Intratumoral genetic and functional heterogeneity 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], Mehdi Zarrei[7,8,9], Jennifer King[2,3], Katrina Ellestad[2,3], Ngoc Ha Dang[2,3], Florence MG Cavalli[10], Michelle M Kushida[10], Fiona J Coutinho[10], Betty Luu[10], Yussanne Ma[11], Andrew J Mungall[11], Richard Moore[11], Marco A Marra[11], Michael D Taylor[10], Trevor J Pugh[12], Peter B Dirks[8,10], Stephen Scherer[7,8,9], Donna L Senger[2,3,14], Benjamin D Simons[4,5,6], Jennifer A Chan[2,3,15], A Sorana Morrissy[1,2,3], Marco Gallo[1,2,3,16]*.

[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 Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada.
[8]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3H7, Canada.
[9]McLaughlin Centre, University of Toronto, Toronto, ON M5S 3H7, Canada.
[10]Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada.
[11]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC VSZ 4S6, Canada.
[12]Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 2C4, Canada.
[13]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3H7, Canada.
[14]Department of Oncology, University of Calgary, Calgary, AB T2N 4N1, Canada.
[15]Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada.
[16]Department of Physiology and Pharmacology, University of Calgary, Calgary, AB T2N 4N1, Canada.

*Correspondence: marco.gallo@ucalgary.ca.

Running title: Evolution of genetic subclones in pediatric glioblastoma

Competing interests
No competing interests.

Abstract

Pediatric glioblastoma (pGBM) is a lethal cancer with no effective therapies. To understand mechanisms of tumor evolution in this cancer, we performed whole genome sequencing with linked reads on longitudinally resected pGBM samples. Our analyses showed that all diagnostic and recurrent samples were collections of genetically diverse subclones. Clonal composition rapidly evolved at recurrence, with less than 8% of non-synonymous single nucleotide variants being shared in diagnostic-recurrent pairs. In order to track the origins of the mutational events we observed in pGBM, we generated whole genome datasets for two patients and their parents. These trios showed that genetic variants could be (i) somatic, (ii) inherited from a healthy parent, or (iii) arose *de novo* in the germlines of pGBM patients. 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. Interestingly, radiation and anti-mitotic chemotherapeutics did not increase overall tumor burden upon recurrence. These findings support an important role for slow-cycling stem cell populations in contributing to recurrences, since slow-cycling cell populations are expected to be less prone to genotoxic stress induced by these treatments and to accumulate few mutations. Our results highlight the need for new targeted treatments that account for the complex functional hierarchies and genomic heterogeneity of pGBM.

Statement of significance

Our work challenges several assumptions on the genetic organization of pediatric GBM, and highlights mutagenic programs that start during early prenatal development.

Introduction

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 (5–8), 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 (6,9,10). For the majority of pGBM patients who harbor no 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 (11). Furthermore, *H3F3A* mutations were shown to be clonal in pGBM (12), 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 recent report identified differences in mutational signatures between matched primary and recurrent pGBM samples using exome sequencing (12).

Evidence of subclonal architecture in diffuse intrinsic pontine glioma and pGBM was recently provided (13). 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 pGBM 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) (14) 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.

Materials and Methods

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 tumor material was also allocated for patient-derived xenograft preparation (see below). Anagraphical information of sample donors are in **Table S2** and **Table S3**.

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

3

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. All xenografts were identity matched to the original patient material by short tandem repeat analysis using the AmpFLSTR Identifiler Plus kit (Applied Biosystems).

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, Calgary, AB). 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 (15) 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 (16).

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 run 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*

putative deleted region (ATRXdel probe) was first normalized to the fluorescence signal from the control diploid region probe. Fluorescence was additionally normalized to the diploid control DNA sample. TaqMan® Probes were 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.

Data and software availability
The linked-read WGS datasets we generated have been deposited to the European Genome-Phenome Archive (EGA) with accession number EGAS00001003432. scRNA-seq datasets have been deposited to the Gene Expression Omnibus (GEO) under accession number GSE117599. Bulk RNA-seq data from the Toronto cohort have been deposited to the EGA with accession number EGAS00001003070.

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 (**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 (nsSNVs), our data are in agreement with previous reports of low mutation burden (3,17,18) 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 nsSNVs shared between longitudinal samples from the same patient. This translated to an average of 7.8% nsSNVs 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 (19), in contrast with the much higher proportion shared between primary and recurrent adult glioma samples

(54%) (20). 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 (19) as well as in adult gliomas (20). 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 the number of mutational events. Sample 2 had 126 somatic nsSNVs, 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 nsSNVs in *TP53*. This patient was treated with temozolomide and local radiation (**Table S2**), which is the standard of care for adult GBM patients. However, the overall number of somatic nsSNVs 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 (12). Samples 1 and 2 from patient 3 only shared a mutation in *PCNX* (**Fig. 1C**). Patient 5 had 26 nsSNVs in the diagnostic sample and in sample 2, 38 in sample 3, and 39 in sample 4 (**Fig. 1D**). 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**). 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 nsSNVs 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. 1E, Fig. S1E,F**). 41.3% of these genes are associated with syndromes of the central nervous system (CNS). Similarly, nsSNVs 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. 1F**). These observations suggest that disruption of multiple unrelated networks of genes may be sufficient for pGBM initiation and progression. The identification of nsSNVs 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 nsSNVs 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 nsSNVs as either unique to the pGBM patient germline (ie not found in either parent's 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, nsSNVs in 9 genes were identified uniquely in the germline of this patient (**Fig. 1G**). Among these genes, *RASAL2* encodes a RasGAP and is a tumor suppressor (21); *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 (22); *HUWE1* encodes an E3 ubiquitin ligase that was shown to be a tumor suppressor in mouse models of skin cancer (23) and colorectal cancer (24). Similarly, we called 4 *de novo* nsSNVs in the germline of patient 3 (**Fig. 1H**). Among these genes, *KIAA1551* was recently identified as a candidate tumor suppressor and is frequently deleted in different cancer types (25). 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 corresponded to an 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 (26). These phylogenies show that each pGBM and its recurrence(s) are characterized by extensive intratumoral genomic heterogeneity and complex subclonal architecture (**Fig. 2A-E**). These results support the presence of multiple genetically distinct subclones that change significantly in frequency throughout the evolution of a tumor. Importantly, each recurrent tumor 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.

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**). 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 (27,28). 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 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) (20). 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 (29). Gene fusions involving *NTRK2* have been previously reported in pGBM (28). Patient 3 had a somatic *ETV6-NTRK3* gene fusion in both primary and recurrent tumors. This gene fusion has been reported in pGBM before (28), and it was also shown to have oncogenic function in other malignancies, including leukemia (30) and breast cancer (31). We also identified a translocation predicted to truncate the proto-oncogene *MET* in patient 4. *MET* amplifications are relatively common in adult GBM (32), and lesions in this gene are known to occur in pGBM (18,28).

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* (**Fig. S3**). Interestingly, *BTNL8* encodes a cell surface protein involved in T cell activation (33), whereas the *APOBEC3* genes encode cytidine deaminases with roles in antiviral immunity (34). 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 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 GROC-SVs package (35). We subset SV predictions to those for which we had sufficient physical coverage in all samples (see Supplemental Methods), thus retaining SVs supported by at least 200 barcodes (i.e. DNA molecules) spanning each breakpoint (**Supplemental file 3**). 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 variation in pediatric cancers (**Fig. 4B,C,G,I**) (36). 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 samples
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 (3,37), 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. 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**; see Materials and Methods). We then performed copy number TaqMan assays using genomes from both the germline and the tumors of our pGBM patients. 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 (**Table S3**). 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, 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.

Evidence for the existence of slow-cycling 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 (38). 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-like cells give rise to a rapidly cycling, self-renewing, progenitor-like population (**Fig. 6A**) (Extended Data Figure 4 of (38)). 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. 6B-E**, **Fig. S7A-I, Table S7**), suggesting pGBM tumor cells may be defined by a similar proliferative hierarchy (see Supplemental 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 Supplemental 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 (39), 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). Interestingly, the pediatric brain tumor diffuse intrinsic pontine glioma has recurrent somatic mutations in *ACVR1* (28,40–42),

and germline mutations in this gene lead to the hereditary syndrome fibrodysplasia ossificans progressiva (43). 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 (44), or if they arise 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 (45,46), and was observed in both adult and pediatric patients with non-hematologic cancers (47). 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 (47). The low allelic frequency 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* nsSNVs 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* nsSNVs is very low, based on recent literature evidence on trios that included probands affected by disorders with high heritability (48,49). 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 in these studies, 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 (50). Because of the established involvement of *ATRX* in gliomagenesis, we speculate that *ATRX* deletions might be a predisposing genetic lesion for pGBM.

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 (19) and diffuse intrinsic pontine glioma (13), 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 slow-cycling 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 (38). Although adult GBM and pGBM are molecularly different, hierarchical organization may be a common feature of both. If slow-cycling, 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. Whereas temozolomide and radiation treatment significantly increased the mutational burden in recurrences in adult GBM, it did not have this effect in pGBM. We propose that this might be because the cancer stem cell population might be more quiescent in pGBM than in adult GBM. Alternatively, it is possible that other biological differences between pGBM and adult GBM play a role.

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.

Acknowledgements

References

1. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DTW, Konermann C, et al. Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. Cancer Cell. 2012;22:425–37.

2. Cohen KJ, Pollack IF, Zhou T, Buxton A, Holmes EJ, Burger PC, et al. Temozolomide in the treatment of high-grade gliomas in children: A report from the Children's Oncology

Group. Neuro Oncol. 2011;13:317–23.

3. Schwartzentruber J, Korshunov A, Liu XY, Jones DTW, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482:226–31.

4. Khuong-Quang DA, Buczkowicz P, Rakopoulos P, Liu XY, Fontebasso AM, Bouffet E, et al. K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. Acta Neuropathol. 2012;124:439–47.

5. Bjerke L, Mackay A, Nandhabalan M, Burford A, Jury A, Popov S, et al. Histone H3.3 mutations drive pediatric glioblastoma through upregulation of MYCN. Cancer Discov. 2013;3:512–9.

6. Lewis PW, Müller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science. 2013;340:857–61.

7. Chan KM, Fang D, Gan H, Hashizume R, Yu C, Schroeder M, et al. The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev. 2013;27:985–90.

8. Bender S, Tang Y, Lindroth AM, Hovestadt V, Jones DTW, Kool M, et al. Reduced H3K27me3 and DNA Hypomethylation Are Major Drivers of Gene Expression in K27M Mutant Pediatric High-Grade Gliomas. Cancer Cell. 2013;24:660–72.

9. Funato K, Major T, Lewis PW, Allis CD, Tabar V. Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation. Science. 2014;346:1529–33.

10. Pathania M, De Jay N, Maestro N, Harutyunyan AS, Nitarska J, Pahlavan P, et al. H3.3 K27M Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas. Cancer Cell. 2017;32:684–700.

11. Jones C, Baker SJ. Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma. Nat Rev Cancer. 2014;14:651–61.

12. Salloum R, McConechy MK, Mikael LG, Fuller C, Drissi R, DeWire M, et al. Characterizing temporal genomic heterogeneity in pediatric high-grade gliomas. Acta Neuropathol Commun. 2017;5:78.

13. Vinci M, Burford A, Molinari V, Kessler K, Popov S, Clarke M, et al. Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. Nat Med. 2018;24:1204–15.

14. Ostrom Q, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. Neuro Oncol. 2015;17:iv1-iv62.

15. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.

16. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

17.   Zhang J, Wu G, Miller CP, Tatevossian RG, Dalton JD, Tang B, et al. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nat Genet. 2013;45:602–12.

18.   Bender S, Gronych J, Warnatz HJ, Hutter B, Gröbner S, Ryzhova M, et al. Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. Nat Med. 2016;22:1314–20.

19.   Morrissy AS, Garzia L, Shih DJH, Zuyderduyn S, Huang X, Skowron P, et al. Divergent clonal selection dominates medulloblastoma at recurrence. Nature. 2016;529:351–7.

20.   Johnson BE, Mazor T, Hong C, Barnes M, Aihara K, McLean CY, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science. 2014;343:189–93.

21.   McLaughlin SK, Olsen SN, Dake B, De Raedt T, Lim E, Bronson RT, et al. The RasGAP Gene, RASAL2, Is a Tumor and Metastasis Suppressor. Cancer Cell. 2013;24:365–78.

22.   Ikeda Y, Kiyotani K, Yew PY, Kato T, Tamura K, Yap KL, et al. Germline PARP4 mutations in patients with primary thyroid and breast cancers. Endocr Relat Cancer. 2016;23:171–9.

23.   Inoue S, Hao Z, Elia AJ, Cescon D, Zhou L, Silvester J, et al. Mule/Huwe1/Arf-BP1 suppresses Ras-driven tumorigenesis by preventing c-Myc/Miz1-mediated down-regulation of p21 and p15. Genes Dev. 2013;27:1101–14.

24.   Myant KB, Cammareri P, Hodder MC, Wills J, Von Kriegsheim A, Győrffy B, et al. HUWE1 is a critical colonic tumour suppressor gene that prevents MYC signalling, DNA damage accumulation and tumour initiation. EMBO Mol Med. 2017;9:181–97.

25.   Cheng J, Demeulemeester J, Wedge DC, Vollan HKM, Pitt JJ, Russnes HG, et al. Pan-cancer analysis of homozygous deletions in primary tumours uncovers rare tumour suppressors. Nat Commun. 2017;8:1221.

26.   Andor N, Harness J V., Müller S, Mewes HW, Petritsch C. Expands: Expanding ploidy and allele frequency on nested subpopulations. Bioinformatics. 2014;30:50–60.

27.   Giunti L, Pantaleo M, Sardi I, Provenzano A, Magi A, Cardellicchio S, et al. Genome-wide copy number analysis in pediatric glioblastoma multiforme. Am J Cancer Res. 2014;4:293–303.

28.   Wu G, Diaz AK, Paugh BS, Rankin SL, Ju B, Li Y, et al. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. Nat Genet. 2014;46:444–50.

29.   Douma S, Van Laar T, Zevenhoven J, Meuwissen R, Van Garderen E, Peeper DS. Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. Nature. 2004;430:1034–40.

30.   Eguchi M, Eguchi-Ishimae M, Tojo a, Morishita K, Suzuki K, Sato Y, et al. Fusion of ETV6 to neurotrophin-3 receptor TRKC in acute myeloid leukemia with t(12;15)(p13;q25). Blood. 1999;93:1355–63.

31.   Li Z, Tognon CE, Godinho FJ, Yasaitis L, Hock H, Herschkowitz JI, et al. ETV6-NTRK3

Fusion Oncogene Initiates Breast Cancer from Committed Mammary Progenitors via Activation of AP1 Complex. Cancer Cell. 2007;12:542–58.

32. Snuderl M, Fazlollahi L, Le LP, Nitta M, Zhelyazkova BH, Davidson CJ, et al. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. Cancer Cell. 2011;20:810–7.

33. Chapoval AI, Smithson G, Brunick L, Mesri M, Boldog FL, Andrew D, et al. BTNL8, a butyrophilin-like molecule that costimulates the primary immune response. Mol Immunol. 2013;56:819–28.

34. Harris RS, Liddament MT. Retroviral restriction by APOBEC proteins. Nat Rev Immunol. 2004;4:868–77.

35. Spies N, Weng Z, Bishara A, McDaniel J, Catoe D, Zook JM, et al. Genome-wide reconstruction of complex structural variants using read clouds. Nat Methods. 2017;14:915–20.

36. Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat Genet. 2016;48:4–6.

37. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;555:321–7.

38. Lan X, Jörg DJ, Cavalli FMG, Richards LM, Nguyen L V., Vanner RJ, et al. Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. Nature. 2017;549:227–32.

39. Amir RE, Van den Veyver I. B, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999;23:185–8.

40. Taylor KR, Mackay A, Truffaux N, Butterfield YS, Morozova O, Philippe C, et al. Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. Nat Genet. 2014;46:457–61.

41. Buczkowicz P, Hoeman C, Rakopoulos P, Pajovic S, Letourneau L, Dzamba M, et al. Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. Nat Genet. 2014;46:451–6.

42. Fontebasso AM, Papillon-Cavanagh S, Schwartzentruber J, Nikbakht H, Gerges N, Fiset PO, et al. Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma. Nat Genet. 2014;46:462–6.

43. Shore EM, Xu M, Feldman GJ, Fenstermacher DA, Brown MA, Kaplan FS. A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. Nat Genet. 2006;38:525–7.

44. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, et al. Timing, rates and spectra of human germline mutation. Nat Genet. 2016;48:126–33.

45. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012;44:651–8.

46. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. 2012;44:642–50.

47. Coombs CC, Zehir A, Devlin SM, Kishtagari A, Syed A, Jonsson P, et al. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. Cell Stem Cell. 2017;21:374–82.

48. O ' Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485:246–50.

49. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet. 2011;43:864–8.

50. Danussi C, Bose P, Parthasarathy PT, Silberman PC, Van Arnam JS, Vitucci M, et al. Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling. Nat Commun. 2018;9:1–15.

FIGURE LEGENDS

Fig. 1. Single nucleotide variants in pediatric GBM samples.
(A-D) Sankey plots for patients 1 (A), 2 (B), 3 (C), and 5 (D). 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.

(E) Analysis of all exonic, non-synonymous single nucleotide variants (nsSNVs) 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.

(F) Analysis of all exonic, nsSNVs in all recurrent tumors. Out of 85 genes with links to hereditary disease, 30 were specifically linked to CNS syndromes.

(G) 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.

(H) *De novo* non-synonymous SNVs in four genes were identified in patient's 3 germline.

Fig 2. Subclonal architecture and evolution in pGBM.
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 sample 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
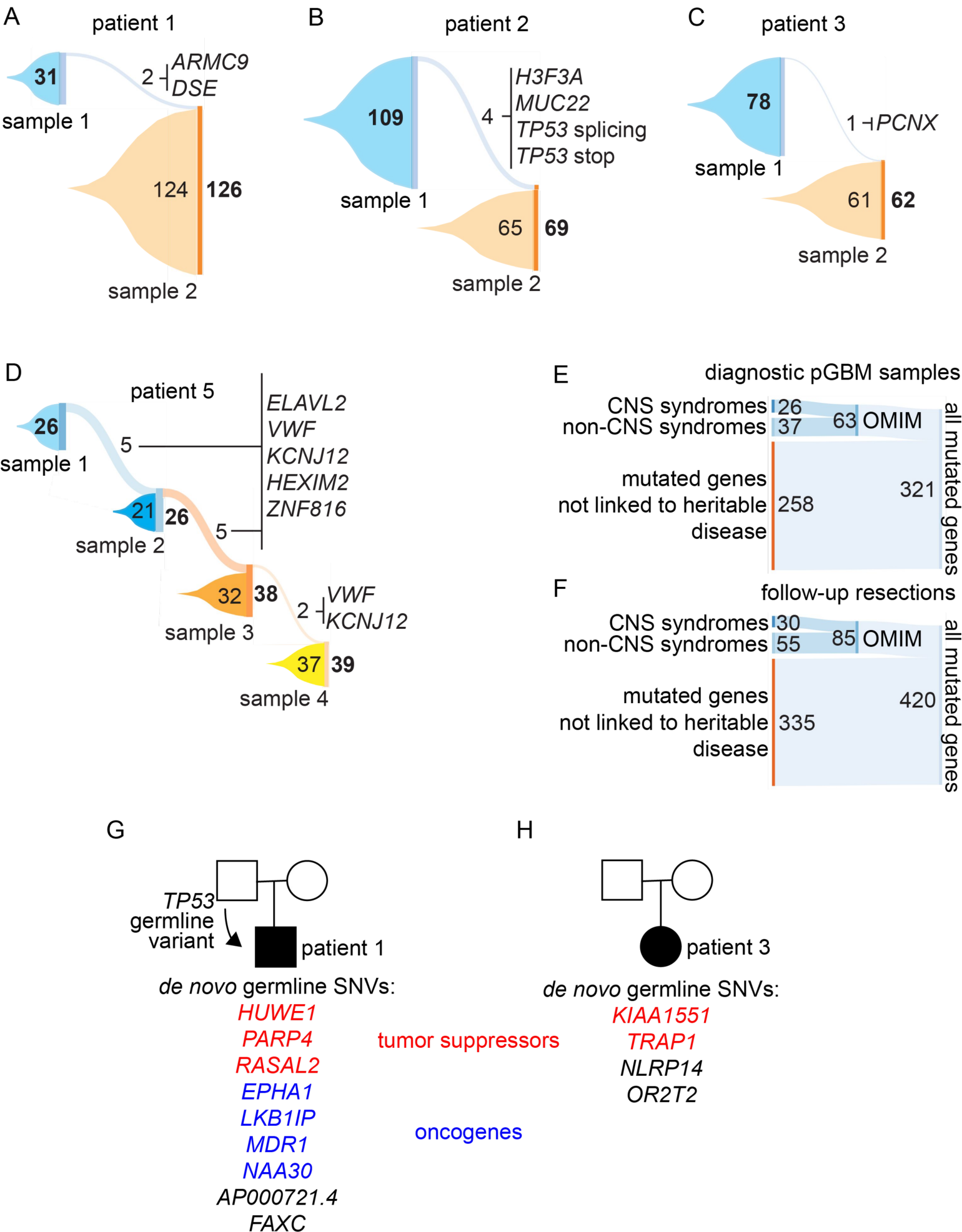
shown in panels B, C, G, and I. Samples include the germline for primary tumors and their recurrences, and germlines of the parents of patient 1 and patient 3. In these patients, germline events are stratified into inherited vs *de novo* events, and those observed in both parental germlines were excluded 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 (colored 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.
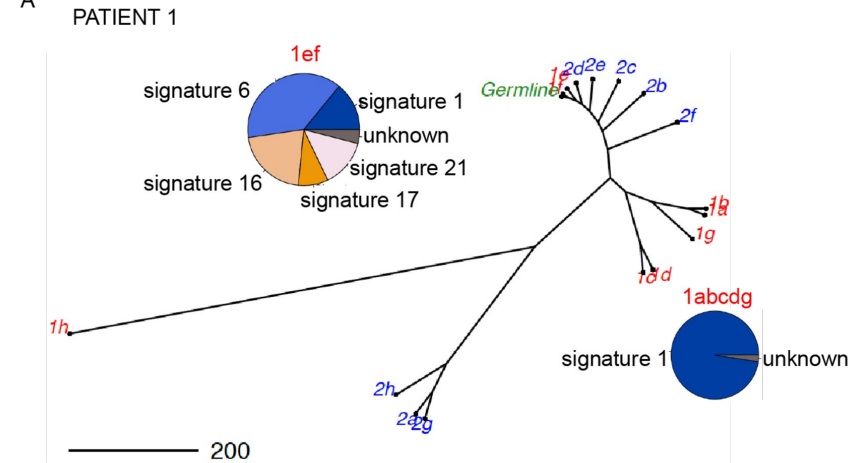
Fig. 5. Subclonal *ATRX* deletions in the germline and tumors of pGBM patients.
(A) Putative subclonal deletions at the *ATRX* locus in a tumor from a pGBM patient.
(B) Diagram describing the binding locations of a TaqMan probe (ATRXdel) designed to detect copy number variations at the *ATRX* locus.
(C) TaqMan assays to detect copy number variation at the *ATRX* locus on samples for patient 4. T-test statistics was used to obtain p-values.
(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 pGBM samples (n = 3) and their matched primary cultures (n = 3). *ATRX* expression was normalized to the average for non-neoplastic brain tissues. P-values were derived from t-test statistics.
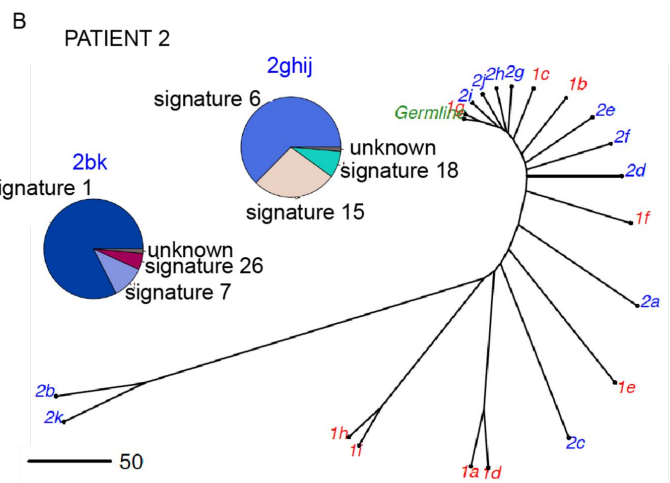
Fig. 6. Evidence for a hierarchical organization of pGBM.
(A) Diagram illustrating a testable model for pGBM organization and growth. Under a hierarchical model that includes slow-cycling cancer stem cells giving rise to proliferating progenitor cells and non-proliferating cells, the incomplete moment 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).
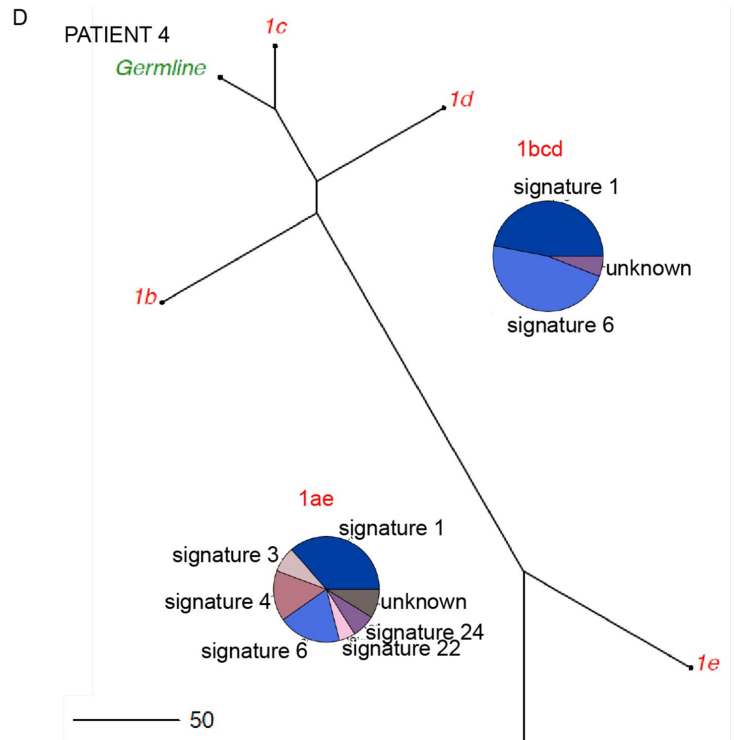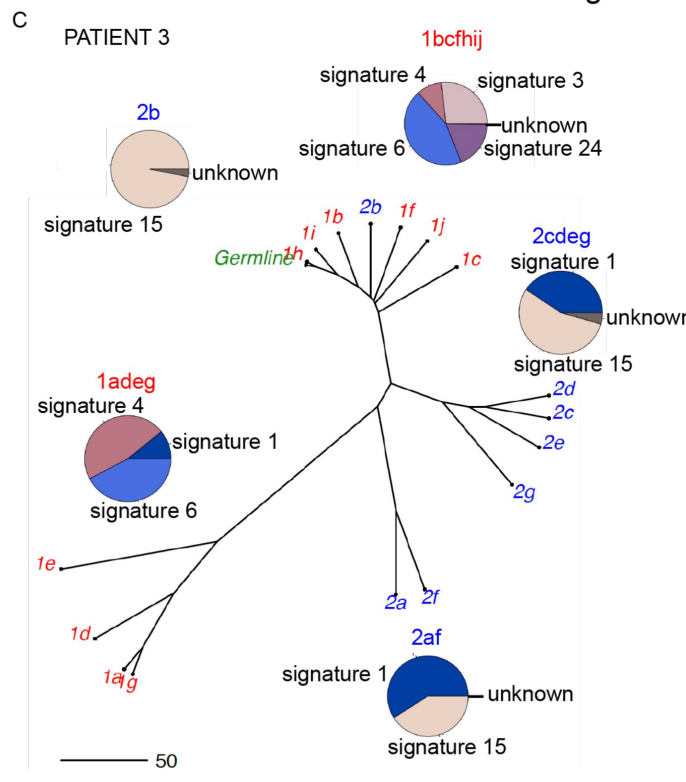
Figure 1

Figure 2

A PATIENT 1

1ef: signature 6, signature 1, unknown, signature 21, signature 16, signature 17

1abcdg: signature 1, unknown

Germline

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

scale: 200

B PATIENT 2

2ghij: signature 6, unknown, signature 18, signature 15

2bk: signature 1, unknown, signature 26, signature 7

Germline

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

scale: 50

C PATIENT 3

2b: signature 15, unknown

1bcfhij: signature 4, signature 3, unknown, signature 24, signature 6

2cdeg: signature 1, unknown, signature 15

1adeg: signature 4, signature 1, signature 6

2af: signature 1, unknown, signature 15

Germline

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

scale: 50

D PATIENT 4

Germline

1bcd: signature 1, unknown, signature 6

1ae: signature 1, signature 3, signature 4, signature 6, unknown, signature 22, signature 24

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

scale: 50

E PATIENT 5

4ef: signature 1, unknown

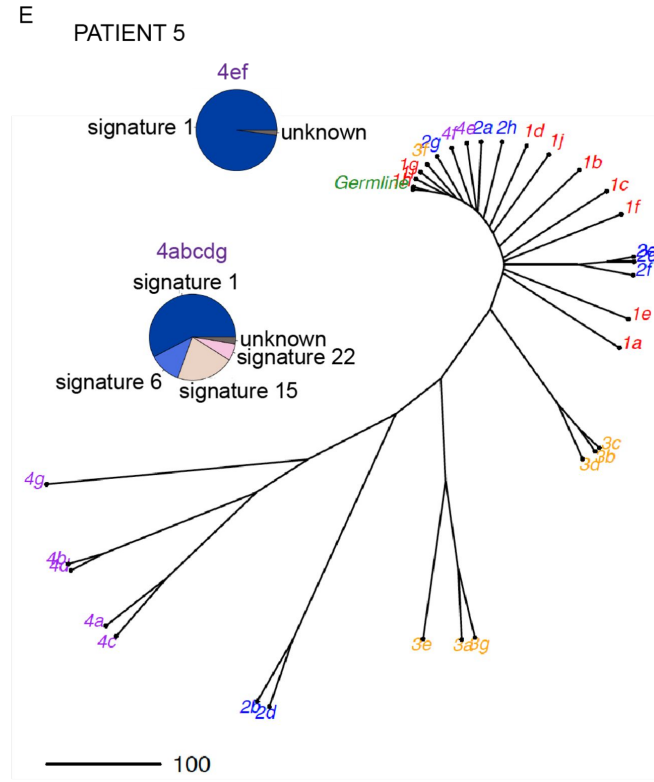4abcdg: signature 1, signature 6, signature 15, unknown, signature 22

Germline

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

scale: 100

Figure 3

A patient 1

B patient 2

C patient 3

D patient 4

E patient 5

Figure 4

Figure 5

Figure 6

A

slow-cycling
cancer stem cell

progenitor-like
cell

differentiated
progeny

cell death

B **patient 5 sample 1**

C **patient 5 sample 2**

D **patient 5 sample 3**

E **patient 5 sample 4**