs in the diagnostic sample (**Fig. 1D**), and no recurrence was available for sequencing. Patient 5 had 26 non-synonymous SNVs in the diagnostic sample and in sample 2, 38 in sample 3, and 39 in sample 4 (**Fig. 1E**). SNVs in 5 genes were shared between sample 1, sample 2 and sample 3 (*ELAVL2*, *VWF*, *KCNJ12*, *HEXIM2* and *ZNF816*). Only SNVs in *VWF* (encoding the von Willebrand factor) and *KCNJ12* (encoding an ion channel) were conserved between all samples for this patient.

No SNV was shared among all samples. Only 4 genes were mutated in at least two primary tumors (**Table S5**), including *VWF*. Six genes were mutated in at least two recurrences from different patients (**Table S6**). No established oncogene or tumor suppressor was recurrently mutated in our cohort, with the exception of *TP53* (altered in two patients). Our data therefore show that pGBM is characterized by large intertumoral genetic heterogeneity and by large evolutionary changes following surgical resection and treatment.

No obvious pattern of gene mutations was observed in our dataset. Gene ontology analysis did not identify any significantly enriched class of mutated genes. However, we found that non-synonymous SNVs in diagnostic samples were significantly enriched for genes with an OMIM entry and linked to hereditary syndromes (19.6% of the 321 mutated genes, permutation test p = 0.00082; **Fig. 1F, Fig. S1E,F**). 41.3% of these genes are associated with syndromes of the central

nervous system (CNS). Similarly, non-synonymous SNVs in subsequent resections had a significant enrichment for genes with OMIM entries (20.2% of the 420 mutated genes, permutation test $p < 10^{-5}$; **Fig. S1E,F**), and 35.2% of these genes were associated with hereditary CNS syndromes (**Fig. 1G**). These observations suggest that disruption of multiple unrelated networks of genes may be sufficient for pGBM initiation and progression. The identification of deleterious SNVs in genes linked to hereditary syndromes is intriguing, given the early-life onset of this malignancy.

Germline sequencing of mother-father-patient trios. Intrigued by the identification of somatic non-synonymous SNVs in genes associated with hereditary disorders, we asked whether pGBM behaves as a developmental syndrome. We reasoned that if we assumed cancer was a developmental disorder of progressive mutagenesis, individual patients might undergo systemic accumulation of mutations over time. We decided to investigate this possibility by assembling trios composed of mother, father and patient, implementing an approach similar to what is done with genetic disorders. We were able to obtain the germlines (peripheral blood) of the parents of patient 1 and patient 3, and used them for WGS with linked reads. Analysis of these trios enabled us to categorize non-synonymous SNVs as either unique to the pGBM patient germline, or inherited from one of the parental genomes. Our data show that a *TP53* germline variant identified in patient 1 (**Fig. S1A and S1B**) was inherited from one of the parents. In addition, non-synonymous SNVs in 9 genes were identified uniquely in the germline of this patient (**Fig. 1H**). Among these genes, *RASAL2* encodes a RasGAP and is a tumor suppressor *(17)*; *PARP4* encodes a member of the poly(ADP)-ribose family, and was recently shown to be mutated in the germlines of patients with thyroid and breast cancer *(18)*; *HUWE1* encodes an E3 ubiquitin ligase that was

shown to be a tumor suppressor in mouse models of skin cancer *(19)* and colorectal cancer *(20)*. Similarly, we called 4 *de novo* non-synonymous SNVs in the germline of patient 3 (**Fig. 1I**). Among these genes, *KIAA1551* was recently identified as a candidate tumor suppressor and is frequently deleted in different cancer types *(21)*. Overall, we noticed a significant increase in the number of short (30-50,000 bp) deletions in the germlines of pGBM patients (n = 8) compared to the control (parental) germlines (n = 4) we sequenced (Mann Whitney test p = 0.0040; **Fig. S1G**). This collection of sequenced trios is a unique dataset in the context of pGBM and offers a new perspective on the emergence of germline variants in pediatric patients with sporadic cancer.

Subclonal architecture of pGBM at diagnosis and recurrence. We next asked whether the small proportion of SNVs shared between each diagnostic sample and its recurrence were a reflection of unexpected intratumoral genomic complexity and consequent clonal selection. We assessed the presence of genetically distinct subclonal cell lineages in each tumor before and after treatment by using somatic mutations and copy number alterations to build a genetic phylogeny for each patient using the EXPANDS algorithm *(22)*. These phylogenies show that each pGBM and its recurrence(s) are characterized by extensive intratumoral genomic heterogeneity and complex subclonal architecture (**Fig. 2A-E**). For instance, in the diagnostic sample from patient 1 we detected 8 subclonal lineages that correspond to three main groups or clusters (**Fig. 2A**). Subclones e and f (1ef) represent ancestral genetic lineages not far diverged from the germline, while subclones 1abcdg and 1h form two major phylogenetic branches, exhibiting moderate to extreme divergence from the germline. In all patients and all samples we observe distinct phylogenetic branches. These support the presence of multiple genetically distinct subclones that change significantly in frequency throughout the evolution of a tumor. Importantly, each recurrent tumor

10

also had a few subclones clustering close to the germline, likely indicating the stable presence of early (ancestral) clones which are not actively accumulating somatic alterations. This observation is best illustrated in patient 5 where three recurrences were available (**Fig. 2E**). Throughout this patient's tumor progressive divergence from the germline occurs at each successive recurrence in a number of genetic lineages, yet a number of ancestral subclones remain detectable at all timepoints (e.g. 2agh at the first recurrence; 3f at the second recurrence; 4ef at the third recurrence). While these subclones likely had ancestral roles in the early part of the evolutionary path of this patient's tumor, each subsequent recurrence is marked by distinct lineages that have diverged significantly from this baseline.

In addition to clonal composition, somatic SNVs were further used to infer mutational signatures. Having already inferred the phylogenetic lineages present in each sample let us select SNVs corresponding to distinct clusters of subclones, such that we could then compare the range of mutation signatures from subclones within each sample. In each patient we see evidence for multiple mutation signatures, indicating the influence of more than one mutagenic process. For instance, in the diagnostic sample of patient 2, the subclones clustering near the germline (1bcfhij) were characterized by mutational signatures 3,4,6, and 24, while the more evolutionarily divergent subclones accumulated mutations dominated by signatures 4, 6 and 1 (**Fig. 2C**). Similar patterns of divergence between mutational signatures specific to clusters of subclones were observed for all other patients (**Fig. 2B-E and Fig. S2**). Signature 1 results from the spontaneous deamination of 5-methylcytosine, is the most commonly seen across cancer types, and was observed in our cohort in all samples though at different levels in each subclone (**Fig. S2**). Signatures 6 and 15, prevalent in cancers with defective mismatch repair, were frequently observed in many subclones, with a handful of additional signatures (including 8; previously observed in the pediatric brain

tumor medulloblastoma) detectable in individual subclones. All together, these data indicate that different mutational processes may operate at different times during the evolution of individual tumors. Alternatively, different mutational processes are acting concurrently at the onset, and the changing selective pressure present at each treatment stage enables different genetic lineages to become dominant within the tumor.

Identification of structural variants in pGBM. pGBM is known to harbor somatic structural variants (SVs), some of which recur at low frequency in the patient population *(23, 24)*. Because WGS with linked reads data are optimal for identification and visualization of structural variants and subclonal events, we explored the behavior of SVs during tumor evolution. SVs were called with the longranger package. We defined large SVs as variants affecting > 30 kbp of the genome. We processed our linked-read WGS data for germlines and tumors independently of each other. This approach allowed us to retain SVs that might be important for tumor etiology, but that would otherwise have been filtered out because they are also present in the germlines. With this approach, we identified 20 to 30 SVs in the germlines of each pGBM patient compared to the reference genome (**Fig. 3A-E; Supplemental file 2**). The overall number of large SVs did not change dramatically between the primary and recurrent tumors. Patient 1 was the only one displaying a large increase in the number of SVs in the primary tumor compared to the germline (**Fig. 3A**). This was probably because of chromosome 6 chromothripsis events, which represented 62% of total SVs for this patient. No trace of chromosome 6 chromothripsis was observed in the second resection and the total number of SVs was similar to the germline. This was unexpected because most patients were treated with radiation, which is known to induce SVs. However, this absence is in accord with the observed divergent clonal structure of pGBM (**Fig. 2**), and suggests that the

12

chromothriptic event may have been present in a subclone. Adult gliomas experience a large number of mutational events (SVs and SNVs) during standard of care treatment (radiation plus temozolomide) *(16)*. Patient 2 underwent radiation and temozolomide treatment, yet the overall mutational burden did not drastically increase (**Fig. 3B**). We detected SVs affecting previously known cancer genes as well. Sequencing data provide evidence of a duplication of *NTRK2*, the overexpression of which was previously shown to contribute to tumor metastasis *(25)*. Gene fusions involving *NTRK2* have been previously reported in pGBM *(24)*. Patient 3 had a somatic *ETV6-NTRK3* gene fusion in both primary and recurrent tumors. This gene fusion has been reported in pGBM before *(24)*, and it was also shown to be have oncogenic function in other malignancies, including leukemia *(26)*, breast cancer *(27)* and papillary thyroid carcinoma *(28)*. We also identified a translocation predicted to truncate the proto-oncogene *MET* in patient 4. *MET* amplifications are relatively common in adult GBM *(29)*, and lesions in this gene are found in several cancer types *(30)*, including pGBM *(24, 31)*.

We have identified several SVs that recurred in more than one patient in our cohorts. Among these, we found germline deletions affecting the coding sequences of *BTNL3*, *BTNL8*, *APOBEC3A* and *APOBEC3B*. Interestingly, *BTNL8* encodes a cell surface protein involved in T cell activation *(32)*, whereas the *APOBEC3* genes encode cytidine deaminases with roles in antiviral immunity *(33)*. These deletions abrogated gene expression (**Fig. S3**).

Germline and somatic structural variants during tumor evolution. In some instances, tumors had lower numbers of SVs than their matched germlines (see diagnostic sample in **Fig. 3C** and samples 2-4 in **Fig. 3E**). We found that every patient carried some germline SVs that were not detected in their diagnostic samples and their recurrences. There were 8 such germline-specific

13

SVs for patient 1 (**Fig. 3A**), 1 for patient 2 (**Fig. 3B**), 7 for patient 3 (**Fig. 3C**), 9 for patient 4 (**Fig. 3D**) and 3 for patient 5 (**Fig. 3E**). Each SV call was made with longranger and was manually reviewed, checked for coverage in the region of the call and confirmed. Some SVs with allelic frequencies between 0.4 and 0.7 in the germline also underwent fluctuations in the tumors, while others were stable (**Fig. S4A-D**). SVs with allelic frequencies > 0.7 in the germline, which we considered clonal, tended to be detected at stable levels in primary and recurrent tumors (**Fig. S4E-H**). The subclonal nature of some SVs in the germline, and their fluctuations in primary and recurrent tumors, were unexpected findings and we set out to further substantiate these observations.

In order to explore the observed SVs with an independent computational method, we re-called SVs with the Genome-wide Reconstruction of Complex Structural Variants (GROC-SVs) package *(34)*. We subset SV predictions to those for which we had sufficient physical coverage in all samples (see Methods), thus retaining SVs supported by at least 200 barcodes (i.e. DNA molecules) spanning each breakpoint (**Supplemental file 3**). GROC-SVs models noise specific to 10X linked-read data and assigns both a binary state corresponding to the presence or absence of a breakpoint in each sample, as well as a p-value to the likelihood of detecting that breakpoint given the local read support. We stratified SVs in each patient based on their pattern of presence or absence across samples. For instance, somatic SVs were absent from the germline and appeared in one or more tumor samples (**Fig. 4**). Germline events were detected in all samples, and in patients 1 and 3 could be further stratified into those inherited from a parent or occurring *de novo*. In all pGBM patients, we identified changes in SV allelic frequencies between tumor resections. Allelic frequency changes could represent loss or acquisition of somatic SVs in the tumors (**Fig. 4H**), and are often observed in our cohort to occur in genes with previous evidence for structural

14

variation in pediatric cancers (**Fig. 4B,C,G,I**) *(35)*. For patient 1, we identified several *de novo* germline SVs (which were not detected in the parents) with allelic frequencies in the tumor samples (**Fig. 4C**). These data point to complex patterns of propagation of SVs from the germline to tumor samples, and highlight for the first time the presence of putative *de novo* germline SVs in pGBM patients.

Subclonal deletions at the *ATRX* locus in germlines and pGBM. The data organization principles underlying linked-read WGS data facilitated identification and visualization of subclonal SVs. Among these SVs, we identified large deletions in *ATRX* that removed almost the entire coding region of the gene (**Fig. 5A**). Somatic SNVs in *ATRX* are known to occur in pGBM *(3, 36)*, and loss of ATRX protein is routinely tested in GBM specimens in the clinic with immunostaining techniques. Our data indicate that *ATRX* harbors both SNVs and SVs in pGBM. Just outside the deletion breakpoints, there are a series of Alu and SINE repetitive elements, and recombination between these repetitive sites may be a mechanism behind the generation of the *ATRX* deletion. Of note, we detected putative subclonal *ATRX* deletions in the germline of pGBM patients. In order to validate the presence of putative *ATRX* deletions with an independent approach, we designed a TaqMan probe for copy number assays that targets the putatively deleted region ("ATRXdel" probe; **Fig. 5B**). A control TaqMan probe was designed to target a region on chromosome 5 predicted to be diploid based on our WGS data ("control" probe; see Supplemental Material). We then performed copy number TaqMan assays using genomes from both the germline and the tumors of our pGBM patients. TaqMan signal for the ATRXdel probe was internally normalized to signal for the control probe and number of X chromosomes. A second normalization was performed against the control diploid genome. This TaqMan assay confirmed the *ATRX*

deletion in both germlines and tumors of all pGBM patients with the exception of patient 5, where the deletion was only detected in sample 2 (**Fig. 5C and Fig. S5**). In order to validate these findings in an independent cohort, we performed WGS with linked reads on three more pGBM samples and their matched germlines collected in Toronto. Our data show evidence of subclonal deletions at the *ATRX* locus in all samples (**Fig. S6A**). These findings confirm our WGS and copy number data and show that subclonal deletions in *ATRX* exist in both the germline and tumors of most pGBM patients.

We expected that subclonal *ATRX* deletions may result in heterogeneous expression of the gene in populations of tumor cells. Since our longitudinal collection was assembled over a long period of time, storage of pGBM samples was not compatible with single-cell RNA-seq (scRNA-seq) methods. We therefore decided to assess transcriptional heterogeneity using patient-derived xenografts (PDXs), which were available for sample 2 of patient 3 and for sample 4 of patient 5. We successfully generated scRNA-seq data for 1,327 cells from patient 3's sample 2, and for 2,204 cells patient 5's sample 4. As predicted from the WGS data and TaqMan assays, scRNA-seq data showed heterogeneous expression of *ATRX* in both PDXs (**Fig. 5D,E**). In contrast to other genes, for instance *H3F3A* and *H3F3B* (both of which encode the histone variant H3.3), were more homogeneously expressed (**Fig. S6B,C**). Widespread decreased *ATRX* expression in pGBM compared to non-neoplastic brain tissue (n = 4) was confirmed by RNA-seq of bulk pGBM tissue (n = 3) and their matched primary cultures from an independent cohort collected in Toronto (**Fig. 6F**). Furthermore, analysis of an independent genomic dataset confirmed that subclonal *ATRX* mutations are detected in the tumors of pGBM patients (**Fig. S6D**). Our data therefore show that in addition to somatic SNVs in *ATRX*, pGBM patients may also carry subclonal *ATRX* deletions in both their germlines and tumors.

Identification of conserved intergenic SVs in pGBM samples and germlines. Genome-wide coverage is a strength of our data compared to previously published datasets. We decided to leverage this asset by looking for intergenic SVs that were conserved in our pGBM samples. We found only one intergenic SV that was shared among all pGBM samples and their germlines, a deletion immediately upstream of *NEGR1* (neural growth regulator 1; **Fig. 6A**). This upstream deletion was heterozygous or homozygous in different patients, but the breakpoints were absolutely conserved. Some haplotypes displayed evidence of subclonal deletions in part of the gene (black box in **Fig. 6A**). Analysis of the GTEx database *(37)* showed that *NEGR1* is expressed at higher levels in brain tissues - including frontal cortex and the cerebellar hemisphere - than in other anatomical locations (**Fig. S6E**). *NEGR1* mutations have been linked to obesity *(38)* and dyslexia *(39)*. In mice, *Negr1* enhances neurite outgrowth in cortical neurons, and *Negr1* knockout animals have axonal defects and behavioral deficits *(40)*. Expression of this gene is downregulated in different cancer types, and downregulation of *NEGR1* promotes migration *(41)*. Our scRNA-seq show that PDX samples with deletions upstream of *NEGR1* have undetectable expression of this gene in nearly all cells (**Fig. 6B**), suggesting that the SV affects regulatory regions. We observed strong repression of *NEGR1* in an independent set of pGBM samples and their matched pGBM primary cultures from Toronto (**Fig. 6C**). Importantly, our analysis of a previously published pGBM cohort *(42)* shows that *NEGR1* expression can stratify pediatric glioma patients based on overall survival, with tumors expressing lower levels of the gene having worse prognosis (log-rank p = 0.040; **Fig. 6D**).

Evidence for the existence of relatively quiescent cancer stem cells in pGBM. We questioned whether the observed variability of the mutational profile might correlate with proliferative heterogeneity in the respective contributions made by subclones to tumor growth and recurrence. Recently, cell tracing studies based on the genetic barcoding of freshly-isolated adult GBM cells transplanted into PDX mouse models have suggested that tumor cells may adhere to a common fate behavior, reminiscent of a normal neurogenic program *(43)*. Specifically, a signature negative binomial dependence of the distribution of variant allele frequencies (VAFs) provides evidence of a conserved proliferative hierarchy in which slow-cycling tumor stem cell-like cells give rise to a rapidly cycling, self-renewing, progenitor-like population (**Fig. 7A**) (Extended Data Figure 4 of *(43)*). Applied to the patient samples in the current study, we find that the VAF distribution obtained from tumor samples is also largely consistent with a negative binomial dependence (**Fig. 7B-E** and **Fig. S7A-I**), suggesting pGBM tumor cells may be defined by a similar proliferative hierarchy (see Methods). This conclusion is reinforced by focusing on the ensemble of *de novo* point mutations that are acquired, or only rise above the detection threshold, during recurrence (**Fig. S7J-N** and Methods). By contrast, mutations that are shared between the primary and recurrent tumor samples show VAFs that are peaked at larger values, consistent with the predominance or fixation of subclones within the tumor population (**Fig. S7O-R**). Our data from our pGBM patient cohort are therefore consistent with a hierarchical model for pGBM, characterized by slow-cycling cancer stem cells that give rise to fast-proliferating progenitor-like cells, which eventually produce non-dividing "differentiated" cells.

Discussion

Cancer is often compared to a caricature of developmental processes. Molecular profiling of pGBM suggests that this cancer has features that make it appear like a caricature of hereditary disorders. Notwithstanding large differences in terms of mutated genes among tumors, in every patient we identified somatic SNVs in several genes linked to hereditary conditions, including MeCP2, which is linked to Rett syndrome *(44)*, and autosomal dominant genes linked to severe intellectual (*ZMYND11*, OMIM number 608668) and developmental syndromes (*DSE*, linked to Ehlers-Danlos syndrome; *REEP1*, OMIM number 609139). Anecdotally, we found somatic SNVs in numerous genes associated with blood disorders, such as the recurrent SNVs in *VWF* (a gene linked to von Willebrand disease), *LMO2* (leukemia), *CR2* (systemic lupus erythematosus), *ASXL1* (myelodysplastic syndrome), *ITAGA2B* (bleeding disorder), and *MLL1* (leukemia). Genetic lesions did not cluster in specific pathways or gene classes, but this might reflect our small sample size. Interestingly, the pediatric brain tumor diffuse intrinsic pontine glioma has recurrent somatic mutations in *ACVR1 (24, 45–47)*, and germline mutations in this gene lead to the hereditary syndrome fibrodysplasia ossificans progressiva *(48)*. The detection of germline variants in pGBM patients, and the identification of somatic SNVs in genes associated with hereditary syndromes, give pGBM a characteristically different molecular flavor from adult GBM. Because of the early onset of pGBM, studies of larger cohorts will be required to determine if germline and somatic events occurring during embryonic and/or fetal development cooperatively contribute to pGBM initiation and progression. These studies will be particularly important given that currently no genetic or environmental factor has been associated with sporadic pGBM. It would be important to determine whether the *de novo* germline variants identified in this study arise in the germ cells of the parents during the normal mutagenic processes of genome replication *(49)*, or if they arise

19

during early embryonal or fetal development. Our data showing that some variants have low (< 0.2) allelic frequency in the germline of pGBM patients suggest that these individuals might display somatic mosaicism resulting from mutational processes during postzygotic stages of development. Interestingly, clonal hematopoiesis has recently been shown to be associated with a 10-fold increase in risk of developing hematologic cancer *(50, 51)*, and was observed in both adult and pediatric patients with non-hematologic cancers *(52)*. In pediatric cancer patients, clonal hematopoiesis was identified in 5-10% of individuals, with highest frequency in children aged 0-9 years. There is therefore a possibility that some subclonal SVs with < 0.2 allelic frequencies in the germline that became barely detectable or undetectable in tumors might reflect clonal hematopoiesis. In this scenario, the frequency of germline SVs in the tumor may depend on the extent of tumor infiltration of hematopoietic cells carrying that specific SV. However, we identified this phenomenon in all pGBM samples, which is more than expected based on available published data *(52)*. The low AF of some germline SVs is therefore more likely the result of developmental processes that selected them over time.

By sequencing the germlines of trios, our data provide an important perspective on the developmental origins of pGBM. Our data show that some potentially deleterious germline variants (like a *TP53* exonic deletion in patient 1) can be inherited from a healthy parent. At the same time, we observed *de novo* non-synonymous SNVs in the germlines of both patient 1 and patient 3, as well as *de novo* germline SVs in patient 1. The expected rates of emergence of *de novo* non-synonymous SNVs is very low, based on recent literature evidence on trios that included probands affected by disorders with high heritability. Trio studies in the context of sporadic autistic spectrum disorder identified 181 *de novo* non-synonymous SNVs in 189 probands, which corresponds to an average of 0.96 *de novo* non-synonymous SNV per child *(53)*. A second trio

study found 32 *de novo* non-synonymous SNVs in 53 probands with schizophrenia, corresponding to a rate of 0.60 *de novo* non-synonymous SNVs per individual *(54)*. In the same study, 7 out of 22 unaffected individuals (control group) had *de novo* SNVs, with 4 out of 7 SNVs being non-synonymous. While acknowledging that our sample set is small and that there may be differences in numbers of SNVs detected and called based on sequencing platform and computational pipelines used, our data offer proof-of-principle that mutational processes are at play in the parents' germ cells or in the early developing zygote leading to *de novo* germline variants in at least some pGBM patients. Because of the rare nature of pGBM, and difficulties in re-contacting families of pGBM patients, the trios we generated represent extremely valuable resources. In the future, it will be important to assemble germline WGS data on trios of pGBM probands in a systematic fashion, in order to assess and investigate genomic predisposing events.

Using WGS data with linked reads, we have identified a set of germline SVs that tend to occur at high frequency in our cohort. One of these examples is subclonal deletions at the *ATRX* locus in both germlines of four out of five patients and tumor tissues of all patients. Recent work showed that deletion of *Atrx* in mouse neural progenitor cells or *ATRX* in adult glioma stem cells alters patterns of chromatin openness genome-wide and leads to enhanced cell migration *(55)*. Because of the established involvement of *ATRX* in gliomagenesis, we speculate that *ATRX* deletions might be a predisposing genetic lesion for pGBM. Germline deletions upstream of *NEGR1* were present in every pGBM patient, in both the Calgary and in the Toronto cohorts. These deletions had conserved breakpoints, and were either homozygous or heterozygous in different patients and resulted in loss of *NEGR1* expression. Given that this gene has an important role in neuronal development in mouse models *(40)*, and its decreased expression positively affects migration *(41)*, it is tempting to speculate that loss of *NEGR1* expression may confer a competitive advantage at

early stages of tumor development. Lesions in *ATRX* and *NEGR1* may cooperate with other SVs we identified in the germlines of pGBM patients to determine a certain level of predisposition to the development of cancer. For instance, we found that 4 out of 5 patients also harbored germline deletions at the *BTNL8/3* locus (**Fig. S3A**), causing loss of expression of the affected genes (**Fig. S3B,C**). *BTNL8* encodes a protein that activates immune cells *(32, 56)*. We hypothesize that multiple lesions in genes involved in brain development and function (eg *ATRX* and *NEGR1*) and in activation of the immune system (eg *BTNL8*) might provide a selective advantage to cells that participate in the early steps of the tumorigenic process.

Importantly, our analyses show that pGBM is characterized by intratumoral genetic heterogeneity and subclonal architecture, which rapidly evolves at recurrence. Our work, together with previous reports on pediatric medulloblastoma *(15)* and diffuse intrinsic pontine glioma *(11)*, indicates that pediatric brain cancers have more complex genetic architectures than previously thought. The genetic divergence between diagnostic and recurrent samples should be kept into consideration when designing personalized treatment approaches. In addition to genetic heterogeneity, our data provide evidence for the existence of a functional hierarchy in pGBM. VAF distributions in pGBM samples are consistent with a hierarchy comprising relatively quiescent cancer stem cells, fast-proliferating progenitor cells and non-self-renewing and non-dividing cells destined to undergo cell death. Single-cell barcoding experiments with xenografts derived from adult GBM patients supported a similar hierarchy *(43)*. Although adult GBM and pGBM are molecularly different, hierarchical organization may be a common feature of both. If relatively quiescent, self-renewing cancer stem cells reside at the apex of a pGBM hierarchy, these cells might be less sensitive to radiation and antimitotics used to treat this cancer, explaining why recurrences did not have a significantly higher mutational burden than the diagnostic samples.

22

Overall, our data provide strong evidence for the existence of subclonal architecture in pGBM and relatively rapid evolution at recurrence, irrespective of treatment. Reconstruction of mother-father-patient trios revealed the potential for inherited and *de novo* germline variants and somatic mutations to contribute to the etiology of this pediatric cancer.

Methods

Contact for reagent and resource . Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Marco Gallo (marco.gallo@ucalgary.ca). Sharing of primary samples, based on availability, may be subject to MTAs and will require appropriate research ethics board certifications.

Human samples. All samples were collected and used for research with appropriate informed consent and with approval by the Health Research Ethics Board of Alberta and the Research Ethics Board of the Hospital for Sick Children. Fresh tumor tissue and blood samples from each patient were collected by the Clark Smith Brain Tumour biobank at the University of Calgary and the Hospital for Sick Children and preserved at either -80°C or in vapor phase liquid nitrogen. For selected cases, a portion the fresh tumou material was also allocated for patient-derived xenograft preparation (see below). Anagraphical information of sample donors are in **Table S2** and **Table S3**.

Genomic DNA extraction and library preparation. High molecular weight genomic DNA (gDNA) was extracted from frozen tumor tissue and frozen blood samples with the Qiagen MagAttract® HMW DNA Kit (catalog # 67563). gDNA fragment size and distribution were quantified with the Agilent Technologies 2200 TapeStation Genomic DNA Assay. Samples with gDNA fragment size > 50 kb were used for library construction. The 10xGenomics Chromium™ Genome Library Kit & Gel Bead Kit v2, 16 reactions (catalog # PN-120258) was used for library preparation. Size and distribution of all sequencing libraries were quantified with the Agilent Technologies 2200 TapeStation D1000 Assay.

Whole-genome sequencing and linked-read data analysis. Whole-genome sequencing was performed at The Centre for Applied Genomics at the Hospital for Sick Children (Toronto, ON) with a HiSeq X Series (Illumina) instrument. Libraries were validated on a Bioanalyzer DNA High Sensitivity chip to check for size and absence of primer dimers, and quantified by qPCR using Kapa Library Quantification Illumina/ABI Prism Kit protocol (KAPA Biosystems). Validated libraries were paired-end sequenced on an Illumina HiSeq X platform following Illumina's recommended protocol to generate paired-end reads of 150-bases in length. Each library was loaded on a single lane. All libraries were sequenced at 30x coverage. Post-sequencing, fastq files were analyzed with the package LongRanger-2.1.6 (10xGenomics) on servers hosted by the Centre for Health Genomics and Informatics (University of Calgary). LongRanger performed alignment of reads to the reference genome (GRCh38), assembled linked-reads, generated haplotype blocks, and called single nucleotide polymorphisms and structural variants. LongRanger output files were visualized with Loupe browser 2.1.2 (10xGenomics). SNVs were called using both Mutect2 *(57)*

24

and Strelka2 (https://github.com/Illumina/strelka). Only SNVs called by both methods were included for downstream analysis. Annotation of SNVs was performed using ANNOVAR *(58)*.

Structural variant calling using GROC-SVs. GROC-SVs was run for all samples in each patient independently using default parameters. SVs in patients with parental germlines available were further analyzed in order to empirically infer a cutoff for barcode coverage that was sufficient to call an SV. For instance, events observed in both parental germlines and the patient tumor samples but not in the patient germline must be false negative calls in the patient germline as they are inherited. A threshold of 200 barcodes (physical coverage) was sufficient to minimize false negative calls across the cohort, without significantly compromising our ability to detect all true positive events. For patients 1 and 3, events observed in both parents were excluded from further analysis, unless they were somatically altered (i.e. absent in the primary or recurrent tumor). The overlap of SV breakpoints and genes was assessed using bedtools intersect, and these breakpoints were visualized using the ProteinPaint software (pecan.stjude.cloud/proteinpaint) along with known pediatric aberrations from a number of databases. These include PCGP (St. Jude - WashU Pediatric Cancer Genome Project), TARGET (Therapeutically Applicable Research To Generate Effective Treatments), SCMC (Shanghai Children's Medical Center pediatric ALL project), UTSMC (UT Southwestern Medical Center Wilms' tumor study), and DKFZ (German Cancer Research Center Wilms' tumor study) *(35)*.

Somatic SNV detection and filtering using varscan2. Somatic variants were also called genome-wide using varscan2 (v2.4.3). Varscan2 tumor-normal variant calling produced variants categorized as somatic, germline or loss of heterozygosity (LOH). Briefly, alignments from 10X

25

Genomics LongRanger underwent tumor-normal variant calling using samtools mpileup (v1.7), to produce a single tumor-normal mpileup file *(59)*. The mpileup files were processed to yield variants using the varscan2 <u>somatic</u> function and high confidence variants were isolated using the varscan2 <u>processSomatic</u> function. Bam read counts files were created using a combination of tumor or normal bam and the appropriate high confidence variant coordinates; one for LOH, one for germline and one for somatic variants. Variants were filtered using the <u>fpfilter</u> function under dream-3 settings to produce a high confidence and filtered set of variants.

Somatic copy number alteration (CNA) calling, filtration and segmentation. Somatic copy number alterations were estimated using varscan2. Tumor-normal mpileup files (see SNV detection, above) were used to generate raw copy numbers using the <u>copynumber</u> function, and further adjusted for for GC content using the <u>copycaller</u> function. Adjusted calls underwent segmentation using a recursive algorithm, circular binary segmentation as implemented in the package DNAcopy *(60)*. Segmented CNA was analyzed and readjusted using <u>copycaller</u> with <u>recenter-up</u> or <u>recenter-down</u>, for segments showing a consistent deviation from a baseline (0.0). Adjacent segments of similar copy number log ratios (<0.25 difference) were merged using a varscan2 affiliated perl script and a custom bash script.

Somatic SNV classification using mclust. Variants were classified based on their variant allele frequencies (VAFs) into distinct clusters using the R package mclust *(61)*. Mclust uses machine-learning model-based estimation based on finite normal mixture modelling using Expectation-Maximization (EM) and the Bayesian Information Criterion, essentially allowing for the grouping

of variants based on allele frequency into distinct distributions. These can be manually assigned into 'homozygous', 'clonal' or 'subclonal'.

Cancer Cell fractions (CCF). Variant cancer cell fractions (CCF) indicates the prevalence of a mutation throughout the tumor sample, with higher CCF values indicating variants clonal in nature. CCFs were calculated as follows *(62)*:

CCF = VAF * (1/Purity)(CN*Purity +2(1-Purity))

Where VAF is the variant allele frequency, CN is the local copy number of the mutation and purity is the estimation as calculated by EXPANDS. Both VAF and CN were calculated using varscan2 (see above).

Phylogenetic reconstruction from combined SNV and CNA data. CNA and loss-of-heterozygosity (LOH) data were combined with somatic variants to infer subclonal architecture using EXPANDS *(22)* independently for each sample. EXPANDS (V2.1.1) was run using the runExpands function with default settings. As per the default, only subpopulations that had a minimum cellular frequency (CF) of 0.1 or greater were taken into consideration. Mutations that were not clusterable were excluded from reconstruction. Mutations were assigned to all appropriate populations and subpopulations, (ie a mutation can be assigned to a parent population (trunk) and a 'daughter' subpopulation (branch).

Phylogenetic reconstruction of the subpopulations in primary and recurrent tumors was performed by calculating the Manhattan distance metric between all subpopulations, followed by hierarchical

clustering using full linkage. Visualization of the phylogenies were created using the as.phylo function from the ape package *(63)*.

Copy number spectrum plots. CNA data from varscan2 copy number analysis was used to visualise copy number gains and losses. The copy number spectrum function for the GenVisR package provides a method of visualizing primary and recurrence CNA changes *(64)*. The genome boundaries were created using BSgenome.Hsapiens.UCSC.hg38 reference.

Loss of heterozygosity plots. Germline and loss of heterozygosity (LOH) variants called by varscan2 were used to visualize LOH regions of the genome using the lohspec function in the GenVisR package. This enabled detection of copy-neutral LOH regions where one parental locus is deleted and the other is duplicated. LOH regions were found using a 2.5 Mbp sliding window to calculate difference in variant allele frequency of germline and LOH heterozygous variants in tumor and normal samples. Within each window the absolute VAF difference is calculated and a mean is obtained. The algorithm continues recursively by moving the windows forward 1 Mbp and calculating the mean tumor-normal VAF difference of LOH and germline variants genome-wide, until a stable solution is found.

Mutational signatures. To determine if different mutational processes were occurring we analyzed mutational signatures. Sets of somatic variants from distinct subclonal populations in primary or recurrent tumors identified by EXPANDS were analyzed separately, and the mutation signatures were determined using the deconstructSigs (v1.8.0) package *(65)*. Mutational signatures were generated for each subclonal population group, all groups were above 50 variants, with the

28

exception of 3f (patient 3), 1ef (patient 1) and 2b (patient 2). DeconstructSigs was run using 30 signatures from the Catalogue Of Somatic Mutation In Cancer public repository as the reference signatures *(66)*. Default settings were used when generating mutational signatures with the exception of genome normalization and weight signature cutoff, which was changed to 0.05.

Allele frequency distribution. Previously, it has been argued that the dynamics of tumor cells may find a signature in hallmark statistical clone size dependences *(67, 68)*. Specifically, for systems characterized by a proliferative hierarchy in which slow-cycling tumor stem-like cells give rise to rapidly cycling, self-renewing, progenitor-like cells that in turn generate short-lived non-cycling cells, the distribution of variant allele fractions (VAFs) is predicted to acquire a negative binomial frequency dependence *(43, 67)*:

$$p\left(f\right) = \frac{1}{N_0\, f} e^{-\frac{f}{f_0}} \qquad (1)$$

Here, the leading *1/f* dependence, highlighted by Williams et al. 2016, represents a generic signature of neutral clone competition. To explore whether the current tumor WGS data can be embraced within the same paradigm, we considered the VAF distribution from patient samples. To this end, we considered a fit to the first incomplete moment of the VAF distribution, which for a negative binomial size dependence takes a pure exponential form *(43, 67)*.

Due to the sequencing resolution limit (~150x physical coverage) we introduced a lower cut-off on the allele frequencies at $f_{min} = 0.12$ *(69)*. At the upper end of the scale, we set a cut-off at $f_{max} = 0.4$, noting that, in the absence of copy number variation, mutations with an allele fraction of 0.5 and above will include those that are fixed across the population. We therefore

adapted the expression for the first incomplete moment to account for this limited allele frequency interval of $[f_{min}, f_{max}]$, defining the measure,

$$\mu_1(f) = \frac{1}{\langle f \rangle} \int_f^{f_{max}} x P(x)\, dx \approx \frac{e^{-\frac{f}{f_0}} - e^{-\frac{f_{max}}{f_0}}}{e^{-\frac{f_{min}}{f_0}} - e^{-\frac{f_{max}}{f_0}}} \qquad (2)$$

where $< f >$ denotes the average allele frequency within the interval and f_0 is the fitting parameter. For details on how the VAF distribution is derived, we refer to ref. *(67)*.

The *NonlinearModelFit* function in *Mathematica 11.0.1.0* was then used to perform a least-squares fit of the first incomplete moment to the sample (allele frequencies corrected for the modest degree of copy number variation and purity). The results of the fit can be found in Table S7, Figure 7B-E and Figure S7A-D. Overall, the findings are largely consistent with a negative binomial tumor clone size dependence. These results suggest that tumor cells are organized in a conserved proliferative hierarchy, with tumor growth characterized by neutral competition between constituent tumor subclones.

For completeness, we then considered a direct fit of the VAF distribution to a negative binomial size dependence (1) over the same interval. Grouping the data into bins of size $(f_{max} - f_{min})/30$, the results of a least-squares fit, obtained using the *NonlinearModelFit* function in *Mathematica 11.0.1.0*, is shown in **Fig. S5E-I**.

Evolution of allele fractions across recurrences. To further challenge the hierarchical model tumor cell dynamics, we took advantage of the availability of recurrent tumor samples. Within the framework of the model, competition between individual tumor stem cell clones is predicted to be neutral, allowing clones to readily adjust, up or down, their fractional tumor contribution. In this

case, the ensemble of clones bearing mutations that are acquired, or rise above detection threshold, during recurrence provides a pure sample on which to test the model through the predicted negative binomial dependence of the VAF distribution. By contrast, mutations that are shared between primary and recurrent tumors are expected to depart from this dependence, either because they are clonally fixed within the tumor population (with VAFs around 0.5), or because they occupy a significant fraction of the tumor at the start of the recurrent period of tumor growth. Notably, for *de novo* mutations that arise at recurrence, using the same least-squares fitting procedure, we find a strikingly good agreement with a negative binomial dependence (**Fig. S5J-N**), while the distribution of shared mutations is typically peaked around larger values, as expected (**Fig. S5O-R**).

Copy number assays for the *ATRX* locus. Custom TaqMan® Copy Number Assay Probes (catalog # 4400294) from Applied Biosystems™ by Thermo Fisher Scientific were used to detect changes in copy number at the *ATRX* locus in all patients. Quantification was performed according to manufacturers' protocols. Briefly, for each patient, a control DNA sample, patient germline DNA, primary tumor DNA, and recurrent tumor DNA were diluted to 5 ng/μL. DNA was quantified in triplicates with the Qubit dsDNA HS Assay Kit (catalog # Q32851). The TaqMan® Genotyping Master Mix and TaqMan® Copy Number Assay was added to each well of a 96-well plate. 20 ng of DNA was added in triplicate wells of a 96-well plate for each assay condition. The samples were then ran on a Bio-Rad CFX Connect™ Real-Time PCR Detection System with the following parameters: hold at 95°C for 10 mins, 40 cycles of 95°C for 15 sec and 60°C for 60 sec. Fluorescence signal for the *ATRX* deleted region (ATRXdel probe) was first normalized to the fluorescence signal from the control diploid region probe. Fluorescence was additionally

31

normalized to the diploid control DNA sample. TaqMan® Probes are labeled with FAM at 5' end and MGBNFQ at the 3' end. For the probe in the *ATRX* deleted region; probe sequence: CACACCCAAATATTGGTAAAAAT, forward primer: TCCAGGACTTAGCAGGATGTGA, reverse primer: ACCCCATCAAGTAGATGGTAAGAAACT. For the probe in the control diploid region; probe sequence: ACACGGGTGTTGAAGACGC, forward primer: AGGCGCTGTGGAAACAATTATAGTA, reverse primer: TGTCCCCGTCAACGATCAC.

Patient-derived xenografts. All mice were used with approval from the Animal Care Committee of the University of Calgary. Fresh tumor tissue was gently dissociated by trituration, filtered using a 0.7 µm filter, washed and resuspended in sterile DPBS (Gibco). Tumor cells were injected into the right striatum of the brains of CB17 SCID mice (Charles River Laboratory). Animals were monitored for tumor growth and tumors were isolated when signs of morbidity were observed. Initial tumor establishment for patient specimens SM4021 (recurrence of patient 3) and SM4058 (third recurrence of patient 5) took 59 and 41 days respectively. Tumors were then maintained by serial *in vivo* passaging in the brain of SCID mice every 1-2 months or cut into 1 mm x 1 mm tissue pieces and cryo-stored in 1 mL CryoStor CS10 solution (Stem Cell) for 10 minutes on ice prior to being stored at -80ºC. In experiments where cryo-stored tissue was used, tissue was thawed, dissociated into a single-cell suspension, filtered using a 0.7 µm filter and implanted into the brain of SCID mice. For single-cell sequencing, established xenografts were isolated and dissociated as described below. All xenografts were identity matched to the original patient material by short tandem repeat analysis using the AmpFLSTR Identifiler Plus kit (Applied Biosystems).

Single-cell RNA-seq (Calgary cohort). Xenograft tumor tissue was dissociated into single cells using the following protocol: 10 mL of ACCUTASE™ (catalog # 07920, Sigma-Aldrich) was added to tumor tissue for manual dissociation with surgical scissors. The sample was then incubated with sterile glass beads at 37°C, rotating continuously for 30 min. The glass beads were removed, and the sample was spun at 1000 rpm for 5 minutes at 25°C. The cell pellet was resuspended in neural stem cell media (NeuroCult™ NS-A Proliferation Kit (Human), catalog # 05751). Mouse cells were removed with the Miltenyi Biotec Mouse Cell Depletion Kit (catalog # 130-104-694) and the autoMACS® Pro Separator and columns (catalog # 130-021-101). Protocol was followed as per manufacturers' instructions. Viability and concentration of the purified human cells were quantified. Cell samples with viability > 80% were used for subsequent single-cell RNA-sequencing. ~2500 cells were loaded into the Chromium™ console and were used for library preparation. The 10xGenomics Chromium™ Single Cell 3' Library & Gel Bead Kit v2, 16 reactions (catalog # PN-120237) was used for library preparation. Size and distribution of all sequencing libraries were quantified with the Agilent Technologies 2200 TapeStation High Sensitivity D1000 Assay. Single-cell RNA-sequencing was performed at the Centre for Health Genomics and Informatics at the University of Calgary. All libraries were sequenced on mid-output 150 cycle NextSeq500 runs, with ~130 million reads each. Sequencing (BCL) files were processed with the Cell Ranger 2.1.0 package (10xGenomics). Reads were aligned to the GRCh38 transcriptome. Data standardization and normalization was performed with the Seurat package *(70)*.

Bulk RNA-seq (Toronto cohort). RNA from tissue and cells was extracted using Qiagen AllPrep DNA/RNA/miRNA Universal Kit (catalog # 80224). Strand-specific RNA-seq (ssRNA-seq)

libraries were constructed from total RNA samples using plate-based protocols. Libraries were sequenced at 75 bp PET using V4 chemistry on a HiSeq 2500 instrument (Illumina) at the Genome Sciences Centre (Vancouver, BC). Reads were aligned with the STAR aligner *(71)* v2.4.2a to hg38 human reference genome (from iGenome). R Bioconductor DESEq2 package *(72)* was used for normalization and vst transformation of the gene expression matrix

Graphing software. GraphPad Prism 7.0c was used for graph generation. The Venn diagrams were generated using the eulerr R package version 4.0.0 (www.cran.r-project.org/web/packages/eulerr/ index.html). Sankey Diagrams were generated using SankeyMATIC (www.sankeymatic.com) and the outputs were modified to be of appropriate format for figures.

External datasets. Identification of genes included in the OMIM database of genetic inheritance was done with DAVID bioinformatics resources 6.8 *(73, 74)*. Survival analysis of pediatric glioma patients was done with R2: Genomics analysis and visualization platform (www.r2.amc.nl) using a previously published dataset by Paugh et al *(42)*. Assessment of allelic fractions for *ATRX* mutations was done using an independent dataset *(36)* and available at www.pedpancan.com. GTEx data was assessed at https://www.gtexportal.org/home/ *(37)*. The Database of Genomic Variants was accessed at http://dgv.tcag.ca/dgv/app/home *(75)*. Developmental human brain gene expression data was extracted from BrainSpan (http://www.brainspan.org).

Statistical analysis of enrichment of OMIM genes in the pGBM mutational dataset. The number *X* of genes harboring non-synonymous, stop gain/stop loss, frame shift, non-frame shift

SNVs and indels was identified for primary and recurrent tumors. From these, genes with a phenotype in OMIM were selected (*Y*). The OMIM genes (n = 13,520) with a clinical phenotype were identified using the genemap2.txt file (containing gene and phenotype information) downloaded from OMIM (https://www.omim.org/downloads/). For p-value analysis, *X* number of random genes were selected from all protein coding genes in Ensemble (n = 20,345) and among these the number of genes with related OMIM phenotype were selected (Rand_OMIM). This analysis was repeated 100,000 times. The p-value was identified as the number of times 'Rand_OMIM' genes were higher than *Y* number of genes in the 100,000 iterations.

References

1. D. Sturm, H. Witt, V. Hovestadt, D. A. Khuong-Quang, D. T. W. Jones, C. Konermann, E. Pfaff, M. Tönjes, M. Sill, S. Bender, M. Kool, M. Zapatka, N. Becker, M. Zucknick, T. Hielscher, X. Y. Liu, A. M. Fontebasso, M. Ryzhova, S. Albrecht, K. Jacob, M. Wolter, M. Ebinger, M. U. Schuhmann, T. van Meter, M. C. Frühwald, H. Hauch, A. Pekrun, B. Radlwimmer, T. Niehues, G. Von Komorowski, M. Dürken, A. E. Kulozik, J. Madden, A. Donson, N. K. Foreman, R. Drissi, M. Fouladi, W. Scheurlen, A. von Deimling, C. Monoranu, W. Roggendorf, C. Herold-Mende, A. Unterberg, C. M. Kramm, J. Felsberg, C. Hartmann, B. Wiestler, W. Wick, T. Milde, O. Witt, A. M. Lindroth, J. Schwartzentruber, D. Faury, A. Fleming, M. Zakrzewska, P. P. Liberski, K. Zakrzewski, P. Hauser, M. Garami, A. Klekner, L. Bognar, S. Morrissy, F. Cavalli, M. D. Taylor, P. van Sluis, J. Koster, R. Versteeg, R. Volckmann, T. Mikkelsen, K. Aldape, G. Reifenberger, V. P. Collins, J. Majewski, A. Korshunov, P. Lichter, C. Plass, N. Jabado, S. M. Pfister, Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma, *Cancer Cell* **22**, 425–437 (2012).

2. K. J. Cohen, I. F. Pollack, T. Zhou, A. Buxton, E. J. Holmes, P. C. Burger, D. J. Brat, M. K. Rosenblum, R. L. Hamilton, R. S. Lavey, R. L. Heideman, Temozolomide in the treatment of high-grade gliomas in children: A report from the Children's Oncology Group, *Neuro. Oncol.* **13**, 317–323 (2011).

3. J. Schwartzentruber, A. Korshunov, X. Y. Liu, D. T. W. Jones, E. Pfaff, K. Jacob, D. Sturm, A. M. Fontebasso, D. A. K. Quang, M. Tönjes, V. Hovestadt, S. Albrecht, M. Kool, A. Nantel, C. Konermann, A. Lindroth, N. Jäger, T. Rausch, M. Ryzhova, J. O. Korbel, T. Hielscher, P. Hauser, M. Garami, A. Klekner, L. Bognar, M. Ebinger, M. U. Schuhmann, W. Scheurlen, A. Pekrun, M. C. Frühwald, W. Roggendorf, C. Kramm, M. Dürken, J. Atkinson, P. Lepage, A. Montpetit, M. Zakrzewska, K. Zakrzewski, P. P. Liberski, Z. Dong, P. Siegel, A. E. Kulozik, M. Zapatka, A. Guha, D. Malkin, J. Felsberg, G. Reifenberger, A. Von Deimling, K. Ichimura, V. P. Collins, H. Witt, T. Milde, O. Witt, C. Zhang, P. Castelo-Branco, P. Lichter, D. Faury, U. Tabori, C. Plass, J. Majewski, S. M. Pfister, N. Jabado, Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma, *Nature* **482**, 226–231 (2012).

4. D. A. Khuong-Quang, P. Buczkowicz, P. Rakopoulos, X. Y. Liu, A. M. Fontebasso, E. Bouffet, U. Bartels, S. Albrecht, J. Schwartzentruber, L. Letourneau, M. Bourgey, G. Bourque, A. Montpetit, G. Bourret, P. Lepage, A. Fleming, P. Lichter, M. Kool, A. Von Deimling, D. Sturm, A. Korshunov, D. Faury, D. T. Jones, J. Majewski, S. M. Pfister, N. Jabado, C. Hawkins, K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas, *Acta Neuropathol.* **124**, 439–447 (2012).

5. L. Bjerke, A. Mackay, M. Nandhabalan, A. Burford, A. Jury, S. Popov, D. A. Bax, D. Carvalho, K. R. Taylor, M. Vinci, I. Bajrami, I. M. McGonnell, C. J. Lord, R. M. Reis, D. Hargrave, A. Ashworth, P. Workman, C. Jones, Histone H3.3 mutations drive pediatric

glioblastoma through upregulation of MYCN, *Cancer Discov.* **3**, 512–519 (2013).

6. P. W. Lewis, M. M. Müller, M. S. Koletsky, F. Cordero, S. Lin, L. A. Banaszynski, B. A. Garcia, T. W. Muir, O. J. Becher, C. D. Allis, Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma, *Science (80-.).* **340**, 857–861 (2013).

7. K. Funato, T. Major, P. W. Lewis, C. D. Allis, V. Tabar, Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation., *Science* **346**, 1529–33 (2014).

8. M. Pathania, N. De Jay, N. Maestro, A. S. Harutyunyan, J. Nitarska, P. Pahlavan, S. Henderson, L. G. Mikael, A. Richard-Londt, Y. Zhang, J. R. Costa, S. Hébert, S. Khazaei, N. S. Ibrahim, J. Herrero, A. Riccio, S. Albrecht, R. Ketteler, S. Brandner, C. L. Kleinman, N. Jabado, P. Salomoni, H3.3 K27M Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas, *Cancer Cell* **32**, 684–700.e9 (2017).

9. C. Jones, S. J. Baker, Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma*Nat. Rev. Cancer* **14**, 651–661 (2014).

10. R. Salloum, M. K. McConechy, L. G. Mikael, C. Fuller, R. Drissi, M. DeWire, H. Nikbakht, N. De Jay, X. Yang, D. Boue, L. M. L. Chow, J. L. Finlay, T. Gayden, J. Karamchandani, T. R. Hummel, R. Olshefski, D. S. Osorio, C. Stevenson, C. L. Kleinman, J. Majewski, M. Fouladi, N. Jabado, Characterizing temporal genomic heterogeneity in pediatric high-grade gliomas, *Acta Neuropathol. Commun.* **5**, 78 (2017).

11. M. Vinci, A. Burford, V. Molinari, K. Kessler, S. Popov, Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells, *Nat. Med.* **epub ehead** (2018), doi:10.1038/s41591-018-0086-7.

12. Q. Ostrom, H. Gittleman, J. Fulop, M. Liu, R. Blanda, C. Kromer, Y. Wolinsky, C. Kruchko, CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006-2010, *Neuro Oncol.* **17**, iv1-iv62 (2015).

13. J. Zhang, G. Wu, C. P. Miller, R. G. Tatevossian, J. D. Dalton, B. Tang, W. Orisme, C. Punchihewa, M. Parker, I. Qaddoumi, F. A. Boop, C. Lu, C. Kandoth, L. Ding, R. Lee, R. Huether, X. Chen, E. Hedlund, P. Nagahawatte, M. Rusch, K. Boggs, J. Cheng, J. Becksfort, J. Ma, G. Song, Y. Li, L. Wei, J. Wang, S. Shurtleff, J. Easton, D. Zhao, R. S. Fulton, L. L. Fulton, D. J. Dooling, B. Vadodaria, H. L. Mulder, C. Tang, K. Ochoa, C. G. Mullighan, A. Gajjar, R. Kriwacki, D. Sheer, R. J. Gilbertson, E. R. Mardis, R. K. Wilson, J. R. Downing, S. J. Baker, D. W. Ellison, Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas, *Nat. Genet.* **45**, 602–612 (2013).

14. International Cancer Genome Consortium PedBrain Tumor Project, Recurrent MET fusion genes represent a drug target in pediatric glioblastoma., *Nat. Med.* **22**, 1314–1320 (2016).

15. A. S. Morrissy, L. Garzia, D. J. H. Shih, S. Zuyderduyn, X. Huang, P. Skowron, M. Remke, F. M. G. Cavalli, V. Ramaswamy, P. E. Lindsay, S. Jelveh, L. K. Donovan, X. Wang, B. Luu, K. Zayne, Y. Li, C. Mayoh, N. Thiessen, E. Mercier, K. L. Mungall, Y. Ma, K. Tse, T. Zeng, K. Shumansky, A. J. L. Roth, S. Shah, H. Farooq, N. Kijima, B. L. Holgado, J. J. Y. Lee, S. Matan-Lithwick, J. Liu, S. C. Mack, A. Manno, K. A. Michealraj, C. Nor, J. Peacock, L. Qin, J. Reimand, A. Rolider, Y. Y. Thompson, X. Wu, T. Pugh, A. Ally, M. Bilenky, Y. S. N. Butterfield, R. Carlsen, Y. Cheng, E. Chuah, R. D. Corbett, N. Dhalla, A. He, D. Lee, H. I. Li, W. Long, M. Mayo, P. Plettner, J. Q. Qian, J. E. Schein, A. Tam, T. Wong, I. Birol, Y. Zhao, C. C. Faria, J. Pimentel, S. Nunes, T. Shalaby, M. Grotzer, I. F. Pollack, R. L. Hamilton, X. N. Li, A. E. Bendel, D. W. Fults, A. W. Walter, T. Kumabe, T. Tominaga, V. P. Collins, Y. J. Cho, C.

Hoffman, D. Lyden, J. H. Wisoff, J. H. Garvin, D. S. Stearns, L. Massimi, U. Schüller, J. Sterba, K. Zitterbart, S. Puget, O. Ayrault, S. E. Dunn, D. P. C. Tirapelli, C. G. Carlotti, H. Wheeler, A. R. Hallahan, W. Ingram, T. J. MacDonald, J. J. Olson, E. G. Van Meir, J. Y. Lee, K. C. Wang, S. K. Kim, B. K. Cho, T. Pietsch, G. Fleischhack, S. Tippelt, Y. S. Ra, S. Bailey, J. C. Lindsey, S. C. Clifford, C. G. Eberhart, M. K. Cooper, R. J. Packer, M. Massimino, M. L. Garre, U. Bartels, U. Tabori, C. E. Hawkins, P. Dirks, E. Bouffet, J. T. Rutka, R. J. Wechsler-Reya, W. A. Weiss, L. S. Collier, A. J. Dupuy, A. Korshunov, D. T. W. Jones, M. Kool, P. A. Northcott, S. M. Pfister, D. A. Largaespada, A. J. Mungall, R. A. Moore, N. Jabado, G. D. Bader, S. J. M. Jones, D. Malkin, M. A. Marra, M. D. Taylor, Divergent clonal selection dominates medulloblastoma at recurrence, *Nature* (2016), doi:10.1038/nature16478.

16. B. E. Johnson, T. Mazor, C. Hong, M. Barnes, K. Aihara, C. Y. McLean, S. D. Fouse, S. Yamamoto, H. Ueda, K. Tatsuno, S. Asthana, L. E. Jalbert, S. J. Nelson, A. W. Bollen, W. C. Gustafson, E. Charron, W. A. Weiss, I. V. Smirnov, J. S. Song, A. B. Olshen, S. Cha, Y. Zhao, R. A. Moore, A. J. Mungall, S. J. M. Jones, M. Hirst, M. A. Marra, N. Saito, H. Aburatani, A. Mukasa, M. S. Berger, S. M. Chang, B. S. Taylor, J. F. Costello, Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma, *Science (80-.).* **343**, 189–193 (2014).

17. S. K. McLaughlin, S. N. Olsen, B. Dake, T. De Raedt, E. Lim, R. T. Bronson, R. Beroukhim, K. Polyak, M. Brown, C. Kuperwasser, K. Cichowski, The RasGAP Gene, RASAL2, Is a Tumor and Metastasis Suppressor, *Cancer Cell* **24**, 365–378 (2013).

18. Y. Ikeda, K. Kiyotani, P. Y. Yew, T. Kato, K. Tamura, K. L. Yap, S. M. Nielsen, J. L. Mester, C. Eng, Y. Nakamura, R. H. Grogan, Germline PARP4 mutations in patients with primary thyroid and breast cancers, *Endocr. Relat. Cancer* **23**, 171–179 (2016).

19. S. Inoue, Z. Hao, A. J. Elia, D. Cescon, L. Zhou, J. Silvester, B. Snow, I. S. Harris, M.

Sasaki, W. Y. Li, M. Itsumi, K. Yamamoto, T. Ueda, C. Dominguez-Brauer, C. Gorrini, I. I. Christine Chio, J. Haight, A. You-Ten, S. McCracken, A. Wakeham, D. Ghazarian, L. J. Z. Penn, G. Melino, T. W. Mak, Mule/Huwe1/Arf-BP1 suppresses Ras-driven tumorigenesis by preventing c-Myc/Miz1-mediated down-regulation of p21 and p15, *Genes Dev.* **27**, 1101–1114 (2013).

20. K. B. Myant, P. Cammareri, M. C. Hodder, J. Wills, A. Von Kriegsheim, B. Győrffy, M. Rashid, S. Polo, E. Maspero, L. Vaughan, B. Gurung, E. Barry, A. Malliri, F. Camargo, D. J. Adams, A. Iavarone, A. Lasorella, O. J. Sansom, *HUWE1* is a critical colonic tumour suppressor gene that prevents MYC signalling, DNA damage accumulation and tumour initiation, *EMBO Mol. Med.* **9**, 181–197 (2017).

21. J. Cheng, J. Demeulemeester, D. C. Wedge, H. K. M. Vollan, J. J. Pitt, H. G. Russnes, B. P. Pandey, G. Nilsen, S. Nord, G. R. Bignell, K. P. White, A. L. Børresen-Dale, P. J. Campbell, V. N. Kristensen, M. R. Stratton, O. C. Lingjærde, Y. Moreau, P. Van Loo, Pan-cancer analysis of homozygous deletions in primary tumours uncovers rare tumour suppressors, *Nat. Commun.* **8** (2017), doi:10.1038/s41467-017-01355-0.

22. N. Andor, J. V. Harness, S. Müller, H. W. Mewes, C. Petritsch, Expands: Expanding ploidy and allele frequency on nested subpopulations, *Bioinformatics* **30**, 50–60 (2014).

23. L. Giunti, M. Pantaleo, I. Sardi, A. Provenzano, A. Magi, S. Cardellicchio, F. Castiglione, L. Tattini, F. Novara, A. M. Buccoliero, M. de Martino, L. Genitori, O. Zuffardi, S. Giglio, Genome-wide copy number analysis in pediatric glioblastoma multiforme, *Am. J. Cancer Res.* **4**, 293–303 (2014).

24. G. Wu, A. K. Diaz, B. S. Paugh, S. L. Rankin, B. Ju, Y. Li, X. Zhu, C. Qu, X. Chen, J.

Zhang, J. Easton, M. Edmonson, X. Ma, C. Lu, P. Nagahawatte, E. Hedlund, M. Rusch, S. Pounds, T. Lin, A. Onar-Thomas, R. Huether, R. Kriwacki, M. Parker, P. Gupta, J. Becksfort, L. Wei, H. L. Mulder, K. Boggs, B. Vadodaria, D. Yergeau, J. C. Russell, K. Ochoa, R. S. Fulton, L. L. Fulton, C. Jones, F. A. Boop, A. Broniscer, C. Wetmore, A. Gajjar, L. Ding, E. R. Mardis, R. K. Wilson, M. R. Taylor, J. R. Downing, D. W. Ellison, J. Zhang, S. J. Baker, The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma, *Nat. Genet.* **46**, 444–450 (2014).

25. S. Douma, T. Van Laar, J. Zevenhoven, R. Meuwissen, E. Van Garderen, D. S. Peeper, Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB, *Nature* **430**, 1034–1040 (2004).

26. M. Eguchi, M. Eguchi-Ishimae, a Tojo, K. Morishita, K. Suzuki, Y. Sato, S. Kudoh, K. Tanaka, M. Setoyama, F. Nagamura, S. Asano, N. Kamada, Fusion of ETV6 to neurotrophin-3 receptor TRKC in acute myeloid leukemia with t(12;15)(p13;q25)., *Blood* **93**, 1355–63 (1999).

27. Z. Li, C. E. Tognon, F. J. Godinho, L. Yasaitis, H. Hock, J. I. Herschkowitz, C. L. Lannon, E. Cho, S. J. Kim, R. T. Bronson, C. M. Perou, P. H. Sorensen, S. H. Orkin, ETV6-NTRK3 Fusion Oncogene Initiates Breast Cancer from Committed Mammary Progenitors via Activation of AP1 Complex, *Cancer Cell* **12**, 542–558 (2007).

28. N. Agrawal, R. Akbani, B. A. Aksoy, A. Ally, H. Arachchi, S. L. Asa, J. T. Auman, M. Balasundaram, S. Balu, S. B. Baylin, M. Behera, B. Bernard, R. Beroukhim, J. A. Bishop, A. D. Black, T. Bodenheimer, L. Boice, M. S. Bootwalla, J. Bowen, R. Bowlby, C. A. Bristow, R. Brookens, D. Brooks, R. Bryant, E. Buda, Y. S. N. Butterfield, T. Carling, R. Carlsen, S. L. Carter, S. E. Carty, T. A. Chan, A. Y. Chen, A. D. Cherniack, D. Cheung, L. Chin, J. Cho, A. Chu, E. Chuah, K. Cibulskis, G. Ciriello, A. Clarke, G. L. Clayman, L. Cope, J. A. Copland, K.

Covington, L. Danilova, T. Davidsen, J. A. Demchok, D. DiCara, N. Dhalla, R. Dhir, S. S.

Dookran, G. Dresdner, J. Eldridge, G. Eley, A. K. El-Naggar, S. Eng, J. A. Fagin, T. Fennell, R.

L. Ferris, S. Fisher, S. Frazer, J. Frick, S. B. Gabriel, I. Ganly, J. Gao, L. A. Garraway, J. M.

Gastier-Foster, G. Getz, N. Gehlenborg, R. Ghossein, R. A. Gibbs, T. J. Giordano, K. Gomez-

Hernandez, J. Grimsby, B. Gross, R. Guin, A. Hadjipanayis, H. A. Harper, D. N. Hayes, D. I.

Heiman, J. G. Herman, K. A. Hoadley, M. Hofree, R. A. Holt, A. P. Hoyle, F. W. Huang, M.

Huang, C. M. Hutter, T. Ideker, L. Iype, A. Jacobsen, S. R. Jefferys, C. D. Jones, S. J. M. Jones,

K. Kasaian, E. Kebebew, F. R. Khuri, J. Kim, R. Kramer, R. Kreisberg, R. Kucherlapati, D. J.

Kwiatkowski, M. Ladanyi, P. H. Lai, P. W. Laird, E. Lander, M. S. Lawrence, D. Lee, E. Lee, S.

Lee, W. Lee, K. M. Leraas, T. M. Lichtenberg, L. Lichtenstein, P. Lin, S. Ling, J. Liu, W. Liu,

Y. Liu, V. A. LiVolsi, Y. Lu, Y. Ma, H. S. Mahadeshwar, M. A. Marra, M. Mayo, D. G.

McFadden, S. Meng, M. Meyerson, P. A. Mieczkowski, M. Miller, G. Mills, R. A. Moore, L. E.

Mose, A. J. Mungall, B. A. Murray, Y. E. Nikiforov, M. S. Noble, A. I. Ojesina, T. K.

Owonikoko, B. A. Ozenberger, A. Pantazi, M. Parfenov, P. J. Park, J. S. Parker, E. O. Paull, C.

S. Pedamallu, C. M. Perou, J. F. Prins, A. Protopopov, S. S. Ramalingam, N. C. Ramirez, R.

Ramirez, B. J. Raphael, W. K. Rathmell, X. Ren, S. M. Reynolds, E. Rheinbay, M. D. Ringel, M.

Rivera, J. Roach, A. G. Robertson, M. W. Rosenberg, M. Rosenthal, S. Sadeghi, G. Saksena, C.

Sander, N. Santoso, J. E. Schein, N. Schultz, S. E. Schumacher, R. R. Seethala, J. Seidman, Y.

Senbabaoglu, S. Seth, S. Sharpe, K. R. M. Shaw, J. P. Shen, R. Shen, S. Sherman, M. Sheth, Y.

Shi, I. Shmulevich, G. L. Sica, J. V. Simons, R. Sinha, P. Sipahimalani, R. C. Smallridge, H. J.

Sofia, M. G. Soloway, X. Song, C. Sougnez, C. Stewart, P. Stojanov, J. M. Stuart, S. O. Sumer,

Y. Sun, B. Tabak, A. Tam, D. Tan, J. Tang, R. Tarnuzzer, B. S. Taylor, N. Thiessen, L. Thorne,

V. Thorsson, R. M. Tuttle, C. B. Umbricht, D. J. Van Den Berg, F. Vandin, U. Veluvolu, R. G.

W. Verhaak, M. Vinco, D. Voet, V. Walter, Z. Wang, S. Waring, P. M. Weinberger, N. Weinhold, J. N. Weinstein, D. J. Weisenberger, D. Wheeler, M. D. Wilkerson, J. Wilson, M. Williams, D. A. Winer, L. Wise, J. Wu, L. Xi, A. W. Xu, L. Yang, L. Yang, T. I. Zack, M. A. Zeiger, D. Zeng, J. C. Zenklusen, N. Zhao, H. Zhang, J. Zhang, J. Zhang, W. Zhang, E. Zmuda, L. Zou, Integrated Genomic Characterization of Papillary Thyroid Carcinoma, *Cell* **159**, 676–690 (2014).

29. M. Snuderl, L. Fazlollahi, L. P. Le, M. Nitta, B. H. Zhelyazkova, C. J. Davidson, S. Akhavanfard, D. P. Cahill, K. D. Aldape, R. A. Betensky, D. N. Louis, A. J. Iafrate, Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma, *Cancer Cell* **20**, 810–817 (2011).

30. E. Gherardi, W. Birchmeier, C. Birchmeier, G. Vande Woude, Targeting MET in cancer: Rationale and progress*Nat. Rev. Cancer* **12**, 89–103 (2012).

31. S. Bender, J. Gronych, H. J. Warnatz, B. Hutter, S. Gröbner, M. Ryzhova, E. Pfaff, V. Hovestadt, F. Weinberg, S. Halbach, M. Kool, P. A. Northcott, D. Sturm, L. Bjerke, T. Zichner, A. M. Stütz, K. Schramm, B. Huang, I. Buchhalter, M. Heinold, T. Risch, B. C. Worst, C. M. Van Tilburg, U. D. Weber, M. Zapatka, B. Raeder, D. Milford, S. Heiland, C. Von Kalle, C. Previti, C. Lawerenz, A. E. Kulozik, A. Unterberg, O. Witt, A. Von Deimling, D. Capper, N. Truffaux, J. Grill, N. Jabado, A. M. Sehested, D. Sumerauer, D. H. Ben Brahim, S. Trabelsi, H. K. Ng, D. Zagzag, J. C. Allen, M. A. Karajannis, N. G. Gottardo, C. Jones, J. O. Korbel, S. Schmidt, S. Wolf, G. Reifenberger, J. Felsberg, B. Brors, C. Herold-Mende, H. Lehrach, T. Brummer, A. Korshunov, R. Eils, M. L. Yaspo, S. M. Pfister, P. Lichter, D. T. W. Jones, Recurrent MET fusion genes represent a drug target in pediatric glioblastoma, *Nat. Med.* **22**, 1314–1320 (2016).

32. A. I. Chapoval, G. Smithson, L. Brunick, M. Mesri, F. L. Boldog, D. Andrew, N. V. Khramtsov, E. A. Feshchenko, G. C. Starling, P. S. Mezes, BTNL8, a butyrophilin-like molecule that costimulates the primary immune response, *Mol. Immunol.* **56**, 819–828 (2013).

33. R. S. Harris, M. T. Liddament, Retroviral restriction by APOBEC proteins*Nat. Rev. Immunol.* **4**, 868–877 (2004).

34. N. Spies, Z. Weng, A. Bishara, J. McDaniel, D. Catoe, J. M. Zook, M. Salit, R. B. West, S. Batzoglou, A. Sidow, Genome-wide reconstruction of complex structural variants using read clouds, *Nat. Methods* (2017), doi:10.1038/nmeth.4366.

35. X. Zhou, M. N. Edmonson, M. R. Wilkinson, A. Patel, G. Wu, Y. Liu, Y. Li, Z. Zhang, M. C. Rusch, M. Parker, J. Becksfort, J. R. Downing, J. Zhang, Exploring genomic alteration in pediatric cancer using ProteinPaint*Nat. Genet.* (2015), doi:10.1038/ng.3466.

36. S. N. Gröbner, B. C. Worst, J. Weischenfeldt, I. Buchhalter, K. Kleinheinz, V. A. Rudneva, P. D. Johann, G. P. Balasubramanian, M. Segura-Wang, S. Brabetz, S. Bender, B. Hutter, D. Sturm, E. Pfaff, D. Hübschmann, G. Zipprich, M. Heinold, J. Eils, C. Lawerenz, S. Erkek, S. Lambo, S. Waszak, C. Blattmann, A. Borkhardt, M. Kuhlen, A. Eggert, S. Fulda, M. Gessler, J. Wegert, R. Kappler, D. Baumhoer, S. Burdach, R. Kirschner-Schwabe, U. Kontny, A. E. Kulozik, D. Lohmann, S. Hettmer, C. Eckert, S. Bielack, M. Nathrath, C. Niemeyer, G. H. Richter, J. Schulte, R. Siebert, F. Westermann, J. J. Molenaar, G. Vassal, H. Witt, M. Zapatka, B. Burkhardt, C. P. Kratz, O. Witt, C. M. V. Tilburg, C. M. Kramm, G. Fleischhack, U. Dirksen, S. Rutkowski, M. Frühwald, K. Von Hoff, S. Wolf, T. Klingebiel, E. Koscielniak, P. Landgraf, J. Koster, A. C. Resnick, J. Zhang, Y. Liu, X. Zhou, A. J. Waanders, D. A. Zwijnenburg, P. Raman, B. Brors, U. D. Weber, P. A. Northcott, K. W. Pajtler, M. Kool, R. M. Piro, J. O. Korbel, M. Schlesner, R. Eils, D. T. W. Jones, P. Lichter, L. Chavez, S. M. Pfister, The landscape of

genomic alterations across childhood cancers, *Nature* **555**, 321–327 (2018).

37. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, M. Moser, E. Karasik, B. Gillard, K. Ramsey, S. Sullivan, J. Bridge, H. Magazine, J. Syron, J. Fleming, L. Siminoff, H. Traino, M. Mosavel, L. Barker, S. Jewell, D. Rohrer, D. Maxim, D. Filkins, P. Harbach, E. Cortadillo, B. Berghuis, L. Turner, E. Hudson, K. Feenstra, L. Sobin, J. Robb, P. Branton, G. Korzeniewski, C. Shive, D. Tabor, L. Qi, K. Groch, S. Nampally, S. Buia, A. Zimmerman, A. Smith, R. Burges, K. Robinson, K. Valentino, D. Bradbury, M. Cosentino, N. Diaz-Mayoral, M. Kennedy, T. Engel, P. Williams, K. Erickson, K. Ardlie, W. Winckler, G. Getz, D. DeLuca, Daniel MacArthur, M. Kellis, A. Thomson, T. Young, E. Gelfand, M. Donovan, Y. Meng, G. Grant, D. Mash, Y. Marcus, M. Basile, J. Liu, J. Zhu, Z. Tu, N. J. Cox, D. L. Nicolae, E. R. Gamazon, H. K. Im, A. Konkashbaev, J. Pritchard, M. Stevens, T. Flutre, X. Wen, E. T. Dermitzakis, T. Lappalainen, R. Guigo, J. Monlong, M. Sammeth, D. Koller, A. Battle, S. Mostafavi, M. McCarthy, M. Rivas, J. Maller, I. Rusyn, A. Nobel, F. Wright, A. Shabalin, M. Feolo, N. Sharopova, A. Sturcke, J. Paschal, J. M. Anderson, E. L. Wilder, L. K. Derr, E. D. Green, J. P. Struewing, G. Temple, S. Volpi, J. T. Boyer, E. J. Thomson, M. S. Guyer, C. Ng, A. Abdallah, D. Colantuoni, T. R. Insel, S. E. Koester, A Roger Little, P. K. Bender, T. Lehner, Y. Yao, C. C. Compton, J. B. Vaught, S. Sawyer, N. C. Lockhart, J. Demchok, H. F. Moore, The Genotype-Tissue Expression (GTEx) project*Nat. Genet.* **45**, 580–585 (2013).

38. C. J. Willer, E. K. Speliotes, R. J. F. Loos, S. Li, C. M. Lindgren, I. M. Heid, S. I. Berndt, A. L. Elliott, A. U. Jackson, C. Lamina, G. Lettre, N. Lim, H. N. Lyon, S. A. McCarroll, K. Papadakis, L. Qi, J. C. Randall, R. M. Roccasecca, S. Sanna, P. Scheet, M. N. Weedon, E. Wheeler, J. H. Zhao, L. C. Jacobs, I. Prokopenko, N. Soranzo, T. Tanaka, N. J. Timpson, P.

Almgren, A. Bennett, R. N. Bergman, S. A. Bingham, L. L. Bonnycastle, M. Brown, N. P. Burtt, P. Chines, L. Coin, F. S. Collins, J. M. Connell, C. Cooper, G. D. Smith, E. M. Dennison, P. Deodhar, P. Elliott, M. R. Erdos, K. Estrada, D. M. Evans, L. Gianniny, C. Gieger, C. J. Gillson, C. Guiducci, R. Hackett, D. Hadley, A. S. Hall, A. S. Havulinna, J. Hebebrand, A. Hofman, B. Isomaa, K. B. Jacobs, T. Johnson, P. Jousilahti, Z. Jovanovic, K.-T. Khaw, P. Kraft, M. Kuokkanen, J. Kuusisto, J. Laitinen, E. G. Lakatta, J. Luan, R. N. Luben, M. Mangino, W. L. McArdle, T. Meitinger, A. Mulas, P. B. Munroe, N. Narisu, A. R. Ness, K. Northstone, S. O'Rahilly, C. Purmann, M. G. Rees, M. Ridderstråle, S. M. Ring, F. Rivadeneira, A. Ruokonen, M. S. Sandhu, J. Saramies, L. J. Scott, A. Scuteri, K. Silander, M. A. Sims, K. Song, J. Stephens, S. Stevens, H. M. Stringham, Y. C. L. Tung, T. T. Valle, C. M. Van Duijn, K. S. Vimaleswaran, P. Vollenweider, G. Waeber, C. Wallace, R. M. Watanabe, D. M. Waterworth, N. Watkins, J. C. M. Witteman, E. Zeggini, G. Zhai, M. C. Zillikens, D. Altshuler, M. J. Caulfield, S. J. Chanock, I. S. Farooqi, L. Ferrucci, J. M. Guralnik, A. T. Hattersley, F. B. Hu, M.-R. Jarvelin, M. Laakso, V. Mooser, K. K. Ong, W. H. Ouwehand, V. Salomaa, N. J. Samani, T. D. Spector, T. Tuomi, J. Tuomilehto, M. Uda, A. G. Uitterlinden, N. J. Wareham, P. Deloukas, T. M. Frayling, L. C. Groop, R. B. Hayes, D. J. Hunter, K. L. Mohlke, L. Peltonen, D. Schlessinger, D. P. Strachan, H.-E. Wichmann, M. I. McCarthy, M. Boehnke, I. Barroso, G. R. Abecasis, J. N. Hirschhorn, Six new loci associated with body mass index highlight a neuronal influence on body weight regulation, *Nat. Genet.* **41**, 25–34 (2009).

39. A. M. Veerappa, M. Saldanha, P. Padakannaya, N. B. Ramachandra, Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity, *J. Hum. Genet.* **58**, 539–547 (2013).

40. K. Singh, D. Loreth, B. Pöttker, K. Hefti, J. Innos, K. Schwald, H. Hengstler, L. Menzel, C.

J. Sommer, K. Radyushkin, O. Kretz, M.-A. Philips, C. A. Haas, K. Frauenknecht, K. Lilleväli, B. Heimrich, E. Vasar, M. K. E. Schäfer, Neuronal growth and behavioral alterations in mice deficient for the psychiatric disease-associated negr1 gene, *Front. Mol. Neurosci.* **11**, 1–14 (2018).

41. H. Kim, J.-S. Hwang, B. Lee, J. Hong, S. Lee, Newly Identified Cancer-Associated Role of Human Neuronal Growth Regulator 1 (NEGR1), *J. Cancer* **5**, 598–608 (2014).

42. B. S. Paugh, C. Qu, C. Jones, Z. Liu, M. Adamowicz-Brice, J. Zhang, D. A. Bax, B. Coyle, J. Barrow, D. Hargrave, J. Lowe, A. Gajjar, W. Zhao, A. Broniscer, D. W. Ellison, R. G. Grundy, S. J. Baker, Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease, *J. Clin. Oncol.* **28**, 3061–3068 (2010).

43. X. Lan, D. J. Jörg, F. M. G. Cavalli, L. M. Richards, L. V. Nguyen, R. J. Vanner, P. Guilhamon, L. Lee, M. M. Kushida, D. Pellacani, N. I. Park, F. J. Coutinho, H. Whetstone, H. J. Selvadurai, C. Che, B. Luu, A. Carles, M. Moksa, N. Rastegar, R. Head, S. Dolma, P. Prinos, M. D. Cusimano, S. Das, M. Bernstein, C. H. Arrowsmith, A. J. Mungall, R. A. Moore, Y. Ma, M. Gallo, M. Lupien, T. J. Pugh, M. D. Taylor, M. Hirst, C. J. Eaves, B. D. Simons, P. B. Dirks, Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy, *Nature* **549**, 227–232 (2017).

44. R. E. Amir, I. . B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, H. Y. Zoghbi, Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2., *Nat. Genet.* **23**, 185–188 (1999).

45. K. R. Taylor, A. Mackay, N. Truffaux, Y. S. Butterfield, O. Morozova, C. Philippe, D. Castel, C. S. Grasso, M. Vinci, D. Carvalho, A. M. Carcaboso, C. De Torres, O. Cruz, J. Mora,

N. Entz-Werle, W. J. Ingram, M. Monje, D. Hargrave, A. N. Bullock, S. Puget, S. Yip, C. Jones, J. Grill, Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma, *Nat. Genet.* **46**, 457–461 (2014).

46. P. Buczkowicz, C. Hoeman, P. Rakopoulos, S. Pajovic, L. Letourneau, M. Dzamba, A. Morrison, P. Lewis, E. Bouffet, U. Bartels, J. Zuccaro, S. Agnihotri, S. Ryall, M. Barszczyk, Y. Chornenkyy, M. Bourgey, G. Bourque, A. Montpetit, F. Cordero, P. Castelo-Branco, J. Mangerel, U. Tabori, K. C. Ho, A. Huang, K. R. Taylor, A. Mackay, A. E. Bendel, J. Nazarian, J. R. Fangusaro, M. A. Karajannis, D. Zagzag, N. K. Foreman, A. Donson, J. V. Hegert, A. Smith, J. Chan, L. Lafay-Cousin, S. Dunn, J. Hukin, C. Dunham, K. Scheinemann, J. Michaud, S. Zelcer, D. Ramsay, J. Cain, C. Brennan, M. M. Souweidane, C. Jones, C. D. Allis, M. Brudno, O. Becher, C. Hawkins, Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations, *Nat. Genet.* **46**, 451–456 (2014).

47. A. M. Fontebasso, S. Papillon-Cavanagh, J. Schwartzentruber, H. Nikbakht, N. Gerges, P. O. Fiset, D. Bechet, D. Faury, N. De Jay, L. A. Ramkissoon, A. Corcoran, D. T. W. Jones, D. Sturm, P. Johann, T. Tomita, S. Goldman, M. Nagib, A. Bendel, L. Goumnerova, D. C. Bowers, J. R. Leonard, J. B. Rubin, T. Alden, S. Browd, J. R. Geyer, S. Leary, G. Jallo, K. Cohen, N. Gupta, M. D. Prados, A. S. Carret, B. Ellezam, L. Crevier, A. Klekner, L. Bognar, P. Hauser, M. Garami, J. Myseros, Z. Dong, P. M. Siegel, H. Malkin, A. H. Ligon, S. Albrecht, S. M. Pfister, K. L. Ligon, J. Majewski, N. Jabado, M. W. Kieran, Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma, *Nat. Genet.* **46**, 462–466 (2014).

48. E. M. Shore, M. Xu, G. J. Feldman, D. A. Fenstermacher, M. A. Brown, F. S. Kaplan, A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic

fibrodysplasia ossificans progressiva, *Nat. Genet.* **38**, 525–527 (2006).

49. R. Rahbari, A. Wuster, S. J. Lindsay, R. J. Hardwick, L. B. Alexandrov, S. Al Turki, A. Dominiczak, A. Morris, D. Porteous, B. Smith, M. R. Stratton, M. E. Hurles, Timing, rates and spectra of human germline mutation, *Nat. Genet.* **48**, 126–133 (2016).

50. K. B. Jacobs, M. Yeager, W. Zhou, S. Wacholder, Z. Wang, B. Rodriguez-Santiago, A. Hutchinson, X. Deng, C. Liu, M. J. Horner, M. Cullen, C. G. Epstein, L. Burdett, M. C. Dean, N. Chatterjee, J. Sampson, C. C. Chung, J. Kovaks, S. M. Gapstur, V. L. Stevens, L. T. Teras, M. M. Gaudet, D. Albanes, S. J. Weinstein, J. Virtamo, P. R. Taylor, N. D. Freedman, C. C. Abnet, A. M. Goldstein, N. Hu, K. Yu, J. M. Yuan, L. Liao, T. Ding, Y. L. Qiao, Y. T. Gao, W. P. Koh, Y. B. Xiang, Z. Z. Tang, J. H. Fan, M. C. Aldrich, C. Amos, W. J. Blot, C. H. Bock, E. M. Gillanders, C. C. Harris, C. A. Haiman, B. E. Henderson, L. N. Kolonel, L. Le Marchand, L. H. McNeill, B. A. Rybicki, A. G. Schwartz, L. B. Signorello, M. R. Spitz, J. K. Wiencke, M. Wrensch, X. Wu, K. A. Zanetti, R. G. Ziegler, J. D. Figueroa, M. Garcia-Closas, N. Malats, G. Marenne, L. Prokunina-Olsson, D. Baris, M. Schwenn, A. Johnson, M. T. Landi, L. Goldin, D. Consonni, P. A. Bertazzi, M. Rotunno, P. Rajaraman, U. Andersson, L. E. B. Freeman, C. D. Berg, J. E. Buring, M. A. Butler, T. Carreon, M. Feychting, A. Ahlbom, J. M. Gaziano, G. G. Giles, G. Hallmans, S. E. Hankinson, P. Hartge, R. Henriksson, P. D. Inskip, C. Johansen, A. Landgren, R. McKean-Cowdin, D. S. Michaud, B. S. Melin, U. Peters, A. M. Ruder, H. D. Sesso, G. Severi, X. O. Shu, K. Visvanathan, E. White, A. Wolk, A. Zeleniuch-Jacquotte, W. Zheng, D. T. Silverman, M. Kogevinas, J. R. Gonzalez, O. Villa, D. Li, E. J. Duell, H. A. Risch, S. H. Olson, C. Kooperberg, B. M. Wolpin, L. Jiao, M. Hassan, W. Wheeler, A. A. Arslan, H. B. Bueno-De-Mesquita, C. S. Fuchs, S. Gallinger, M. D. Gross, E. A. Holly, A. P. Klein, A. Lacroix, M. T. Mandelson, G. Petersen, M. C. Boutron-Ruault, P. M. Bracci, F. Canzian, K.

Chang, M. Cotterchio, E. L. Giovannucci, M. Goggins, J. A. H. Bolton, M. Jenab, K. T. Khaw, V. Krogh, R. C. Kurtz, R. R. McWilliams, J. B. Mendelsohn, K. G. Rabe, E. Riboli, A. Tjønneland, G. S. Tobias, D. Trichopoulos, J. W. Elena, H. Yu, L. Amundadottir, R. Z. Stolzenberg-Solomon, P. Kraft, F. Schumacher, D. Stram, S. A. Savage, L. Mirabello, I. L. Andrulis, J. S. Wunder, A. P. García, L. SierrasesÃmaga, D. A. Barkauskas, R. G. Gorlick, M. Purdue, W. H. Chow, L. E. Moore, K. L. Schwartz, F. G. Davis, A. W. Hsing, S. I. Berndt, A. Black, N. Wentzensen, L. A. Brinton, J. Lissowska, B. Peplonska, K. A. McGlynn, M. B. Cook, B. I. Graubard, C. P. Kratz, M. H. Greene, R. L. Erickson, D. J. Hunter, G. Thomas, R. N. Hoover, F. X. Real, J. F. Fraumeni, N. E. Caporaso, M. Tucker, N. Rothman, L. A. Pérez-Jurado, S. J. Chanock, Detectable clonal mosaicism and its relationship to aging and cancer, *Nat. Genet.* **44**, 651–658 (2012).

51. C. C. Laurie, C. A. Laurie, K. Rice, K. F. Doheny, L. R. Zelnick, C. P. McHugh, H. Ling, K. N. Hetrick, E. W. Pugh, C. Amos, Q. Wei, L. E. Wang, J. E. Lee, K. C. Barnes, N. N. Hansel, R. Mathias, D. Daley, T. H. Beaty, A. F. Scott, I. Ruczinski, R. B. Scharpf, L. J. Bierut, S. M. Hartz, M. T. Landi, N. D. Freedman, L. R. Goldin, D. Ginsburg, J. Li, K. C. Desch, S. S. Strom, W. J. Blot, L. B. Signorello, S. A. Ingles, S. J. Chanock, S. I. Berndt, L. Le Marchand, B. E. Henderson, K. R. Monroe, J. A. Heit, M. De Andrade, S. M. Armasu, C. Regnier, W. L. Lowe, M. G. Hayes, M. L. Marazita, E. Feingold, J. C. Murray, M. Melbye, B. Feenstra, J. H. Kang, J. L. Wiggs, G. P. Jarvik, A. N. McDavid, V. E. Seshan, D. B. Mirel, A. Crenshaw, N. Sharopova, A. Wise, J. Shen, D. R. Crosslin, D. M. Levine, X. Zheng, J. I. Udren, S. Bennett, S. C. Nelson, S. M. Gogarten, M. P. Conomos, P. Heagerty, T. Manolio, L. R. Pasquale, C. A. Haiman, N. Caporaso, B. S. Weir, Detectable clonal mosaicism from birth to old age and its relationship to cancer, *Nat. Genet.* **44**, 642–650 (2012).

52. C. C. Coombs, A. Zehir, S. M. Devlin, A. Kishtagari, A. Syed, P. Jonsson, D. M. Hyman, D. B. Solit, M. E. Robson, J. Baselga, M. E. Arcila, M. Ladanyi, M. S. Tallman, R. L. Levine, M. F. Berger, Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes, *Cell Stem Cell* **21**, 374–382.e4 (2017).

53. B. J. O ' Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith, E. H. Turner, I. B. Stanaway, B. Vernot, M. Malig, C. Baker, J. M. Akey, E. Borenstein, M. J. Rieder, D. A. Nickerson, R. Bernier, J. Shendure, E. E. Eichler, Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations, *Nature* **485**, 246–250 (2012).

54. B. Xu, J. L. Roos, P. Dexheimer, B. Boone, B. Plummer, S. Levy, J. A. Gogos, M. Karayiorgou, Exome sequencing supports a de novo mutational paradigm for schizophrenia, *Nat. Genet.* **43**, 864–868 (2011).

55. C. Danussi, P. Bose, P. T. Parthasarathy, P. C. Silberman, J. S. Van Arnam, M. Vitucci, O. Y. Tang, A. Heguy, Y. Wang, T. A. Chan, G. J. Riggins, E. P. Sulman, F. Lang, C. J. Creighton, B. Deneen, C. R. Miller, D. J. Picketts, K. Kannan, J. T. Huse, Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling, *Nat. Commun.* **9**, 1–15 (2018).

56. J. Aigner, S. Villatoro, R. Rabionet, J. Roquer, J. Jiménez-Conde, E. Martí, X. Estivill, A common 56-kilobase deletion in a primate-specific segmental duplication creates a novel butyrophilin-like protein, *BMC Genet.* **14** (2013), doi:10.1186/1471-2156-14-61.

57. K. Cibulskis, M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel,

M. Meyerson, E. S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, *Nat. Biotechnol.* **31**, 213–219 (2013).

58. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data, *Nucleic Acids Res.* **38** (2010), doi:10.1093/nar/gkq603.

59. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, G. P. D. P. Subgroup, The Sequence Alignment, *Bioinformatics* **25**, 2078–2079 (2009).

60. V. E. Seshan, A. B. Olshen, DNAcopy : A Package for Analyzing DNA Copy Data, *Bioconductor Vignette* , 1–7 (2014).

61. L. Scrucca, M. Fop, T. B. Murphy, A. E. Raftery, mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models, *R J.* **8**, 289–317 (2016).

62. P. J. Stephens, P. S. Tarpey, H. Davies, P. Van Loo, C. Greenman, D. C. Wedge, S. Nik-Zainal, S. Martin, I. Varela, G. R. Bignell, L. R. Yates, E. Papaemmanuil, D. Beare, A. Butler, A. Cheverton, J. Gamble, J. Hinton, M. Jia, A. Jayakumar, D. Jones, C. Latimer, K. W. Lau, S. McLaren, D. J. McBride, A. Menzies, L. Mudie, K. Raine, R. Rad, M. S. Chapman, J. Teague, D. Easton, A. Langerød, R. Karesen, E. Schlichting, B. Naume, T. Sauer, L. Ottestad, M. T. M. Lee, C. Y. Shen, B. T. K. Tee, B. W. Huimin, A. Broeks, A. C. Vargas, G. Turashvili, J. Martens, A. Fatima, P. Miron, S. F. Chin, G. Thomas, S. Boyault, O. Mariani, S. R. Lakhani, M. Van De Vijver, L. Van't Veer, J. Foekens, C. Desmedt, C. Sotiriou, A. Tutt, C. Caldas, J. S. Reis-Filho, S. A. J. R. Aparicio, A. V. Salomon, A. L. Børresen-Dale, A. L. Richardson, P. J. Campbell, P. A. Futreal, M. R. Stratton, The landscape of cancer genes and mutational processes in breast cancer, *Nature* **486**, 400–404 (2012).

63. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R

language, *Bioinformatics* (2004), doi:10.1093/bioinformatics/btg412.

64. Z. L. Skidmore, A. H. Wagner, R. Lesurf, K. M. Campbell, J. Kunisaki, O. L. Griffith, M. Griffith, GenVisR: Genomic Visualizations in R, *Bioinformatics* **32**, 3012–3014 (2016).

65. R. Rosenthal, N. McGranahan, J. Herrero, B. S. Taylor, C. Swanton, deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution, *Genome Biol.* **17** (2016), doi:10.1186/s13059-016-0893-4.

66. S. A. Forbes, D. Beare, H. Boutselakis, S. Bamford, N. Bindal, J. Tate, C. G. Cole, S. Ward, E. Dawson, L. Ponting, R. Stefancsik, B. Harsha, C. YinKok, M. Jia, H. Jubb, Z. Sondka, S. Thompson, T. De, P. J. Campbell, COSMIC: Somatic cancer genetics at high-resolution, *Nucleic Acids Res.* **45**, D777–D783 (2017).

67. B. D. Simons, Deep sequencing as a probe of normal stem cell fate and preneoplasia in human epidermis, *Proc. Natl. Acad. Sci.* **113**, 128–133 (2016).

68. M. J. Williams, B. Werner, C. P. Barnes, T. A. Graham, A. Sottoriva, Identification of neutral tumor evolution across cancer types, *Nat. Genet.* **48**, 238–244 (2016).

69. H. T. Shin, Y. La Choi, J. W. Yun, N. K. D. Kim, S. Y. Kim, H. J. Jeon, J. Y. Nam, C. Lee, D. Ryu, S. C. Kim, K. Park, E. Lee, J. S. Bae, D. S. Son, J. G. Joung, J. Lee, S. T. Kim, M. J. Ahn, S. H. Lee, J. S. Ahn, W. Y. Lee, B. Y. Oh, Y. H. Park, J. E. Lee, K. H. Lee, H. C. Kim, K. M. Kim, Y. H. Im, K. Park, P. J. Park, W. Y. Park, Prevalence and detection of low-allele-fraction variants in clinical cancer samples, *Nat. Commun.* **8** (2017), doi:10.1038/s41467-017-01470-y.

70. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, *Nat. Biotechnol.* **36**

(2018), doi:10.1038/nbt.4096.

71. 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).

72. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* **15** (2014), doi:10.1186/s13059-014-0550-8.

73. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc.* **4**, 44–57 (2009).

74. D. W. Huang, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res.* **37**, 1–13 (2009).

75. J. R. MacDonald, R. Ziman, R. K. C. Yuen, L. Feuk, S. W. Scherer, The Database of Genomic Variants: A curated collection of structural variation in the human genome, *Nucleic Acids Res.* **42** (2014), doi:10.1093/nar/gkt958.

Acknowledgements

Authors contribution

MG and MH conceived the experiments; MH, JK, KE and NHD performed the wet lab experiments; MH, AHG, DJK, MJ, AN, KN, MG and SM analyzed the WGS and scRNA-seq data; FMGC, MMK, FJC, BL, YM, AJM, RM, MAM, MDT, GB, TJP and PBD generated the bulk RNA-seq data; JAC, MMK, FJC and PBD provided samples; MG, SM, JAC, BDS and DS provided supervision to the work described hereby; MG, MH and JAC wrote the original manuscript; all authors provided editorial and experimental feedback.

Competing interests

No competing interests.
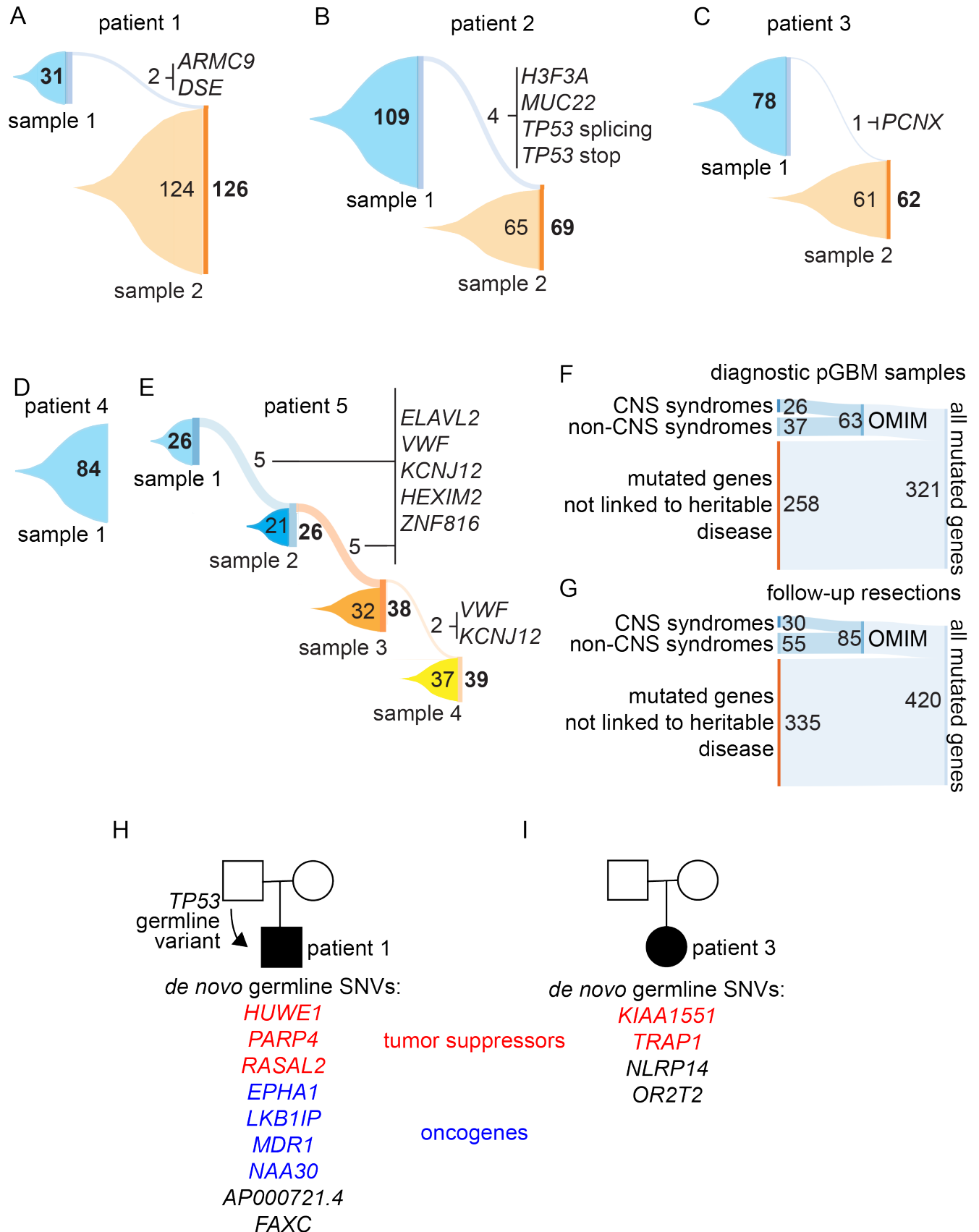
Fig. 1. Single nucleotide variants in pediatric GBM samples.

Sankey plots for patients 1 (A), 2 (B), 3 (C), 4 (D) and 5 (E). Each plot illustrates changes in the mutational landscapes between primary tumors and recurrences. Bold numbers represent total number of mutations in each profiled tumor. Gene lists represent the mutated genes that are shared between consecutive tumors.

(F) Analysis of all exonic, non-synonymous single nucleotide variants (SNVs) in all primary tumors reveals a fraction of genes are associated with hereditary syndromes, according to the OMIM database. Out of the 63 genes with links to hereditary disease, 26 were specifically linked to central nervous system (CNS) syndromes.

(G) Analysis of all exonic, non-synonymous SNVs in all recurrent tumors. Out of 85 genes with links to hereditary disease, 30 were specifically linked to CNS syndromes.

(H) Germlines were sequenced for patient 1, and the patient's mother and father. A deletion of the last exon of *TP53* was detected in the father, and it was inherited by patient 1. Non-synonymous SNVs in 9 other genes were identified in patient's 1 germline, but not in the parents.

(I) *De novo* non-synonymous SNVs in four genes were identified in patient's 3 germline.
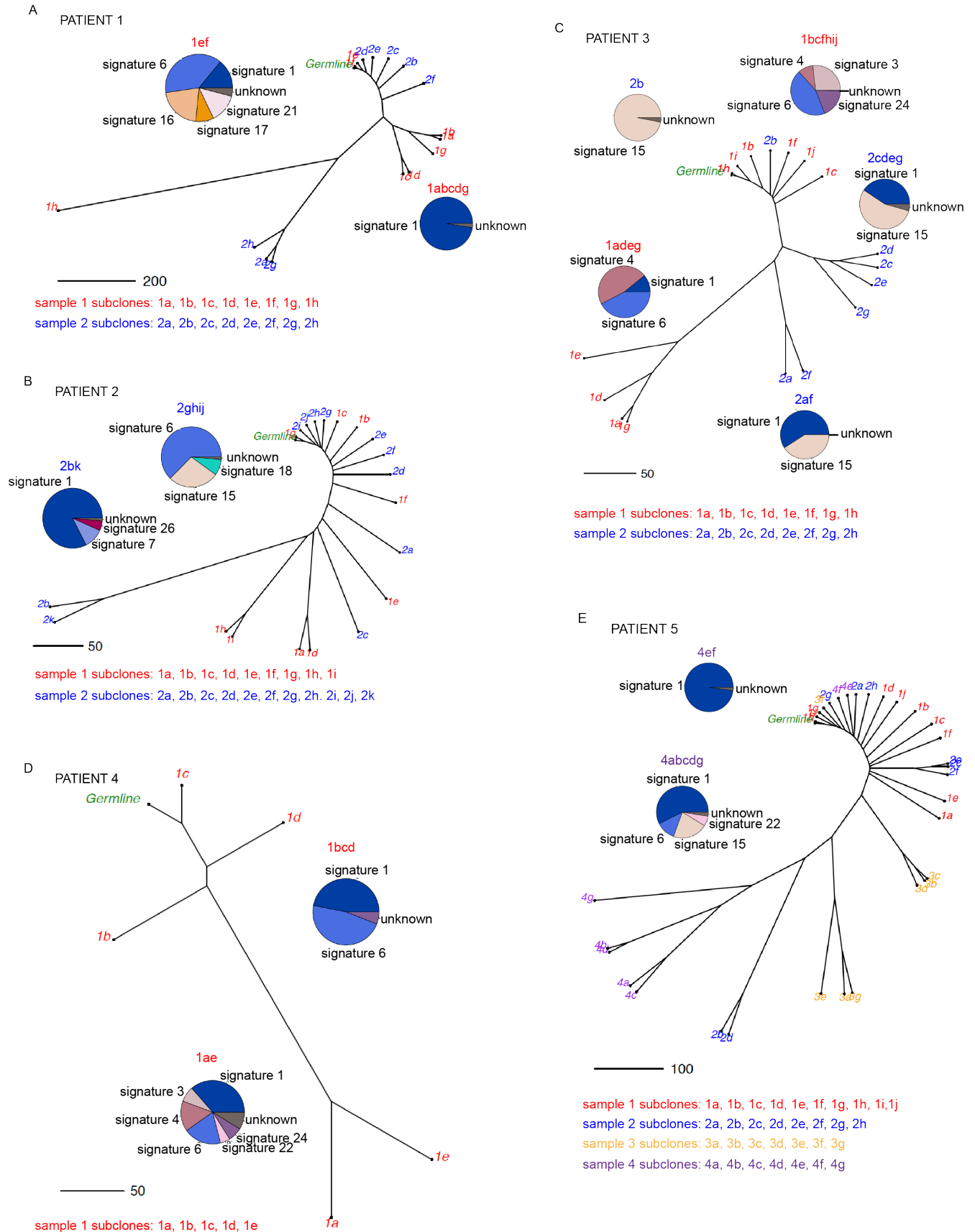
A — PATIENT 1

sample 1 subclones: 1a, 1b, 1c, 1d, 1e, 1f, 1g, 1h
sample 2 subclones: 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h

B — PATIENT 2

sample 1 subclones: 1a, 1b, 1c, 1d, 1e, 1f, 1g, 1h, 1i
sample 2 subclones: 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h. 2i, 2j, 2k

C — PATIENT 3

sample 1 subclones: 1a, 1b, 1c, 1d, 1e, 1f, 1g, 1h
sample 2 subclones: 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h

D — PATIENT 4

sample 1 subclones: 1a, 1b, 1c, 1d, 1e

E — PATIENT 5

sample 1 subclones: 1a, 1b, 1c, 1d, 1e, 1f, 1g, 1h, 1i,1j
sample 2 subclones: 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h
sample 3 subclones: 3a, 3b, 3c, 3d, 3e, 3f, 3g
sample 4 subclones: 4a, 4b, 4c, 4d, 4e, 4f, 4g

58

Fig 2. Subclonal architecture and evolution in pediatric GBM.

Phylogenetic trees reflect intermixing of genetic lineages in each sample from patients 1 (A), 2 (B), 3 (C), 4 (D), and 5 (E). The subclonal structure of each tumor sample was inferred using somatic mutations and copy number alterations. Distance between nodes is proportional to the number of genetic differences between subclones. Branch tips are color-coded according to sample number, and labeled according to biopsy number (1, 2, 3) and clonal lineage (a, b, c). For selected clusters of genetic subclones, pie charts display the contribution of mutational signatures detectable using the somatic mutations specific to those phylogenetic branches.
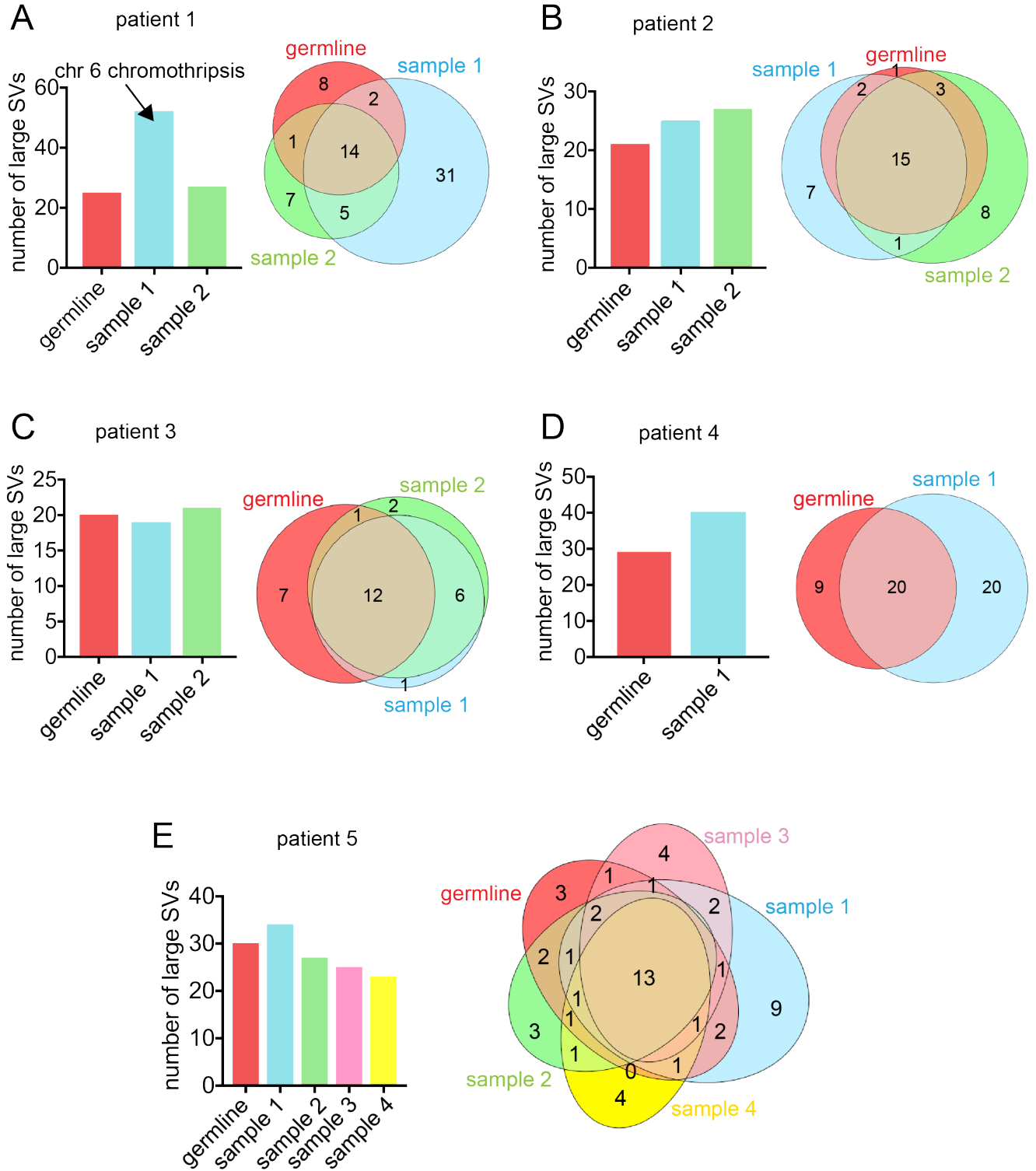
Fig. 3. Number of large structural variants (SVs) in the germlines and pGBM tumor tissue.

Total number of large SVs in the germline, primary and recurrent tumors for patients 1 (A), 2 (B),

3 (C), 4 (D), and 5 (E). SV calls were performed with the LongRanger pipeline.
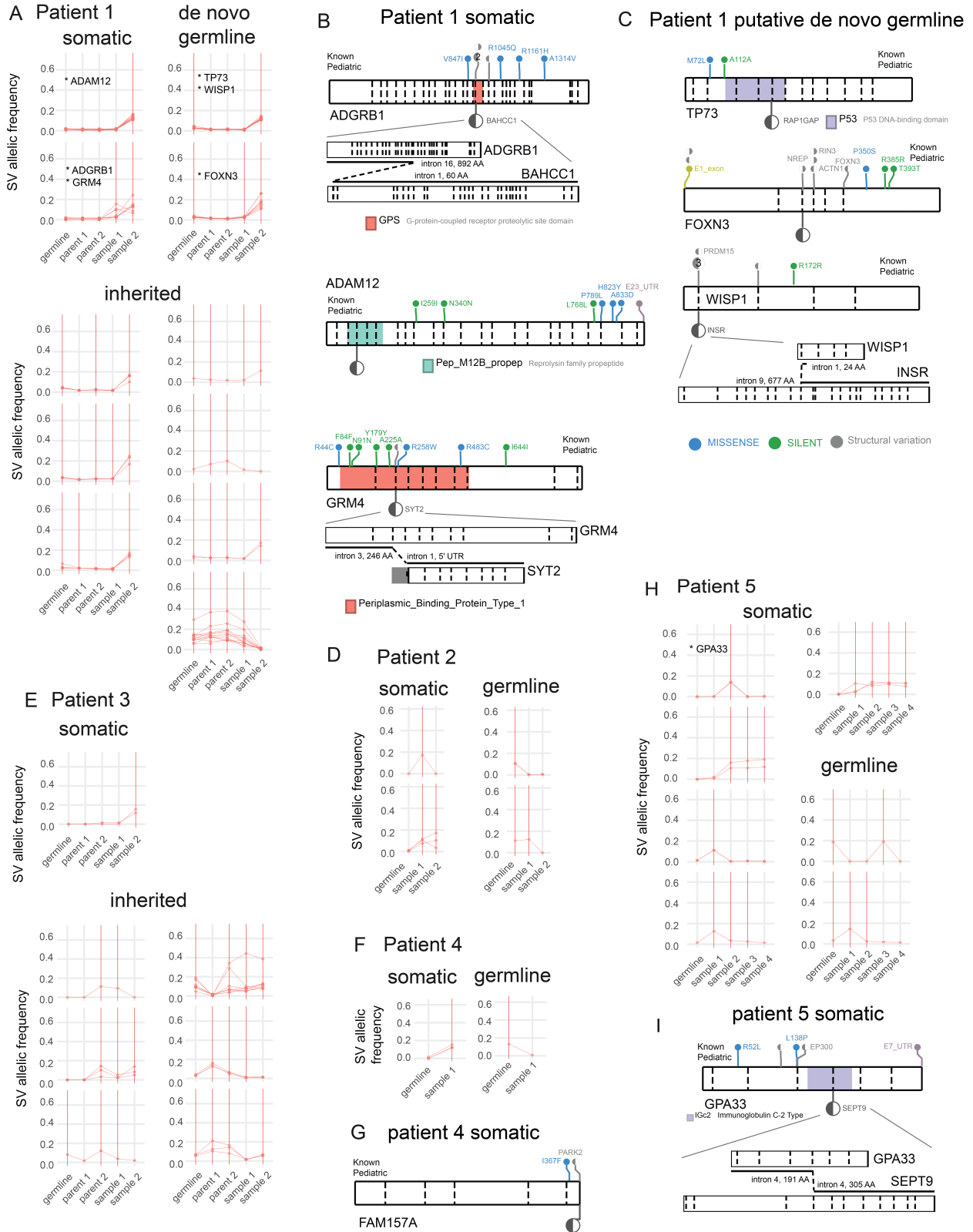
Fig. 4. Dynamic contributions of structural variants to tumor genomes.

Allelic frequencies for structural variants called by GROC-SVs are plotted across samples for all patients (patient 1 (A), patient 2 (D), patient 3 (E), patient 4 (F), and patient 5 (H)). Red lines connect dots that represent the allelic frequency of 10x barcodes supporting a particular SV, such that higher values represent a higher level of support for the SV. Only SVs with a minimum of 200 barcodes in each patient sample are shown. Red vertical lines highlight samples for which an SV was confidently called in the corresponding sample. Inset gene names represent genes shown in panels B, C, G, and I. Samples include the germline for primary tumors and their recurrences, and in germlines of the parents of patient 1 and patient 3. In these patients, germline events are stratified into inherited vs *de novo* events, and those observed in both parental germlines were excluded unless they were absent in another sample. Panels (B), (C), (G), and (I) contain gene level diagrams of select SVs that occur in genes previously associated with pediatric cancers. Gene diagrams include exon boundaries (dashed horizontal lines), protein domains (coloured boxes, only including domains overlapping the SV breakpoint), previously annotated somatic alterations in pediatric cancers (lollipops above the gene diagram), and the breakpoint observed in a particular patient (larger lollipop below the gene diagram). For SV events spanning genes ADGRB1-BAHCC1, GRM4-SYT2, and GPA33-SEPT9, the predicted effect of the structural variant on the structure of the genes is shown below the main gene model.
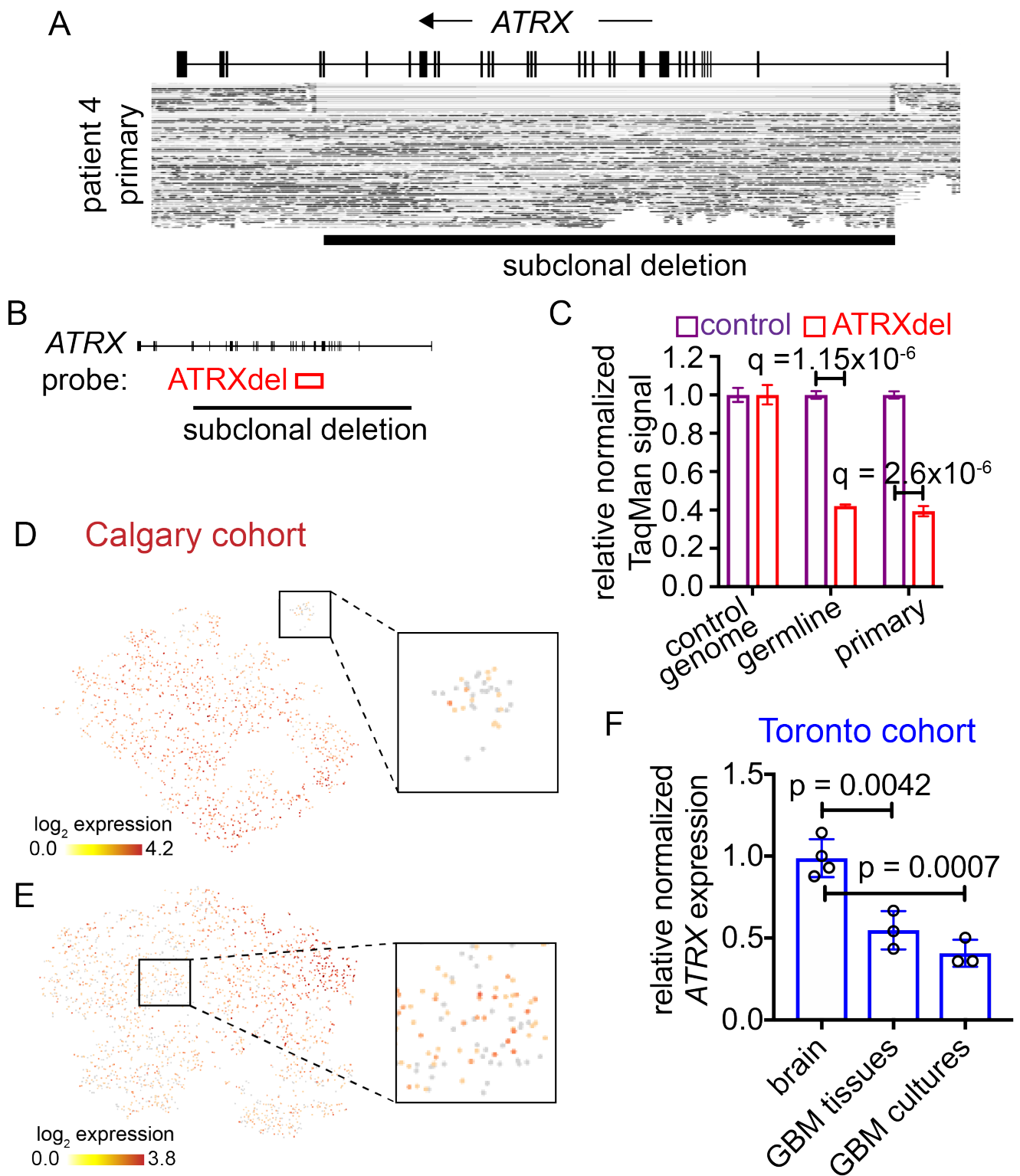
Fig. 5. Subclonal *ATRX* deletions in the germline and tumors of pediatric GBM patients.

(A) Putative subclonal deletions at the *ATRX* locus in tumors from a pediatric GBM patient.

(B) Diagram describing the binding locations of a TaqMan probes (ATRX del) designed to detect copy number variations at the *ATRX* locus.

(C) TaqMan assays to detect copy number variation at the *ATRX* locus on samples for patient 4.

(D) Single-cell RNA-seq in a xenograft derived from patient's 3 recurrence. *ATRX* expression levels are represented in this tSNE plot.

(E) Single-cell RNA-seq data for recurrence 3 of patient 5. This tSNE plot displays expression levels for *ATRX*.

(F) RNA-seq data for *ATRX* expression in non-neoplastic brain tissues (n = 4), bulk pediatric GBM samples (n = 3) and their matched primary cultures (n = 3). *ATRX* expression was normalized to the average for non-neoplastic brain tissues.
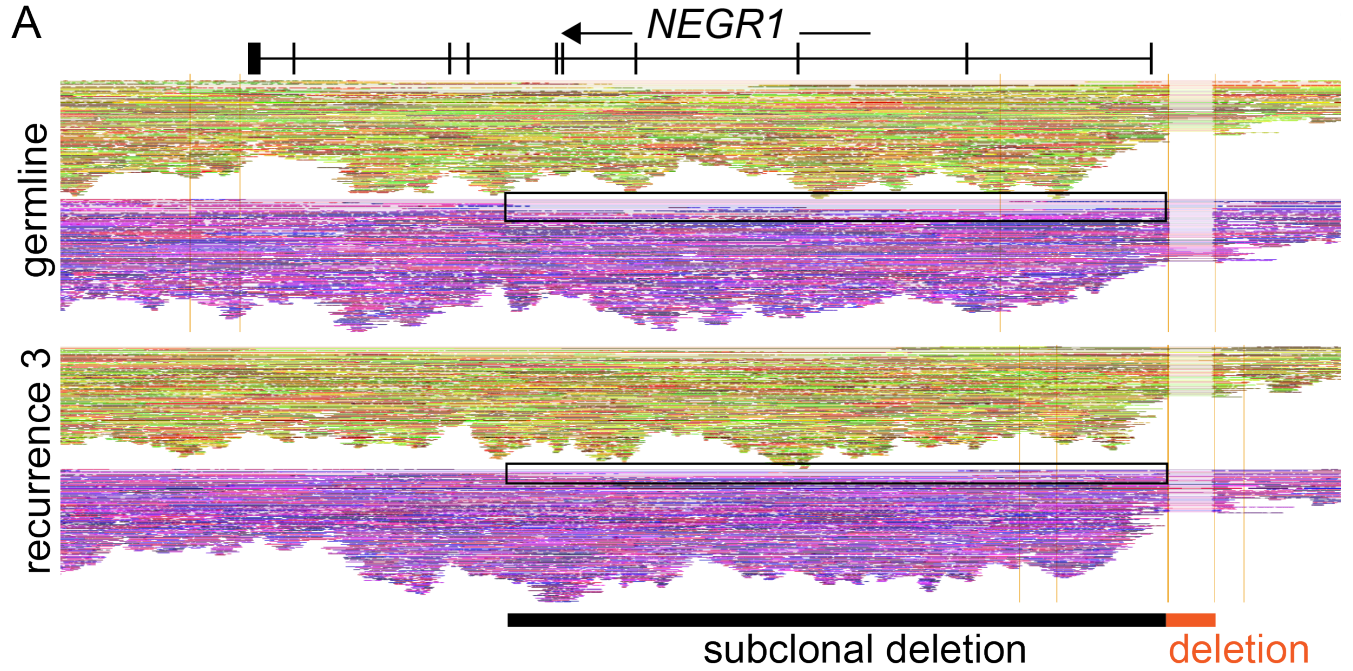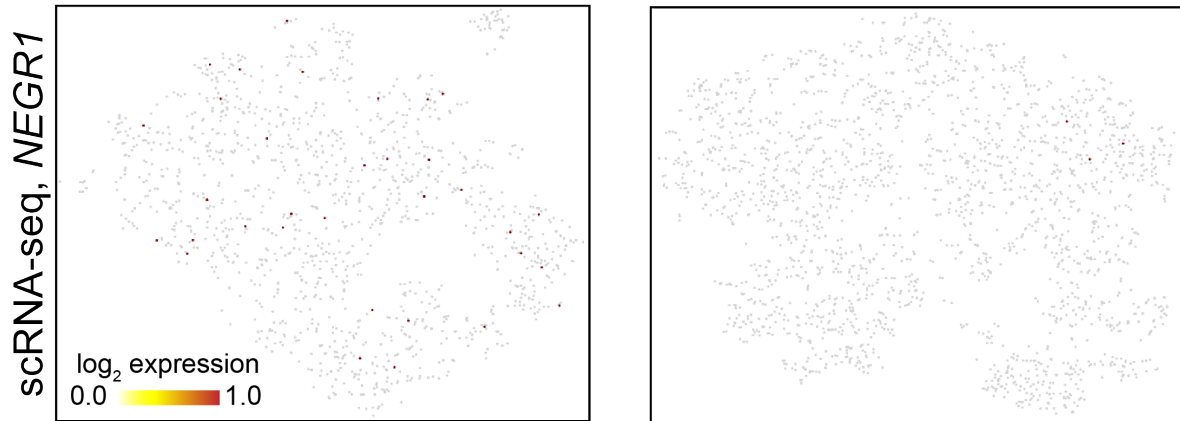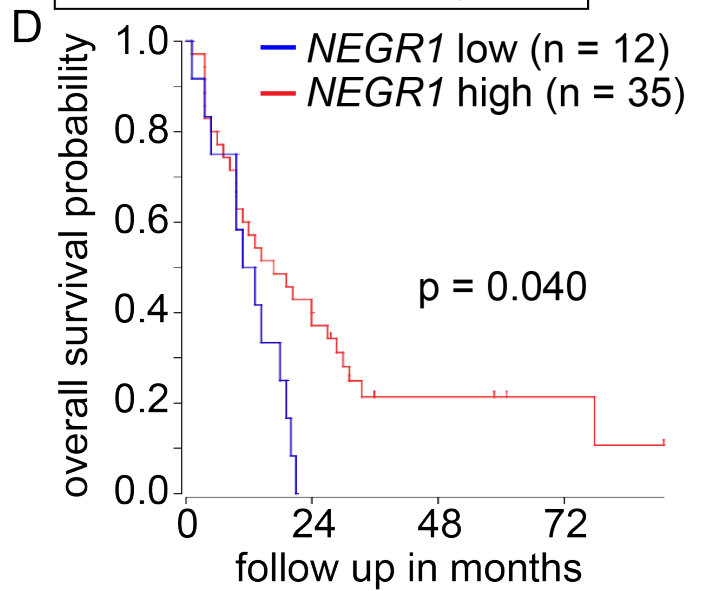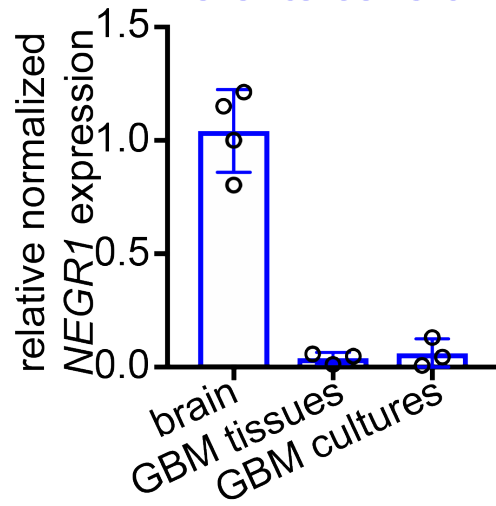
Fig. 6. Germline deletions at the *NEGR1* locus are common to all patients.

(A) Coverage of linked-reads at the *NEGR1* locus. Linked-reads are grouped based on haplotypes (red/green hues for one haplotype, purple/violet for the second). A deep deletion is observed in the region immediately upstream of the gene for the germline and the third recurrence for a pediatric GBM patient. A subclonal deletion that extends into the gene body is also visible (black box).

(B) Single-cell RNA seq for a pGBM xenograft derived from the recurrence of patient 3, and a xenograft derived from the third recurrence of patient 5. Expression levels of *NEGR1* are shown.

(C) RNA-seq data for *NEGR1* expression in non-neoplastic brain tissue (n = 4), bulk GBM tissues (n = 3) and their matched GBM primary cultures (n = 3).

(D) Kaplan Meier curve for pediatric GBM patients in the Paugh dataset, stratified by expression of *NEGR1*. Log-rank p = 0.040.
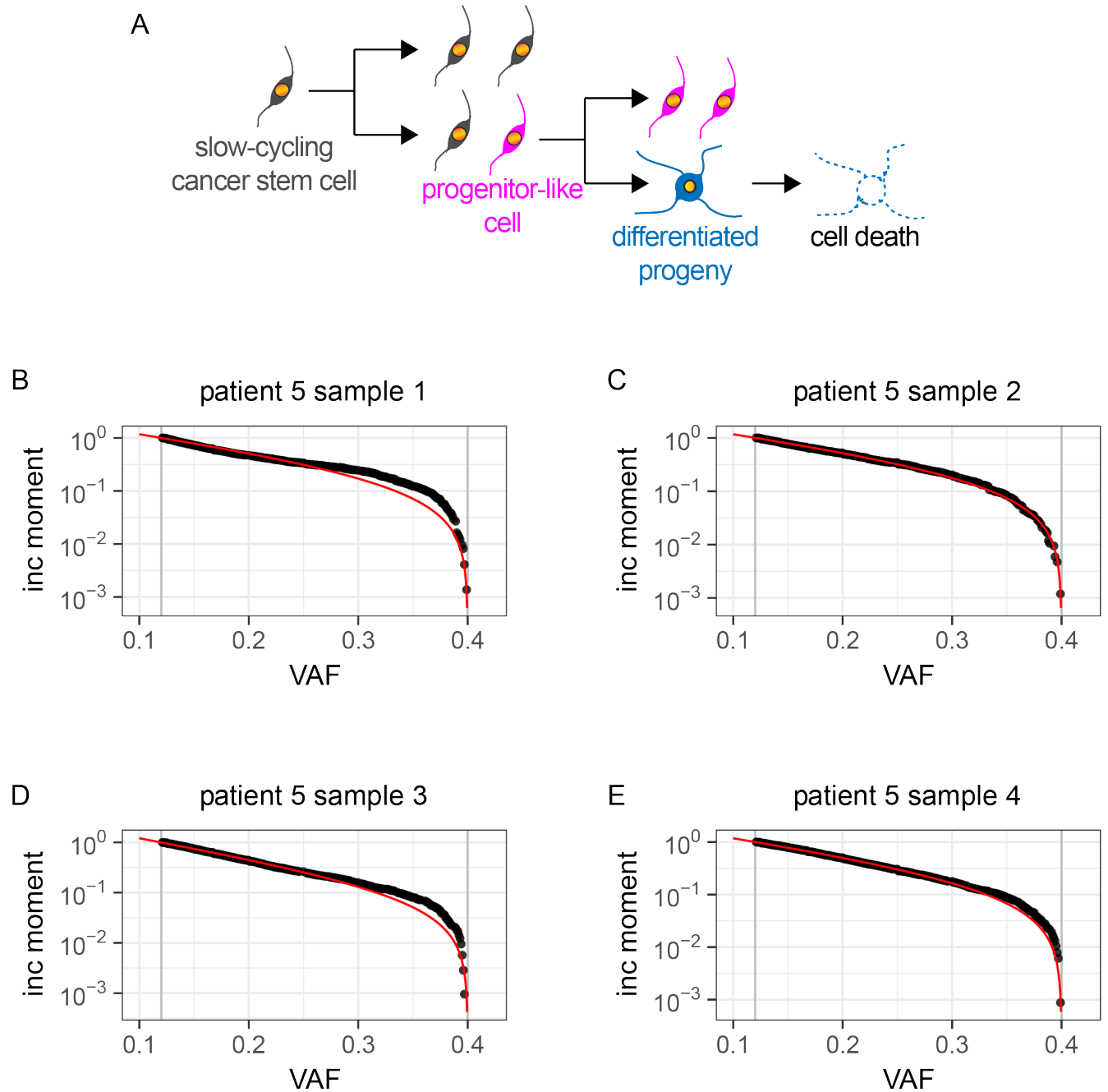
Fig. 7. Evidence for a hierarchical organization of pGBM.

(A) Diagram illustrating a testable model for pGBM organization and growth. Under a hierarchical model that includes relatively quiescent cancer stem cells giving rise to proliferating progenitor cells and non-proliferating cells, the incomplete moment of variant allele fractions (VAFs) would follow a negative binomial distribution.

(B-E) Fit of the first incomplete moment of the VAF distribution based on a neutral hierarchical model of tumor cell dynamics (red line) to data obtained from patient 5 (black points). The grey lines indicate allele frequency cut-offs due to sequencing resolution (lower) and ploidy (upper).